



PROTEUS 6G

PROJECT: **PROTEUS-6G** GRAND AGREEMENT NO: **101139134**

**PROGRAMMABLE RECONFIGURABLE OPTICAL TRANSPORT FOR
EFFICIENTLY OFFERING UNCONSTRAINED SERVICES IN 6G**

INFORMATION & COMMUNICATION TECHNOLOGIES (ICT)

START DATE: 01/01/2024
END DATE: 31/12/2026
DURATION: 36 MONTHS
PROJECT URL: WWW.PROTEUS-6G.EU

Deliverable D2.4 Preliminary Network Optimization Studies

WORK PACKAGE	WP 2
TASK	TASK 2.4
DUE DATE	30/08/2025
SUBMISSION DATE	25/08/2025
DISSEMINATION LEVEL	PUBLIC
DELIVERABLE LEAD	UPAT
VERSION	3.0
AUTHORS	CHRISTOS CHRISTOFIDIS, DIMITRIS UZUNIDIS (UPAT), DAN MAROM (HUJI), JOSE RIVAS MOSCOSO, VANESSA ZANNELLA (TID), DAVID LARRABEITI, GABRIEL OTERO, CELIA LOPEZ (UC3M), ANA LARRAÑAGA, SANDRA LAGÉN (CTTC), PABLO TORRES-FERRERA (INF)
REVIEWERS	ANTONIO NAPOLI (INF) ASMA TAHIR (CNIT)



Document revision history

Version	Date	Description of change	List of contributor(s)
V0.0	11/02/2024	Initial table of contents	CTTC
V1.0	11/04/2025	UPAT has inserted the final table of contents	UPAT
V2.0	25/04/2025	UPAT, HUJI, CTTC, UC3M,TID fill its contributions	UPAT, HUJI, CTTC
V2.1	25/07/2025	First complete draft	ALL
V2.2	04/08/2025	Comments from First revision included	INF
V2.3	14/08/2025	Comments from Second revision included	CNIT
V3.0	24/08/2025	Final version released	UPAT, ALL

Disclaimer

The content of the publication herein is the sole responsibility of the publishers, and it does not necessarily represent the views expressed by the European Commission or its services.

While the information contained in the documents is believed to be accurate, the authors or any other participant in the PROTEUS-6G consortium make no warranty of any kind with regard to this material including, but not limited to the implied warranties of merchantability and fitness for a particular purpose.

Neither the PROTEUS-6G Consortium nor any of its members, their officers, employees or agents shall be responsible or liable in negligence or otherwise howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing neither the PROTEUS-6G Consortium nor any of its members, their officers, employees or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein

Copyright notice

© 2024 – 2026 The PROTEUS-6G Consortium. This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both. Reproduction is authorized provided the source is acknowledged.



Executive summary

Deliverable D2.4 presents the preliminary network optimization studies for the PROTEUS-6G architecture, addressing the transport-network challenges posed by future 6G Radio Access Network (RAN). As RANs evolve toward disaggregated and virtualized architectures, fronthaul, midhaul, and backhaul segments must meet extreme requirements for capacity, latency, flexibility, and energy efficiency. This work investigates and quantifies key design and optimization strategies through analytical modeling, real-world traffic analysis, and high-fidelity simulations.

The deliverable is structured around three main technical pillars:

1. **Reference Transport Network Designs (TID)**
Several optical transport architectures for fronthaul are analyzed, from point-to-point dedicated fibre links to advanced wavelength-division multiplexed passive optical networks (WDM-PONs). The study examines suitability for high- and low-layer functional splits, detailing trade-offs in scalability, latency, cost, and energy efficiency. Special focus is given to the PROTEUS-6G Spatially Diverse Point-to-Multipoint fronthaul node (SDPtMP), which enables efficient multi-site connectivity without wavelength contention.
2. **Machine Learning-Based Dynamic Network Optimization (University of Patras (UPAT), Hebrew University of Jerusalem (HUJI), Infinera (INF))**
Using real traffic traces from a large operator and scaling them to 6G scenarios, multiple machine learning models were trained to predict hourly traffic demands at midhaul nodes. Random Forest achieved the best accuracy (Mean Absolute Percentage Error (MAPE) = 2.29%). These predictions drive dynamic transceiver activation strategies, achieving up to 20% fewer active transceivers compared to static provisioning, with 21% total energy savings when transceivers are placed in sleep mode. Dynamic allocation also reduced over-provisioning from over 86% in semi-static cases to under 29% in the (ML-)driven approach.
3. **Flexible Functional Splits (Centre Tecnològic de Telecomunicacions de Catalunya (CTTC))**
The concept of Flexible Functional Splits (Flexible Functional Split (FFS)) is explored as a means to adapt RAN centralization to varying traffic patterns (e.g., day/night scenarios). Using the 5G-LTE-EPC Network Simulator (LENA) simulator, functional splits 6, 7.1, 7.2, and 7.3 were evaluated under different fronthaul capacity constraints. Results show that highly centralized splits (e.g., 7.1) can significantly increase fronthaul load and may require adaptive reconfiguration to prevent throughput degradation when fronthaul capacity is limited.
4. **Cell-Free Fronthaul Network Planning (Universidad Carlos III de Madrid (UC3M))**
Cell-free architectures, in which multiple distributed radio units jointly serve each user, are analyzed for their potential to boost throughput and fairness via Coordinated Multi-Point (CoMP) techniques. Because native cell-free modeling is not supported in 5G-LENA, a novel simulation-aggregation methodology was developed to emulate multi-association scenarios. This approach revealed significant combinatorial complexity, addressed through high-performance computing and data aggregation pipelines. The resulting datasets enable Artificial Intelligence (AI) training for real-time inference of optimal user-to-base-station associations.

Key Findings:

- (AI-)driven traffic prediction enables substantial energy and equipment savings in 6G transport networks.
- FFS can adapt network function placement to match temporal traffic variations, but requires careful tuning under fronthaul constraints.



- Cell-free networks promise large performance gains but introduce high computational complexity, motivating hybrid simulation-AI methodologies.



TABLE OF CONTENTS

LIST OF TABLES.....	7
ABBREVIATIONS	8
1 Introduction.....	11
2 Reference network design.....	12
1.1 Reference network designs for the 6G transport	12
Network Optimization	18
1.2 Deep Learning based network optimization studies	18
Network under study	18
Real-world dataset and assumptions	20
ML algorithms and training procedure	20
Results	22
Future Work and Recommendations	25
1.3 Flexible functional splits	26
1.4.1. State of the Art of FFS	26
1.4.2. FFS use case for energy efficiency optimization	28
1.4.3. Impact and initial evaluation of various Functional Splits using 5G-LENA network simulator	29
1.4.4. Conclusions and Future work	33
2. Towards cell free Fronthaul Network planning and optimal configuration.....	34
2.1 On cell-free mobile networks.....	34
2.2 Planning networks for cell free fronthaul traffic	36
2.3 Utility Functions and Fairness Models	36
2.4 Optimal cell free fronthaul associations	38
Simulation Data Aggregation and Combination Management	42
Conclusions	44



LIST OF FIGURES

Figure 1 Comparison of (a) DRAN, (b) C-RAN with low layer split, and (c) C-RAN with high layer split. Note the maximum fibre lengths which can be employed based on the latency requirements of the different technologies	13
Figure 2 P2P with dedicated fibre.....	13
Figure 3 Mapping of CU/DU/RU to PON in optical fronthaul architecture. (a) high layer split; (b) low layer split; (c) cascaded split; (d) parallel split [ITU-T, "5G wireless fronthaul requirements in a passive optical network context", Series G, Supplement 6, September 2020].	14
Figure 4 FBA operation of TDM-PON.....	14
Figure 5 P2MP fronthaul topology allowing wavelength overlay for (a) high layer split and (b) high layer split, and (c) mixed high layer and low layer split.....	16
Figure 6 (a) Wavelength-routed and (b) wavelength-selected P2P WDM PON.	16
Figure 7 The network under study; inset: operation of SDPtMP node element.	19
Figure 8 Representative example of true traffic pattern (red) and predicted pattern (blue) using ML over ten days.	23
Figure 9 Over-allocated transceivers [%] compared with the true needs for the three operational cases.	24
Figure 10 Calculated number of transceivers estimated per hour for all 130 rings of TEFNET24 topology (red triangles: average value for each hour across all rings).	24
Figure 11 Relative power consumption estimated per hour when: a) subcarriers are deactivated or b) entire transceivers are put in sleep mode.....	25
Figure 12 FFS day and night scenario.	28
Figure 13 FH throughput for different functional splits when FH capacity is 1Tbps.	31
Figure 14 Application throughput for different functional splits when FH capacity is 1Tbps.	32
Figure 15 FH throughput and Application throughput for different functional splits and FH capacities.	32
Figure 16 Dynamic beam forming in a single-user MIMO smart antenna array.....	34
Figure 17 Coordinated multi-point for interference avoidance (a)(b)(c) and throughput boosting (d) (adapted from)).....	35
Figure 18 Cell free capacity aggregation from multiple RUs	35
Figure 19 Coverage map and dual user associations in the custom 5G MATLAB simulator. Voronoi regions define primary coverage; users (blue crosses) connect to two base stations based on sequential coverage re-evaluation.	38
Figure 20 A sample simulation case with 5G-LENA.....	39



LIST OF TABLES

Table 1. Details of TEFNET24 and total number of required transceivers and interlacer elements for each ring topology. 15

Table 2. Optimal hyperparameter values for the six methods..... 22

Table 3 Possible combinations to simulate and post-processing choices 40



ABBREVIATIONS

Abbreviation	Definition
AI	Artificial Intelligence
AWG	Arrayed Waveguide Grating
BBU	Baseband Unit
BS	Base Station
BWP	Bandwidth Part
CA	Carrier Aggregation
CB	Coordinated Beamforming (in CoMP context)
CC	Channel Coding / Control Channel (context dependent, here Control Channel)
CG	Cloud Gaming
CNN	Convolutional Neural Network
CPRI	Common Public Radio Interface
CPU	Central Processing Unit
CR	Control and Scheduling signaling rate
CS	Coordinated Scheduling (CoMP)
CU	Central Unit
CU-DU	Central Unit - Distributed Unit interface
DBA	Dynamic Bandwidth Allocation
DL	Downlink
DNN	Deep Neural Network
DRAN	Distributed Radio Access Network
DT	Decision Tree
DU	Distributed Unit
ECOC	European Conference on Optical Communication
EPC	Evolved Packet Core
FBA	Fixed Bandwidth Allocation
FFS	Flexible Functional Split
FH	Fronthaul
FS	Functional Split
FSR	Free Spectral Range
HARQ	Hybrid Automatic Repeat Request
HPC	High Performance Computing



IEEE	Institute of Electrical and Electronics Engineers
ITU-R	International Telecommunication Union - Radiocommunication Sector
ITU-T	International Telecommunication Union - Telecommunication Standardization Sector
JT	Joint Transmission (in CoMP)
KNN	k-Nearest Neighbors
LENA	LTE-EPC Network Simulator
LTE	Long-Term Evolution
LTE-EPC	LTE Evolved Packet Core
MAC	Medium Access Control
MAPE	Mean Absolute Percentage Error
MATLAB	Matrix Laboratory (simulation software)
MCS	Modulation and Coding Scheme
MEC	Multi-access Edge Computing
MIMO	Multiple Input Multiple Output
ML	Machine Learning
MOPA	Micro-Optics Passive Alliance (MOPA Alliance)
NG-RAN	Next-Generation Radio Access Network
NGI	Next Generation Interface
NN	Neural Network
NR	New Radio
NR-U	New Radio User
ODN	Optical Distribution Network
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OLT	Optical Line Terminal
ON	Optical Node
ONU	Optical Network Unit
PDCP	Packet Data Convergence Protocol
PF	Proportional Fair (scheduler)
PHY	Physical
PON	Passive Optical Network
QAM	Quadrature Amplitude Modulation
RAN	Radio Access Network
RB	Resource Block



RF	Random Forest
RLC	Radio Link Control
RRC	Radio Resource Control
RRM	Radio Resource Management
RU	Radio Unit
SINR	Signal to Interference plus Noise Ratio
SVR	Support Vector Regression
TDD	Time Division Duplexing
TDM	Time Division Multiplexing
TEFNET24	Telefónica Reference Network Topology 2024
TTI	Transmission Time Interval
UE	User Equipment
VR	Virtual Reality
WDM	Wavelength Division Multiplexing
PON	Passive Optical Network
XG-PON	10 Gigabit-capable Passive Optical Network
XGS-PON	10 Gigabit-capable Symmetric Passive Optical Network



1 Introduction

The evolution The transition from 5G towards 6G represents not only an incremental upgrade in wireless communication performance but also a paradigm shift in how transport networks are designed, managed, and optimized. Beyond sheer capacity growth, 6G transport networks must accommodate stringent requirements in terms of latency, scalability, flexibility, energy efficiency, and trustworthiness. This requires a fundamental rethinking of the x-haul infrastructure (fronthaul, midhaul, and backhaul), which will serve as the backbone of future wireless services.

In particular, the Radio Access Network (RAN) evolution towards disaggregated and virtualized deployments introduces unprecedented challenges for optical transport. Unlike traditional architectures, where baseband functions were tightly coupled to radio units, 6G requires a highly dynamic orchestration of Central Units (CUs), Distributed Units (DUs), and Radio Units (RUs), interconnected via heterogeneous transport technologies. The ability to flexibly decide the location of functions, adapt to varying traffic conditions, and ensure synchronization across diverse network domains is essential for guaranteeing service continuity and quality.

The **PROTEUS-6G project** directly addresses these challenges by proposing a programmable, reconfigurable optical transport infrastructure tailored to the needs of 6G. The overarching vision is to deliver unconstrained services in terms of capacity, energy efficiency, and adaptability, leveraging advanced optical, networking, and AI-driven optimization technologies. This **Deliverable (D2.4)** presents the **preliminary network optimization studies**, serving as a foundation for the full design and validation of the PROTEUS-6G architecture.

Three key research pillars are pursued in this work:

- **Reference transport network designs:** Comparative evaluation of optical transport architectures for fronthaul, ranging from conventional point-to-point fibre deployments to advanced wavelength-division multiplexed passive optical networks (WDM-PONs) and the novel Spatially-Diverse Point-to-Multipoint (SDPtMP) approach. These solutions are assessed in terms of scalability, latency, resilience, and energy performance.
- **Machine Learning-based dynamic optimization:** Exploiting real-world datasets and predictive models such as Random Forest (RF), Deep Neural Networks (DNNs), and Support Vector Regression (SVR) to forecast traffic dynamics. These predictions are translated into resource allocation strategies that reduce over-provisioning and enable energy-efficient transceiver and subcarrier management across large-scale transport topologies.
- **Flexible functional splits and cell-free planning:** Investigating how Flexible Functional Splits (FFS) can dynamically adapt to traffic fluctuations and infrastructure constraints, while cell-free fronthaul planning explores joint transmission schemes and Coordinated Multipoint (CoMP) to enhance throughput and fairness. These methodologies are supported by ns-3/5G-LENA simulations and scalable data aggregation pipelines.

The results presented demonstrate that AI-driven and flexible transport solutions are essential for achieving the dual goals of **sustainability and performance** in 6G. Preliminary findings already indicate significant improvements in energy efficiency (up to 21% reduction via transceiver sleep modes), lower over-provisioning, and improved adaptability to diverse service requirements



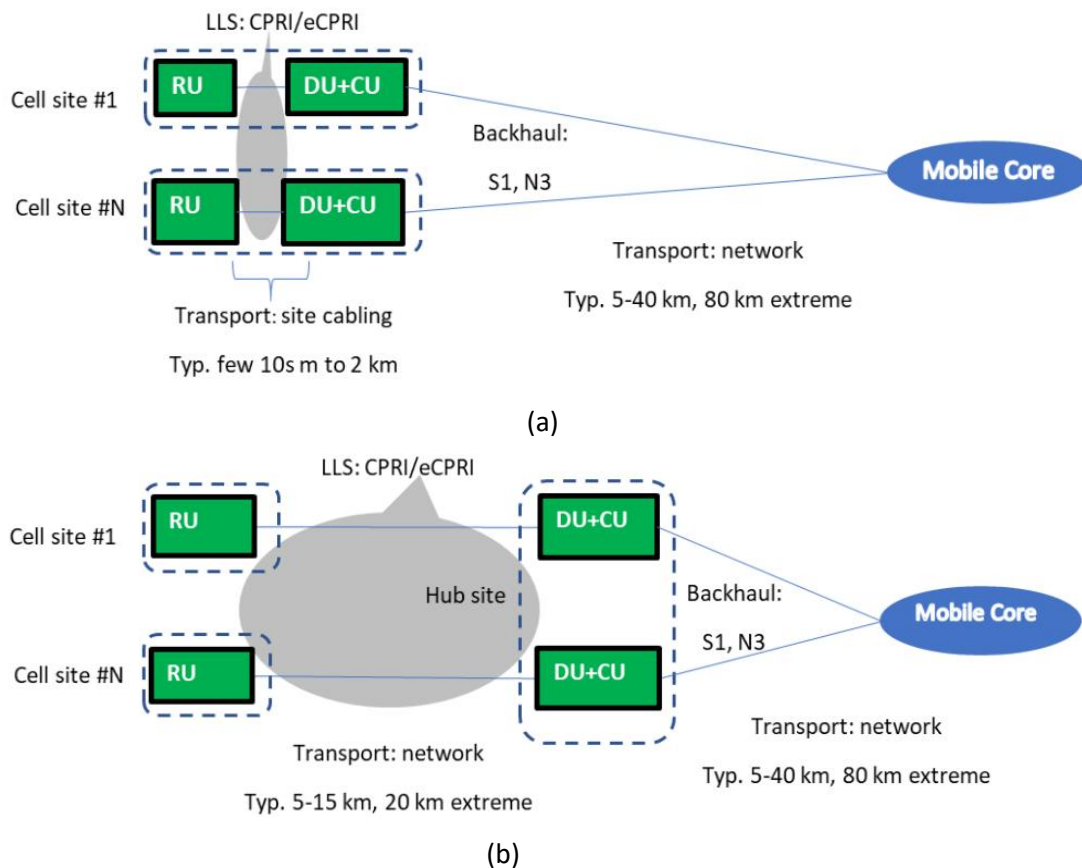
2 Reference network design

1.1 Reference network designs for the 6G transport

The evolution of mobile access networks towards 6G imposes substantial requirements on the transport infrastructure that connects Radio Units (RUs), Distributed Units (DUs), and Central Units (CUs). These requirements include ultra-low latency, high throughput, synchronisation precision, and increased energy efficiency. As the RAN becomes increasingly disaggregated and centralised, the role of the transport network is now critical, particularly for fronthaul segments. This section presents an overview of transport network designs, beginning with Distributed RAN (DRAN) and culminating in advanced fronthaul solutions aligned with the requirements of 6G.

Traditional deployments have relied on DRAN architectures, where baseband processing is collocated with radio heads at each cell site. While this model simplifies transport requirements, it suffers from high operational cost and limited coordination across cells. The introduction of centralised RAN (C-RAN) moved digital units to aggregation points or central offices, allowing better resource pooling and CoMP. In turn, virtualised RAN (vRAN) further abstracts the DU and CU functions, enabling flexible deployment of software-defined baseband functions over commodity hardware.

The transformation from DRAN to C-RAN and vRAN introduces the concept of functional splits between RUs, DUs, and CUs. These splits determine the transport bandwidth and latency requirements (see Figure 1). High-layer splits (e.g., Option 2 or F1 interface) allow the use of standard Ethernet-based transport with relaxed latency constraints, while low-layer splits (e.g., Option 7 or Common Public Radio Interface (CPRI)/eCPRI interfaces) require deterministic, high-speed links with tight jitter and synchronisation guarantees.



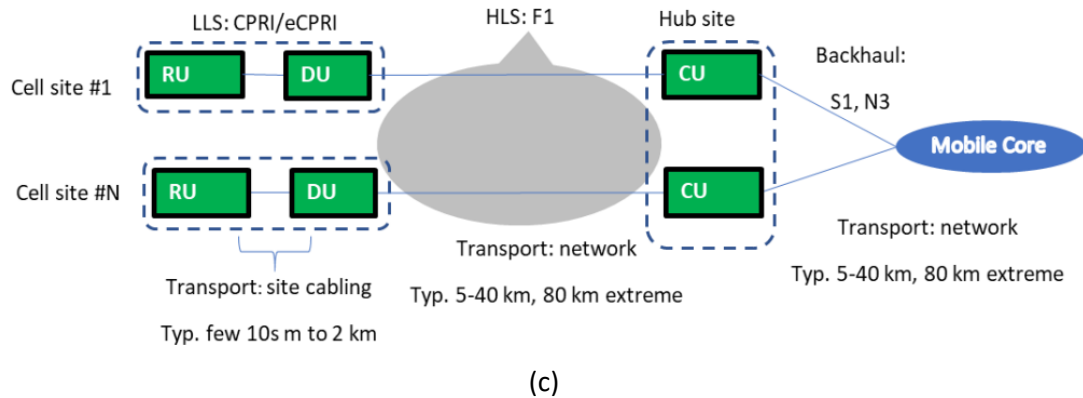


Figure 1 Comparison of (a) DRAN, (b) C-RAN with low layer split, and (c) C-RAN with high layer split. Note the maximum fibre lengths which can be employed based on the latency requirements of the different technologies¹.

Several network architectures have been proposed for the fronthaul segment of the transport network. They range from Point-to-Point (P2P) solutions over dedicated fibre to shared optical networks.

- P2P dedicated fibre topology

In areas with abundant dark fibre, the most straightforward fronthaul design is a P2P architecture where each RU is directly connected to a DU or Optical Line Terminal (OLT) via dedicated fibres. This setup ensures high performance, simplicity of deployment, and strong fault isolation. It is especially suitable for low-layer splits requiring constant bit rates and low latency. Figure 2 illustrates this basic layout.

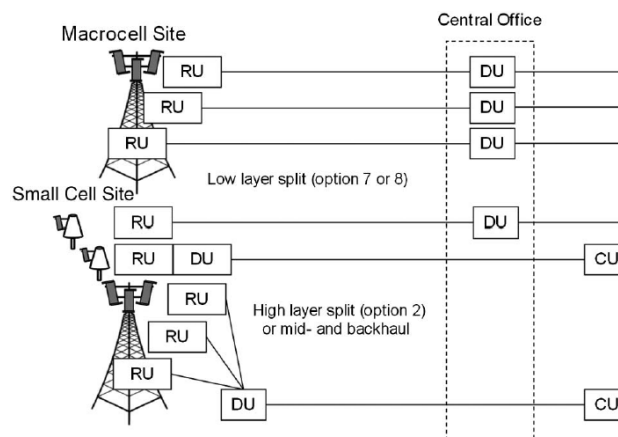


Figure 2 P2P with dedicated fibre²

However, P2P fibre becomes economically and physically unfeasible in dense urban environments or rural areas with limited fibre infrastructure. Therefore, shared optical access solutions become important in these scenarios. Figure 3 shows how PON topologies can be used to connect CU/DU/RU through the Optical Distribution Network (ODN).

¹ MOPA, Technical paper, Version 3.2, April 1, 2025: https://mopa-alliance.org/wp-content/uploads/2025/03/MOPA_Technical_Paper-v3.2-Final.pdf

² Philippe Chanclou, Hiroo Suzuki, Jin Wang, Yiran Ma, Mauro Renato Boldi, Kazuki Tanaka, Seungjoo Hong, Claudio Rodrigues, Luiz Anet Neto, and Jiang Ming, "How Does Passive Optical Network Tackle Radio Access Network Evolution," J. Opt. Commun. Netw. 9, 1030-1040 (2017)



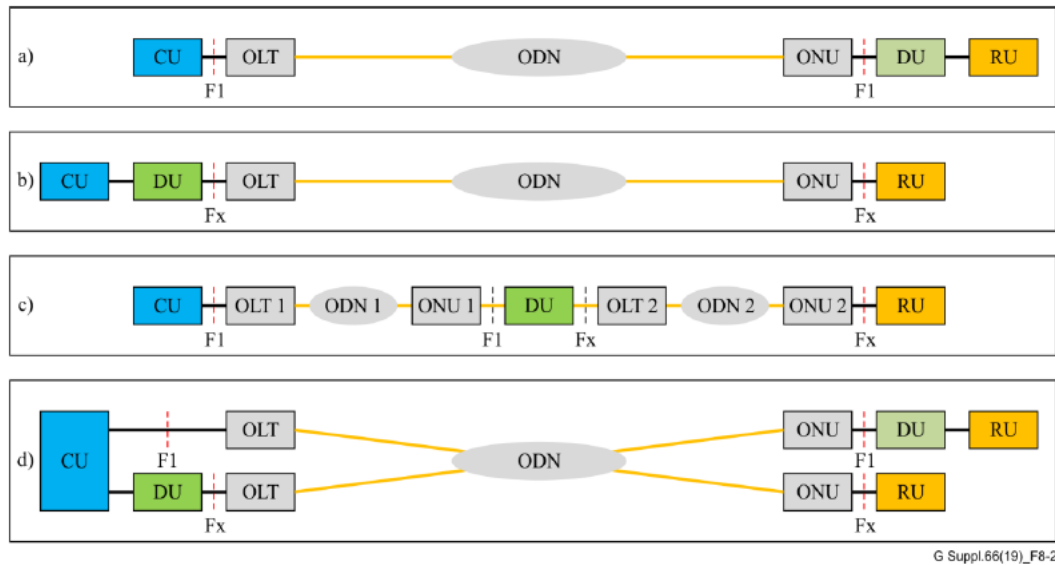


Figure 3 Mapping of CU/DU/RU to PON in optical fronthaul architecture. (a) high layer split; (b) low layer split; (c) cascaded split; (d) parallel split [ITU-T, "5G wireless fronthaul requirements in a passive optical network context", Series G, Supplement 6, September 2020].

- TDM-PON for fronthaul

Time Division Multiplexing PONs (TDM-PONs), such as G-PON, 10 Gigabit-capable Passive Optical Network (XG-PON), and 10 Gigabit-capable Symmetric Passive Optical Network (XGS-PON), offer cost-effective solutions by allowing multiple RUs to share a single optical fibre through time-based scheduling. For high-layer splits, Dynamic Bandwidth Allocation (DBA) mechanisms can efficiently adapt to traffic fluctuations, supporting statistical multiplexing.

For low-layer splits, however, standard DBA is unsuitable due to the constant bit-rate requirement of (CPRI-)like interfaces. In such cases, Fixed Bandwidth Allocation (FBA) must be employed, whereby the OLT grants periodic upstream windows without waiting for Optical Network Unit (ONU) requests. This is essential to reduce latency and jitter, as shown in Figure 4. Low-latency DBA, which coordinates wireless scheduling in the DU with PON scheduling in the OLT, is another emerging enhancement, described in ITU-T G.989.3 Amendment 1.

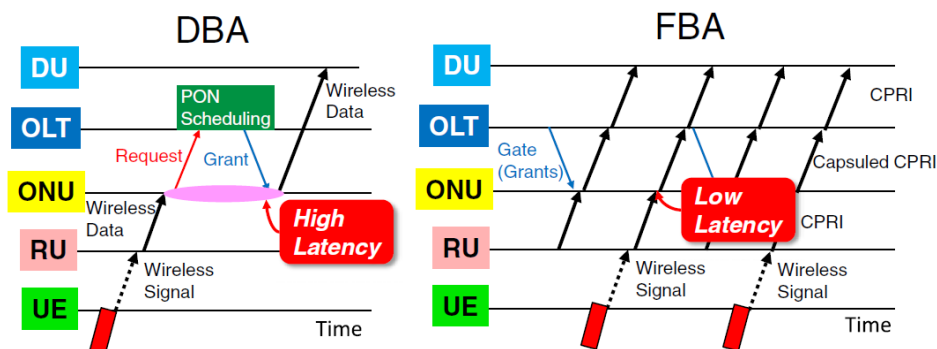


Figure 4 FBA operation of TDM-PON³.

³ Philippe Chanclou, Hiroo Suzuki, Jin Wang, Yiran Ma, Mauro Renato Boldi, Kazuki Tanaka, Seungjoo Hong, Claudio Rodrigues, Luiz Anet Neto, and Jiang Ming, "How Does Passive Optical Network Tackle Radio Access Network Evolution?", J. Opt. Commun. Netw. 9, 1030-1040 (2017)

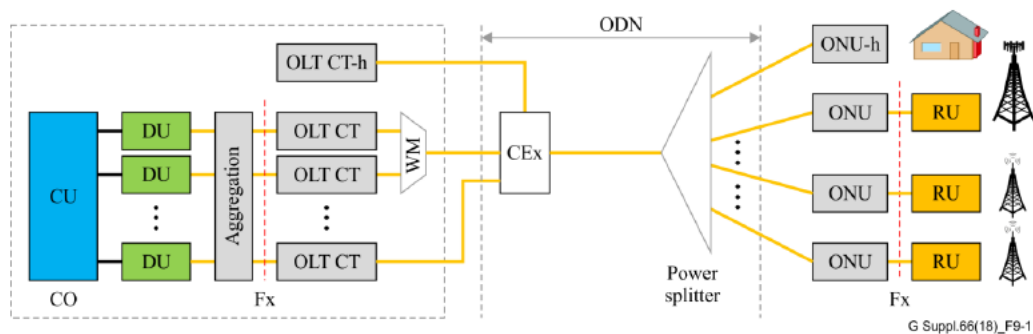


TDM-PON remains attractive due to its maturity and reuse potential of existing FTTx infrastructure. However, its limitations in bandwidth scalability and tight latency control present challenges for next-generation fronthaul.

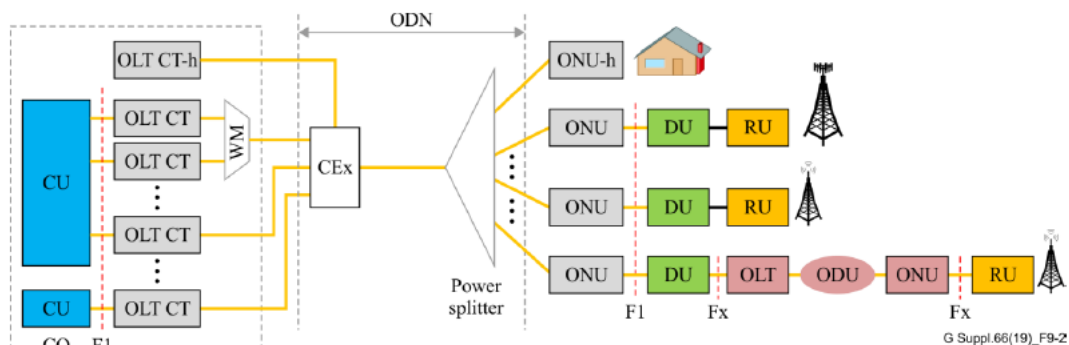
- (TWDM-PON) and (NG-PON2)

To overcome the limitations of basic TDM-PON, Time and Wavelength Division Multiplexed PONs (TWDM-PON), such as NG-PON2 (ITU-T G.989), introduce multiple wavelength channels per direction. This allows for virtual PONs with dedicated wavelengths per RU or DU, thus increasing aggregate capacity to 40 Gbps or more.

TWDM-PON can simultaneously serve both mobile and fixed broadband users, offering improved resource isolation and flexibility. Each wavelength can be assigned a low-latency scheduling profile, extending the benefits of coordinated DBA into the wavelength domain. Figure 5 provides a visual representation for this architecture, where (WM) is the wavelength multiplexer and CEx the coexistence element enabling the overlay of technologies.

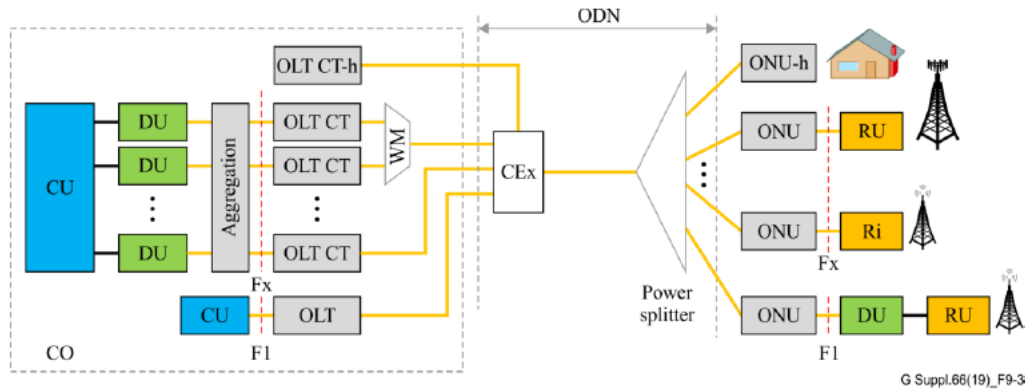


(a)



(b)





(c)

Figure 5 P2MP fronthaul topology allowing wavelength overlay for (a) high layer split and (b) high layer split, and (c) mixed high layer and low layer split⁴.

This model aligns well with 6G visions of converged fixed-mobile access and multi-service integration, particularly in urban environments with high RU density. Alternatively, a dedicated PON for mobile services can be implemented.

- WDM-PON and P2P wavelength solutions over shared fibre

P2P WDM-PON provides dedicated wavelength pairs (upstream/downstream) for each (ONU-RU) connection. It is the most deterministic optical transport solution and supports both high and low layer splits without requiring statistical multiplexing. P2P WDM-PON is particularly relevant in greenfield deployments or in cases where ultra-low latency and isolation are essential. There are two subtypes:

- a) Wavelength-Routed: Uses fixed wavelength filters or Arrayed Waveguide Grating (AWGs) to allocate wavelengths statically (Figure 6(a)).
- b) Wavelength-Selected: Uses passive splitters or filters that expose a range of wavelengths to ONUs, allowing dynamic tuning (Figure 6(b)).

The main technical challenge lies in the use of colorless ONUs, which must dynamically select and stabilize their wavelengths.

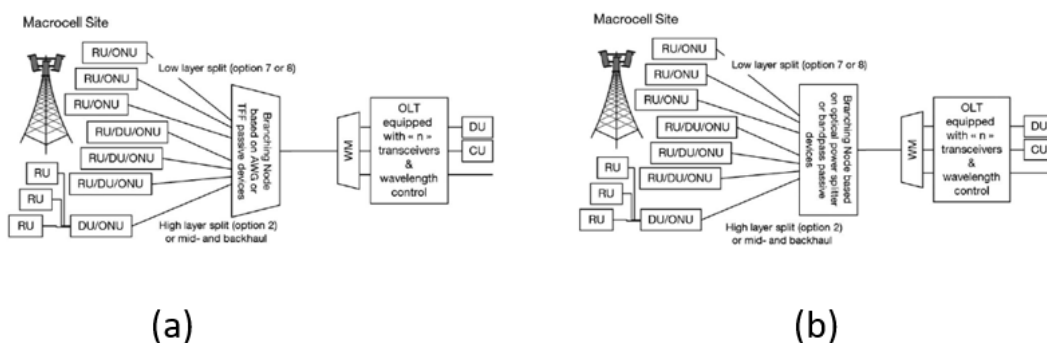


Figure 6 (a) Wavelength-routed and (b) wavelength-selected P2P WDM PON⁵.

⁴ ITU-T, “5G wireless fronthaul requirements in a passive optical network context”, Series G, Supplement 66, September 2020

⁵ Philippe Chanclou et al, "How Does Passive Optical Network Tackle Radio Access Network Evolution?," J. Opt. Commun. Netw. 9, 1030-1040 (2017)



P2P WDM-PON supports transparent and transcoded transport modes for fronthaul. In transparent mode, CPRI or e-CPRI streams are directly transmitted. In transcoded mode, user and control data are multiplexed under a unified framing, as described in Institute of Electrical and Electronics Engineers (IEEE) 1914.3. These features make WDM-PON highly compatible with 6G requirements for deterministic, high-capacity, and synchronized transport. PROTEUS-6G aims at a wavelength-routed fronthaul network architecture enabling P2MP transmission.



Network Optimization

1.2 Deep Learning based network optimization studies

Network under study

The X-haul network architecture under investigation, illustrated in Figure 7, comprises of three segments: the fronthaul, the midhaul, and the backhaul. In TEFNET24⁶, the midhaul is organized in hierarchical horseshoe ring topologies that aggregate traffic from metro and access domains. Each midhaul node typically serves 50–100 Base Stations (BSs) and each ring comprises 2 and 10 midhaul nodes.

Central to the mid-haul node we consider the SDPtMP node element⁷, which enables a single transceiver to serve four BSs by distributing traffic across four spatial feeds. The subcarrier interlacing is implemented passively using cascaded sharp optical interleavers, arranged in two stages (inset of Figure 7). The first stage, with a 64 GHz Free Spectral Range (FSR), yields complementary 30 GHz-wide channels, while the second stage—phase-shifted by $\frac{1}{4}$ of the FSR—generates four distinct subcarrier groups, each 14 GHz wide, sufficient to carry three 4 GHz subcarriers.

It is important to note that a single wavelength may be present on several inputs, which makes the SDPtMP node free from wavelength contention. The best practice is to initially utilize the same wavelength for all spatial inputs before adding more wavelength channels. This approach enables the receiver to identify all four spatial channels using fixed subcarrier mappings, resulting in a fourfold increase in incoming capacity compared to conventional PON systems.

In this study, we assume that each node hosts High-speed Coherent Transceiver Prototype (XR400) prototype transceivers⁸, transmitting up to 16 digital subcarriers towards the BSs, each operating at 4 GBaud and utilizing dual-polarization 16 Quadrature Amplitude Modulation (QAM) modulation, leading to a net data rate of 25 Gb/s per subcarrier and up to 75 Gb/s per BS. Using the assumption that each node typically serves 75 BSs on average, the traffic of one midhaul node at peak hours yields 5.6 Tb/s.

⁶ J. M. Rivas-Moscoso et al., "TEFNET24: reference packet optical network topology for edge to core transport," in *Journal of Optical Communications and Networking*, vol. 16, no. 11, pp. G28-G39, November 2024, doi: 10.1364/JOCN.533131

⁷ C. Christofidis et al, "Spatially-Diverse Point-to-MultiPoint Optical Distribution Network for Enhanced 6G Fronthaul," 2023 International Conference on Photonics in Switching and Computing (PSC), Mantova, Italy, 2023, pp. 1-3, doi: 10.1109/PSC57974.2023.10297167.

⁸ A. Rashidinejad et al., "Real-Time Point-to-Multipoint for Coherent Optical Broadcast and Aggregation – Enabled by Digital Subcarrier Multiplexing," in *Optical Fiber Communications Conference and Exhibition (OFC)*, San Diego, CA, USA, 2023, pp. 1-3, DOI:10.1364/OFC.2023.W3H.1



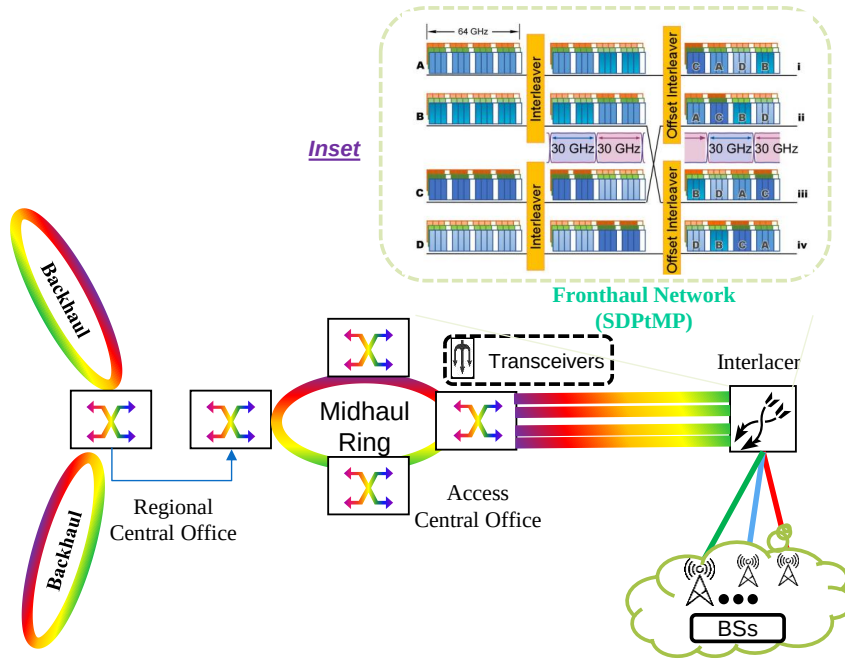


Figure 7 The network under study; inset: operation of SDPtMP node element.

To provide more insights into TEFNET24, in Table 1 we perform a statistical breakdown for the different topologies, illustrating how network resources scale across different configurations. We can observe 6 single-ring and 47 multi-ring topologies comprising 130 rings containing 725 nodes. The data reveals significant variability in ring structures, from small 2-node rings to larger 10-node configurations (averaging 5.5 nodes per ring). This heterogeneity directly impacts resource requirements, with substantial scaling in both transceiver counts (13,594 total) and base station connections (54,375 total) by assuming that each node serves 75 BSs. The 2-ring topology dominates the deployment (26 occurrences), while 4-ring configurations show the highest average transceivers per ring (130), demonstrating how topology choices significantly influence equipment density and resource requirements, directly affecting the adopted energy optimization strategies. Moreover, the number of interlacers deployed across these topologies plays a crucial role in determining the network's flexibility and resilience. Topologies with more interlacers can better accommodate dynamic traffic patterns but may also introduce additional complexity and cost in network management.

Table 1. Details of TEFNET24 and total number of required transceivers and interlacer elements for each ring topology.

#-Ring	Occurence	N _{rings}	N _{nodes}	Avg. N _{Nodes} / Ring	N _{BSs}	N _{TRxs}	Avg. N _{TRxs} / Ring	Interlacer elements	Avg. interlacer elements / ring
1	6	6	33	5.50	2,475	619	103	155	31
2	26	52	266	5.12	19,950	4,988	96	1247	24
3	16	48	274	5.71	20,550	5,138	107	1284	25
4	3	12	83	6.92	6,225	1,556	130	389	30
6	2	12	69	5.75	5,175	1,294	108	323	24
Total	53	130	725		54,375	13,594		3,398	



Real-world dataset and assumptions

The dataset under study contains real data collected during 2022 and 2023 from Telefónica's Network in a city in South America, representing the total traffic of 600000 subscribers at four network gateways. The data spans over 358 days, and each day contains 24 measurements (one per hour), leading to a total of 8592 values. Maximum traffic is observed between 19:00 and 22:00, and the lowest traffic between 1:00 and 4:00. The dataset contains mainly 4G (and a small fraction of 5G) traffic. To scale the traffic to the midhaul capabilities, we make the following assumptions. First, our target is a city of more significant scale, e.g. 3 million subscribers, leading to a 5-times multiplication factor. Next, by scaling the 4G peak capacity of 1 Gb/s⁹ to the peak capacity of 6G, which might reach up to 400 Gb/s for the long-term evolution scenario¹⁰, there is also another 400 times multiplication factor. These two factors increase the maximum traffic by 2,000 times, scaling the traffic during the peak hours from about 14.05 Gb/s to 28.1 Tb/s, which is the estimated traffic for the average midhaul ring scenario.

ML algorithms and training procedure

1.3.3.1 Decision Tree

The Decision Tree (DT) organizes the data in a tree structure starting from the root node towards the leaf nodes by asking a question with a binary answer, e.g. yes/no, each time. Every time a question is answered, new branches are created, and the feature space is segregated into disjoint regions. The optimal feature selection in each node and the proper question are usually based on the feature with the highest information gain. As a consequence, by selecting the appropriate tree depth, a decision tree with a large number of nodes that can provide estimations for the value y can be generated. DTs are highly interpretable and provide a clear visual representation of the decision-making process, making them easy to understand. However, they are prone to overfitting, especially when the tree is deep or the dataset contains noise.

1.3.3.2 Random Forest

The Random Forest (RF) algorithm is in essence a forest that incorporates an ensemble of decision trees. This forest is usually trained using the bagging method. The main advantage of RF over DT is the use of multiple trees instead of a single tree since many uncorrelated trees operating as an ensemble outperform an individual tree. For this reason, RF typically provides a higher accuracy than DT; however, it comes at the cost of higher complexity. Furthermore, RFs are less prone to overfitting, as the averaging of multiple trees helps to smooth out anomalies and noise in the data. Each tree in the forest is trained on a random subset of the data, and at each split, a random subset of features is considered, which increases diversity among the trees. This diversity is crucial for robust performance and generalization to unseen data. Additionally, RFs also provide measures of feature importance, which can be valuable for understanding which variables most influence the model's predictions.

1.3.3.3 k-Nearest Neighbors

The k-Nearest Neighbors (k- (NN)) algorithm first calculates the Euclidean distances between a new data point and each of the data points of the training set as follows

$$d(y_i, \hat{y}) = \sqrt{(x_{i,1} - x_1)^2 + (x_{i,2} - x_2)^2 + \dots + (x_{i,n} - x_n)^2} \quad (x1)$$

⁹ Recommendation ITU-R M.1645, June. 2003.

¹⁰ NetWorld2020, "Smart networks in the context of NGI," Strategic Research and Innovation Agenda 2021-27, 2020.



Then, it selects the k y points with the smaller Euclidean distance to predict the quantity by taking their average. In general, k -NN is more time costly than DT and RF, especially in large datasets, e.g., that comprise thousands of samples, as it needs to calculate the distance between the input vector and all other vectors in the dataset and afterward select the k vectors with the lowest distance.

1.3.3.4 Support Vector Regression

Support Vector Regression (SVR) allows defining the value of acceptable error ε in the resulting model to find the proper hyperplane to fit data and estimate the value. The mathematical target of the SVR algorithm is to

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + CC \sum_{i=1}^n |\xi_i| \\ & \text{subject to } |y_i - w_i x_i| \leq \varepsilon + |\xi_i| \end{aligned} \quad (x2)$$

where w_i is the value of the i th coefficient, ξ the deviation of every value that falls outside of ε and Channel Coding / Control Channel (CC) is a value that trades the accuracy of estimations with the tolerance for points that fall outside ε . As CC increases, the tolerance is higher whilst as CC approaches zero, the tolerance is minimized. The SVM can also be employed to solve regression problems, and in this case, the algorithm is named SVR.

1.3.3.5 Deep Neural Network

A Deep Neural Network (DNN) exploits simple processing elements, called “neurons,” to process data and exchange information. Each neuron in the network can receive incoming signals, process them, and send an outgoing signal. Each neuron is connected to at least one other neuron, and each connection is evaluated with a real number called a weight coefficient. The weight coefficient reflects the degree of importance of that particular connection in the neural network. For example, we have a deep neural network consisting of neurons, which are organized into layers. The first layer is called the “input layer,” the intermediate layers are called “hidden layers,” and the last one is the “output layer.” Each layer processes the information it receives and passes the results to the next layer. The hidden layers are responsible for extracting complex features and patterns from the input data. As the data moves through the layers, the network learns to represent increasingly abstract concepts. During training, the weights of the connections are adjusted using algorithms such as backpropagation, allowing the network to improve its performance. The output layer produces the final prediction or classification based on the processed information.

1.3.3.6 Convolutional Neural Network

Deep Convolutional NNs (CNNs) consist of a number of neurons with learnable weights and biases. Typically, a Convolutional Neural Network (CNN) consists of a number of different layers: input, output as well as a number of hidden layers which can perform various operations, i.e. convolution and pooling. Convolution refers to the mathematical operation where the input signal (e.g., array) is multiplied element-wise with a weight W to extract features (e.g., edges) in an autonomous way and, consequently, to produce the feature maps. A pooling layer is employed to reduce the data dimensions by composing a cluster of many input neurons to a single output neuron. Finally, the output layer typically employs an activation function which classifies the input data based on the calculated probability distribution.

To predict the traffic by recognizing patterns and extracting features automatically from the data, we employed a custom lightweight deep CNN (of 33,569 parameters) with five layers:

- **Layer 1:** 16 convolutional filters with a size of (1,96). This is followed by a ReLU activation function, a (1,4) strided max-pooling operation, and a dropout probability equal to 0.25.
- **Layer 2:** 32 convolutional filters with a size of (1,24). This is followed by a ReLU activation function, a (1,4) strided max-pooling operation, and a dropout probability equal to 0.25.
- **Layer 3:** 48 convolutional filters with a size of (1,6). This is followed by a ReLU activation function, a (1,3) strided max-pooling operation, and a dropout probability equal to 0.25.



- **Layer 4:** 64 convolutional filters with a size of (1,2). This is followed by a ReLU activation function, a (1,2) strided max-pooling operation, and a dropout probability equal to 0.25.
- **Layer 5:** 64 hidden units with a linear activation function.

1.3.3.7 Training Procedure

During the training, we divided the dataset into train, validation, and test subsets. The data from the first 212 days formed the test set, the next 68 days the validation set and the final days the test set. Each input sample has 96 values (data of 4 consecutive days), and the ML output was the traffic for the next hour. An overlap of 95/96 was considered as an attempt to estimate the traffic of all possible hours on the test set as well as to create more training samples. Before being injected into the ML, the data were normalized using Z-score normalization. Finally, for the CNN, Adam was the optimizer, the maximum number of epochs was set to 500 and the training stopped if the loss in the validation set did not improve after 50 epochs.

Results

1.3.4.1 ML evaluation

The six methods were applied to the dataset to predict the traffic on an hourly basis. During the training, the optimal hyperparameters derived through random grid search for each method, leading to the estimation of the Mean Absolute Percentage Error (MAPE), for all six methods (reported in Table 2). From this table, we can observe that RF outperforms the other methods, mainly due to its ensemble design combining multiple decision trees. Next, SVR, shows a similar performance, while DT and DNN attain a lower accuracy, with DT's simplicity and DNN's layered architecture providing still reliable but less optimal results. Next, k-NN shows a lower performance, while CNN underperforms significantly, suggesting its complex 5-layer structure requires a larger amount of training data. Overall, the error in all six methods is considered acceptable for this task. These predictions can be exploited to estimate the actual number of required transceivers on an hourly basis, putting some of them into sleep mode or deactivating subcarriers¹¹ to reduce the overall network power consumption.

Table 1. Optimal hyperparameter values for the six methods.

Method	Optimal hyper-parameters	MAPE
DT	depth = 10	2.88%
RF	depth = 10, number of estimators = 10	2.29%
k-NN	10	3.54%
SVR	C=1, epsilon=0.1	2.39%
DNN	hidden layers = 2, neurons/layer = 32	2.65%
CNN	5 layers	3.75%

Next, a representative five-day prediction is shown in Figure 8, where the ML's output (blue curve) tracks the true traffic pattern (red curve) effectively.

¹¹ C. Castro et al., "Power and Spectral Savings in Metro-Aggregation Networks Exploiting Coherent Point-to-Multipoint Transceivers," ECOC 2024; 50th European Conference on Optical Communication, Frankfurt, Germany, 2024, pp. 519-522.



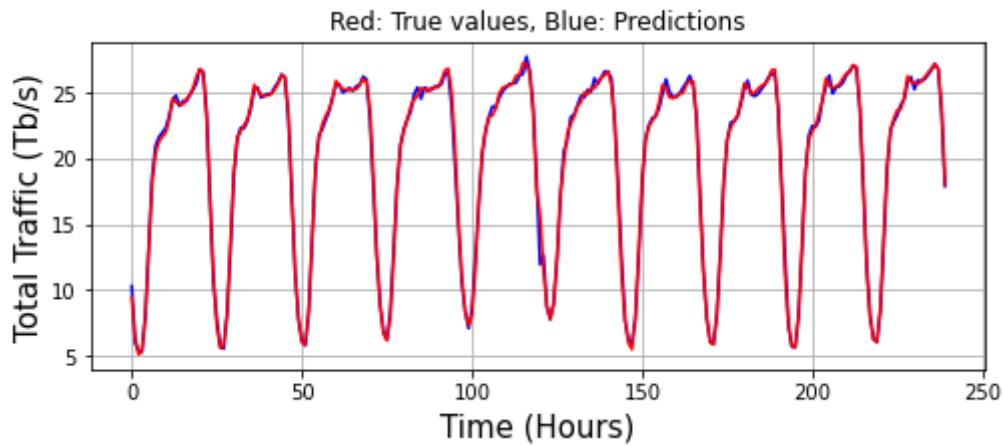


Figure 8 Representative example of true traffic pattern (red) and predicted pattern (blue) using ML over ten days.

1.3.4.2 Transceiver savings over the average ring

In this section, we estimate the number of transceivers required on an hourly basis for the average midhaul ring comprising 5 nodes in the test set for: a) a static case, considering that all 95 transceivers are exploited during all hours and days; b) a semi-static case, considering that the number of transceivers is adjusted every 8 hours, according to the average value of these 8 hours and scaled up to ensure zero under-allocation in the validation set; c) a dynamic case, calculating the number of transceivers on an hourly basis using the ML, and slightly scaling it up to about 15% to avoid resource under-allocation in the validation set. Figure 9 shows the calculated number of transceivers on an hourly basis for the three cases. The dynamic case illustrates the number of transceivers averaged over all days in the test set for the different hours. The first observation is that the dynamic case leads to a reduced number of transceivers, especially for the hours between 23:00 and 6:00, where the requirements are minimized. However, even in the hours 6:00 to 18:00, the number of transceivers is between 2.5% and 22% lower than the peak hour. The semi-static case provides 10% lower transceivers during off-peak hours than the static case, leading to a 5% lower average number of transceivers over the entire day. Overall, the dynamic case leads to an average of about 20% lower number of transceivers than the static case and 16% lower number of transceivers than the semi-static case. These gains can be directly translated to up to a total of 20% lower power consumption in case of transceiver sleep mode, as it is shown in the following section.

The second set of results illustrated in Figure 9 shows the percentage of overallocated transceivers. This compares the number of transceivers for the three cases against the actual needs on the test set, when estimated for all 24 hours. As expected, there is a high overallocation in all three cases, especially in off-peak hours (between 01:00 and 06:00). This happens because, during these hours, the traffic variance is the highest, making the predictions harder than the peak hours. This, in turn, increases the number of assigned transceivers to avoid any potential under-allocation which can negatively impact the Quality of Service (QoS). Overall, the average over-allocation over all days of the test set for the static, semi-static and dynamic cases is 92.5%, 86.7% and 28.8%, respectively, demonstrating the superiority of the dynamic case, especially during off-peak hours.



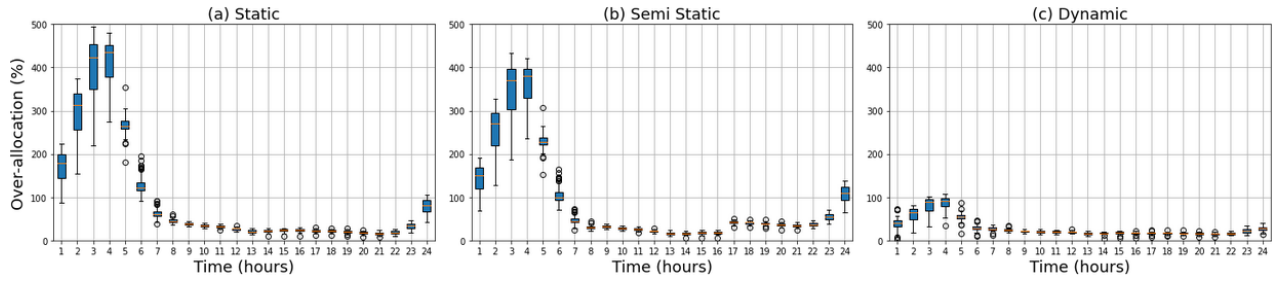


Figure 9 Over-allocated transceivers [%] compared with the true needs for the three operational cases.

1.3.4.3 Energy consumption savings over the entire network topology

As in the previous section we estimated the maximum number of transceivers for the average ring, Figure 10 illustrates the transceiver requirements across all 130 rings in the TEFNET24 architecture over a 24-hour period. To perform our estimations, we exploit the outputs of the RF, and we slightly increase this traffic to about 15% to avoid any potential resource under-allocation. Then the number of required transceivers per hour and per ring is quantified. According to Figure 10, during off-peak hours (0:00-5:00), the lowest number of transceiver requirements (approximately 25-70) is observed, while peak hours (10:00-21:00) necessitate substantially higher resources (approximately 70-130 transceivers). The wide interquartile ranges and extended whiskers highlight considerable variability, suggesting diverse traffic conditions throughout the network topology of 130 midhaul rings. This is justified by the significant variability in the number of nodes per ring, which ranges from 2 to 10, with an average of 5.6 nodes. This network-wide perspective confirms substantial opportunities for dynamic transceiver management to reduce energy consumption, as illustrated in the next set of results.

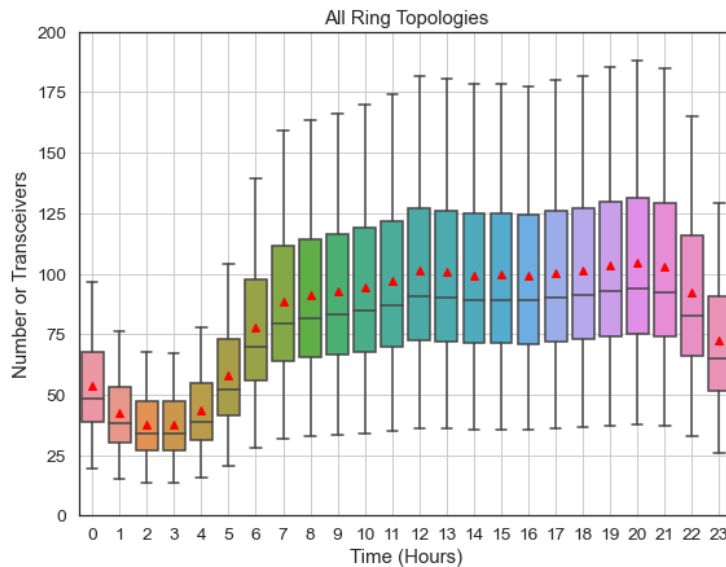


Figure 10 Calculated number of transceivers estimated per hour for all 130 rings of TEFNET24 topology (red triangles: average value for each hour across all rings).

Figure 11 illustrates the average relative energy consumption across all rings of the TEFNET24 topology under two different scenarios. The first scenario considers the case where subcarriers are deactivated during hours of low demand. The second scenario involves putting transceivers in sleep mode, based on the actual hourly traffic needs. When the transceivers are put in sleep mode, the energy consumption can drop down to almost zero, as only a few components (e.g., internal clocks) are active¹², so the assumption made here is that the transceivers consume zero energy. We then compare the energy consumption results from these two energy-

¹² L. Kundu, X. Lin and R. Gadiyar, "Toward Energy Efficient RAN: From Industry Standards to Trending Practice," in IEEE Wireless Communications, vol. 32, no. 1, pp. 36-43, February 2025, doi: 10.1109/MWC.010.2400061.



saving strategies with a baseline scenario in which the maximum number of transceivers, as determined by peak-hour requirements, always remains active.

For the case of subcarrier deactivation, we consider a 92%, 85% and 77% relative power consumption when the number of active subcarriers is 12, 8 and 4, respectively. As the SDPtMP node exploits at most three subcarriers out of each group of four, we consider that the worst case is to have 12 subcarriers active. This strategy, shown in Figure 11(a), demonstrates moderate energy savings with consumption ranging from 83% to 93% relative to the baseline, exhibiting lesser fluctuation throughout the day. Meanwhile, Figure 11(b) demonstrates transceiver sleep mode, which delivers significantly higher energy savings during off-peak hours (0:00-5:00), where energy consumption drops down to approximately 35% to 40%. During peak hours (11:00-21:00), both strategies show almost similar energy consumption levels and close to maximum capacity. Overall, the transceiver sleep mode approach clearly outperforms subcarrier shutdown in terms of energy efficiency, particularly during nighttime hours, achieving total savings of 21% compared to 10% for subcarrier deactivation.

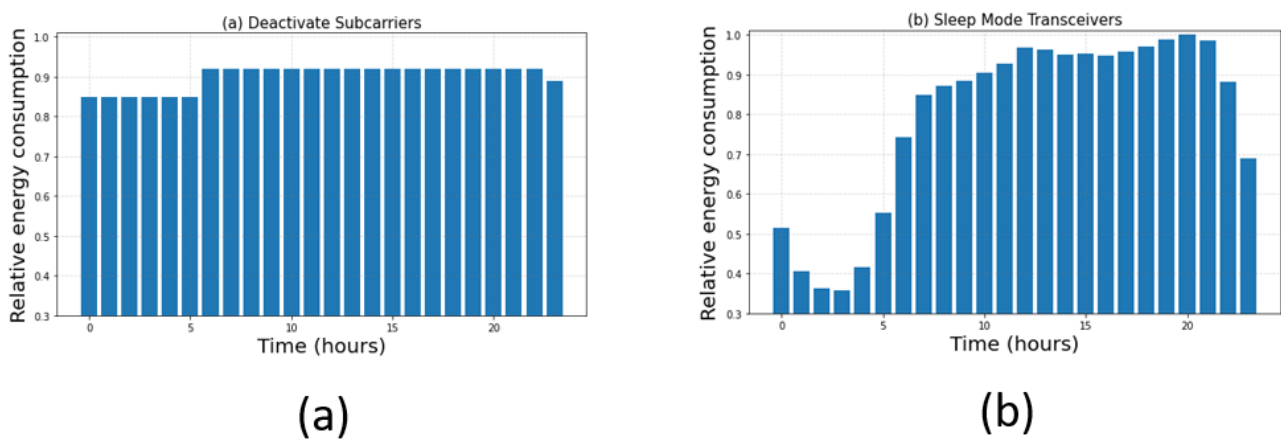


Figure 11 Relative power consumption estimated per hour when: a) subcarriers are deactivated or b) entire transceivers are put in sleep mode.

Future Work and Recommendations

The network wide study showed that integrating ML algorithms into resource allocation for 6G transport networks delivers significant improvements in both operational efficiency and scalability. By using real-world traffic data from a large urban network and scaling it to future 6G scenarios, the research demonstrates that dynamic, ML-driven strategies can accurately predict hourly traffic loads and adjust the number of active transceivers accordingly. This dynamic approach resulted in up to 20% fewer transceivers required compared to static allocation, directly reducing energy consumption. Notably, placing transceivers into sleep mode during off-peak hours yielded up to 21% total energy savings, outperforming subcarrier deactivation, which achieves about 10% savings. Dynamic allocation also minimized resource over-allocation, with average excess dropping from over 86% in static cases to just 28.8% when using ML-based methods, especially during periods of low demand.

Based on these findings, it is recommended that network operators adopt ML-driven resource management, prioritize transceiver sleep mode as the primary energy-saving strategy, and invest in robust data collection and real-time analytics. Additionally, future network designs should support modular upgrades and granular resource control to adapt to variable traffic and topological conditions. These strategies are essential for meeting the high capacity and sustainability demands of 6G networks, ensuring both operational efficiency and environmental responsibility as traffic volumes and network complexity continue to grow.



1.3 Flexible functional splits

The 6G wireless network is currently in its early stages. Like 5G, it is expected to support ultra-high data rate transmissions, ultra-low latency communications, and ultra-dense connectivity. A key approach will be to virtualize the RAN architecture by splitting the protocol layers across different units; RU, DU, and CU. These splits are known as functional splits. The units are interconnected via transport networks (fronthaul and midhaul).

Each functional split offers different network performance. Functional split options 8, 7.1 and 7.2x centralize baseband functions, which bring several benefits for network operators. For instance, coordinating the same protocol layers across different cells can reduce interferences, enhance reliability, or simplify handover procedures, while reducing operational costs¹³. However, achieving maximum centralization is challenging due to limitations in the fronthaul and midhaul network capacity, diverse traffic delay requirements, functional split constraints, and asymmetric computational resources, among others^{14,15}. Therefore, flexibility in choosing the optimal functional split is crucial to meet the diverse throughput and latency demands of different traffic types and reduce operational costs, while optimizing resource utilization and network performance¹⁶. In this context, one step forward to the 3GPP/O-RAN vision is the possibility of using FFS^{17,18}. In this case, the main idea is that the functional split is dynamically adapted based on the given requirements (e.g., delay) of the traffic load, the CU/DU computational capacity, and the midhaul/fronthaul conditions.

This section is divided into three parts. First, we provide a summary of the state of the art related to flexible functional splits. Next, we introduce the use case that will be employed throughout the PROTEUS-6G project to test FFS technology. Finally, we describe the analysis and implementation carried out in the open-source end-to-end 5G-LENA network simulator, where we have analyzed the different functional splits defined by 3GPP to assess their impact when different FH capacities and air interface bandwidths are chosen. This serves as a preliminary step toward a comprehensive analysis of FFS in ns-3 5G-LENA.

1.4.1. State of the Art of FFS

Currently, some studies focus on exploiting the concept of FFS selection. For instance, articles focus on the search for the optimal functional split in order to reduce the monetary cost. In particular,¹⁹ proposes an optimization problem that jointly selects the optimal Functional Split (FS) and the most suitable routing path between each RU and CU, aiming to balance performance and associated monetary costs. The article

¹³ L. M. P. Larsen, A. Checko, and H. L. Christiansen, "A survey of the functional splits proposed for 5G mobile crosshaul networks," in *IEEE Communications Surveys and Tutorials*, vol. 21, no. 1, pp. 146–172, Jan. 2019.

¹⁴ N. Sen and A. Antony Franklin, "Slice Aware Baseband Function Placement in 5G RAN Using Functional and Traffic Split," in *IEEE Access*, vol. 11, pp. 35556–35566, 2023.

¹⁵ F. Z. Morais et al., "PlaceRAN: Optimal Placement of Virtualized Network Functions in Beyond 5G Radio Access Networks," in *IEEE Transactions on Mobile Computing*, vol. 22, no. 9, pp. 5434–5448, Sep. 2023.

¹⁶ L. H. Shen, Y. T. Huang, and K. T. Feng, "CoMP-Enhanced Flexible Functional Split for Mixed Services in Beyond 5G Wireless Networks," in *IEEE Transactions on Communications*, vol. 71, no. 7, pp. 4133–4150, Jul. 2023.

¹⁷ D. Harutyunyan and R. Riggio, "Flex5G: Flexible Functional Split in 5G Networks," in *IEEE Transactions on Network and Service Management*, vol. 15, no. 3, pp. 961–975, Sep. 2018.

¹⁸ D. Harutyunyan and R. Riggio, "Flexible functional split in 5G networks," in *13th International Conference on Network and Service Management (CNSM)*, pp. 1–9, 2017.

¹⁹ A. Garcia-Saavedra, X. Costa-Perez, D. J. Leith, and G. Iosifidis, "FluidRAN: Optimized vRAN/MEC Orchestration," in *IEEE Conference on Computer Communications, Honolulu, HI, USA*, pp. 2366–2374, 2018.



²⁰addresses cost reduction by optimizing the placement of baseband units (BBUs) within a mixed functional split scheme. Their approach minimizes the number of BBUs and fibers required. Furthermore, they also propose a heuristic algorithm for large networks. Results indicate that this mixed split strategy enhances Baseband Unit (BBU) centralization compared to fixed split setups. The authors in²¹ address the joint optimization of base station sleeping, functional split selection, and routing to minimize the operator's total cost while satisfying User Equipment (UEs) requirements. Similarly,²² focuses on reducing operational costs by optimizing the functional split, slice request admission, CU/DU placement, and traffic routing.

Furthermore, articles also focus on finding the optimal FS option, with an emphasis on maximizing the degree of centralization whenever possible. In²³, the optimization problem dynamically selects the FS in a scenario with heterogeneous traffic and limited midhaul bandwidth. The objective is to maximize the centralization of the RAN, and the results show that a dynamic FS selection provides up to 90% more centralization compared to a static configuration. The study in²⁴ selects the optimal functional split by considering traffic, slice specifications, and various placement options, aiming to maximize centralization while minimizing the number of active servers. The results show that slices with more stringent requirements do not support highly centralized functional splits. Similarly,²⁵ proposes an optimization problem that aims to find the optimal functional split to reduce computational resource usage and simultaneously maximize the centralization or aggregation level of network functions.

Most research studies in this field are based on optimization problems, where they model the network and simplify its behavior in order to find optimal solutions as efficiently and quickly as possible. However, when translating these solutions into real-world deployments, they often fall short. Therefore, it is essential to first validate these solutions through network simulators, and subsequently implement them on physical testbeds. Some works, such as²⁶, have already taken this approach, but there is still significant progress to be made in this direction. In case of, they evaluate the performance of flexible functional split concept using a lab-based testbed environment.

²⁰ M. Ahsan, A. Ahmed, A. Al-Dweik, and A. Ahmad, "Functional Split-Aware Optimal BBU Placement for 5G Cloud-RAN Over WDM Access/Aggregation Network," in *IEEE System Journal*, vol. 17, no. 1, pp. 122–133, Mar. 2023.

²¹ Y. Xu et al., "Joint Optimization of Base Station Sleeping, Functional Split, and Routing Selection in Virtualized Radio Access Networks," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Glasgow, United Kingdom, pp. 1-6, 2023.

²² M. Mushtaq, M. Golkarifard, N. Shahriar, R. Boutaba, and A. Saleh, "Optimal Functional Splitting, Placement and Routing for Isolation-Aware Network Slicing in NG-RAN," in *19th International Conference on Network and Service Management (CNSM)*, Niagara Falls, ON, Canada, pp. 1-5, 2023.

²³ H. Gupta, A. Antony Franklin, M. Kumar, and B. R. Tamma, "Traffic-Aware Dynamic Functional Split for 5G Cloud Radio Access Networks," in *IEEE 8th International Conference on Network Softwarization (NetSoft)*, Milan, Italy, pp. 297-301, 2022.

²⁴ N. Sen and A. Antony Franklin, "Slice Aware Baseband Function Placement in 5G RAN Using Functional and Traffic Split," in *IEEE Access*, vol. 11, pp. 35556–35566, 2023.

²⁵ F. Z. Morais et al., "PlaceRAN: Optimal Placement of Virtualized Network Functions in Beyond 5G Radio Access Networks," in *IEEE Transactions on Mobile Computing*, vol. 22, no. 9, pp. 5434–5448, Sep. 2023.

²⁶ Y. Alfadhli, M. Xu, S. Liu, F. Lu, P. Peng, and G. Chang, "Realtime demonstration of adaptive functional split in 5G flexible mobile fronthaul networks," in *Optical Fiber Communications Conference and Exposition (OFC)*, pp. 1–3, 2018.



1.4.2. FFS use case for energy efficiency optimization

In this subsection, we present the use case defined for evaluating FFS within the 5G-LENA simulator (see Figure 12). This use case will serve as the basis for testing and analyzing the potential improvements enabled by FFS technology. In this context, we consider a scenario with multiple gNBs, where a central gNB provides macro cell coverage, while additional gNBs serve as small cells to cover hotspot areas with highly demanding and concentrated traffic during day hours. During the day (Figure 12(a)), the number of UEs connected to the small cells is expected to be high, as these cells are located in buildings such as universities, shopping malls, and similar facilities, to cover such densely populated areas. However, during the night (Figure 12(b)), User Equipment (UE) density in these areas significantly decreases. In this context, small cell gNBs can be deactivated when UE density is low. This use case allows us to analyze how dynamically turning small cell gNBs on and off, as well as reconfiguring the splits as needed, helps to reduce costs and enhance operational efficiency, particularly in scenarios with daily variations such as day and night. This use case objective is in line with articles from ²⁷ to ²⁸.

Initially, the fronthaul capacity is configured based on the maximum number of UEs that are connected to the network during the day, representing the worst-case scenario and proper dimensioning to cover it. However, during the night, several UEs do not generate traffic and/or are disconnected from the network, so there is a significant reduction of traffic in the network. Therefore, during the night, unnecessary RUs can be turned off to reduce energy consumption. At the same time, the fronthaul capacity can be better utilized, configuring it with the optimal functional split to ensure the link capacity is maintained while maximizing the advantages of centralization. During the day, due to the higher number of UEs, the RUs need to be activated. In this situation, the available capacity with a highly centralized split will not be sufficient. Therefore, the functional split must be shifted to a more distributed configuration to meet the traffic demands. The main question we want to solve through this use case is how many small cells need to be deactivated, and which functional split can be used, so that the energy consumption is reduced while performance is guaranteed for given user densities and variations during day and night.

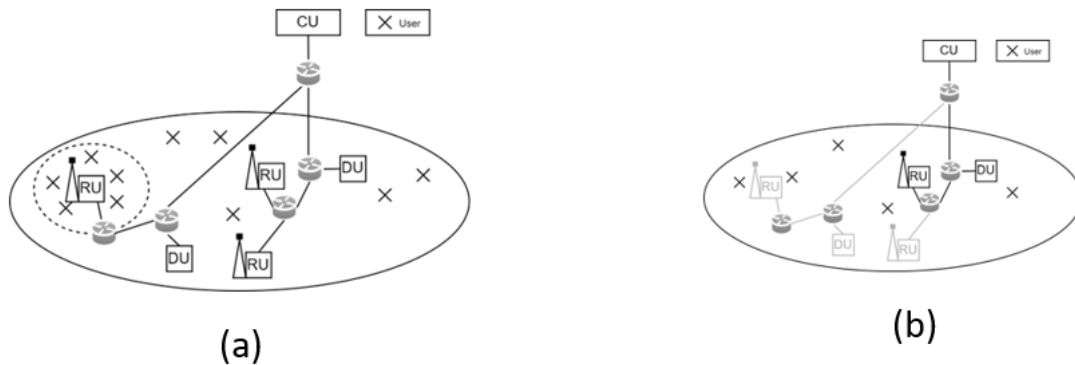


Figure 12 FFS day and night scenario.

²⁷ A. Garcia-Saavedra, X. Costa-Perez, D. J. Leith, and G. Iosifidis, "FluidRAN: Optimized vRAN/MEC Orchestration," in IEEE Conference on Computer Communications, Honolulu, HI, USA, pp. 2366-2374, 2018.

²⁸ M. Mushtaq, M. Golkarifard, N. Shahriar, R. Boutaba, and A. Saleh, "Optimal Functional Splitting, Placement and Routing for Isolation-Aware Network Slicing in NG-RAN," in 19th International Conference on Network and Service Management (CNSM), Niagara Falls, ON, Canada, pp. 1-5, 2023.



1.4.3. Impact and initial evaluation of various Functional Splits using 5G-LENA network simulator

In this subsection, we evaluate different functional splits (namely 6, 7.3, 7.2, and 7.1) within the ns-3 5G-LENA simulator. This evaluation represents an initial step toward exploring the potential of flexible functional splits. To support this, we have extended the current implementation in the 5G-LENA simulator, which already considers FH capacity constraints for the 7.2 functional split with modulation compression technique, and we have added the constraints for alternative functional split options 6, 7.1 and 7.3. Moreover, we have included the possibility to simulate functional split option 7.2 without modulation compression technique.

1.4.3.1 Implementation in 5G-LENA

5G-LENA implements a module called NrFhControl to manage and control the transmitted traffic over the fronthaul network and adjust the related centralized Medium Access Control/Physical (MAC)/ (PHY) layers. To calculate the amount of transmitted traffic, it uses the equations presented in²⁹, which were previously explained in D2.1 of PROTEUS-6G. Each functional split option corresponds to a specific equation. This logic is implemented in GetFhThr function, within NrFhControl module. Therefore, to incorporate the FH capacity requirements associated with each functional split, we have extended the logic of GetFhThr algorithm.

During the simulation, GetFhThr function is called in every slot to calculate the amount of data that will be transmitted over the fronthaul. Specifically, it calculates the FH throughput for each UE. This calculation depends on the selected functional split. To support multiple functional split configurations, we introduced a new attribute named FunctionalSplit, which allows us to specify the desired split option in the simulation example.

The throughput calculation process is detailed in Algorithm I. First, the selected FS is identified using the value provided by the FunctionalSplit attribute. Next, the algorithm iterates over all the UEs. Since UEs are assigned to potentially different Bandwidth Parts (BWPs), the throughput is calculated separately for each BWP transmitted over the FH. Finally, the total FH throughput is obtained by summing the individual values (line 10 from Algorithm I).

Algorithm 1: Total Fronthaul throughput calculation

Input: *FunctionalSplit*, set of $\{UE_i\}$ with $i = 1, \dots, n^{UE}$, set of $\{BWP_j\}$ with $j = 1, \dots, n^{BWP}$

Output: Thr^{FH}

Initialize to 0 and thr_j^{BWP} , thr_i^{UE} and Thr^{FH}

1. **For** $i = 1$ to n^{UE}
 2. thr_i^{UE} = calculate the amount of throughput transmitted for using *GetFhThr* considering selected FS
 3. **For** $j = 1$ to n^{BWP}
 4. **if** UE_i is within the BWP_j
 5. $thr_j^{BWP} = thr_j^{BWP} + thr_i^{UE}$
 6. go to next UE_i
 7. **end**
 8. **end**
 9. **end**
 10. $Thr^{FH} = \sum_{j=1}^{n^{BWP}} thr_j^{BWP}$
-

²⁹ A. Larrañaga *et al.*, "Fronthaul/Midhaul Networks: Capacity and Latency Requirements Imposed by 6G Disaggregated RANs," in *IEEE Communications Magazine*, vol. 63, no. 5, pp. 86-93, May 2025.



As a result, the primary factor influencing the calculation of throughput across different functional splits is the amount of data to be transmitted for each UE. This value accounts for both user plane data and the required control information at the RU, particularly in the case of Downlink (DL) transmission. Next, we will explain how this value is calculated in particular for split 6.

- FS option 6

In the case of functional split 6, it centralizes RRC, PDCP, RLC and MAC layer at the CU and DU. Its total FH throughput is calculated as:

In this context, PR denotes the maximum user data peak rate, and Control and Scheduling signaling rate (CR) represents the control and scheduling signaling rate. This signaling is crucial for radio resource allocation and Hybrid Automatic Repeat Request (HARQ) processes. According to³⁰, reference value for a 4G configuration is set to 150 Mbps for downlink PR and 5 Mbps for CR. To adapt these parameters for a 5G network, the sum of PR and CR should be scaled based on differences in bandwidth, number of MIMO layers, and modulation order between 4G and 5G configurations. This scaling uses the ratio of the 5G parameters to the corresponding values used in the 4G reference configuration; bandwidth, number of MIMO layers, and modulation order. These 5G values are user-specific and may vary dynamically in each time slot.

While and are directly provided by the simulator, the 5G bandwidth must be computed based on the number of Resources Blocks (RBs) assigned to each UE. In Orthogonal Frequency Division Multiple Access (OFDMA), each RB spans all OFDM symbols allocated to a single user. Given the numerology index, the subcarrier spacing is KHz. Since each RB consists of 12 subcarriers, the effective bandwidth for a user is calculated as:

$$BW_i = \frac{N_i^{RB}}{N_i^{sym}} \cdot 12 \cdot 2^\mu \cdot 15k$$

1.4.3.2 Evaluation scenario

We have analyzed the different splits from option 6 to 7.1 considering a scenario with one gNB and 14 randomly distributed UEs. The evaluation has been carried out considering a different capacity in the FH in each simulation, from no limiting to highly limiting capacity: 1 Tbps, 100 Gbps and 1 Gbps. The amount of data transmitted through the FH is controlled by the postponing FH control method. This method checks if the scheduled data fits in the FH, and if it doesn't, postpones it and tries to schedule it again in the next slots. Below we list the parameters we have configured in 5G-LENA network simulator to evaluate the different functional splits:

- The channel bandwidth is equal to 400 MHz, and gNB transmission power is equal to 30 dBm.
- The frames are configured with 120 kHz of sub-carrier spacing and they follow a Time Division Duplexing (TDD) pattern.
- The Modulation and Coding Scheme (MCS) is adaptive, and it follows Table 2 in TS 38.214 which supports up to 256-QAM.
- Seven UEs run the Virtual Reality (VR) application and the other seven the Cloud Gaming (CG) application. The applications specifications:
 - VR: The 5QI is set to 89 which corresponds to high-priority (DGBR_VISUAL_CONTENT_89) and the average data bit rate offered is equal to 45 Mbps.
 - CG: The 5QI is set to 3 which means medium-priority (GBR_GAMING) and the average data bit rate offered is equal to 30 Mbps.

³⁰ 3GPP TSG RAN WG3, "CU-DU split: Refinement for Annex A (Transport network and RAN internal functional split)," R3-162102, Sophia Antipolis, France, 2016.



- Antenna configuration: It is configured with uniform planar arrays.
 - gNB: It is configured with 4 rows and 8 columns of dual-polarized antenna elements, organized into 4 horizontal and 2 vertical antenna ports, respectively, being a total of 16 antenna ports.
 - UE: It is configured with 2 rows and 4 columns of dual-polarized antenna elements, spread in 2 horizontal and 1 vertical antenna port, respectively, with dual-polarized configuration, being a total of 4 antenna ports.
- Scheduling and radio access mode: We use the Proportional Fair (PF) MAC scheduler along with OFDMA access mode to allocate radio resources to transmit the packets.

1.4.3.3 Simulation results

shows the comparison of the throughput transmitted through the FH when different splits are configured, and the FH capacity does not limit the data transmission. In the case of split 6, the maximum transmitted throughput is 10 Gbps, while in the case of split 7.1 the maximum throughput can reach up to 170 Gbps. This figure shows that the more the protocol layers are centralized, the higher the throughput through the FH is, and it also shows that the transmitted throughput with splits 7.2 and 7.1 is much higher with respect to split 6 or 7.3.

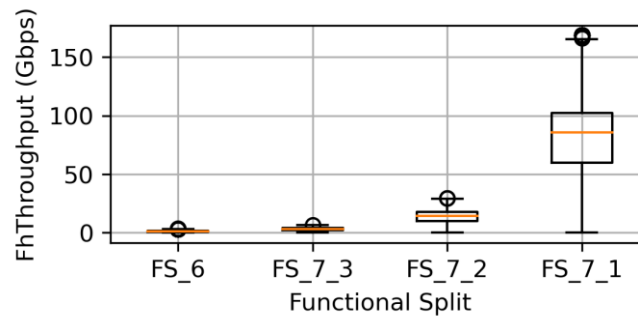


Figure 13 FH throughput for different functional splits when FH capacity is 1Tbps.

shows the end-to-end application throughput when there is no limitation in the FH nor in the channel bandwidth. For the VR application (blue) the average offered rate is 45Mbps while for the CG application (red) the average offered rate is 30Mbps. The received application average rate or AppThroughput is almost equal to the offered average rate for both applications because the amount of transmitted data fits within the available capacity without needing to postpone or discard data.



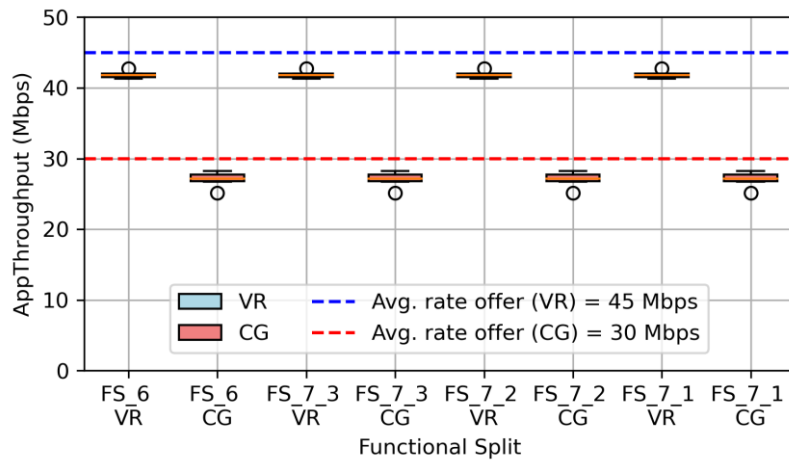


Figure 14 Application throughput for different functional splits when FH capacity is 1Tbps.

Figure 14 shows a comparison of end-to-end application throughput when the capacity at the FH is either A) 100 Gbps or B) 1 Gbps. With a 100 Gbps FH capacity, the received application throughput nearly matches the offered throughput for all functional splits except split 7.1. This is because the data rates for those splits fit within the available FH capacity, allowing applications to operate at their maximum performance. In the case of split 7.1, however, the potential FH throughput without fronthaul constraints can reach up to 170 Gbps. Since the FH capacity is limited to 100 Gbps, this results in packet delays or packet losses, causing the UE to receive fewer packets than were originally generated. When the FH capacity is reduced to 1 Gbps, it becomes a bottleneck for all splits. As a result, the end-to-end application throughput is lower than the offered throughput.

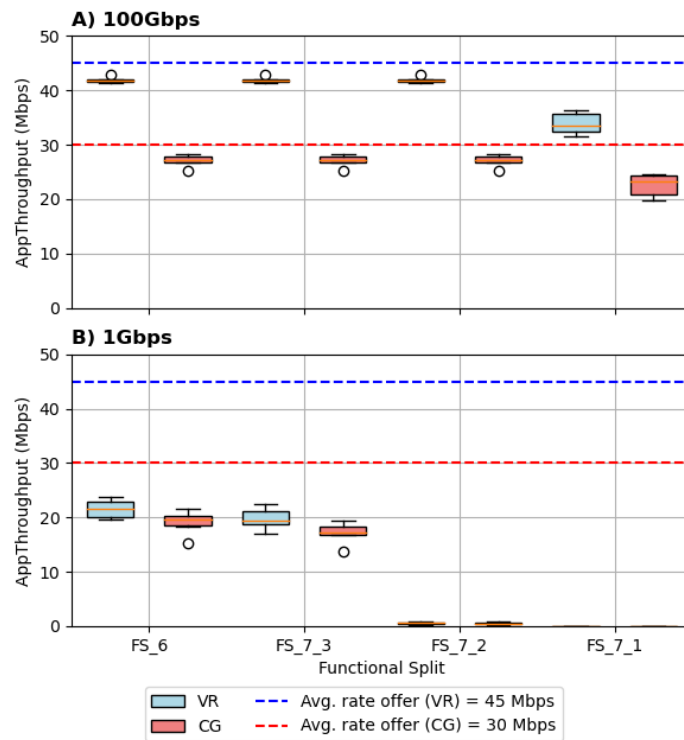


Figure 15 FH throughput and Application throughput for different functional splits and FH capacities.



1.4.4. Conclusions and Future work

In this section, we first reviewed the state of the art of FFS. We identified that one of the main needs is to perform FFS simulations using more advanced and comprehensive simulators, rather than relying solely on optimization-based approaches that model the 5G network in a simplified way. This is essential to realistically evaluate the behavior and performance of flexible functional splits in real-world scenarios. For this reason, for this work we use the well-known Open-source end-to-end 5G-LENA simulator that simulates 5G NR advanced technology with 3GPP-aligned protocols. This tool provides a detailed and comprehensive simulation of the different layers of the 5G network, making it well-suited for evaluating FFS in realistic RAN scenarios. We then outlined the specific use case we will focus on within the PROTEUS-6G project to assess the benefits of implementing flexible functional splits. Finally, we presented the initial steps taken using the 5G-LENA simulator, where we analyzed different functional splits individually. This serves as a first step for future work toward an implementation in which the functional split can adapt dynamically to changing conditions such as channel quality, deployment scenario, and network requirements.



2. Towards cell free Fronthaul Network planning and optimal configuration

2.1 On cell-free mobile networks

Cell-free in next-generation cellular networks refers to a network architecture where the concept of traditional cells is removed or significantly relaxed. Instead of dividing the coverage area into cells served by individual base stations, a large number of distributed RUs jointly serve all users in the area over the same time-frequency resources in a coordinated and coherent manner. The fact that the RUs generate fronthaul flows using a functional split below option 6 (splits 7.x (intra-Phy) or 8 (Phy)) is essential for the cell-free coordination to take place, as those splits enable the centralization of the MAC layer of the multiple RUs into a common DU. In the cell free paradigm for 6G, the Central Processing Unit (CPU) that coordinates the transmission and reception across RUs, is an entity that embeds or controls the DU function associated to each RU.

In cell-free users are not attached to a specific cell or base station (that is why it is also called user-centric communication) but dynamically and simultaneously served by multiple RUs distributed across the area. Although RUs do not perform any association function, since normally not all RUs in an area will be serving a single user, we may consider that virtually cell-free performs multiple dynamic associations, which may change every Transmission Time Interval (TTI). The reason for this association to just a subset of RUs may be due to the lack of received power for the communication to be effective (low coverage), or simply because the scheduling mechanism decides so at the current instant. For instance, it is not effective to try to associate and use the radio resources of a neighboring overloaded RU.

In cell-free CoMP is applied at scale. This requires intensive use of the latest dynamic beamforming single-user or multi-user MIMO technologies. Sticking to current single-user MIMO systems, a RU has the capability to direct a beam toward a UE at a given time, with a given frequency f_1 which it can't reuse in the same antenna array (see Figure 16).

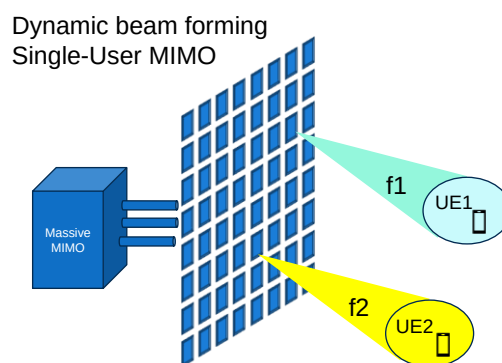


Figure 16 Dynamic beam forming in a single-user MIMO smart antenna array

Another antenna array of the same tower or in a neighboring cell (see Figure .17(c)) as long as the beams do not interfere.

With this technology several CoMP schemes can be assumed as part of cell free. Figure 17 describes four kinds of CoMP.



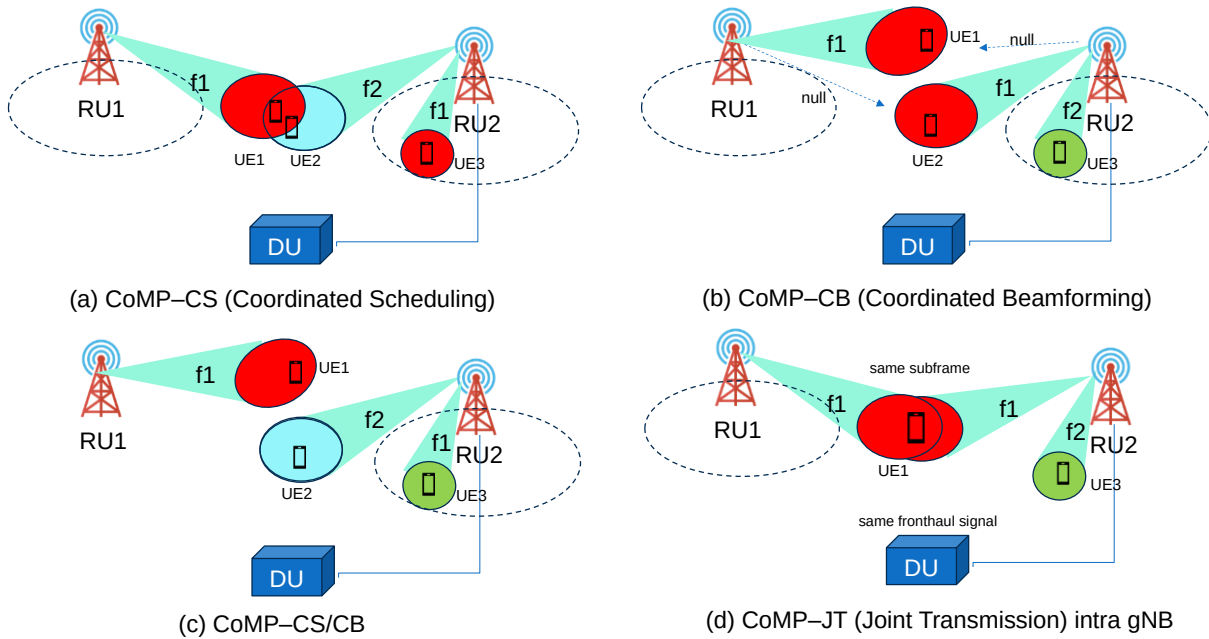


Figure 17 Coordinated multi-point for interference avoidance (a)(b)(c) and throughput boosting (d) (adapted from³¹)

In CoMP-Coordinated Scheduling (CS) (Figure 17 (a)) at a given instant (RU1) and (RU2) direct, in a coordinated way, beams in different frequencies f_1 and f_2 to (UE1) and (UE2), located in the interfering region between two cells. At the same time RU2 may be re-using frequency f_1 for another UE, (UE3), that is within the RU2 central non-interfering zone (marked in dashed line) where lower power is required.

The capability to generate radiation nulls by the antenna arrays allows to re-use the frequency at a given instant in the interfering area in the so-called CoMP-Coordinated Beamforming (CB) technique (figure 17. (b)). CoMP-CS/CB spans the previous two concepts (Figure 17. (c)). The main advantage of schemes (a)(b)(c) is the reutilization of frequencies among neighboring cells and interference avoidance.

Finally Figure 17. (d) describes CoMP-Joint transmission where the same signal for the same UE is radiated from/to several RUs expecting to reach the UE at the same time (imposing the most strict time synchronisation among RUs of any of the schemes), providing a constructive redundancy that leads to a higher throughput. It should be noted that CoMP-Joint Transmission (JT) is the case generating the highest fronthaul traffic of all four schemes.

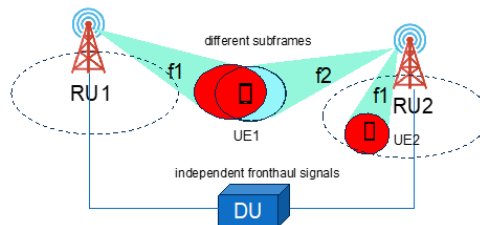


Figure 18 Cell free capacity aggregation from multiple RUs

³¹ <https://www.netmanias.com/en/post/blog/6558/comp-lte-lte-a/comp-1-comp-types-cs-cb-jt-and-dps>

In cell free, UEs are assumed to be potentially served by multiple RUs (Figure 18), so that, the coordination allows UE1 and UE2 to use both frequencies f_1 and f_2 to carry data for the same UE in a coordinated way. This can also be seen as a CoMP mechanism, which is enabled by the centralization of the MAC layer of the DUs serving both RUs. A special frequent case in cell free where a UE can take advantage of two RUs is when RU2 is a micro cell within a macro cell RU1. Normally, the UE in the micro cell would use only the capacity of that cell, but the cell free scheme allows to use at the same time resources of the container macro cell.

This enhanced throughput mechanism, along with interference avoidance thanks to CoMP, is the main feature of cell free and the target of the algorithms described in this deliverable and that will be developed in the final one (D2.5). It should be noted that thanks to the centralized beam forming and MAC controls we may assume perfect coordination among overlapping cells to serve each user, yielding no interference. In other words, for the sake of modelling the target aspects of cell free that we wish to address (throughput and fronthaul traffic), we can assume that the neighboring cells use completely separate carrier frequencies. This way we can use the 5G-LENA simulator, as described below.

2.2 Planning networks for cell free fronthaul traffic

The fronthaul traffic rates generated by different functional splits required for planning for single associations (non-cell-free) can be obtained from D2.1 as a function of the load from splits 7.2 and above. The expected rates for cell-free will be simulated under the assumptions that follow and are expected to be the same for high loads (as the UEs will very likely only associate with a single gNB though a single RU). For low and medium UE loads, however, we expect a multiplying factor n in the fronthaul traffic, where n is the average RUs within coverage. This is the expected behavior because at lower loads, UEs will attach to multiple RUs, leading to higher individual rates than without cell free.

Another key point in the methodology is which traffic profile UEs have. To provide results as general as possible, we shall assume the full buffer assumption. In other words, each active UE tries to obtain as much throughput as possible without any pause. Under the full buffer assumption, the load is defined as average load of the cells in an area, since the load in a cell becomes 100% as soon as an active user equipment enters the cell.

2.3 Utility Functions and Fairness Models

In cellular traffic works, finding optimal configurations needs to define a utility function. If a fairness scheme is not in place, the natural objective would be to maximize the overall throughput. However, this would translate into an unfair scheme where the RUs serve their respective closest UEs, leading to starvation of distant UEs suffering from low coverage. To avoid this situation, scheduling algorithms apply proportional fair share either to radio resources or in terms of rate within the cell. Here comes another open issue in cell free: if the concept of cell does not exist, how should we apply fair sharing? on a global base or on a per-cell base? Given the fact that the LENA scheduler works on the per-cell fairness.

To model user satisfaction under different fairness regimes, our study will use the **α -fair utility function framework**, a well-established family of utility functions in network resource allocation. These functions allow us to flexibly represent trade-offs between overall throughput and fairness among users by tuning a single parameter α .

The **utility function** of a user u , as a function of its aggregated rate R_u , is defined as:



$$U_{\alpha}(R_u) = \begin{cases} \frac{R_u^{1-\alpha}}{1-\alpha}, & \text{if } \alpha \geq 0, \alpha \neq 1 \\ \log(R_u), & \text{if } \alpha = 1 \end{cases}$$

Key fairness interpretations for specific values α of are:

- $\alpha = 0$: **Max throughput** (linear utility).
- $\alpha = 1$: **Proportional fairness** (log utility).
- $\alpha \rightarrow \infty$: **Max-min fairness**, favouring the least-served users.

The total **network utility** is the sum of utilities across users, for both uplink and downlink:

$$f_{\alpha} = \sum_u [U_{\alpha}(R_u^{\text{DL}}) + U_{\alpha}(R_u^{\text{UL}})]$$

The aggregated **rates** R_u^{DL} and R_u^{UL} are defined as:

$$R_u^{\text{DL}} = \sum_{b \in \mathcal{B}_D} r_{ub}^{\text{DL}} \cdot y_{ub}^{\text{DL}}$$

$$R_u^{\text{UL}} = \sum_{b \in \mathcal{B}_U} r_{ub}^{\text{UL}} \cdot y_{ub}^{\text{UL}}$$

Where:

- $\mathcal{B}_{\text{DL}}, \mathcal{B}_{\text{UL}}$ are the sets of downlink and uplink base stations,
- y_{ub} is the fraction of resources allocated from base station b to user u ,
- r_{ub} is the **achievable data rate** between user u and base station b , given by:

$$r_{ub} = W \cdot \log_2(1 + \text{SINR}_{ub})$$

In this expression:

- W is the transmission bandwidth,
- SINR_{ub} is the signal-to-interference-plus-noise ratio for the link between user u and base station b .

This formulation links physical layer performance (through Signal to Interference plus Noise Ratio (SINR)) with higher-level resource allocation and fairness mechanisms. By using precomputed SINR values from the network simulation, we are able to evaluate utility under different user–base station association configurations and resource allocation policies. This enables the study of how cell-free architectures can impact fairness and performance at scale.

In a complementary approach (to the simulations with LENA) to evaluating user associations in cell-free scenarios, we developed a custom 5G network simulator in MATLAB, implemented at UC3M, which supports dual associations per user. This simulator enables each user to be simultaneously connected to a maximum of two base stations. The figure illustrates a snapshot of a simulation, showing the coverage map generated through Voronoi tessellations. Each colored region represents the coverage area of a



base station. The blue crosses denote user positions, and the lines extending from each user indicate their active associations to two different base stations. The association algorithm follows a two-step strategy: (1) the first association is determined by the user's position within the coverage map. Each user initially connects to the base station covering its current location; (2) to determine the second association, the base station involved in the first association is temporarily removed from the topology. A new coverage map is then computed, and the user is associated with the base station now covering its position. This technique ensures spatial diversity in user connections while maintaining a coverage-aware logic in the association process.

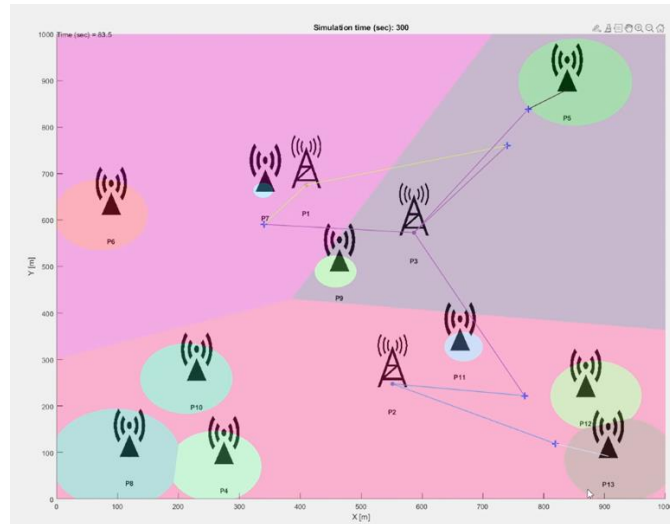


Figure 19 Coverage map and dual user associations in the custom 5G MATLAB simulator. Voronoi regions define primary coverage; users (blue crosses) connect to two base stations based on sequential coverage re-evaluation.

2.4 Optimal cell free fronthaul associations

The focus of the algorithm will be on finding optimal associations in a cell-free scenario using either an analytical or a much more realistic simulation performed by 5G-LENA. The target is finding the configuration at a given time that maximizes the aggregate utility. Let us first review the characteristics of LENA/5G-LENA.

2.4.1 5G-LENA and its application to model cell free scenarios

LENA (Long-Term Evolution (LTE) Evolved Packet Core (LTE-Evolved Packet Core (EPC)) Network Simulator) is an open-source simulation framework developed and maintained by the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC). Built as a module within the ns-3 discrete-event network simulator, LENA provides a comprehensive and realistic end-to-end simulation environment for 4G LTE networks, including both the EPC and the LTE Radio Access Network.

LENA implements the full protocol stack—from the physical layer up to application-level traffic generation—enabling the study of link-level, system-level, and network-level behaviors. It supports key LTE features such as dynamic scheduling, Radio Resource Management (RRM), handover procedures, Quality of Service (QoS), heterogeneous networks (HetNets), Self-Organizing Network capabilities, and mobility modeling. It includes support for multiple cells, inter-cell interference coordination, spectrum reuse, and realistic channel models.

Originally released in 2011 and later integrated into the official ns-3 LTE module (starting from ns-3 version 3.19), LENA has been widely adopted in both academic and industrial contexts for testing, benchmarking, and validating novel algorithms, protocols, and network configurations under a wide range of scenarios.

To address the increasing complexity of next-generation networks, CTTC has also developed 5G-LENA, an evolution of the LENA simulator aimed at modeling 5G NR networks. 5G-LENA expands the LTE



capabilities by supporting advanced 5G features such as flexible numerology, beamforming, massive MIMO, carrier aggregation, (NR-V2X) (vehicular communications), (NR-U) (unlicensed spectrum), and dynamic TDD, among others. It enables researchers and developers to simulate complex 5G use cases, including ultra-reliable low-latency communications (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communications (mMTC).

Both LENA and 5G-LENA are actively developed and maintained by CTTC's OpenSim research unit and are distributed under the GNU GPLv2 license. Extensive documentation, tutorials, and example scenarios are provided to facilitate reproducibility and adoption by the research community.

While 5G-LENA offers a detailed and realistic simulation framework for 5G NR networks, it does not natively support cell-free architectures or scenarios where users are simultaneously associated with multiple base stations. To overcome this limitation and still benefit from 5G-LENA's high-fidelity MAC and physical layer modeling, we devised an alternative simulation strategy for our studies on cell-free connectivity and multi-association scenarios. Specifically, we computed all possible user-to-base-station association combinations within a given scenario. Assuming a full-buffer traffic model—where users always have data to transmit—we conducted a series of individual simulations, each representing a single-user-to-single-base-station association. This approach allows us to approximate the performance of cell-free systems by aggregating the throughput results from these independent simulations. In doing so, we estimate the overall throughput of a hypothetical cell-free configuration by summing the throughput achieved in each corresponding single-association simulation, thereby emulating the effect of users being simultaneously connected to multiple base stations.

2.4.2 Modelling cell-free with the support of 5G-LENA

To find the optimum for a new situation (distribution of UEs in an area) we shall apply two main strategies. Both of them rely on the generation of cases such as the one depicted in Figure 20, which will be input to 5G-LENA, in order to generate data with which apply learning schemes.

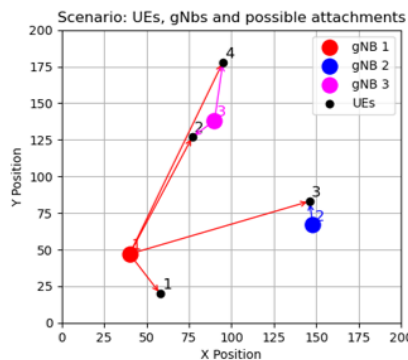


Figure 20 A sample simulation case with 5G-LENA

Such scenario leads to a number of single associations with their respective throughputs proportional fair sharing within each cell as per the estimation obtained from 5G-LENA as exemplified by the next table. In this table, combination [1,1,1,1] means that the four UEs attach to the macro cell 1, whereas [1,1,1,3] means that UE1, UE2 and UE3 are associated to cell 1 and UE 4 is associated to cell 3. The other columns represent the respective throughputs obtained through those associations. To model cell free configurations, it is necessary to add coherently the throughputs as depicted in the post processing part of the table.



Table 2 Possible combinations to simulate and post-processing choices

Possible combinations to simulate	1	2	3	4
[1,1,1,1]	62 Mbps	58 Mbps	20 Mbps	20 Mbps
[1,1,1,3]	85 Mbps	22 Mbps	51 Mbps	255 Mbps
[1,1,2,1]	85 Mbps	22 Mbps	255 Mbps	51 Mbps
[1,1,2,3]	127 Mbps	76 Mbps	255 Mbps	255 Mbps
[1,3,1,1]	85 Mbps	255 Mbps	4 Mbps	51 Mbps
[1,3,1,3]	127 Mbps	127 Mbps	47 Mbps	127 Mbps
[1,3,2,1]	127 Mbps	255 Mbps	255 Mbps	47 Mbps
[1,3,2,3]	255 Mbps	127 Mbps	255 Mbps	127 Mbps
Post-Processing				
UE 1 = {1} UE 2, UE 4 = {1,3} UE 3 = {1,2}	62 Mbps	58 + 127	20 + 255 Mbps	20 + 127 Mbps
UE 1, UE 2 = {1} UE 3 = {1,2} UE 4 = {1,3}	62	58	20 + 255 Mbps	20 + 255 Mbps
...				

It should be noted that this scheme is not capable of modelling unbalanced resource allocations within a cell due to the current availability only of fair schedulers. It might be the case that such unbalanced configurations at the cell scale, provide maximal utility at the larger area scale.

2.4.3 Optimal configuration finding methodology

Each simulation run will generate data in this sample format.

```
[
  { "combination": [
    [ 1, 0, 0 ],
    [ 1, 0, 0 ],
    [ 1, 0, 0 ] ],
    "BSs": [
      { "x": 69, "y": 99, "power_dBm": 35 },
      { "x": 118, "y": 41, "power_dBm": 25 },
      { "x": 51, "y": 170, "power_dBm": 25 } ],
    "UEs": [
      { "distancesBS": [112.2007130102122, 41.97618372363071, 175.2369823981228],
        "throughputs": [37.819733, 0, 0 ]},
      { "distancesBS": [60.530983801686226, 51.19570294468082, 111.01801655587259],
        "throughputs": [84.48, 0, 0 ]},
      { "distancesBS": [...
        ....
      ]
    }
  ]
}
```



} ,
...
]

To make the system find optimal configurations, two main approaches will be applied

1. Regression mechanisms: feed in the post-processed optimal and predict the optimal associations for the new cases with classic ML regression schemes such as linear regression, k-Nearest Neighbors (KNN), Random Forests, etc. This will use the LENA simulator as depicted above.
2. Reinforcement Learning: the objective is to try to find optimum associations from variations in the raw cases, instead of learning from the optimal. In this case an analytical model of the channel will be used due to the time required for each simulation (several minutes).

2.4.4 Combinatorial problem and approach

This simulation methodology, while effective in approximating cell-free behaviour using a tool that does not support it natively, introduces significant combinatorial complexity. As each user in the scenario may potentially be associated with multiple base stations, the total number of possible user-to-base-station combinations grows exponentially with the number of users and base stations. More precisely, if each user can connect to any subset of the available base stations, the number of combinations can scale as 2^B , where U is the number of users and B the number of base stations. Even when limiting associations to a fixed number of serving nodes per user, the total number of required simulations remains high due to the need to evaluate each possible combination individually. As a result, the total simulation time increases linearly with the number of cases, while the memory and computational demands grow significantly, especially when each simulation involves detailed physical layer modelling. This combinatorial explosion can rapidly exceed the capabilities of a single-machine setup, requiring either High Performance Computing (HPC) infrastructure or the use of parallel execution frameworks to keep simulation times within practical limits. Therefore, although this method provides a way to exploit LENA's precision in scenarios beyond its native design, it imposes substantial computational overhead, which must be carefully managed to ensure the feasibility of the overall study.

Empirical tests performed on simplified scenarios consisting of 3 users and 3 base stations show that each individual simulation—representing a single user-to-base-station association configuration—can take between **30 and 120 seconds each** to complete, depending on system parameters and channel conditions. Given the combinatorial growth in the number of possible associations, the total simulation workload can quickly escalate to hundreds or thousands of cases. For this project, we have leveraged a dedicated high-performance server to run simulations continuously over a period of more than one month. This extensive simulation campaign has generated a rich dataset of user scenarios and their corresponding optimal cell-free association combinations. These datasets serve as the foundation for training artificial intelligence models, which learn to approximate optimal network configurations from the simulated data.

It is important to distinguish between the AI training phase and the inference phase: training involves computationally intensive processes that require large amounts of data and time to develop accurate models, whereas inference—the application of trained models to new user positions—can be performed rapidly and in real time. This separation allows the system to benefit from the high accuracy of physics-based simulations during training, while enabling fast decision-making when deployed in operational environments.

Managing the simulations and the cell-free association combinations arising from a scenario defined by user and base station positions presents an extremely complex challenge, both mathematically and from a software development perspective. From a mathematical standpoint, the core difficulty lies in representing and manipulating the vast combinatorial space of user-to-base-station associations. Each



cell-free configuration corresponds to a unique subset of possible associations, and identifying which individual simulation results must be aggregated to reconstruct the performance of that configuration requires a precise mapping between these subsets and the simulation outputs. This involves defining suitable data structures and indexing schemes capable of efficiently enumerating and referencing the exponential number of possible combinations, especially when multiple users may be connected simultaneously to multiple stations. On the programming side, implementing this logic demands designing robust, scalable, and maintainable code that can automatically generate the required sets of individual simulations, manage their execution, and then correctly aggregate the results to form composite performance metrics for each cell-free configuration of interest. The software must handle the complexity of tracking which single-association simulations correspond to which higher-order combinations, ensuring no overlaps or omissions in the aggregation process. Additionally, this system must be flexible enough to incorporate constraints (such as limiting the number of associations per user) and to efficiently prune infeasible or low-impact configurations to reduce computational overhead. Achieving all this requires careful architectural planning, modular design, and extensive testing, as well as optimization strategies such as parallelization and caching to manage runtime and memory consumption. Overall, the apparatus of mathematical modelling and code engineering needed to navigate this combinatorial space and produce accurate aggregated results represents a significant technical endeavour critical to the success of simulating realistic cell-free wireless network scenarios.

Simulation Data Aggregation and Combination Management

The management of simulation results and the reconstruction of cell-free scenarios are handled through a structured workflow implemented in Python. The core idea is to first gather results from individual simulations (which must be performed first), each representing a specific assignment of UEs to base stations BSs. These simulation outputs include user assignments and their corresponding throughput values, which are parsed and stored in memory in a structured format.

For each simulation file, user-to-base-station assignments are extracted and represented as binary vectors, where each vector indicates which base station a given user is connected to. Throughput data per user is similarly extracted. This data serves as the building blocks for more complex cell-free association scenarios.

To explore multiple user-to-base station associations simultaneously—typical of cell-free architectures—the system generates all possible combinations of user connections as binary matrices. Each matrix row corresponds to a user, and each column corresponds to a base station; a value of 1 indicates an active connection. These combinations represent candidate cell-free configurations.

The aggregation process involves iterating over these binary combinations and summing the throughput contributions of the relevant individual simulations to estimate the performance of each candidate cell-free scenario. Specifically, for each user and base station pair, the code checks which individual simulation result matches the combination's assignment pattern and retrieves the corresponding throughput. These per-user, per-base-station throughputs are summed to produce an overall throughput estimate for the combination.

Additionally, spatial data such as the positions of users and base stations and distances between them are incorporated to enrich the output, allowing further analysis of spatial effects on performance.

This approach requires careful and precise handling of assignment patterns and throughput data to ensure that only relevant simulations are aggregated, avoiding overlaps or omissions. It also reflects the substantial complexity of managing combinatorial explosion in the number of possible associations, highlighting the need for optimized data structures and algorithmic checks.

The resulting aggregated dataset captures the performance of a wide range of cell-free configurations based on physically accurate, individual simulation results. This dataset then serves as the foundation



for training machine learning models that can infer optimal associations in new scenarios with significantly reduced computational effort.

The following steps, summarize the used methodology:

Throughput Estimation Using 5G-LENA for Cell-Free Setups

- 1. Start**
- 2. Input: Network Scenario Configuration**
 - a. Define UE and base station (gNB) positions
 - b. Set simulation parameters (e.g., power, full-buffer traffic model)
- 3. Enumerate All Unique UE-to-gNB Assignment Scenarios**
 - a. For a given number of UEs and gNBs, generate all possible combinations where each UE is assigned to exactly one gNB
 - b. Each combination represents a complete scenario (with all UEs present)
 - c. These combinations define the set of individual simulations to run
- 4. Run Individual Simulations (5G-LENA)**
 - a. Each simulation corresponds to a specific one-to-one assignment of all UEs
 - b. Output includes throughput per UE and the gNB assignment
- 5. Parse Simulation Results**
 - a. Extract throughput and assignment data
 - b. Convert assignments into binary matrix form
 - c. Store results in a structured dataset for combination
- 6. Define Target Cell-Free Configurations**
 - a. Generate cell-free association patterns:
 - i. Each UE can connect to multiple gNBs simultaneously
 - ii. Represented as binary matrices
- 7. Aggregate Throughput for Cell-Free Combinations**
 - a. For each target cell-free combination:
 - i. Identify matching elements across the precomputed individual simulations
 - ii. Aggregate throughputs from relevant entries
 - iii. Compute total throughput per UE
- 8. Save Aggregated Results**
 - a. Store each cell-free configuration and its throughput performance
 - b. These form the dataset to train AI models for fast inference in future scenarios



Conclusions

This deliverable outlines the foundational steps toward optimized 6G transport networks using a combination of algorithmic, architectural, and simulation-based approaches. Key findings include:

- 1) Dynamic resource allocation using ML, particularly Random Forest-based predictions, enables up to 20% transceiver savings and 21% energy reduction, compared to static allocation.
- 2) Flexible Functional Splits (FFS) offer improved operational efficiency when dynamically adapted to day/night traffic patterns. Initial evaluation using the 5G-LENA simulator demonstrates the feasibility and performance of splits 6, 7.1, 7.2, and 7.3 under various fronthaul constraints.
- 3) Cell-free network planning shows that advanced coordination and multi-association schemes can significantly increase throughput while maintaining fairness. However, modeling these scenarios requires high computational effort due to the combinatorial explosion of user-to-base-station associations.

These studies validate the need for intelligent, modular, and adaptable transport systems in 6G, where AI-driven orchestration and scalable network architectures can support extreme performance demands. Future efforts will expand simulation coverage, explore hybrid learning strategies, and refine orchestration mechanisms for real-time deployment.



