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

Deliverable D5.2

Final Considerations on Synchronous Control Functions and Agile Resource Management for 5G

Version: v1.1

2017-03-31

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

Deliverable D5.2

Final Considerations on Synchronous Control Functions and Agile Resource Management for 5G

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/D5.2
Document Title:	Final Considerations on Synchronous Control Functions and Agile Resource Management for 5G
Version:	v1.1
Delivery Date:	2017-03-31
Editor(s):	Ömer Bulakci (Huawei Technologies GRC), David Gutierrez Estevez (Samsung), Athul Prasad (Nokia Bell Labs)
Authors:	Jakob Belschner, Paul Arnold (Deutsche Telekom), Jan Christoffersson, Mårten Ericson (Ericsson), Ömer Bulakci, Emmanouil Pateromichelakis (Huawei Technologies GRC), Yang Yang, Ahmed M. Ibrahim (Intel), Haris Celik (KTH), Athul Prasad, Fernando Sanchez Moya (Nokia Bell Labs), David Gutierrez Estevez (Samsung), Roberto Fantini, Giorgio Calochira, Sergio Barberis (Telecom Italia), Luis M. Campoy (Telefonica), Ivan Seskar, Francesco Bronzino, Parishad Karimi (WINLAB)
Keywords:	5G, control functions, RAN, resource management
Status:	Final
Dissemination level:	Public

Abstract

This deliverable contains final concepts, findings, and analyses on synchronous control functions and the associated agile resource management (RM) framework for the fifth generation (5G) radio access network (RAN) developed by the 5G PPP METIS-II project. Also, the deliverable is deemed to serve as a means for fostering discussions with other 5G Infrastructure Private Public Partnership (5G PPP) projects on relevant aspects.

Revision History

Revision	Date	Description
1.0	2017-03-31	D5.2 Release v1.0
1.1	2017-05-31	Typo corrected in DR-5 title

Executive summary

This Deliverable D5.2 presents the final concept of the agile resource management (RM) framework for 5G, which is one of the key innovation pillars that are developed by METIS-II. It further highlights final considerations on synchronous control functions of the fifth generation (5G) radio access network (RAN) along with analyses and results. Additionally, it aims to serve as a solid basis for discussions and collaborations with other 5G PPP projects.

Herein, we first outline the scope of the work comprising the motivation and the linkage of D5.2 with other METIS-II deliverables that describe the envisioned 5G radio access network (RAN) design and landscape. A high-level conceptual description of the agile RM framework investigated in METIS-II is then presented. The agile RM framework provides holistic RM solutions and air interface (AI) abstraction models that consider and exploit the novel aspects of 5G systems, such as, very diverse service requirements, existence of multiple AI variants (AIVs) in the overall 5G AI, dynamic topologies, and novel communication modes. 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. 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 topology, as well as hardware and software resources. For the optimum resource mapping, traffic steering schemes are proposed not only to be more dynamic as compared to the legacy systems, i.e., without the need for hard handovers, but also to factor in the peculiarities of different AIVs as well as multi-connectivity to these AIVs. In terms of the AIVs, the optimum mapping also takes into account the tight interworking between novel 5G AIVs and evolved legacy AIVs, e.g., long-term evolution (LTE). Further, agile RM enables the sharing of a common RAN by multiple network slices. In terms of novel interference mitigation techniques, advanced cooperative techniques can improve cell-edge user throughput in very dense deployments, considering both fixed and dynamic radio topologies. Cooperative techniques can be also utilized to moderate RAN to attain energy efficiency gains taking into account flexible frequency duplex options. Consequently, functional extensions and changes in the device measurement context are needed to enable such new functionalities tailored to different 5G use cases (UCs), while considering device performance.

The agile RM framework is constructed by functionality frameworks, which are defined in terms of sets of identified building blocks (BBs), which will be detailed in the following sections. A functional decomposition is applied to harmonize the BB descriptions and to determine the essential logical entities (LEs) and logical entity groups (LEGs) of the functional RM architecture. Further, possible mappings of the functional RM architecture onto RAN protocol and deployment architecture are analyzed and depicted. Finally, enabling technologies along with the key design recommendations, RAN design implications, and qualitative as well as quantitative analyses are presented to realize the developed agile RM framework.

Contents

List of Abbreviations and Acronyms

3GPP	Third Generation Partnership Project
5G	Fifth Generation
5G PPP	5G Infrastructure Private Public Partnership
AaSE	AIV agnostic Slice Enabler
AC	Admission Control
ACK	Acknowledgement
AI	Air Interface
AIV	AI Variant
ARP	Allocation Retention Priority
AN	Access Network
AN-I	Access Network-Inner
AN-O	Access Network-Outer
API	Application Programming Interface
BB	Building Block
BH	Backhaul
BLER	BLock Error Rate
BS	Base Station
BSS	Business Support Systems
BSR	Buffer Status Report
C-RAN	Centralized/Cloud RAN
CA	Carrier Aggregation
CAC	Central Access Controller
CBR	Constant Bit Rate
CCH	Control Channel
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CH	Cluster Head
CLI	Cross-link interference
CN	Core Network
CoMP	Coordinated Multi-Point
CP	Control Plane
CPNF-H	Higher layer Control Plane Network Function
CPNF-L	Lower layer Control Plane Network Function
CPNF-M	Medium layer Control Plane Network Function
CQI	Channel Quality Indicator
C-RS	Cell specific Reference Signal
CSI	Channel State Information
D2D	Device-to-Device communications
D-RAN	Distributed Radio Access Network
DL	Downlink
DPB	Dynamic Point Blanking
DPS	Dynamic Point Selection
DR	Design Recommendation
DRX	Discontinuous Reception
DTX	Discontinuous Transmission
E2E	End-to-End
eICIC	Enhanced Inter-Cell Interference Coordination
eMBB	enhanced Mobile Broadband
eNB	Evolved Node B
EPC	Evolved Packet Core
F-OFDM	Filtered OFDM
FBMC	Filterbank Multi-Carrier
FBR	Front to Back Ratio
FDD	Frequency Division Duplex
FEC	Forward Error Correction
FFS	For Further Study
FQAM	Frequency and Quadrature Amplitude Modulation
FSK	Frequency Shift Keying
GBR	Guaranteed Bit Rate
GFDM	Generalized Frequency Division Multiplexing
gNB	Gigabit Node B
GNSS	Global Navigation Satellite System
GSM	Global System for Mobile

	communication
HARQ	Hybrid Automatic Repeat Request
HSPA	High Speed Packet Access
HTTP	Hypertext Transfer Protocol
I/O	Input/Output
ICI	Inter-Cell Interference
ICIC	Inter-Cell Interference Coordination
ICR	Increased Contamination Regime
IDC	In-Device Coexistence
IM	Interference Management
ISD	Inter-Site Distance
ITU	International Telecommunication Union
JR	Joint Reception
JT	Joint Transmission
KPI	Key Performance Indicator
LAA	Licensed Assisted Access
LBT	Listen Before Talk
LE	Logical Entity
LEG	Logical Entity Group
LL	Low Latency
LOS	Line of Sight
LTE	Long-Term Evolution
LTE-A	Long-Term Evolution Advanced
LWA	LTE-A WLAN Aggregation
MAC	Medium Access Control (Layer)
MANO	Management and Orchestration
MDT	Minimization of Drive Test
METIS-II	Mobile and wireless communications Enablers for the Twenty-twenty Information Society-II
MIMO	Multiple-Input Multiple-Output
mMTC	Massive Machine-Type Communications
mmW	millimeter Wave
MS	Mobile Station
MU-MIMO	Multi-User Multiple-Input Multiple-Output

NACK	Negative Acknowledgement
NFV	Network Function Virtualization
NLOS	Non Line of Sight
NN	Nomadic Node
OFDMA	Orthogonal Frequency Division Multiple Access
O&M	Operations and Maintenance
OOB	Out-Of-Band
OSS	Operations Support Systems
PCEF	Policy and Charging Enforcement Function
PCR	Pilot Contamination Regime
PDCCH	Physical Downlink Control Channel |
PDCP	Packet Data Convergence Protocol
PDN-GW	Packet Data Network – Gateway
PGIA	Pre-emptive Geometrical Interference Analysis
PHY	Physical (Layer)
PPI	Power Preference Indicator
PRACH	Physical Random Access CHannel
QAM	Quadrature Amplitude Modulation
QCI	QoS Class Identifier
QoS	Quality of Service
RAN	Radio Access Network
RAP	Radio Access Point
RAT	Radio Access Technology
RCR	Reduced Contamination Regime
REST	Representational State Transfer
RLC	Radio Link Control
RLF	Radio Link Failure
RM	Resource Management
RRH	Remote Radio Head
RRC	Radio Resource Control
RRM	Radio RM
RSC	Resource Sharing Cluster
RSCP	Received Signal Code Power
RSRP	Reference Signal Received Power

RSRQ	Reference Signal Received Quality
RSSI	Received Signal Strength Indicator
RTT	Round-Trip Time
sBH	Self-Backhauling
SBS	Small-cell Base Station
SDN	Software Defined Networking
SDWN	Software Defined Wireless Network
SINR	Signal-to-Interference-plus-Noise Ratio
SLA	Service Level Agreement
SON	Self-Organizing Network
SotA	State of the Art
SRS	Sounding Reference Signal
TA	Tracking Area
TCE	Trace Collection Entity
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TeC	Technology Component
TN	Transport Network
TRxP	Transmission Reception Point

TTI	Transmission Time Interval
Tx	Transmission
UC	Use Case
UDN	Ultra-Dense Network
UE	User Equipment
UF-OFDM	Universal filtered OFDM
UL	Uplink
uMTC	Ultra-reliable Machine-Type Communications
UP	User Plane
UPNF-H	Higher layer User Plane Network Function
UPNF-L	Lower layer User Plane Network Function
UPNF-M	Medium layer User Plane Network Function
UR	Ultra Reliable
URLLC	Ultra-Reliable and Low-Latency Communications
UTRA	Universal Terrestrial Radio Access
xMBB	Extreme Mobile Broadband
WP	Work Package

1 Introduction

The explosive growth in capacity and data rate demands, together with novel and challenging service types envisioned for the time frame beyond 2020 towards extreme Mobile Broadband (xMBB), aka enhanced Mobile Broadband (eMBB), massive Machine-Type Communications (mMTC), and ultra-reliable Machine-Type Communications (uMTC), aka Ultra-Reliable and Low-Latency Communications (URLLC)[1], are the main motivation for the development of the fifth generation (5G) mobile communication technologies. 5G will consist of technologies that have to fulfill ambitious requirements driven not only by the extensions to today's broadband applications but also by the vertical industries, which are interested in using radio networks for wireless connectivity and control. It is envisioned that, in 5G, the overall air interface (AI) comprises different AI variant(s) (AIVs)[2] that are optimized, e.g., for the specific frequency bands of operation (below 6 GHz, centimeter wave, millimeter wave, etc.) and for one or more target use cases (UCs) [MET-II16-D11][MET-II16-D22][MET-II16-WP]. While there can be different AIVs, the resource management (RM) framework should be agile to operate in an AIV-overarching manner [MET-II16-D51][BPB+15]. That is, AIV-overarching RM functionality framework shall remain agnostic to the design of the physical layer (PHY) of the AIVs that are involved, and, thus, can also operate over a newly introduced AIV. At the same time, mechanisms pertaining to Intra-AIV RM functionality framework shall be tailored to the peculiarities of a given AIV, e.g., interference management and dynamic resource scheduling. On this basis, the METIS-II project [MET-II] has developed an agile RM framework comprising the aforementioned functionality frameworks.

Along with the ambitions for 5G networks, new challenges will also rise. In terms of interference conditions, 5G networks will be more challenging due to the ultra-dense deployment of cells and dynamic radio topologies, e.g., consisting of new types of access nodes (such as, nomadic nodes (NNs), which are movable access nodes mounted on vehicles). Besides, a versatile wireless access landscape is foreseen originated by a much wider use of beamforming, flexible uplink/downlink (UL/DL) frame configuration, in-band self-backhauling (sBH) with access and backhaul networks using the same carrier or channel, and device-to-device (D2D) communications. New dimensions to the problem of efficient resource allocation within 5G networks further include the need to support ultra-reliable and latency-critical applications and

[1] For the sake of consistency, xMBB, mMTC, and uMTC are primarily utilized herein to refer to generic 5G service types, as these are the initial terms introduced by the METIS project and further adopted by the METIS-II project.
[2] An AIV is defined as the RAN protocol stack (i.e., PHY/MAC/RLC/PDCP/RRC or 5G equivalents, or subset thereof) and all related functionalities describing the interaction between infrastructure and device, and covering, e.g., a subset of services, bands, cell types that characterize the overall 5G system.

the support for network slicing concept[3]. The agile RM framework takes those aspects into account and leverages the degrees of freedom offered by the availability of multiple links and layers (in terms of frequency bands, AIVs, cell types, etc.). In this report, we present the final description of the developed agile RM framework in terms of functional, protocol and deployment perspectives, and highlight the key considerations and design recommendations to enable the development of such a framework. The details of enabling technologies, along with the defined building blocks (BBs) within the agile RM framework and state of the art (SotA), are also provided in the following sections of this report.

1.1 Scope

As part of the overall 5G RAN design, METIS-II has developed an agile RM framework that operates over a multi-AIV environment comprising both novel 5G AIV(s) and evolved legacy AIV(s) [MET-II16-D22][MET-II16-D41]. A fundamental part of the framework is the allocation of resources based on availability and service needs. The term “resource” may not only refer to the classical definition as in the context of radio RM (RRM), e.g., available licensed spectrum and time resources, but can also incorporate an extended definition, such as hardware resources (e.g., number/type/configuration of antennas, existence of nomadic nodes in an area or mobile terminals that can be used as D2D relays) and software resources (e.g., software capabilities of network nodes and devices, such as data storage and processing). Apart from these resources, the need for additional context information (e.g., AIV-specific congestion levels), the frequency of context acquisition (e.g., real-time AIV-specific radio link feedback) and functional extensions to user equipment (UE) context (e.g., new measurement configurations) are also analyzed. In the considered heterogeneous 5G environment, an appropriate modeling of these resources is required. The main objectives of this work highlighted in this document can be outlined as:

- Efficient use of any available resources when and where needed considering the extended notion of a resource beyond conventional RRM,
- The optimum mapping of 5G services to the resources taking into account target performance metrics, e.g., energy consumption and latency, as well as AIV capabilities and characteristics, and
- Multi-slice RM ensuring service level agreements (SLAs) of network slices.

These aforementioned objectives are tightly coupled with the user plane (UP) design. For instance, considering service-tailored user plane (UP) design in novel 5G AIVs [MET-II16-D41], e.g., flexible frame structure, it is required to revisit existing functionalities or to introduce new functionalities, such as, new multi-service scheduling paradigms. The overall control plane (CP) design in METIS-II is based on the development of *synchronous* control functions that require radio frame/slot/sub-frame or any time-domain level synchronization, e.g., the aforementioned

[3] A network slice, namely “5G slice,” supports the communication service of a particular connection type with a specific way of handling the CP and UP for this service and is seen from a customer perspective as a separated logical network [NGM15][MET-II16-D22][MET-II16-WP].

multi-service scheduling, and *asynchronous* control functions that do not require radio frame/slot/sub-frame or any time-domain level synchronization, e.g., mobility and initial access functions. On this basis, the design of synchronous control plane (CP) functions is targeted in a holistic way along with the aim of further advancing their capabilities, e.g., modifications relative to legacy functions. Accordingly, the work considered herein substantially contributes to the overall CP design (see [MET-II17-D62][4] for the final considerations on asynchronous CP functions and overall CP design). On this basis, two main directions for the work have been followed in work package 5 (WP5) of METIS-II:

The first direction (within Task 5.1 (T5.1)) is the *Design of Synchronous Control Functions for Resource & Traffic Management*. In this direction, the essential synchronous CP mechanisms for the 5G radio access network (RAN) have been designed and evaluated in order to define an agile RM framework. These mechanisms are used in multi-link, multi-layer and multi-node mode of operation (including multi-hop self-backhauling communication).

The second direction (within Task 5.2 (T5.2)), which is closely coupled with the previous one, is the *Design of a Resource Abstraction Framework.* In this direction, AIV-overarching RM has been designed as a functionality framework, which also includes paradigm changes when defining the delimitation between the synchronous and asynchronous control functions, e.g., designing synchronous control functions that have been conventionally applied in an asynchronous manner, such as traffic steering. To this end, the necessary context information to support the AIV-overarching RM mechanisms is determined.

The agile RM framework is thus, a harmonization of the different enabling technologies that are developed in the above-mentioned directions. Figure 1-1 illustrates the development of this framework starting from technology components (TeCs), which relate to one or more synchronous control functions, going towards the final framework via BBs and functionality frameworks. As indicated by the horizontal focus bars, the previous deliverable D5.1 [MET-II16-D51] focused on the development of TeCs and the associated BBs while providing the foundations for functionality frameworks along with the first vision on the agile RM framework. The current deliverable D5.2 places the focus on the further refinement of functionality frameworks and harmonization of the BBs towards the final design based on the functional decomposition, where TeCs and BBs are described in terms of fundamental functional elements, namely, logical entities (LEs) and logical entity groups (LEGs). The associated key design recommendations and RAN implications to enable this final design, along with the corresponding qualitative and quantitative analyses are also provided.

[4] The publication date of METIS-II deliverable D6.2 is one month later than that of this deliverable D5.2.

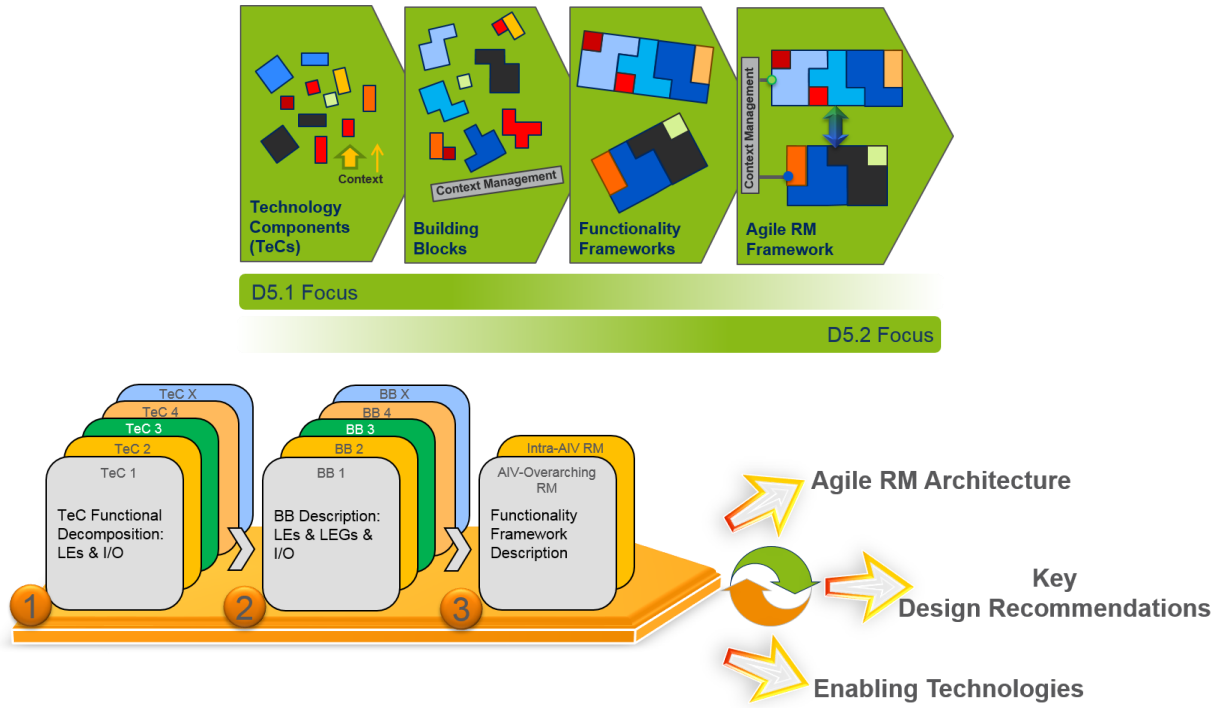

Figure 1-1 Development of agile RM framework; focal points of deliverables (upper) and functional decomposition to harmonize enabling technologies (lower).

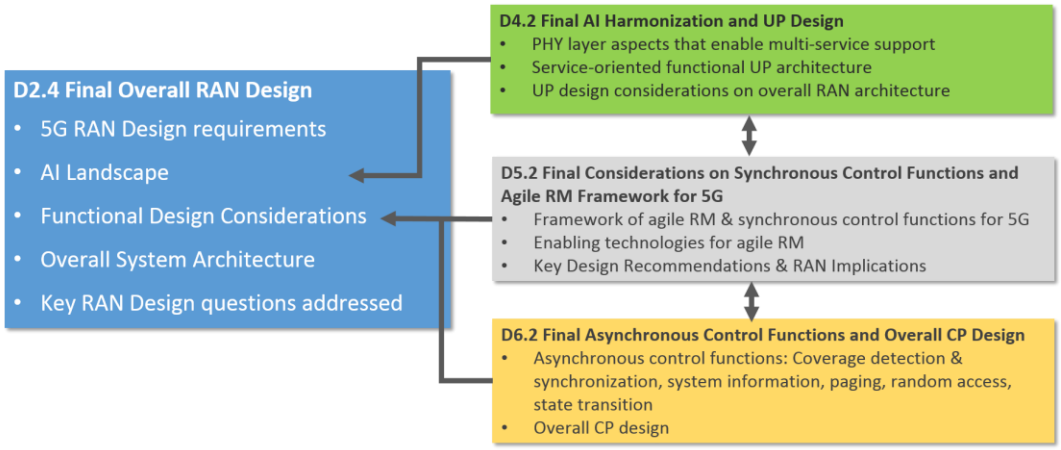

Figure 1-2 Positioning of D5.2 with respect to other METIS-II deliverables.

Figure 1-2 illustrates the role of D5.2 with respect to other METIS-II deliverables while specific details on the inter-relations are highlighted in different sections herein. In particular, *D2.4 Final Overall RAN Design* [MET-II17-D24][5] provides an overall picture of the final 5G RAN design,

[5] The publication date of METIS-II deliverable D2.4 is three months later than that of this deliverable D5.2.

linking and integrating the technical issues, concepts, and solutions covered in detail in D4.2 [MET-II17-D42][6], D5.2, and D6.2 [MET-II17-D62].

In this document, the term evolved Node B (eNB) is used to indicate a generic LTE base station (BS) connecting to the evolved packet core (EPC), whereas gigabit Node B (gNB), aka 5G-gNB, refers to the 5G BS connecting to the 5G core. The terms in general are used to refer to a wireless access node. Furthermore, the terms UE and MS are used as synonyms.

1.2 Objective of the Document

The objective of the deliverable is to present the final BBs and enabling technologies required for developing the 5G agile RM framework as highlighted in the scope section. The document provides the final description of the framework from functional, protocol, and deployment perspectives for resource and traffic management in 5G.

1.3 Structure of the Document

The structure of the document is as follows. In Section 2, the high-level description of the framework for agile RM and related synchronous control functions is provided. Section 3 provides the description of the agile RM framework from functional, protocol, and deployment perspectives. In Section 4, the enabling technologies for the agile RM framework are described in detail along with the associated key design recommendations, RAN design implications as well as qualitative and quantitative justifications. The conclusions and future outlook of the work are presented in Section 5. Finally, Annex A includes further details and analyses of various TeCs mapped onto design recommendations.

[6] The publication date of METIS-II deliverable D4.2 is one month later than that of this deliverable D5.2.

2 Framework of Agile RM and Synchronous Control Functions

A conceptual illustration for the agile RM framework is shown in Figure 2-1. In particular, the framework operates over the 5G landscape consisting of different and novel deployment options (e.g., ultra-dense networks, UDNs, and dynamic radio topology based on NNs), novel communication modes (e.g., unicast or multicast D2D communication), 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. Besides, network slicing is seen as a key enabler for new 5G businesses that support one or more services with their associated service level agreements (SLAs) and, therefore, the agile RM framework includes RM for network slices considering the overall AI, as well.

Based on the methodology in Figure 1-1, the agile RM framework is built upon TeCs, where, herein, each TeC can comprise one or more 5G synchronous control functions. A synchronous control function [MET15-D64] is coupled with a time-frame structure of the target radio interface. Accordingly, the framework of agile RM comprises

- synchronous control functions, i.e., within TeCs, that ensure the fulfillment of service requirements as well as network slice requirements,
- paradigm changes for efficient operation and improved performance of typical *synchronous* control functions (e.g., interference management (IM) mechanisms to adapt to new dynamic radio topologies), and for fast operation of typical *asynchronous* control functions (e.g., dynamic traffic steering applied on a synchronous level rather than legacy hard handovers),
- a resource abstraction that enables AIV-overarching RM where real-time context is collected from novel AIV(s) and the legacy ones (e.g., those of evolved Long-Term Evolution (LTE) and WiFi) to determine whether a service flow can be mapped onto a given AIV (see multi-AIV resource mapping in Section 3),
- AIV-specific and AIV-agnostic RM functionalities that construct the Intra-AIV and AIV-overarching RM functionality frameworks, respectively, and
- design recommendations (DRs) to enable the envisioned agile RM framework along with their RAN design implications and analyses.

In particular, developed RM schemes enable mapping of resources to diverse service types with contradicting requirements that exploit specific AIV characteristics, e.g., flexible numerology of the PHY. Further, the boundaries between synchronous and asynchronous control functions may vary to ensure 5G service requirements, e.g., dynamic traffic steering to reduce latency. Given the latency-critical services to be enabled by 5G networks, the efficiency of such mechanisms shall be clearly improved to be agile enough to react sufficiently to service needs.

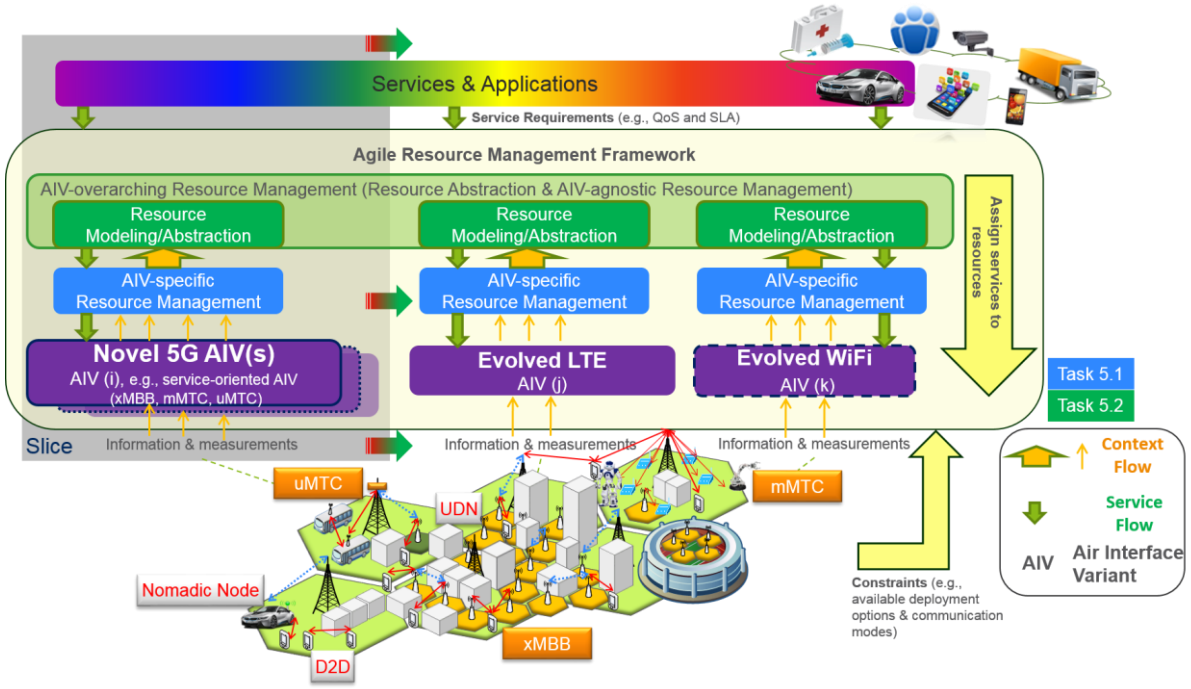

Figure 2-1 High-level illustration of the agile RM framework[7] that is developed in METIS-II. The horizontal arrows on the grey slice rectangle indicate that network slicing can be extended towards evolved legacy AIVs, as well.

Another important design criterion is the overall energy efficiency of the networks particularly considering the envisioned UDNs, where the number of active access nodes should be optimized, while fulfilling service requirements. Consequently, the implications of supporting the developed RM schemes on the user equipment (UE) context should be assessed in terms of the functional extensions needed, while taking into account the trade-off between device complexity and performance.

In 5G, the notion of a resource cannot be anymore confined to be strictly related to RRM, but instead the considered realm of resources can expand beyond the conventional options[8], such as frequency (i.e. a licensed band), time, and power, as visualized in Figure 2-2. Accordingly, the extended notion of resource includes the following five basic types, in addition to the conventional radio resources:

1. Spectrum Resources with other kinds of licensing regimes, e.g. radio resources in unlicensed or shared bands
2. Base Station Transmission points (Tx points) antennas along with radio frequency (RF) equipment and the UE capabilities

[7] A previous version of this illustration has been published in [BPB+15].
[8] A conventional radio resource is defined as the resource that is typically used by well-established RRM schemes, e.g., in legacy networks.

3. Soft capabilities, such as processing, storage and memory resources
4. Transport network resources
5. Energy

A radio resource is already part of the conventional notion of resource. It is characterized by time (the duration of the transmission), frequency (the carrier frequency and the bandwidth) and the transmit power. A radio resource can be located in a licensed, in an unlicensed spectrum band (in both low and high frequencies [LPV+17]) or in bands that are subject to spectrum sharing (including secondary spectrum usage).

A radio resource is associated with a Tx point to perform a radio transmission. It is possible that a Tx point allocates a radio resource to multiple devices, e.g., by separating them with the help of codes (code division multiple access, CDMA) or spatially (multi-user multiple-input multiple-output, MU-MIMO). The Tx point can be of various kinds: it can be a typical static access node (e.g., a small cell, a macro cell or a remote radio head, RRH, in case of centralized/cloud RAN, C-RAN, deployments), a dynamic access node (e.g., NN) or a device in case of D2D communication.

Processing, memory, and storage resources, which are referred to as soft capabilities herein, are required to perform the processing on the different protocol levels. There is a tendency towards specialized signal processing resources at lower layers, whereas higher layer processing can be executed in general purpose processing resources. The processed data is transmitted and received using RF equipment (such as digital / analog converters, amplifiers, and antennas) through one or multiple radio resources.

Data from and to the core network (CN) needs to be transported to / from the antenna site using a transport network (TN). Transport capacity (e.g., expressed by a bandwidth and a latency requirement) needs to be reserved for this purpose.

The operation of any network element consumes energy, which needs to be managed in order to reduce the overall energy consumption of the network as well as of the devices. In particular, efficient energy consumption is important for nomadic (battery-powered) devices.

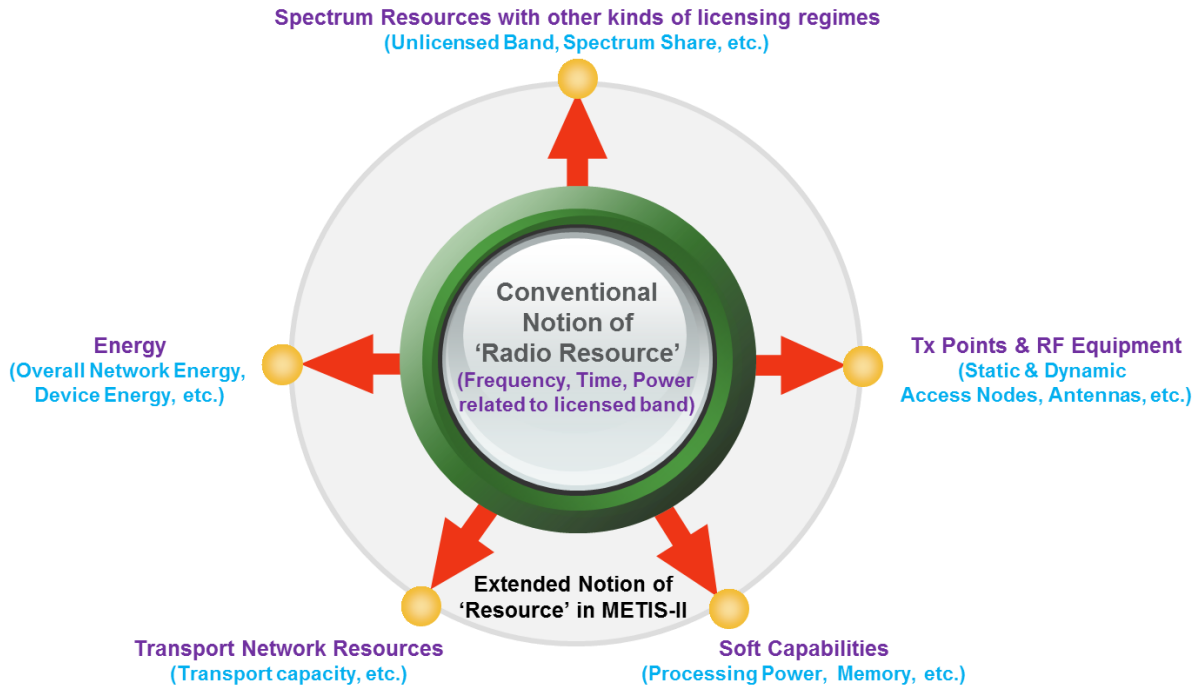

Figure 2-2 Extended notion of a resource in METIS-II including conventional RRM resources.

As mentioned before, the agile RM framework comprises functionality frameworks that are supported by context management. A functionality framework is defined by a set of BBs that are based on one or more TeCs (see also Figure 1-1). The AIV-overarching functionality framework takes into account the overall AI comprising various AIVs and, thus, encompasses the considerations on resource abstraction and AI-agnostic RM schemes, while the Intra-AIV RM functionality framework operates on a target AIV. It is worth noting that a TeC[9] may relate to one or more synchronous control mechanisms. Furthermore, the functions in Intra-AIV RM functionality framework can be tailored for specific AIVs and, thus, are resulting in the AIV-specific RM illustrated in Figure 2-1.

The BBs of the agile RM framework are briefly described below and will be detailed in Section 4 while the essential LEs and LEGs pertaining to BBs will be highlighted in Section 3.

- **Interference Management:** The techniques for interference management are essential for 5G networks not only to ensure high capacity and wide coverage of high end-user data rates but also to ensure high reliability for uMTC services. The developed schemes cope with the aforementioned 5G landscape and the associated new interference

[9] TeCs are tagged by dedicated IDs, i.e., T5.*x*-TeC*y* where *x* and *y* indicate the numbering, for the sake of easy tracking in the Annex. Each TeC has a relation to a provided design recommendation (DR). Furthermore, when an enabling technology has implications on both research directions in T5.1 and T5.2 (see Section 1.1), two IDs are assigned to that enabling technology.

conditions, e.g., due to dynamic radio topology or duplexing flexibility. Also, to materialize the theoretical gains that network coordination can bring to interference management in practice, the dependency and impact of the interference management schemes on the practical limitations of the X2* interface (see [MET-II16-D22][MET-II16-WP] for architectural details) between access nodes shall be described.

- **Flexible Short-term Spectrum Usage:** This BB discusses the flexible usage of unlicensed bands in UDNs, by taking into account dynamic radio topology that may change because of moving access nodes, such as NNs. Further, the duration and the extent of the usage of unlicensed bands depend on the uncontrolled source of interferers and the dynamicity in topology changes; hence, unlicensed spectrum usage may be adapted in short-time intervals.
- **RAN Moderation:** RAN moderation takes into account the existing network topology and self-backhauling, and aims at enabling fast on/off switching of small cells to reduce energy consumption while ensuring that service requirements are met. The building block aims to leverage lean system design and tighter interworking of RAN and backhaul traffic aggregation nodes to provide higher energy efficiency in 5G networks using cooperative sleep strategies between the RAN and BH links.
- **Multi-Slice and Multi-Service Holistic RM:** Considering the envisioned 5G AI which consists of multiple AIVs, new services with challenging requirements and the vision of network slicing [MET-II16-D22][MET-II16-WP], RM techniques for inter-slice and intra-slice resources are captured in this BB. Furthermore, the agile RM framework operates on a complex multi-link, multi-layer, and multi-AIV 5G landscape, and, accordingly, this BB also constructs a holistic view of the RM. On one hand, RM approaches to deal with the very different service requirements and novel communication variants in a flexible way are highlighted. On the other hand, the protocol stack implications of the split between AIV-specific and AIV-agnostic RM functionality (see Section 3.2) are tackled by this BB.
- **Dynamic Traffic Steering:** The mechanisms to enable mapping of a service flow through the right AIV(s) taking into account the envisioned 5G landscape along with the evolved legacy networks are captured in this BB. It is important to note that, herein, mechanisms towards traffic steering aim at fast time-scale RM rather than, e.g., slow time-scale handovers between access nodes and/or AIVs as in legacy networks. The BB takes into account the challenges introduced by novel 5G design aspects, such as an antenna beam-based system, for the service flow mapping and for achieving the 5G KPIs.
- **Tight Integration with Evolved Legacy AIVs:** This BB analyzes the interworking between novel 5G AIV(s) and evolved LTE at RAN level, as opposed to legacy networks, where interworking is at the CN level. The envisioned resource abstraction framework aims at efficient interworking.

- **RM for Inter-Network Collaboration:** This BB utilizes software defined networking (SDN) principles towards RM for inter-network coordination among different operator networks and/or different AIVs considering various usage scenarios.
- **Context Management** schemes need to be in place for the synchronous mechanisms to operate efficiently in the demanding 5G environment. Means to support effectively and efficiently a fast RM are required, where the context data include the ones from users, access nodes, and CN. The trade-off between the amount of data to be gathered, the complexity of RM algorithms and the corresponding enhancements in terms of network performance are considered.

3 Agile RM Architecture and Design Paradigms

Section 2 has provided the motivation and high-level framework of the agile RM. Herein, the agile RM framework will be described from functional architecture, protocol architecture, and deployment architecture perspectives. These architectural descriptions are the outcome of the methodology outlined in Figure 1-1, where a functional decomposition is applied to TeCs, though which BB descriptions are harmonized and the most essential LEs and LEGs are identified. These LEs and LEGs, and their inter-relations constitute the functional RM architecture which is provided in Section 3.1. Thereafter, Section 3.2 presents the options on how the functional RM architecture can be mapped onto the RAN protocol architecture and how the protocol stack can be split among RAN entities.

3.1 Functional RM Architecture

The functional architecture of the agile RM framework is depicted in Figure 3-1. It is worth noting that for the sake of simplicity AIVs are not depicted on the 5G-UE. A 5G-UE may support one or more AIVs. Furthermore, only the RAN and UE sides are illustrated, where the interface to the CN is discussed on Section 3.2. Section 3.1.1 describes AIV-overarching RM that comprises LEs and LEGs, which are operating over multiple AIVs to map the data flows to appropriate AIVs based on the context received. The overall functional architecture is formed by functionality frameworks of AIV-overarching RM and Intra-AIV RM. LEs and LEGs that need to be tailored to each AIV construct the Intra-AIV RM which will be described in Section 3.1.2. Here the AIV-overarching functionalities are mainly located with the access network – outer (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 limited by the AIV design and hence are assumed to be located within the AN – inner (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.

The functional architecture of the agile RM framework is designed to provide flexibility in order to extend the framework functionality for future needs. For example, the framework can be extended by adding a new LE to an existing LEG, or by adding a new LEG. In addition, whenever a new AIV is added, the framework functionality can be extended by modifying the relevant LEs.

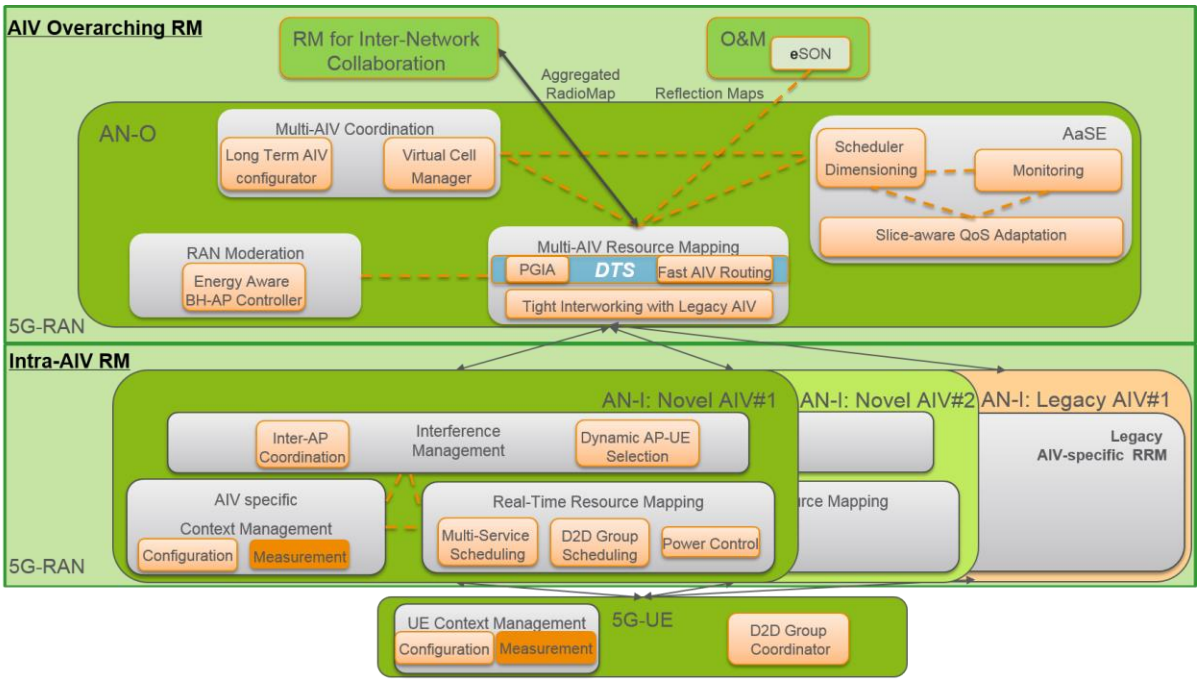

Figure 3-1 The functional architecture of agile RM framework.

3.1.1 AIV-overarching RM functionality framework

The AIV-overarching RM functionality framework enables the 5G RAN to implement various virtual functions which are AIV-agnostic potentially in a centralized unit, in order to achieve various features and KPIs required for the system. The functionalities could be located within a BS as well, depending on the deployment strategy and use case targeted by the network operator. The main focus of the functionality framework is to enable the system to achieve the system level target KPIs with the help of various LEGs. AN-O layer is assumed to be logically similar to the central unit / upper layer of the 5G-gNB considered currently in 5G RAN architecture [3GPP16-38801], with the LEGs located in this layer having the ability to control the operational paradigms of multiple AIVs. The AN-O layer is assumed to consist of some key LEGs that are,

- **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 4.3),
- **AIV agnostic Slice Enabler (AaSE)** enables performance guaranteeing multi-slice RM with real-time SLA monitoring (see Section 4.4),
- **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 Section 4.4),

- **RM for Inter-Network Collaboration** provides needed context and control mechanisms to enable inter-working among different networks, e.g., 5G, WiFi, and WiMAX (see Section 4.7), and
- **Multi-AIV Resource Mapping** capitalizes on the interfaces with the aforementioned LEGs 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 4.5 and Section 4.6).

These LEGs tightly interwork with each other within the AN-O, and LEs such as RM for inter-network collaboration and enhanced self-organizing network (SON) functions within the operation and management (O&M), to enable a resource management framework which can achieve xMBB data rates, as well as uMTC and mMTC type of communication in 5G RAN, for ensuring ultra-reliability, low latency, and massive densification of the network.

3.1.2 Intra-AIV RM functionality framework

One of the main goals of METIS-II is to develop synchronous control functions which could be used for resource and traffic management over multiple AIVs that form the 5G RAN [MET-II16-WP]. The LEGs presented in this section are employed considering a single AIV, i.e., they are functionality groups that need to be specifically tailored to each AIV. Furthermore, schemes that operate in a very fast time scale such as adaptive interference management techniques may be required. This is due to the dynamic nature of the 5G environment and the added constraints such as the frequent reconfiguration of the radio links introduced by the use of multiple AIVs in UDNs. Similar constraints are applicable for spectrum and traffic management which need to be implemented in a dynamic and adaptive manner in order to satisfy 5G requirements. The key LEGs that are part of the Intra-AIV functionality framework are the following:

- **Interference Management** 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 nomadic nodes (NNs). The coordination and selection of the nodes that are part of the schemes are two key LEs as noted in Figure 3-1. Please see Section 4.1 for further details.
- **Real-time Resource Mapping** is a collection of LEs that deals 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) uplink power control such that the users transmit in a cooperative way without causing much interference to each other [FAN17-D42]. Further details can be found in Section 4.4.
- **AIV-specific Context Management** handles all context parameters that are bound to one specific AIV. Both measurements and configurations aspects are contained in this LEG. Further details can be found in Section 4.8.

This set of LEGs is present for each AIV that co-exists in the RAN, hence there are multiple instances of the functionality framework depicted in Figure 3-1. A more detailed description of all involved LEGs and the corresponding functionality decomposition is discussed in detail in the following section of this deliverable.

3.2 Radio Protocol Architecture and Deployment Options

This section deals with the RM functionality that is applicable for a multi-AIV environment, comprising both 5G and evolved legacy AIVs [MET-II16-D41], as it was shown in Figure 2-1. Thus, the RM functionality is AIV-overarching, which implies that the TeCs described here remain agnostic to the AIV-related physical layer design or operation. Furthermore, the AIV-overarching RM functionality should consider the synchronization requirements of the AIVs, as well as different deployments including dynamic topology settings, and the imposed limitations of the legacy networks.

As shown in Figure 1-2, D5.2 takes into account the research described in two additional technical deliverables within METIS-II, namely D4.2 and D6.2. While in D4.2 potential CP and UP functional split options are explored and necessary interfaces are identified [MET-II17-D42], in D6.2 a set of CP and UP architecture options are investigated with different RAN functional split options between centralized and distributed network functions [MET-II17-D62]. In this section, D5.2 complements the outcome with a possible split between a common AIV-overarching CP/UP and AIV-specific CP/UP with respect to RM functionalities. One split option at Packet Data Convergence Protocol (PDCP) level for a particular deployment scenario is also illustrated in Figure 3-4.

Based on the detailed view on available split options and the corresponding information exchange between CP and UP layers [MET-II17-D42], functional extensions are introduced to both layers from an RM perspective. In the common CP depicted in Figure 3-2, all identified LEGs of the RM framework, such as AaSE for network slice handling, RAN moderation to take into account energy efficiency, and the Multi-AIV Resource Mapping to support traffic steering are introduced as additional AIV-independent functionalities. For the AIV-specific CP, extended IM and real-time resource mapping functionalities are introduced. Basically, state of the art scheduling and inter cell interference coordination (ICIC) schemes are extended by innovative advanced means, further described in detail in Section 4. The overarching common CP might be located in an aggregation point in the RAN.

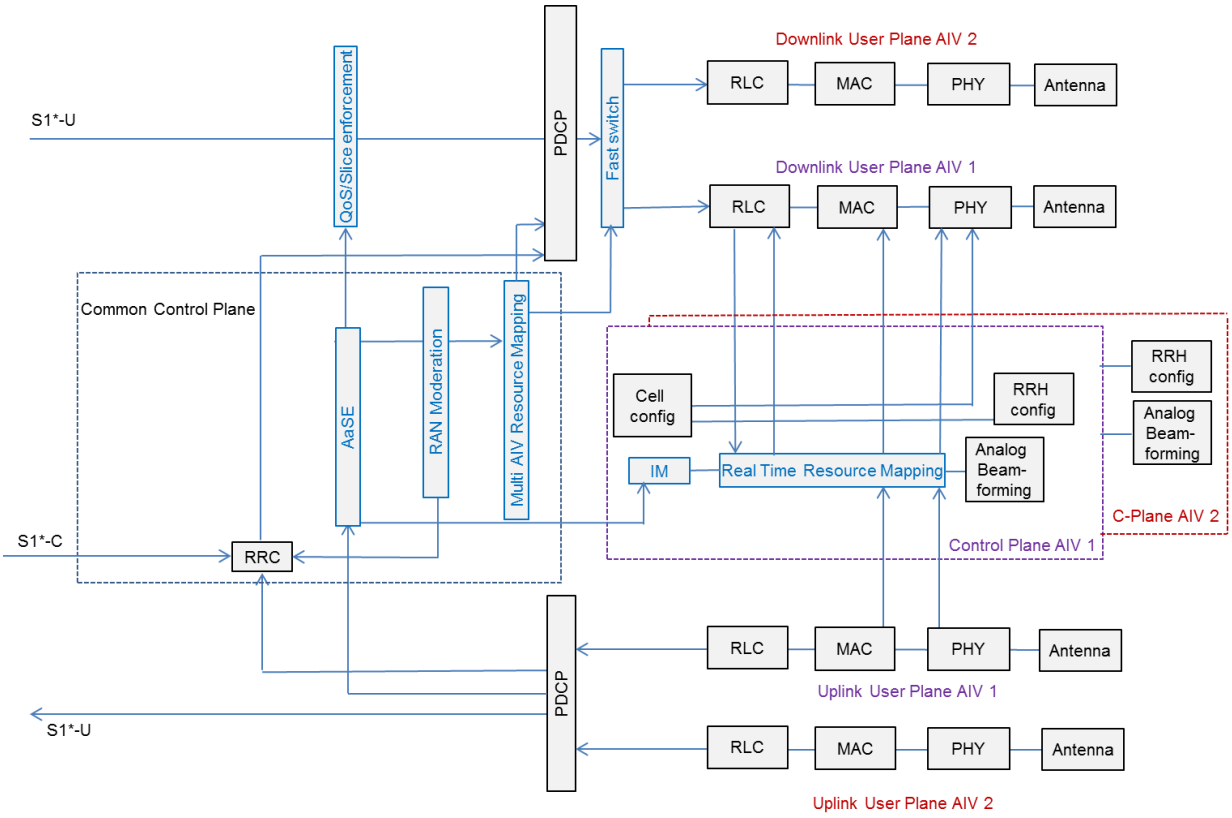

Figure 3-2 Extension of the radio protocol architecture

In Figure 3-4 the extended protocol architecture is now mapped to the deployment option, depicted in Figure 3-3, considering a CP/UP split (M8) [MET-II16-D22] at PDCP level. As an example the RM framework could be deployed by the introduction of a central access controller (CAC) to the RAN [MET-II17-D62]. Here, the overarching RM functionality is placed at the CAC instance, with the split done at PDCP level. Figure 3-3 gives an indication which part of the CP is harmonized at the CAC. The higher layer UP network functions (UPNFs-H), such as QoS/Slice enforcement and PDCP plus higher and medium CP network functions (CPNFs-H/M), such as Radio Resource Control (RRC) and AaSE, traffic steering or RAN moderation, are deployed at the CAC. Lower and medium UPNFs-M/L (Radio Link Control (RLC) to PHY) as well as lower layer CP processing (e.g. MAC scheduling, IM defined as CPNFs-L) will be done at the BSs.

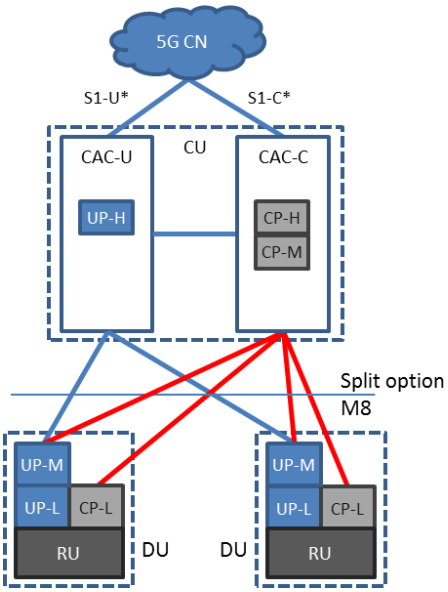

Figure 3-3 Example of a possible deployment option with central access controller based on [MET-II17-D62]

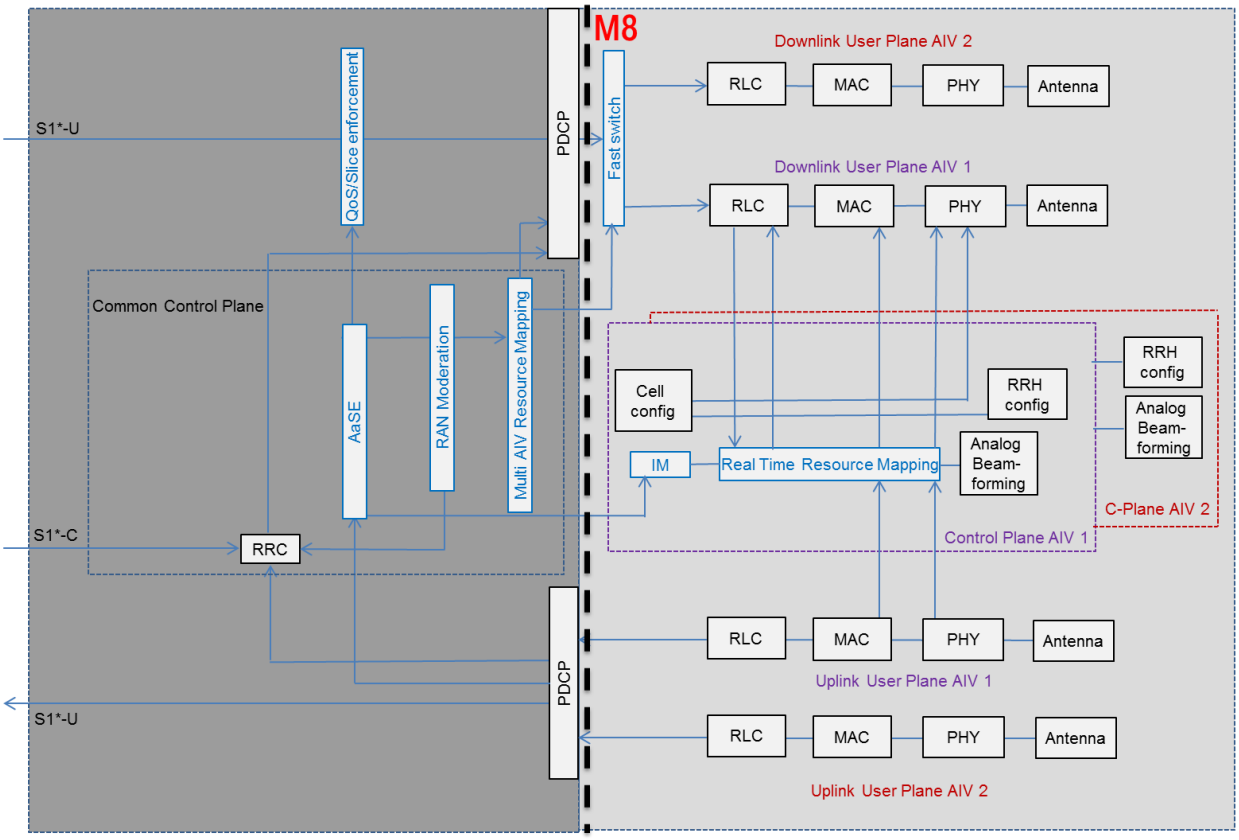

Figure 3-4 Example of mapping a protocol split option to the RM framework

In Figure 3-5 a detailed example is given based on a message sequence chart, how the RM framework will exchange information and extend data processing, to transmit in an optimized manner. The detailed consideration on how the CP and UP functions interact with each other is used to give the reader further insides how a data flow which arrives from the CN is treated when the RM framework is applied in a 5G RAN.

During the orchestration process [5GNORMAD32] [MET-II16-D22] the AaSE receives SLA and QoS policies for each network slice which is instantiated by the management plane of the system as well as which RRM network functions (e.g. MAC scheduler) can be used for each slice. The AaSE then activates correspondingly orchestrated functions and might adapt the usage during runtime of the system. After the orchestration procedure is done, data packets which arrive from the 5G CN via S1*-U interface are received by the AaSE enforcement function. Each packet already contains a tag which indicates the affiliation to the network slices [3GPP16-23799]. AaSE knows about SLA and QoS policies of each network slice (step 1) and is aware of the instantaneous condition of each data flow by monitoring QoS KPIs, such as data rate, delay, and BLER on the AIVs (step 19). Based on the control entity of AaSE the data packets' QoS tag might be adapted when business driven SLA requirements are in danger to be violated. The AaSE enforcement function then changes the QoS tag of the data packet to improve the SLA of the corresponding network slice (step 3). The SLA/QoS mapping information of AaSE is also provided to the RAN Moderation functionality to use the adapted information to decide which AIV might be used the best from energy efficiency perspective (step 2). Besides that, AaSE gives information to the AIV-specific IM block which schemes to be used the best for adapted QoS fulfillment (step 7). In step 4 the control information derived from previous steps is passed to the Multi-AIV Resource Mapping block which then decides on which AIV the data flow is steered, by sending the decision to the harmonized PDCP layer (step 5). To enable fast switching on a common AIV, such as evolved legacy (e.g., 3GPP LTE-A Pro) and 3GPP 5G NR which will be both CP-OFDM based [3GPP17-38802], the Multi-AIV Resource Mapping can even react faster, e.g. to follow fast fading (step 6). Step 9 or 10 further processes the data to RLC which gives information to the AIV specific CP Real Time Resource Mapping. It is responsible to map the data to radio resources under consideration of enabled IM schemes (steps 7 and 8) and service type of the data flow on a Transmission Time Interval (TTI) (≤ 1ms) basis (steps 12,13,15). To derive optimized decisions the Real Time Resource Mapping entity needs measurement information, such as Channel State Information (CSI) reports and Hybrid Automatic Repeat Request (HARQ) Acknowledgement (ACK)/ Negative Acknowledgements (NACKs), which are received in steps 14, 16 and 17. In step 18, the received data flow by the UE is sent back to the 5G CN over S1*-U interface to be further processed. The RRC entity is informed about performance of lower layers in step 20.

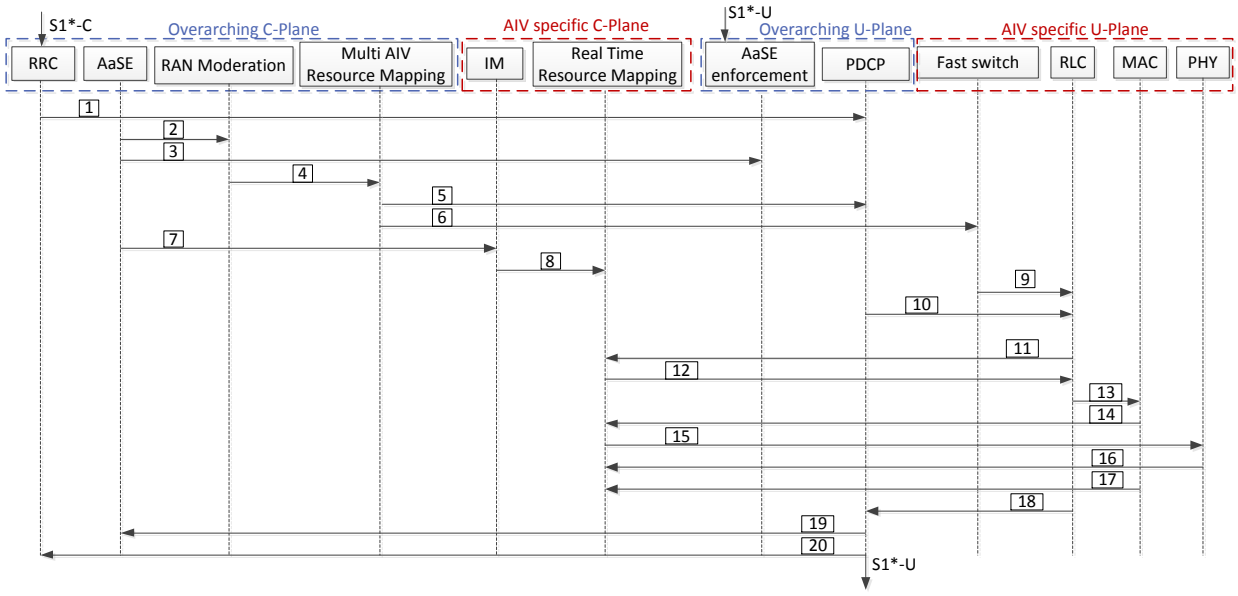

Figure 3-5 Example of a message sequence chart for the RM framework, where the processing of a data flow is depicted

4 Enabling Technologies for Agile RM

In this section, an overview of the enabling technologies for the envisioned agile RM framework is provided organized in BBs – a combination of logical entities - and resultant 5G RAN architecture design recommendations. The rationale behind each design recommendation is also provided, along with the implications on the 5G RAN and detailed performance analysis. In order to provide a comprehensive overview of the BBs within the agile RM framework, a functional decomposition of each BB and the possible deployment architecture from a network operators' perspective, is described. Each BB proposes various design recommendations[10] which are essential for achieving 5G RAN design targets from an RM perspective. In addition, it equips the system with requisite agility and flexibility to provision advanced features, while tightly interworking with legacy AIVs.

4.1 Interference Management

Co-channel interference is an inherent limitation of wireless systems employing universal frequency reuse, i.e., a frequency reuse factor of one. While such feature maintains high spectrum utilization, the interference limits spectral efficiency and thereby network and user performance, unless some type of interference management is employed. Overcoming interference is therefore essential in ensuring high capacity and wide coverage of high end-user data rates, as well as robust and efficient communication.

A common objective of most interference management schemes relates to improve cell edge user throughputs, coupled with varying levels of network coordination [PSQ+13][HRT+14]. For instance, centralized scheduling as well as Inter-Cell Interference Coordination (ICIC) introduced in LTE Release 8 aims to improve cell edge signal-to-interference-plus-noise ratio (SINR) through frequency and power allocation. In later LTE releases, the backward compatible enhanced ICIC (eICIC) framework for heterogeneous networks provides the ability to mitigate interference on data and control channels in frequency or time domain. In the time domain, interference avoidance within eICIC is facilitated using Almost Blank Subframes (ABS) by scheduling the intended and interfering signals on different subframes. These schemes were complemented by further enhanced ICIC (FeICIC) from Release 11 [3GPP15-36300], mainly focusing on interference handling on user side (via interference cancellation schemes). In frequency domain, Carrier Aggregation (CA) introduced additional degrees of freedom that can be exploited for interference management purposes, as interference is partly avoided by scheduling the control channels of the macro cell and small cells on different carriers. Fast

[10] Design Recommendations are tagged by dedicated IDs, i.e., DR-*x* where *x* indicates the numbering.

cross-carrier scheduling of the data can also help to reduce interference when there is a strong aggressor cell present while with Coordinated Multi-Point (CoMP) it is possible to take a network-wide approach at interference management by considering a larger both dynamic and fixed set of cooperative radio nodes [3GPP13-36819]. 5G RAN will enable multi-layer multi-node connectivity over one or more AIVs resulting in much higher data rates and additional degrees of freedom for the fast synchronous control. The performance gain of CoMP depends to a large extent on the tightness of the coordination, i.e., the level of synchronization that is needed. It is known that CoMP is sensitive to backhaul latency for the signaling, knowledge and accuracy of channel state information (CSI), and large communication overhead, which is the reason why much of the theoretical gains are hard to materialize in practice. In the end, designing a proper interference management scheme depends on the UC, deployment scenario, and size of the cooperative set, while maintaining a reasonable degree of flexibility for the RM. The cooperative design recommendations presented in this section focus on the innovative mechanisms proposed for increasing both: the user throughput KPI (required in generic eMBB scenarios) and the coverage KPI (required in mMTC scenarios) specifically addressed on DR-5.

4.1.1 BB Description

Functional Decomposition

The internal functionalities of the BB, i.e., LE representation, are provided in Figure 4-1. A brief description of each LEG or LE is provided as follows:

- **Measurements (LEG)**: 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 access points, APs (as needed for cross-link interference management in dynamic TDD scenarios).
- **Measurement Signaling (LE)**: Handles all the signaling coming from different network elements as explained above.
- **Configuration Signaling (LE)**: Handles all the information messages needed to configure all network elements involved in any interference management scheme, i.e., it takes care of the network coordination part.
- **Node Selection (LEG):** 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 FQAM. This LEG also provides an interface to the RAN moderation BB as the set of serving NNs needs to be selected from the available set of candidate NNs.
- **Coordination & Scheduling (LEG)**: This LEG contains the intelligence related to all network coordination algorithms and schemes necessary to apply the interference management schemes. Nodes involved in this LEG include both static and dynamic APs (i.e., BSs and NNs) as well as BS clusters.

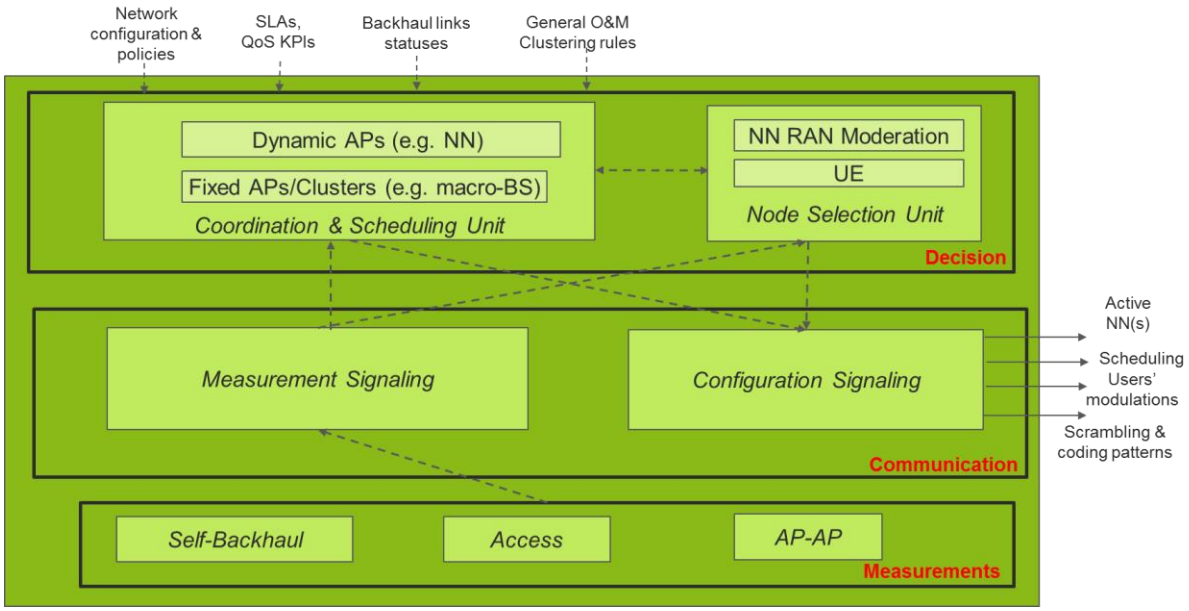

Figure 4-1 Interference management functional decomposition

Deployment Architecture

In Figure 4-2, we show some deployment considerations of the interference management BB. In particular, the picture is intended to provide insight to where in the physical architecture of the network each of these LEGs would sit. The illustration shows how the macro BS would contain part of the functionality of each LEG, while small cells would be able to manage most of the functionality themselves (except for certain node selection functions). The UEs would mostly provide measurements to the network. In addition, NNs would need to be part of the coordination algorithms triggered at the coordination LEG, and an interface needs to be provided to the node selection unit (in particular, to the RAN moderation BB) to appropriately switch on or off NN service as instructed. It is worth mentioning that the interference management set of functions can be also implemented in a centralized manner via e.g., a C-RAN deployment, and the deployment option shown below does not exclude an additional level of centralized processing.

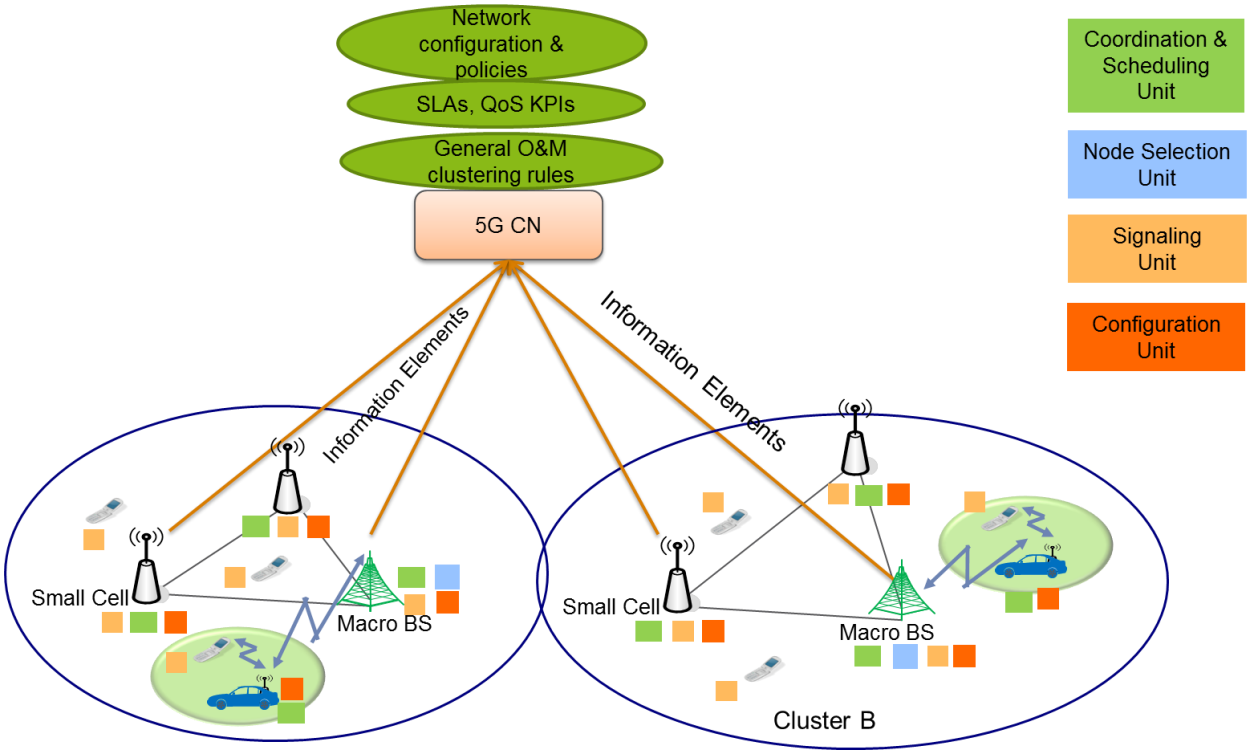

Figure 4-2 Deployment considerations for interference management

4.1.2 Design Recommendations

DR-1: Enable adaptive interference mitigation to cope with dynamic radio topologies

Rationale

A key deployment characteristic in UDNs, which was also discussed in [MET-II16-D51], is the employment of dynamic APs (e.g. NNs, unplanned small cells) to further improve spectral efficiency by enhancing the spatial reuse on demand. To this end, shared spectrum among APs could be a potential solution in order to enable more efficient handling of resources. However, in such case, a holistic inter-cell RM framework is highly required to allocate RAN resources in a way that interference is mitigated while keeping the spectrum utilization high. Furthermore, the full operation of such dense fixed small cell deployment is not always needed due to the inhomogeneous distribution of traffic over time and space. Hence, fixed small cell / relay deployment has the disadvantage of increased capital expenditure (CAPEX), e.g., due to deployment of additional wireless access nodes, and operational expenditure (OPEX), e.g., due to the incurred site leasing costs, although small cells need to be operated only partially. Towards the 5G system, the concept of dynamic radio topology has emerged as a complementary enhancement to fixed small cells [BRZ+14][BRZ+15][MET15-D66][NGM15]. Within the framework of dynamic radio topology, the network shall react quickly and dynamically

to fulfill the increased service requirements in a certain time period and at a target service region. On this basis, one component to enable dynamic radio topology is the introduction of NN aka vehicular NN (VNN) operation. A VNN is a movable access node with a wireless backhaul link, which can provide coverage extension and/or capacity improvement on demand. VNNs can be integrated into vehicles, e.g., within a car sharing fleet. The notion of “nomadic” implies the availability of the VNNs over time and space, i.e., VNNs may change location and be available in another location. Further, it is assumed that a VNN is only activated when the car is parked; thus, the VNN is static during its operation.

The concept of this work provides an UE-centric interference management in heterogeneous UDNs complemented through the notion of the dynamic radio topology by means of selecting appropriate access nodes. Here, two case studies for a hotspot area and a 5G RAN consisting of NNs under macro BS coverage are evaluated. In the first case study, a dense NN deployment is considered, where NNs are mounted on cars that are parked along a road side (see Annex A.1) [PCF+16]. To overcome increased interference among NNs due to close proximity, coordinated resource allocation and joint transmission (JT) are applied adaptively based on backhaul conditions (i.e., between access node and serving BS), load constraints, and service type. In the second case study, the flexibility of NNs is exploited by selecting the closest NNs to a hotspot area. The performance of such a dynamic radio topology is compared to a fixed small cell deployment, namely, pico cells [SBS+17]. In this case, a minimum distance of 50m is set between any active NNs, to reduce the impact of interference among selected NNs and to increase the spatial diversity.

RAN Implications

In order to benefit from NN operations, new inter-cell RM schemes are needed for the coordination of the access nodes in dynamic radio topology by a centralized coordinator (e.g. at the macro cell site). Also, the backhaul link measurements and activation/deactivation commands may possibly imply new signalling elements on the wireless backhaul link.

Results and Analyses

In the first case study, the key interference management mechanisms applied are JT between the access links of NNs (i.e., between NNs and UEs) when ideal BH is provided. The selection of candidate UEs for JT is based on the difference of reference symbol received power (RSRP) measurements from serving and neighboring NNs. Given the number of UEs experiencing low channel quality, a number of resource blocks (RBs) are reserved for JT, and resource allocation between different NNs is done. For remaining UEs, interference management is applied, where dynamic frequency partitioning or muting of resources for some NNs is performed [MET-II16-D51].

In [MET-II16-D51], it was shown that when employing more NNs, without Interference Management the throughput gain decreases significantly; nevertheless when we provide interference management on top, we can have much higher gains. Additional evaluation results are depicted in Figure 4-3 for the Cumulative Distribution Function (CDF) of user throughput for

different activations to better capture the gain trends to both cell center and cell edge users by interference management. Evaluation results show that 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.

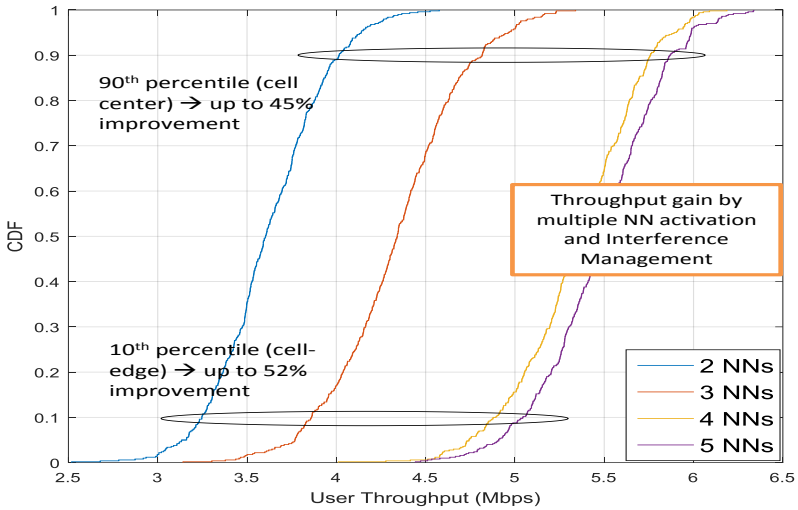

Figure 4-3 CDF of user throughput for different number of activated NNs in dense NN deployment (case study 1).

In the second case study, there are 20 randomly located and inactive, parked NNs, present at each macro cell, in average (see Annex A.1.1). UEs are distributed in the network as random hotspots (1 per macrocell, in average). Constituting each hotspot, UEs are randomly distributed within an annulus region bounded by the radii of 10 and 50 m, centered at a random point in the macro cell. The number of UEs per hotspot is 25 and 50 in the uplink and downlink simulations, respectively. A specific number (1, 2, or 4 NNs) of the closest NNs to the center of each hotspot are activated by the network [SBS+17]. UEs are not forced to connect to the activated NNs but they still attach to the node with the largest RSRP value, as the conventional cell selection scheme. For comparison, the same number of pico cells is considered, where pico cells are at fixed locations. Yet, due to dynamicity of the hotspot (e.g., a street event), from hotspot perspective pico cells are randomly located and there is no correlation between the hotspot and pico cell locations. The downlink throughput gains at 10%-ile and 50%-ile user throughput with respect to pico cell deployment are shown in Figure 4-4 (while the uplink results are provided in Annex A.1.1). It can be seen, for example, by activating one NN closest to the hotspot, 10%-ile throughput gain is around 150% compared to one pico cell.

In addition, although VNNs aka NNs are assumed to be stationary during their operation, their availability changes with respect to time and space according to their battery state or driver needs (hence, the term “nomadic”). Furthermore, as VNNs are integrated into vehicles, due to low height of 1.5 m like the one of UEs, severe fading characteristics can be expected on the wireless backhaul link as opposed to well-elevated small cells (e.g., at 5-10 m height for fixed

relay nodes). Accordingly, to ensure the expected benefits of VNN operation, active VNNs shall be properly selected such that the backhaul link quality is optimized. The proposed strategies along with the associated analyses are provided in Annex A.1.2.

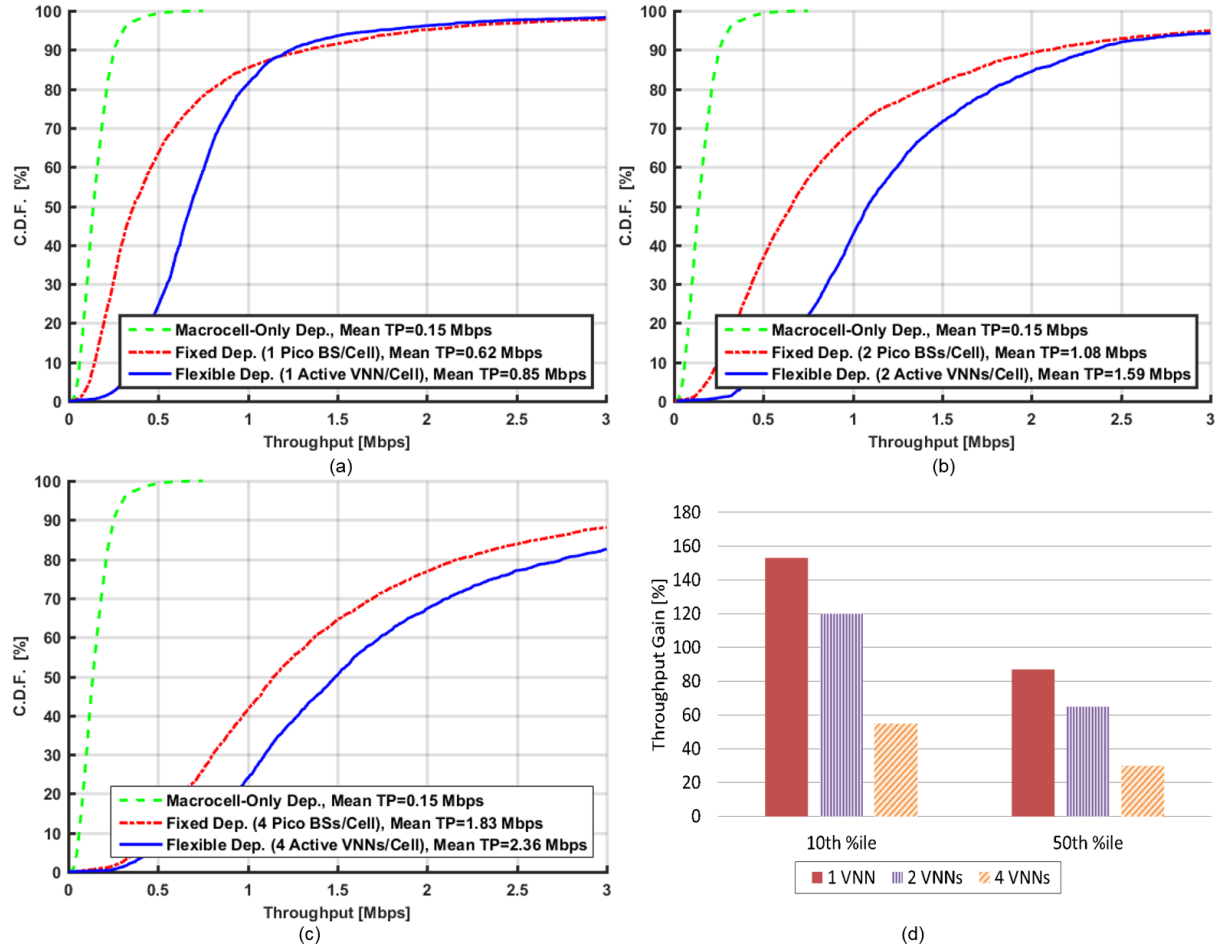

Figure 4-4 User throughput CDFs on downlink with different deployment scenarios (a), (b), and (c), together with the throughput gain of the VNN deployments compared to pico cell deployments at the 10th and the 50th percentiles of the CDFs (d). Mean user throughput (Mean TP) levels are provided in the legends.

DR-2: Enable adaptive interference mitigation exploiting interference-resistive design via advanced modulation techniques

Rationale

Conventional interference management approaches assume that interference follows a Gaussian distribution. It is known that the worst-case additive noise in wireless networks, in the sense of the channel capacity, has a Gaussian distribution. From this observation, one can expect that the channel capacity can be increased by designing radio transmissions to result in inter-cell interferences (ICIs), which are non-Gaussian. In practice, the distribution of ICI

depends on the modulation schemes of the interfering BSs. Therefore, an active interference design improving channel capacity in the presence of interference, particularly in the low SINR regime, can be achieved by applying a new type of modulation called FQAM [MET-II16-D41]. This modulation scheme is a combination of frequency shift keying (FSK) and quadrature amplitude modulation (QAM) compatible with OFDMA systems, as described in [HSL+14].

In this design recommendation, a resource partitioning scheme to support FQAM in high interference scenarios is evaluated. The proposed scheme partitions radio resources into orthogonal parts for QAM and FQAM along extended radio resource dimensions, namely space and frequency. This can be achieved by incorporating advanced beamforming algorithms or performing a frequency-based split of FQAM resources to effectively improve the data rate of the edge users experiencing heavy interference.

In the space domain, the proposed resource partitioning concept can be easily applied in the following scenarios. Firstly, all beams from all cells causing interference to their respective neighboring cells would employ FQAM to contribute to a deviation of the interference distribution from a Gaussian distribution. However, it may not, necessarily, optimize the performance of the entire system since, even though the cell edge users may benefit from the employment of FQAM, other users located near the BSs should still apply QAM modulation to achieve higher throughput. In this regard, FQAM should be applied to only a subset of the interfering beams in the network with the purpose of maximizing spectral efficiency of the entire system, where the number and selection of beams utilizing FQAM could be obtained using many different methods such as those employed in interfering nulling schemes.

The details of the frequency-based approach can be found in Annex A.2 and [MET-II16-D51].

RAN Implications

This TeC requires some level of coordination among BSs to apply the corresponding interference management techniques. The coordination of interferences between the cells when using the frequency-based FQAM method can be done by simple exchange of information, such as LTE-A X2 relative narrowband transmit power or any X2AP messages [3GPP16-36423] in general. No explicit exchange of interference management information is needed in cases when it is not suitable or critical for performance, thereby reducing overall signaling and related delays. Still, signaling information exchange among BSs via supporting protocols is required to determine the beams or the sub-frames where FQAM should be applied, and this exchange can be implemented using either a distributed or a centralized approach. The necessary notification between adjacent cells can be exchanged on X2* or can be facilitated by multi-connectivity (e.g., using low frequency AIV) in case one of the connections can achieve a higher visibility of interference pattern per contending zone via UE reports. In summary, the three main RAN implications are shown below:

1. Flexible signaling exchange needed among BSs
2. Information exchanged can be facilitated via X2* or multi-connectivity legs

3. Specification implication is minimal

Results and Analyses

Evaluation results focusing on the spatial dimension are presented, showing significant performance improvements when compared to regular QAM. The system level performance of this TeC is evaluated using a 21-cell grid, 3GPP LTE radio frame structure with short cyclic prefix, 20 MHz bandwidth, 43 dBm Tx power at each BS, constant bit rate (CBR) traffic sources, a proportional fairness scheduler, and full-buffer assumption. For FQAM, we consider M-ary turbo coding with Complex General Gaussian decoding [HSL+14]. Using beamforming-based space domain resource partitioning and centralized switching mechanism, the addressed 5G KPI is the experienced user throughput. The scenario we investigate is described as follows: Two beams are assumed for each BS. The first beam, which uses QAM or FQAM, is directed to an UE located at the center of three adjacent BSs and thus experiencing high level of interference. In addition, there is one more UE per cell that experiences low level of interference and is always covered by another beam with QAM. In case FQAM and QAM co-exist in the same cell, we call this modulation distribution hybrid (i.e., QAM+FQAM).

Figure 4-5 (left-most) shows that the cell edge UE throughput can be still significantly improved by applying FQAM to those UEs experiencing high level of interference. This is because the availability of radio resources for the UEs in high SINR regimes also depends on the resources consumed by UEs in low SINR regimes and FQAM can improve throughput of the latter. Figure 4-5 (center, right) shows the average and peak UE throughput and as observed, if the hybrid scheme is applied to all beams, the average and 95%-ile throughput when using QAM are still lower than with the hybrid scheme. However, applying FQAM does not affect average UE throughput significantly and only causes a minor peak UE throughput improvement since peak rate is normally achieved by those UEs experiencing low level of interferences and thus QAM should be used.

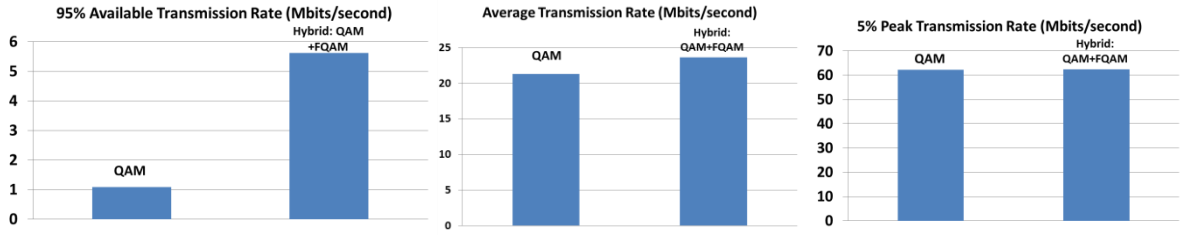

Figure 4-5 5%-ile UE throughput, average UE throughput, and 95%-ile UE throughput results.

DR-3: Employ transmit precoding to mitigate same- and other-entity interference for dynamic TDD in UDN

Rationale

Dynamic TDD is a promising duplexing technique for handling fast-changing traffic, especially in short-range communication where transmit powers for UL and DL tend to be similar. At the same time, it generates new interferences (UE-to-UE and BS-to-BS) in addition to existing interferences (BS-to-UE and UE-to-BS). In UDNs where probability of line-of-sight (LoS) between a UE and its interferers increases, managing these interferences becomes even more important.

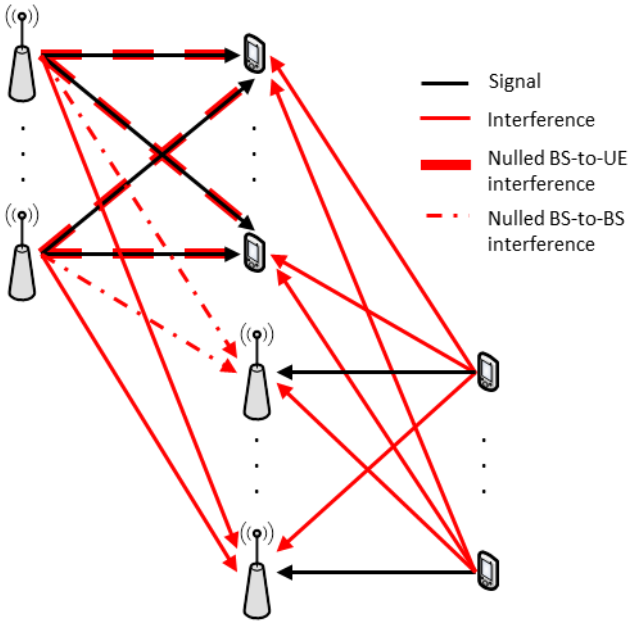

Figure 4-6 Multi-cell dynamic TDD network employing JT-DS.

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. Since the proposed scheme focuses on transmit precoding, the aspect of BS decoding which can null UE-to-BS interferences is outside the scope of this work and therefore not treated here. 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, see Figure 4-6. Since DL BSs are not aware of which symbols UL UEs will transmit beforehand, dummy symbols are transmitted virtually with zero power. The dummy symbols can therefore be thought of as a mathematical construct corresponding to making certain columns zero after the precoder matrix has been generated to enable zero forcing also to UL BSs. The proposed scheme is denoted as

joint transmission with dummy symbols (JT-DS), and relies on both DL and UL traffic for its implementation. The number of UL BSs that can participate in the precoding is also constrained by the number of cooperating DL BS antennas. For the selection of UL BSs, we consider those serving UEs with worst UL rates calculated according to our chosen baseline [CS15] which provides a lower bound on performance if interference is left unmitigated.

For the performance evaluation, we consider a simplified two-dimensional environmental layout based on UC2 described in [MET-II16-D21], which includes walls but excludes furniture. Results for an open indoor grid deployment are found in Annex A.3. For channel propagation, the average path loss of the WINNER II A1 indoor model [WINNER08] is selected together with fast fading to model local scattering. Channel estimation is assumed to be ideal, and operating frequency is set to 2 GHz as form factor is not considered to be an issue here for the omnidirectional single-antenna BSs. UEs are dropped uniformly over the office area constrained on that each small cell serves no more than a single active user due to the ultra-densification. The shorter UE-to-BS distances means that we assume the same maximum transmit power of 100 mW for both DL and UL. While UL UEs transmit at full power and independently of all other transmissions, in the DL a convex optimization problem is solved for the power allocation with, as noted earlier, zero transmit power for the dummy symbols. Average traffic demand is distributed evenly between UL and DL, though instantaneous traffic demand may be highly asymmetric. Sum-rate based on Shannon's capacity formula is taken as objective function, and the ensemble averaging relies on a total of 1000 snapshots. For active users, queues are infinitely backlogged so that bits are always available for transmission. Finally, we assume single antennas for BSs and UEs to ensure low cost. Another approach taking into account network throughput and energy efficiency is provided in Annex A.9.

RAN Implications

Accurate estimation of the gain and phase between BS-to-UE and BS-to-BS links is assumed in order to construct precoders. Pilots should therefore be known to all BSs in the cooperating set to avoid contamination in the channel estimation process. The amount of instantaneous UL and DL traffic demand should be known by the coordinating entity in charge of precoder design as no UL traffic or too much DL traffic will render the proposed scheme unused. Moreover, in the absence of same-entity interference, phase estimation for BS-to-BS links may become redundant.

Results and Analyses

Figure 4-7 shows the average UL and DL sum-rate, respectively, as a function of system utilization defined here as the ratio between number of scheduled UEs and number of deployed BSs. For UL, JT-DS significantly improves performance thanks to the mitigation of BS-to-BS interference, but uncontrolled UE-to-BS interferences limit further gains. The number of UL BSs that can participate also decreases with utilization as DL traffic increases. In comparison, JT which only nulls inter-user interference between DL UEs will have marginal effect on UL performance. For DL, nulling BS-to-UE interferences will significantly improve performance with JT, except for very low utilizations where the already low interference is further reduced by

walls. The limited gains of JT-DS are attributed to a more ill-conditioned precoder compared to JT due to the inclusion of UL BSs, resulting in lower DL transmit powers in order to not violate the BS power constraint. In addition, transmission of dummy symbols will not contribute to the DL sum-rate, and UE-to-UE interferences further limit DL performance. At full (100%) utilization, UL BSs can no longer participate in the precoding, in which case the precoder matrices of JT and JT-DS are identical and square. Based on the results, the best transmission scheme depends on the utilization percentage and the preference for UL or DL sum-rate maximization.

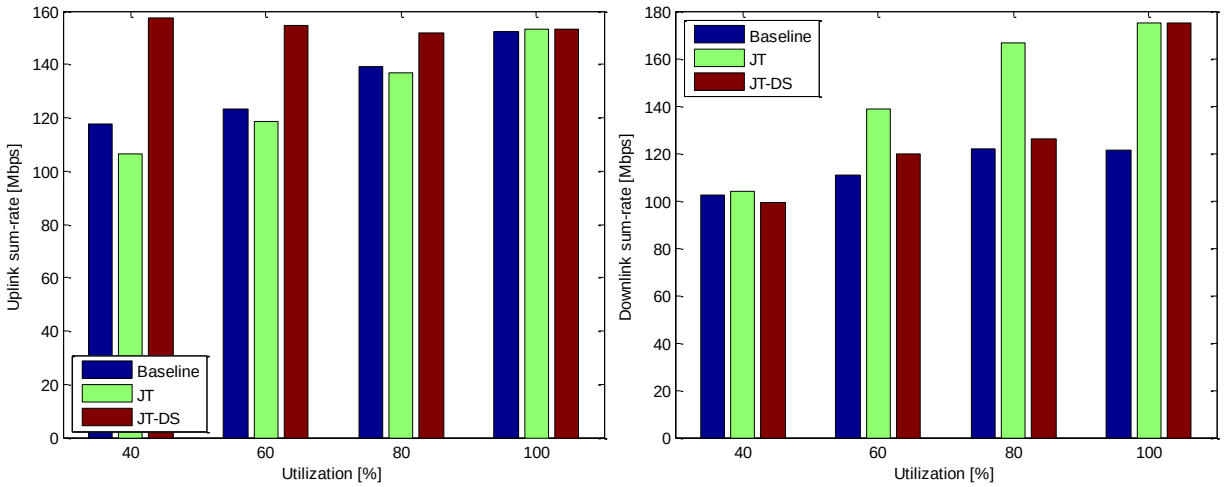

Figure 4-7 Average UL (left) and DL (right) sum-rate.

DR-4: Support dynamic selection of transmission path (DL or UL) for pilots in dynamic TDD systems

Rationale

In this contribution, a practical novel TDD design principle is proposed for massive MIMO HetNets that leverages the inherent features of a flexible TDD design to mitigate both the beamformed interference caused by the pilot contamination effect and BS-to-BS interference. The design is based on the key observation that the transmission path chosen for training by the non-massive MIMO base stations plays an important role in the interference behavior of the network. This means that the data slots need to be configured accordingly. Figure 4-8 shows a sketch of the main concept. The pilot of the macro cell may be contaminated by another pilot transmitted in the downlink (S_D) or in the uplink (S_U), depending on the transmission path that is used for training at the small cell. This creates two possible cases of beamformed interference, either directed to the user or to the small-cell base station (SBS). In that context, the small cell needs to decide which transmission path to use during that time slot. The obvious answer is the configuration that avoids receiving signal at a network element while the beamformed interference is being directed to that same network element. However, the freedom to choose is limited by the load distribution, which imposes the number of UL and DL slots in the cell. We provide answers to the following two questions:

- Which transmission path (uplink or downlink) should be used for training at the small cell tier?
- In which order should UL/DL slots be allocated to prevent both *cross-link* and *beamformed* interference while matching the load distribution?

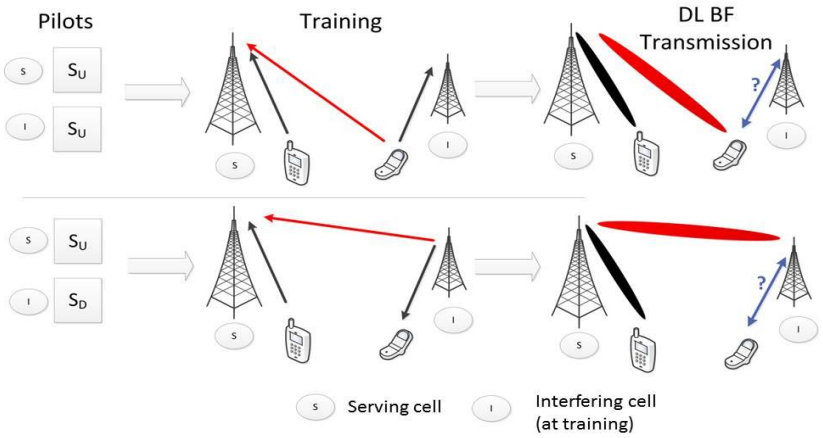

Figure 4-8 Sketch of main observation upon which the design rules are based

To answer the previous questions the following observations have been made:

- The pilot overhead introduced by employing downlink pilots (S_D) in a TDD massive MIMO system is very high. Hence, when using a large number of antennas, channel training should be performed in the uplink (S_U) to avoid the costly overhead.
- The rest of single-antenna base stations or those equipped with a smaller number of antennas may choose either uplink or downlink to perform channel training in TDD systems.
- Beamformed interference to users at other cells caused by pilot contamination effect may be avoided by right selection of transmission path at each time slot.
- Strong cross-link interference (in particular, BS-to-BS) generated by flexible TDD deployments must be prevented as LOS channels between BSs would cause interference to be highly damaging because of the received interference power.

Hence, following the above observations, the following classification can be established when PCR (Pilot Contamination Regime) is in place, i.e., when pilot contamination effects happens due to the TDD configuration of the network:

- According to the reception of beamformed interference
  - RCR (Reduced Contamination Regime): node not listening when beamformed interference is present
  - ICR (Increased Contamination Regime): node listening when beamformed interference is present

- According to the contaminating pilot
  - PCR-Downlink (PCR-D): Massive MIMO pilots contaminated by interfering DL training sequence
  - PCR-Uplink (PCR-U): Massive MIMO pilots contaminated by interfering UL training sequence

The problem of possible beamformed interference adds to the problem of cross-link interference, hence making a joint solution essential. The Figure 4-9 shows how the above observations can be implemented in simple TDD design recommendations. A detailed explanation of these observations can be found in the Annex A.4.

	Cell (Figure 4 8)	RCR config (interf avoidance)		ICR config (beamformed intf)	
PCR-D (DL training)	S	D	U	D	U
	I	D	U	U	D
PCR-U (UL training)	S	D	U	D	U
	I	U	D	D	U

Figure 4-9 Interference-based TDD configurations classification

RAN Implications

This scheme shows that interference can be avoided in a dynamic TDD system with at least one massive MIMO base station 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 so that ICR configurations are avoided and CLI is mitigated. 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 frame so that the number of ICR configurations at each time instant is minimized, despite the number of slots being 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. Therefore, as RAN implication it should be mentioned that specifications' support for dynamic selection of UL or DL pilot transmission for measurement purposes should be enabled. This is in addition to dynamic selection of data slots, a way how the harmful effect of beamformed interference caused by pilot contamination is avoided.

Results and Analyses

The resulting UL and DL rate distributions are shown in Figure 4-10, where the rates are expressed as spectral efficiency quantities measured in bps/Hz. DL rates represent the rates of the interfered users (i.e., all small cell users) while UL rates are measured at the macro cell since no pilot contamination effect appears in single-antenna base stations. We compare TD-

LTE with TDFLEX, our proposed scheme explained in detail in Annex A.7. Several interesting observations can be made. First, managing the interference caused by the pilot contamination effect in a HetNet by means of the TDD configuration makes a very positive impact on the attainable user rates: By selecting the appropriate receiver when beamformed interference is present, the user rates can fully benefit from the advantages of massive MIMO systems. This advantage can be easily missed if the TDD configuration is not designed with this objective in mind, as in the case of dynamic TD-LTE. The effect, although troublesome for both DL and UL communications, is particularly bad for small cell users who see their rates very limited by the beamformed interference coming from the eNB.

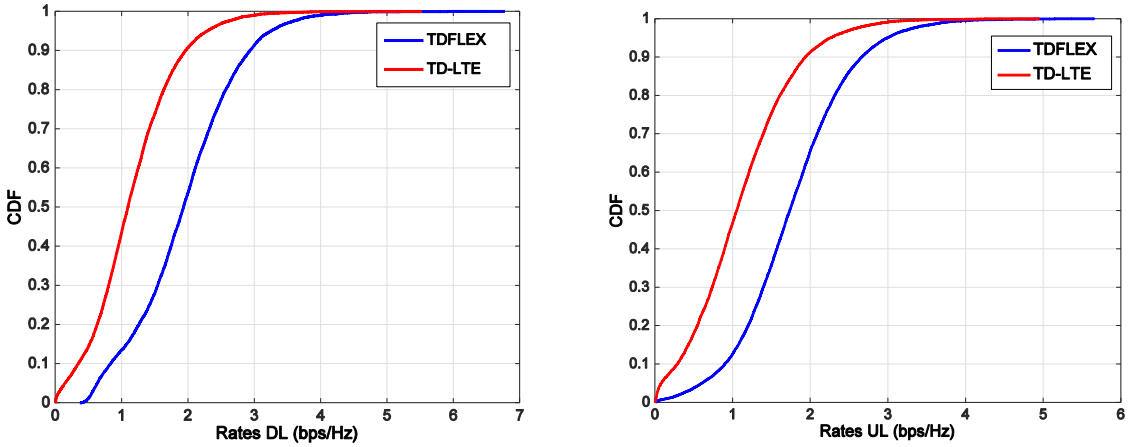

Figure 4-10 Rate distributions

DR-5: Enable Interference Avoidance in high-interference scenarios

Rationale

In mMTC usage scenarios, that will enable a plethora of new IoT services, one of the most relevant KPIs is the percentage of area coverage, while others such as latency and throughput are less relevant for the service provision.

Currently the LTE strategies for interference mitigation are based on:

- Static interference management, based on static partitioning and assignment of radio resources.
- Control plane interference management, i.e. exchanging information on radio resource utilization between BSs, providing more flexible and variable assignments.
- User plane interference management, in which actual UE data plane information is exchanged between the different access points, as well as the appropriate control plane.

However, static and control plane interference management are both suboptimal solutions, based on static radio resource partitioning, and focused on avoiding interference from one single interferer BS. Data plane interference management requires very low latency and high

backhaul throughput, as well as additional processing at the BSs and thus, is significantly increasing the costs associated to network deployment.

Therefore, 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.

The TeC aims to configure a procedure for orthogonalizing neighbor BSs transmissions, by means of control plane based interference management information (spreading & scrambling codes) between base stations grouped in a base station cluster.

In order to keep frequency band usage limited, the approach will be 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.

RAN Implications

From RAN architecture and logical entities implications, this design recommendation will require that 5G system includes mechanisms for:

- **BS clustering**. The clustering procedure must focus on selecting the members of BS cluster so that the number of UEs in high interference situations being served by one single cluster is maximized, i.e. aiming the less possible number of high interfered UEs are in the border between two clusters.
- **BS cluster coordinator,** will lead the clustering procedure, initiating the code patterns generation to be used by all BSs in the cluster. An X2 interface between neighbor BSs is needed for the pattern usage communication.
- **Signaling to UEs,** with indication of the spreading and scrambling codes used by its serving BS, in order to enable the despreading and descrambling procedures.

The logical entities involved in the deployment of this RAN design recommendation are:

- **UE interference measurement**, which performs the measurement of the level of interference present at the UEs, in order to evaluate base station clustering.
- **UE control plane signaling**, which includes the base station activation/deactivation of the spreading procedure in the UE control plane signaling, as well as the spreading and scrambling codes associated.
- **Precoding manager**, which is in charge of communicating the spreading and scrambling codes that will be used over the complex symbols transmitted by the different base stations inside the cluster to the UEs.

Results and Analyses

Link-level simulations have been conducted for evaluation of this design recommendation, with three scenarios taken into account: BSs within the same coordination cluster, BSs from different

clusters, and uncoordinated BSs not following the proposed precoding scheme. The assumptions taken into account in these simulations are:

- **Interference from a BS within the same coordination cluster** is modelled by means of a fully constructed OFDM signal carrying random payload bits. The signal to interference ratio (S/I) expresses the average ratio between the received signal power and the interference power coming from neighbor BSs within the cluster. Both the desired and interfering signals pass through multi-tap Rayleigh channels characterized by independent power delay profiles. Bit and block error rates versus SNR have been obtained for several S/I operating points.
- **Interference from a BS belonging to a different cluster** has been modelled by means of a fully constructed OFDM signal carrying random payload bits. The beneficial effect of the scrambling pattern has been analyzed with an S/I ratio defined between the received signal power and the interference power coming from a BS within another cluster. A different Rayleigh channel is also used for this interference, and bit and block error rates obtained for several S/I points.
- **Interference from an uncoordinated BS** has been modeled by means of a standard OFDM signal carrying random information.

Since the interference coming from within a cluster has similar effects for all the BSs belonging to it, it is reasonable to simplify the evaluations and group together the interference from multiple BSs into a single interfering link.

The annex A.5 includes the summary of the throughput versus SNR and SINR, for different interference scenarios, modulation and coding and number of BSs in the cluster. In each simulation the SNR or SINR (simulating interference as a decorrelated OFDM signal) values are compared with the baseline scenario (i.e. without this design recommendation implemented), therefore clearly showing the regions for which there is an advantage in its usage, the lowermost part of the SNR and SINR values.

One of the major advantages of this design recommendation is foreseen, when the modulation and coding schemes (MCS) are already in the lowest values, such as QPSK1/3, and an increase in coverage is needed to provide service to low throughput nodes (IoT scenarios). Results for a cluster of 4 BSs (offering around 6 dB gain taking into account the lower value at which a few kbps connection is feasible) as well as for a cluster of 8 BSs (further improving the gain 3 dB in addition), in which the coverage is significantly increased, are shown in Figure 4-27 where the 3GPP Extended Pedestrian channel model A at 3km/h (EPA3) was assumed.

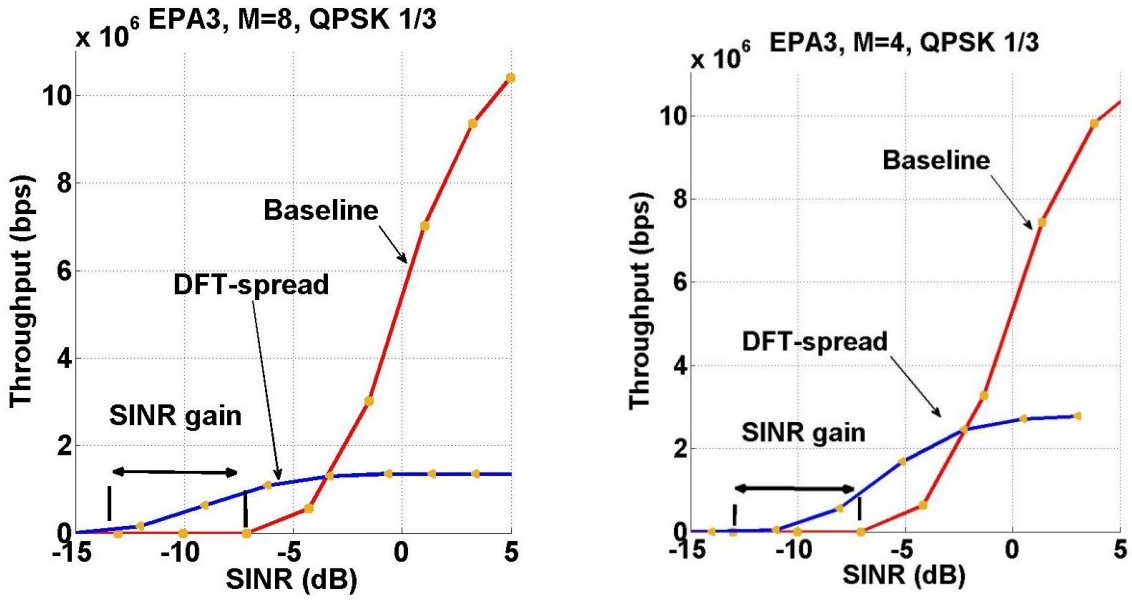

Figure 4-11 Throughput comparison between DFT-spread and baseline techniques for EPA3, QPSK 1/3, M = 4 & M=8 (M: Cluster size), as functions of SINR

4.2 Flexible Short-term Spectrum Usage

In this BB, novel control mechanisms are discussed and evaluated by examining new solutions of fast spectrum usage, such as the introduction of a virtualization layer that achieves a fast and seamless fragmented spectrum usage, including low and high unlicensed frequency bands. Adaptations of the control mechanisms are necessary to support efficient and on-demand resource allocation of unlicensed spectrum, by taking into account the dynamicity of the RAN environment (e.g. dynamic radio topologies), as well.

4.2.1 BB Description

Functional Decomposition

The internal functionalities of the BB, i.e. LE representation, are provided in Figure 4-12. The measurements are performed by the NNs and the UEs in both licensed and unlicensed carriers and sent to the BS via measurement signaling. From the set of NNs and the 'remaining KPIs' (e.g. remaining capacity per UE or achieved reliability), we select which NNs to act as Licensed Assisted Access (LAA) [3GPP15-36889] node and to which user they can be associated. Here, remaining KPI can be defined as the performance gap between the achieved KPI and the target. This can take the form of additional capacity requirement (by utilizing additional bandwidth) or reliability (by scheduling additional redundant links). Other inputs are the RRC indication and the WiFi AP location awareness which can affect the decision for the operation in unlicensed bands.

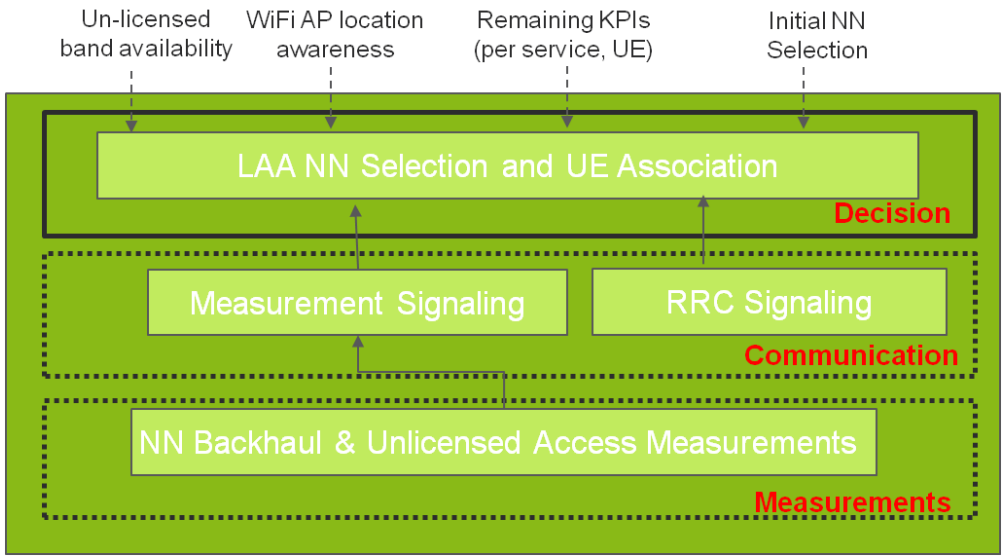

Figure 4-12 Flexible Short-term Spectrum Usage functional decomposition

Deployment Architecture

Figure 4-13 presents the functional logic: NNs perform the required measurements on the wireless BH link and sends this set of information to the serving BS. The selection of the NN as the serving NN is preferably done by the BS. Therefore, part of functionality is deployed at the serving BS and part of the functionality is deployed at the NNs. Furthermore, the UE performs measurements on the unlicensed carrier and sends them to the NN. NN can forward these measurements to the BS which will then select the association of UE to NN.

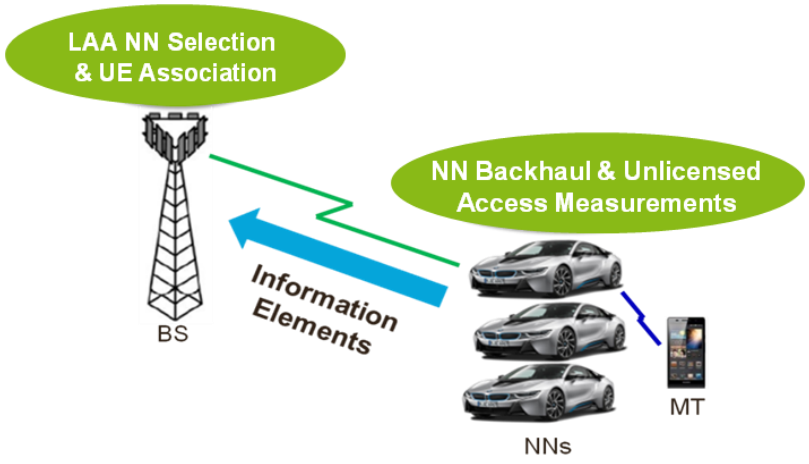

Figure 4-13 Deployment of functionalities in case of dynamic NN selection

4.2.2 Design Recommendation

DR-6: Support LAA for unplanned and Dynamic Radio Topologies (e.g., NNs)

Rationale

In this study, we examine the coordinated scheduling of the RAN resources utilizing unlicensed spectrum. To do this, we need to identify which RAN nodes are utilizing shared unlicensed bands and coordinate them in time and frequency domain so as to minimize interference and maximize the gain. Otherwise, the uncoordinated use of these resources will create severe interference and limit the gains of the extra spectrum assignment. Here, another key parameter is the presence of multiple service types, which may have different KPIs. In other words, the assignment of extra resources is prioritized in a per-service manner so as to ensure that their requirements are met. In particular, the unlicensed spectrum in 5GHz bands via dynamic radio topologies (e.g. Nomadic Nodes, cellular-assisted D2D) is a case study of key importance to be investigated. Multi-connectivity would be possible using both licensed and un-licensed links with supported on demand APs.

In this study, the activation of NNs with capability to operate in unlicensed bands is assumed in a hotspot scenario, as a mean to boost performance due to the interference limitations in licensed carrier. A major challenge is inter-NN interference, which might significantly degrade the users' performance. NNs initially request measurements from users who fail to meet their KPIs in licensed bands (after performing interference management as discussed in [CSCN16]).

A key challenge in LAA is the effect of Listen Before Talk (LBT) to the actual performance, since the dense activation of LAA nodes might lead to severe collisions if there is no centralized LAA coordination. The users provide measurements for the unlicensed channels (per carrier) using the licensed bands, also providing the RSRPs of dominant interferers. For these users we can either perform uncoordinated resource allocation or Interference Management, which can be adapted based on the neighbourhood of activated NNs.

Two different xMBB case studies are shown where we have different remaining KPIs given, the resource situation (e.g. inter-node interference and resource availability) and user allocation in licensed bands. For the first case where the remaining KPI is throughput, we perform Coordinated Scheduling/ JT as discussed in *DR-7,* taking into account a pre-defined energy detection threshold for WiFi activated access points. The second case is when some users cannot achieve the target reliability in licensed bands (according to the service requirements) and need to enhance it by utilizing in parallel LAA. For the latter, together with the NNs in licensed carriers, a set of LAA NNs provide multi-connectivity by parallel redundant links to the user at a reserved (but low) number of RBs, which can be changed subject to the number of users demanding unlicensed bandwidth for reliability improvement.

Further details are provided in Annex A.6.

RAN Implications

Mainly interfaces between NNs and macro cells will be required. For the effective and efficient utilization of the unlicensed spectrum, AIVs should allow interference estimation also on the unlicensed bands.

Results and Analyses

Here, a particular deployment of 5 NNs is investigated (as can be seen in Annex A.6).The NNs operate mainly in licensed carriers, but they can also operate in unlicensed bands (5GHz band) subject to the access conditions and the availability of WiFi access points. For the channel models in the simulation scenarios, 3GPP-compliant parameters were used (outdoor LAA scenario 4, in [3GPP15-36889]).

For the first case (throughput as KPI), as illustrated in Figure 4-14 (left), the more NNs appear in unlicensed band without coordination, the gains are limited. On the other hand, if coordination is performed (interference coordination by muting at cell edges or multi-connectivity) we can observe huge gains in terms of DL throughput. Here to mention that this figure mainly shows the trend of capacity gains when no interference from WiFi is assumed, within the given scheduling period. Here to mention, that since one of the limitations in LAA is the foreseeable delay due to HARQ, we assume that the re-transmissions are handled by the primary cell (e.g. macro or NN) in licensed carrier.

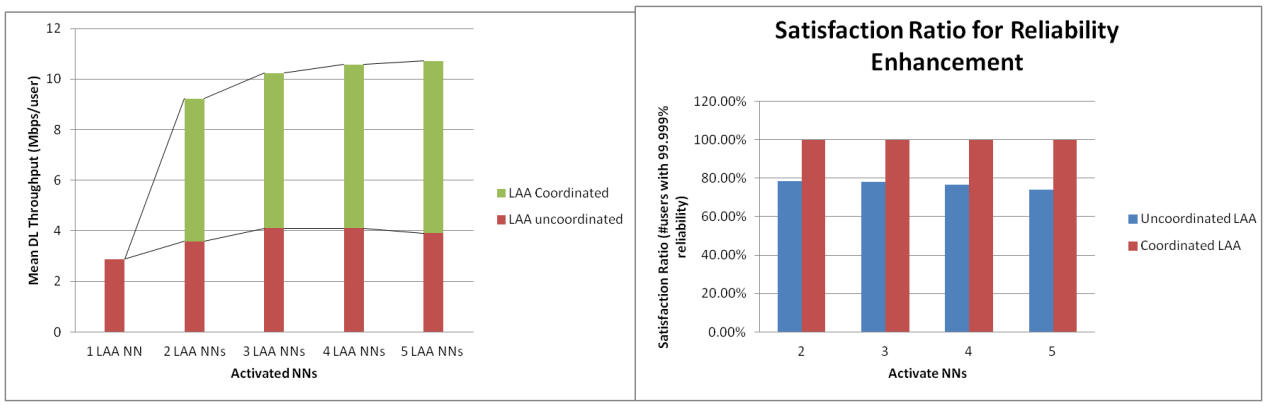

Figure 4-14 Mean User Throughput for different LAA NN Activations (left) and Satisfaction Ratio (right)

In Figure 4-14 (right), we also provide the satisfaction ratio (the proportional number of users which achieve 99.999% reliability) for the second case study. Satisfaction in terms of reliability can be measured at the 10^{-5} percentile of CDF of SINR where the threshold of 0dB is assumed as exemplary target reliability (according to [PSL+15]). As can be seen from Figure 4-14, when coordinating the LAA operation in a reliability-driven fashion, so to improve the worst case SINR (by allowing multi-connectivity with limited number of reserved RBs), the reliability requirement is met and the proportion of users with satisfied KPI increases from 80% to 100%. On the other hand, without coordinated operation, the satisfaction ratio decreases as the number of NNs increases, due to Inter-NN interference.

4.3 RAN Moderation

The main aim of this BB is to enable the energy-efficient moderation of RAN nodes, i.e. the optimal active-mode operation, with the help of quality of service and channel quality aware operation of the 5G network. The energy saving gains of the moderated network is obtained due to the unique lean system design of 5G which enables the gNBs to be in sleep mode when there is no traffic to be served. Here we consider the RAN network to also include self-backhauling (sBH) nodes where access and backhaul links are provided using the same 5G radio, possibly using the same frequency band for operation. In this section, we will provide an overview of the BB and the possible 5G RAN design recommendation which impacts the system design.

4.3.1 BB Description

Functional Decomposition

The functional decomposition of the BB shown in Figure 4-15 provides an overview of the operational paradigms involved in the moderated 5G RAN network. For the efficient operation of the BB, the assumption is that the radio resource control information, context information, real-time traffic information (including UL and DL buffer status information), and QoS information are available as inputs. The channel quality and traffic measurement information are used as possible inputs to determine the optimal activation of the nodes in the 5G-RAN using an energy-aware backhaul-access point (BH-AP) controller. The BH-AP controller configures the APs using a switch-on/off command and the sBH nodes using DRX configurations.

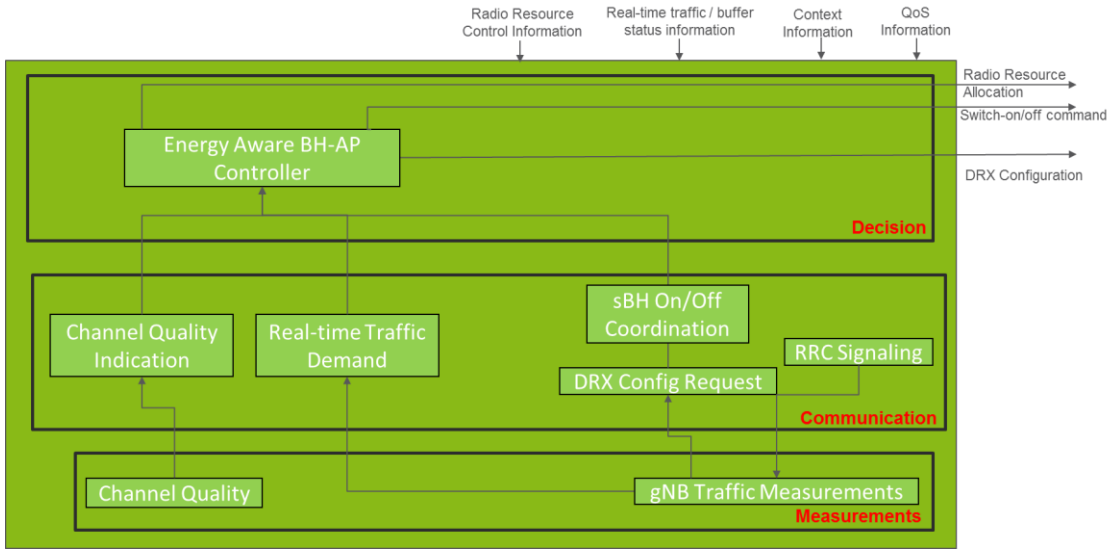

Figure 4-15 Functional decomposition diagram for RAN moderation

Deployment Architecture

The possible deployment architecture of the BB is as shown in Figure 4-16. It is envisioned that the control functions that enable this BB are implemented in the AN-O layer of the network possibly as virtual functions, which operate with assistance from information elements from the AN-I layer. The information elements transport various measurement information, in order to enable the energy aware BH-AP controller to make RAN moderation decisions. External LEs such as O&M, network configuration policies, QoS information, etc., are assumed to assist the RAN moderation function.

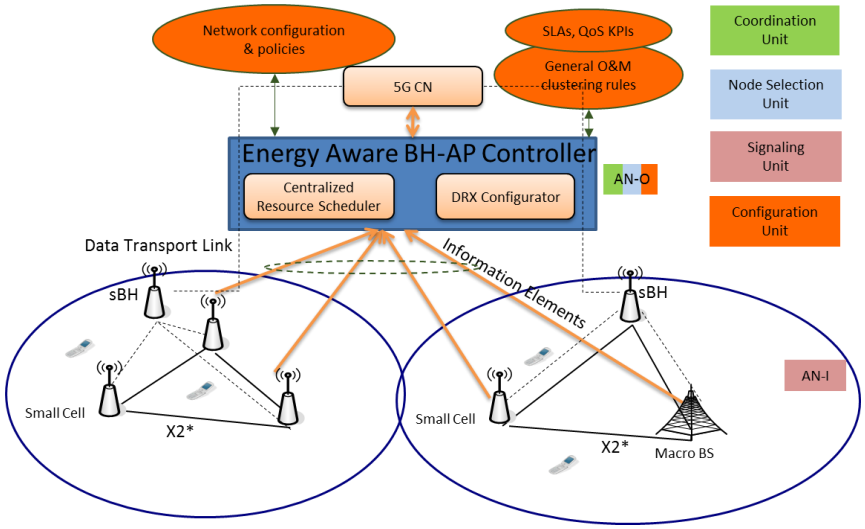

Figure 4-16 Possible deployment architecture for RAN moderation

Here, the dashed lines to the sBH node indicates the possible Un* interface between the BS and sBH aggregation node, which could be used to transport data between the 5G-RAN and CN. The interface is optional, depending on whether the deployment architecture is using sBH or using direct BH links to transport the data to the 5G CN.

4.3.2 Design Recommendation

DR-7: Centralized resource scheduling or distributed coordination for maximizing energy efficiency

Rationale

While the building block has a broader scope in terms of enabling mechanisms for RAN moderation, the main focus of this design recommendation is the coordination of sBH nodes. The main rationale behind this design recommendation is to enhance the energy efficiency for UDNs leveraging lean carrier design, based on the concept presented in [PUM17]. The coordination is achieved when 5G-gNB informs the backhaul aggregation node about the intended sleep decision and about the time periods where it would be in sleep mode. The backhaul aggregation node is assumed to be the traffic aggregation point which could control

the active / sleep mode operation of the backhaul links. This would be required for the backhaul node to adjust the scheduling of downlink data from the CN accordingly. If the backhaul node provides a dedicated backhaul link for the gNB with one sBH link provisioned per gNB, the backhaul node enters sleep mode in a synchronized manner with the gNB. This would be essential in the fast wake up of the gNB and avoiding further re-synchronization procedure from the gNB, which can significantly reduce the amount of wake-up delay. Here the key idea is that since the 5G-gNB – BH link is essentially a stable link (both gNB and BH node are static), achieving synchronicity through this technique would avoid any unnecessary delays. Potential enhancements could be applied to moving gNBs in 5G (using different UE categories which require achieving synchronization again with the backhaul node). Further details are provided in Annex A.8.

A simple diagram illustrating this operation is shown in Figure 4-17 [PUM17]. Here a new message needs to be defined for coordinating the sleep cycles between the gNBs. Here we assume that a fixed set of configurations are available for the gNB sleep modes, where the DTX subframes are indicated to the gNB / sBH node using a new signaling message over RRC. Such a configuration can enable a sleep mode configuration negotiation between the 5G-gNB and sBH node (which appears as a logical UE node). In a general case, where the BH node is a proprietary node, then this information exchange occurs over the proprietary interface.

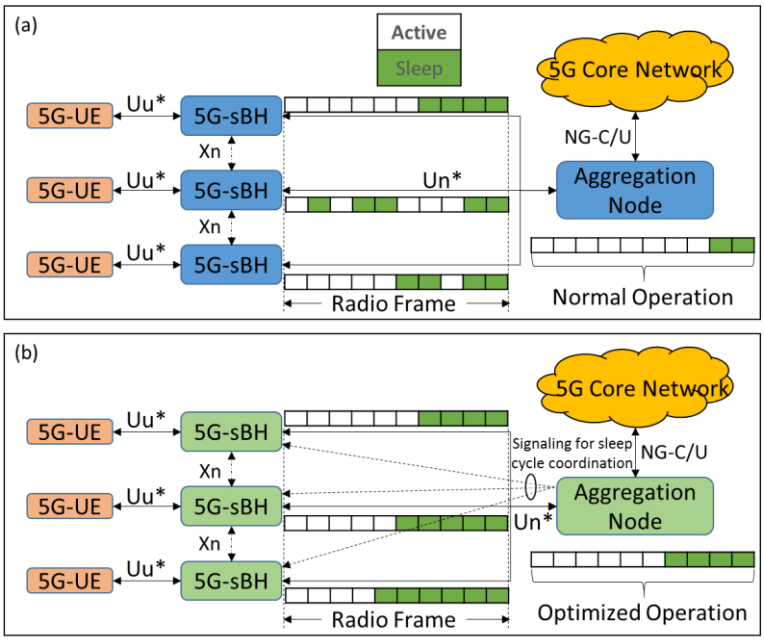

Figure 4-17 Operation diagram of coordinated RAN-BH sleep [PUM17]

RAN Implications

Additional RRC signaling for coordinating the sBH node operation: Since the sBH node appears as a UE from the aggregation node perspective, there needs to be additional RRC signaling for discontinuous reception configuration so that the sBH node along with the aggregation node

enters energy efficient / sleep mode during the time instances (subframes) where no data is scheduled.

Channel measurement coordination for Energy Aware BH-AP Controller: New information elements to be defined between AN-O and AN-I, in order to enable channel measurement coordination for the BH-AP controller. The information elements are proposed to be transported through the open interface between the AN-O and AN-I.

Results and Analyses

Here, we do a conceptual evaluation of the proposed method, in a network consisting of 1000 small cells / 5G-gNBs with both dedicated sBH link and aggregated BH node. Here we assume fully loaded cells, with possibly full-buffer traffic type, in order to evaluate the maximum potential energy saving gains that could be obtained by the proposed mechanism. For the non-sBH link, we assume fiber access BH using GPON technology, with active and sleep mode power consumption values similar to the ones used in [FMM+11]. For the legacy LTE technology, we assume the presence of fiber access links without sleep modes with 125 aggregation nodes (one aggregation node per 8 backhaul links), each consuming 100 W in active mode. For the 5G-gNB with dedicated BH link, we assume similar BH architecture, with the GPON sleep mode power consumption and 50 W per aggregation node in sleep mode. 5G-gNB with sBH link has similar power consumption assumptions as the gNB, with different sleep modes considered in power model presented in [DDL15]. Thus, the energy saving gains based on GPON are applicable to LTE and 5G with dedicated fiber access links. 5G-Dedicated BH has the additional advantage of sleep modes based on [DDL15].

The power saving potentials are as shown in Figure 4-18, with significant power saving possibilities using the proposed method in 5G, as compared to legacy LTE. Here, the gains come from the synchronized sleep mode between the gNB and BH link, along with the possibility for the aggregation node to enter coordinated sleep modes as well for the sBH case. Here sleep modes 1 - 4 indicate the level of power savings that could be achieved at the BS, depending on the available sleep mode duration (of 0.071 – 1000 ms), when the BS is inactive with no traffic scheduled. The power savings are achieved due to the different components within the BS being inactive during the sleep mode.

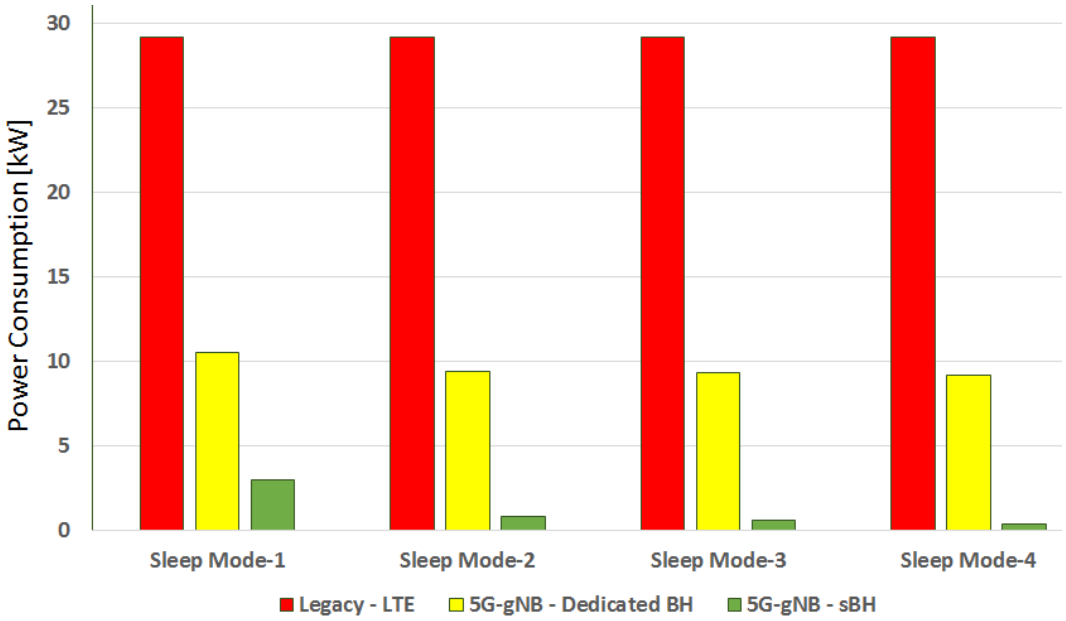

Figure 4-18 Various power saving potential with different backhaul types

In a parallel work detailed in Annex A.9, non-coherent JT/JR with dynamic TDD was considered for the virtual indoor office scenario with the objective of maximizing either throughput or energy efficiency. In short, results indicate minuscule throughput gains at even medium traffic loads, and therefore little to no gains for the energy efficiency compared to the distributed system where UEs only associate to their strongest BS [CS15]. From this we infer that in the case of dynamic TDD, power gain alone is not enough to improve the energy efficiency without also managing the interference by using more sophisticated schemes.

On the other hand, considering an FDD scenario, higher energy efficiency can be reached when traffic is below its peak, if one designs scheduling algorithms that exploit coordination approaches like JT, so that the number of cells that should be activated to satisfy the existing traffic can be reduced. This approach has been studied in the past for 4G networks, and has been further investigated in Annex A.7 considering enhanced power models that will be available with 5G BSs.

4.4 Multi-Slice and Multi-Service Holistic RM

Full support of network slicing is one of the key innovative features of upcoming 5G standards. To fulfill multi-dimensional KPIs for different service types, such as xMBB, mMTC, uMTC across AIVs the RAN needs to support optimized short term resource allocation within a single AIV, but also introducing common control functionality for fast data flow handling across AIVs, as well as long term common control enablers.,

Network slicing enables end to end service chain optimization for different services. While the CN optimizes the placement of virtualized network functions, the RAN needs to handle slice specific configuration rules [5GNORMAD32] 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.

4.4.1 BB Description

Functional Decomposition

To decompose the functionalities to enable multi-slice and multi-service support by the RM framework, the necessary logical entities are grouped and mapped to the AIV agnostic and the AIV specific layers of the RM framework. Figure 4-19 shows an overview of the functional decomposition. The AaSE consists of three LEGs: an AIV overarching monitoring entity, a logical entity to control and dimension RAN slicing as well as an LE 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. In addition an LEG for multi-AIV coordination is defined which enables virtual cell deployments in multi-AIV scenarios as well as a long term AIV configurator. Further information on the long term AIV configurator is given in the Annex A.15 and in [MET-II17-D62]. Both AN-O LEGs are interacting with the BB Dynamic Traffic Steering (DTS) which further processes data flows to the AN-I specific functionalities. For fast short term RRM an AIV specific LEG is defined which supports multi service scheduling and multiplexing as well as D2D group scheduling. D2D group scheduling needs also support in the 5G UE. The BB needs tight interworking with the BB Context Management to make use of necessary extended measurements.

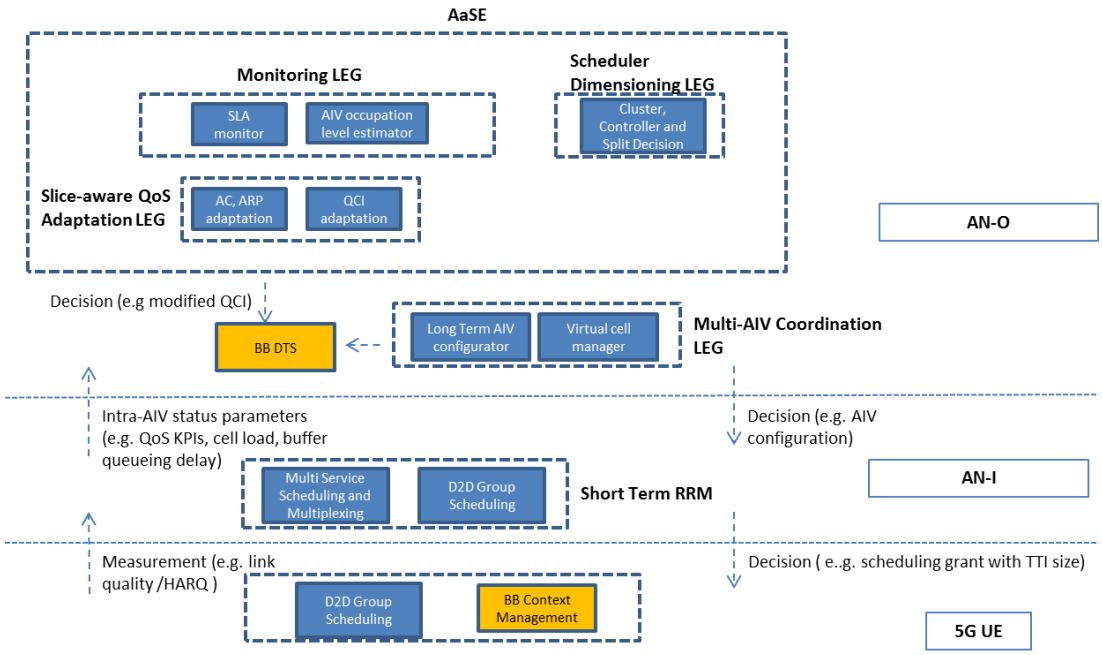

Figure 4-19 Functional Decomposition for Multi-slice and Multi-service Holistic RM

Deployment Architecture

Since the LEs of the considered BB are distributed across the whole framework, as depicted in Figure 4-19, an example for a possible architecture can be found in Section 3.2 considering Figure 3-3.

4.4.2 Design Recommendations

DR-8: Enable AIV-Agnostic Network Slicing Support by using SLA based QoS adaptation and slice-adaptive RRM placement

Rationale

Network slicing is a new 5G concept in which multiple logical networks run as virtually independent business operations on a common physical infrastructure. In contrast to the operation of dedicated physical networks, sharing of available resources on a dynamic basis allows exploitation of pooling gains. With respect to the RAN, an efficient sharing of scarce radio resources among the network slices is the key challenge, which is achieved by RRM for network slicing. It is realized by introducing a logical entity in the RAN, AaSE, that is responsible for monitoring and enforcing SLA 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 to all corresponding data flows that are part of it a certain QoS class. Using ARP, the importance of individual data streams can be configured. It is then a task of the LEGs multi-AIV resource mapping, interference management, and real-time resource mapping to realize the corresponding QoS (for more details on the interworking on the LEGs see Section 3.1). More details on the proposed solution can be found in [MET-II16-D51] and in Annex A.10.

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). More analysis on this technology candidate can be found at Annex A.11 as well as [PP17].

RAN Implications

1. A new logical RAN entity, the AaSE, is proposed. It performs SLA monitoring, Slice-aware QoS Adaption, formation of slice-aware BS clusters for RRM purposes, and an adaptive placement of intra-slice RRM functionalities to the RAN nodes of these clusters.
2. AaSE has to interact with the CN and the orchestrator as described in Section 3.2.

Results and Analyses

The simulation results in Figure 4-20 show a comparison of two RANs (subnetworks) with best effort traffic in terms of user throughput. In the first case (red curves), two dedicated networks

with 10 MHz system bandwidth each are operated for independent businesses. The dedicated network 1 serves hundred users with a low demand, such that the network is low loaded. In contrast, the dedicated network 2 serves 710 users causing a fully loaded system with lower performance per user. In the second case (blue curves), a common RAN for both networks is operated on 20 MHz system bandwidth. The detailed simulation assumptions can be found in Annex A.10. The pooling of resources enables a gain in user throughput as can be depicted from Figure 4-20 showing that the probability for users in both slices to miss a certain throughput figure is always well below that of users in both networks (solid curves). By means of an SLA, it is targeted that users of the virtual network 1 (network slice 1) reach a similar capacity as in the case of dedicated networks. As the dedicated network 1 reached a mean network throughput (averaged over time) of 218 Mbps, an SLA was used to a guaranteed network capacity of 220 Mbps. Network slice 1 achieves a network throughput of 209 Mbps. This is slightly below the guaranteed capacity due to variations in the traffic pattern that cause a demand of less than 220 Mbps at some time instances.

The simulation results show that network slicing can achieve performance gains due to pooling of resources while protecting the performance of individual network slices.

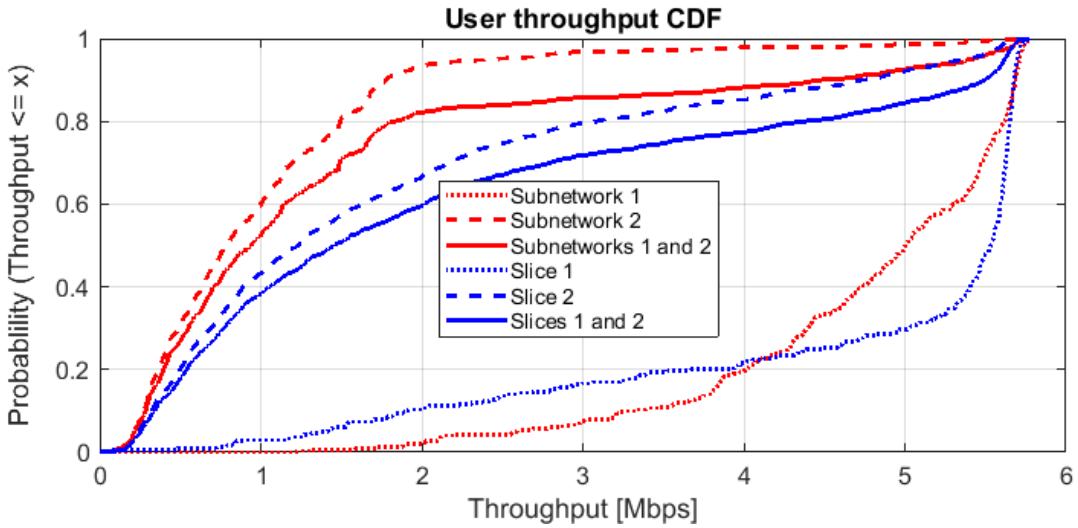

Figure 4-20 Simulation results of RRM for network slicing

One additional evaluation study is shown for the scheduler dimensioning and placement of RRM functionalities. For different slices we may have different requirements for spectral efficiency and different RRM centralization requirement. For the example shown in a practical scenario (see Annex A.11), for uMTC (termed also as URLLC) more than 1bps/Hz is an acceptable level, while for eMBB more than 2.5bps/Hz spectral efficiency 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 multi-connectivity at cell edges). The per-AP Spectral Efficiency for this particular simulation setup can be seen in Figure 4-21.

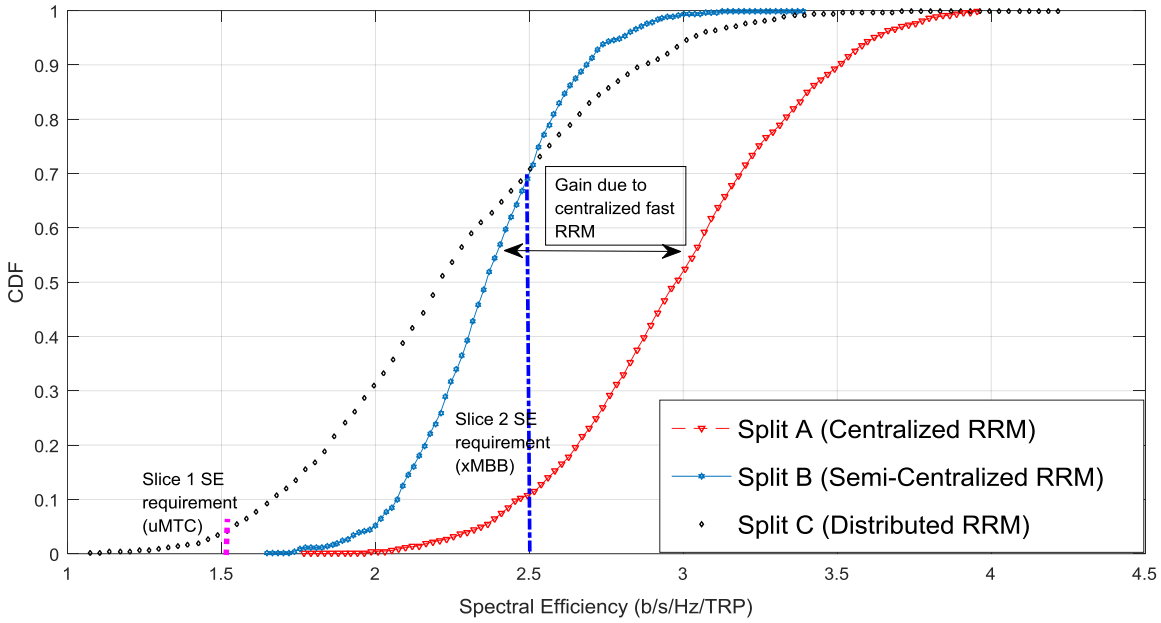

Figure 4-21 CDF of Spectral Efficiency – Comparison of different splits

As we can observe from the CDF of spectral efficiency, 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.

DR-9: Enable multi-service resource allocation in a flexible manner

Rationale

The AN-O layer decides on the most appropriate AIV and configuration to serve a specific data flow in a particular instance, whereas it is the responsibility of the AN-I layer and the short term RRM to assign specific radio resources and multiplex different data flows within a single AIV. These data flows can come from the same or different slices and can potentially have very different performance requirements. It is therefore desirable to be able to accommodate them in a flexible manner that dynamically adapts to the user demands, rather than requiring a strict or more static separation and reservation of resources for different services.

It is well known from the existing literature that there are fundamental tradeoffs between scheduling users to maximize their spectral efficiency, coverage, latency or reliability [SMP+14]. This calls for flexible scheduler functionality that allows scheduling each user in accordance with its desired optimization target. One option for this is to design the 5G system to support scheduling with different TTI sizes [PBF+16], which is the foundation of the scheme proposed in the following, based on a flexible frame structure that dynamically allows variable TTI size configuration per user and per scheduling instance.

In this way, for instance, the following scheduling decisions are possible: use short TTI for uMTC users to optimize their latency, at the expense of increased control overhead and lower channel coding gains; eMBB users can be scheduled with longer TTIs and wider frequency allocations to cope with the high data rate demands; and mMTC users can benefit from narrow bandwidth allocation and long TTIs, which are attractive characteristics from cost and coverage perspectives. In addition, the possibility to set the TTI size per scheduling instance enables optimizing eMBB services that make use of Transmission Control Protocol (TCP). Short TTI duration can be used in the first transmissions to reduce the round-trip time of the flow control mechanism in the slow start phase of TCP, and later longer TTIs can be configured to maximize the spectral efficiency when steady operation is reached.

RAN Implications

The flexible multi-service scheduling framework requires that the PHY/MAC layers and scheduling functionality support the following features:

- Dynamic allocation of time (i.e., TTI size) and frequency (contiguous/non-contiguous) resources per user and per scheduling instance, including options for asymmetric link operation (i.e. different TTI sizes in DL/UL)
- Flexible and highly scalable scheduling grant design for efficient radio resource usage (i.e. possibility to multiplex a highly variable number of users per carrier, ranging from one to a potentially large number, avoiding control channel blocking problems)
- Dynamic resource sharing between xMBB and uMTC services, without “hard resource partition”, for maximum resource efficiency and achievable throughput while guaranteeing uMTC requirements (e.g. via derived schemes from this framework that allow scheduling an uMTC transmission on resources being used by an ongoing xMBB transmission)

Results and Analyses

System-level performance evaluations have been carried out to compare the performance of several TTI size configurations, in order to estimate the most suitable TTI size that should be dynamically chosen per UE depending on service requirements, traffic type, radio channel quality, and system load. The evaluation is performed in a 3GPP Urban Macro scenario with 7 BSs, each having 3 sectors, 500 m inter-site distance (ISD) and using 10 MHz band [3GPP10-36814]. In-resource control channel (CCH) scheduling grants with link adaptation are assumed, which allows to model different degrees of CCH overhead (i.e. aggregation levels or number of resource elements dedicated to CCH) depending on the UE radio conditions [PBF+16] [PNS+16].

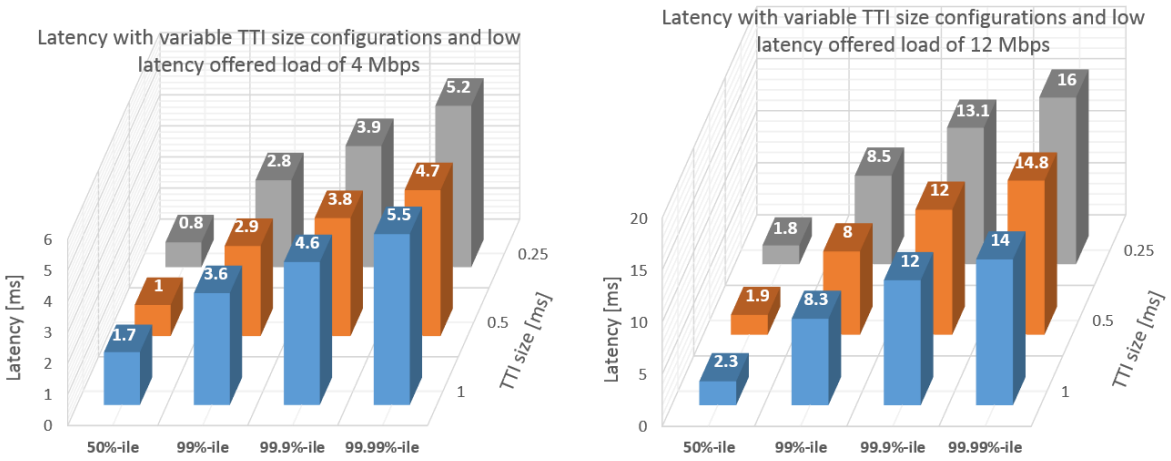

Figure 4-22 Latency values from packet latency CDF with variable TTI configurations and offered loads for a mix of xMBB and uMTC traffic.

The packet latency (i.e. MAC layer one-way user plane latency) achieved with different TTI sizes and system loads with a mix of xMBB and uMTC services is shown in Figure 4-22. The xMBB traffic is modelled with a single user full buffer download, whereas higher priority uMTC low-latency traffic follows a Poisson arrival process with 1 kB payload and varying total cell offered load. More details can be found in [PPS+16]. As depicted in Figure 4-22, at low system loads, using a short TTI is in general a more attractive solution to achieve low latency communications. However, looking at the tail values (99.9%-ile and above), a 0.5 ms TTI size offers better latency than the 0.25 ms TTI, even for low loads.

As the load increases, longer TTI configurations with lower relative CCH overhead (and therefore higher data spectral efficiency) provide better performance, as these better cope with the non-negligible queuing delay. The 1 ms TTI configuration is beneficial from a latency point of view for high loads and above the 99.9%-ile, due to queuing delay. As the offered load increases or as we consider UEs with the worst channel conditions, the queuing delay becomes the most dominant component of the total latency, therefore it is beneficial to increase the spectral efficiency of the transmissions (by using a longer TTI) in order to reduce the experienced delay in the queue. The observed trends are relevant for uMTC use cases, which require latency guarantees of a few milliseconds with reliability levels up to 99.999%.

The results presented above and detailed in [PPS+16], as well as related studies performed in [PNS+16] and summarized in Annex A.12, indicate that the optimum TTI size varies depending on multiple factors. Therefore, it is beneficial to be able to dynamically adjust the TTI size per user's service requirements and scheduling instance, rather than operating the system with a fixed TTI. Ongoing and future studies will focus on analysing and optimizing the performance when multiple TTI sizes coexist in the same simulation, as well as on evaluating efficient schemes derived from this framework for dynamic sharing of xMBB and uMTC resources (e.g. puncturing ongoing xMBB transmissions by shorter and low-latency uMTC transmissions).

DR-10: Enable UL group transmissions using D2D groups

Rationale

The high frequency bands for 5G enable high bit rates due to the vast amount of available bandwidth, but are subject to high path loss, diffraction, and are prone to spotty coverage. The proposed group transmission concept uses D2D communication within a group of UEs to distribute the UL user data to the group members, which then in a coordinated fashion simultaneously transmit the UL data to the eNB. The group transmission based on D2D is appealing to use since even for a relatively small group, it can be expected that at least some of the UEs will have a good link to the eNB. Especially for the UL, the importance of enhancing the radio link is clear due to the low Tx power of the UE, compared to that of the eNB. This concept can be envisioned to be successful in several scenarios such as massive MTC, where devices can form groups and aid each other. This can be useful when some of the devices are deployed indoors, e.g., in basements with bad UL coverage while other devices have better coverage. Another scenario is that of emergency services in remote areas or where part of the infrastructure has been damaged, i.e. situations where coverage is bad but good UL coverage is vital.

RAN Implications

Distributing data within a group using D2D and jointly transmitting the data requires tight control of the D2D transmissions within the group and synchronization of the UL transmission and maintenance of the transmission buffers within the group. This requires reliable resource control for UE groups preferably via a coordinating UE.

Results and Analyses

The potential benefits of using coordinated UL transmissions follow from that being the received SINR at the BS of the group transmission the sum of the individual UEs' SINR. By applying Shannon's theorem in a coverage limited scenario, the maximum achievable bit rate grows linearly with the number of UEs in the group. If Shannon's theorem is applied in a capacity limited scenario, the maximum achievable bit rate grows at a slower rate, i.e. proportional to $\log_2 M$. This indicates that the group transmission concept has its main advantages in scenarios where coverage is the limiting factor. In this scheme, the gains in maximum achievable bit rate are assessed using a path loss analysis based on single cell setup with the Okumura-Hata channel model environment being used between the UEs and the BS and assuming that the D2D transmissions are ideal. Figure 4-23 shows the gain for maximum achievable bit rate when groups of 5 users randomly dropped in hotspots transmit as a group compared to the maximum achievable bit rate if they would transmit on their own. This is shown for hotspots of different radius at a distance of 2000m from the BS. With varying hotspot size, the differences between the users' path loss increase and the differences in gain for the UEs increase.

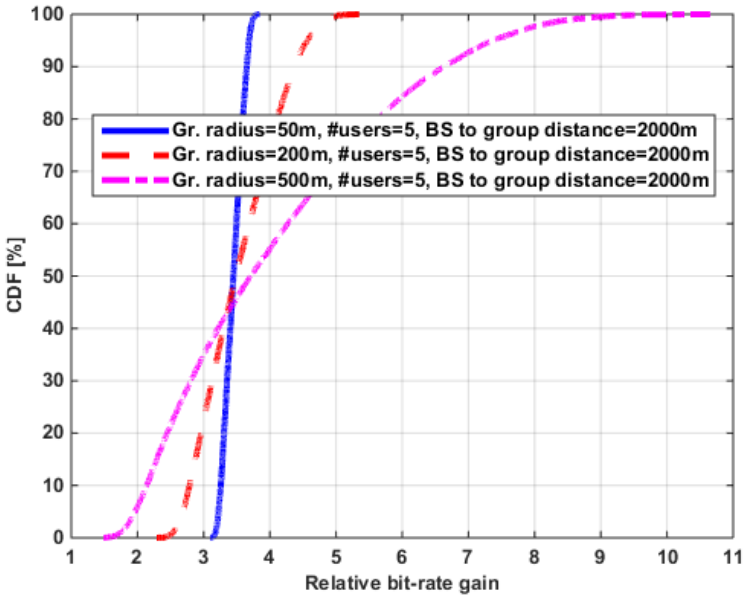

Figure 4-23 Relative bit rate gain for groups of 5 users.

As illustrated in Figure 4-23, gains in maximum achievable bit rate have a median of approximately 3.5, while for the larger hotspot, the 90th percentile gain is 6.5. This follows since when the radius grows, the probability that some of the other UEs in the group have a better link to the base station increases. Further analyses in the Annex A.13 show that the gain increases with the distance between hotspot and BS since hotspots far from the BS experience worse coverage. It should be noted that the evaluations assume ideal UE to UE communication. In reality, there would be some degree of efficiency loss and increased latency which is not accounted for here. Designing a group communication scheme would also require careful analysis of complexity versus gain. If not, there is a risk of designing either an inefficient scheme, or a scheme which is too complex to reach a wide deployment

DR-11: Support virtual cell deployments in multi-AIV scenarios

Rationale

A legacy user-centric cell system is essentially based on all BSs being equal peers wherein each BS has its own decision capability (i.e. a fully distributed and network-scalable system). In this virtual cell based scheme, to simplify inter-BS control, including conflict resolution among the equal BSs, a logical hierarchy is introduced among the user-centric BSs where one of them assumes the role of a master and other BSs in the virtual cell assume the role of a slave (see Annex A.14). In this master-slave cooperation protocol, the master is a layer 2 control anchor, i.e. it (a) generates all layer 2 control signaling, (b) performs all scheduling and radio resource control, for example, the UL multi-cell multi-user power control as described in [FAN17-D4.2], (c) manages addition/deletion of slave BSs, (d) functions as the point of contact to the network, and (e) resolves all conflicts. The slave BS's role is to cooperate in data transmission to serve

users based on the master's instructions. Despite the imposition of this logical hierarchy in the form of master and slave BSs, the user-centric concept is fully distributed in that any BS can become a master. A BS that is a master in one MS's user-centric cell is generally a slave in another MS's user-centric cell.

Two procedures (i) slave BS update and (ii) master BS change are central to the virtual cell reformation. In an MS's virtual cell, the master BS constantly updates the slave BSs by adding/deleting slave BSs so that the MS can always be served by BSs with good enough SINR. Furthermore, any BS should be ready to take on the role of a master BS without incurring a service interruption to "avoid" the handover experience. When a master BS decides to include a slave BS in its virtual cell, it provides/shares the MS's context over any available interface. This pre-sharing of user context at an early stage of joining is essential in eliminating the signaling messages that would otherwise be incurred at the last moment prior to handover. Figure 4-24 shows an example of such reformation procedure.

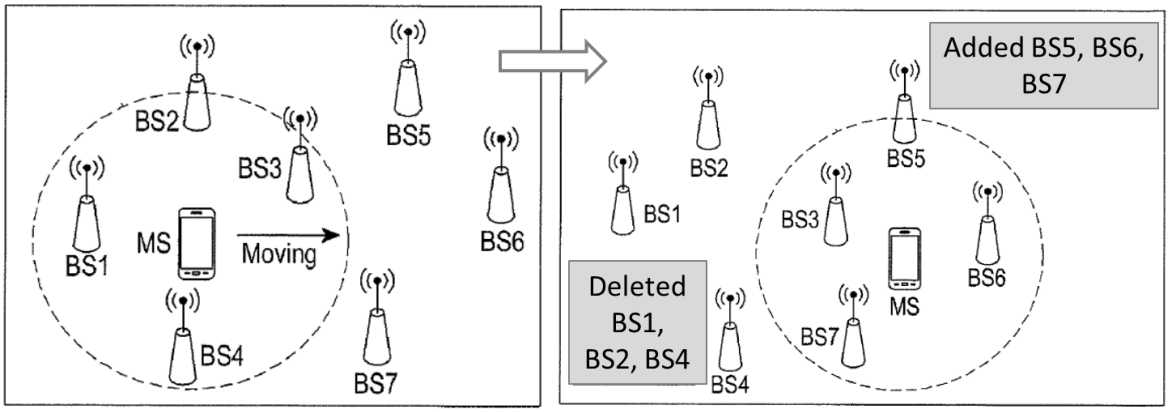

Figure 4-24 Illustration of dynamic virtual cell reformation.

For dynamic reformation of a 5G virtual cell where BSs may be serving using different AIVs or multiple AIVs, principles for 5G cell re-selection are proposed, as Annex A.14 explains. This enables the formation a "heterogeneous" virtual cell where the BSs are connected even to different cores. To form an effective 5G virtual cell out of different AIVs such as LTE and mmW 5G, the master node should naturally know the trade-off between the different AIVs.

RAN Implications

This scheme provides non-negligible impact on the RAN protocol design of current cellular systems as there is currently no abstraction model implemented to handle heterogeneous AIVs. However, the impact will be significantly lower once it is agreed the level of harmonization and related AIV-agnostic vs. AIV-specific split that will be required to manage different 5G AIVs. The level of impact will also depend on the coordination mechanism that is employed among the nodes: No significant impact should be incurred for techniques operating on the same layer, but other solutions may operate at different layers of aggregation for different AIVs and this should be taken into account. In addition, introducing the logical entity of the virtual cell functionality will

require some modifications at the network management level, both on a distributed architecture (D-RAN) and a centralized architecture (C-RAN). Regarding the interface needs, the exchange of messages between physical BSs can take place on the X2* interface, or can be facilitated by multi-connectivity (e.g. using a lower-frequency AIV). This applies for both forming and managing the virtual cell as well as for serving users. Some level of modifications may be required to exchange messages using a harmonized protocol stack, but these changes may need to happen for 5G independently of this TeC.

Results and Analyses

The benefits of virtual cells in terms of throughput is shown in Figure 4-25. SINR fluctuations are experienced by the MS as it moves through the cell centers and edges of the different BSs, hence causing variations on the corresponding TCP segment error rates, which differ significantly in the case of independent conventional cells and virtual cells. This is because in the typical conventional cell operation, the TCP window shrinks when the symbol error rate increases and causes TCP timeouts at the edge of the cell, but then recovers when the trajectory approaches the center of the cell. It is worth noting that even if the SINR values increase rapidly, the TCP window takes a while to get to its maximum size. This performance impact, however, is much lower in the case of user-centric virtual cells, keeping the TCP stable throughout the MS trajectory. The reason is that a virtual cell is reconfigured continuously to keep the MS at the center of it at all times, as opposed to letting the SINR (and hence TCP throughput) drop before performing a handover to a different conventional cell.

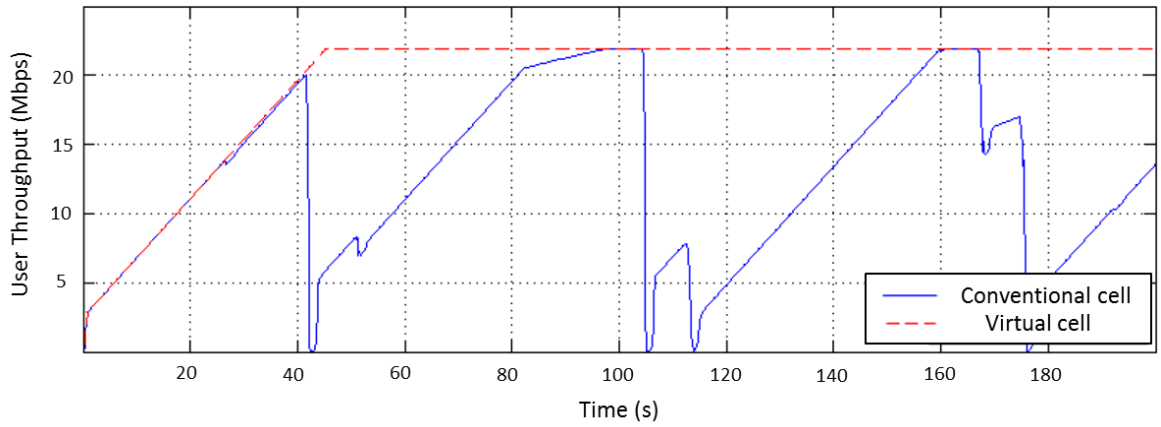

Figure 4-25 Virtual cell user throughput

4.5 Tight Integration with Evolved Legacy AIVs

3GPP is now developing 5G, and the Dual Connectivity (DC) concept from LTE Release 12, see [3GPP13-36842], is being used as a basis for a tighter integration between LTE and 5G. It will enable the UE to be connected to LTE and 5G AIVs (UP and CP) simultaneously. In DC, the node which is responsible for the CP is the Master NB (MeNB) and the node that primarily handles UP is the secondary NB (SeNB) (referring by NB to both eNB and/or gNB). The PDCP layer of the MeNB is common to both AIVs. DC can increase the UE throughput due to UP

aggregation (receiving data from both NBs simultaneously) and make the connection more reliable via CP diversity. The increased reliability also comes from the case when UE handovers to another SeNB, in this case it can still be connected to the MeNB and reliably receive RRC signaling from the MeNB. One disadvantage of DC is that it may be less resource efficient than to be connected only to the best cell, since DC will transmit packets also on links with relatively bad quality and thereby taking resources from users who could utilize the link better.

An alternative to DC is Fast user plane Switching (FS), where the CP is connected to both AIVs at the same time but the UP is only transmitted via one of the AIVs, i.e. the PDCP level routes the packets to one of the AIV nodes. If the CP is connected to both the LTE node and the 5G node, no signaling is required when switching AIV and the UP switch can be almost instantaneous. The FS can be based on normal handover measurements such as RSRP, but also more advanced and faster types of measurements are possible and advantageous. Further details are provided in Annex A.16.

4.5.1 BB Description

Functional Decomposition

The main functional entities in the tight integration BB are depicted Figure 4-26. The main functions are:

- Coordination and configuration unit combined with a user plane splitting function which determines which link the UP should be sent over in case of FS.
- Radio quality measurement signaling over one or both links to eNB/gNB node or UP splitting node. Measurement signaling possibly over the X2* link.
- Radio quality measurements performed in the UE.

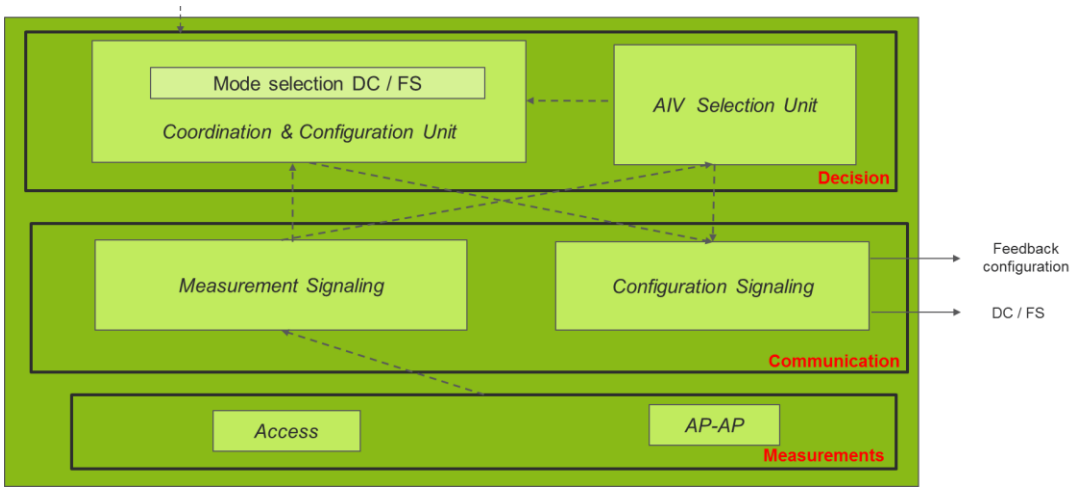

Figure 4-26 LTE-5G AIV Tight Integration Functional Decomposition

A coordination and configuration unit controls which mode is used, i.e. DC or FS. This is based on for example the QoS requirement or the load on the AIVs. There is also an AIV selection node that controls which link transmits the UP packets in case of FS.

Radio quality measurements are performed on both AIVs and signaled to the different AIV nodes and the AIV selection unit in case of FS is applied.

Deployment Architecture

There are two main scenarios for tight integration. The first case is when the AIV nodes are co-located. The other case which can be expected to be common is when there are stand-alone nodes, connected via the X2* interface. In the latter case, the capacity of the X2* interface will have implications on how well the integration can be done, e.g., how fast the UP switching can be performed. The split will in all cases be done on the PDCP protocol layer.

There are also two main options for placement of the PDCP splitting node. Either there is a split before the PDCP packets reaching the MeNB and SeNB or the split is conducted by the MeNB which forwards selected packets to SeNB. In the former case the PDCP protocol entity is moved up to this splitting function from the MeNB.

4.5.2 Design Recommendation

DR-12: Enable PDCP level fast switch between novel and legacy AIVs

Rationale

In cases when DC is configured, two links will be utilized by each UE. This is very often beneficial since it allows for aggregation and hereby increased bit rate compared to if only one AIV is used. However, in many cases, this will make UEs transmit packets over relatively bad links. In case of high load, these links could be better used by other UEs whose connection to the BS is better. In these situations, it is better to employ FS where the UEs only utilize their best link. In addition to use best link and lower the total interference, the FS can also perform rather fast load balancing between the AIVs. In order to make the FS efficient, it is important that the switch can be made relatively fast, in order to always be able to utilize the best link for every UE. Ideally, the switch could in some scenarios be so fast that it can follow the fast fading of a low frequency AIV.

RAN Implications

To enable tight interworking of AIVs, a common S1* CN/RAN interface for LTE and 5G AIV is needed. This means that the S1* can be terminated at the MeNB. The MeNB then handles the setup and teardown of the SeNB. This reduces the signaling to the CN. To handle the FS, new fast measurements will be needed so that the splitting node can make fast and accurate switch decisions. These new measurements can be UE measurements of both AIVs' radio link quality, preferable faster than today's mobility measurements (i.e. RSRP measurements each 480 ms) or even UE measurement normally used for scheduling decisions (CQI).

Results and Analyses

Figure 3-17 in [MET-II16-D51] shows simulations of FTP traffic which illustrates that DC and FS increase the robustness, i.e. for 10th percentile bit rate at low loads compared to stand-alone deployment of NR. For the 90-th percentile, a similar performance between standalone and DC was observed while the FS gave lower bit rate. At higher loads, the performance was more similar between the techniques for the 10th percentile and a slight advantage for FS at the 90th percentile was revealed.

In a second analysis, video traffic has been analyzed and it is now seen that the FS has advantages compared to DC in terms of higher average bit rate at higher loads. It is also demonstrated that using a short switching time which can account for much of the fast fading will enhance performance in terms of shorter delays compared to using a longer switch time. By allowing a faster UP switch, the link can be chosen to better follow the fast fading of LTE. With a link switch possible every 20 ms, the gain is over 20 percent in effective throughput at high load as illustrated in Figure 4-27. How fast the switching must be carried out to follow the fast fading depends on the radio frequency and speed of the UE. This is described further in Table A-7 in Annex A.16. Effective load means that bit rate is calculated only over the time when there is data transmission ongoing, i.e. from the time the video packet arrives in the PDCP buffer in the eNB until it is successfully received at the UE. A more detailed description of the simulation is found in Annex A.16.

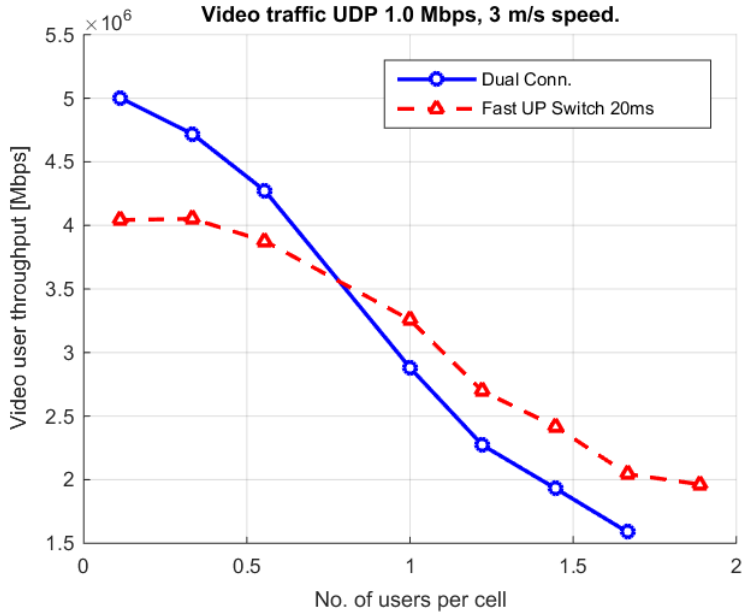

Figure 4-27 User throughput for FS is higher than for DC at high load. At lower load, DC has the advantage of aggregation which gives better performance than FS.

Allowing switches between AIVs on a fast time scale requires that implementation is done efficiently. When the AIV is changed, PDCP packets must be moved to the new node. This will

need to be done efficiently not to cause delays or interruptions. As an example HARQ re-transmissions must be handled. This can be done either by waiting for them to complete or by cancelling them. Also scheduling will be affected, since measurements, e.g. CQI reports must be available at the new AIV. If the packets cannot be scheduled immediately due to missing measurements, FS will be subject to interruptions and less than optimal performance.

4.6 Dynamic Traffic Steering

Dynamic traffic steering BB is a key enabler for achieving the key 5G KPIs such as required by use cases xMBB and URLLC (uMTC). This BB enables the utilization of a dynamic QoS framework to provide additional fast traffic steering functionalities to the 5G RAN, based on the initial considerations presented in [MET-II16-D51]. The key enablers for such a traffic steering function would be the availability of UE link measurements and mmW node localization information to serve the UEs with an optimal number of AN-I nodes, in order to satisfy the requirements of the served traffic. In this section, we provide an overview of the functional decomposition of the BB, the deployment architecture and various design recommendations that arise from the traffic steering framework.

4.6.1 BB Description

Functional Decomposition

The functional decomposition diagram of the dynamic traffic steering framework is as shown in Figure 4-28. Here the key measurement elements required for functioning of the framework are the mmW node localization information and UE link quality measurement information, which is then utilized to enable scheduler signaling and optimizing the QoS function. The traffic steering framework is enabled using two key decision elements – the Pre-emptive Geometrical Interference Analysis (PGIA) and Resource Sharing Cluster (RSC) management function as well as the dynamic traffic steering / serving beam selection function, which includes deployments in mmW and below 6 GHz frequency bands.

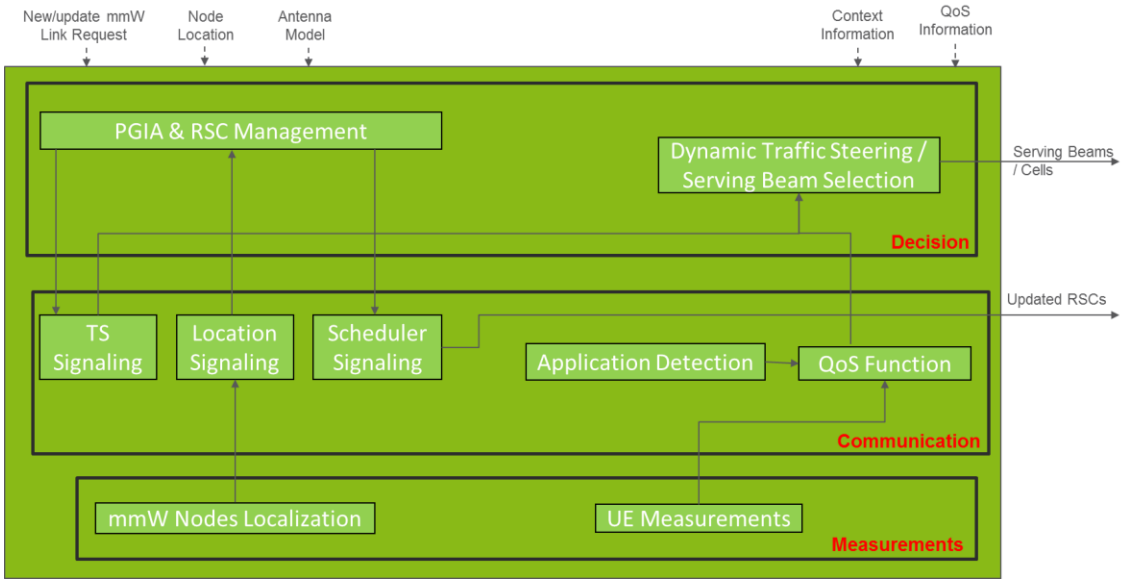

Figure 4-28 Functional decomposition diagram for dynamic traffic steering

Deployment Architecture

The possible deployment architecture of the BB is shown in Figure 4-29, with two key virtual functions implemented in the AIV-agnostic AN-O layer – dynamic QoS function and dynamic traffic steering function enabling key 5G requirements. The information elements from the AN-I layer transport link-level quality information and mmW node localization information to the AN-O layer, which assists in the dynamic traffic steering decision, in terms of optimizing flows and selecting the optimal beams and cells for delivering the traffic to the end users.

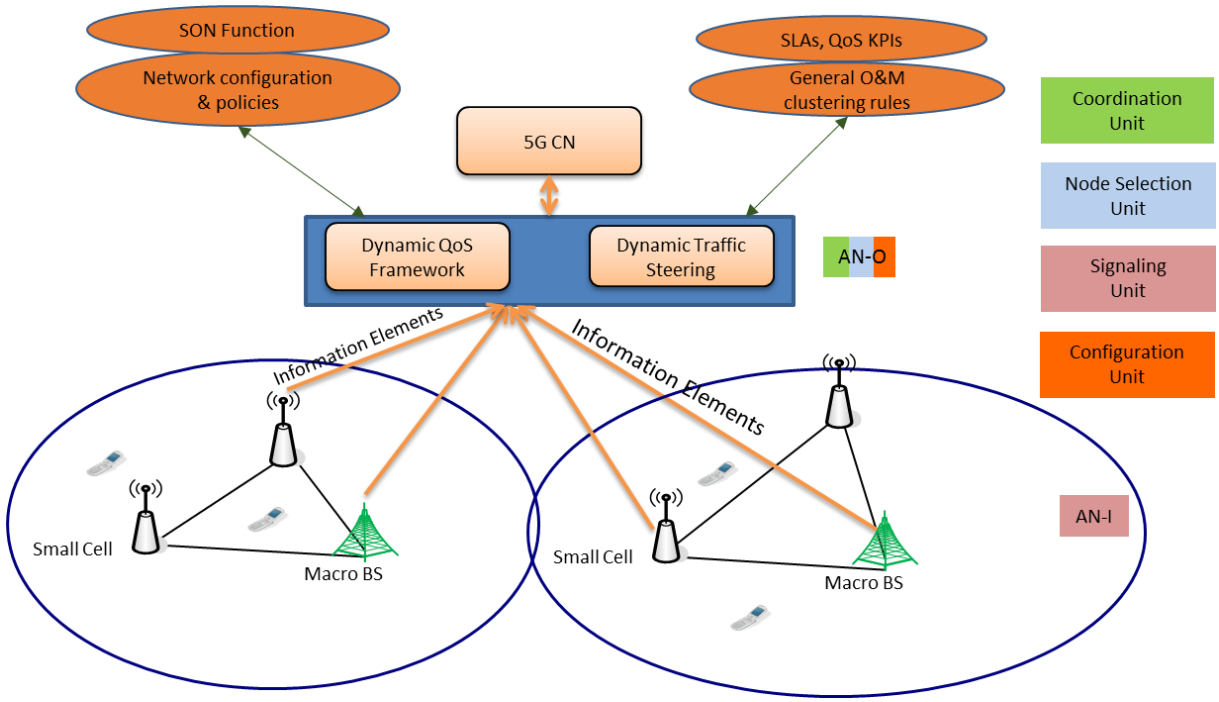

Figure 4-29 Possible deployment architecture for dynamic traffic steering

4.6.2 Design Recommendations

DR-13: Enable Faster Traffic Steering

Rationale

The service flow delivery mechanism using the proposed traffic steering mechanism is as shown in Figure 4-30 [PME+16]. Here the traffic steering framework, based on the link-level quality feedback and with the knowledge of the priority levels of the various service flows to be delivered to the users, routes traffic accordingly to various APs. Due to the relative low reliability of the involved links, the feedback is required in real-time, in order to do a fast traffic rerouting, in case a link failure is detected. Current LTE radio link failure detection and recovery mechanisms would take several seconds in order to re-establish the radio bearer, and since this is unacceptable for high-priority, high-reliability traffic, the dynamic traffic steering framework will ensure that the QoS policies received from the CN are successfully enforced. Further details are provided in Annex A.17.

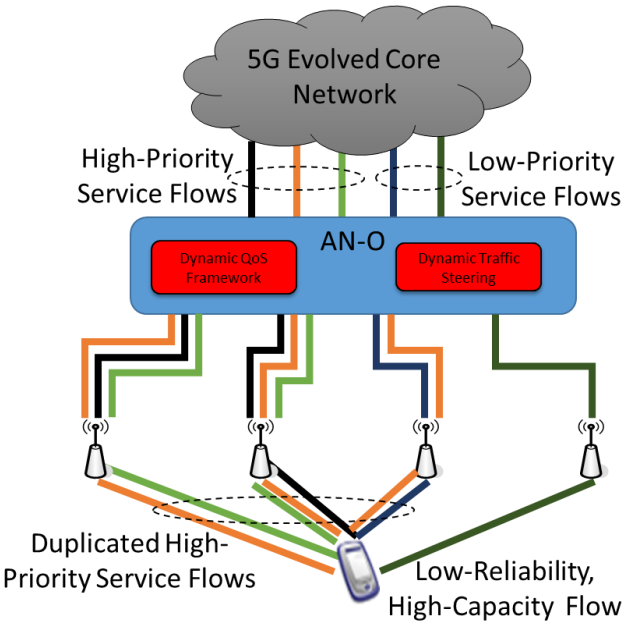

Figure 4-30 Service flow delivery mechanism enabling low-latency in 5G

RAN Implications

New Information Elements: To convey link level AIV-specific measurements from AN-I to AN-O using abstract and quantized values, in order to determine fast traffic steering criteria. New information elements conveying the instantaneous link status in a beam-based system would also be required to be defined over the open interface between the centralized and distributed units.

Fast Routing: The dynamic traffic steering function should be able to route traffic over multiple AIVs instantaneously, depending on the link-level feedback. The routing should essentially be independent of the level of aggregation done in the AN-O (in terms of the protocol stack split within the AN-I relative to the AN-O). Fast routing should also include the ability to do packet duplication over multiple AIVs, depending on the cumulative link reliability in a multi-connectivity system.

Dynamic QoS on the RAN: The RAN should have dynamic QoS functionality, whereby the QoS enforcement capability should be located in the AN-O. This would enable dynamic QoS modification based on the real-time link conditions of the UE, thereby enabling fast QoS adaptation in case there are drastic changes in the link conditions.

Results and Analyses

The proposed concept was implemented using a system level simulator, in order to evaluate the technology potential and to quantify the potential gains. The main focus here has been on reliability improvements through packet duplication over the AI. The scenario assumes an area of 1 sq.km, with 1000 UEs connected to 200 gNBs, randomly deployed, which could resemble a massive indoor industrial environment. Here we consider 2 or 3 of the strongest gNBs and their

corresponding strongest beams are cooperating, while the remaining gNBs are interfering with the UE. Here cooperation indicates that duplicated packets are sent over such links, which is seen as a single transmission from the UE perspective (similar to single frequency network concept). From Figure 4-31, we can observe that the SINR experienced by the UEs is significantly higher for the proposed mechanism as compared to the baseline scheme, which is similar to the current LTE systems, where none of the gNBs are cooperating. From the figure, it can also be noted that significant gains are observed for the fifth percentile or cell-edge users, as seen from the SINR improvement in the lower-percentile users. With the SINR improvement, there would be corresponding improvements in the physical layer parameters that needs to be adopted for transmitting the URLLC data, while minimizing the need for retransmissions, which is the most critical contributor regarding delay within the RAN. Since the URLLC services for augmented / virtual reality and robotic control, etc. requires significant data rates, apart from reliability and low-latency, the proposed mechanism could be a key enabler for provisioning such services in 5G.

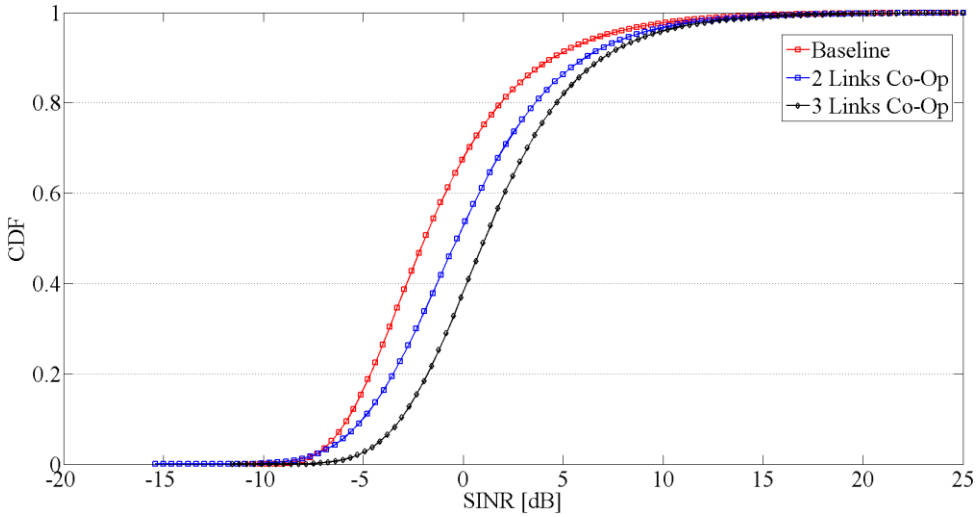

Figure 4-31 SINR distribution for the legacy LTE-A baseline transmissions as compared to cooperative transmissions with duplicated packets sent over multiple links.

DR-14: Enable Pro-active DTS rather than being reactive

Rationale

In order to increase the overall capacity of 5G cellular systems, network radio nodes can be designed to operate at frequencies in mmW bands (e.g., 60 GHz). The mmW band usage will be, nevertheless, coupled with traditional network connections at cellular bands below 6 GHz in heterogeneous environments [STS+14]. In such a heterogeneous environment, as depicted in Figure 4-32, a RM system should be able to manage resources (e.g., steer traffic to the proper transmission link), taking into account the specific characteristics of the various potential transmission links and the actual condition (e.g., interference) at the time of transmission.

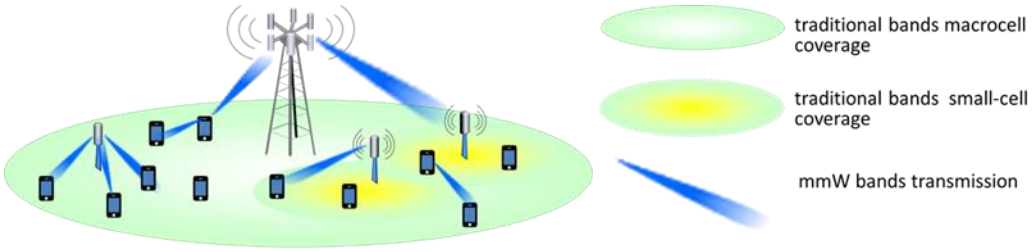

Figure 4-32 Example of heterogeneous environment with mmW band transmissions

The pro-active PGIA mechanism allows to limit transmission collisions (intended here as transmissions creating a very high mutual interference with neighbor transmission links which make the communication impossible) and subsequently signaling overhead by means of a geometrical-based interference analysis that is able to define and update, prior to the establishment of a new mmW transmission link, all the existing sets of prospective interfering mmW transmission links. In other words, each one of these sets called RSC is a set of mmW transmission links that need to be coordinated to avoid potential collisions and should be revised periodically or on an event-triggered basis. The method allows the network to proactively take educated actions before a new transmission link in the mmW band is to be initiated such as implementing a suitable resource partitioning mechanism (e.g., Frequency- or Time-Division Multiplexing), in the AN-I layer (scheduler), between the new transmission link and all the existing and potentially colliding transmission links. Alternatively, other measures in the AN-O layer such as traffic steering (e.g., establishing a transmission link on a lower frequency) can be taken. Details of the PGIA method for avoiding transmission collisions prior to a new mmW transmission are described in the Annex A.18

RAN Implications

A prerequisite is that the **geometrical position** of all the mmW transmitting/receiving nodes (including UEs) of the cellular network should be known by the network at any time regardless of their mobility. Moreover, a model of the antenna of each mmW node (e.g., at least the main beam angle and the Front-to-Back Ratio -FBR-) is needed for the analysis.

PGIA can be decomposed to the following essential LEs. Since the location measurements can be performed by the mmW nodes and sent to the serving BS via measurement signaling there are:

- **mmW Nodes Localization** Function: performs one-time (for *Fixed* mmW nodes), long-term (for *Nomadic* mmW nodes) or short-term (for *Mobile* mmW nodes) localization of mmW nodes (e.g., small cell, terminals, BSs);
- **Location Signaling** Function: sends/updates mmW node location measurements to decision LE.

The decision mechanism is the following:

- **PGIA and RSCs management:** carries out the actual analysis and decision making on pre-emptive geometrical interference analysis and related updating of the RSCs. This is preferably performed inside a serving macro BS (AN-O layer).

It takes into account the internal received location updates when available and alternatively the mmW nodes' locations input, as well as the other input flows, such as, new/update mmW link request and mmW node antenna models. The decision is taken accordingly and the scheduler is informed of the updated RSCs or, alternatively, the Dynamic Traffic Steering Function is informed of the suggested traffic steering via the following functions:

- **Scheduler Signaling:** the updated RSCs are conveyed to the scheduler at the inner layers (AN-I)
- **Traffic Steering Signaling (TS Signaling):** traffic steering suggestions are dispatched to the Dynamic Traffic Steering function (in AN-O)

Figure 4-33 presents the functional logic: mmW BSs, NNs and UEs may perform the required localization assessment and send this information to the serving BS. Upon a new or updated mmW link request the PGIA analysis will convey the updated RSCs to scheduler or traffic steering suggestions to Dynamic Traffic Steering module. Therefore, almost all of the functionality capabilities are deployed at the serving BS and only the localization part can be deployed at mmW BS, NN or UE level.

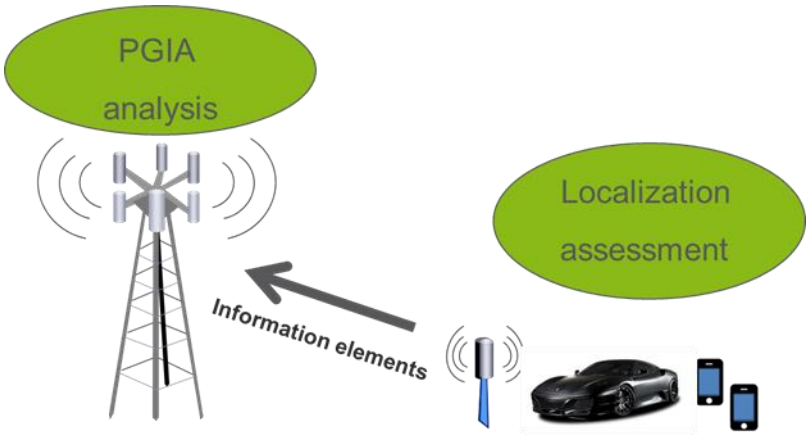

Figure 4-33 PGIA functional logic

Results and Analyses

Preliminary results described in section 3.2.5 of [MET-II16-D51] show that PGIA with a simple resource sharing mechanism can significantly reduce the number of interfered links as the number of concurrent mmW links in 1 km^2 increases. In subsequent work, the average link (user) throughput as a percentage of the total throughput achievable by one single user/link without any resource sharing (whenever the link is the only and single member of an RSC) was added as a KPI. Moreover, through the analysis of the distribution of RSCs w.r.t. their size (number of links within them) the algorithm was found too conservative. In the original algorithm the size of the interfering link cluster was not taken into account in the calculation leading to overestimated interferences and formations of big RSC. Two variants were introduced:

1. PGIA w/ Clusters: in the geometrical analysis, the size of the interfering link cluster is considered, nevertheless interference contributions by different links of the same cluster are summed up and checked against the S/I threshold;

2. PGIA w/ Cluster & Sum: as in PGIA w/ Clusters plus, to count for all interference contributions, all different interferences coming from different clusters are summed up and checked against the S/I threshold; if there's a collision the most interfering cluster is merged with the interfered

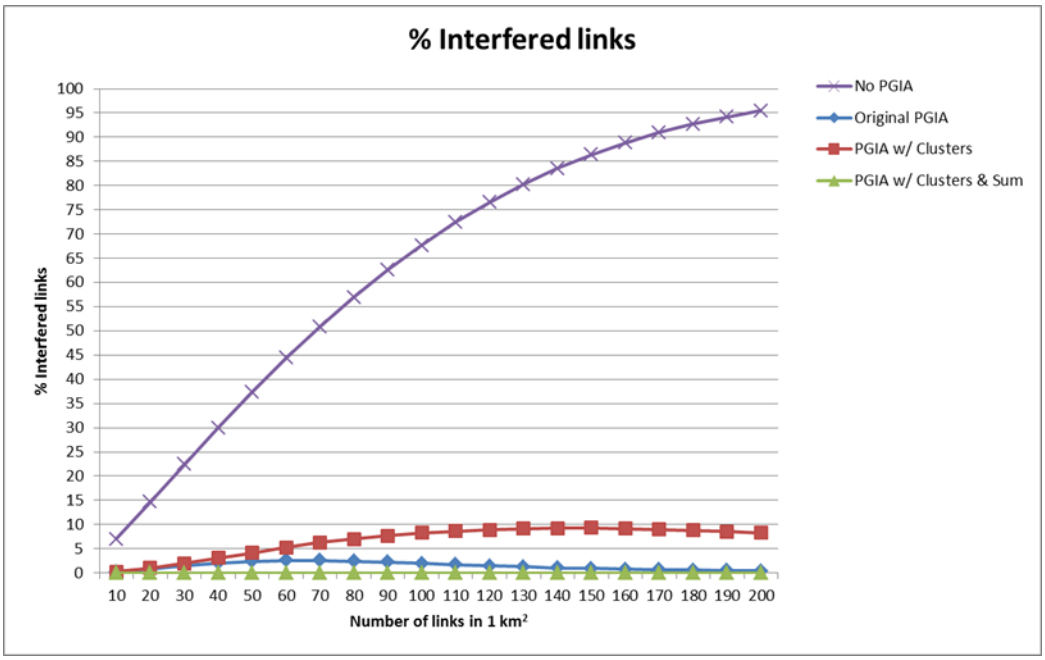

Figure 4-34 Average % of interfered links as a function of the number of concurrent links in 1 km^2, without and with different PGIA algorithms

In Figure 4-34, the fairness of the network with no algorithm or different PGIA algorithms is compared. Under the simulation assumptions described in Annex A.18 the different variants of PGIA can significantly reduce the number of interfered links as the number of concurrent mmW links in 1 km^2 increase: with no PGIA mechanism, the average percentage of interfered mmW links rises over 95% as the number of concurrent links grows to 200, while with the original PGIA the average percentage of interfered mmW links is capped around a very low 2.5%. The PGIA w/ Clusters variant, since it looks for interference only on a cluster by cluster basis keeps the interference capped a little higher but below 9.3%. The PGIA w/ Cluster & Sum variant hunts down every possible collision considering the sum of the interfering signals coming from all the different RSCs: the result is a 0% of interfered links regardless of the number of concurrent mmW links in the area paid by bigger RSCs and lower average throughput than the PGIA w/ Cluster variant as explained in Annex A.18. In fact, other results notably regarding the new KPI based on the average link throughput and a new scenario with obstacles are shown in the Annex A.18

DR-15: Increase Environmental Awareness for RM Mechanisms

Rationale

Due to lower link reliability of mmW networks with beam-based system design, increased radio environmental awareness is essential for the optimal operation of the network. We propose the creation and use of a radio reflection environment map (RefMaps) in ultra-dense mmW deployments. The main aim of RefMaps is to overcome the high probability of link failure in mmW bands, through the fast link recovery which is enabled by dynamic traffic steering. A detailed study of the proposed RefMaps mechanism is presented in [PUL+16]. Here, as illustrated in Figure 4-35, the 5G-gNB configures the UE to report a list of reflected beams with a configured level of signal strength that the UE receives. The signal strength could be configured using RRC signaling similar to the RSRP measurement configurations available currently in LTE. The RefMaps mechanism can be configured in a distributed manner locally at the gNB, with the RefMaps database consisting of the reflected beams and their strength at various physical locations. Centralized deployment of RefMaps could also be done with enhancements in the 5G-SON function, to store the RefMap information in a central database, with the gNBs fetching the information in case of a link failure. This could be done with the assistance of minimization of drive test (MDT) entities such as trace collection entity (TCE) used along with advanced 5G data visualization and optimization tools. This enables enhanced and reliable 5G deployments, irrespective of the frequency band of operation. The mechanism is expected to operate during two phases – the initial or training phase when the RefMaps information is collected through UE measurements and database information is compiled. The second phase is the fast recovery phase, during which link recovery actions are taken by the gNB, when a link breakage is detected, in terms of connecting to the UE using reflected beams. Further details are highlighted in Annex A.19.

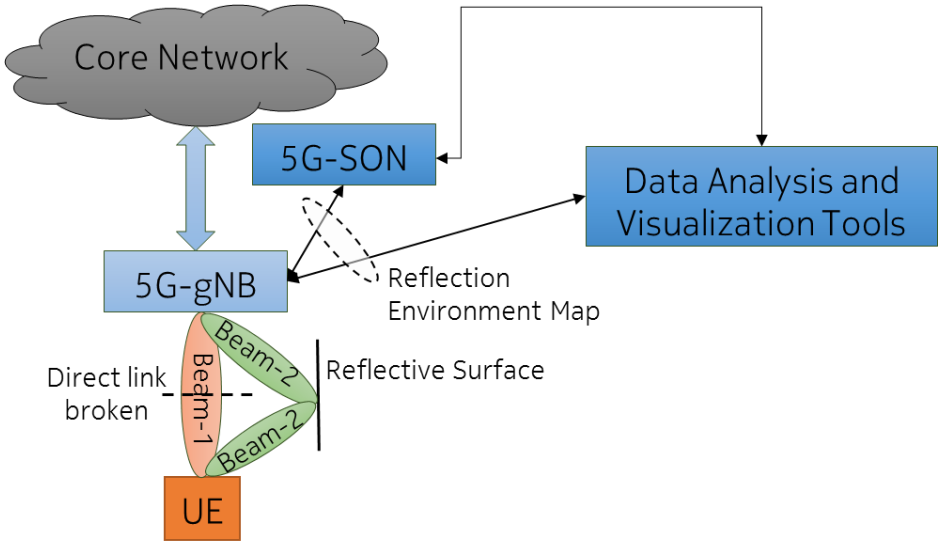

Figure 4-35 Basic working of the RefMaps mechanism [PUL+16]

During the initial / training phase, the UE could be configured to report additional measurements, during which alternate beams by which the UE could be reached, could be calculated, along with the possible loss in signal strength due to the attenuation faced by the reflected beams. The information is then optimized to generate the RefMaps indicating to the gNB about the fast link reestablishment procedure in case direct beams are blocked. The high probability of link failure could also be due to the body loss caused by the link blockage by the UE user as well. The optimizations could be done with the help of enhanced 5G SON functions, which is further studied in [MET-II17-D62].

RAN Implications

New IEs: To transport the reflection environment map information to the AN-O and 5G-SON entities. This would implicitly require new RRC configuration signaling messages to the UE, in order to report direct and reflected beam measurement.

Generate RefMaps: Based on new measurements configurations and database creation algorithms (either at the AN-O or 5G-SON entities), with assistance from data analytics and visualization tools have to be provided.

Results and Analyses

System level simulations were performed to evaluate the performance of RefMaps, using ultra-dense deployment of mmW small cells, with simulation details as presented in [PUL+16], with slight changes in the deployment scenario. We consider an ultra-dense deployment scenario, with small cells operating in mmW band randomly deployed within an area of 500 m x 500 m. The UEs are randomly dropped within the scenario and each is assumed to connect with the strongest cell in its vicinity. We evaluate two cases RefMaps-5 dB and RefMaps-10 dB, where the SINR of the received signals at the UE through reflections is deteriorated by 5 and 10 dB, respectively. This was done in order to estimate the gains that can be achieved using the mechanism, under ideal conditions. The user throughput distribution for the proposed RefMaps mechanism in comparison to the baseline mechanisms is shown. The performance is compared to the dynamic traffic steering solution (DR-13) and baseline case with no enhancements. Here δ_{th} is that perceived loss in SINR due to data transmissions over reflected beams rather than direct beams and δ_{NLOS} is the additional non-line of sight factor considered for the ultra-dense indoor deployment scenario used for evaluations.

The results are shown in Figure 4-36, for the proposed RefMaps enhancement as compared to baseline cases. The results give an indication of the advantage in terms of user throughput and active link users. Here due to the full-buffer traffic assumptions, RefMap scenario causes a slight lowering of user throughput compared to the reference mechanisms, due to the higher number of users being active in the network. Thus, the results indicate that more UEs can be active with the proposed RefMaps enhancement, thereby receiving active downlink data.

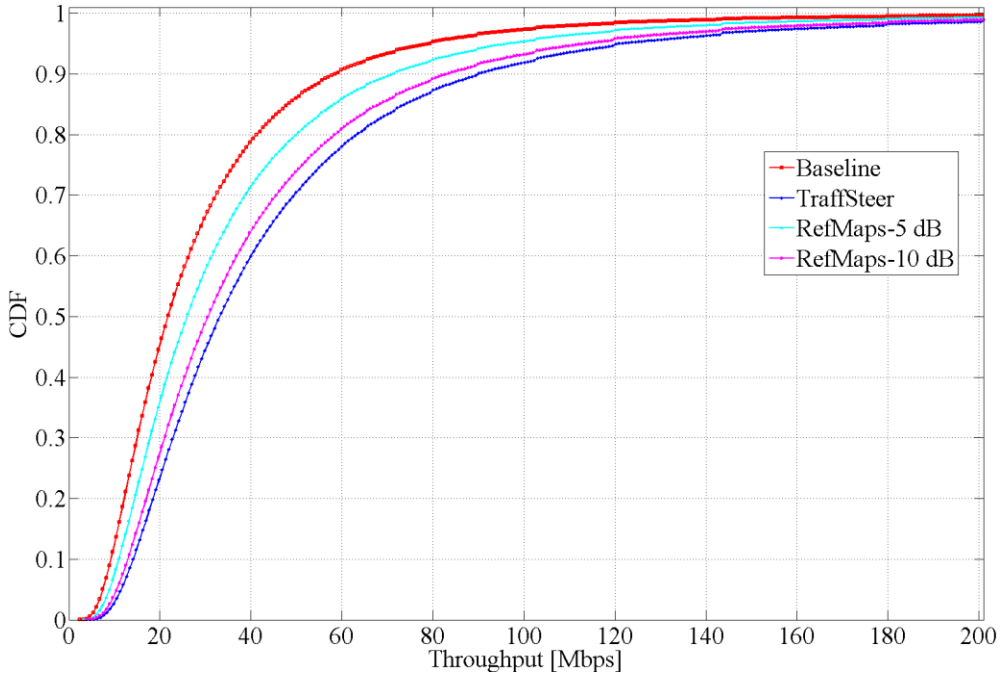

Figure 4-36 Normalized Active UEs, δ_{th} = -6 dB, δ_{NLOS}=0.2

4.7 RM for Inter-Network Collaboration

The aim of this section is to elaborate network assisted resource management for inter-network collaboration and defining a set of control API that can be applied to emerging 5G scenarios. Building on top of the concepts introduced by Software Defined Wireless Networking (SDWN) [YLJ+15][ONF+16], the focus is on providing easy-to-write and globally optimal algorithms to control heterogeneous wireless networks and pave the way for their co-existence. The challenge to address is designing a distributed framework, which facilitates the coordination between these various radio technologies in a scalable manner.

4.7.1 BB Description

Functional Decomposition

In order to develop a fully technology agnostic framework for intra- and inter-network cooperation, the separation of desired functionalities from the underlying infrastructure is necessary. In order to achieve this, the availability of a fully IP-based network among the edge of AIVs of different technologies is required.

The control APIs are implemented on top of common IP based protocols with well-known properties and requirements (e.g., HTTP based REST Interface [RR+08]); these requirements can be identified mostly across reliability and performance characteristics. For example, REST APIs rely on HTTP to reliably perform any operation guaranteeing in order delivery of requests/responses.

The essential logical entities are as follows:

- **Technology-agnostic API: The API serves the** function of providing the necessary abstractions for the logical distribution of functionalities in 5G (universal abstraction of parameters like frequency, weighted average on channel quality measurement parameters like MCS, etc.)
- **RadioMap: The RadioMap** allows the collection of the neighborhood view/state, necessary towards fully deploying coordination techniques in inter-network and inter-technology scenarios

It should be noted that after collecting the RadioMap for multiple wireless domains (ranging from 5G networks to WiFi and WiMAX), different algorithms ranging from integer programming, optimizations and machine learning algorithms can run on the controller. The specification of these algorithms is out of the scope of this document.

Deployment Architecture

Main deployment for the inter-network cooperation case is assumed to be in terms of stand-alone controllers, implementing southbound control interfaces using hardware specific protocols (called control API for simplicity, e.g. OpenFlow, SNMP, etc.) and those expose a northbound REST API to control all exposable parameters of such hardware. Further details are given in Annex A.20.

4.7.2 Design Recommendation

DR-16: Enable technology-agnostic RM framework for inter-network collaboration

Rationale

Separation of desired functionalities from the underlying infrastructure and heterogeneous deployments of wireless access technologies is conducted by introduction of tech-agnostic control API and RadioMap. In order to ensure the coexistence of various wireless technologies like 5G, WiMAX and WiFi this framework is required as it is both infeasible and highly costly to define specialized interfaces for the interaction of these technologies. Moreover, the controller will have the neighborhood view required for globally optimum optimization. Further details are given in Annex A.20.

RAN Implications

The general assumption is that new technology-agnostic control APIs are implemented on top of common IP based protocols with well-known properties and requirements (e.g. HTTP based REST Interface); these requirements can be identified mostly across reliability and performance characteristics. For example, REST APIs rely on HTTP to reliably perform any operation guaranteeing in order delivery of requests/responses. Also these APIs will exist for every RAN and will mainly consist of extensions to SDN.

As the move to a fully IP-based world would be quite drastic, our solution can be adapted to incrementally be deployed over existing protocols. With the exception of the introduction of the

new distribution of resources and entities (e.g. employment of a controller), we envision a minimum impact on the architecture design, with a higher impact on logic distribution of functionalities. As a result, the inter-network collaboration framework can interface with AN-O though a control API, which will enable reporting of an aggregated RadioMap to the controller. Moreover, the ultimate goal will be to have a SON framework, where by plugging a compatible RAN to the aforementioned framework, various parameters will be set. An overview of the framework is shown in Figure 4-37.

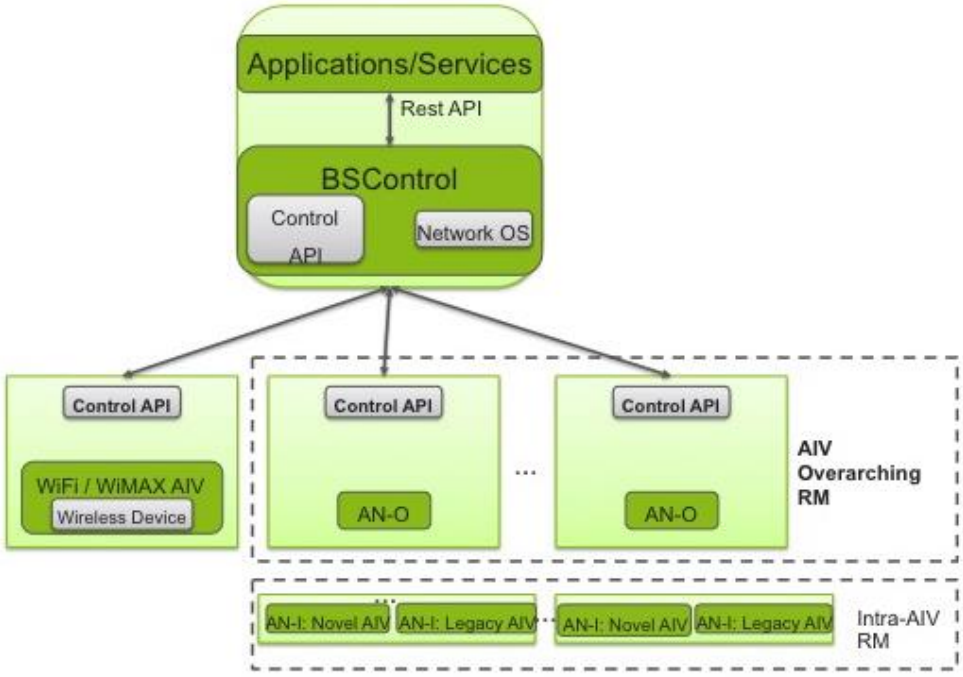

Figure 4-37 Overview of the BS control framework to enable inter-network cooperation

Results and Analyses

Evaluation is being carried out using software prototypes deployed on the ORBIT [RSO+05] wireless testbed, which enables realistic evaluation using software-defined radios and spectrum measurement. The initial evaluations have been done for the specific case of channel assignment in WiFi networks. The channel assignment algorithms can be broadly classified as centralized and distributed. Distributed channel assignment algorithms are mainly based on individual AIVs sensing the spectrum and making sub-optimal decisions. Exploiting SDWN properties enables deployment of optimal centralized algorithms for frequency assignment in ANs (Access Networks).

In our experiments, a centralized channel assignment algorithm, “HSum” algorithm [MBA+05], is used to showcase the necessity of inter-network cooperation. HSum is a weighted graph coloring algorithm, which can be implemented in a centralized manner and conducts frequency assignment to minimize interference.

We deployed the channel assignment algorithm for two scenarios using 8 AIV nodes in ORBIT lab main grid enabled with the Interface API distributed across the grid. The APs are grouped into two logical networks. A dedicated controller that queries the resources on the nodes, as depicted in Figure 4-38, controls each network.

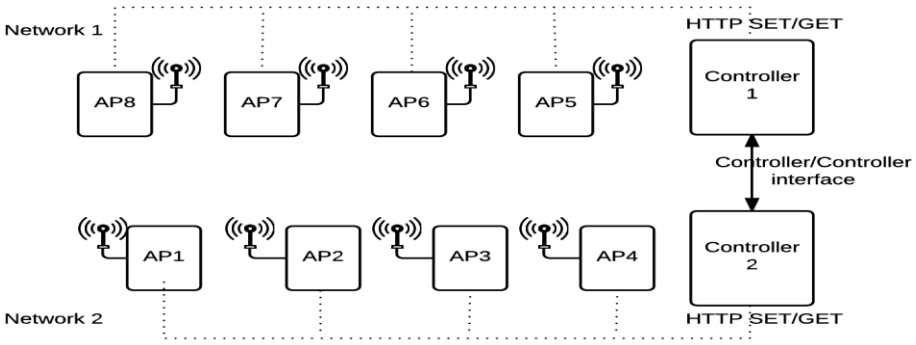

Figure 4-38 Experiment Setup on Orbit Testbed

In the first scenario, each controller independently runs the HSum algorithm. In the second scenario the two controllers exchange RadioMap information after running the HSum channel assignment algorithm, and each controller will rerun the algorithm and modify the assigned parameters (in this case frequency) if better average client throughput is achieved. The result of this experiment is illustrated in Figure 4-39. As can be seen, the throughput improves from ~ 5.5 Mbps to ~8 Mbps with inter-network cooperation.

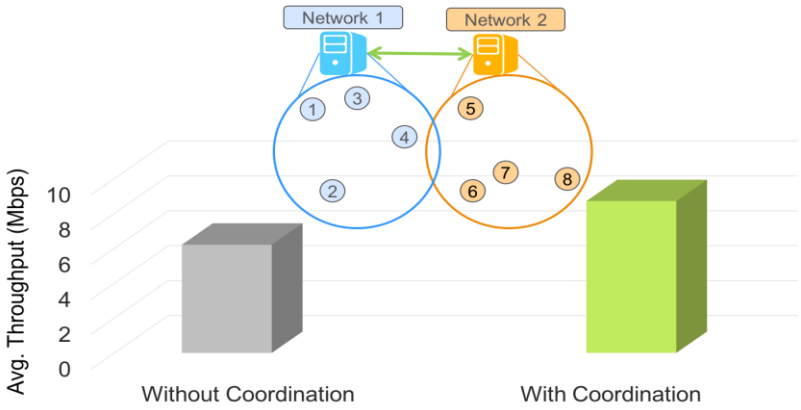

Figure 4-39 Experimental results without and with inter-network cooperation for frequency assignment

4.8 Context Management

The identification of the context information required from different sources (e.g., UE) is of key importance to support efficient RRM required by 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 [MET-II17-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 tracked 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. For example, since multiple use cases possibly with contradicting KPIs are identified in 5G, different AIVs may be used in different use cases. Consequently, more frequent inter-AIV switches (i.e., the switch from a certain AIV to another AIV) are foreseen. For example, inter-AIV switch can occur due to a change in running applications and the related AIVs, or due to the switching between bands and related AIVs. So, to enable the efficient switching from one AIV to another, the UE may need to perform separate measurements for each AIV.

4.8.1 BB Description

Functional Decomposition

The considered BB can be decomposed to 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.

The functional decomposition is given in Figure 4-40.

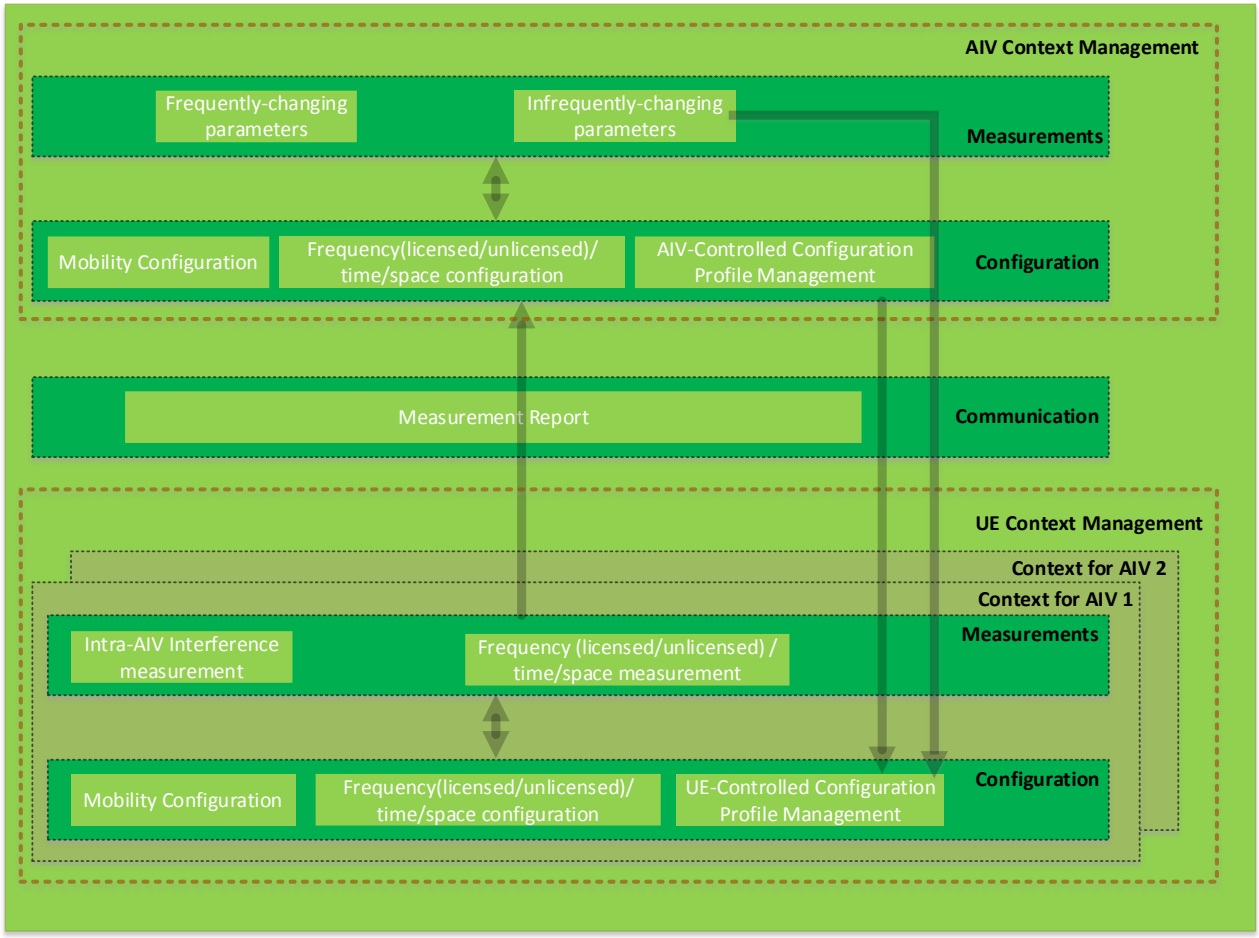

Figure 4-40 Functional decomposition of context management

In what follows, we describe how the different logical entities 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 logical entities 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).

4.8.2 Design Recommendation and RAN Design Implication

DR-17: Adaptive UE Context Management to support novel RM mechanisms within the agile RM framework

The UE measurement context should be adapted to assist in reporting existing information (such as location, CQI, and HARQ) with a more accurate estimate (as required by the BB "Dynamic Traffic Steering" in Section 4.6 and the BB “Multi-Slice and Multi-Service Holistic RM” in Section 4.4), and reporting new information such as the inter-AIV interference (as required by the BB "Interference Management" in Section 4.1). Moreover, the exploitation of high frequencies is based on directional transmission schemes via beam-forming (as required by the BB “RAN moderation” in Section 4.3), which may imply updating the measurement context to support such new configurations (e.g., space tailored configurations as required by the BB "Interference Management" in Section 4.1). The UE context management should also be adaptive and suitable for high mobility, as envisioned in [MET-II16-D61]. The measurements in unlicensed bands should be supported to trigger a more flexible use of radio resources, such as

in LAA and LTE-WLAN-Aggregation (as required by the BB “Flexible Short-Term Spectrum Usage” in Section 4.2). Further details are provided in Annex A.21.

In addition, in order to satisfy various mobility requirements in different 5G use cases, measurement configurations should be adapted to support mobility based configurations (such as a mobility based measurement interval). Furthermore, the UE could be able to maintain multiple measurement contexts, such as multiple configurations for multiple AIVs (as required by the BB "Tight Integration with Evolved Legacy AIVs" in Section 4.5).

The indicated functional extensions and changes on the UE measurement context compared to 3GPP Release 13 are summarized in Figure 4-41.

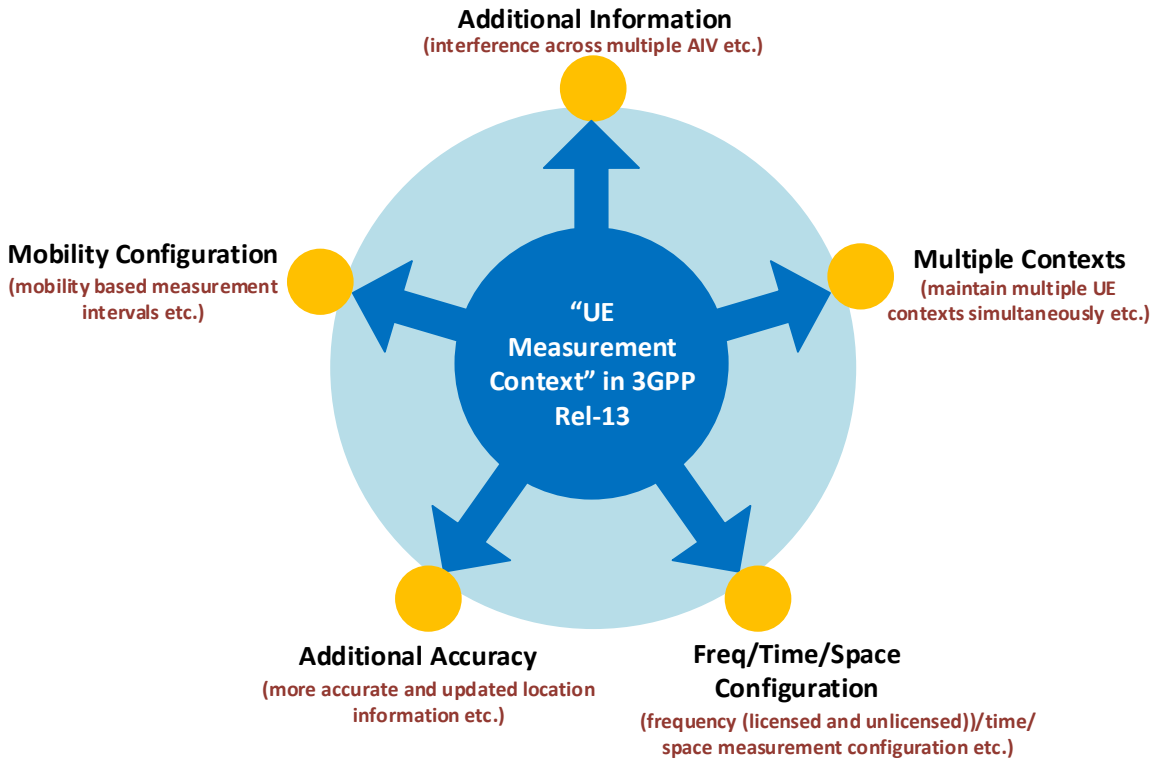

Figure 4-41 Possible changes on UE Measurement context

5 Conclusions and Outlook

In this deliverable, we have provided the design of the agile RM framework comprising two main functionality frameworks, namely, AIV-overarching RM and Intra-AIV RM, and the associated enabling technologies in terms of BBs. The essential LEs and LEGs of the BBs construct these functionality frameworks. The agile RM framework operates over 5G AIVs and legacy AIVs, which take into account diverse service requirements, dynamic radio topologies, and novel communication modes. Furthermore, the BBs of this framework are described along with the qualitative and quantitative analyses, where various design recommendations are highlighted along with the RAN design implications. In particular, the 5G RM framework shall

- adapt to peculiarities of multiple AIVs that constitute the overall 5G AI,
- take into account dynamic radio topologies with new types of access nodes, such as NNs,
- capitalize on new interference-resistive designs to mitigate inter-cell interference especially for cell-edge users,
- be able to operate over extended notion of a resource, such as unlicensed band,
- enable tight interworking with evolved legacy networks, such as on RAN level, and operate on a faster time scale, e.g., by dynamic traffic steering, to avoid hard handovers,
- enable RAN support for end-to-end slicing considering novel 5G aspects, and
- enable energy-efficient network operation while fulfilling service requirements.

To this end, the UE context shall be extended to support the developed synchronous control functions. Consequently, the proposed agile RM framework is seen as an enabler for a multi-AIV, multi-service, and multi-slice 5G system, fostering timely and efficient standardization. This deliverable is also expected to provide means for joint discussions with other 5G PPP projects. The timeline is illustrated in Figure 5-1 along with the various milestones and project deliverables that are the most relevant to D5.2 (see, for example, Section 1.1).

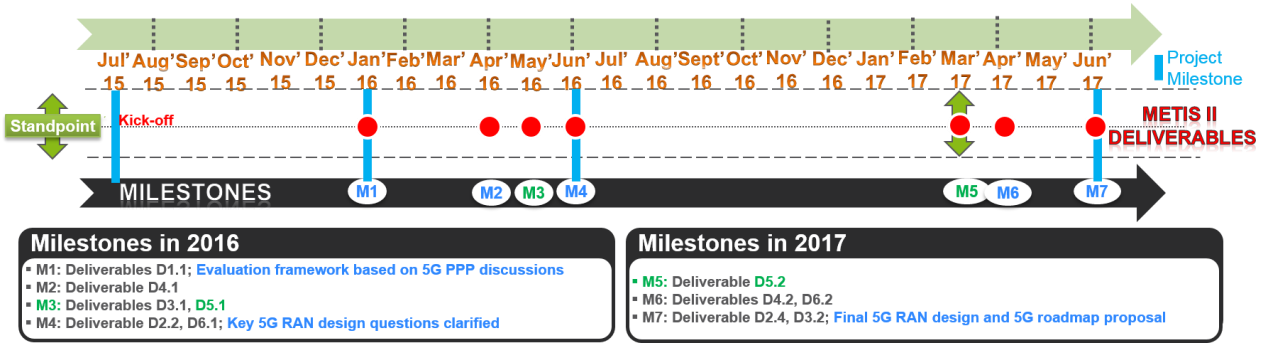

Figure 5-1 Timeline including METIS-II deliverables and milestones.

6 References

[3GPP05-05] 3GPP TS 05.05, Technical Specification Group GSM/EDGE Radio Access Network, “Radio transmission and reception,” Annex C.3 Propagation models, Release 1999, v8.20.0 (2005-11).

[3GPP10-36814] 3GPP TR 36.814, “Further advancements for E-UTRA physical layer aspects,” Release 9, v9.0.0 (2010-03).

[3GPP13-36819] 3GPP TR 36.819, “Coordinated multi-point operation for LTE physical layer aspects,” Release 11, v11.2.0 (2013-09).

[3GPP13-36842] 3GPP TR 36.842 “Study on Small Cell enhancements for E-UTRA and E-UTRAN” Release 12, v12.0.0 (2013-12).

[3GPP13-36874] 3GPP TR 36.874 “Coordinated multi-point operation for LTE with non-ideal backhaul,” Release 12, v12.0.0 (2013-12).

[3GPP14-23203] 3GPP TS 23.203, Technical Specification Group Services and System Aspects, “Policy and charging control architecture,” Release 13, v13.0.1, (2014-09).

[3GPP14-36314] 3GPP TS 36.314, “E-UTRA: Layer-2 Measurements,” Release 13, v13.0.0 (2014-12).

[3GPP15-36300] 3GPP TS 36.300, Technical Specification Group Radio Access Network Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN), “Overall description,” Release 13, v13.2.0 (2015-12).

[3GPP15-36889] 3GPP TR 36.889, “Feasibility Study on Licensed-Assisted Access to Unlicensed Spectrum”, Release 13, v13.0.0 (2015-07)

[3GPP16-36213] 3GPP TS 36.213, “Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures,” Release 13, v13.1.1 (2016-03).

[3GPP16-23799] 3GPP TR 23.799, “Study on Architecture for Next Generation System”, Release 14, v2.0.0 (2016-11)

[3GPP16-36423] 3GPP TS 36.423, Evolved Universal Terrestrial Radio Access Network (E-UTRAN), “X2 Application Protocol (X2AP),” R8, R10, R11, R12+.

[3GPP16-38801] 3GPP TR 38.801 “Study on New Radio Access Technology; Radio Access Architecture and Interfaces,” Release 14, v1.0.0 (2016-12).

[3GPP17-38802] 3GPP TR 38.802, “Study on New Radio (NR) Access Technology; Physical Layer Aspects”, R14 (2017-01)

[3GPP16-RP160043] 3GPP RP-160043, Summary of “[5G-AH-06] Fronthauling,” Telecom Italia, et al.

[3GPP16-RP160600] 3GPP RP-160600, “Enhanced LTE-WLAN Aggregation”, Work Item.

[5GREEN] 5GREEN Project, see http://wireless.kth.se/5green/.

[5GNORMA15-D31] 5G NORMA project deliverable 3.1 “Functional network architecture and security requirements,” Dec 2015.

[5GNORMA17-D32] 5G NORMA project deliverable 3.2 “5G NORMA network architecture – Intermediate report”, Jan 2017.

[ABH11] I. Ashraf, F. Boccardi, and L. Ho. "Sleep Mode Techniques for Small Cell Deployments." *IEEE Communications Magazine*, Vol. 49, No. 8, pp. 72-79, August 2011.

[ACO+13] H. Ali-Ahmad, C. Cicconetti, A. La Oliva, V. Mancuso, M. Reddy Sama, P. Seite, and S. Shanmugalingam, (2013, November), “An SDN-based network architecture for extremely dense wireless networks,” In Future Networks and Services (SDN4FNS), 2013 IEEE SDN.

[ADF+12] D. Astely, E. Dahlman, G. Fodor, S. Parkvall and J. Sachs, “LTE Release 12 and Beyond,” *IEEE Communications Magazine*, Vol. 51, No. 7, pp. 154-160, July 2013.

[AFT+15] A. Abrardo, G. Fodor, B. Tola, “Network Coding Schemes for D2D Communications Based Relaying for Cellular Coverage Extension,” IEEE Signal Processing Advancements for Wireless Communications (SPAWC), Stockholm, June 2015.

[AHT+11] S. B. Ali, Y. Chia-Hao, O. Tirkkonen, and C. Ribeiro, “Optimization of Dynamic Frame Offset in Time Division Duplex System,” in *IEEE 73rd Vehicular Technology Conference (VTC Spring),* pp.1-5, 15-18 May 2011.

[ALS+14] M.R. Akdeniz, Y. Liu, M.K. Samimi, S. Sun, S. Rangan, T.S. Rappaport, E. Erkip “Millimeter Wave Channel Modeling and Cellular Capacity Evaluation”, *IEEE Journal on Selected Areas in Communications*, Vol. 32, No 6, June 2014.

[ART10-D51] ARTIST4G D5.1, “Scenarios, Key Performance Indicators and Evaluation Methodology for Advanced Cellular Systems,” May 2010, https://ict-artist4g.eu.

[BFJ+15] M. Belleschi, G. Fodor, D. D. Penda and M. Johansson, “Benchmarking Practical RRM Algorithms for D2D Communications in LTE Advanced”, *Wireless Personal Communications*, Volume 82, Issue 2, pp 883-910, May 2015.

[BI09] C.S. Bontu, and E. Illidge. "DRX mechanism for power saving in LTE." *IEEE Communications Magazine*, Vol. 47, No. 6, pp. 48-55, June 2009.

[BKE+17] Ö. Bulakci, A. Kaloxylos, J. Eichinger, C. Zhou, “RAN Moderation in 5G Dynamic Radio Topology,” IEEE VTC 2017- Spring, Sydney, Australia.

[BPB+15] Ö. Bulakci, A. Prasad, J. Belschner, M. Ericson, I. Karls, H. Celik, M. Tesanovic, R. Fantini, L.M. Campoy, E, Pateromichelakis, F.S. Moya, G. Zimmermann, I.D. Silva, “Agile Resource Management for 5G – A METIS-II Perspective,” IEEE CSCN, Tokyo, Oct. 2015.

[BRZ+14] Ö. Bulakci, Zhe Ren, Chan Zhou, et al., "Dynamic Nomadic Node Selection for Performance Enhancement in Composite Fading/Shadowing Environments," 2014 IEEE VTC Spring, Seoul, 2014, pp. 1-5.

[BRZ+15] Ö. Bulakci, Zhe Ren, Chan Zhou, et al., “Towards flexible network deployment in 5G: Nomadic node enhancement to heterogeneous networks,” IEEE ICC, London, 2015, pp. 2572-2577.

[BSR+11] G. Bhanage, I. Seskar, D. Raychaudhuri, “A virtualization architecture for mobile WiMAX networks,” ACM SIGMOBILE Mobile Computing and Communications Review, vol. 15, no. 4, October 2011.

[BSS+12] A. Baid, M. Schapira, I. Seskar, J. Rexford, D. Raychaudhuri, “Network cooperation for client-AP association optimization,” Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2012.

[CS15] H. Celik and K. W. Sung, “On the Feasibility of Blind Dynamic TDD in Ultra-Dense Wireless Networks,” in IEEE VTC (Spring), pp.1-5, 11-14 May 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.

[EAR12-D23] ICT-247733 EARTH project, “Energy efficiency analysis of the reference systems, areas of improvements and target breakdown”, Deliverable D2.3 v2.0, Jan. 2012.

[EARTH] EARTH Project, see http://www.ict-earth.eu.

[ETSI-NFV] ETSI, “Network Functions Virtualisation (NFV); Management and Orchestration” GS NFV-MAN 001 V1.1.1, Dec. 2014, see http://www.etsi.org/deliver/etsi_gs/NFV-MAN/001_099/001/01.01.01_60/gs_nfv-man001v010101p.pdf.

[FAN17-D42] ICT-671660 Fantastic5G, Deliverable 4.2 Version 1 “Final results for the flexible 5G air interface multi-node/multi-antenna solution”, May 2017.

[FMC+13] P. Fotiadis, M. Polignano, L. Chavarria, I. Viering, C. Sartori, A. Lobinger, K. Pedersen "Multi-Layer Traffic Steering: RRC Idle Absolute Priorities & Potential Enhancements." IEEE 77th Vehicular Technology Conference (VTC Spring), 2013.

[FMM+11] P.l. Frenger, et al. "Reducing Energy Consumption in LTE With Cell DTX." IEEE 73rd Vehicular Technology Conference (VTC Spring), 2011.

[FPS+14] G. Fodor, S. Parkvall, S. Sorrentino, P. Wallentin, Q. Lu, N. Brahmi, "Device-to-Device Communication for National Security and Public Safety"**,** *IEEE Access*, Vol. 2, pp. 1510-1520, December, 2014.

[GSMA14] GSM Association (GSMA), “The Impact of Licensed Shared Use of Spectrum,” Report produced by Deloitte for the GSMA, Jan. 2014.

[HRL+14] K. T. Hemachandra, N. Rajatheva, M. Latva-aho, “Sum-Rate Analysis of Full-Duplex Underlay Device-to-Device Networks”, *IEEE Wireless Communications and Networking Conference* (WCNC), Istanbul, Turkey, April 6-9, 2014.

[HRT+14] E. Hossain, et al., “Evolution Toward 5G Multi-tier Cellular Wireless Networks: An Interference Management Perspective,” IEEE Wireless Communications, 21(3), pp. 118-127, 2014.

[HSL+14] S. Hong et al., “Frequency and Quadrature-Amplitude Modulation for Downlink Cellular OFDMA Networks,” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp.1256-1267, Jun. 2014.

[HVM14] L. Haeyoung, S. Vahid and K. Moessner, "A Survey of Radio Resource Management for Spectrum Aggregation in LTE-Advanced," in IEEE Communications Surveys & Tutorials, vol.16, no.2, pp. 745-760, 2014.

[ITU-R15] ITU-R WP5D, Revision 2 to Document 5D/TEMP/469-E, Chapter 5.3.8.

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

[JLW11] N.T.K. Jørgensen, D. Laselva, and J. Wigard. "On the potentials of traffic steering techniques between HSDPA and LTE." In IEEE 73rd Vehicular Technology Conference (VTC Spring), 2011.

[JRK12] P. Jänis, C. B. Ribeiro, and V. Koivunen, “Flexible UL-DL Switching Point in TDD Cellular Local Area Wireless Networks,” MONET 17(5): 695-707 (2012).

[KTC15] P. Kela, J. Turkka, M. Costa, “Borderless Mobility in 5G Outdoor Ultra-Dense Networks”, IEEE Access, Vol. 3, pp. 1462-1476, 2015.

[LBS+15] J. Li, E. Bjornson, T. Svensson, T. Eriksson and M. Debbah, “Joint precoding and load balancing optimization for energy-efficient heterogeneous networks”, IEEE Transactions on Wireless Communications, vol. 14, no. 10, pp. 5810-5822 2015.

[Lei14] Y. Leiba, “Systems and Methods for Improving the Quality of Millimeter-wave Communication”, US. Patent No. 2014/0370925 Al, 18 December 2014.

[LPV+17] Y. Li, E. Pateromichelakis, N. Vucic, J. Luo, W. Xu, G. Caire, “Radio Resource Management Considerations for 5G Millimeter Wave Backhaul and Access Networks”, IEEE Communications Magazine, June 2017 (to appear)

[MBA+05] A. Mishra, S. Banerjee, W. Arbaugh “Weighted coloring based channel assignment for WLANs”, ACM SIGMOBILE Mobile Computing and Communications Review, 2005

[MNK+07] P. Mogensen et al., “LTE Capacity Compared to the Shannon Bound,” IEEE 65th Vehicular Technology Conference Spring, Dublin, 2007, pp. 1234-1238.

[MBL+13] P. Munoz, R. Barco, D. Laselva, P. Mogensen, "Mobility-based strategies for traffic steering in heterogeneous networks." IEEE Communications Magazine, Vol. 51, No. 5, pp. 54-62, 2013.

[MET13-D11] METIS project, “Scenarios, requirements and KPIs for 5G mobile and wireless system”, Deliverable 1.1 Version 1, April 2013.

[MET14-D32] METIS project, “First performance results for multi-node/multi-antenna transmission technologies”, Deliverable D3.2, Apr. 2014

[MET15-D15] METIS project, “Updated scenarios, requirements and KPIs for 5G mobile and wireless system with recommendations for future investigations”, ICT-317669 METIS Deliverable D1.5, Version 1, April 2015.

[MET15-D33] METIS project, “Final performance results and consolidated view on the most promising multi-node/multi-antenna transmission technologies,” ICT-317669 METIS Deliverable D3.3, Feb. 2015.

[MET15-D42] METIS project, “Final Report on Network Level Tradeoff Investigations”, Deliverable D4.2, August 2014.

[MET15-D43] METIS project, “Final Report on Network Level Solutions”, Deliverable D4.3, February 2015.

[MET15-D64] METIS project, “Final report on Architecture,” ICT-317669 METIS Deliverable D6.4, Version 2, January 2015.

[MET15-D66] METIS project, “Final report on the METIS 5G system concept and technology roadmap,” ICT-317669 METIS Deliverable D6.6, April 2015.

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

[MET-II16-D11] METIS-II, “Refined scenarios and requirements, consolidated use cases, and qualitative techno-economic feasibility assessment,” Deliverable 1.1, Jan 2016.

[MET-II16-D21] METIS-II, “Performance evaluation framework”, Deliverable D2.1, Jan. 2016.

[MET-II16-D22] METIS-II, “Draft Overall RAN Design”, Deliverable D2.2, June 2016.

[MET-II16-D41] METIS-II, “Draft air interface harmonization and user plane design,” Deliverable 4.1, April 2016.

[MET-II16-D51] METIS-II, “Draft Synchronous Control Functions and Resource Abstraction Considerations”, Deliverable D5.1, May 2016.

[MET-II16-D61] METIS-II, “Draft asynchronous control functions and overall control plane design”, Deliverable D6.1, June 2016.

[MET-II16-WP] METIS-II, “Preliminary views and initial considerations on 5G RAN architecture and functional design”, White Paper, March 2016.

[MET-II17-D24] METIS-II, “D2.4 Final overall 5G RAN design”, June 2017

[MET-II17-D42] METIS-II, “D4.2 Final air interface harmonization and user plane design”, April 2017

[MET-II17-D62] METIS-II, “D6.2 Final asynchronous control functions and overall control plane design”, April 2017

[NGM15] Next Generation Mobile Networks (NGMN) Alliance, “5G White Paper”, Feb. 2015, https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf.

[NGM_CoMP15] Next Generation Mobile Networks (NGMN) Alliance, “RAN EVOLUTION PROJECT COMP EVALUATION AND ENHANCEMENT”, March 2015, https://www.ngmn.org/uploads/media/NGMN_RANEV_D3_CoMP_Evaluation_and_Enhancement_v2.0.pdf.

[OFE+13] Brief, ONF Solution, "OpenFlow™-Enabled Mobile and Wireless Networks." September 2013, https://www.opennetworking.org/images/stories/downloads/sdn-resources/solution-briefs/sb-wireless-mobile.pdf

[ONF+16] "Third Wireless Transport SDN Proof of Concept White Paper", December 2016, https://www.opennetworking.org/images/stories/downloads/sdn-resources/technical-reports/Third-Wireless-Transport-SDN-Proof-of-Concept-White-Paper.pdf

[OPA+15] A. De la Oliva, X.C. Perez, A. Azcorra, A. Di Giglio, et al., "Xhaul: Toward an integrated fronthaul/backhaul architecture in 5G networks," IEEE Wireless Communications, vol. 22, no. 5, 2015.

[ORB] ORBIT Testbed: http://www.orbit-lab.org.

[PBF+16] K.I. Pedersen, G. Berardinelli, F. Frederiksen, A. Szufarska, "A Flexible 5G Frame Structure Design for Frequency-Division Duplex Cases", in IEEE Communications Magazine, Vol. 54, No. 3, pp. 53-59, March 2016.

[PCF+16] E. Pateromichelakis, H. Celik, R. Fantini, et al., "Agile Interference Management Framework for 5G," IEEE CSCN 2016, October 2016.

[PFT+00] J. Padhye, et al., "Modeling TCP Reno Performance: A Simple Model and Its Empirical Validation", in IEEE/ACM Trans. on Networking, Vol. 8, No. 2, pp. 133-145, April 2000.

[PGR09] M. Park, P. Gopalakrishnan, R. Roberts, "Interference Mitigation Techniques in 60 GHz Wireless Networks", IEEE Communications Magazine, Vol. 47, No. 12, pp. 34-40, December 2009.

[PLS+15] G. Pocovi, B. Soret, M. Lauridsen, K. I. Pedersen and P. Mogensen, "Signal Quality Outage Analysis for Ultra-Reliable Communications in Cellular Networks," IEEE Globecom Workshops (GC Wkshps), San Diego, CA, 2015, pp. 1-6.

[PLV+15] A. Prasad, P. Lundén, K. Valkealahti, M. Moisio, M.A. Uusitalo, "Network Assisted Small Cell Discovery in Multi-Layer and mmWave Networks," IEEE Conference on Standards for Communication and Networking (CSCN), Oct. 2015.

[PM15] A. Prasad, and A. Maeder. "Backhaul-Aware Energy Efficient Heterogeneous Networks with Dual Connectivity," Telecommunication Systems, Vol. 59, No. 1 pp. 25-41. 2015.

[PME+16] A. Prasad, F.S. Moya, M. Ericson, R. Fantini, Ö. Bulakci, "Enabling RAN Moderation and Traffic Steering in 5G Radio Access Networks," 84th IEEE Vehicular Technology Conference, Montreal, Sept. 2016.

[PNS+16] K.I. Pedersen, M. Niparko, J. Steiner, J. Oszmianski, L. Mudolo, S. Khosravirad, "System Level Analysis of Dynamic User-centric Scheduling for a Flexible 5G Design", IEEE Globecom, Washington D.C., Dec. 2016

[PP17] E. Pateromichelakis, C. Peng, "Selection and Dimensioning of slice-based RAN Controller for adaptive Radio Resource Management", IEEE Wireless Communications and Networking Conference (WCNC), March 2017.

[PPS+16] G. Pocovi, et al., "On the Impact of Multi User Traffic Dynamics on Low Latency Communications", 13th International Symposium on Wireless Communication Systems (ISWCS), Sept. 2016

[PSQ+13] E. Pateromichelakis, et al., “On the Evolution of Multi-Cell Scheduling in 3GPP LTE / LTE-A,” IEEE Communications Surveys & Tutorials, Vol. 15, No. 2, pp. 707-717 2013.

[PSQ+14] E. Pateromichelakis, M. Shariat, A. U. Quddus and R. Tafazolli, “Graph-Based Multicell Scheduling in OFDMA-Based Small Cell Networks,” in IEEE Access, vol. 2,pp. 897-908, 2014.

[PSW+13] C. S. Park et al., "Carrier aggregation for LTE-advanced: design challenges of terminals," in Communications Magazine, IEEE, vol.51, no.12, pp.76-84, December 2013.

[PUL+16] A. Prasad, M.A. Uusitalo, Z. Li, P. Lundén, "Reflection Environment Maps for Enhanced Reliability in 5G Self-Organizing Networks," IEEE Globecom Workshops (GCW), Washington D.C., Dec. 2016.

[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.

[RFC6824] IETF, TCP Extensions for Multipath Operation with Multiple Addresses, Jan. 2013.

[RR+08] L. Richardson, S. Ruby “RESTful web services,” O’Reilly Media, Inc.”, 2008.

[RSD+15] S.Redana et al, “5G NORMA: A NOvel Radio Multiservice adaptive network Architecture for 5G networks,” EuCNC 2015.

[RSO+05] D. Raychaudhuri, I. Seskar, M. Ott et al., “Overview of the ORBIT radio grid testbed for evaluation of next-generation wireless network protocols,” IEEE WCNC 2005

[SA05] M. Simon and M. Alouini, Digital communication over fading channels, 2nd ed. New York: John Wiley & Sons Inc., 2005.

[SA13] I. Shomorony and A. S. Avestimehr, “Worst-case additive noise in wireless networks,” IEEE Trans. Inf. Theory, vol. 59, no. 6, pp. 3833–3847, Mar. 2013.

[SAM15] Samsung 5G Vision White Paper, March 2015 (available at http://www.samsung.com/global/business-images/insights/2015/Samsung-5G-Vision-0.pdf).

[SAV] SAVANT Project at http://www.winlab.rutgers.edu/projects/savant/Index.html.

[SBS+15] S. Sagari, S. Baysting, D. Saha, I. Seskar, W. Trappe, D. Raychaudhuri, Coordinated dynamic spectrum management of LTE-U and Wi-Fi networks, IEEE Dynamic Spectrum Access Networks (DySPAN), 2015.

[SBS+17] T. Sahin, Ö. Bulakci, P. Spapis, A. Kaloxylos, “Flexible Network Deployment in 5G: Performance of Vehicular Nomadic Nodes,” VTC 2017- Spring, Sydney, Australia.

[SFM+14] J. M. B. da Silva Jr, G. Fodor, T. F. Maciel, "Performance Analysis of Network Assisted Two-Hop D2D Communications", 10th IEEE Broadband Wireless Access Workshop, Austin, TX, USA, December 2014.

[SMM11] S. Singh, R. Mudumbai, U. Madhow, “Interference analysis for highly directional 60 GHz mesh networks: the case for rethinking medium access control”, IEEE/ACM Transactions on Networking, October 2011.

[SMP+14] B. Soret, et al., “Fundamental Tradeoffs among Reliability, Latency and Throughput in Cellular Networks,” IEEE GLOBECOM, December 2014.

[SSJ+13] A. Galindo-Serrano, B. Sayrac, S. B. Jemaa, J. Riihijärvi, and P. Mähönen, “Harvesting MDT data: Radio Environment Maps for Coverage Analysis in Cellular Networks,” in 8th International Conference on Cognitive Radio Oriented Wireless Networks, 2013, pp. 37–42.

[STR+15] S. Sun, T.A. Thomas, T.S. Rappaport, et al. “Path Loss, Shadow Fading, and Line-Of-Sight Probability Models for 5G Urban Macro-Cellular Scenarios,” IEEE Globecom Workshops (GCW), Dec. 2015.

[STS+14] K. Sakaguchi, G.K. Tran, H. Shimodaira, S. Nanba, T. Sakurai, K. Takinami, I. Siaud, E. Calvanese Strinati, A. Capone, I. Karls R. Arefi, T. Haustein, “Millimeter-wave Evolution for 5G Cellular Networks,” arXiv: 1412.3212 [cs.NI], 11 Dec 2014.

[USRP] Universal Software Radio Peripheral (USRP) B210: https://www.ettus.com/product/details/UB210-KIT.

[VHM+14] V. Venkatasubramanian, M. Hesse, P. Marsch, and M. Maternia, “On the Performance Gain of Flexible UL/DL TDD with Centralized and Decentralized Resource Allocation in Dense 5G Deployments,” in IEEE PIMRC, pp.1840-1845, 2-5 Sept. 2014.

[WHY14] R. Wang et al., “Potentials and Challenges of C-RAN Supporting Multi-RATs toward 5G Mobile Networks,” Access, IEEE, vol.2, no., pp.1187, 1195, 2014.

[WINNER08] “WINNER II Channel Models”, WINNER II, D1.1.2, 2008.

[WN+03] D. Wu and R. Negi, “Effective capacity: A wireless link model for support of quality of service,” IEEE Transactions on Wireless Communications, vol., 2, no. 4, pp. 630-643, July 2003.

[ÖAS+15] P. Ökvist, H. Asplund, A. Simonsson, et al., “15 GHz Propagation Properties Assessed with 5G Radio Access Prototype”, IEEE PIMRC, 2015.

[YLJ+15] M. Yang, Y. Li, D. Jin, et al., “Software-Defined and Virtualized Future Mobile and Wireless Networks: A Survey”, ACM Journal Mobile Networks and Applications-2015.

A Annex

A.1 T5.1-TeC3 5G user-centric interference management in UDNs *(DR-1)*

A.1.1 Evaluation Methodology of Case Study 1 and Case Study 2

For the first case study, system-level Monte Carlo simulations were performed to evaluate the performance in the proposed scenario with different number of activated NNs (1-5) and. 25 UEs were randomly dropped in a hotspot area at the edges of the macro BS coverage area. For the simulation set-up, we used the Madrid Grid deployment (see Figure A-1) and radio channel models from 3GPP [3GPP10-36814] (UMa for macro BS and UMi for NNs). Macro BS operates at 2 GHz carrier frequency, while the NNs utilize spectrum at 3.5 GHz with full frequency re-use. Both access node types operate with 20 MHz bandwidth available. Ideal backhaul is assumed for the NN-macro BS links (with 100% LOS probability) with half duplex operation and round robin algorithm used for scheduling.

In order to benefit from NN operations new inter-cell resource management schemes are needed for the coordination of the access nodes in dynamic radio topology. Also, the backhaul link measurements and activation commands may possibly imply new signaling elements on the wireless backhaul link.

Figure A-1 Visualization of user-centric interference management in Madrid Grid deployment considering dynamic radio topology based on NNs.

In the second case study, a dynamic system-level HetNet simulator is employed [SBS+17]. Pico cells and Vehicular NNs are deployed on top of a wrapped-around hexagonal grid consisting of 19 tri-sectored macro BSs. A section of the considered network is shown in Figure A-2. The simulator also features the enhanced inter-cell interference coordination (eICIC). This coordination involves two controllable parameters which are the almost blank subframes (ABS) ratio of the macro BSs and the cell range extension (CRE) bias of the pico BSs or the NNs. ABS ratio is the percentage of muted subframes at the macro BS to reduce the interference on DL and CRE allows the deployed small nodes to extend their coverage by increasing the offset value which increases the attachment probability of the UEs in order to offload macro BSs. The selected parameters for the experiments are provided in Table A-1 [SBS+17]. Furthermore, the uplink throughput gains for different number of access node activations are provided in Figure A-3. It can be seen that by activating the NNs, it is possible to provide a high capacity gain of 183% at the lower (10th) percentile of the user throughput CDFs, compared to the pico cell deployment.

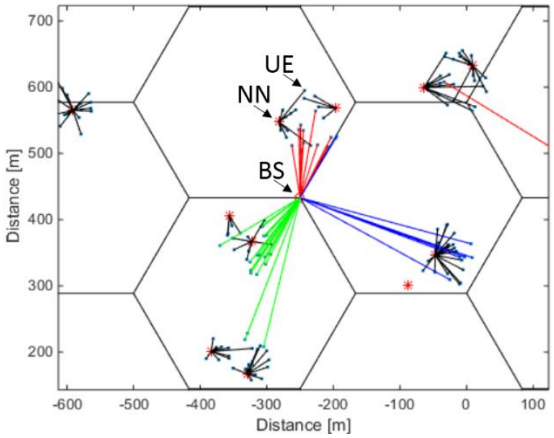

Figure A-2 A section of the network showing the links between the UEs (blue dots) and the NNs (red stars) as black lines, and the links between the UEs and the macro BSs (red circles at the center of each site) as colored lines.

Table A-1 Simulation Parameters

Feature	Implementation
Network Topology	Wrapped-around hexagonal grid of 19x3=57 macro cells; randomly placed pico cells and VNNs, the latter being 20 per macro cell in average
UE Layout and Load	Randomly dropped hotspots, 1 per macro cell in average, containing 25 and 50 UEs in UL and DL, respectively; indoor UEs (20 dB penetration loss)
Inter-Site Distance	500 m
System Bandwidth	20 MHz and 10 MHz in UL and DL, respectively; centered at 2.6 GHz; FDD
Frequency Reuse	1
eICIC Parameters	CRE offset of 12 dB and ABS ratio of 25% and 50% in UL and DL, respectively [SBS+17]

Traffic Type	FTP
Scheduler	Proportional fair
Shadowing	Log-normal shadowing fading with standard deviations 8 dB macro BS to UE, 10 dB pico BS to UE and 7 dB VNN to UE; Shadowing decorrelation distance of 50 m
Tx Powers	Macro BS: 46 dBm; Pico BS: 30 dBm; VNN: 30 dBm; UE: max 23, min -40 dBm with UL power control
Antennae	Gains: macro BS 14 dBi, pico BS 5 dBi, VNN 5 dBi and UE 0 dBi; Heights: macro BS 32 m, pico BS 5 m, VNN 1.5 m and UE 1.5 m
Receiver	1x2 Maximal Ratio Combiner
Modulation	QPSK, 16QAM and 64QAM

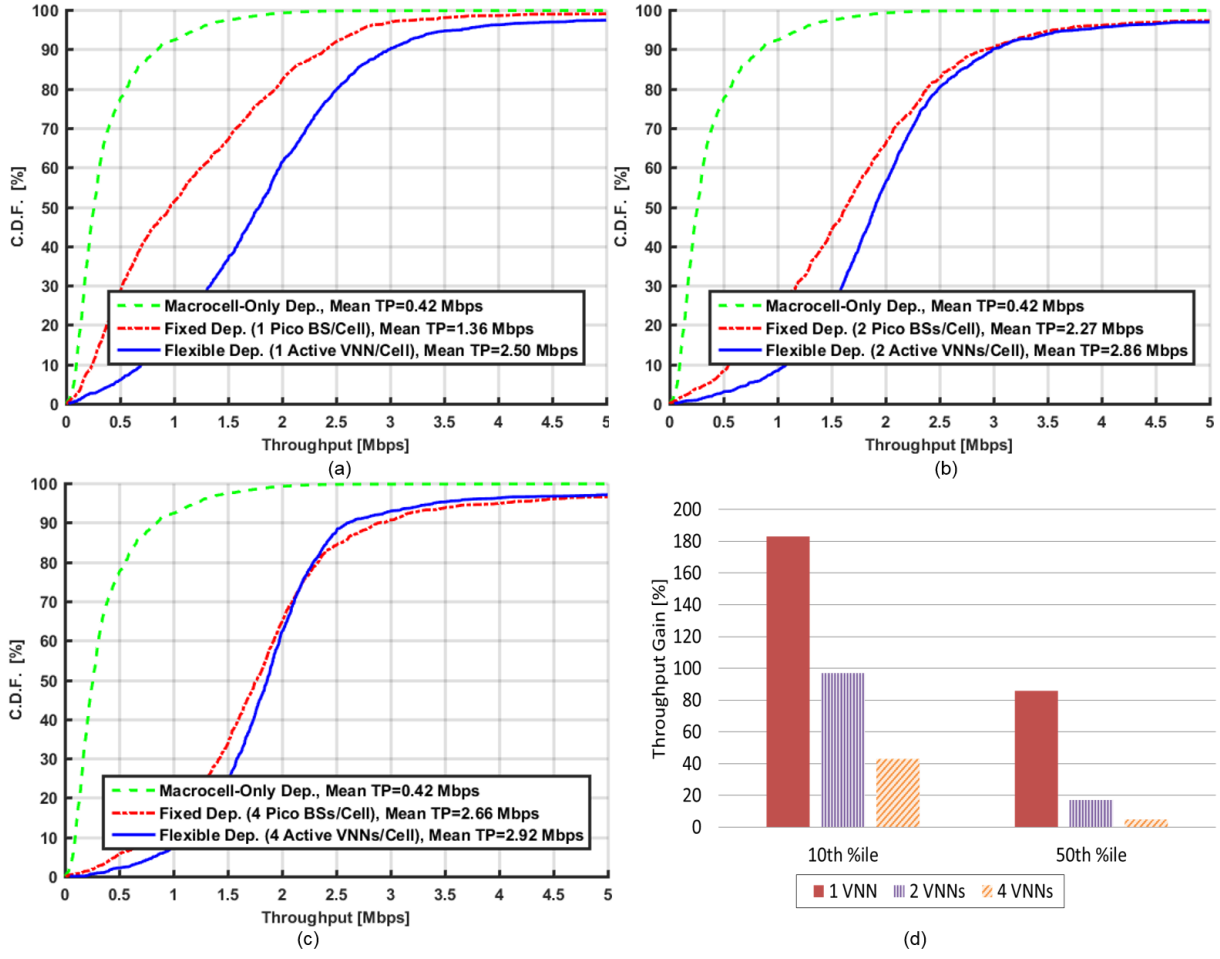

Figure A-3 User throughput CDFs on uplink with different deployment scenarios (a), (b), and (c), together with the throughput gain of the VNN deployments compared to pico cell deployments at the 10th and the 50th percentiles of the CDFs as a column chart (d). Mean user throughput (Mean TP) levels are provided in the legends (case study 2).

A.1.2 Dynamic NN Selection Strategies

The wireless backhaul links of the NNs provide a means for flexible deployment. Further, the backhaul link quality plays a crucial role for the end-to-end performance especially when an in-band half-duplex operation is employed. The advantage of the in-band half-duplex operation is the low complexity and cost-efficiency.

The considered system model is depicted in Figure A-4. In the exemplary illustration, there are three candidate VNNs available in the target service region, i.e., three vehicles are parked in the region and can be activated to serve the UEs in the proximity of the active VNN on the access link (VNN-UE link). It is worth noting that the target service regions can be determined by the operators, e.g., based on UE density. In order to take into account the uncertainty for the availability of VNNs, we utilize the parking lot model given in [BRZ+14], which is based on continuous-time Markov chains. We consider a parking lot where a maximum of $M_{\max}$=5 places are available on a line road. Further results pertaining to $M_{\max}$=15 and 25 are presented in [BKE+17]. The distance between two nearby VNNs is taken as 6 m. Moreover, we set the parking lot model parameters, e.g., departure and arrival rates, such that a regular day time is simulated.

The backhaul and access links are modelled by Nakagami-lognormal and Rician-lognormal composite distributions, respectively, which are the two common models in the literature [SA05]. Interfering signals on the backhaul link are assumed to be subject to Rayleigh-lognormal (a.k.a. Suzuki) composite fading/shadowing. Further, a single UE is connected to a single VNN on the access link and is communicating via this VNN with a BS. The considered network is represented by a regular hexagonal layout with seven macro cells. The rest of the system parameters are in line with [BRZ+15]. Besides, the simulations are conducted using MATLAB as the computational environment.

NN selection strategies take into account the backhaul link qualities at different candidate VNNs towards the available BSs. At a given time instant, there are M available VNN candidates in cell k out of which we select the VNN m^* and associate it with the BS k^* such that downlink SINR γ on the backhaul link is maximized as

$$\gamma_{m^*,k^*}^{\text{opt}} = \max_{m,k}\{\gamma_{m,k} : m = 1, 2, \ldots, M \;\&\; k = 1, 2, \ldots, K\}$$
$$\text{subject to } M \le M_{\max} \tag{A.1}$$

Accordingly, the serving BS may not necessarily be the closest BS to the candidate VNNs. For instance, the closest BS may be shadowed due to a large obstacle and, thus, a neighbor BS may provide the best backhaul link conditions. Therefore, analyses are conducted such that location and serving BS selections are jointly optimized, and the impact of serving BS selection is highlighted. Further, in case of optimum VNN selection, shadowing and multi-path fading are considered, while in case of coarse VNN selection only shadowing is factored in. That is, optimal selection takes into account short-term changes in radio conditions, whereas coarse

selection is focusing on the long-term radio conditions. Consequently, the optimal selection requires more frequent channel quality indications to be sent.

The cumulative distribution function (CDF) plots of end-to-end rate (in terms of bps/Hz) are illustrated by in Figure A-5 for M_{max}=5. The end-to-end rate is the resultant rate on the two-hop relayed communication, where it is equal to the lower rate of those on the backhaul and access links for a given UE. The Nakagami fading parameter is set to one (Rayleigh fading) to simulate more severe fading characteristics on the backhaul link along with a shadowing standard deviation of 8 dB. Thus, the channel model parameters are the same for the backhaul link and direct link (BS-UE link). Random VNN selection is taken as reference. It is first noticed that VNN selection is vital because without VNN selection (see random selection), the VNN performance becomes worse than that of the direct link due to half-duplex constraint. When serving BS is jointly determined with the VNN selection, clear gains can be observed, where these gains are higher in case of optimal VNN selection (compare single BS and multi BS). In case of multi BS (Optimal) the performance can be improved by 40 times at 5%-ile rate level compared to the direct link. VNN performance is worse than the direct link as of 80%-ile; yet, by optimal cell selection the shown hull curve performance can be approached.

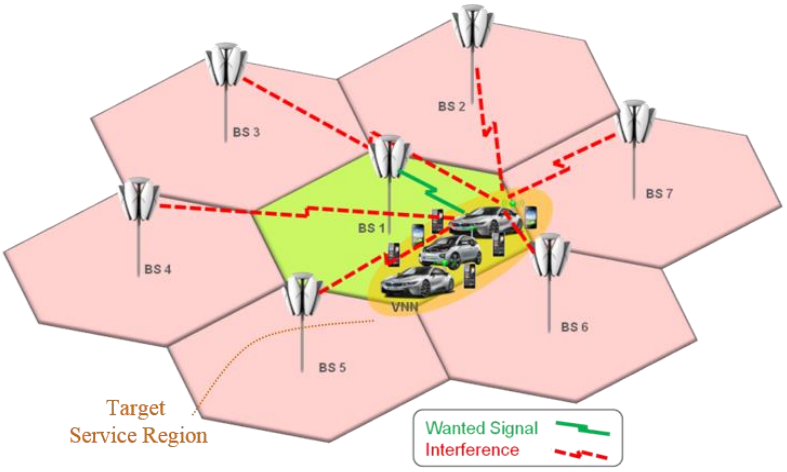

Figure A-4 The considered system model with UEs, VNNs and macro BSs. The active VNN to serve the target service region and the associated serving BS are selected based on the backhaul link SINR.

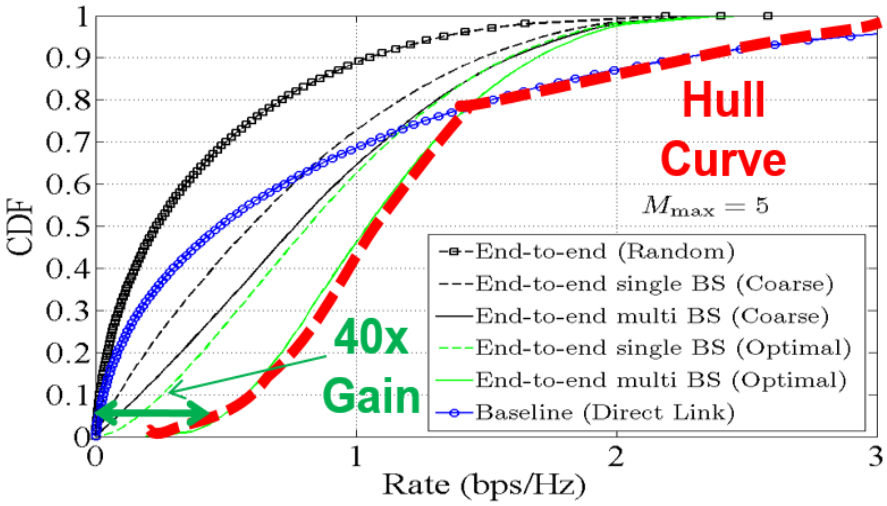

Figure A-5 End-to-end rate gains with different NN selection strategies. The Hull Curve indicates the upper bound of achievable gains when the cell selection for the UE (BS or VNN) is optimized. Single BS implies that the activated VNN is served by the midmost BS only, whereas multi BS implies that VNN and serving BS selections are performed jointly.

A.2 T5.1-TeC-8 Flexible Interference Management for 5G AIVs *(DR-2)*

Interference management via FQAM-based frequency partitioning can be simply achieved by allocating a dedicated spectrum subband to FQAM transmissions in interfering cells where the UEs in victim cells will be served from. An efficient and agile resource management strategy on interference management can then be applied to enable interference control between clusters of mutually interfering cells (or users therein). To achieve the benefits of FQAM in the cell edge of victim cells while maintaining high throughout in interfering cells, low-SINR users are scheduled from a flexible and adaptive reserved resource pool, negotiated between neighboring cells.

Our algorithm presented below implements an efficient and flexible resource management strategy on top of FQAM to enable fast yet flexible overhead interference control between clusters of mutually interfering cells (or users therein). The steps of this algorithm are as outlined below:

1. Per target cell, users are split into high-SINR versus low-SINR ones; this can be done e.g. per TTI or on longer intervals.
2. Low-SINR users in a target cell need an active interference management from interfering cells to improve the performance.

3. The interference management can be realized by utilizing FQAM in neighboring interfering cells on resources (in frequency domain) that are reserved for low-SINR users in a target cell.
4. The way of applying different modulation schemes to different frequency ranges can additionally be affected by the type of synchronization / co-ordination that exists between cells in question.

For the sake of clarity, Figure A-6 illustrates the concept of frequency partition of FQAM, where users with low SINR (i.e. high interference) are allocated the FQAM resources.

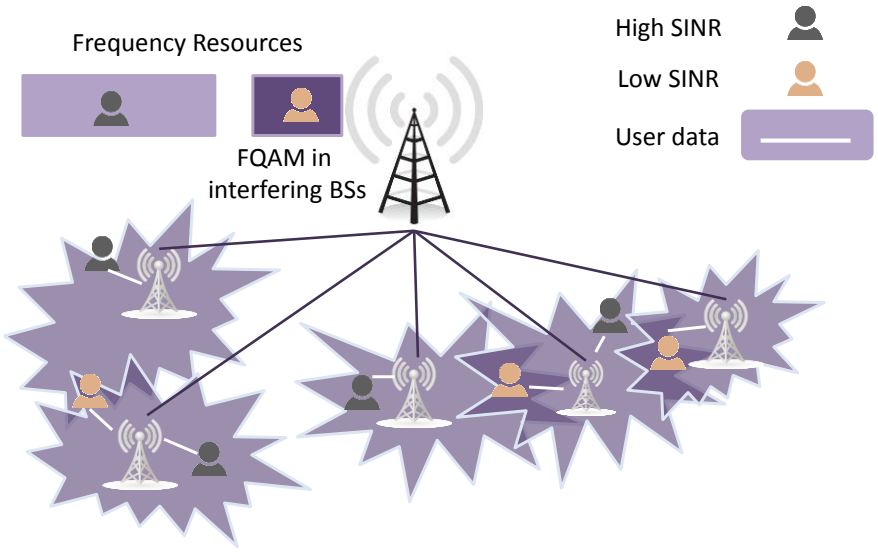

Figure A-6 Frequency-based FQAM for flexible interference management.

Results for frequency-based FQAM are shown in Figure A-7. In comparison of space versus frequency FQAM shown in Section 4.4.1 (although the scenarios are not directly comparable) space partitioning requires additional antennas while frequency partitioning requires a larger system bandwidth. However, the results point to a more spectrally efficient use of the spatial dimension for FQAM, as average transmission rates for spatial partitioning are indeed larger than in frequency partitioning. This is due to the large synergistic benefits of employing beamforming and FQAM simultaneously.

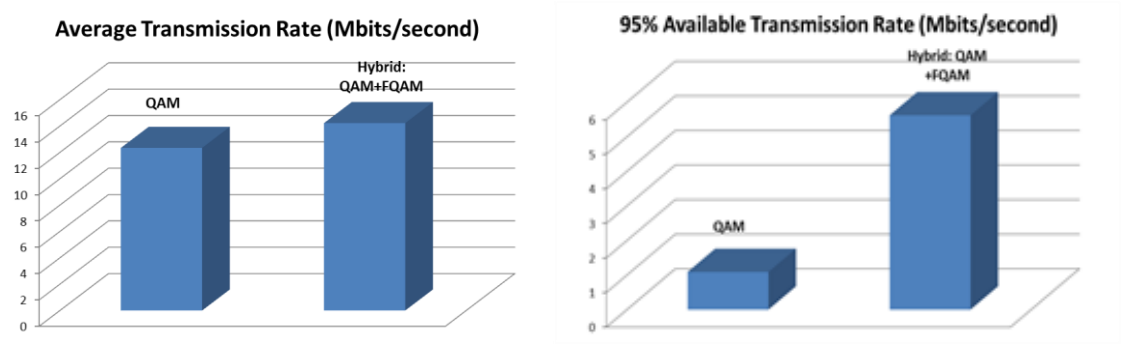

Figure A-7 Frequency-based FQAM rate results.

A.3 T5.1-TeC15 Joint Transmission with Dummy Symbols for Dynamic TDD (*DR-3*)

A.3.1 Further Results

In this section, we provide supplementary results to the ones given in Section 4.1.2 for DR-3 by considering an open indoor office scenario with a 4x4 grid BS deployment. Due to the short-range, cell radius is set to 5 meters, resulting in an inter-site distance of 10 meters. The motivation for considering this scenario is to study how the proposed scheme copes with additional interference as the network size in terms of area and number of deployed BSs grows. For this supplementary scenario, we therefore exclude walls that may otherwise attenuate the interference. In addition, we elaborate on certain trends in the results that become more evident for larger network sizes.

As we shall see, the number of UL BSs that participate in the precoding will inflict a trade-off between DL and UL performance. For simplicity, let V denote the maximum number of UL BSs that can participate. One can control the number of participating UL BSs W with the parameter δ such that $W=\max\{0,V-\delta\}$. We study the effect of this below.

Figure A-8 shows the average UL and DL sum-rate, respectively, as a function of system utilization. The lowest utilization point (0%) reflects an interference-free environment with a single user, and is therefore omitted here. Because the interference is lower at low utilization, the sum-rate may drop initially when more interference is suddenly introduced.

As a starting point, we begin by considering the case when $\delta=0$. It is shown that, at low utilization, both JT and JT-DS provide substantial performance gains in the DL thanks to the nulling of BS-to-UE interferences, but is limited by (uncontrolled) BS-to-UE interferences. In the UL, JT-DS significantly improves sum-rate by including UL BSs in the precoding which reduces BS-to-BS interference, but similar to the DL case it is constrained by uncontrolled UE-to-UE interference.

As utilization and thereby UL (and DL) traffic increases, so does unmitigated UE-to-BS and UE-to-UE interference. In the UL, more UL BSs are thus able to participate in the precoding. As a

result, UL sum-rate increases in part thanks to more UL traffic, and in part thanks to the nulling of more BS-to-BS interference. At some utilization point however, DL demand will be high enough resulting in fewer participating UL BSs. Despite this, an UL performance gain is still achievable by including UL BSs corresponding to worst performing UL UEs.

In DL, however, performance of JT-DS starts to diminish as adding more receivers also increases ill-conditioning of the precoder matrix. As a result, precoder elements may become overly large, and transmit powers need to be lowered to compensate for the difference in order to not violate the BS power constraint. While this is observed for both JT and JT-DS, the difference is more pronounced for JT-DS as the participating UL BSs do not contribute in increasing DL sum-rate, in addition to the more severe ill-conditioning. DL performance can therefore become a bottleneck for JT-DS at higher traffic loads. On the other hand, the lower DL powers help reduce BS-to-BS interference and improve UL performance.

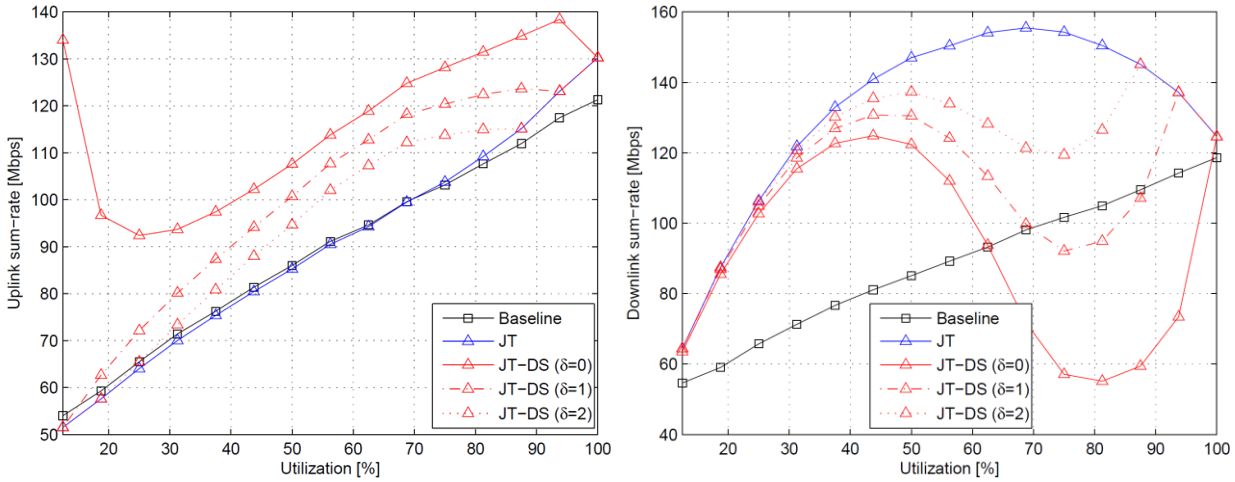

Figure A-8 Average UL (left) and DL (right) sum-rate

Because of the ill-conditioning, we expect a trade-off between DL and UL performance for JT-DS when varying the size of the precoder via δ. When utilization is low, the effects of increasing δ are especially noticeable in the UL where the number of UL BSs is already small due to exceedingly low user diversity. This implies that few if any UL BSs will be able to participate in the precoding even for small values of δ. In DL, however, the size of the precoder is still fairly small and the effects of ill-conditioning less severe compared to high utilization regime. In contrast, at high utilization even small changes in the number of participating UL BSs make a large difference to DL performance. At the same time, it is evident that including fewer UL BSs in JT-DS will effectively increase BS-to-BS interference to UL BSs excluded from the precoding and lower overall UL sum-rate.

A.4 T5.1-TeC14 Flexible TDD design for NR *(DR-4)*

The following is an explanation of Figure 4-9: Assuming pilot contamination exists in the system due to pilot reuse, each table cell represents a transmission path (DL or UL) in one single data slot. The table is structured as two sets of two rows each, each set (PCR-D and PCR-U) representing the communication path over which the pilots for channel training were sent: DL in case of PCR-D, and UL in case of PCR-U. Furthermore, each row in each set represents a cell's TDD configuration during data transmission slots, where the S rows represent the serving cell and I rows represent the interfering cells during training phase. The columns identify which of the TDD configurations correspond to a well-managed interference case (RCR) that avoids beamformed interference and which correspond to a beamformed interference case (ICR). The division by quadrants of Figure 4-9 visually allows a prioritized classification of the different possible TDD modes. RCR is obviously preferred over ICR for the pilot contamination related reasons previously stated. Within RCR, the PCR-D mode (yellow) does not require any modification of the transmit powers while PCR-U (green) requires additional CLI interference management to counteract the TRP-to-TRP interference characteristic of reverse TDD mode. In the case of ICR, the PCR-U mode (orange) suffers from beamformed interference but it is preferred over PCR-D (red) as the latter adds TRP-to-TRP interference on top of pilot contamination interference. Hence, the TDD configuration of the cells in a HetNet must be set following this colored priority order, also pointed by the arrow: 1) yellow, 2) green, 3) orange, and 4) red.

In the following preliminary evaluation, a simple case of two cells operating under pilot contamination is shown where one cell acts as the serving cell and the second cell as interferer. A two-base station scenario is assumed with one user each sharing the same pilot sequence. The serving base station is equipped with a very large array of antennas. The simulation is restricted to two time slots: A training phase followed by a data transmission phase. Trying to assess the performance difference between RCR and ICR when pilot contamination exists, it is assumed that the serving base station gets its channel estimate contaminated by interfering pilots carried in the downlink (PCR-D). Then, the signal-to-interference ratio (SIR) is measured during the data transmission slot at the receiving ends, namely the interfered user in the case of downlink transmission, and the serving base station in the case of uplink transmission. The results are displayed in Figure A-9. The SIRs (y-axis) are measured for different contamination ratios (x-axis), where the contamination ratio is defined as the quotient between the received serving power and the received interfering power during the training phase.

The main conclusions that can be extracted from the graphs are as follows. Clearly, the pilot contamination effect degrades SIRs both in the downlink and uplink when beamforming and combining are respectively employed. More interestingly, selecting the RCR configuration over ICR greatly increases the SIR of the downlink and uplink transmissions. Furthermore, the contamination ratio plays an important role: When the power level of the contamination is high, the beamformed interference experienced at the users increase, hence enlarging the SIR gap between RCR and ICR. This observation is crucial when designing a TDD configuration for a

HetNet since beamformed interference coming from high-power elements is much more dangerous than the interference coming from low-power elements. In summary, it can be observed that the design of the TDD configuration is a critical parameter to control interference in massive MIMO systems.

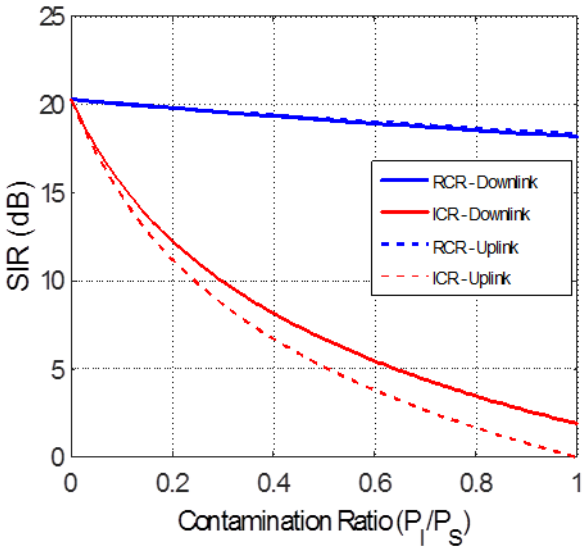

Figure A-9 Pilot contamination regime analysis.

A.4.1 TDFLEX Heuristic

We propose a heuristic solution to the problem of configuring the TDD frames of a HetNet called TDFLEX. The problem can be described as follows: Given a separate dynamic load for each cell, we need to dynamically fill out the time slots of the TDD frames with the objective of minimizing the interference caused by the pilot contamination effect. Assuming training occurs in the first time slot of every frame, a main constraint is the need for an SU mode during the training slot of the massive MIMO MBS. Then, the rest of training slots at every small cell must be set to either SU or SD. Furthermore, data slots need to be filled with either U (uplink) or D (downlink) modes to minimize pilot contamination. The TDFLEX is a low-complexity heuristic that dynamically matches the load distribution while minimizing the number of ICR collisions C. For that, it induces either PCR-D or PCR-U modes, whichever is more suitable, with the purpose of minimizing the impact of the pilot contamination effect. The working principle of TDFLEX is as follows: By sorting the U and D slots, the number of collisions both in PCR-D and PCR-U modes can be easily computed given the load. In particular, Eq. (A.2) provides the number of collisions for a given load distribution if PCR-D is chosen, while Eq. (A.3) does the same in the case of PCR-U. In both equations, the superscript M represents the macro cell and S the contaminating small cell. In addition, power considerations are taken into account to favor PCR-D when no collision exists while favoring PCR-U if a collision is unavoidable. Once the computations are performed and the slots filled, columns of data slots in the A matrix can be rearranged if certain slot sorting is preferred. The complexity of the algorithm is O(LN), i.e.,

linear in the product of number of cells times number of time slots within a subframe. This means that a huge complexity reduction is achieved in comparison with the exact optimization problem.

$$C^{PCR-D} = N - [\min(n_D^S, n_D^I) + \min(n_U^S, n_U^I)] \quad \text{(A.2)}$$
$$C^{PCR-U} = N - [\min(n_D^S, n_U^I) + \min(n_U^S, n_D^I)] \quad \text{(A.3)}$$

The complete TDFLEX algorithm can be found in [WCNC17]. The main steps of TDFLEX are summarized as follows:

- Calculate n_U and n_D for each cell.
- Fill out the pre-sorted macro cell TDD row of A.
- For each small cell:
  - Calculate collisions with PCR-D using Eq. (A.2). If B2B interference appears, discard.
  - Calculate collisions with PCR-U using Eq. (A.3).
  - Select mode with lesser number of collisions, prioritizing PCR-D for equal Cs as less power is needed.
  - Fill out sorted data slots accordingly.
  - If PCR-U is selected, enhance uplink power in U slots under RCR.
  - Rearrange data slot columns if desired.

A.4.2 Further Results

Finally, we show a very interesting effect observed in the downlink of our massive MIMO-enabled HetNet: A very different performance dependence on the number of antennas for the two presented TDD designs, namely TDFLEX and TDLTE. The results are shown in Figure A-10. The most important observation that can be extracted from these results is the difference in the evolution of the rate distribution curve when the number M of antennas at the MBS is increased and beamforming is present. If the pilot contamination effect is well managed via TDFLEX, the users' rates benefit from the increase of the number of antennas. However, if the underlying TDD architecture does not account for this effect such as the standard TD-LTE, the increase of antennas can be counterproductive and greatly damage the attainable rates of users subject to receive beamformed interference in the downlink.

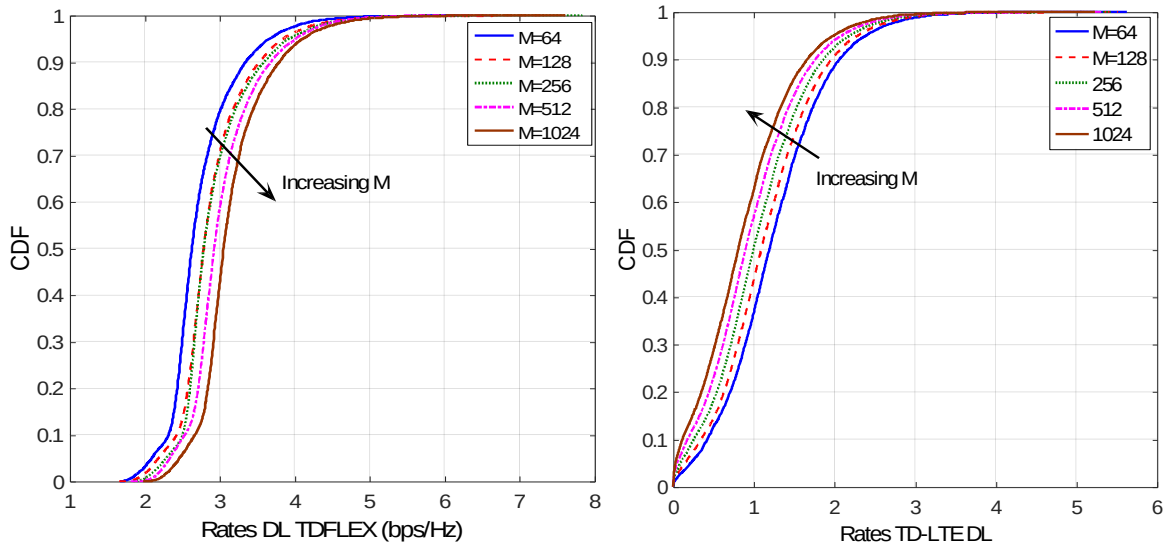

Figure A-10 Rate distributions for different number of antennas at the MBS.

A.5 T5.1-TeC11 Interference Coordination / Cancellation Strategy (*DR-5*)

A.5.1 Detailed description

This design recommendation is based on the clustering of base stations, accordingly with the presence of UEs being in the coverage area of several base stations, and therefore susceptible of being interfered by them.

The main idea is to create a mechanism to orthogonalize the signals transmitted from a specific time/frequency zone grid from all the base stations in the cluster. This time/frequency grid zone will be determined for a number of subcarriers and a number of symbols duration, and will contain all the symbols used for the UEs at base station edge, i.e. UEs that could be interfered by one or several base stations in the cluster.

The names of the values used in this TeC are:

M as the number of base stations composing the cluster to be coordinated,

N as the number of subcarriers in which the orthogonalization process will be carried out,

T as the number of time slots (in the time/frequency grid) in which the orthogonalization process will be carried out.

The TeC is based on applying to the complex symbols, in the selected time/frequency grid of the OFDM like signal, a precoding composed by the application of a scrambling and spreading mechanism. This precoding is based on a scrambling pattern (common to all base stations in a cluster) and one spreading patterns for each base station in the cluster, so that they are

orthogonal between each two base stations in the cluster. The spread however takes place over time in different frames, not as CDMA which uses frequency spreading, therefore allowing to keep low the total amount of bandwidth assigned to low SINR UEs in each TTI.

The spreading mechanism is based on M·T-length DFT of the complex symbols which are selected in the time/frequency grid as graphically shown in Figure A-11. The complex symbols originally scheduled for the UEs in the base station edge (namely a[i,n]) are spread over M consecutive TTIs, being the M·T symbols transmitted named as b[i,m] obtained as:

$$b[i,m] = M \sum_{n=0}^{T-1} a[i,n]\, e^{\left[-j2\pi(M\cdot n+k_i)\frac{m}{(M\cdot T)}\right]} \quad \text{(A.4)}$$

Being "i" the carrier frequency and k_i natural numbers $0< k_i < (M-1)$, comprising the spreading pattern of each specific base station. Therefore, the spreading pattern are orthogonal coordination vectors, one for each of the M base stations included in the coordination cluster. These coordination vectors are composed by N pseudorandom values, each one being any value from 0 to M-1, i.e., the M coordination vectors are (TRxP: Transmission Reception Point):

$$k_i^1 = [k_0^1, k_{12}^1, \ldots\ldots, k_{N-1}^1] \quad \text{TRxP 1 (A.5)}$$

$$k_i^2 = [k_0^2, k_{12}^2, \ldots\ldots, k_{N-1}^2] \quad \text{TRxP 2 (A.6)}$$

$$\ldots\ldots\ldots$$

$$k_i^M = [k_0^M, k_{12}^M, \ldots\ldots, k_{N-1}^M] \quad \text{TRxP M (A.7)}$$

The orthogonality is achieved by selecting patterns that guarantee that:

$k_i^a \neq k_i^b$ for different base stations a≠b , and for any value of subcarrier i

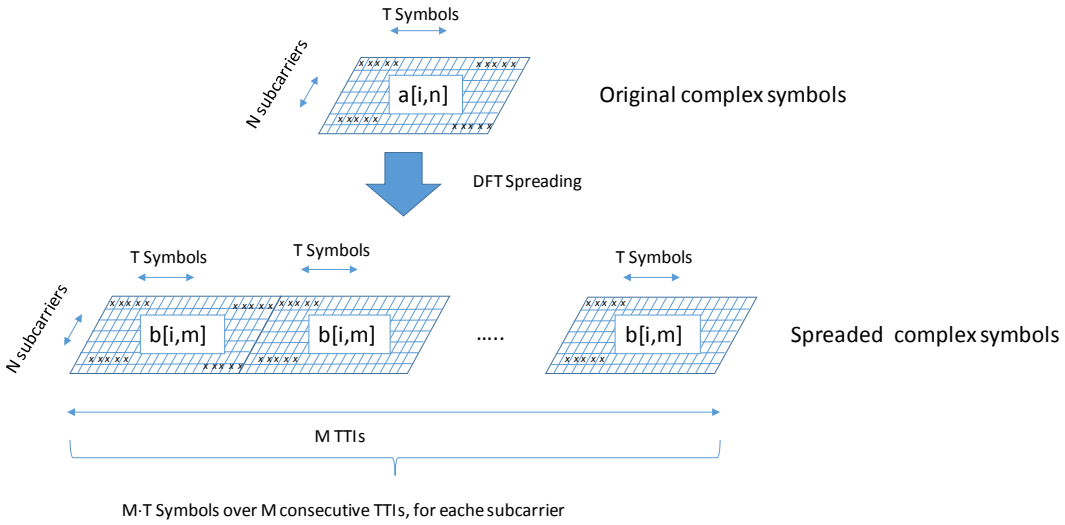

Figure A-11 Complex symbols spreading over successive TTIs

It should be taken into account that the number of base stations of the cluster M is identical to the number of TTIs in which the original complex symbols are spread by the DFT, and therefore the additional latency included.

After obtaining the spread complex symbols, a scrambling operation given by the randomizing pattern common to all the base stations within the coordination cluster, randomizes these symbols over M consecutive TTIs in the time domain for each subcarrier. This scrambling operation (based on a pseudo-random pattern characteristic of each coordination cluster) avoids the occurrence of the periodical pattern associated with the DFT nature of the spreading operation.

A.5.2 Evaluation assumptions

Link-level simulations are conducted for evaluation accordingly with the description in this report, being the main details of these simulations:

Table A-2 Simulation Setting for DR-5

Parameter	Setting
Carrier frequency	2.6 GHz
Bandwidth	20 MHz (1200 active subcarriers)
Subframe structure	LTE-like subframe, 14 OFDM symbols, no control information (only payload data)
Cluster size *M*	2, 4 and 8
Interference scenarios	Intra-cluster (scenario 1); inter-cluster (scenarios 2/3); and baseline

	(scenario 4)
Desired channel model	AWGN and 3GPP EPA (3 km/h)
Desired modulation and coding rates	QPSK 1/3; 16QAM 2/3; 64QAM 4/5
Interference channel model	AWGN
Interference modulation	QPSK
Detector type	SISO MMSE, downlink
Channel estimation	Ideal
HARQ	Not present

A.5.3 Further Results

This design recommendation has been analyzed for different SNR and SINR scenarios, for different MCS from the base station and for different number of base station clustering, being the main outcomes showed in the following table.

Table A-3 Simulation Results for DR-5

Channel type	Modulation and coding	*Cluster size*	Min. SINR gain (dB)
AWGN	QPSK 1/3	2	2.98
		4	5.99
		8	8.95
	16QAM 2/3	2	2.6
		4	5.58
		8	8.55
	64QAM 4/5	2	0.61
		4	3.3
		8	6.35
3GPP EPA3	QPSK 1/3	2	2.4
		4	3.37
		8	5.34
	16QAM 2/3	2	1.11
		4	2.94
		8	4.46
	64QAM 4/5	2	0.66
		4	2.87
		8	4.31

A.6 T5.1-TeC4 Support LAA for unplanned and Dynamic Radio Topologies *(DR-6)*

A.6.1 Evaluation Methodology

For the evaluation, system-level Monte Carlo simulations were performed to evaluate the performance in proposed scenario with different number of activated NNs (1-5) and 25 UEs were randomly dropped in a hotspot area at the edges of the macro BS. The NNs utilize spectrum at 3.5 GHz with full frequency re-use. Both access node types operate with 20 MHz bandwidth available. Ideal backhaul is assumed for the NN-macro BS links and round robin algorithm is used for scheduling.

Below, we show a case study where 5 NNs are deployed in licensed mode, and given the access conditions and WiFi APs (which are randomly activated at each snapshot) in this area we decide to activate in unlicensed mode (5GHz band) a subset of them only if the unlicensed carrier is sensed clear. For the channel models are the simulation scenarios, 3GPP-compliant parameters were used (outdoor LAA scenario 4, in [3GPP15- 36.889].

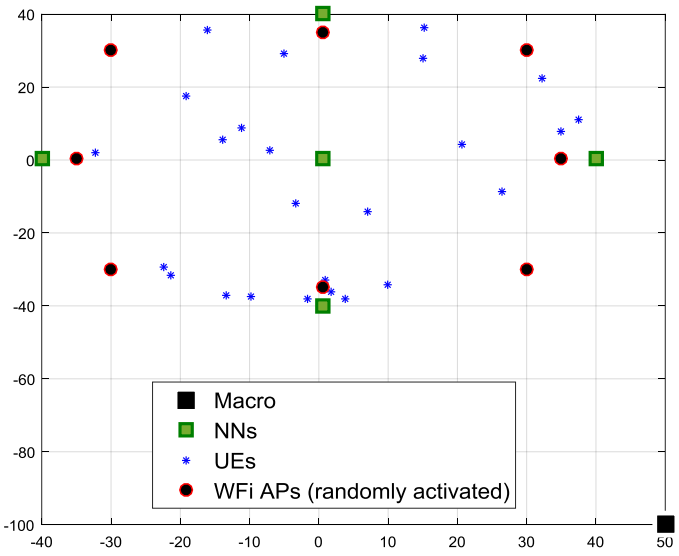

Figure A-12 Deployment

One of the key challenges in LAA is to ensure fair co-existence of LTE in unlicensed bands with other technologies like WiFi. To this end, mechanisms like Listen-Before-Talk (LBT) could effectively enable fair co-existence with WiFi and allow for efficient LAA operation. With LBT, the BS can more dynamically occupy the channel based on the detected medium status, which can both alleviate the delay issue and effectively balance the channel occupancy among co-existing transmitters. For the NNs and UEs we performed Energy Detection (ED) using the threshold of -70dB to ensure that no interference is created to potentially activated WiFi access points at the

scheduled frames. Here, no hidden terminal problem is assumed since the central coordinator is aware of the locations of the WiFi access points (which can be either active or in-active). Therefore, for the scheduling period it was assumed that no activated WiFi APs are in close vicinity. To support this, we assume that the WiFi AP locations are known by the macro which performs the decisions on LAA.

A.7 T5.1-TeC10 Dynamic cell switch off (*DR-7*)

A.7.1 Evaluation Methodology

At present two different simulators are available for performance evaluation:

- A simulator developed in MATLAB to study resource allocation issues, where link level performance and effects of higher layers are highly simplified in order to reduce both complexity and simulation time. This was used to evaluate scenarios defined in METIS (a simplified Dense Urban Information Society and the Stadium Test Cases [MET14-D32]).
- A simulator developed in C++ for LTE-Advanced simulations where channel effects, link layer, and higher layer aspects are modeled with higher detail (implementing classical 3GPP-like scenarios for both homogeneous and heterogeneous networks, see [3GPP10-36814]).

Both simulators are semi-static [ART10-D51], meaning that the positions of active users are modeled by a random uniform distribution, are fixed over a simulation run, and multiple simulation runs are carried out in order to collect significant statistics of the behavior of the system. Even if users are not moving during one simulation run, some degree of mobility is included in the system in the form of fast fading on the channel. Full buffer and non-full buffer traffic sources are available in both cases.

A.7.2 Detailed Analyses and Further Results

Models for the overall BS power consumption, based on the amount of radio resources used in transmission were firstly proposed in the EARTH project [EARTH]. According to these power models, the overall energy consumption of a macro or micro BS increases in first approximation linearly with the total amount of radio resources that are used for transmission. According to the analysis in [EAR12-D23] this part changes dynamically with traffic load and such variation is mainly due to the power amplifier and base band modules in the BS. The remaining elements in the BS (antenna interface, RF small-signal transceiver, DC-DC power supply, active cooling system, and AC-DC power supply unit) usually show an almost constant power consumption regardless of the traffic load, representing a static power consumption portion in the overall power consumption of the node. In some BSs a power saving mechanism is available, so that when no transmission is performed, the BS can enter in a sleep-mode that further reduces its

consumption. This is shown with a discontinuous point in the Figure A-13 for 0% resource usage.

It is expected that future transmission nodes will be able to scale their consumption, based on the actual amount of served traffic, in a more efficient way than nowadays system does. The 5GREEN project [5GREEN] estimated that an improvement of 8% every year can be achieved in the dynamic part of power models, so that the overall power consumption will scale more significantly with the actual radiated power every year, and also “sleep” mechanism in the nodes will become more and more efficient. Figure A-13 shows the power models for 2010 nodes based on the EARTH project, and compares them with the 2020 power models proposed in 5GREEN.

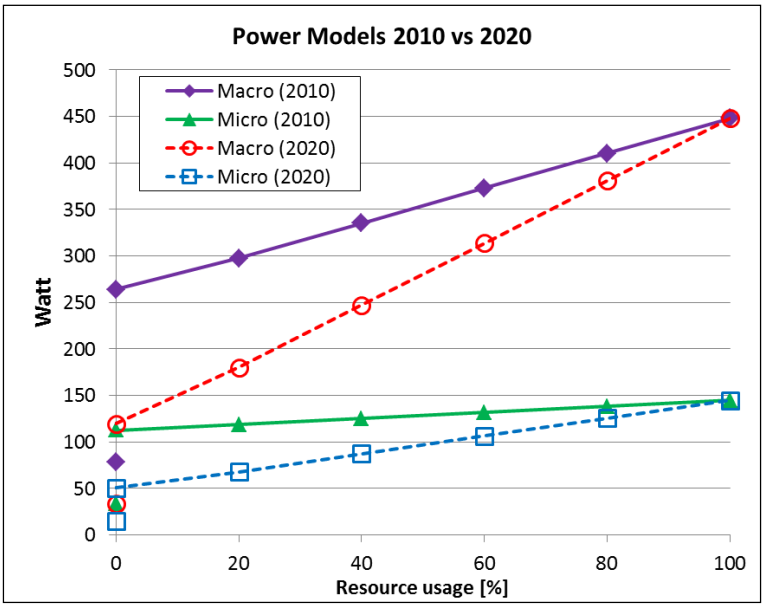

Figure A-13 5GREEN power models for macro and micro BS in 2010 and 2020.

Using these power models, it is possible to evaluate the overall energy consumption of a system when different traffic loads, and consequently different amount of radio resources are considered. In [MET15-D33], 2010 power models were used to evaluate the improvement in energy consumption that could be achieved using the proposed centralized coordination scheme. Figure A-14 shows the impact of power models suitable for 2020 transmission nodes considering the proposed scheme, comparing them to results that were reported in [MET15-D33] with 2010 power models. Results are obtained under the simplified Madrid Grid scenario described in [MET14-D32]. This scenario comprises 3 macro BS and 9 micro BS, with 10 users connected to each BS. A signal bandwidth of 10 MHz is assumed, and different traffic loads have been evaluated using CBR traffic sources, which generate data at a given rate for each user. As a further reference also the traditional full buffer traffic condition has been simulated. Power consumption when no coordination between nodes (NoCoord) is exploited is compared with results obtained assuming the centralized scheduler that exploits Joint Transmission (JT)

and Dynamic Point Selection/Dynamic Point Blanking (DPS/DPB) for Energy Efficiency (referred to as EE JT in the figure) here discussed.

As shown in Figure A-14, the higher dynamicity of future nodes can be even better exploited by the proposed scheme, which is able to deliver energy savings up to 51%, while only up to 27% savings were achieved with 2010 power models.

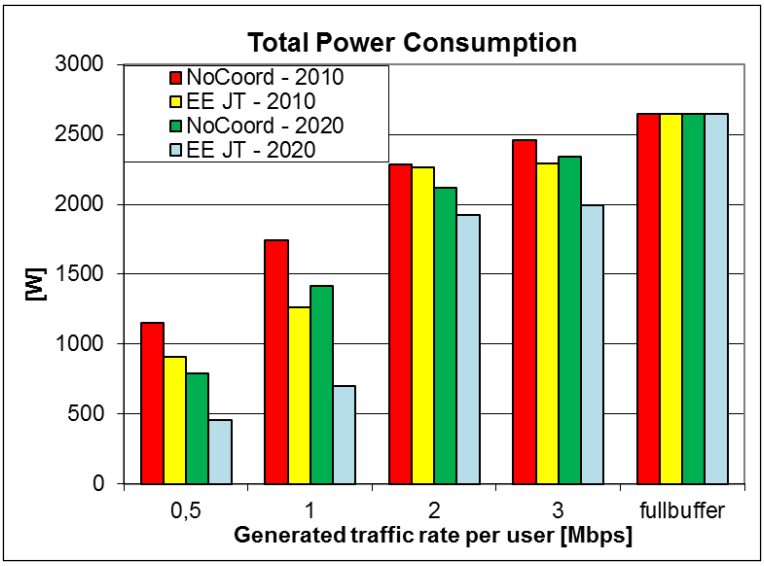

Figure A-14 Comparison of power consumption with and without coordination with 2010 and 2020 5GREEN power models.

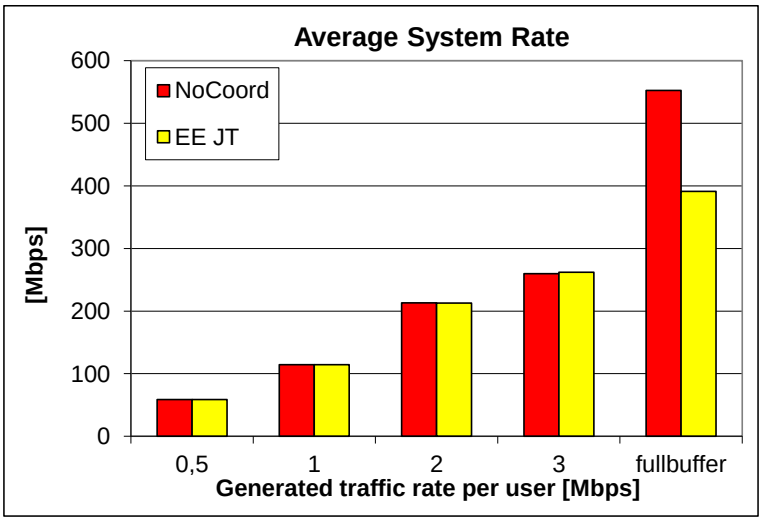

Figure A-15 Average system rate achieved with the proposed solution.

Note that, as shown in Figure A-15, the main focus of the proposed scheme is to reduce energy consumption through the reduction of active nodes, and this could be done with no impact on the achievable rate as long as the amount of traffic is not too large. In full-buffer condition,

where all the transmission nodes should be active in order to face the large traffic request, the proposed scheme becomes sub-optimal and should not be used.

METIS-II recently proposed similar power models for 2010 and 2020 equipment [MET-II16-D21] as shown in Figure A-16. According to METIS-II power models, moving from years 2010 to 2020, not only the equipment will have better scalability of the energy consumption as a function of resource usage, and more effective sleep mode states, but the overall energy consumption at full resource usage will also be drastically lower.

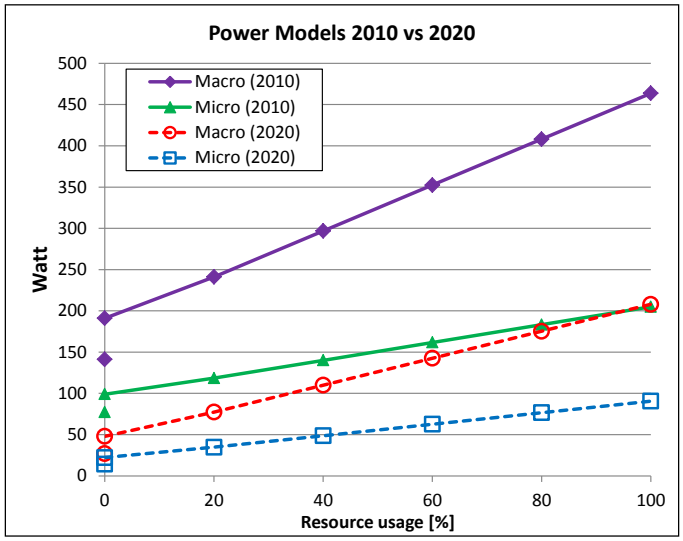

Figure A-16 METIS-II power models for Micro and Macro Base Station in 2010 and 2020.

Simulation results considering METIS-II power models are shown in Figure A-17. As it was expected also in this case the higher energy efficiency of 2020 equipment reflects in a drastic reduction of the power consumption, both with and without the centralized entity for coordination. In this case, also the portions of power consumption due to the static consumption part in the power model, and that due to the dynamic part, are shown through different colors in the bar representing the overall power consumption. The higher dynamicity in power consumption that METIS-II power models show the better it can be exploited with the proposed solution, so that in 2020 the power consumption reduction that can be achieved using the EE JT scheme can be as high as 51%, whereas in 2010 only savings up to 31% could be achieved.

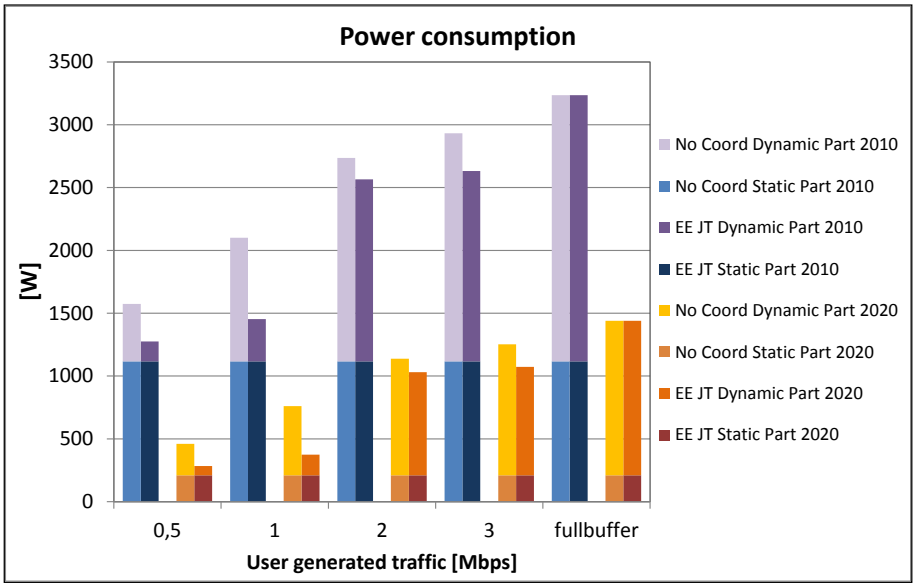

Figure A-17 Comparison of power consumption with and without coordination with 2010 and 2020 METIS-II power models.

A.8 T5.1-TeC12 Coordinated Sleep Mode for self-Backhauling (*DR-7*)

A.8.1 Evaluation Methodology

The requirements for higher data rates and ultra-reliable communication in 5G are expected to be enabled using ultra-dense deployment of 5G-base stations. The 5G-gNBs are expected to support higher bandwidths in order to support extreme mobile broadband use cases and is also expected to operate at higher frequency bands (centimeter / millimeter wave, cmW / mmW) where such higher bandwidths would be available. Having such dense deployment of 5G-gNBs and related backhaul links would lead to a significant increase in the network power consumption, which should be avoided. Lean carrier design is a key design enhancement in 5G, where the idea is to not have always-on signals any more, but rather rely on optimized control signaling. This would enable better energy saving strategies in gNBs whereby with loads less than 100 %, the gNBs can enter micro-discontinuous transmission (DTX) modes, thereby saving valuable power.

Self-Backhauling (sBH) technology is yet another key enhancement that is currently being studied in 5G, where the sBH-gNBs would have dedicated backhaul links (Uu*) with a traffic aggregation node / donor gNB, sharing the same wireless channel as the access link. The aggregation node would then have a direct link with the 5G CN over the S1* link providing connectivity to the sBH nodes. sBH could improve the deployment efficiency of the 5G networks, with potentially unplanned and random deployment of the BSs. The join optimization

of gNBs and sBH is considered in this technology component as a novel enabler for energy efficiency in 5G through intelligent RAN moderation.

A.8.2 Detailed Analyses and Further Results

The gNB operation is as shown in Figure A-18, based on the work presented in [PUM17]. Depending on the DL buffer status, with various QCIs requiring different priority and handling, and based on the UL buffer status reports (BSRs), the 5G-gNB determines the sleep pattern it can support for energy efficiency, while supporting the QoS requirements of the served traffic. The gNB then sends this information to the BH aggregation node or sBH link. If the gNB has a dedicated BH link / node, then the BH node also follows the same sleep pattern as the gNB in order to achieve high-energy savings. If the BH traffic aggregation node is serving multiple gNBs, then it evaluates the various potential patterns and negotiates with the gNBs to select the pattern, which enables the highest amount of power savings at the BH node.

The negotiation could be done depending on the QoS of the active traffic, minimizing the potential delay that the energy saving actions could cause. There could also be preconfigured energy saving targets at the 5G-gNBs depending on the time-of-day, location, etc. Once the control plane signaling is done, and the sleep patterns are finalized, the DRX configurations are done on the UE, in order to synchronize the sleep cycles between all the nodes involved in the network. The gNB could reject the sleep mode reconfiguration request from the aggregation node if the additional delay involved would affect the QoS of the served traffic. Another criterion for the configuration accepting / rejecting could be the control / discovery signaling periodicity of the gNB. Since the sleep modes are assumed to be configured based on the gNB dependent discovery and control signaling periodicity, the synchronization of sleep signals should not affect this mechanism, since it would have impacts on the mobility and discoverability of the gNB.

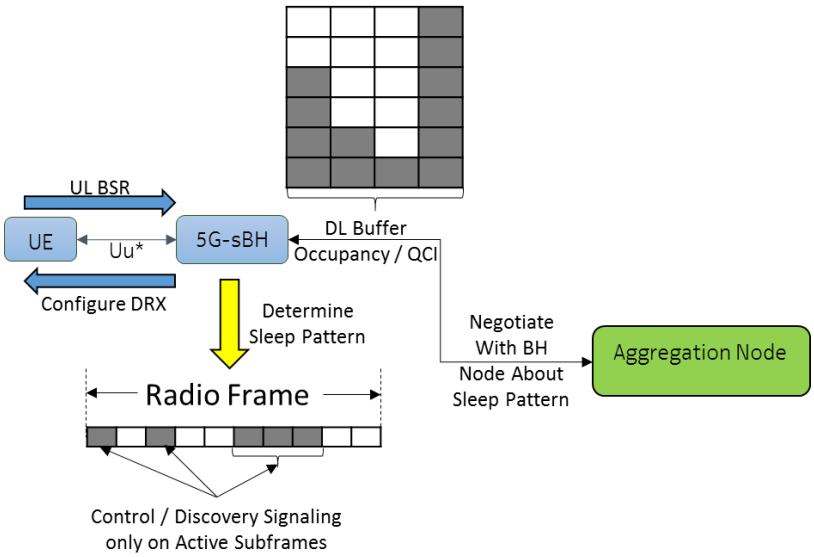

Figure A-18 5G-gNB Operation during Low Load Conditions [PUM17].

Figure A-19 presents the functionalities / signaling involved for the RAN-BH joint operation mechanism for energy savings over the possible Un* interface.

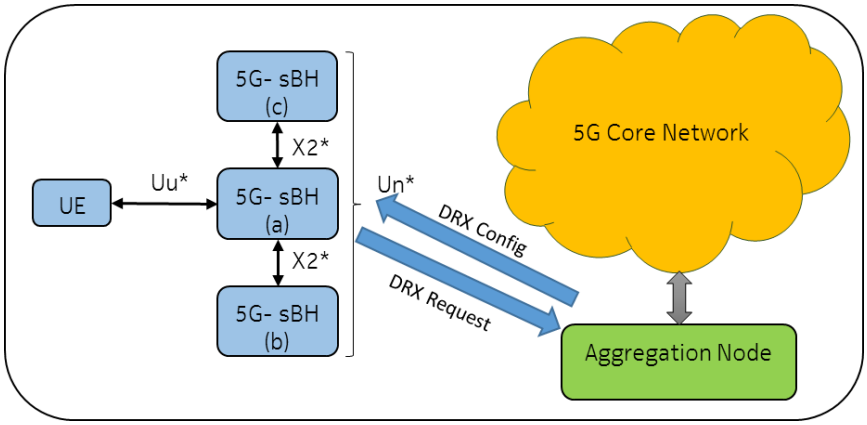

Figure A-19 Functionalities / Signaling involved in case of RAN-BH joint operation.

A.9 T5.1-TeC6 RM for Multi-cell Coordination for Ultra-Dense Network Employing Dynamic TDD (*DR-7*)

A.9.1 Evaluation Methodology

This section considers a dynamic TDD network employing non-coherent JT/JR illustrated in Figure A-20. In order to investigate achievable gains, the evaluation framework employs exhaustive search with the objective to maximize system performance. By considering both a fixed and random BS deployment, we are also able to take into account the effects of a more uneven interference distribution and its effect on the proposed scheme.

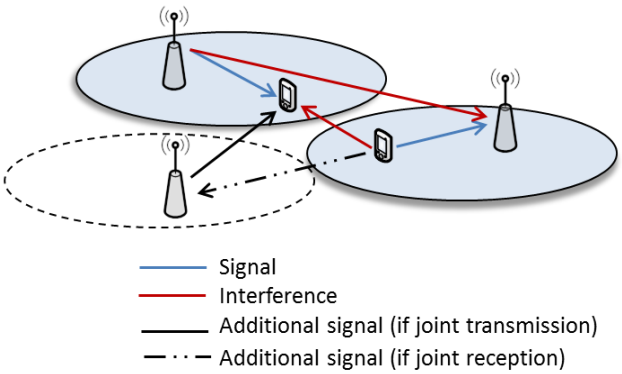

Figure A-20 Multi-cell dynamic TDD network employing JT/JR.

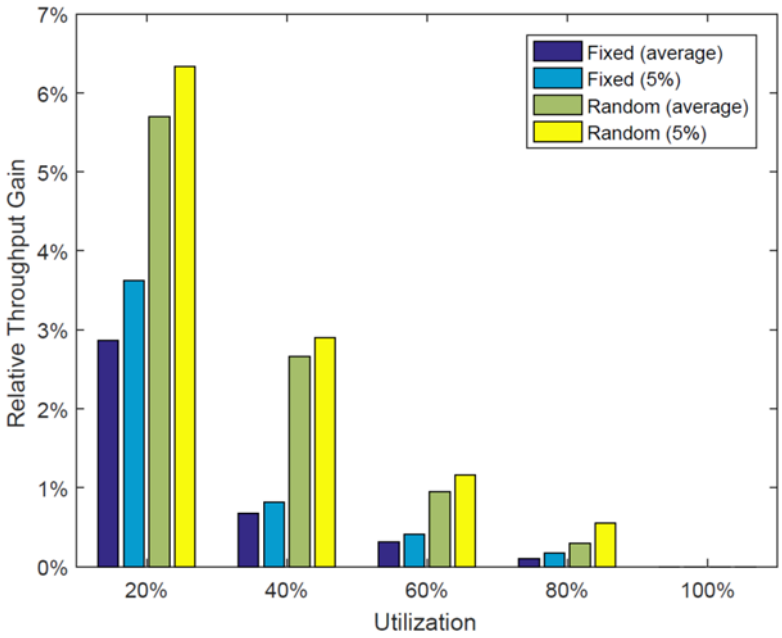

Figure A-21 Relative throughput gain for the considered scenario.

Relevant KPIs include throughput and energy efficiency. Throughput is calculated using a modified version of Shannon's formula [MNK+07, Eq. (3)] that takes into account bandwidth (Beff=0.84) and SINR inefficiencies (SNReff=1.25) when the SINR is between -7dB and 30dB.

A.9.2 Detailed Analyses and Further Results

The evaluation scenario is based on the BS deployment and 2D geometric layout of the virtual indoor office scenario described in [MET-II16-D21], accounting for walls but not chairs or tables. Channel modeling therefore assumes the WINNER II A1 indoor path loss model [WINNER08], which includes shadowing between all entities. Operating frequency is set to 2 GHz. For a given snapshot, users are dropped uniformly over the office area constrained on that each small cell serves no more than a single active user. In total, 100 snapshots are generated for the ensemble averaging. While multiple users can be associated with the same BS, only of one of them is assumed to have traffic demand and therefore called active. For active users, queues are infinitely backlogged so that bits are always available for transmission. Furthermore, we define system utilization as the ratio between number of active users and number of BSs.

The cell radius of UDNs is typically much shorter than traditional outdoor networks, 10-100 meters, which allows BSs and users to employ similar transmit powers when cell radii are in the lower end of that range. This evaluation is based on the power models in [DDL15] for femtocells with 250 mW transmit power and sleep mode 4 in idle mode, compared to 100 mW transmit power for user devices. Once scheduled, a BS or UE transmits at full power. Single antennas with no beamforming capabilities are assumed for both BSs and UEs to keep hardware costs low. It is noted that in UDN where the cell radius can be small, the notion of cell edge user can be interpreted more loosely.

As shown by the results in Figure A-21, the power gain is able to improve average system throughput by 3-4%, though gains diminish quickly as traffic increases and more interference is generated. Somewhat larger gains are attained for worst (5th) percentile performance, especially at low utilization where throughout increase is about 6%. The proposed scheme can therefore help to mitigate the effects of an instantaneously bad propagation environment to these UEs. At higher traffic loads, potential gains become inferior to the increase in interference. At full traffic load (100% utilization), no gains can be made as there are no more idle BSs to add. So, while non-coherent JT/JR provides some gains in low utilization regime, more sophisticated schemes are needed to achieve meaningful gains at higher utilization.

As a consequence, we observe in our evaluation that energy efficiency does not improve with non-coherent JT/JR for the use case considered here. For brevity, we do not show those results there. Based on this, we infer that additional interference management techniques are needed in order to motivate powering on idle BSs and improve both throughput and energy efficiency.

A.10 T5.1-TeC1 Multi-dimensional RM for 5G & Legacy AIs (*DR-8*)

A.10.1 Proposed Approach for RRM for Network Slicing

The basis for allocating resources in a slice aware manner is monitoring the current status of the network slices with respect to their SLAs. AaSE has to be aware of the existing network slices and their SLAs, as well as which data stream belongs to which network slice. This can be realized through signaling from the CN.

The enforcement of the network slice specific requirements happens with the help of existing QoS mechanisms of the 5G AIVs. Based on the outcome of the SLA monitoring, the QCIs of the individual data streams are adjusted. If, for example, the SLA of a network slice guaranties a data rate of 1 Mbit/s per data stream, any data stream could be mapped to a corresponding QCI class. This mapping is a dynamic process which is supposed to solve conflicts between network slices in a way that all SLAs can be fulfilled.

Figure A-19 visualizes this process. The basis for a slice aware RM is that data flows from the CN are tagged with either information on the corresponding network slice or the corresponding SLA. An entity called Air interface agnostic Slice Enabler (AaSE) is responsible for monitoring and enforcing SLAs. As stated before, this could be part of an access controller. Based on the information from the CN and from the AIVs (e.g. from AIV specific schedulers), AaSE monitors the SLA status. An enforcement of SLAs happens by adapting the QoS classes of individual data streams. For example, a data stream from a network slice with high data rate guarantees can be configured to have a QoS class of a specific AIV with a guaranteed bit rate. For monitoring the SLA status, AaSE reads QoS KPIs of the AVIs. A feedback to the CN (whether SLAs are currently fulfilled) is important to monitor SLA status also there as well as to trigger network changes in case of constant SLA violations.

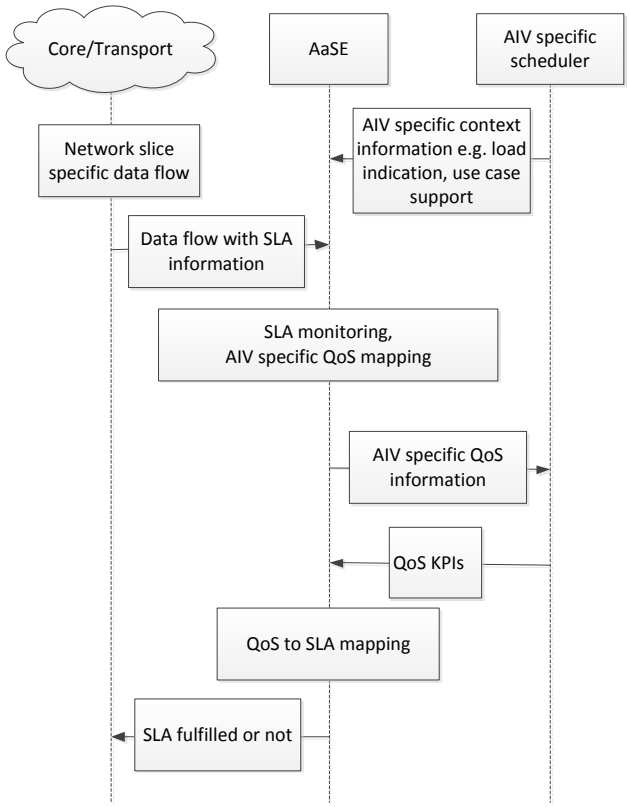

Figure A-22 Implementation option for RRM for network slicing

A.10.2 Simulation assumption for Figure 4-20

Table A-4 Simulation assumptions for Figure 4-20

Parameter	**Setting**
Scenario	Madrid grid
Number of users	810
User drops	3
Snapshots	180 000 of 1ms
System Bandwidth	20 MHz / 2 * 10 MHz for the two subnetworks
Performance estimation	Through lookup-tables from LTE-A simulator
UE speed	3 km/h
Offered load	2.2 Mbit/s per UE

A.11 T5.2-TeC3 Slice-oriented AI-agnostic Resource Abstraction for 5G RAN *(DR-8)*

A.11.1 Detailed Analysis

In this section, mechanisms are presented that decide on how to form the clusters, what is the role of the nodes in each cluster (e.g. RRM Controller or Slave AP) and who is forming the clusters (e.g., management entity (e.g., network management entity (NMS)) based on physical deployment, long term statistics of the load of APs and slice characteristics or KPIs). In CU, the formation of appropriate clusters is selected. Subsequently, the level of centralized RRM is decided for each access node, given the RAN limitations and the level of slice awareness. The level of centralization is translated as a flexible split of RRM functions, which can be slice-tailored and cell-specific. The heterogeneous split of RRM functions will provide new requirement for signaling between the APs. For example, as mentioned in state-of-the-art, by centralizing only slower RRM functions like IM and Load Balancing (LB), signaling should be exchanged for the resource restrictions and cell re-selections between the centralized and distributed nodes for the dynamic resource allocation. In case of having distributed allocation of IM and LB, e.g. due to slice requirements for fast IM, we need to exchange new messages regarding the dynamic resource restrictions in order to allow for centralized LB (taking into account and the other RRM Splits). In this section, we propose some low complexity heuristic approaches, with low signaling cost to solve these sub-problems.

Cluster and Controller Selection

As mentioned above, we try to find sets of feasible solutions (e.g. multiple maximal cliques) to form the clusters. From the set of maximal cliques, we aim to find nodes with the highest occurrences to become candidates for RRM Controllers. The proposed heuristic algorithm is presented below.

Below, in Figure A-23, the flowchart is illustrated:

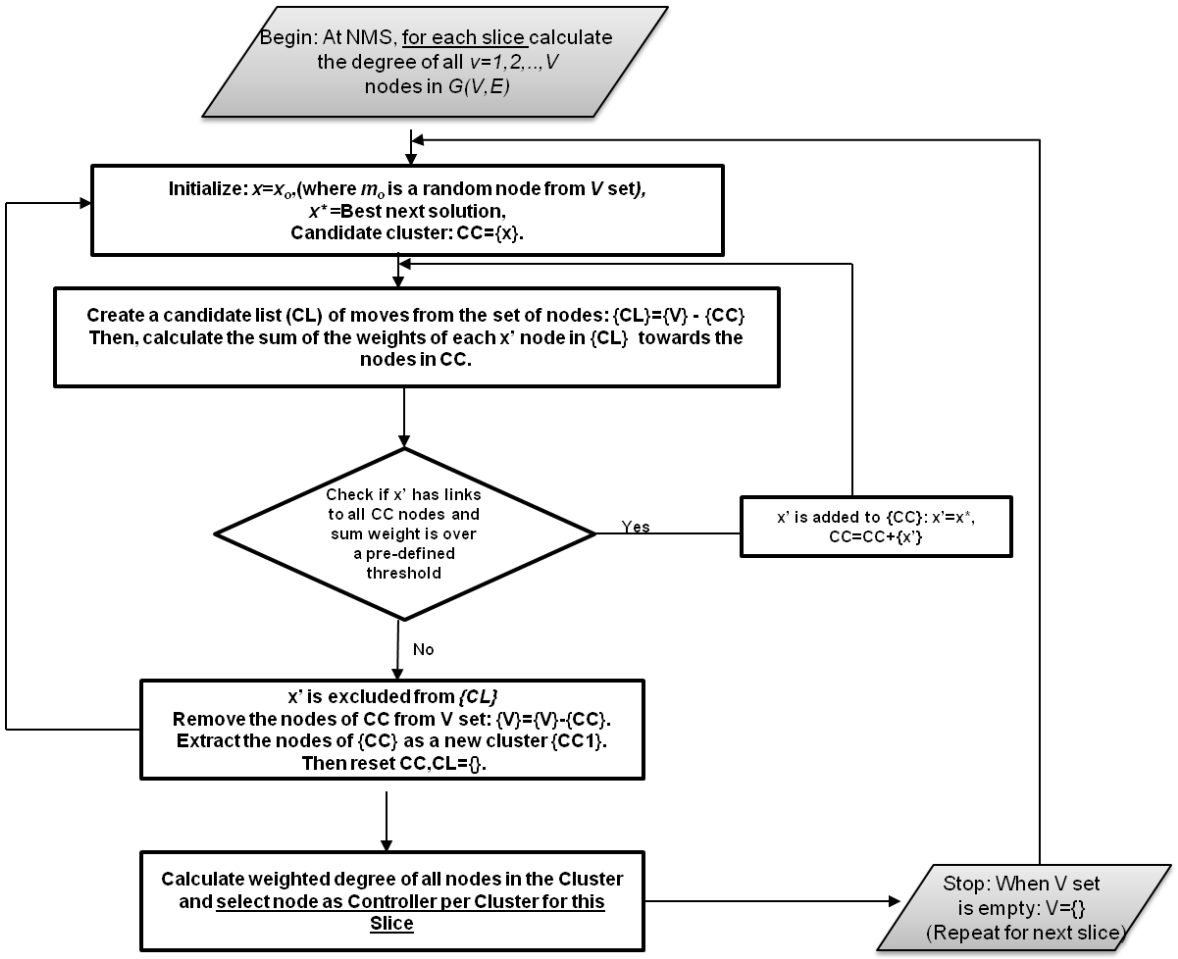

Figure A-23 Flowchart for Controller and Cluster Selection

RRM Split Configuration

In this solution, we aim to find what the best split is for each controller-slave AP pair based on the parameters as mentioned above. Here, by taking into account the BH constraint, the per slice preference on certain split and the load of the controller, we decide whether to use Split A (centralized), Split B (semi-centralized) or Split C (distributed). Below, in Figure A-24, we briefly show the flowchart for this selection.

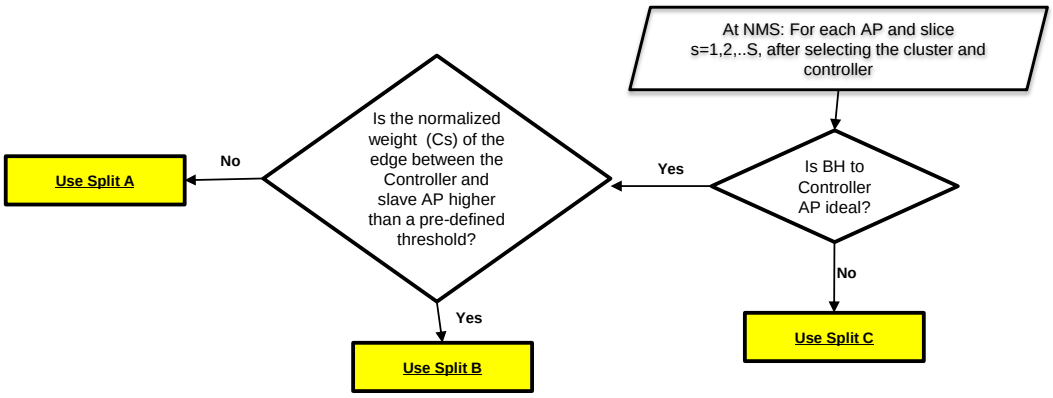

Figure A-24 Flow-chart for RRM option configuration

An example system, where 2 clusters (complete sub-graphs) consisting of APs, which are selected as the cliques with the minimum cardinality, can be seen in Figure A-25. In this case, we select one controller per cluster and the RRM Split. In Cluster 1, we can observe that Slow RRM (e.g. Cell Selection) happens centrally, whereas the Dynamic RRM is performed in distributed way. In similar manner in Cluster 2, the Controller performs centralized RRM for some APs, whether one AP can also perform dynamic RRM.

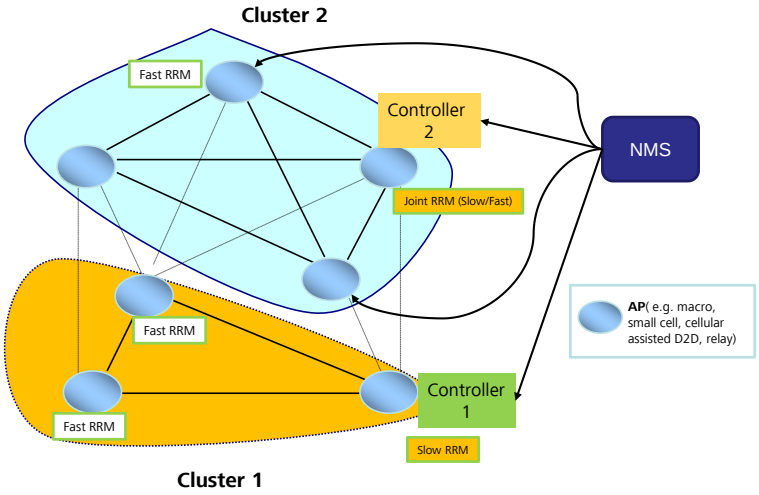

Figure A-25 Exemplary scenario with 2 Clusters and 2 Controller

A.11.2 Evaluation Methodology

System level Simulations were performed to show the tradeoff between Centralized and Distributed Interference Management. The deployment is a cluster of 9 APs using 3GPP LTE as baseline for our simulations (40 users uniformly distributed, 3GPP UMi channel, ideal BH). In case of Centralized RRM we perform Centralized CoMP (coherent JT), in case of Split B we

perform only centralized eICIC and in case of Split C we perform single-cell scheduling in each AP without interference management.

At first, a CU selects the RRM Controller and the cluster size, the characteristics of the physical nodes (e.g., BS, (non) ideal BH link, available spectrum) in case of slice support, the KPIs of every slice, and also, based on long term statistics for load, BH conditions per deployed slice. The Network Management System (NMS) according to the previous data will decide on an Initial RRM Split. NMS configures each AP as on its cluster membership, its operation mode as an RRM Controller (or simple member), and which initial RRM split will be used between the AP and the RRM Controller.

Initially, an event (e.g. slice instantiation request) triggers the action from management entity to decide on how to form the clusters/who is forming the clusters (e.g., CU based on physical deployment, long term statistics of the load of BSs) and which RRM Split to be used. This can be configured by CU. The selection of cluster, controller, and RRM split for a certain slice, will divide the total set of access nodes into orthogonal AP clusters. Thereafter, network will provide one or more access nodes as RRM Controller candidates based on the following parameters: AP general processing capabilities, average load information, number of neighbouring APs with good / ideal BH, slice KPIs. Then, based on the slice requirements different RRM controller candidates can be mapped to different slices. In Figure A-26, we can also observe the message sequence chart for the process. In addition to the typical operation between the APs and CU, which involves the feedback of measurement and long term statistics periodically, the Controller assignment message is forwarded from CU to the Controller AP and the cluster member assignment is then forwarded to other APs. Here, two new messages can be defined for the controller assignment and cluster notification.

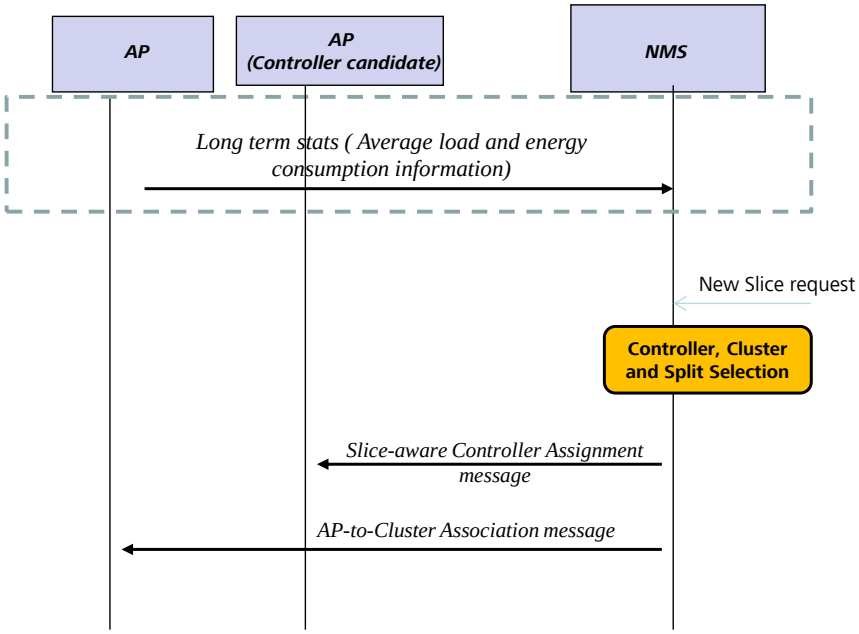

Figure A-26 Message Sequence Chart for Controller and Cluster Selection

A.12 T5.2-TeC5 Flexible Multi-Service Scheduling Framework (*DR-9*)

A.12.1 Further Analyses and Results

In this section, additional results to those presented in the main body are included. The study focuses on the impact of the traffic type, protocol used (e.g. TCP), file size, system load and user channel conditions on the selection of the optimal TTI size to schedule a user in a particular instance. The simulation methodology and assumptions are similar to those presented in the main body, and for more details the reader is referred to [PNS+16]. The modelling of TCP follows the Reno model [PFT+00]. When a TCP packet (with the maximum segment size of 1500 B) is generated at the traffic source, it is subject to a CN latency of 2 ms before arriving at the BS. The corresponding TCP acknowledgment (ACK) from UE in the UL is transmitted with the same TTI size as in the DL. Conveying the TCP ACK from the BS to the traffic source is again subject to the CN latency. The traffic model follows a Poisson arrival process with file sizes of 50 kB and 500 kB, and variations in the total offered load are simulated to assess the end-user throughput.

The evaluation results shown in Figure A-27 indicate that for low offered loads, with both small file size downloads (dominated by the TCP slow-start phase) and with large file downloads, the best performance is achieved with a short TTI that minimizes the round-trip time (RTT) of TCP ACKs/negative ACKs (NACKs). Higher system load means higher inter-cell interference, which

requires extra CCH overhead (i.e. more redundant encoding to guarantee scheduling grants reliability), that especially impacts shorter TTIs (larger relative CCH overhead) and UEs in less favourable radio conditions. In addition, higher loads imply longer queuing delays at the BS, which makes longer TTIs with higher data spectral efficiency a more attractive option. For large file sizes, as the load increases, it is more efficient to schedule with a short TTI during the slow-start phase and later, when reaching steady state operation, transmit over longer TTI duration to achieve higher data spectral efficiency from reduced CCH (e.g. scheduling grants) overhead.

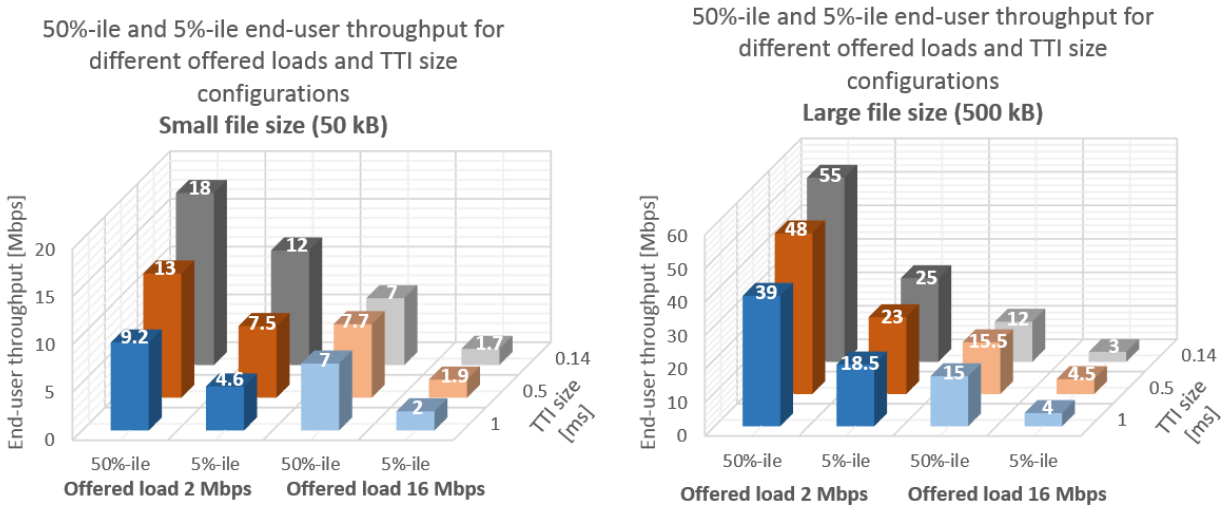

Figure A-27 5%-ile and 50%-ile end-user throughput for TCP file download with different TTI size, file size and offered cell load.

These results are in line with the observations made in the main body, which indicate that the optimal TTI size to schedule a user depends on several factors, being therefore desirable to have the flexibility to dynamically choose the TTI duration to satisfy the 5G requirements.

A.13 T5.2-TeC4 Group Transmission Concept using D2D (*DR-10*)

The benefit from UL group transmission stems from the fact that the SINR from the different UEs can be added. This follows since the UL transmissions are done jointly in a synchronized manner so that the received signals can be added at the eNB. Hence, from a network perspective, the group is seen as a single UE transmitting from different points. This is especially useful in low SINR scenarios or when the original UE that wants to transmit has a low SINR (or is even out of UL coverage) and the other group members have higher SINR. Hence, the SINR from the group transmission is typically considerably higher than for a specific UE within the group.

As we concluded above, when we form a group of UEs and transmit simultaneously, the received *SINR* at the base station is the sum of the individual UEs *SINR*. By applying Shannon's theorem on this, the maximum bit rate *BR* that can be obtained from a group of *M* UEs is given by

$$\boldsymbol{BR = BW \cdot log_2(1 + \sum_{m}^{M} SINR_m)} \quad \mathbf{(A.8)}$$

In order to investigate the potential (ideal) gains with the group transmission a small simulation study has been conducted. The simulation parameters are given in Table A-5 Group transmission parameters. Five UEs are randomly dropped into a hotspot area in a cell. This models the spatial distribution of the UEs that form the group. For these UEs, the path loss to the base station is used to calculate the combined path loss that results from simultaneous group transmission and finally the maximum bit rate that can be achieved from group transmission.

Table A-5 Group transmission parameters

Parameter	Value
Carrier Frequency	15 GHz
Attenuation constant	33.1 dB
Attenuation factor	3.7
Channel model	Okumura-Hata model
Bandwidth	1 RB (200 KHz)
Deployment	Single cell
Cell radius	2000 m
Hot spot radius	Varying
Number of users in group (*M*)	5
No of seeds	3000

This simulation setup uses a single cell setup with the Okumura-Hata channel model environment between the UEs and the base station, but assumes that the D2D transmissions are ideal. That is, no losses occur when data is distributed between the users in the group. The varying hot spot radius will model the varying differences in propagation that will occur when different UEs are located with various spatial distances to the base station. A larger radius will give a larger difference between the users in the group and therefore possibly larger gain for the users with bad connection. This will of course also result in less gain for the users with the best

connections. A large radius would also impact the D2D radio quality negatively, but this is not modelled in this study.

The UEs are all assumed to transmit with equal (maximum) power and all UEs experience the same (inter-cell) interference. The SINR expression for UE_m is

$$SINR_m = Pathloss_m * P/(N + I) \quad (A.9)$$

Inserting this into expression for maximum bit rate, we obtain the maximum achievable bit rate for the group transmission.

This is done for groups of 5 users that cooperate in an ideal way (i.e. no delay or losses within the group). The group radius is the size of the hotspot where the users in the group are dropped, i.e. a small radius means that the UEs are dropped close to each other and hence that the path loss is more similar between the UEs than when the radius is large.

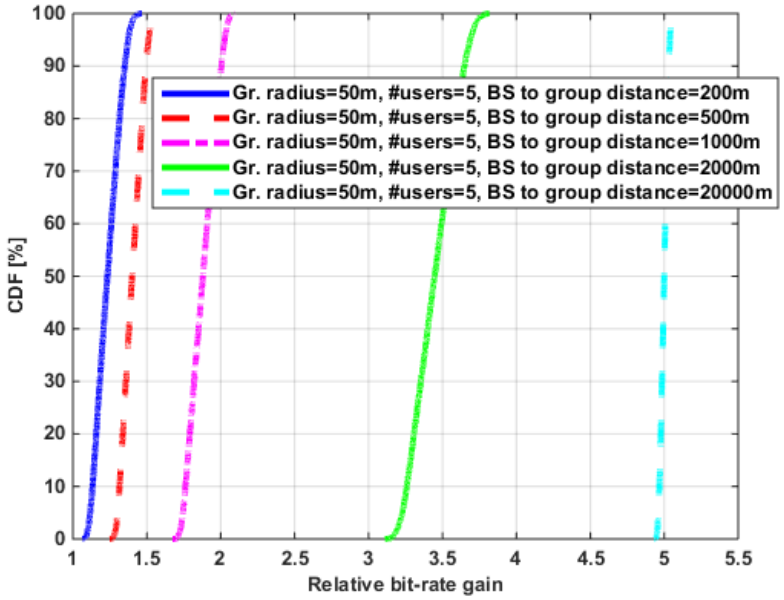

Figure A-28 Relative bit rate gains for hotspots of various distances to base station.

The relative bit rate gains for UEs in groups formed from UEs within a 50m hotspot, where the hotspots are at different distances from the base station is shown in Figure A-28 Relative bit rate gains for hotspots of various distances to base station. The median relative bit rate gain is in this case between 1.25 for hotspots close to the base station and up to 5 times for hotspot far away from the base station, i.e. for groups in bad coverage.

A.14 T5.2-TeC6 Abstraction Models for 5G multi-AIV Systems *(DR-11)*

The main objective of this TeC is to enable the networks to increase the flexibility of the RAN and meet the highly diverse requirements of 5G networks. Today's network deployments have cell edges and these edges give rise to undesirable effects such as low throughput, high interference, service interruptions, and call drops. Denser deployments lead to even smaller cells which in turn may lead to a larger number of edge-related challenges as the ones stated above. An additional important goal for next generation systems therefore will be to eliminate this edge problem, which can be achieved by employing user-centric virtual cells**.** A primary goal of these virtual cells is to provide a uniform quality of experience to users anywhere in the system, by "eliminating" the edge, i.e. to provide uniform SINR, eliminate handover, and provide a sustained TCP throughput for a uniform service experience regardless of the user location. In contrast to static configurations with predefined central controllers, a user-centric virtual cell achieves this by utilizing a group of cooperating nodes wherein a user is served by one or more dynamically assigned nodes, and the virtual cell is continuously reformed trying to keep the user at the center of the cell. In contrast to static configurations with predefined central controllers, a user-centric virtual cell consists of a group of cooperating BSs, wherein a user is served by one or more dynamically assigned BSs, and this virtual cell is continuously reformed so that the user "always" finds itself at the "center" of the cell. Figure A-29 illustrates this concept.

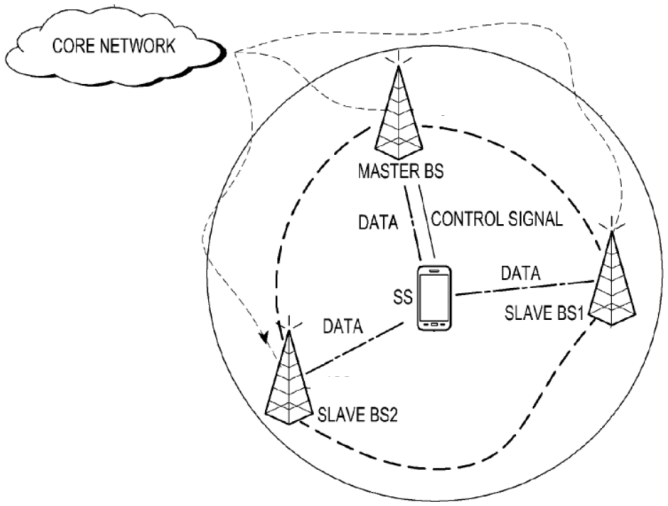

Figure A-29 Master and Slave roles in user-centric cell

An MS's user-centric cell is formed when the MS performs network entry. Figure A-30 shows a flow diagram where the process of a MS joining a virtual cell composed of two BSs is succinctly depicted. The BS that receives an access request from the MS becomes the Master BS of that MS's virtual cell. The MS may decide to include a list of BSs based on measurements such as RSRP when sending the access request. The list then forms the basis for the Master BS to select the slave BSs for forming the virtual cell. The Master BS communicates with the other BSs in the received list to confirm participation in the virtual cell. The confirmed BSs become the Slave BSs in the MS's virtual cell. BSs in the virtual cell are updated dynamically depending on the MS's location and/or the network situation.

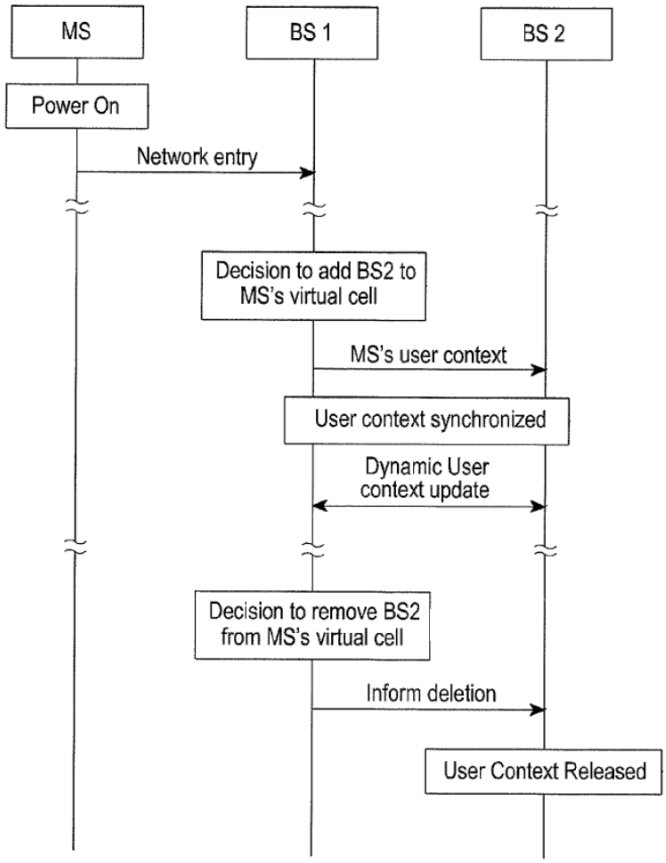

Figure A-30 Message exchange for virtual cell reformation

A.14.1 Cell Reselection and Dynamic Reformation for multiple AIVs

One key aspect of 5G is the coexistence of multiple AIVs, which may have a large impact in the selection of the appropriate BSs to form the virtual cell. To understand the process of dynamic reformation for virtual cells, we propose a method for cell reselection in 5G.

In LTE, the MS selects the strongest suitable cell during initial cell selection and it does not check if the cell supports specific service (e.g., D2D, MBMS, and MTC). When camped on a single BS, the MS initiates the process of establishing connection with another BS. The MS regularly searches for a better cell according to the cell reselection criteria. If a better cell is found, that cell is selected. The MS performs inter frequency cell reselection to a cell on a different frequency if:

- The new frequency has higher priority than the serving frequency and the quality of cell on the new frequency is greater than a threshold for certain duration
- The new frequency has lower priority than the serving frequency, quality of serving cell is below a threshold and quality of cell on the new frequency is greater than a threshold for certain duration

Priority of a frequency for cell reselection is indicated in system information and/or dedicated signaling. In certain cases the MS is allowed to consider the priority of certain new frequency to be highest irrespective of priority of frequency signaled by network. For example, if UE is interested in D2D communication and D2D communication can be performed by camping on a frequency then UE can consider the frequency supporting D2D communication as highest priority. This enables UE to reselect and camp on cell supporting the desired service even if quality of serving cell is above a threshold and serving frequency has higher priority than frequency supporting the desired service.

In 5G, the different access technologies shall support broad range of use cases including eMBB, mMTC, and uMTC. Furthermore, 5G will also consider frequency ranges up to 100 GHz. 5G access technology is also expected to support various services such proximity services (i.e. D2D), V2X, MBMS, etc. All services may not be supported on all frequencies. For example, eMBB may be supported on F1 (e.g. mmWave) while mMTC or uMTC may not be supported on F1. mMTC or uMTC may be supported on F2 (e.g. low frequency). In another example, public safety (PS) services may be provided on frequencies dedicated for PS operation. If a service is supported on a frequency, then it is supported on all cells of that frequency. Therefore, in a typical deployment all services may not be supported on all frequencies.

Therefore, in 5G the MS should (re-)select a cell if it is suitable and supports the service in which the MS is interested. Suitability criterion can be similar to LTE. On a frequency, if the MS finds multiple suitable cells supporting the service in which the MS is interested, then the MS (re-)selects the cell with best cell quality. If multiple frequencies support the service in which the MS is interested, then the MS prioritizes reselection to frequency with high priority.

In order to assist the MS in cell (re-)selection, the BS broadcasts the service(s) supported by it. The BS may also broadcast services supported in each neighboring frequencies. During the cell reselection the MS can prioritize the frequency which supports the service in which the MS is interested.

In deployments of 5G, 5G cells may coexist with legacy LTE cells (connected to EPC, not to 5G CN). 5G cells may or may not support all the services supported by LTE cells. 5G cells together with LTE cells can provide various services in an operator's network. For example, D2D discovery, D2D communication, V2X communication, MBMS services, etc. may be supported via LTE cells while enhanced mobile broadband, ultra-reliable and low latency services may be supported via 5G. In such a scenario, if the MS is camped on high priority 5G cell and is interested in service supported by low priority LTE cell but not supported by 5G cell, then the MS can obtain the desired service by prioritizing low priority RAT and performing inter RAT cell reselection. Therefore, the MS camped on high priority 5G cell and interested in service not supported by 5G cell but supported by low priority LTE cell can perform inter RAT cell reselection to low priority LTE cell. Figure A-31 shows an example of a heterogeneous virtual cell where the BSs are connected even to different cores.

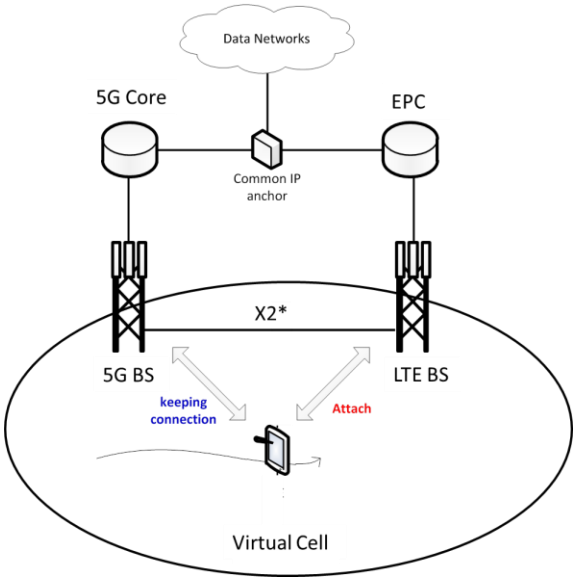

Figure A-31 Multiple-AIV 5G virtual cell

A.15 T5.2-TeC07 RAN Unified Management (DR-8, DR-11)

Future 5G systems will be composed by different AIVs and RAN slices, being needed a common framework for radio resource coordination between them, enabling mobile operators a single operational and maintenance systems. This unified management system may enable the adaptation of Radio Resources usage by different AIVs accordingly with the actual customers' service requirements.

The concept under the Unified Management is to provide the enablers for over the air coexistence of different encapsulated AIVs technologies, each one attending a specific

service/frequency band, by means of a “radio encapsulation” procedure with the following advantages:

- Reducing operators Operations Support Systems & Business Support Systems (OSS/BSS) costs. By means of unifying different AIV OSS/BSS tools and systems in one single virtual AI framework.
- Flexible adaptation of radio resources to services demand, enabling the dynamic assignment of radio resources to different AIVs, accordingly with the needs of service been provided by the operator in different areas.
- Provision of future proof framework for the deployment of new AIVs, decreasing the time to market of radio innovations.
- Inclusion of legacy systems as LTE-A, as part of the services.

Therefore, providing a framework for inter-AIV collaboration, allowing the inclusion of AIVs with different characteristics and numerology, maintaining as much as possible the freedom of design of multiple AIVs, encapsulated in their assigned radio resources.

An AIV coordinator, with access to the full context information, and able to establish requirement priorities and therefore scheduling of resources among the different AIVs, needs to be defined and characterized.

The Figure A-32, shows different logical entities involved in this TeC, presenting the Common Broadcast Channel as the key mechanism for AIVs coordination. The steps foreseen for the operation of this TeC, are:

1. The AIV Occupation level estimator is constantly monitoring the usage of different AIVs served by the TRxP, and reporting radio resource occupancy and requirement estimations from different UE traffic.
2. The AIV Scheduler coordinator, based on the reported used and required radio resources from different AIVs, and on the O&M policies established by the operator decides the amount of radio resources available for each AIV the TRxP is capable to generate.
3. The TRxP broadcasts the information related to the radio resources associated to different AIVs.
4. The Common Broadcast Control Channel is used for signaling to UEs (in different AIVs) their handover mechanisms. Enabling the reassignment of radio resources, and informing UEs of the new AIVs frequency band assignment, as well as the timestamp foreseen for this new radio resource allocation.

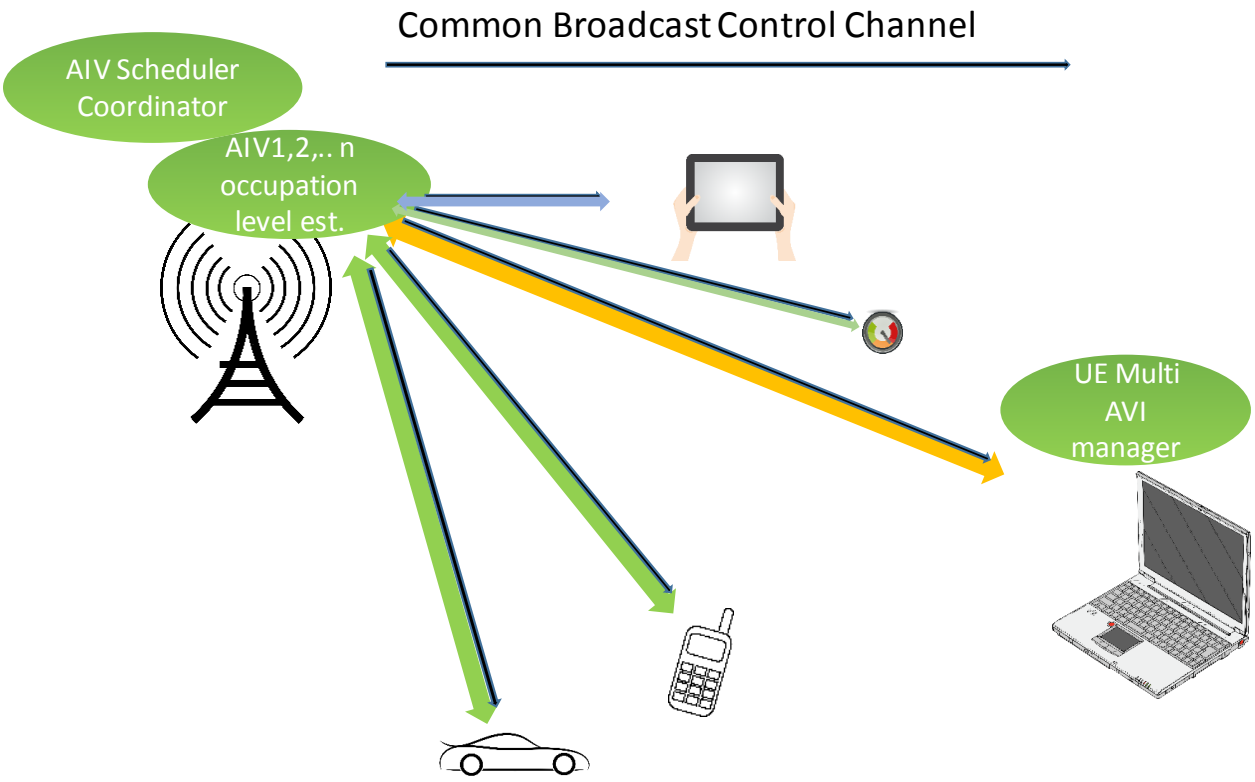

Figure A-32 Unified Management Functional deployment

All the AIVs work in an independent way and the additional Common Broadcast Control Channel is used in order to dynamically manage the radio resources allocated to different AIVs.

UEs being served by one or several AIVs used by the TRxP may incorporate a UE Multi-AIV manager, capable of selecting the most appropriate AIV, depending on the traffic pattern requirements and radio conditions (including the influence of the UE speed), or decide to use simultaneously several AIVs, for the same or different services.

A.16 T5.1-TeC2 LTE and 5G Tight Integration (*DR-12*)

Fast user plane switching (FS) is a scheme when the control plane is connected to both AIVs at the same time but the user plane is only transmitted via one of the AIVs, i.e. the PDCP level routes the packets to one of the AIV nodes. If the control plane is connected to both the LTE node and the 5G node, no signaling is required and the user plane switch may be almost

instantaneous. The fast user plane switch can be based on normal handover measurements such as RSRP, but also more advanced and faster type of measurements are possible and advantageous. The assumed protocol stack in this TeC is depicted in Figure A-33.

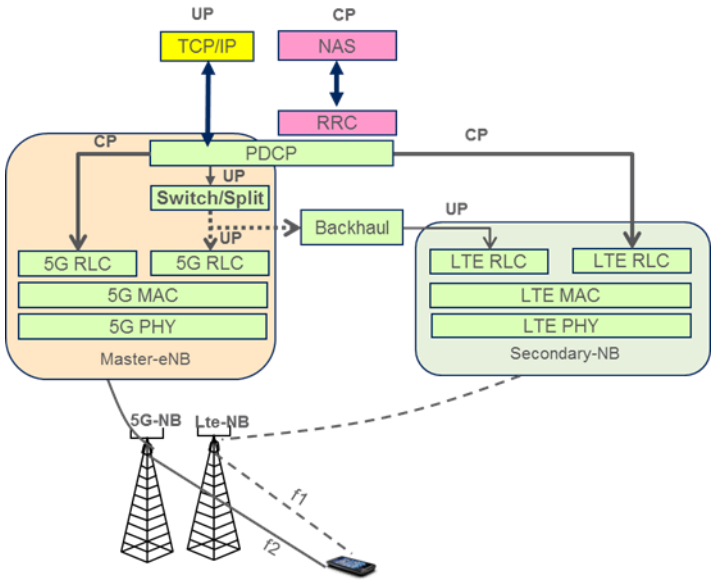

Figure A-33 Protocol stack assumed for tight integration simulations.

The different LTE and 5G tight integration concepts are evaluated using a system simulator.

Table A-6 Simulation parameters summarizes the most important parameters used in the simulations. In this paper, the new 5G AIV utilizes 0.2 ms TTI and 20 subbands (which corresponds to a LTE resource block) per 20 MHz compared to 100 subbands for LTE. The reason the number of subbands decrease is that the symbol length is decreased for the 5G AIV. This is done in order to keep the number of symbols per TTI the same for 5G AIV and LTE, and due to the inherent properties of OFDM when the symbol length decreases the sub-carrier spacing in frequency domain becomes larger, see [ÖAS+15] for more information. The LTE and 5G nodes are co-sited (see Figure A-34) and the frequency bands investigated are 2 GHz for LTE and 15 GHz for 5G. The channel model is the 3GPP Case 1 with typical urban channel model [3GPP05-05] where the attenuation constant is modified based on the carrier frequency. The bandwidth is 20 MHz per radio access.

One of the main characteristics of the novel 5G AIV is beamforming, to compensate the propagations effects at higher frequencies but this is not modeled in the simulations. Note that all signaling are ideal, i.e., all RRC signaling is always received correctly. This means that there are no handover failures due to RRC transmission failures. The LTE and 5G nodes are co-sited as depicted in Figure A-34.

Table A-6 Simulation parameters

Parameter	LTE	5G AIV
Carrier Frequency	2 GHz	15 GHz
Bandwidth	20 MHz	20 MHz
TTI	1 ms	0.2 ms
Subbands per 20 MHz	100	20
BS Tx power	40 W	40 W
Deployment	LTE and NR co-sited	
Traffic	FTP download of one 10 MB object per user and Video traffic using UDP, 1Mbps	
User speed	3, 10 m/s	
Backhaul	Ideal	
AIV selection	RSRP or SINR	
Dual connectivity selection	RSRQ or SINR	
AIV switch threshold	3 dB better	
DC add SeNB threshold	Maximum 1 dB worse than Master	
DC release of SENB threshold	More than 3dB worse than Master	

In [MET-II16-D51] a comparison of LTE-5G AIV dual connectivity vs. Fast UP Switch vs. 5G AIV stand-alone when 5G AIV is using 15 GHz carrier frequency was done. The differences between LTE-5G AIV dual connectivity (20+20 MHz) vs. 5G AIV stand-alone on 40 MHz were relative small on high percentiles. At low loads, the 90% of FS did not perform as good as DC or stand-alone since it cannot aggregate or utilize the same bandwidth. However, for 10%-ile user throughput, LTE-5G AIV dual connectivity (DC, blue solid line) performed much better, almost 300% better user throughput at low load compared to 5G AIV stand-alone and the FS also performed better than 5G AIV stand-alone. At higher loads, there is higher interference and not enough resources to allow sufficient aggregation making the FS perform as good as both stand-alone 5G AIV and DC.

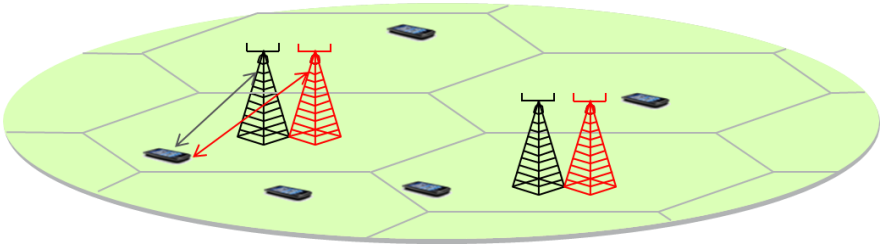

Figure A-34 3GPP case 1 hexagonal cell layout, LTE and 5G are co-sited.

In the results for FTP traffic in [MET-II16-D51], the possible switching times where 110ms. This may be sufficient in several scenarios, such when UEs go in and out of good 5G AIV coverage, or to handle shadow fading. In other situations, it is beneficial if the switching could follow the fast fading of at least an AIV on a lower frequency. How fast switching should be done in a more general case can be derived from the wave length of the fast fading for different UE speeds and carrier frequencies. In Table A-7, the time for the UE to travel one fourth of the wave length is given which means that if feedback is given at this rate, the FS can follow the fast fading relatively well.

Table A-7 How often switching must be possible (in ms time distance) to reasonably well follow the fast fading for different carrier frequencies and UE speeds.

	Carrier Frequency (GHz)			
UE speed (m/s)	**2 GHz**	**6 GHz**	**15 GHz**	28 GHz
0.1	375 ms	125 ms	50 ms	26.7 ms
0.833	45 ms	15 ms	6 ms	3.2 ms
3	12.5 ms	4.2 ms	1.7 ms	0.9 ms
10	3.37 ms	1.25 ms	0.5 ms	0.3 ms

From this it can be concluded that a fast switch or scheduling decision made every 10 ms would be sufficient to capture the fast fading for UE speeds up to 3m/s for e.g. LTE on 2GHz or up to 0.833 (3km/h) for an 5G AIV on 6GHz. To follow the fast fading for faster UEs or on higher frequencies would require faster measurement feedback for the FS. If the fast fading can be followed, even with the same average SINR level for LTE and 5G AIV, the FS would be able to ride on the fading tops and increase the SINR and thereby the performance.

To illustrate the effects of shorter minimum switch time, the 10-th percentile user throughput at higher loads is depicted in Figure A-35. Here it is clearly seen that the throughput increases for the worst users as the minimum switch time is decreased. The throughput with DC is also shown for comparison.

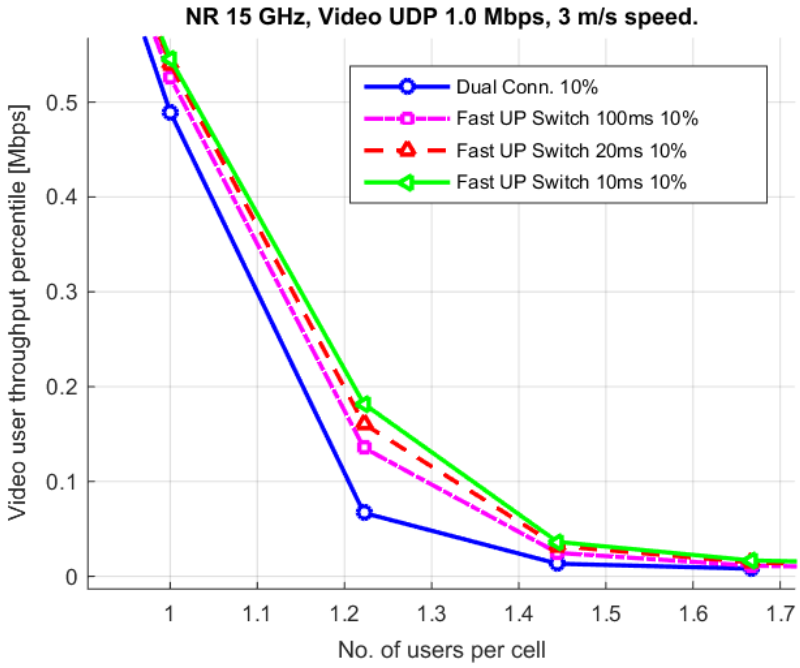

Figure A-35 10th-percentile user throughput.

To further illustrate the effects of shortening the minimum switch time, the video packet delay is depicted in Figure A-36, where it is seen that the mean delay is reduced by close to 20% at high load when the minimum possible switching time is reduced from 100ms to 10ms.

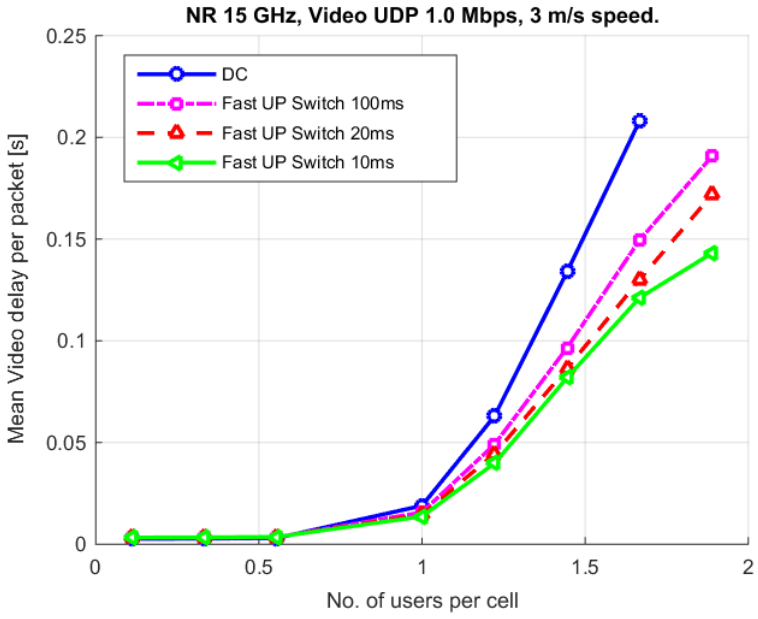

Figure A-36 Mean video delay decreases as minimum possible AIV switch times is reduced

A.17 T5.1-TeC7 Multi-AIV Dynamic Traffic Steering Framework (*DR-13*)

A.17.1 Further Description

In this section, we describe the evaluation methodology used in terms of simulation details, models used, etc., and provide some further simulation results and related analyses. The fundamental mechanism is similar to the one described in [MET-II16-D51].

A.17.2 Evaluation Methodology

The basic scenario used is as shown in Figure A-37, with a 5G-UE engaged in multi-connectivity with multiple AIVs. Here we consider the case where the UE has simultaneous multi-connectivity with various 5G carriers. Each UE is connected to the strongest link from each of the AIVs after considering the possible link losses due to path loss and shadow fading. The scenario used is similar to the Madrid scenario considered in [PLV+15], with only outdoor deployment assumptions. The main focus of the evaluations was on SINR optimization using packet duplication over the air interface.

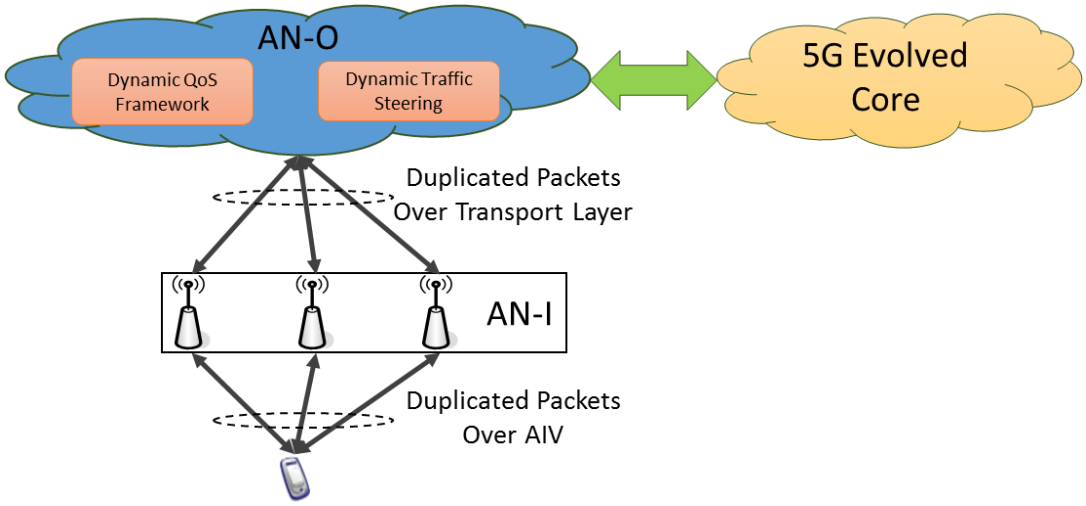

Figure A-37 Scenario Overview.

A.17.3 Detailed Analyses and Further Results

The set of basic assumptions used in this work for bearer architecture and dynamic QoS framework is similar to the one presented in [MET-II16-D51], and shown in Figure A-38. In the simulations results presented, conceptual beam coordination is assumed, with evaluations done using multiple time synchronized cells creating an SFN area by transmitting duplicated packets over same physical resource blocks while generating the SINR curves.

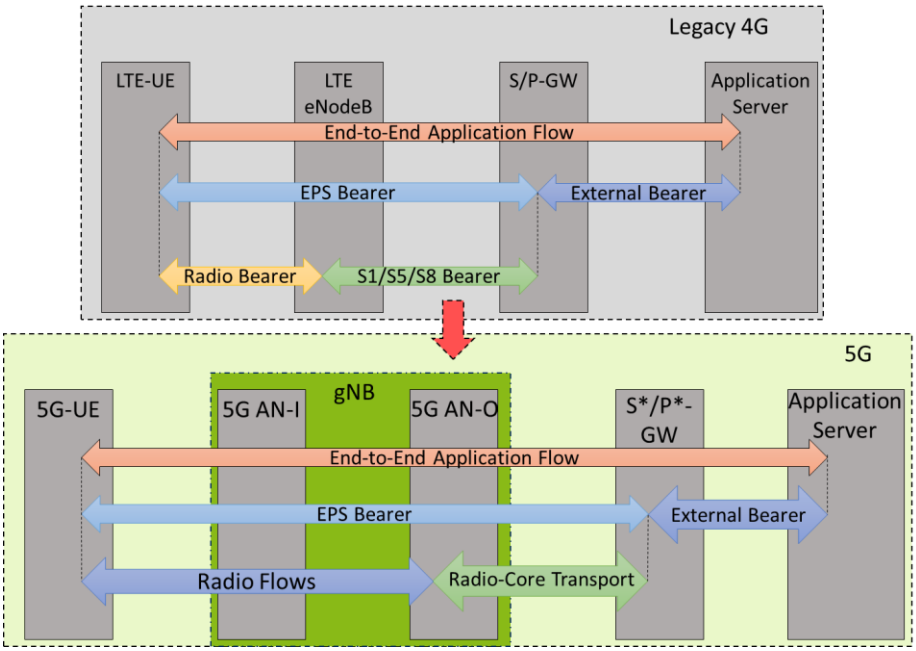

Figure A-38 LTE bearer architecture and evolution towards 5G service flow architecture [PME+16], [MET-II16-D51].

The possible options for a flexible functionality or protocol split between the 5G gNB and possible Cloud RAN deployments, is as shown in Figure A-39. The options are based on the proposals made in [3GPP16-RP160043], and also presented in [MET-II16-D51]. Here the key assumption is that the 5G xhaul that is assumed to link the cloud RAN and gNB is an open interface, with standardized information elements transporting the control and user plane data between the nodes. The various options described here are based on the work done in [MET-II16-D51], with the main difference being the focus on improving reliability through packet duplication rather than throughput improvements. As discussed in [MET-II16-D51], the Option-A could be considered similar to the LTE dual connectivity feature and the Option-D represents the fully centralized C-RAN deployment with xHaul transporting information similar to currently deployed fronthaul. The dynamic traffic steering framework for improving reliability would require limited impacts for Options C-D, whereas Options A-B would require faster AN-I link layer measurement information exchange, while supporting fast traffic re-routing between different nodes within the AN-I. Such considerations were discussed in [MET-II16-D51] as well.

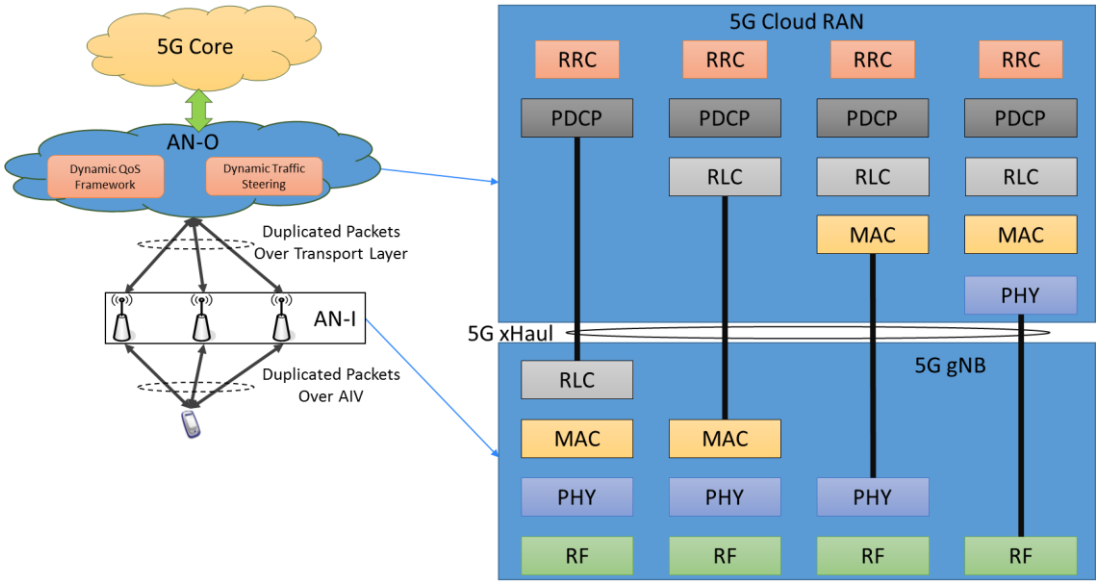

Figure A-39 Application of potential 5G protocol split options [3GPP16-RP160043] to AN-O and AN-I for C-RAN deployments [MET-II16-D51].

A.18 T5.1-TeC9 RM and Traffic Steering in Heterogeneous Environments (*DR-14*)

A.18.1 Detailed description of the solution

The following definitions are used in the proposed solution description.

- A network node is defined by its coordinates and its mobility type. The mobility type defines the mobility characteristic of the network node by three categories, from the highest to the lowest mobility: *Mobile*, *Nomadic* and *Fixed*[11].
- A transmission link is defined by its transmitting node and its receiving node and by the type of used transmission band. The transmission link may be of three types: a *Downlink*, an *Uplink* or a *Common* (specifically, *Downlink* or *Uplink* bands are used if a FDD technique is adopted while the *Common* band is used if a TDD technique is adopted)[12].

[11] As an example, by adopting a Cartesian reference system, a fixed network node A, can be described by a couple $A_{3D}=[(x_A,y_A,z_A)$, *Fixed*] or $A_{2D}=[(x_A,y_A)$, *Fixed*] if respectively a 3D or a 2D geometrical analysis is adopted.

[12] For example a transmission link L_1 can be described by a triplet L_1=[A, B, *Downlink*] where the network node A is the transmitting node (TxNode(L_1)=A), the network node B is the receiving node (RxNode(L_1)=B) and the transmission band is of the downlink type.

- A collision is defined as the situation occurring when the establishment of a new transmission link between a transmitting and a receiving node or the modification of an existing transmission link creates an extra interference toward neighbor receiving nodes in such a way that at the receiver of a neighbor node (S/I) < $T_{S/I}$, having defined S/I as the ratio between the useful signal power and the interference power associated to the transmission links affected by the extra interference and $T_{S/I}$ as the minimum allowed S/I threshold (e.g., 12 dB as specified in [PGR09]).
- A Resource Sharing Cluster (RSC) is defined as the set of transmission links that need to be coordinated (e.g. by a suitable Time-Division Multiplexing mechanism) in order to avoid potential collisions between each other's. Each RSC may be defined by the set of potentially colliding transmission links, the transmission band, the center of gravity (defined, as an example, by the average values of the coordinates of all the transmitting/receiving network nodes involved in the transmission links of the cluster) and the mobility category of a RSC corresponding to the mobility category of the most mobile network nodes among all the transmitting and receiving network nodes of the transmission links of the RSC[13]. The data associated to the RSCs need to be updated by taking into account that:
    - a RSC containing a transmission link L should be updated at the end of the transmitting period of the transmission link by the deletion of the link from the RSC;
    - RSCs having a mobility category corresponding to *Nomadic* needs periodically updates of the transmission links involving all the transmitting and/or receiving nodes of the *Nomadic* type;
    - RSCs having a mobility category corresponding to *Mobile* need of constant updates of the transmission links involving all the transmitting/receiving network nodes of the *Mobile* type as soon as the location information of at least one of said nodes change and it is delivered to or known to the system.

[13] As an example, a Resource Sharing Cluster RSC_1 can be defined by a quadruple RSC_1=[{L_1,L_2,L_3,…},*Downlink*,(x_{CG},y_{CG}),*Nomadic*] where the set of transmission links that form the cluster is {L_1, L_2, L_3, …}, all the transmission links of the cluster are of the downlink type, the center of gravity (CG) of the cluster has coordinates (x_{CG},y_{CG}) (in a 2D) and at least one network node among all the transmitting and receiving network nodes connected by the communication links of the cluster has a mobility category defined as *Nomadic* (all the other network nodes are classified either *Nomadic* or *Fixed* and none network node of *Mobile* category is present).

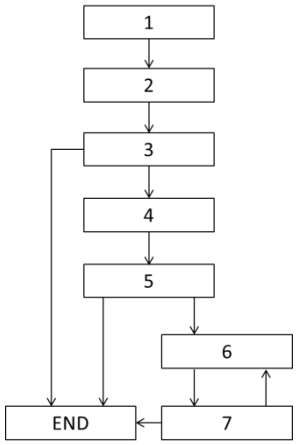

Figure A-40 Steps of the solution

The method for avoiding transmission collisions prior to a new mmW transmission in a cellular network by means of a geometrical analysis is described by the following steps (see Figure A-40).

1. The process starts when a new transmission link needs to be activated[14].
2. The second step of the method provides for the discovery of a list of candidate RSCs to be considered for performing the geometrical analysis in the activation of the new transmission link. The criteria for the inclusion in the list depends on the transmission band category and possibly on the distance between the center of gravity of the candidate RSCs and the transmitting network node of the new transmission link (i.e. since transmission links are limited by a maximum range, for example, a two times maximum range could be considered). The list of candidate RSCs involved in the activation of the new transmission link is named CAND_RSC_LIST
3. If no RSC is found and the CAND_RSC_LIST is empty, a new RSC containing the new transmission link is created, and the method ends[15]. Otherwise, if the CAND_RSC_LIST is not empty, a geometrical interference analysis for identifying possible collisions between the new transmission link and all the candidate RSCs belonging to the CAND_RSC_LIST is performed in the following step 4.
4. For each RSC in the CAND_RSC_LIST the set of transmission links of the RSC is considered and each incumbent transmission link of the set is geometrically compared (by an algorithm described below) with the new transmission link to determine if a

[14] $L_{NEW}=[N_1,N_2,Band\_Cat]$ is defined by the band category Band_Cat (that can be *Downlink*, *Uplink* or *Common*), by the transmitting network node $N_1 = [(x_{N1},y_{N1}), Mobility\_Cat_{N1}]$ and the receiving network node $N_2=[(x_{N2},y_{N2}), Mobility\_Cat_{N2}]$ with $Mobility\_Cat_{N1}$ and $Mobility\_Cat_{N2}$ that can be *Mobile*, *Nomadic* or *Fixed*.

[15] For example, by considering the new transmission link L_{NEW} defined above, a new RSC would be $RSC_{NEW}=[\{L_{NEW}\},Band\_Cat,(x_{CG},y_{CG}),Mobility\_Cat_{NEW}]$ where $x_{CG}=x_{N1}+x_{N2}/2$, $y_{CG}=y_{N1}+y_{N2}/2$ and $Mobility\_Cat_{NEW}$ would be the most mobile category between $Mobility\_Cat_{N1}$ and $Mobility\_Cat_{N2}$.

possible transmission of the new link could interfere or be interfered by the incumbent link. If at least one collision is found, the search in the current RSC terminates and the RSC is inserted in a list, named 2MOD_RSC_LIST of RSCs to be modified[16].

5. If no RSCs containing a colliding link are found (i.e. the 2MOD_RSC_LIST is empty) a new RSC containing the new transmission link is created and the method ends. Otherwise, the following step 6 is performed.
6. Prior to the insertion of the new transmission link in at least one of the RSCs listed in the 2MOD_RSC_LIST, an evaluation procedure (specified below) can be performed in order to assess whether the considered RSC will have enough resources to include the new transmission link, without having an unacceptable degradation of the performances of the new link or of one or all of the incumbent links.
7. If an unacceptable degradation is foreseen, a number of different actions (described below) could be performed and the process ends. Otherwise, the new transmission link is added to the considered RSC (that is deleted from 2MOD_RSC_LIST) and resources will be re-usable e.g. in time and/or frequency domain within the RSC. The steps of the evaluation procedure are then repeated (going back to step 6) for all the others RSCs listed in the 2MOD_RSC_LIST (in the event of a multiple insertion in different RSCs a coordination between the resource sharing mechanisms of the different RSCs will be needed e.g., by simply merging the two RSCs) until the list is empty and the process ends.

A.18.2 Detailed description of step 4

The geometrical-based interference analysis of step 4 is here described with reference to Figure A-41.

[16] For example if a RSC_k=[{L_1,L_2,L_3,…L_n}, Band_Cat,(x_{CG},y_{CG}),Mobility_Cat] is the RSC considered against the new transmission link L_{NEW}=[N_1,N_2,Band_Cat] , each L_i of the RSC_k is tested in two ways for determining a potential collision between L_i and the new transmission link: whether the transmitting node N_1 of L_{NEW} may possibly interfere the receiving node RxNode(L_i) of L_i and whether the transmitting node TxNode(L_i) of L_i may possibly interfere the receiving node N_2 of L_{NEW}. If, after having performed this test, a potential collision is found, (e.g. with L_3) the search in RSC_k is stopped (links L_4 through L_n are not tested) and RSC_k is inserted in the 2MOD_RSC_LIST.

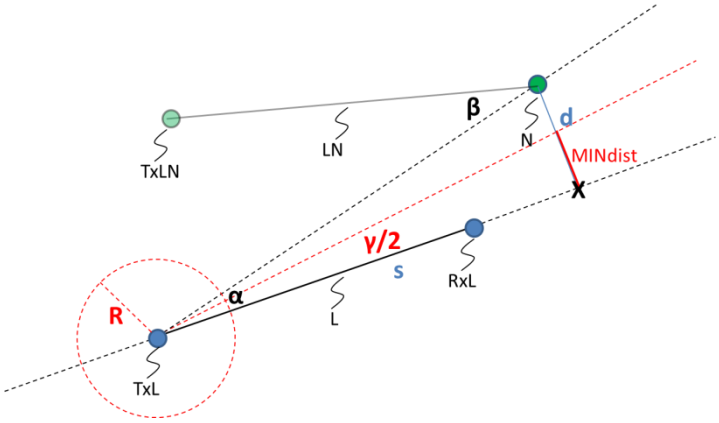

Figure A-41 Geometrical-based interference analysis on whether link L interferes the receiving node N.

The approach, described in a 2D analysis, can be extended to 3D. The aim is to check whether a transmission link L can interfere a receiving network node N of a transmission link LN. In Figure A-41, the transmitting network node (TxNode(L)) of the transmission link L is designated with TxL and the receiving network node (RxNode(L)) of the transmission link L is designated with RxL while the second transmitting network node of the transmission link LN is designated with TxLN. The geometrical-based interference analysis comprises the following three main steps:

I. computing the distance d between the straight line connecting the nodes TxL, RxL and the node N; the angle α between the straight line connecting the nodes TxL, RxL and the straight line connecting the node TxL and the node N; the distance s between the node TxL and a projection x of the node N on the straight line connecting the nodes TxL, RxL. As an example, according to an embodiment of the present invention, given the coordinates (x_{TX},y_{TX}) of the node TxL, (x_{RX},y_{RX}) of the node RxL and (x_N,y_N) of the node N,

- the slope of the straight line connecting the nodes TxL, RxL can be computed as: $m_L=(y_{RX}- y_{TX})/(x_{RX}- x_{TX})$;
- the distance d can be computed as: $d = |y_N-(m_L(x_N- x_{TX})+ y_{TX})|/\sqrt{(1+m_L^2)}$;
- the slope of the straight line connecting the node TxL and the node N can be computed as: $m_N = (y_N- y_{TX})/(x_N- x_{TX})$;
- the angle α can be computed as: $\alpha = \arctan(|(m_L-m_N)/(1+ m_L\, m_N)|)$;
- the distance s can be computed (using the cotangent function) as: $s = d \cot(\alpha)$;

II. determining the minimum distance MINdist within which an interference between the two transmission links L, LN is possible, according to the following considerations:

- the distance s;
- the location of the projection x with respect to nodes TxL and RxL that is whether the projection x is or not on the same half–line of the node RxL (i.e. α is larger than 90°);
- whether the antenna of the node TxL is directive (i.e. conveying the main part of the radiated energy in a specific direction) or not (in such a case, angle α is essential for determining if the node N is within the main beam of transmission of the antenna of the node TxL);
- whether the antenna of node N is directive or not (in such a case, angle β between the straight line connecting the nodes TxLN, N and the straight line connecting the node TxL and the node N, that can be computed in a similar way as angle α, is essential for determining if the receiving antenna of the node N is within the main beam of transmission of the antenna of the node TxL);
- the transmission power, and other specific radio parameters (e.g. radiation pattern, sensitivity, etc.) of nodes TxL and N;
- safety margins driven by concerns about technology constraints (e.g., non-ideal radiation pattern), type of environment (open range against canyon, free Line-Of-Sight against a lot of obstacles), etc.

For example, given the S/I threshold $T_{S/I}$, a simple way to determine MINdist could take into account the distance s, the main lobe (or beam) angle of the antenna of the node TxL, the transmission range r_L of the transmission link L, i.e. the distance between the nodes TxL and RxL. Under the assumption that the antenna at the receiving node is omnidirectional and that at each receiving node the received signal power is always of the optimal magnitude, the delta power level reached by the transmission link L in the projection x compared to the power level in the node RxL can be calculated by the following formula that determines a dB increase/decrease of the transmission power level of the transmission link L: $10 \log(r_L^2/(s/\cos(\alpha))^2)$. Moreover, defined γ as the angle of the main lobe (or beam) of the antenna of the node TxL, the projected distance d_{MAIN} "under" the main lobe can be calculated as $d_{MAIN}=s \tan(\gamma/2)$. Defined a Front to Back Ratio (FBR) as the delta in dB between the transmission gain of the main lobe of the antenna of the node TxL and its side lobe, and a circle whose radius R is determined by the C /I threshold such that $-(10 \log(r_L^2/R^2) - FBR) = T_{S/I}$, we can then define:

- MINdist=0 for all distances s such that $s>R\cos(\gamma/2)$ and $-10 \log(r_L^2/(s/\cos(\alpha))^2) > T_{S/I}$;

- MINdist=d_{MAIN} for all distances s such that s>Rcos(γ/2) and -10 log(r_L^2/(s/cos(α))^2) ≤$T_{S/I}$;
- MINdist= R sin(α) for all distances s such that –R ≤ s ≤ Rcos(γ/2);
- MINdist=0 for all distances s < -R (in the opposite half–line of the node RxL);

III. Determining by the comparison between d and MINdist whether the transmission link L eventually interferes with the node N. The collision check is negative if and only if the distance d is higher than MINdist, otherwise it is positive. For example, the collision check shown in Figure A-41is negative and no interference is foreseen by the transmission of link L toward the receiving node N.

The geometrical comparison between the new link L_{NEW} and L_i (one of the links of the considered RSC) in step 4 is therefore performed in both ways:

1. L= L_{NEW} and N is the receiving node of L_i to check if the new link interferes
2. L=L_i and N is the receiving node of new link L_{NEW} to check if the new link is interfered

A.18.3 Detailed description of step 6

Step 6 can also be further detailed. In particular, the evaluation procedure, prior to the insertion of the new transmission link in a RSC to assess if the RSC has enough resources to include the new transmission link, without having an unacceptable degradation of the performances of the new transmission link or of one or all of the incumbent transmission links, can take into account the following aspects.

A. limits on the available resources per RSC; for example, a simple evaluation could consists in a comparison between the number of incumbent transmission links of a considered RSC and a fixed threshold MAX_RSC_Link_No (maximum number of transmission links per RSC): if RSC_k=[{L_1,L_2,L_3,…L_n}, Band_Cat,(x_{CG},y_{CG}),Mobility_Cat] is the RSC considered, the check would be n < MAX_RSC_Link_No;

B. service level agreements and operator QoS policies for the new transmission link or one or all of the incumbent transmission links in the considered RSC, possibly in conjunction with the type of transmission (or type of service that is the cause of the transmission) of the new transmission link or of one or all of the incumbent transmission links; for example, a gateway/backhauling purpose of the new transmission link or of one of the incumbent transmission links has in itself a sort of priority over other transmissions;

These aspects can be clarified with an information exchange with the lower and upper RRM/SON layers. Specifically, for the information of type A an exchange with the lower layer (scheduler) is expected, while for information regarding type B aspects an interaction with upper RRM/SON layer can be predicted.

A.18.4 Detailed description of step 7

In step 7, if indeed an unacceptable degradation is found, a number of possible actions available to the network could be performed with the help of AN-O and AN-I layers of Figure 3-1. For examples, the followings actions can be performed:

- an alternative transmission link on a lower frequency can be set up for the new transmission link or for one or some of the incumbent links (in this case, the AN-O layer with its traffic steering mechanisms is involved in the solution);
- in case all of the transmitting/receiving nodes involved in the activation of the new transmission link are mobile nodes a random postponement of the transmission can be applied, looking forward to more favorable new conditions of the cellular network due to new positions of the mobile nodes (in this event the AN-I layer with its scheduling functionality is involved); a similar action could be adopted by one or more of the incumbent transmission links, provided that the same mobility condition of their respective nodes applies;
- in case the new transmission link has gateway/backhauling purposes for a group of network nodes (for example, in vehicular and NN services) an alternative network node within the group can be chosen. In such a case the foreseen transmission will not take place and a new analysis will be performed on another transmission link involving the alternative network node (in such solution, the AN-O layer is involved).

A.18.5 Detailed description of algorithm variants

In simulations with the original PGIA, through the analysis of the distribution of RSCs w.r.t. their size, it was discovered that the algorithm was too conservative. In the original geometric analysis, the size of the interfering link cluster is not taken into account in determining potential interferences/collisions. This leads to overestimated interference values and the developments of RSC of large size. In the example of Figure A-42a) the interference of link L_1 upon the receiving node RX is considered by the Original PGIA always on and on the entire band, while actually, due to the resource sharing, the link interferes for 1/4 of the time or band only. In the first variant of the algorithm (PGIA w/ Clusters) the interference analysis takes into account the size of the RSC which the interfering link belongs to, in particular the formulae defining the contribution (in dB) of the interference in the main and secondary lobes turn into $-10\log(r_L^2/((s/\cos(\alpha))^2 n_{RSC}))$ and $-(10\log(r_L^2/((s/\cos(\alpha))^2 n_{RSC})) - FBR)$ where n_{RSC} is the number of links in the RSC containing link L. Besides at step 4 of the algorithm (see A.18.2) the interference contributions of all the links belonging to the considered cluster are summed up and checked against the $T_{S/I}$. In particular, given RSC-K the cluster in CAND_RSC_LIST under scrutiny, links that not cause collisions are nevertheless added to $S_{RSC\text{-}K}$ which is the linear sum of the interference that cluster RSC-K causes as a whole to the new link L_{NEW}. Given the conversion function from linear to dB $Linear2dB(x)=10\log_{10}(x)$, if no link belonging to RSC-K causes a collision, at the end of the process a new check $-\ Linear2dB(S_{RSC\text{-}K}) \leq T_{S/I}$ is introduced: if positive RSC-K is added to 2MOD_RSC_LIST. In the example of Figure A-42 b)

the interfering links L_1, L_2 and L_3 interfere each for 1/4 of the time or band and are all over the $T_{S/I}$ threshold (they do not cause collisions) but the sum of their interference contributions S_1 could cause a collision and it is checked against the $T_{S/I}$ threshold.

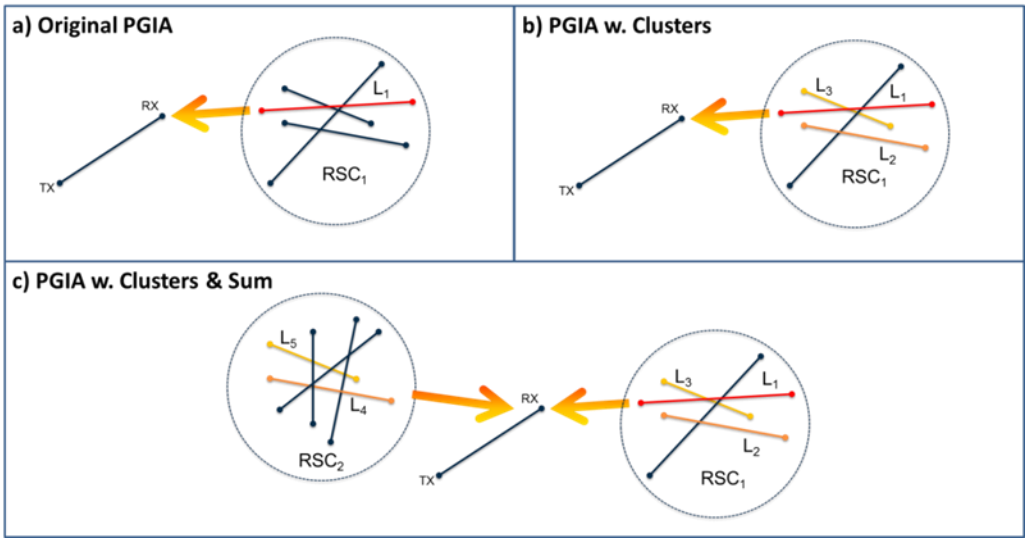

Figure A-42 Examples of the different variants of the PGIA algorithm

In the second variant of the algorithm (PGIA w/ Clusters & Sum) the interference analysis takes into account the size of the RSC which the interfering link belongs to and the sum of the interference contributions by RSC as PGIA w/ Cluster but in addition the interference contributions of the different clusters are summed up and the sum is checked against the threshold $T_{S/I}$: in case of a collision the cluster of the interfered link is merged with the cluster that causes the highest interference. In particular at step 4 let CAND_RSC_LIST={RSC-1, RSC-2, …, RSC-N}, if no single link of any single RSC-K causes a collision and no RSC-K has a S_{RSC-K} such that – *Linear2dB*(S_{RSC-K}) ≤ $T_{S/I}$ let $\Sigma_{CAND\_RSC\_LIST}$= S_{RSC-1}+ S_{RSC-2}+… S_{RSC-N}. If – *Linear2dB*($\Sigma_{CAND\_RSC\_LIST}$) ≤ $T_{S/I}$ is verified, let RSC-M be the cluster of CAND_RSC_LIST such that S_{RSC-M} = max{ S_{RSC-1}, S_{RSC-2},…, S_{RSC-N}}, RSC-M is added to 2MOD_RSC_LIST. In the example of Figure A-42c) the interfering links L_1,…,L_5 and S_1 and S_2 do not cause a direct collision but the sum of the interference contributions of RSC_1 and RSC_2 could cause a collision and it's checked against the $T_{S/I}$ threshold.

A.18.6 Detailed description of PGIA in a scenario with obstacles

To test the PGIA algorithm in a scenario where obstacles are taken into account, a simple statistical model has been introduced [ALS+14]. This model presents two formulae regarding the mmW links:

1. P_{OUT}, outage probability (that impedes communication) as a function of the distance d between transmitter and receiver nodes: $P_{OUT}(d)=\max(0,1-e^{a\_out \cdot d + b\_out})$

2. P_{LOS}, Line-Of-Sight probability (visibility between the nodes) as a function of the distance d between transmitter and receiver nodes: $P_{LOS}(d)=(1- P_{OUT}(d))\cdot e^{a\_los\ d}$

Parameters used in the formulae are based on empirical measurements within the New York University campus (urban environment), in particular a_out=1/30, b_out=5.2 and a_los=1/67.1. In Figure A-43 the two probabilities are plotted in a transmitter-receiver distance range between 5 and 200 meters.

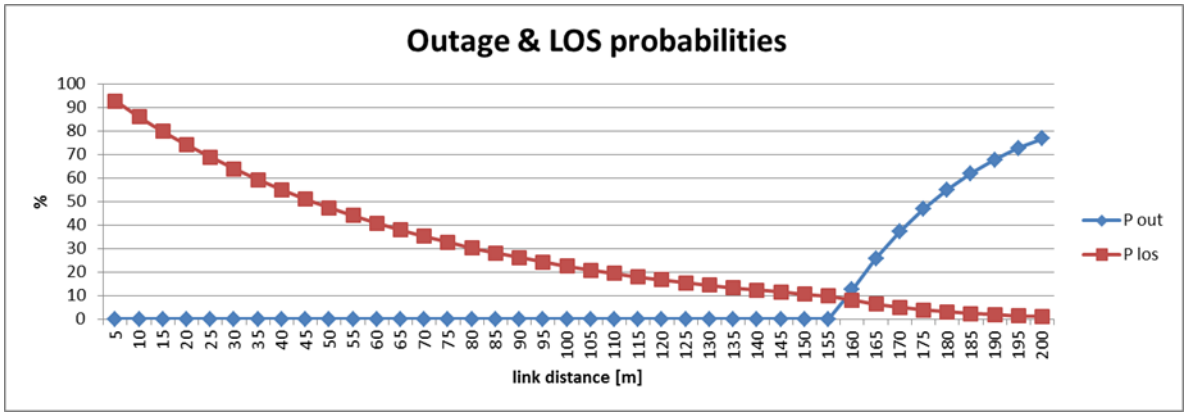

Figure A-43 P_{OUT} and P_{LOS} plotting for mmW link distances between 5 and 200 meters

As the distance between transmitter and receiver nodes grows the LOS probability decrease and at around 155 meters the outage probability starts to rise.

When taking into account buildings (as obstacles and reflectors) through P_{OUT} and P_{LOS} probabilities, a mmW link or an interference, depending on its distance, could be blocked by an obstacle, be in LOS (Line-Of-Sight), when transmitting and receiving/interfered nodes are visible to each other, or be in NLOS (Non-Line-Of-Sight), when the signal of transmitter node reaches the receiver/interfered node through one or more reflections. From [ALS+14] measurements at 28 GHz and 73 GHz it can be derived that path loss values of links in NLOS are, on average, 25 dB higher than those in LOS. Thus, when considering interference from a transmitter node to a potentially interfered receiver node of two different links belonging to different RSCs, the signal should be checked to understand if it can reach the receiver (there is no outage, so the nodes are not so far apart). Subsequently the status LOS/NLOS of the original interfering mmW link and of the interference "link" (between the interfering transmitter and the potential interfered node) must be compared. Under the assumption of ideal power control the power of the interference signal is subject to a delta stemming from the LOS/NLOS status of the original link and the interference. Values are shown in the following table:

Table A-8 Power delta [dB] applied in the different LOS/NLOS situations

Original link	Interference	Pow. delta [dB]
NLOS	NLOS	0
LOS	NLOS	-25
LOS	LOS	0
NLOS	LOS	25

In the upper two rows where interference is in NLOS (so the interferer node is at a certain distance from the potentially interfered one) interference could be depowered if the original link was in LOS as its receiver and transmitter are quite near to each other. On the other hand, in the lower two rows where interference is in LOS (so the interferer node is quite near to the potentially interfered one) interference could be overpowered if the original link was in NLOS as its receiver and transmitter are at a certain distance from each other.

The above randomly overpowered/depowered interferences could undermine the effectiveness of the original clustering mechanism. To include, to an extent, such effect in the algorithm, the statistical characterization of P_{OUT} and P_{LOS} should be added to PGIA. With this implementation, for each mmW link and each interference, the distance d between corresponding transmitting and receiving nodes is calculated and two random numbers are generated and checked against the $P_{OUT}(d)$ and $P_{LOS}(d)$ to ascertain whether the radio transmission/interference is possible and it is in LOS or NLOS. This characterization of links and interferences will be used in the creation of the concurrent set of mmW links (since links in outage are eliminated) and to determine the above-mentioned power deltas in the computing of interference values within the RSCs formation. Such calculations will be independently repeated for interference evaluation during KPIs analysis to simulate the actual “real” positions of obstacles leading to blocked signals or signals received through reflections in NLOS situations.

A.18.7 Evaluation Methodology

Most of the simulations are performed under the assumptions of a 1 km^2 flat area with no obstacles and random positioning of transmitting and receiving nodes (here the network is supposed to be highly dynamic and nodes classical distinction between access point/base station and user equipment/terminal is blurred due to a widespread use of opportunistic mmW communication/relaying). For receiving nodes omnidirectional antennas are used. Transmitters are equipped with directional antennas characterized with the main lobe of 4 degrees and a FBR of 30 dB. An ideal power control mechanism is active: at each receiving node the received signal power is always of the optimal magnitude, moreover, a maximum range of a mmW radio link is set to 200 m with a S/I threshold of 12 dB (below this threshold the transmission link cannot be established and the throughput of the link is zero). The overhead of the resource sharing mechanisms is negligible as, in mmW, the band is thought to be quite broad: throughput

is divided equally between the links of the same RSC. Simulations are Monte Carlo with 10000 runs for every studied scenario where the number of mmW links per km^2 varies from 10 to 200 with 10 link steps.

In simulations without PGIA, after the random deployment of nodes in the area (maximum range withstanding), the receiving nodes with an S/I below the S/I threshold are counted regardless of their source of interference (main or secondary lobe, one or several interferer).

In simulations with PGIA, initially every random link is a member of its own cluster (RSC). A geometrical analysis is then performed to determine collisions by the main or secondary lobes: a collision (with the S/I of the receiving node below the S/I threshold) of at least one of a RSC members with a member of a different RSC causes the merging of the two RSCs. At the end of the process the receiving nodes with an S/I below the S/I threshold are counted. In such a scenario, the remaining interfered nodes will be the ones interfered by several transmitting links belonging to different RSCs from their own (each interferer producing a S/I above the threshold if considered standalone and therefore undetected by the PGIA analysis). The contribution of each interferer is divided by the size of its cluster under the hypothesis of a (time) sharing of the resources.

In final simulations obstacles are modeled via a statistical characterization of P_{OUT} and P_{LOS}. For all the mmW links the distance between transmitter and receiver nodes is determined and P_{OUT} and P_{LOS} computed (see A.18.6 for details). The same check is also applied to all sources of interference: the first time in the RSCs forming process to simulate the statistical inclusion of P_{OUT} and P_{LOS} in the PGIA algorithm; the second time in the KPIs analysis to simulate the actual “real” positions of obstacles that lead to interference blockage or NLOS interference through reflections.

A.18.8 Further results

In subsequent work to [MET-II16-D51], the average link (user) throughput as a % of the total throughput achievable by one single user/link without any resource sharing (whenever the link is the only and single member of an RSC) was added as a KPI.

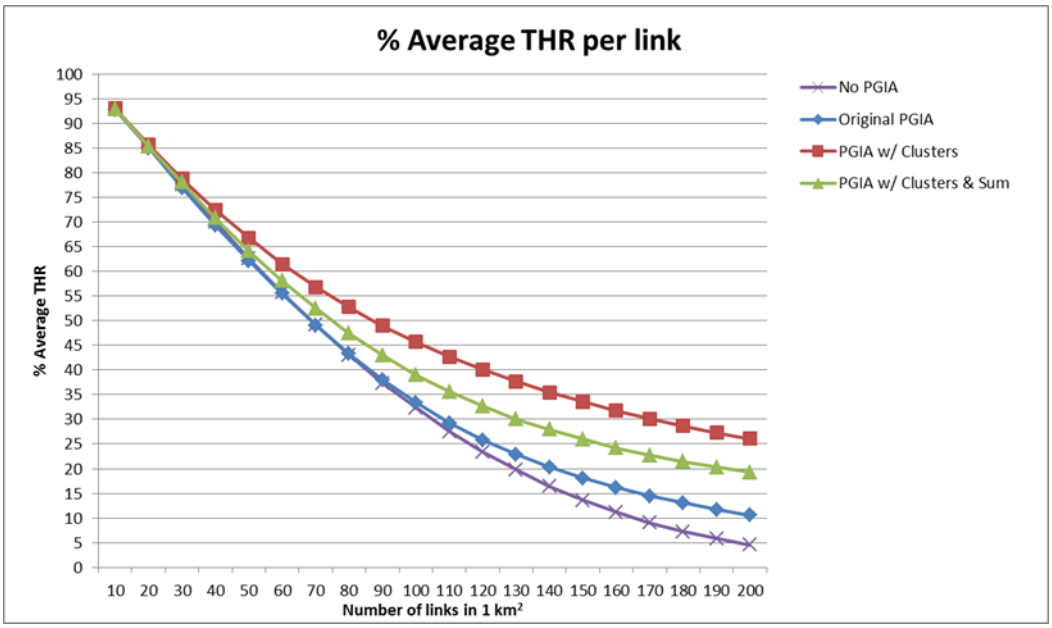

Figure A-44 Average throughput (THR) per link as a % of the total throughput achievable by one single link as functions of the number of concurrent links in 1 km^2, without and with different PGIA algorithms

In Figure A-44 the performance of the network with no algorithm or different PGIA algorithms is compared. Under the simulation assumptions described above, the different variants of PGIA mechanism not only keep the number of interference links very low, but also achieve a better average throughput per link with a consistent gain in areas with many concurrent mmW links in 1 km^2: both new variants perform better than the original PGIA algorithm (which nonetheless doubles the average throughput in the most loaded scenario), the PGIA w/ Clusters variant, by considering potential collision only on a cluster by cluster basis keeps the size of the RSC quite limited achieving the best performance of all the algorithms. On the other hand, in the PGIA w/ Cluster & Sum variant, the absence of interference is paid by a lower throughput for all the concurrent links, none of them experiencing mutual interference.

Other simulation results regarding less substantial KPIs are shown below.

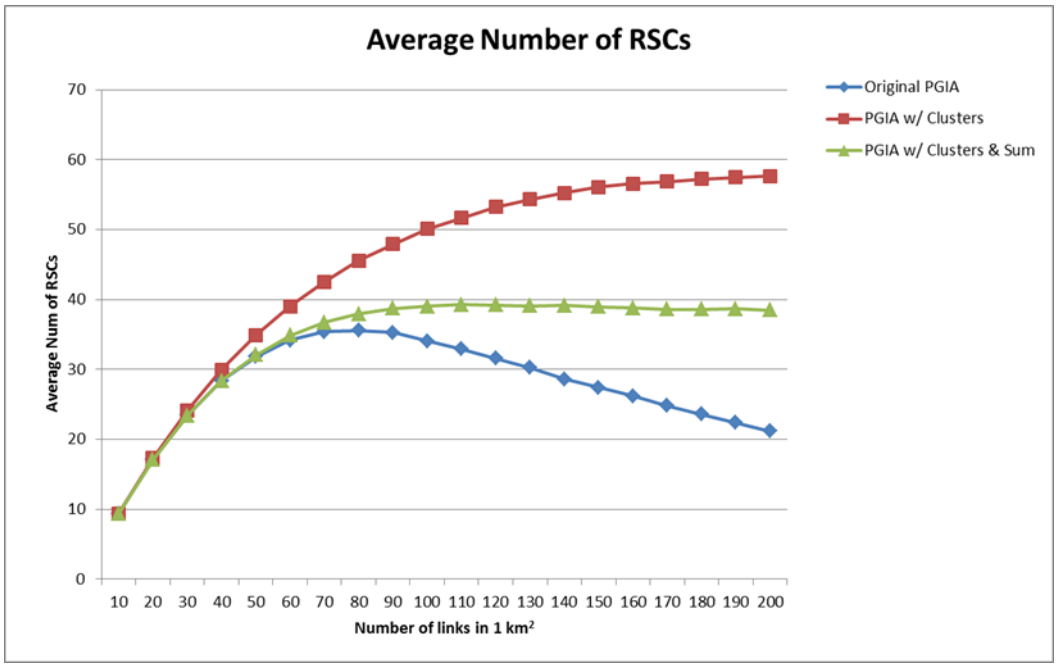

Figure A-45 Average number of RSCs as function of the number of concurrent links with different PGIA algorithms

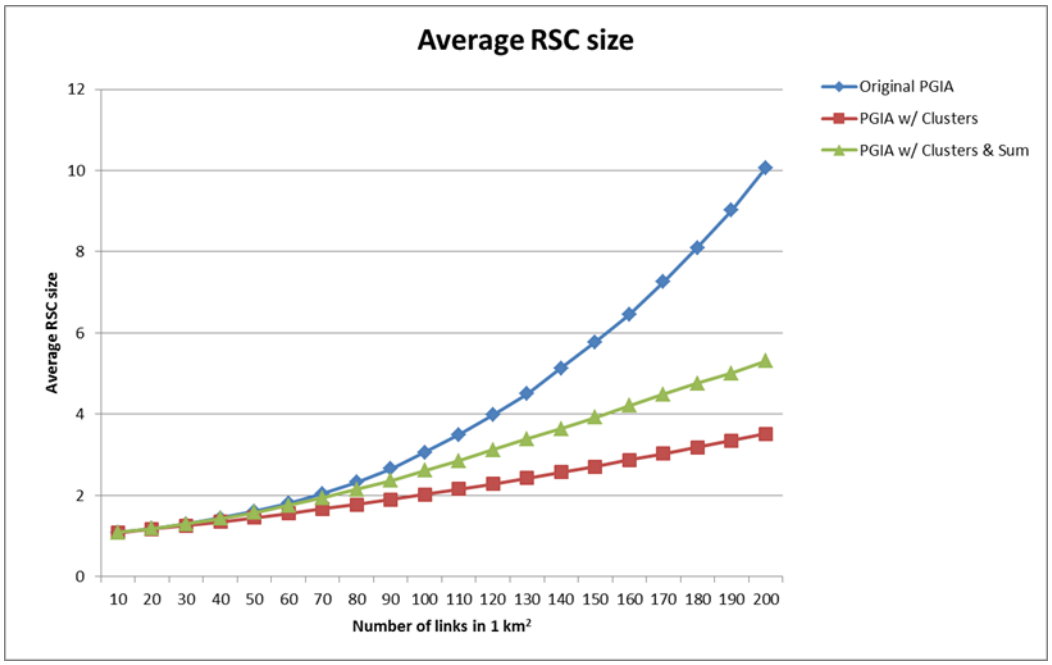

Figure A-46 Average RSC size as function of the number of concurrent links with different PGIA algorithms

Figure A-45 depicts the performance of the various PGIA algorithms with the respect to the average number of RSCs as the number of links per km^2 increases. In the original PGIA the

average number of RSCs increases to around 35 and then decreases as a tendency emerges toward the formation of a big *supercluster* containing most of the links of a scenario. This tendency is far less evident with the PGIA w/ Cluster & Sum variation and absent with PGIA w/ Cluster variation where the merging between clusters is more focused a far less conservative than the original algorithm. Since more mergers of clusters diminish the average number of RSCs and increase their average size the same consideration can be applied to Figure A-46 which shows the average number of links per cluster (RSC size) as the number of links per km^2 increases. Here the tendency to form a big *supercluster* of the original PGIA algorithm is particularly evident in the sharp rise in the average RSC size, while the growth is the most restrained with the PGIA w/ Cluster variation.

Furthermore, in the final step of the analysis a scenario with obstacles and a suitable model was introduced based on [ALS+14]. In an urban environment were buildings can be obstacles to mmW transmission or help through reflections for NLOS connections, statistical characterizations of the P_{OUT} (outage probability) and the P_{LOS} (Line-of-Sight probability) in relation to the distance between transmitter and receiver of an mmW link were introduced. In the hypothesis of mmW transmissions operational in LOS only, the performance of the algorithms improves since interference is kept low by obstacles. Adding the statistical characterization of P_{OUT} to the PGIA algorithms, by taking into account, in the RSCs formation, the probability that interferences successfully reach potentially interfered nodes, improves performance. This is achieved because interference is not overestimated leading to a better and leaner cluster formation with consequent improvements in terms of throughput. On the other hand, taking into account both LOS and NLOS mmW transmissions, results in additional interference from nearby transmitters in LOS occurring (randomly) when they have to power up their transmissions in order to compensate for their NLOS links to their distant receivers. Conversely, interference from distant transmitters in NLOS could be lower than expected (randomly) when they have to power down their transmissions to their near receiver in LOS. Such randomly increased/decreased interferences can offset the effectiveness of RSCs formation unless statistical characterization of P_{OUT} and P_{LOS} are taken into account in the PGIA mechanisms (see A.18.6 for details). In Figure A-47, once P_{OUT} and P_{LOS} are added to the scenario and to the PGIA mechanism, a comparison of the average % of interfered links is shown with and without a PGIA w/ Cluster & Sum mechanism.

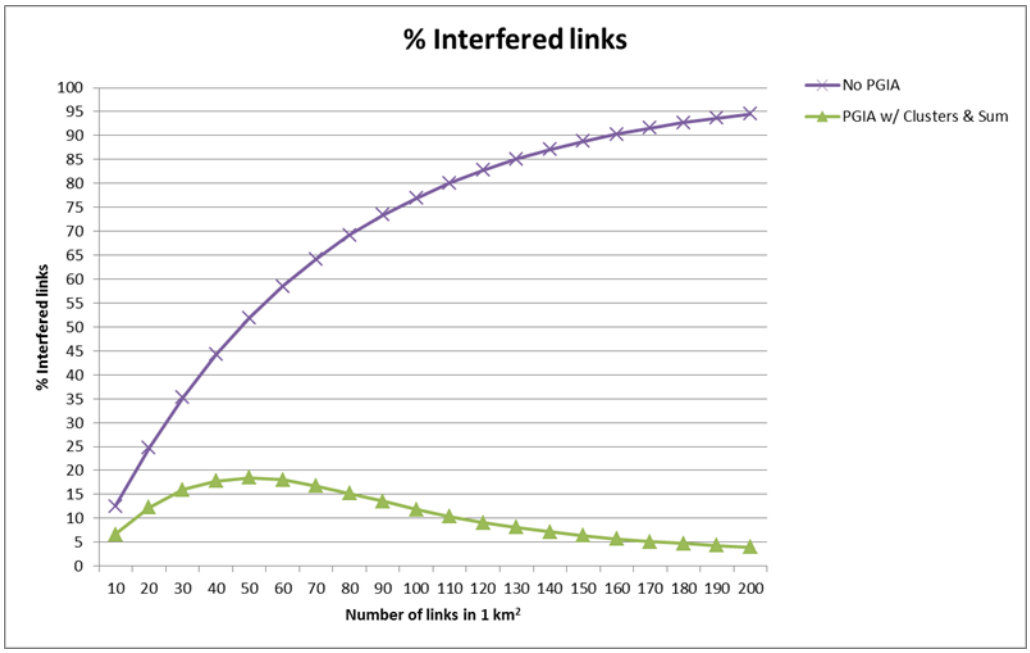

Figure A-47 Average % of interfered links as functions of the number of concurrent links in 1 km², without and with PGIA w/ Clusters & Sum algorithm in an urban environment with obstacles (modeled by P_{OUT} and P_{LOS} statistical characterizations)

It can be noted that, without the PGIA mechanism, the percentage of interfered link gets quite high even with few concurrent links in the area, exceeding the 50% threshold with less than 50 concurrent mmW links while between 80 and 95 percent of links will be interfered in a crowded scenario. On the other hand, with the PGIA w/ Clusters & Sum algorithm and the statistical characterization of P_{OUT} and P_{LOS} interference is kept quite low (with only between 5 and 10 percent of interfered links in crowded scenario). In this case, interference is capped at around 18% with 50 concurrent mmW links when the actual random increased/decreased interference instances are maybe too few to be properly captured by the statistical characterization.

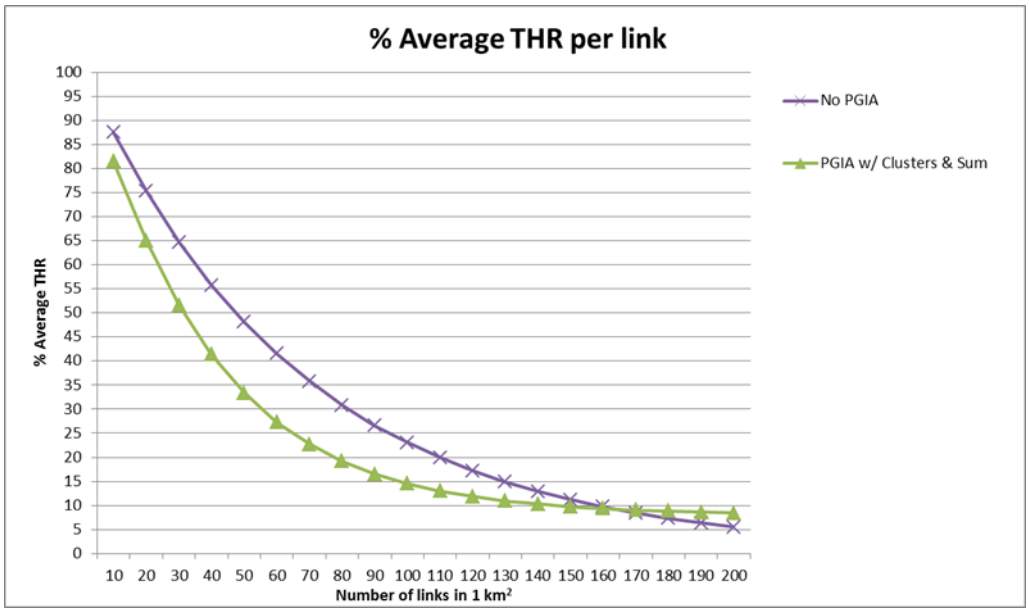

Figure A-48 Average throughput (THR) per link as a % of the total throughput achievable by one single link as functions of the number of concurrent links in 1 km^2, without and with PGIA w/ Clusters & Sum algorithm in an urban environment with obstacles (modeled by P_{OUT} and P_{LOS} statistical characterizations)

The outstanding results in fairness of the PGIA w/ Clusters & Sum and statistical characterization algorithm are not quite equated in the average throughput per link (see Figure A-48) but tend to be almost constant around 10% of the total throughput achievable by a single user in very crowded scenario whereas, without PGIA, average throughput would have plunged to a very low 5% of the total throughput achievable.

A.19 T5.1-TeC13 Reflection Environment Maps for Enhanced Reliability and Traffic Steering (*DR-15*)

A.19.1 Further Description

One of the key challenges currently faced by 5G radio access networks is to enable extreme mobile broadband and ultra-reliable communication in a practically viable manner. In order to achieve this, currently defined self-organizing and traffic steering function considerations for legacy radio access technologies need to be enhanced. In this TeC, we consider the use of a novel concept called reflection environment maps (RefMaps) to improve reliability, coverage and data rates in 5G ultra-dense small cell deployments, by acting as an essential assisting

function for dynamic traffic steering. The 5G SON functions are enhanced to support the creation and usage of such maps, with additional measurement configurations done in the UE.

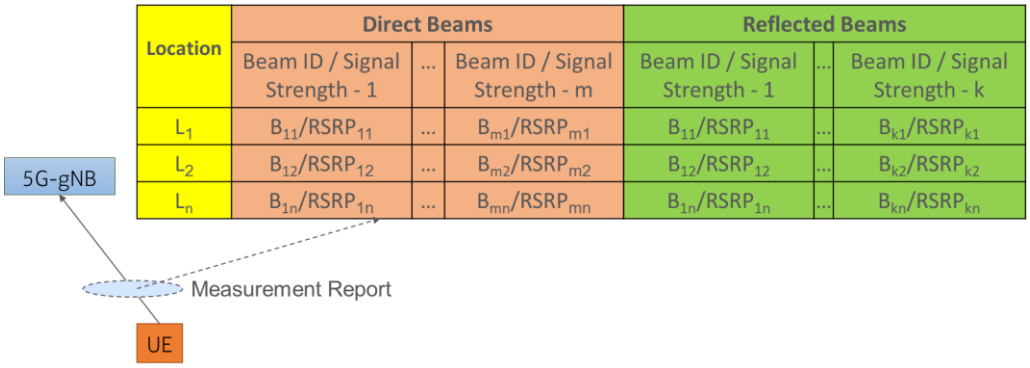

Figure A-49 Proposed Reflection Environment Map Database Structure [PUL+16]

Detailed performance evaluations of the benefits of having ideal RefMaps are evaluated using 5G mmW ultra-dense deployments, and it is shown that the proposed mechanism can provide significant improvements in terms of coverage and capacity of the network. The evaluations indicate that such enhancements would be essential for enabling reliable and coverage enhanced 5G deployments, especially in the higher frequency bands. The possible RefMap database structure is as shown in the figure above.

A.19.2 Evaluation Methodology

The possible RefMap database structure generated based on the UE measurement report is as shown in Figure A-49, containing the strongest direct and reflected beam information. The key assumption here is that the UE is sending frequent uplink signals which could be similar to the uplink sounding reference signals (SRS) configuration available in LTE-Advanced systems [3GPP16-36213]. The gNB could configure the periodicity of the measurements or the events based on which UE should report measurements, depending on the deployment scenario, mobility conditions, network architecture, etc. The key operation methodology is that, when the gNB detects that a direct link with the UE is broken due to dynamic radio conditions, the UE is signaled using alternate, reflected beams, based on the information available in RefMaps database. Here the assumption is that the initial / training phase is complete and a detailed RefMaps database is available. The gNB could also query the reflected beam information from the database using the location information of the UE when the link was broken. Depending on the attenuation faced by the reflected beams, the gNB could send duplicated data over multiple beams in order to utilize the soft-combining gains available at the UE. The assumption is that the combined signal received at the UE would have sufficient strength for reliable data reception.

The measurement configuration used to generate the RefMaps could be based on a modified RRCConnectionReconfiguration message [3GPP15-36300], using which the UE could additionally report the reflected beams it receives, along with the direct beams. The possible signaling diagram during the initial / training phase described earlier where the RefMap

database information is collected, is as shown in Figure A-50. The optimization functions that are required for generating the RefMap database could be collocated with the gNB, in case of the distributed scenario. The signaling diagram for the fast recovery phase is as shown in Figure A-51, indicating the actions taken by the gNB when a link breakage is detected. In the considered example, the gNB detects that Beam-1 is blocked, due to which the direct link with the UE is broken. But, based on the RefMaps information queried by the gNB, it is detected that the UE could be still reached using Beam-2, Beam-3, but with additional attenuation due to reflections. Thus, the gNB sends the data to the UE using the alternate beams, thereby ensuring that the UE can continue to be in active state and receive data from the gNB. The gNB could send the data using same modulation and coding scheme (MCS) using the same physical resource blocks, which could then be combined by the UE to enable better data reception. This would be similar to concept presented in TeC7 related to packet duplication for improving reliability. Here the added constraint is that the duplication is done over multiple beams rather than cells. The use of reciprocal beamforming technique in uplink ensures that the UE can send UL data and control information to the gNB using this mechanism.

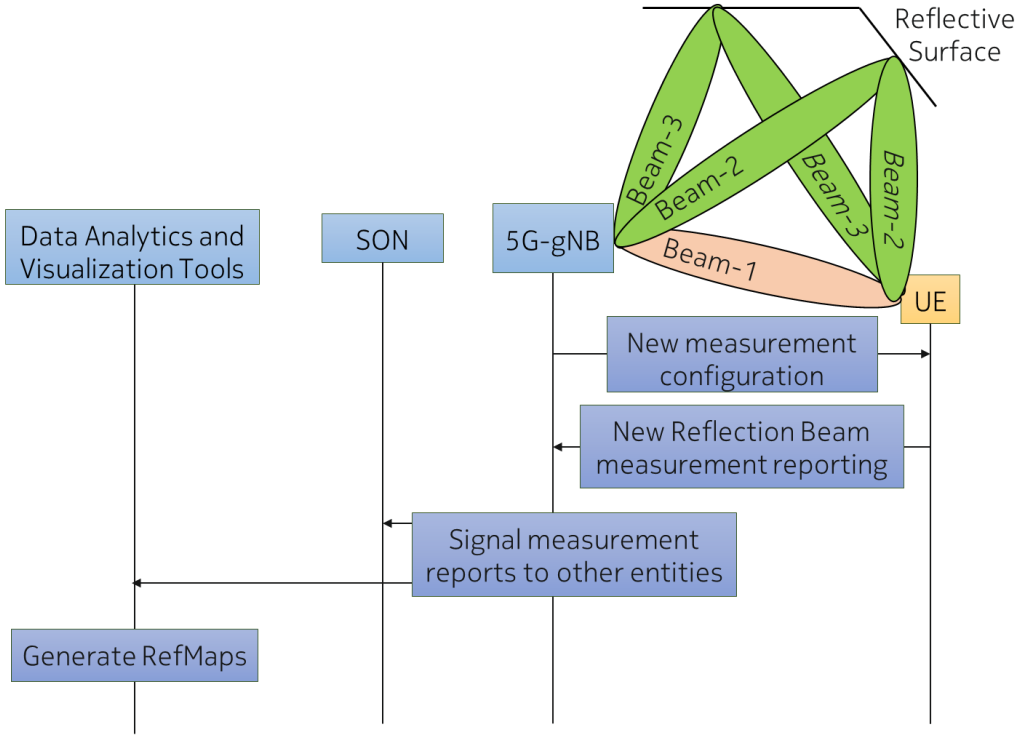

Figure A-50 Signaling Diagram for the Initial / Training Phase [PUL+16].

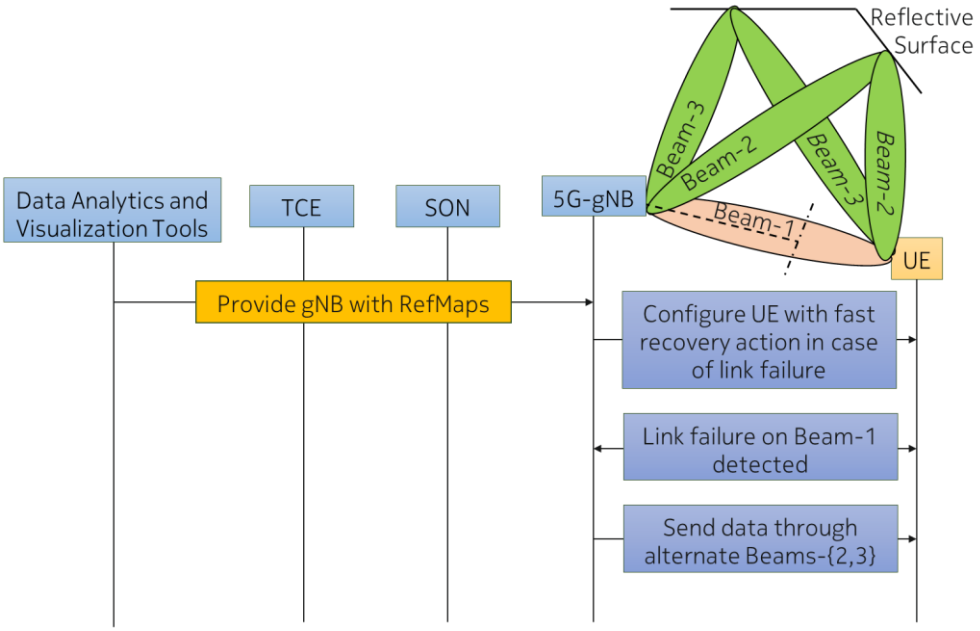

Figure A-51 Signaling Diagram for the Fast Recovery Phase [PUL+16].

A.19.3 Detailed Analyses and Further Results

The architectural assumptions made by the RefMaps mechanism are as listed below, similar to the system model presented in [PUL+16]:

- We consider a system where the 5G-UE reports measurements to the gNB, based on the measurement configurations, and it is assumed that the BS is connected to the SON node using a direct interface, with possible collocation with the BS. The mechanism is not limited by the location of the RefMaps database.
- The TCE assumed in this work could be similar to the one used in MDT feature [SSJ+13], which can even be used to collect the idle mode UE measurements.
- For the connected mode UEs, the measurements can be collected by the BS as part of RRC measurement reporting, and reported to the SON function.
- The TCE is linked with the data analysis and visualization tools which can generate the RefMaps.
- We consider a beam-based system design with each beam having a unique beam ID adding an additional dimension to cell IDs available in 4G system. Such design is essential for operation in mmW bands and gains increasing relevance in UDN deployments.

- For mobility procedures, each an enhanced neighbor cell measurement concept available in LTE-A systems is assumed, with each gNB and UE maintaining the list of strongest neighboring cell and beam IDs.
- Due to indoor deployment assumptions, the amount of information stored in the RefMaps database can be limited, based on the possible mobility paths of the UE.

For the simulation, similar to the assumptions in [PUL+16] we consider the users that are not in NLOS conditions and with SINR > δ_{th} dB to be active UEs. We also evaluate the performance of using the RefMaps with an additional NLOS probability of δ_{NLOS}=0.2, apart from the distance dependent probability values. Full-buffer traffic was considered for the evaluations, with round-robin scheduler for data transmissions. Macro cells are not considered here for the evaluations and small cells are assumed to be deployed in a dedicated frequency layer. Here, the main focus here is on improving the reliability within an ultra-dense indoor environment supporting high data rates. The assumptions used are similar to the one used in [PUL+16], with models used based on the work done in [STR+15], [CGK+13].

The number of active users normalized to the RefMaps-5dB case is as shown in Figure A-52. Here RefMaps-5dB and RefMaps-10dB indicate the scenario where the gNB reaches the UE through alternate beams, with a signal strength attenuation of 5 and 10 dB respectively. The signal strength attenuation would mainly impact the user throughput of the UE, when the direct link is broken. From the figure, we can observe that the proposed radio environmentally aware mechanism enables significantly higher number of users to be active, as compared to the dynamic traffic steering mechanism based on enhanced QoS framework and baseline LTE scenario. The gains are due to the gNB reaching the UE through alternate beams with a finite attenuation in signal levels, if the primary beam has been blocked due to changes in the radio environment.

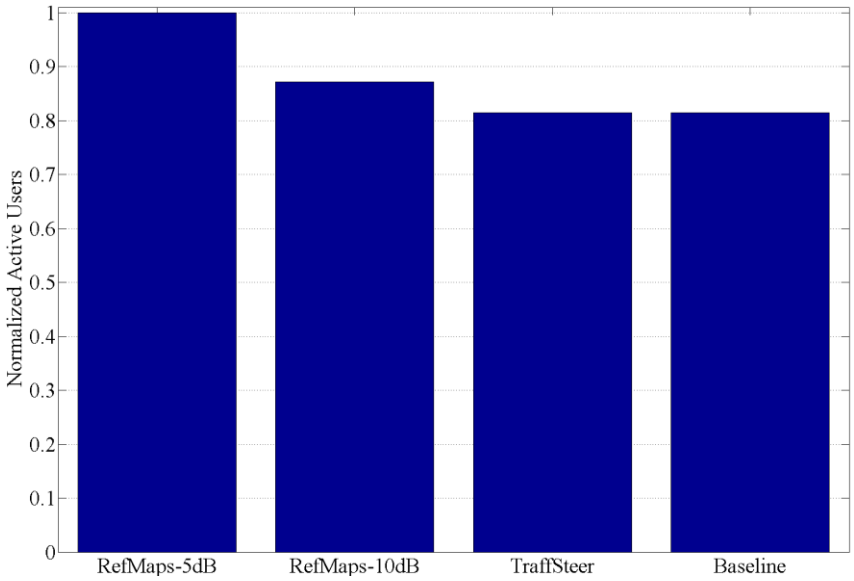

Figure A-52 Number of active connected mode users normalized to that of RefMaps-5dB case.

It has been shown that the proposed enhancement enables ultra-dense standalone deployment of mmW small cells network with extreme mobile broadband and ultra-reliability. There are some additional measurement enhancements required from the UE perspective, and related configuration and infrastructure enhancements expected from the network perspective for RefMap creation and storage. Such enhancements, when combined with the dynamic traffic steering framework and with the assistance of dynamic QoS framework in the 5G-RAN discussed in Annex A.17, will act as key enablers for currently envisioned 5G KPIs.

A.20 T5.2.TeC-8 RM for Inter-Network Coordination (*DR-16*)

A.20.1 Evaluation Methodology

Evaluation has been carried out using software prototypes deployed on the ORBIT wireless testbed [ORB] which enables realistic evaluation using software-defined radios and spectrum measurement infrastructure (the testbed was recently used to support the “DARPA Spectrum Challenge” conducted in 2013-14 which provided a competitive evaluation of peer-to-peer spectrum cooperation and dominance techniques). The prototype architecture is depicted in Figure A-53.

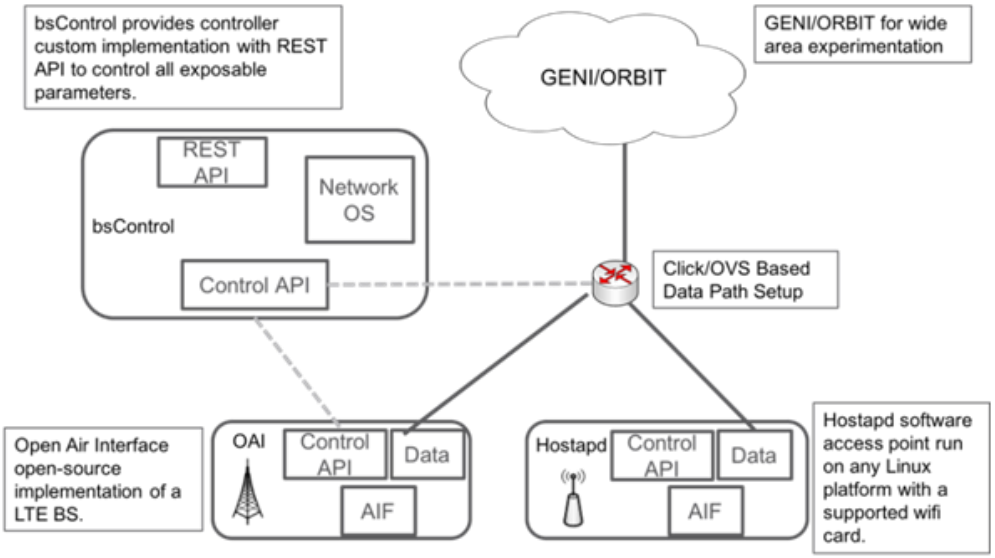

Figure A-53 Prototype architecture.

Previous simulation based work showcases benefits of spectrum coordination across different wireless networks. The prototype based deployment has been exploited to both corroborate the results obtained in simulation, and to further push the introduction of new coordination algorithms. First, we have demonstrated results, with particular focus on WiFi spectrum coordination [BSS+12] and then on WiFi/LTE coordination [SBS+15]. Based on these set of use cases, we have evaluated the potential of the proposed API framework.

The prototype consists of the following components extended to support external control APIs:

- OpenAirInterface (OAI): Open-source implementation of a LTE BS and UE. In the ORBIT testbed, USRP B210 [USRP] are used for LTE BS and UEs and Android devices as commercial grade UEs.
- Hostapd: Software access point running on any Linux platform with a supported WiFi card.
- bsControl: A custom controller implementing southbound control interfaces using hardware specific protocols (e.g. OpenFlow, SNMP, etc.) and that exposes a northbound REST API to control all exposable parameters of such hardware.

Use cases such as interference management and data path management across multiple technologies have been deployed on the presented prototype to evaluate the proposed design.

A.21 T5.1-TeC 5 UE Measurement Context (*DR-17*)

In this annex, we explain in detail the rationale for the functional extensions of UE context envisioned in Figure 4-41.

Additional mobility configurations: In order to satisfy various mobility requirements in different UCs (e.g. high mobility in “connected cars”, medium mobility in “dense urban information society”, or low mobility in “massive distribution of sensors and actuators”), the network may need to change the measurement configurations to be based on the device speed. For example, the network can configure the measurement intervals to be related to the device velocity.

Band/time specific configuration for neighbor cell measurements: As illustrated in [MET-II16-D61], neighbor link measurements in LTE are currently performed over cell specific reference signals (C-RSs). C-RSs are constantly broadcasted by all the cells and transmitted over the whole bandwidth. In the case of energy efficiency it is obvious that having signals broadcasted all the time and over the whole bandwidth disables the possibility of applying DTX cycles by the network to its power amplifiers, in order to save energy and control the level of interference being generated.

Consequently, although the current LTE C-RS approach gives the device flexibility in time/frequency measurement for neighbor links, it is expected that the network may configure the device to perform the channel signal measurements to be on a specific pair of time and frequency resources.

Space measurement configuration: In high frequencies, the signals are transmitted along a specific direction using the concept of beamforming, so the UE need to detect a beam that may have a narrow coverage, leading to a direction measurement configuration (in addition to time and frequency measurement configuration). Thus, the access node shall configure the UE with the angular space configuration (e.g., antenna directionality) in order to achieve better beam scanning.

Reporting interference across multiple AIVs: In current 3GPP standards, in order to allow users to access various networks and services ubiquitously, a UE may be equipped with multiple radio transceivers. For example, a UE may be equipped with LTE, WiFi, Bluetooth transceivers, and Global Navigation Satellite System (GNSS) receivers. In 5G, several AIVs may coexist and operate simultaneously for different network slices. Due to the extreme proximity of multiple radio transceivers within the same UE operating on adjacent frequencies or sub-harmonic frequencies, the interference power coming from a transmitter of the collocated radio may be much higher than the actual received power level of the desired signal for a receiver. This situation causes In-Device Coexistence (IDC) interference and is referred to as IDC problems. As a result, when a UE experiences IDC problems that it cannot solve by itself, a network intervention is required, that is, the UE sends an IDC indication via dedicated RRC signaling to report the IDC problems to the eNB.

However, this IDC interference mitigation scheme may not be efficient in future deployment scenarios. For example, it is not always possible for the access node to switch the device’s configured measurement frequency to another frequency which suffers from less interference. In such a case, the access node may need from the UE the report on the measurements of the interference among different active (running) AIVs as an important factor in RM. For example,

this can be accomplished by changing the operating resources (time, frequency, and space) assigned for each AIV, or by selecting the best suitable set of AIVs that achieves less interference given the current situation.

Maintaining multiple UE contexts: Due to multi-connectivity, the UE may need to establish multiple connections with multiple APs at the same time. Consequently, the device may need to maintain multiple UE contexts simultaneously. This may also be the case when several AIVs are activated simultaneously but each AIV may have its own specific template for the UE context (e.g., measurement reporting scheme).

Conceptually, permitting the device to maintain simultaneously multiple UE contexts with different templates may negatively impact both the device complexity and the efficiency of the inter-AIV switching (especially for more frequent inter-AIV switching, when the UE context transfers among the different AIVs). Consequently, a further study may be needed to assess the benefits of having unified UE context across different AIVs.

More accurate and updated information on positioning, CQI, and HARQ: In 5G, as required by the building block "Dynamic Traffic Steering" in Section 4.6 and the building block “Multi-Slice and Multi-Service Holistic RM” in Section 4.4, it is expected that the information about the position, CQI, and HARQ of the UE collected from the device should be known by the network to enable a more dynamic traffic steering and RM scheme.