Deliverable 2.2

Evolved Architecture and Power Enhancement Mechanisms

July 31, 2024

Co-funded by
the European Union

6GSNS

Contractual Date of Delivery:	**June 30, 2024**
Actual Date of Delivery:	**July 31, 2024**
Editor(s):	**Jose A. Ayala-Romero, J. Xavier Salvat (NEC)**
Author(s):	**Alejandro Blanco, Pablo Suárez-Reyero, Guillermo Bielsa (TSA)** **Jose A. Ayala-Romero, J. Xavier Salvat, Andres García-Saavedra, Lanfranco Zanzi, Xi Li (NEC)** **Vladica Sark, Mert Özates, Jesús Gutiérrez (IHP)** **Miguel Catalán-Cid, Esteban Municio (i2CAT)** **German Castellanos, Revaz Berozashvili, Simon Pryor (ACC)** **Anna Umbert, Juan Sánchez-González, Jordi Pérez-Romero, Oriol Sallent (UPC)** **Joss Armstrong (LMI)** **Keith Briggs (BT)** **Mir Ghoraishi (GIGASYS)**
Work Package:	**WP2**
Target Dissemination Level:	**Public**

This work is supported by the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101097083, BeGREEN project. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or SNS-JU. Neither the European Union nor the granting authority can be held responsible for them.

Revision History

Revision	Date	Editor / Commentator	Description of Edits
0.1	30.01.2024	Jose Ayala Romero, Josep Xavier Salvat (NEC)	Table of contents/ work distribution
0.2	16.04.2024	Anna Umbert (UPC)	Added B5G RAN enhanced with relay nodes
0.3	25.04.2024	Jose Ayala Romero, Josep Xavier Salvat (NEC)	Added first version of Sec. 3.4
0.31	27.04.2024	Alejandro Blanco Pizarro, Pablo Suarez Reyero (TSA)	Added first version of Sec. 3.1
0.32	28.04.2024	Jesús Gutiérrez, Vladica Sark (IHP)	Added first version of Sec. 3.2
0.4	28.04.2024	Jesús Gutiérrez (IHP)	Added first version of Chapter 2
0.5	05.05.2024	Miguel Catalan (i2CAT), Jesús Gutiérrez (IHP), Juan Sánchez (UPC)	Added first version of Chapter 4
0.6	14.05.2024	Anna Umbert (UPC)	Revision and final version of Sec. 3.3
0.62	20.05.2024	Jose Ayala Romero, Josep Xavier Salvat (NEC)	Revision and final version of Sec. 3.4
0.63	26.05.2024	Alejandro Blanco Pizarro, Pablo Suarez Reyero (TSA)	Revision and final version of Sec. 3.1
0.64	29.05.2024	Jesús Gutiérrez (IHP), Vladica Sark (IHP)	Revision and final version of Sec. 3.2
0.7	30.05.2024	Jesús Gutiérrez (IHP)	Revision and final version of Chapter 2
0.8	31.05.2024	Miguel Catalan (i2CAT), Jesús Gutiérrez (IHP), Juan Sánchez (UPC)	Revision of Chapter 4
0.81	15.06.2024	Jesús Gutiérrez (IHP), Anna Umbert (UPC)	Revision and final version of Chapter 2
0.82	28.06.2024	Esteban Municio (I2CAT), Jesús Gutiérrez (IHP),	Revision and final version of Chapter 4
0.9	15.07.2024	Keith Briggs (BT)	Overall review
1.0	31.07.2024	Mir Ghoraishi (GIGASYS) Simon Pryor (ACC)	Final review, proof-reading, and submission to the EC

Table of Contents

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Project
6G-IA	6G Infrastructure Association
AI	Artificial Intelligence
AP	Access Point
AWGN	Additive White Gaussian Noise
B5G	Beyond 5G
BER	Bit Error Rate
BS	Base Station
CDF	Cumulative Distribution Function
CF-mMIMO	Cell-Free massive Multiple-Input Multiple-Output
CPU	Central Processing Unit
CSI	Channel State Information
DL	Downlink
D-MIMO	Distributed Multiple Input Multiple Output
DoW	Description of Work
DU	Distributed Unit
FEC	Forward Error Correction
GPU	Graphics Processing Unit
IAB	Integrated Access and Backhaul
ISAC	Integrated Sensing and Communication
ISM	Industrial Scientific and Medical
ITU	International Telecommunication Union
KPI	Key Performance Indicator
KV	Key Value
KVI	Key Value Indicator
L2	Layer 2
LDPC	Low Density Parity Check
LLC	Last Level Cache
LMMSE	Linear Minimum Mean Square Error
LS	Least Squares
LTE	Long Term Evolution
MIMO	Multiple-Input Multiple-Output
ML	Machine Learning
mmWave	Millimetre Wave
near-RT	near Real-Time
NFV	Network Function Virtualisation
non-RT	non-Real-Time
O-DU	Open Distributed Unit
O-RAN	Open RAN
O-RU	Open Radio Unit
OFDM	Orthogonal Frequency-Division Multiplexing
QAM	Quadrature Amplitude Modulation
PHY	Physical layer
PoC	Proof of Concept
RAN	Radio Access Network
RIC	Radio Interface Controller

RIS	Reconfigurable Intelligent Surface
RT-RIC	Real Time RAN Intelligent Controller
RU	Radio Unit
SDR	Software Defined Radio
SINR	Signal-to-Interference-plus-Noise Ratio
SMO	Service Management and Orchestration
SMT	Simultaneous multithreading
SNS JU	Smart Networks and Services Joint Undertaking
TF	Task Force
TMV	Test Measurement and Validation
TRL	Technology Readiness Level
UC	Use Case
UE	User Equipment
UL	Uplink
UPF	User Plane Function
vBS	Virtual Base Station
vRAN	Virtualized RAN
WP	Work Package
ZF	Zero Forcing

Executive Summary

Current document, BeGREEN D2.2, builds upon the groundwork established in BeGREEN D2.1. The latter offered an extensive overview of the State-of-the-Art (SotA) for Radio Access Networks (RANs) from an energy-efficiency standpoint, while proposing solutions for RAN optimization at hardware, link, and system levels.

BeGREEN D2.2 opens the discussion on the shift from purely performance targets previously adopted by the 5G-PPP to a more societal vision, including sustainability, security, and inclusivity. The ongoing efforts aligned with these values are explored and their integration in BeGREEN project are justified. Then, four mechanisms to enhance energy efficiency in RANs are proposed, culminating in an evolved architecture that enables their integration.

Firstly, advanced forms of MIMO technologies and related energy efficiency factors are studied. Traditional MIMO boosts throughput by transmitting multiple data streams simultaneously but struggles at cell edges due to weaker signals and interference. Distributed MIMO (D-MIMO) improves performance by combining MIMO with distributed antenna systems, ensuring consistent data rates. Cell-Free massive MIMO (CF-mMIMO), an advanced form of D-MIMO, dynamically selects access points (APs) based on signal quality, eliminating cell boundaries and improving energy efficiency by reducing the number of transmitting antennas. An empirical study to evaluate the performance of CF-mMIMO compared to traditional MIMO systems is performed. The second mechanism is Integrated Sensing and Communication (ISAC), whereby we propose extracting useful information from the sensing data that can lead to the optimization of resource allocation. The third mechanism is relaying transmitted data in RANs using relay nodes. Strategic placement of fixed or mobile relays can significantly enhance energy efficiency by optimizing network energy consumption. Relays improve signal conditions for user equipment (UE), allowing base stations (BS) to transmit with lower power. However, the power consumption of relays must also be considered. BeGREEN D2.2 explores relay deployment in two scenarios: one with single-antenna entities and another with massive MIMO at the BS. The fourth mechanism is computing allocation in virtualised RANs (vRANs) and its impact on energy consumption. The deployment of multiple virtual base stations (vBS) on shared computing infrastructure is analysed, and the "noisy neighbour" problem where energy overhead increases exponentially is characterised. The study explores CPU pinning, Last Level Cache (LLC) allocation, and various parameters such as traffic load, BS bandwidth, and modulation and coding schemes. The findings highlight trade-offs that present opportunities for developing optimisation algorithms to significantly improve RAN energy efficiency.

Finally, leveraging the mechanisms proposed in WP3 and WP4, and based on the O-RAN architecture and the initial BeGREEN architecture from BeGREEN D2.1, the evolved BeGREEN architecture is introduced and discussed. This involves architectural extensions for integrating AI/ML support, D-MIMO, ISAC, Reconfigurable Intelligent Surfaces (RISs), relays, and the Edge domain. BeGREEN D2.2 serves as input to BeGREEN D2.3 as well as to BeGREEN WP3, WP4, and WP5.

1 Introduction

This document, BeGREEN D2.2, builds upon the foundation laid out in BeGREEN D2.1 [1], which provided a comprehensive overview of the State-of-the-Art (SotA) for Radio Access Networks (RANs) from an energy efficiency perspective. BeGREEN D2.1 also proposed potential solutions for RAN optimization at hardware, link, and system levels, while establishing the overall project framework. This framework encompassed the architecture, Key Performance Indicators (KPIs), and specific technical targets.

Expanding upon this groundwork, BeGREEN D2.2 provides an initial analysis of the project's KPIs, key value indicators (KVIs), and societal key values (KVs). Subsequently, this deliverable proposes a suite of mechanisms designed to bolster energy efficiency within the RAN. Finally, we propose an evolved architecture specifically designed to integrate these mechanisms seamlessly.

By undertaking these steps, WP2 aims to pave the way for the successful implementation of the technical innovations envisioned throughout the BeGREEN project's various work packages (WPs).

While the 5G-PPP framework has focused on achieving specific performance targets (KPIs), more recently, the EU's 6G vision project, Hexa-X, identified a crucial shift: moving beyond pure performance to a value-oriented approach. This value encompasses societal and human well-being, including sustainability, security, and inclusivity. This deliverable explores ongoing efforts aligned with these values and justifies their integration within the BeGREEN project. Four mechanisms to enhance the energy efficiency in RANs are proposed as explained below. Finally, an evolved architecture is developed that brings together the four mechanisms.

Firstly, a study is conducted on Multiple-Input Multiple-Output (MIMO) technology. MIMO allows the transmission of multiple data streams simultaneously, boosting throughput. Additionally, MIMO utilizes beamforming to focus signals on specific devices, minimizing interference. While traditional MIMO offers advantages, it struggles at cell edges due to weaker signals and interference. Distributed MIMO (D-MIMO) tackles these issues by combining MIMO with distributed antenna systems. D-MIMO offers more consistent data rates throughout the coverage area. CF-mMIMO, an evolved version of D-MIMO, takes it further by dynamically selecting access points (APs) to serve users based on signal quality, eliminating cell boundaries. CF-mMIMO also boasts about improved energy efficiency compared to traditional MIMO. This is because fewer antennas are transmitting at any given time, as user requests are handled by access points closer to the user. In this document, an empirical study is conducted which demonstrates that a D-MIMO system provides superior performance using less transmit gain compared with a traditional MIMO system.

Secondly, the processing of ISAC information at the RAN is analysed, particularly at the Radio Units (RUs), thus achieving optimal coverage in the area of interest, making use of the available RUs at their minimal transmit power. This is carried out by processing the resulting heatmap, therefore optimizing the transmissions of the RU to reduce energy consumption while serving the UEs in the area.

Thirdly, the impact of the deployment of relay nodes is studied. Research suggests that equipping a RAN with relay nodes can significantly improve its energy efficiency. By strategically placing fixed or even mobile relays, the network can optimize energy consumption across its various components. The benefits come from two main factors. UEs connected through relays experience better signal conditions. This allows the BS to transmit with lower power compared to directly connected UEs. However, it is important to consider that the relays themselves also consume power when operational. In this deliverable we explore this concept in two scenarios: one with all entities using a single antenna, and another with a more complex setup where the BS utilizes a massive MIMO antenna system.

Fourthly, a comprehensive analysis of the computing allocation in vRANs and their impact on energy consumption is presented. The analysis scenario consists of multiple vBS deployed on a shared computing infrastructure. In this context, the phenomenon known as the noisy neighbour problem by which the energy

overhead increases exponentially. The effects of CPU pinning and Last Level Cache (LLC) allocation on computational resource utilization are analysed as well. The influence of various parameters such as traffic load, BS bandwidth, and the employed modulation and coding scheme are also studied. BeGREEN D2.2 findings reveal a number of intriguing trade-offs that present opportunities for the development of optimisation algorithms that could significantly improve the energy efficiency of RANs. Finally, considering the mechanisms to enhance the energy efficiency of the network being proposed in WP3 and WP4, and using as baseline the O-RAN architecture and the initial BeGREEN architecture that was presented in BeGREEN D2.1 [1], the evolved BeGREEN architecture is introduced and the required architectural extensions related to the integration of AI/ML support, D-MIMO, ISAC, Reconfigurable Intelligent Surfaces (RISs), relays, and the Edge domain are discussed**.**

BeGREEN D2.2 is structured in five chapters. Other than current chapter that is the introduction:

Chapter 2 presents an analysis of the project's key performance indicators (KPIs), key value indicators (KVIs), and societal key values (KVs).

Chapter 3 proposes a set of mechanisms to optimise the energy consumption in the network.

Chapter 4 proposes an evolved architecture specifically designed to seamlessly integrate these aforementioned mechanisms.

Finally, chapter 5 puts forward BeGREEN D2.2 summary and conclusions.

2 Key Metrics in BeGREEN

Over the last 10 years, 5G networks in Europe have been developed under the umbrella of the 5G Infrastructure Public Private Partnership (5G-PPP). Indicative new network characteristics, the so-called Key Performance Indicators (KPIs), were the 5G-PPP quantitative metrics to be achieved at an operational level [2]. The involvement of vertical customers in the ongoing deployment of 5G networks demanded methods and strategies for performing tests and measurements. This need has been dealt with recently by the Test, Measurement and KPIs Validation (TMV) 5G-PPP Working Group (WG) [2][3][4].

Concurrently, the EU flagship project for 6G vision, Hexa-X, recognized the necessity to expand the fundamental network design paradigm from mainly performance-oriented to both performance- and value-oriented [5]. Here value entails intangible yet important human and societal needs such as sustainability, trustworthiness, and inclusion [6].

This chapter goes over the current work that is linked to these values and motivates their use in the framework of the BeGREEN project.

2.1 Terminology

The following terminology are important to understand the remainder of the chapter [7]:

- **Key Performance Indicator (KPI)**: A quantifiable measurement, agreed to beforehand, that reflects the critical success factors of a proposed solution. There exist network KPIs with defined target values for 5G systems. These are well-known KPIs that will be evolved to 6G by changing target values.
- **Key Value (KV)**: societal KVs (Table 1 in [6]) mean values important to people and society that may be directly addressed or indirectly impacted by future network technology, like environmental sustainability, digital inclusion, etc. A specific KV could be estimated through multiple KVIs.
- **KV as criteria and goals**: define the categories of KVs that are affected in a more high-level view, driving how a given society defines progress and public goods. Once KVs as criteria are defined, they need to be further analyzed to KVs as outcome, prior to being quantified or qualified into KVIs.
- **KV as outcome**: define the specific impact on the KV when considering the use of an artefact or system, and the KVIs aim to define quantifiable or qualitative metrics that assess progress towards such impact.
- **Key Value Indicator (KVI)**: A KVI should be a measurable quantity or requirement that in some form provides an estimate of an affected KV. A specific KV could be estimated through multiple KVIs. KVIs differ from KPIs in that they provide deeper insight into human-related factors and can require conversations and creativity to emerge.
- **KV Technical Enabler**: an aspect of future network technology that enables a usage related to a KV.

The relation between these terms is depicted in Figure 2-1.

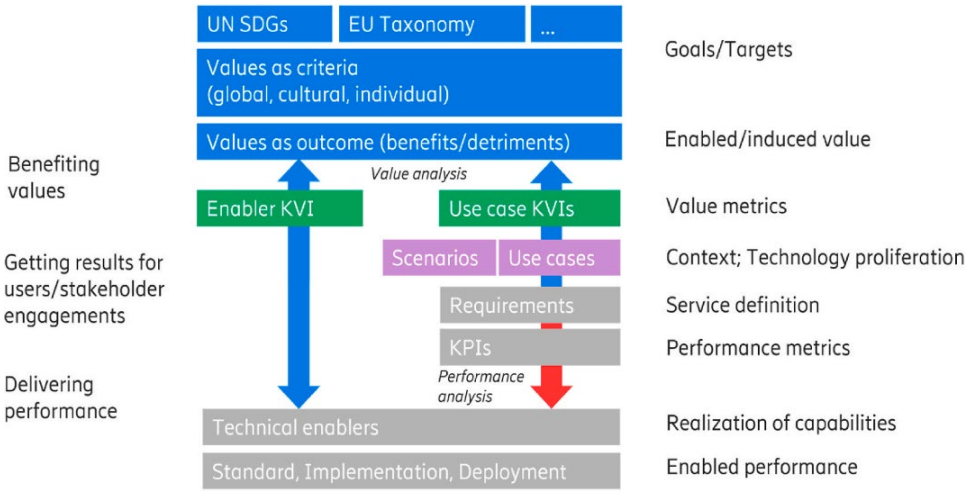


Figure 2-1 Overview of the employed terminology [7]

Table 2-1 BeGREEN Use Cases Defined in BeGREEN D2.1 [1]

Physical layer reference use cases
PHY#1 - B5G energy efficiency enhanced RAN through relay nodes
PHY#2 - Densification of the radio access architecture
PHY#3 - General mMIMO incentives
PHY#4 - Reconfigurable intelligent surfaces (RIS)
System-level reference use cases
SYS#1 - Energy efficiency in vRAN deployments with shared computing infrastructure
SYS#2 - Joint Orchestration of vRAN and Mobile Edge AI Services
SYS#3 - Traffic-aware management of NFV user-plane functions
SYS#4 - RIC driven energy-efficient RU on/off control

2.2 From use cases to key value indicators (KVIs)

The road to 6G demands analyses in multiple dimensions with the aim at KVs and redefining as well as advancing on an existing set of capabilities. The capabilities in turn can be mapped to measurable KPIs, whereas KVs are mapped to KVIs. The KVIs are related to the impact of the use cases, hence they shall be defined on a per UC basis, and they involve also foresight view of the proliferation of use cases.

2.2.1 BeGREEN use cases revisited

Table 2-1 enumerates the use cases BeGREEN D2.1 [1]. Given that the project seeks to promote energy efficiency in B5G/6G RAN components, these use cases refer to those strategies and additional capabilities provided to the RAN that could lead to an energy efficient design.

A broader view of the use cases listed in Table 2-1 can be provided by translating the requirements of these to prove whether they can satisfy the demand of current and envisioned 6G-like use cases [8] [9] [10]. The physical- and system-level capabilities provided by the use cases listed in Table 2-1 will be able to a set of the 6G requirements being considered by the following use cases (some related to those use case families included in the Hexa-X-II Flagship project [18]):

1. **Ubiquitous access**, mapped to PHY#1, PHY#2, PHY#3, PHY#4.
2. **Digital Twin**, capable of establishing a faithful replica of physical networks, providing a risk-free environment for preliminary research on innovative technologies, facilitates the verification of

extended 6G network architectures before deployment. This use case is mapped to PHY#4.

3. **Physical and environmental awareness**, enabling a better understanding of the surrounding environment, which can improve operational safety and efficiency, and contribute to goals such as sustainability or climate action. This use case is mapped to PHY#2, PHY#3, PHY#4, SYS#1, SYS#4.
4. **Towards Disaggregated and Virtualized Networks** [19]. This use case is mapped to SYS#1, SYS#2 and SYS#3.

2.2.2 Key Value Indicators (KVIs)

The United Nations (UN) SDGs and EU directives and frameworks provide general normative frameworks, to which the KVIs need to be linked to. BeGREEN is involved in the Smart Networks and Services Joint Undertaking (SNS JU) Test Measurement and Validation (TMV) Work Group and in the 6G-IA Sustainability Task Force (TF) to follow up the work on these topics at the framework level.

Given the demand of generating values-driven technology developments, the challenge is now how to integrate the societal challenges and values in the development of technology for 6G [7]. To extend the view of the conventional KPIs towards a more comprehensive value-driven approach, KVIs have been introduced to complement the KPIs and are able to better capture the spirit of the SDGs.

2.2.2.1 Framework followed by BeGREEN

A structured KVI framework tailored to the ICT research and development (R&D) sector is proposed in [7] which will be followed by BeGREEN. This framework builds upon the familiar engineering approach of using KPIs to find a process that systematically integrates considerations of economic, social and ecological impact into technological development projects. Figure 2-2 depicts the process of defining, eliciting, analysing, and assessing KVIs.

- Step I) Definition of **scenario** and **use cases**: In this step the global scenario that is addressed by the technological solution is outlined and instantiated to specific use cases.
- Step II) Elicitation: Identification of **KVs as criteria** and **KVs as outcome** to be associated with the use case.
- Step III-a) **Analysis of outcome on value domains**, leading to **formulation of KVIs as metrics** associated with the KVs for the use case. In this step, KVIs are formulated as absolute or relative measures of the impact of technology solution on the identified KVs. Use case KVIs will need to capture KVs that are both benefited as well as those that are determined. This gives a possibility for formulating targets based on societal impact and get a first-glance view of the expected outcome.

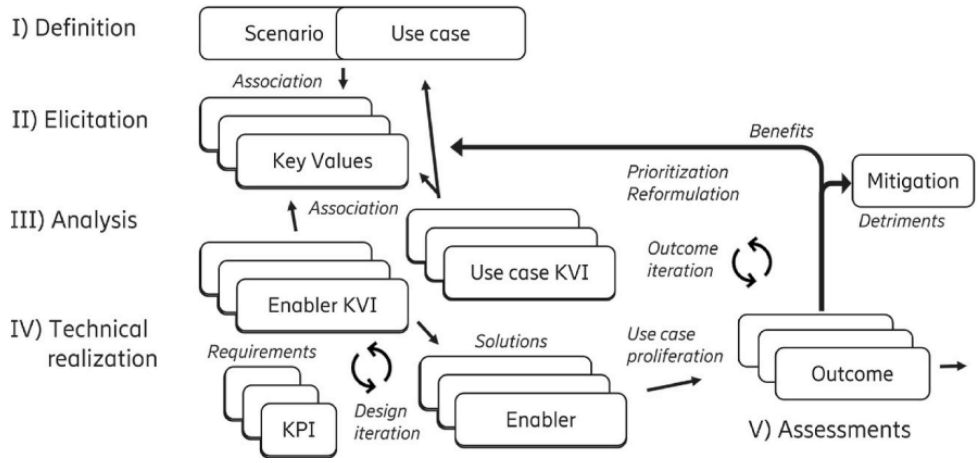


Figure 2-2 Work process of the KVI framework [7]

- Step III-b) **Analysis of value proposition** and estimation of use-case proliferation, leading to an established view on the KVI outcome.
- Step III-c) **Analysis involving prioritization and balance between KVIs and KPIs**, leading to target formulations based in policy.
- Step IV) **Technical realization**, involving identification and specification of enablers (i.e., ICT solutions), identification of KVs and formulation of KVIs associated to those, and a design iteration with respect to the KPIs and KVIs. In a more mature phase of development, the technology and system effects of providing a use case is better understood in terms of enablers.
- Step V) **Assessments**, where KVIs can subsequently be gauged through a progression of evaluations, notably (Va) expert assessments, (Vb) simulations, and (Vc) twinned systems. Assessment should be done with defined methodologies and can lead to a reformulation of KVIs and associated targets based on an increased understanding. Mitigating actions, related to the technical enablers and service solution of the use case, can also be launched as an outcome of these assessments to avoid negative impacts and regain positive impacts.

2.2.2.2 Initial work on the definition of KVIs

The mapping of the 6G use cases and the expected KVs and KVIs to be considered in BeGREEN is captured in Table 2-2.

Table 2-2 KVIs for the UCs Addressed by BeGREEN

6G UC	BeGREEN UC (as of D2.1 [1])	KV as Criterion and Goal	KV as Outcome	UC KVIs
#1	PHY#1, PHY#2, PHY#3, PHY#4	Environmental Sustainability Digital Inclusion	**KV1**: Services are accessible and available to everyone **KV2**: Network capabilities are environmentally responsible, energy efficient **Stakeholder**: Experimenters, Researchers, MNOs, Verticals	Service Availability
#2	PHY#4	Societal Sustainability, Environmental Sustainability	**KV1**: Improvements to customer experience **KV2**: Ability to test changes before they are implemented **Stakeholder**: Customers, Experimenters, Researchers	Network optimization
#3	PHY#2, PHY#3, PHY#4, SYS#4	Economical Sustainability and innovation, Environmental Sustainability	**KV1**: Energy efficiency optimization **KV2**: Network capabilities are environmentally responsible, energy efficient **Stakeholder**: MNOs, Verticals	Energy efficiency
#4	SYS#1, SYS#2, SYS#3	Economical Sustainability and innovation, Digital Inclusion, Environmental Sustainability	**KV1**: Cost-efficient networks **KV2**: Energy efficiency optimization **Stakeholder**: MNOs	Cost, Resource utilization

2.2.3 Key Performance Indicators (KPIs)

BeGREEN has proposed in the DoW several KPIs that are relevant to the quantitative outcomes that will evaluate the degree of achievement of the project results. Table 2-3 lists the mapping of the KPIs expected in the project with the following fields:

- **KPI Source (DoW)**: KPI as defined in the DoW (with nomenclature mapped to the Objectives).

- **KPI-ID**: revised nomenclature (general) to address the KPIs throughout the duration of the project.
- **6G Use Cases**: relevance of the KPI to the identified 6G use cases.
- **WP**: Work Package for which the KPI is relevant.
- **Verification Method**: level of assessment of the KPI – from simulation to evaluation in demonstrations in laboratory environment – Technology Readiness Level 4 (TRL4) or relevant environment (TRL5).
- **Tech Spec and reference values for 6G use cases**: specification of the quantitative metrics associated to the KPI. Expected values vs. those values that are expected in the above-mentioned 6G use cases (when applicable).
- **Description**: description of the KPI evaluation and details on how to evaluate it. Reference to past 5G-PPP documents or SNS agreed KPI evaluation methods may be included.

These KPIs will be evaluated by means of different verification methods pertaining to the work carried out in BeGREEN WP2, WP3, WP4 and WP5, and reported in previous deliverables.

Table 2-3 BeGREEN KPIs

KPI NO	KPI Source (DoW)	KPI-ID	6G Use Case	WP	Verification Method	Tech Spec and Reference Values for 6G Use Cases	Description
#1	DoW-KPI-O1.1	BGR-KPI-01	No specific UC, relevant for all	2	Simulation	Energy consumption, N/A	Energy consumption model for 5G/B5G base-stations and proposing energy efficiency enhancements
#2	DoW-KPI-O1.2	BGR-KPI-02	No specific UC, relevant for all	2	Simulation	Energy consumption, N/A	System level simulator (AIMM) to enable area-wide assessment of energy consumption over time.
#3	DoW-KPI-O1.3	BGR-KPI-03	No specific UC, relevant for all	2	Simulation	Energy efficiency optimization across BeGREEN components, N/A	Balance between different network evolution strategies to optimize energy efficiency in different target service areas
#4	DoW-KPI-O2.1	BGR-KPI-05	No specific UC, relevant for all	3	In-Lab PoC4	Power reduction of at least 15%, N/A	Optimization of the implementation of accelerated LDPC and sphere decoder
#5	DoW-KPI-O2.2	BGR-KPI-06	No specific UC, relevant for all	3	Simulation vs. In-Lab PoC4	Implementation loss of up to 2 dB, BLER and throughput comparisons	Accelerated processing, performance compared to Matlab system level simulations
#6	DoW-KPI-O2.3	BGR-KPI-07	No specific UC, relevant for all	3	In-Lab PoC4	Utilization and processing time	Comparison CPU (x86, arm) vs GPU for the LDPC and sphere decoder
#7	DoW-KPI-O3.1	BGR-KPI-08	No specific UC, relevant for all	3,4	PoC2, Adastral PoC	Functionality, latency	Cell on/off scheme by xApp based in geolocation information
#8	DoW-KPI-O3.2	BGR-KPI-09	No specific UC, relevant for all	3	Simulation: Envelope tracking and DPD, Lab Test & Final Demo (BT Testbed): RF blanking - real	RU energy saving better than 40%	Demonstrated energy reduction of RU of >=40% for emulated day/night (busy/idle) scenario
#9	DoW-KPI-O3.3	BGR-KPI-10	No specific UC, relevant for all	3	PoC5, Adastral PoC	RU energy saving of 40%	Demonstrated energy reduction of RU of >=40% by use of dynamic AGC, and combined AI-based DPD and Envelope Tracking
#10	DoW-KPI-O4.1	BGR-KPI-12	#2, #3	3	Simulation and In lab test (PoC2)	Angle and range precision for localisation	Precision of the developed sensing algorithm for detecting potential users
#11	DoW-KPI-O4.2	BGR-KPI-13	#2, #3	3	Simulation	Use identification with 20% less wireless medium (channel) usage	Sensing assisted beam search – 20% performance improvement with respect to extensive search and hierarchical search.
#12	DoW-KPI-O4.3	BGR-KPI-14	#2, #3	3	Simulation	50% accuracy improvement in user estimation	Detection of users in order to estimate the presumed network load – at least 50% accuracy of estimation of potential mobile users
#13	DoW-KPI-O5.1	BGR-KPI-15	#4	4	PoC1	Working version of the Intelligence plane	BeGREEN Intelligence Plane

#14	DoW-KPI-O5.2	BGR-KPI-16	#4	4	PoC1	Energy rating	Overall energy rating is provided at use case and deployment level.
#15	DoW-KPI-O5.3	BGR-KPI-17	#4	4	PoC1	Energy rating	Orchestration strategy to achieve improved energy rating
#16	DoW-KPI-O6.1	BGR-KPI-18	#4	4	In-Lab PoC		>20% bare metal server energy consumption reduction at low load with respect to bare metal server energy consumption at peak load, without noticeable impact on user plane traffic performance.
#17	DoW-KPI-O6.2	BGR-KPI-19	#4	4	In-Lab PoC, PoC1		At least 2 AI/ML adaptation strategies for energy efficiency demonstrated in laboratory environment with TRL4.
#18	DoW-KPI-O6.3	BGR-KPI-20	#4	4	In-Lab PoC, PoC1		An energy rating that effectively rates the VNFs (e.g., in the scale of A to E where A is highly efficient, and E is least efficient). A scoring method (value between 0 to 100) measure the energy efficiency of VNFs with its influencing factors including traffic.
#19	DoW-KPI-O7.1	BGR-KPI-21	#4		Simulations and in-lab tests (NEC)		>20% power consumption reduction on the server that runs the edge AI service AI service power consumption
#20	DoW-KPI-O7.2	BGR-KPI-22	No specific UC, relevant for all		Analytical analysis based on lab tests with x86 and ARM servers with the same test scenario. The metric used will be Joules/bit to include the performance improvement of the system in this test. Both Energy Consumption and Energy Performance (Joules/bit) must fulfil the 20% reduction.		>20% power consumption reduction on running CU on ARM and HW accelerating PDCP of CU-UP
#21	DoW-KPI-O7.3	BGR-KPI-23	#4	3	Analytical analysis based on lab test with x86 and ARM servers with same test scenario.		>20% power consumption reduction on running Near-RT RIC on ARM and HW accelerating xApp
#22	Not in DoW	BGR-KPI-24	No specific UC, relevant for all	4			O-RAN compliant sub-millisecond control interface for RIS
#23	Not in DoW	BGR-KPI-25	No specific UC, relevant for all	4			An extension of the O-RAN RT-RIC that can configure RIS surfaces to optimize energy consumption for edge tasks
#24	Not in DoW	BGR-KPI-26	#1	4	Simulation	Energy consumption	Energy consumption reduction, spectral efficiency improvement and coverage hole mitigation through a deployment of fixed relays.
#25	Not in DoW	BGR-KPI-27	#1	4	Simulation	Energy consumption	Energy consumption reduction through smart relay activation/deactivation algorithms based on AI/ML.

3 Mechanisms to Enhance Power Efficiency

In BeGREEN D2.1 [1] the RAN was identified as the most power-consuming element with up to 73% of the total power consumption. Therefore, reducing the energy consumption of the RAN will significantly reduce the total consumption. To this end, four mechanisms are identified, i.e., (1) D-MIMO deployments, (2) Sensing-aided resource allocation of radio resources, (3) B5G RAN enhanced through relay nodes, and (4) Energy-aware computing allocation in vRAN. We explain each of them in detail below.

3.1 Distributed MIMO deployments

Traditional MIMO has become one of the main features of 4G (LTE) and 5G to fulfill the need for improved data rates, leveraging multipath propagation. MIMO can send multiple data streams at the same time; thus, the throughput is (under ideal conditions) proportional to the number of streams compared with a single-stream system. MIMO uses large antenna arrays to separate the streams in space. The first release of the MIMO technology was in LTE where the BSs had up to 4 or 8 antennas. This architecture allows the BS to use array-processing techniques like beamforming to create narrow beams toward the UE, therefore the signal quality is maximized while minimizing the interference. 5G currently supports massive MIMO with 64 or 256 antennas at the BS. Despite the large number of antennas, traditional MIMO or collocated MIMO (MIMO), shows poor performance at the edge of the cell because of two main factors: (i) the signal quality is limited at the edge of the cell, and (ii) the interference from adjacent cells. To deal with that, Distributed MIMO (D-MIMO) combines the advantages of MIMO and distributed antenna technologies to provide consistent throughput over the cell. CFmMIMO [26] inherently uses D-MIMO to erase cell boundaries where some distributed antennas are selected in terms of signal quality to maximize the UE performance (user-centric system). In the context of cell-free or D-MIMO systems, the distributed antennas are denoted as Access Points (APs).

D-MIMO significantly improves energy efficiency compared with MIMO due to two main reasons. First, the throughput is distributed almost uniformly as it is quite likely that a specific user is close to an AP or more APs. Hence, a small group of selected APs serve a user which considerably reduces the number of transmitted antennas compared with MIMO where all the antennas radiate. Second, the BS with a MIMO configuration presents low azimuth spread since the height of the BS is usually high, therefore it has limited radio wave scattering. The azimuth spread has much influence on the correlation among individual MIMO channel responses [12], it has been studied that the greater the azimuth spread, the lower the correlation and the bigger the MIMO channel capacity. Under this assumption, we can then anticipate that D-MIMO can provide the same level of quality link – in terms of Bit Error Rate (BER) – as MIMO even using low transmit power. To the best of our knowledge, some researchers have tackled energy efficiency using cell-free architecture as compared with MIMO. In [13] the authors investigated the use of cell-free massive MIMO with conjugate beamforming and max-min power control to double the energy efficiency. In [14] a closed-form expression for the spectral efficiency was derived considering the effects of backhaul power consumption, the number of APs, and the number of antennas per AP on the total energy efficiency. With this expression, the authors analysed their proposed AP selection scheme and power allocation algorithm in terms of energy efficiency. Finally, the work in [15] aims to solve a combinatorial problem of selected APs and associated UEs to minimize energy consumption while providing a constant data rate in DL. Although there are a number of works that tackle D-MIMO (cell-free) from the energy reduction perspective, their evaluations do not consider the realistic scenarios that an operator needs to deploy.

In this section, we present an empirical study that quantifies how much energy can be saved using a D-MIMO system compared with a MIMO system. To this end, we implement a D-MIMO system using the NVIDIA code Sionna [17]. Sionna is an open-source Python library for link-level simulations of digital communication systems built on top of the open-source software library TensorFlow. We extensively evaluate our D-MIMO

implementation in a *realistic* and simulated scenario in Madrid that mimics an operator-like deployment using a ray-tracing tool. We deploy 32 distributed antennas to form a D-MIMO system and compare its performance with a 32-antenna MIMO system. We measure the performance in terms of BER for different transmit gains, different modulations, and different numbers of MIMO streams. Our findings show that D-MIMO offers superior performance compared with MIMO while utilizing less transmit gain. We also observe that MIMO cannot deal with 4 and 8 streams while D-MIMO provides functional BER.

This evaluation only assesses a D-MIMO system with MIMO. The cell-free implementation based on the O-RAN architecture is explained in Chapter 4. It is planned to delve deeper into AP selection for cell-free purposes in the upcoming BeGREEN D2.3.

3.1.1 Background

Let the number of antennas at the transmitter and at the receiver be in a general MIMO system as N_t and N_r, respectively. The channel state information (CSI) for the i -th and j -th transmitting and receiving antennas is denoted as $h_{i,j}$. The general OFDM distributed wireless MIMO transmission in frequency domain is described as

$$y[k] = \boldsymbol{H}[k]x[k]g + z[k], \qquad k = 0,1,\dots,K-1, \tag{3.1}$$

where $x[k]$ is the N_t-dimensional transmitting signal at the subcarrier k, $\boldsymbol{H}[k]$ is the $N_t \times N_r$ CSI matrix at the subcarrier *k*, $y[k]$ is the N_r-dimensional receiving signal at the subcarrier k, $z[k]$ is an N_r-dimensional Additive White Gaussian Noise (AWGN), *g* is a transmit gain, and *K* is the number of subcarriers.

Usually, the CSI $h_{i,j}$ is expressed as the product of the shadowing and fast fading factors

$$h_{i,j} = h_i^s \cdot {h_{i,j}}^w,$$

where h_i^s and ${h_{i,j}}^w$ are the shadowing and fast fading factors, respectively.

Regarding the MIMO model, the propagation paths from the BS to the UE antennas are consistent; hence $h_i^s = h^s$. Therefore, D-MIMO and MIMO are unified in terms of mathematical model, the only difference is in the shadowing.

To recover the information that was sent by the transmitter, we need to estimate the channel. To do so, we apply Least Square (LS) estimation

$$\hat{\boldsymbol{H}}[k] = y[k](x[k])^{-1}. \tag{3.2}$$

Once the channel is estimated, we apply Linear Minimum Mean Square Error (LMMSE) equalization to recover the transmitting signal. To this end, we first estimate

$$\boldsymbol{G}[k] = \hat{\boldsymbol{H}}^H[k]\left(\hat{\boldsymbol{H}}[k]\hat{\boldsymbol{H}}^H[k] + N_0\boldsymbol{I}\right)^{-1}, \tag{3.3}$$

where N_0 is the noise power and $\boldsymbol{I}$ is the identity matrix. Finally, the estimated transmitting signal is

$$\hat{x}[k] = diag\left(\boldsymbol{G}[k]\hat{\boldsymbol{H}}[k]\right)^{-1}\boldsymbol{G}[k]y[k]. \tag{3.4}$$

Note that the term $diag\left(\boldsymbol{G}[k]\hat{\boldsymbol{H}}[k]\right)^{-1}$ is applied as a scaling factor to help the modulation demapper; see [40] for more information.

3.1.2 Experimental evaluation

This section explains the software for our simulation environment, Sionna, and why we chose it. We then described the scenario and its technical characteristics. And finally, we show the result of the comparison between D-MIMO and MIMO in terms of energy reduction and BER. This evaluation only assesses a D-MIMO

system with MIMO, we explain how a cell-free can be implemented in Chapter 4. In the upcoming BeGREEN D2.3, we will delve deeper into AP selection for cell-free purposes.

3.1.2.1 Platform

6G will change the paradigm of wireless communications by integrating AI/ML natively, hence there is a need to develop AI/ML-oriented wireless communication frameworks, and this is where Sionna [16] comes into play. Sionna is a TensorFlow-based open-source Python library designed for researchers to perform link-level simulations with native integration of neural networks, state-of-the-art ML algorithms, and a ray tracing tool.

Sionna provides an extensive selection of Python functions essential for 5G communication systems. In contrast to MATLAB and many other simulators, Sionna includes differentiable functions and tools that can efficiently perform algorithms based on gradient descent, with most variables represented as tensors. This feature makes Sionna highly attractive, allowing for faster workflows due to reduced processing times, as its functions are optimized for GPU execution. However, it can also operate on a CPU if a GPU is not available.

Although Sionna provides a wide range of functionalities, not all new technical enablers for 6G are considered in the first place. That is why we had to slightly tune tensor-based variables to trick the simulator, to implement some wireless environments such as D-MIMO and CFm-MIMO.

After analyzing Sionna´s source code, we observed that we need to modify the CSI matrix ($\boldsymbol{h_{freq}}$) given by the ray tracing. We deploy several independent APs and one UE, and the ray tracing computes all the paths from UE and APs and the other way around. Then, the code estimates the channel matrix in the frequency domain based on the output of the ray tracing. Since D-MIMO requires a join processing of all the APs, we need to modify the size of the channel matrix to consider a single entity rather than independent APs. Let consider the size of the channel matrix as

$$\boldsymbol{h_{freq}} \mapsto \boldsymbol{H_{freq}} \in \mathbb{C}^{batch \times \#_{rx} \times \#_{rx,ant} \times \#_{tx} \times \#_{tx,ant} \times \#_{OFDM} \times \#_{subcarriers}},$$

where:

- $batch$ denotes the size of the batch. In this case, this dimension is only introduced to avoid compatibility issues in further computations.
- $\#_{rx}$ represents the number of receivers deployed on the scene.
- $\#_{rx,ant}$ reflects the number of antennas incorporated in each of the receiver arrays.
- $\#_{tx}$ represents the number of transmitters deployed on the scene.
- $\#_{tx,ant}$ reflects the number of antennas incorporated in each of the transmitter arrays.
- $\#_{OFDM}$ denotes the number of OFDM symbols, in our case we set it to 14.
- $\#_{subcarriers}$ corresponds to the number of subcarriers, in our case; 72.

After determining the dimensions of $\boldsymbol{h_{freq}}$, we need to reshape it to implement a D-MIMO scenario. Specifically, we configured the tensor such that the number of transmitters (APs in a DL scenario) is set to one, with this single transmitter being equipped with $\#_{tx} \cdot \#_{tx,ant}$ antennas. Therefore, the resulting size of $\boldsymbol{H_{freq}}$ is

$$\boldsymbol{H_{freq}} \in \mathbb{C}^{batch \times \#_{rx} \times \#_{rx,ant} \times 1 \times (\#_{tx} \cdot \#_{tx,ant}) \times \#_{OFDM} \times \#_{subcarriers}}.$$

If UL simulation is desired, the transmitter and the receiver are swapped, and we need to apply the same procedure to the receiver side.

Technical configuration of the link. A compliant massive MIMO 5G system based on Orthogonal Frequency-Division Multiplexing (OFDM) in the frequency domain is used. Both UL and DL are simulated, where the block diagram can be observed in Figure 3-1. These blocks are introduced briefly as:

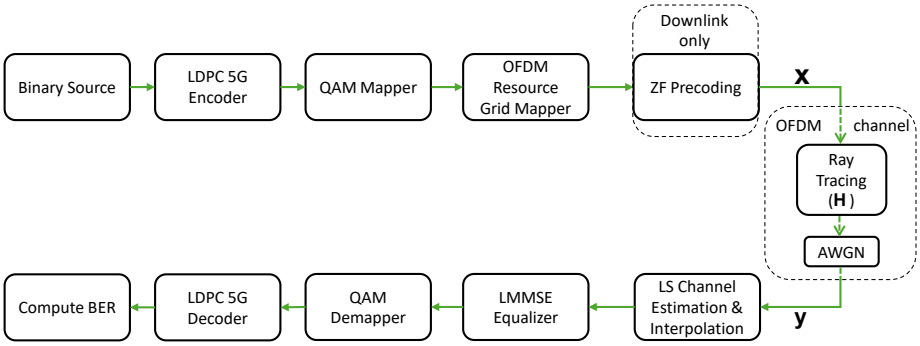


Figure 3-1 Communication's system block diagram

1. Binary source. It is the block that creates random bits to be sent.
2. Low-Density Parity Check (LDPC) 5G encoder/decoder. It is the linear correction method for 5G
3. QAM Mapper/Demapper. It transforms bits of information into symbols and vice versa.
4. OFDM Resource Grid Mapper. It creates a resource grid 5G compliant and maps the symbols into the resource grid.
5. ZF Precoding [40]. It applies precoding for DL-only transmission using the zero-forcing approach.
6. Ray tracing channel. It computes all the possible paths between the transmitter and receiver and creates an OFDM channel in the frequency domain.
7. Additive White Gaussian Noise (AWGN). It adds the AWGN noise to the signal.
8. LS Channel estimation and interpolation. It applies Least Square (LS) for channel estimation and interpolates the channel.
9. LMMSE Equalizer. It applies LMMSE to obtain a better estimate of the transmitted signal.
10. Compute Bit Error Rate (BER). It computes how many received bits are wrong because of the channel.

3.1.2.2 Scenarios

Before digging into the deployed scenarios, we briefly explain the key aspects of Sionna's ray tracing tool, highlight its importance in this contribution, and mention the most important parameters.

Ray tracing is a mature technology capable of accurately predicting the propagation of electromagnetic waves in many scenarios of interest [17]. The accuracy of ray tracing predictions is fundamentally reliant on two key elements: the scene geometry and the characteristics of the materials involved, e.g., their conductivity σ and relative permittivity ε_r [17]. The scenario geometry is derived from a selected geographical area sourced from OpenStreetMap and rendered in Blender, where we assign the corresponding materials, whose properties are stipulated by the ITU and coded in Sionna. With this in mind, we designed our scenarios to include three materials, assigned as indicated in Table 3-1. The scenario contains everything necessary for the simulation and representation of radio propagation and supports wave scattering and reflection, but not refractions.

Table 3-1 Used Sionna Materials and Some of its Radio Properties

Structure	Material	$\varepsilon_r\,[F/m]$	$\sigma[S/m]$
Ground plane	itu_concrete	5.24	$1.349 \cdot 10^6$
Building facade	itu_marble	7.074	$3.803 \cdot 10^6$
Rooftops	itu_metal	1	10^7

Moving on to the available scenarios provided by Sionna, there are two main scenarios: Arc de Triomphe and Munich. However, with the assistance of one of its tutorials and Blender, we recreated two scenarios in Madrid: one at the Puerta del Sol central plaza and another in the Salamanca district. The first scenario depicts a typical European-style central plaza, with adjacent pedestrian streets converging into a wide-open plaza. Meanwhile, the Salamanca district scenario features a geometric grid pattern, exemplifying a Manhattan-type setting.

In these two Madrid-based scenarios, we mimic an operator-like 5G massive MIMO deployment in the FR1 band. For each scenario, we deploy 16 distributed APs, and each AP has 2 antennas to form the D-MIMO system, while the MIMO system consists of 32 collocated antennas. In addition, we place the UE randomly in 100 different positions and evaluate the performance of both systems per location of the UE. The deployment procedure we followed consists of generating a coverage map, from which we retrieve the attenuation of a given scenario. Then, we randomly generate 100 UE positions, by establishing some attenuation and distance restrictions.

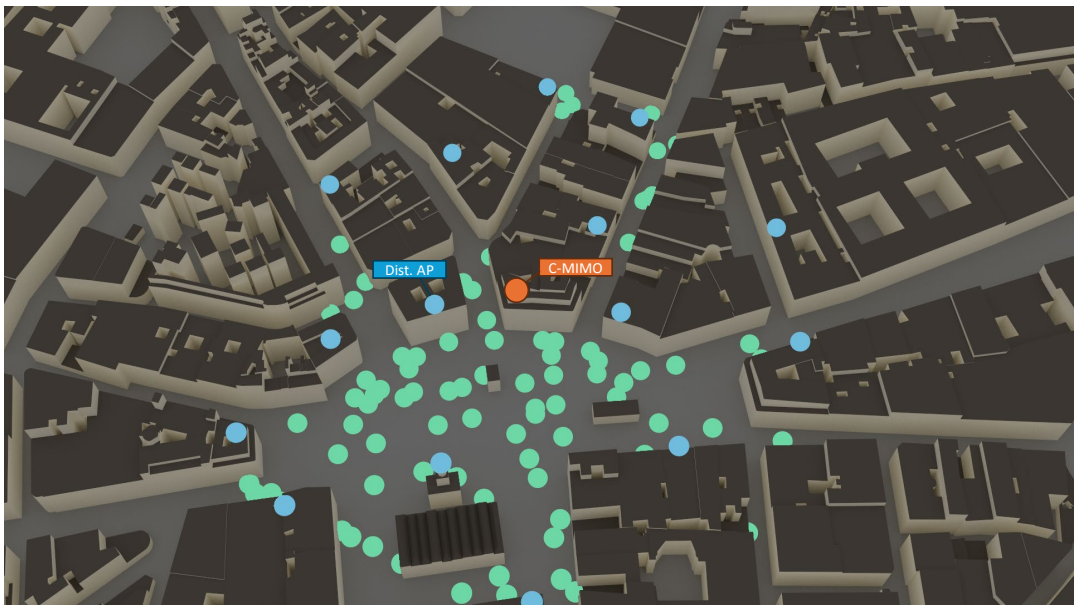


Figure 3-2 Ray tracing rendering of UEs and APs deployed at Puerta del Sol plaza

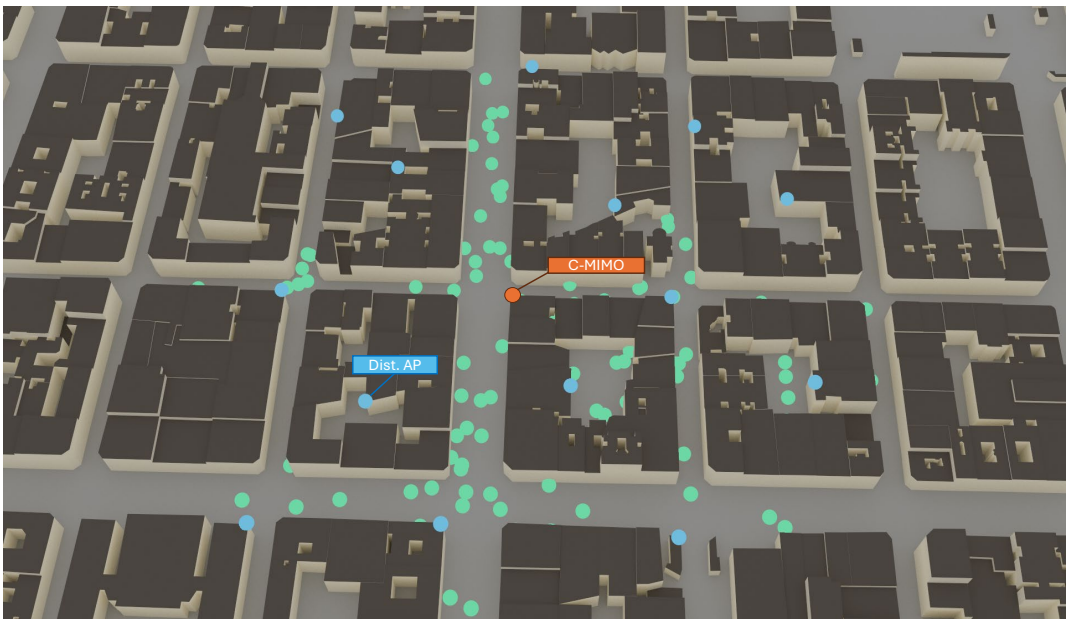


Figure 3-3 Ray tracing rendering of UEs and APs deployed at Salamanca district

The deployment of the distributed APs as well as the positions of the UE are discussed next. In particular, the blue dots represent the AP locations, the green dots indicate the UE locations, and the orange dot is the MIMO location.

3.1.2.3 Numerical results

As shown in Figure 3-1, we transmit bits in a 5G-compliant system and compute the BER. To do so, we simulate the wireless link at a central frequency of 3.5 GHz with a bandwidth of 10 MHz and a subcarrier spacing of 15 kHz. The antennas use cross-polarization. We vary the transmit gain, the number of antennas at the UE, and the number of bits for QAM. We evaluate the performance of both systems, MIMO and D-MIMO for 100 positions of the UE. We apply 4 transmit gains, g1, g2, g3, and g4. g1 is the lowest gain, g2 is 10 dBs higher than g1, g3 is 10 dBs higher than g2 and the same for g4. Regarding the number of antennas at the UE and modulation, we evaluate performance for 2, 4, and 8 antennas and 4-QAM, 16-QAM, and 64-QAM, respectively.

Madrid-Sol.

The results for Sol are shown in Figure 3-4 and in Figure 3-5, for both DL and UL transmissions, respectively. The rows represent the modulation, the first row indicates 4-QAM, the second row indicates 16-QAM, and the third row indicates 64-QAM. The columns represent the number of antennas at the UE, the first column represents 2 antennas, the second column represents 4 antennas, and the third column represents 8 antennas. 2 antennas at the UE means that the system transmits 2 streams, 4 antennas mean 4 streams and so on. Every subfigure includes 8 lines, the solid lines illustrate MIMO, and the dashed lines indicate D-MIMO. As explained above, every system is evaluated using four transmit gains (g1, g2, g3, and g4). Since we are evaluating the performance of both systems for 100 UE locations that are randomly located, we visualize the plots in a kind of Cumulative Distribution Function (CDF) way to show how the BER decays.

In general, we notice that the higher the modulation, the higher the BER for both systems. This behaviour is expected since higher modulation translates into smaller boundaries for discriminating symbols. In addition, increasing the number of streams also results in higher BER.

Regarding DL and the most robust case, 4-QAM and 2 antennas, MIMO for g1 has the worst performance, and from 30% of the cases, the BER is zero. If the transmit gain increases, the BER decreases. For instance, MIMO for g4 has zero BER except for a few cases. Regarding D-MIMO, it outperforms MIMO for all the cases. It is worth highlighting that D-MIMO shows even better performance for g1 than MIMO for g2, i.e., *D-MIMO performs better than MIMO using 10 dB of less transmit gain*. We also have the same behaviour for 16-QAM and 64-QAM. We believe that D-MIMO comes into play for a higher number of streams since *D-MIMO outperforms MIMO even with 40 dB less transmit gain for 4 and 8 streams.* We want to emphasize that in the worst case (64-QAM and 8 streams), around 90% of the cases have a high BER for MIMO, while D-MIMO provides much better performance. This points out that D-MIMO outperforms MIMO for a higher number of streams as the rank of the D-MIMO channel is usually higher. In a nutshell, the rank is a way to measure how many independent streams can be sent.

For UL, the general behaviour of both systems is like DL however, the UL BER tends to be a bit worse. This tendency is because of the precoding in DL. Precoding is used to remove interference from transmitting streams towards other UEs. To this end, it utilizes narrow beams to steer the signal into the desired UE.

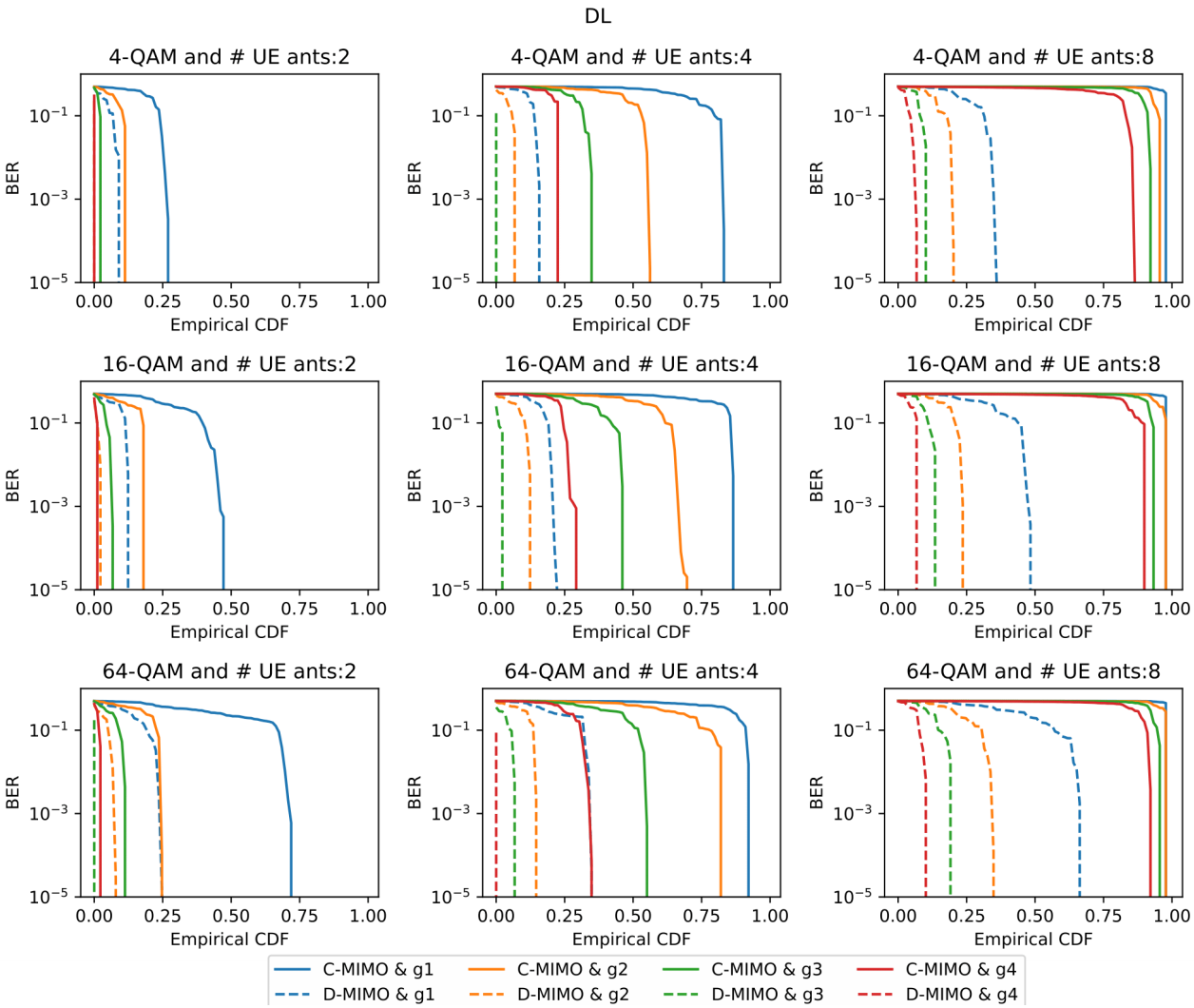


Figure 3-4 Madrid-Sol (DL results)

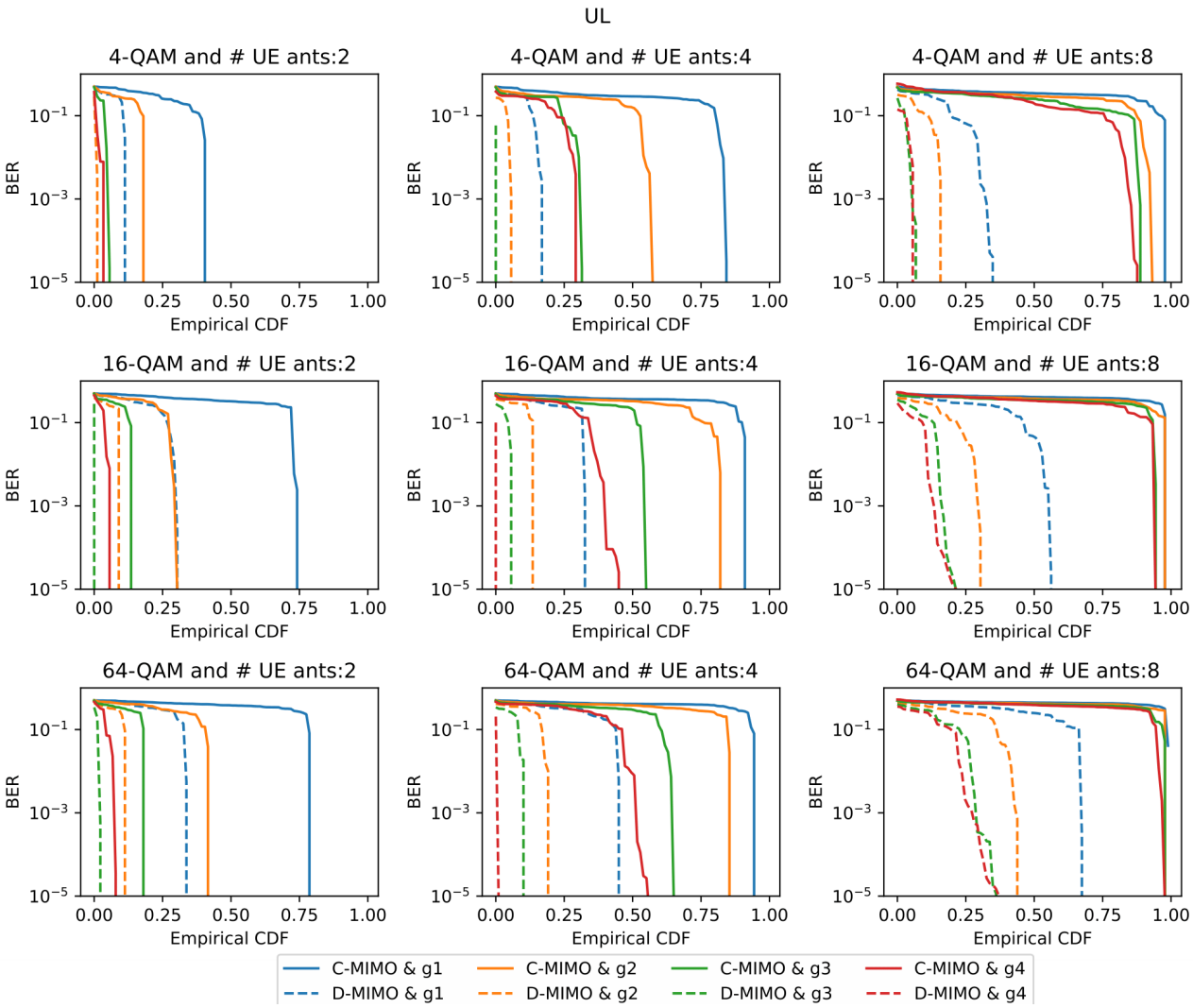


Figure 3-5 Madrid-Sol (UL results)

Madrid-Salamanca evaluation.

The results for this evaluation are shown in Figure 3-6 and Figure 3-7, for both DL and UL, respectively. Regarding DL and the most robust case, 4-QAM and 2 antennas, *D-MIMO with 10dB less transmit gain exhibits better performance than MIMO* which follows the same behaviour as Madrid-Sol for both systems. If the number of antennas at the UE increases, the BER increases as well. For 4 antennas and 4-QAM, D-MIMO with ten dB less of transmit gain overcomes MIMO while for 16-QAM and 64-QAM, D-MIMO and MIMO for g1 have similar performance. We would like to emphasize that D-MIMO is the only system that can provide functional performance with a very low BER for 4 antennas. Moving to 8 antennas and 4-QAM, MIMO for g1 cannot provide a BER of zero as its curve does not reach the bottom. This means that MIMO for g1 provides really poor performance. D-MIMO for g4 is the only one that provides reasonable BER. For 16-QAM and 64-QAM, D-MIMO outperforms MIMO, but its performance is really limited, and it cannot be used in practice.

For UL, both systems behave similarly to DL, but the BER is slightly worse. As explained before, the DL results are better because of the precoding.

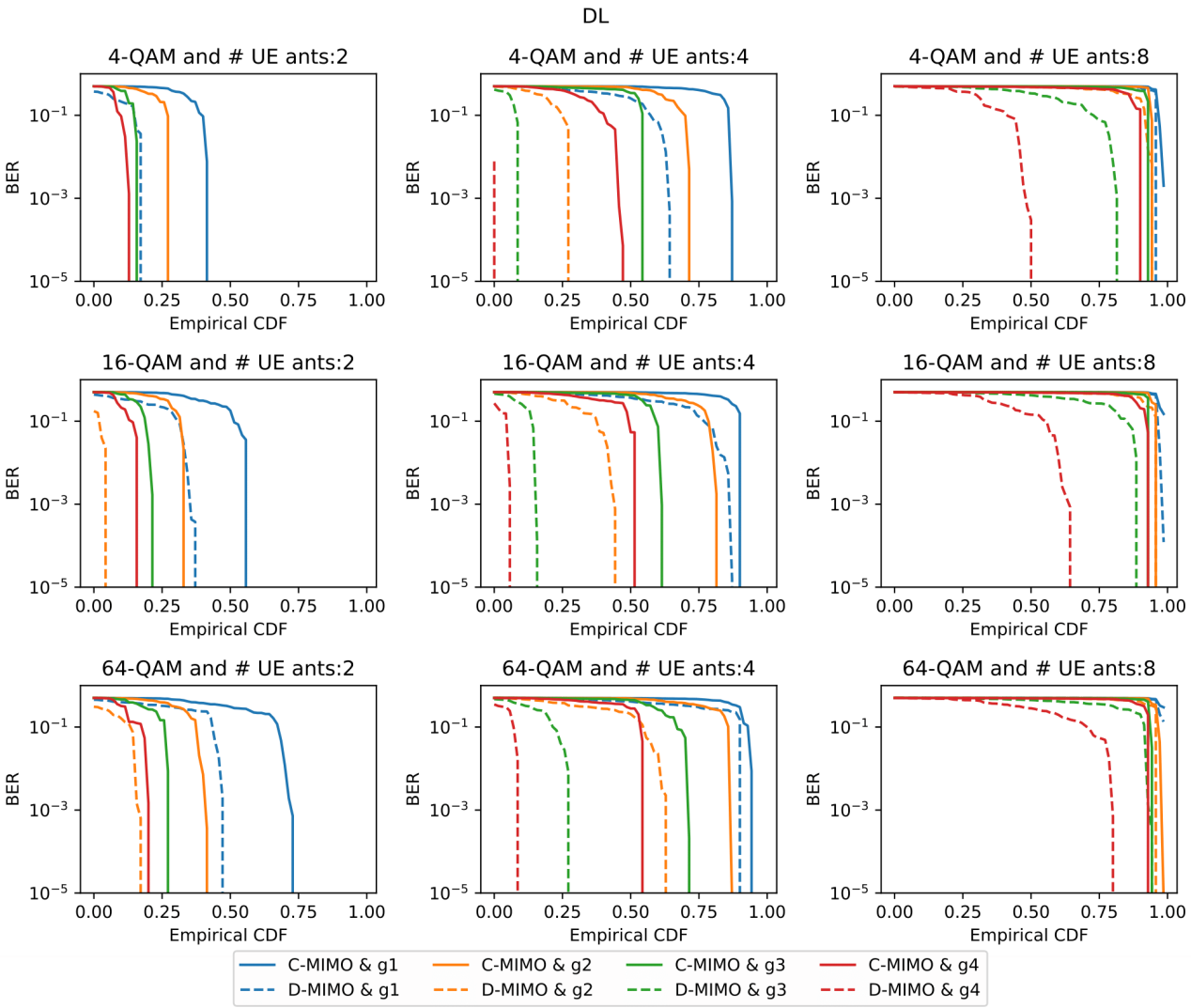


Figure 3-6 Madrid-Salamanca (DL results)

The degradation in the BER for Madrid-Salamanca could be due to the geometry of Madrid-Salamanca compared with Madrid-Sol. In Madrid-Salamanca, we have long and narrow streets which make the channel sparse in multipath components and the paths are very similar which increases correlation. In Madrid-Sol, the channel is richer in multipath components because of the big square which makes the scenario more diverse, and the paths are less correlated. Therefore, Madrid-Salamanca has a smaller number of independent paths resulting in worse BER.

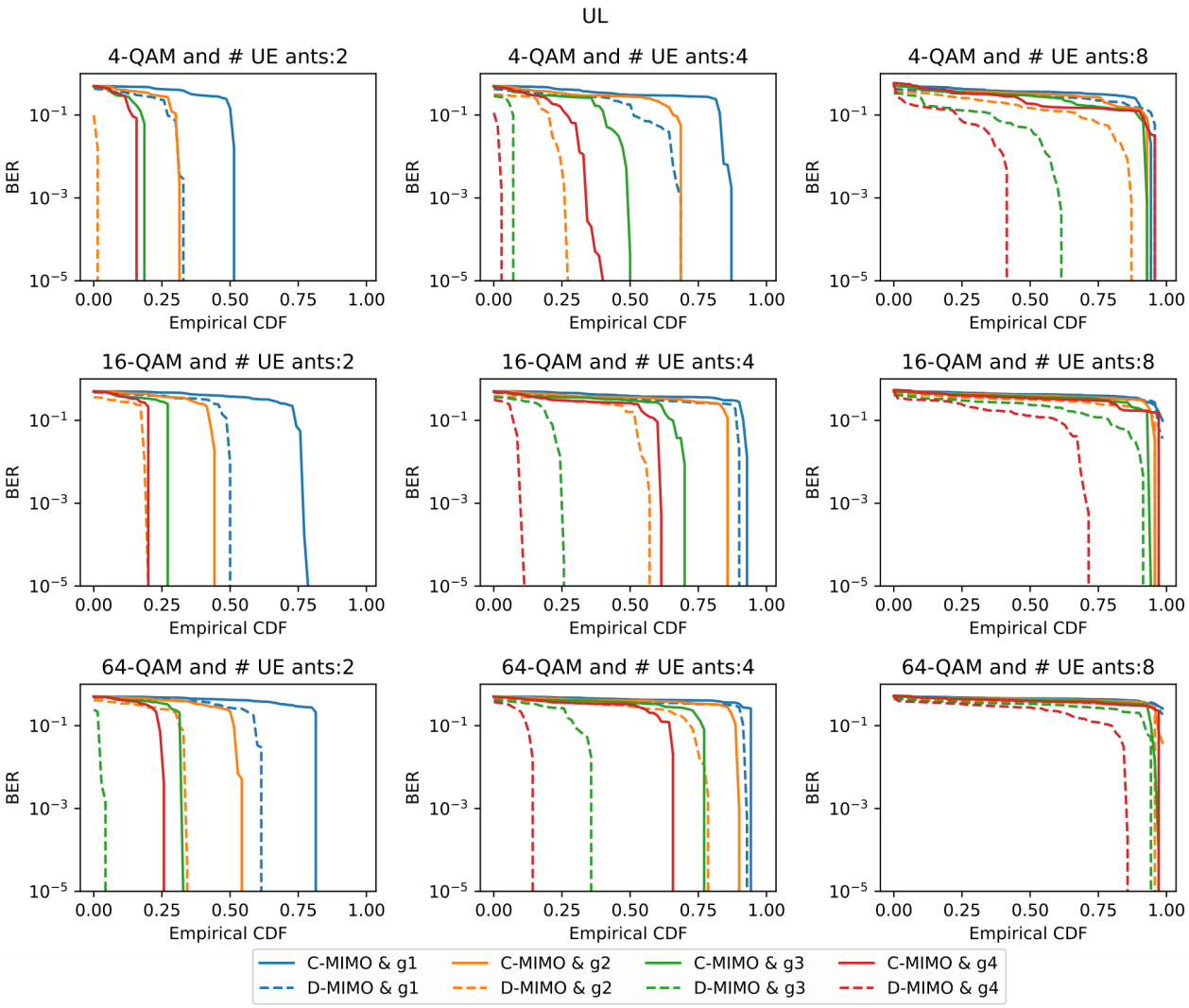


Figure 3-7 Madrid-Salamanca (UL results)

3.1.2.4 Conclusion and discussion

This section introduced the empirically evaluated D-MIMO system and its comparison with a MIMO system in terms of energy savings. The NVIDIA tool 'Sionna' was used for this purpose which allows to implement a MIMO 5G-OFDM-like system under realistic settings using its ray tracing module. Both systems were extensively evaluated in two areas of Madrid, in Sol Square and in the Salamanca neighborhood. The findings show that D-MIMO needs less transmit gain to achieve better performance compared to a traditional MIMO setup. The results also show that D-MIMO provides superior performance for a higher number of antennas at the UE, i.e., for a higher number of streams. This is because the D-MIMO channel has higher rank which provides more independent paths compared with MIMO. It is also observed that the geometry of the scenario plays a substantial role. Narrow and long streets make the channel sparser in multipath components which degrades the channel rank and performance, while big squares make it higher and provide better results.

BeGREEN D2.3 will explore strategies to select certain APs, in other words, determined APs for energy-saving purposes are witched off, while the same performance is preserved. This analysis was carried out for research purposes and so the cost of deploying more antennas is not analysed.

3.2 Sensing-aided resource allocation of radio resources

The RAN resources are managed to optimize different parameters of the network. Among them one can

consider system throughput, spectral efficiency, quality of service (QoS), quality of experience (QoE), fairness, energy efficiency, etc. In addition, interference management strategies are deployed to minimize the interference between the different radio devices in the network, which can significantly increase the overall network performance.

The networks mainly have the possibility to dynamically manage the radio resources to minimize energy consumption. Nevertheless, to achieve this, information such as the spatial density of the users is of paramount importance. This section presents related work which can be carried out when extracting ISAC information from RUs that feature both communication and sensing capabilities.

3.2.1 Description of the steps for processing sensing data

The mechanisms for extracting useful information from the sensing data and to post-process this data with the goal of optimizing resource allocation do follow the steps listed below:

1. Availability of several RUs with ISAC capabilities (waveform choice depending on the RU capabilities, waveform analyses are treated in WP3 deliverables, such as in BeGREEN D3.2).
2. Extract sensing data making use of the ISAC algorithms presented in WP3 deliverables.
3. From the available sensing data, we proceed to represent it in the form of a heatmap (see an example in Figure 3-8).
4. Post-process the heatmap and optimize the transmissions of the RU to reduce energy consumption while serving the UEs in the area.

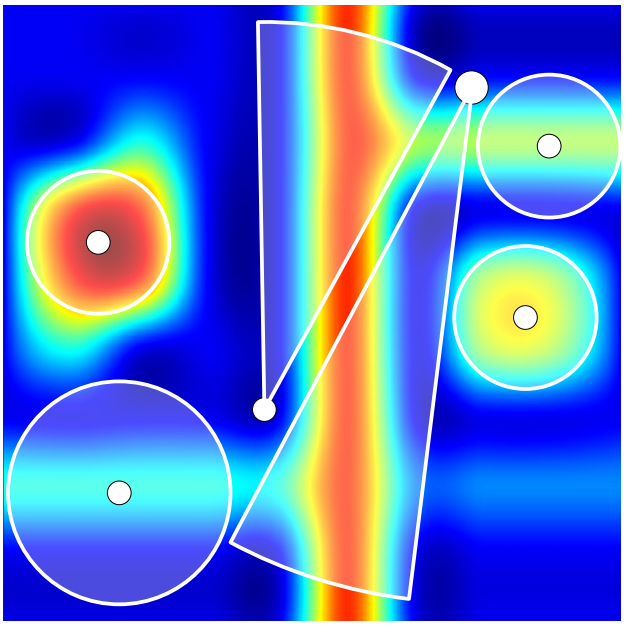


Figure 3-8 Example heatmap obtained by sensing data

3.2.2 Heatmap analysis and optimization

A heatmap representing the position of the obstacles of interest is obtained using the sensing functionality of the network, more concretely thanks to the sensing information offered by ISAC-capable APs or O-RUs (in O-RAN nomenclature). The heatmap can also represent the probable concentration of users and can therefore be used for the estimation of the spatial distribution of the traffic density. The expected spatial traffic density is proportional to the spatial user density. The small white circles observed in Figure 3-8 represent the RUs, and their coverage is given with the large white arc pies or circles.

In this scenario we are trying to solve an optimization problem, where we aim to achieve optimal coverage in the area of interest, making use of the available APs/RUs at their minimal transmit power. The coverage of the APs/RUs can be either modelled as a circle, with a radius of coverage depending on the transmit power, or as an arc pie, where the coverage area depends on the transmit power. These are the simplest models assumed for RAN devices.

3.2.3 Conclusions and Future Work

The high-level simulation we are going to perform will be used to investigate the possible power savings that can be obtained by optimizing the RU coverage, based on the available demand. Results related to this work will be included in BeGREEN D2.3, with the objective of extending the work in this direction until the end of the project.

3.3 B5G RAN enhanced through relay nodes

As presented in BeGREEN D2.1 [1], a RAN enhanced with relay nodes can reduce energy consumption while maintaining performance, thus providing a more energy efficient network. More specifically, the deployment of fixed or moving relays in a RAN can alter the amount of energy consumed by the different entities involved. On the one hand, the BS consumes less power since the users connected through the relay experience better propagation conditions, and thus they require less transmitted power from the BS than if they were directly connected to it. On the other hand, the relays themselves (fixed or mobile) consume power when they are active.

First, we studied a scenario with all the radio entities with a single antenna (i.e., BS, relay and UEs), and then we studied a more complex scenario where the BS uses massive MIMO.

3.3.1 Assessments of energy efficiency improvements through relays nodes

To evaluate the energy efficiency improvement using relay nodes, we formulate a power consumption model and then we assess the performance at the scenario of the University Campus of Polytechnic University of Catalonia (UPC) in Barcelona, which is described in Section 3.1.1 of BeGREEN D2.1 [1].

3.3.1.1 Power consumption model

This section presents the power consumption model considered in the presented studies. For this purpose, the required transmission powers of the BS and the relay to achieve a given bit rate for the UE and the total power consumption and energy efficiency computations are formulated. Figure 3-9 depicts the two connectivity situations considered by the model, namely 1) when a UE is connected directly to the BS, and 2) when it is connected through the relay. The downlink direction is considered in a single cell scenario. Out-of-band relaying is assumed, which means that the BS-Relay and the Relay-UE links operate in different frequencies.

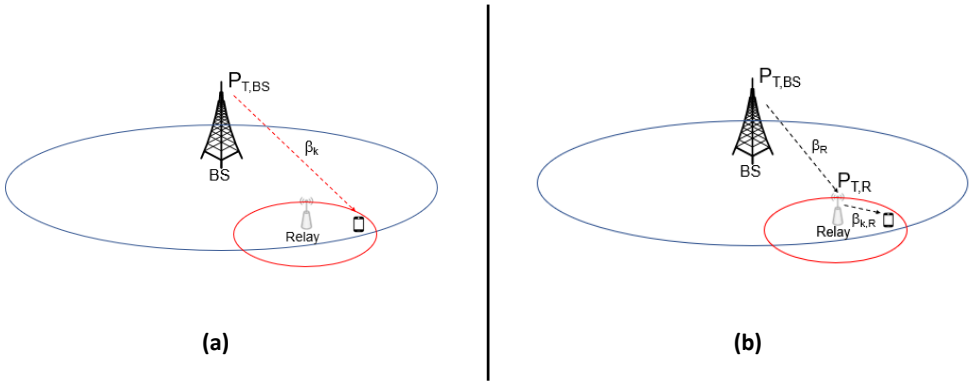


Figure 3-9 Considered connectivity situations: a) direct connection, b) connection through a relay

Let us start with the characterisation of the transmitted power at the BS. When a generic UE, denoted as UE k, connects directly to the BS, the Signal-to-Interference-plus-Noise Ratio (SINR) of the link BS-UE is defined as:

$$SINR_k = \frac{P_{T,BS}\beta_k}{P_N} \quad (1)$$

In this model, $P_{T,BS}$ refers to the BS transmission power, P_N refers to the noise power over a given bandwidth B, and β_k corresponds to the BS-UE link channel gain, taking into account antenna gains, propagation losses and shadowing, defined as:

$$\beta_k = \frac{G_T G_R}{L_k} \quad (2)$$

where G_T and G_R are the transmit and receive antenna gains and L_k is the total propagation loss in the link BS-UE that includes the pathloss and the shadowing.

And the bit rate achieved by this UE is given by:

$$R_k = B \cdot \varepsilon \, log_2(1 + SINR_k) \quad (3)$$

where the term ε is an efficiency factor $0<\varepsilon\leq 1$ that accounts for the overheads associated to cyclic prefix, reference signals, control plane signaling, etc.

Using (3) the minimum required SINR to achieve the bit rate R_k of the UE is given by:

$$SINR_{min,k} = 2^{\frac{R_k}{B\cdot\varepsilon}} - 1 \quad (4)$$

Consequently, substituting (1) in (4), we obtain that the BS transmission power for the direct link with the UE is defined as:

$$P_{T,BS} = \left(2^{\frac{R_k}{B\cdot\varepsilon}} - 1\right)\frac{P_N}{\beta_k} \quad (5)$$

When the UE is connected through a relay, and assuming that the BS and the relay operate at different frequencies, the total transmitted power consists of the transmitted power by the BS to the relay, and the transmitted power by the relay to the UE. Therefore, as shown in Figure 3-9 (a), there is a β_R that characterizes the BS-Relay link, and a $\beta_{k,R}$ that characterizes the Relay-UE link.

Following the same procedure as before, by means of equations (1) and (3), the obtained transmission powers are:

$$P_{T,R} = \left(2^{\frac{R_k}{B_R\cdot\varepsilon_R}} - 1\right)\frac{P_{N_R}}{\beta_{k,R}}, \quad (6)$$

$$P_{T,BS} = \left(2^{\frac{R_k}{B\cdot\varepsilon}} - 1\right)\frac{P_N}{\beta_R}, \quad (7)$$

where B_R is the relay bandwidth and ε_R is the efficiency factor at the relay-UE link.

To assess the power consumption, we consider the model proposed in [20], which assumes that the relation between the BS transmission power and the power consumption of a BS is nearly linear. Then, for the case that the UE is directly connected to the BS, the total power consumption is:

$$P_{TOT_D} = a_{BS}P_{T,BS} + P_{0,BS}, \quad (8)$$

where $P_{0,BS}$ represents the BS power consumption at zero RF output power associated to circuits, signal processing, etc., and a_{BS} corresponds to the linear dependency between the total BS power consumption and the radiated power $P_{T,BS}$.

Similarly, for the case that the UE is connected through the relay, the total power consumption is:

$$P_{TOT} = a_{BS}P_{T,BS} + P_{o,BS} + a_R P_{T,R} + P_{o,R} \quad (9)$$

where a_R and $P_{o,R}$ are the corresponding power model parameters for the relay.

Based on the power consumption expressions, the energy consumption (J/bit) for the direct link (E_D) and for the case of using the relay (E_R) are given by:

$$E_D = \frac{a_{BS}P_{T,BS}+P_{0,BS}}{R_k} \quad (10)$$

$$E_R = \frac{a_{BS}P_{T,BS}+P_{0,BS}+a_R P_{T,R}+P_{0,R}}{R_k} \quad (11)$$

And the energy efficiency measured in bit/J is defined as:

$$EE_D = \frac{1}{E_D} \quad (12)$$

$$EE_R = \frac{1}{E_R} \quad (13)$$

From these expressions, we can obtain the energy-saving value, by comparing the energy consumption values obtained in the direct connection to the BS case in the case that the connection is done by means of the relay:

$$EnergySaving(\%) = 100 \cdot \left(1 - \frac{E_R}{E_D}\right) \quad (14)$$

3.3.1.2 Performance assessment

The scenario where the described situations have been studied corresponds to the UPC campus, shown in Figure 3-10. This area contains several buildings and outdoor spaces and has 5G NR coverage provided by three outdoor macrocells which operate in band n78. To analyse the possible energy efficiency improvements, the impact of locating fixed relays at some selected indoor positions has been studied. The defined parameter values for the different tests are shown in Table 3-2.

Table 3-2 Parameters for Relay-Enhanced RAN Simulation

Parameter	BS	Relay	UE
Propagation model	UMa - 3GPP TR 38901	InH - 3GPP TR 38901	-
Antenna gain	10 dB	3 dB	3 dB
Bandwidth	20 MHz	20 MHz	-
Efficiency Factor BS	0.59	0.59	-
Power consumption parameters	a_{BS}=28.4, $P_{o,BS}$=156.38 W [21]	a_R=20.4, $P_{o,R}$=13.91 W [21]	-

Figure 3-10 Scenario under study: UPC campus

Regarding the term ε, it is defined as follows:

$$\varepsilon = \frac{14}{15}(1 - OH)\alpha_{TDD}, \quad (15)$$

where the fraction 14/15 reflects the useful time of the OFDMA symbols discounting the cyclic prefix. The term *OH* corresponds to the overhead associated to physical control signals (e.g. reference signals, physical downlink control channels, etc.). For the DL case in sub-6 GHz frequencies this term is given by *OH*=0.14 according to [22]. The term α_{TDD} is the ratio of DL OFDMA symbols in a TDD system. For example, in the recommended TDD configuration from [23], in which there are 3 DL slots of 14 symbols, one special slot with 10 DL symbols and one UL slot, the factor α_{TDD} equals 52/70.

The first performed analysis has been carried out at two specific buildings of the campus, which have been chosen based on the spectral efficiency maps of the campus obtained by means of simulations for an initial case in which the transmission power at the BS is fixed at 38 dBm. These maps are shown in Figure 3-11.

The selected buildings for the study are building A2 (at the ground floor) and building D5 (at the first floor), since they have significant areas where the spectral efficiency (defined as $\log_2(1+SINR)$) is lower than 1 b/s/Hz. In addition to this, they have an area in which the spectral efficiency is higher, which is the area where the relay will be located. For the studies performed at buildings D5 and A2, a relay has been placed at the locations shown in Figure 3-12.

Considering now the relay, the previously described power consumption model has been applied to determine the required transmitted powers at the BS and the relay to serve a UE that could be located in any of the pixels of the considered floors in the two buildings. Then, for each pixel, the total power, the energy consumption and the energy efficiency were computed for the cases that a UE located at that pixel was connected directly to the BS or was connected through the relay.

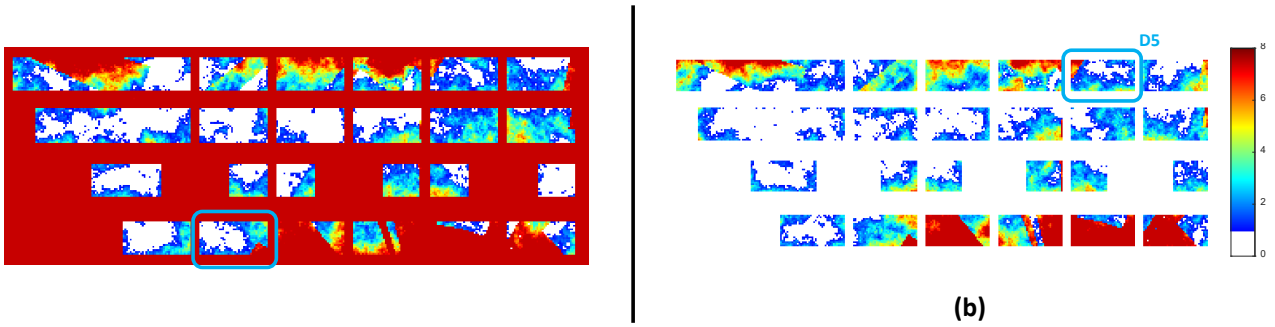


Figure 3-11 Map of the campus ground floor (a) and first floor (b) of simulated spectral efficiency $\log_2(1+SINR)$ without any relay and if transmitted power at the BS was fixed

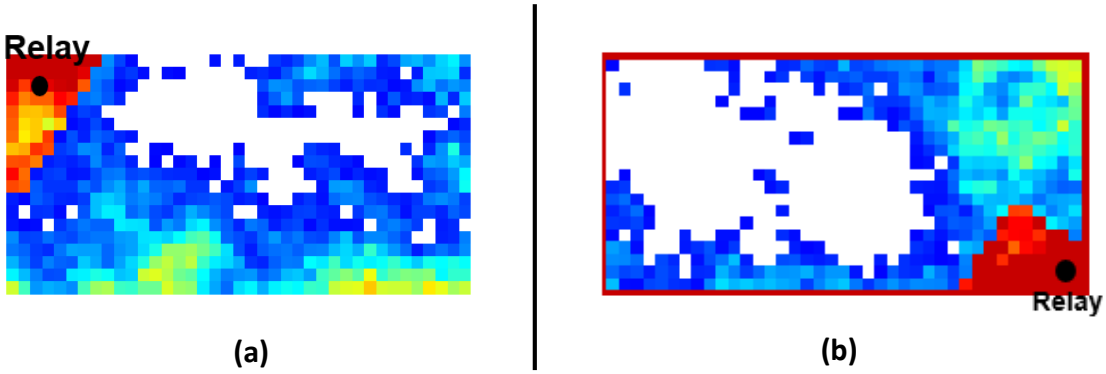


Figure 3-12 Spectral efficiency map and relay position of, (a) building first floor of D5; and (b) ground floor of A2

Figure 3-13 shows the CDFs of the energy efficiency at both buildings for a required bit rate R_k = 50Mbps, with and without the relay. We can see how, without the relay, about 85% of the positions of building D5 and 90% of the positions of building A2 have an energy efficiency lower than $1x10^5$ bit/J. However, when the relay is introduced, the whole floor obtains an energy efficiency of $2.6x10^5$ bit/J in the case of the D5 and $2.75x10^5$ bit/J in the case of the A2.

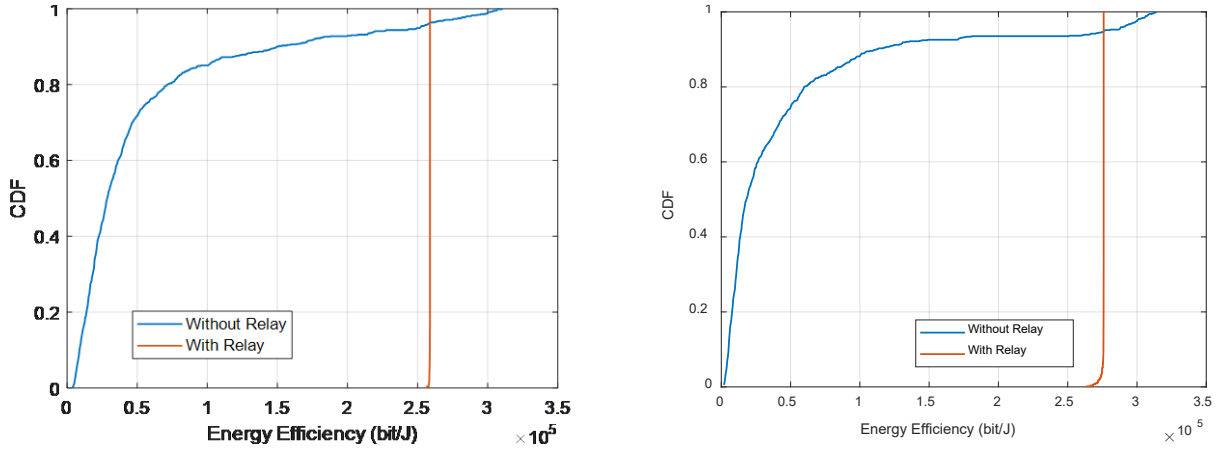


Figure 3-13 Energy Efficiency CDF for bit rate Rk=50 Mbps of, (a) first floor of D5, and (b) ground floor of A2

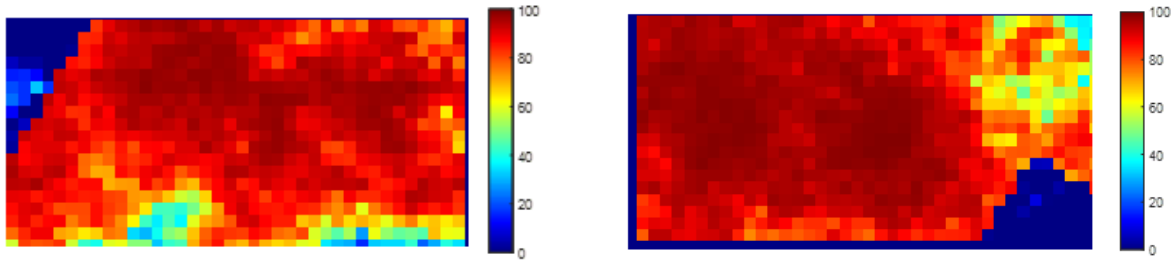


Figure 3-14 Energy saving map for required bit rate Rk=50Mbps of (a) first floor of D5, and (b) ground floor of A2

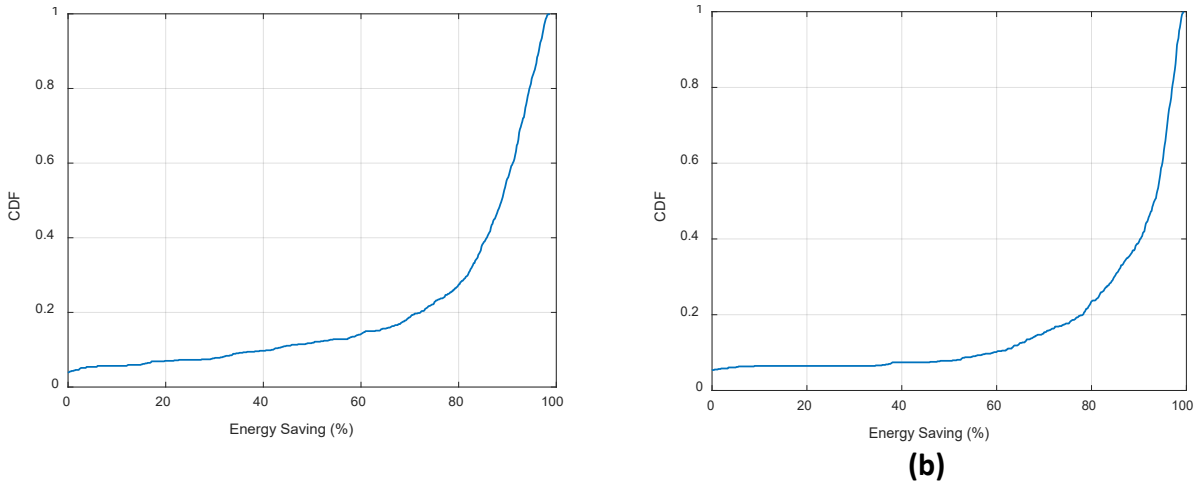


Figure 3-15 Energy saving CDF of first floor of D5, (a) and ground floor of A2, (b) for required bit rate R_k=50Mbps

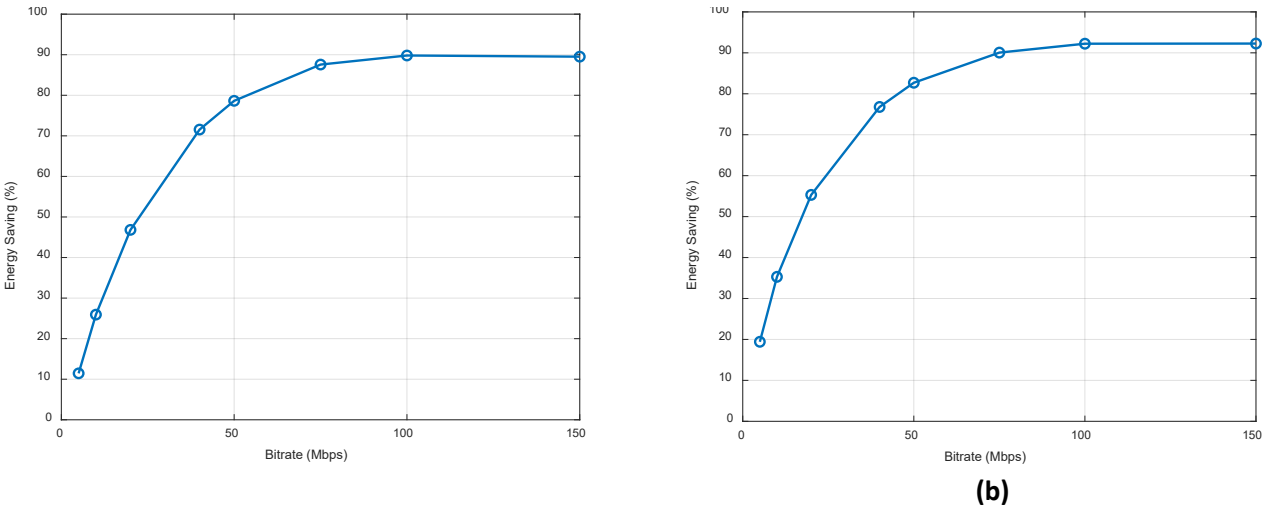


Figure 3-16. Average energy savings for different bit rates in, (a) first floor of D5, and (b) ground floor of A2

Checking the energy saving percentage values Figure 3-14, we observe how, in both cases, almost the whole floor obtains savings thanks to the introduction of the relay. The only exception are the pixels close to the position of the relay because in these positions the spectral efficiency with the direct connection is already high, so it is better to connect directly to the BS. Looking at the CDFs in Figure 3-15, it can be seen how about 90% of the positions at building D5 and 95% at building A2 obtain energy savings over 40%.

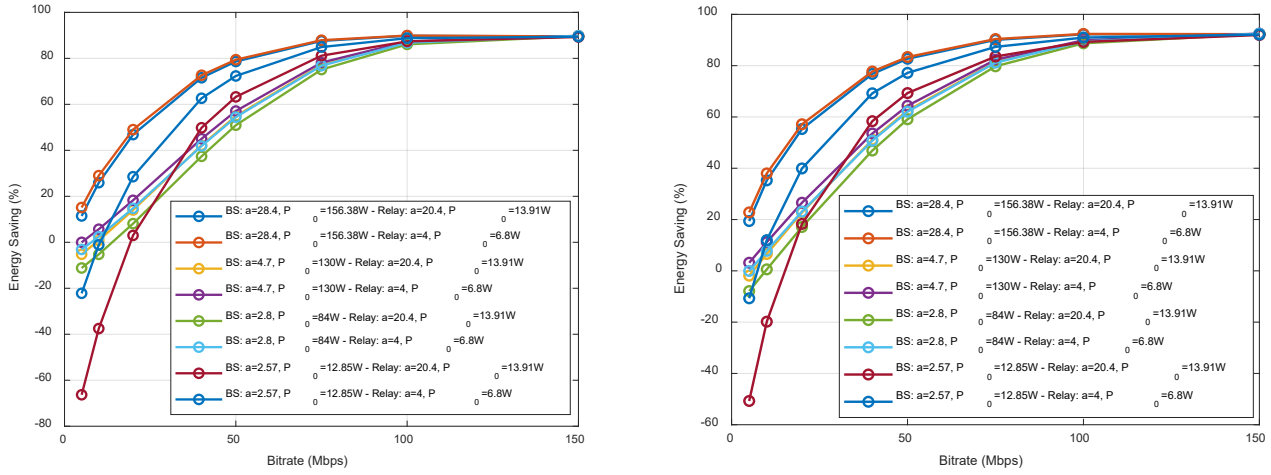
Figure 3-16 plots the energy saving values averaged for all the pixels with different required bit rates in both buildings. It is observed that energy savings increase with the required bit rate, going from about 10% in D5 and 20% in A2 when the required bit rate is 5 Mbps, up to 90% for a required bit rate of 100 Mbps.

As seen in equation (9), the total power consumption obtained in simulations strongly depends on the values chosen for the a and P_0 parameters for both the BSs and the relays. Therefore, various values obtained from different references have been considered and analysed, in order to determine their impact on the obtained energy-saving values. The parameter values that have been considered are shown in Table 3-3.

Table 3-3 a and P_o Parameter Value Combinations

BS			
Scenario	***a***	***P_0***	**Reference**
Macro	2.57	12.85 W	[24]
Macro	28.4	156.38 W	[21]
Macro	4.7	130 W	[25]
Micro	2.8	84 W	[25]
Relay			
Scenario	***a***	***P_0***	**Reference**
Fixed Relay	20.4	13.91 W	[21]
Pico cell	4	6.8 W	[25]

Figure 3-17 shows the plots of the energy saving percentages for buildings D5 and A2 for the different combinations of the power consumption model parameters. It can be seen how the lower the required bit rate, the bigger is the difference between the results obtained with different combinations. Checking the results for a required bit rate of 50 Mbps, for example, the savings range between around 50% and 80% in building D5 and between 60% and 85% in building A2. Besides, some combinations do not offer any gain at the lowest bit rates and even present losses. At a required bitrate of 5 Mbps, for example, only two combinations show remarkable energy savings of about 20%, which correspond to the parameters that involve a higher scaling factor and circuit power consumption per antenna.



(b)

Figure 3-17 Energy saving percentages for different α and P_0, (a) on the first floor of D5, and (b)ground floor of A2

The rest of the parameter combinations, however, show very little gains or even losses. For example, the combinations that involve the lowest circuit power consumption per antenna (P_0=12.85 W) present losses between 10% and 70%.

For higher bit rates such as 100 or 150 Mbps, however, all combinations show practically the same savings.

3.3.1.3 Conclusions

The power consumption improvements that can be obtained by locating fixed relays in certain buildings of a scenario corresponding to the UPC campus with 5G NR coverage are analysed. Several conclusions can be derived from the study performed:

- The insertion of a relay at an indoor position with high enough spectral efficiency can bring important energy savings which increase with the required bit rate and depend on the power consumption model parameters. For a required bit rate of 50 Mbps, energy savings range between 50% and 80% depending on the model parameters. For larger bit rates, energy savings of up to 90% have been found.
- The usage of relays brings notable improvements in energy efficiency. Specifically, a properly placed relays allows passing from a situation with over 85% of the positions of a floor with an energy efficiency under $1x10^5$ bit/J to a situation where almost the whole floor achieve an energy efficiency of $2.6x10^5$ bit/J.

Based on the obtained results, we plan to extend the analysis to other buildings in the campus in order to consolidate these conclusions.

3.3.2 Assessment of energy efficiency improvements through the use of relays in a mMIMO equipped base station

This section analyses the energy efficiency resulting in relay-enhanced B5G scenarios when massive MIMO is adopted at the BS. To evaluate the energy efficiency improvement using relay nodes, we formulate the power consumption model and we then assess the performance in a simulated scenario.

3.3.2.1 Power consumption model with massive MIMO

The model assumes a 5G NR mMIMO system with M antennas at the BS and K single-antenna UE devices numbered as k=1,...,K, with M>>K. The downlink direction is considered in a single cell scenario. This is illustrated in Figure 3-18.

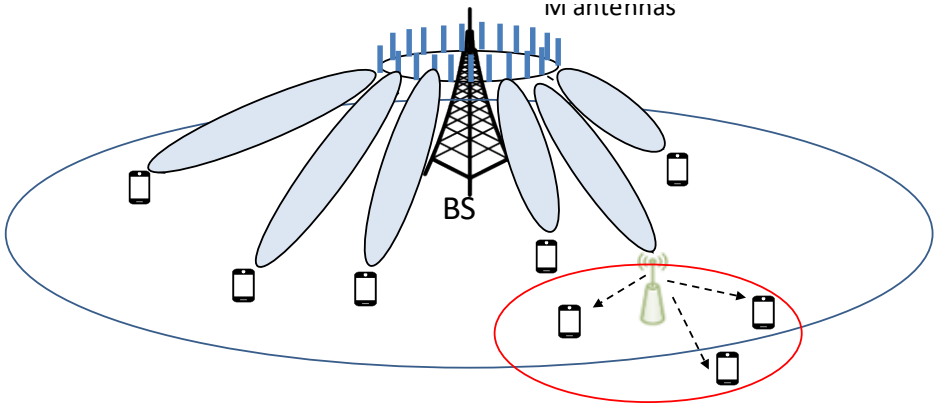


Figure 3-18 Considered relay-enhanced scenario when adopting mMIMO

Exploiting the mMIMO capabilities, the transmissions to the K UEs connected to the BS occupy the same bandwidth B. The channel gains of the different UE-BS links are independent.

A relay is also connected to the BS. The relay does not use massive MIMO and serves a total of K_R UE devices. From the perspective of the connectivity BS-relay, the relay is just modeled as one of the other UEs. Without loss of generality, it can be assumed that the K-th UE is the relay. The link between the relay and the UEs which it serves operates at a different frequency than the BS and uses a total bandwidth B_R. Moreover, the K_R UEs that are connected through the relay are numbered as k=K+1,...,K+K_R and each UE uses a different portion of the total bandwidth B_R.

Each UE connected to directly the BS or through the relay requires a certain bit rate denoted as R_k for k=1,...,K+K_R. The bit rate of the relay is given by the aggregate of all its served UEs, that is:

$$R_K = \sum_{k=K+1}^{K+K_R} R_k. \tag{16}$$

Below we formulate the required transmitted power at the BS and the relay to support the bit rate of the UEs. Then, the total power consumption in the scenario is presented.

Using the expressions of chapter 3.3 of [26] adapted to the notation used in this document and assuming the use of Maximum Ratio (MR) precoding, the SINR obtained by the k-th UE connected to the BS is given by:

$$SINR_k = \frac{M \cdot P_{T,k}\, \beta_k c_{CSI}}{P_N + \sum_{k'=1}^{K} P_{T,k'} \beta_k} \tag{17}$$

where $P_{T,k}$ is the transmit power of the BS to the k-th UE, k=1,...,K. This power is split across all the antennas in accordance with the precoding matrix. P_N=N_o·B is the noise power measured over bandwidth B with N_o the noise power spectral density and β_k is the large-scale fading in the link between the BS and the k-th UE. It accounts for the pathloss, shadowing and antenna gains, that is:

$$\beta_k = \frac{G_T G_R}{L_k}, \tag{18}$$

where G_T and G_R are the transmit and receive antenna gains, and L_k is the total propagation loss in the link BS-UE k which includes the pathloss and the shadowing.

Moreover, the term c_{CSI} ranges between 0 and 1 and reflects the quality of the channel estimation that impacts on the computation of the precoding matrix. Following [26] and [27], it is defined as:

$$c_{CSI} = \frac{\gamma_k}{\beta_k}, \tag{19}$$

where γ_k is the mean square of the channel estimate that includes both fast fading and large-scale fading. MMSE channel estimation is assumed. The mean square error of the channel estimation is (β_k-γ_k). For ideal channel estimation the error is zero, so β_k=γ_k and c_{CSI}=1.

The term in the denominator from (17) accounts for the noise and the total intra-cell interference. It is worth noting that this interference results, on the one hand, from the non-orthogonality of the channels of the different users, so that each UE *k'* different from *k* will generate interference to the UE *k*. On the other hand, the interfering term in the denominator corresponding to the same user *k'=k* is due to the beamforming gain uncertainty [26].

The spectral efficiency of the *k*-th UE can be estimated as:

$$S_k = log_2(1 + SINR_k). \tag{20}$$

And the bit rate achieved by this UE is given by:

$$R_k = B \cdot \varepsilon \, log_2(1 + SINR_k), \tag{21}$$

where the term ε is an efficiency factor that accounts for the overheads due to cyclic prefix, reference signals, control plane signalling, etc.

Using (21) the minimum required SINR to achieve the bit rate R_k of the *k*-th UE is given by:

$$SINR_{min,k} = 2^{\frac{R_k}{B \cdot \varepsilon}} - 1. \tag{22}$$

From (17) the required transmitted power by the BS to the *k*-th UE is given in terms of the SINR requirement $SINR_{min,k}$ as:

$$P_{T,k} = \frac{SINR_{min,k}\left(P_N + \beta_k P_{T,BS}\right)}{M \cdot \beta_k \, c_{CSI}} \tag{23}$$

where $P_{T,BS}$ is the total transmitted power of the BS for all the UEs, that is:

$$P_{T,BS} = \sum_{k=1}^{K} P_{T,k}. \tag{24}$$

By substituting (23) in (24) we get:

$$P_{T,BS} = \sum_{k=1}^{K} \frac{SINR_{min,k}\left(P_N + \beta_k P_{T,BS}\right)}{M \cdot \beta_k \, c_{CSI}} \ . \tag{25}$$

From this relationship, the total required power at the BS to support all the UEs is given by:

$$P_{T,BS} = \frac{\sum_{k=1}^{K} \frac{SINR_{min,k} P_N}{M \cdot \beta_k \, c_{CSI}}}{1 - \sum_{k=1}^{K} \frac{SINR_{min,k}}{M \cdot c_{CSI}}} \ . \tag{26}$$

By substituting (26) in (23) the required transmitted power of the BS to the *k*-th UE is given by:

$$P_{T,k} = \frac{SINR_{min,k} P_N}{M \cdot \beta_k \, c_{CSI}} \frac{M \cdot c_{CSI} - \sum_{k'=1}^{K} SINR_{min,k'} \left(1 - \frac{\beta_k}{\beta_{k'}}\right)}{M \cdot c_{CSI} - \sum_{k'=1}^{K} SINR_{min,k'}}. \tag{27}$$

Focusing now on the relay, since it does not support massive MIMO, it is assumed that the transmissions of the K_R UEs are multiplexed in frequency. Then, considering that the total bandwidth available at the relay is denoted as B_R, the bandwidth occupied by the transmission to the *k*-th UE is a fraction of the total bandwidth denoted as $B_k = \alpha_k B_R$, where $0 < \alpha_k < 1$. The general term α_k allows modelling different strategies. For example, an equal bandwidth allocation per UE can be modelled by setting $\alpha_k = 1/K_R$, or an allocation proportional to the bit rate requirement can be modelled by setting $\alpha_k = R_k/(R_{K+1} + ... + R_{K+K_R})$, etc.

The *SINR* requirement of the *k*-th UE in the link relay-UE is given by:

$$SINR_{R,min,k} = 2^{\frac{R_k}{B_k \cdot \varepsilon_R}} - 1, \tag{28}$$

where the term ε_R is the efficiency factor in the relay-UE link.

The required *SINR* relates with the transmitted power by the relay to the *k*-th UE as follows:

$$SINR_{R,min,k} = \frac{P_{TR,k}\beta_{R,k}}{N_o B_k}, \quad (29)$$

where $\beta_{R,k}$ is the large scale fading in the link relay-UE$_k$ accounting for antenna gains, pathloss and shadowing, and $N_o \cdot B_k$ is the noise power in the band used by the k-th UE. Then, the required transmitted power by the relay to the k-th UE is given by:

$$P_{TR,k} = \frac{SINR_{R,min,k} N_o B_k}{\beta_{R,k}}. \quad (30)$$

And the total power transmitted by the relay is the aggregate of the transmitted powers of all the UEs:

$$P_{T,R} = \sum_{k=K+1}^{K_R+1} \frac{SINR_{R,min,k} N_o B_k}{\beta_{R,k}}. \quad (31)$$

Following a similar model like in section 3.3.1.1, the total power consumption at the BS and the relay will be linearly dependent on the transmitted powers $P_{T,BS}$ and $P_{T,R}$ including a constant term per antenna (T_{RX}) that accounts for the circuit power consumption. This yields:

$$P_{TOT} = a_{BS} P_{T,BS} + M \cdot P_{o,BS} + a_R P_{T,R} + P_{o,R}, \quad (32)$$

where a_{BS}, a_R are the scaling factors of the transmitted power at the BS and relay, respectively. Similarly, $P_{o,BS}$ and $P_{o,R}$ are the circuit power consumption per antenna at the BS and relay, respectively.

Besides the transmission power at BS and relay, and the total power consumption, some other useful metrics to analyse the impact of the relay on the system are the energy efficiency and the energy saving percentage. Energy efficiency is computed as the total bit rate divided by the total power consumption:

$$EE(bit/J) = R_{TOT}/P_{TOT}. \quad (33)$$

Where R_{TOT} is the sum of the bit rates required by all the UEs.

Energy saving percentage is defined as the energy consumption relative difference between the total power consumption for the cases with and without relay, denoted, respectively as $P_{TOT,R}$ and $P_{TOT,NR}$. Then, the energy saving in percentage is given by:

$$EnergySaving(\%) = 100\left(1 - \frac{P_{TOT,R}}{P_{TOT,NR}}\right) \quad (34)$$

3.3.2.2 Performance assessment with mMIMO

To determine the impact that the usage of a relay has on power consumption, we have performed an analysis that considers a scenario in which both the cases with and without the relay have been studied. A base configuration has been defined with the parameters shown in **Error! Reference source not found.**. Based on this, the effect of varying different parameters has been studied.

The considered scenario as seen in Figure 3-19 sets out a situation in which there are K=15 UEs. 14 of them have the pathloss L_{BS} with the BS and the UE 15 has a higher pathloss value L_{BS}+L_{incr}. The pathloss increase L_{incr}(dB) could be due for example to having an obstacle between the BS and the user. With this configuration, two cases have been studied, namely the case where the UE 15 connects to a fixed relay (Figure 3-19 a) and the case where it connects directly to the BS (Figure 3-19 b). The relay has also an associated pathloss value L_{BS} in the link BS-Relay, and the distance between it and the 15th UE is denoted as $d_{UE\text{-}Relay}$. This distance is associated to a pathloss $L_{R\text{-}UE}$ in the link Relay-UE, according to the defined propagation model, which is the UMi model with LOS, in this case.

The values assigned initially to the parameters a_{BS}, a_R, $P_{o,BS}$ and $P_{o,R}$ in the given case are shown in the Table 3-5 and have been obtained from [21].

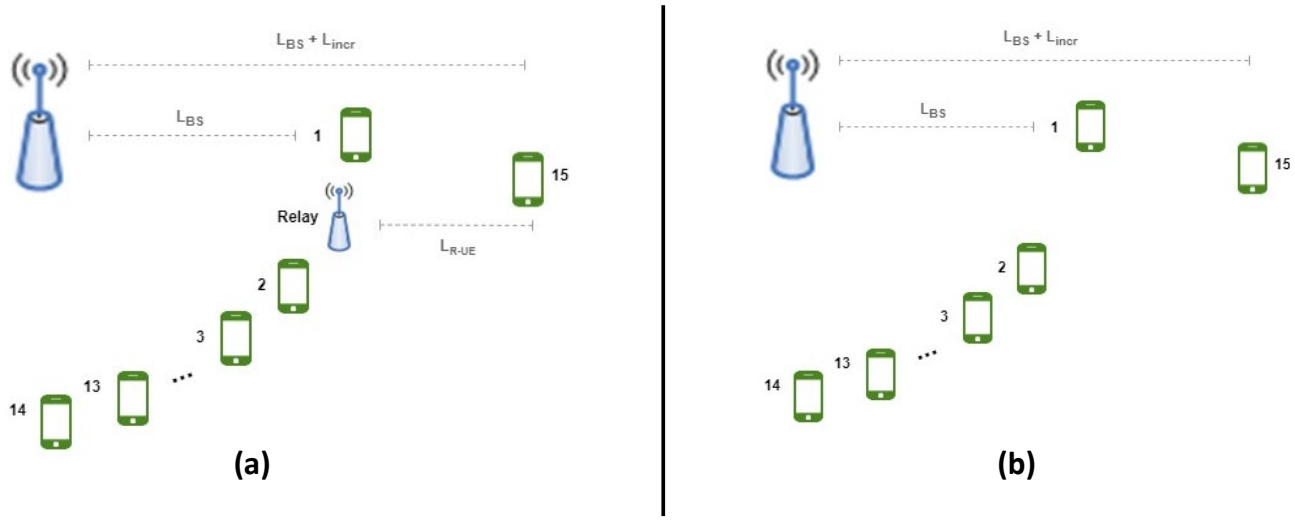


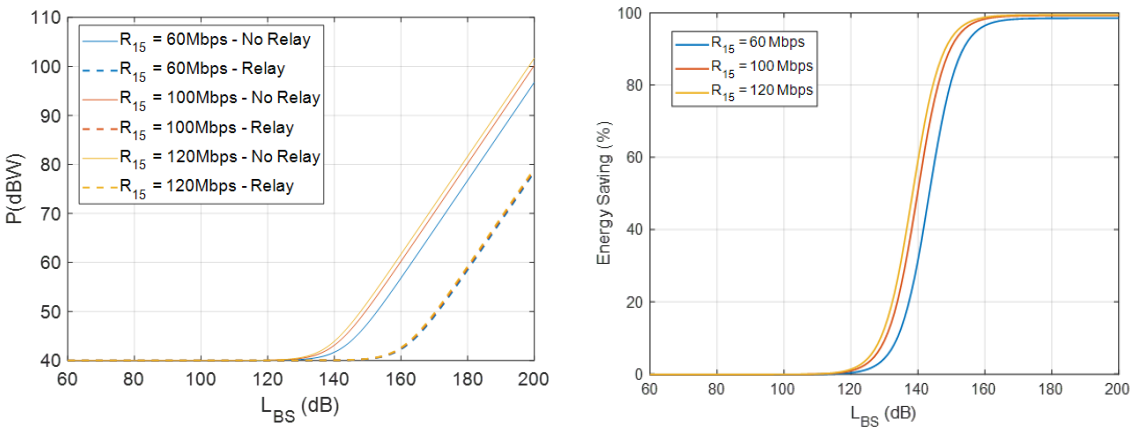
Figure 3-19 Considered scenario with, (a) relay, and (b) without relay

Table 3-4 mMIMO BS with Relay Scenario Configuration Parameters

BS		Relay		UE	
f_{BS}	3.7 GHz	f_R	3.5 GHz	G_U	3 dB
M	64	B_R	100 MHz	R_k	60 Mbps
B	100 MHz	G_R	3 dB	L_{BS}	140 dB
G_B	10 dB	K_R	1	L_{incr}	30 dB
K	15	α_k	$1/K_R$	$d_{UE\text{-}Relay}$	100 m
c_{CSI}	1	N_0	-168 dBm/Hz	N_0	-168 dBm/Hz

Table 3-5 Initial Power Consumption Parameters

a_{BS}	28.4
$P_{0,BS}$	156.38 W
a_R	20
$P_{0,R}$	13.91 W



(a) **(b)**

Figure 3-20 For different R_{15} values, (a) total power consumption, and (b) energy saving

The first performed analysis involves the variation of the pathloss value L_{BS} between the BS and the UEs for different values of required bit rate for the 15th UE, denoted as R_{15}.

Figure 3-20(a) shows the total power consumption at the BS and the relay defined in eqn. (32) for three different R_{15} values as a function of the pathloss value L_{BS}. The continuous lines show the power consumption when there is no relay while the discontinuous lines show the power consumption when the relay is introduced.

It can be observed that for small pathloss values, there is no difference of power consumption between the cases with and without relay. This happens because the constant term corresponding to the circuit power consumption has the biggest weight on the total power consumption and the changes on the transmission power do not have significant impact on the total value. The total power consumption is 40 dBW in that case, which corresponds to the value obtained from the constant power consumption of $P_{0,BS}$=156.38W for each one of the 64 antennas. However, when the pathloss is higher than 120 dB, the total power consumption starts to grow, and so does the difference between both cases, achieving a lower consumption when the relay is introduced.

It can also be observed that the difference between both cases is more significant when the value of the bit rate required by the 15th UE is higher. For example, the difference in the case of R_{15} = 60 Mbps at L_{BS} = 140 dB is around 1 dB, while in the case of R_{15} = 120 Mbps is around 4 dB. We can also mention that, for example, in the case that the 15th UE requires 120 Mbps, the total power consumption is the same with a pathloss of 148 dB and no relay and with a pathloss of 170 dB and a relay. On the other hand, it can be seen that the difference in the required bit rate has very little effect on the total power consumption in the relay case. The difference in power consumption between the highest and lowest required bitrate in this case is about 0.7dB. This difference is mainly caused by the difference in the BS transmission power, which is also quite small in the cases where the relay is present. The relay transmission power does change more significantly with the required bitrate, from 9 dBm to 14 dBm. However, the relay power consumption has very little effect on the total power consumption since the power consumed by the BS is much bigger, so this difference is not reflected in the total power consumption.

Checking the energy saving Figure 3-20(b) we can also see that when L_{BS} is higher than 120 dB, the energy saving starts to grow quite fast. For example, for L_{BS}=140 dB, values of 31% for R_{15} = 60 Mbps and 59% for R_{15} = 120 Mbps are obtained. Furthermore, with a pathloss L_{BS}=160 dB, savings of 96% and 98% are achieved for the cases R_{15} = 60 Mbps and R_{15} = 120 Mbps, respectively.

Another study has been to vary the pathloss increase between the BS and the last UE. Figure 3-21 depicts the total power consumption results for three different values of the pathloss between the BS and the UEs (L_{BS}) and for two different bit rates. It is observed that, the higher the pathloss increase, the bigger is the difference in the total power consumption. For example, in the case of R_{15}=60 Mbps and L_{BS}=140 dB, a gain of about 8 dB is achieved with L_{incr} = 40dB, and the gain goes up to about 25 dB with that same L_{incr} and L_{BS} = 160 dB. Also, when the pathloss between the BS and the UEs is bigger, a smaller pathloss increase is needed to achieve a significant gain. For example, in the case of R_{15}=60 Mbps when the BS pathloss is 160 dB, a gain of about 6 dB is achieved with a pathloss increase of 20 dB, while that same gain is achieved with an increase of about 38 dB when the pathloss is 140 dB.

In the case of R_{15}=120 Mbps the gains obtained are higher, since the higher required bit rate involves a higher transmission power at the link with the pathloss increase. This time, when the pathloss L_{BS} = 160 dB, a gain of about 10 dB is achieved for a pathloss increase of L_{incr}=20 dB, and the gain goes up to 30 dB with the same L_{BS} and L_{incr}=40 dB.

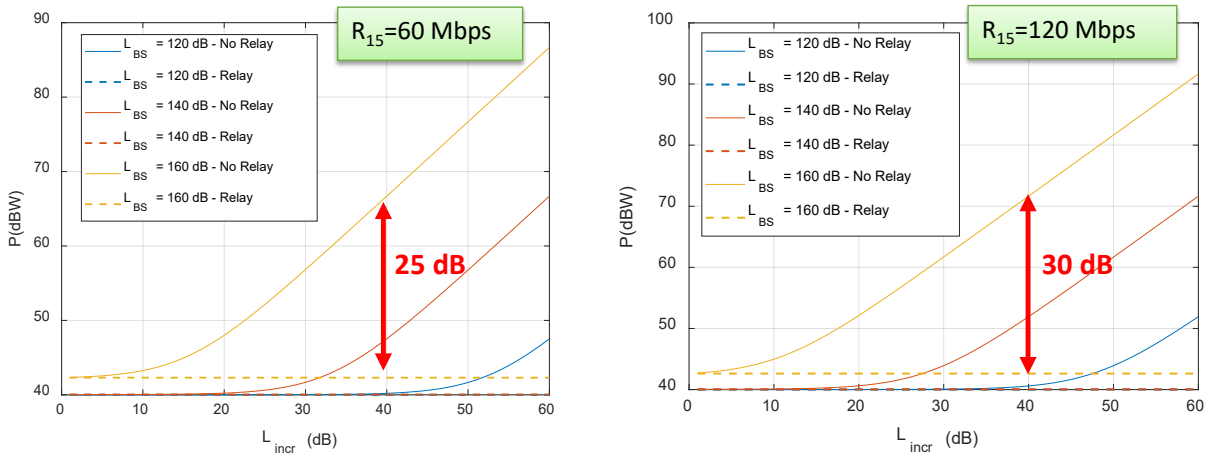
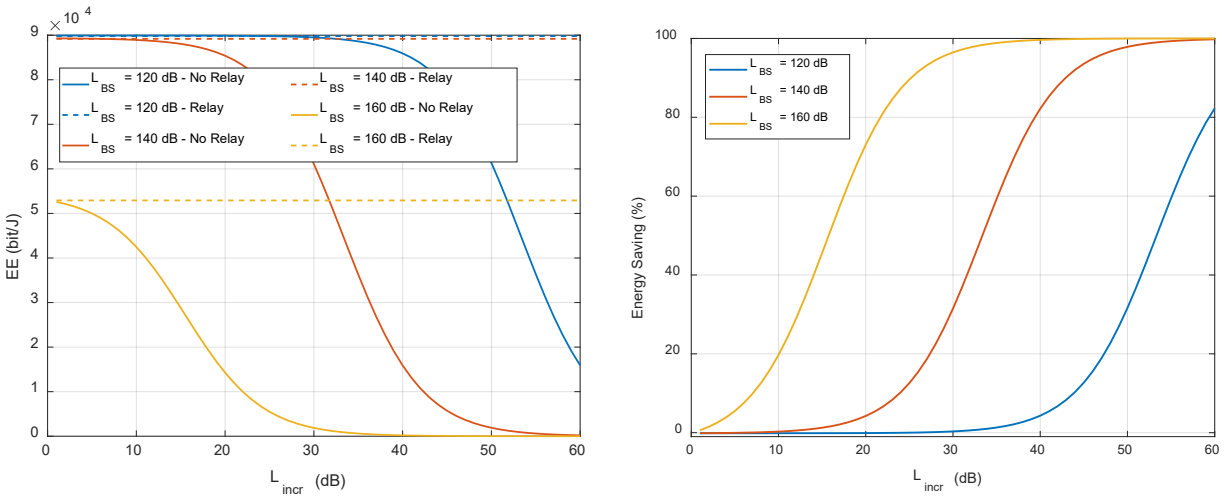


Figure 3-21 Total power consumption for different L_{incr} values.

Figure 3-22 plots the energy efficiency and energy saving values. If we observe the energy efficiency, we can see that, when there is no relay, it decreases as L_{incr} grows, while it is maintained constant in the case of having a relay, since L_{incr} does not affect in this case. This means that the higher L_{incr}, the bigger is the difference between both cases. Also, when L_{BS} is higher the decrease of the case without relay starts at lower values of L_{incr}. In the case of the achieved energy saving, we can also see how it grows with L_{incr}, and how that growth is achieved for lower values of L_{incr} when L_{BS} is higher. For example, an energy saving of 20% is achieved with a L_{incr} of 10 dB for an L_{BS} of 160 dB, while that same gain is achieved with a L_{incr} of 47dB when L_{BS} is 120 dB.



(b)

Figure 3-22 Energy efficiency (a) and energy saving (b) for different L_{BS} values

3.3.2.3 Conclusions

This section has analysed the energy efficiency and energy savings obtained in a relay enhanced scenario with a BS with Massive MIMO. The performed study allows to draw several conclusions:

- The higher the pathloss between the BS and the UEs, the bigger is the achievable energy saving percentage by means of the relay, which grows from about a 10% with L_{BS}=130 dB to over a 95% with L_{BS} = 160 dB. Below a pathloss value of about 110 dB, however, there is no energy saving.
- The bit rate required by the UE that connects to the relay also has an impact on the energy efficiency and power consumption, making the energy savings bigger when the required bit rate is higher.

- The obtained energy savings grow when increasing the additional pathloss (L_{incr}) experienced by the UE connected to the relay. For example, when the pathloss between the BS and the rest of users including the relay is 140 dB, the energy saving ranges between 5% and 35% when the pathloss increase L_{incr} varies between 20 and 30 dB. This growth is even bigger when the pathloss between the BS and the rest of users increases.

Based on the obtained conclusions in the considered study, it is planned to consolidate the results to further assess the conditions under which the benefits of relays with massive MIMO can be more beneficial.

3.4 Experimental analysis of energy-aware computing allocation in vRANs

Traditionally, the major cause behind energy consumption in cellular networks has been the power needed to amplify and transmit wireless signals from BSs [28][29][30]. Based on this, most of the research endeavors have been mainly focused on the analysis of the transmitted power.

However, the landscape is rapidly changing. The current trend towards network densification involves deploying a multitude of smaller BSs strategically placed throughout an area. These small cells operate with much lower transmission power, as they only need to cover a shorter range. This reduces the overall energy consumption associated with signal transmission [31].

Moreover, this shift towards densification brings with it a new challenge: the processing power required to handle high-bitrate signals. As data rates soar, the network infrastructure needs to be able to process this information efficiently. This has led to a surge in the computing capacity demanded by the RAN. Network virtualization, where network functions are software-defined and run on centralized off-the-self hardware, further amplifies this demand [32][33].

The result of these trends is a shift. While transmission power is still a significant factor in many use cases, the computational power needed for processing high-speed data is becoming an increasingly relevant component of the RAN's overall energy consumption. This change requires new research efforts that prioritize energy efficiency not just in transmission hardware but also in the processing and virtualization technologies that power modern cellular networks.

This section addresses the use case where several vBS are deployed in the same shared computing infrastructure. Nonetheless, while shared computing platforms offer enhanced flexibility and cost-effectiveness, they also introduce challenges for 5G BSs, as they compromise the predictability offered by dedicated platforms [34] [35].

The term *noisy neighbour problem* has been coined to refer to the issue when shared resources are consumed in extremis, meaning that another function restricts one virtualized function’s resources. This problem has motivated substantial research over the years [34] [36].

In this section, we study the noisy neighbour problem and how the computing resources and memory resources impact on the noisy neighbour problem and the computing usage in general. The optimization of the computing usage is very important since, as shown in Figure 3-23, there is a linear relation between the computing usage and the energy consumption. Specifically, Figure 3-23 shows the relationship between the normalized energy consumption on top of the system's baseline (i.e., idle) consumption of our experimental platform (described in Sec. 3.4.3) as a function of the total computing load. Based on this, we can conclude that computing usage minimization and energy minimization are equivalent problems.

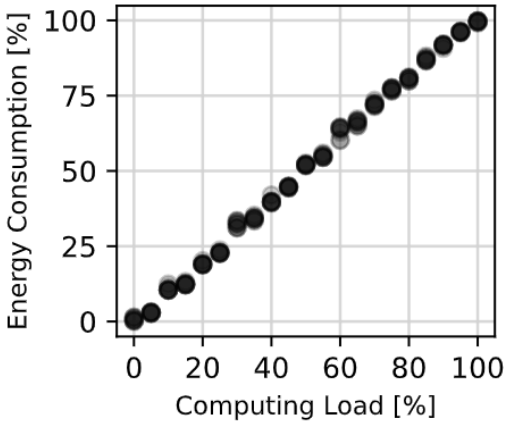


Figure 3-23 Energy consumption as a function of the computing load

3.4.1 Motivation

As mentioned in the previous section, the term noisy neighbour has been coined to refer to the issue when shared resources are consumed *in extremis*. We confirm this by deploying multiple vBS instances using Docker containers into a pool of computing cores in a shared off-the-shelf server. In these experiments, we configured the pool of computing cores to achieve as much predictability as possible.

Figure 3-24 depicts the aggregated computing usage per core in our vRAN platform as a function of the number of deployed vBS under maximum traffic load in uplink (UL) and downlink (DL). The bars in yellow show the expected usage assuming perfect resource isolation in place. We compute these by linearly scaling up the CPU usage of a single vBS instance. The red bars show the actual CPU consumption, which unveils an exponentially growing overhead induced by the aforementioned noisy neighbours problem. The increased computing overhead has its roots in *the lack of cache memory isolation*. The computing overhead poses an issue of the *increase of the total energy consumption* of the vRAN platform.

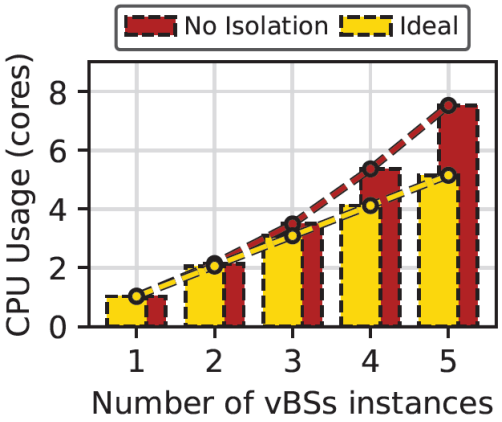


Figure 3-24 Aggregated per-core usage as a function of the number of vBS instances in our experimental vRAN platform

3.4.2 Background on general-purpose computing platforms

In this section, we provide an overview of the structure of cache memory in general-purpose computing platforms. Modern processors have different computing cores which can execute one thread at a time. The Operating System (OS) scheduler decides where to allocate the different running threads according to different criteria.

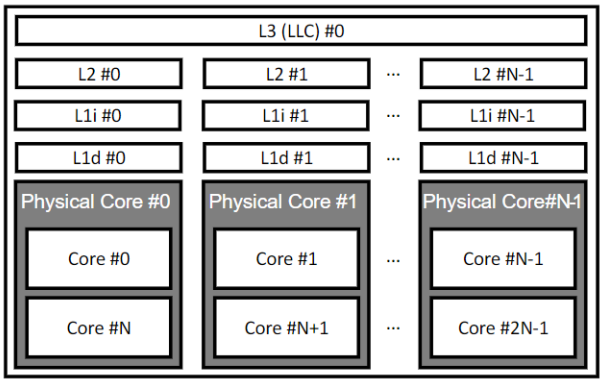


Figure 3-25 General-purpose CPU platform architecture

Figure 3-25 presents the CPU architecture of a radio computing platform. Modern superscalar processors leverage on simultaneous multithreading (SMT), also known as Hyper-threading in Intel CPUs. This technology enables a physical core to execute multiple threads concurrently, essentially presenting each computing core to the operating system as two distinct cores. However, half of the total number of cores are virtual and share the same physical computing core.

Cache memory bridges the speed gap between RAM and the CPU itself and it is organized into various levels based on speed and size. The first two cache memory levels called L1 and L2 cache are the closest and fastest to the system, although its capacity is limited. Each physical core has its dedicated L1 and L2 cache. L1 is faster than L2 but smaller in size.

In modern processors, L1 cache memory comes in two types: (i) a memory for storing instructions (L1i) and (ii) a memory for storing data (L1d). Each physical core has its dedicated L1 cache. L2 cache, on the other hand, is larger but slower than L1, and it is also dedicated to each physical core. Unlike L1, L2 cache is predominantly used for storing data rather than instructions.

Lastly, the L3 cache, also known as the *Last Level Cache (LLC)*, is the slowest CPU's cache and is shared among all computing cores [37][38]. Table 3-6 shows the access latency and the available size per core of different cache levels in the Intel Skylake architecture.

Table 3-6 Access Latency at Different Cache Levels

Memory Type	Access Latency	Size per Core
L1 cache	4-6 cycles	64 KB
L2 cache	14 cycles	256 KB
L3 (LLC) cache	50-70 cycles	2 MB
RAM	120-600 cycles	2-4 GB

Cache memory is organized into several associative sets also known as cache ways. A cache way is a set of lines or blocks that can store data from a specific location in the main memory. Tools such as Intel CAT[1] (Cache Allocation Technology) enable allocating LLC cache ways to specific cores or processes.

In a general-purpose computing platform, when a core executing a thread requires frequently used memory blocks, it loads them into the cache for quicker access. Consequently, if a thread references a memory block that is not present in the cache, the core initiates an interrupt known as a “cache miss” and searches for the data in a higher memory level. This incurs additional CPU cycles which increase the total execution time of a

[1] Intel CAT is an open-source tool and can be found online as a part of Intel RDT Software Package: https://github.com/intel/intel-cmt-cat

computing thread.

3.4.3 Experimental methodology

To evaluate the impact of cache memory resources on the noisy neighbour problem on the vRAN energy consumption, we measure the computing usage and low-level cache metrics by deploying a different number of vBS instances in diverse scenarios and different cache memory configurations in our vRAN platform.

Figure 3-26 shows our experimental vRAN platform. Therein, we deploy the different vBS instances in an isolated pool of computing cores from an Intel Xenon E5-2650 v4 CPU @ 2.20GHz in a shared off-the-shelf server. Each vBS has its dedicated RF radio head connected to one UE, which is used to emulate the aggregated cell load. We configured the pool of computing cores to retain as much predictability as possible: (i) we isolated 12 cores with 12 dedicated Last Level Cache (LLC) ways, (ii) the system can only use C0/C1 C-states and we turned off hardware P-states, and (iii) we deactivated Hyper-threading. We then initiate bidirectional data flows, both uplink (UL) and downlink (DL), with maximum load and good wireless channel conditions between each vBS instance and different user equipment (UE). We consider the following scenarios:

- *Ideal*. We compute the CPU usage scaling linearly the usage of a single vBS instance assuming that the cache memory size also scales linearly.
- *No isolation*. We deploy an increasing quantity of vBS instances without any cache memory isolation mechanism.
- *Pinning*. L1 and L2 cache levels are dedicated per core. We pin different vBS instances to distinct cores to assess the impact on L1 and L2 cache isolation. To facilitate comparisons, the number of cores assigned to each vBS is given by:

  $$cores\ per\ vBS\ = \left\lfloor \frac{total\ cores}{deployed\ vBS} \right\rfloor,$$

  where the total number of cores equals 12 and the number of deployed vBS increases from 1 to 5. Note that when 5 vBSs are deployed, each vBS is assigned to 2 cores and two free cores are left.
- *Pinning + LLC isolation*. We perform the L3 cache allocation in the same manner as the CPU pinning. We allocate the total number of cache ways equally for every single vBS, i.e., the number of cache ways per vBS is given by:

  $$cache\ ways\ per\ vBS = \left\lfloor \frac{total\ cache\ ways}{deployed\ vBS} \right\rfloor.$$

In our platform, there are a total of 12 cache ways. We use Intel CAT to allocate the LLC cache ways to the corresponding computing cores. To measure the computing usage, we use the */proc* filesystem to read and store periodically the computing time for the different threads of a vBS on the computing cores allocated. On the other hand, we used the tool *perf* to measure the number of instructions per cycle (IPC) and the number of cache misses per 1k instructions (MPKI) by one vBS.

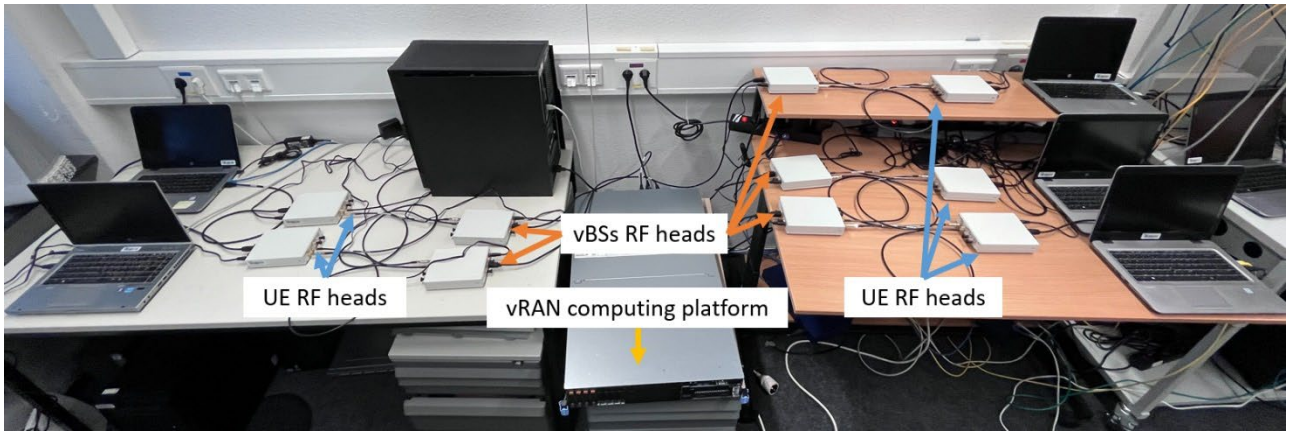


Figure 3-26 vRAN experimental platform

3.4.4 Experimental characterization of memory allocation

Figure 3-27 shows the measured computing usage in the different scenarios. We observe that, for the no isolation configuration, the computing usage increases by approximately 50% compared to the ideal case, increasing the energy consumption. This also impacts the IPC and MPKI, as shown in Figure 3-28 and Figure 3-29, respectively. We observe that the IPC decreases, and the cache misses increase, as more instances are deployed. There is a sixfold increase in the cache misses when transitioning from 1 to 5 vBSs. This increase also impacts the number of cycles required to execute the same number of instructions, decreasing the IPC.

The *Pinning* configuration shows a lower computing usage than the *No isolation* but still higher than the ideal case. The computing usage decrease is significant considering that L1 cache ways are approximately 100 times lower in size than L3 cache ways and L2 cache ways are 10 times lower in size than L3 cache ways. In Figure 3-28 we can observe a higher IPC across any number of deployed vBSs compared to the *No isolation* scenario. However, Figure 3-29 shows the same number of cache misses for all the cases. This might seem shocking at first glance, but as we have previously detailed L1 and L2 cache way size is very low compared to L3.

Finally, we study the effect of the L3 cache isolation. In Figure 3-27, the *Pinning + LLC isolation* configuration does not improve computing usage. Specifically, the computing usage is the same for all the cases except for the case with 5 vBS, in which it is slightly higher. The reason behind this behavior is that we only allocate 10 out of 12 L3 cache ways to all vBSs for the case of 5 vBSs, while in the *Pinning* configuration, all vBSs have all the L3 cache memory available, using on average 2.4 L3 cache ways.

Next, we study how the allocation of LLC cache ways impacts the computing usage of a vBS instance. First, we measure the percentage of LLC cache memory used of the total LLC memory allocated to a vBS as a function of the traffic demand. Figure 3-30 shows the LLC occupancy with 12 and 2 LLC cache ways for different SNR values and traffic demands in uplink and downlink. The *high* series depicts the LLC occupancy with a high SNR environment while the *low* series depicts the LLC occupancy with a low SNR environment. We can see that the total LLC occupancy is above 80% for uplink and downlink. Also, the LLC occupancy is almost orthogonal to the demand of the vBS, yielding an increase of 2-6% when the demand goes from 10% to 100%.

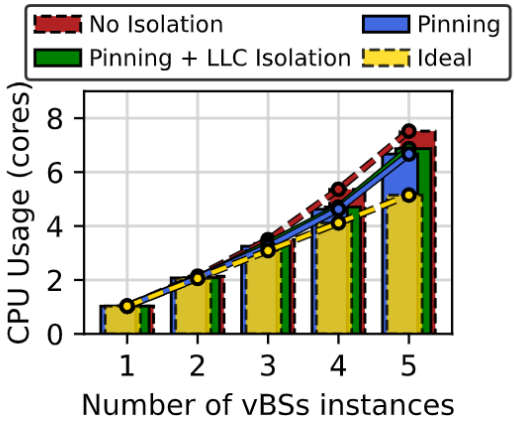


Figure 3-27 Comparison of the aggregated per-core usage with number of vBS instances showing the "No isolation", the "Pinning", and the "Pinning + LLC isolation" scenarios

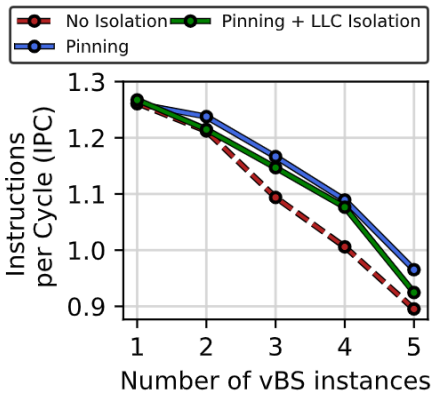


Figure 3-28 Instructions per cycle (IPC) as a function of the number of deployed vBSs

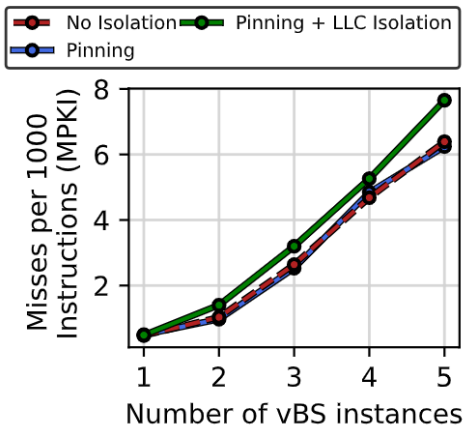


Figure 3-29 Misses per 1000 instruction (MPKI as a function of the number of deployed vBSs

Finally, Figure 3-31 depicts the computing usage of a vBS as a function of the L3 cache ways when we deploy it using three cores. Figure 3-31 depicts the computing usage for the different Modulation Coding Schemes (MCSs) with a low traffic demand (i.e., 20% of the total demand) and high traffic demand (i.e., 100% of the total demand) for uplink and downlink. We observe that there are significant differences between the achievable computing usage reduction. For high traffic both in uplink and downlink, we can achieve a more significant computing usage reduction. On the contrary, a vBS that processes a low traffic demand can attain lower gains in terms of computing usage. We conclude that LLC resources have different *utility* depending on the vBS context.

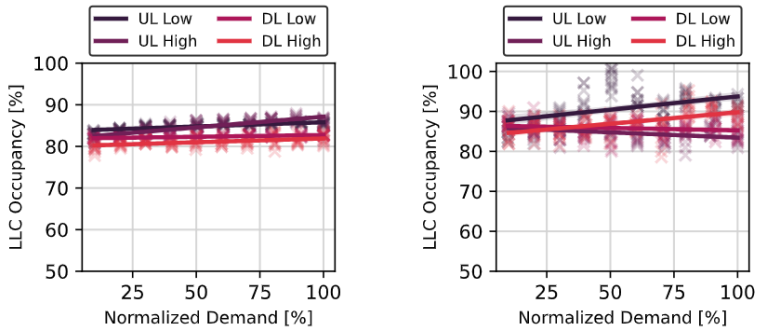


Figure 3-30 LLC occupancy in percentage as a function of the total demand for different SNR cases in UL and DL

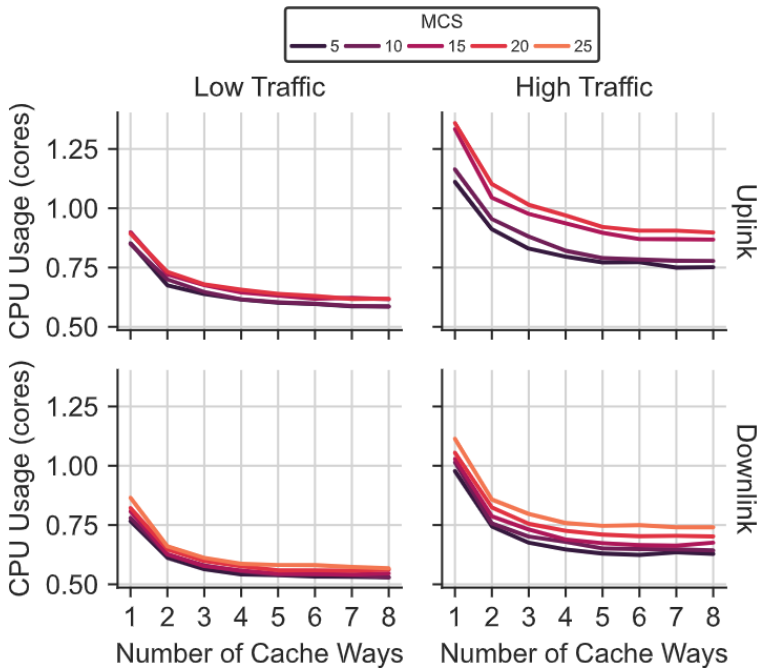


Figure 3-31 Computing usage as a function of the LLC allocated cache ways for different SNR and in UL and DL

This contrasts with Figure 3-30 which shows that the vBS makes full use of the cache memory regardless of the demand. Thus, if there is no LLC cache allocation mechanism, all vBSs deployed will be using the same amount of LLC cache memory on average. It is key to strategically distribute the LLC cache ways among the vBSs deployed boosting its utility to minimize the computing usage and therefore the energy consumption depending on the traffic demands.

Quantifying the LLC cache memory utility is a challenging task. The LLC cache utility depends on the computing demands of the vBS which are influenced by various factors, including traffic demand in both the downlink (DL) and uplink (UL), the signal-to-noise ratio (SNR) of each wireless link, and the specific MCS utilized for communication. All these elements interact in a complex manner. Figure 3-32 (a) depicts the relative mean core usage of the vBS, and shows that, given an MCS, lower SNR regimes demand a higher amount of computing resources. The underlying reason is the iterative nature of the forward error correction (FEC) algorithms -- signals received with lower SNR require a higher number of FEC iterations to decode the transported codeword successfully. This is confirmed by Figure 3-32 (b), which shows the amount of time taken by the decoder to finish its task for every transport block.

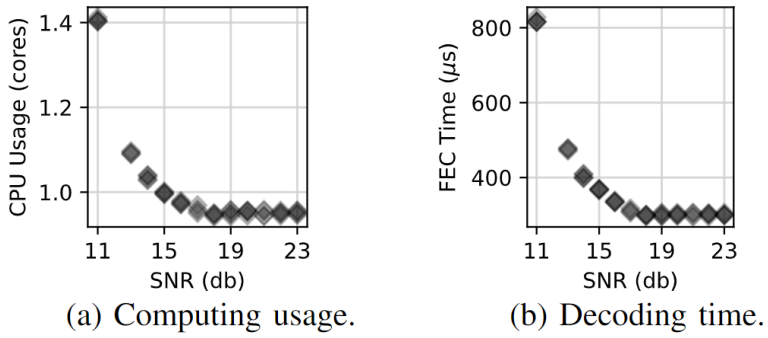


Figure 3-32 Computing usage and decoding time of a vBS with max. UL and DL load with mild MCS over different SNR conditions

3.4.5 Experimental characterization of computing requirements

In light of the above characterization of the cache memory requirements, the computing requirements of vBSs depending on their radio conditions or networking load are studied.

Given a bandwidth configuration, the radio scheduler of the vBS is free to allocate radio resources (physical radio blocks or PRBs) to individual users. The impact that different scheduling decisions in a single vBS, in both uplink and downlink, have on computing consumption is depicted in Figure 3-33. As expected, computing consumption is linearly dependent on the number of PRBs allocated. Moreover, processing uplink data is costlier than downlink data, which is also expected as FEC operations are more compute-involved when decoding data (which may require several iterations over the received data) than when encoding data.

The maximum eligible number of PRBs that can be allocated per TTI depends on the bandwidth settings of the vBS. The selected bandwidth can influence the workload of the vBS in two ways. First, Fast Fourier Transform (FFT) and its inverse IFFT counterpart are tasks required by a vBS every TTI to convert time-domain data into spectrum-domain data for signal processing (and vice versa). Larger bandwidth settings incur a higher computational burden on these tasks with a linear relationship. Second, larger bandwidth settings enable larger transport blocks (chunks of user data), which need to be encoded (downlink) or decoded (uplink) every TTI. It is well-known that the computing complexity of a 4G/5G decoder scales linear with the amount of information to encode/decode.

Figure 3-34 depicts the CPU usage for four different bandwidth settings ranging from 1.4 MHz to 10 MHz in a setup with 5 vBS instances. We compare our measurements with an ideal case of perfect isolation. For the latter case, we measure the CPU consumption of one single vBS with the appropriate settings and linearly increase the rendered consumption by 5. The results show that larger bandwidth settings glowingly increase the aggregate consumption, likely caused by an increase on cache memory contention when processing the aforementioned tasks.

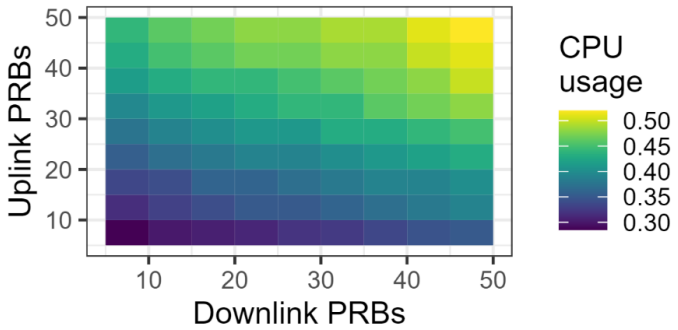


Figure 3-33 Average CPU consumption for a range of scheduling policies in a 10-MHz vBS

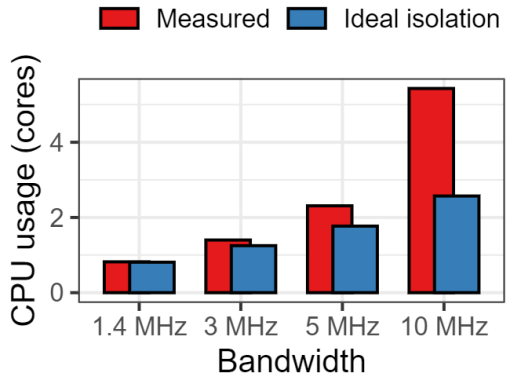
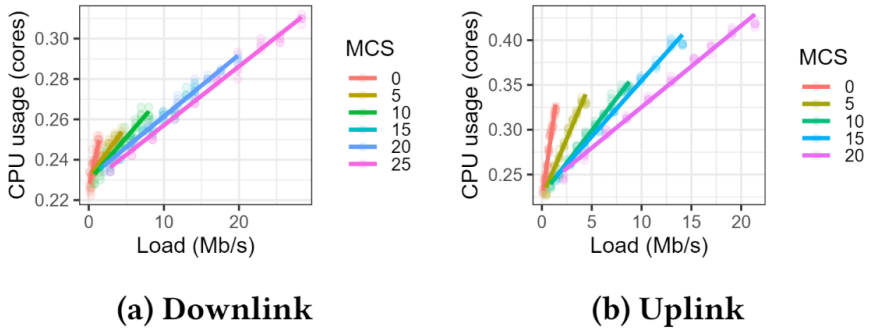


Figure 3-34 95th percentile of aggregated per-core usage of a vRAN with 5 vBSs for different bandwidth settings



(a) Downlink **(b) Uplink**

Figure 3-35 Mean CPU consumption for a range of network load settings and different MCS indexes

The computing consumption of a vBS also depends on the network load it must process. Figure 3-35 shows the mean CPU usage for different network load and MCS settings. These results confirm previous results where the CPU consumption grows linearly with the amount of data it processes, which requires a larger amount of PRBs to carry this data. Larger MCS indexes encode a larger amount of bits per PRB and hence it is expected that the CPU consumption grows in the same way.

3.4.6 Conclusions

This section presents a comprehensive analysis of the computing allocation in vRANs and their impact on energy consumption. We begin by investigating a scenario where multiple vBS are deployed on a shared computing infrastructure. In this context, we characterize the phenomenon known as the noisy neighbour problem. This problem shows an exponentially increasing energy overhead.

To understand the experimental analysis, we first provide background on the architecture of general-purpose computing platforms. We then detail the methodological approach used in our experiments. The, we explore and analyse the effects of CPU pinning and LLC allocation on computational resource utilization. We further examine the influence of various parameters such as traffic load, BS bandwidth, and the employed modulation and coding scheme.

Our findings reveal several intriguing trade-offs that present opportunities for the development of optimization algorithms. These algorithms could significantly improve the energy efficiency of RANs by dynamically allocating computing resources based on real-time network conditions. By mitigating the noisy neighbour problem and optimizing resource allocation, significant energy savings can be achieved within vRAN deployments. The development and evaluation of such algorithm will be reported in WP4.

4 Evolved BeGREEN architecture

This chapter outlines the evolution of the BeGREEN architecture based on the findings from WP2, WP3, and WP4. Compared to the previous version detailed in BeGREEN D2.1 [1], this update offers a more precise specification of the necessary components and interfaces, along with a clearer definition of their required functionalities and interrelationships. Figure 4-1 depicts the main architecture, whose main novelties will be further described in the next subsections of this chapter.

As shown in Figure 4-1, and considering O-RAN as the baseline architecture, several architectural extensions have been already integrated to the BeGREEN architecture (i.e., AI Engine, RIS, Relays and Edge), while others such as distributed MIMO and ISAC are still under study and will be incorporated in the next BeGREEN D2.3. Note that one of the objectives of BeGREEN is to integrate these architectural extensions under the umbrella of the Intelligence Plane, which consists of the non-RT and near-RT RICs, plus the AI Engine.

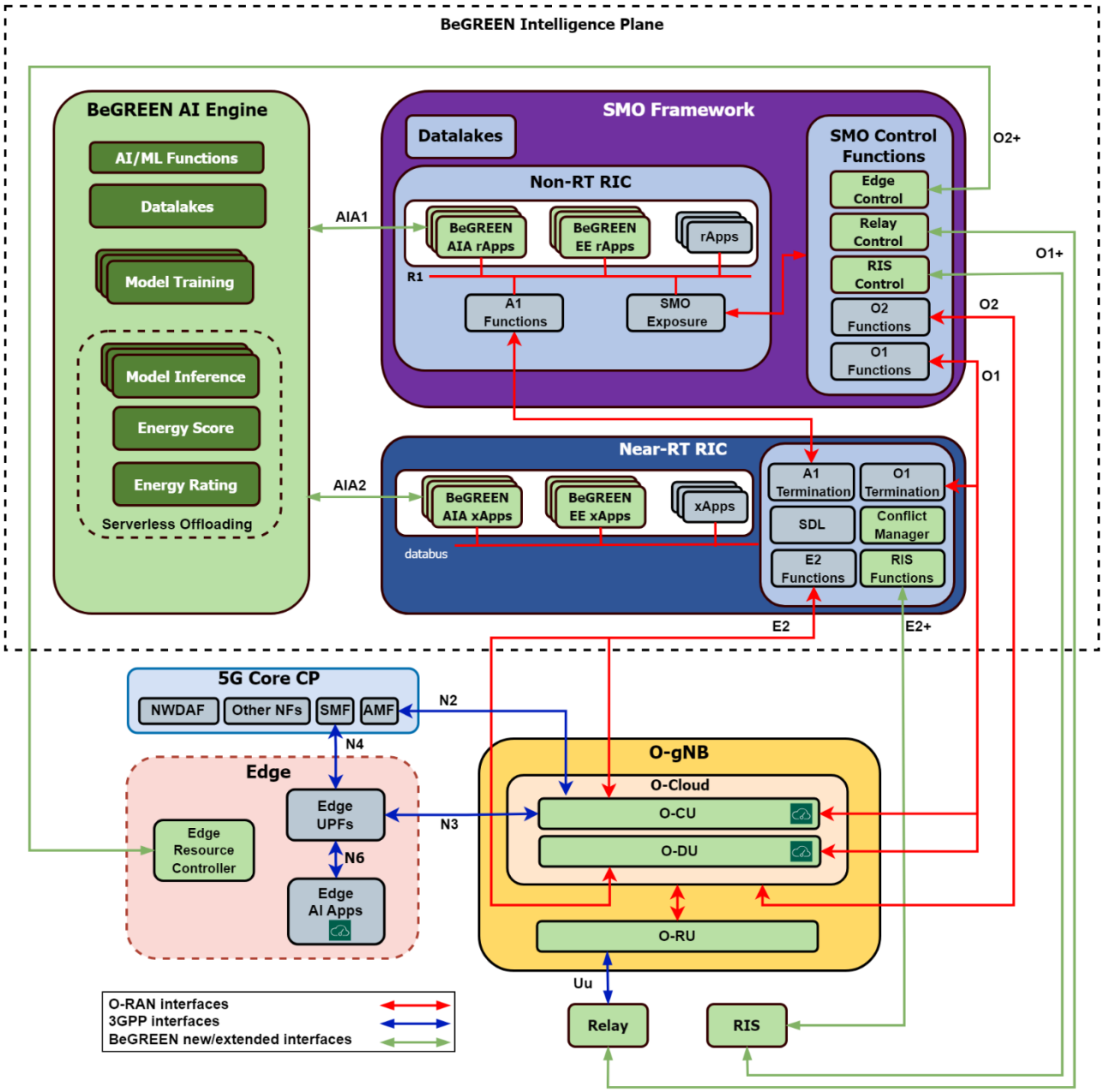


Figure 4-1 Evolved BeGREEN architecture

Therefore, the next subsections mainly discuss on the integration of these technologies within the Intelligence Plane, and, how this will impact actual O-RAN and/or 3GPP architecture by requiring of novel/extended components and/or interfaces.

4.1 Architecture extensions for AI/ML support

The support of AI-driven optimisations is key aspect to enhance decision-making by learning from historical data, adapting to evolving network dynamics, and making informed decisions to enhance network performance and energy efficiency. The last specifications of O-RAN include the support of AI/ML services and workflows in the RICs [41], which shall be exposed to the rApps and xApps, for instance to allow model management and training operations. However, multiple implementation options shall be supported: e.g., by the RICs themselves, by the Service Management and Orchestration (SMO) or even by external frameworks. For instance, the O-RAN Software Community (OSC) is working on an AI/ML framework implementation at the SMO level based on Kubeflow (data processing, training, and monitoring), which relays on Kserve adapters in the rApps to implement model inference [42].

The design of the BeGREEN Intelligence Plane adopts a model-based and loosely coupled approach, based on the introduction of the AI Engine component [43]. This AI/ML framework provides a serverless execution environment hosting the AI/ML models, offloading inference, and training services from the rApps/xApps. Each ML model implements its own data processing, training, monitoring and inference services in the AI Engine, which are then exposed by the AI Engine Assist (AIA) xApps/rApps (i.e., ML model producers) to any control rApp/xApp (i.e., ML model consumers). This approach allows ML developers to focus on model development and optimization, while RAN control experts can concentrate on control-loop decision logic. Additionally, it increases the reusability of ML models, which shall be exposed to any control rApp/xApp requiring their outputs. Finally, the serverless offloading of inference services allow to apply HW acceleration through GPUs or edge near-RT inference when required.

Further details on this extension and its implementation and validation can be found in BeGREEN WP4 deliverables. An initial implementation was demonstrated at the EuCNC 6G Summit 2024 conference, involving the AI Engine and the non-RT RIC components. It is expected that a TRL4 or TRL5 will be reached through its validation within the scope of WP5.

4.2 Architecture extensions for cell-free systems

A cell-free architecture is conceived as a user-centric architecture that aims to provide the best link quality at any time. To this end, distributed APs are pervasively deployed, and the best ones are selected to serve each of the user. Hence, a cell-free system requires coordination among the different APs, a joint processing, and dynamically adapting the selected of APs over time according to the position of the UEs and their needs. It is also worth mentioning that the joint processing is usually computed on a common server.

In the scope of O-RAN [44], O-RUs are the cell-free APs and the O-DUs compute the joint processing. A cell-free system makes use of the distributed APs to erase cell boundaries. However, in an O-RAN deployment, there are logical boundaries as one UE can be served by different O-RUs that are not connected to the same O-DU. We refer to this user as the edge user. Therefore, we need the O-DUs to cooperate and collaborate in executing the joint processing. This coordination is done by enabling an interface that connects different O-DUs.

When serving an edge user, one O-DU will be the main O-DU and will receive user-related data from O-RUs that are connected to the other O-DUs, and the latter will share this data through the Inter-O-DU interface as illustrated in Figure 4-2. Regarding the near-RT RIC, it can determine whether the UE is a regular or an edge user and assign the best RUs to this UE. This process needs to be updated over time as the user moves.

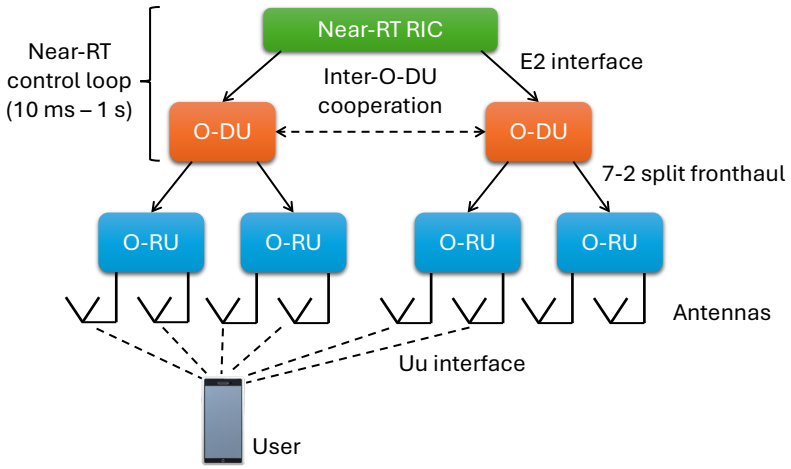


Figure 4-2 Architectural modification for cell-free system [44].

In conclusion, out of the box, O-RAN cannot fully support cell-free for all the cases as the interface that interconnects different O-DUs is not implemented in O-RAN. We illustrate this interface in the O-RAN scope in Figure 4-2. A low TRL of 2 is expected for this architecture extension. BeGREEN D2.3 is planned to discuss this extension with additional details.

4.3 Architecture extensions for ISAC

Though it has not been reported in BeGREEN DoW, the project is seeking means to integrate the ISAC work carried out in WP3 (see BeGREEN D3.1 [45] and BeGREEN D3.2 [46]) with the BeGREEN architecture shown in BeGREEN D4.1 [47]. If we recall back on the ISAC work defined in the DoW there were two main technologies used for ISAC purposes:

- Sub-6 ISAC, using Ettus N321 equipment.
- mmWave ISAC, using Xilinx RFSoC-based devices.

These developments do not rely on 3GPP devices, therefore their integration in the BeGREEN architecture is solely considered as producers of sensing information that can be used by the BeGREEN system.

The capabilities for a 3GPP/O-RAN network for extracting sensing information is dependent on the capabilities of the O-RAN RU for providing this information. The possibilities to process the received I/Q samples that carry the ISAC information are diverse.

When possible, and to save data rate over the interfaces towards the Core Network (CN), any feasible pre-processing of the raw measurement data (e.g., I/Q samples) should happen as early as possible. This pre-processing of the data can take place in the O-RU (if this element has the flexibility of pre-processing the data), alternatively in the O-DU or in the O-CU. Moreover, the pre-processing can take place in the Near-RT RIC, taking care of its AI/ML functionalities in xApps. Not all these possibilities are feasible with the current available technology and efforts need to be made to allow the possibility of processing the sensing data in all the above-mentioned elements. As an example, commercial RUs do not support processing of the sensing data and work needs to be carried out to integrate these capabilities into Software Defined Radio (SDR)-based RUs.

The work related to the definition of these concepts and to propose a lean implementation of ISAC in SDR-based systems will be put forward in BeGREEN D2.3, with the objective of extending the work in this direction

until the project end.

4.4 Architecture extensions for RIS

The proposed updates on the integration of RIS into the O-RAN enabled BeGREEN architecture are discussed here. The preliminary description provided in BeGREEN D2.1 (Section 4.2) [1] and BeGREEN D4.1 (Section 3.4.3) [47] considered that the RIS was controlled by both the gNB (through the F1-x interface) and the near-RT RIC (through the Ra interface). This F1-x interface is required for uses cases that aim to perform joint gNB and RIS beamforming. In such scenarios, the gNB needs to provide configurations to the RIS according to the corresponding beamforming configurations for a given user at a given time, which requires a very low latency that the near-RT RIC cannot provide. However, this approach has the problem of not having a direct interface between the non-RT RIC and the RIS, which may be required for non-RT management functions such as offline optimization, or logical split of the RIS.

In this sense, the RIS-aided use cases considered in BeGREEN do not have the need of such real-time control of the RIS. The BeGREEN RIS-aided use cases rather consider near-real time scales, which implies that it is more suitable to involve the near-RT RIC (and non-RT RIC) for controlling the RIS. Therefore, the actual integration of RIS into the BeGREEN architecture assumes that the RIS is both directly controlled by the near-RT RIC for near-RT operations such as RIS per-element configuration or load a precomputed configuration from a codebook, and the non-RT RIC for operations such as training or RIS logical splitting.

To integrate the RIS (i.e., the RIS Actuator which is the element that locally controls the antenna elements) with the non-RT RIC, we will use the O1 interface. This interface may be built on the NETCONF protocol or an equivalent one and employs YAML to model the necessary configuration and data collection commands for interactions. Key tasks anticipated for the non-RT RIC include managing the RIS logical split and optimizing/training the codebook, as these tasks are suited for a non-RT time scale. As depicted in Figure 4-1, we call O1+ to the standard O-RAN interface O1 plus the RIS-specific management models.

On the other hand, in order to integrate the RIS (i.e., the RIS Actuator) with the near-RT RIC, we will leverage the E2 interface. Two new service models will be used, the E2 Service Model Smart Surface Control (E2SM-SSC) to modify and initiate RIS control commands that may result in change of RIS control behaviour in a near-RT time scale, and the E2 Service Model Smart Surface Monitoring (E2SM-SSM), to monitor RIS state in a near-RT time scale. Since RIS are not E2 nodes, E2SM-SSC and E2SM-SSM service models will need to deprecate some E2SM common IEs that other standard, gNB-oriented service models have. For this reason, we refer to the E2 interface implementing these two service models as the E2+ interface, which need to be supported both in the RIS Actuator and in the near-RT RIC. More information about RIS-enabled O1 and E2 interfaces will be provided in future deliverable D4.2 [48].

Finally, the addition of a gNB-RIS interface (i.e., the F1-x interface), will be revisited in BeGREEN D2.3 and BeGREEN D4.3, where the final and consolidated version of the BeGREEN architecture will be provided.

4.5 Architecture extensions for Relays

The development of AI/ML-based functionalities for relay control leads to the improvement of decisions related to the deployment, activation and deactivation of the relays, etc., resulting in a better system performance and a reduction in energy consumption. As mentioned in BeGREEN D2.1 [1] and BeGREEN D4.1 [47], these relay functionalities cover different aspects. The coverage hole detection functionality, described in BeGREEN D4.1 [47], focuses on the identification of periods of time and geographical regions with high traffic demands and poor propagation conditions, that is, coverage holes. These coverage holes may be addressed in different ways. On the one hand, the fixed relay placement functionality determines the most adequate location for the placement of new relays and their initial configuration parameters to address the identified coverage holes. On the other hand, these coverage holes may also be addressed with the use of

RUEs (UEs with relaying capabilities that may act as a relay between the network and other neighbor UEs located in these coverage holes). For this purpose, it is necessary to properly identify UEs with some specific characteristics that may indicate that they can be good candidate RUEs. This is done by the candidate RUE identification functionality that will be described in BeGREEN D4.2 [48]. Finally, BeGREEN also considers a functionality that deals with the dynamic activation and deactivation of these relays/RUEs [47]. The aim of this functionality is to activate the relays or RUEs to serve UEs located in these coverage hole regions and deactivate these relays when they are not necessary, with the aim to minimising the energy consumption.

For the implementation of these previously mentioned relay control functionalities, several components are necessary in the BeGREEN architecture. On the one hand, as shown in Figure 4-1, a *Relay Control* entity is placed at the SMO and it is in charge of the interaction with the relays for the collection of the network measurements and the relay reconfigurations through the O1/O1+ interface. On the other hand, the previously mentioned relay control functionalities are sustained by different rApps in the non-RT RIC. The required AI/ML models are hosted in the AI Engine. Concerning the connection between the non-RT-RIC and the relays there are two possibilities. As stated in BeGREEN D2.1 [1], the relay can be either an Integrated Access and Backhaul (IAB) node or a RUE. In the case of the relay being an IAB-node, the Uu interface is used to connect the IAB-MT entity of the IAB-node with the RU of the gNB. In addition, the interfaces F1-C/F1-U are used to connect the IAB-DU entity of the relay with the CU-CP/CU-UP of the IAB donor. The physical realization of these interfaces is done on top of the Uu interface between the IAB-MT and the RU. In the case of the relay being as IAB-node the collection of measurements and relay reconfiguration commands can be transmitted between the relay control and the IAB-DU of the relay through the O1+ interface. In the case of the relay being a RUE, it is connected with the RU through the Uu interface. The collection of measurements and the relay reconfigurations are transmitted between the relay control and the RUE through the O1+ interface.

A more detailed description of these processes will be presented in BeGREEN WP4 deliverables. A TRL2 is expected for this architecture extension.

4.6 Architecture extensions for Edge

The management of Edge resources not dedicated to vRAN or O-Cloud is actually not being considered by O-RAN specification. However, in recent publications, O-RAN has studied collaboration and convergence across domains to enable cross-domain AI optimizations [49]. Indeed, the utilization of general-purpose servers for deploying virtualized RAN functions opens the door to the joint management of RAN, Core, and service resources.

To this end, as shown Figure 4-1, the BeGREEN Intelligence Plane includes an Edge Control function in the SMO conceived to dynamically control the resources of edge servers hosting 5G Core User Plane Functions (UPFs) and/or Edge AI services. This function is exposed to the rApps aiming at optimizing edge resources to enhance energy efficiency, also enabling the creation of control-loops focused on joint optimizations. Since this approach has several similarities with the O-cloud management performed by the SMO, an interface leveraging O2, denoted as O2+ in Figure 4-1, has been specified between the SMO and the edge server. Through this interface, the SMO shall be able to obtain the configuration of the edge, monitor its resources, and apply configurations related to energy saving modes, CPU frequency control, CPU allocation or GPU allocation.

This final design of this extension will be reported in BeGREEN WP4 deliverables, being related to the design and validation of two different energy efficiency optimization strategies: Traffic-aware CPU state management to enhance UPF Energy Efficiency and the joint orchestration of vRANs and Edge AI services. A low TRL2 or TRL3 is expected since focus will be on demonstrating the benefits of the optimization mechanisms.

5 Summary and Conclusions

BeGREEN aims to provide innovative solutions on the RAN for energy-saving purposes. In this document, BeGREEN D2.2, four different mechanisms are discussed and demonstrated through an extensive evaluation that each mechanism saves energy in the RAN. Chapter 2 identifies the KPIs, the KVs, and KVIs for the scope of BeGREEN, while Chapter 3 presents the four mechanisms and extensively evaluates them. Firstly, an empirical study is presented which confirms that a D-MIMO system provides superior performance using less transmit gain compared with a traditional MIMO system. Secondly, the intended work on the optimization of the resource allocation given a certain sensing data is introduced, which is to be carried out and evaluated in BeGREEN D2.3. Thirdly, the impact of the deployment of relay nodes are studied. The evaluations demonstrate that equipping a RAN with relay nodes can significantly improve its energy efficiency by strategically placing fixed or even mobile relays, the network can optimize energy consumption across its various components. Further, a comprehensive analysis of the computing allocation in vRANs and their impact on energy consumption are discussed, the related findings reveal several intriguing trade-offs that present opportunities for the development of optimization algorithms that could significantly improve the energy efficiency of RANs. Finally, Chapter 4 proposes the O-RAN architecture extensions that are needed to deploy the four mechanisms. This includes architectural extensions for integrating AI/ML support, D-MIMO, ISAC, RIS, relays, and the Edge domain. BeGREEN D2.2 serves as input to BeGREEN D2.3 as well as to BeGREEN WP3, WP4, and WP5.

6 References

[1] BeGREEN deliverable D2.1, "BeGREEN Reference Architecture", July 2023, https://www.sns-begreen.com/deliverables?id=971369

[2] 5G-PPP Key Performance Indicators (KPIs), https://5g-ppp.eu/kpis/

[3] 5G-PPP TMV White Paper, "Test Measurement and KPIs Target Values", June 2022, https://zenodo.org/records/6577506

[4] 5G-PPP TMV White Paper, "Beyond 5G/6G KPI Measurement", June 2023, https://5g-ppp.eu/wp-content/uploads/2023/06/white_paper_b5g-6g_kpi_measurement_FINAL.pdf

[5] Hexa-X deliverable D1.1 "6G Vision, use cases and key societal values", 2021, https://hexa-x.eu/wp-content/uploads/2021/02/Hexa-X_D1.1.pdf

[6] G. Wikström, "What societal values will 6G address?". Zenodo, May 17, 2022, 6G IA SNS. URL: https://doi.org/10.5281/zenodo.6557534

[7] G. Wikström, N. Bledow, M. Matinmikko-Blue, H. Breuer, C. Costa, G, Darzanos, A. Gavras, T. Hossfeld, I. Mesogiti, K. Petersen, P. Porambage, R.-A. Stoica, S. Wunderer, "Key value indicators: A framework for values-driven next-generation ICT solutions", Telecommunications Policy, 2024, 102778, ISSN 0308-5961, https://doi.org/10.1016/j.telpol.2024.102778.

[8] 3GPP 22.x series are dedicated to Use Cases and their requirements. https://www.3gpp.org/dynareport?code=22-series.htm

[9] NGMN, "6G Use Cases and Analysis", February 22, 2022, url: https://www.ngmn.org/wp-content/uploads/NGMN-6G-Use-Cases-and-Analysis.pdf

[10] ETSI GR THz 001, "TeraHertz modelling (THz); Identification of use cases for THz communication systems", January 2024, https://www.etsi.org/deliver/etsi_gr/THz/001_099/001/01.01.01_60/gr_THz001v010101p.pdf

[11] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson and T. L. Marzetta, "Cell-Free Massive MIMO Versus Small Cells," in IEEE Transactions on Wireless Communications, vol. 16, no. 3, pp. 1834-1850, March 2017, doi: 10.1109/TWC.2017.2655515.

[12] Da-Shan Shiu, G. J. Foschini, M. J. Gans and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," in IEEE Transactions on Communications, vol. 48, no. 3, pp. 502-513, March 2000, doi: 10.1109/26.837052.

[13] H. Yang and T. L. Marzetta, "Energy Efficiency of Massive MIMO: Cell-Free vs. Cellular," 2018 IEEE 87th Vehicular Technology Conference (VTC Spring), Porto, Portugal, 2018, pp. 1-5, doi: 10.1109/VTCSpring.2018.8417645.

[14] H. Q. Ngo, L. -N. Tran, T. Q. Duong, M. Matthaiou and E. G. Larsson, "On the Total Energy Efficiency of Cell-Free Massive MIMO," in IEEE Transactions on Green Communications and Networking, vol. 2, no. 1, pp. 25-39, March 2018, doi: 10.1109/TGCN.2017.2770215.

[15] Demir, Ö. T., Méndez-Monsanto, L., Bastianello, N., Fitzgerald, E., & Callebaut, G. (2024). Energy Reduction in Cell-Free Massive MIMO through Fine-Grained Resource Management. *arXiv preprint arXiv:2405.07013*.

[16] Hoydis, J., Aoudia, F. A., Cammerer, S., Nimier-David, M., Binder, N., Marcus, G., & Keller, A. (2023). Sionna RT: Differentiable ray tracing for radio propagation modeling. *arXiv preprint arXiv:2303.11103*.

[17] Hoydis, J., Aoudia, F. A., Cammerer, S., Euchner, F., Nimier-David, M., Brink, S. T., & Keller, A. (2023). Learning radio environments by differentiable ray tracing. *arXiv preprint arXiv:2311.18558*.

[18] Hexa-X-II deliverable D1.2, "6G Use Cases and Requirements", https://hexa-x-ii.eu/wp-content/uploads/2024/01/Hexa-X-II_D1.2.pdf

[19] Green Future Networks Metering for Sustainable Networks, NGMN, https://www.ngmn.org/wp-content/uploads/220125-GFN-Metering-White-Paper-v1.0.pdf

[20] EARTH project, "D2.3 - Energy efficiency analysis of the reference systems, areas of improvements and target breakdown", January, 2012.

[21] R. Fantini, D. Sabella, M. Caretti, "An E3F based assessment of energy efficiency of Relay Nodes in LTE-Advanced networks", PIMRC, 2011.

[22] 3GPP TS 38.306 v17.1.0, "NR; User Equipment (UE) radio access capabilities (Release 17)", June, 2022.

[23] "5G TDD Synchronisation. Guidelines and Recommendations for the Coexistence of TDD Networks in the 3.5 GHz Range", GSMA, White Paper, Apr. 2020, Accessed: Apr. 6, 2023. Available: https://www.gsma.com/spectrum/wp-content/uploads/2020/04/3.5-GHz-5G-TDD-Synchronisation.pdfAvailable: https://www.gsma.com/spectrum/wp-content/uploads/2020/04/3.5-GHz-5G-TDD-Synchronisation.pdf

[24] E. Björnson, M. Kountouris and M. Debbah, "Massive MIMO and small cells: Improving energy efficiency by optimal soft-cell coordination," ICT 2013, Casablanca, Morocco, 2013, pp. 1-5, doi: 10.1109/ICTEL.2013.6632074.

[25] G. Auer et al., "How much energy is needed to run a wireless network?", in IEEE Wireless Communications, vol. 18, no. 5, pp. 40-49, October 2011, doi: 10.1109/MWC.2011.6056691.

[26] T. L. Marzetta, E. G. Larsson, H. Yang, H. Q. Guo, Fundamentals of Massive MIMO, Cambridge University Press, 2016

[27] E. Björnson, "A Basic Way to Quantify the Massive MIMO Gain", March, 2018. Online: https://ma-mimo.ellintech.se/2018/03/23/a-basic-way-to-quantify-the-massive-mimo-gain/Online: https://ma-mimo.ellintech.se/2018/03/23/a-basic-way-to-quantify-the-massive-mimo-gain/

[28] Oliver Arnold, Fred Richter, Gerhard Fettweis, and Oliver Blume. 2010. Power consumption modeling of different base station types in heterogeneous cellular networks. In 2010 Future Network & Mobile Summit. IEEE, 1–8.

[29] Gunther Auer, Vito Giannini, Claude Desset, Istvan Godor, Per Skillermark, Magnus Olsson, Muhammad Ali Imran, Dario Sabella, Manuel J Gonzalez, Oliver Blume, et al. 2011. How much energy is needed to run a wireless network? IEEE wireless communications 18, 5 (2011), 40–49.

[30] Josip Lorincz, Tonko Garma, and Goran Petrovic. 2012. Measurements and modelling of base station power consumption under real traffic loads. Sensors 12, 4 (2012), 4281–4310.

[31] Jeffrey G Andrews, Stefano Buzzi, Wan Choi, Stephen V Hanly, Angel Lozano, Anthony CK Soong, and Jianzhong Charlie Zhang. 2014. What will 5G be? IEEE Journal on selected areas in communications 32, 6 (2014), 1065–1082.

[32] Jose A. Ayala-Romero et al. 2021. Bayesian Online Learning for Energy-Aware Resource Orchestration in Virtualized RANs. In IEEE INFOCOM 2021 - IEEE Conference on Computer Communications. 1–10.

[33] Jose A Ayala-Romero, Ihtisham Khalid, Andres Garcia-Saavedra, Xavier Costa-Perez, and George

Iosifidis. 2021. Experimental Evaluation of Power Consumption in Virtualized Base Stations. In 2021 IEEE International Conference on Communications (ICC).

[34] Garcia-Aviles, Gines, et al. "Nuberu: Reliable RAN virtualization in shared platforms." Proceedings of the 27th Annual International Conference on Mobile Computing and Networking. 2021.

[35] Kumar, Praveen, et al. "PicNIC: predictable virtualized NIC." Proceedings of the ACM Special Interest Group on Data Communication. 2019. 351-366.

[36] Park, Jinsu, Seongbeom Park, and Woongki Baek. "CoPart: Coordinated partitioning of last-level cache and memory bandwidth for fairness-aware workload consolidation on commodity servers." Proceedings of the Fourteenth EuroSys Conference 2019. 2019.

[37] Jacob, Bruce, David Wang, and Spencer Ng. Memory systems: cache, DRAM, disk. Morgan Kaufmann, 2010.

[38] Drepper, Ulrich. "What every programmer should know about memory." Red Hat, Inc 11.2007 (2007): 2007.

[39] NGMN 5G White Paper, https://www.ngmn.org/wp-content/uploads/NGMN_5G_White_Paper_V1_0.pdf

[40] A. Wiesel, Y. C. Eldar and S. Shamai, "Zero-Forcing Precoding and Generalized Inverses," in IEEE Transactions on Signal Processing, vol. 56, no. 9, pp. 4409-4418, Sept. 2008, doi: 10.1109/TSP.2008.924638.

[41] O-RAN Alliance, “O-RAN Non-RT RIC Architecture 5.0”, WG2, February 2024

[42] O-RAN Software Community Wiki, “https://wiki.o-ran-sc.org/display/AIMLFEW/Design", accessed June 2024

[43] M. Catalan-Cid, J.Pueyo, J. Sanchez-Gonzalez, J. Gutierrez, M. Ghoraishi, “BeGREEN Intelligent Plane for AI-driven Energy Efficient O-RAN management”, EuCNC/6G Summit, Antwerp, June 2024

[44] V. Ranjbar, A. Girycki, M. A. Rahman, S. Pollin, M. Moonen and E. Vinogradov, "Cell-Free mMIMO Support in the O-RAN Architecture: A PHY Layer Perspective for 5G and Beyond Networks," in *IEEE Communications Standards Magazine*, vol. 6, no. 1, pp. 28-34, March 2022, doi: 10.1109/MCOMSTD.0001.2100067.

[45] BeGREEN deliverable D3.1, “State-of-the-Art on PHY Mechanisms Energy Consumption and Specification of Efficiency Enhancement Solutions”, January 2024

[46] BeGREEN deliverable D3.2, “Initial developments and evaluation of the proposed enhancements and optimization strategies”, work in progress.

[47] BeGREEN deliverable D4.1, “State-of-the-art review and initial definition of BeGREEN O-RAN intelligent plane, and AI/ML algorithms for NFV user-plane and edge service control energy efficiency optimization”, December 2023

[48] BeGREEN, D4.2, “Initial evaluation of BeGREEN O-RAN intelligent plane, and AI/ML algorithms for NFV user-plane and edge service control energy efficiency optimization”, work in progress.

[49] O-RAN Alliance, “Research Report on Native and Cross-domain AI: State of the art and future outlook”, nGRG group, September 2023.