Beyond 2020 Heterogeneous Wireless Network with Millimeter-Wave Small-Cell Access and Backhauling

Grant agreement n°619563

Deliverable D2.4

Dynamic self-organising network functional description and algorithms

Date of Delivery:	31 Dec. 2015 (contractual)	23 Dec. 2015 (actual)
Editor:	Nokia	
Participant(s):	IMC, NOKIA, UNIS, VTT, TST	
Work package:	WP2 – Networking functions and algorithms	
Dissemination:	Public (PU)	
Version:	1.0	
Number of pages:	138	

Abstract: This deliverable of MiWaveS project describes and specifies the networking functions and algorithms envisioned in the millimetre-wave (mmW) network including backhaul and access links. An overall heterogeneous network design is proposed to address a set of selected use cases and motivate the important aspects of control and data plane split and multi-hop transmission. A mathematical model for radio resource management in mmW mesh backhaul networks is formulated and potential algorithms are analysed. Dynamic routing with semi-static scheduling algorithm and practical mmW small-cell initial attachment procedure are studied for mmW backhaul network. Several important aspects related to the access link are presented, and factors affecting the energy efficiency and security in 5G HetNet analysed.

Keywords: heterogeneous network, system functionalities, radio resource management, routing, implementation requirements, complexity, power consumption, millimetre-waves.

Executive Summary

This document covers the study of the following objectives of MiWaveS WP2 (Networking functions and algorithms):

1. Scenarios and requirements for the self-organized heterogeneous network with millimetre-wave (mmW) small cells, integrated backhaul and access link;
2. Overall heterogeneous network design with flexible small cell-macro cell –relationships and dedicated mmW multi-hop backhaul network;
3. Radio resource management (RRM) theoretical aspects and algorithms with respect to heterogeneous networks (HetNet) with multi-hop backhaul link;
4. Function description and algorithms for the backhaul link and multi-hop relay;
5. Function description and algorithms for the access link;
6. Principles and functions for power consumption optimization;
7. Communications security and public safety aspects;
8. Smart beam steering in backhaul and access.

This deliverable describes and specifies the networking functions and algorithms envisioned in the mmW network including backhaul and access links. An overall heterogeneous network design is proposed to address a set of selected use cases and motivate important aspects of control data plane split and multi-hop transmission. A mathematical model for the radio resource management problem in mmW mesh backhaul networks is formulated and potential algorithms are analysed. Dynamic routing with a semi-static scheduling algorithm and a practical mmW small-cell initial attachment procedure for wireless backhaul networks is presented. Several important aspects related to the access link such as user equipment (UE) random access and optimal UE-cell association for load balancing, are also presented. Factors affecting the energy efficiency and security in 5G HetNet are analysed. Principles for traffic off-loading and small-cell power efficient operation are provided, and beam cell management method to support multi-connectivity and mobility is studied. In addition, the fundamentals of beam steering in both access and backhaul are covered.

Authors

IMC	Michael Faerber	michael.faerber@intel.com
IMC	Honglei Miao	honglei.miao@intel.com
Nokia/NSN	Jyri Putkonen	jyri.putkonen@nokia.com
Nokia/NSN	Jouko Kapanen	jouko.kapanen@nokia.com
TUD	Hsiao-Lan Chiang	hsiao-lan.chiang@tu-dresden.de
TUD	Tobias Kadur	tobias.kadur@ifn.et.tu-dresden.de
UNIS	Mehrdad Dianati	m.dianati@surrey.ac.uk
UNIS	Mehrdad Shariat	m.shariat@surrey.ac.uk
VTT	Jorma Kilpi	jorma.kilpi@vtt.fi
VTT	Kari Seppänen	kari.seppanen@vtt.fi
VTT	Tapio Suihko	tapio.suihko@vtt.fi

Table of Contents

List of Figures

List of Tables

List of Acronyms and Abbreviations

Term	Description
5G	5th generation mobile network, after LTE/LTE-A
AoA	Angle of Arrival
AoD	Angle of Departure
AP	Access Point
APRS	AP-specific Reference Signal
BS	Base Station
BH	Backhaul
BRS	Beam-specific Reference Signal
BW	Bandwidth
CA	Carrier Aggregation
CB	Conversion Block
CC	Centralized Controller
CDF	Cumulative Distribution Function
C-RNTI	Cell-Radio Network Temporary Index
CRS	Cell-specific Reference Signal
CSI	Channel State Information
CSI-RS	CSI Reference Signal
DB	Database
DeNB	Donor eNB
DMTC	Discovery Signal Measurement Timing Configuration
DO	Discovery Occasion
E2E	End-to-End
eNB	eNodeB
EC	European Commission
EPC	Evolved Packet Core
ESM	Energy Saving Management
FP7	Seventh Framework Program
FRR	Fast Re-route
FST	Fast Session Transfer
GBR	Guaranteed Bit Rate

GPRS	General Packet Radio Service
GTP	GPRS Tunnelling Protocol
HetNet	Heterogeneous Network
HSS	Home Subscriber Server
KPI	Key Performance Indicator
LOS	Line-Of-Sight
LTE	Long Term Evolution
LTE-A	LTE Advanced
MB	Multi-Band
MBSFN	Multicast-Broadcast Single-Frequency Network
MCS	Modulation and Coding Scheme
MIMO	Multiple Input Multiple Output
MME	Mobility Management Entity
mmW	Millimetre Wave
mmW-NCT	Millimetre Wave New Carrier Type
MPC	Multi-Path Component
MPLS	Multiprotocol Label Switching
MSC	Message Sequence Chart
MT	Multi-Technology
MTC	Machine Type Communication
NE	Network Element
NLOS	Non-Line-Of-Sight
NRT	Non Real-Time
OAM	Operations Administration and Maintenance
PCell	Primary Cell
PDCCH	Physical Downlink Control Channel
PRACH	Physical Random Access Channel
PSS	Primary Synchronization Signal
PtP	Point-to-Point
QoE	Quality of Experience
QoS	Quality of Service
RACH	Random-Access Channel

RAT	Radio Access Technology
RN	Relay Node
RF	Radio Frequency
R-PDCCH	Relay PDCCH
RRC	Radio Resource Control
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RT	Real-Time
SAP	Service Access Point
SCell	Secondary Cell
SDMA	Space Domain Multiple Access
SLA	Service Level Agreement
SSS	Secondary Synchronization Signal
SVD	Singular Value Decomposition
TDD	Time Division Duplex
TDMA	Time Domain Multiple Access
TTI	Transmission Time Interval
UC	Use Case
UE	User Equipment
UT	User Terminal
WiGig	Wireless Gigabit [Alliance]
WG	WMN gateway
WMN	Wireless Mesh Network
WP	Work Package

1. Introduction

MiWaveS project is structured in several work-packages (WPs) covering different aspects of mmW system definition, study of novel techniques related to the different parts of the resultant network, development of key hardware modules, testing some critical prototypes and disseminating the work. WP2, named Networking functions and algorithms, focuses on defining the network topology, networking functions and algorithms needed for the implementation of next-generation heterogeneous mobile networks.

This deliverable (D2.4, “Dynamic self-organising network functional description and algorithms”) elaborates on the system scenarios defined in earlier MiWaveS deliverables by more detailed functional descriptions and algorithms. These definitions and outcomes may be utilized in algorithm implementation for MiWaveS system demonstrator.

This document is structured in the following four main technical sections.

Section 2 presents the overall heterogeneous network design. A generic heterogeneous network architecture, comprised of multi-connectivity and multi-hop backhauling, is proposed in this section. The multi-connectivity in the support of control and data plane split is elaborated to address the roles of macro cell and mmW small cells for various network functions. The multi-hop backhauling with envisioned important features is summarized to motivate the detailed solutions. The section ends with discussion of security and public safety aspects.

Section 3 explores the radio resource management in heterogeneous networks. In particular, a decomposition methodology is introduced and subsequently its relevant sub-problems (interacting algorithms) are derived. The pros and cons of different alternatives are discussed and the proposed algorithms are evaluated through various simulations. In addition, a practical joint beam-frequency multiuser scheduling algorithm is proposed.

Section 4 defines the system functions for mmW-based backhaul (BH) networks. Self-organization, routing and scheduling solutions with the associated algorithms are proposed for the complementary contexts of a dedicated standalone mmW mesh backhaul and self-backhauling with small-cell relay nodes. Furthermore, we evaluate the impact of practical issues and hardware limitations on the design and performance of both hybrid beamforming and analogue beam switching in the backhaul links.

Section 5 defines the system functions for the access link. Firstly, a mmW user equipment (UE) random access method is proposed to enable the uplink transmit beam alignment with the small cell access point (AP). Secondly, different UE-small cell association approaches are investigated to provide traffic offloading benefits at different degrees. Next, a small cell on-off procedure is presented and analysed in terms of power saving efficiency. Then, a beam-cell management method to support multi-connectivity and mobility is presented, and lastly, it is shown that an efficient beam alignment algorithm for access links can be implemented even within a system having limited hardware resources.

Finally, a conclusion section closes the document, highlighting the key concepts discussed along the document, summarizing the most important outputs from the above-mentioned sections.

2. Overall heterogeneous network design

A heterogeneous network consisting of macro cells and ultra-densely deployed small cells is considered to be widely employed for the future 5G wireless network, wherein the mmW communication bands are adopted for both backhaul and access links. Dual or multi connectivity concepts [1] supporting control and data plane split have been standardized already in the LTE system [9], but those are envisioned to be further developed for 5G HetNet.

A generic 5G HetNet architecture comprised of overlay macro cells and dedicated mmW BH networks and mmW small cells is proposed in Section 2.1, where deployment examples of selected use cases based on the generic HetNet architecture are also detailed. The roles of macro cells and mmW small cells are further elaborated in Section 2.2, especially in light of control and data plane split; mmW specific functionalities are addressed to cope with different challenges envisioned in considered use cases. Finally, in Section 2.3, the important features of multi-hop backhaul networks with a mesh structure are discussed to motivate the detailed solutions proposed to mitigate the challenges and achieve identified objectives.

2.1 Heterogeneous network architecture

Heterogeneous global cellular mobile networks are envisioned to evolve in phases and to be based on 3GPP technologies. In the first phase, the current LTE network capacity will be increased using a new mmW small cell (SC) layer; this phase is the main scope of the MiWaveS project. In a second phase, the RAN architecture will change into a new 5G architecture. LTE will remain and complement new 5G technologies in areas where there is no 5G macro coverage. The capacity as well as mmW access coverage will be further expanded. 5G will not only provide totally new radio interfaces but also introduce new types of services and enable new applications. The network will provide seamless mobility across radio interfaces using multi-connectivity, and new 5G specific services.

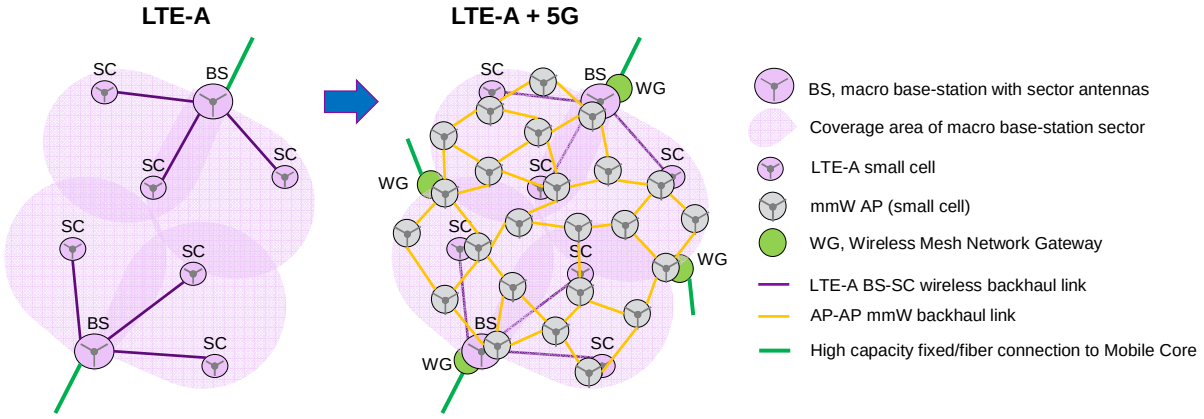


Figure 2-1: HetNet architecture evolution

Generic heterogeneous network architecture

Figure 2-2 illustrates the heterogeneous network (HetNet) structure consisting of underlying macro BS network with small cell overlay implemented by the MiWaveS' mmW access and backhaul network, possibly in combination with existing LTE small cell technology. mmW backhaul links may be implemented either with dedicated backhaul radios (yellow BH-BH links in the figure) or inband by sharing the common radio resources between the backhaul link (grey AP-AP links) and the access link (blue AP-UE links). Dedicated BH links and shared (inband) AP/access links may operate either in 60 GHz or in 70/80 GHz band.

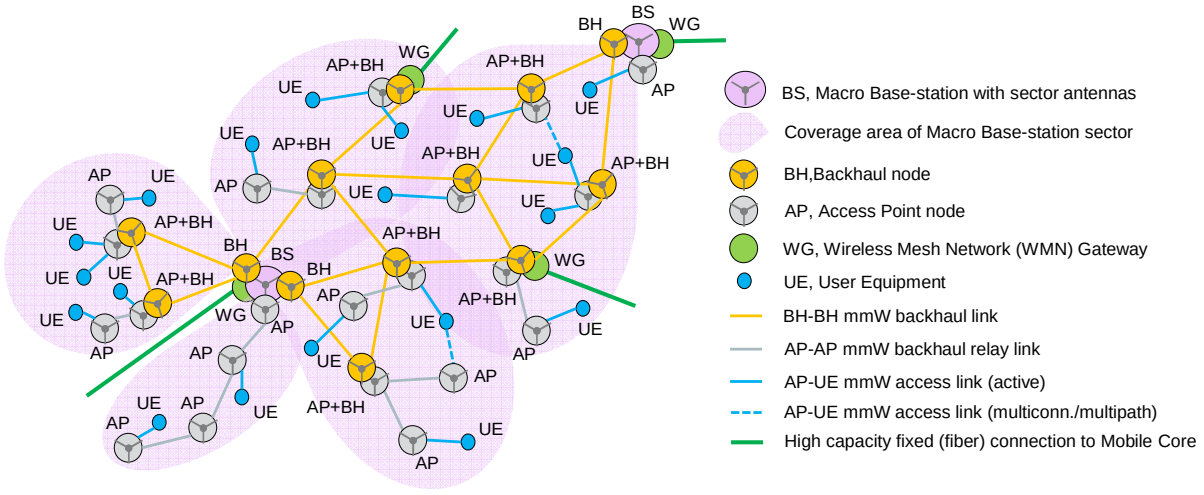


Figure 2-2: HetNet architecture with mmW small-cell layer.

High capacity backhaul transport connection between mobile core gateways and Wireless Mesh Network (WMN) gateway (WG) points carries the aggregated macro and SC layer traffic. Macro radio layer provides reliable ubiquitous control plane for all NEs and user equipment, multi-connectivity as well as overall network coverage and user plane connectivity for the fast moving UEs. The aspect of control and data plane split is more detailed in Section 2.2. Macro BS controlled dual/multi-connectivity is utilized to realize the bearer split and traffic offloading via the SC-layer APs.

The common HetNet architecture and macro-SC relationship apply to all MiWaveS use cases, see Annex C, even though certain differences of emphasis and challenges may occur, see **Table 2-1** below.

Table 2-1: Applicability of the proposed architecture to MiWaveS use cases.

Use Case	Focused advantages	Challenges
UC1: urban street-level outdoor mobile access and backhaul system	Ultra high capacity provision for backhaul and access network	Reliable backhaul network, especially with multi-hop BH connection; UE mobility and relevant cell/beam handover.
UC2: massive public events and gatherings	Ultra high capacity provision for backhaul network	Self-organized installation and possibly moving backhaul network.
UC3: indoor wireless networking and coverage from outdoor	High capacity provision for backhaul and access network	Seamless handover between indoor and outdoor.
UC4: rural detached small-cell zones and villages	High capacity provision and economically viable installation of backhaul network	Reliable backhaul network and self-organized installation. Needed macro coverage.
UC5: hotspots in shopping malls	High capacity provision of backhaul and access network	Combination of UC2 and UC3.

2.1.1 Network deployment

Due to the high number of network elements (NE) in 5G network, any manual on-site configurations or adjustments are economically infeasible. This means that when a new AP has been installed and the power switched on, the new AP shall automatically discover the neighbors and perform the required self-configuration actions to become part of the backhaul network, as well as

fetch operator-specific parameters for access links. Furthermore, a 5G small cell network should not only automatically establish the links but also actively reconfigure them, e.g. in case of NE or link failure, and compensate for any deflections (e.g. caused by mast twist or sway).

The point in time and urgency to set-up a new AP, i.e. discover the neighbours, reconfigure the routes etc., depends on the use case in question: usually it may be timed to happen during the low traffic hours (e.g. at night time), but e.g. in case of moving sport events (UC2), more immediate and continuous actions are needed.

The method used for neighbour discovery (BH/AP) depends e.g. on the beam steering capabilities of the nodes and possibilities to utilize the macro layer assistance. These issues are examined and alternative procedures are proposed in section 4.

Deployment examples

Several deployment examples with respect to some UCs are presented in the following. The basic principle is to take a two-step approach, where the first step is to determine the cell coverage based on the required traffic demand and the data transport capacity, i.e., link budget, of the access link, and the second step is to determine BH link coverage, especially the BH hop length, based on the aggregated BH traffic demand and BH link capacity. The main focuses of UC3 are not about deployment method, but rather a specific link installation technique to mitigate the severe attenuation caused by windows/walls and a seamless handover between indoor and outdoor. From the deployment point of view, UC5 is in synergy with UC1 and UC2. As such, this section focuses on UC1, UC2 and UC4. According to [2], the envisioned traffic demand and densities are summarized in the following table.

Table 2-2: Use case relevant traffic requirements [2].

Use Case	User experienced data rate (also at cell edge)	Connection density	Traffic density (Connection density * user data rate)	Remark
UC1: urban street-level outdoor mobile access and backhaul system	DL: 300Mbps UL: 50Mbps	200-2500/km^2	DL:750Gbps/km^2 UL: 125Gbps/km^2	No.1 in [2], Device density is 2000-25000/km^2, 10% activity factor
UC2: massive public events and gatherings	DL: 25Mbps UL: 50Mbps	150,000/km^2 (30,000/stadium)	DL: 3.75Tbps/km^2 UL: 7.5Tbps/km^2	No.3 in [2], Stadium of typical area 0.2km^2, 100000 persons, 30% activity factor.
UC4: rural detached small-cell zones and villages	DL: 50Mbps UL: 25Mbps	100/km^2	DL: 5Gbps/km^2 UL: 2.5Gbps/km^2	No.4 in [2]

Assuming the UC1 with 50 m AP-to-AP distances in the HetNet architecture of **Figure 2-2**, the network configuration of **Figure 2-3** can be defined for capacity calculation and transport dimensioning purposes.

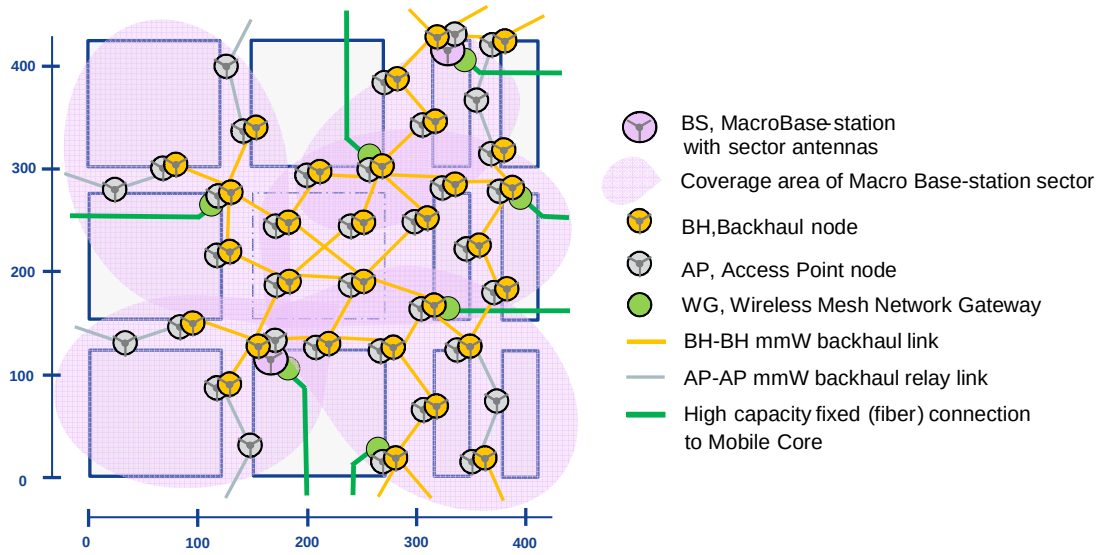


Figure 2-3: Network configuration example for Use Case 1 dimensioning.

The maximum radio system capacity and traffic density in **Figure 2-3** example depend e.g. on the backhaul (BH) radio characteristics, like power (energy), antennas (type and gain), modulation and coding scheme (complexity) and channel bandwidth (regulation). When each BH node is connected to its neighbour nodes with directional point-to-point mmW radio links and using a concept of shared radio resources (one common RF unit per sector, see section 4.2 for further details) the BH link capacity, e.g. 10 Gbps (Annex A, BE1), is divided between the connected links. The system capacity is also proportional to the number of WMN Gateway (WG) points. The system capacity can be optimised from BH point of view by setting the number of BH sectors (links) to such that required capacity can be transported to and from WGs in a certain area.

It is shown in the link budget calculations in Annex A that the cell radius (r) of 25 m can fulfil the user data rate demand of 300 Mbps in downlink and 50 Mbps in uplink (**Table 2-2**, UC1), respectively. Aggregating 8 carriers in DL (AV4) and UL (AV2) leads to maximum capacities of 5280 Mbps and 1406 Mbps within a sector, respectively. We can assume that UEs are evenly distributed within a round cell and adaptive MCS is in use, so the average UE data rate can be used to estimate network load. From the envisioned traffic demand in **Table 2-2**, the cell coverage of about 2000 m^2 (r=25m) may need to serve the aggregated traffic of up to 1.5 Gbps (750 Gbps/km^2 x 0.002 km^2) in downlink and 250 Mbps (125 Gbps/km^2 x 0.002 km^2) in uplink. Due to the BH capacity limit of 10 Gbps (BE1) and TDD (Time Division Duplex) operation with downlink to uplink traffic ratio of 6:1, the maximum number of multi-hop nodes, each aggregating with local traffic, will be 5 or 6. Moreover, when considering the maximum number of hops one needs to be aware of the end-to-end (E2E) latency. For instance, the overall radio access network latency comprised of all the latencies from macro BS via a number of BH nodes to UE, cannot exceed 10% of the E2E latency. Such requirement has implications on air interface design, especially the transmission time interval. It should be noted that the maximum number of hops due to both BH capacity and RAN latency requirement also determines the distance from the UE to the core network. In the deployment example of UC1, the maximum number of hops of 6 leads to 300 m distance of UE to macro BS or gateway. However, if a dedicated BH node with a large antenna array is adopted, the hop length between BH nodes can be relatively large, even about 300 m according to Annex A, which would naturally increase the maximum possible distance from the UE to the core network.

Similarly, the traffic requirements relevant to the UC2 are given in **Table 2-2**. Considering aggregation of 8 carriers and smaller than 10 m cell radius (Annex A), will lead to downlink and uplink capacity limits of 5280 Mbps (AV3) and 2640 Mbps (AV1). In TDD operation this will mean that the optimal downlink and uplink ratio given will be 1:4; namely 1/5 time slots are scheduled for downlink, and 4/5 time slot for uplink. The maximum link capacity and traffic demand result in the cell radius of 8 m, namely cell coverage of 200 m^2 with the support of 1.056 Gbps (5.28 Gbps x 1/5) and 2.112 Gbps (2.64 Gbps x 4/5) aggregated traffic for downlink and uplink, respectively. As such, the stadium of 0.2 km^2 may need 1000 APs, i.e., 100 persons served per AP assuming 30% user activity. With the BH capacity (link budget) limit of 10 Gbps (BE1), and considering the traffic requirements of 25 Mbps in downlink and 50 Mbps in uplink per active user, each aggregated BH branch can support at most 4 (10 Gbps/ (30 x 0.075) Gbps) APs accordingly. If 15 core network anchor nodes or macro BSs, each with the coverage of 13300 m^2, are installed in the stadium, then each core network anchor node needs to support 5 dedicated BH branch simultaneously.

As shown from **Table 2-2**, it is plausible that the main usage of mmW transport capability is to provide the BH link capacity. Due to the relatively long distance from the rural area in UC4 to more populated area where core network infrastructure or macro BS is normally available, dedicated mmW BH node equipped with a large antenna array is typically preferred to provide long hop length. According to link budgets in Annex A, the BH link with specific antenna array can offer data rates of up to 1.5 Gbps for the communication path of 1400 m with BPSK modulation scheme. One possible deployment for UC4 is to evenly partition the rural area of 1 km^2 into 5 sub-regions, each of which would serve 1.5 Gbps aggregated traffic density including downlink and uplink traffic. Then each dedicated BH node needs at least one BH link of 1.5 Gbps with the hop length up to 1.4 km towards the core network.

2.2 Multi-connectivity with control and data plane split

As outlined in Section 2.1, macro BS controlled dual/multi-connectivity is envisioned to be utilised to facilitate mmW small cell layer and the corresponding capacity boost. This is crucial due to directional transmissions in mmW bands; In particular, the possibility of multi-connectivity and separation between control and data channel can be leveraged to address many challenges identified in **Table 2-1**, for different UCs.

In this section, first, we elaborate challenges in the realisation of different UCs; then, we explore how (if possible) multi-connectivity featured with control and data plane split could be exploited to address such challenges.

2.2.1 Challenges of realising MiWaveS use cases

An Urban street–canyon use case (UC1) has been envisioned to support ultra-high capacity in dense environment comprising several mmW small cell APs. Backhauling is a key challenge in this UC as many APs may not have direct connection to the core network GWs (**Figure 2-3**). As a result, BH-BH (yellow) or AP-AP (grey) mmW radio links should be utilised in a multi-hop / mesh manner to connect mmW APs to the core network. Similarly, mobility is a challenging task in UC1 at the access link as a UE orientation change could trigger cell switching or radio link failure (RLF). This is due to the narrow beams required in mmW bands, impacting the perceived Quality of Experience (QoE). Cell switching trigger will not be exclusive to geographical cell boundaries but can happen in any locations within mmW small cells. Therefore, the cell switching on the access link should be complemented with relevant backhaul re-routing of traffic (in case required).

Supporting massive events (UC2) requires agility in self-alignment of mmW APs to maintain ultra-high capacity connection on the backhaul between mmW APs and the core network GW. In particular, portable APs may be required in this UC to support mass movements in some events. This requires efficient beam discovery and synchronisation with backhaul nodes or GWs.

Indoor coverage from outdoor via wireless connection (UC3) introduces another potential use case to support both backhaul and access mmW with high capacity. Seamless handover and intelligent offloading from outdoor to indoor or vice versa in this UC can be challenging. In particular, the indoor traffic (whether uplink or downlink) should be forwarded across (at least) two hops, i.e. one from outdoor AP to indoor AP and another between indoor AP and UE.

When it comes to supporting rural areas with high-capacity backhaul connections (UC4), self-configuration and reliability of backhaul link play key roles. In particular, long-range coverage in backhaul is required to make this UC a viable and cost-effective option.

In indoor hotspots (UC5, e.g. shopping malls), a combination of challenges as outlined in UC2 and UC3 shall be overcome. Therefore, integrating the corresponding solutions may be required.

2.2.2 Control- and data-plane split for multi-connectivity design

To address the challenges in heterogeneous environments comprising both macro cell and mmW small cells, multi-connectivity can be considered where macro layer provides control signalling to UEs (or mmW APs) while mmW APs focus on information delivery in data plane. In such cases, the network can benefit from (at least) two types of connectivity per node:

- Macro-cell connectivity (e.g. LTE air interface)
- Small-cell connectivity (mmW air interface)

The multi-connectivity can be partly characterised as outlined in 3GPP TR 36.842 [1] where UE connects and consumes resources at different levels: Master Cell Group, Secondary Cell Group.

Regarding the user plane architecture of multi-connectivity, two alternatives, 1A (no bearer split with independent PDCPs) and 3C (bearer split, S1-U terminates in Master Macro cell, independent RLCs) can be considered. In particular, option 1A is required for standalone mmW operation in case macro BS support is not available on data plane.

Regarding control plane architecture of multi-connectivity, alternative C1 can be used where single Radio Resource Control (RRC) entity residing in Master Macro cell will communicate with UE regarding RRC messages. Nevertheless, Master cell will coordinate with secondary mmW Cells before generating any RRC messages.

Multi-connectivity can also be exploited in another level to further coordinate and orchestrate the operations of mmW small cells in HetNet architectures via the parent macro cell(s). In this direction, in addition to co-tier interfaces between mmW APs, cross-tier interfaces can be defined between parent macro BS and small cell mmW APs.

In UC1, in the backhaul side, macro BSs can perform both link scheduling and backhaul traffic routing. For link scheduling, a parent macro BS can assist with implementation of semi-static scheduling sets (i.e. links that can be active simultaneously per time instance). This is due to network-wide information that can be aggregated in macro BSs unlike the mmW small cell APs with limited local visibility. Some possible information, to be collected, includes average channel information on the

backhaul beams between adjacent mmW APs or BH nodes as well as average queue size information, reflecting the quality of channel or the level of congestion in the APs / backhaul nodes. For backhaul routing (either BH-BH or AP-AP), a macro BS can help with route discovery to shortlist low-cost routes between different pair of nodes. Furthermore, macro BS could help out in fault recovery and congestion management. On the access side, a macro BS can coordinate the synchronisation of UEs with different mmW small cells (or more specifically to different beam-specific pilots) to support mobility. This facilitates fast and smooth beam switching between neighbouring mmW APs.

In UC2, deployment of portable mmW APs requires rapid backhaul self-organization based on discovery of neighbouring backhaul network nodes, by using beam steering, somewhat similarly to cell discovery between UEs and APs. Multi-connectivity can support coordination on timing and location of beam-specific pilots between APs in this UC.

For mobility support in UC3, macro BSs may have an active role via multi-connectivity to monitor the QoE or Energy consumption of supporting UEs on Macro links versus multi-hop connectivity on mmW band. This could enable intelligent QoE-aware / Energy-aware offloading strategies between the two.

The support of multi-connectivity in UC4 depends on the macro BSs coverage in rural areas. If there is stable and reliable control-level connectivity between mmW APs and macro BSs, similar procedures as UC2 can be employed. Otherwise, control-level connectivity should be implemented through standalone self-configuration / self-alignment features to establish a reliable backhaul link.

In UC5, similar procedures as outline for UC2 / UC3 can be incorporated in the system design.

2.3 Multi-hop transmission in the backhaul network

In the light of the proposed heterogeneous network architecture illustrated in **Figure 2-2**, multi-hop mmW backhaul network is a key building block for the future 5G wireless network. As summarized in **Table 2-1**, the main performance challenge for the backhaul data plane operation for the use cases considered in MiWaveS is throughput in general, and especially in the street–canyon use case (UC1) and the massive events use case (UC2). This, along with 5G's latency requirements, limits the feasible hop lengths. Besides mmW link degradation and breakage that occur in a (quasi) static backhaul, UC2 may introduce moving APs and BH link accordingly, which further exacerbates the issues in maintaining data plane integrity. As a result, a comprehensive backhaul networking solution is required to address all these challenges. Such a solution shall aim at achieving high throughput as well as providing flexible QoS with minimum possible power consumption. This will naturally contribute to the overall energy saving and reduced EMF objectives of the HetNet design. Taking these facts into account, the respective solutions to initial backhaul network establishment and backhaul network scheduling problems are presented in this section.

As described in the previous section, nodes in the backhaul network establish mmW links between each other and can achieve synchronization by using a neighbour discovery procedure depending on the beam steering capability of the nodes. In the backhaul self-organization, the nodes form a partial mesh structure, which enables resilience against node and link failures as far as the mmW mesh, overlaid by routing paths, has the required redundancy. The route diversity, inherently supported by multi-connectivity described in Section 2.2, enables protection against node and link degradations or failures.

In the inband self-backhauling solution, the network connection management (i.e. AP topology formation, link establishment between APs) is assisted by a macro BS as specified in Section 4.1. The routing topology will then be composed of semi-static AP relay chains between AP end nodes and the macro BS. The multi-hop paths are backed up by macro-level connectivity, using the LTE air interface. The dynamic scheduling on the mmW backhaul relay links can reuse the standard Relay PDCCH (R-PDCCH) procedures defined in LTE Rel-10. Here, the control channel may be established between the macro BS and AP (cross-carrier scheduling) or between neighbouring APs. User data can be tunnelled on the data plane by using GTP-U. It should be noted that the R-PDCCH may need to be redesigned, especially in terms of transmission time interval, to be compatible with the mmW RAT.

In the dedicated standalone backhaul, BH nodes search for new candidate member nodes for the backhaul mesh by using controlled beam steering during dedicated time slots and deliver reports of the discovered neighbours to a centralized controller. The controller then re-computes routes and link schedules for the new topology and re-configures the network. This type of mmW mesh routing and scheduling solution is described in more detail in Section 4.2. In the BH node mesh, virtual connections among BH nodes, and between the nodes and the macro BS, are provisioned by the centralized controller, which can be located in the macro cell. With centralized pre-computation of the routing paths and static contention-free link schedule, hop-by-hop forwarding in the mesh is deterministic in a stable and lightly loaded network. This combined with prioritized packet queuing in the nodes allows bounding the delay of the highest priority traffic. Load balancing of lower priority traffic can be achieved by applying link-state and congestion aware selection of alternative routing paths for traffic flows at the traffic source nodes. These features make it possible to constrain the end-to-end delay, increase network reliability, and provide QoS in general, which pertain to the common KPIs in the use cases.

Due to the traffic fluctuation over the time and according to the actual traffic demand, certain AP or BH nodes may be turned on and off to reduce the network power consumption. For instance, a two-hop BH connection with short hop lengths during peak traffic period may convert to a one-hop BH connection with longer hop length. The supported data rate on a single hop naturally decreases due to the rate drop caused by the larger coupling loss. This can be controlled by the macro BS which has the full knowledge of the routing topology of the network. As a result, the centralized scheduler/controller in the macro BS determines the routes of BH packets by taking into account all the above factors including traffic priority, latency requirement, load balancing and potential power consumption.

In the future 5G networks, mobility management may become integrated as part of the backhaul routing functionality. In the light of control and data plane split, UE mobility, and UE-specific traffic steering in general, is assumed to be taken care of by the macro cell layer, especially this may impact routing within the backhaul network. For inter-AP communications, the backhaul topology supports both “X2 mesh” (i.e., shortest path) connections and “X2 star” (i.e., via macro BS or edge gateway) connections. The high density of SCs, and the directionality and fragility of mmW access links, may imply pronounced requirements for efficient multicast forwarding in the backhaul, in order to support UEs’ multipoint connectivity and smooth handovers.

2.4 Security and public safety

MiWaveS heterogeneous network concept can contribute to security and safety in several ways. In the context of mobile communications, security addresses information security and the availability

of critical infrastructure. High antenna directivity, state-of-the-art technology required and short hop lengths make eavesdropping and interfering (jamming) of mmW radio links difficult, which contributes to increased security. High encryption and privacy standards of LTE and later 5G network will apply to MiWaveS network. Multi-layered heterogeneous network with possibility for mesh-type backhaul topology increases redundancy and reliability. Small energy consumption and sleep-mode (Sec 5.3) enable longer backup times in case of power grid outage. Easy installation and small size of network elements and self-organising deployment features enable network extensions and topology changes if needed for public safety or network security purposes.

MiWaveS use cases UC2 "Massive public events and gatherings" and UC4 "Rural detached small-cell zones and villages" can be considered the most relevant to tackle the challenge of public safety emergency communications missions. In UC2, the reference network topology is represented by a self-configurable backhaul network (SON) able to dynamically steer the network capacity towards the areas where users are. Radio links between nodes are assumed to be highly directional steerable beams while the same can apply to mobile user access. A low-frequency macro layer provides the control plane to a large area. It is assumed that, in the large scale, the telecommunications network is operational and high capacity is available to the area through long-haul optical fibres or via microwave link. In the context of public security, UC4 can be considered as an extension of UC2 to emergency communications cases where ad-hoc wireless backhaul chain is needed to a detached pocket.

Safety can be defined to be the control of recognized hazards to achieve an acceptable level of risk. This can take the form of being protected from the event or from exposure to something that causes health problems. One highly debated safety issue in the context of mobile communications is possible health hazards of radio wave emissions.

In case of small cell mmW networks, the hypothesis of MiWaveS is that total exposure of general public to radio frequency radiation is lower because of high-directivity beam-steering mmW antennas that offload transmissions from lower frequencies to mmW bands. With highly directive antennas, RF power is not radiated to all directions but only to the wanted direction and only when transmission is needed. So the total dose of tissue of general public is lower. Also the skin penetration of mmW is short (< 2 mm), which implies the energy for tissue heating is lower than at lower frequencies. While macro network is still active and needed in MiWaveS concept, its power density can be lower because it handles mainly low volume data and control traffic. This is an indicative hypothesis and will be verified in WP4 investigations; however, it is possible that conclusive results are not achieved during MiWaveS project.

3. Radio resource management in mmW networks

In Section 3.1 we propose utilisation of decomposition techniques to systematically design radio resource management (RRM) algorithms for backhaul and access in mmW heterogeneous networks. In addition, a practical joint beam-frequency multiuser scheduling algorithm is proposed in Section 3.2.

3.1 Radio resource management in heterogeneous networks

In the following subsections, we firstly provide an overview of the decomposition methodology and evaluation for RRM algorithm. Subsequently, we derive relevant sub-problems (interacting algorithms) addressing the pros and cons of different alternatives and finally explore the procedures in the devised algorithms and provide different sets of simulation analyses.

3.1.1 Performance requirements of radio resource management

The problem of RRM for mmW backhaul can be formulated as an optimization problem, which then can be systematically decoupled to a number of sub-problems. The performance of RRM for heterogeneous networks can be evaluated using different key performance indicators (KPIs) depending on the case under study, including:

- System capacity and stability metrics (e.g., area throughput, cell throughput and relevant cumulative distribution function (CDF) curves) to evaluate how close an algorithm operates to the boundary of joint capacity region between backhaul and access links.
- Fairness measures, to assess how fair an algorithm divides resources among different UEs (flows). Here, fairness can be evaluated based on the normalized CDF of per user throughput or delay. The CDF curve(s) can be compared based on a predefined fairness criterion or a utility function (e.g. proportional fairness).
- Quality of Service (QoS) measures that aim to ensure satisfactory performance based on target application services. These can include different packet delay measures and relevant CDF curves.
- Energy efficiency measures (e.g. bits per joule) that relate to the energy costs and environmental impacts of target solutions.

Additionally, each RRM algorithm may have certain costs in terms of complexity and signalling overhead that can be further analysed based on the required level of interactions among different network entities.

3.1.2 Alternative decomposition frameworks

In this subsection we present two alternative decomposition frameworks, each leading to development of different RRM algorithms based on the optimization model, and explore solution frameworks and related sub-problems. **Table 3-1** summarizes the notations that are used in formulations of RRM.

3.1.2.1 Node-centric case

The RRM optimization problem in node-centric case is defined as ($\forall d \in \boldsymbol{D}, \forall m \in \boldsymbol{M}, m \neq d$):

$$\max_{\boldsymbol{a},\boldsymbol{r}^d} \sum_s U_s(a_s), \quad (1)$$

$$a_s \leq \sum_{n:(m,n)\in \boldsymbol{L}} r^d_{m,n} - \sum_{o:(o,m)\in \boldsymbol{L}} r^d_{o,m}, \quad (2)$$

$$\left\{\sum_d r_{m,n}^d\right\} \in \mathbf{\Pi}. \quad (3)$$

It is worth noting $\left\{\sum_d r_{m,n}^d\right\}$ in (3) is equivalent to $\boldsymbol{r} = \{r_{m,n}\}$. As the capacity region $\mathbf{\Pi}$ will be near convex set [3] in presence of efficient beamforming strategies, we can employ dual decomposition techniques:

Table 3-1: Overview of mmW RRM optimization parameters

Parameters	Definition
$\boldsymbol{M}$	Set of nodes (APs or UEs)
$\boldsymbol{L}$	Set of links
$\boldsymbol{D}$	Set of destination nodes
$\mathbf{X}$	Set of paths (routes)
$\mathbf{\Pi}$	Capacity region
$\mathbf{\Omega}$	Stability region
$r_{m,n}$	Data rate on link $(m,n) \in \boldsymbol{L}$
$G_{m,n}$	Channel gain including beamforming effect
$p_{m,n}$	Allocated power to link $(m,n) \in \boldsymbol{L}$
$p_{max,m}$	Maximum power of node m
a_s	Admitted flow in source node $s \in \{1,2,3,\dots,S\}$
$r_{m,n}^d$	Data rate allocated on link (m,n) towards destination d
$\boldsymbol{r} \coloneqq \{r_{m,n}\}$	Transmission rate vector
$\boldsymbol{a} \coloneqq \{a_s\}$	Admitted source flow vector
$U_s(a_s)$	Utility function of source s
$\boldsymbol{r}^d \coloneqq \{r_{m,n}^d\}$	Routing / scheduling vector towards destination d
a_s^χ	Fraction of admitted flow at s (towards d) via path χ
$B_{m,n}^\chi$	Incidence matrix of link (m,n) and path χ

Here, each source node like s can be characterized as: $s = (m,d)$. Therefore, considering the problem formulation as in (1)-(3), we can decompose the problem by forming partial Lagrangian with respect to constraint set (2), subject to (3):

$$L(\boldsymbol{a},\boldsymbol{r^d},\boldsymbol{\lambda}) = \sum_s U_s(a_s) - \sum_s \lambda_s a_s + \sum_{m,d} \lambda_m^d \left(\sum_{n:(m,n)\in\boldsymbol{L}} r_{m,n}^d - \sum_{o:(o,m)\in\boldsymbol{L}} r_{o,m}^d\right), \quad (4)$$

where $\boldsymbol{\lambda} \coloneqq \{\lambda_s\} = \{\lambda_m^d\}$ denotes the set of Lagrangian multipliers (price variables) corresponding to (2). It is worth noting the notation for price variables λ_s and λ_m^d are equivalent to each other and are differentiated in (4) for the sake of interpretation when the problem decouples at later subsections. The dual function and dual problem Φ can be formulated as below subject to feasibility constraint (3):

$$g(\boldsymbol{\lambda}) \coloneqq \max_{\boldsymbol{a}\geq\boldsymbol{0},\boldsymbol{r^d}\geq\boldsymbol{0}} L(\boldsymbol{a},\boldsymbol{r},\boldsymbol{\lambda}), \quad (5)$$

$$\Phi = \min_{\boldsymbol{\lambda} \geq \mathbf{0}} g(\boldsymbol{\lambda}). \tag{6}$$

By applying standard decomposition, problem (5) can be decoupled at two levels based on the admitted traffic and rate variables:

3.1.2.1.1 Admission control level

To maximize Lagrangian in (5) , by partial derivative of Lagrangian with respect to a_s, we have:

$$U'_s(a_s^*) - \lambda_s = 0. \tag{7}$$

In other words,

$$a_s^* = U'^{-1}_s(\lambda_s). \tag{8}$$

This shows that the admitted level of traffic per source node s should be adjusted based on the price variables λ_s. Furthermore, the admitted source flow will be dependent on the inverse of marginal utility function, i.e. $U'^{-1}_s(.)$.

3.1.2.1.2 Packet forwarding /scheduling level

To maximize Lagrangian in (5) with respect to $r_{m,n}^d$, we have:

$$\max_{\boldsymbol{r^d} \geq 0} \sum_{m,d} \lambda_m^d \left(\sum_{n:(m,n)\in \boldsymbol{L}} r_{m,n}^d - \sum_{o:(o,m)\in \boldsymbol{L}} r_{o,m}^d \right). \tag{9}$$

The above sub-problem can be equivalently formulated [4] as:

$$\max_{\boldsymbol{r^d} \geq 0} \sum_{m,n,d} r_{m,n}^d \left(\lambda_m^d - \lambda_n^d \right). \tag{10}$$

This results in the following packet forwarding/ scheduling policy with respect to transmission variables:

$$\boldsymbol{r}^* \in \arg\max_{\boldsymbol{r}\in \boldsymbol{\Pi}} \sum_{(m,n)\in \boldsymbol{L}} w_{m,n} r_{m,n}, \tag{11}$$

$$w_{m,n} = (\lambda_m^{d^*} - \lambda_n^{d^*}) \text{ where } d^* \in \arg\max_{\boldsymbol{d}} (\lambda_m^d - \lambda_n^d). \tag{12}$$

In other words, node m should allocate the achievable data rate on link (m,n) to the traffic towards destination d^* with maximal differential prices as (12), (i.e. $r_{m,n}^{d^*} = r_{m,n}$), such that (11) is satisfied. This solution can be interpreted in two sub-routines:

- Node m needs to serve the traffic associated with maximal differential prices as in (12) provided that the associated link (m,n) gets scheduled: $r_{m,n}^{d^*} = r_{m,n}$.
- Link (m,n) will get scheduled provided that it fits into the maximiser vector $\boldsymbol{r}^*$ in (11).

Here, the link scheduling sub-problem (11) can couple the problem across the network links in order to find the maximum weighted sum-rate among the feasible vectors ($\boldsymbol{r} \in \boldsymbol{\Pi}$); each candidate vector should still satisfy the duplexing constraints even in the absence of inter-link interference. On the other hand, the packet forwarding strategy according to (12) is a local optimization, only requiring the pricing variables of each node with respect to the adjacent neighbours. It should be noted that in this form of decomposition routing has also effectively been absorbed in admission control of (8) and

links scheduling and packet forwarding of (11) and (12). Therefore, no explicit route discovery or incidence matrix is required.

3.1.2.1.3 Sub-gradient update

The dual problem as in (6) can be solved based on sub-gradient update. The updating sequence per set of dual variables can be formulated as follows:

$$\lambda_m^d(t+1) = \left\{\lambda_m^d(t) - \alpha\left(\sum_{m:(m,n)\in \boldsymbol{L}} r_{m,n}^d(t) - \sum_{o:(o,m)\in \boldsymbol{L}} r_{o,m}^d(t) - a_s(t)\right)\right\}^+. \quad (13)$$

Here, $\alpha > 0$ is the positive step size of the update and should be either diminishing or sufficiently small to guarantee convergence. The evolution of the dual prices as in (13) is highly correlated to the evolution of queues q_m^d per backhaul nodes (or access nodes) towards different destinations:

$$q_m^d(t+1) = \left\{q_m^d(t) - \left(\sum_{m:(m,n)\in \boldsymbol{L}} r_{m,n}^d(t) - \sum_{o:(o,m)\in \boldsymbol{L}} r_{o,m}^d(t) - a_s(t)\right)\right\}^+ . \quad (14)$$

As it can be seen, $\lambda_m^d(t) = \alpha\, q_m^d\,(t)$. In other words, the evolution of queue sizes per node towards different destinations implicitly acts in line with the dual prices. Intuitively, higher queues sizes (prices) lead to reducing the level of admitted traffic per node (towards a destination) according to (8) whereas it will increase the chance of the corresponding link (and the traffic) to be scheduled according to (11)-(12). This leads to stabilization of prices (queue sizes) in long term at an operation point that satisfies (5)-(6). As a result, the efficiency and stability are jointly achieved.

3.1.2.2 Path-centric case

The RRM optimization problem in path-centric case is defined as:

$$\max_{\boldsymbol{a}^\chi, \boldsymbol{r}} \sum_s U_s(a_s), \quad (15)$$

$$a_s = \sum_\chi a_s^\chi, \quad (16)$$

$$\{\sum_{s,\chi} B_{m,n}^\chi a_s^\chi\} \in \boldsymbol{\Pi}. \quad (17)$$

If we replace a_s from (16) into (15) and considering $\sum_{s,\chi} B_{m,n}^\chi a_s^\chi \le r_{m,n}$, we can form partial Lagrangian subject to feasibility constraint ($\{r_{m,n}\} \in \boldsymbol{\Pi}$), as below:

$$L(\boldsymbol{a}^\chi, \boldsymbol{r}, \boldsymbol{\psi}) = \sum_s U_s\left(\sum_\chi a_s^\chi\right) - \sum_{(m,n)\in \boldsymbol{L}} \psi_{m,n} \sum_{s,\chi} B_{m,n}^\chi a_s^\chi + \sum_{(m,n)\in \boldsymbol{L}} \psi_{m,n} r_{m,n}\,, \quad (18)$$

where $\boldsymbol{\psi} \coloneqq \{\psi_{m,n}\}$ denotes the set of Lagrangian multipliers (prices) $\psi_{m,n}$ corresponding to (17). The dual function and dual problem Φ can be formulated as below subject to feasibility constraint:

$$g(\boldsymbol{\psi}) \coloneqq \max_{\boldsymbol{a}^\chi \ge \boldsymbol{0}, \boldsymbol{r} \ge \boldsymbol{0}} L(\boldsymbol{a}^\chi, \boldsymbol{r}, \boldsymbol{\psi}), \quad (19)$$

$$\Phi = \min_{\boldsymbol{\psi} \ge \boldsymbol{0}} g(\boldsymbol{\psi}). \quad (20)$$

By applying standard decomposition, problem (19) can be decoupled at two levels based on the fraction of admitted flows (routing variables) and rate variables:

3.1.2.2.1 Routing level

To maximize Lagrangian in (19) with respect to $a_{s\chi}$, we have:

$$\max_{a_s^{\chi}\geq 0}[U_s\left(\sum_{\chi} a_s^{\chi}\right) - \sum_{\chi} a_s^{\chi} \sum_{(m,n)\in \boldsymbol{L}} \psi_{m,n} B_{m,n}^{\chi}\,]. \quad (21)$$

In (21), the second term $\sum_{(m,n)\in \boldsymbol{L}} \psi_{m,n} B_{m,n}^{\chi}$ can be interpreted as the cost of choosing path χ out of possible paths (routes) [5].

3.1.2.2.2 Scheduling level

To maximize Lagrangian in (19) with respect to $r_{m,n}$, we have:

$$\boldsymbol{r}^{*} \in \arg\max_{\boldsymbol{r}\in \boldsymbol{\Pi}} \sum_{(m,n)\in \boldsymbol{L}} \psi_{m,n} r_{m,n}. \quad (22)$$

As it can be seen, in this form of decomposition, the weighted sum-rate maximization reappears similar to (11). However, this time the weighting factor reflects the dual prices $\psi_{m,n}$ based on the cost of each route.

3.1.2.2.3 Sub-gradient update

The dual problem as in (20) can be solved based on sub-gradient update. The updating sequence per set of dual variables can be formulated as follows:

$$\psi_{m,n}(t+1) = \left\{\psi_{m,n}(t) - \alpha\left(r_{m,n}(t) - \sum_{s,\chi} B_{m,n}^{\chi} a_s^{\chi}(t)\right)\right\}^{+}. \quad (23)$$

3.1.2.3 Practical issues

Node-centric decomposition leads to a problem formulation with local price variables $\lambda_s(\lambda_m^d)$ that regulate the admission level of traffic per node (either in backhaul or access links). The price variables can be associated with the queue sizes per node towards different destinations. This enables calculation of differential prices that are required for packet forwarding (via local exchange of messages between adjacent nodes). Furthermore, the routing is effectively absorbed in admission control, packet forwarding and scheduling sub-algorithms. As a result, no explicit route discovery is required.

However, the scheduling sub-problem as (11) couples the optimization across different nodes in the system. This requires adopting efficient implementations to simplify this integrated problem. Another consideration for this form of decomposition is related to the separate queues required towards different destinations as outlined in the algorithm. In other words, the hardware complexity of backhaul (or access) nodes will be scaled based on the total number of destinations. Therefore, it suits for scenarios with limited destination sets or if the number of destinations is reduced via clustering.

Path-centric decomposition, on the other hand, leads to a different set of prices $\psi_{m,n}$ per link. This set of prices is associated with the cost of each path according to the incidence matrix, i.e. $\sum_{(m,n)\in \boldsymbol{L}} \psi_{m,n} B_{m,n}^{\chi}$. Therefore, route discovery and the outcome incidence matrix $B_{m,n}^{\chi}$ are crucial elements in this form of decomposition and they require network-wide knowledge. Destination-based traffic differentiation (separate queues) is not required in this case. Nevertheless, the scheduling problem will be still coupled across different nodes according to (22).

3.1.3 Devised RRM algorithm(s) and performance analyses

As outlined above, node-centric decomposition leads to a solution framework for the RRM problem without explicit route discovery. Assuming that we reduce the destination set via clustering the nodes into super nodes, node-centric decomposition can facilitate promising and flexible RRM algorithms in particular for the mmW backhaul links. In this case, each super node will represent all the attached nodes (e.g. attached UEs) per backhaul node or mmW AP.

In this section, we discuss different steps of such RRM algorithms, taking into account the signalling / architectural implications. Next, we will evaluate the performance of the algorithms in different settings via simulations.

3.1.3.1 Steps of the RRM algorithm

As discussed above, node-centric RRM algorithms comprise two sets of procedures:

Local Procedures:

- Each backhaul node upon joining to the network should initialize an arbitrary set for the price variables $\boldsymbol{\lambda}$. This will be set based on the buffer size of each queue after initialization stage.
- Each backhaul node will adjust the admitted source flows considering the prices (queue sizes) at the previous stage as in (8).
- Each backhaul node will communicate with adjacent nodes to identify their corresponding prices per specific traffic flow per time interval.
- Each backhaul node will locally calculate the differential prices and identifies the maximum values per time interval as in (12).
- Each backhaul node estimates the quality of channel w.r.t. adjacent neighbouring nodes.

Network-wide procedures:

- Each backhaul node will communicate the by-product of maximal differential prices and the channel qualities to a central scheduling entity. This by-product reflects the reward of scheduling the corresponding link with the optimal traffic flow per backhaul node.
- The central scheduling entity calculates the optimum set out of all links taking into account the duplexing constraints or interference constraints (if any) as in (11).
- The corresponding schedule is conveyed back to the backhaul nodes for the next time interval.

Finally, the packets are forwarded per link based on the link scheduling decision and the corresponding prices are locally updated as in (13) and (14).

Remark1: The central scheduling entity can reside within the macro BS supporting a cluster of mmW BH nodes or APs. In this case, the C-plane between macro BS and mmW APs (or backhaul node) can be utilized to exchange signalling information as outlined in Section 2.2. The size of cluster can be carefully set to limit the scale of required signalling. Alternatively, one of the BH nodes or APs with good visibility w.r.t. the rest of a cluster can take the master role for link-scheduling via C-plane. Fully decentralized alternatives can also be possible. However, the cost of synchronization / signalling might become prohibitive.

Remark 2: As shown above, the local procedures are more associated with routing and packet forwarding whereas network-wide procedures are more related to link scheduling in node-centric case. To save on the cost of either routing or link scheduling (in terms of signalling overhead and

complexity), it is possible to combine their dynamic procedures with static but random selection. Therefore, 3 alternative solutions can be explored:

- Case 2: Static Link Scheduling- Dynamic Routing
- Case 3: Dynamic Link Scheduling – Static Routing
- Case 4: Dynamic Link Scheduling – Dynamic Routing

Remark 3: Another possible alternative is where each backhaul node relies on its own prices to calculate the by-product reward, i.e. $d^* = \mathrm{argmax}_d(\lambda_m^d)$. In this manner, the signalling cost of local procedures can be saved and a Modified Dynamic Link Scheduling –Dynamic Routing (Case 5) will form.

3.1.3.2 Simulation study

Here, we provide an evaluation of the performance of different types of RRM cases for mmW backhauling as outlined above. The performance of fully Static Link Scheduling- Static Routing (Case 1) is also included for the benchmarking purpose. We consider a two-flow scenario as shown in **Figure 3-1**. It is assumed that the traffic flows are received in Ingress Node and then transferred across the defined paths to the corresponding users. Per flow traffic is generated with independent exponentially distributed average waiting time between the packets for both flows.

The simulation system model is adopted in line with urban street-level outdoor mobile backhaul system as one of the outlined use case scenarios (UC1) in Annex C where mmW APs are deployed on light poles in a chain / mesh topology. The inter-site distance of the light poles is set to 100 m with the street width of 10 m.

The antennas are assumed to cover sectors of 40 degrees towards each neighbouring AP in azimuth. Each Antenna is equipped with one RF chain for backhaul and supports up to 10 beams with gain of 30 dBi. We consider independent RF Unit and frequency of operation for access beams in this study and mainly focus on the dedicated BH part. With respect to channel modelling two Scenarios are adopted:

- Scenario 1: Assuming a perfect LoS channel between neighbouring mmW APs / BH nodes.

- Scenario 2: A more detailed channel model as outlined in [6] where we assume up to three scatters per link. The link budgets and path loss are adjusted in line with option BV2 of Annex A.

The link capacities are assumed to be stable per scheduling interval due to very short transmission time intervals (TTIs) envisioned for mmW networks. However, in line with RF constraints, we assume each BH node will either transmit or receive (in one designated direction) per interval. Here, the possible directions of a mmW beam are shown in dashed lines whereas the scheduled beams are in solid lines.

Here, we investigate the impact of different levels of traffic load on BH End to End (E-E) delay for all cases. The reported values are averaged over 10,000 snapshots.

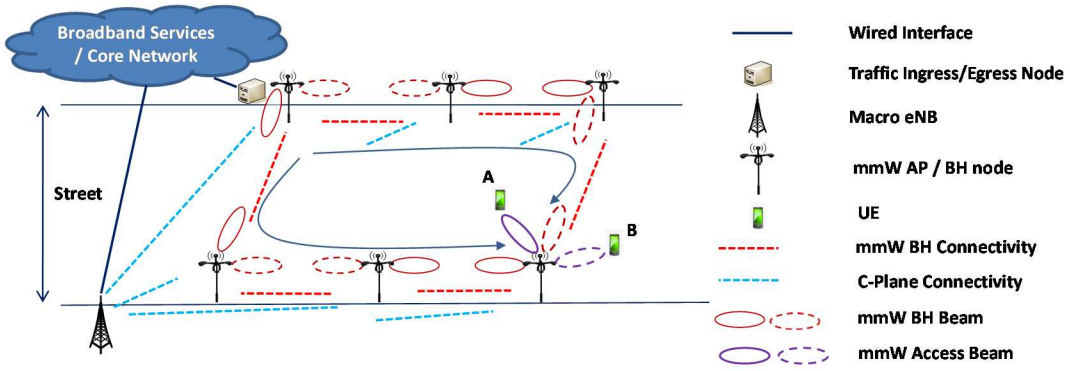


Figure 3-1: Simulation topology.

Figure 3-2 shows the performance of different RRM algorithms based on the backhaul E-E delay vs. loading factor. **Figure 3-3** shows the CDFs of E-E delay in medium loading (50%) for similar algorithms. Both figures are based on perfect channel assumption (Case1).

As it can be seen, by increasing the level of load, the average E-E delay increases for all cases. For Case 1, due to rigidity of both RRM functions, the E-E delay significantly increases in response to the traffic load. In particular, the delay distribution between the flows also suffers from unfairness in higher levels of load. For Case 2 and Case 3, dynamic routing (or dynamic link scheduling) can substantially improve the E-E delay compared to Case 1. However, the slope of delay expansion can be larger in highly-loaded cases (70%+). Case 4, provides the best performance compared to the rest. The E-E delay increases slowly versus loading level, in this case, due to dynamic adaptation of both link scheduling and routing functions. Case 5 presents a very promising performance compared to Case 4 in this scenario and it can be considered as a more viable option in terms of signalling.

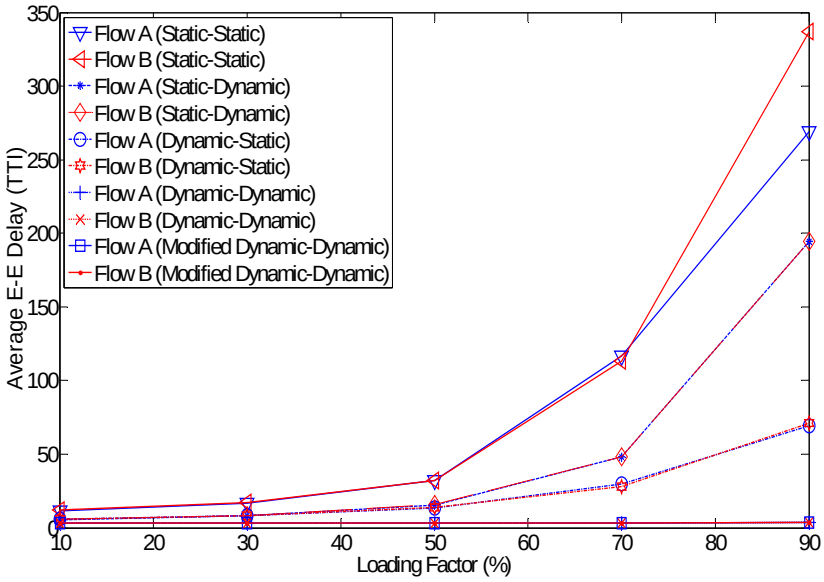


Figure 3-2: Delay vs. loading factor of different algorithms for perfect channel (scenario 1).

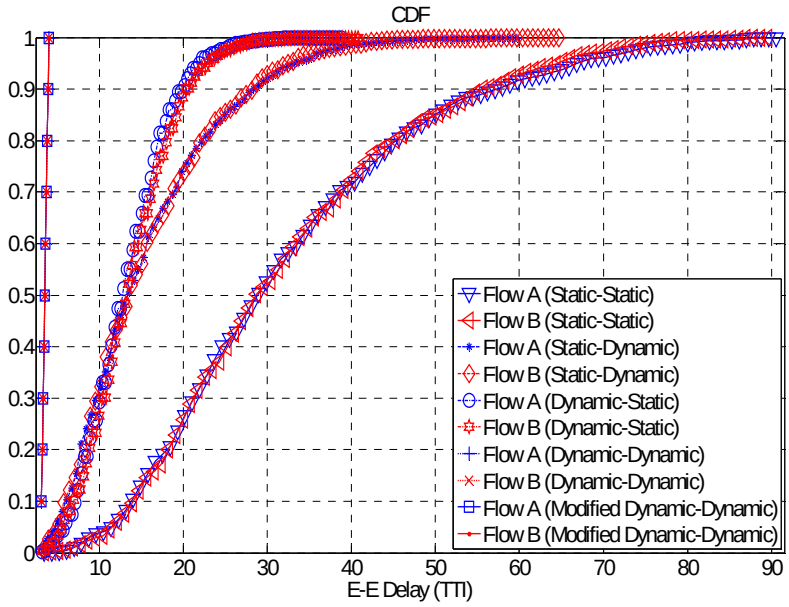


Figure 3-3: CDFs of delay for 50% loading for perfect channel (scenario 1).

Figure 3-4 shows the performance of different RRM algorithms based on the backhaul E-E delay vs. loading factor for the detailed channel (Scenario 2). **Figure 3-5** also shows the CDFs of E-E delay in medium loading (50%) for a similar scenario.

As it can be seen, similar trends as the perfect channel (Scenario 1) is observed here. However, due to occasional degradations in channel condition, higher levels of E-E delay will be perceived for similar traffic loading points compared to Scenario 1.

Considering the simulation results, Case 2 (Static Routing–Dynamic Scheduling) can be a commendable option for low to medium network loading due to substantial reduction in network-wide signalling whereas Cases 3 and 5 can better adapt to medium to highly loaded cases at the cost of more network-wide signalling with the central scheduling entity.

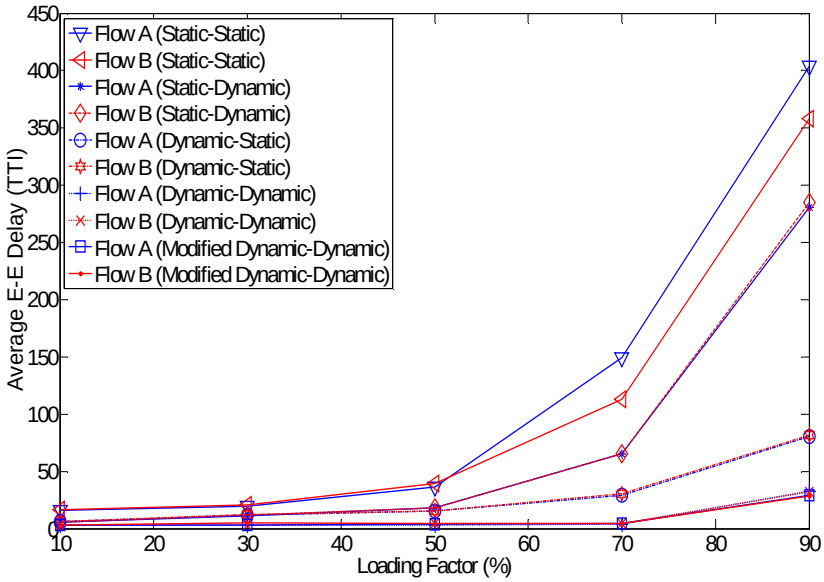


Figure 3-4: Delay vs. loading factor of different algorithms for detailed channel (scenario 2).

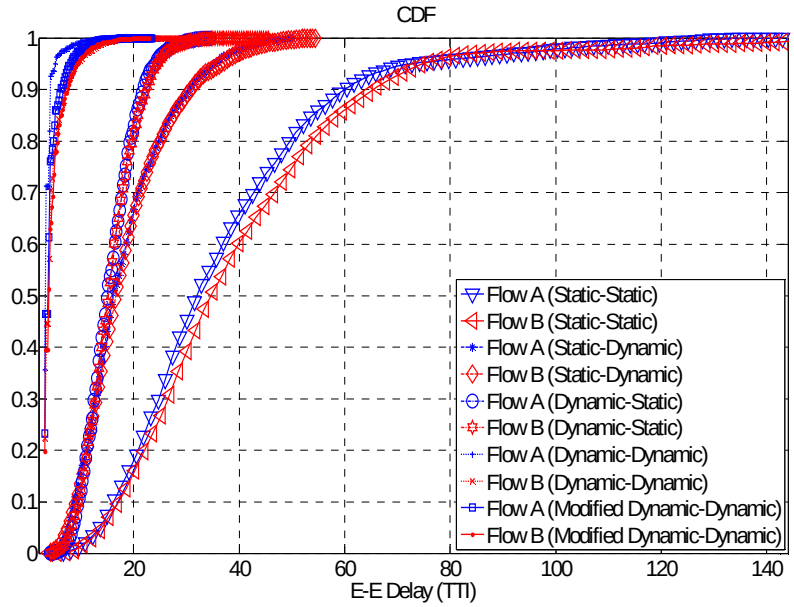


Figure 3-5: CDFs of delay for 50% loading for detailed channel (scenario 2).

3.2 Joint beam-frequency multiuser rate scheduling for downlink multiplexing

To mitigate the severe path loss caused by the very high frequency, beamforming techniques become very critical to achieve target link budget. It is envisioned that the pencil beam with very narrow beamwidth, e.g. 5 to 10 degrees in access downlink offering high beamforming gain, is to be widely used in mmW systems. To transmit multiple parallel beams, mmW AP can be equipped with multiple beamformers, each of which illuminates a different beam direction.

In addition to higher throughput, the future 5G system also targets to lower latency and better spectrum efficiency. As such, it is desired that the mmW system can schedule as many users as possible in the same transmission time interval (TTI). In addition to the shared frequency resources, the mmW system is also limited by the restrictions of physical analogue beamformer.

The notations used in this section are as follows. Uppercase and lowercase boldface letters denote matrices and vectors, respectively, $(.)^{\mathrm{T}}$ denotes the transpose operator. While the letter-like symbol is used for the definition of set, in some cases a vector also stands for a set, and $|\boldsymbol{a}|$ defines the size of the set $\boldsymbol{a}$.

This section is organized as follows. Section 3.2.1 includes the system model and problem description of beam-frequency scheduling in mmW system. Section 3.2.2 presents the optimal beam-frequency scheduling algorithm. An efficient beam-frequency scheduling procedure is detailed in Section 3.2.3. Numerical results are presented in Section 3.2.4, and summary remarks are given in Section 3.2.5.

3.2.1 Beam frequency scheduling in mmW system

3.2.1.1 System model

Let N define the number of UEs being served in the mmW small cell. The target data rate for the nth UE is defined by $r_{\mathrm{T},n}$, and the associated scheduling priority factor is denoted by w_n, which determines the priority of the UE to be scheduled in the current TTI, and ensures the radio resources to be shared in a fair manner among multiple UEs served by the system. A practical mmW system equipped with three analogue beamformers is illustrated in **Figure 3-6**. Let N_{B} define the number of beams which can be illuminated by a mmW AP, and N_{PB} the number of parallel beams which can be simultaneously transmitted by the mmW AP. The set of total supported beam indices is defined as $\boldsymbol{\mathcal{B}} = (0,1,\dots,N_{\mathrm{B}}-1)$, and the beam space covered by the ith analogue beamformer is denoted as $\boldsymbol{\mathcal{B}}_i, 0 \le i \le N_{\mathrm{PB}}-1$. As shown in **Figure 3-6**, the mmW AP is equipped with three analogue beamformers, each of which is capable of transmitting a pencil beam with the beamwidth of $10°$ within a dedicated sector of $120°$. It follows from the architecture that $N_{\mathrm{B}} = 36$, $N_{\mathrm{PB}} = 3$, and $N_{B_0} = N_{B_1} = N_{B_2} = 12$. In other words, each physical analogue beamformer can only transmit in 12 beam directions within its sector, and only one beam direction is illuminated at each TTI. As a result, the beam space sets in this setup are $\boldsymbol{\mathcal{B}} = (0,1,\dots,35)$, and $\boldsymbol{\mathcal{B}}_i = (12(i-1),\dots,12i-1), i = 0,1,2$.

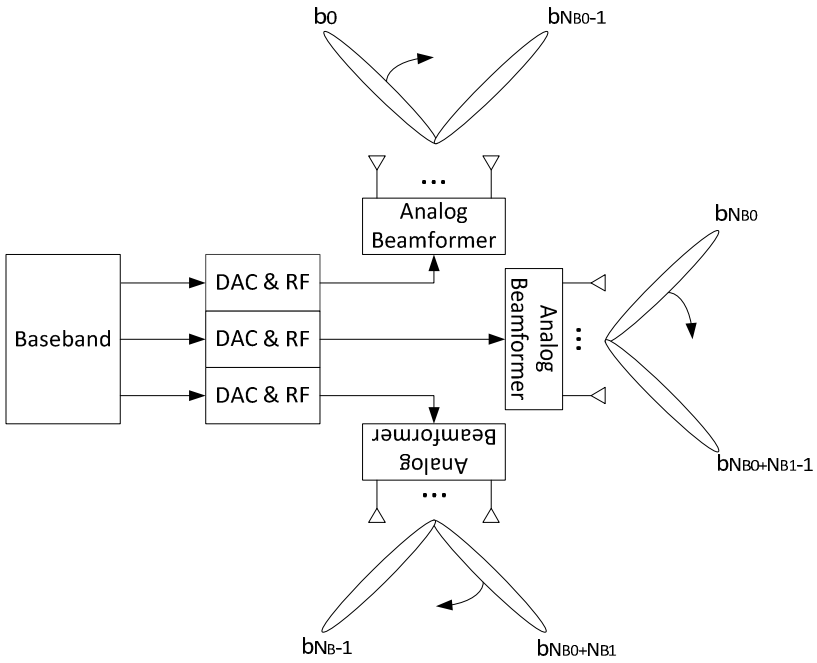


Figure 3-6: mmW system architecture with analogue beamformers.

3.2.1.2 Channel state information at the transmitter

The channel state information (CSI) feedback from the n^{th} UE includes the set of preferred beam indexes $\boldsymbol{p}_n = \left(p_{n,0},\dots,p_{n,M-1}\right)^{\mathrm{T}}$, where M defines the maximum number of preferred beam indexes reported by a UE, and $p_{n,b} \in \mathcal{B}$ stands for the preferred beam index. For each reported beam index $p_{n,b}$, the UE further reports the channel quality index (CQI) vector $\boldsymbol{q}_{n,b} = \left(q_{n,b,0},\dots,q_{n,b,Q-1}\right)$ for all the subbands, where Q refers to the number of subbands in the system bandwidth, and $q_{n,b,f}, 0 \le b \le M-1; 0 \le f \le Q-1$ defines the channel quality of the nth UE with respect to the bth reported beam index at the fth subband of the system bandwidth. To ease the elaboration of the developed

algorithms in the subsequent sections, we introduce the definitions $q_{n,-1,f} = 0, f = 0,\dots,Q-1, n = 0,\dots,N-1$ for the CQI value of all the unreported beam directions from the UEs.

For the sake of simplicity, the reported CQIs from the nth UE can be further formulated as the matrix $\boldsymbol{Q}_n = (\boldsymbol{q}_{n,0};\ \dots;\boldsymbol{q}_{n,M-1}) \in R^{M\times Q}$, where each row vector is comprised of the subband CQIs corresponding to the reported beam index defined in $\boldsymbol{p}_n$.

3.2.1.3 Beam-frequency scheduling problem

Typically, a radio resource scheduler at the MAC layer determines the radio resource allocation to each scheduled UE in the current TTI. In the LTE system, the radio resource is defined in a time-frequency resource grid. In addition to the frequency resource allocation, the mmW AP also needs to appropriately select the transmitted beams for the scheduled UE. Hence, the radio resources managed in the mmW system can be characterized as a resource grid of three dimensions including time, frequency and beam domains.

Specifically, the beam-frequency scheduler in the MAC determines an appropriate beam-frequency rate allocation matrix $\boldsymbol{R}_n$ for the n^{th} UE which can be expressed as

$$\boldsymbol{R}_n = \begin{pmatrix} r_{n,0,0} & \cdots & r_{n,0,Q-1} \\ \vdots & \ddots & \vdots \\ r_{n,M-1,0} & \cdots & r_{n,M-1,Q-1} \end{pmatrix} \in R^{M\times Q}, \quad (1)$$

where $r_{n,b,f}$ stands for the data rate allocated for the nth UE transmitted with the beam index $p_{n,b}$ on the fth frequency subband, and it is bounded in the range

$$0 \le r_{n,b,f} \le q_{n,b,f}, \quad (2)$$

$r_{n,b,f} = 0$ implies that no data for the UE is allocated at the associated beam direction and frequency subband for the scheduled TTI. It is envisioned that $\boldsymbol{R}_n$ is a sparse matrix. Given the rate allocation result $\boldsymbol{R}_n$, the allocated sum rate of each UE can be obtained by

$$r_n = \sum_{b=0}^{M-1}\sum_{f=0}^{Q-1} r_{n,b,f},$$

which is subject to the following constraint

$$r_n \le r_{\mathrm{T},n}. \quad (3)$$

In addition to the maximum rate constraints in (1) and (2), the beam-frequency scheduler must also respect the following constraints of the beam and frequency allocation. Let $\boldsymbol{b}_n \subseteq \boldsymbol{p}_n$ define the set of selected beam directions for the n^{th} UE in a TTI, then $\boldsymbol{b}_n$ can be obtained as

$$\boldsymbol{b}_n = \{p_{n,b} | b: \textstyle\sum_f r_{n,b,f} > 0\}.$$

Firstly, in one TTI at most N_{PB} parallel beams can be transmitted from the mmW system due to the limited number of RF chains. This leads to the following constraint on the feasible beam allocation

$$\left|\cup_{n=1}^{N} \boldsymbol{b}_n\right| \le N_{\mathrm{PB}}. \quad (4)$$

Secondly, only one beam direction can be transmitted by any analogue beamformer in a TTI to achieve the maximum beamforming gain. This implies that

$$\left|(\cup_{n=1}^{N} \boldsymbol{b}_n) \cap \boldsymbol{\mathcal{B}}_i\right| \leq 1,\ i = 1,2,\ldots,N_{\mathrm{PB}} \tag{5}$$

Thirdly, a certain subband transmitted from a particular beam direction can only be allocated to a single UE. For any two UEs, namely the lth and mth UE, the common allocated beam set is denoted as $\boldsymbol{c}_{l,m} = \boldsymbol{b}_l \cap \boldsymbol{b}_m$ which can be further defined as

$$\boldsymbol{c}_{l,m} = \left\{c_{l,m}[k] \middle| \exists k_l, k_m, \text{s.t.}, b_l[k_l] = b_m[k_m] = c_{l,m}[k]\right\}.$$

The above constraint that any frequency subband in the shared beam direction can be allocated to only one UE, leads to the following expression

$$r_{l,p_l^{-1}(c_{l,m}[k]),i} \cdot r_{m,p_m^{-1}(c_{l,m}[k]),i} = 0, \tag{6}$$

where $k = 1,2,\ldots,\left|\boldsymbol{c}_{l,m}\right|, i = 1,2,\ldots,Q$, and function $p_n^{-1}(x)$ is defined as

$$p_n^{-1}(x) = \begin{cases} b, & if\ \exists b, \text{s.t.}, x = p_{n,b}; \\ -1, & \text{otherwise.} \end{cases}$$

The optimal multi-user beam-frequency scheduler finds the solution to the following optimization problem subject to the constraints in (2)-(6):

$$(\boldsymbol{R}_1^*, \ldots, \boldsymbol{R}_N^*) = \arg \max_{\boldsymbol{R}_1,\ldots,\boldsymbol{R}_N} \sum_{n=1}^{N} w_n r_n, \tag{7}$$

More constraints can be added to the scheduling problem due to further restrictions with respect to the UE multi-stream reception capability and resulted control signalling overhead to assist the UE data packet reception. However, in this section, we focus on the above constraints only. The developed method can be straightforwardly extended to more constrained scenarios.

3.2.2 Optimal multiuser beam-frequency rate scheduling algorithm

The beam-frequency rate allocation problem in (7) subject to the constraints (2)-(6) is a constrained convex optimization problem. Apparently it is challenging to obtain a close-form solution to this problem. However, an exhaustive search method over the feasible solution space can provide an optimal solution. This section presents the procedure and algorithms to find the optimal solution.

Let $\mathcal{P} = \cup_{n=1}^{N} \boldsymbol{p}_n$ define the set of all reported beam directions from all the UEs served by the mmW system. We further partition $\mathcal{P}$ into N_{PB} subsets, namely, $\mathcal{P}_i, i = 0,1,\ldots,N_{\mathrm{PB}} - 1$, each of which contains all the reported beam directions steered by the respective analogue beamformer. Without loss of generality, we assume that $|\mathcal{P}_0| \leq |\mathcal{P}_1| \leq \ldots \leq \left|\mathcal{P}_{N_{\mathrm{PB}}-1}\right|$. It is clear from the constraint (5) that the feasible solution of a selected beam set can only contain one beam per analogue beamformer. As a result, the total number of feasible beam sets is equal to $\Psi = \prod_{i=0}^{N_{\mathrm{PB}}-1} |\mathcal{P}_i|$, and each feasible beam set can be defined as

$$\bar{\boldsymbol{p}}^k = \left(\bar{p}_0^k, \bar{p}_1^k, \ldots, \bar{p}_{N_{\mathrm{PB}}-1}^k\right), \tag{8}$$

where $\bar{p}_i^k \epsilon \mathcal{P}_i, i = 0,1,\ldots N_{\mathrm{PB}} - 1; k = 0,\ldots,\Psi - 1$ denotes the beam direction of the i th analogue beamformer in the kth feasible beam set.

Algorithm 1 Beam-frequency resource allocation for a feasible beam set

0. **Initializations**:
1. $\boldsymbol{\psi} = \mathbf{0}_{N_{\mathrm{PB}} \times Q}$;
   \\ Set all the beam-frequency resources to be unscheduled.
2. $\tilde{r}_{\mathrm{T},n} = r_{\mathrm{T},n}, n = 0..N-1$;
   \\ Initialize the target scheduling rate of each UE.
3. $\Phi = 0$;
   \\ Initialize the resulted scheduling metric.
4. **for** $i = 0 : N_{\mathrm{PB}} Q - 1$ **do**
5. **for** $b = 0{:}\, N_{\mathrm{PB}} - 1;\ f = 0{:}\, Q-1$ **do**
6. $r(b,f) = 0$;
7. $u(b,f) = -1$;
8. **if** $\psi(b,f) = 0$ **do**
9. $u(b,f) = \underset{n=0:N-1}{\operatorname{argmax}} w_n \cdot \min\{\tilde{r}_{\mathrm{T},n}, q_{n,p_n^{-1}(\bar{p}_b),f}\}$;
   \\ Select the UE with the maximum scheduling metric contribution for the subband.
10. $r(b,f) = \min\left\{\tilde{r}_{\mathrm{T},u(b,f)}, q_{u(b,f),p_{u(b,f)}^{-1}(\bar{p}_b),f}\right\}$;
    \\ Store the scheduling metric for the resource element.
11. **end**
12. **end**
13. $(b^*, f^*) = \underset{b=0:N_{\mathrm{PB}}-1, f=0:Q-1}{\operatorname{argmax}} w_{u(b,f)} r(b,f)$;
    \\ Select the resource element with the maximum scheduling metric**.**
14. $\bar{U}(b^*, f^*) = u(b^*, f^*)$;
    \\ Allocate the resource element to the selected UE.
15. $\bar{R}(b^*, f^*) = r(b^*, f^*)$;;
    \\ Determine the data rate for the selected resource element.
16. $\Delta = \Delta + w_{u(b^*,f^*)} \bar{R}(b^*, f^*)$;
    \\ Increase the scheduling metric of the beam by the contribution of the allocated subband.
17. $\psi(b^*, f^*) = 1$;
    \\ Mark the scheduled resource element to be 1.
18. $\tilde{r}_{\mathrm{T},u(b^*,f^*)} = \tilde{r}_{\mathrm{T},u(b^*,f^*)} - r(b^*, f^*)$;
    \\ Decrease the target data rate for the selected UE by the allocated data rate for this UE.
19. **end**

For each feasible beam set $\bar{\boldsymbol{p}}^k$, the multiuser beam-frequency scheduler performs a local optimal rate allocation $\bar{\boldsymbol{R}}_k \in R^{N_{\mathrm{PB}} \times Q}$ and $\bar{\boldsymbol{U}}_k \in {I_N}^{N_{\mathrm{PB}} \times Q}, I_N = \{0,1,\ldots,N-1\}$ which define the rate and UE allocation to beam-frequency resource grid, respectively, so as to maximize the target scheduling metric**,** and calculates the associated local optimal scheduling metric $\Delta_k(\bar{\boldsymbol{R}}_k, \bar{\boldsymbol{U}}_k)$**.** With the above results for each feasible beam set, the global optimal beam-frequency rate allocation solution can be readily obtained as follows

$$\{\bar{\boldsymbol{R}}^*, \bar{\boldsymbol{U}}^*\} = \arg \max_{\bar{\boldsymbol{R}}_k, \bar{\boldsymbol{U}}_k} \Delta_k(\bar{\boldsymbol{R}}_k, \bar{\boldsymbol{U}}_k). \tag{9}$$

The beam-frequency resources with respect to the beam set $\bar{\boldsymbol{p}}^k$ can be organized as a beam-frequency resource matrix $\boldsymbol{E}^k$ of N_{PB}-by-Q dimensions, each element $e_{b,f}^k, b = 0,..N_{\mathrm{PB}} - 1, f =$

$0.., Q-1$ of which can be allocated to only one UE with the rate subject to the constraints of (2) and (3). The local optimal rate and UE beam-frequency allocation $\overline{\boldsymbol{R}}_k^*, \overline{\boldsymbol{U}}_k^*$ can be found by Algorithm 1 described above. For the sake of conciseness, we drop k from Algorithm 1.

The kernel computations of Algorithm 1 for the UE beam-frequency rate allocation in the most inner loop, namely calculations of line 9 and 10, need to be calculated $N_{\text{PB}}Q!$ times. As a result, the overall complexity of the exhaustive search based optimal resource allocation is in the magnitude of $(N_{\text{PB}}Q!)^{\Psi}$.

3.2.3 Efficient multiuser beam-frequency rate scheduling algorithm

It is clear that the optimal beam-frequency allocation algorithm described in Section 3.2.2 has a very high complexity. This section presents an efficient beam-frequency rate allocation procedure to achieve a similar performance as the optimal algorithm while considerably reducing the complexity. The basic principle of the efficient algorithm is to sequentially find the optimal beam direction belonging to the unscheduled beamformers to achieve maximum scheduling metric. After an optimal beam direction is scheduled, the corresponding analogue beamformer is set to be scheduled, and all the reported beam directions thereof are removed from the subsequent beam search space. Such beam search procedure is performed until either the target scheduling metric is achieved or all the beamformers are scheduled. The concrete steps are elaborated as follows.

Step 0 (Initialization): The beam-frequency scheduler sets all the beamformers to be unscheduled. Let $\tilde{\mathcal{P}} = \bigcup_{n=1}^{N} \boldsymbol{p}_n$ further define the set of all the reported beam directions steered by unscheduled beamformers. At the initial step, $\tilde{\mathcal{P}}$ includes all the reported beams. We further partition $\tilde{\mathcal{P}}$ into N_{PB} non-overlapped subsets, namely $\tilde{\mathcal{P}}_i, i = 1,2, \dots, N_{PB}$, each of which contains all the reported beam directions steered by a particular beamformer.

Step 1 (Start of iteration): For each beam direction $\tilde{p} \in \tilde{\mathcal{P}}$, the scheduler determines the rate allocation of each frequency subband to a certain UE so as to maximize the contribution to the target scheduling metric. The procedure described in Algorithm 1, by setting N_{PB} to 1, can be employed to find the maximum achieved target scheduling metric. The outcomes of the calculation include the optimal UE-beam-frequency allocation $\overline{U}(\tilde{p})$, the associated rate-beam-frequency allocation $\bar{R}(\tilde{p})$, the achieved scheduling metric $\Delta(\tilde{p})$, and the residual target data rate $\tilde{r}_{\text{T},n}(\tilde{p})$ of each UE.

Step 2: Based on the above outputs of each unscheduled beam direction, the beam direction $\tilde{p}^*$with maximum scheduling metric $\Delta(\tilde{p}^*)$ is selected and the frequency resources thereof are scheduled according to the UE-beam-frequency allocation $\overline{U}(\tilde{p}^*)$ and rate-beam-frequency allocation $\bar{R}(\tilde{p}^*)$.

Step 3: The beamformer steering the selected beam direction $\tilde{p}^*$ is marked as scheduled beamformer, and all the reported beam directions, which are exclusively steered by this scheduled beamformer, are removed from the set $\tilde{\mathcal{P}}$ of unscheduled beam directions.

Step 4: The target scheduling metric is updated by removing the contribution of achieved scheduling metric $\Delta(\tilde{p}^*)$.

Step 5 (End of iteration): If the updated target scheduling metric is 0, that is, all the UEs have been allocated with target data rate, or all the beamformers are scheduled, the beam-frequency scheduler stops the operation. Otherwise, the scheduler repeats steps 1 to 4 until the one of the two stop conditions are met.

It should be noted that the above step 3 only removes those beams exclusively steered by the scheduled beamformer. If all analogue beamformers steer the same beam space, then only the selected beam in step 2 is removed from the set of unscheduled beam directions. It is obvious that the most computations take place in step 1, and the kernel computations in Algorithm 1 need to be performed by $!Q$ times for each reported unscheduled beam. As a consequence, the overall complexity of the efficient scheduling algorithm is upper bounded by $(Q!)^{\sum_{i=0}^{N_{\mathrm{PB}}-1}(i+1)|\boldsymbol{P}_i|}$, which is significantly less than that of the optimal solution in Section 3.2.2.

3.2.4 Numerical examples

In this section, some numerical examples are given to demonstrate the performances of the proposed scheduling algorithms. Let us focus on the mmW AP with the architecture illustrated in **Figure 3-6**, i.e., 3 analogue beamformers, each of which exclusively steers 12 beams of 10° beamwidth, with beams numbered from 0 to 35. It is assumed that 3 UEs are in the coverage of the mmW AP. It is further assumed that the system bandwidth is divided into 4 subbands, and each UE feedbacks at most the 3 strongest beams with respect to the subband CQI values. Let $\mathrm{r_D}$ define the basic data rate in bits/s, which is used as the unit for the CQI and UE target data rate definitions. The CQI value ranges from 1 to 8 for each subband. The UEs target data rate and scheduling priority are listed in **Table 3-2** and the beam-subband CQI feedbacks are described in

Table 3-3 and illustrated in **Figure 3-7**.

Table 3-2: UE target data rate and scheduling priority

UE-Scheduling targets	**UE 0**	**UE 1**	**UE 2**
Target data rate $r_{\mathrm{T},n}$ $[\mathrm{r_D}]$	16	24	32
Scheduling priority [1-8]	7	5	3

Table 3-3: Subband CQI reports from 3 UEs

Beam-subband CQI feedback $[\mathrm{r_D}]$		**Subband 0**	**Subband 1**	**Subband 2**	**Subband 3**
UE0	b10	1	2	3	2
	b11	2	3	4	3
	b12	1	2	3	2
UE1	b11	3	4	3	2
	b12	4	5	4	3
	b13	3	4	3	2
UE2	b11	6	8	4	2
	b12	5	6	4	3

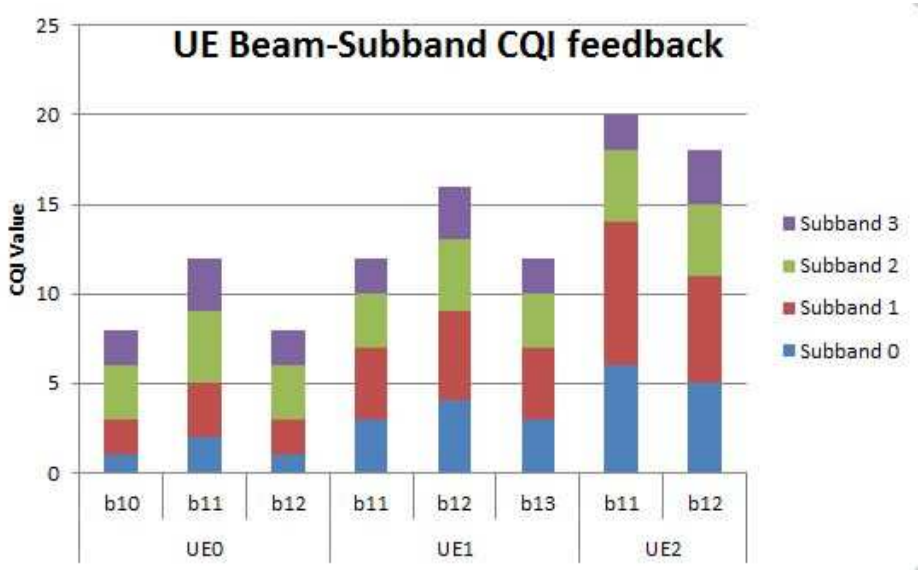


Figure 3-7: UE beam-subband CQI report.

The optimal algorithm in Section 3.2.2 and the efficient scheduling algorithm in Section 3.2.3 obtain the same UE-beam-frequency allocation results. The obtained UE-beam-frequency allocation results are illustrated in **Figure 3-8**. Moreover, different UEs can be transmitted from the same beam by being multiplexed in frequency domain. It is further observed that one UE, namely UE0, can also be scheduled simultaneously with multiple beam transmission. Such multiple beam allocation assumes that the UE has the capability of performing multibeam reception, e.g., by virtue of multiple RF receive chains.

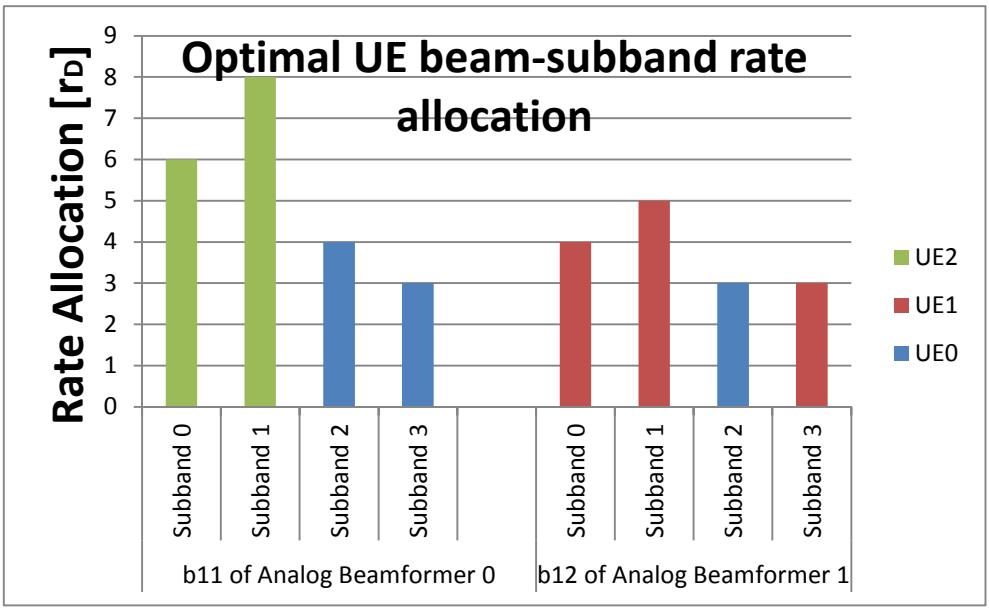


Figure 3-8: UE-beam-frequency multiplexing results.

It is clear from the results obtained by our simulations that even though the efficient scheduling method manages to achieve the same scheduling metric as the optimal scheduling algorithm, it does so, as illustrated in **Figure 3-9**, with a significantly smaller computation complexity in terms of the number of kernel calculations, i.e., a figure coming out of the formula $w_n \cdot \min\{\tilde{r}_{\mathrm{T},n}, q_{n,p_n^{-1}(\bar{p}_b),f}\}$.

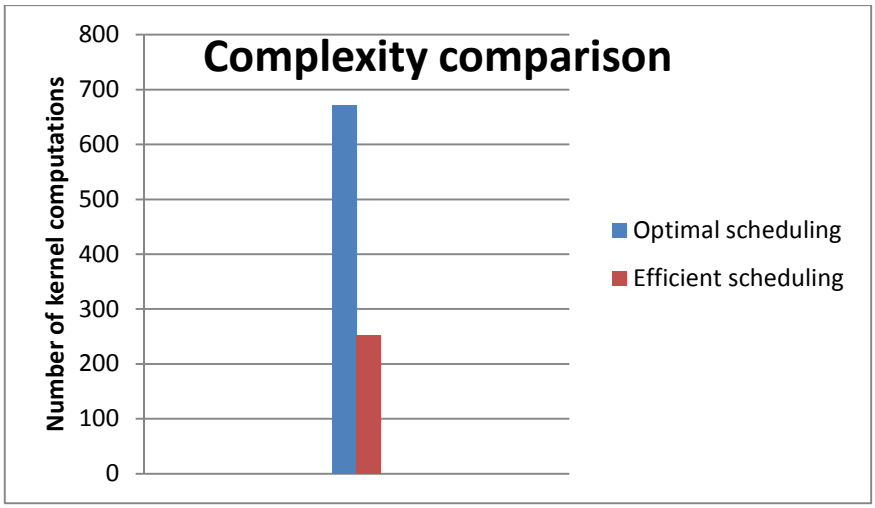


Figure 3-9: Complexity comparison of optimal algorithm and efficient algorithm.

3.2.5 Summary

An optimal scheduling algorithm and an efficient scheduling procedure have been developed for the mmW downlink system. It is demonstrated that the computation efficient scheduling procedure with considerable reduced complexity can achieve the same target scheduling metric as the optimal algorithm. It has been shown from simulation results that the optimal scheduling solution may multiplex different UEs in both frequency and beam domain, and some UEs may be scheduled with multiple beams simultaneously.

4. Functions and algorithms for the backhaul

mmW communication is envisioned to be a promising technology to cater for the continuous growth of backhaul capacity demand driven by the ever-increasing end-user traffic. Due to the challenging propagation environment for mmW signals, the communication distance between two mmW entities is typically quite limited, e.g., ranging from several tens of metres to one or two hundred metres depending on the availability of a LOS transmission over the radio channel. As such, the backhaul is typically formed by multi-hops, i.e., the forwarding paths are formed of concatenated BH nodes or AP relays, to reach the final destination, usually a small cell AP serving the UEs. Such meshed multi-hop backhauls, in the form of a **dedicated standalone backhaul** or as an **inband backhaul**, or consisting of interworking partitions of both types of backhaul, seem to be required for all the use cases defined in Annex C.

In AP installation, the mmW backhaul provisioning becomes challenging due to the need for antenna beam alignment between ends of the backhaul links. Therefore, it is desirable to implement self-organized backhaul link establishment so that the beam alignment can be achieved automatically with minimal human effort. This requires efficient neighbour search algorithms that utilize the beam steering capabilities in the nodes.

Due to the time-varying nature of wireless channels and user traffic demand, it is also very beneficial to have dynamic backhaul resource allocation and packet routing to optimize the network resilience, network efficiency and power consumption.

In the light of above backhaul design objectives, this section defines system functions and algorithms for two complementary types of mmW backhaul, which may be deployed independently or combined as partitions in the same backhaul network, as described in Section 2.1.

Section 4.1 describes a practical method for the initial link establishment to support dynamic routing in a self-backhauling inband backhaul composed of AP relay nodes. Here, the neighbouring AP nodes discover each other based on assistance from a macro BS, by utilizing dual connectivity between the layers. That is, a major part of the control plane functions in the support for the mmW backhaul network formation relies on LTE-A wide area signalling between a macro BS and its subordinate APs that form the relay network.

Section 4.2 presents functions to support self-organization, routing, and link scheduling in a dedicated standalone backhaul. The BH nodes discover and establish mmW links to their neighbours autonomously and report of the new links to a centralized controller, which computes and configures a routing overlay and link schedules into the network. In this case, as opposed to the above-mentioned inband backhaul, the backhaul is not reliant on the availability of wireless wide area control-plane connectivity. Still, it is required that the mmW backhaul mesh is reachable from the controller, via a gateway node.

Section 4.3 elaborates on beam steering strategies for the backhaul links, by proposing algorithms for hybrid beamforming and analogue beam switching.

4.1 Inband backhaul with small-cell relay nodes

4.1.1 Background and motivation

This section focuses on mmW technology deployment for the backhaul link by expanding on the existing AP relay concept defined for LTE-A. In the proposed inband self-backhauling solution, mmW

radio resources are shared between backhaul and access links. The AP relay mesh formation is supported by a macro eNB by using wide-area connectivity for the control plane. The backhaul deployment scenario is illustrated in **Figure 4-1**.

The targeted frequency ranges have difficult propagation conditions, mainly caused by the high propagation losses in these bands, as well as various atmospheric effects. On the other hand, the small wavelength allows the creation of small-size, highly directional, high-gain antennas. These high antenna gains can be used to compensate the additional path loss, and, in our research task, we take high gain steerable directional antenna patterns as a working assumption for both ends of the link, i.e. the donor eNB (DeNB) and the small cell AP.

The problem in using high gain antennas is the fact that these antennas have to exactly point to each other, i.e., TX and RX beam alignment. High gain antennas are typically composed of an array of radiating elements fed through a beam steering device, e.g., a phase shifting device, and, by adjustment of these phase shifts, the beam can be electronically steered over a certain range, w/o mechanical work. A deeper look on electronic beam steering is given in Section 4.3.

However, it is important minimise manual adjustment during the installation by human interaction, i.e. the service personnel. Further, we have to consider that the mmW radio links will vary depending on short-term variations, e.g. rainfall and wind induced beam misalignment or long-term variations, e.g. seasonal foliage changes. Therefore, the initial antenna adjustment may be degraded with time and the backhaul quality loss may result in:

- increased delay by increased re-transmission and packet losses, impacting delay sensitive services,
- loss in the available channel capacity,
- total loss of mmW connectivity.

Therefore, we need means for self-organized antenna adjustment for directional point-to-point mmW backhaul links (e.g., between a DeNB and a small cell AP), both for the initial set-up and for optimizations during operation.

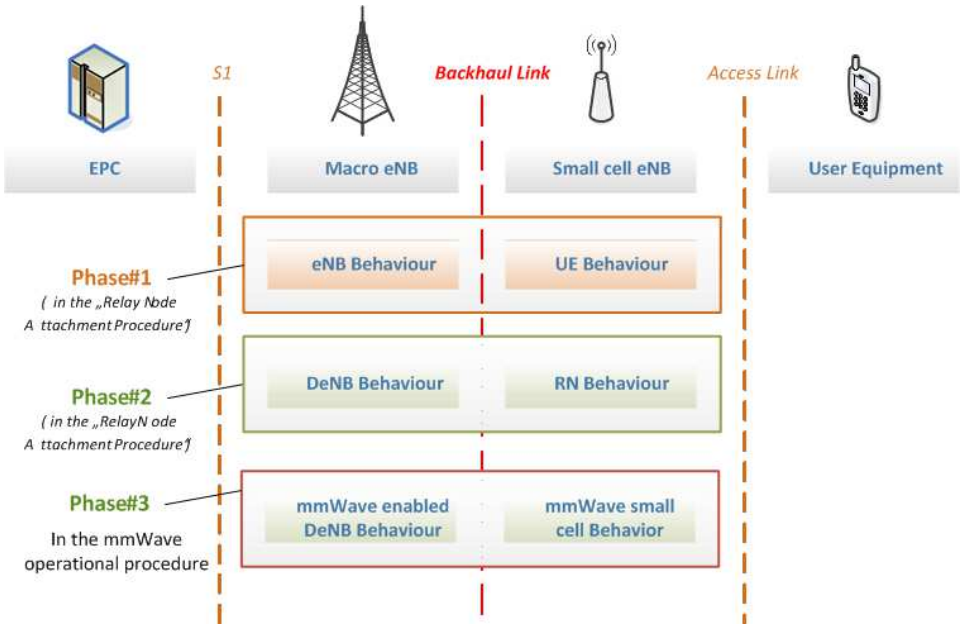


Figure 4-1: Deployment scenario for mmW backhaul links.

4.1.2 Description of proposed innovations

In most of the operational cases, the mmW connected small cell is additionally covered by a mobile broad band wide area technology in a heterogeneous network (HetNet). If the overlaying technology is LTE-A, the standard already provides a procedure for inband relays to connect to the DeNB. This is described in 3GPP Rel-10 specifications.

The HetNet operational case allows combining existing technology with the mmW backhaul, and procedures and messages can get introduced and enhanced in future 3GPP releases to enable a small cell AP or relay node initial identification. This procedure can enable

- initial and operational geographical location of the small cell,
- initial and operational antenna adjustment for the backhaul link (i.e. the red interface in **Figure 4-1**).

The known 3GPP Un interface and the Multicast-Broadcast Single-Frequency Network (MBSFN) frame format allow establishing transport channels for the inband relay backhaul and can be enhanced to establish a directional mmW connection.

In principle, we can also consider search algorithms opportunistically steering beams and using low rate signalling with highly robust coding to establish an initial connection and antenna adjustment for the scenario of mmW band standalone operation. This is a way to go if no mobile broadband system co-exists, which is the assumption in the dedicated standalone backhaul case described in Section 4.2. In cases of co-existence, we can utilize the wide area system to aid mmW link establishment. The methods studied in this section focus on the heterogeneous network scenario where mmW small cells are deployed in the coverage of a macro cell base station.

4.1.2.1 Reuse of LTE relay initial attachment procedure

As shown in phase 1 in **Figure 4-1**, the intention is to reuse the known relay node attachment procedure to identify the small cells presence under cellular coverage. However, additional messages and procedures shall be introduced to establish and maintain the mmW backhaul link.

Figure 4-2 shows the set-up principle and terminology for an inband relay scenario, in which the backhaul is using the same spectrum as the access (AP-UE interface). Therefore, a solution to deal with the interference between backhaul and access is needed. This is solved by the time domain multiplexing with an MBSFN subframe as shown in **Figure 4-3**, which allows a temporary blanking of the relay node transmission. The MBSFN subframes can be used for the backhaul link transmission between the donor eNodeB (DeNB) and the relay node (RN), because the UE does not expect to have UE intended transmissions when receiving this type of subframe. In the MiWaveS use cases, the small cell AP can be seen identical as the relay in 3GPP release 10.

3GPP LTE-A [9] itself provides a 2-phase initial establishment procedure. In the first phase, the relay node, in our case this is a mmW small cell AP, attaches to the infrastructure network by using a Random-Access Channel (RACH) procedure and initially behaving like a UE (cf. Phase #1 in **Figure 4-1**). In the second phase (cf. Phase #2 in **Figure 4-1**), the infrastructure network is informed that the attached device is a relay node. In our case, it is in fact a mmW small cell AP, which is revealed in additional Phase #3. The basic relay node attachment procedure illustrated in **Figure 4-4** is known and described in [9].

The intended concept uses the relay node attachment procedure for mmW small cell AP. This requires having an LTE-A air interface support (at least for an initial backhaul link operation), to establish a connection like a relay node to the DeNB via a Un interface. The capability to execute a system attach by an LTE-A procedure does not necessarily imply that the attaching entity, regardless whether this is a small cell AP or a legacy Rel-10 relay node, provides an LTE access link to UEs via a Uu interface.

The use of mobile broadband technology allows an initial access without having the small cell's mmW antenna directions a priori adjusted. It further provides a fallback solution, in case the mmW backhaul gets lost or impaired, and a re-establishment is needed. When the mmW link is established, the LTE backhaul can be reduced to a small bandwidth to enable a closed loop antenna optimization signalling.

Once a Radio Resource Control (RRC) connection has been established, the small cell AP has to inform the macro cell about its nature being a small cell which is capable of supporting mmW communication, so as to furnish a higher data rate for the BH link when demanded. This will need new information elements or RRC messages in the 3GPP LTE-A standards, and respective modification can be introduced in releases 13 or 14 as a new relay node category. Specifically, the small cell can inform the macro cell base station about its capabilities, e.g. if it is a legacy relay node with LTE-A access capabilities, or a small cell AP with mmW access only, or a small cell AP with combined mmW and LTE-A access functions. The primary task after completing phase 3, is the RRC establishment to provide a control channel connection and feedback loop, which supports the adjustment of the backhaul mmW antennas.

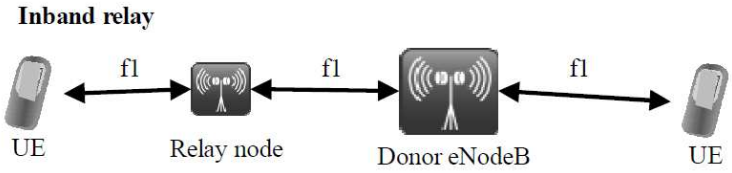


Figure 4-2: Inband-relay scenario.

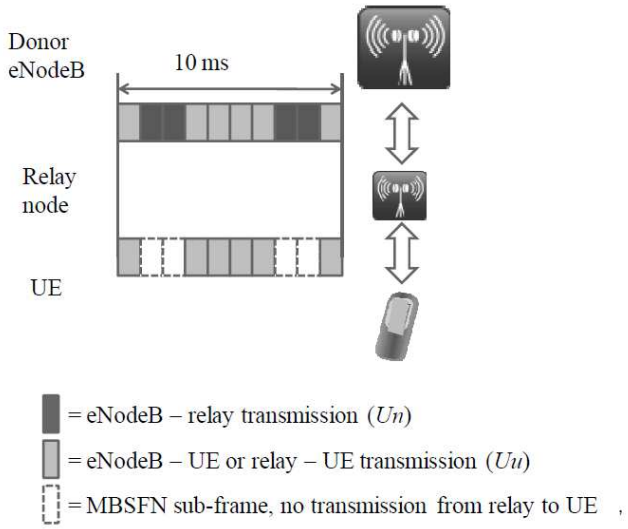


Figure 4-3: Time multiplex on backhaul link and access link to avoid interference.

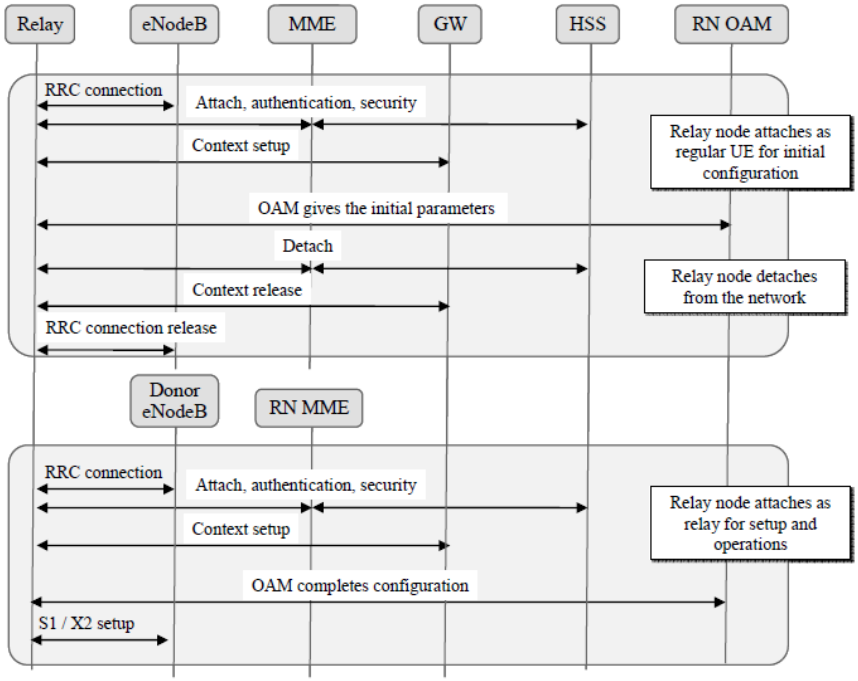


Figure 4-4: Relay Node (RN) attachment procedure according to 3GPP TS 36.300 Rel-10.

4.1.2.2 mmW backhauling discovery and establishment

It is widely known that the mmW communications suffer from severe path losses due to their very high signal frequency. Especially the path loss in the non-line-of-sight (NLOS) scenario can be 20-30 dB higher than in LOS transmission. As such, the coverage of the mmW small cell is typically quite limited, e.g., 100m-200m [7], especially in urban scenarios where LOS transmission can be quite often obstructed by the environment. To have sufficient high data rate coverage in a hot-spot area, mmW small cells may be ultra-densely deployed to form clusters of small cells. The backhaul link between the cluster of small cells and the macro eNB serving as the mobility anchor connecting to the core network can be furnished by the mmW communication connection if the cluster of small cells is in the mmW coverage of the donor eNB. The UEs could get access to the donor eNB also via mmW relay nodes that are along the way from the donor eNB to the destination small cell APs.

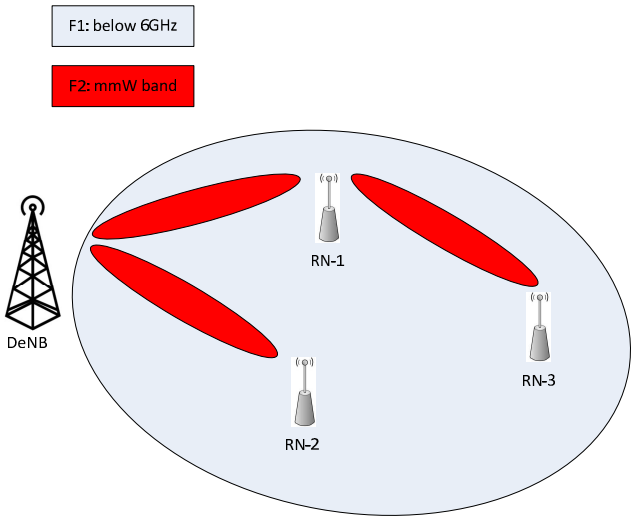


Figure 4-5: mmW small-cell APs under the coverage of donor eNB.

As shown in **Figure 4-5**, three mmW small cell relay nodes, i.e., RN-1, 2 and 3, are deployed in the cellular spectrum coverage of the macro donor eNB, i.e., DeNB. RN-1 and RN-2 are in the mmW coverage of the DeNB while RN-3 is out of the mmW coverage of the DeNB. However, the RN-3 is in the mmW coverage of RN-1 and RN-2. As a result, the mmW backhauling from the RN-3 to the DeNB can be established via RN-1 and/or RN-2.

As mentioned above, the high directional transmission is critical for the mmW communication to provide a meaningful coverage while respecting the regulated transmission power limit. The beam scanning approach has been adopted in the WIGig standard [8], where a number of beam-formed preamble signals targeted to different directions are transmitted prior to each data packet to allow the beam alignment between transmitter and receiver. In principle, the beam alignment between the transmitter and receiver can be achieved by a closed loop channel state information (CSI) feedback scheme in which the receiver firstly estimates the angle of departure (AoD) and the angle of arrival (AoA) for each dominant multipath component of the channel impulse response, secondly configures the receive beamformer according to the AoA estimates, and lastly reports the AoD estimates to the transmitter to facilitate the transmit beamformer settings. However, such schemes suffer from the challenge of the limited coverage of the omnidirectional mmW transmission, resulting in very poor channel state estimates, and prohibitive complexity of the accurate massive MIMO channel estimation algorithms. Therefore, it is beneficial to study some mechanism with practical complexity, to achieve the beam alignment in the course of the initial backhaul link establishment. In the light of this objective, a self-organized multi-hop backhauling establishment procedure based on cell discovery signal measurement is developed for the mmW BH link establishment for two scenarios:

(1) small cells in the mmW coverage of the DeNB, defined as in the hop-1 mmW coverage of the DeNB; and

(2) small cells out of the mmW coverage of the DeNB, that can however be connected to the DeNB with multi-hop mmW links formed by a chain of mmW small cell relay nodes, defined as in the hop-n mmW coverage of the DeNB, where n denotes the number of relay hops from the DeNB to the destination small cell AP.

As shown in **Figure 4-5**, RN-1 and RN-2 are in hop-1 mmW coverage of the DeNB, while RN-3 is in hop-2 mmW coverage of the DeNB.

It is noted that this section focuses on the signalling procedure to enable transmit beam alignment, and the detailed receive beam alignment algorithm will be covered in future study. The organization of this section is as follows. Section 4.1.2.2.1 presents the discovery signal design, based on which initial mmW small cells are detected. Section 4.1.2.2.2 details the self-organized backhauling initial establishment procedure, to support dynamic path switching and potentially cooperative transmission. Section 4.1.2.2.3 describes the control channel design to support cross carrier scheduling for multi-hop mmW backhaul link.

4.1.2.2.1 mmW discovery signal design

To facilitate the mmW small cell discovery, one needs to design the relevant discovery signal. Moreover, in LTE Release-12, the energy efficient small cell operation has been actively discussed. As a result, the small cell on-off procedure is to be standardized to enable the small cell eNB to switch off the entire downlink transmission signals for a long time period when no UE traffics are transported in the cell. To aid the UE to discover the small cell with on-off procedure, a new concept of discovery

signal has been agreed to be specified in the Rel-12 standard, and it can be configured to each Rel-12 UE by an UE specific RRC signalling. The discovery occasion (DO) defines a signal period comprised of up to 5 transmission time intervals (TTIs) in which the synchronization signals (primary synchronization signal and secondary synchronization signal), cell-specific reference signal (CRS), and possibly channel state information reference signal (CSI-RS) are transmitted. The discovery occasion is transmitted periodically, and the periodicity can be 80/160ms or even longer. The UE is configured by a so-called discovery signal measurement timing configuration (DMTC) parameter to measure the discovery signal receive power (RSRP) or discovery signal receive quality (RSRQ). The measured discovery RSRP and/or RSRQ shall be reported to the macro eNB to enable the dual connectivity operation to achieve high capacity or load balancing by virtue of the small cell layer. In addition, the discovery signal measurement can also be used for the secondary connectivity handover of a particular UE with mobility.

With the state of the art in the LTE development mentioned above, it is plausible to assume that the mmW-RAT (Radio Access Technology) will also support on-off operation to save network energy. This means the mmW base station can turn off the downlink signals when no traffic is transmitted to the served UEs in the cell. The design principle of the LTE discovery signal can be enhanced to support the mmW small cell discovery, and to enable the mmW small cell AP/relay node to establish a mmW backhaul link to the DeNB.

As mentioned above, each mmW small cell should also transmit discovery signals periodically. These discovery signals can be implemented by a signal sequence having a good auto-correlation property, e.g. Zadoff-Chu sequence as in the LTE system, to aid the discovery signal detection. Given the transmit power limit, targeted beamforming gain and afforded implementation complexity, the mmW small cell AP/relay node can transmit one or several parallel directional RF beams at the same time. Let $n_b \in \{1,2,\dots,N_b\}$ define the number of supported parallel RF beam sectors by the mmW small cell AP. The mmW small cell shall transmit n_b discovery signals using the same time-frequency resources with different sequence signatures, which is a function of the sequence ID, with respective beam direction.

For instance, a mmW small cell node is capable of transmitting 3 RF beams at the same time, and the discovery signals transmitted from the mmW small cell node can be illustrated in **Figure 4-6**. As shown in **Figure 4-6** (a), (b) and (c), the mmW small cell node transmits 3 discovery signals at the same time, each discovery signal uses its respective signal sequence. In each of these discovery signals, a so called discovery cluster is transmitted periodically with the same period of n_D frames, e.g., 80ms, 160ms or even longer. Each discovery cluster consists of n_o discovery occasions, and each discovery occasion is comprised of N TTIs, among which one or several TTIs are reserved for Tx only in order to transmit discovery signal. The discovery occasions in one discovery cluster can be transmitted successively with different beam directions, as such the beam scanning can be performed by the same physical RF beamformer. For example, the mmW small cell AP has 3 antenna arrays, each of which is driven by its own RF beamformer and serves a sector of 120 degrees. With the assumption of $n_o = 8$, each sector can be scanned by 8 beam directions with beam direction spacing of 15 degrees in the time period of one discovery cluster. In case of one TTI of 100us, and one discovery occasion of 10 TTIs, one sector can be scanned in 8ms.

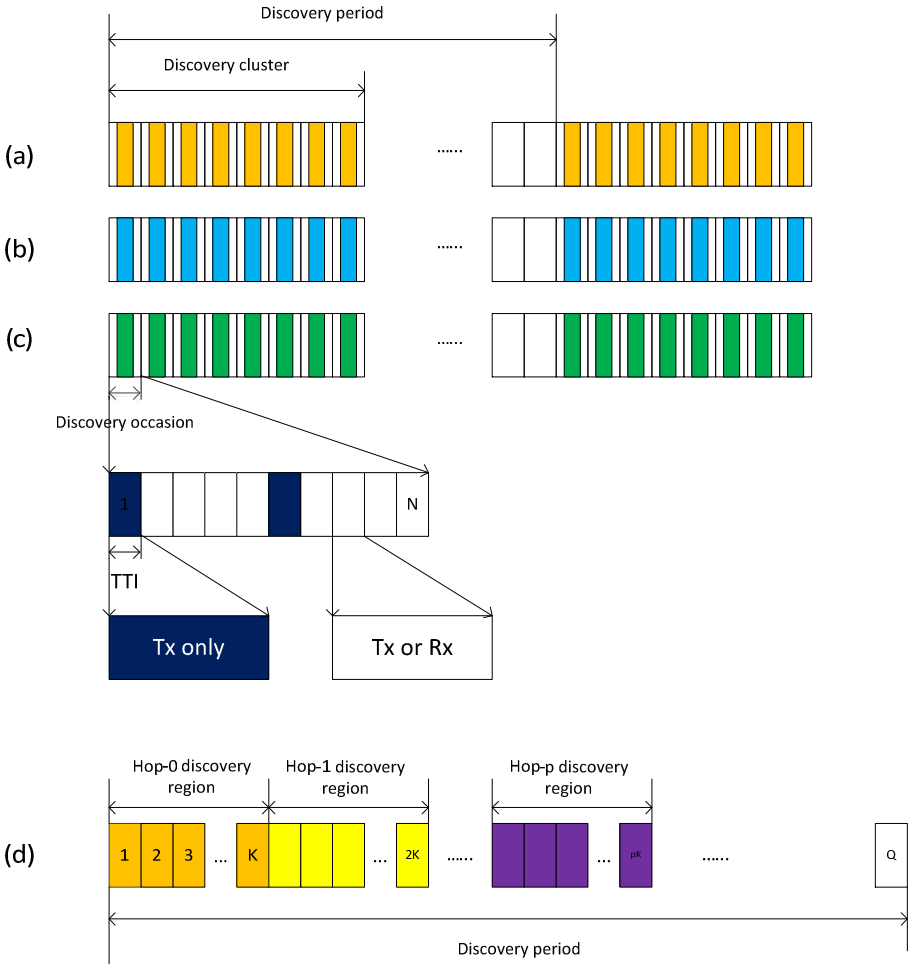


Figure 4-6: Discovery signal frame structure.

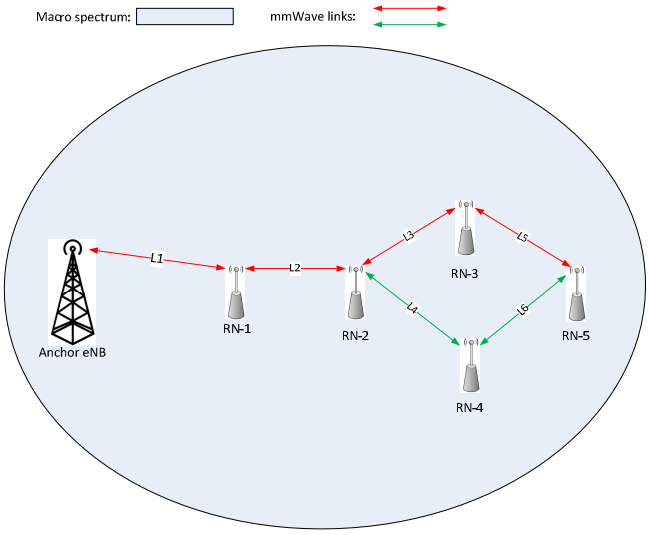


Figure 4-7: Multi-hop mmW backhaul link.

To support the multi-hop backhaul link, each mmW small cell node needs to receive/track the discovery signal from upstream small cell nodes and transmit its own discovery signal for the downstream small cell nodes and/or its served UEs. To avoid the needs of transmitting and receiving simultaneously, the small cell nodes at different hops need to transmit discovery signals at different time, otherwise, full duplex would be required. Therefore, at the beginning

of each discovery period shown in **Figure 4-6**-(d), p discovery regions are defined, each of which consists of K discovery occasions. It is required that $n_o < K$. In case of the network shown in **Figure 4-7**, the mmW-RAT of the anchor eNB shall transmit discovery signal at hop-0 region, RN-1 and RN-2 are transmitting discovery signals at hop-1 and hop-2 regions, respectively. RN-3 and RN-4 in **Figure 4-7** use hop-3 region, and RN-5 the hop-4 region. Another approach is to only design two discovery regions, one is used for those small cell nodes at even-number hop locations, and the other is used for those at odd-number hop locations. It should also be noted that mmW-RAT signal should be synchronous to the primary cell (PCell) signal transmitted by the macro cell eNB in the sense that the frame boundary or discovery period boundary of the mmW small cell should be aligned with the corresponding boundary in the PCell. The synchronicity can be potentially leveraged by the PCell assisted small cell discovery detailed in the next section.

The discovery signal transmission of each small cell node can be configured by the anchor eNB or DeNB according to the backhaul link topology and the transmitter capability of the small cell node, e.g., how many parallel beams can be transmitted, and how many sequential beams for beam scanning shall be supported. The detailed configuration method is presented in the next section.

4.1.2.2.2 Self-organized multi-hop backhauling initial establishment

As shown in **Figure 4-7**, five relay nodes/small cells, namely RN-x, x=1,2..,5, are in the coverage of the macro cell created by the anchor eNB with the mobile broadband spectrum. The term relay node is used, to underline the function of establishing a backhaul connection. Moreover, the anchor eNB and all the relay nodes are equipped with mmW-RAT, and all the possible mmW link connections in the network are illustrated by the arrow lines with the respective label Ly, where y=1,2,.., 6. In particular, the mmW backhaul link between RN-5 and the anchor eNB can be formed by at least two possible multi-hop backhaul link paths, i.e., P1 = (L1, L2, L3, L5) or P2= (L1, L2, L4, L6).

The proposed mmW backhaul link establishment procedure is illustrated in **Figure 4-8**. As shown in **Figure 4-8**, this procedure is built by three phases. The first phase is for the upstream node selection/configuration and relevant downlink transmission beam alignment. Phase 2 is for the discovery signal configuration for the small cell node. Phase 3 is for the uplink transmission beam alignment. This procedure is detailed in the following.

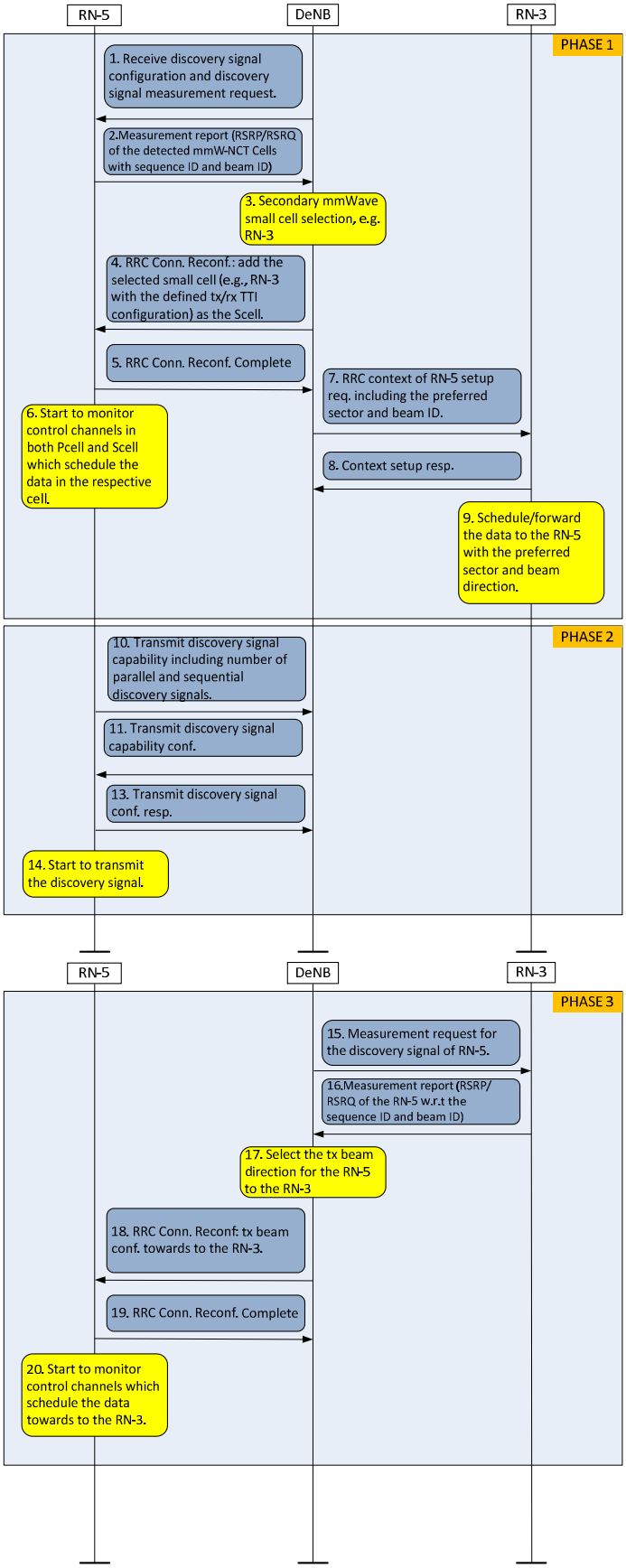


Figure 4-8: Signalling diagram for the mmW backhaul link establishment.

Phase 1: Once an RRC connection is established between a small cell node and the anchor eNB (at the end of the process described in the previous section), the anchor eNB serves as the PCell for the small cell node. With the knowledge of the discovery signal configuration for the existing mmW small cell nodes, the anchor eNB signals the receive discovery signal configurations to the newly camped mmW small cell node. In particular, the starting time position of each discovery cluster, number of discovery occasions in each discovery cluster and the periodicity of the discovery cluster shall be informed to the small cell node. With this information, mmW small cell node only turns on its mmW receiver during those time periods with the possible presence of discovery signals. After some discovery signals are detected by the mmW small cell node, the receive power and/or quality of each detected discovery signal, which is identified by the sequence ID and the discovery occasion index within the discovery cluster, shall be measured. Only those measurements metrics above a certain threshold will be reported to the anchor eNB. Based on the reported discovery signal RSRP or RSRQ, the anchor eNB will configure one or several mmW small cells as the secondary cells (SCells) to the newly camped mmW small cell node. Moreover, the anchor eNB signals the ID of the new small cell node and its preferred beam directions indicated by the discovery occasion index within the discovery cluster to the upstream small cell node, to create a RRC context related to the new camped small cell node in the upstream small cell node. For example, in **Figure 4-7**, RN-5 is the new small cell node. Assuming the discovery signal RSRP from RN-3 is greater than that from RN-4, the anchor eNB can configure RN-3 as a SCell to RN-5, and informs RN-3 to serve as a SCell for RN-5 as well. From this point, the new camped small cell node is able to receive the traffic from the upstream small cell node.

Phase 2: As the mmW-RAT is envisioned to be employed for both the access link and backhaul link, it is more likely that the time-domain duplex for the mmW small cell node transmission and reception will be adopted. Thus, the TX/RX (downlink/uplink) TTI configurations for adjacent small cell nodes with established communication path would be complementary to each other. For example, a TX TTI of RN-3 should correspond to a RX TTI of RN-5 in **Figure 4-7**. With such complementary TX/RX TTI configuration, the same air interface for the backhaul link and access link can be used, i.e. mmW-Uu interface can be used for both backhaul link and access link, in contrast to the LTE inband relay where Uu and Un interfaces are used for the access link and backhaul link, respectively.

In phase 2, the new camped small cell node firstly signals its discovery signal transmission capability to the anchor eNB. With the received discovery signal transmission capability from the mmW small cell node, the anchor eNB configures the proper discovery signals to the new camped small cell node, and the small cell node starts to transmit its own discovery signal.

Phase 3: The anchor eNB requests the upstream small cell node, e.g., RN-3 in **Figure 4-7**, to detect and measure the discovery signals transmitted from the new camped small cell node, i.e. RN-5. With the reported discovery signal receive power and/or quality transmitted by RN-5, from RN-3, the anchor eNB further informs the preferred beam direction for the RN-5 to transmit data to RN-3. Hence, the beam alignment in the uplink direction has also been accomplished. It should be noted that the discovery signal measurements are typically performed by all the small cell nodes periodically, and measurements results are regularly reported to the anchor eNB. Consequently, the beam tracking is inherently enabled by this method.

Upon the physical layer connection being established by the above signalling method, the backhaul link traffic can be transported by using the S1 and X2 interfaces of the LTE.

mmW backhaul link reconfiguration

It is envisioned that in certain circumstances, the multi-hop backhaul link path can be changed for various reasons. For example, for the sake of load balancing or to improve the power saving, the anchor eNB in **Figure 4-7** may decide to switch the backhaul link to RN-5 from P1 to P2 defined above. This would require changing the SCell configuration for RN-5 from RN-3 to RN-4. In this case, the signalling in phase 1 and phase 3 in **Figure 4-8** needs to be performed by substituting RN-3 with RN-4 to achieve the downlink and uplink beam alignment for the new backhaul link L6.

To achieve dynamic path switching, it is also possible that in phase 1, the anchor eNB shall configure both RN-3 and RN-4 as the SCells for RN-5. In a particular backhaul link TTI, the backhaul link packet to RN-5 can be routed from either RN-3 or RN-4 depending on the routing decision made by the anchor eNB based on certain criteria. By doing so, the dynamic path switching or routing can be implemented without backhaul link reconfiguration and also transparently to the destination small cell node. In addition, when both RN-3 and RN-4 are configured as the upstream small cell nodes for RN-5, it also enables the cooperative transmission/reception operation for the backhaul link.

4.1.2.2.3 Control channel schemes for multi-hop mmW backhaul link

In LTE system, physical downlink control channel (PDCCH) is employed to schedule data packet transmission in a TTI, and every UE in a cell monitors its search space of PDCCHs to detect if any downlink data packet is scheduled or uplink assignment signalling is transmitted in the current TTI. Furthermore, two types of control channel scheduling approaches are specified, namely same carrier scheduling and cross carrier scheduling, for data transmission with carrier aggregation (CA). Especially the cross carrier scheduling approach in which the control channel schedules a data packet transmitted from a different component carrier, is specified with the aim of supporting inter-cell interference coordination for the PDCCH in the heterogeneous network. As such, it can be also beneficial to develop control channel cross-carrier scheduling schemes for the mmW multi-hop backhaul link data transmission. In particular, if the control channel scheduling the data packet in the mmW link is transmitted in the macro cell broadband spectrum, the detection performance of the control channel is clearly better than that transmitted in the mmW spectrum due to better channel condition and all those well-established PDCCH design benefits. This section presents several alternatives to support cross carrier scheduling for the control channel in the mmW multi-hop backhauling.

Cross carrier scheduling in mmW multi-hop backhauling: For example, in the scenario shown in **Figure 4-5**, when a data packet is forwarded from RN-1 to RN-3 on the mmW SCell of the backhaul link, a PDCCH is typically required to inform RN-3 about this particular data packet transmission. It is assumed that the control channel is transmitted in the PCell while the scheduled data is transmitted in the mmW-based SCells. Moreover, to benefit from the existing robust design and maximize the backward compatibility, the R-PDCCH defined in LTE Rel-10 for relay nodes is transmitted from the PCell to convey the downlink data allocation and/or uplink scheduling request information for the data packet in the mmW SCell. This means that the cross carrier scheduling is adopted for the backhaul link with carrier aggregation of mmW SCell established by the procedure described in the previous section. The cross carrier scheduling for the mmW small cell is illustrated in **Figure 4-9**. With this approach, we do not need to design a completely new downlink control channel for the mmW CA operation, and all the design benefits offered by the Relay PDCCH (R-PDCCH) are naturally available to the mmW CA communication. The example in **Figure 4-9** illustrates the use of a single secondary cell, to underline the principle. The available bandwidths of the mmW are not yet fully standardized, and the bandwidth

needs may vary. Therefore, several SCells may be combined with a contiguous and non-contiguous band allocation. The following descriptions showing single SCell for mmW use cases are also applicable for multiple SCells.

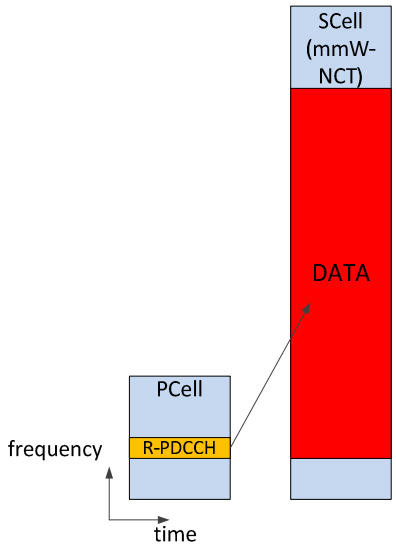


Figure 4-9: Cross-carrier scheduling with R-PDCCH in the mmW backhaul link.

Alternative 1: R-PDCCH transmitted from the centralized DeNB

This alternative is to transmit the R-PDCCH from the DeNB of the macro cell, to schedule the data packet in the mmW SCell along the backhaul links. For the hop-n small cell AP in case of n>1, the R-PDCCH and associated data packet are transmitted from different nodes: the R-PDCCH is transmitted from the DeNB, and the scheduled data packet in the mmW SCell from the upstream small cell AP. As shown in **Figure 4-10**, RN-3 receives the data packet in the mmW SCell from RN-1 while the R-PDCCH scheduling the data packet is transmitted from the DeNB. In this approach, the radio resource of mmW SCell for the link from RN-1 to RN-3 is managed by the DeNB. And RN-3 views RN-1 as the part of the DeNB, e.g., a remote radio head, in the sense that RN-3 sees the DeNB to act as its only donor eNB towards to the mobile core network. Because the R-PDCCH and associated data packet are transmitted from different cells and sites, the DeNB needs to ensure the timing alignment between the R-PDCCH and associated data packet.

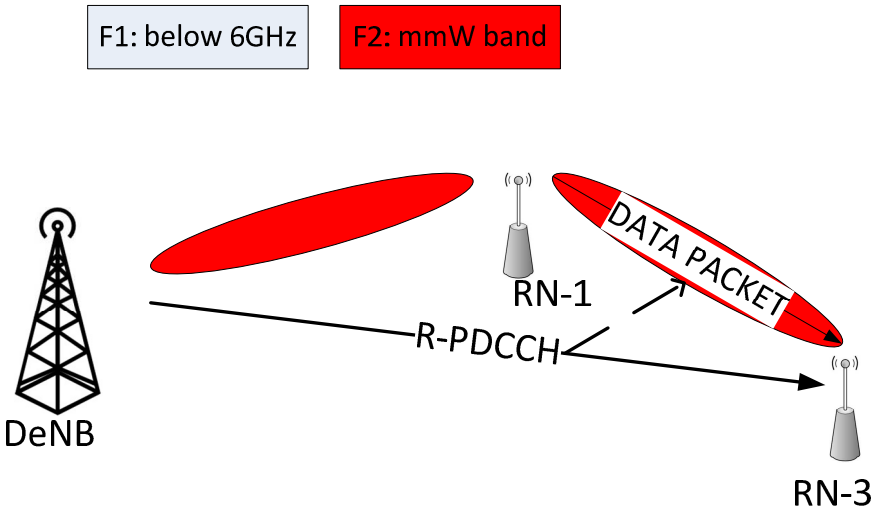


Figure 4-10: The R-PDCCH and data packet from different sites.

Alternative 2: R-PDCCH transmitted from the upstream small cell AP

This alternative is to configure multiple R-PDCCH sets to be monitored by the multi-hop small cell APs so that the R-PDCCH and the associated data packet can be transmitted from the same site. As shown in **Figure 4-11**, RN-3 is configured by the DeNB to monitor two R-PDCCH sets. The R-PDCCHs from one set are transmitted from the DeNB, while the other set is transmitted from RN-1. When the data packet is transmitted from the mmW SCell served by RN-1, the R-PDCCH from the set corresponding to the RN-1 is used to convey the downlink control information to RN-3. Due to the colocation of R-PDCCH and the data packet, the timing alignment problem in the "Alternative 1" no longer exists. Similar to the "Alternative 1", RN-1 is also viewed as the part of the DeNB by RN-3.

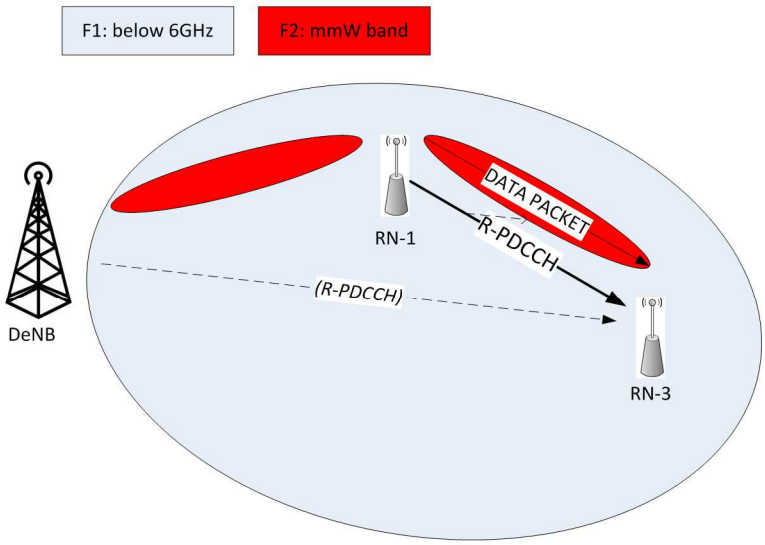


Figure 4-11: Two R-PDCCH sets configured, R-PDCCH and scheduled data packet are transmitted from the same node.

Alternative 3: upstream small cell AP serving as the donor eNB of the downstream small cell AP

This alternative is to employ a two-step DeNB handover approach to camp one small cell to another small cell according to the possible multi-hop routing path of the mmW coverage. As shown in **Figure 4-12**, in the first phase, the DeNB serves as the donor eNB for both RN-1 and RN-3. After the mmW coverage detection by RN-1 and RN-3, such information is reported to the DeNB. The DeNB has learnt that two hop-1 mmW relay links can be formed by DeNB-to-RN-1 and RN-1-to-RN-3, and that RN-3 is out of the mmW coverage of the DeNB. Consequently, the DeNB decides to handover RN-3 to RN-1 so that RN-1 becomes the donor eNB of RN-3. After camping on RN-1, RN-3 only monitors the R-PDCCH transmitted from RN-1 to determine whether data packet is scheduled in the mmW SCell served by RN-1.

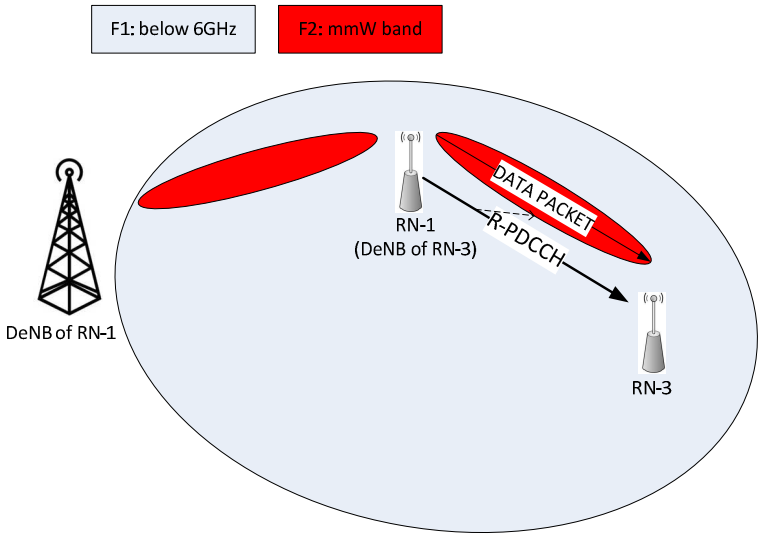


Figure 4-12: One small-cell AP serves as the DeNB of another small-cell AP.

4.2 Dedicated standalone WMN backhaul

4.2.1 Background and motivation

The standalone mmW backhaul is a transparent sub-network that offers first-mile connectivity, that is, Ethernet transport service, between APs and WMN gateways towards the higher aggregation network. The primary design targets are to maximise resilience by exploiting link redundancy in the wireless mesh topology and to keep the forwarding delay minimal and predictable by using an optimized and deterministic link schedule.

The backhaul nodes (BH nodes) and Access Points (APs) may be combined together in the same physical equipment, but it is assumed that BH nodes and AP nodes do not share radio resources and that AP nodes do not participate in the BH network's routing and scheduling scheme. This allows use of mmW frequency bands and development of MAC and PHY layer for the BH link independently of the UE access link. This is in contrast to the inband backhauling solution, described in Section 4.1, which reuses the existing eNB relay concept in LTE-A.

A macro BS may include the WMN gateway functionality (WG) and centralized WMN control functions. The network structure and the network elements involved in the backhaul are illustrated in **Figure 4-13**.

BH nodes are connected to each other with directional point-to-point mmW radio links. Thus, inter-link interference is assumed to be negligible and tractable. Communication among the nodes assumes a specific Time Domain Multiple Access (TDMA) concept: a BH node can communicate with only one neighbour node at a time such that each radio frame can be individually routed to a wanted direction (neighbour) with a beam-steerable antenna, or with separate antennas that cover different sectors. This basic concept of RF resource sharing in a node can be extended to deployments in which a node may have more than one RF units, but a single RF unit is assumed in the following.

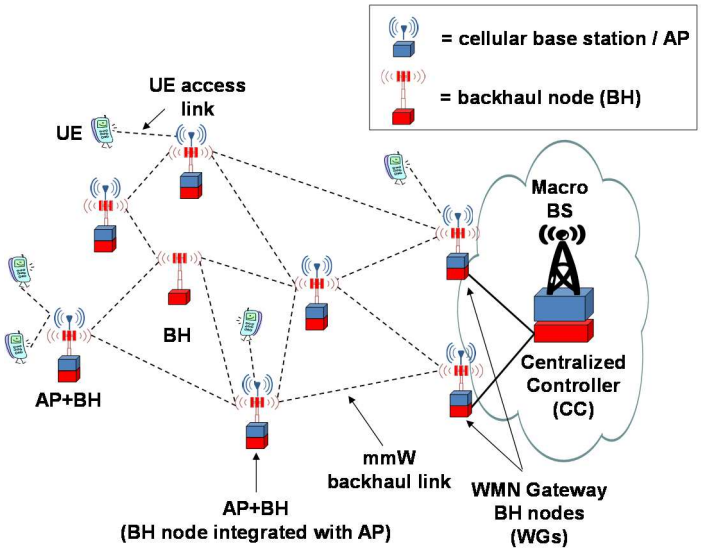


Figure 4-13: The structure of the dedicated backhaul network.

The main features for the mmW WMN include

- Self-organization
  - neighbour discovery
  - automatic network reconfiguration when nodes are added or removed
- Self-optimization
  - adaptation to traffic load and link capacity changes, for efficient use of network resources
- Self-healing
  - fast recovery from node and link failures
- QoS awareness
  - predictable delay for high priority traffic
- Security
  - node and network authentication in node attachment
- Energy efficiency
  - shutting down unused resources / reducing power in lightly loaded resources

One of the main challenges is to meet the envisioned requirements for 5G data plane latency. The requirements imply the need for fast beam alignment: it must be possible to change the beam direction in around 100 µs intervals. This interval also defines the length of the radio frame. In TDD operation, the frame is assumed to consist of variable-length transmission opportunities for the two directions on a link.

The WMN functions include:

- Node synchronization
  - a prerequisite for neighbour discovery and link scheduling
- Neighbour discovery
  - support for autonomous incremental network build-up
- Routing
  - path computation and path selection
- Link scheduling

  - schedule computation and link activation timing
  - dynamic optimisation of the TX/RX ratio between the directions of the link
- Fault and congestion management
  - fault detection and recovery (protection and restoration)
  - load balancing
- QoS provisioning and enforcement
  - traffic prioritization by using path diversity and link/queue scheduling

In the categorization of a routing or scheduling approach to either “static” or “dynamic” with respect to its dynamic adaptivity in the face of changing network and traffic conditions, the solution described in the following applies “dynamic routing” and “static scheduling”. This can be contrasted with the theoretical routing and scheduling framework and algorithms presented in Section 3.1.3. The related simulations demonstrated the advantages of a fully dynamic routing and scheduling solution. Nonetheless, selecting between a dynamic or (semi)static link scheduling approach is not straightforward but involves a trade-off between flexibility and adaptivity on the one hand and complexity and computation/signalling overheads on the other hand.

In the following subsections, the focus is on the neighbour discovery (Section 4.2.2) packet routing (Section 4.2.3), link scheduling (Section 4.2.4), and link duplexing (Section 4.2.5). Here, routing functions embody various fault management, congestion management, and QoS provisioning that drive traffic steering decisions within the backhaul.

4.2.2 Neighbour discovery

Neighbour discovery is a prerequisite for autonomous incremental BH node deployment and an enabler for self-organisation of the BH WMN in general. A BH node that has already been configured as a member of the WMN tries to find its neighbouring nodes by sending beacon messages and waiting for responses from its neighbours. The first WMN member is always a gateway node, which reports itself to the centralized controller (CC).

In the following description of the neighbour discovery procedure, we assume a single RF chain in a BH node and beam switching with discrete pencil-shaped beams, with the constraint that neighbour discovery can only be performed during pre-assigned time periods during the cyclic link schedule (for the link scheduling concept, see Section 4.2.4). The elementary exhaustive search scheme described below could be combined and enhanced with the ideas for search methods introduced in Section 4.3, which assume multiple RF chains in a node with beamforming capability, and exploit multi-path components (MPCs).

For successful discovery, beacon sending and listening must happen at the same point in time while the transmitting and receiving antennas are aligned. Therefore, each BH node that is already a WMN member (“existing node”), acting as an “initiator”, sends beacons to sequential antenna directions during a dedicated control slot that is reserved in the link schedule for the purpose of neighbour discovery (for the description of link scheduling in the BH, see Section 4.2.4). Correspondingly, a BH node that is not yet a WMN member (“new node”), acting as the “responder”, listens for beacons, from sequential antenna directions. When the responder receives a beacon, it responds with an acknowledgement. The time periods of beacon sending and listening in each direction should be adjusted such that the initiator’s and responder’s beaming directions are bound to coincide within a definite maximum discovery time. Because an existing node has to interleave neighbour discovery

operation with normal data transmission according to the configured cyclic link schedule, the node must maintain state related to its current beaconing direction.

Eventually, after having received an acknowledgement to the beacon, the initiator informs the CC of the found link. The centralized controller then adds the link to its conception of the physical WMN topology. Finally, the controller may include the link to the set of active links and reconfigure the WMN with a new routing topology.

Assuming that beacons are sent to only one beam direction during one control slot in each schedule cycle, the worst case discovery time T_W of the exhaustive search depends on the number of possible beam directions and on the length of the schedule cycle:

$$T_W \cong {N_D}^2 T_C \quad (1)$$

where N_D is the number of discrete beam directions in a node (assumed to be the same for all nodes) and T_C is the length of the link schedule cycle.

For example, if the number of beam directions N_D = 64 and the link schedule cycle length is T_C = *1 ms*, and assuming a "perfect channel", the discovery time is about 4 s, at most, which can be considered sufficiently fast for the alignment of semi-static backhaul links. The average discovery time is half of the maximum time.

In the above, an existing WMN member node was discovering a new candidate member node, of which only the former was bound to following the WMN-wide link schedule. However, it is also possible that existing neighbour nodes have not yet discovered each other. If the existing nodes share the common link schedule, the discovery can be accomplished by (temporarily) assigning one of the nodes the responder role and following the above-described discovery procedure. Because in his case also the responder is bound to go along with the schedule, it is able to listen for beacons only during the control slot. This, however, is not a disadvantage because beacons are sent only during the control slot anyway. Therefore, the above-mentioned search procedure and the discovery time estimations apply also in this case.

4.2.3 Routing

Because mmW links are fragile compared to wired links, the routing solution cannot simply reuse existing routing protocols and mechanisms designed for wired networks, much less those for wireless ad hoc networks. Instead, packet routing must be prepared for rapid reaction to frequently occurring link quality impairments (abrupt changes in link capacities) and traffic fluctuation.

One of the most important performance metrics in 5G small cell backhaul is delay, which depends not only on routing but also heavily on link scheduling. Our approach to delay optimization is to separate routing and link schedule computations and to make the latter one easier by optimizing edge-to-edge delays only for a certain set of paths called primary paths. These primary paths are the best paths resulting from the route computation and they are given as an input to the link schedule optimization procedure.

The interaction between the computations is outlined in **Figure 4-14**, which also suggests an optional iteration: if link schedule computation cannot fit a satisfactory schedule on the routing overlay, route computation should be retried.

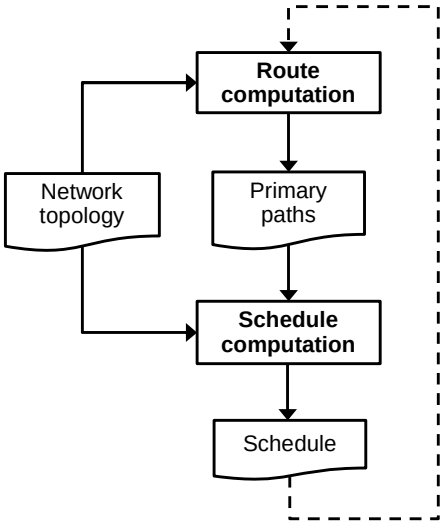


Figure 4-14: Route and schedule computation pipeline.

Both routing and link scheduling involve: 1) prior CPU-intensive computations in a centralized controller, which configures the BH with the routing paths and the network-wide link schedule; and 2) eventual in-network runtime operations, which make use of the pre-computed configurations. In this section, the route pre-computation algorithm is specified first, followed by a description of the network's run-time functions that support resilience in the face of node/link failures and capacity/traffic variation. Link scheduling is described in Section 4.2.4.

4.2.3.1 Route computation

The requirement for fast route recovery in case of failures necessitates a path protection scheme. Therefore, sets of alternative directional routing paths are pre-computed for each pair of BH nodes. A source BH node can then select a path for a traffic flow on demand based on the paths' current states and preferences, and based on the traffic flow's priority. To avoid fate sharing among the paths, route pre-computation aims at producing alternative paths that are as far node-disjoint as possible, while still keeping path lengths (hop counts) in reasonable limits.

The majority of the traffic can be assumed to flow via WGs, which become the most likely points of congestion. Thus, load balancing and other congestion mitigation methods should be able to move traffic from one WG link to another. To this aim, the routing structure is based on WG-link specific spanning trees (ST). Each ST is rooted at a WG and the root has always only one child, which is one of the WG's neighbours. Thus, the total number of STs in the network is the same as the number of WG links of all WGs, and each non-WG node belongs to every ST. Then moving traffic away from a congested WG link is equivalent to allocating the traffic to another ST. Furthermore, if these STs are sufficiently non-overlapping, they should then provide close to the maximal number of disjoint paths to the WG for any BH node.

The first step in the computation of the "multiple disjoint spanning trees" (MDST) is to generate *stems* for each ST for one WG. A stem is a part of a ST that does not share any nodes or links with other stems that originate from the same WG (except for the WG itself). Each non-WG node in the network should belong to one of the stems rooted at each WG. To get balanced stems that provide short routes for all nodes, all the stems originating from a single WG are grown (i.e., nodes are allocated to the stem) in parallel. This growing procedure progresses hop by hop so that:

- If an unallocated node neighbours only one stem, it is added to that stem.

- If an unallocated node neighbours two or more stems, a tie-breaking procedure is used to decide which stem takes that node.

The tie-breaking procedure takes into account the number of nodes allocated to stems and the path cost to the WG. The results of the stem-generating phase for a 5 × 4 regular grid network are shown in **Figure 4-15**.

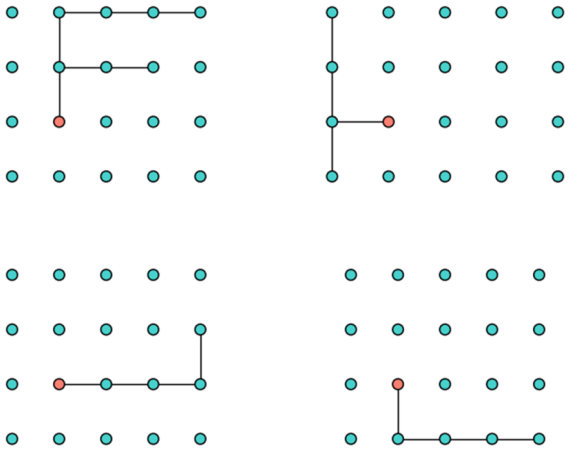


Figure 4-15: Stems generated for a 5×4 regular grid network with one WG (the red node).

As the next step, each stem is expanded to a full ST by growing "branches" from the nodes that are already included into the stem. Here hop counts to the WG and the aggregation of the estimated traffic load on WG links can be used to steer the ST expansion. The final STs for the example 5 × 4 regular grid network are shown **Figure 4-16**.

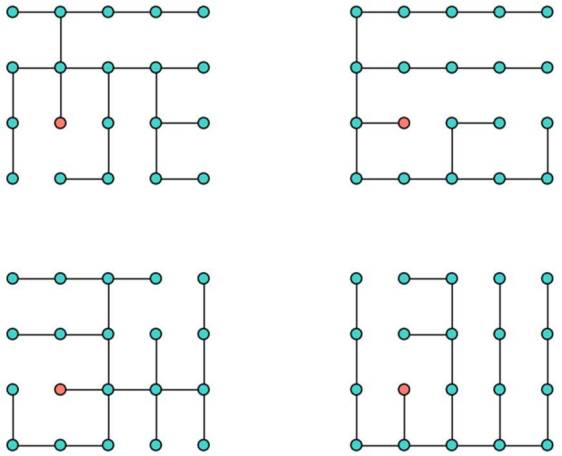


Figure 4-16: Stems grown to full spanning trees.

The final step in route computation is primary path allocation, in which the preferred WG and routing path (i.e., preferred ST) to that WG are assigned for each node. First, the nodes that have only one clearly best path to one WG are allocated to those paths and WGs. Secondly, the nodes that have two or more equal-cost paths are allocated so that the expected traffic demand, e.g., based on traffic statistics, is divided among STs (i.e., among WG links) as equally as possible.

4.2.3.2 Resilience and load balancing

The goals in route computation and path selection are to minimize path delay and to avoid packet loss caused by congestion or failures – at least for the highest priority traffic classes. To that aim, the states of the paths need to be continuously monitored. Based on these states, a traffic flow may need

to be moved to an alternative path. Diversion of a flow from its original path to an alternative path may be caused by recovery or congestion avoidance decisions that operate in different time scales. In the order of increasing reaction and recovery time, the operations may consist in: (i) fast local rerouting to by-pass a single failed BH node or link; (ii) path reselection at the source BH node when there is a failure or congestion on the current path; (iii) local repair of the routing structure; (iv) full recomputation and reconfiguration of the routing paths. Fault recovery and congestion management are illustrated in **Figure 4-17**, which highlights the different time scales in the operations.

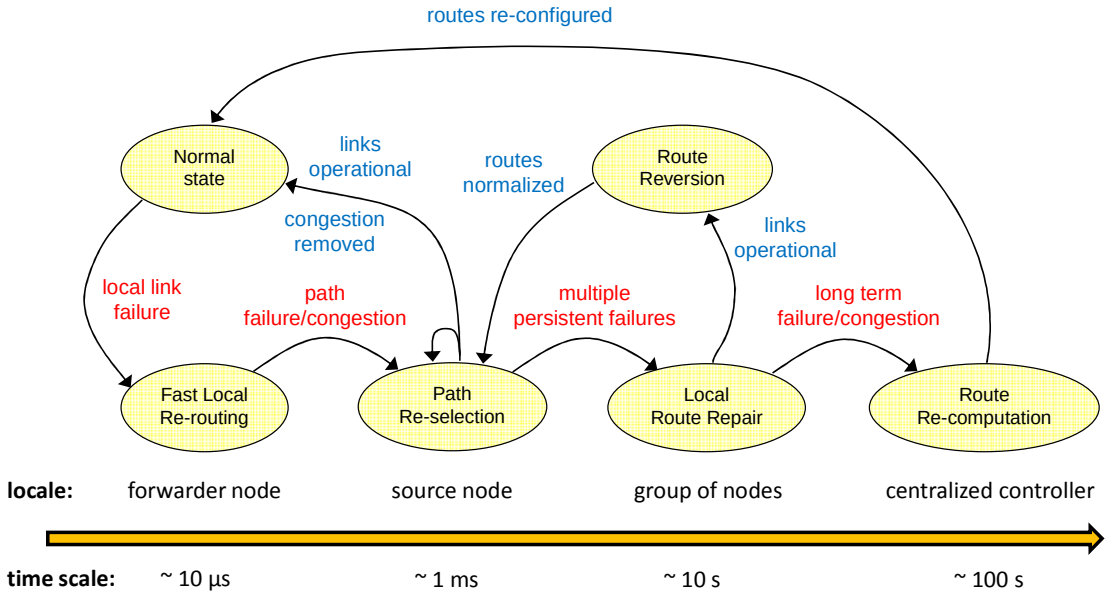


Figure 4-17: Fault recovery and congestion management.

A link failure, which may be a consequence of a node failure, is first locally recovered at the end node(s) of the link by fast re-routing in order to maintain as hitless forwarding behaviour as possible. This protection operation is analogous to Fast Reroute (FRR) in Multiprotocol Label Switching (MPLS). The fault is then signalled in the network, which makes traffic-originating end nodes to reselect new paths for their flows, taking into account the flow priorities. The re-steering of a flow may cause packet re-ordering, which requires re-sequencing capability at the destination end nodes.

If the failure persists, the network nodes may initiate the repair of the routing paths in order to restore connectivity. If the links later revive, the routing paths can be reversed back to their original configuration. In long-lasting failures or congestion, the whole network needs to be reconfigured with new routes and schedules. Because route and schedule computations require substantial computing capacity, these computations are assumed to be performed in a centralized entity, which may reside in the macro base station.

One challenge in the proposed solution is to verify that the use of distributed decision algorithms that dynamically select forwarding paths for traffic flows will result in optimised overall network resource usage.

4.2.4 Link scheduling

Scheduling means allocation of time slots to BH links during which they can transmit or receive. The link schedule is determined by a network-wide semi-permanent configuration of a sequence of (maximal) transmission sets. Each transmission set defines which links in the network can be active simultaneously, with the constraint that any two links in a single transmission set cannot have common

end points. This constraint, which originates from the assumption of BH nodes having a single RF unit, could be relaxed with multi-sector BH nodes equipped with more than one RF units.

The transmission sets are cyclically recurrent, with a cycle length of about 1 ms. Besides the network-wide link schedule, the transmission opportunities in the two transmission directions may be determined by a "link-internal schedule", according to a dynamic Time-Division Duplex (TDD) scheme, which will be examined in Section 4.2.5. Yet another dynamic aspect of the essentially static scheduling is entailed by the Modulation and Coding Scheme (MCS). However, this link rate adaptation is beyond the scope of the WMN-wide schedule computations, although awareness of the link capacity is used in dynamic routing decisions.

Both route and schedule computations take the physical network topology, including the locations of the gateway nodes, as input. The physical topology view builds up as an outcome of the incremental neighbour discovery (described in Section 4.2.2). The schedule computation further needs input from the route pre-computations (described in Section 4.2.3), in order to become aware of the routing paths that would be primarily used for the highest-priority traffic classes.

The computed transmission sets are always assumed *maximal*, which means that it is not possible to add any more links to a single transmission set. A "good" transmission set is also close to *maximum* in the sense that the number of links in it is close to the maximum possible.

The node degrees (number of scheduled links starting from a node) should be limited in order to meet the path delay requirements. If large node degrees are allowed, then some of the paths can still be selected to be fast but with a consequence that some other paths will be impractically slow and thus useless.

An example of a schedule with 8 transmission sets (which can also be considered as link colours) in a grid-shaped network is shown in **Figure 4-18**.

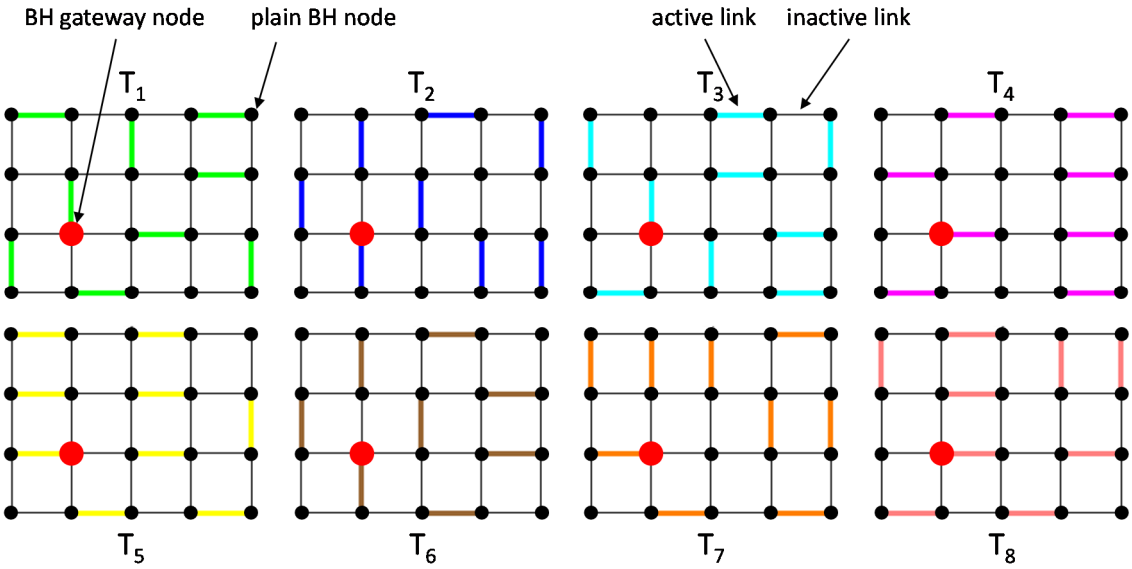


Figure 4-18: An example of a schedule with 8 transmission sets (colours) in a grid network.

As shown in **Figure 4-18**, a node can communicate with only one neighbour at a time. On the other hand, the transmission sets are not disjoint: a single link may belong to several transmission sets (i.e., to have several colours), especially if it carries any number of primary paths. This can be considered as optimizing the spatial reuse of the links: important links are active as often as possible. Link activity

can be tuned by adjusting the length of the time slots. However, it is assumed in the following that all time slots (T_i) are of equal length.

The 8 colours of the example in **Figure 4-18** form a cycle that is repeated in the order described in **Figure 4-19** below.

$$\underbrace{\ldots,T_7,T_8,\underbrace{T_1,T_2,T_3,T_4,T_5,T_6,T_7,T_8}_{cycle},T_1,T_2,T_3,T_4,T_5,T_6\ldots}_{schedule}$$

Figure 4-19: Transmission sets in the cyclic schedule.

Once a presumably representative set of transmission sets has been generated, the actual schedule optimization is performed by trying to find the optimal subset of the transmission sets and their order of appearance in the schedule cycle. The comparison between schedule candidates is made by applying a score (fitness) function. Because the main concern is network latency, a suitable target is to find a schedule that provides the minimum of the worst-case path delays across all primary paths. If several schedule candidates give the same minimum, a secondary criterion is applied among the candidates: select the schedule that provides the minimum average path delay.

Design issues related to the schedule computation include: determination of suitable schedule cycle lengths for different network topologies and development of heuristic algorithms for collecting a good enough set of transmission sets and for finding their optimal order in the link schedule. Initial solutions for these issues are proposed in the following.

4.2.4.1 Link schedule computation

The path delays need to be optimised bi-directionally between WGs and plain BH nodes, which suggests constructing a link schedule cycle that contains a number of time slots (*N*) that is twice the number of links in the WG. If there are multiple WGs, the one with the largest number of links is considered.

In the first step of link schedule computation, sets of feasible transmission sets are generated. At least links on the primary paths between plain BH nodes and WGs should be scheduled such that the paths satisfy the network delay constraints. Therefore, each link is given a weight according to how many primary paths traverse the link. The weight can then be used to guide the assignment of time slots to the links. The goal is to generate a set of *N* transmission sets that is populated by links according to their weights, by satisfying the following viability criteria:

1) Each link must be included in at least 1 transmission set.
2) Each link with non-zero weight should be included in at least 2 transmission sets.

Further minor eligibility criteria include:

a) The transmission sets should be as maximal as possible (e.g., measured as the overall link activity percentage across all transmission sets).
b) The transmission sets should be as far link disjoint as possible, e.g., measured by computing the Jaccard distances (the sizes of the normalized symmetric difference between sets: $|A \triangle B|/|A \cup B|$) across all pairs of transmission sets.

If the success criterion 1) cannot be satisfied, the target value for the number of transmission sets (*N*) needs to be increased.

After generating the set of transmission sets, their cyclic order needs still to be optimised with respect to the previously mentioned delay targets (maximum and average path delay). The path delay can be simply obtained by emulating packet forwarding along the path assuming link scheduling according to the computed transmission sets (as was indicated in **Figure 4-19**). In the delay estimation, an unloaded network carrying high-priority traffic is assumed (i.e., a packet never misses its transmission opportunity).

Finding the optimised order for the transmission sets is computationally hard because there are $(N-1)!$ permutations to be considered. Some speed-up in the computing can be gained by handling the paths in the longest first manner: if the maximum path delay exceeds the current best, the permutation can be dropped immediately. One approach is to first make a collection of cycle candidates by applying the rather arbitrary transmission set order, and then to select the one of the candidates that produces the best delay result as input for the subsequent actual combinatorial optimisation. This is a trade-off between computation time and link schedule performance — it is possible that some candidate with worse initial score would give a better schedule. In any case, the computation time required by the optimisation depends on how close the selected initial transmission set order is to the optimum.

For short link schedule lengths, extensive search across all permutations is feasible. However, with long schedules, say, longer than 10, metaheuristics are required. For example, with simulated annealing based optimization, new schedule candidates can be generated by using pair-wise swap of transmission sets.

4.2.4.2 Performance of link schedule computation and scheduling

Link schedule optimization speeds were evaluated using a regular 9×9 grid network with one WG. **Figure 4-20** shows a density plot, smoothed for visualisation, containing all permutations of a cycle, mapped to of average vs. maximum path delays.

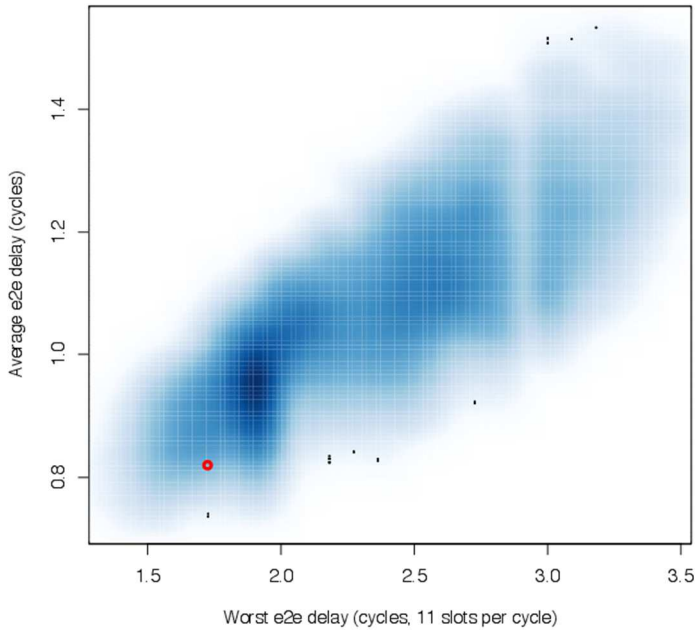


Figure 4-20: Smoothed scatter plot showing “density” of schedule candidates.

Darker regions in **Figure 4-20** represent areas where most of the cyclic orders lie. The red ring shows the location of the initial cycle candidate. As a new cyclic order is rejected immediately when a

worse path delay compared to the current best is found, and paths are evaluated starting from the longest ones, it is likely that only a few paths have to be evaluated for those orders that lie at the right side of the current winner in the density plot. Thus, the "initial guess" shown in **Figure 4-20** has been quite fortunate when considering the required computation time.

Computation performance was tested with a mid-2009 MacBookPro having 3 GHz Intel Core 2 Duo CPU and 8 GB memory. The computation times for different schedule lengths are shown in **Table 4-1**.

Table 4-1: Link schedule computation times.

Schedule length	Extensive search	Simulated annealing
8	7.1 s	
9	8.7 s	
10	21.2 s	3 m 20 s
11	2 m 37 s	3 m 14 s
12	15 m 35 s	3 m 47 s
13	> 2h	4 m 06 s

Various parameter combinations and cooling strategies for simulated annealing were examined. In the end, slow exponential cooling turned out to be the most suitable considering the need to have reasonably short computation times. The progress of the optimization in one computation run is shown in **Figure 4-21**.

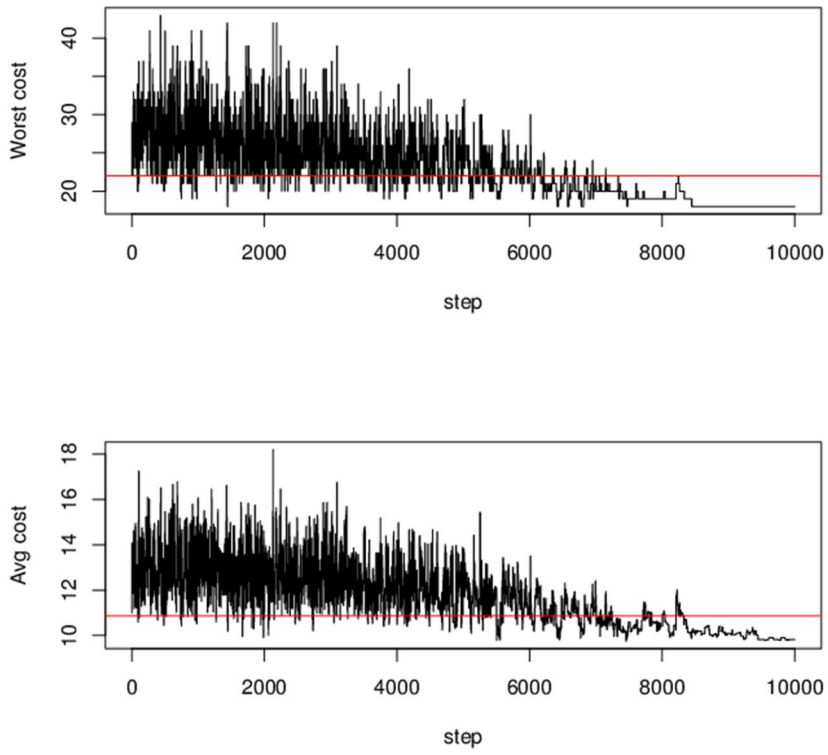


Figure 4-21: Optimizing link schedule with simulated annealing.

The results of the route and link schedule computations have been analysed with a regular 6 × 6 grid network (36 nodes and 60 links) with a single WG at location (2,2). **Figure 4-22** shows a boxplot for expected path delay between BH nodes and the WG.

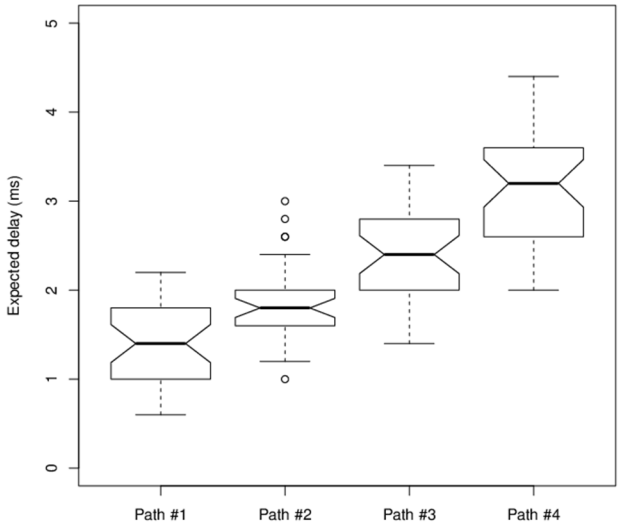


Figure 4-22: Expected maximum delays along paths, in decreasing order of path preference.

Delays were calculated assuming worst-case waiting time at the ingress node (that is, delay until the activation of the first-hop mmW link according to the schedule). Four alternate paths for each node pair were computed by assuming a schedule cycle length of 1.6 ms containing 8 transmission sets. As the maximum hop count in this example is 6, which is rather extreme in the typical use cases, the results show that it is possible to optimize the link schedule such that even in the worst case, the longest delay for the lowest priority path is much less than the cycle length multiplied by the hop count.

Furthermore, the longest (6-hop) path can be travelled in less than 1.4 cycle lengths in the best case. It can be also noticed that primary and secondary paths most often offer rather similar performance and thus load balancing between them would not usually affect QoS. The optimization (both routes and link schedule) took 4.4 s in total. Hence, new routes and link schedules can be computed on demand, e.g., if traffic load moves from one part of the network to another.

4.2.5 Dynamic TDD

TDD is applied locally on a link between two BH nodes at the ends of the link. Dynamic TDD (dTDD) enables adaptation of the directional link capacity to the variation of the traffic load on the link. This section examines means of applying dTDD in the BH WMN that at the network-level follows a cyclic link schedule.

In the network-wide link schedule, the capacity of the link between two BH nodes is shared in time domain using fixed-length time slots during which both nodes have a sending opportunity. Let us call these two opportunities “Slot A” and “Slot B” so that Slot A is the first slot and Slot B the second slot (time-wise). Further, let us assume that the nodes are alternating in sending the slots. That is, if node X is sending in Slot A (“A-sender”) and node Y in slot B (“B-sender”) in the current time slot then node Y will be A-sender and node X B-sender in the next time slot, and so on, as shown in **Figure 4-23**. The principle of alternating the A/B-sender roles enables schemes for fair sharing of the link capacity between the nodes without explicit signalling.

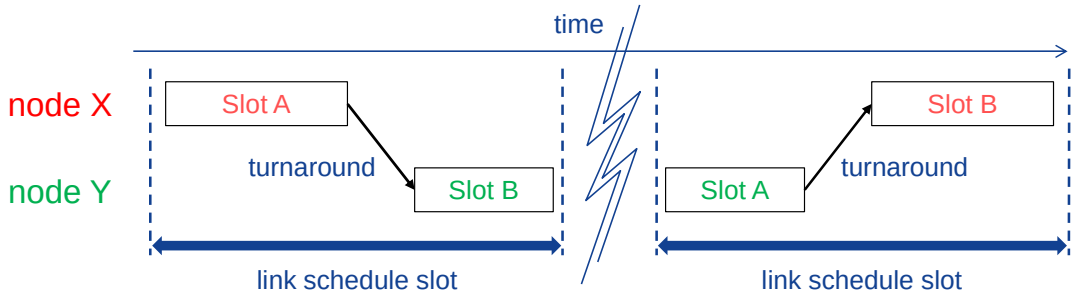


Figure 4-23: Alternation of Slot A/B sender roles between the end nodes of a link.

4.2.5.1 dTDD policies

Different policies for dividing the scheduled time slot between the end-points of a link can be considered. In any case, the division ratio of the current slot may depend on the nodes' previous and current capacity demands and on their actual prior data transmissions. These elements are illustrated in the TDD model in **Figure 4-24**.

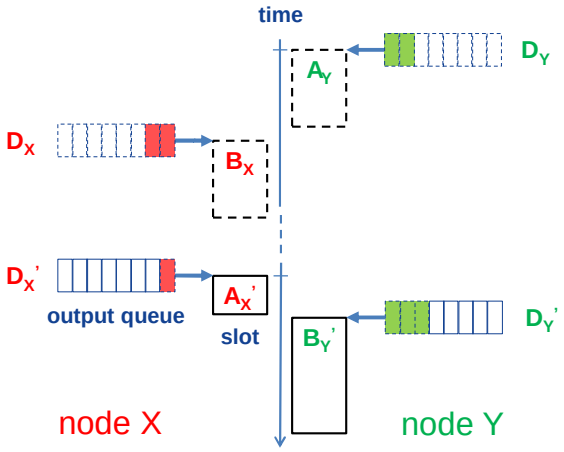


Figure 4-24: dTDD model.

The model introduces the parameters listed in **Table 4-2**:

Table 4-2: Parameters in the dTDD model.

Parameters	Definition
D_X and D_Y	the traffic demands (the number of bytes in the output queue) of node X and Y at the time of their starting to send the previous Slot B and Slot A, respectively
D_X' and D_Y'	the traffic demands of node X and Y at the time of their starting to send the current Slot A and Slot B, respectively
B_X and A_Y	the amount of data (in bytes) that nodes X and Y had sent in the previous Slot B and Slot A, respectively
A_X' and B_Y'	the amount of data (in bytes) that nodes X and Y will send in the current Slot A and Slot B, respectively.

Further, let us define $S = A_X' + B_Y' = B_X + A_Y$ as the total pre-known fixed capacity of the link schedule slot. The problem is to determine the amount of data that node X in its role as A-sender will send in Slot A, that is, to determine A_X' (which also implies the maximum of B_Y'). Possible policies for computing A_X' include:

Gross: The slot capacity division is fixed: $A'_X = \frac{S}{2}$, where A'_X includes the amount of $S/2 - D'_X$ padding if $D'_X < S/2$. This is a kind of baseline TDD operation, without any adaptation to the traffic variation.

Guarded: The slot capacity *S* is divided equally between the nodes: $A'_X = min\left(D'_X, \frac{S}{2}\right)$. In this case, B-sender will get more than the half of the link capacity ($B'_X > S/2$) if A-sender sends less than $S/2$.

Greedy: Node X uses as much of the slot capacity as it needs and can, that is, $A'_X = min(D'_X, S)$. Node Y will be compensated for the temporal unfairness by the node itself being greedy on its turn as A-sender.

Gracious: Node X remembers its previous use of Slot B (B_X) and sends $A'_X = min\left(D'_X, max\left(B_X, \frac{S}{2}\right)\right)$. This means that if a node does not have much to send it can relinquish capacity to its neighbour for a longer term, but when the node gets more data to send, it can reclaim its share of the capacity next time when it is A-sender as it can use at least one half of the time slot.

Generous: Node X anticipates that its neighbour would like to go on sending as much data as it sent in its previous turn as A-sender (i.e., A_Y). Therefore, node X tries to balance between its own and its neighbours assumed interests, by fair sharing of the capacity: $A'_X = min\left(D'_X, \frac{D'_X}{D'_X + A_Y} S\right)$.

Granting: Node X knows its neighbour's previous demand D_Y and tries to reconcile between the competing demands by sending $A'_X = min\left(D'_X, \frac{D'_X}{D'_X + D_Y} S\right)$. Because node X cannot deduce D_Y based on the data plane transmissions, node Y would need to send an explicit D_Y capacity request out of band within its own Slot A. As a drawback, D_Y may be rather outdated at the time when node X starts transmitting.

Godly: Node X knows its neighbour's traffic demand at the time of initiating its own Slot A transmission (let us call this demand $\widehat{D}'_Y$, which is a slightly anachronistic, and thus possibly underestimating, approximation of the eventual D_Y'). Node X tries to balance between the demands by sending $A'_X = min\left(D'_X, \frac{D'_X}{D'_X + \widehat{D}'_Y} S\right)$. In practice, it would be impossible for X to learn $\widehat{D}'_Y$. Therefore, this dTDD policy can be only used as a theoretical reference to "ideal" dTDD behaviour.

According to the above-described model, any dTDD policies that do not require explicit signalling between the nodes would be of the form $A'_X = f(D_X, B_X, D'_X, A_Y)$. The policies could be refined by analysing the longer-term directional balance of the traffic on the links, e.g., by computing the Exponential Moving Average (EMA) of the data rates in both directions, and possibly (in the "guarded" and "gracious" cases) offsetting the slot's point of fair division away from the default *S/2* middle point. Further, packet loss could be used to indicate of an urgent need for more capacity. If some minimum Slot B allowance (B_{min}) should be defined then node X has to be constrained to sending at most $A''_X = \min(A'_X, S - B_{min})$. In practice, the effective capacity is also reduced by various overheads (e.g., guard times and frame preambles), which have been ignored in the above.

4.2.5.2 dTDD performance

The performance of the dTDD policies introduced in Section 4.2.5.1 was analysed by simulations on a simple BH network having the tree topology shown in **Figure 4-25** (in the Figure, the digit on each link indicates the number of transmission sets in the link schedule cycle, assigned to the link).

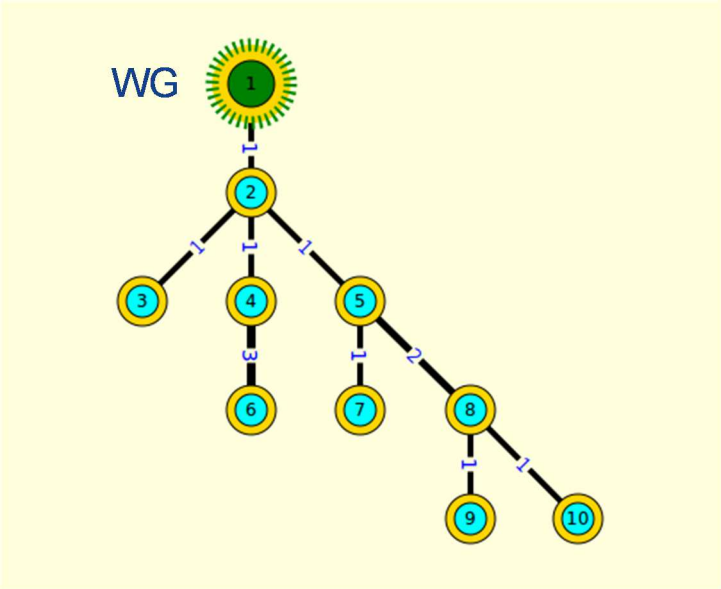


Figure 4-25: Backhaul topology used in the dTDD simulations.

In the network, node 1 is the WG, the downlink of which was in the focus of the measurements.

The simulations were carried out by using ns-3 network simulator. The most relevant simulation parameters are listed in **Table 4-3**.

Table 4-3: Simulation configuration parameters.

Context	Parameter					Value
link	bit rate (perfect channel)					1 Gbit/s
	output queue size					200 packets
	encapsulation overhead (IP over PPP)					4 bytes
link schedule	cycle length					4 slots
	number of control slots in the schedule cycle					0 slots
link schedule time slot	slot length					100 µs
	"mini slots" in a slot (Slot A/B granularity)					100 (x 1 µs)
	minimum time allocation for B-sender					8 µs
	frame overhead (guard times, etc.)					8 µs
traffic	UDP	packet size (UDP header + payload)				255 bytes
		distribution	Pareto	on	mean	2.0 s
					shape	2.0
					bound	5.0 s
				off	mean	0.5 s
					shape	2.0
					bound	1.5 s
		duration				30 s
	TCP	version				NewReno
		application				bulk
		duration				300 s

The Slot A/B size can vary in the range $8\ \mu s \ldots 84\ \mu s$, which corresponds to $1 kB \ldots 10{,}5\ kB$ of data.

In the simulations with UDP, on/off UDP traffic streams with Pareto-distributed on and off times were generated bi-directionally between each BH node and the WG. The Pareto parameters shown in **Table 4-3** were used in ns-3 for constructing random numbers from a fixed Pareto distribution with the specified mean value, shape (alpha), and upper bound. In order to examine the dTDD behaviour in the face of diverse levels of traffic asymmetry, the relative upstream and downstream loads were varied while keeping the total network-wide traffic load constant. A series of 51 mixes of the upstream

and downstream rates were generated. That is, the WG sent one UDP stream down to each plain BH node and each plain BH node sent one UDP stream up to the WG such that, in the "on" state, for each stream:

$$\left.\begin{matrix} GW\ node\ down\ rate = ND \\ BH\ node\ up\ rate = (K - N + 1)D \end{matrix}\right\}, N \in [1 \dots K], K = 51$$

The lowest stream rate *D* was selected appropriately to keep the packet drop rates reasonable. At $N = 26$, the upstream and downstream rates were equal.

The packet loss on the WG link and the 99th percentile delays on the WG downlink were observed, with the results shown in **Figure 4-26**.

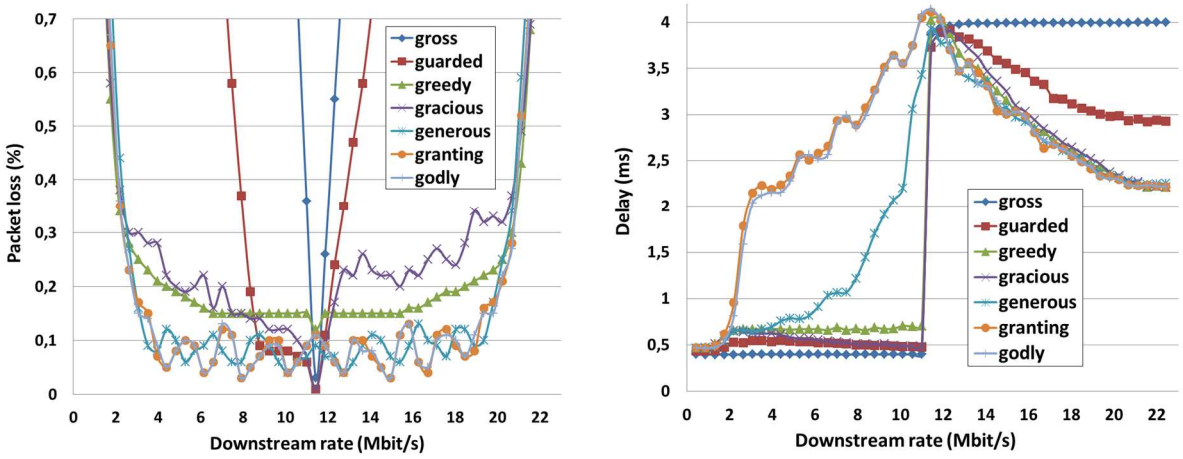


Figure 4-26: UDP: packet loss on the WG link and 99th percentile delays on the WG downlink.

The packet dropping is rather similar in all TDD policies that try to fully exploit the available transmission time.

The WG downlink delay profile in **Figure 4-26** (of which the corresponding uplink delay profile would, roughly speaking, be a mirror image) points out the "gracious" policy as the overall winner, though the "greedy" policy performs almost equally. With the fairly large 200 packet queue size, a full queue contains $200 \times 255\ bytes = 51\ kB$ of UDP payload (not including encapsulation overheads), which takes at least 5 Slots A/B to transmit. In the worst case, this translates to 5 full schedule cycles (400 µs each) which takes a total of 2 ms.

In the simulations with TCP, traffic streams were generated by using ns-3's built-in TCP bulk send application that sends data as fast as possible, essentially keeping a constant flow of data. The WG sent two TCP streams down to each plain BH node and each plain BH node sent one UDP stream up to the WG. The directional throughputs and packet losses on the WG link were observed, with the results shown in **Figure 4-27**.

TCP congestion control mechanisms tend to share the available path capacity fairly among the streams, thus smoothing out the performance differences in the dTDD policies.

This analysis seems to indicate that the alternation of the A/B-sender roles alone provides adaptivity to the traffic variation without a need for devising elaborate load-balancing algorithms for dynamically adjusting the TDD division ratio. That is, even the simple stateless methods, like "Greedy"

and “Gracious”, provide flexibility and performance comparable to the more sophisticated adaptive algorithms.

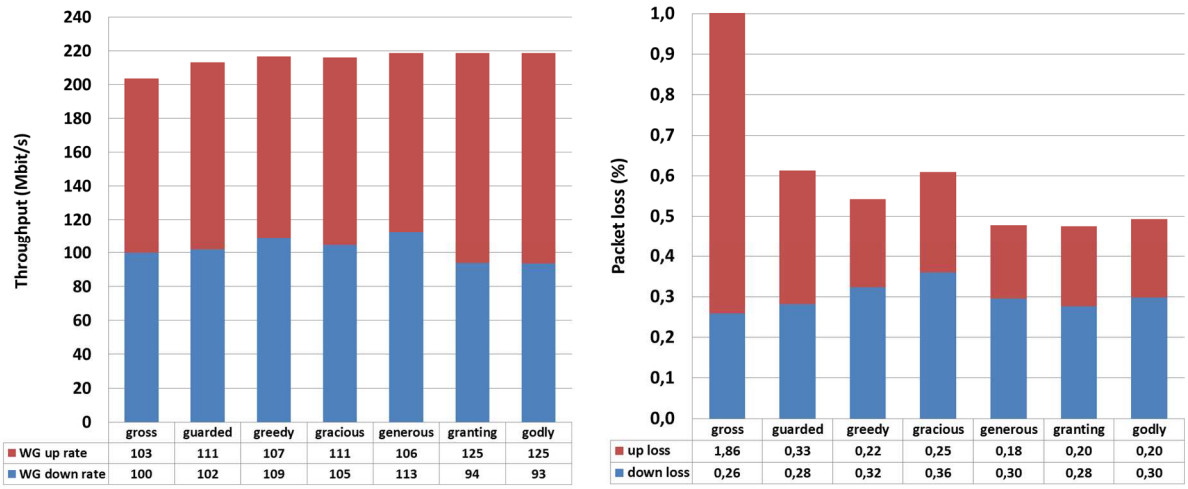


Figure 4-27: TCP: throughput and packet loss on the WG link.

4.3 Beam steering in backhaul

This section expands on the ideas for backhaul neighbour discovery described in Section 4.2.2 (which assumed simple beam switching with a single-RF system) by proposing efficient beam-steering algorithms for scenarios in which the node architecture includes multiple RF chains, and which support beam forming, instead of only switching of pencil-shaped beams.

Due to the nature of electromagnetic waves’ propagation, directional transmission via large antenna arrays or high-directivity antennas and employment of suitable beamforming algorithms is crucial to implement practical communication links in mmW bands. Designing such beamforming algorithms is typically constrained by practical issues related to CSI acquisition and hardware limitations on the required level of power, number of RF chains and RF phase shifter design. In this section, we briefly explore hybrid beamforming and analogue beam switching strategies to combat the above impairments.

4.3.1 Hybrid beamforming

Exhaustive beam search algorithms are generally power-efficient due to employing maximum array gain on both Tx / Rx ends although the number of training steps can substantially increase by the number of supported beams (at either side). If the receiver is equipped with multiple RF chains, it can simultaneously scan multiple directions, leading to a reduction in the training time. The number of training steps can also be reduced to a manageable level via limiting the steering range in the beam search space.

Assuming higher level of access to phase shifters to shape / tune beam patterns, it is possible to reduce the training time via multi-resolution beamforming. In this case, the estimation algorithm starts with lower directivity gain (but wider beams) at the early stages of channel estimation and it gradually narrows down the beams over multiple levels. However, this will be at the price of high transmit power during the estimation phase.

In light of the above discussion, we aim to design a flexible approach between the two (i.e. exhaustive versus multi resolution) with limited training steps and lower training power requirements. In this direction, we developed an enhanced one-sided search algorithm applicable to both analogue

and hybrid beamformers. The estimation algorithm consists of a two-stage handshake between devices d1 and d2 as outlined below. **Figure 4-28** schematically shows the channel estimation algorithm.

At the first stage of channel estimation, only one antenna at d1 transmits a quasi Omni-directional signal whereas d2 will directionally scan the channel as it is shown in **Figure 4-28.a** and **Figure 4-28.b**.

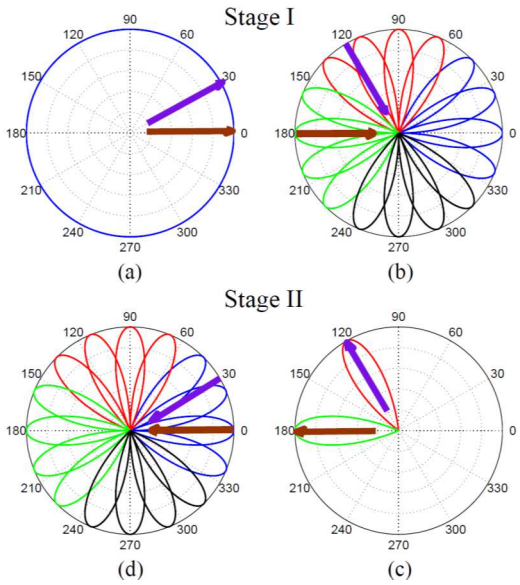


Figure 4-28: Exemplary channel estimation procedure for two multi-path components (MPCs) (purple and brown arrows) – stage I: a) d1 transmits omnidirectional b) d2 scans multiple directions, stage II: c) d2 transmits in the direction of its AoDs, d) d1 scans multiple directions.

At the second stage, d2 transmits the responding pilot signal in the direction of $\psi_{\mathrm{d2},l}$, $l = 1, \ldots, L$. This could be done by applying the hybrid precoding algorithm as shown in **Figure 4-28.c**. Then, d1 applies the stage I approach to detect its AoAs as in **Figure 4-28.d**. After this handshake, one extra transmission is required to determine the relative values of channel gains. Once all the AoAs, AoDs and the relative path gains are derived, the channel matrix can be calculated at both sides.

Next, we present the simulation results of the proposed channel estimation algorithm. The ergodic spectral efficiency is calculated over 2000 channel realizations. For the sake of simplicity, all paths have the same total transmit power over noise power $\gamma = P_t/P_n$. The default system parameters are set as pilot power P_{p} = 35 dBm, data power P_{d}= 20 dBm, phase shifter resolution (number of the bits of the digital phase shifters) B = 2, number of RF chains M = 4, number of antenna elements N = 32.

We evaluated the system performance of hybrid beamforming when multiple streams are transmitted. **Figure 4-29** presents the performance of the proposed channel estimation in the hybrid case when multiple streams are simultaneously transmitted. As shown, the estimation algorithm provides promising performance compared with ideal (unconstrained) SVD (Singular Value Decomposition). The performance gap, however, widens at higher number of MPCs. Nevertheless, increasing the number of RF chains can diminish the performance gap at higher γ.

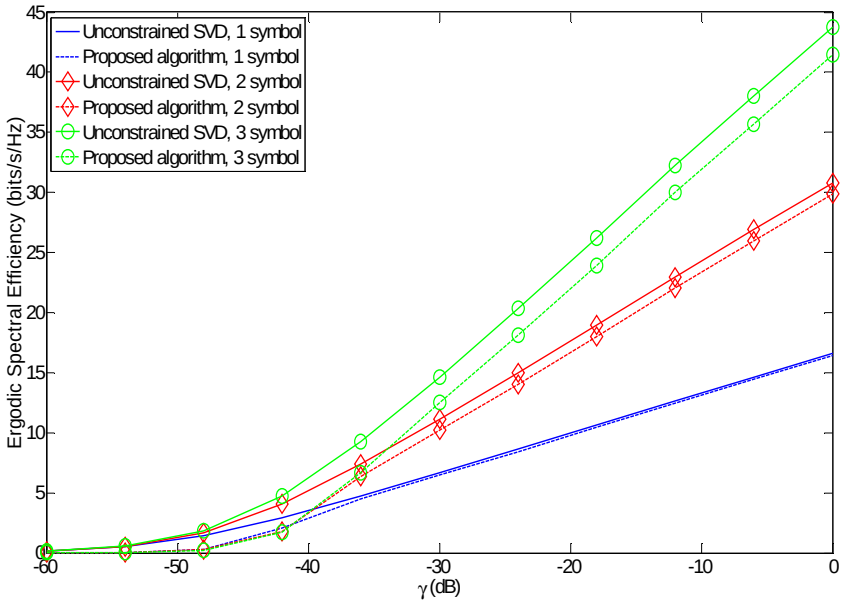


Figure 4-29: Impact of the number of the transmitted symbols on spectral efficiency in multi-stream scenario.

4.3.2 Analogue beam switching

While hybrid beamforming can provide promising performance compared to optimal case (unconstrained SVD), analogue beam switching (with exhaustive-like search) might also be chosen as a viable option for beam alignment in backhauling scenarios to save on cost and complexity of transmitters and receivers within mmW APs / backhaul nodes; in particular, when the steering range and the resulting beam search space is limited. In this case, one MPC could be exploited contrary to hybrid scenario. To evaluate the impact of this strategy in line with link budgets defined in Annex A, here, we provide some guideline performance evaluations. One example link budget option (Annex A, Option BV1/2) is considered, assuming unconstrained capacity. This option reflects the performance of E-band backhaul with Tx / Rx directivity gain of 30 dBi assuming (8 beams in -10 /+10 steering range). The powers P_{p} and P_{d} are set to 13 dBm, respectively.

As shown in **Figure 4-30**, analogue beam switching can provide promising performance in low number of MPCs compared with SVD. However, the performance can notably degrade with increasing MPCs.

So, assuming low number of MPCs, beam-switching with limited steering range can be a viable option for the backhaul. However, for scenarios with richer channel characteristics (higher MPCs), a finer resolution of beam-switching in spatial domain or utilizing hybrid strategies is suggested to compensate the performance loss.

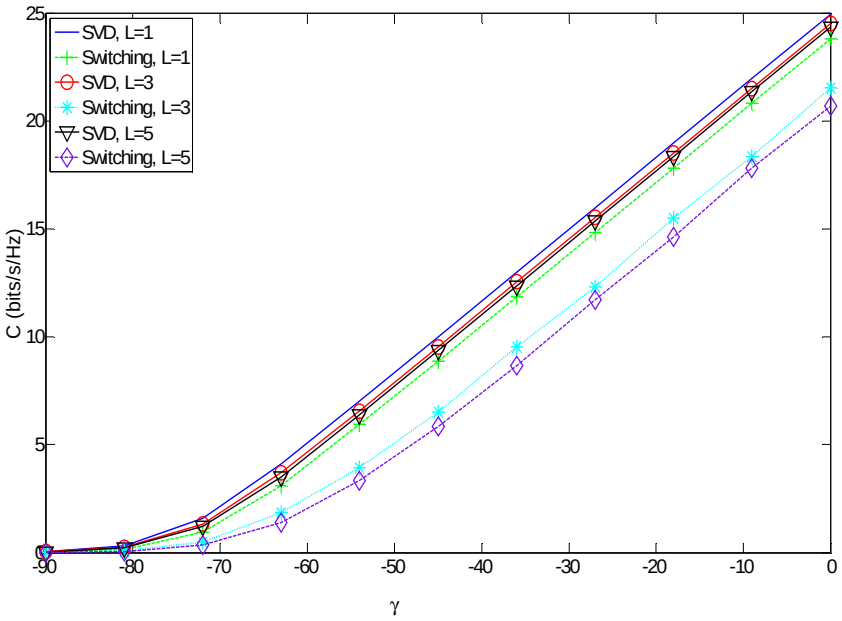


Figure 4-30: Performance evaluation of BV1/2 (unconstraint capacity); L is the number of multipath components of the channel.

5. System functions for the access link

Compared to the backhaul link presented in Section 4, the access link, i.e., the communication path between the UE and the mmW small cell AP, is envisioned to be more challenging due to more dynamic channel condition caused by the mobility of the UE and/or surrounding environment. As such, the probability of the LOS connection in the access link should be smaller than that in the backhaul link, and the resulted average receive signal power is supposed to be lower and vary more frequently. Furthermore, the power consumption of the portable device significantly affects user experience. It is therefore very important to develop energy efficient schemes for UE's communication.

Due to the small wave length of mmW signals, it is feasible to equip the UE with an antenna array with small form factor as well. Hence, it is beneficial as well as realistic to apply the beamforming algorithm at the UE. As mentioned above, due to the UE mobility, it is also important to develop optimal UE-Cell association schemes to improve the overall network efficiency in terms of load balancing and achieved average UE QoS. With the observation of traffic demand time variation, small cell on/off operation has been actively studied in 3GPP, and was supported in LTE Rel-12. As such, it is important that mmW small cell supports the dynamic on/off operation to improve the network energy efficiency.

Given the above motivation, this section is organized as follows. Section 5.1 presents a mmW UE initial access method in the support of the uplink beam alignment to improve the reliability of the subsequent communication. Section 5.2 describes several UE-cell association strategies to improve the overall network efficiency. Section 5.3 analyses with simulations the energy efficiency of the small cell on/off operation in the detail. Section 5.4 presents a beam cell management method to support multi-connectivity and mobility. Finally, a practical beamforming algorithm for restricted antennas for the access link is studied in Section 5.5.

5.1 mmW UE random access to small cell

5.1.1 Background and motivation

The random access channel (RACH) design is very important for a UE to start the initial RRC connection request and/or establish/maintain the uplink synchronization for the purpose of uplink data traffic transmission and/or control information feedback. For example, in LTE the physical uplink control channel is used for ACK/NACK response to downlink HARQ transmission. As mentioned above, efficient beamforming technique is very crucial for the mmW system to mitigate the severe path loss and save the energy consumption by focusing the transmit beam to the desired direction. The downlink beam alignment, i.e., steering the downlink beam to the optimal direction from the small cell node to the UE, can be assisted by macro cell control signalling, e.g., similar to the one detailed in Section 4, taking use of directional discovery signals and its relevant measurements feedback. To initiate an efficient uplink transmission in the mmW spectrum, it is beneficial for the UE to perform an uplink beam alignment procedure as well. Without proper uplink beam alignment, it can be very challenging for the base station to correctly detect the uplink random access signals, especially in case of relatively large TX-RX distance. As such, it is desired or even required to accomplish the uplink beam alignment in the course of random access procedure for the subsequent uplink data transmission.

This section describes a macro cell assisted random access method for the mmW UE, to accomplish the transmit direction alignment in addition to the timing advance information with the small cell node to enable an efficient uplink communication. In particular, based on the existing LTE Physical Random

Access Channel (PRACH) preamble signal design, a multi-preamble PRACH signal is developed. The multi-preamble PRACH probe method enables the UE to achieve the uplink beamforming alignment so that the subsequent uplink data can be transmitted more reliably and efficiently. With the proper uplink beamforming, the PRACH sensitivity of the mmW small cell can be considerably improved.

5.1.2 mmW UE random access in the support of uplink beam alignment

The proposed uplink beam alignment method requires that each mmW small cell supports multiple PRACH configurations (PRACH_Configs). Each configuration corresponds to a certain directional PRACH probe capability, and is allocated to a set of uplink time-frequency resources. The preamble sequences for different PRACH_Configs should be different. For example, a mmW small cell can support 3 PRACH configurations whose preamble sequences are illustrated in Figure 5-1. As shown in **Figure 5-1**, each PRACH probe opportunity of the PRACH_Config 0 includes only one preamble sequence. As such, the PRACH_Config 0 intends to be mainly utilized by those UEs which are not using directional PRACH transmission, i.e., with omnidirectional PRACH transmission. However, the PRACH probe opportunities of the PRACH_Configs 1 and 2 contain two and three preamble sequences, respectively. Different preamble sequences in the same PRACH probe opportunity can be transmitted to different beam directions. The PRACH_Configs 1 and 2 are intended to be used by those UEs with beamformed PRACH transmission capability. It should be noted that multiple preamble sequences in the same PRACH probe opportunity may not be transmitted continuously in case of TDD based downlink and uplink duplex, and when furthermore no adjacent uplink TTIs are configured. In this case, the multiple PRACH preamble sequences in one PRACH probe opportunity can be transmitted in logical continuous uplink TTIs. The time-frequency resource allocations of the three PRACH_Configs are illustrated in **Figure 5-2** and **Figure 5-3**. **Figure 5-2** shows that three PRACH_Configs resource allocations are aligned in time domain, whereas they may use different frequency resources. **Figure 5-3** shows another time-frequency resource allocations for different PRACH_Configs, in which the different PRACH_Configs are allocated in different uplink TTIs. To mitigate the PRACH co-channel interferences at the eNB receiver, the time domain multiplexing shown in **Figure 5-3** can be preferred. However, it is up to the eNB to determine the resource allocation for different PRACH_Configs based on the envisioned RACH capacity and RACH receiver complexity. It should be also noted that different PRACH_Configs may furnish different RACH probe density depending on the envisioned PRACH demand of the particular kind, i.e., the periodicity of RACH probe opportunity can be different for different PRACH_Configs.

The directional PRACH procedure for a UE to accomplish uplink beam alignment is illustrated in the **Figure 5-4**.

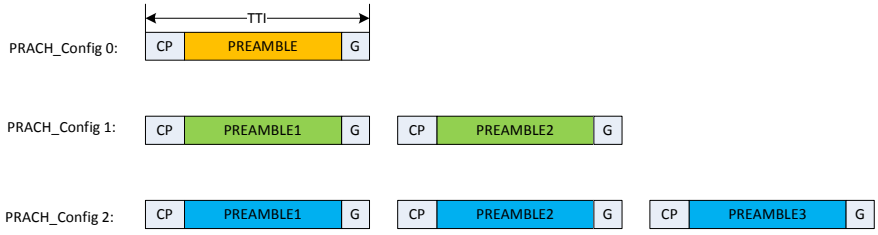


Figure 5-1: Preamble sequences of multiple PRACH_Configs.

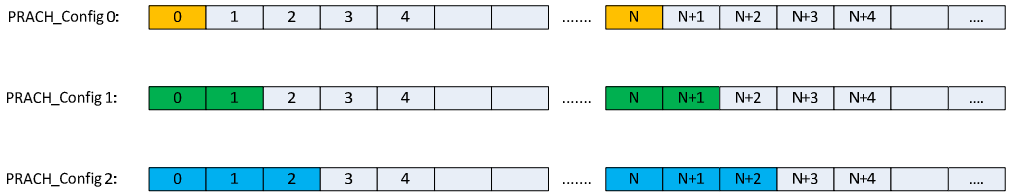


Figure 5-2: Resource allocation of multiple PRACH_Configs: frequency domain multiplexing.

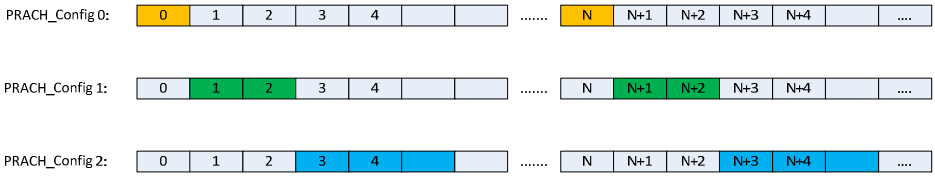


Figure 5-3: Resource allocation of multiple PRACH_Configs: time domain multiplexing.

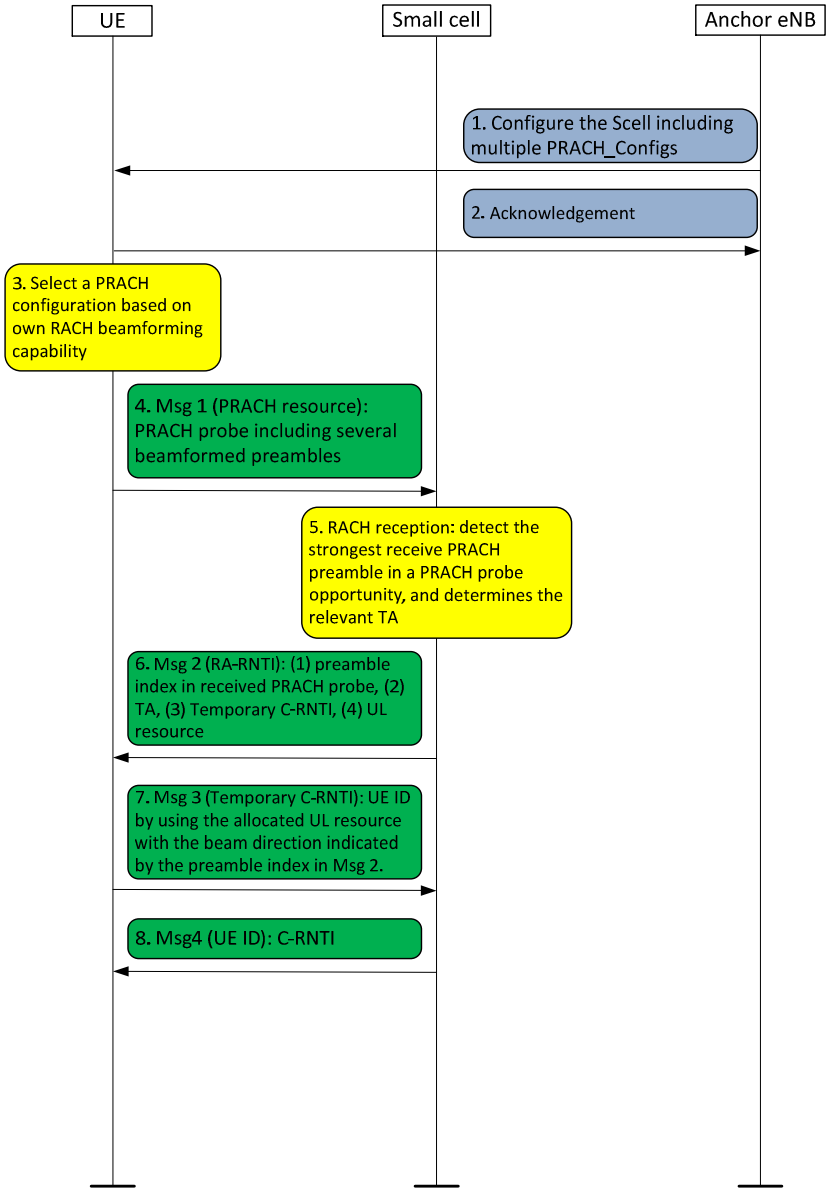


Figure 5-4: Directional PRACH procedure to support uplink beam alignment.

As shown in **Figure 5-4**, the anchor eNB first configures a mmW small cell as the SCell to the UE, and the relevant configuration message includes the supported PRACH configurations, e.g., 3 PRACH_Configs shown in **Figure 5-1**, **Figure 5-2**and **Figure 5-3**, by the small cell. The detailed PRACH configuration contains the information described above. Then the UE acknowledges the configuration.

Assuming that the UE can transmit PRACH using three directions as shown in **Figure 5-5**, each covering an uplink sector of 120 degree, the UE selects the "PRACH_Config 2" in **Figure 5-1**, and starts the directional PRACH procedure. The UE transmits a PRACH probe, including three PRACH preambles, using the allocated time-frequency resources for the "PRACH_Config 2". For each probe opportunity of a supported PRACH_Config, the mmW small cell RACH receiver tries to detect all the preamble sequences of the PRACH probe opportunity, and if any preamble sequence of a multi-preamble probe is detected, the small cell node responds to the UE with the "Msg 2". This is the PRACH response message, in which the strongest received preamble index in the PRACH probe opportunity is signalled in addition to the timing advance (TA), temporary C-RNTI and uplink (UL) resources for the following UL message transmission. As the example shown in **Figure 5-5**, the PRACH response message includes the index of the "PRACH preamble 2" to indicate the preferred uplink beam direction for UE's subsequent data transmission.

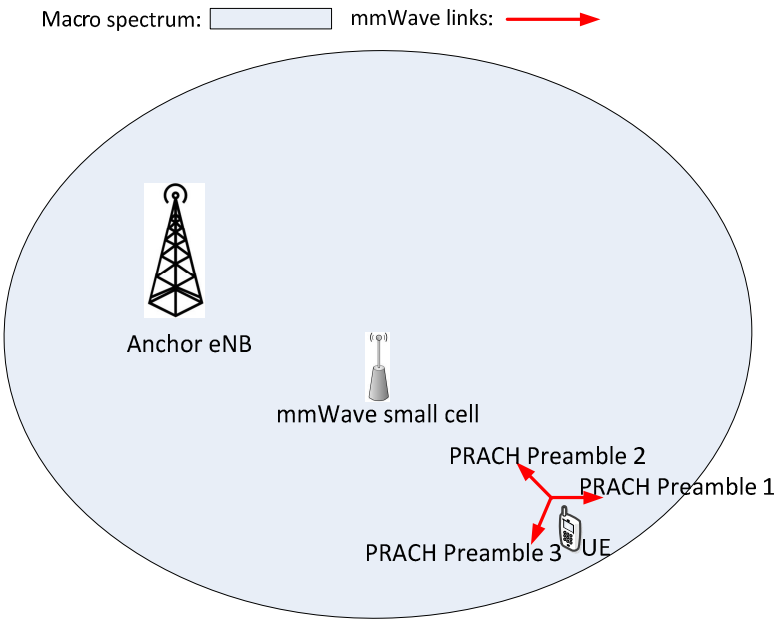


Figure 5-5: Directional PRACH transmission.

After correctly receiving the PRACH response message, the UE transmits the "Msg 3" to the small cell node by using the preferred beam direction, i.e., the beam direction corresponding to the PRACH preamble 2, and allocated uplink resource signalled in the "Msg 2". The "Msg 3" includes the UE identity so as to enable the small cell to resolve the possible UE collision. Upon correctly receiving the UE ID in the "Msg 3", the small cell node sends back the "Msg 4" including the C-RNTI for the UE to identify its subsequent uplink data packet, to the UE. The "Msg 4" is addressed by the UE ID signalled by the "Msg 3". After completing the procedure depicted in **Figure 5-4**, the directional UE PRACH is finished. As a result, the UE achieves both the timing and spatial synchronization with respect to the target mmW SCell. It is envisioned that small cell may also need a beamformer for the RACH receiver to improve the RACH sensitivity, as such detailed RACH receive beamforming algorithm can be an interesting topic for future study as well.

5.2 Small-cell traffic offloading

Usually the networks are configured to serve moving users' needs and allocate them on the best cell regarding the available information gathered. These moving users shall be kept satisfied and, as far as possible, guarantee their requirements of QoS and coverage regardless of the path followed.

The envisaged deployment of small mmW cells on top of the 4G macro and also 4G small cells provides as output a complex grid of BSs covering dense areas. It is critical to know the best cell to allocate each user to at each point, as in some situations there could be several cells within the coverage area.

The data currently used to perform this allocation of users consists of only the user power received from its attached cell and the subscriber info (SLA, bearer class, etc.). At user side, there is a plethora of very valuable information available that may help enhancing the way users are allocated in cells. The proposed solution includes both the definition of a control plane and IP level middleware able to transfer this information to the network and a simulation tool has been designed to compare theoretically the potential gains from QoS perspective.

The main considered scenario layout considered is shown in **Figure 5-6**. A grid of 4G macro cells is used as basis for the coverage, and then a hot-spot area is dealt with by the inclusion of mmW small cells, with reduced area and enhanced coverage.

Using this kind of mmW small cells will help cutting costs. In addition, it also helps improving the QoS at user level, making sure the connection requirements in high-density areas are always fulfilled. Finally, the coverage of the overall network is also enhanced.

The idea behind the middleware consists in developing a system able to detect the subset of reachable BSs for each user given their location. This information, combined with the new parameters introduced by communications via the control plane, enables a more efficient handover control.

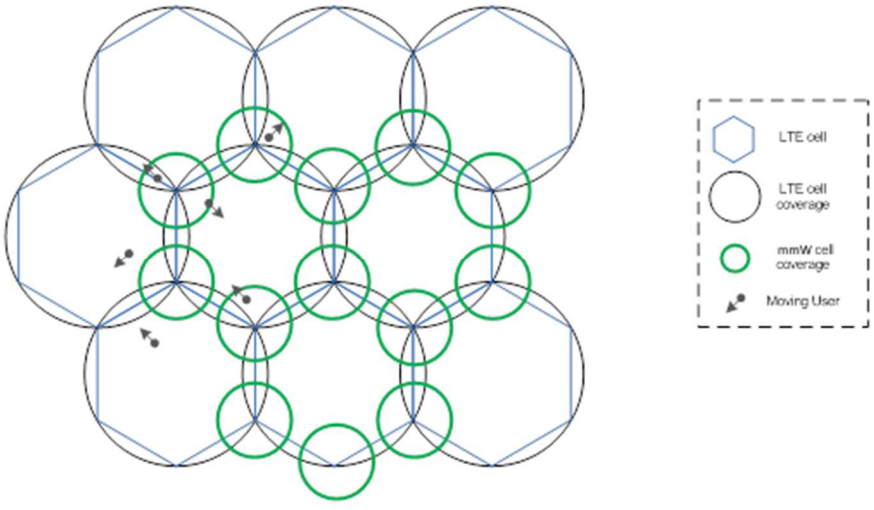


Figure 5-6: mmW Small cell load balancing layout.

Establishing such a communication as introduced by this envisaged middleware requires a continuous exchange of data between upper and lower layers both locally and remotely (offering different Service Access Points, SAP), as well as at a global network scale. The high-level interactions between middleware entities can be seen in **Figure 5-7**. Lower entities represent transceivers, users and/or BSs. Upper entities are managed at Core Network level.

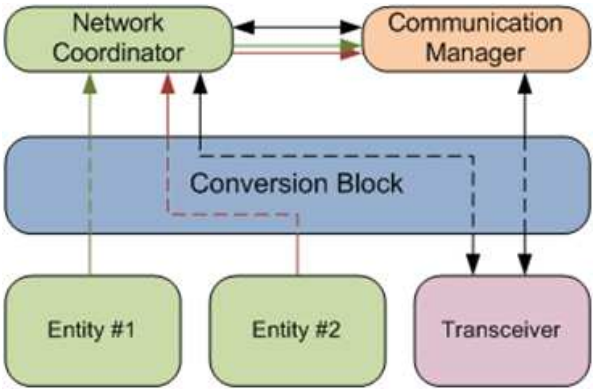


Figure 5-7: Basic block diagram of the traffic offloading middleware.

The middleware intelligence will be embedded in the entity depicted in the middle blue box, and it is in charge of receiving all information from the different external entities and producing and delivering associated primitives to the proper destination, avoiding if possible replicated paths. It will be based on a publisher/subscriber architecture, so it will be possible to support event subscription functionalities for upper layers and remote functional blocks at both local and remote levels.

The use of this middleware will allow technology agnostic information management. Some fine tuning of critical parameters and measurements will be enabled in order to properly characterize the envisaged RATs to be used. A common data model will be designed so as to define and accommodate the received data.

All in all, the Conversion Block will imply a novel and an efficient method to perform inter and intra-RAT duties, as required by the mobility conditions so typical nowadays. Its internal structure will be similar to the one shown in **Figure 5-8**, distinguishing two main parts: the ACCESS section, directly connected to the desired entity, and the CORE section, where a dedicated database is deployed and the information processed and conveniently dispatched.

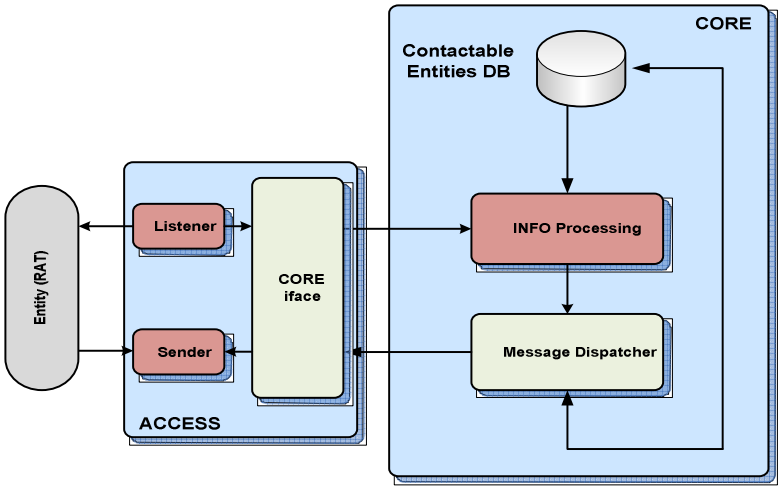


Figure 5-8: Conversion Block structure.

5.2.1 Common data model definition

A Common Data Model will be detailed, which enables the exchange of information among local and/or remote entities through three main types of primitives:

- Events: used to indicate dynamic changes (spectrum, link, quality).
- Configuration and Commands: physical configuration, resource allocation, mobility

- Exchange Information: framework to exchange and discover info about opportunistic spectrum usage, spectrum portfolio, QoS features in communications, etc.

Within this context, up to three types of communication modes are considered:

- Request, in order to handle petitions
- Acknowledgement, to notify an entity its message has been properly received
- Indication, to provide information considered as important

The common data model will be based on custom primitives built using IEEE802.21 standard as reference. The fields used by those primitives are described in the following **Table 5-1**.

Table 5-1: Middleware primitives.

Primitive	Semantics
ENTITY_REG	Registers an entity at the offloading middleware. To be registered an entity will need to provide certain info which will identify it and thus will be able to communicate.
GET_INFO	Request information on other entities present in the system
INFO_DETAILS / INFO	Specifies the information required by the entity which initiated the request
EVENT_SUBSCRIPTION	Remote Event Subscription request managed by CB
EVENT	Generated due to an event occurrence. Some parameter's value has crossed a threshold previously established.
COMMIT	Once received some information on neighbouring access points, a User Terminal decides the new Base Station to be connected to and sends a primitive to its current BS. This will manage a new connection process.

The interaction among the diverse entities and this middleware can be roughly depicted in the Message Sequence Chart (MSC) presented in **Figure 5-9**, where a MiWaveS entity (namely, a user equipment) initiates a registration process, in order to make itself known within the system. At this moment, the middleware acknowledges its presence, proceeds to include it in its own database and to confirm the entity this event. From this moment on, the middleware is capable of sending information to the entity when necessary or interesting.

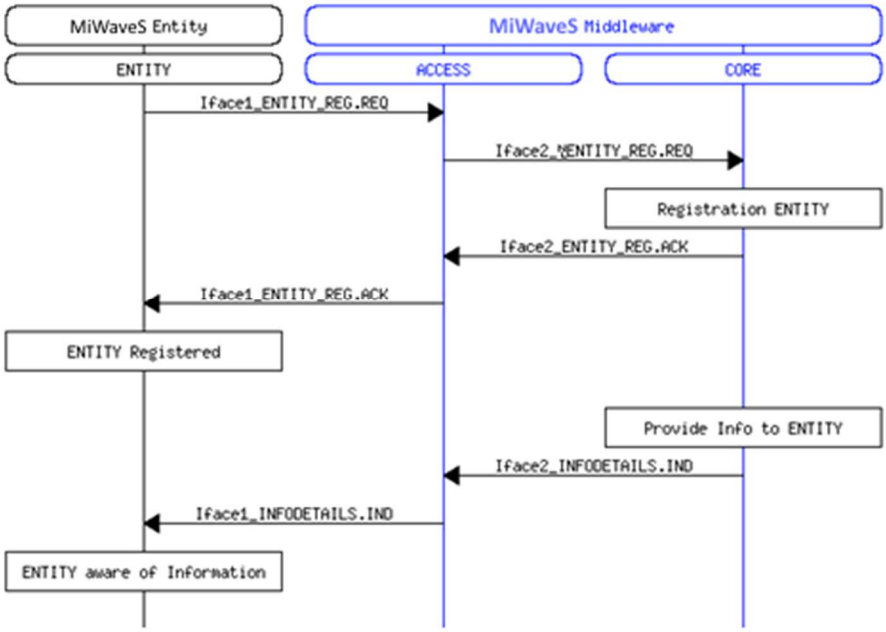


Figure 5-9: Interactions between Entities and Middleware MSC.

5.2.2 Simulator parameters

Once defined the middleware basis, the info being transferred at both ends of the communication and the required interactions between the network elements, it is possible to start programming the simulator in charge of the proof of concept.

The idea consists of a two-stage approach:

- First of all, a software tool has been developed to create the scenario of use. The number of macro BSs can be specified, as well as the main characteristics of their radio parameters. In case the Inter-RAT offloading is considered, some extra access points using different technologies can be also deployed. The resultant output will be a coverage matrix determining, at each point of the scenario, the received signals from each nearby BS.
- With this coverage matrix obtained, a second simulating software tool has been developed so as to create a 24-hour set of users moving randomly among the scenario created previously. Assuming the network knows all user-related data, it will allocate users on different BS depending on the user needs and BS occupancy.

The intelligence programmed is based on the premise that the network is aware of some user parameters. These parameters are transferred from the users to the network by using the primitives depicted at Section 5.2.1, using the control plane. In case the network needs to relocate or notify some user, it will proceed the same way, using the proper primitive via the control plane.

The data known by the network at each moment is summarized on the following **Table 5-2** and **Table 5-3**, taking into account the infrastructure (data from all BS deployed in the area) and the user info (UE characteristics and user profile).

Table 5-2: Base Station parameters.

Base Station Parameters	
Max_Load	Maximum Load in Mbps allowed by the cell
X_position	X position in the layout (m)
Y_position	Y position in the layout (m)
Coverage_radius	Coverage radius of the BS (m)
Max_users	Maximum number of users allowed
Power_cons	Power output of the BS
Type	True if it is LTE or false if it is a small cell
Name	Unique Identifier of the BS into the layout
CurrentUsers	Current users supported by the BS
CurrentLoad	Current load handled by the BS
CurrentLoadGBR	Current load of just GBR users

With these previously mentioned parameters, it is possible to characterize the whole network. But some hierarchy is needed in order to determine which users shall be prioritized and which ones can be served with lower quality than required in a potential situation of lack of resources.

This hierarchy is performed via the use of LTE bearer classes. Thus the following rules are defined:

- Distinction between Guaranteed Bit Rate (GBR) users (the ones which are demanding a certain amount of bandwidth which cannot be neglected) and Non-Guaranteed Bit Rate (Non-GBR) users (low priority ones whose demanded QoS can be diminished if needed).
- Distribution of users depending on the type of data demanded to the network, differentiating between (in decreasing order of bandwidth demand):

- Gaming
- Streaming
- Video call
- Mail
- Voice

Combining these two hierarchies and assigning each of the resultant types of users an estimated occurrence percentage and an estimated bandwidth demand, the following **Table 5-4** can be proposed to serve as basis to program the network intelligence of the system.

Table 5-3: User information shared with the network.

User Parameters	
RSSI	BS and power received at each instant
Interfaces	True if it supports attachment to small cells and false otherwise
Battery	Charge status of the UE battery
SLA	Mbps required by the user
X_position	X position in the grid (m)
Y_position	Y position in the grid (m)
X_next	Next X position expected from a random mobility model
Y_next	Next Y position expected from a random mobility model
Session	Minutes left of the user session
QoS	QoS experienced at each moment (measured as the mean percentage of SLA per user served by the network with respect to user SLA originally requested)
Selected_EB	BS to which the user is attached in this moment
Priority	Bearer class

Table 5-4: Bearer classes.

Service	BW (Mbps)	Class	Priority	Probability
Voice	0.1	GBR (reserved BW)	1	10%
Videocall	0.5		3	5%
Streaming	1.5		4	5%
Gaming	2		2	2%
Voice	0.1	Non-GBR	9	30%
Videocall	0.5		8	10%
Streaming	1.5		5	15%
Mail	0.5		6	20%
Gaming	2		7	3%

With all this information available at network level at each instant, it is possible to allocate each user at the best place, offloading the traffic at dense areas to less crowded BS if possible. The results, at each minute of the 24-hour simulation period, are gathered in the following format (see **Table 5-5**).

Table 5-5: Results gathered.

Results Parameters	
N_Users	Total number of users on the network at each moment
Load	Total amount of data served by the network at each moment (Mbps)
N_Users_Bad_SLA	Number of users with QoS < 100%
QoS	Average QoS of all users in the network
Consumption	Power consumption of the network
QoS_GBR	Average QoS of GBR users in the network

5.2.3 Simulation assumptions and strategies

This algorithm assumes the network is available to relocate users at any time using a typical handover mechanism. This handover is currently activated in case a user is moving from one BS to another BS. The network is the entity in charge in this case to trigger this handover in case it detects that the user is losing coverage from one cell and there is another one more suitable in terms of signal strength.

The novelty introduced by this algorithm is the extra information being sent by the user terminal. This new information includes the number of cells detected by the user (including all connection technologies supported by the terminal) and the QoS requested both by user contract and by the application currently running in the device. This way, the network will be able to relocate users not only due to changes on their physical location but also due to service requirements.

Taking into account the previously mentioned bearer classes, the following list contains the extra and potential situations where a given user might be forced to handover to other cells as a result of applying the proposed algorithm:

- A non-GBR user is in a cell which is crowded and a GBR user appears. In case the former non-GBR users is able to connect to a different BS, a handover will be requested.
- A GBR user changes between a high demanding application (i.e. video streaming) and a low demanding one (i.e. voice call). If the original location was a cell from which the signal received is strong but it is crowded and there is another cell with less users within the range, a handover will be requested. It also applies the other way around, changing from less demanding application to high demanding application. If a closer BS is detected, a handover will be requested to enhance the QoS.
- The previous behaviour is also applicable to non-GBR users but just in case the BSs are not crowded and they have still room for more users.

These new considerations are expected to bring QoS enhancements to the overall network. So as to test them, three main strategies will be studied to compare the network performance with no pre-known data available, using just some initial information and finally including the full set of parameters aforementioned to allocate users at the best access point.

It should be noted that this algorithm could be used to maximize any of the result parameters. In this case, the goal is to enhance the QoS of the overall network.

The three strategies used are the following ones:

- **Simple**: Users are attached to the BS from where more power is received.

- **Intermediate**: If the BS with more power is crowded, it looks for another one with less number of users.
- **Complex**: Uses bearer classes to allocate users:
    - Users with GBR are always allocated in the best BS. If they demand high bit rates, in the most powerful one. If they require voice/mail services, in the less crowded one.
    - Users with non-GBR may be forced to handover if a GBR user comes.
    - All users look for the best choice of BS.

The general layout for this simulation considers a grid of 5x5 macro BS able to handle 250 users and 120 Mbps each. The BS coverage radius is 50 metres and the Inter Site Distance (ISD) $\sqrt{3}\cdot 50$ metres. If Inter-RAT offloading is required, extra mmW access points can be also considered, up to 50 users and 500 Mbps with 10 m coverage radius.

Two main simulation rounds were run using this scenario, and differentiating in the coverage considerations:

- In a conservative first approach (**Figure 5-10** and **Figure 5-11**), the 50- and 10-metres coverages are taken as a limiting factor. Further than those limits, it is considered that the user cannot listen to the BS.
- In a more realistic approach (**Figure 5-12** and **Figure 5-13**), this coverage radius is extended by a 1.5 factor. This way, the user is able to listen to a BS even though it is not close to it. Then the network has more flexibility, as the user can be allocated in more BS at each point. It is important to note that this scenario does not break the frequency reuse planning used in LTE. When using this approach, the user terminal battery must also be considered, as if a user is forced to handover to a distant cell it shall transmit with higher power and the battery might not be able to handle it for a long period of time. If a user battery is below 40%, the user cannot be forced to handover to a distant cell.
- This is a layout created to handle the maximum number of users, but with slightly less capacity than required at peak time, to force QoS < 100%. This forces the situation in which, at peak hour (in the afternoon), the available resources in terms of Mbps are not enough to cover the user demand. It would be important to see the evolution of the QoS of users depending on the selected strategy.
- Users appear and move randomly. The number of users is limited by the voice and data traffic patterns obtained from the FP7 EARTH project [17] models. At each moment, a user has equal probability to stay still or move randomly 1 metre on each direction (as can be seen in **Figure 5-14**).
- The duration of the user session is random, lasting between 1 and 25 minutes. The rest of user-related parameters such as priority and bearer classes are determined by the occurrence percentage presented at bearer classes table.
- Users have 20% more probability to be placed at the region where small cells are deployed, in case they are placed on the layout. This way it is possible to simulate hot zones (as depicted in **Figure 5-15**).
- 30% of BW is reserved for GBR users in complex strategy. This way, it is possible to guarantee that they will be always served with the required bandwidth.

- The mmW access points are considered in this example to be located in cell edges. They could be placed anywhere in the layout. They are located this way so as to cover the hot zone of users. The assumption of mmW placed inside macro cells and not on their edges is equivalent to the algorithm, as they are just influencing the number of cells received at each location in the layout. It is just important to consider deploying inside the hot zone as many small cells as needed so as to have at each point at least the choice to connect to the macro or the small cell.

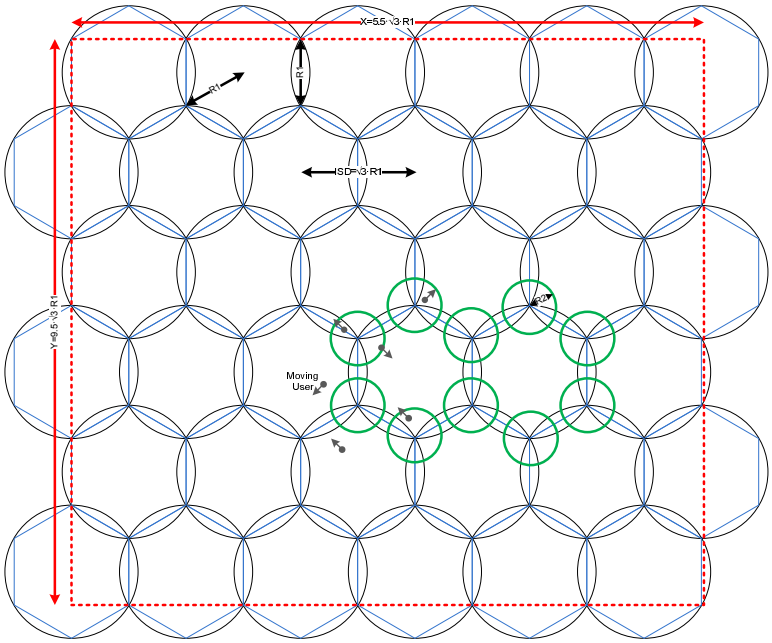


Figure 5-10: Conservative layout.

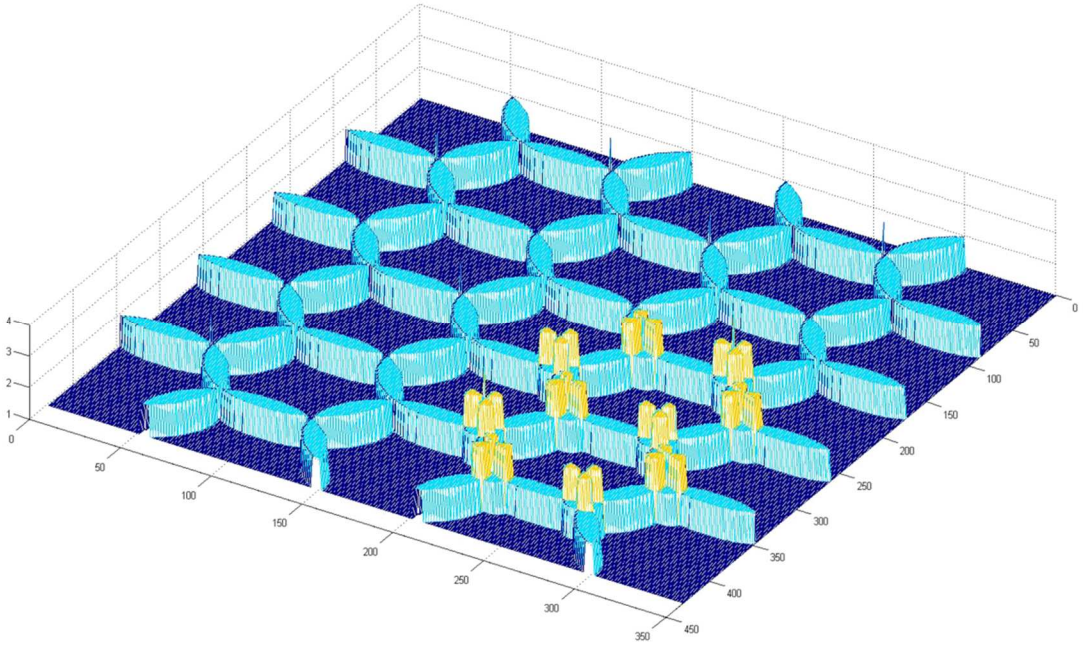


Figure 5-11: Number of BS seen at each point with a conservative approach.

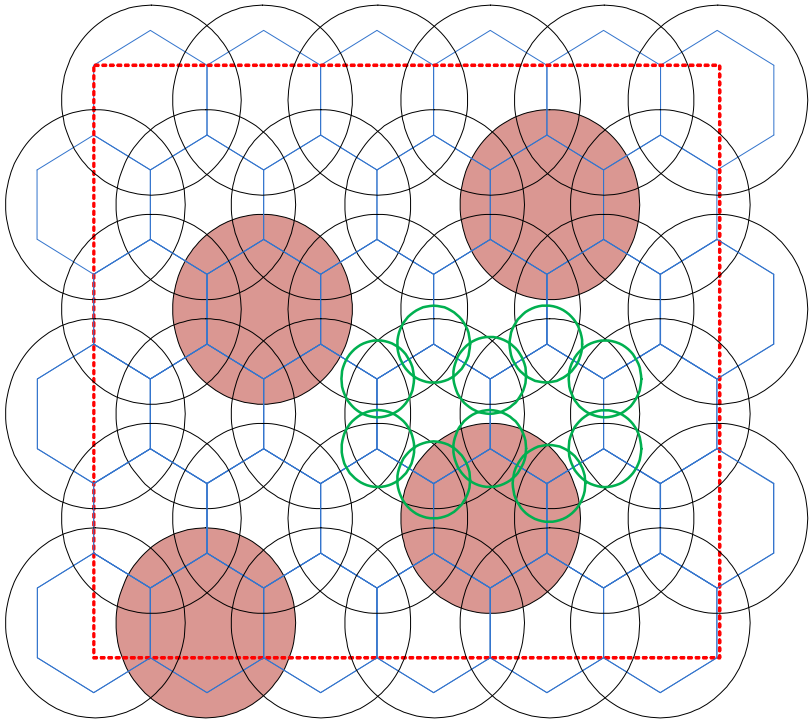


Figure 5-12: Realistic approach (1.5 coverage radius extended).

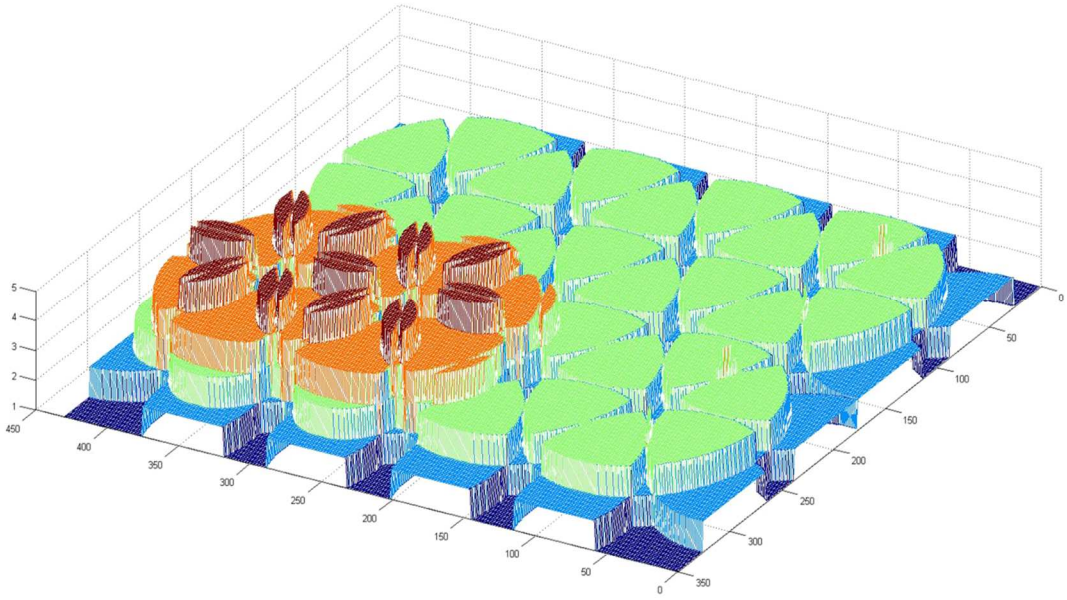


Figure 5-13: Number of BS seen at each point with a realistic approach.

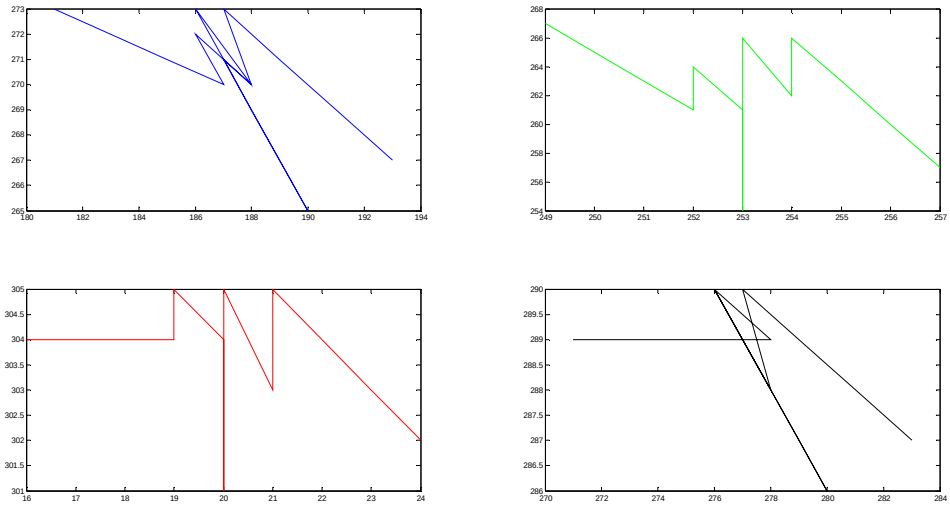


Figure 5-14: Examples of user movement patterns (axis in metres).

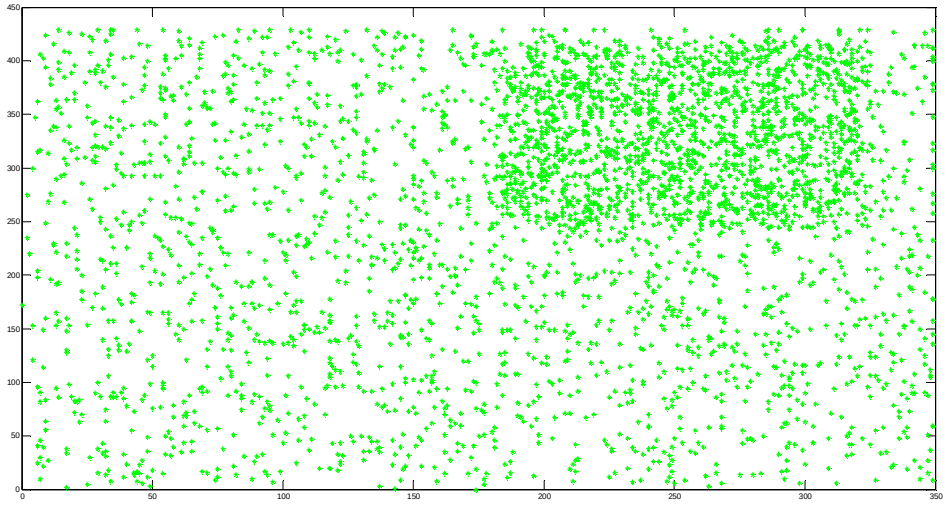


Figure 5-15: User position at peak hour (axis in metres).

5.2.4 Simulation results

In this subsection, the results gathered due to several 24-hour simulations performed per scenario are presented. They are analysed and explained to throw the final conclusions regarding this traffic offloading mechanism.

The goal of the whole middleware+simulator tool definition is to enhance the resultant QoS of the overall network by knowing in advance some user related parameters not used currently.

The simulator interfaces for both the network planning and the 24-hour period simulator can be seen at **Figure 5-16** and **Figure 5-17**.

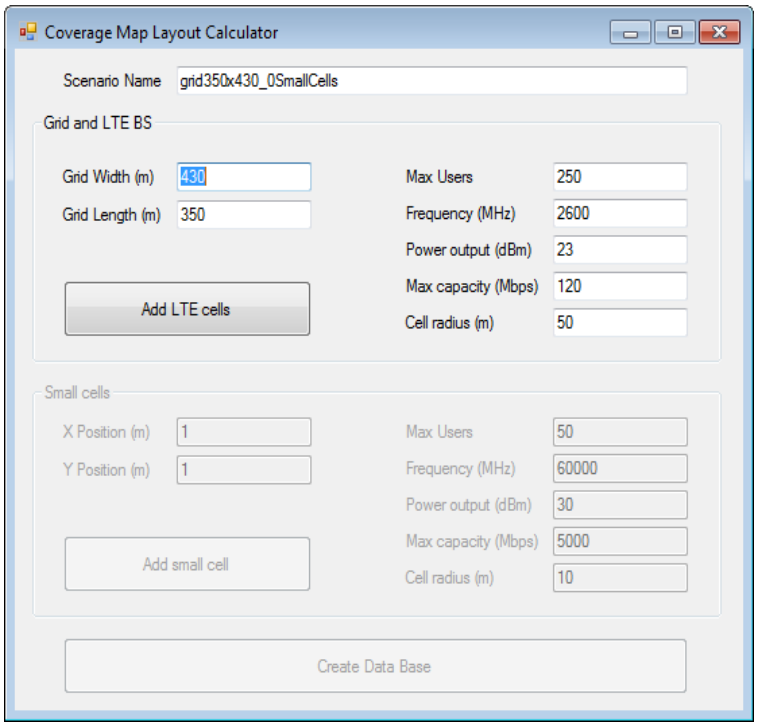


Figure 5-16: Network planning tool.

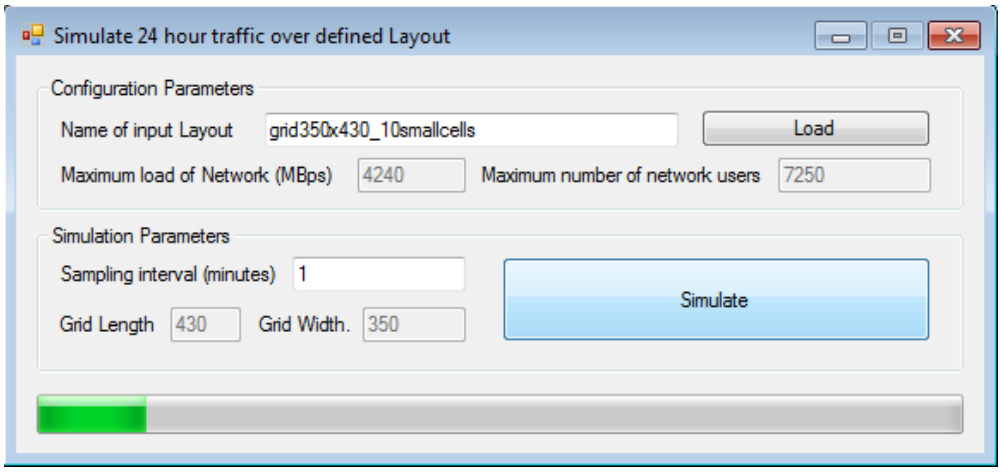


Figure 5-17: 24-hour simulator interface.

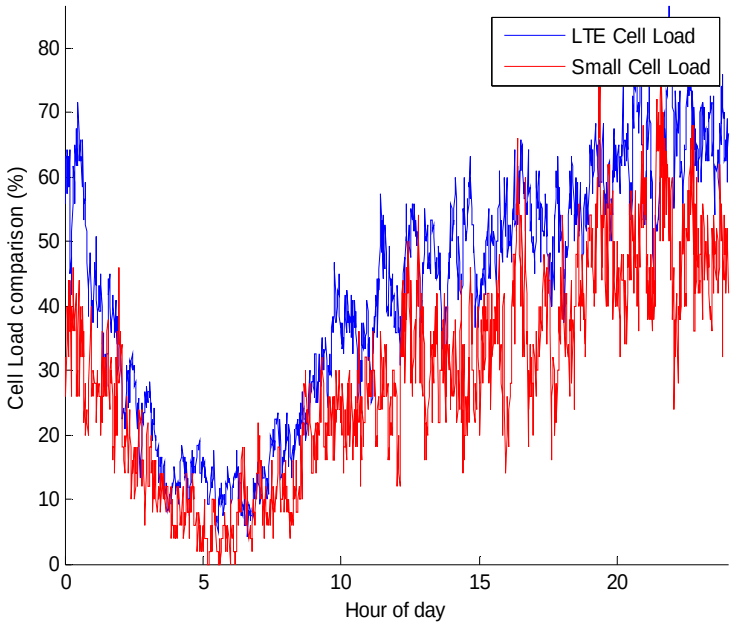
Depending on the policy and layout used, the following results are obtained. In all cases three curves are presented: total amount of users registered at each moment (green curve), traffic load associated to each user (in Mbps, blue curve) and the number of users with QoS below 100% (experiencing throughput under their SLA, red curve. The goal will be reducing this last red curve to the most by allocating as many users as possible in nearby cells fulfilling their needs at each moment.

- Using the policies (intermediate or complex) is always better than using the simple approach. This can be noted on when comparing **Figure 5-19** with **Figure 5-20** or comparing **Figure 5-21** with **Figure 5-22**. The number of users with less than 100% QoS (red curve) is always lower on the latter ones.

- The conservative scenario presents always worst results than the realistic one. This can be also seen by comparing **Figure 5-19** with **Figure 5-21** or **Figure 5-20** with **Figure 5-22**. The first figure of each pair is using the conservative scenario while the second one uses the realistic one. The

first pair compares the simple policy and the second pair uses the complex one. The red curve is always lower on realistic scenario simulations.

- The traffic load balancing between small cells and macro BS is better when using realistic scenarios. This is shown at **Figure 5-18**. On the left hand side, the realistic scenario shows cell occupancy levels very similar on LTE and small cells, while the conservative scenario (right) has most of small cells under used.
- The non-GBR QoS can be maintained above 75% for conservative scenarios and complex strategies (**Figure 5-20**), and above 85% for realistic scenarios and complex strategies (**Figure 5-22**).
- When using the complex policy in conservative scenarios (**Figure 5-22**), 15% peak and 30% mean enhancements can be obtained in terms of overall QoS with respect to simple policy (**Figure 5-21**).
- The complex policy in realistic scenarios (**Figure 5-22**) is able to balance the traffic of the network at each moment except when the traffic exceeds the overall network capacity.



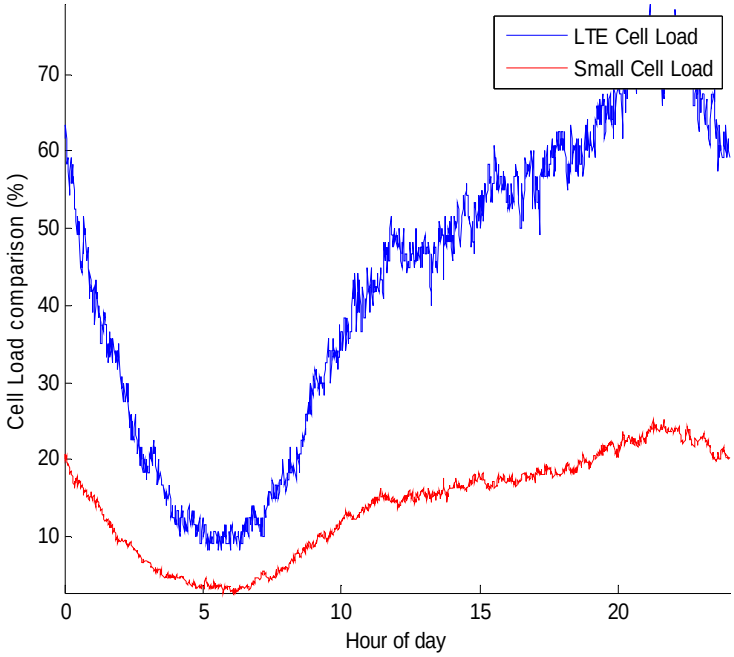


Figure 5-18: Cell load comparison between realistic (the upper graph) and conservative (the lower graph) scenarios.

Given the results presented, it is possible to say that the defined and simulated algorithm is able to enhance the network performance significantly. By deploying the IP-based middleware and interchanging information between users and the network at control plane, it is possible to trigger network centralized decisions to balance the traffic and allocate users at the best BS at each moment.

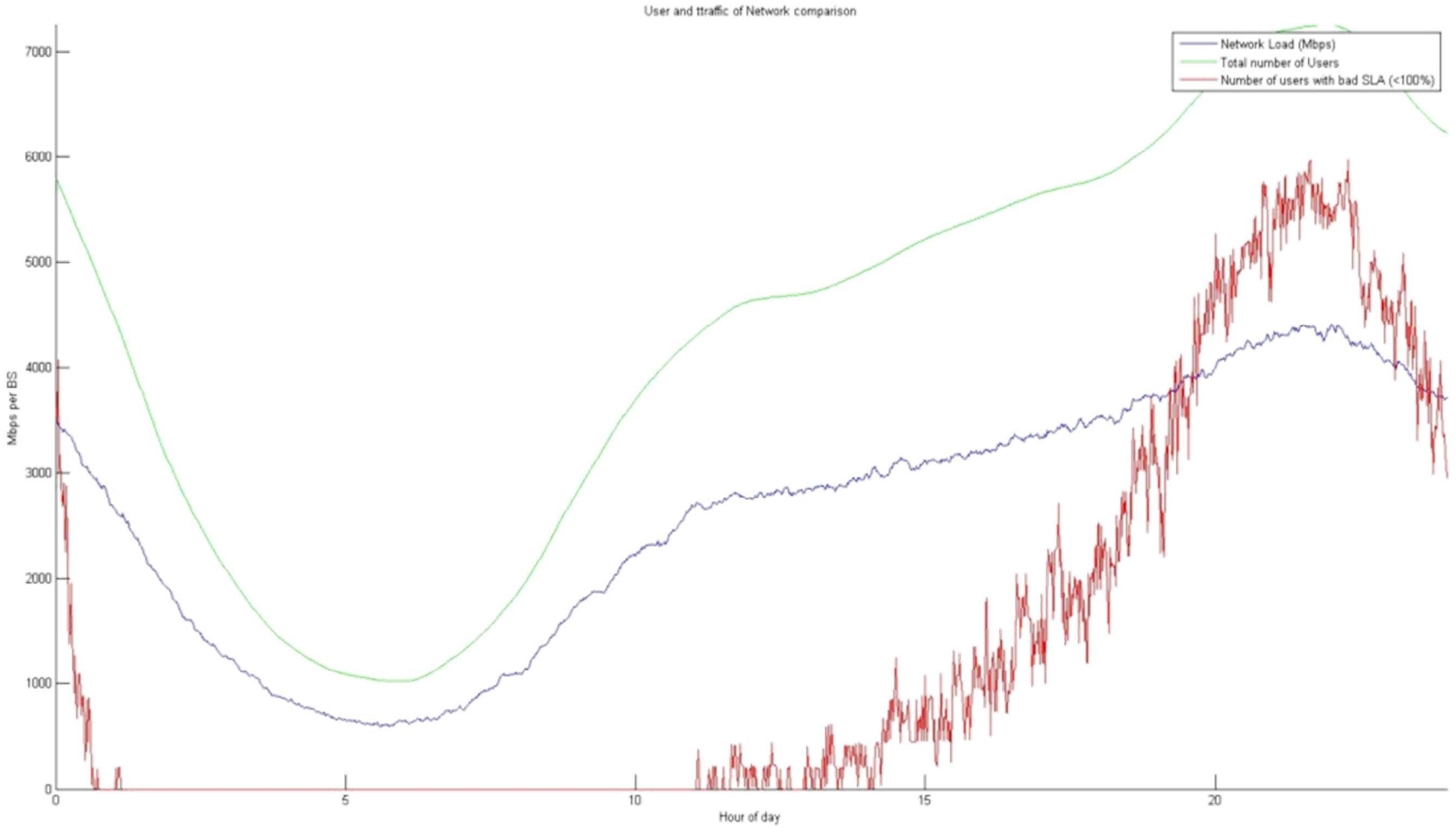


Figure 5-19: QoS (red curve) of the network using simple policy in conservative scenario.

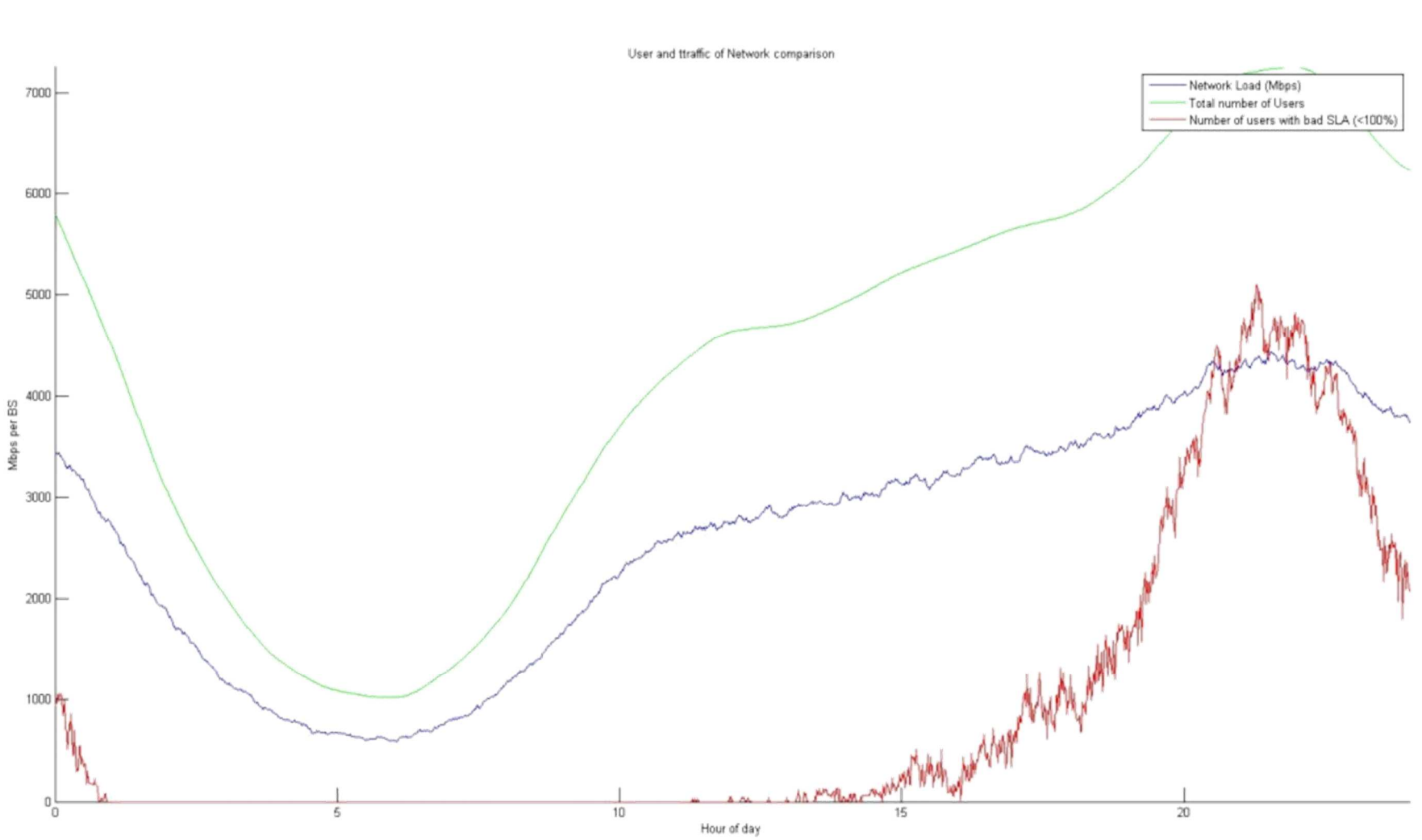


Figure 5-20: QoS (red curve) of the network using complex policy in conservative scenario.

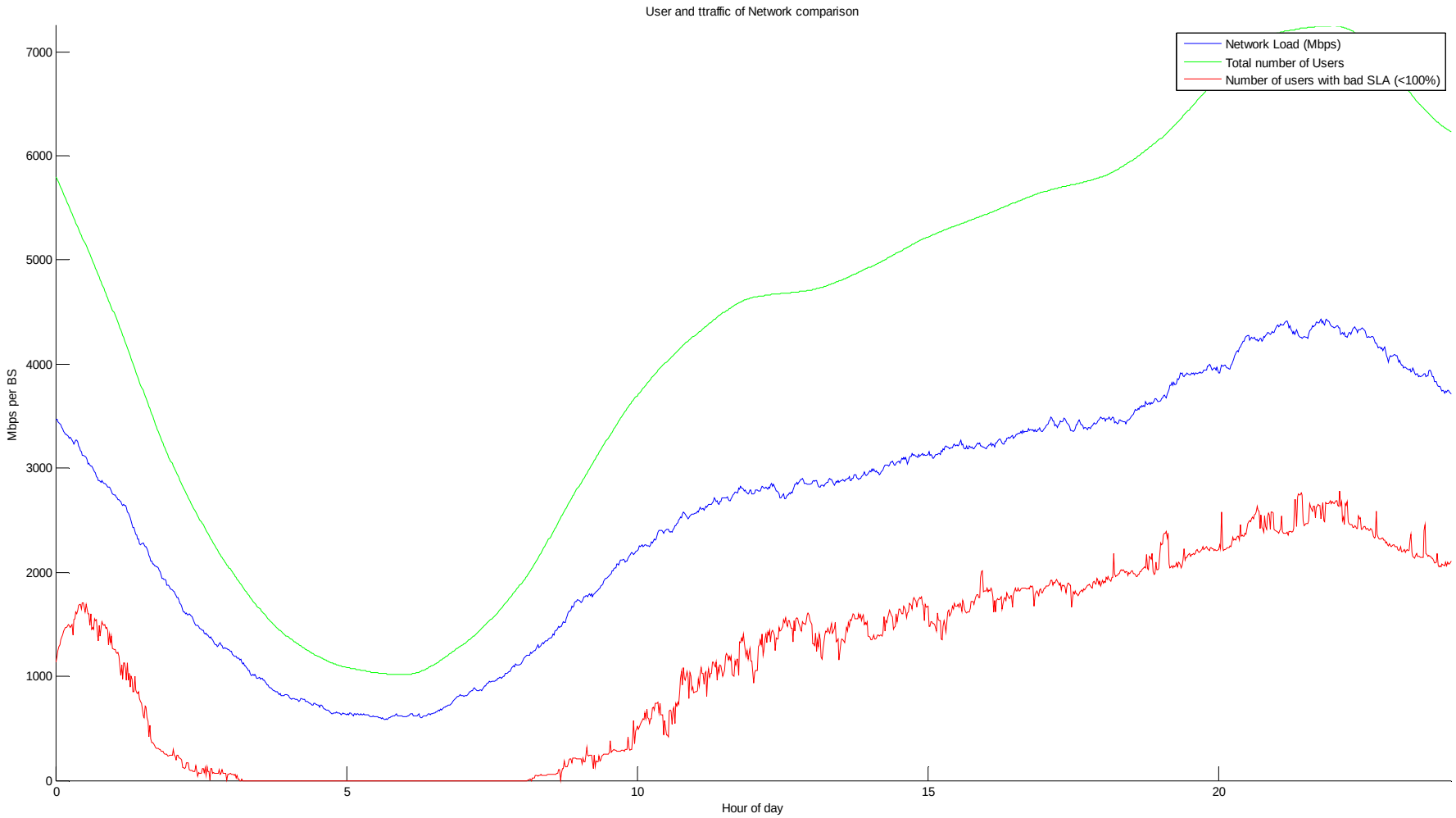


Figure 5-21: QoS (red curve) of the network using simple policy in realistic scenario.

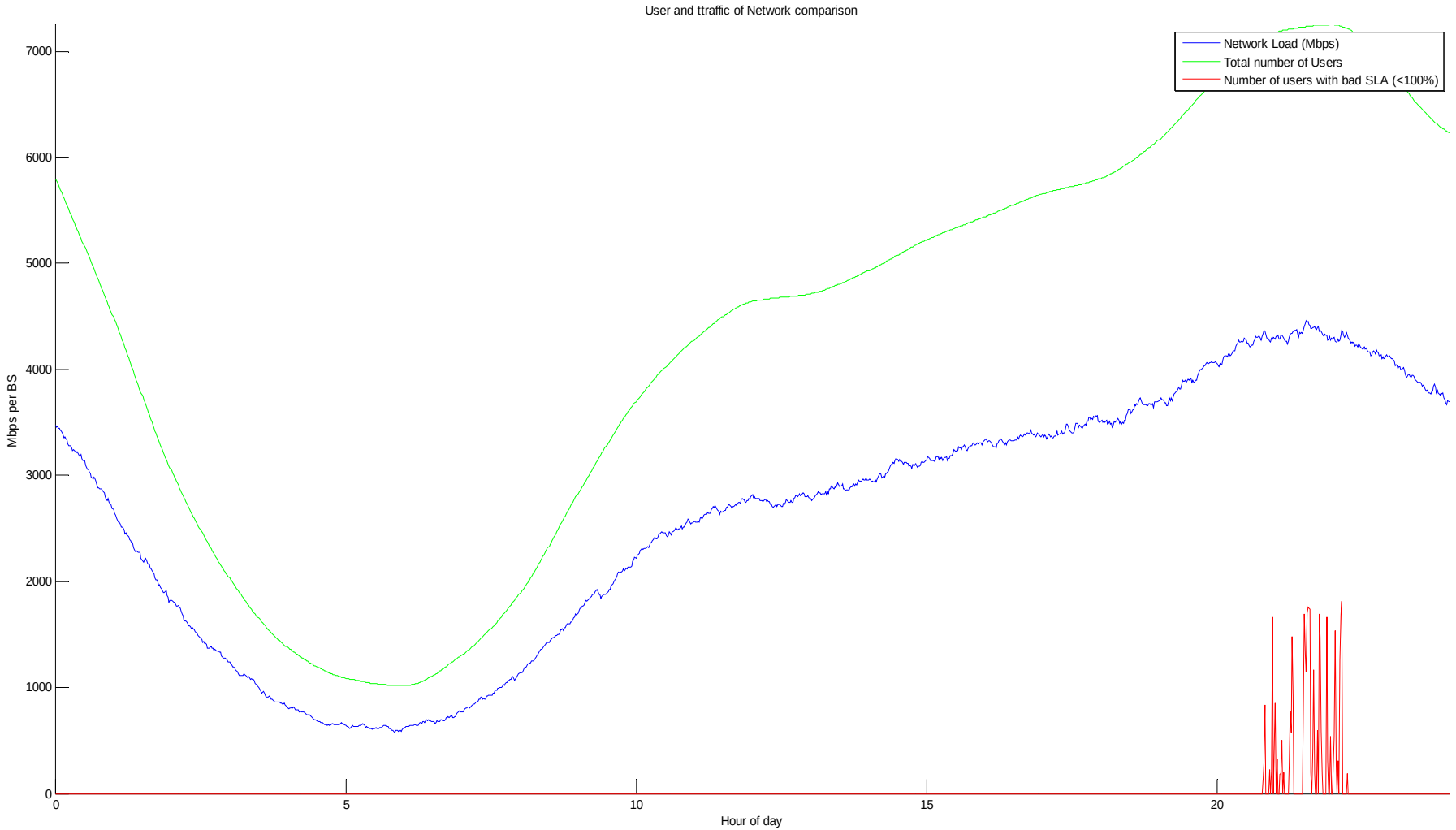


Figure 5-22: QoS (red curve) of the network using complex policy in realistic scenario.

5.3 mmW small-cell energy efficient operation

This section presents two analysis and concepts for evaluating and controlling MiWaveS network energy consumption. First a simplified broadcasting (downlink) scenario is considered. Because data rates

Second, it evaluates the potential power savings related to intelligent switching ON/OFF of a number of BSs in certain time periods when the network load is such that it can be served just with a subset of the BSs. This is possible in dense areas, where lots of BSs are deployed to be able to cover the user traffic demand at peak hours. As mmW small cells are envisaged to be placed inside a macro cell coverage area, it is very important to control them when they are inefficient due to small amount of user demand. In that situation, it is possible to switch off the small cells and leave just the macro cell connected to serve the remaining users.

Multiple management schemes have been studied recently in the literature. One of the most relevant methods is Energy Saving Management (ESM, proposed by ETSI and 3GPP). In addition, improvement proposals can be found such as cell breathing, QoS scheduling or this proposed ON/OFF method. Means and methods to identify the nodes/APs that can be switched OFF (on again ON) has been studied and described in Section 5.4.2.

5.3.1 Energy efficiency in broadcasting

MiWaveS system concept is based on assumption of heterogeneous network and multi-connected UEs. Underlying large-area network operating on lower frequencies provide the ubiquitous mobile connection while the mmW small-cell network with highly directive antennas provide the ultrahigh-capacity data connections. The ultimate data rates are not needed continuously especially when users are moving by foot or car.

Control connections between UEs, APs and macro BSs are established with overlaying wide area network. Control traffic data rate is relatively low, however connection must always exist and be relatively reliable. If overlaying network is used mainly for control traffic and low-rate data services like browsing, short-messages and voice, it can be dimensioned to use lower energy per user than with full service. When a high instantaneous data rate is needed or a large amount of data is to be sent, the control channel activates mmW links on-need basis. This applies to both access links between UE and AP as well as backhaul links between APs/BSs. mmW link activation rules may be different depending on network QoS targets as discussed later in this section.

Table 5-6: Time (in seconds) to send one minute of video with different quality.

Service	Mbits/ 1 minute	Radio capacity		
		100Mbps	2Gbps	10Gbps
YouTube 480p video using H.264, 1 Mbps	60	0.6	0.03	0.01
HD-video 720p using MPEG2 compression, 19 Mbps	1140	11.4	0.57	0.11
Ultra-HD 2160p (4K) using HEVC, 35 Mbps	2100	21.0	1.05	0.21
Ultra-HD 4320p (8K) using H.265/MPEG-4, 90 Mbps	5400	54.0	2.70	0.54

Table 5-6 shows how many seconds it takes to send one minute of video with different quality. Using these figures, we can make an energy saving scenario for video delivery type of case. This scenario assumes that data transfer is (almost) one-way and can be buffered smartly at differed levels of network.

Assuming a simplified network topology in **Figure 5-23**, mmW backhaul network delivers traffic from AP's to UE's via a hierarchical tree-topology. Over longer time in MiWaveS concept, the tree-structure of a mesh backhaul network can change when routings change for example due to load-balancing. This makes BH traffic more symmetric between nodes and thus the energy consumption of BH/AP nodes more equal but the calculation ere is still valid. In our example there are 10 users (UE) per one AP with maximum data rate of 2 Gbps watching continuously HD-video or consuming similar amount of non-real-time data. To calculate how long time transmitter and receiver must be active, we assume similar additional overhead time due to TDD receive period, beam steering and re-transmissions activities, which leads to multiplier 2. So, the tranceiver is on (10 x 19 Mbps x2) / 2Gbps = 0.19 of the time.

Conclusion is that individual mmW AP and UE radios can be turned off i.e. stay in power-saving mode about 80% of time. In case of broadcasting or downloading the UE's mmW transceiver operates mainly in receiving mode that consumes less energy than transmitter and thus consumes even less energy and saves battery.

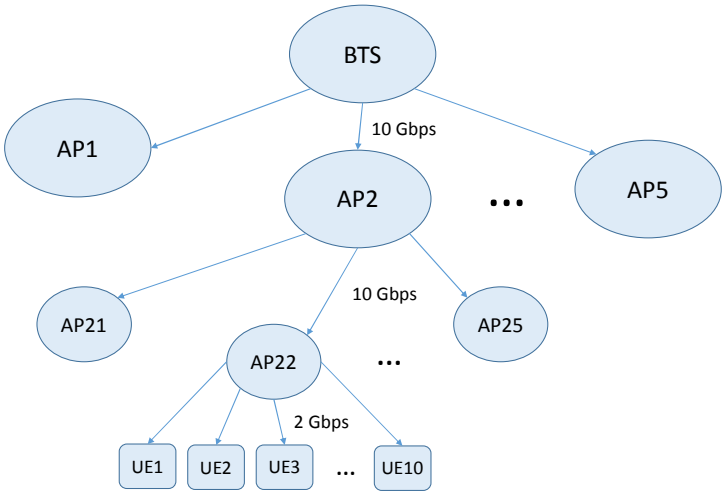


Figure 5-23: A simplified hierarchical topology to calculate duty cycles in downlink.

We assume a two-layer AP hierarchy or two-hop BH network (APmn - APm - BS) with span of five nodes, i.e. each BH node have five sub-stations. In this conceptual study, a backhaul link can be either an integrated inband or a dedicated link. The next level of BH nodes (APm) distributes traffic to five APs at the speed of 10 Gbps and the duty cycle is (10 x 19 Mbps x 5 x 1.5) / 10 Gbps = 0.14.

The transmission takes only 14% of the total time while the rest 86% can be spent in power saving mode. Because BH nodes are more stable the overhead energy consumption time can be assumed to be only 50% on top of traffic, hence the multiplier is 1.5. The next, i.e. the highest, level of BH aggregates all the traffic and is active (10 x 19 Mbps x 5 x 1.5 x 5 x 1.5) / 10 Gbps = 1.07.

Hence, the radio is continuously transmitting and it is even slightly overloaded.

This example indicates that significant energy savings are possible with mmW small-cell topology especially in video delivery type of cases.

When offloading high-volume data traffic to mmW small-cell network, energy saving can be achieved especially in non-realtime or non-conversational traffic. To achieve savings we assume that data can be buffered in BH and AP nodes as well as in UE. The further assumption is that radio circuits, like amplifiers and modems can be turned off or at least put to power saving mode. So, mmW parts of UE, AP or BH radio are continuously operating, but should be able to hibernate as much as possible. This has an effect for example on beam steering. Also the radio switching-on and beam-search times must be short, otherwise the overhead can consume a relatively large amount of energy compared to actual communications that may be only a short burst.

5.3.2 Energy saving with ON/OFF method

In case small cells are considered, the network management must be done in a centralized manner. It is important to note that the clusters to be switched ON or OFF are fixed, as just the small cells inside a macro can be taken into account. Each macro BS should be related to the inside mmW small cells and their total load should be checked with respect to a customized threshold so as to determine if the traffic is such that it can be covered just with the macro BS.

As this concept is studied for MiWaveS deployment, the related layouts can also be considered for conventional HetNet scenarios, and even just 4G macro BSs. In those cases, the assumption is that, as all cells are initially equal in size and coverage radius, when applying the ON/OFF scheme, the cell that is still connected should increase its transmitted power to be able to cover the whole area previously covered by switched OFF cells. It is assumed that this increase in power is negligible with respect to the total BS energy consumption.

On the other hand, if just 4G macro-BSs are considered, as all cells are equal in size and coverage area, apart from the centralized method, it is possible to introduce as well the decentralized management. In that case, the decision is not made from a core network perspective but individually at each cell level. That implies each cell is aware of the current traffic load of neighbour cells. If one cell detects that the cluster in which it is the centre cell has loads below the predefined threshold, it can directly trigger the ON/OFF switching. This way, the way clusters are powered ON and OFF varies in time and the efficiency of the power savings is increased.

It is important to stress that real world deployments have certain layout restrictions due to cell geometry and deployment density and schemes forcing that just some ON/OFF layouts can be applied. For instance, hexagonal omnidirectional or tri-sectorial cells allow for 1/7 or 1/9 schemes (noting with X/Y the layout in which just X of the total Y cells remain switched ON), but Manhattan cell deployment can consider 1/2 or 1/3 layouts. The layouts considered in MiWaveS are 1/4, 1/7, 1/9 and 1/19 for hexagonal cells and 1/2 and 1/3 for Manhattan (as depicted on **Figure 5-24** and **Figure 5-25**). It is possible to use just one scheme per simulation or combining several together in a two-step energy saving procedure.

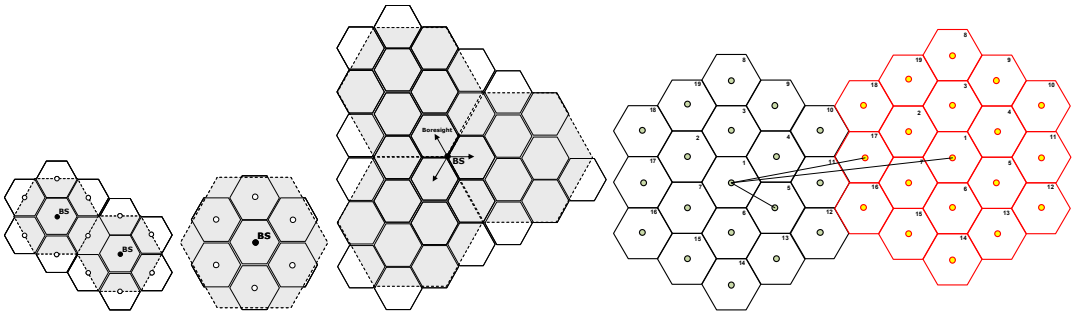


Figure 5-24: Layouts for hexagonal cells (from left to right 1/4, 1/7, 1/9 and 1/19).

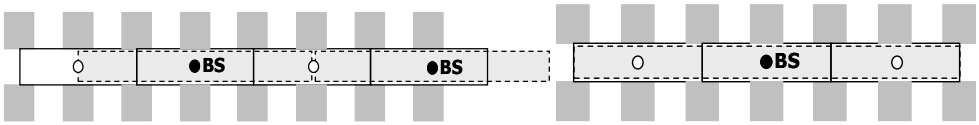


Figure 5-25: Layouts for Manhattan deployments, 1/2 (left) and 1/3 (right).

Once defined the layouts to apply depending on the deployment considered, the next logical step is establishing the related thresholds to trigger the ON/OFF mechanisms, depending on the traffic load detected. When fixing those thresholds, it is important to consider the potential situation in which, as the traffic load in real networks is very bursty (peaks and valleys may occur very fast), the systems may remain for some time triggering ON and OFF orders when the traffic load is near the threshold value. In order to avoid such situations, the first consideration introduced is forcing thresholds on 80% of the theoretical value. That is, for instance, for 1/4 schemes, instead of switching OFF the sector when the load goes under 25% of peak traffic, triggering the order just when it is under 0.8·0.25=20% of total traffic. In addition, linear predictors are considered to foresee the future behaviour of traffic. These predictors are launched every time the threshold is overcome, to provide information about the probability of a potential network traffic increase in the following 30 minutes. This way, it is possible to avoid inefficient order that makes the network unstable.

Finally, different traffic models are considered to evaluate the potential savings. The first approach was to model the traffic using theoretical data:

- Linear approximation. It consists on a curve with flat maximum, flat minimum and a transition linear connection with slope "d". Thus, the amount of time during which the BSs could be switched off will be comprised between τ_E and τ_M where $L(\tau_E)=L(\tau_M)= a$ (point where a BS can absorb the traffic, i.e. load (L), from switched off ones) with $\tau_E < \tau_M$. It considers just a peak period during daytime and a valley slot at night.

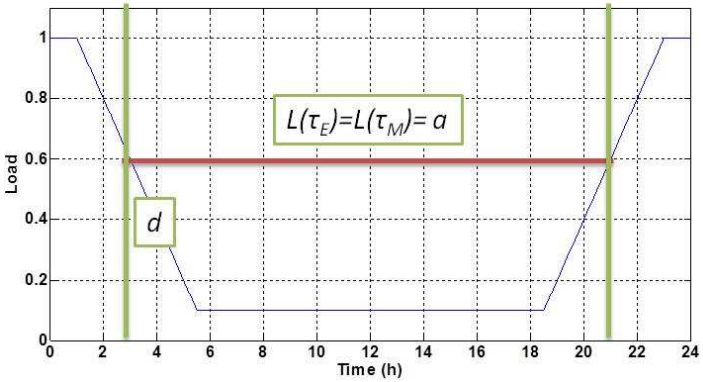


Figure 5-26: Linear approximation.

- Gaussian approximation. Models two peak hours, at lunch time and during dinner time. It uses two combined Gaussians with same mean and different standard deviation (being σ'=3σ).

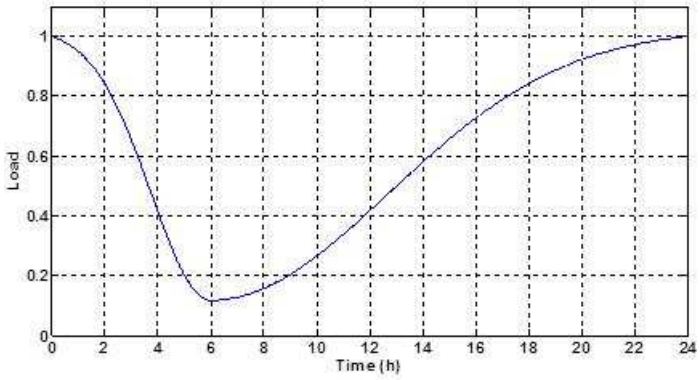


Figure 5-27: Gaussian approximation.

In addition to theoretical curves, real data has also been considered. FP7 EARTH project [17] dense urban mean traffic curves have been used to verify the proper response with respect to real data curves. As this is averaged input, in an attempt to introduce realistic curves into the simulator, some bursty noise has been added on top of the model to force some fast peaks and valleys to model real world daily patterns in urban cells.

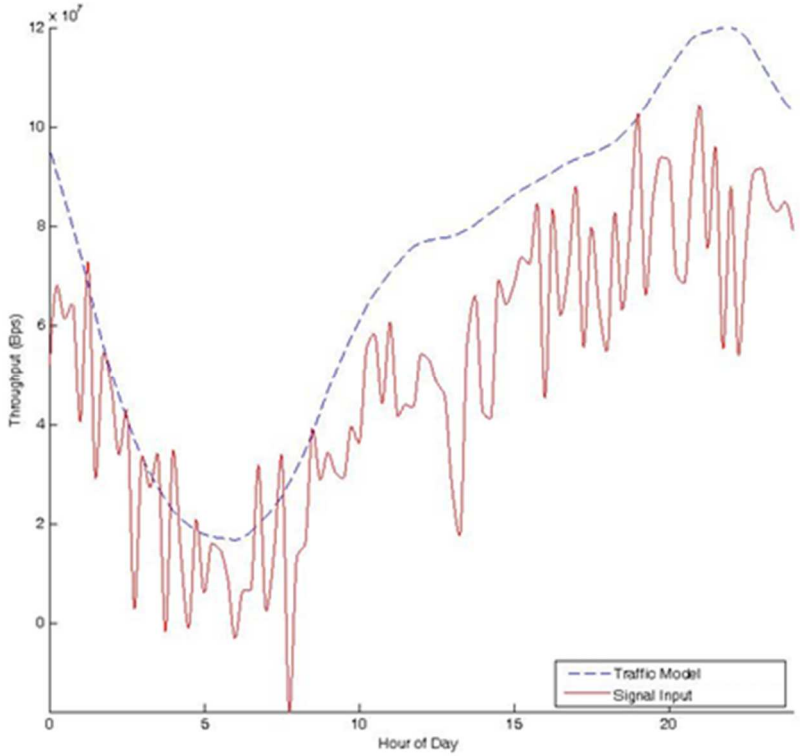


Figure 5-28: EARTH traffic model and real curve introduced in the system.

5.3.3 Simulation results

In this subsection, the results obtained due to the simulations performed using all the aforementioned parameters are presented. For each use case considered, 1.000 simulations were run, taking the mean savings as result.

The following **Table 5-7** summarizes each use case and the obtained savings. In "Curve" column the traffic model is shown (L for Linear, G for Gaussian and R for real data). "Ctrl" column denotes if fixed sectors are considered (mmW small cells approach, C) or each cluster independently takes their own decisions (4G cells, D). "Static" columns is used to identify if the simulation considers single scheme approach (S) or multiple layouts (M). "Scheme" contains the specific layouts used for the simulation. "Pred" is used to show if the simulation uses linear prediction on thresholds (Y) or not (N). Finally, the results column shows the savings in % obtained in mean for all simulations.

Table 5-7: Summary of ON/OFF results.

Curve	Ctrl	Static	Scheme	Pred	Results
L	C	Y	1/2,1/3, 1/4,1/7, 1/9	N	When d = 2/T, where T is the time period corresponding to one day, the energy savings would be the lowest, around 30%. However, when d = 0.5, the highest energy savings are achieved, reaching nearly 70% of power savings.
G	C	Y	1/2,1/3, 1/4,1/7	N	The lower the value of sigma (=gaussian deviation) is, the greater the gains obtained. For sigma values of 5 up to 16% gains are obtained. Sigma equals to 8 throws just 12% savings.
R	C	Y	ALL	Y	Savings from 22% (Macro Rural 1/2) up to 47% (PICO URBAN 1/9)
R	C	N	1/4,1/9	N	20.6% mean savings
R	D	N	1/4,1/9	N	23% mean savings
R	C	N	1/7,1/19	N	14% mean savings
R	D	N	1/7,1/19	N	25% mean savings

5.4 Dynamic beam cell management to support multi-connectivity and mobility

5.4.1 Background and motivation

As described in Section 2.1, mmW small cells ultra-densely deployed in the coverage of a macro cell (operating in below 6GHz band), is envisioned to be an important heterogeneous network structure for the future 5G mobile system. The use of pencil beams with very narrow beamwidth, e.g.: 5 to 15 degrees, offers high antenna/beamforming gain. This is a crucial technique to be adopted by mmW small cell AP, to achieve the wanted link budget target and enables energy efficient communication. In principle, each pencil beam can be viewed as a pencil cell, which can be detected or identified by its dedicated signature signal sequence. In an ultra-densely deployed mmW small cells environment, it is very likely that an UE can be served by multiple pencil beams, experiencing a variety of different path losses, which originate from one or several mmW small cell APs. To establish reliable, energy efficient and high rate data communication path between a mmW UE and the radio access network (RAN), it is very important to select and allocate one or multiple pencil beams to serve the UE. Due to the mobility of mmW UEs and variations within the surrounding environment, it is natural that the set of optimal pencil beams varies according to the dynamics of the propagation channels between the UE and RAN. As a result, it is very important for the RAN to keep tracking on such

dynamics, to provide a consistent high data rate service to the UE without user perception of service quality fluctuations. Apparently, this is a challenging task due to the severe pathloss of mmW signal in general and unpredicted channel variation.

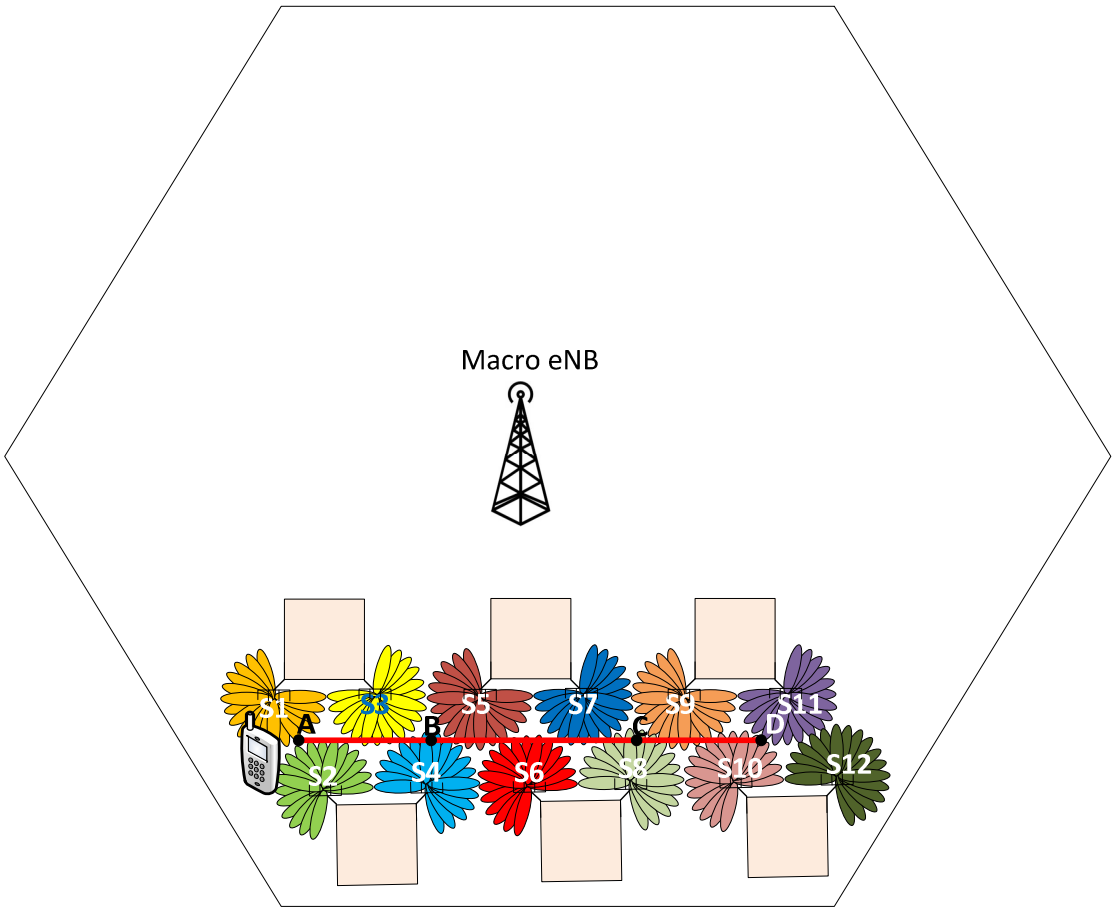


Figure 5-29: Heterogeneous network with macro eNB and mmW small cells.

One example of the studied scenario is shown in **Figure 5-29**, wherein a heterogeneous network consists of a macro cell served by a macro eNB, and a number of mmW small cells in the coverage of the macro cell, named s1 to s12, which are served by respective small cell APs. Moreover, in this Example these small cell APs are installed on house block corners along a street. Each small cell offers own coverage by pencil beams, and a UE is moving from point A, via points B and C to point D along the red line. It is shown in **Figure 5-29** that point A is in the coverage of small cells S1 and S2, point B in the coverage of S3, S4 and S5, point C in S7, S8 and S9, and point D in S10 and S11. It is obvious that optimal set of pencil beams for the UE varies along with the movement.

Some beam tracking mechanisms have been developed and standardized for WiGig systems [8]. However, these mechanisms are mainly suitable for the medium access method of WiGig systems, which has a less optimal radio resource utilization, than the scheduled systems such as LTE and its evolution. In addition, the mobility support of WLAN and WiGig is more limited than the cellular system. Direct adoption of WiGig beam tracking methods will not be optimal for the future 5G mmW radio access network which is very likely largely based on a scheduled system like LTE.

In the following section, we propose a new method for pencil beam/cell management in heterogeneous network with ultra-densely deployed mmW small cells. The proposed method provides an efficient mechanism to track and determine the optimal pencil beams/cells for the moving UE. The

selected and allocated beams can be transmitted from one or several small cell APs. As a result, multi-connectivity or coordinated multiple point transmission (CoMP) of multi-stream data packets to a UE is naturally supported as well.

5.4.2 Description of proposed innovations

The basic principle of the proposed methods is to employ a beam cell management approach, to enable downlink beam acquisition and scheduling, and further support CoMP operation and mobility.

The rest of the section is organized as follows. In Section 5.4.2.1, a novel method for mmW beam cell discovery is presented. In Section 5.4.2.2, based on carrier aggregation in LTE, beam cell aggregation procedure is proposed to support multi-connectivity and CoMP operation. In Section 5.4.2.4, beam cell reconfiguration approach is developed to enable the support of mobility. And an overall signalling example is given in Section 5.4.2.5.

5.4.2.1 Pencil beam cell discovery

As mentioned above, beamforming is an essential technology for mmW systems to achieve a reasonable communication range. In addition to some specific beam alignment procedure, such as those defined in WiGig standard [8], operating in the physical layer, the MAC layer approach can also be adopted similar to the method described in Section 4.1.2. The principle of the MAC based approach for beamforming alignment is described as follows. For example, each mmW AP can transmit a number of beam-specific reference signals (BRS), each of which is characterized with its identity and illuminating a narrow beam sector. The BRS identity can be defined by the sequence itself or time-frequency resource allocation, the details of which will be addressed in the following. When UEs in the mmW small cell detects the BRSs transmitted from the AP, a relevant BRS receive power (BRSRP) or receive quality (BRSRQ) will be transmitted back to the AP. As a result, the AP has a priori knowledge about the preferable downlink transmit beam for a specific UE, and so the beam alignment is inherently achieved.

To achieve the above MAC type beam alignment, the UE needs to search all the possible BRSs potentially transmitted in the system. Due to the potential large variety of potential mmW AP implementations, it is plausible to have the BRS structure being configurable, rather than having a fixed structure in the standard, like the primary synchronization signal (PSS) and secondary synchronization signal (SSS) structure in LTE. When configurable BRSs are employed, the UE needs to be informed about the configurations of BRSs, used in the coverage zone, which can be realized by some macro cell assisted signalling. In the case of ultra-densely deployed mmW small cells, we face a small coverage of each BRS. To create seamless connected mmW coverage, it is envisioned that a large number of BRSs are required to be transmitted from the cluster of mmW APs in the coverage. The UE may face exhaustive search space for BRSs.

To reduce the potential huge search space of BRSs, an AP-specific reference signal (APRS) with low transmission power and transmitted in low frequency band is proposed to reduce the number of BRSs to be searched by the UE. This requires that each mmW AP has the ability of transmitting the APRS in a low frequency band, e.g. by using the same frequency as the macro eNB. The APRS transmission power is determined in a way, that the APRS coverage of is approximately equal to the total coverage of BRSs originated from the AP. As such, when a UE detects an APRS and reports the RSRP or RSRQ measurements to the macro eNB, the macro eNB acquires proximity information, and can determine

which mmW AP for the UE is currently best suited. In a further step the UE can receive signalling by the macro eNB to search those BRSs transmitted from the best detected mmW AP. At the first glance, this seems to increase the complexity of mmW APs, by adding a dedicated RF circuit to transmit low frequency signals. However, the small cell enhancement for ultra-densely deployment scenario has been extensively developed in LTE, and it is consequently assumed that such small cells are first widely deployed in practical environment, before the mmW RAT is added to further boost the capacity. As such, it is very practical to assume that the mmW AP shall have a dual RAT capability, including both low frequency RAT such as LTE and a mmW RAT. Moreover, the proposed method does not require the mmW AP to have full LTE functionality, in principle this can be limited to the capability to transmit the APRS in low frequency band. The concrete design approach of APRS and BRS is presented in the following.

APRS design

As mentioned above, the mmW AP is transmitting both the omnidirectional APRS in macro cell spectrum/low frequency band and the directional BRSs in mmW spectrum. The APRS is not required to be transmitted in an omnidirectional pattern. In principle, the APRS can also be transmitted in a sector, when the AP and the pencil beams have only to serve a sector.

As illustrated in **Figure 5-30**, the APRS is designed to support similar coverage as that of aggregated pencil cell BRSs transmitted from the AP;

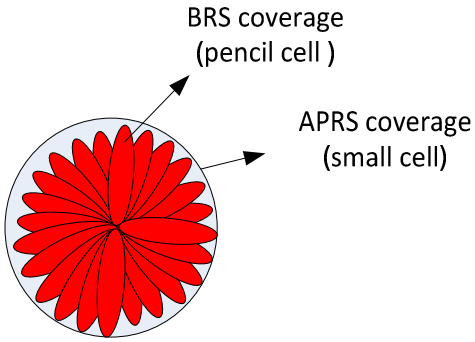


Figure 5-30: Coverage of APRS and BRSs from the same mmW AP.

Several LTE compatible options can be reused for APRS signal design as follows.

Alternative 1: LTE cell specific reference

The CRS port 0 of LTE system can be reused for the APRS due to the fact that the sole purpose of APRS is to enable the UE to detect the proximity to a mmW AP. As such, the APRS can be transmitted within a narrow bandwidth of 1.4MHz. The detailed CRS port 0 design can be referred to [16]. Small cell discovery signal has been developed in LTE release 12 [16] and is comprised of synchronisation signals, CRS port 0 and CSI-RS. The intention of introducing discovery signal is to enable small cell on-off operation for network power consumption optimization. Specifically as presented in Section 5.3, when there is no traffic demand in the cell, e.g. during the night, small cells with low traffic demand can be switched off, except that the discovery signals are transmitted periodically. However, in the LTE system based on previous releases, the CRS signals are always transmitted regardless of actual traffic demand. Discovery signal can be considered as an enhanced alternative for the APRS to enable mmW AP to be turned off during the time with less traffic demand.

Alternative 2: LTE sidelink synchronization signal

In LTE Release 12, the proximity-based services (ProSe) have been developed and standardized to support D2D communication and discovery functions. As a result, a new set of physical sidelink channels and signals is defined in the standard [16]. From macro eNB point of view, the mmW AP can be treated as a UE in the coverage of macro cell, which can further provide high data rate communication with the normal UEs in the proximity. Therefore, the physical sidelink synchronization signal (PSSS) can be naturally configured for the mmW AP to serve as the APRS to be discovered by the UE in the proximity.

BRS design

The BRS is periodically transmitted to illuminate respective pencil beam cell. The periodicity and time-frequency resources of each BRS shall be configured by the macro eNB. Some desired design aspects of BRS are in the following:

Design Aspect 1: The BRS design can reuse similar approach for LTE discovery signal. For instance, the BRS can be a plurality of beam specific synchronization signals and/or CSI-RSs.

Design Aspect 2: The different BRSs of the same AP may be transmitted with different periodicities. Specifically, those beam cells with more served UEs or traffic demand may transmit respective BRS more frequently than those with less traffic demand.

Design Aspect 3: Some BRSs may be transmitted simultaneously provided that different RF chains/analogue beamformers are employed for these beams.

Design Aspect 4: Those BRSs sharing the same RF chain are transmitted in different transmission time. Macro eNB needs such constraints from the mmW AP to configure a feasible BRS set for the mmW AP.

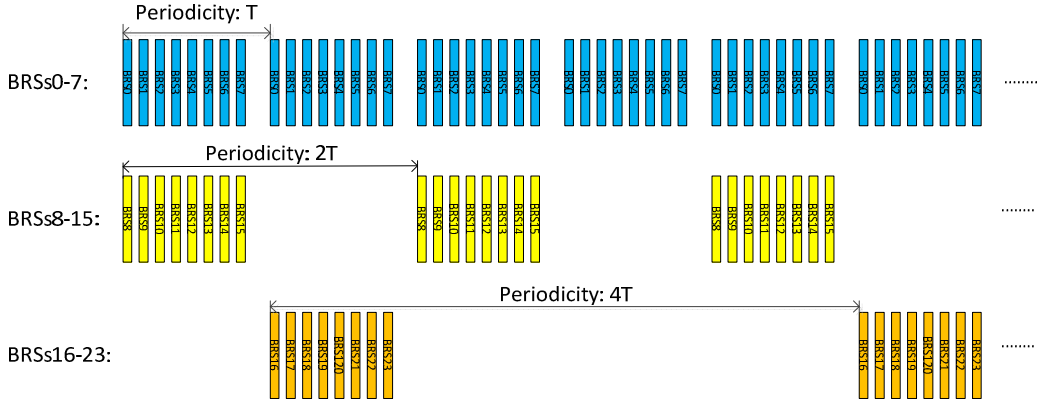


Figure 5-31: Timing arrangement of configured BRSs.

One timing configuration of BRSs is illustrated in **Figure 5-31** as an example configuration. With the assumption that the mmW small cell AP is equipped with three analogue beamformers, and that each beam of 15° beamwidth can be transmitted from respective analogue beamformer, at most three beams can be simultaneously transmitted from the mmW AP. As shown in **Figure 5-31**, three groups of BRSs with different periodicities can be configured to the mmW AP. Assuming that more traffic demands and associated beam tracking measurements are required in the beam cells 0 to 7, the BRSs 0 to 7 have smaller periodicity than other BRSs. On the contrary, there is least number of mmW UEs served in the beam cells 16 to 23, the respective BRSs, namely BRSs16-23, have been configured with relatively longer periodicity to save the transmission power.

APRS and BRS detection and measurement

A UE performs a two-stage cell detection and measurements. In the first stage, UE detects the proximity of mmW APs by searching APRSs, and reports the measurement results of detected APRSs to macro eNB. With the measurement results from the UE about the detected APRSs, macro eNB configures the UE to further detect and measure a set of BRSs associated with those selected mmW AP. In the second stage, UE detects and measures the BRSs configured by macro eNB, and reports the BRS measurement results either periodically or as a response to a macro eNB instant request.

The first stage cell detection enables UE to discover a mmW AP without the need of beam acquisition and avoids exhausted initial beam cell search from all possible mmW APs. The second stage cell detection and measurement requires the UE only to measure those BRSs signalled by the macro eNB. It is plausible that the macro eNB selects those BRSs originated from the mmW AP whose APRS RSRP or RSRQ reported from the UE are above certain defined threshold, and further configures them to the UE. Such selection strategies shall significantly reduce the BRS search and detection burden of the UE.

5.4.2.2 Beam cell aggregation for multi-connectivity and CoMP

After being configured with a measurement set of BRSs, the UE reports the BRSRP or BRSRQ periodically or as per macro eNB request. Based on the reported BRS receive quality from the UE, macro eNB configures some selected mmW beam cells to the UE as secondary component cells. As a result, the downlink beam acquisition is inherently realised by the secondary cell aggregation in the beam granularity. The configured mmW beam cells may be transmitted from one or several mmW APs so that coordinated multiple transmission is naturally supported as well.

It is envisioned that the number of configured secondary cells (SCells), which may possibly include small cells in low frequencies and mmW beam cells, can be larger than the number of SCells supported in the current LTE standard. To cope with this problem, we can continue to increase the number of supported SCells in the standard. For example, the carrier aggregation up to 32 component carriers is to be supported in LTE Rel-13, and the number can be further increased in future releases. We can also adopt another solution to allow secondary cell to be further aggregated with own secondary cells, named as “order-2 secondary cell” to support hierarchical carrier aggregation. This would increase the number of supported SCells exponentially, and without the limitation of the hierarchy level, the number of supported SCells can be infinite. Certainly some more signalling capability of the inheritable SCell which can aggregate own SCells, may be required to support some RRC functions.

5.4.2.3 Selection of beam cells for secondary cell configuration

In this section, two methods, named as “Method C1” and “Method C2”, are proposed for beam cell selection. In “Method C1”, macro eNB defines a threshold T_{BRS_RSRP} of BRS RSRP/RSRQ, and all the beam cells with the reported BRS RSRP/RSRQ results above T_{BRS_RSRP} can be selected as the secondary cells. Due to channel variation and intermittent characteristic of mmW signal propagation caused by unpredicted blocking, the BRS RSRP/RSRQ can vary rapidly and it may cause quite frequent beam SCell reconfiguration.

To avoid frequent SCell reconfiguration due to the channel variation caused by the UE or surrounding environment movement, “Method C2” is proposed as follows. In “Method C2”, multiple thresholds are defined, e.g., T_{BRS_RSRP1}, $T_{BRS\_RSRP2, \ldots,}$ $T_{BRS\_RSRPK.}$, without loss of generality, we can assume $T_{BRS\_RSRP1} > T_{BRS\_RSRP2} > \ldots > T_{BRS\_RSRPK.}$ .And each threshold T_{BRS_RSRPk} where $k = 1..K$, is used to define a beam

cell cluster which comprises B_k adjacent beam cells centred about the beam cell with reported BRS RSRP/RSRQ greater than $T_{BRS\_RSRPk.}$ It is plausible to define $B_1 > B_2 > \ldots > B_K$. Let p_i where i=1,..,L, define the L BRSs reported by the UE to macro eNB, each of which has the reported RSRP/RSPQ greater than one of those defined thresholds $T_{BRS\_RSRPj,}$ where j=1,..,K. Macro eNB selects all the beams in L beam clusters defined as P_i where i=1,..L, as the secondary cells to the UE. Specifically P_i contains a number of adjacent beam cells which are centred around the beam cell p_i, and the number of beam cells in P_i, is dependent on which threshold the RSRP/RSRQ of p_i is above. The other beam cells than p_i in P_i can be called as "guard beams". Thanks to the guard beams, the SCell reconfiguration needs due to mobility reduce significantly.

"Method C1" can be considered as a special case of "Method C2". With proper selection of these design parameters, a good trade-off between the size of beam cells to be monitored for the possible data transmission and the required SCell reconfiguration signalling overhead, can be achieved.

5.4.2.4 Beam cell reconfiguration to support mobility

When the updated APRS and BRS receive qualities are reported from the moving UE to macro eNB, reconfiguration/reselection of mmW beam SCells can be performed for the UE by macro eNB. To avoid the excessive beam cell reconfiguration overhead, the above "Method C2" is proposed to include some guard beams around each reported anchor beam as the configured SCells. By virtue of the guard beams, a beam cell set reconfiguration/reselection with low signalling overhead is proposed as follows.

Step 1: Let $P_{current}$ define the current configured secondary beam cells. Based on updated BRS RSRP/RSRQ measurement results, a new set P_{new} of beam cells can be determined according to Method C1 or Method C2.

Step 2: Create a beam cell set P_{work} which is the intersection set of $P_{current}$ and P_{new}.

Step 3: Calculate the selection metrics O_{work} and O_{new} of P_{work} and P_{new} respectively. For example, each selected beam cell in the set can be given a weight factor according to its reported RSRP/RSPQ, or some minimum weight factor if the beam cell is not reported but selected as a "guard beam" in Method C2.

Step 4: If the ratio of O_{work} to O_{new} is below a certain defined threshold T_{Re}, the beam cell reconfiguration is performed. Otherwise, the configuration of SCells remains the same.

It is noted that mmW pencil cells "handover" is inherently implemented as the result of reconfiguration of secondary mmW component beam cells.

5.4.2.5 Overall signalling example of proposed method

The signalling diagram of the proposed method is illustrated in **Figure 5-32**, were all control signalling are communicated between UE and macro eNB by using macro cell spectrum. This also reflects the current design trend of control- user-plane split for the future 5G mobile network. After a UE is attached to macro eNB, the UE is configured by macro eNB with a list of APRSs to be measured and associated reporting mechanism. Upon the reception of APRS measurement configuration, the UE starts to detect and measure the receive power or quality of configured APRSs. The UE periodically reports the receive power/quality and/or channel state information with respect to detected APRSs according to the configured feedback mode. Based on the receive power/quality feedback of APRSs from the UE, macro eNB determines a set of BRSs, associated with those APRSs with receive signal power or quality higher than a defined

threshold. The eNB further configures them as a measurement set of BRSs with respective feedback mode to the UE. With the further configuration of the measurement set of BRSs, the UE detects/measures and further reports the received quality of detected BRSs. Based on the feedback w.r.t. BRSs, the macro eNB determines and configures a set of beam cells associated with those strongly received BRSs as secondary components cells for the UE. With the repeated information updates about receive quality feedback of APRSs and BRSs, the macro eNB may reconfigure the measurement sets of APRSs and BRSs to the UE. Accordingly, the secondary component cells can be reconfigured as well in the case of UE moving to different locations with different preferred pencil cells.

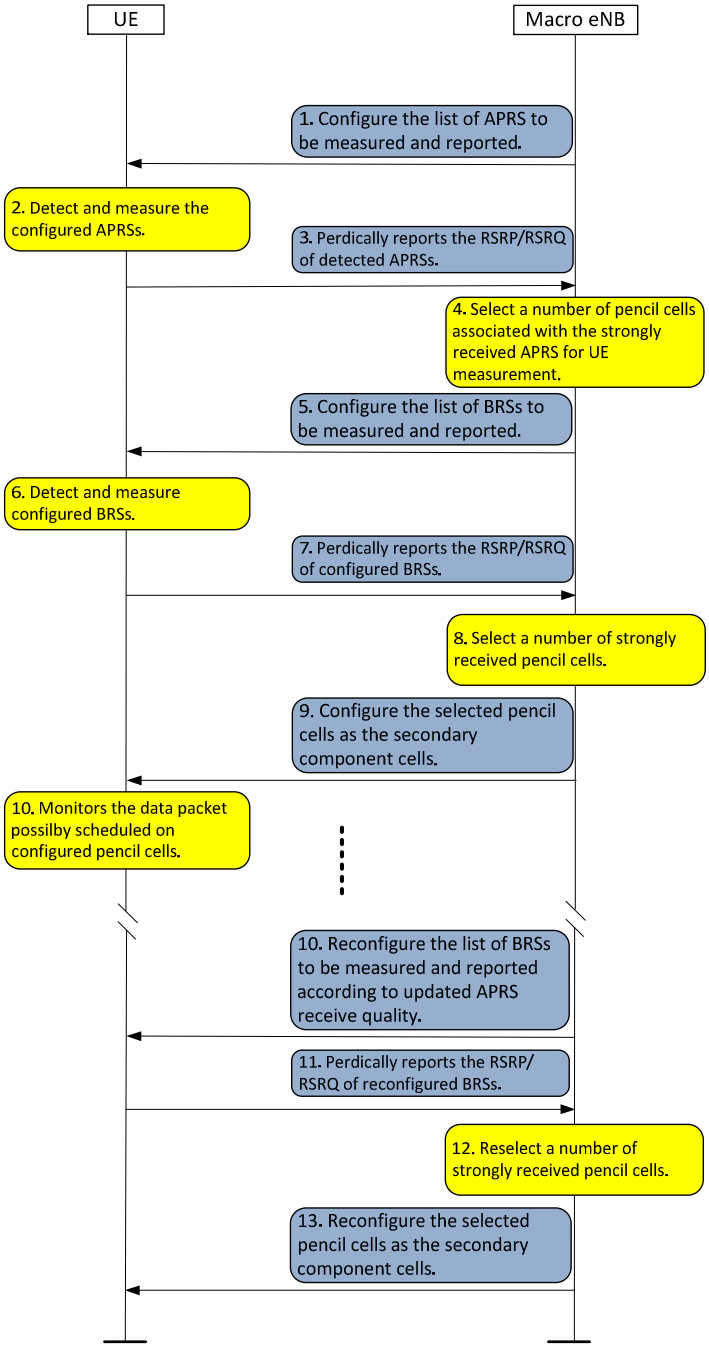


Figure 5-32: Signalling diagram of proposed method.

In the example of **Figure 5-29**, the UE can be first configured with a measurement set comprised of APRSs in regard to all 12 small cells, i.e., S1 to S12. At the time when UE moves to point A, UE may report the receive quality of all configured APRSs, where APRSs of S1 and S2 have the strongest receive power by the UE. As a result, macro eNB may configure the UE with a BRS measurement set, including those BRSs transmitted from S1 and S2. Upon receiving feedback of configured BRSs, macro eNB selects and configures a set of received beam cells with good receive quality, as the secondary cells of the UE. In the course of UE moving from point A to point D, the secondary beam cells for the UE may come from the following sets of small cells: (S1, S2) at point A, (S3, S4, S5) at point B, (S7, S8, S9) at point C, and (S10, S11) at point D. As a consequence, at each point, the UE can receive mmW data packets from multiple beams even with different locations. This would further increase the throughput by employing multi-stream transmission or improve the reliability by virtue of transmit diversity.

5.4.2.6 Summary of the proposed method

To implement the proposed method, the current LTE system needs to be equipped with multiple antenna ports for measurement set and report set. It is envisioned that a relatively large number of mmW APs, where each of them transmits its own APRS and a number of BRSs, will be deployed in the 5G mobile network. Based on the UE reports of detected beam cells, all those beam cells with sufficient receive quality can be configured as secondary cells so that the existing carrier aggregation framework in LTE can be reused to implicitly achieve beam alignment. To avoid frequent reconfiguration of secondary cells due to the channel dynamics, some guard beams centred around those strongly reported beams can be also configured as the secondary cells as “Method C2”.

Apparently UE may need to be (re)configured with a measurement set of relatively large amount of APRSs and BRSs. Moreover, the number of aggregated carriers/cells may also need to be increased beyond the current LTE system. The carrier aggregation of up to 32 component carriers shall be supported by LTE Rel-13 system. However the proposed method of beam cell aggregation requires that a mmW component carrier may further aggregate a number of beam cells. In case of UE aggregating component carriers ranging from cmW to mmW bands, the total number of aggregated cells/carriers may reach a value beyond the current supported range. As proposed in Section 5.4.2.2, a more futureproof solution is to employ hierarchical cell/carrier aggregation that allows a secondary component carrier to be further aggregated with its own secondary cells. For example, a UE can aggregate two component carriers, one in cmW spectrum as primary cell, and another in mmW spectrum as secondary cell (SCell). And the SCell can further aggregate multiple beam cells as its SCells.

The proposed method of two-stage cell detection and measurement reduces the initial cell search burden and feedback overhead significantly. The proposed beam cell aggregation approach achieves the mmW downlink beam alignment, without defining specific beam alignment protocols. As such, this approach simplifies the system design and relevant standardization effort, and further reduces the physical layer latency. In addition, the channel state information feedback and beam cell (re)configuration are performed by the control plane signalling in macro cell which is more robust and power efficient.

The mmW downlink CoMP operation can be also supported in the format of a joint beam cell allocation or a dynamic beam cell selection, determined by the macro eNB. Moreover, the beam

cell reconfiguration based on periodically reported receive quality of configured BRSs during the movement of UEs, supports the UE mobility so that a moving UE is always configured with a number of beam cells with strong receive signals.

5.5 Beam steering in access

5.5.1 System model

We consider single user systems with beam-switching antennas (modeled as ULA) at the transmitter and the receiver as shown in **Figure 5-33**. In addition to the mmW high gain antennas, we consider a secondary system, which could be an existing cellular system like LTE, to perform the initial synchronization and to exchange control information independently from the status of the beam-alignment process. The secondary system is proposed through a parallel mmW connection via omnidirectional antennas, where the poor SNR value can be mitigated by intense coding, or a parallel lower frequency connection (e.g. 4G connection).

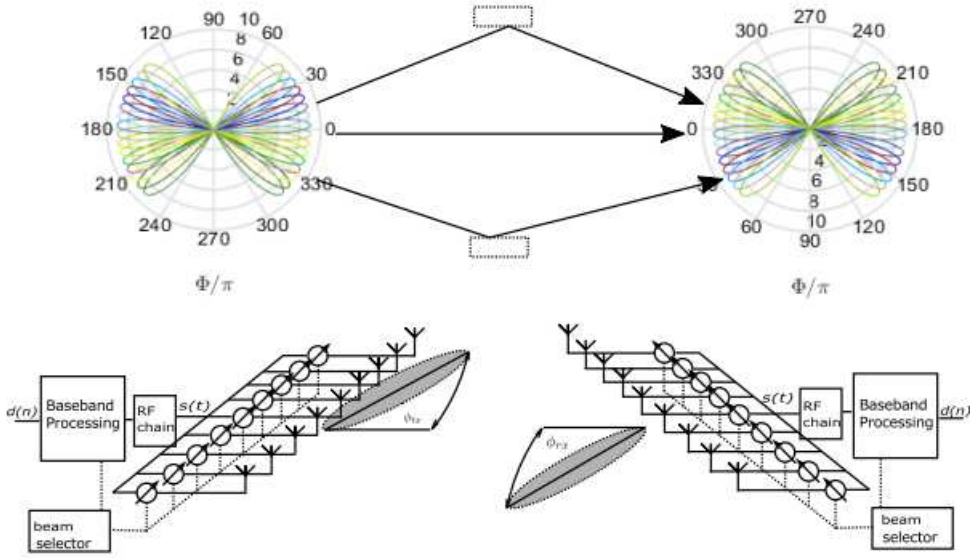


Figure 5-33: System model.

The channel is stimulated via a transmitted beam of the initiator, which could be AP or UE. The introduced multipath signals, due to reflection from surrounding objects, can be "collected" via the receiver beam. The transmitter and receiver consist of an ULA, one RF chain and digitizer, and the baseband processing. The beams are generated by applying discrete phase shifts to each antenna element, also known as codebook based beamforming. The baseband can thus only select a beam index and does not have to directly access the phase shifters.

The TDD system is adopted to evaluate beamforming performance because of the reciprocity of the spatial characteristics of the channel and more flexible scheduling strategies for downlink and uplink. One device will serve as an initiator, to initiate the beam alignment procedure and the other as a responder, to receive the data and respond to the request. Because of the assumption that both initiator and responder have identical antenna designs and number of RF chains and in order to specifically introduce the beamforming schemes for access links, in the following introduction in this section, UE will be regarded as initiator and the AP as responder device, but this can be changed according to the protocol's need. A single carrier modulation scheme with frequency domain

equalization is considered to be used in this system, because of its lower PAPR compared to the well-known multicarrier schemes such as OFDM. Thus, less expensive and less power consuming transceivers can be used, which is an important issue at higher frequencies. Each device consists of a baseband unit, one analogue to digital converter (ADC) and digital to analogue converter (DAC) and the ULA including a beam-forming network. The production of accurate and low power consuming phase shifters for the mmW band is too complex, thus most antennas use alternative approaches e.g. Butler matrices. These allow only to steer the main lobe in a certain set of predefined directions, and not to scan the lobe continuously as with phase shifters and are commonly named beam-switching antennas. For the sake of easiness of the analytical description, the beam forming network is modeled here via phase shifters allowing only certain phase shifts for each antenna element for each direction. In general, the received signal of a multiple in multiple out (MIMO) system with N_{tx} transmit antennas, N_{rx} receive antennas and N_S RF chains can be mathematically described as

$$\mathbf{y}(t) = \mathbf{W}^{\mathrm{H}}\mathbf{H}\mathbf{F}\mathbf{s}(\mathrm{t}) + \mathbf{W}^{H}\mathbf{n}(\mathrm{t}) \quad (1)$$

with $\mathbf{F} \in \mathbb{C}^{N_{rx}\times N_s}$, describing the precoding matrix, $\mathbf{W} \in \mathbb{C}^{N_{tx}\times N_s}$, describing the receive combining matrix at the responder and $\mathbf{H} \in \mathbb{C}^{N_{tx}\times N_{rx}}$, is the channel matrix. If, as in our case only one RF chain is available, the precoding matrix **F** and combining matrix **W** reduce to vectors, so that

$$\mathrm{y}(t) = \mathbf{w}^{H}\mathbf{H}\mathbf{f}s(t) + \mathbf{w}^{H}\mathbf{n}(\mathbf{t}) \quad (2)$$

with $f \in \mathbb{C}^{N_{tx}\times 1}$, describing the complex phase weighting vector of the initiator, $w \in \mathbb{C}^{N_{rx}\times 1}$, describing the complex phase weighting vector of the responder. As we consider a beam-switching antenna, only discrete directions and its corresponding weighting vectors for the phase shifters can be chosen from a predefined codebook. The codebook is designed for discrete values of the sum and delay beam former,

$$\mathbf{f} \in [\mathbf{f}_{tx,1}(\phi_{tx,1}), \mathbf{f}_{tx,2}(\phi_{tx,2}), \ldots, \mathbf{f}_{tx,K_{tx}}(\phi_{tx,K_{tx}})] \quad (3)$$

$$\mathbf{w} \in [\mathbf{w}_{rx,1}(\phi_{rx,1}), \mathbf{w}_{rx,2}(\phi_{rx,2}), \ldots, \mathbf{w}_{rx,K_{rx}}(\phi_{rx,K_{rx}})] \quad (4)$$

with a discrete set of steering angles ϕ_{tx} and ϕ_{rx} and

$$\mathbf{f}_{tx}(\phi_{tx}) = \frac{1}{\sqrt{N_{tx}}}\left[1, e^{\frac{j2\pi d}{\lambda}\sin(\phi_{tx})}, \ldots, e^{\frac{j2\pi d}{\lambda}(N_{tx}-1)\sin(\phi_{tx})}\right]^{T} \quad (5)$$

and $\mathbf{w}_{rx}$ respectively.

5.5.2 Simplified channel model

As stated in [10], a mmW channel consists only of a few multipath components. Because of this effect, traditional models based on a Rayleigh or Rice distributions are not applicable. Thus, we introduce a simple geometric model, which implies major characteristics of the real mmW channel. We consider the transmitter and receiver antenna array are facing each other. Scatterers are located on a line, placed orthogonally at mid-distance between the transmitter and the receiver. The contribution of the reflection and the additional paths compared to the direct path between the transmitter and the receiver is covered by a stochastic parameter α. Each scatterer corresponds to a different path of the channel. Due to different distances and thus travel times, the arrival time of each path is different. We assume sampling frequencies of 1.5 GHz, distance differences up to 20 cm will be resolved. Because we are limited by the restriction to one beam, which is only steerable in µs range, we can only focus on detecting the strongest beam. Because we consider using SC-FDE or a similar

modulation scheme, we can assume the effect of the multi-tap channel in time domain to be perfectly recovered by the equalization and the taps to be constructively added. For this reason, we focus only on the spatial effects of the channel, taking the receive power in a single beam of one beam combination as quality measure. The channel matrix H is described as

$$\mathbf{H} = \sqrt{\frac{N_{rx}N_{tx}}{\rho\, L}} \sum_{l=1}^{L} \alpha_l \mathbf{a}_{rx}(\phi_{rx,l}) \mathbf{a}_{tx}^{H}(\phi_{tx,l}), \tag{6}$$

where L defines the number of paths, which are physically available and $\alpha_l \in N\mathbb{C}(0,1)$ is the amplitude of each path, which is normal distributed. The array propagation vector of the transmitter is defined as

$$\mathbf{a}_{tx}(\phi_{tx}) = \frac{1}{\sqrt{N_{tx}}} \left[1, e^{\frac{j2\,\pi d}{\lambda}\sin(\phi_{tx})}, \dots, e^{\frac{j2\,\pi d}{\lambda}(N_{tx}-1)\sin(\phi_{tx})}\right]^{T} \tag{7}$$

where $d = \frac{\lambda}{2}$ is the distance between the antenna elements and ϕ_{tx} describes the steering direction of the main lobe. The array propagation vector of the receiver is constructed accordingly. The angles ϕ_{tx} and ϕ_{rx} are uniformly distributed in the range of $\phi_{tx,rx} \in (0 \dots \frac{\pi}{2})$.

5.5.3 Problem formulation

Because of the assumed restrictions of the system, which were discussed in the last section, an effective beam alignment procedure is difficult to design. The ability of only one beam with a defined beam width at one instant precludes all sorts of multi-codebook schemes and the ”coding-the-beam” approach [11]. The unavailability of several RF chains prohibits some preprocessing in the baseband domain. Thus, the problem cannot be solved by traditional adaptive approaches like the steepest descent, because the channel is not explicitly known. We may split the problem in several dependent problems for transmitter and receiver beam search, but this requires access to the antenna signal in order to obtain the covariance matrix of the antenna signals [12][13]. A simple solution is to try all possible combinations of Tx and Rx beams, which would lead to the exhaustive search algorithm, which has the disadvantage of needing $N_{tx} \times N_{rx}$ iterations.

In order to use existing algorithms in literature, one can convert the problem of finding the best matching beams into a so-called ”black box optimization problem”. The receive power of equivalent SNR is often used as a measure of the quality of a link obtained by two beams, which can be achieved by the algorithm [11][14]. We define the impact of the channel and the preprocessing vector $\mathbf{f}(\phi_{tx})$ and post-processing vector $\mathbf{w}(\phi_{rx})$ as black box while the angles ϕ_{tx} and ϕ_{rx} as variables. Additionally a noise term will be added. The system can be described as

$$\mathrm{y}(t) = \underbrace{\mathbf{w}^{H}(\phi_{rx})\mathbf{H}\mathbf{f}(\phi_{tx})\, s(t) + \mathbf{w}^{H}\mathbf{n}(t)}_{\text{black box}} \tag{8}$$

In order to find the best beam pair, the absolute power has to be maximized. This maximization problem can easily be converted to a minimization problem by changing the sign of the expression

$$(\phi_{tx,min}, \phi_{rx,min}) = \underbrace{\operatorname{argmin} c_0 - |y(t)|^2}_{\text{cost function } J(\phi_{tx},\phi_{rx})} \tag{9}$$

This description allows us to take advantage of predefined black box optimization algorithms, which are described in literature.

5.5.4 Proposed algorithm using "black box functions"

In this section, we will briefly describe the idea of the mode pursuing sampling method for a discrete variable space (D-MPS), which is used with small modifications in order to solve our problem effectively. Detailed information about the algorithm can be found in [15]. The concept of D-MPS is to achieve a balance between exploration of possible new optima and exploitation of discovered optima regions by using a concept of two circle areas, which define the range of search. Following this approach, one circle starts with a small initial radius increasing, if evaluations in this sphere hold an improvement, while the initial large radius of the second circle area will do the contrary. A linear line spline fitting function is used in order to build up the model, which serves to find possible optima without actually evaluating the real function. It should be noted that the evaluation of the cost function represents a one beam change of initiator and responder, sending and receiving the training data, while the term iteration corresponds to one bigger part of the algorithm, containing m evaluations and one feedback over the secondary link. Table 5-8 gives a list of the used parameters. The algorithm is summarized as follows:

- Step 1: All parameters, especially the radii have to be initialized to default values (Table 5-8).
- Step 2: For initial points of several values of ϕ_{tx} and ϕ_{rx}, the cost function has to be evaluated in order to get a basis for the model. In contrary to the original model, this is done here using a predefined pattern, which allows to explore the variable space equidistantly. The minimum of the found values serves as center of the two circle areas from now on.
- Step 3 performs the basic D-MPS algorithm.
    - Step 3.1: From each of circle regions of valid arguments ϕ_{tx} and ϕ_{rx}, denoted as D1 and D2, N/2 arguments are chosen randomly. They now form the set of cheap points.
    - Step 3.2: The model of the two dimensional curve is approximated by using the existing expensive points, which are actually evaluated. This is done by using line spline interpolation, where evaluated arguments are connected via plain surfaces, avoiding oscillations of higher order interpolation techniques.
    - Step 3.3: In this step, the expensive points, which actually will be evaluated by actual real function are chosen among the cheap points, which are only evaluated by the model function. This is called discriminative sampling. After the decision, the resulting expensive point will be evaluated by the cost function. This corresponds to perform some channel measurements by using specific setups of antenna steering angles at the transmitter and receiver. The detailed procedure will be explained below.
- Step 4: The evaluated m expensive points from the last step will be added to the list of known points, which are called $E(J) = [\phi_{tx,1}, \phi_{rx,1}, J(\phi_{tx,1}, \phi_{rx,1}); \dots; \phi_{tx,M}, \phi_{rx,M}, J(\phi_{tx,M}, \phi_{rx,M})]$. From this list, the current minimum of the cost function $J(\phi_{tx}, \phi_{rx})$ has to be specified and the corresponding arguments $\phi_{tx,1}$ and $\phi_{rx,1}$. This point will serve as the center for the radii furthermore.
- Step 5: This step actually performs the double sphere strategy.

- If the last step discovered a new minimum: increase radius of smaller hypersphere to 1/α,
- Reduce the radius of bigger hypersphere by α. If there was no improvements in the last n_α iterations: reduce the smaller radius by 1/α, increase the bigger radius by α .
- Check borders (minimum radius and maximum radius) of the radii and restrict them to the nearest one.

- Step 6: Define new hyperspheres D_1 and D_2 by using new radii and center point. Check whether the maximum iteration is reached. Otherwise go to step 3.

Table 5-8: Parameters of the access link simulations.

Parameters	Definition	Unit	Value
N_{rx}	Number of receive antennas	-	8
N_{tx}	Number of transmit antennas	-	8
B_{tx}	Number of beams in the defined angle space	-	16
B_{rx}	Number of beams in the defined angle space	-	16
N_s	Number of RF chains	-	1
ϕ_{tx}	AoD	degree	-
ϕ_{rx}	AoA	degree	-
L	Number of paths	-	-
N	Number of cheap points per iteration using model	-	30
m_{init}	Number of initial expensive points/evaluations	-	24
m	Number of expensive points/evaluations per iteration	-	4
n_{it}	Number of iterations	-	4
$J(\phi_{tx,k}, \phi_{rx,k})$	Cost function	-	-
$\hat{J}(\phi_{tx,k}, \phi_{rx,k})$	Model of the cost function by interpolation	-	-
D_1	Variable space of the smaller circle area	-	-
D_2	Variable space of the larger circle area	-	-
α	coefficient for increasing and decreasing the radii	-	-

Figure 5-34 shows a contour plot of an example cost function. Several scatterers implicate minima and it can be seen, that due to the double hypersphere, a local extremum was dismissed due to the global optimum (channel with 3 paths). The evaluations show, that the algorithm is able to step from a non-optimal minimum to the best minimum. This explains why strict gradient based methods are avoided here.

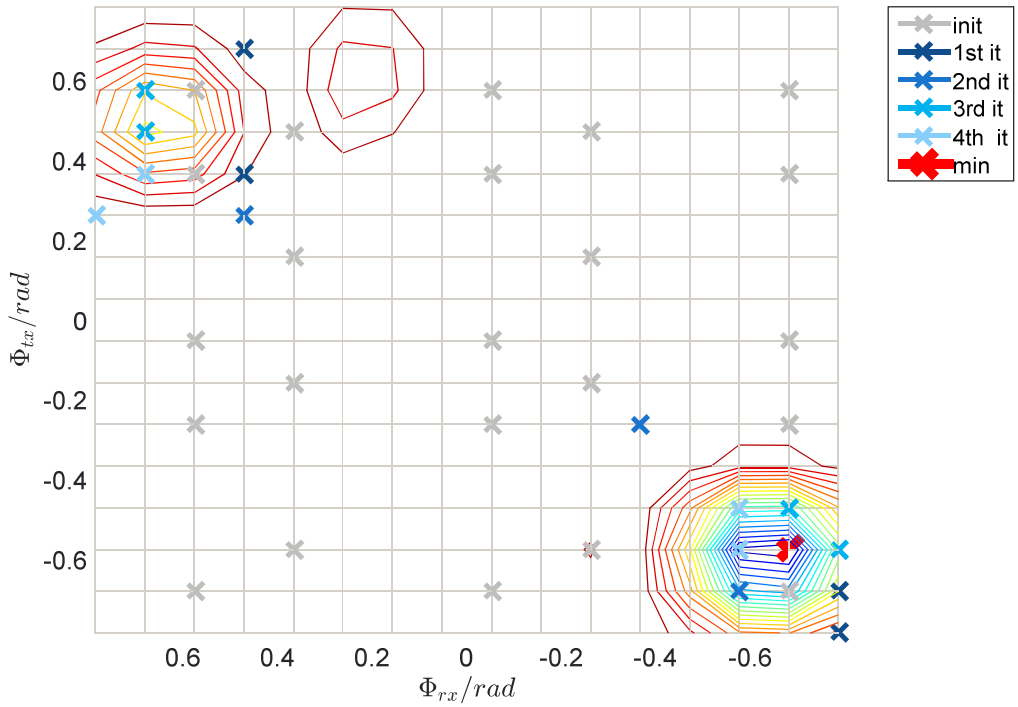


Figure 5-34: Cost function $J(\phi_{tx}, \phi_{rx})$ and evaluations (black) and final result (red). It can be seen, that due to the double hypersphere, a local extremum was dismissed due to the global optimum (channel with 3 paths).

The timing diagram of the algorithms is depicted in Figure 5-35. After applying the time synchronization by the secondary link, the initial exploration of the channel performed. Therefore, the initiator sends a PN sequence in a SC-FDE symbol subsequently with beam $i_{i,1} \dots i_{i,end}$, while the responder receives data subsequently with beam $i_{r,1} \dots i_{r,end}$. After this initial beam-searching, the responder generates an interpolation model of the channel. Based on the described algorithm, it generates the next m beam pairs for the initiator and the responder ands sends the corresponding initator beam indexes to the initiator via the secondary link. Then the initiator again transmits a PN sequence with now m different beams, while the responder also receives the data with m different beams. The responder uses the new channel measurements to improve the interpolation model. Based on this, it sends again the next beam indexes to the initiator. This is repeated until the maximum number of iterations is reached. Then the best beam pair is chosen and communicated to the initiator via the secondary link. After this, data transmission starts.

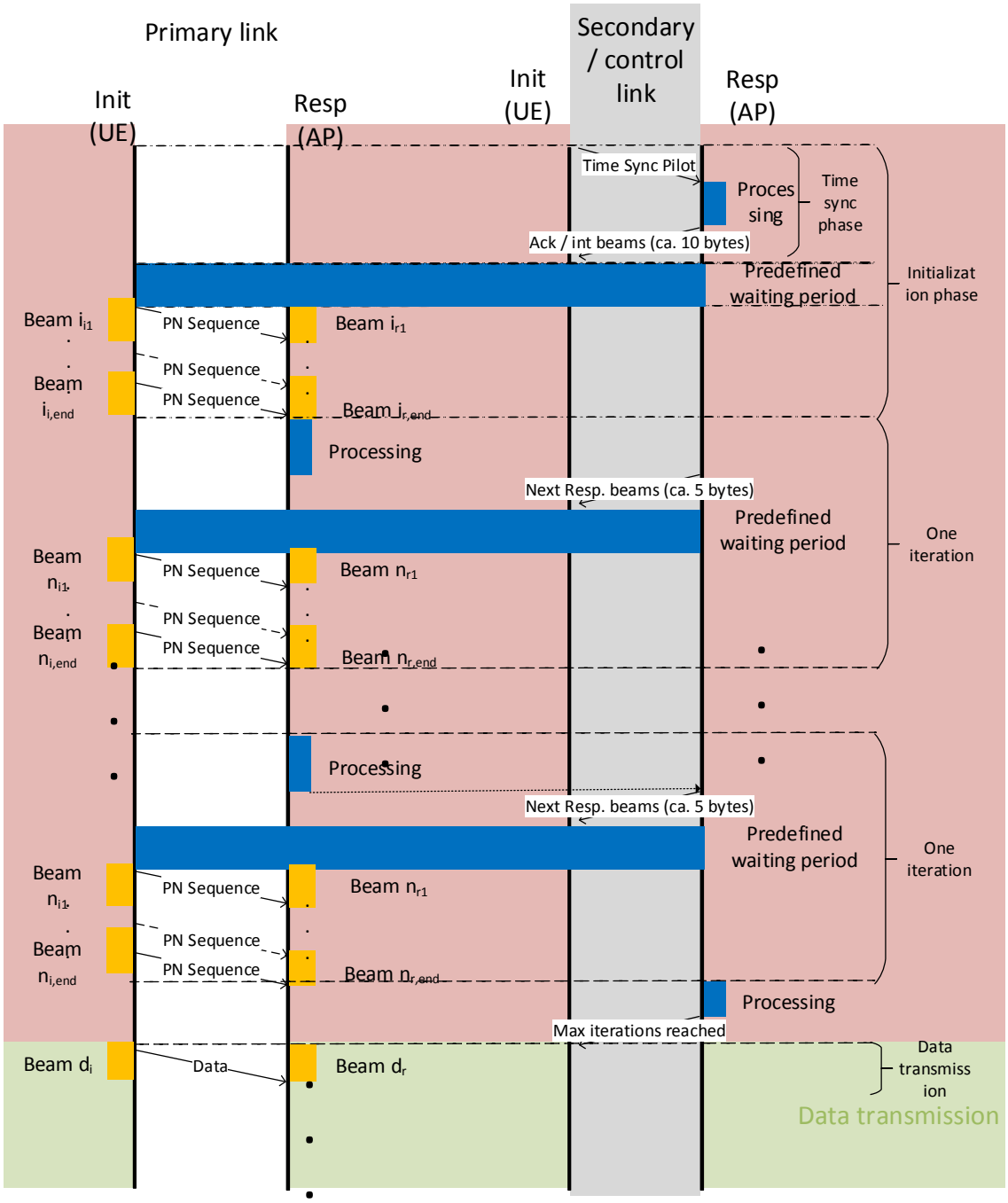


Figure 5-35: Timing diagram of the algorithm.

5.5.5 Simulation results for static scenarios

In order to show the performance and ability as also the suitable case of applications of the algorithm, some simulations without the influence of noise have been performed. As performance measure in absence of noise serves the normalized gain, which is defined as the received power of beam combination found with the proposed algorithm, divided by the best achievable received power by determining the best beam combination by using exhaustive search

$$norm.\,gain = \frac{achieved\ power\ with\ iterative\ beam\ alignment}{\max achieved\ power\ with\ exhaustive\ search}$$

The simulation neglecting the noise allows us also to fix a valid working point for the parameters of the proposed algorithm, especially the number of evaluations and the number of iterations.

5.5.5.1 Performance of the algorithm without noise

At first, the behavior of the system without the influence of noise is studied. At first a range of $\in -\frac{\pi}{4}..\frac{\pi}{4}$. Figure 5-36 - Figure 5-38 show the influence of the number of iterations and the number of the evaluations (*m*) of the cost function at one iteration, which is equivalent to the number of beam changes in one iteration. The three subplots differ in the number of initial evaluations, which serve as input for the black box algorithm. The initial evaluations can be seen as the performance of a very rough exhaustive search algorithm. If using only three equidistant points (Figure 5-36) the achievable performance in terms of the achieved power of a beam pair increases if either the number of iterations or the number of evaluated points per iteration, *m* is increased.

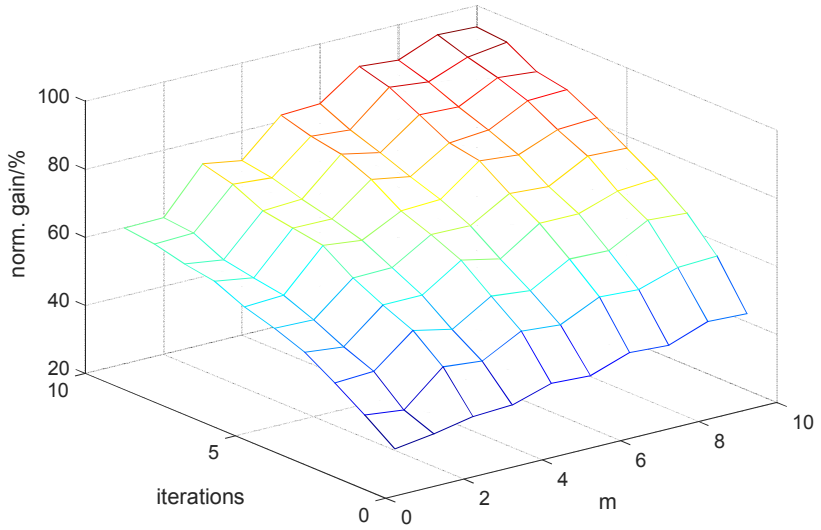


Figure 5-36: The normalized achieved gain in dependency of the number of the iterations and the number of evaluations per iteration (2 paths and 3 initial evaluations).

Figure 5-36 shows that with 3 initial beam-search values only a very bad performance of 10% of the possible gain is achieved. By running the algorithm with some iterations and evaluations, this performance can be increased continuously.

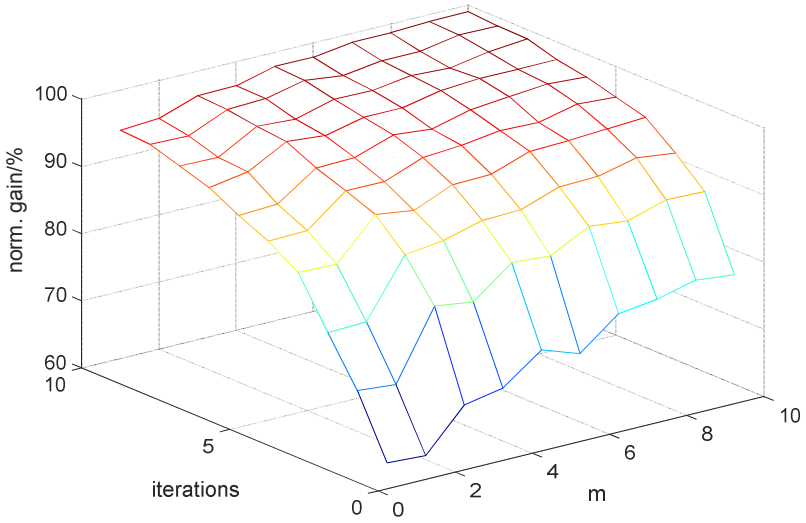


Figure 5-37: The normalized achieved gain in dependency of the number of the iterations and the number of evaluations per iteration (2 paths and 14 initial evaluations).

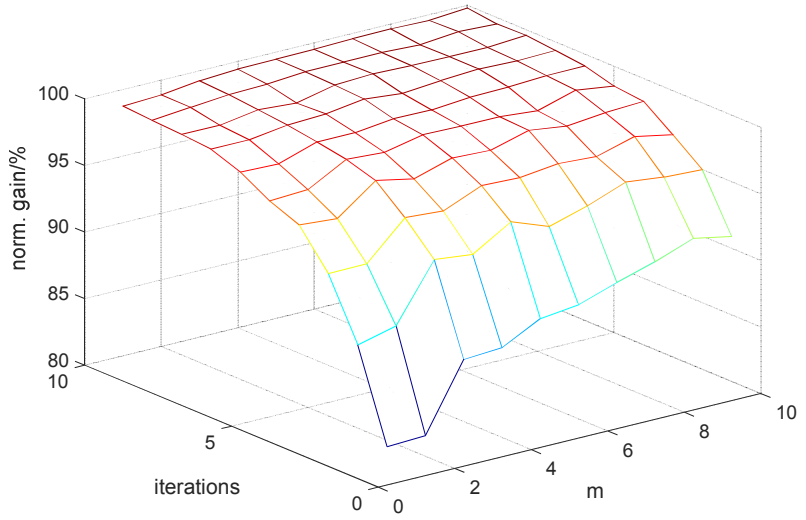


Figure 5-38: The normalized achieved gain in dependency of the number of the iterations and the number of evaluations per iteration. (2 paths and 24 initial evaluations).

Raising the initial evaluations to 14 or 24, shows, that there is a saturation reached very soon and the performance doesn't increase relevant after 4 iterations. Good performance of ca. 94% normalized gain can be achieved in this case (without noise). The total number of evaluations is calculated by adding the number of initial evaluations to the number of expensive points per iterations times the number of iterations

$$N_{eval} = N_{init\_eval} + N_{exp\_points} \cdot N_{it} \qquad (10)$$

According to these figures and considering the fact that initial evaluations are "cheaper" in terms of time consumption than increasing the number of iterations, where feedback is needed, the normal working point is chosen to be $N_{init\_eval} = 24$, $N_{\exp\_points} = 4$, $N_{it} = 4$ and thus a total number of evaluations of $N_{eval} = 40$ using the standard values in Table 5-8. This corresponds to a fraction of the

$B_{tx} \cdot B_{rx} = 256$ evaluations, which are needed if the exhaustive search algorithm is used. It should be noted, that, unlike as in exhaustive search, $m = 4$ feedback signals are needed.

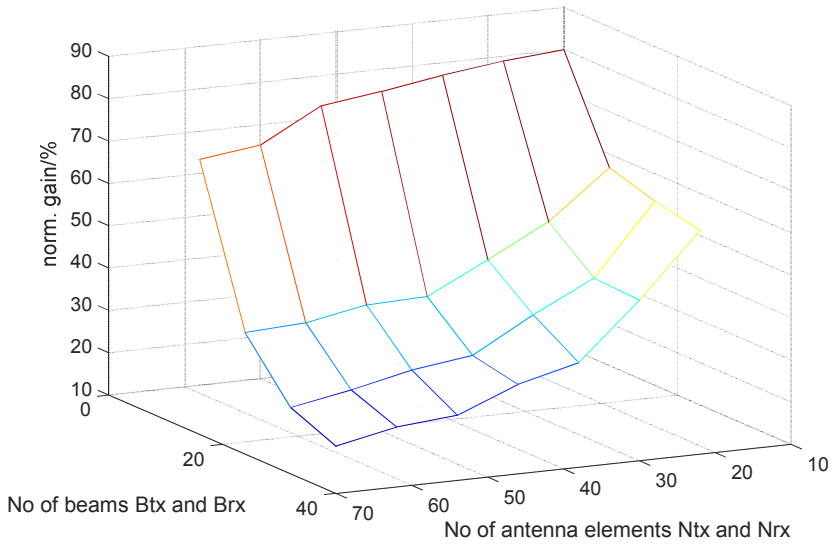


Figure 5-39: The influence of the number of beams vs the number of antenna elements.

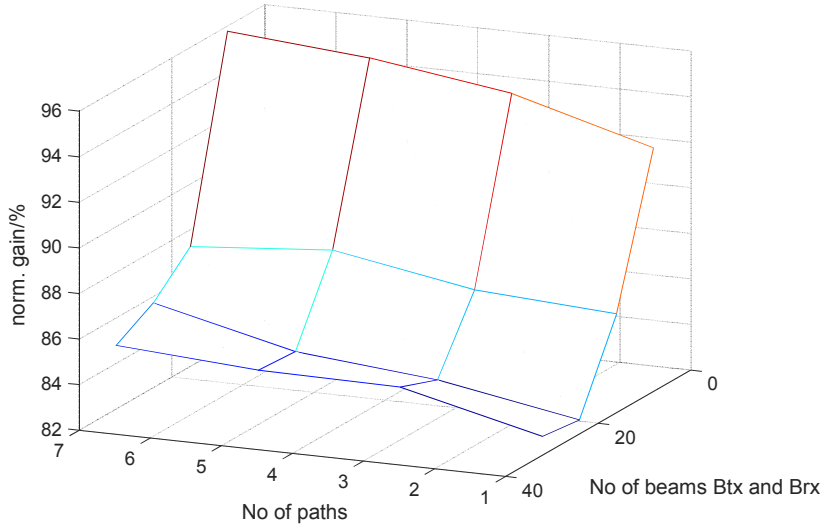


Figure 5-40: The influence of the number of beams and the number of paths.

Another influence we have to consider is the number of paths of the channel, the number of the antenna elements, which has some influence to the beam width and the number of beams in the allowed range. Figure 5-39 shows the influence of the number of element antennas and the number of beams to the performance of the system. As can be seen, the algorithm works better if there are moderate number of antenna elements and beams. Increasing the number of antenna elements decreases the beam width and thus the correlation between the neighboring points of the function, which is needed in order to estimate the value of a point from the neighboring point. Increasing the number of beams will accordingly increase the number of valid points of the function and thus raise the need of evaluations in order to get the best solution. Figure 5-40 shows, that increasing the number of paths also increases the performance. This shows that the algorithm is able to detect the best solution among several extrema.

5.5.5.2 Performance of the algorithm with noise

In order to test the algorithms performance in reality, the influence of the noise has to be taken into account. Consequently the measure of the impact of the noise is the signal to noise ratio (SNR). Using the parameters described above, the performance measure is the normalized gain, as also above or the ergodic capacity in [bit/s/Hz], defined as

$$C = \log_2\left(1 + \frac{P}{N_{tx}}(\mathbf{w}^H\mathbf{w})^{-1}\mathbf{w}^H\mathbf{H}\mathbf{f}\,\mathrm{f}^H\mathbf{H}^H\mathbf{w}\right) \quad (11)$$

The values of the SNR ∈ (−40dB...20dB) corresponding to a distance of d ∈ (0.23m...165m) if the RF output power is considered to be 5dBm. **Figure 5-41** shows the influence of the noise to the capacity. It shows that the algorithm proposed here provides almost the same capacity as the exhaustive search solution, which chooses the best possible beam combination, while using only a fraction of the evaluations and corresponding time. **Figure 5-42** shows the normalized gain as described above dependent from the SNR, by using the parameter derived in the previous section. The algorithm converges at -20dB to its maximum value of 90% of the normalized gain for all cases. Following from the simulation results, one can follow that the number of paths, i.e. the number of scatterers, in the channel has hardly no influence of the performance of the system. This shows that the algorithm is able to deal with multiple paths or scatterers, which results in different extrema in the cost function.

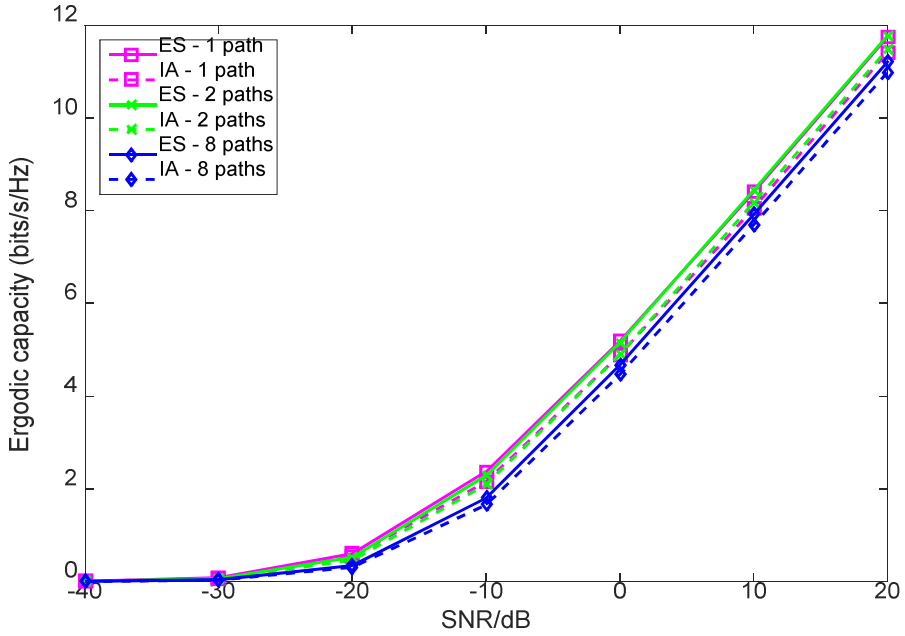


Figure 5-41: Ergodic capacity for channels with 1, 2, and 8 scatterers (multipath components) in dependency of the SNR. Monte-Carlo simulation of 2000 channel realizations per SNR point.

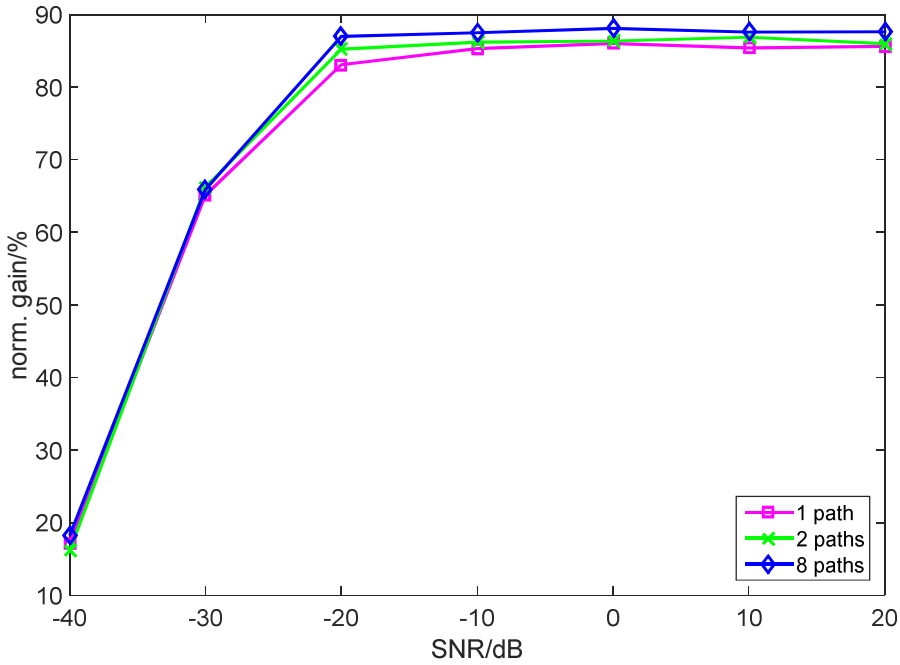


Figure 5-42: Mean of the normalized gain for several scatterers in dependency of the SNR. Monte-Carlo simulation of 2000 channel realizations per SNR point.

6. Conclusions

This document describes algorithms and networking functions for heterogeneous networks using millimetre-wave wireless communications for mobile access in small cells and backhauling between small-cells access points and macro base-stations. The overall heterogeneous network design is described in Section 2. The proposed generic HetNet structure can support all the use cases, in particular several deployment examples are provided to demonstrate the network dimensioning. Multi-connectivity with control and data plane split is presented in Section 2.2 to address the roles of macro cells and mmW small cells. The relevant challenges in HetNet comprised of mmW small cells are analysed to motivate the required solutions, some of which are provided in this document. The important functionalities of a multi-hop backhaul network with a mesh structure are studied in Section 2.3 by taking into account the relevant challenges and design objectives, especially the backhaul network initial establishment and scheduling/routing issues are identified as the main topics to be investigated in this document. MiWaveS network concept also has special features that improve security and therefore make these networks suitable for emergency communications. These results are promising but are preliminary findings with specific assumptions and require further study and measurements to verify.

As discussed in Section 3, the RRM for HetNet mmW small cells can be mathematically modelled as an optimization problem with some constraints. Decomposition techniques facilitate decoupling the RRM optimization problem across several sub-problems. Each sub-problem can be associated with certain functionality in a specific layer or network entity. We outlined two possible decomposition alternatives, i.e. node-centric and path-centric, and discussed the pros and cons of each method from an implementation perspective. Furthermore, we developed practical algorithms based on this vision to address the identified challenges and provided various simulation analyses. In addition, a joint beam-frequency multiuser rate scheduling algorithm for mmW downlink system has been developed. Two algorithms including the optimal algorithm and the reduced complexity algorithm are presented. It is shown that the reduced complexity algorithm can achieve the same scheduling metric with considerable reduced complexity.

We have proposed a new self-organized link establishment procedure for a multi-hop inband AP relay backhaul in Section 4.1. The proposed procedure reuses the existing LTE relay initial attachment framework, on top of which the carrier aggregation technique is further employed to configure the mmW backhaul link as the secondary link of the respective connection. The proposed new mmW small cell discovery signals enable the mmW link detection and configuration. Moreover, it also facilitates the downlink beam alignment operation via the established RRC connection operating in the conventional broadband spectrum. It should also be noted that the multiple mmW BH link configuration to a single mmW small cell node furnishes the capability of dynamic BH routing.

In Section 4.2, we have described a dedicated standalone mmW backhaul which builds up autonomously without assuming wireless wide area signalling on the control plane (as opposed to the solution in Section 4.1). The standalone backhaul is a transparent sub-network that offers first-mile connectivity between small-cell APs and macro BS. The proposed solution provides self-organization, self-optimization and self-healing capabilities, by means of fast protection and restoration and QoS-aware congestion management with load balancing. Overall, the adaptation to dynamically changing traffic demands and link capacity changes is managed by the flexibility in routing. On the other hand, transmission opportunities are proposed to be allocated to the mmW links semi-permanently by

applying a pre-computed cyclic schedule, with dynamic dTDD. The routing and link scheduling schemes need to be further developed in order to support sectorised antenna architectures with multiple RF chains in the nodes.

In Section 4.3, we proposed a low complexity channel estimation algorithm based on an enhanced one-sided search for hybrid beamforming with a high level of accuracy, limited power and limited number of measurements. We additionally analysed the performance of analogue backhaul beam switching in different possible options. As discussed in low number of MPCs, beam-switching with limited steering range can be a viable option for the backhaul. However, for scenarios with richer channel characteristics, a finer resolution of beam-switching in spatial domain or utilizing hybrid strategies is recommended.

Based on the existing LTE RACH procedure, a macro-cell assisted mmW UE random access method has been introduced in Section 5.1. The proposed random access method employs a new RACH channel consisting of multiple preamble sequences. The RACH response message from mmW small cell signals the preferred uplink transmit beam direction. As a result, the uplink beam alignment is achieved as the complementary to downlink beam alignment setup by the method in Section 4.1.

By deploying IP-based middleware and interchanging information between users and the network at control plane, it is possible to trigger network-centralized decisions so as to balance the traffic and allocate users at the best BS at each moment. Given the background explained in section 5.2, it is possible to foresee high potential gains in terms of overall QoS of the network. We simulated a system where a 24-hour load period is generated and tested over several possible scenarios: baseline, intermediate and complex policies combined with just macro cell and mmW cells inside macro cells in conservative or realistic scenarios in terms of cell coverage.

Section 5.3 describes with an example how ultra-high speed mmW network layer in heterogeneous network can contribute to energy savings. In many parts of the topology network elements can spend most of the time in low-power sleep mode. Pursuing energy efficient performance of mmW small cells is a key objective to build green networks. As small cells are very prone to present low occupancy periods, it is important to study and assess the best layout and policy to apply at each moment. Some potential situations are described in Section 5.3, modelling the network organization, traffic load and cell layouts. A complete study evaluating the potential and theoretical savings of applying multiple combinations of the aforementioned parameters is carried out to provide advice on the best way to deploy and manage these mmW small cells.

The proposed method in Section 5.4 of two stage cell detection and measurement reduces the initial cell search burden and feedback overhead significantly. The proposed beam cell aggregation approach achieves the mmW downlink beam alignment, without defining specific beam alignment protocols. In addition, the channel state information feedback and beam cell (re)configuration are performed by the control plane signalling in macro cell which is more robust and power efficient. The mmW downlink CoMP operation can be also supported in the format of a joint beam cell allocation or a dynamic beam cell selection, determined by the macro eNB. Moreover, the beam cell reconfiguration based on periodically reported receive quality of configured BRSs during the movement of UEs, supports the UE mobility.

The proposed beam alignment algorithm in Section 5.5 for access links can be implemented within a system suffering from limited hardware constraints. The algorithms show significant time reduction of the alignment period compared to exhaustive search.

7. References

[1] 3GPP TR 36.842 V12.0.0 (2014-01) 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Study on Small Cell enhancements for E-UTRA and E-UTRAN; Higher layer aspects; (Release 12).

[2] NGMN Alliance, "NGMN 5G White Paper," 2015.

[3] S. Singh, R. Mudumbai and U. Madhow, "Interference Analysis for Highly Directional 60-GHz Mesh Networks: The Case for Rethinking Medium Access Control," *IEEE/ACM Transactions on Networking*, vol. 19, no 5, pp. 1513-1527, 2011.

[4] L. Chen, S. Low, M. Chiang and J. Doyle, "Cross-Layer Congestion Control, Routing and Scheduling Design in Ad Hoc Wireless Networks," *IEEE International Conference on Computer Communications*, 2006.

[5] X. Lin and N. B. Shroff, "Joint rate control and scheduling in multihop wireless networks," *IEEE Conference on Decision and Control*, 2004.

[6] M. Kyrö, et. al., "Experimental propagation channel characterization of mm-wave radio links in urban scenarios," *IEEE Antennas and Wireless Propagation Letters*, Vol. 11, 2012.

[7] M. R. Akdeniz, Y. Liu, M. Samimi, S. Sun, S. Ragan, T. S. Rappaport and E. Erkip, "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE Journal on Selected Area in Communication*, vol. 32, no 6, pp. 1164 - 1179, 2014.

[8] IEEE Std 802.11ad, Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band, 2012.

[9] 3GPP TS 36.300 V10.0.0 (2010-06) 3rd Generation Partnership Project;Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 10).

[10] T.A. Thomas, et al., "3D mmWave Channel Model Proposal," *IEEE Vehicular Technology Conference Fall*, 2014, pp. 1 - 6.

[11] Y.M. Tsang, A.S.Y Poon, S. Addepalli, "Coding the Beams: Improving Beamforming Training in mmWave Communication System," *IEEE Global Telecommunications Conference*, pp. 1-6, December 2011.

[12] J. McGarrity, J. Soraghan, T. Durrani, and S. Mayrargue, "The ESPRIT and MUSIC algorithms using the covariance matrix," *International Conference on Acoustics, Speech, and Signal Processing*, Apr 1991, pp. 3285–3288, vol.5.

[13] A. H. Sayed, "Adaptive Filters", John Wiley & Sons, 2008.

[14] A. Alkhateeb, Jianhua Mo, N. Gonzalez-Prelcic and R.W. Heath, "MIMO Precoding and Combining Solutions for Millimeter-Wave Systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122-131, December 2014.

[15] B. Sharif, G. G. Wang, and T. Y. ElMekkawy, "Mode pursuing sampling method for discrete variable optimization on expensive black-box functions," *Journal of Mechanical Design*, vol. 130, 2008.

[16] 3GPP TS 36.211 V12.6.0 (2015-07) 3rd Generation Partnership Project;Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved

Universal Terrestrial Radio Access Network (E-UTRAN); Physical channels and modulation; (Release 12).

[17] FP7-ICT EARTH project (247733), “D2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown”, 2011.

[18] J.M. Correira, D. Zeller, O. Blum, D. Ferling, Y. Jading, I. Godor, G. Auer, L. Van der Perre, “Challenges and Enabling Technologies for Energy Aware Mobile Radio Networks”, *IEEE Communications Mag.*, vol. 48, no. 11, Nov. 2010, pp. 66–72.

A. Annex: Link budgets

This Annex, and **Table A-1**, is provided as a collection of link budgets used to evaluate the feasibility of MiWaveS use cases and achievable capacities vs. link distances. Grey fields are results of an equation while white ones are input parameters. 'Channels' indicate how many concatenated or aggregated channel or carriers are assumed. 'Bandwidth' in the (noise-) bandwidth of each channel. 'Number of parallel amplifiers' indicates how many PAs is assumed in TRX architecture, and it is a multiplier for EIRP. Antenna size gives an estimate of the antenna size if it would be of circular or rectangular form. In path loss calculation, a heavy rain (tropical) about 100 mm/h rain rate zone is assumed.

AV1...5 gives link budgets for mobile access (here 'A') link (downlink and uplink) implemented with 60 GHz (here 'V') TRX. AP antenna is of modest gain while UE antenna has low gain. BV1...2 gives two examples of V-band backhaul links with dedicated hardware with much higher capacity and larger antennas.

'BH-BH' columns refer to links between dedicated backhaul radios and 'AP-AP' refer to inband backhaul cases. BE1...2 gives link budget examples for dedicated 80 GHz band (here 'E') backhaul (here 'B') link with strongest performance in hop length and capacity. Integrated E-band case BE3...4 utilise smaller antennas and larger amount of parallel transmitters.

Link margin LOS indicates how much stronger than required the signal is in ideal line-of-sight case. Margin NON-LOS assumes that in severely obstructed case the signal is not totally blocked but additional free-space loss attenuation is 20 dB/decade i.e. total attenuation 40 dB/decade. OBS-LOS (Obstructed LOS) is in between.

Table A-1: MiWaveS' link budgets

MiWaveS link budget

	BH - BH				AP - AP		AP - UE UPLINK		AP - UE DOWNLINK		
	BV1	BV2	BE1	BE2	BV3	BV4	AV1	AV2	AV3	AV4	AV5
Modulation	16QAM	64QAM	128QAM	16QAM	16QAM	64QAM	QPSK	BPSK	16QAM	16QAM	QPSK
Transmitter											
Carriers	8	8	8	8	8	8	1	1	2	1	1
Bandwidth, noise (GHz)	1,76	1,76	2	2	1,76	1,76	0,22	0,22	0,44	0,22	0,22
P1dB at PA output (dBm)	15	18	22	22	15	15	12	10	15	15	14
PA output back-off (dB)	2	4	8	4	4	7	2	0	2	2	4
Peak-Average ratio (dB)	3	4	4	3	3	4	0	0	3	3	0
Peak power at PA output (dBm)	13	14	14	18	11	8	10	10	13	13	10
Average power at PA output (dBm)	10	10	10	15	8	4	10	10	10	10	10
PA-antenna interconnection loss (dB)	2	2	2	2	2	2	2	2	3	3	3
Emitted mean power (dBm)	8	8	8	13	6	2	8	8	7	7	7
Number of parallel Power Amplifiers	1	1	1	1	8	8	1	1	6	6	6
TX Antenna	Lens2	Lens2	Lens3	Lens3	PlaneAP1	PlaneAP2	PlaneUE1	PlaneUE1	BTSSect2	BTSSect2	BTSSect2
Antenna gain (dBi)	30	30	38	38	15	24	5	5	16	16	16
Antenna efficiency	0,7	0,7	0,7	0,7	0,36	0,36	0,36	0,36	0,7	0,7	0,7
Antenna size, side length (mm)	53	53	99	99	13	37	4	4	11	11	11
Antenna size, diameter (mm)	68	68	126	126	17	47	5	5	14	14	14
Half-power beam width (deg)	6,4	6,4	2,5	2,5	36,0	12,8	113,9	113,9	32,1	32,1	32,1
EIRP (dBm)	38,0	38,0	46,0	51,0	30,0	35,0	13,0	13,0	30,8	30,8	30,8
Channel											
Distance (m)	170	100	350	70	20	60	10	18	10	18	25
Attenuation coefficient	3,4	2,0	7,0	1,4	0,4	1,2	0,2	0,4	0,2	0,4	0,5
Carrier frequency (GHz)	60	60	81	81	60	60	60	60	60	60	60
Gas attenuation (dB)	2,55	1,5	0,175	0,035	0,01	0,03	0,005	0,009	0,15	0,27	0,0125
Availability (time-%)	99,9	99,9	99,9	99,9	99,9	99,9	99,9	99,9	99,9	99,9	99,9
Rain attenuation, tropical (dB)	2,04	1,2	4,2	0,84	0,24	0,72	0,12	0,216	0,12	0,216	0,3
Path loss LOS (dB)	119,9	113,5	128,4	111,2	97,4	107,2	91,4	96,5	91,6	96,7	99,4
Path loss NON-LOS (dB)	157,8	147,9	171,0	143,3	121,3	138,3	110,7	119,7	110,9	119,9	124,7
RX Antenna	Lens2	Lens2	Lens3	Lens3	PlaneAP1	PlaneAP2	BTSSect2	BTSSect2	PlaneUE2	PlaneUE2	PlaneUE2
Antenna gain	30	30	38	38	15	24	16	16	8	8	8
Antenna size, side length (mm)	53	53	99	99	13	37	15	15	4	4	4
Receiver											
LNA-antenna interconnection loss (dB)	2	2	2	2	2	2	3	3	3	3	3
Power at LNA input LOS (dBm)	-53,9	-47,5	-46,4	-24,2	-54,4	-50,2	-65,4	-70,5	-55,8	-61,0	-63,6
Power at LNA input NON-LOS (dBm)	-91,8	-81,9	-89,0	-56,3	-78,3	-81,3	-84,7	-93,7	-75,1	-84,2	-88,9
Signal to Noise Ratio											
Thermal Noise (dBm)	-73,5	-73,5	-73,0	-73,0	-73,5	-73,5	-82,6	-82,6	-79,6	-82,6	-82,6
Implementation loss (dB)	3	5	3	3	3	3	2	1	3	3	2
Interference degradation (dB)	0	0	0	0	0	0	1	1	1	1	1
Noise figure (dB)	8	8	8	8	8	8	8	8	8	8	8
SNR at ADC input LOS (dB)	16,7	21,1	23,6	45,7	16,2	20,4	14,2	10,1	19,8	17,6	16,0
SNR at ADC input NON-LOS (dB)	-21,2	-13,4	-19,0	13,6	-7,7	-10,7	-5,2	-13,1	0,5	-5,6	-9,3
Signal processing											
Required SNR for BER=1e-5 (dB)	14,5	19	22	14,5	14,5	19	9,5	8	14,5	14,5	9,5
Margin LOS (dB)	2,2	2,1	1,6	31,2	1,7	1,4	4,7	2,1	5,3	3,1	6,5
Margin NON-LOS (dB)	-35,7	-32,4	-41,0	-0,9	-22,2	-29,7	-14,7	-21,1	-14,0	-20,1	-18,8
Margin Average/OBS-LOS (dB)	-16,8	-15,2	-19,7	15,2	-10,3	-14,2	-5,0	-9,5	-4,4	-8,5	-6,2
Data rate, FEC (Mbps)	5280	7920	10500	6000	5280	7920	330	176	1320	660	330

B. Annex: Common reference block terminology

This Annex presents an agreed-upon terminology for the reference blocks, wireless network elements and the main components used in all MiWaveS project deliverables.

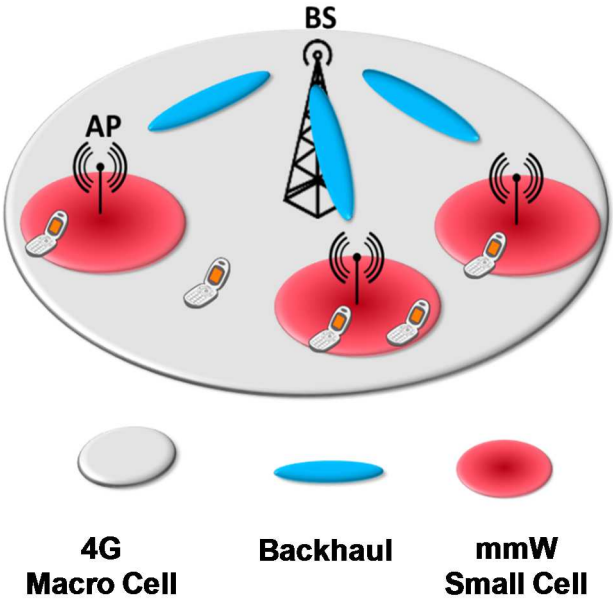


Figure B-1: MiWaveS reference architecture.

To ensure a unified understanding of the different wireless network elements, the main components and the terminology are defined and described in the following, focusing on their main purpose in the transmission chain.

- **User Equipment (UE) or User Terminal (UT)**: these are the devices (e.g. smartphones, tablets, laptops) handled by the users in the system, enabling them to connect to the network and establish voice and data sessions. As an important assumption of the MiWaveS project, these user terminals shall enable mmW connectivity so as to be able to connect to mmW access points. In addition, they will also support LTE connectivity, as well as some other optional interfaces to connect to the rest of the elements in the heterogeneous network.

  The user terminal represents the end of the data link from an application point of view. Its main characteristics are:

  - Low transmit power;
  - Low to medium antenna directivity and gain;
  - Small form factor, i.e. portable or handheld;
  - Highly optimized integration of processing modules;
  - Many user applications running in parallel.
- **Access link**: this term refers exclusively to the link used by the UE to communicate with Access Points (excluding the UE to BS link which could also be considered as access link) and, thus, providing them access to the network. MiWaveS proposes using mmW frequencies for handling these access links. Its main characteristics are:
  - Focus: Flexible multi-user radio access with reduced interference and capacity fairness;

  - Wireless connection in the mmW frequency range;
  - Static and mobile, indoor and outdoor environments;
  - Line-of-sight and non line-of-sight connections.
- **Access point (AP)**: It is a low-power radio access nodes used to provide extra coverage or higher data rates in localized areas. It is small in comparison to a conventional macro base-station, able to provide coverage up to a few tens of kilometres. MiWaveS' scope is to provide a definition of and support for mmW access points, able to help operators to support the forecast growth in mobile data traffic. Its main characteristics are:
  - Range from tens of metres up to a few hundred metres;
  - Statically mounted, indoor or outdoor;
  - Deployed by a network operator or a user;
  - Low to medium transmit power from tens of milliwatts to one watt;
  - Possible relay data to another AP;
  - With or without radio access control;
  - Restricted size, e.g. small boxes, lampposts;
  - Medium to high-gain antennas with beam-steering.
- **Backhaul (BH) link**: if we consider a telecommunications network deployed with the usual hierarchy, the backhaul link can be defined as all intermediate communication links between the core network and the small subnetworks (access points) at the "edge" of the aforementioned hierarchical network. MiWaveS proposes a set of access points linked together with mmW wireless backhaul.

  The main characteristics of backhaul links are:
  - Focus: High throughput to deliver data for multiple users; a peak capacity of 10 Gbps is envisioned;
  - Wireless connection in the mmW frequency range;
  - Static environment (no mobility);
  - Line-of-sight connection;
  - High-directivity antennas at transmitter and receiver;
  - Steerable directive antennas may be considered for automatic beam alignment or backhaul reconfiguration.
- **Base station (BS)**: it manages the low-level operation of all user terminals connected, by sending them signalling messages such as handover commands. Each base station connects with the core network and it can also be connected to nearby base stations, mainly for signalling and packet forwarding during handover.

  The base station represents the anchor point to the core network and is responsible for managing the radio access and the successful transmission of the data coming from the core network to the end user. Its main characteristics are:
  - Range: up to 10 km;
  - Outdoor placement;
  - Statically mounted;
  - Mounted by professional;
  - High transmit power;
  - Multi-user radio access and resource control of macro cell;
  - Interference coordination;

  - Rack size for signal processing;
  - Available space for large antennas;
  - Many parallel control tasks.
- **Core network**: This part of the telecommunications system is not affected by MiWaveS' investigation, so it will remain with the same functionalities and elements as in 4G networks.
- **Macro Cell**: a macro cell provides the largest area of coverage within a mobile network, having typical coverage radio up to 10 km. It corresponds to the coverage of a BS.
- **Small Cell**: a small cell provides improved cellular coverage, capacity and applications for homes and enterprises as well as metropolitan and rural public spaces. It corresponds to the coverage of an AP and has a typical range from 10 metres to several hundred metres.
- **Relay**: A relay in mmW enhanced mobile broadband network is characterized by using mmW spectrum and transmission technology to connect the relay with another AP.
- **Beamforming (BF):** action of fitting the radiation pattern to a specified radiation mask.
  - **Digital beamforming**: In principle, each antenna element is connected to one RF chain. Beam forming or/and beam steering is done in digital baseband domain, therefore more degrees of freedom are existing compared to analogue domain (phase shifts and weights). Because each antenna element transmits/receives one digital signal, it is possible to send several partly correlated (time diversity) or uncorrelated streams with these antennas (spatial multiplexing), which results in either better robustness or higher data rate.
  - **Hybrid beamforming**: This is a mixture of analogue beamforming and digital beamforming for which number of antenna elements > number of RF chains > 1. Consequently, the flexibility and achievable performance is in between these methods.
  - **Analogue beamforming**: One RF chain is existing. The steering of the beam is done by analogue components like phase shifters or delay lines (e.g. Butler matrices). In most cases, only the phase of one antenna element can be adjusted, not the weights.

C. Annex: MiWaveS use cases

In order to base the research activities of the MiWaveS project on solid foundations, five use cases has been defined:

1. **Urban street-level outdoor mobile access and backhaul system**, in which 1000-times higher spatial data consumption is expected by 2020. Users expect to have multi-Gigabit low-latency connections to services almost anywhere,

2. **Large public events, covering massive crowd gatherings**, sports events or vacation resorts. A great amount of users using data-hungry applications are served by the network, but just in some specific periods and in small areas,

3. **Indoor wireless networking and coverage from outdoor**, including the increase of indoor networks capacity and versatility, using indoor or outdoor antennas, and connecting to the operator network by quasi-fixed links,

4. **Rural detached small-cell zones and villages**, using mmW wireless backhaul technologies standalone or combined with wired line connection, to overcome the deployment difficulties of wireline backhaul installations,

5. **Hotspot in shopping malls**, considering ad-hoc deployment of small cells and mmW backhaul as a cost efficient solution to enable high data rate services inside the malls.

These use cases will provide a common background to group the technical solutions proposed by the project partners, and will allow to compare and evaluate such solutions in a very efficient way.

The following **Table C-1** contains the common KPIs defined for all MiWaveS' use cases.

Table C-1: Common KPIs for MiWaveS' Use Cases

#	KPI	Definition
1	Energy Efficiency (EE)/power consumption	An estimation of the EE savings shall be done against the legacy or against an always-on scenario, preferably by using EARTH [17] or OPERA-NET [18] traffic and power profiles, and referring to ETSI EE standards.
2	End user capacity	Capacity offered to the user within the service area: 1. Average capacity everywhere in the service area; 2. Peak capacity in specific places. Strictly related to the KPI #12 on QoS/QoE.
3	Reliability (service/backhaul)	An estimation of the reliability shall be done against the legacy network implementation. 1. Service: out of service rate. 2. Backhaul: outage rate of the BH link. BH outage rate affects the service availability for all the UEs served by the associated small cells. It is envisioned that the BH link can consist of several relay nodes, i.e., multi-hop relays based BH link; in such a case, the reliability of the BH link refers to the overall outage rate of the end-to-end link aggregating all intermediate links.
4	System Coverage	Ability to support the services with defined QoS/QoE thresholds.

5	Network adaptation versatility	Automatic adaptation of the network to the arrival or clearing of a node.
6	Area Throughput/ System Capacity	1. Total maximum system capacity and expected maximum number of connected users per area. 2. Overall offered throughput in the service area. This KPI should be evaluated by considering full buffer traffic.
7	Backhaul range	The maximum distance between two adjacent nodes in the backhaul link for achieving the required overall backhaul reliability and capacity.
8	Efficiency of installation and in operation	Support of cost-efficient roll-out and operation of the small cell networks (WP1 O1.2).
9	HetNet capability/ Dual Connectivity	In addition to the seamless hand-overs between mmW AP´s, interworking between access layers: 1. Between macro-cell access and mmW small-cell access. 2. Between mmW small-cell access and other small-cell RATs, e.g. WiFi 2/5 GHz
10	QoS and QoE	Capability to provide capacity to selected users, services and geographical areas. Suitable capacity to support the services requested by the UEs.
11	Cost efficiency	Optimization of network deployment costs.

Note: Exposure to EM fields as well as other node regulatory-type approval specifications are not considered as system KPIs.

UC1: Urban street-level outdoor mobile access and backhaul system

Characterized by:

- 1000-times higher data volume per area has been envisioned by year 2020.
- End users expect to have high capacity seamless connections to wireless services almost anywhere, especially in urban street areas.
- Area-specific mobile network use and demands depend e.g. on the following variables: time of day, day of the week and seasonal differences.
- In addition to human users, thousands of sensors, detectors and small Machine Type Communication (MTC) equipment at the street level are expected to be connected to the network.

Key assumptions and technical challenges of UC1:

- Network topology: mesh to increase network reliability, chain to relay backhaul capacity to further APs, star topologies will co-exist.
- Type/Size of small cells: In hot-spots/-zones, small cells form a continuous coverage, uniform capacity is preferred.
- AP density: AP at every light pole (ca. 30–50m), depending on the link budget , etc.
- Mobility: stationary or low mobility users.

Figure C-1: Urban street-canyon mmW small cell access and backhaul scenario.

UC2: Massive public events and gatherings

Characterized by:

- Massive crowds gathered for some periods of time in small areas, e.g. for public or sport events.
- Three main types of events primarily considered:
    - "Massive crowd": big events held in locations where the network infrastructure not sized for such gathering, e.g. Assisi during Pope's visit;
    - "Sport event": demand for continuous broadband coverage along the route even if not at the same time, e.g. Tour de France or Turin Marathon;
    - "Vacation resort": huge user densities are concentrated in particular locations, e.g. ski resorts in winter, beaches during summer time.

Key assumptions and technical challenges of UC2:

- Network topology: reconfigurable backhauling features with steering capability.
- Type/Size of small cells: small cells able to connect to the self-configurable backhaul link.
- Backhaul and backhaul re-configurability: ability to follow the traffic in space and time distribution.

Figure C-2: Covering hot spots with small cells in typical mass event environments.

UC3: Indoor wireless networking and coverage from outdoor

Characterized by:

- Evolution of the network in domestic and professional environment toward an increase of transmission capacity and versatility to connect to the network.
- Connection to the external network can be achieved by installing an antenna inside the room (maybe near a window), or outside the house (like current satellite antennas).
- Connected to the operator network by quasi-fixed links activated when needed, e.g.
    - Replacing the current Asymmetric Digital Subscriber Line (ADSL) by offering a larger capacity and more use versatility for different applications;
    - To cover zones where fixed lines or fibre are not deployed (residential area, villages, city in developing countries, etc.).

Key assumptions and technical challenges of UC3:

- User antenna inside the building: no need for additional infrastructures from the user part, but link budget may not allow sufficient transfer capacity.
- User antenna outside the building: antenna can serve also as a relay of a mesh network.

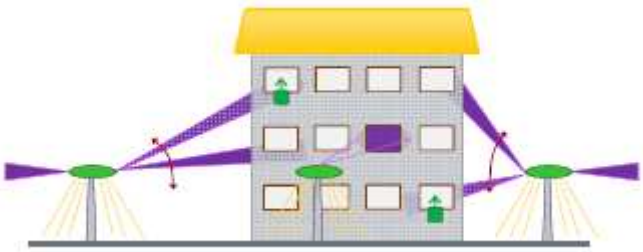


Figure C-3: Providing indoor coverage from outdoors.

UC4: Rural detached small-cell zones and villages

Characterized by:

- Wired line based infrastructure installations (fibre of copper) may be difficult or not feasible, therefore these backhaul traffic demands can be better covered by mmW wireless backhaul.
- There are two basic use cases:
    - Cells are being connected with mmW only;
    - Cells have already a wired line connection, but with limited bandwidth and mmW are used to enhance the backhaul bandwidth.
- A key element of backhaul provision to remote cells is the use of relays, i.e. a chain of relays which finally terminates at a cell (AP).

Key assumptions and technical challenges of UC4:

- Network topology: to support the detached remote cells consists of mmW band multi-hop connections. Hop length estimate 200-800 m.
- Type of cells: remote mobile broadband cells for traffic spots, mmW user access on relay and end points.

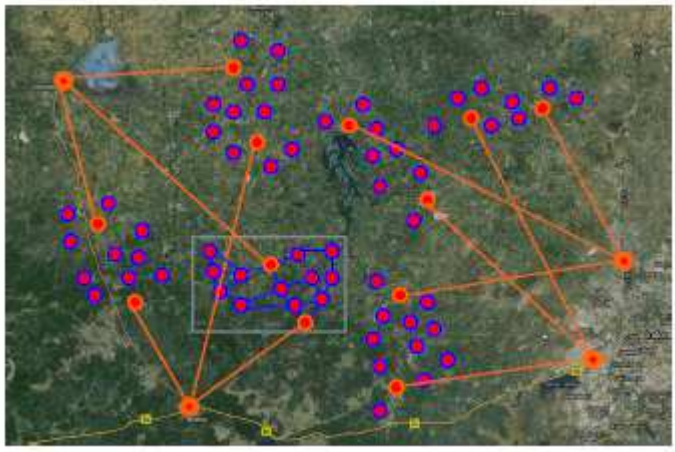


Figure C-4: Rural village use case, orange dots: relays or APs, purple dots: UEs.

UC5: Hotspot in shopping malls

Characterized by:

- Ad-hoc deployment of small cells is an efficient solution to cope with the high data rate traffic services that originate indoors
- mmW backhaul may be preferred as a cost-efficient solution to enable high data rate services inside the malls.
- In the shopping malls, small cells can be deployed either by the shopping mall owner in a semi-planned fashion or by the owners of the shops, e.g. to provide additional services (such as internet connection or advertising)

Key assumptions and technical challenges of UC5:

- Shopping centre will likely be equipped with pico-cells operating at lower frequency bands to provide coverage and minimum data rate services.
- In this deployment scenario, additional small cell nodes are deployed under the coverage of overlaid pico-cells in order to boost the capacity.

- At each shop, one or more small cells can be deployed depending on the used technology, required service, and coverage.Backhaul and backhaul reconfigurability: ability to follow the traffic in space and time distribution.

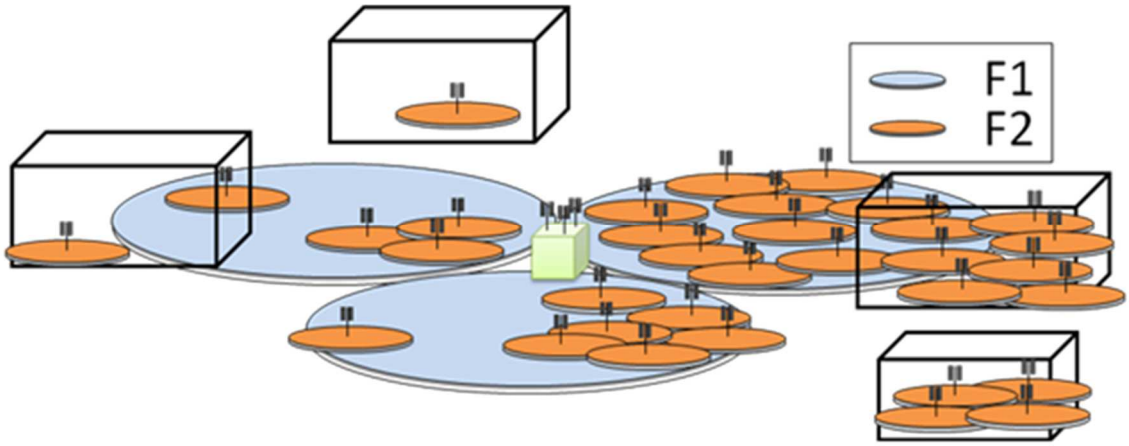


Figure C-5: Deployment scenarios of small cell with/without macro coverage [6].

Use-cases wrap-up:

The defined use cases cover most of the typical environments where small-cell heterogeneous millimetre wave mobile networks (i.e. backhaul and access nodes) will be deployed and where mobile services are used. Even though the proposed environments are very different from one another, they also present some similarities:

- In some cases, maximising mmW backhaul range may be an issue; but, in most cases, short hops (achievable with reasonable antenna sizes, output powers and energy consumption) are enough.
- End-user capacity and total maximum capacity per area: not only high peak and average access or backhaul link capacities matters, but even more important is the area capacity. Data rates determine the ultimate user experience like download durations, seamless operation and feedback time in gaming.
- There must also be quality of service and/or CEM mechanisms to provide capacity to selected users, services and geographical areas in case priority is needed.
- Spectral efficiency must be considered when starting to exploit mmW bands. There is a lot of spectrum available, but frequency regulation's general requirement of interference free and efficient use of spectrum applies also here.
- Ease of small-cell/ultra-dense network planning and deployment is essential for cost-efficiency in many cases. Also, operation, management and configuration changes in high-density networks must be easy and quick to perform, and with minimized service outages.
- Because of challenges in mmW propagation, the backhaul reliability is seen as one KPI.
- Power consumption reduction compared to legacy is an obvious requirement today for all installations. There can be different indicators for energy consumption like consumption per bit, per network element or for the whole network.

The actual usage, i.e. applications requiring mobility and high data rates, would still need more attention to characterise the requirements of the network.
High level classification and summary of the KPIs defined for the MiWaveS use cases is presented in **Table C-2**.

Table C-2: Summary of the KPIs for MiWaveS use cases.

High-level targets	KPIs	Use cases
Capacity	End-user capacity	1, 2, 3
	Area throughput/system capacity	1, 2, 5
Service and reliability	Reliability (service/backhaul)	2, 3, 4
	System coverage	2
	Backhaul range	4
	QoS and QoE	1, 4, 5
Green radio	Energy efficiency / power consumption	1, 2, 3, 4, 5
Network flexibility / TCO reduction	Efficiency of installation and operation	1, 4
	HetNet capability	1, 3
	Network adaptation versatility	3
	Cost efficiency	3