Project no.:	101095738		
Project full title:	6G Short range extreme communication IN Entities		
Project Acronym:	6G-SHINE		
Project start date:	01/03/2023	Duration	30 months

D3.3 – RECOMMENDED PHY AND MAC ENABLERS FOR IN-X SUBNETWORKS

Due date	30/6/2025	Delivery date	30/06/2025
Work package	WP3		
Responsible Author(s)	Baldomero Coll-Perales (UMH)		
Contributor(s)	Syed Morsleen Riaz (UMH), Baldomero Coll Perales (UMH), Javier Gozalvez (UMH), Davide Dardari (CNIT), Giulia Torcolacci (CNIT), Alessandro Mirri (CNIT), Enrico Paolini (CNIT), Wei Liu (IMEC), Meng Li (IMEC), Thijs Havinga (IMEC), Shin Horng Wong (SONY), Yassin Awad (SONY), Fredrik Rusek (SONY), Hamid Reza Hashempour (AAU), Gilberto Berardinelli (AAU), Christos Tsakos (COGN), Jürgen Hupp (FHG), Thomas Haaning Jacobsen (Nokia), Spilios Giannoulis (IMEC), Irfan Jabandzic (IMEC), Filipe Conceicao (IDE)		
Version	V1.0		
Reviewer(s)	Christian Hoffmann (APPLE), Pedro Maia de Sant Ana (BOSCH)		
Dissemination level	Public		

VERSION AND AMENDMENT HISTORY

Version	Date (MM/DD/YYYY)	Created/Amended by	Changes
0.1.	07/01/2025	B. Coll-Perales	ToC started/finalization
0.2	03/03/2025	B. Coll-Perales	Initial contributions from partners
0.3	30/04/2025	B. Coll-Perales	Editor revision and feedback
0.4	15/04/2025	B. Coll-Perales	Editor review completed
0.5	21/05/2025	B. Coll-Perales	Version ready for internal review
0.6	06/06/2025	Christian Hofmann, Pedro M. de Sant Ana	Feedback from the internal reviewers
0.7	20/06/2025	B. Coll-Perales	Updates requested by the reviewers added
0.9	27/06/2025	Berit H. Christensen	Final proofreading and layout check
1.0	30/06/2025	B. Coll-Perales	Submitted version

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS

Acronym	Description
ACRDA	Asynchronous Contention Resolution Diversity ALOHA
ADAS	Advanced Driver-Assistance Systems
ADC	Analog Digital Converter
ALLR	Approximated Log Likelihood Ratio
AP	Access Point
AWGN	Additive White Gaussian Noise
BLE	Bluetooth Low Energy
BLER	Block Error Rate
CDF	Cumulative Distribution Function
CAM	Connected and Automated Mobility
CG	Configured Grant
CNN	Convolutional Neural Network
COT	Channel-Occupancy Time
CPE	Customer Premises Equipment
CRA	Coded Random Access
CRC	Cyclic Redundancy Check
CS	Configured Scheduling
CSI	Channel State Information
CSMA/CD	Carrier Sense Multiple Access with Collison Avoidance
CTI	Cross Technology Interference
DAC	Digital-to-Analog Conversion
DCI	Downlink Control Information
DRL	Deep Reinforcement Learning
DSA	Dynamic Scattering Arrays
DL	Downlink
E/E	Electrical/Electronic
FR#	Frequency Range #
ECU	Electronic Control Unit
EM	ElectroMagnetic
ESIT	Electromagnetic Signal and Information Theory
ESP	Electromagnetic Signal Processing
ESPAs	Electrically Steerable Passive Array radiators
FA-CSA	Frame-Asynchronous Coded Slotted ALOHA
FBE	Frame-Based Equipment
FFP	Fixed Frame Period
HARQ	Hybrid-Automatic Repeat Request
HC	High-Capabilities device
HMD	Head-Mounted Displays
HMI	Human-Machine Interface
HPCU	High-Performance Computer Unit
IBE	In-Band Emissions
InF	Indoor Factory

IPU	Image Processing Unit
KPI	Key Performance Indicators
KVI	Key Value Indicators
LAA	Link Adaptation Aware
LAD	Latency Aware Deconfliction
LBE	Load-Based Equipment
LBT	Listen-Before Talk
LC	Low Capability device
LDPC	Low Density Parity Check
LMMSE	Linear Minimum Mean-Squared-Error
LO	Local Oscillator
LOS	Line-Of-Sight
LR-WPAN	Low Rate – Wireless Personal Area Network
LSTM	Long Short-Term Memory
MAC	Medium Access Control
MAP	Maximum A-Posteriori
MCG	Multiple Configured Grant
MCS	Modulation and Coding Scheme
MDS	Maximum Distance Separable
MIMO	Multiple Input Multiple Output
MINLP	Mixed-Integer Nonlinear Programming
MISO	Multiple Input Single Output
mMTC	massive Machine-Type Communication
MPR	Multi-Packet Reception
MSE	Mean Squared Error
NCC	Network Coded Cooperation
NLOS	Non-Line of Sight
NR	New Radio
OFDMA	Orthogonal Frequency Division Multiple Access
ONNX	Open Neural Network Exchange
OTA	Over-The-Air
PCG	Predictive Configured Grant
PDG	Predictive Dynamic Grant
PDSCH	Physical Downlink Shared Channel
PHY	Physical
PLC	Programmable Logic Controller
PLR	Packet Loss Rate
PUCCH	Physical Uplink Control Channel
QoS	Quality of Service
RAN	Radio Access Networks
RB	Resource Block
ReLU	Rectified Linear Unit
RNTI	Radio Network Temporary Identifier
ROS	Robot Operative System

RU	Resource Unit
RF	Radio Frequency
RIS	Reconfigurable Intelligent Surfaces
RRM	Radio Resource Management
SCM	Self-Conjugating Metasurfaces
SCS	Sub Carrier Spacing
SDP	SemiDefinite Programming
SIC	Successive Interference Cancellation
SINR	Signal-to-Interference Noise Ratio
SIM	Stacked Intelligent Surfaces
SNE	Sub-Network Element
SOA	State-Of-the-Art
SPCA	Sequential Parametric Convex Approximation
SPS	Semi-Persistent Scheduling
SR	Scheduling Request
SVD	Singular Value Decomposition
SW	Sliding Window
TB	Transport Block
TC#	Technology Component number #
TRL	Technology Readiness Level
TSN	Time-Sensitive Networks
UE	User Equipment
UL	Uplink
URLLC	Ultra-Reliable Low-Latency communications
VF	Virtual Frame
XR	eXtended Reality

EXECUTIVE SUMMARY

This deliverable presents the 6G-SHINE project's recommended physical (PHY) and medium access control (MAC) solutions for in-X subnetworks. Building upon the preliminary insights provided in Deliverable D3.1, the document outlines refined and validated technology enablers tailored to the specific characteristics of in-X subnetworks, such as short-range connectivity, low-complexity devices, and static or predictable mobility patterns.

The design of these PHY and MAC enablers has been strongly refined by inputs from other 6G-SHINE's WP2, which contributed detailed use case descriptions, relevant KPIs/KVIs, traffic and propagation models, and architectural design principles. These inputs ensured that the solutions are grounded in realistic subnetwork scenarios and performance expectations.

Rather than aiming for a full-stack PHY/MAC design, 6G-SHINE focuses on identifying and validating individual technology components that demonstrate strong potential for enabling key in-X subnetwork functionalities. PHY-layer solutions include metasurfaces, beamforming and jamming-resilient and sub-THz transceivers. MAC-based solutions focus on proximity grouping, configured grant and coded random access for low latency, flexible duplexing for the multiplexing of services with diverse requirements, and predictive resource allocation to meet the stringent requirements of in-X subnetworks. The MAC solutions also cover unlicensed bands. These PHY and MAC solutions have been selected and refined based on their relevance to specific use case requirements. In addition, their novelty and performance improvements over the current state of the art give them strong potential for standardization impact within bodies such as 3GPP and ETSI.

These developments position the 6G-SHINE PHY and MAC enablers as key building blocks for the development of in-X subnetworks as part of the 6G network of networks [1].

1 INTRODUCTION

This deliverable presents 6G-SHINE's recommended physical (PHY) and medium access control (MAC) solutions for in-X subnetworks. It builds on the foundation established in Deliverable D3.1 [2], which reported preliminary results on PHY and MAC enablers for in-X subnetworks.

The recommended PHY and MAC solutions are based on the initial hypotheses regarding the specific requirements and characteristics of in-X subnetworks —such as short-range wireless connectivity, low-capability subnetwork elements, and static or predictable mobility patterns. More importantly, these solutions have been refined and validated through insights and feedback obtained from other 6G-SHINE work packages. A key contribution to the design of the PHY and MAC enablers has come from WP2, which provided (readers are recommended to see [3] for details):

- Detailed descriptions of in-X use cases
- Associated KPIs/KVIs
- In-X subnetworks traffic and radio propagation models
- Architectural building blocks for subnetwork design

These inputs have been instrumental in grounding the design of PHY and MAC solutions in a solid understanding of in-X subnetwork operations. Wherever applicable, the design process of the PHY and MAC solutions for in-X subnetworks has leveraged this information to ensure that the proposed enablers are aligned with realistic conditions and performance expectations.

Nevertheless, we would like to emphasize that delivering a full and cohesive PHY and MAC design for future in-X subnetworks is not the goal of 6G-SHINE, as such a task would exceed the scope and resources of the project. Rather, our aim is to explore and validate individual PHY and MAC technology components that hold strong potential as key enablers. The solutions that have been investigated in 6G-SHINE were selected [2] and have since been refined based on their relevance for contributing to address individual or group of requirements of specific in-X subnetwork use cases.

The preliminary version of this deliverable, i.e., D3.1 [2] already provided the analysis of the Key Value Indicators that are impacted by the 6G-SHINE PHY and MAC in-X enablers. Readers are referred to D3.1 for detailed information which is summarised as:

- Improved <u>environmental sustainability</u> due to achieved lower energy consumption on the devices thanks to 1) high-energy efficient sensors; 2) optimized transceiver architectures; 3) highly efficient transmissions; .4) low complexity designs; 5) extreme reliability and low latency enabling the replacement of wired connections.
- Improved <u>economic sustainability</u> thanks to 1) jamming-resilient operation avoiding network outages; 2) reduced operational delays; 3) capability of scheduling device with very diverse requirements by a same node, without additional infrastructure; 4) making wireless as deterministic as wired and allowing flexible subnetwork designs; 5) enabling unlicensed spectrum usage with guaranteed service levels (i.e. saving licensed spectrum fees).
- <u>Improved social sustainability</u> 1) by facilitating full automation of systems.

Finally, the PHY and MAC enablers for in-X subnetwork presented in this deliverable retain the potential for standardization impact reviewed in D3.1. This includes 3GPP RAN 1 and RAN 2 (within different releases depending on the required enablers), ETSI THz group and ETSI ISM. A much broader analysis of

the potential and current standardization impact and technology transfer of 6G-SHINE will be provided as part of deliverable D6.3.

The rest of this deliverable is structured as follows:

- Section 2 summarizes what are the main outcomes of the PHY and MAC solutions for subnetworks derived in 6G-SHINE. This summary is done identifying what are the primary 6G-SHINE use-cases that are targeted by the PHY and MAC solutions, and the extent to which the solutions achieve the use cases' key performance indicators.

- Section 3 details the PHY enablers for subnetworks recommended by 6G-SHINE. This section introduces a set of PHY-layer enhancements aimed at improving performance and reducing complexity in in-X subnetworks, addressing key requirements such as high throughput, low latency, and interference resilience. The proposed solutions include innovations in antenna design, metasurface-based beamforming, sub-THz transceiver architectures, and jamming-resilient physical layer techniques.

- Section 4 presents the final outcomes on MAC and service multiplexing mechanisms tailored for in-X subnetworks, focusing on meeting stringent latency, efficiency, data rate, deterministic and scalability requirements. Proposed enhancements include proximity-based retransmission strategies, improvements to configured grant access, and coded random-access schemes. The proposals also show the benefits for subnetworks of flexible duplexing in low-power subnetworks to support diverse service types while mitigating interference. Predictive resource allocation, AI/ML-based mobility awareness, multi-link reliability strategies, and novel channel access techniques for unlicensed spectrum are also proposed to further optimize performance in dynamic and constrained subnetwork environments.

- Finally, Section 5 concludes this deliverable with a discussion on the final recommendations for PHY and MAC enablers for in- X subnetworks.

2   MAIN OUTCOMES OF THE PHY AND MAC ENABLERS FOR SUBNETWORKS

This deliverable aims to show how the PHY and MAC solutions developed in 6G-SHINE fulfil Objectives 3 and 4 of the project:

Objective 3: Design new physical layer (PHY) enablers for scalable requirements in terms of latency, reliability or throughput, tailored to devices with constrained computational capabilities by leveraging the opportunities offered by short range subnetworks.
Objective 4: Develop new effective medium access control (MAC) solutions for efficient multiplexing of diverse traffic types in a subnetwork, including deterministic traffic.

The main outcomes of the PHY and MAC technology components for subnetworks developed in 6G-SHINE have been demonstrated through the following use cases. It should be noted that these are primary examples illustrating the advantages of the proposed technologies; however, the results can often be generalized to other use cases, both within the same subnetwork category and across different ones. Figure 1 presents the primary use cases addressed by the PHY and MAC technology components developed in 6G-SHINE. This section aims to highlight the main results achieved by these enablers and their relevance to the key performance indicators (KPIs) defined for 6G-SHINE use cases. It is important to note that fully meeting the requirements of a given use case may necessitate integrating multiple methods developed in 6G-SHINE. Nevertheless, due to the low Technology Readiness Level (TRL) focus of the project, the methods are presented and evaluated individually. Further details on each of the methods introduced in this section are provided in Section 3 (PHY) and Section 4 (MAC).
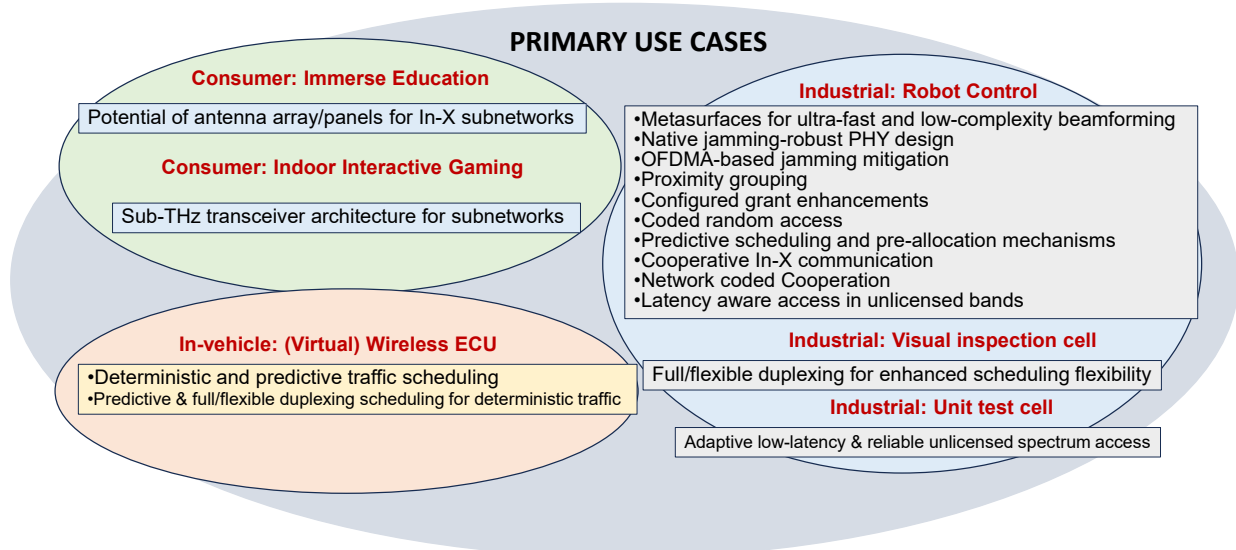
FIGURE 1 PRIMARY USE CASES FOR THE PHY AND MAC TECHNOLOGY COMPONENTS DESIGNED IN 6G-SHINE.

2.1   Immersive Education (Consumer Category)

The Immersive Education use case (D2.2 [3]) focuses on enhancing traditional classroom experiences by leveraging eXtended Reality (XR) technologies, diverse media content, and interactive tools to improve student engagement and knowledge retention. It supports varied learning styles by offering rich stimuli and fosters inclusion of both in-person and remote participants through telepresence and virtual interaction. As shown in Figure 2, the concept aims to create intuitive and engaging learning experiences that go beyond the limitations of a physical classroom. While primarily designed for classroom settings, this use case also extends to field excursions and mobile learning environments.

This use case is primarily envisioned for indoor deployments, where a 6G parent network provides connectivity, internet access, and potentially edge or cloud-based computation resources. Within the subnetwork, devices with higher computational capabilities (i.e., HC) - such as laptops, PCs, or advanced smartphones - can support immersive educational activities by contributing additional communication and processing power. For outdoor scenarios lacking infrastructure, the system should be capable of autonomous operation using locally available content and compute resources. If 6G connectivity is intermittently available, it can supplement local resources. In such cases, it may be necessary to reduce the complexity or visual quality of the experience to align with the available network and computational capacity.
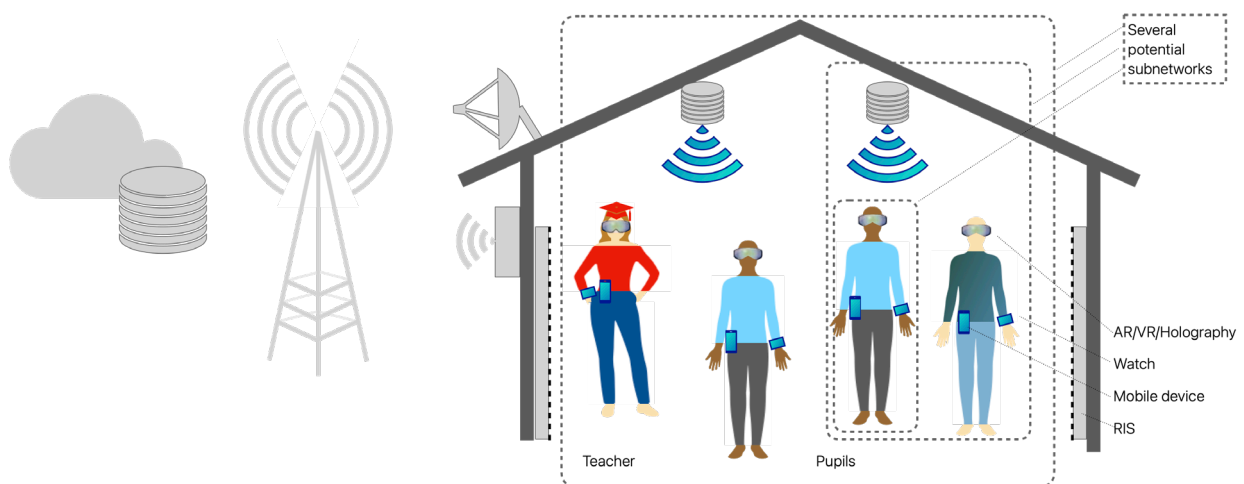
FIGURE 2 ILLUSTRATION OF THE IMMERSIVE EDUCATION USE CASE (D2.2 [3])

Main KPIs of this use case relates to the support of varying latencies, in the range from 10s of milliseconds down to few milliseconds, while high reliability (up to 99.999%) and varying throughput levels from few Mbps to Gbps are achieved. In addition, this use case targets the support of relatively high number of active devices ~50 devices. We refer to deliverable D2.2 for a thorough description of the use case and its requirements.

Antenna array/panels have been identified as a key technology enabler for supporting the stringent requirements of this use case in short-range subnetworks. The proposed reciprocity-based beam alignment solution has shown to be capable of increasing significantly the throughput compared to state-of-the-art schemes at no signalling overhead cost. This has been demonstrated in an indoor classroom scenario, for different antenna array/panel configurations and considering the size and capabilities constraints of low-complex devices such as those that can be present in the immersive education subnetwork: smartphone or customer premises equipment, extended reality glasses, and sensors and actuators. A summary of the benefits of antenna array/panels for in-x subnetworks is shown below, while a detailed description is presented in Section 3.

Antenna array/panels for In-X subnetworks
It is found that the proposed reciprocity-based beam alignment procedure can increase the throughput in the order of +50% to +100% compared to a single wide-beam transmit antenna configuration. The proposed beam alignment procedure boosts average system-wide spectral efficiency from 37Mbps (0.19bps/Hz) to 65Mbps (0.33bps/Hz) (+75%) by only adding an antenna

panel of 4 elements at the HC, and additionally to 81Mbps (0.4bps/Hz) (+119%) when considering antenna panels of 4 elements at the SNE as well. With an even larger panel at the HC (16 elements) this increases to 130Mbps (0.65bps/Hz) (+251%). The best performing beam alignment scheme representing a 5G NR baseline scheme is denoted as a narrow-beam-based alignment procedure can further generates an additional throughput on average between +19% and +30% on top of the proposed reciprocity-based beam alignment procedure, but that comes at the cost of 5x higher signalling overhead, which makes the reciprocity based beam alignment procedure a good compromise between beam forming gains with low overhead.

2.2 Indoor interactive gaming (Consumer Category)

This use case focuses on XR interactive gaming within a controlled indoor environment equipped with specialized infrastructure to enable immersive gameplay. Such environments typically include strategically placed sensors and actuators, a high-capability (HC) node functioning similarly to an access point, and a high-end edge server for local processing. Players are equipped with Virtual Reality (VR) headsets and body-attached sensor/actuator devices.

The tracking of user movements can be done using either outside-in (external sensors tracking the user) or inside-out (headset-embedded sensors). Given the controlled setting, outside-in tracking is more suitable, offering precise user pose and orientation data.

These static sensors communicate with the HC node in a local subnetwork. The XR game content, preloaded from a remote application cloud, is processed on the edge server, which generates the XR scenes. This split-rendering reduces the computational load on the VR headset, although headsets with high processing power may render scenes locally at the cost of higher energy consumption.

To enhance immersion, the edge server also controls actuators on the players (e.g., for vibration or heat), synchronized with the XR content. Additionally, synchronization between multiple players ensures a shared, coherent XR experience. This tightly integrated setup enables high-performance, low-latency XR gaming within a confined, sensor-rich indoor space.
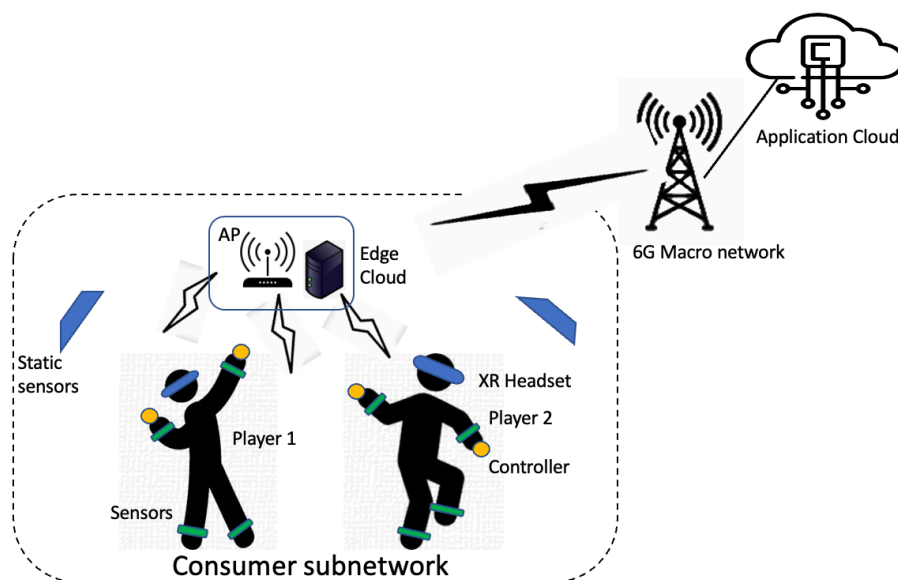
FIGURE 3 ILLUSTRATION OF THE INDOOR INTERACTIVE GAMING USE CASE (D2.2 [3])

This use case's KPIs are defined considering the rich multimedia traffic generated in the scenario, including video, audio and sensors and actuators that contribute to create the interactive gaming experience. This represents then a mix-traffic scenario demanding throughput levels that range from low-medium data rates for traffic generated by sensors/actuators and audio to very hight data rates, 30–150 Mbps for 8K video up to 6.37 Gbps for ultimate VR with 24K quality. At the same time, smooth VR experience also requires demanding latency values, which are below < 20 ms to prevent motion sickness, which define an ideal target for future networks of 1 ms over-the-air latency and tight synchronization required for multi-stream data. This use case also requires support for up to ~50 devices connected per subnetwork. We refer to deliverable D2.2 for a thorough description of the use case and its requirements.

Like the previous use case described in Section 2.1, PHY enhancements are key for supporting the needs of this use case. As shown below and described in detail in Section 3, we have considered **sub-THz transceiver architectures** as key enablers for achieving the low latency and high data rate demands of interactive gaming, while maintaining low power consumption and complexity of devices.

Sub-THz transceiver architecture for subnetworks
We have derived models for characterizing each component of the sub-THz transceiver separately depending on the underlying technology and their interaction with the overall transceiver architecture as well as the signal processing and algorithms to potentially develop robust and low complexity PHY. In the indoor interactive gaming subnetwork use case under study, the sub-THz wireless communication is used as the wireless link. The sub-THz communication has shown its capabilities to support this use case by providing wider bandwidth and date rate (4K and 8K video resolution with different codec rates), power consumption (< 1W for HC (HMD)), ~x2 lower complexity at device side, and high reliability (BLER < 10^{-6}).

2.3   Robot Control (Industrial Category)

A very representative industrial subnetwork use case is the wireless control of robot operations (see Figure 4), including multi-axis robots, force control, and the coordination of moving or rotating parts in machines like printers and packaging systems. Replacing wired connections with wireless can enhance flexibility in movement, enabling greater degrees of freedom in manufacturing processes. Traditionally, these robot control systems rely on wired technologies such as Profinet IRT or EtherCAT, which support extremely short communication cycles to ensure precise and responsive operation. This use case focuses on replicating these capabilities over wireless networks while preserving performance and quality of service.

Robot control subnetworks are typically deployed within factory production cells or mobile robots, occupying a volume of about 3–5 m³. Sensors and actuators are placed on robot components and the surrounding plant area, with controllers located nearby—within 3 meters—or integrated into the robot or cell itself. HC serving as Access points (APs) can be co-located with controllers and connected to a parent 6G enterprise network managing broader factory operations. In sequential production lines, subnetworks may be adjacent, while in matrix production layouts, a minimum spacing of 2 meters between subnetworks is expected.
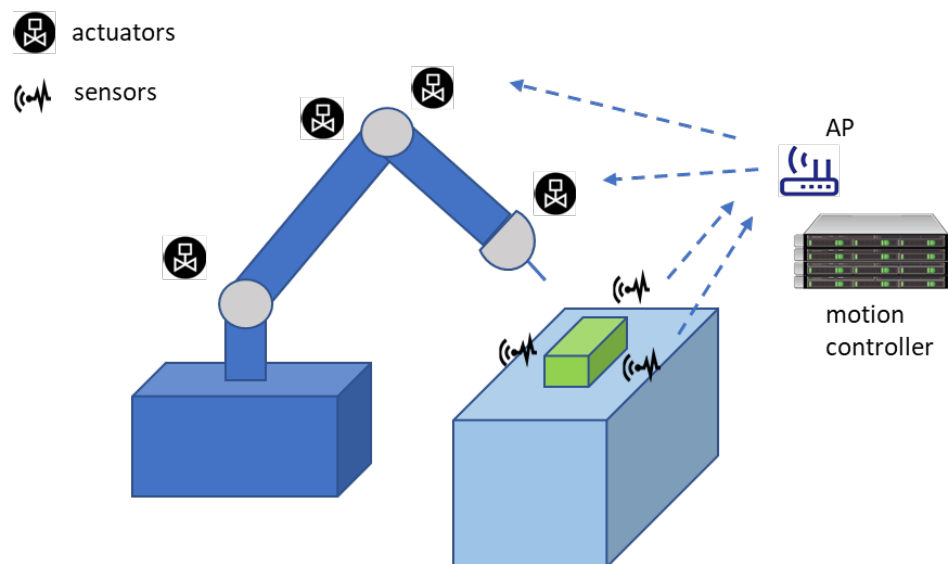
FIGURE 4 ILLUSTRATION OF THE ROBOT CONTROL USE CASE (D2.2 [3])

Main KPIs of this use cases relates to the support of low latency, down to 100 μs, as well as the ability of supporting heterogeneous communication cycles (e.g., ~up to 100 ms); high reliability in the order of 10^{-6} and below; scalability to a relatively large number of sensors, in the order of ~50; support of device mobility up to 20 m/s. We refer to deliverable D2.2 for a thorough description of the use case and its requirements.

Short-range high reliable communication with low latency and fast support of device mobility may benefit from significant enhancements in multi-antenna solutions and beamforming. Specifically, **Metasurfaces for ultra-fast and low-complexity beamforming** solutions, i.e. self-conjugating metasurfaces and dynamic scattering arrays, can support the stringent latency and reliability requirements of this use case with very low part consumption, by moving part of the signal processing to the electromagnetic domain. Communication reliability can be hindered by external interference (e.g., not coming from other robots), including intentionally malicious interference. In this respect, a **Native jamming-robust PHY design** can proactively offer a tier of robustness to such interference; while **OFDMA-based jamming mitigation** techniques can reactively adjust frequency resources according to the classification of external interference.

The support of low latency can be eased by the enhancements of communication procedures as defined by current 3GPP radio technologies. For example, the usage of **Proximity grouping** techniques exploits the similar radio conditions experienced by devices in short-range industrial subnetworks for enabling fast pre-emptive retransmissions. In order to sustain the periodic control traffic at multiple periodicities, **configured grant enhancements** with multiple grants are also foreseen.

Coded random-access enhancements can address grant-free transmission of multiple sensors operating at the same time, supporting improved scalability while coping with very short cycles.

Since the movements of devices within a robot may follow known or predictable patterns (e.g., sensors on a robot arm), **Predictive scheduling and pre-allocation mechanisms** can be used for anticipating the radio channel behaviour and significantly reducing the signalling overhead due to channel state information reporting.

As also highlighted in the propagation studies presented in deliverable D2.3, in spite of the short-range characteristics, industrial subnetworks may suffer from harsh radio conditions due to signal blockages (e.g., due to metallic obstacles). **Cooperative in-X communication** methods rely on the presence of multiple intermediate nodes within the industrial subnetwork to provide link diversity, translating to a significantly large spectral efficiency for tight cycles. **Network coded cooperation** further exploits the principles of network coding to enhance cooperative communication.

In the case of in-robot subnetworks operating on the unlicensed spectrum, latency and reliability can be also affected by other systems operating over the same frequencies. **Latency aware access in unlicensed bands** is therefore needed for counteracting the carrier sense multiple access limitations; our proposed method relies on the introduction of prioritization mechanisms which support deterministic communication cycles without hindering the performance of the underlying traffic.

The ensemble of methods studied here represents a solid portfolio for supporting low latency and high reliable communications for in-robot subnetworks, while also being applicable to other use cases demanding low latency with high reliability. Most of the methods are complementary and can be efficiently integrated in a holistic design; this is left for future higher TRL research.

A summary of the capabilities of the proposed methods is shown below, while their detailed description is presented in Section 3 and Section 4.

Metasurfaces for ultra-fast and low-complexity beamforming
Two different solutions have been investigated: Self-conjugating metasurfaces (SCMs) and Dynamic scattering arrays (DSAs).   • SCMs is a solution for ultra-fast MIMO beamforming with low-complexity SNEs and short packets in high frequency subnetworks. The adoption of SCM-based SNEs provides an energy efficiency gain of at least x10 and no computational effort at the SNE side compared to classical active solutions. SCMs are applied to low-complexity, ultra-fast MIMO beamforming in the presence of uncoordinated interference coming from the same subnetwork or from a neighbouring subnetwork. Results indicate a packet error rate $<10^{-6}$ for a latency <100 us with an interfering traffic load from uncoordinated adjacent subnetworks of G=2 packets per packet time.   • DSA as general-purpose EM signal processor to deal with energy consumption, complexity and latency reduction. Adopted in HC or 6G BS devices and it can hence be applicable in all the use cases of the 6G-SHINE project. DSAs provide unprecedented flexibility in managing the EM field by reducing or circumventing digital processing at the baseband, thereby surpassing the flexibility offered by classical MIMO and recent SIM implementations. The idea is to move signal processing tasks required to implement beamforming and precoding in multiple-antenna multiple-SNEs scenarios to the EM level. Compared to current MIMO digital and hybrid technology adopted in 5G systems, a power reduction > x10 has been achieved with near-zero latency without sacrificing the processing flexibility.
Native jamming-robust PHY design
This technology is of interest for industrial setup with a large number of robots with their own installed subnetwork, carrying out critical operations that might be jeopardized by external interferers. The proposed solution is able to achieve robustness to both legitimate and external interference. Performance is evaluated by aligning to the assumptions on radio propagation to the findings reported in D2.4. The proposed method is able to achieve spectral efficiency losses below

10% in the presence of a jammer with similar power level as the subnetwork power level. A failure probability below 10^{-6} is reachable when the jamming-robust PHY is combined with a centralized packet repetition scheme. Such centralized packet repetition scheme also brings a factor of x10 reduction in terms of failure probability with respect to the case of optimized fixed number of repetitions.

OFDMA-based jamming mitigation

The OFDMA based jamming mitigation technique first classifies the interference and then apply proper scheduling which takes into account the interferer's frequency, time span and physical location. Experimental evaluation is performed and show that in the tested scenario the throughput is maintained almost identical as when interferer is not present, which is a 33% increase when comparing to a baseline state-of-the-art scheme.

Proximity grouping

The proximity grouping method exploits the characteristics of a subnetwork, i.e., small area and known or highly predictable movements, to enable pre-emptive retransmissions and predictive scheduling for transmissions and receptions between HC and SNE. The method is further enhanced with eavesdropping and piggybacking methods to provide transmit and receive diversity and efficient utilisation of resources in a subnetwork. The main target KPI is on low latency (< 1 ms) and high reliability (BLER < 10^{-6}). It is shown that utilising proximity grouping, the pre-emptive retransmission is able to reduce latency of retransmission by at least 50% compared to the current Rel-18 5G network on a typical slot format configuration.

Configured grant enhancements

The Link Adaptation Aware (LAA) – Multiple configured Grant (MCG) method enhances traditional Configured Grant (CG) scheduling by defining multiple CG configurations (i.e., MCS + RB allocations) per SNE, then using lightweight (short-DCI) signals to switch among these as channel conditions change. Compared to single-CG, it adds a small control overhead but provides partial link adaptation while still retaining CG's low-latency advantages. LAA-MCG aims to improve throughput (spectral efficiency) and reduce latency, particular the one coming from retransmissions, versus standard single-CG while keeping signalling overhead lower than a fully dynamic scheduling scheme. Operating on slot-based scheduling and under moderate channel conditions, the spectral efficiency can be increased by about a factor of 1.5–2 compared to single-CG. As for latency, shorter retransmission times (up to ~2–3× improvement in worst-case delays) have been observed coming from reduced retransmissions.

Coded random access

This method focuses on frame-asynchronous coded slotted ALOHA (FA-CSA), an advanced protocol designed to enhance the performance of industrial subnetwork applications. FA-CSA extends traditional CSA by introducing frame-level asynchrony, enabling more flexible access opportunities. Additionally, the protocol can be further enhanced through retransmission mechanisms and scheduled resource allocation, improving contention resolution and optimizing the trade-off between scalability, reliability and latency in both grant-free and scheduled communication scenarios. The target KPIs include ultra-low latency (e.g., communication cycles of 50 µs) and high reliability (e.g., packet error rates below 10^{-6}). The main results demonstrate that FA-CSA provides significant improvements over state-of-the-art (SOA) grant-free access schemes. Starting with the baseline FA-CSA configuration, operating with K=2 subcarriers, the system achieves a packet loss rate of 10^{-7} with a reduction of a factor ~x10 with respect to the target requirements. When a retransmission mechanism is introduced, the system further enhances its scalability-reliability trade-off by recovering lost packets, illustrating its strong performance even in high-load conditions.

Predictive scheduling and pre-allocation mechanisms
This technology/method leverages an LSTM-based approach for CSI prediction and proactive resource allocation. The goal is to proactively allocate resources based on LSTM predictions. To minimize latency to the HC and ensure compliance with the stringent latency requirements of industrial applications, our strategy assigns the LC with the strongest link to each SNE. The target KPI for latencies is below 100 µs and for reliability, it is 10^{-6}, both of which are validated through simulations. A 20% improvement in spectral efficiency is achieved with the predictive scheduling scheme compared to relying on outdated data under the same reliability conditions. Our method achieves up to ~x8 reduction in pilot transmissions using predicted CSI compared to delayed CSI for the reliability of 10^{-6}.
Cooperative In-X communication
This technology/method exploits different cooperative methods, including decode-and-forward and amplify-and-forward, to achieve macro-diversity. Both TDMA and FDMA schemes are considered for uplink transmission, assuming multiple SNEs within the subnetwork and a few LCs acting as relays. Our strategy is to assign the LC with the strongest link to each SNE to minimize latency to the HC, ensuring compliance with the stringent latency requirements of industrial applications. Cooperative communication enhances spectral efficiency and reduces latency by enabling more flexible resource allocation within the same band. A ~x10 reduction in latency compared to 5G is achieved by reducing transmission time from 1 ms to 100 µs. Moreover, a latency reduction of at least ~x2 is achieved for 1 of 4 cooperative communication compared to single-hop transmission under the same reliability conditions and in low SNR scenarios. Another key objective is power efficiency improvement through macro-diversity. Results indicate that increasing the number of APs significantly enhances energy efficiency. Specifically, with four APs (4 APs), emitted power is reduced by up to 7.5 dB compared to single-hop transmission. The proposed solution leads to a ~x2 and ~x4 higher spectral efficiency than single hop transmission when 2 or 4 relays are exploited, respectively, for a same reliability level of is 10^{-6}.
Network coded Cooperation
Network Coded Cooperation in a TDMA/TDD system supporting cycle times of 100-250 µs has been investigated. This achieves a 4-10-fold reduction in cycle time compared to 5G. The most important KPI of NCC is the increase in reliability, especially in the presence of obstacles. The links between HC and SNEs blocked by obstacles can be efficiently compensated for by NCC. Instead of communication breakdowns as would occur with classic error control methods based on retransmission, NCC still enables error-free transmission (in the sense of a probability for 2 consecutive packet failures < 10^{-6}) while keeping the number of necessary resource blocks equal or even lower.
Latency aware access in unlicensed bands
The proposed LA MAC TC introduces a hard prioritization mechanism in MAC schemes allowed in unlicensed bands. We designed a simulation model of the proposed LA MAC to evaluate its possible performance, achieving near deterministic operation independent of network load. Measuring its performance based on KPIs like Application latency, MAC delay and jitter, we were able to achieve up to 93% reduction in average and 95% reduction in maximum latency where it matters, that is in the medium to high network congestion conditions where legacy MAC schemes break down. That is almost a 20x improvement in average and maximum latency. The achieved jitter was from 50% to 93% lower versus the legacy systems as the network load was increased, offering a 2x to 10x improvement to jitter. It is important to point out that the proposed MAC scheme is technology agnostic and can be applied to any application or use case that could benefit from harvesting network capacity from ISM bands with latency and prioritisation requirements in place.

2.4 Visual Inspection Cell (Industrial Category)

This use case describes a visual inspection cell designed for quality assurance in an industrial manufacturing environment. The core functionality involves using video feeds to monitor parts moving along a conveyor belt, which are then analysed for defects or anomalies. Based on this analysis, the system can generate commands for actuators to perform corrective actions, thereby improving the overall quality of the manufacturing process. The visual inspection cell is typically installed adjacent to a conveyor belt that transports components for assembly across various production modules. It includes several interconnected components: 1) cameras, possibly equipped with integrated laser sensors, capture detailed images or video of the components in motion; 2) an Image Processing Unit (IPU) receives this visual data and performs localized analysis; 3) a soft Programmable Logic Controller (PLC) interprets the processed data and triggers appropriate quality control responses; 4) an actuator executes the corrective action if any quality deviation is detected; and 5) a Human-Machine Interface (HMI) device may be included for monitoring purposes. Notably, this HMI can attach or detach from the inspection cell during runtime without halting operations, enabling flexible operator interaction. All devices are connected to a wireless subnetwork, communicating through a local access point (AP). In some implementations, the AP may be co-located with an edge processing server, which takes over the PLC tasks. Alternatively, the edge server may exist as a separate device within the subnetwork. Additionally, the camera and IPU can either be integrated into a single device or connected via a wired link, appearing as one logical node from the subnetwork's point of view. This setup exemplifies a distributed, wireless automation architecture supporting low-latency, high-reliability communication, essential for real-time quality control in smart manufacturing systems.
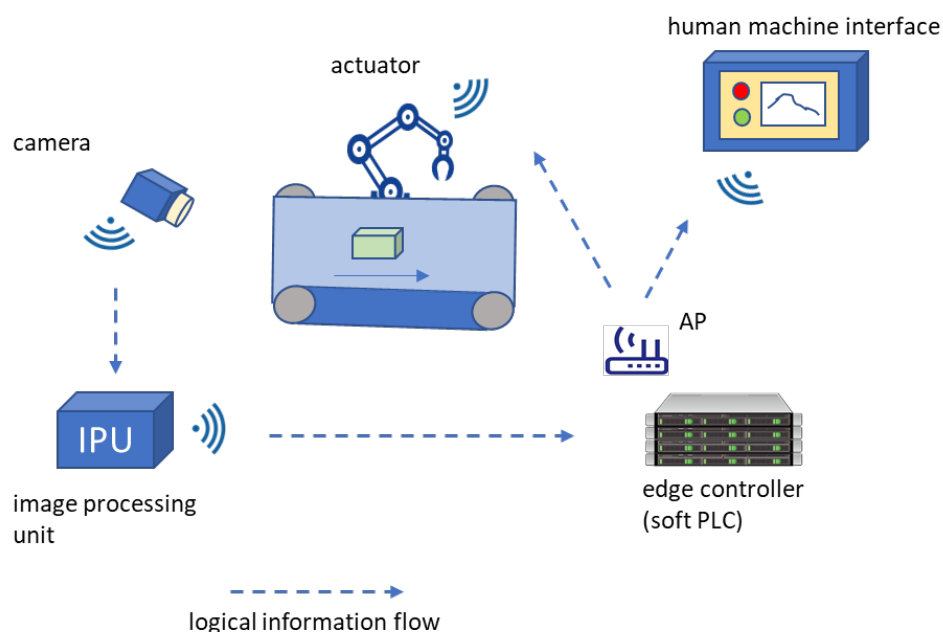
FIGURE 5 ILLUSTRATION OF THE VISUAL INSPECTION CELL USE CASE (D2.2 [3])

This use case can benefit of same low latency enablers as robot control. A peculiar characteristic of the visual inspection cell use case is the need of multiplexing traffic with different characteristics, such as high data rate traffic (up to 50 Mbps) generated by cameras, with frequent small packets transmissions (~100 bytes) for control commands to the actuators.

The characterization of the visual inspection cell traffic has helped on identifying **full/flexing duplexing techniques** as a key enabler for supporting the KPI needs of the visual inspection cell use case. Such techniques enable simultaneous uplink and downlink transmissions over different subbands, therefore efficiently supporting such diverse traffic. By applying power control to reduce the cross-link interference, this technique has shown its capabilities to efficiently support the demands of high-bandwidth video data multiplexed with intermittent UL/D traffic from sensors/actuators. A summary of the capabilities of the proposed method is shown below, while its detailed description is presented in Section 4.

Full/flexible duplexing for enhanced scheduling flexibility
This method targets use cases like the visual inspection cell where uplink (UL) and downlink (DL) traffic of different characteristics has to be multiplexed, i.e. high throughput video streaming from cameras, and periodic transmissions to sensors/actuators. The usage of flexible duplexing enabled HCs allows to efficiently multiplex these traffic types since it avoids forcing devices supporting different traffic to have a same UL/DL switching point. Different power control strategies- including a joint UL/DL power control solutions- are proposed, for the sake of reducing the impact of in-band emissions from simultaneous UL and DL transmissions over neighbour sub bands. Power control is shown to have a larger impact than frequency separation, at least when the joint UL/DL power control solution is applied. Flexible duplexing aims at improving spectral efficiency and reduce latency, by enabling a larger flexibility in resource allocation in a same band. A factor of ~x10 reduction of latency with respect to 5G is obtained by construction, by allowing simultaneous UL and DL transmissions in the order of 100 µs in the same spectrum for the critical traffic. The spectral efficiency gain of the proposed joint UL/DL power control is of a factor up to ~x8 with respect to absence of power control for the critical traffic flow, at the 10^{-4} percentile of the distribution. On the other hand, this might be insufficient for achieving a packet error probability of 10^{-6}, which however can be obtained with a single retransmission.

2.5 Unit Test Cell (Industrial Category)

A unit test cell is meant to perform quality assurance tasks of product parts in the manufacturing process, as well as of devices used in the manufacturing process. For example, it can be used for providing calibration and tolerance figures for actuators to be used in a factory, to make sure they comply with predefined requirements during execution.
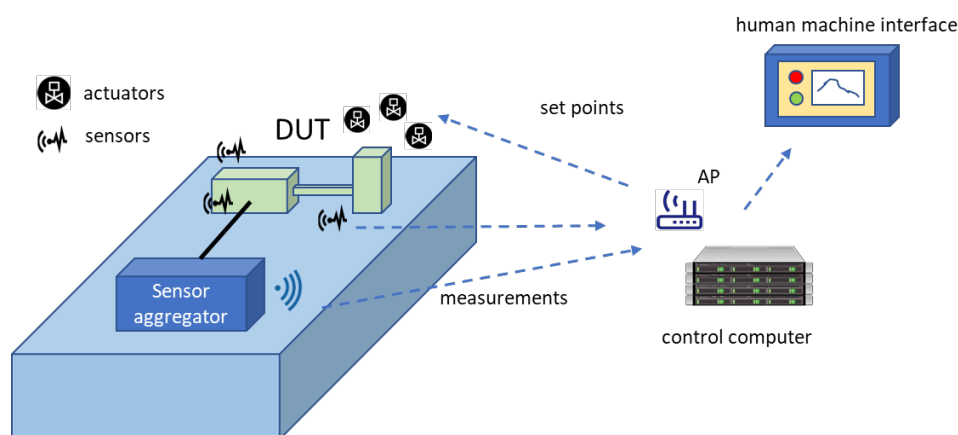
FIGURE 6 ILLUSTRATION OF THE UNIT TEST CELL USE CASE (D2.2 [3])

The KPIs of the unit test cell use case are related to the short communication interval (below 10 ms) the system must support for transmissions of ~200byte packets between the sensors and the subnetwork APs. Reliability targets below 10^{-6} is accounted in this case as the probability of having two consecutive errors.

As a 6G candidate spectrum band, especially for industrial scenarios, we have analysed **adaptive low-latency and reliable unlicensed spectrum access schemes** for addressing this use case' needs. The proposals exploit outcomes from common unlicensed spectrum access schemes, such as the listen-before-talk, to dynamically utilize techniques such as retransmissions and multi-PUSCH that allow meeting the strict reliability requirements, even in harsh operating conditions, that are common in industrial unit test cell scenarios. We are showing below a summary of the capabilities of the proposed method, and full details are available in Section 4.

Adaptive low-latency & reliable unlicensed spectrum access
This scheme identifies as its primary use case the unit-test cell where SNEs in a subnetwork forward periodic or possibly aperiodic traffic to a HC node. SNEs with periodic traffic operate using semi-static channel access combined with UE-initiated channel-occupancy time (COT), which provided predictable transmission opportunities with minimal signalling. However, congestion might arise from aperiodic traffic SNEs which might operate under dynamic channel access. To address this, an adaptive algorithm is proposed that improves how semi-static channel access SNEs utilize the granted COT once access is obtained. Rather on relying on semi-static transmission strategies, the algorithm dynamically switches between Repetitions and multi-PUSCH modes, based on feedback from listen-before talk (LBT) outcomes and HARQ responses. This adaptive behaviour enables more efficient channel usage without adding any signalling overhead and results in fewer packets dropped due to exceeding the latency deadlines. More specifically, a 2-3x decreased packet drop ratio can be observed on average across the full SNR range compared to the two current static schemes.

2.6 Virtual (Wireless) ECU (In-Vehicle Category)

This use case is aimed at enabling wireless zonal ECUs for in-vehicle zonal E/E and computing-centralized architectures [3]. The in-vehicle zone is characterized by the presence of sensors and actuators that support automotive functions and systems from different in-vehicle domains (e.g., ADAS, powertrain, etc.). In the in-vehicle zone, there is also a zone ECU that manages and controls the sensors and actuators that are located in this zone. The zone ECU is connected to the High-Processing Computing Unit (HPCU) which is connected to the 6G parent network.

This use case defines the 6G in-vehicle wireless zone subnetwork that is utilized by some sensors and actuators located in this zone to connect wirelessly to the zone ECU that manages and controls them. The zone ECU includes a wireless communication interface and AP capabilities to wirelessly communicate with the sensors and actuators of this in-vehicle wireless zone, and it maintains the wired communication interface(s) that it uses to communicate with the sensors and actuators that are not part of the 6G wireless subnetwork and with the HPCU. The 6G in-vehicle zone wireless subnetwork is considered to replace cable links utilized to support demanding and critical information, which justifies the need for a 6G wireless technology. For example, this might be the case of cable links utilized to transfer (raw) sensor data generated by sensors like cameras, radars, or lidars, which are utilized in ADAS-based automotive systems.

As an extension of the wireless zone ECU use case, the virtual ECU use case focuses on integrating the in-vehicle (wireless) network with the 6G parent network, following the 6G 'network of networks' paradigm. The goal is to seamlessly extend the in-vehicle embedded computing capabilities to the edge/cloud. The connection from the in-vehicle network to the edge/cloud is exploited in this use case to enable opportunistic offload and vehicle-network-cloud cooperation to support advanced in-vehicle automotive functionalities critical for AD and the continuous evolution and advancement of vehicles. This way the edge/cloud would act as a virtual ECU (or HPCU) by elastically extending the computing and processing capabilities of the vehicle using the 6G network. This includes, for instance, the possibility to opportunistically offload to the edge or cloud the processing of the sensor data generated in an area supported by the wireless zone ECU, or some demanding functions like the machine learning inference. This use case also enables the possibility of data collaboration and machine learning operation based on the data from multiple vehicles. This use case utilizes the high capabilities of the in-vehicle HPCU (acting as communication control unit) to act as the bridge between the in-vehicle network and the 6G parent network.
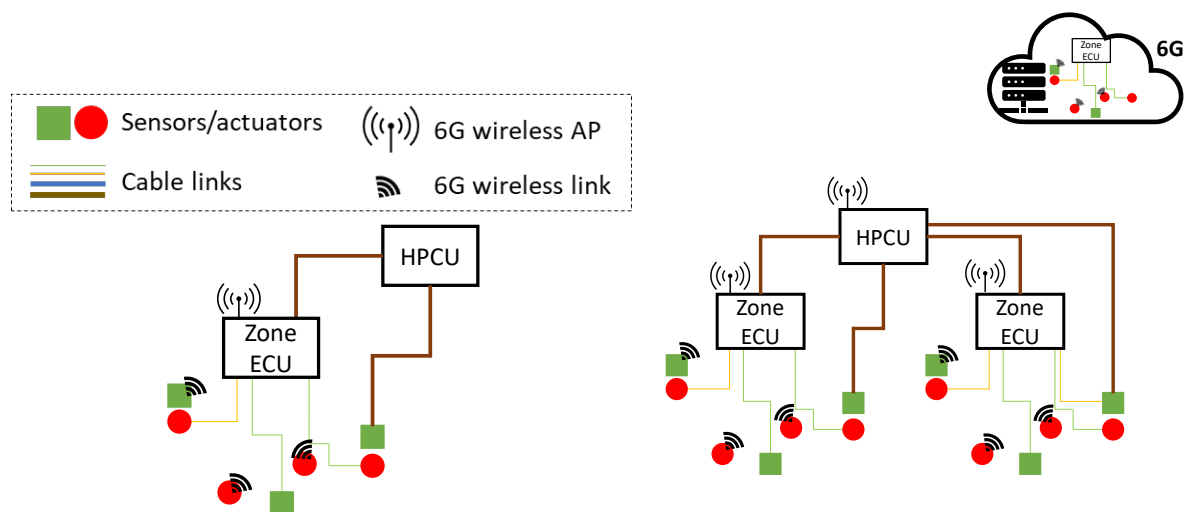
FIGURE 7 ILLUSTRATION OF THE WIRELESS ZONE ECU (LEFT) AND VIRTUAL ECU (RIGHT) USE CASES OF THE IN-VEHICLE SUBNETWORK CATEGORY (D2.2 [3])

The main KPIs associated with this use case are derived from the key automotive functions that these subnetworks must support. They include latency levels below 1 ms, throughput requirements ranging from 2 Mbps to 10 Gbps depending on whether the data is processed or raw, and strict reliability and determinism requirements.

To support the in-vehicle subnetwork requirements, we leverage the knowledge gained in WP2 from the characterization of in-vehicle network traffic and channel propagation characteristics. This has enabled us to define **predictive traffic scheduling schemes** that use this traffic knowledge to anticipate future demands and allocate resources accordingly—based on predicted packet sizes and arrival times— while ensuring that latency constraints (i.e., determinism) are respected for the mixed traffic generated within the in-vehicle network. We have extended the analysis of these proposals to scenarios where in-vehicle devices with high capabilities support **full/flexible duplexing**. We have demonstrated that this approach can achieve superior spectral efficiency and reliability while still meeting the strict requirements of in-vehicle traffic, thanks to the predictive knowledge of traffic patterns.

| **Deterministic and Predictive Traffic Scheduling** |
| The proposed predictive schemes, predictive dynamic scheduling (PDS) and predictive configured grant (PCG), advance conventional 3GPP scheduling by integrating traffic prediction, enhancing resource allocation, and ensuring deterministic service provisioning. The target KPIs include latency, reliability, and determinism, ensuring the sustainability of predictive schedulers under high transmission loads with diverse requirements. A factor of up to ~18x improvement in reliability is observed, particularly for transmissions with stringent latency requirements, while a factor of ~7x improvement is achieved for more relaxed requirements. For all transmissions with varying requirements, an overall improvement of ~2x in reliability is achieved compared to standard 5G schedulers. The same factor applies to determinism, as transmissions using the proposed methods consistently meet their latency constraints, ensuring successful packet delivery within their respective latency bounds. Our approach considers scalability, ensuring that even as transmission load increases, performance remains stable or improves by a factor of ~5x for stringent latency requirements, ~7x for relaxed requirements, and ~1.50x overall when compared with standard 5G schedulers. Additionally, resource utilization efficiency improves by up to ~2x, with significantly better resource utilization to meet latency requirements compared to the standard 5G schedulers. |
| **Predictive & Full/Flexible Duplexing Scheduling for Deterministic Traffic** |
| This scheme leverages predictive scheduling and full/flexible-duplexing mechanisms to further optimize in-vehicle network communications. We extend the PDS algorithm by integrating flexible duplexing, enabling intelligent resource allocation while minimizing interference risks—particularly cross-link interference caused by in-band emissions (IBE). The proposed joint predictive scheduling and flexible duplexing scheme exploits predicted traffic information and the spatial diversity of in-vehicle sensors and actuators (i.e., SNE) to allocate uplink and downlink transmissions to resources that minimize interference impact. Traffic predictions are utilized to prioritize resource allocations that satisfy the requirements of ongoing transmissions, while also accounting for resources likely to be needed by predicted future packets. The flexible duplexing dimension introduces both additional flexibility and new constraints in allocating resources to current transmissions. To minimize cross-link interference—specifically, the interference from uplink (UL) transmissions from a SNE to the HC or gNB that affects downlink (DL) transmissions from the HC or gNB to another SNE—our scheme strategically allocates UL and DL resources to SNEs that are spatially as distant from each other as possible. The proposed predictive method, leveraging spatial diversity to reduce mutual interference, resulted in improved communication reliability. This approach achieves a ~10.4% relative improvement in reliability compared to the benchmark method. |

3   ROBUST AND LOW COMPLEXITY PHY ENHANCEMENTS

This section presents a set of enhancements designed to optimize performance while minimizing complexity designed specifically for in-X subnetworks. The advancements are crucial for supporting the diverse requirements of in-X subnetworks, including high throughput, low latency, and resilience against interference.

This section begins in Section 3.1 by exploring innovative antenna array and panel configurations that offer improved beam alignment techniques, promising enhanced throughput with minimal signalling overhead. It then delves into the potential of metasurfaces and dynamic scattering arrays to revolutionize beamforming processes in Section 3.2, shifting complex signal processing tasks to the electromagnetic level for ultra-fast and low-power communication. In Section 3.3, sub-THz transceiver architectures tailored for VR applications are examined, providing insights into power consumption optimization and system-level design choices. Finally, in Section 3.4, the challenge of interference is addressed through jamming-robust PHY designs, leveraging advanced techniques like fast frequency hopping and OFDMA-based interference classification.

3.1   POTENTIAL OF ANTENNA ARRAY/PANELS FOR IN-X SUBNETWORKS

Beamforming is already today a key technical enabler for coverage enhancements in mid and high frequency ranges in wide-area and indoor deployments with Wi-Fi and 3GPP 5G NR. It is therefore also obvious that there can be benefits of beam focusing and beamforming in in-X subnetworks as captured in technology component 5 (TC5) as defined in the project but given that the in-X subnetworks are predominantly short-range and the devices may be size constrained, the potential of the technology needs to be studied. Additionally, this can set constraints and hints into how a beamforming or beam focusing management procedure should be supported for efficient use of the technology component in in-X subnetworks.

The study is split into three parts:
1. A deep dive into the characteristics that needs to be taken into account to study the benefits and costs related of doing beam forming is presented in Section 3.1.1
2. A static analysis (i.e. devices not moving) summary of what was presented in D3.1 [2] is provided in Section 3.1.2. Here the best obtainable gains for in both a robot control use case and an immersive education use case with ideal beam alignment was conducted, and
3. An analysis of beam alignment techniques with overhead and mobility considerations in an immersive education use case is presented in Section 3.1.3, along as using updated simulation assumptions and KPIs from WP2 in deliverable D2.3 [4] and D2.2 [3], respectively.

Lastly, the conclusions and recommendations of the study are presented in Section 3.1.4.

3.1.1   Beamforming for in-X subnetworks

Beamforming is a widely used technique in wireless communication, and can, potentially also for in-X subnetworks, be used to increase the communication coverage, increase spectral efficiency, reduce power consumption, reduce signal power to keep interference at a minimum and in turn increase spectral area efficiency. Additionally, it can improve resilience to jamming and strengthening privacy by making it harder to intercept and eavesdrop the subnetwork communication links.

Beamforming can be realized, for example, with a set of antennas and phase shifters which may be controlled differently depending on whether they represent a so called analogue beamformer, digital beamformer and/or hybrid beamformer (more in D3.1 [2]). For in-X subnetworks, several different device types or elements are defined in D2.2 [3]. One of them is a high capability (HC) subnetwork device, another of them is referred to as a subnetwork element (SNE), and another is the low capability (LC) subnetwork device. Device types can be enablers of each other, e.g., a HC device can offload some complexity from the LC or SNE subnetwork devices, which on the other hand, can be designed to be lower cost and hence contribute to the subnetwork ecosystem. An LC or SNE device might also be battery powered; this can set requirements to battery size, battery life and eventually power consumption. Regarding antennas, it can be assumed that a LC or SNE device can be equipped with relatively few and small antennas, very few radio frequency chains and with either no capability to perform beamforming or the capability to perform analog beamforming. A HC device can potentially, on the other hand, be equipped with multiple antennas, antenna panels and capability to do either analog, hybrid or digital beamforming. An example of a HC subnetwork device can be a 6G smartphone, or a 6G customer premises equipment (CPE), whereas examples of a LC subnetwork device could be a for example a pair of eXtended Reality (XR) glasses and a SNE could be an even simpler device such as a sensor or actuator.

There are several factors to consider when evaluating the potential of beamforming for in-X subnetworks. The most impacting factors are presented below:

Radio channel characteristics such as whether there is presence of multiple strong paths (i.e., by reflections) between the transmitter and receiver and whether any of these are line-of-sight (LOS) or non-line-of-sight (NLOS). Beam alignment may be a much simpler task when LOS links are present, as this involves an easy-to-detect and strong link (in a particular direction). In an NLOS condition, the sum of the radio links is what matters; if links are very scattered there may not be any directions that are better than others and consequently the benefits of narrow beam alignment may diminish.

The design and implementation of the antennas are hugely important to the beamforming shape that is eventually propagated from the device. For example, a one-dimensional constellation of antenna elements often used for beamforming with a smartphone with 4 antenna elements, should be able to give a beamforming gain of up to 6 dB due to the concentration of the emitted radio energy. However, when it is implemented on a smartphone, considerations such as placement and antenna panel size, and interaction with the smartphone material and components, can result in a much smaller gain.
In FR2 (28-71 GHz) an implementation loss for a typical smartphone is around 4 dB for a monopole antenna, whereas an estimate has been made for antenna arrays to place the implementation loss from 6 dB [5] and up to 8 dB for a 4 element antenna array [5][6]. For an antenna array on a pair of XR glasses the implementation loss is expected to be at least 8 dB and most likely much higher [6]. A reference model of a typical smartphone placement with four antenna panels from [7] is shown in Figure 8, and in Figure 9 antenna panel placement for XR glasses is illustrated. These single row antenna panels are typically considered in smartphone and XR glasses as they are small in one of their dimensions, but this also comes with the constraint that the patterns (as illustrated in Figure 9) from the generated beams can only be made narrow in the panel plane, but not in a perpendicular plane, where these beams will appear very wide.

External interaction with the device is particularly relevant when considering a smartphone device. Here, it has been shown that when the smartphone is handheld (by a human), the placement of the head, hand and fingers can have a dramatic effect on the propagation patterns and amount of energy being emitted. Human interaction causes absorption and influences directly the electromagnetic properties of the antenna element, which is why the patterns are radically different compared to when no human hand interaction is present.

An example with simulations conducted can be found in [5][7], where it has been shown that the human hand impact can be from 2 dB to 8 dB radiated efficiency loss per panel, where it should be noted that in most grip options there is another panel not as affected; this motivates that multiple properly placed panels should be used when human hand grips are expected.

Mobility of the device, external interactors and other entities in the scenario impact the design of the beam management framework as it needs to be designed in a way that can enable the system to adapt the link between the two devices sufficiently fast when the link performance changes, due to e.g., the factors mentioned above. Assuming that this is not an issue for small scale fading (due to larger bandwidth), it can be due to blockage (for example robotics or other humans) causing that some of the dominant paths used by the link are getting absorbed or redirected. Such detection typically requires measurements by the receiver followed by a report to the transmitter, which can then trigger a link adaptation procedure such as a beam "realignment" procedure. While that sounds simple, the overhead of signalling and measurements increase with the dynamic parts within the system and so does the overhead of the link adaptation process.
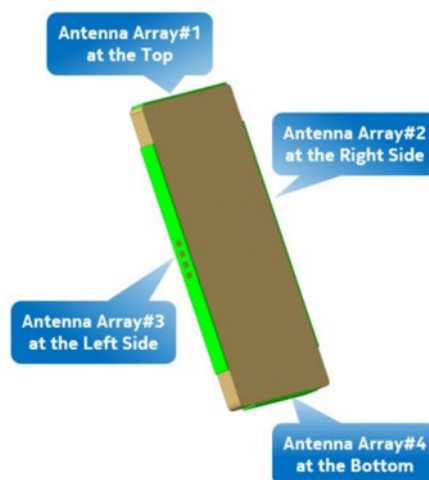
FIGURE 8. ANTENNA PANEL PLACEMENT EXAMPLES FOR A SMARTPHONE. PICTURE FROM [7].

FIGURE 9. POSSIBLE PLACEMENT OF ANTENNA ARRAYS ON XR GLASSES. PICTURE FROM [7].

3.1.2 Static analysis of beam forming gains in in-X subnetworks

This first part of the analysis was disseminated in D3.1 [2], and considered two scenarios, an indoor factory scenario with the robot control use case and an indoor classroom scenario with the immersive education use case. Both scenarios were modelled to take place in a 20m x 20 m x 3m (length, width, height) room consisting of several independent subnetworks with a HC device in the middle communicating with a set of SNE/LC devices in a proximity of maximum 2.5 meters for the robot control use case and 1 meter for the immersive education use case. For HC were considered to use up to 4 antenna panels with up to 16 dual polarized antenna elements and the LC devices were considered to use also up to 4 antenna panels with maximum 4 antenna elements in each panel. An implementation loss of 4, 6 or 8 dB at the antenna elements for the largest antenna panels to model the penalty of constraining the antennas in size in e.g. a handheld device (smartphone or even a pair of XR glasses). A key performance outcome of this first part can be found in Figure 10, where the average downlink SINR for the robot control use case is presented. Clear gains are observed over a theoretical omni antenna, also considering the antenna implementation losses. These implementation losses used in D3.1 [2] were defined in scenarios, such that implementation loss scenario 0 is representing an ideal scenario, scenario 1 represents a typical smartphone implementation loss of 4dB, scenario 2, is similar to 1 but adds a 6dB antenna implementation loss to antenna panels larger than 1x4, and scenario 3 is also similar to 2 but adds an implementation loss of 8dB instead of 6dB to represent a pair of XR glasses. In D3.1 [2] the key take aways can be summarized as follows:

- Antenna panels can provide significant gains in both an indoor factory and indoor classroom scenario, particular in the indoor factory scenario.
- Four antenna panels provide a better coverage than two antenna panels and is needed to meet and exceed the performance that can be achieved with a dipole antenna.
- With the high gains achieved of using antenna panels, there is a potential for scheduling coordination between subnetworks, and for beam aware radio resource management (RRM) mechanisms such as power control to exploit the directivity gains while minimizing generated interference and power consumed.
- The antenna implementation losses do impact the achieved gains, but not to such extend that there are still clear benefits of using antenna panels, also with 4 or more antenna elements.
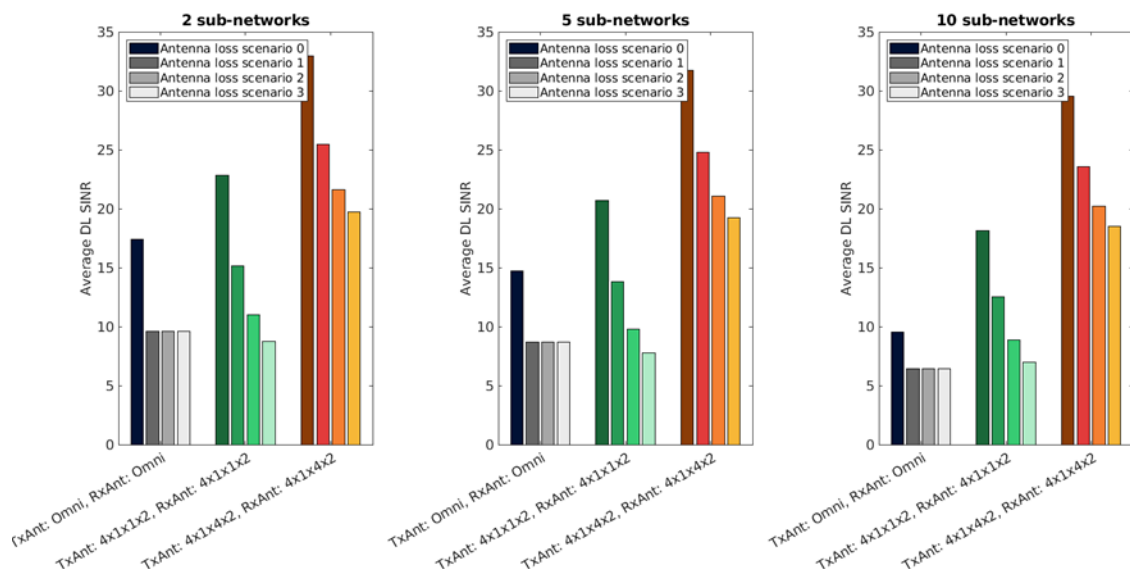
FIGURE 10. AVERAGE SINR IN INDOOR FACTORY SCENARIO, VARYING SUBNETWORKS, ANTENNA SETUPS, AND LOSS SCENARIOS FROM D3.1

3.1.3 Dynamic analysis of beam management gains in in-X subnetworks

The first part of the study gave clear indications that beam forming in in-X subnetworks scenarios has potential, but these do assume ideal beam alignment (here meaning that the best Rx and Tx beam pairs are always used), which is not a realistic assumption. In practice a beam alignment procedure will be put into place and attempt to find the best beam pairs. The challenge of the beam alignment procedure is to be executed as infrequent as possible, and when executing it, it should consume as few resources as possible (i.e. in terms of reference signals to be transmitted and reference signals to be measured (processed) as well as beam reporting to be transmitted as feedback). The less frequent the procedure is executed the lower overhead is imposed, therefore radio resources can be used for something else than control plane data).

In 3GPP NR 5G-Advanced, the state-of-the-art beam alignment procedure to identify the used beam between a base-station and a user equipment (UE) is a three-phase procedure [8]. In phase 1, the base station is transmitting a set of reference signals, throughout its cell. The UE will measure and select the best SSB beam and report this choice to the base station. In phase 2, the base station will transmit a set of narrow beams also referred to as CSI-RS beams, of which the UE will measure, and again report back to the base station which one of them is the best. In phase 3, the UE transmits narrow beams to the base station of which measures the best one and reports this to the UE.

While the procedure has demonstrated to work well for wide-area scenarios (and predominantly as a downlink coverage enhancement feature), it is not clear whether such lengthy procedure is needed for a short-range scenario, and additionally whether this is suitable for a device-to-device driven scenario, where both devices may be mobile in contrast to the base station which is typically stationary. Additionally in the context of a distributed device-to-device scenario, it is important to not change transmit power out of sync with other ongoing transmissions as that can cause AGC issues (as discussed in D4.1 [9]), which in turn means that no transmissions can take longer than a pre-agreed transmission slot (14 symbols in 5G NR), which again means that it may take a lot of time to complete a transmit beam sweep based scheme (as 5G NR).

An alternative beam alignment procedure, which relies on channel reciprocity to significantly simplify the 5G NR beam alignment procedure, has been proposed for 5G NR Sidelink in Release 18 [10]. This procedure, adapted for sub-network terminology, is illustrated in Figure 11, where two devices (HC and SNE) are executing the beam alignment procedure. Whereas phase 1 is in principle similar as to the 5G NR procedure, phase 2 is different as it relies on the wide beam transmission (selected from phase 1) and the sweeping of narrow beams on the receiver side. The best narrow beam is selected (but not reported) and used for transmissions in phase 3, where the directions are now reversed.
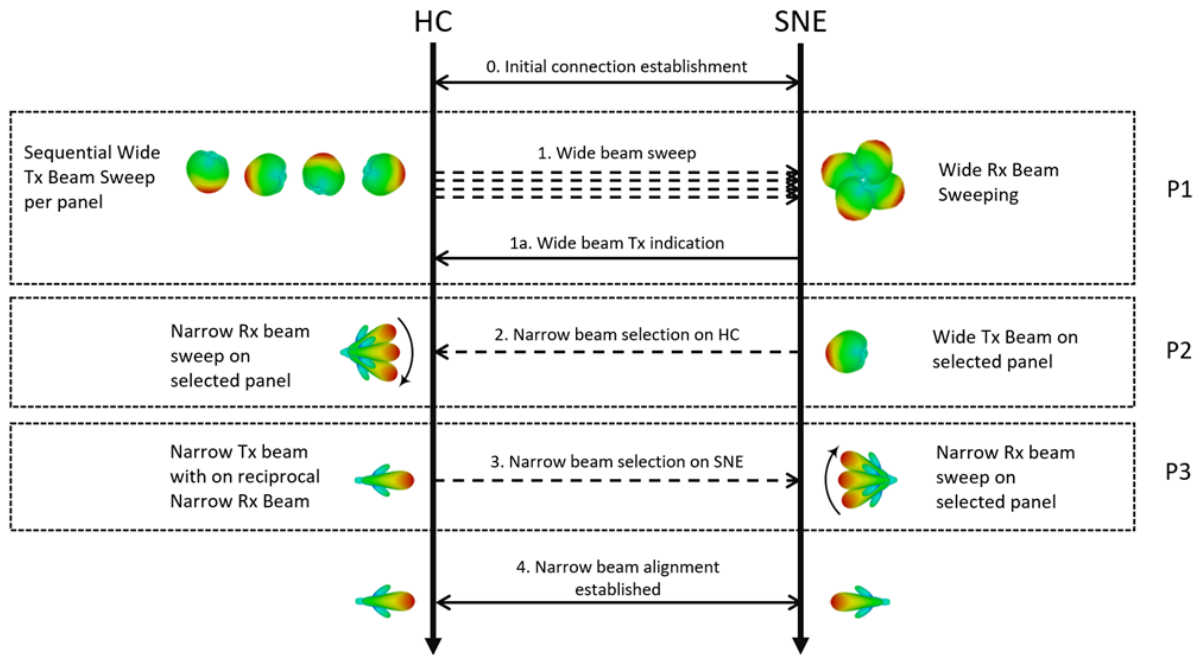
FIGURE 11. ILLUSTRATION OF THE PROCEDURE OF RECIPROCAL BEAM ALIGNMENT.

To evaluate the performance of this beam alignment procedure, it is going to be compared against two other beam alignment schemes: A. Only phase 1 (i.e. relying entirely on wide beam alignment) and B. Tx based narrow beam sweeping using phase 2 and phase 3 (similar as to 5G NR).

Obviously, a phase 1 only beam alignment procedure can have several benefits in a short-range subnetwork, as it is fast to execute and introduces only a small overhead as it only needs a Tx and Rx sweep of all wide beam options (at both HC and SNE) and might not need a frequent re-evaluation when the devices are moving. Whereas a scheme relying on phase 2 and phase 3 narrow beam exhaustive search, is a more lengthy procedure to execute as it needs to sweep all narrow beam options (at both HC and SNE), but on the other hand, it would also find an optimum beam alignment pair given that the devices have not moved out of beam synchronization before the procedure completes. Therefore, it is also clear that such procedure will be more prone to re-evaluation when the devices are mobile. The reciprocity-based scheme, aims to harvest the benefits from both schemes, being fast to execute, only requiring slightly more transmissions than the phase 1 only procedure, but would still be as sensitive to re-evaluation as the narrow beam only alignment procedure. Additionally, the reciprocity-based scheme, also impose the risk of selecting a narrow Tx beam (in phase 2) which is suboptimal compared to a wide Tx beam or even another wide beam pairing, if the narrow beams are capturing a weak ray (reflection or LOS component) but in fact it would have been better to capture all reflections.

To evaluate the performance of these three beam alignment procedures, an extensive system-level-based evaluation has been conducted evaluating the achievable throughput from the three schemes, and then the sensitivity of the schemes as a function to how fast the SNE/LC devices are moving. The simulation assumptions are provided in Table 1.

TABLE 1: SIMULATION ASSUMPTIONS

Parameter	Value
Scenario	6G-SHINE use case: Immersive education [3] A single room of 20 m x 20 m x 3 m (length x width x height), 6 subnetworks of a single SNE/LC are being served by a single HC. The HC is placed in the corner at a height of 2 meters. The SNE/LCs could be a smartphone or XR glasses. When being served by the HC, they are treated equally. A generalized illustration of the considered scenario is provided in Figure 12.
Traffic modelling	Indoor classroom: XR traffic following 3GPP TR 38.835 aligned with the traffic description in D2.2 [3]. From the HC to the SNE/LC a video stream modelled as an interarrival rate between 12.67 ms and 16.67 ms with a packet size between 32 kB and 93 kB following a truncated Gaussian. From SNE to HC a periodicity of 4 ms and a size of 100 B.
Channel model	Indoor factory (InF-DH) from 38.901. It has been chosen to use this in the immersive education, as the model is considered to fit the short-range campus scenario as studied in D2.3.
Antenna panel locations	2 panels: Panel#1-#2 placed on each side of the device, alike the layout depicted in Figure 8 4 panels: Panel#1-#4 placed on all four sides of the device, alike the layout depicted in Figure 8
Panel configuration	1x1x2 (MxNxP): 1 beam per panel at 0deg. 1 dB antenna gain assumed for the antenna element. 1x4x2 (MxNxP): 4 beams configured and equally spaced from -56deg to +56deg. 1 dB antenna gain is used for this configuration at the HC, and -1 dB at the SNE. 1x16x2 (MxNxP): 16 beams configured equally spaced from -60deg to +60deg. -1 dB antenna gain per antenna element is used for this configuration at the HC (this configuration is not used at the SNE. Note 1: The antenna gain losses are obtained from [7] and used in D3.1 [2] as well. These are relative to an antenna gain of 5dB per antenna element. Lower values are applied at the SNE/LC as it is assumed these would have smaller antennas than those on the HC. Note 2: Notation of the number of panels is placed in the front of the antenna panel configuration, e.g. 4x1x4x2 is 4 panels using configuration 3.
Beam alignment procedure	Procedure 1: Phase-1-based beam alignment Procedure 2: Reciprocity-based beam alignment (as illustrated in Figure 11) Procedure 3: Narrow-beam-based beam alignment (Phase 2 and Phase 3 from 5G NR)
Mobility model	Random walk model. 50% chance of direction change per meter, maximum direction change 0.79 radians per change. Speeds considered are 3 kmph (0.83 m/s) and 7.2 kmph (2 m/s)

Carrier frequency and bandwidth	28 GHz carrier frequency with 200 MHz bandwidth
Frame structure	TDD band with DDDUD pattern. Orthogonal frequency division multiplexing (OFDM) with 60 kHz SCS, 14 symbols per slot and 1 slot takes 250 µs
Scheduling	Proportional fair scheduling in both uplink (UL) and downlink (DL).
Link adaptation	Link adaptation targeting 10% error rate, following MCS table 4 from 38.214
Power control	Fixed transmit power of 10 dBm for the HC and SNE
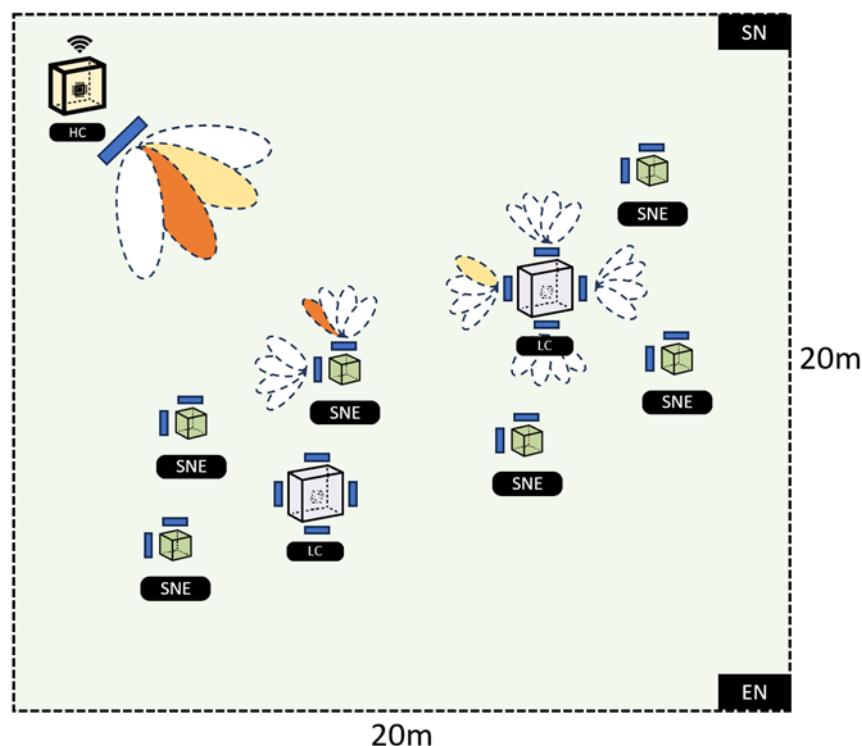
FIGURE 12. VISUALIZATION OF AN IMMERSIVE EDUCATION SCENARIO UTILIZING TECHNOLOGY COMPONENT BEAM FORMING. HERE AN HC DEVICE IS PLACED IN THE CORNER, WITH A SINGLE (POTENTIALLY LARGER) ANTENNA PANEL. SNE/LC DEVICES WITH TWO OR FOUR PANELS ARE IN THE ROOM AND WHILE BEING IN THE ROOM THEY ARE SERVED BY THE HC.

An important aspect of these three beam alignment procedures is the imposed overhead of executing the procedure. Figure 13 provides a comparison of the needed number of transmissions to complete a beam alignment procedure, for four different antenna configurations. These are from left to right; a dual polarized antenna configuration (one at HC and two at SNE), then a single four antenna element panel at the HC paired with four dual polarized antennas at the SNE, then introducing also four antenna elements at the SNE, and lastly introducing a 16-antenna element panel at the HC. The number of needed transmissions includes the signalling of the best wide beam from phase 1 (applicable to the wide-beam alignment procedure and the reciprocity-based procedure), and two reports (one from phase 2 and one from phase 3 for the narrow-beam-based procedure). It is worth noticing that, for antenna configurations with low number of antennas, there is no difference between the reciprocity-based and narrow-beam-based procedure, but this increases by 50% when doubling the number of antennas at the SNE, and almost triples with one and four antenna panels at the HC and SNE respectively. Continuing, the overhead increases to 5 times more transmissions per alignment when introducing the

single 16 element panel at the HC. The phase-1-based procedure on the other hand, maintains a low overhead of scaling only with the number of wide beams where it is assumed that one is used per antenna panel and a single report.
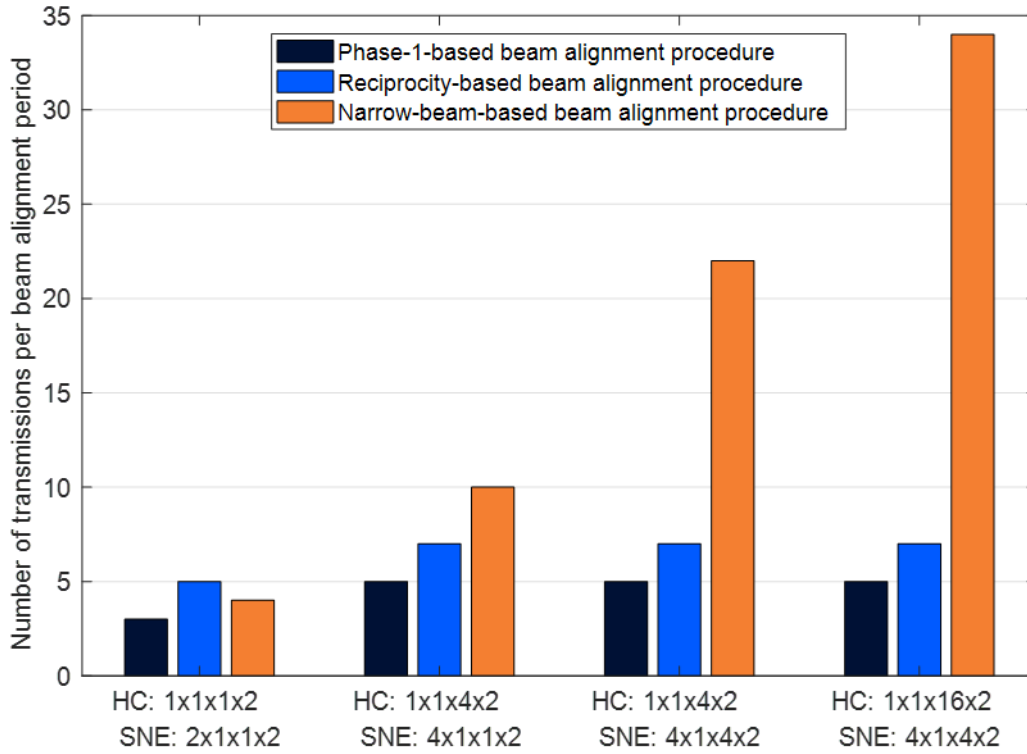
FIGURE 13. OVERHEAD COMPARISON OF THE THREE CONSIDERED BEAM ALIGNMENT PROCEDURES

Figure 14 shows a CDF of the instantaneous experienced throughput per user for the three considered beam alignment procedures. Similar antenna configurations as applied to Figure 13 are used here for each of the beam alignment procedures with the exception for the Phase-1-based procedure as this produce similar results also for larger panel sizes. From Figure 14, it is first observed that the trends observed in D3.1 also applies in this immersive education scenario, that is that there is a significant performance improvement to be obtained by using larger panels and at least two antennas on the SNE/LC. Unsurprisingly, the beam alignment scheme which produces the highest throughput is the narrow-beam-based procedure (blue) by an average throughput of 160Mbps with the 16-antenna element and 116Mbps with the 4-antenna element configuration. The reciprocity-based procedure achieves 130 Mbps and 81 Mbps for the same antenna configurations in comparison corresponding to a 19% and 30% degradation respectively.

At the 5'th percentile of the throughput CDF, the narrow beam only scheme provides a 60Mbps, and the reciprocal-based procedure at 30Mbps and 14Mbps corresponding to a 50% degradation respectively for the same antenna configurations. This likely occurs due to the risk of a sub-optimal narrow beam selection based on the selected wide beam.

It is worth to recall that the narrow beam only beam alignment procedure also needs 3 and 5 times as many transmissions per beam alignment procedure execution which becomes non-data carrying overhead (which is not subtracted in the throughput in Figure 14). This means, that despite of the 1x16

antenna configuration of the reciprocity-based procedure performs similarly to the 1x4 antenna configuration with the narrow-beam-based procedure, it still implies a lower overhead, and performs provides better throughput values on average.
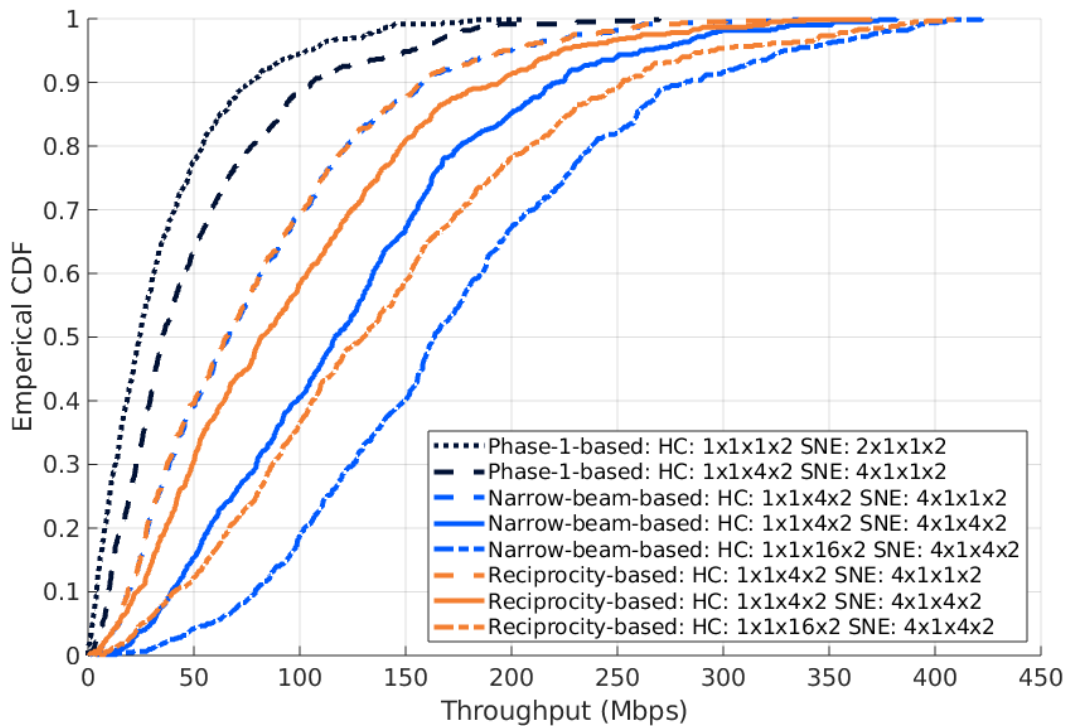
FIGURE 14. CDF OF THE INSTANTANEOUS THROUGHPUT COMPARING THE CONSIDERED BEAM ALIGNMENT PROCEDURES FOR DIFFERENT ANTENNA CONFIGURATIONS.

Figure 15 shows the throughput achieved when the SNE/LCs are moving at 2m/s (7.2kmph) as listed as the maximum velocity range in D2.2 for the immerse education use case, and of which can be compared with 3kmph in Figure 14, and with beam alignment periodicity of 20ms and 100ms. Starting with 2m/s, both the reciprocal-based procedure and the narrow-beam procedures are particular sensitive as they face a significant maximum throughput reduction as the periodicity decreases from around 370Mbps at 20ms periodicity, to 300Mbps at 100ms and then 170Mbps at 1000ms beam alignment periodicity. In general, at both maximum and average, the narrow-beam-based procedure is still superior, but the curves of the narrow-beam-based procedure at 100ms beam realignment, does get close to the reciprocal-based procedure at 20ms, in which case the two would have the same overhead. It may also be worth noticing that the phase-1-based procedure is hardly impacted by the reduced beam alignment periodicity or increased SNE/LC velocity. Comparing SNE velocities of 3kmph and 2m/s from Figure 14 and Figure 15 respectively for 20ms periodicities, there is hardly any difference (within 5Mbps on average), and with this velocity range representing the min and maximum from D2.2 [3], all schemes will be able to support the immersive use case velocities with the considered beam realignment periodicities. Increasing this periodicity further to e.g. 1s, causes both the narrow-beam-based procedure and reciprocal-based procedure to only slightly outperform the Phase-1 based procedure, indicating the optimum beam realignment periodicity for these schemes are within 100ms-1s range in the simulated scenario.
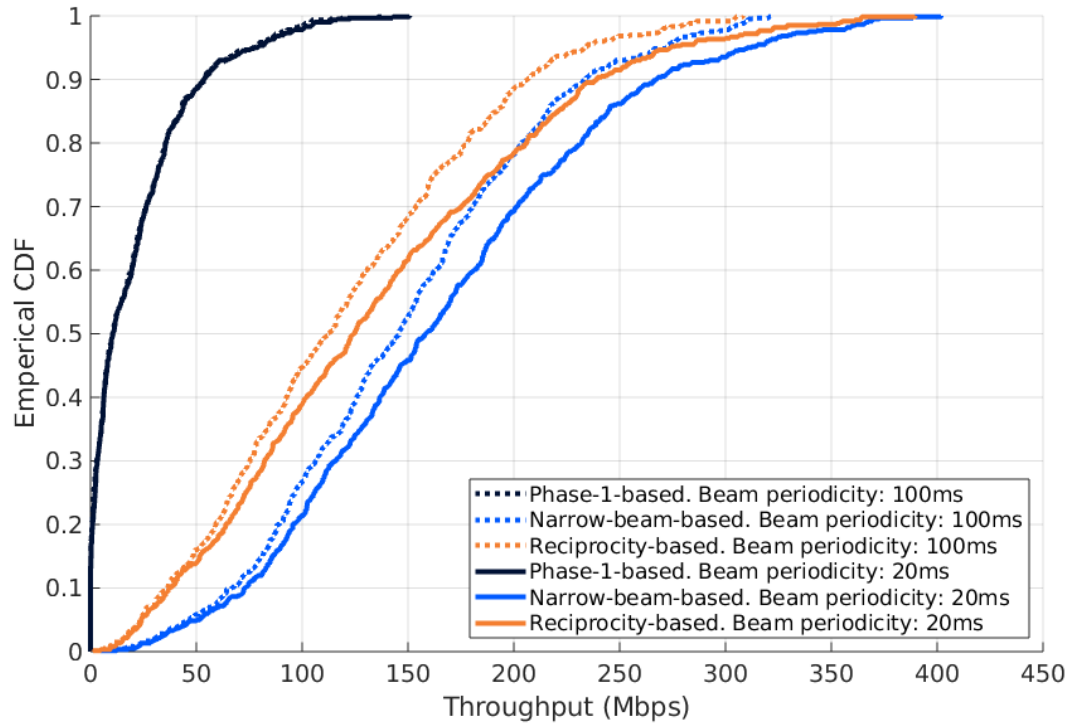
FIGURE 15. THROUGHPUT CDF STUDYING THE IMPACT OF SNE VELOCITY SENSITIVITY TO THE BEAM ALIGNMENT PERIODICITY. THE ANTENNA CONFIGURATIONS USED ARE HC: 1x1x16x2, SNE: 4x1x4x2.

3.1.4 Conclusion and future recommendations

In this section, the impact of beam forming into in-X subnetworks has been studied, considering the indoor classroom scenario. Following the study presented in D3.1 [2] which focused on the potential of beam forming without considering the beam alignment procedure or devices moving, this study here has included these aspects. The main take aways from the study and recommendations is as follows:

- A throughout enhancement in the order of 50% - 100% can be obtained by using aligned narrow beams contrary wide beam alignment only, for both reciprocity-based and narrow-beam-based beam alignment procedures. This makes it clear that despite of in-X subnetworks being predominantly short-range, beam forming can still be beneficial.
- Both the reciprocity-based and narrow-beam-based beam alignment procedures perform well under moving SNEs by the maximum considered velocity for immersive education.
- The 5x additional overhead of the narrow-beam-based alignment procedure generates an additional throughput on average between 19% and 30%.
- A beam alignment procedure focusing on reciprocity may be a good tradeoff between overhead and beam alignment gains, allowing also for reducing transmission power or increased channel capacity.

3.2 EXPLOITING METASURFACES FOR ULTRA-FAST AND LOW-COMPLEXITY BEAMFORMING

The exploration of higher frequency bands such as millimeter waves and sub-THz is expected to open new opportunities in terms of large available bandwidth, massive deployment of devices (e.g., sensors) [11], and the exploitation of the near-field propagation regime [12]. These opportunities can be realized through the adoption of extremely large-scale multiple-input multiple-output (XL-MIMO) systems, thus leading to unprecedented levels of communication and sensing performance, flexibility, and resolution [13][14]. However, we have reached the point where further advancing the capabilities of already

established technologies and incorporating new ones is hindered by issues related to practical implementation, hardware cost, and complexity, as well as excessive processing time and power consumption [15]. These issues often result in the so-called "digital bottleneck", posing a substantial challenge to the feasibility, scalability, and sustainability of future wireless systems, especially those with hardware and energy limitations such as subnetworks. For instance, performing ultra-fast (i.e., with low latency) beamforming becomes challenging if combined with the need for low power consumption and low complexity, especially in SNE devices.

A promising approach toward this goal is to transfer part of the signals' processing directly to the electromagnetic (EM) level [16][17][18]. This approach is known as over-the-air electromagnetic signal processing (ESP), also referred to as electromagnetic signal and information theory (ESIT). Typical EM processors are composed of several reconfigurable passive scattering elements and can be classified in two categories [18]: i) The scattering elements are located in the radiative region of the signal source(s). Reconfigurable intelligent surfaces (RISs) [19][20], addressed in D3.4, and self-conjugating metasurfaces (SCMs) [21], addressed in D3.1 and in the following, fall within this category; ii) The scattering elements are located in the reactive region of the signal source(s) so that inductive coupling takes place between them. In this category we find stacked intelligent surfaces (SIMs) [22] and dynamic scattering arrays (DSAs) [23], described in Section 3.2.2.

In subsection 3.2.1, further results on SCMs applied to low-complexity, ultra-fast MIMO beamforming in the presence of uncoordinated interference are reported. Novel results and application of DSAs as general-purpose EM signal processors to deal with energy consumption, complexity and latency reduction are addressed in subsection 3.2.2.

3.2.1   Self-conjugating metasurfaces for ultra-fast, low-complexity, and low-overhead beamforming

Classical MIMO techniques rely on complete channel state information (CSI) at both the transmitter and receiver to realize a proper beamforming, which involves computationally demanding operations like singular value decomposition (SVD) and significant overhead for training and signalling. While sub-optimal methods, such as V-BLAST and beam-scanning techniques, can reduce this burden, they often lead to performance-latency trade-offs or increased complexity, especially when large antenna arrays are used. These challenges are exacerbated at high-frequency bands like millimetre waves and sub-THz, where technological constraints limit signal processing flexibility, making traditional methods unsuitable for low-cost devices, such as SNEs.

To tackle the above challenging requirements, it has been proposed in Deliverable D3.1 [2] the adoption of (passive/active) modulating self-conjugating metasurfaces (SCM) at the SNE side to avoid RF/ digital-to-analog conversion (DAC) chains and signal processing as well as a blind self-configuring iterative scheme tailored to SCMs, taking advantage of its backscattering nature to obtain ultra-fast beamforming, joint communication and beamforming, no CSI estimation and signalling overhead.

The proposed solution is based on two key pillars [21]:
- The integration of a modulating SCM at the SNE, which enables the reflection of the incoming signal—namely, the interrogation signal—along its direction of arrival, while also allowing modulation of the reflected signal's phase at the electromagnetic level according to SNE data, thereby enabling low-complexity uplink communication.

- The design of an algorithm running on the HC capable of jointly demodulating the backscattered signal and deriving the optimal precoding/combining vector, i.e., enabling joint communication and beamforming. This is realized without any CSI estimation or prior knowledge, thanks to the retrodirectivity property of the SNE antenna. The algorithm also dynamically adjusts the beam orientation to accommodate changes in the SNE's position during the ongoing communication, thereby exhibiting tracking capabilities while communicating.

With respect to the study carried out in Deliverable D3.1 [2], subsequent activity has been focused on the analysis of the effect of static and dynamic clutter, the impact of hardware impairments such as non-reciprocity effects, the extension to a multi-SNE and multi-subnetwork asynchronous scenarios [24]. Specifically, it has been shown in [21] that the effect of clutter on the beamforming algorithm is binary: below a certain threshold, it has no impact, but above the threshold the algorithm fails, making the SNE unobservable. The threshold can be increased by increasing the number of antennas at the HC. While the extension of the proposed SCM-based solution to single-subnetwork with multiple SNEs in the presence of scheduled periodic traffic is straightforward because no packets collision takes place, in the presence of uncoordinated interference caused by asynchronous packets, ad hoc solutions must be designed. In the following, we summarize the main results obtained when applying the proposed scheme to ensure up-link communications in the presence of neighbour uncoordinated subnetworks generating inter-subnetwork interference and between many SNEs (e.g., sensors) generating asynchronous traffic with short packets in a grant-free manner within the same subnetwork (intra-subnetwork interference), which is typical of the industrial scenario, in particular, the Robot control and Unit test cell use cases, as defined in D2.2.
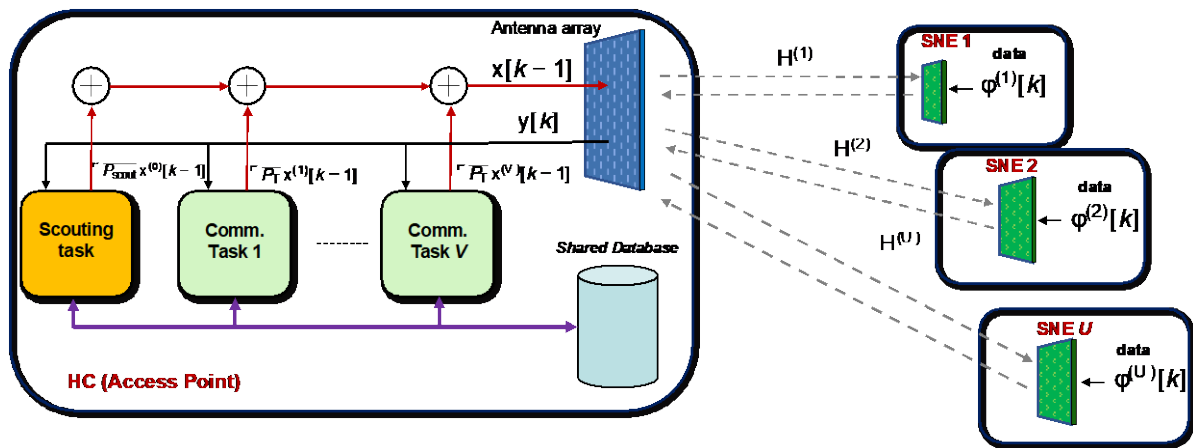
FIGURE 16 – TASKS RUNNING AT THE HC TO DETECT AND BEAMFORM MULTIPLE SNEs SIMULTANEOUSLY.

SCM-based communication in the presence of uncoordinated interference

In multi-SNE scenarios, such as in industrial use cases, the SCM-based beamforming procedure can serve as the foundation for a grant-free channel access method, enabling multiple SNEs to establish parallel communication with the HC without coordination. The same method can be also exploited in a scheduled access scheme in the presence of interference generated by uncoordinated neighbouring subnetworks. The studied method involves executing multiple parallel tasks at the HC to transmit interrogation signals and process SNEs' responses received by its antenna array as illustrated in Figure 16. This facilitates the simultaneous detection of new SNEs and the estimation of the optimal beamforming vectors directed toward them.

Specifically, a dedicated *Scouting Task* must continuously operate at the HC to detect new SNEs intending to transmit data. At each iteration, this task sends interrogation signals and processes backscattered responses to adjust beams toward newly detected SNEs using the SCM-based approach described in Deliverable D3.1 and in [21]. To prevent interference with ongoing communications with already-active SNEs, the interrogation signal must be transmitted in a channel subspace orthogonal to those used by the active links, thereby limiting the searching space. On the other side of the link, previously silent SNEs that need to transmit data respond by backscattering the interrogation signal toward the HC, leveraging the retrodirectivity property of their SCM. After a few iterations, once a beamforming vector converges to the (ideally) optimal configuration for a given SNE, the HC initiates a *Communication Task* for the corresponding SNE. In this task, the HC continues transmitting interrogation signals using the beamforming vector estimated during the *Scouting Task*, while the SNE transmits data by modulating the phase of the retrodirected signal. By operating in this manner, the HC establishes a true MIMO uplink communication with new devices in a few iterations. This process is blind and nearly optimal, eliminating the need for CSI estimation or time-consuming beam alignment schemes, thus requiring no computational burden on the SNEs.

The result is scalable scheme where complexity grows linearly with the number of SNEs and can handle asynchronous random transmissions of packets of any size without explicitly estimating the multi-device MIMO channel. This ensures near-zero latency and eliminates jitter, regardless of the number of connected devices. Further details are available in [21]. Finally, we note that estimating the optimal precoding vector ultimately corresponds to determining the angle-of-arrival of the backscattered signal when operating in line-of-sight (LOS) condition. This information can then be utilized to localize the SNEs within the environment. This aspect is elaborated in [25].
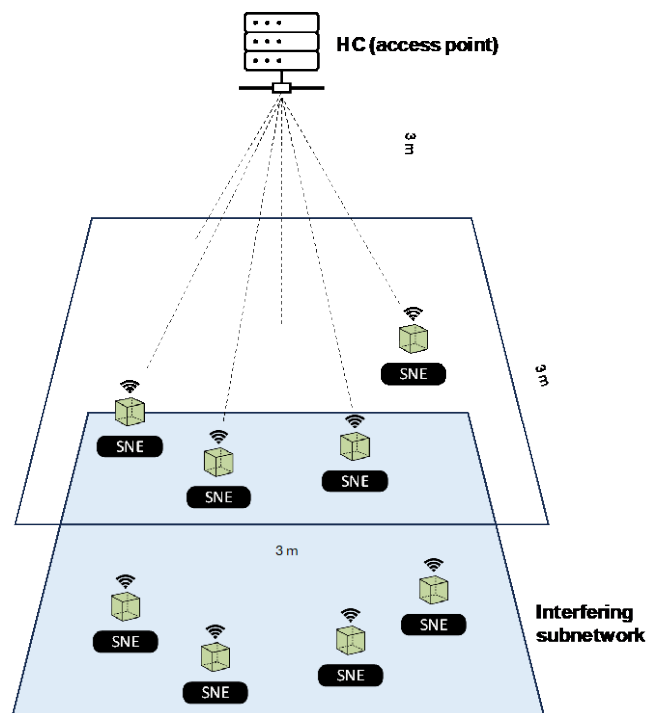
FIGURE 17 –SCENARIO WITH SNEs GENERATING INTRA-SUBNETWORK INTERFERENCE (SAME SUBNETWORK) AND SNEs GENERATING INTER-SUBNETWORK INTERFERENCE (ADJACENT SUBNETWORKS).

Main Results

As anticipated, the proposed method has been applied to address two situations: i) asynchronous multiple access within a subnetwork with grant-free characteristics (intra-subnetwork interference); ii) coexistence of adjacent uncoordinated subnetworks (inter-subnetwork interference) (See Figure 17). In both situations, the subnetwork(s) involves a full-duplex transceiver at the HC, equipped with a conventional antenna array, which is designed to iteratively transmit an interrogation signal and receive data from SNEs equipped with a modulating SCM. The scenario considered applies to the Robot control and unit test cell use cases defined in Deliverable D.2 (Work package WP2). Interfering intra and inter subnetwork asynchronous traffic is modelled as Poissonian with exponential inter-arrival time distribution of mean T_p/G, where G is the offered traffic defined as the mean number of generated packets per packet time, and T_p is the packet time. Very short packets of 100 bytes are considered. The main simulation parameters are reported in Table 2. As suggested in Deliverable D2.3, for the FR2 band the channel model is the 3GPP Indoor Factory (InF) with parameters modified according to the measurement campaigns carried out in WP2 and reported in Table 25 of D2.3.

TABLE 2: SIMULATION PARAMETERS

Parameter	Value
Carrier frequency	28 GHz
Bandwidth (single carrier)	10 MHz
Symbol time	0.1 us
Packet length	100 Bytes
Guard symbols	16 (2 Bytes)
Total TX power	-20 dBm
SCM gain	0 dB
HC antennas	20 x 20 (10 x 10 cm)
SCM antennas (SNE)	10 x 10 (5 x 5 cm)
HC noise figure	3 dB
Path-loss exponent	1.7 (LOS) – from Table 25 Deliverable D2.3
k-factor	1 - from Table 25 Deliverable D2.3
Delay spread	18 ns - from Table 25 Deliverable D2.3
SIC SINR threshold	3 dB

We first investigate the beamforming capability of the proposed method during the Scouting Task. More investigations can be found in D3.1 and [24].

In Figure 18 (left), the empirical cumulative distribution function (CDF) if the beamforming convergence time during the Scouting Task is shown in case of free space propagation and 3GPP channel model for random initial precoding vector. While in free space, the beamforming process is almost deterministic, in realistic propagation conditions in the presence of multipath the convergence time fluctuates but less than 16 symbols times (1.5 us) for 99% of the cases. This suggests that a guard time of 16 symbols set to zero, inserted at the beginning of each packet and followed by the payload of 100 bytes, gives sufficient time to the Scouting Task to converge to a stable beamform and detect the packet. It is worth noting that the overhead is only 2% of the packet's length despite the short duration of the packet, and thanks to the proposed beamforming technique, the MIMO link can be established in an extremely short time without the need for signalling and CSI estimation procedures.
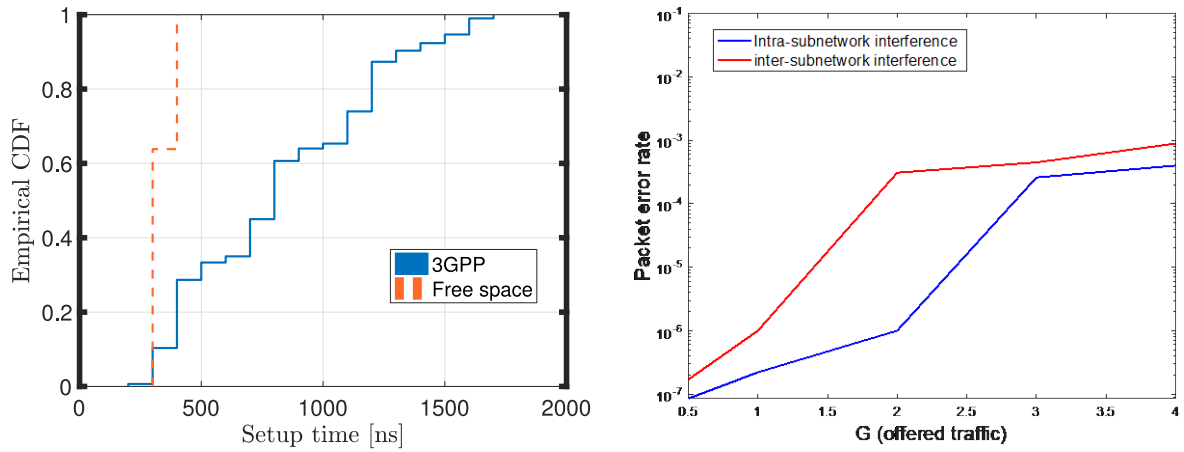
FIGURE 18 – EMPIRICAL CDF OF THE BEAMFORMING CONVERGENCE TIME (LEFT); PACKET ERROR RATE AS A FUNCTION OF THE OFFERED TRAFFIC (RIGHT).

The impact of the intra and inter-subnetwork interference on the packet error rate as a function of the offered traffic load G is reported in Figure 18 (right). The effect of the interference in case of packet collision is mitigated by the joint combination of beamforming (spatial filtering) and the processing adopted at the receiving HC. The numerical results have been obtained supposing a successive interference cancellation (SIC) technique is adopted at the HC. In the intra-subnetwork scenario, SNEs are supposed to be uniformly distributed within a 3x3m^2 area at 3 m from the HC, whereas in the inter-subnetwork scenario, the interference comes from SNEs belonging to an adjacent subnetwork located at 6m from the useful HC. In both cases, the total offered traffic is G.

Numerical results indicate that the proposed method is less sensitive to interference when it comes from the same subnetwork. This can be attributed to the higher angle span of closer SNEs with respect to far SNEs from adjacent subnetworks that allows for a more effective spatial filtering during the beamforming. A packet error rate $<10^{-6}$ is obtained for G<2 with a total latency <100 us (a few microseconds for beamforming + the packet duration).

Comparison with other MIMO Implementations

To gain a complete understanding of the advantages and disadvantages of SCM-enabled MIMO communications, a comparison with alternative solutions is necessary, and it is provided below.
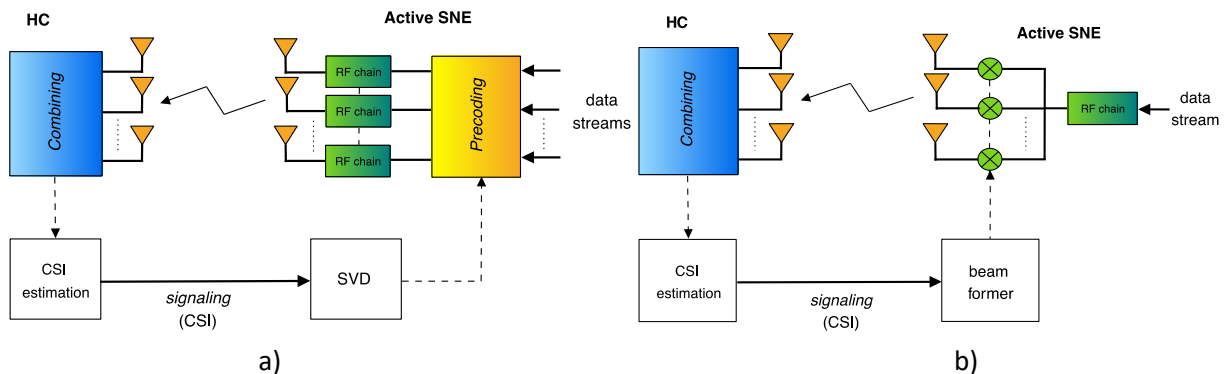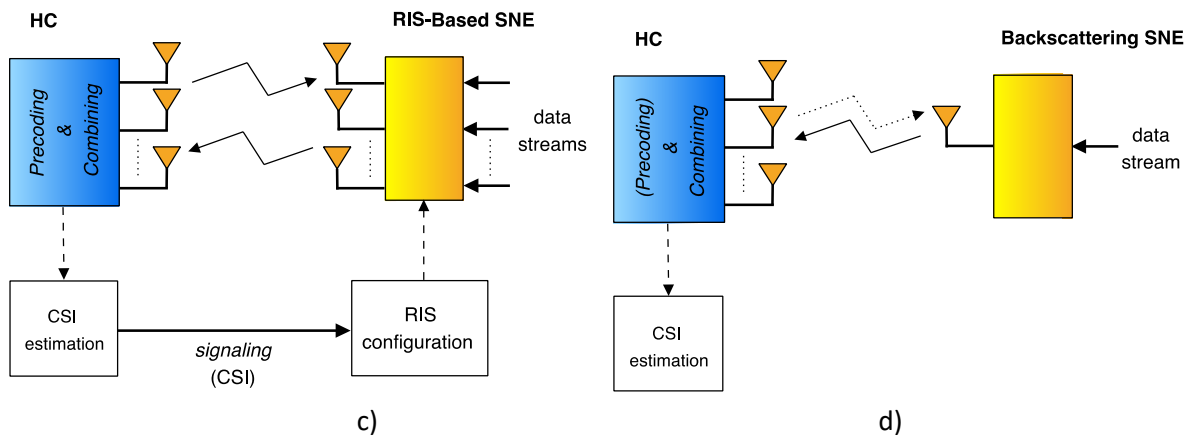
FIGURE 19 - COMPARISON OF MIMO IMPLEMENTATIONS: A) FULL DIGITAL MIMO; B) SINGLE-RF MIMO (ANALOG BEAMFORMER); C) RIS-BASED MIMO BACKSCATTER; D) SINGLE ANTENNA (ACTIVE OR BACKSCATTER).

- Full Digital MIMO (Figure 19a): The full digital MIMO scheme includes multiple antennas at both the HC and the SNE, with each antenna being equipped with its own RF chain.
  Pros wrt the SCM-based scheme: It offers multiplexing gain and a more favourable link budget, as the SNE does not operate in backscatter mode.
  Cons wrt the SCM-based scheme: It necessitates multiple RF/DAC chains at the SNE along with a digital processing unit for computing precoding and combining vectors, as well as demodulating data. For instance, assuming that both the transmitter and receiver are equipped with the same number N of antennas, the computational complexity of the SVD scales proportionally to the third power of N. Additionally, it involves a CSI estimation process, resulting in signalling overhead and increased latency, whose outcome must be available at both the SNE and HC. Moreover, the SNE is intrinsically power-hungry (active transceiver).

- Single-RF MIMO (Figure 19b): A single-RF MIMO system, such as the one proposed in [26], or a conventional analog-beamformer scheme, considers beamforming operated in the analog domain at the SNE (or a similar strategy leading to a single RF chain). Thus, it entails significantly lower complexity than a full digital MIMO.
  Pros wrt the SCM-based scheme: Similarly to the SCM-based scheme, it performs beamforming resulting in a single-layer optimal communication (thus, without exploiting spatial multiplexing). However, since it employs active devices, it offers a more favourable link budget.
  Cons wrt the SCM-based scheme: It still necessitates a CSI estimation process, leading to signalling overhead, and the estimation must be available at both the SNE and HC. Similarly to full digital MIMO, the SNE is power-hungry (active transceiver).

- RIS-based MIMO Backscatter (Figure 19c): RIS-based backscatter communication systems incorporate a reconfigurable intelligent surface (RIS) at the SNE side, aiming to effectively reduce SNE complexity and minimize power consumption [27].
  Pros wrt the SCM-based scheme: Since a RIS is composed of multiple cells, different data streams can be transmitted at the same time, thus exploiting spatial multiplexing, and beamforming gain can be exploited to improve the link budget.
  Cons wrt the SCM-based scheme: To configure the RIS, CSI estimation is required in addition to a dedicated control channel between the RIS and the HC. Moreover, the RIS configuration speed

may become a bottleneck, limiting the system's ability to operate in highly dynamic environments and deliver low-latency beamforming

- Single-antenna (Active or Backscatter) (Figure 19d): To avoid the CSI estimation process and achieve an SNE with complexity comparable to the SCM-based scheme, an active or backscatter-based single-antenna SNE can be employed [28].
  Pros wrt the SCM-based scheme: In the case of active transmission the link budget is more favorable, at the expense of increased power consumption.
  Cons wrt the SCM-based scheme: The single-antenna scheme, although single-layer like the SCM-based scheme, does not benefit from the beamforming and diversity gains. This is a primary issue when considering a backscatter-based solution working at high frequency, such as mmWave, which has an unfavourable link budget.

3.2.2   Dynamic scattering arrays for ultra-low power and complexity beamforming

We illustrate the dynamic scattering array (DSA), a versatile 3D structure capable of performing joint wave-based computing and radiation by moving the processing from the digital domain to the EM domain, thus minimizing the required number of radiofrequency (RF) chains, which are the main responsible for energy consumption, complexity and latency in multi-antenna wireless systems. The DSA represents a possible enabler for the technology component TC5 (Analog/hybrid beamforming/beamfocusing) to tackle simultaneously the energy consumption, computational complexity, and latency reduction KPIs in beamforming procedures and it can be applied to HC/6G-BS nodes in all the subnetwork categories defined in 6G-SHINE, as described in D2.2.
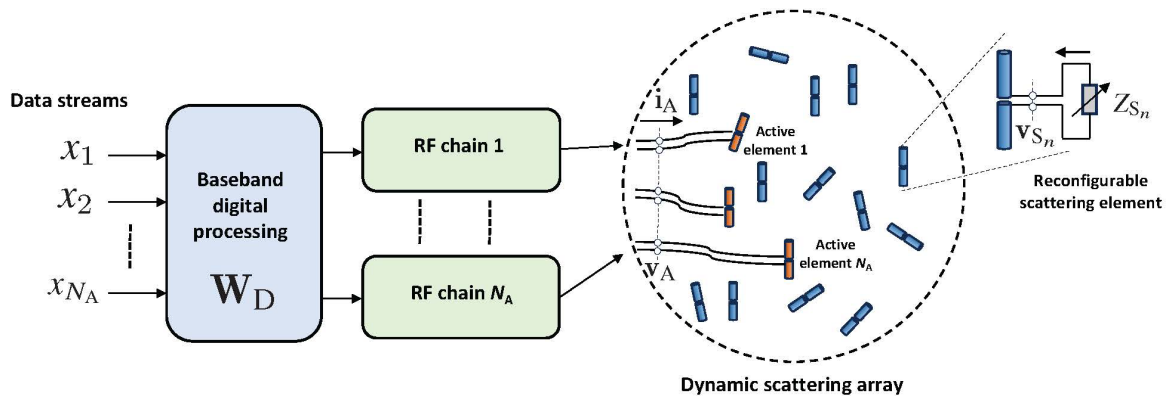
FIGURE 20 – ILLUSTRATION OF A DSA WITH N_A RF CHAINS AND ACTIVE ANTENNAS.

A DSA consists of a limited number of active antenna elements, each associated with an RF chain, surrounded by a cloud of reconfigurable passive scatterers that interact with each other in the reactive near field so that EM processing and radiation are performed jointly "over the air" [23]. A DSA can be considered as a generalization of reactively controlled arrays, originally proposed in [24], and their evolution toward electrically steerable passive array radiators (ESPARs), in which only a single RF chain is considered [29]. Specifically, as illustrated in Figure 20, a DSA is composed of N_A active antenna elements surrounded by N_S reconfigurable passive scatterers. The nth element (active or passive) of the DSA, with $n = 1,2 \dots N, N = N_A + N_S$. The nth scatterer, with $n = 1,2 \dots N_S$, is itself an antenna element terminated by a reconfigurable impedance load $Z_{S_n}(\theta_n)$. To minimize the power dispersion, the resistive part of the load must be negligible so that it is mainly reactive, i.e., $Z_{S_n}(\theta_n) = j\theta_n$, where $\theta_n \in$

$\mathbb{R}$ is the reconfigurable reactance. Denote with $\theta = [\theta_1, \theta_2, \ldots, \theta_{N_s}]^{\mathrm{T}}$ the set of reconfigurable reactances and with $Z_S(\theta) = \mathrm{diag}\left(Z_{S_1}(\theta_1), Z_{S_2}(\theta_2), \ldots, Z_{S_{N_s}}(\theta_{N_s})\right)$ the corresponding set of loads.

The DSA can be modelled as a linear N-port network. Specifically, in a compact description, we define $v_A \in C^{N_A \times 1}$ and $i_A \in C^{N_A \times 1}$ the complex voltage and current envelopes (in the following denoted simply as voltages and currents) at the ports of the N_A active antennas (i.e., the input ports of the DSA), and $v_S \in C^{N_S \times 1}$ and $i_S \in C^{N_S \times 1}$ the voltages and currents at the ports of the N_S scatterers. We collect all the currents and voltages at the N ports of the DSA in the vectors

$$i = \begin{bmatrix} i_A \\ i_S \end{bmatrix} \quad v = \begin{bmatrix} v_A \\ v_S \end{bmatrix}$$

At the scatterers' ports it is $v_S = -Z_S(\theta)i_S$. All the interactions between the elements of the DSA are captured by the impedance matrix $Z \in C^{N \times N}$, which does not depend on the reconfigurable loads, and relates the voltages and currents of the N ports as $v = Zi$. In particular, the (n, m)th element of Z represents the mutual coupling coefficient between the nth and mth elements obtained as the ratio between the open-circuit voltage observed at the m-th port and the excitation current applied to the nth port supposing that the remaining ports are kept open (no current flow). It can be analytically for simple structures, like Hertzian or half-wave dipoles, or through EM simulations for more complex structures. The N_A active elements are connected to N_A RF chains, including digital-to-analog converters (DACs) and up-conversion stages, through a $2\,N_A$-port power matching network whose purpose is to maximize the power transfer between the RF chains and the DSA. The matching network depends on the DSA's input impedance and hence on the parameters θ. Therefore, the matching network should be reconfigured for a given DSA configuration to maximine the power transfer. It is also foreseen an optional baseband linear processing block implemented in the digital domain (digital precoder), described by the matrix W_D, whose input is represented by the information vector x to be transmitted. The matrix W_D relates the transmined information vector x and the voltages v_T at the input of the matching nctwork i.e., $v_T = W_D x$. The reconflgurable reactances θ along with the elements of matrix W_D represent the set of parameters to be oprimized to obtain the desired processing functionality. The relationship between the information vector x and the total current i flowing in the DSA is:

$$i = W_{\mathrm{EM}}(\Theta)W_D x$$

where $W_{\mathrm{EM}}(\Theta)$ depends on it is $Z_S(\theta), Z$ (details can be found in [23]) and accounts for the EM-level signal processing operated by the reconfigurable DSA as a function of the parameters θ.

Consider now K test points located in positions t_k. with $k = 1.2\ldots, K$. The useful component (i.e., without noise) of the received signal at the test positions is

$$y = H_c i = H_c W_{\mathrm{EM}}(\theta)W_D x = H(\theta, W_D)x$$

where H_c is the transimpedance matrix of the radio channel accounting for the propagation effects and $H(\theta, W_D)$ is the end-to-end baseband equivalent channel matrix as commonly defined in signal processing. For a given propagation scenario characterized by the transimpedance matrix H_c and K test points, suppose we fix a specification H_{opt} on the desired end-to-end channel matrix, i.e., the desired signal processing task. The reconfigurable parameters of the DSA and the digital precoder can be designed by solving the following constrained optimization problem:

$$\underset{\theta, W_{D, \alpha}}{\mathrm{minimize}} \| \alpha H_c W_{\mathrm{EM}}(\theta)W_D - H_{\mathrm{opt}} \|_{\mathrm{F}}$$
$$\text{s.t. } \|W_D\|_{\mathrm{F}}^2 = R N_A$$

Where R is the internal resistance of the RF chain, the scalar $\alpha \in \mathbb{R}$ accounts for the possible lack in the link budget to achieve H_{opt} that has to be eventually coped with increased transmitted power.

In the following, we present two examples of signal processing tasks that can be realized through a proper optimization of the DSA. The following parameters have been considered in the numerical evaluations if not otherwise specified: $f_0 = 28\text{GHz}, R = 50\,\text{Ohm}$. Hertzian dipoles with vertical polarization have been considered for an analytical evaluation of the impedance matrix Z.

Beam Forming and Superdirectivity – We first illustrate the beam forming capabilities of the DSA using one RF chain ($N_A = 1$ active antenna element) and the deployment of the scattering elements according to $L = 5$ concentric circles with incremental radius of step with $\Delta_L = \lambda/4$, being λ the wavelength, corresponding to $N_S = 121$ scattering elements. The disk-like structure appears very appealing especially for its use in HC devices (e.g., access point) thanks to its circular symmetry. Obviously, other 2D or 3D structures can be considered as well depending on the specific application. The target end-to-end channel matrix H_{opt} with $K = 108$ test points deployed on the $x - z$ plane was considered for 4 different steering angles, $0°, 30°, 60°$, and $90°$, respectively (in this case W_D is a scalar and no digital processing takes place).
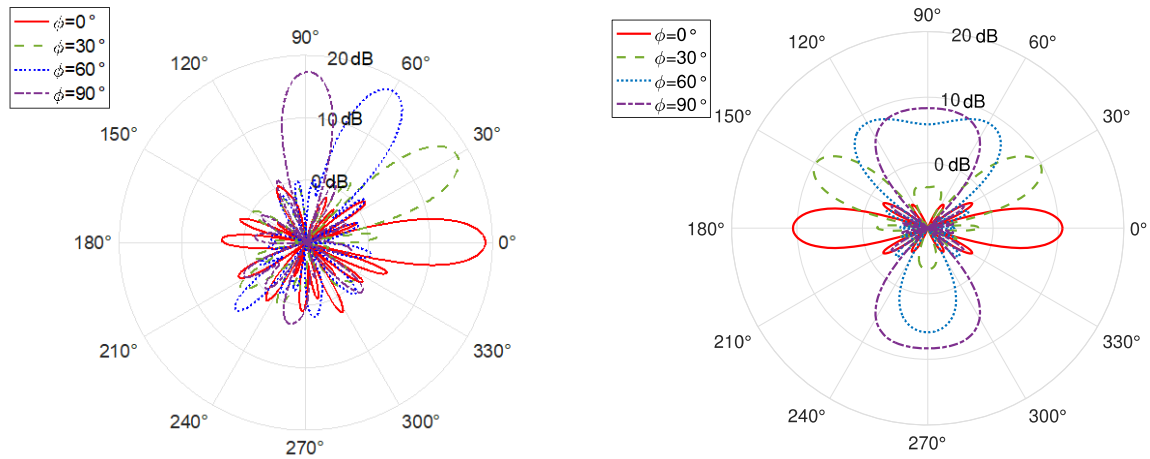
FIGURE 21 – (LEFT) RADIATION DIAGRAM OF A DISK-SHAPED DSA, EXHIBITING A SUPER DIRECTIVE CHARACTERISTIC. (RIGHT) RADIATION PATTERN OF A CONVENTIONAL ULA WITH N_A = 6 ACTIVE ELEMENTS.

In Figure 21-left, the radiation diagrams for the 4 steering angles are shown. As it can be observed, the gain of the DSA is independent of the angle, and it is about 18.6 dB. In addition, limited back radiation is obtained without the need to insert a ground plane that would impede the steering in the angle range $[90° - 270°]$. For the sake of comparison, we consider the radiation diagrams obtained using a standard full-digital ULA with the same aperture of the DSA (3.2 cm diameter) in the $x - z$ plane. The corresponding radiation diagrams are reported in Figure 21-rigth. This array requires $N_A = 6$ active antennas compared to only one active antenna of the DSA. As well-known, in ULAs the gain degrades when approaching $90°$ from a maximum value of 10.6 dB (about $10 \log_{10}(N_A G_d)$, with $G_d = 3/2$ the gain of the Hertzian dipole) and symmetric back radiation is present. The proposed DSA exhibits superdirectivity capability with an additional gain of 8 dB. Contrary to standard arrays, where superdirectivity is obtained only in the end-fire direction, here notably superdirectivity is equally obtained in all steering directions.
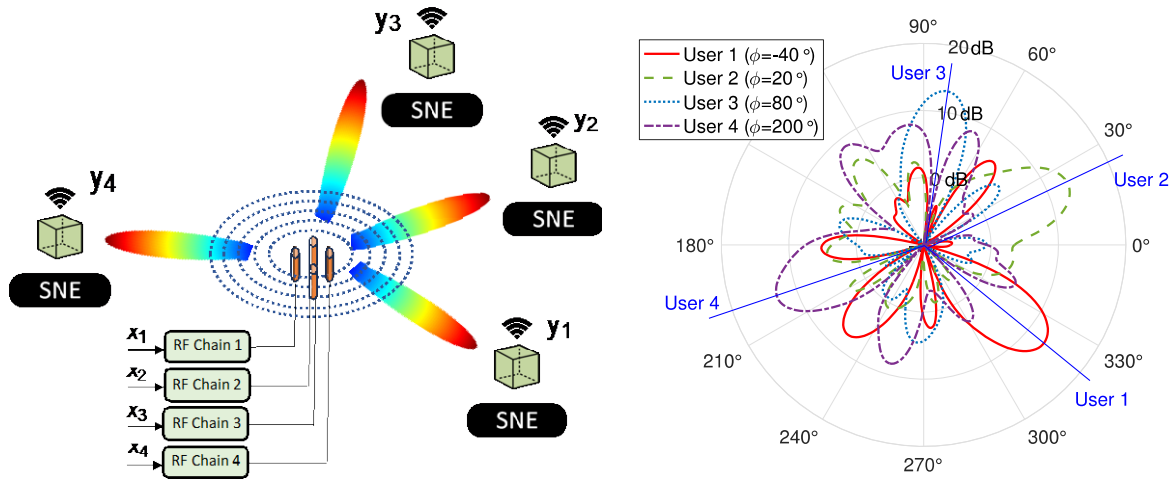
FIGURE 22 – FOUR-SNE DOWNLINK MISO WITH A DSA (LEFT) – RADIATION DIAGRAM OF A ZERO-FORCING FOUR-SNE MISO SYSTEM BASED ON A DSA (RIGHT).

Multi-SNE MISO downlink – In this second example, we consider a multi-SNE multiple-input single-output (MISO) downlink scenario shown in Figure 22-left, where a HC device is equipped with a DSA with N_A = 4 active antennas and N_A RF chains to communicate simultaneously with $K = N_A$ single-antenna SNEs located at angles −40∘, 20∘, 80∘, and 200∘ in the $x − z$ plane, by maximizing the signal-to-interference noise ratio (SINR) at each SNE according to the zero-forcing criterium. Given the channel transimpedance matrix $\mathrm{H_c} = \left[\mathrm{h_{c1}; h_{c2}; …; h_{c}}_{N_A}\right]$, where $\mathrm{h_{ck}}$ is the $N \times 1$ channel vector associated with the k th SNE, the precoding matrix is $\mathrm{V} = \beta \mathrm{H_c^\dagger}$, where β is a normalization factor to make V unitary, i.e., $\mathrm{V^H V = I}_N$. This is equivalent to setting $\mathrm{H_{opt}} = \mathrm{H_c H_c^\dagger}$ in the optimization problem in (25), corresponding to a DSA and a digital precoder satisfying

$$\mathrm{W_{EM}}(\hat{\theta})\mathrm{W_D} = \mathrm{V}$$

The corresponding radiation diagrams are shown in Figure 22-right. The coupling between users is impressive (< −100 dB) at the expense of increased leakage in unwanted directions and a reduction of the gain of some dBs compared to the pattern in Figure 21-left corresponding to a single-SNE link. Further examples of EM signal processing tasks, such as MIMO EM precoder, using a DSA can be found in [23]. Extension to frequency-selective processing and more realistic models for the reconfigurable loads has been introduced in [30].

Compared to a full digital MIMO implementation with *N* antennas, there is a reduction of at least a factor of *N/NA* in power consumption using a DSA. In fact, in case of reconfigurable scatterers using varactor diodes, their power consumption is <100nW, which is negligible with respect to the rest of the system. In the examples reported above, this corresponds to a factor of about 30 times less power consumption. In addition, at the EM level, the processing is performed at the speed of light, i.e., with near-zero latency. Compared to hybrid digital-analog solutions employing a *N* phase shifters associated with each RF chain. i.e., a total of $N \cdot N_A = 484$ phase shifters, the advantage depends on the specific technology adopted for the phase shifters. From the power consumption viewpoint, no significant difference is present if passive vector-sum phase shifters are used [31]. However, these shifters introduce a very large insertion

loss (typically 10-15 dB) that must be compensated by a corresponding increase of the transmitted power. Negligible insertion loss is possible using active phase shifters [32], but each of them typically requires about 10-15 mW at millimeter waves, resulting in an increased power of about 4.8 Watt in the example above.

3.2.3   Summary and recommendations

Summarizing, the proposed scheme has the following characteristics with respect to the project KPIs and KVIs:

- *Improved environmental sustainability thanks to lower energy consumption enabled by the high-energy efficient sensors:* The energy consumption at the SNE, being based on backscattering technology, is expected to be in the same order of magnitude as traditional backscattering systems (e.g., RFID) but with the advantage of MIMO and beamforming capabilities. Specifically, backscattering technologies have a power consumption <100uW compared to several mW required by conventional active systems (energy efficiency gain of at least x10 at the SNE side). Therefore, SNEs might eventually be power supplied using energy harvesting techniques (no battery).
- *Reduction of computational complexity:* Despite the SCM can be composed of hundreds of elements, no RF chains, signal processing, beam alignment schemes are needed at SNE side, so that no computational effort is required at the SNE. At the HC, the computational effort increases linearly with the number of antennas N compared to N^3 required to compute the SVD of the channel. In addition, no CSI estimation is needed thus further reducing the computational complexity.
- *Signalling overhead reduction*: Ultra-fast beamforming is obtained with no signalling and minimal overhead (2%) with short packets (100 Bytes).
- *Reliability:* Results indicate a packet error rate $<10^{-6}$ for a latency <100 us (a few us for initial beamforming + 80 us for packet transmission) when the traffic load is G<2. Future investigations could be addressed to further improve the reliability.

In conclusion, the approach requires minimal computational effort and introduces negligible latency, enabling an efficient communication link that swiftly adapt to dynamic conditions, such as those induced by mobility. Additionally, it minimizes hardware complexity and power consumption, enabling low-cost MIMO-enabled SNEs, making it a practical solution for subnetworks operating at high frequency while maintaining high-performance, low-latency communication.

For what the DSA is concerned, it provides unprecedented flexibility in managing the EM field by reducing or circumventing digital processing at the baseband, thereby surpassing the flexibility offered by classical MIMO and recent SIM implementations. It worth to remark that the performance of any SIM is constrained by the final layer surface of the structure, therefore it is upper bounded by that one can obtain by replacing the SIM with a full digital planar or linear array with the same layout and number of elements. The reduction in RF chains and digital computation needs enabled by the use of DSAs facilitates the realization of energy-efficient, low-latency, and cost-effective holographic MIMO systems, then addressing the sustainability concerns of future wireless subnetworks. In the set up analysed, a power reduction of x30 compared to an equivalent full digital implementation has been achieved with near-zero latency.

3.3 SUB-THZ TRANSCEIVER ARCHITECTURES & ANALYSIS FOR SUBNETWORKS

Previous sections have outlined the advantages of using the mmWave spectrum for in-X subnetworks, including an evaluation of various antenna panel performances and the proposal of self-conjugating metasurfaces and dynamic spectrum arrays for low complexity SNEs. In this section, the focus shifts to an even higher frequency range—the sub-THz band (100–300 GHz)—which emerges as a promising candidate for subnetworks due to the abundance of available spectrum (several GHz), which allows for the support of ultra-high data rates. The smaller wavelengths associated with sub-THz frequencies enable the use of compact antenna designs, where the large antenna array and the beamforming technique increase the link budget to ensure high data rate transmission. In the previous deliverable D3.1 [2], we focused on the throughput and power analysis for sub-THz transceiver with different beamforming architectures at HC devices for the indoor interactive gaming scenario. In this deliverable, we explore power co-optimization for the power consumption for the access point (AP) and head-mounted displays (HMDs) in the subnetwork system using sub-THz communication under different user experience levels considering different video resolutions and codec compression rates.

3.3.1 Methodology

The power consumption of the HMD is modelled considering analog components and digital functions. We focus on optimizing the power consumption in the downlink since the throughput in the uplink direction is negligible. A power model tailored to sub-THz frequencies is considered along with different hybrid and digital beamforming architectures [33]. In this work, fully digital beamforming architectures are considered for AP and HMD but the analysis can be extended to hybrid architectures as well. We denote the AP power consumption by P_{TX} computed by using the model

$$P_{TX} = N_t(P_{PA} + P_{mix} + P_{LO,d} + P_{DAC}) + P_{LO} + P_{bb,dig}(N_t)$$

where P_{PA}, P_{mix} $P_{LO,d}$ P_{DAC} and $P_{bb,dig}$ is the power consumption of the PA, mixer, local oscillator (LO) distribution, digital-to-analog converter (DAC), LO, and digital baseband, respectively. N_t is the number of antennas at the transmitter side. PA power consumption P_{PA} is a function of the PA output power $P_{PA}^{(out)}$. The TX digital baseband section includes IFFT, baseband filtering, MIMO processing, modulation, and channel encoding operations.

On the other hand, the receiver's power consumption is more critical since it is a battery powered device with limited resources. We estimate the power consumption of the receiver as

$$P_{RX} = N_r(P_{LNA} + P_{LO,d} + P_{ADC}) + P_{LO} + P_{bb,dig}(N_r)$$

where P_{LNA}, and P_{ADC} is the power consumption of the LNA, and analog-to-digital converter (ADC), respectively. N_r is the number of antennas at the receiver side. In particular, the RX digital baseband section includes FFT, baseband filtering, MIMO processing, demodulation, and channel decoding operations.

In order to improve energy efficiency, the system can use duty cycling. It transmits short bursts at a high throughput $T(\gamma_k, B)$ larger than the throughput T_{req}, and switches to sleep mode in-between [34]. The system is hence active a fraction $D = \frac{T_{req}}{T(\gamma_k,B)}$ of the time and consumes only a fraction D of the peak power on average. The average power consumption of the transmitter and receiver can be calculated considering the duty-cycling ratio as

$$\{P_{TX,avg}, P_{RX,avg}\} = D\{P_{TX}, P_{RX}\}$$

The objective is to identify the AP and HMD architecture parameters that jointly minimize the $P_{TX,avg}$ and $P_{RX,avg}$ considering the throughput requirements with respect to different video compression rates. The number of AP antennas N_t, the number of HMD antennas N_r, and operating bandwidth are the architecture parameters of interest. Therefore, we write the following optimization problem with the throughput constraint:

$$\{B^\star, N_t^\star, N_r^\star\} = \underset{\{B,N_t,N_r\}}{arg\ min}\left(\beta P_{RX,avg} + (1-\beta)\frac{P_{TX,avg}}{P^{ref}}\right)$$
$$T(\gamma, B) \geq T_{req}$$

where $\beta \in [0,1]$ is a tuning parameter that gives priority to minimization of $P_{RX,avg}$ or $P_{TX,avg}$, P^{ref} is the normalization factor. In particular, $\beta = 0$ corresponds to pure AP energy efficiency optimization while $\beta = 1$ corresponds to pure HMD power consumption optimization. The normalization factor P^{ref} is set so that $P_{RX,avg}$ and $P_{TX,avg}$ are scaled in a way that they are comparable. Next, we present the simulation results where the impact of different VR realizations on the AP and HMD architectures and the computational load-off are analysed.

TABLE 3: UPDATED PARAMETERS FOR SUB-THZ TRANSCEIVER POWER ANALYSIS SIMULATION FOR XR INDOOR GAMING SCENARIOS.

Parameters	Value
Bandwidth [GHz]	{1,2,4,8,10,16}
Number of Tx antennas N_t	{4,8,16,24,...,128}
Number of Tx antennas N_r	{1,4,8,16}
Number of channel realizations	500
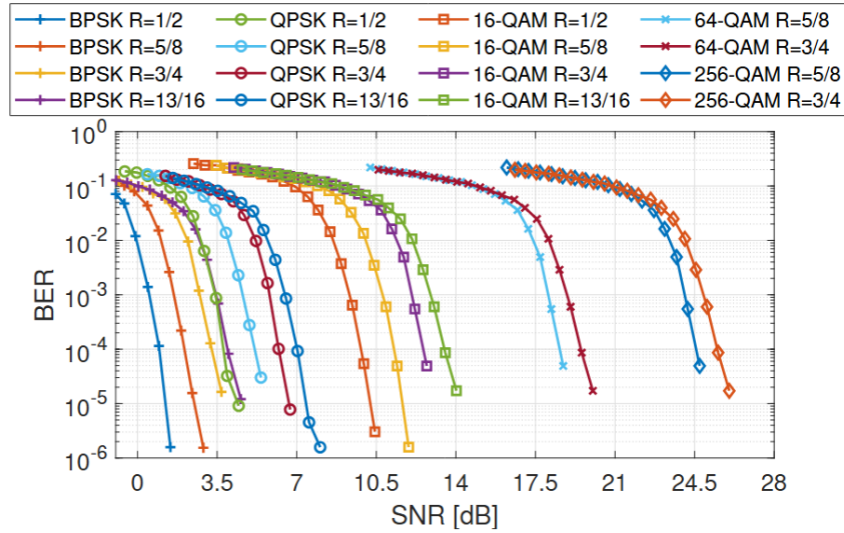
FIGURE 23. BER PERFORMANCE OF THE LDPC CODE WITH BLOCK LENGTH 672 FOR BPSK, QPSK, 16-QAM, 64-QAM, AND 256-QAM MODULATIONS UNDER AWGN CHANNEL.

Figure 23 depicts the BER versus SNR performance of the IEEE.802.11ad/ay [35] compatible Low Density Parity Check (LDPC) code with block length 672 bits for BPSK, QPSK, 16-QAM, 64-QAM, and 256-QAM modulations under additive white Gaussian noise (AWGN) channel for 4 different code rate. The required SNR $\overline{\gamma_i}$ for MCS_i is selected targeting 10^{-5} BER.

Throughput requirement T_{req} for different VR generations with varying data compression schemes are listed in Table 4. Please note that higher compression rates lead to higher latency and additional power consumption considering the video decompression.

TABLE 4: THROUGHPUT REQUIREMENT T_{req} FOR DIFFERENT VR USE CASES

Requirement	4K-VR	8K-VR
Video Resolution	3840 × 1920	7680 × 3840
[1:1] Uncompressed	10.62 Gb/s	63.7 Gb/s
[5:1] Ultra low latency compression	2.12 Gb/s	12.74 Gb/s
[20:1] Low latency compression	530 Mb/s	3.18 Gb/s
[300:1] Lossy latency compression	35 Mb/s	210 Mb/s

We identify the favourable architecture configurations that minimize the joint power consumption for each VR generation considering (i) achieved throughput per user, (ii) AP power consumption, and (iii) HMD power consumption. We set the scaling value P^{ref} depending on the compression rate. We report the average power consumption of the HMD based on the average throughput per user. On average, we expect all users to experience the same throughput due to the high number of channel realizations, hence the analysis is valid for all users in the system.

3.3.2 Simulation result

Figure 24 illustrates the average throughput versus the number of transmit antennas N_t for varying bandwidth B and different number of receive antennas N_r. The increase of N_t, N_r and B leads to higher average throughput per user.
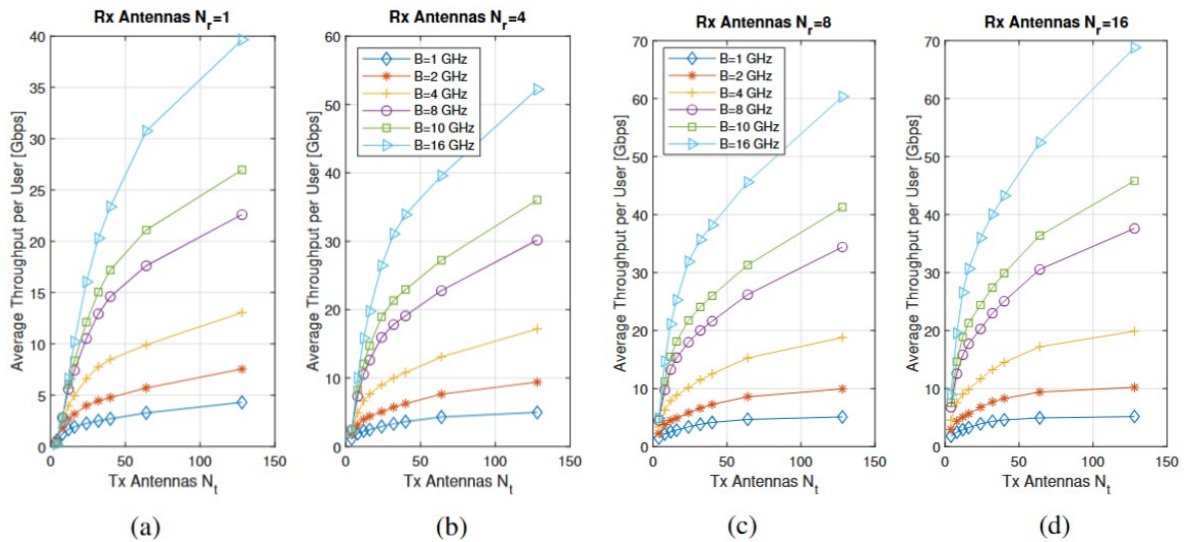
Figure 24. Average throughput per user versus the number of transmit antennas comparison with respect to operating bandwidth and the number of receive antennas.

TABLE 5: POWER CONSUMPTION OF THE HMD WITH RESPECT TO RX ANTENNA AND BANDWIDTH IN WATT

Bandwidth [GHz]	$N_r = 1$	$N_r = 4$	$N_r = 8$	$N_r = 16$
1	0.18	0.4	0.69	1.26
2	0.25	0.68	1.24	2.35
4	0.39	1.26	2.4	4.67

8	0.69	2.48	4.86	9.53
10	0.88	3.12	6.14	12.03
16	1.36	5.11	10.05	19.87

The power consumption of the HMD P_{RX} in the reception phase is listed in Table 5 for varying numbers of receive antennas N_r and bandwidth B. We consider 28nm technology for the realization of digital baseband power consumption. Figure 24 and Table 5 report the performance metrics for the solution space without considering any objective function or constraints. Next, we apply the objective function along with throughput constraint to further select some of the favourable architecture configurations from the solution space.

In the 4K-VR scenario, we focus on the transmission of the 4K video quality with different compression rates. Figure 25 illustrates the average power consumption of the HMD $P_{RX,avg}$ versus the average power consumption of the AP $P_{TX,avg}$ for (a) uncompressed data (b) ultra-low latency transmission considering 4K-VR generation, which the architecture configurations that satisfy the throughput requirement constraint.

The results demonstrate that the average power consumption of HMDs decreases significantly as the number of antennas at the AP increases. Only HMD configurations with $N_r = 1$ achieve an average power consumption below 1 W for uncompressed data transmission, as shown in Figure 25(a). This highlights the potential to offload computational complexity to the AP side, thereby reducing the power burden on the HMD. This trade-off is particularly beneficial for scenarios where the HMD's power resources are constrained, such as battery-operated devices.

A similar trend can be observed in Figure 25(b) where more architecture configurations satisfy the throughput requirement T_{req} due to the video compression reduces the throughput requirement.
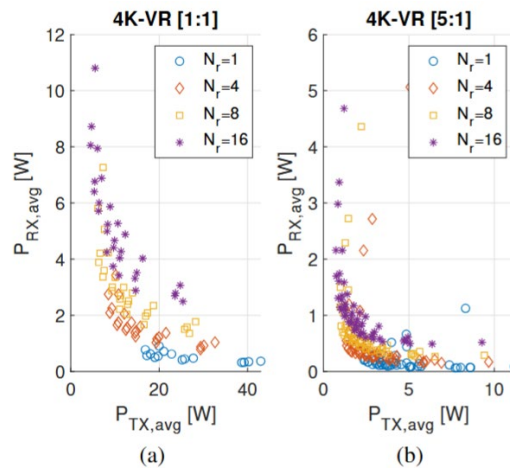
FIGURE 25 AVERAGE POWER CONSUMPTION OF THE HMD $P_{RX,avg}$ VERSUS THE AVERAGE POWER CONSUMPTION OF THE AP $P_{TX,avg}$ A.) UNCOMPRESSED DATA TRANSMISSION B.) ULTRA-LOW-LATENCY COMPRESSION

TABLE 6: OPTIMIZED ARCHITECTURE CONFIGURATION FOR 4K-VR SCENARIO

Uncompressed transmission [1:1], T_{req}=10.6 Gb/s, P^{ref} =0.1					
β	$P_{TX,avg}$ [W]	$P_{TX,avg}$ [W]	B	N_t	N_r
1	38.6	0.31	4	128	1

0.5	17.2	0.56	8	32	1
0	4.7	8.05	8	8	16
Ultra-low latency transmission [5:1], T_{req}=2.12 Gb/s, P^{ref} =0.1					
β	$P_{TX,avg}$ [W]	$P_{TX,avg}$ [W]	B	N_t	N_r
1	7.71	0.06	4	128	1
0.5	2.88	0.12	4	24	1
0	0.73	2.15	4	4	16
Low latency transmission [20:1], T_{req}=0.53 Gb/s, P^{ref} =0.1					
β	$P_{TX,avg}$ [W]	$P_{TX,avg}$ [W]	B	N_t	N_r
1	1.92	0.01	4	128	1
0.5	0.72	0.03	4	24	1
0	0.18	0.53	4	4	16
Lossy transmission [300:1], T_{req}=35.6 Mb/s, P^{ref} =0.01					
β	$P_{TX,avg}$ [W]	$P_{TX,avg}$ [W]	B	N_t	N_r
1	0.12	0.001	4	128	1
0.5	0.083	0.001	8	64	1
0	0.012	0.03	4	4	16

Table 6 summarizes the optimal architecture configurations for different values of tuning parameter β. In particular, we observe that the tuning parameter $\beta = 0$ and $\beta = 1$ find the lowest average power consumption configurations for AP and HMD, respectively. On the contrary, $\beta = 0.5$ offers a balance between $P_{TX,avg}$ and $P_{RX,avg}$. In particular, the highest number of transmit antennas N_t is selected when $\beta = 1$ to improve the throughput, thus increasing the duty cycling ratio D, neglecting the average power consumption of the AP $P_{TX,avg}$. On the other hand, the highest number of receive antennas N_r is selected when $\beta = 0$ to improve SINR γ_k, neglecting the power consumption of the HMD.
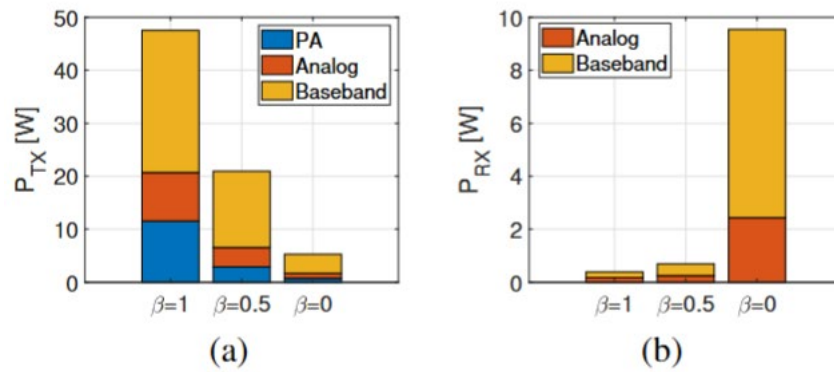
FIGURE 26 POWER CONSUMPTION BREAKDOWN OF OPTIMAL ARCHITECTURES FOR THE UNCOMPRESSED 4K TRANSMISSION CASE (A) AP AND (B) HMD

Figure 26 depicts the power consumption breakdown of optimal AP and HMD architectures for the uncompressed 4K transmission case, respectively. Optimized configurations have duty cycling ratios of $D = 0.81$ for $\beta = 1$, $D = 0.82$ for $\beta = 0.5$, and $D = 0.88$ for $\beta = 0$. Please note that the digital baseband power consumption dominates the power where many antennas are deployed and high bandwidth is utilized.

3.3.3   Summary and recommendations

The challenges in a VR system operating in sub-THz are identified as throughput, latency, and power consumption. A power model and PHY simulation framework were employed to analyse the impact of key parameters, such as antenna configurations, bandwidth, video resolution, and compression rates, on power consumption. We demonstrate that for different power optimization goals, a tailored transceiver design with different numbers of antennas is required for different VR realization. Additionally, the results reveal that strategic offloading of computational complexity from HMDs to APs can significantly reduce HMD power consumption, enhancing the feasibility of battery-constrained VR devices. This research offers guidance for the design of energy-efficient VR systems in the sub-THz band, contributing to the advancement of next-generation VR in the 6G in-X subnetworks.

3.4   JAMMING-ROBUST PHY DESIGN

Jamming and external interference pose significant threats to communication reliability, particularly in use cases requiring ultra-low latency. In this respect, the 6G physical layer must be designed with built-in robustness against such type of malicious interference. This is critical for applications demanding highly reliable communication, including industrial robot control, Wireless Zone ECU, Collaborative Wireless Zone ECU, Inter-subnetwork Coordination, Virtual ECU in automotive systems, and AR navigation for consumers [3]. In this section, we present two approaches for improving robustness to jamming, by relying on physical layer features.

The first approach considers a design of the air interface and transmission procedure for providing a native *proactive* robustness to such interference. We extend the design principle already presented in D3.1 [2] and includes simulation results and a new centralized packet repetition scheme proposal for reducing the impact of legitimate collisions introduced by such native design.

The second approach is instead *reactive*, as it is based on classifying the external interference via convolutional neural networks and relies on scheduling to mitigate its effect.

3.4.1   Native Jamming Robust Design

Jamming and external interferers are a major thread for critical communication in subnetworks, e.g. for robot control in industrial subnetworks, as it can jeopardize the possibility of achieving the required very low packet failure probability.

Though techniques based on detection and mitigation of jammers can alleviate the problem, it is acknowledged that the physical layer should feature some degree of robustness to such external interference, in order to provide a tier of protection to it, at least for critical communication.

In deliverable D3.1 [2], we have presented the main design criteria for a jamming-robust physical layer, which can be summarized as follows:

- Consecutive transmissions to and from a given device should occur over different frequency channels or subbands, preferably in a manner that is as unpredictable as possible. The hopping pattern can be generated using pseudorandom sequences known only to the transmitter and receiver. To ensure proper alignment, the transmitter and receiver must share the seeds for the pseudorandom sequence generation during a phase such as authentication, before operation begins. This approach is comparable to the one employed by ad hoc technologies operating in unlicensed spectrum, such as Bluetooth, but it is not commonly used in cellular technologies.

Importantly, channel hopping should apply not only to data packets but also to control information, including broadcast channels. This is because jamming broadcast channels could disrupt communication across the entire subnetwork, rather than just individual links. The usage of frequency hopping can also reduce the implementation cost, thanks to the instantaneous narrowband transmissions.

- Reliability can be further improved by encoding a packet across multiple hops or by repeating the packet. However, this comes at the cost of reduced spectral efficiency, so this method should be reserved for scenarios requiring ultra-reliable communication. Additionally, the system must be designed to ensure all repetitions fit within an acceptable latency budget. This may necessitate the use of extremely short transmission units in combination with a wide bandwidth.

A critical point in dimensioning the radio resources, is setting up the number of packet repetitions for each transmission. A higher number of repetitions increases the chances of correct detection but also raises the likelihood of collisions with neighbouring subnetworks. Therefore, the optimal number of repetitions should strike a balance between efficient resource utilization and the interference generated. The optimal number of repetitions was analytically derived in D3.1 [2], assuming a packet erasure channel model.

However, forcing all subnetworks to adopt the same number of packet repetitions is inefficient since it disregards the actual impact of the interference on the individual subnetworks. Certain subnetwork can indeed experience weaker interference and/or having a stronger desired link, therefore benefitting from a low number of packet repetitions for the sake of energy saving and reduce interference footprint to the other subnetworks.

Here, we first evaluate the effective impact of the jammer on the average failure probability, and then propose a novel centralized method for adjusting the number of packet repetitions on a subnetwork basis.

3.4.1.1   Effect of jammers

We consider a deployment of N subnetworks, in a 20 x 20 m^2 area. Following 6G-SHINE use cases in D2.2 [3], this can represent the scenario of multiple robot control cells involved in local control tasks, located at minimum distance, as presented in deliverable D2.2. Each subnetwork includes an HC and a single SNE, located at a random distance between 1 m and 3 m. Subnetworks operate over a bandwidth of 100 MHz, divided in 10 channels of 10 MHz each. A transmission frame consists of 14 slots, with each slot having a duration of 66.67 μs. We refer to [36] for a detailed description of all simulation parameters. Subnetworks operate using the transmission protocol presented above, i.e. repeating their transmissions multiple times over different frequency channels by using a pseudo-random hopping pattern. Transmit power at each device is 0 dBm. The number of repetitions is selected according to the rule identified in [36]. A transmission fails in case none of the packet repetitions is correctly detected.

A jamming device located at a random position within the deployment area is considered: it jams persistently over a single frequency channels, or jumps periodically over different channels, with the instantaneous bandwidth equal to a single channel only. We limit to our analysis to such narrowband jammer as more difficult to identify than wideband jammers, and therefore potentially more disruptive.
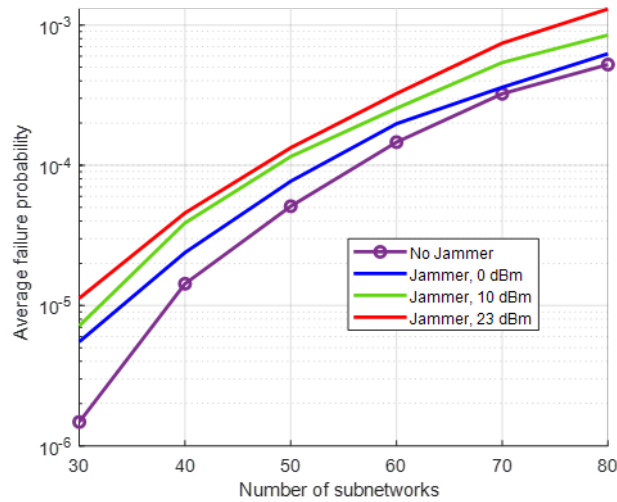
FIGURE 27. EFFECT OF JAMMERS IN THE AVERAGE FAILURE PROBABILITY.
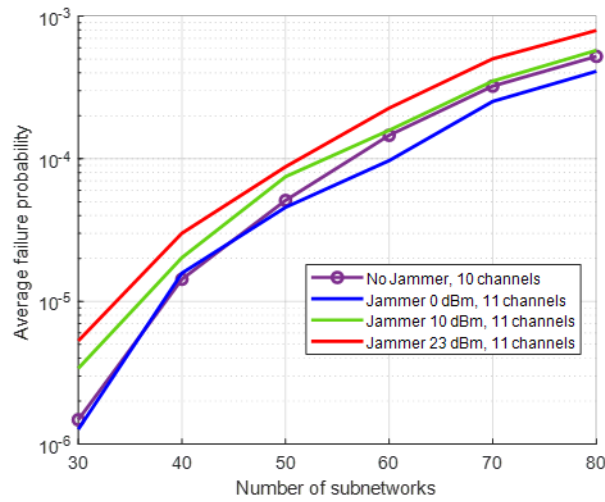
FIGURE 28. EFFECT OF JAMMERS IN THE AVERAGE FAILURE PROBABILITY OF THE JAMMING-ROBUST PHY, WITH A 10% INCREASE OF RADIO RESOURCES

Figure 27 shows the average failure probability performance as a function of the number of subnetworks in the considered deployment, considering the jamming device operating at different power levels. It is assumed here a persistent jammer. As expected, failure probability increases as the power of the jammer increases, especially for low subnetwork density. In case of jammer with the same power as the subnetworks' transmit power (0 dBm), degradation tends however to be negligible for dense subnetworks.

Among the 6G-SHINE verifiability targets, we set a maximum 10% spectral efficiency degradation in the presence of jammers. Figure 28 shows the average failure probability for the case in which the available bandwidth is increased of 10%, by setting to 11 the number of frequency channels in the presence of jammers. Results show that, in the case of a jammer with the same power as the subnetworks' transmit power, the addition of an additional channel can lead to slightly better performance than the case of 10 channels with absence of jammers; while in the case of a jammer with a 10 dBm transmit power, degradation is negligible in the case of dense networks.

3.4.1.2   Centralized packet repetition scheme

The proposed frequency-hopped communication protocol allows to achieve robustness to both legitimate and external interference, such as jammers. Nonetheless, the selection of a same number of packet repetitions at each subnetwork is suboptimal since its value is derived based on the assumption of a packet erasure channel model, where each collision is a disruptive event. However, in practice, a transmission can be successful even when collisions happen, in case the SINR is sufficient for decoding the packet; therefore, the assumption of collisions as disruptive events is pessimistic since the desired link power might be very strong or the interference received by subnetworks located far apart very weak. Moreover, forcing all subnetworks to operate with the same number of packet repetitions is inefficient as disregards the specific SINR conditions experienced at each subnetwork.
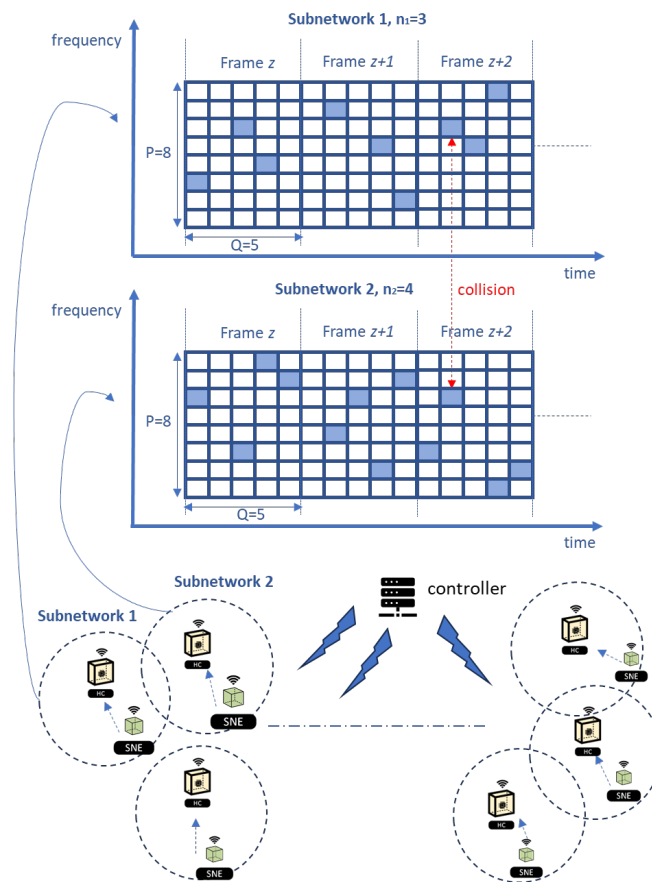
FIGURE 29. DENSE SUBNETWORKS OPERATING IN JAMMING-ROBUST NATIVE MODE. A COLLISION EVENT IS ALSO HIGHLIGHTED.

We then propose a novel centralized packet repetition schemes aiming at improving the packet error probability in dense subnetworks, where a central controller decides for the number of packet repetitions to be used at each subnetwork based on their desired power and mutual interference conditions.

It is assumed that the HC in each subnetwork is able to measure the interfering power from the neighbour subnetworks, and report it to a central controller (see Figure 29). By concatenating all the information received by each subnetwork, the central node is able to build the matrix H, having size NxN, where N is the number of subnetworks, and the diagonal is the vector of the desired channel power. The setup is reported in Figure 29.

Let us denote by $n = [n_1 \dots n_N]$ the vector of the number of packet repetitions to be used at each subnetwork. The controller aims at solving the following integer programming problem:

$$\min_n \; C(H, F, Q, n)$$
$$\text{s.t. } 1 \leq n_i \leq \min(F, Q)$$

where F and Q are the number of frequency channels and time slots of the resource grid, and $C(.)$ is a cost function related to the network performance.

We consider two cost functions:

- Optimal and near-optimal average failure probability. The optimal average failure probability in the network can be calculated by considering, for each subnetwork, all the possible subsets of combination of interfering subnetworks whose cumulative interference lead to a transmission failure. By denoting as V_i the number of subsets of subnetworks whose cumulative interference power leads to a transmission failure, the calculation of the optimal average probability requires an exorbitant number of $2^{V_i} - 1$ calculation per subnetwork. Such number of calculations can be reduced by restricting the subsets of subnetworks with cumulative disruptive interference to $\overline{N} \ll N - 1$ subnetworks having the highest channel gain with respect to the i-th subnetwork. This can reduce the number of calculations to $N \cdot 2^{\overline{N}} + \sum_{i=1}^{N} V_i$. We refer to this option as near-optimal failure probability, and refer to [36] for all the mathematical details.

- Empirical cost function. In spite of the simplifications, the calculation of near-optimal failure probability is still highly cumbersome as it scales exponentially with the network size. We propose here a low complexity empirical cost function obtained by correcting the average failure probability for the simplified case of a packet erasure channel model, with an empirical factor that includes the impact of the interference. The proposed cost function is the following:

$$C(H, F, Q, n) \doteq \frac{1}{N} \sum_{i=1}^{N} \left(1 - \prod_{j=1, j \neq i}^{N} \left(1 - \alpha_{i,j} \frac{n_j}{FQ} \right) \right)^{n_i},$$

where $\alpha_{i,j}$ is a weight factor that depends on the interference from subnetwork I to subnetwork j. factor $\alpha_{i,j}$ should be designed such that $\alpha_{i,j} \to 0$ in case of low interference between subnetwork i and subnetwork j, and $\alpha_{i,j} \to 1$ in case of high interference. We suggest then to define $\alpha_{i,j}$ as follows:

$$\alpha_{i,j} \doteq \tanh\left(\frac{\overline{\beta_i} h_{ij}}{h_{ii}} \right)$$

where $\overline{\beta_i} \doteq 2^{\frac{B_i}{wt}} - 1$, with B_i being the packet size transmitted by subnetwork I, and w and t being the size of the frequency channel and duration of a time slot, respectively. $\overline{\beta_i}$ is therefore the minimum SINR for achieving correct detection. The usage of such empirical cost function reduces the number of calculations to N^2.

The integer programming problem can therefore be solved by using the near-optimal failure probability or the empirical cost function. We propose to solve the problem via a genetic algorithm, being a population-based approach that explore the search space using multiple solutions simultaneously; this leads to a faster convergence time than single solution-based approaches such as simulated annealing.

Performance of the proposed centralized packet repetition scheme is evaluated in a 20x20 m² industrial scenario with multiple subnetworks located in close proximity, reflecting a scenario of dense robot control cells. Each subnetwork consists of an HC and a served device, located at a random distance between 1 and 3 m. Subnetworks are operating over a bandwidth of 100 MHz, divided in F=10 equal-sized channels. Devices are transmitting at each frame a packet of 100 bytes, repeated n_i times by hopping over randomly selected time slots and channels. Operational carrier frequency is 10 GHz. Radio propagation assumptions are aligned with the findings in deliverable D2.4 [37]. We refer to [36] for further details on the simulation assumptions.

The performance of the proposed scheme is evaluated against the following benchmarks: a) Random selection, referring to the case where each subnetwork selects randomly the number of repetitions to be used at each frame; b) Fixed 1 repetition, where all subnetworks are only transmitting once per frame; c) Fixed 10 repetitions, where all subnetworks are repeating the packet for the maximum number of times, i.e. 10; d) Optimized Fixed repetitions, obtaining by exhaustively searching over the 10 possible fixed repetition options, the one leading to the lowest average failure probability; e) Erasure scheme, where the controller minimizes the empirical cost function assuming $\alpha_{i,j} = 1$, i.e. assuming a packet erasure channel model.
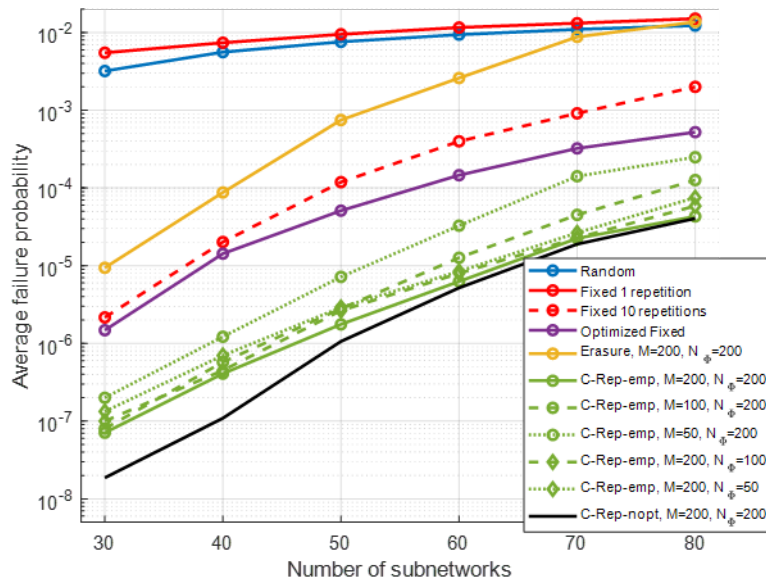
FIGURE 30. AVERAGE FAILURE PROBABILITY PERFORMANCE OF C-REP SOLUTIONS.

Figure 30 shows the average failure probability in the network as a function of the number of subnetworks in the deployment. C-Rep scheme is evaluated considering different values for number of iterations (M) and population size (N_Φ), considering the empirical cost function (C-Rep-emp), as well as the near-optimal failure probability (C-Rep-nopt), calculated assuming $\overline{N} = 12$. C-Rep solutions clearly outperforms all the benchmark schemes, of a factor of up to $\sim$ ×10 with respect to Optimized Fixed repetitions for M = 200 and N_Φ = 200. C-Rep-nopt shows the best performance especially for a low number of subnetworks, while its benefits are limited for a high number of subnetworks. However, such benefits come at the expense of a significantly high number of calculations with respect to C-Rep-emp, estimated to be of a factor of up to ~x70 higher [36].

3.4.1.3   Summary and recommendations

In this section, we have presented out guidelines for the design of a jamming robust air interface, leveraging frequency hopping over multiple subbands, combined with packet repetitions. This can be applied to industrial use cases demanding very low packet error rates, such as robot control as identified in D2.2 [3]. Such design is able to significantly reduce the impact of jammers in the error probability, with spectral efficiency losses lower than 10% in case of a jammer (or an external interferer) having similar transmit power as the subnetwork itself. In order to further reduce the overall failure probability in dense uncoordinated subnetwork deployments, it is recommended to use a centralized packet repetition strategies where the number of packet repetitions to be used at each subnetwork is optimized by a central controller, while subnetwork operations at each transmission (i.e., selection of resource unit for each retransmission) can still be uncoordinated; such solution can reduce the failure probability of a factor of ~x10 with respect to an optimized fixed number of repetitions.

3.4.2   OFDMA based jamming mitigation

In the previous section, we have presented the design principles of an air interface able to natively (and therefore proactively) provide protection to external interference and jamming attacks, along with a method to further improve packet error rate via optimized selection of packet repetitions. Here, a reactive approach to jamming mitigation is instead presented, where the external interference is first classified, and its impact is counteracted via scheduling.

In our previous study of D3.1 [2] and D5.2 [38], we have shown that the impact of a wide band jammer on receiver's decoding performance is equivalent to the increased noise floor. An approximated log likelihood ratio (ALLR) based demapper is designed in the hardware to improve physical layer's robustness against a wide band jammer. In this study, we aim to improve the robustness against narrow band jammer in two steps: first step is to identify and classify the jammer, and the second step is to take effective scheduling to mitigate the jammer's impact. The key points are summarized below, for more detailed explanation of this work please refer to [39]. The related usecase defined in WP2 is the robotic control in an industrial environment. Please note that the concept of OFDMA based mitigation is validated via Wi-Fi due to practical constraints, however since OFDM is also used in 5G and 6G, it is readily applicable to 6G subnetworks.

3.4.2.1   CSI based technology detection and classification

Given a known transmitted OFDM signal in the frequency domain x (namely the HE-LTF in 802.11ax)[40], a transmitted interference signal *i* and noise vector *n*, the received OFDM symbol *y* in a frequency selective fading environment is then formulated as:

$$y = h_w \odot x + h_i \odot i + n$$

Where h_w, h_i are the channel responses of Wi-Fi and jammer signal respectively, and $\odot$ represents elementwise multiplication. All symbols are vectors of complex values with dimension M, as the number of active subcarriers in the OFDM symbol. The observed Channel State Information (CSI) by the receiver is then formulated as:

$$\hat{h} = y \oslash x = (h_w \odot x + h_i \odot i + n) \oslash x$$

Where $\oslash$ represents the elementwise division. From the characteristics of *i*, such as its presence on different subcarriers, as well as its magnitude and phase change due to different modulation types, the spectral location and technology can be derived. Given a newly observed CSI h , the task is thus to

estimate *i* and extract features from *i* to classify the type of cross technology interference (CTI), if any. This is a typical task for a Convolutional Neural Network (CNN). By training it on different ˆh with and without interference, the model learns each of the individual components of the received signal in order to determine *i* from a new CSI ˆh. The benefit of using machine learning over purely rule-based methods is that it does not require threshold tuning and it can easily be retrained with real-life data.

As the in-X subnetworks foresees to also operate in unlicensed spectrum, a typical scenario where this theory can be applied is for a Wi-Fi channel in the 2.4 GHz ISM band, which overlaps with multiple LR-WPAN and BLE channels, as shown in Figure 31. The use of Wi-Fi is however merely a choice of validation tool, the technique of OFDMA based interference mitigation is applicable for all OFDM based technology, hence also 6G subnetworks.

With OFDMA enabled, Wi-Fi 6 allows dividing subcarriers into several Resource Units (RUs), the boundaries of the RUs comprised of 106 subcarriers (106-tone RUs) are highlighted by dashed blue lines.
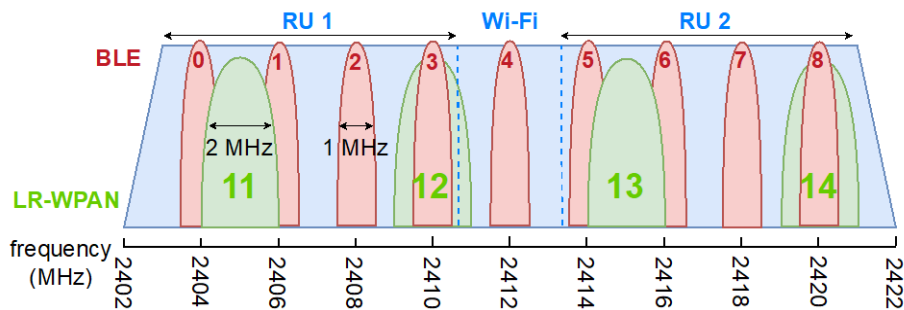
FIGURE 31. SPECTRUM USED BY WI-FI CHANNEL 1 (DOTTED LINES SHOW BORDERS BETWEEN TWO 106-TONE RUS), LR-WPAN CHANNELS 11-14 AND BLE CHANNELS 0-8.

Figure 32 shows the magnitude and phase of three typical CSI snapshots captured by ESP32-C6 on Wi-Fi channel 1, either with no interference, or with Jammer from Low Rate – Wireless Personal Area Network (LR-WPAN) or Bluetooth Low Energy (BLE). In real-life scenarios, packets with different SNR and SIR levels will be received. The SNR represents the power of $h_w \odot x$ as compared to *n* and thus influences the quality of the CSI. Similarly, the SIR refers to the power ratio of $h_w \odot x$ with respect to $h_i \odot i$. The higher the SIR, the weaker the interference signal, the more difficult it is to detect and characterize the cross-technology interference (CTI), which is the usual suspect of jammer.
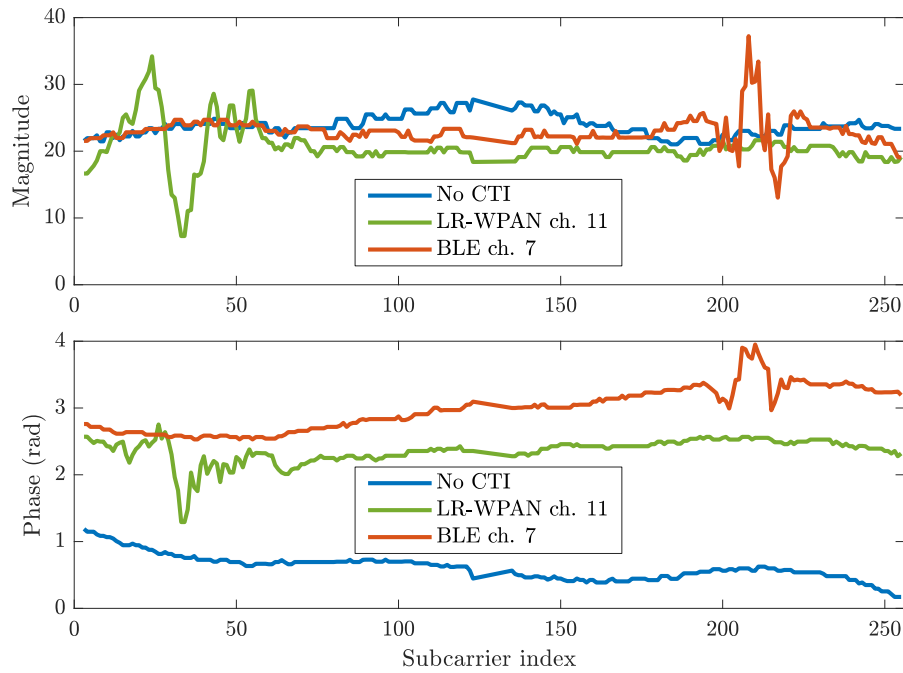
FIGURE 32. MAGNITUDE AND PHASE OF CSI SNAPSHOTS CAPTURED AT HIGH SNR BY THE ESP32-C6 WHEN NOT INTERFERED (BLUE), INTERFERED WITH LR-WPAN CHANNEL 11 (GREEN), OR BLE CHANNEL 7 (RED) WITH AN SIR OF 1 DB.

We create a CNN that has a matrix input $\widehat{H} \in \widehat{A}^{2x242}$, representing the real and imaginary part of the 242 active subcarriers of the CSI. Note that the CSI is thus not converted to magnitude and phase components to avoid this computational step on-device. The output is 14 classes $C_1, C_2, ..C_{14}$, where C_1 is no interference, C_2 to C_5 are the interference of the LR-WPAN channels and the remaining classes, C_6 to C_{14}, pertain to the interference of the BLE channels (see Figure 31). In order to keep the size of the CNN small, and since there are less features to be extracted compared to e.g. image recognition, it consists of only two convolutional layers. Afterwards, in order to understand the complex relationships between features like the shape of the overall CSI and the distortions due to interference, three fully-interconnecting layers with generalized matrix multiplication (Gemm) are applied. Each layer uses Rectified Linear Unit (ReLU) activation and at the final stage, a LogSoftmax function generates the probability distribution across the classes. See Figure 33 for the full model.
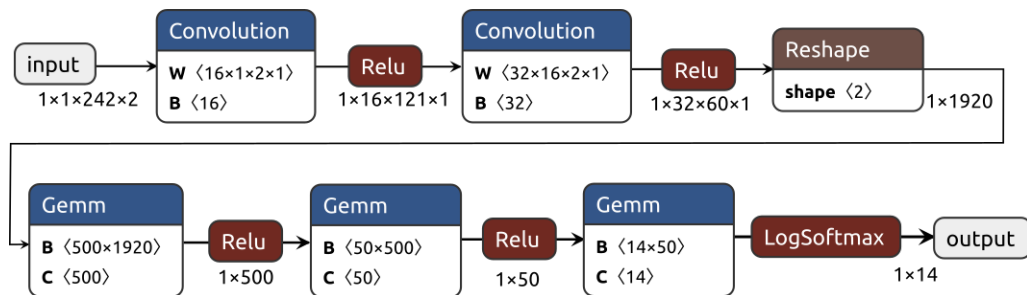
FIGURE 33 SCHEMATIC OF THE CNN USED TO DETECT CTI.

In order to create a general model that is not dependent on specific transmitter, receiver or environmental characteristics, we opt to use artificial data to train the model. The data is generated based on the physical layer specifications and multipath channel models of the relevant standards. Namely, we prepare a 242-tone HE-LTF field (20MHz bandwidth) in MATLAB and generate several LR-WPAN and BLE waveforms using the Communications and Bluetooth Toolbox, respectively. We shift the

LR-WPAN and BLE signals in the frequency domain to one of the overlapping channels within Wi-Fi channel 1 as shown in Figure 31. The HE-LTF and the interfering signals are convoluted with a different realization of the IEEE 802.11ax indoor spatial channel models B or C [41], suitable for small to medium indoor environments such as factory floor or warehouse. No carrier and sampling frequency offset are added to the HE-LTF symbol, because the receiver will compensate for this already using the legacy preamble training fields. Afterwards, we scale the IQ samples of the interference according to a certain SIR of the full 20 MHz and superpose them with the HE-LTF. Then we add white Gaussian noise according to the SNR, which is based on the signal power without channel response and interference. Following, we perform the 256-point FFT on it, and compensate it with the known HE-LTF sequence to get the CSI at the active subcarriers.

The CNN is implemented in Python using the PyTorch library. Models are initially trained based on CSI obtained at specific SNR and SIR to determine the range at which sufficient accuracy can be achieved. The general model is trained on data generated with an SNR range of 14 dB to 24 dB and an SIR range of 1 dB to 15 dB, varied in 1 dB steps. The generated data is divided into 80% training and 20% validation data. We use the Adam optimizer with a learning rate of 0.001 and a batch size of 256 for 200 epochs. Training was performed on an 8-core Intel Xeon CPU E5-2620 v4 at 2.10GHz with 32GB RAM, which took about 10 hours. We convert the PyTorch model to the Open Neural Network Exchange (ONNX) format. It is then pre-processed and optimized without quantization for the ESP32.

Three different tests are done to evaluate the performance of the classification model, namely an over-the-air (OTA) test, a test using coaxial cable, and a test with artificial data on a host PC. The OTA test includes hardware impairments as well as affects from unseen channel responses. The cable test eliminates the channel response, while the artificial data shows the performance without any hardware impairments on a non-constraint device.

In order ensure 100% interference ratio, we utilize the R&S CMW270 wireless connectivity tester as an IEEE 802.11ax AP, which simultaneously replays clean IQ samples of the waveforms of LR-WPAN channels 11-14 or BLE channels 0-8 via the general-purpose RF output. The AP is set to use channel 1 and create HE-SU packets using MCS 0. The Wi-Fi tester is set to its maximum transmit power of -3 dBm with antennas connected, while the ESP32-C6's antenna port is connected to an RF terminator. In this way, the ESP32-C6 is only able to receive the strong signal from the nearby Wi-Fi tester. By moving the ESP32-C6 around in approximately a meter range, we control the SNR relative to the AP's power. Still in some cases a clear unintentional interference was visible in the CSI. These measurements were manually discarded. The SNR is estimated based on RSSI reports during the measurement and the receiver sensitivity of the ESP32-C6. SIR is determined by the power difference between the RF ports for Wi-Fi and the interference signal. For the cable test, the ESP32-C6's antenna port is connected to the output of a 2-way power combiner, one input is connected by coaxial cable to the tester's RF port for Wi-Fi and one to the RF port for interference. The different RSSI values are obtained by changing the tester's output power. When deploying the model on the ESP32-C6, it could be seen that at lower SNR, the model often judges interference at BLE channel 4, which is exactly in the middle of Wi-Fi channel 1. A possible reason for this is that leakage of the transmitter's or receiver's Local Oscillator (LO) can create distortion in this region. From here on, we use filtered accuracy to refer to results without misdetection at BLE channel 4.
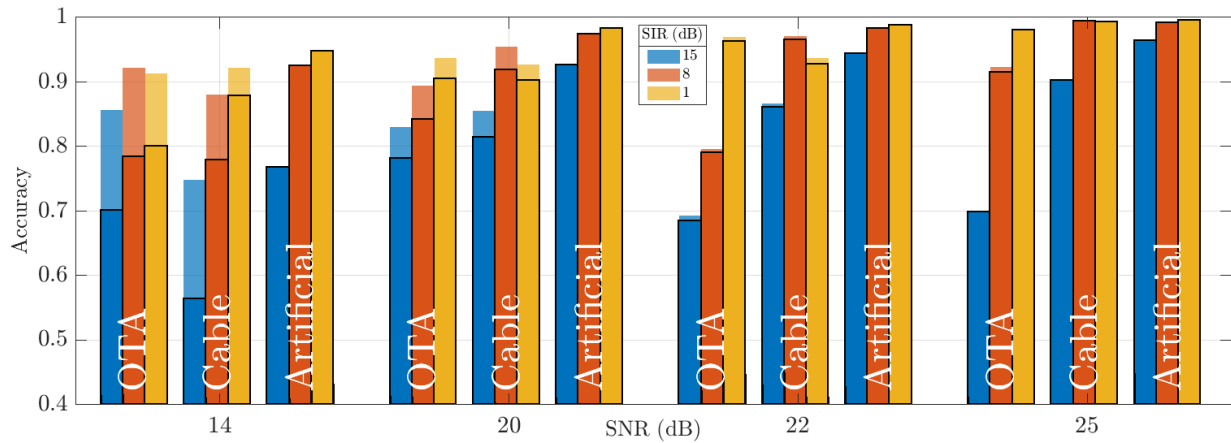
FIGURE 34 CLASSIFICATION ACCURACY OF THE OTA, CABLE AND ARTIFICIAL TESTS FOR DIFFERENT SNR AND SIR (NOTE THAT THE ACCURACY STARTS AT 0.4 FOR BETTER VISIBILITY).

Figure 34 shows the accuracy of CSI-based technology classification for each combination of SNR and SIR of the OTA, cable and artificial test. For the OTA and cable test we show both the original accuracy (bars with borders) and filtered accuracy (bars without borders). In general, the lower the SNR and the weaker the interference (higher SIR), the lower the accuracy. Two interesting outliers from this general rule can be identified. The accuracy of the OTA test at SIR 15dB peaked at 20dB SNR; there is a downward trend at higher SNR. This may be because the model has learned that for a smooth CSI (often happens at high SNR), there needs to be a clear distortion visible to classify it as interference. However, in the dynamic environment with unseen channel responses, weak CTI might be easily confused as part of the fading channel. Furthermore, in the cable test at 20dB SNR and above, detecting weaker interference (SIR of 8dB) is better than detecting stronger interference (SIR of 1dB). A reason for this might be that due to the very strong CTI, the synchronization, frequency or sampling offset estimation of the Wi-Fi receiver is more likely to be wrong, which will hurt the quality of the CSI, hence affecting the accuracy.

TABLE 7: CONFUSION MATRIX PER TECHNOLOGY.

Actual \ Predicted		No CTI	LR-WPAN	BLE
No CTI	OTA	94.75	3.0	2.25
	Cable	91.5	3.75	4.75
	Artificial	85.53	5.03	9.44
LR-WPAN	OTA	15.17	81.46	3.38
	Cable	7..54	87.5	4.91
	Artificial	2.27	95.82	1.91
BLE	OTA	9.79	1.80	88.42
	Cable	2.42	3.68	93.9
	Artificial	2.16	0.80	97.03

Next, we analyse the confusion matrix per technology in Table 7. It shows the percentage of predicting either no CTI, LR-WPAN or BLE in case of each actual class, obtained for all SNR and SIR using the filtered results. In general, it can be seen that from the OTA to the cable and artificial test, the model went from low to high false positive rate (predicting LR-WPAN or BLE when there was no CTI), and from low to high true positive rate. Also, the presence of LR-WPAN is more often classified as ``no CTI'' than BLE, which can be explained by the fact that the power of BLE is concentrated on a smaller bandwidth, meaning that at the same SIR, BLE has a higher power spectrum density than LR-WPAN, hence its distortion on

the CSI is clearer. To assess the accuracy of the spectral location, we divided the spectrum into four 52-tone RUs, each overlapping with two BLE channels and one LR-WPAN channel. When interference is detected in the OTA test, the correct overlapping RU is detected in 99.8% of the cases.

3.4.2.2   CTI aware OFDMA scheduling

In order to show the benefit of CTI-aware OFDMA scheduling, we implement a scheduler on an openwifi AP and let the ESP32-C6 perform CTI detection. Openwifi is an open-source implementation of the IEEE 802.11 standard running on a System-on-Chip. The baseband processing and low-MAC is realized on the Field Programmable Gate Array (FPGA); the driver and high-MAC run as Linux kernel modules on the on-chip ARM processor. The baseband processing for downlink OFDMA support following the IEEE 802.11ax standard has been implemented in our previous work [42]. In this work, we make the necessary addition to the driver to control the RU allocation and corresponding parameters like MCS. The STAs perform the CTI classification and send the result to the AP using Wi-Fi frames. Since the two-user OFDMA frames consisting of 2x106-tone RUs have a different CSI dimension than the 242-tone RU used previously, the model has been retrained to incorporate the new CSI dimension.

For this specific OFDMA scheduler we are only interested in whether CTI is detected and its spectral location. Therefore, the AP holds a historical record of 64 interference detection results per RU and per STA within the driver. First, the AP decides upon an RU allocation based on the number of users for which it has data available. Then, users will be assigned to RUs where they experienced the least interference during the record. By using a single CSI snapshot combined with a historical record, the CTI detection time is relatively low, hence the scheduler can capture transient interference and quickly adapt to it. To validate the CTI-aware OFDMA scheduler, the experimental setup as shown in Figure 35 is used.
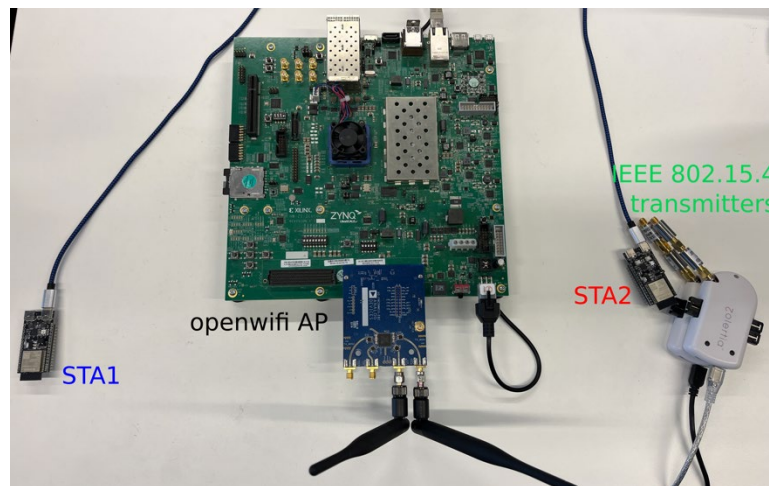
FIGURE 35 EXPERIMENTAL SETUP TO VALIDATE THE **CTI**-AWARE **OFDMA** SCHEDULER ON THE OPENWIFI **AP,** WHEN SENDING DOWNLINK TRAFFIC TO **STA1** AND **STA2;** ONLY **STA2** IS INTERFERED.

Namely, STA1 and STA2 (ESP32-C6 boards) are located on opposite sides of the openwifi AP, which runs on a Xilinx Zynq UltraScale+ MPSoC ZCU102 with an Analog Devices FMCOMMS3 RF front-end. Next to STA2, the two IEEE 802.15.4 transmitters[1] are are placed on top of each other. The openwifi AP has only limited output power, thus the STAs need to be close to the AP to stay connected. In order to create the

topology where only STA2 experiences interference, the RF outputs of the LR-WPAN transmitters are connected to two 20dB attenuators and their output power is set to -24dBm. The LR-WPAN devices continuously send request and reply to packets of 80 bytes to each other, which equals 3.83ms of airtime per packet. Meaning that if each request is received correctly and answered with an equally long reply packet, around 51% duty cycle is achieved. Due to packet losses, in reality the duty cycle of the interference is only around 35%.

We compare the CTI-aware scheduler to a naive MU scheduler using a fixed RU allocation, which happens to always allocate a user to the RU where it experiences the worst interference. Furthermore, we evaluate the performance when using SU packets. This packet format always occupies the full bandwidth, which relies on the standard Carrier Sense Multiple Access with Collison Avoidance (CSMA/CA) for CTI mitigation. iPerf is used to start UDP traffic to the two STAs simultaneously, with the default payload length of 1470 bytes. The open WIFI AP is set to use a fixed MCS 7. The high MCS is chosen to clearly show the impact of CTI. For each scheduler, five tests of one minute are executed and the average throughput reported by the STAs is recorded. In addition, the measurements are repeated also when the CTI is disabled, to obtain a performance baseline. Figure 36 displays the average throughput for each STA with error bars showing the standard deviation.
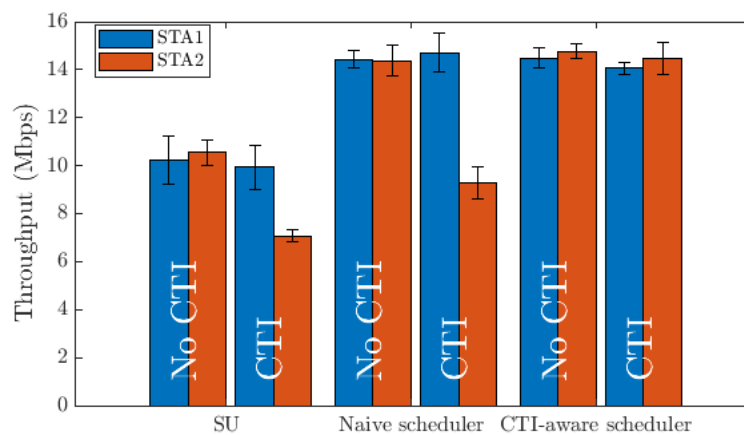
FIGURE 36. DOWNLINK THROUGHPUT OF TWO STAS USING SU PACKETS ONLY, MU PACKETS USING A NAIVE RU ALLOCATION, AND MU PACKETS BY CTI-AWARE RU ALLOCATION.

In general, using SU packets results in lower total throughput, which can be partially explained by the fact that it has more overhead due to the preamble and contention time. For SU packets, it can be seen that when CTI is applied, the throughput towards STA1 (the blue bars) remains roughly stable, while for the STA2, it decreases by about 33% (indicated by the red bar that drops from 10.5Mbps to {7Mbps). Roughly the same observation for the naive MU scheduler can be made, where the throughput of STA1 stays above 14 Mbps but STA2's throughput dropped from 14Mbps to 9Mbps} (about 35 % drop). On the other hand, the throughput reduction for STA2 is negligible when using the CTI-aware scheduler, because the data is now sent on a portion of the spectrum that does not overlap with the CTI.

3.4.2.3   Summary and recommendations

This work shows that a common module in the physical layer of a radio receiver can be applied to identify narrow band jammer, and a scheduler can take into account of physical topology and the spectrum occupancy to optimize the network performance so that it becomes interference resilient. When applying the proposed scheme, throughput remain nearly stable in the presence of external interference, while it may drop of above 30% in the absence of CTI scheduler.

4   MEDIUM ACCESS ENHANCEMENTS AND SERVICE MULTIPLEXING

This section presents the final project results on medium access mechanisms and service multiplexing for in-X subnetworks, building upon the preliminary findings presented in deliverable D3.1. Supporting stringent latencies for a large number of devices requires indeed significant improvements in medium access control; improvements based on proximity-based grouping for proactive retransmissions and scheduling, enhancements to the 5G configured grant framework, and the use of coded random access schemes are proposed for these purposes.

Due to their low-power nature compared to wide-area networks, in-X subnetworks are also well-suited for implementing flexible duplexing nodes; it is explored how these capabilities can support service multiplexing with diverse requirements while counteracting the impact of in-band emissions over neighbour subbands.

The predictable behaviour of industrial entities (e.g. robot motion) and in-vehicle generated data traffic offers the opportunity to subnetworks to leverage predictive mechanisms for efficient resource allocations. The presented solutions leverage channel state information prediction for robotic control and camera-generated data for in-vehicle applications, employing AI/ML techniques to enhance context and mobility awareness.

Further, we present our results on multi-link enhancements for boosting reliability of communication in the presence of obstacles. In particular, solutions based on cooperative communication and network coding techniques are described. Finally, we introduce novel approaches to improve channel access and reliability for subnetworks operating in unlicensed spectrum bands.

4.1   LOW LATENCY MAC ENHANCEMENTS

Medium access solutions for reducing latency while supporting high reliability includes the usage of novel proximity grouping schemes able to exploit similar radio conditions in the short-range subnetwork environments, as well as configured grant solutions for improved support of deterministic communications, and the usage of novel coded random access enhancements for addressing grant-free transmissions. These methods are presented in this section.

4.1.1   Proximity grouping

A proximity group contains two or more devices that experience similar radio channel conditions [2]. This property in a proximity group can be beneficial since a transmission or reception outcome of one device in the proximity group is likely to be the same as other devices in the same proximity group within a time window. Proximity grouping is more likely to be performed in in-X subnetworks consisting of a small area where the elements are closer together and more likely to experience LOS channel conditions than in a cellular network in an urban environment. In the previous deliverable [2], we introduced pre-emptive retransmissions for a TDD system, that utilises the property of proximity grouping. In this section, we will evaluate the latency reduction analytically. In addition, we introduce two methods that further utilises the property of proximity grouping.

4.1.1.1 Latency evaluation of pre-emptive retransmissions

In pre-emptive retransmissions, the network schedules an SNE downlink resources for a retransmission of a physical downlink shared channel (PDSCH) before receiving any Hybrid Automatic Repeat Request (HARQ) feedback from that SNE. The network determines whether an SNE in a proximity group, requires a PDSCH retransmission or not, from the HARQ feedback from other SNEs in the same proximity group, since they have similar or the same radio propagation channel.

The analytical method used to evaluate the latency follows that used in Section A.4 of 3GPP TR38.824 [43]. The parameters used for this evaluation is summarised in Table 8. The TDD slot pattern of {DDDDU} consists of four downlink slots followed by one uplink slot and the pattern repeats, where each slot is 0.125 ms duration. With this TDD slot pattern, UL resource is available every 70 OFDM symbols or 0.625 ms. The scheduling information for a PDSCH is contained in a Downlink Control Information (DCI) which is carried by the physical downlink control channel (PDCCH), which in this evaluation we assumed that it occupies two OFDM symbols and there is one PDCCH per slot. The PDSCH carrying the data is assumed to occupy 7 OFDM symbols or half a slot. The HARQ feedback is carried by the physical uplink control channel (PUCCH), which we assumed to be two OFDM symbols long and there can be 7 PUCCH transmission per slot, i.e., the network may schedule up to 7 PUCCH's per uplink slot.

TABLE 8: LATENCY EVALUATION PARAMETERS

Item	Parameters	Value	Comments
1	Subcarrier Spacing (kHz)	120	
2	OFDM symbol duration (ms)	0.008928571	
3	TDD Slot Pattern	{DDDDU}	
4	UL resources periodicity (symbols)	70	UL resource every 5 slots (70 symbols)
5	Start UL resource in each TDD pattern (symbols)	56	
6	PDCCH duration (symbols)	2	
7	PDSCH duration (symbols)	7	
8	PUCCH duration (symbols)	2	
9	UE (SNE) PDSCH processing, N_1 (symbols)	20	Table 5.3-1 of Section 5.3 in 38.214
10	gNB (HC) PUCCH processing, N_3 (symbols)	18	Assume half of UE PUSCH processing time, i.e. $N_2/2$
11	PDCCH per slot	1	
12	PUCCH per slot	7	
13	Number of HARQ transmission	2	Initial + 1 retransmission

We follow the processing capabilities in the 3GPP specification [44], that is, the SNE requires $N_1 = 20$ OFDM symbols to process a PDSCH and issue a HARQ feedback in a PUCCH. For the PUCCH processing time by the gNB or HC N_3, we use the same assumption in [43], that is, $N_3 = N_2/2$, where N2 is the time required by a UE to process the PUSCH. Figure 37 shows these two processing times, N_1 and N_3, where here SNE1 receives a DL grant in Slot $n+2$ which schedules a PDSCH in Slot $n+3$. SNE1 takes $N_1 = 20$ OFDM symbols to decode and construct a PUCCH, which it sends immediately at the end of Slot $n+4$. The HC then takes $N_3 = 18$ OFDM symbol to decode the PUCCH which indicates a NACK, in which case, it

immediately schedules a retransmission by sending a DL grant in Slot $n+6$, which schedules the retransmission for the PDSCH in Slot $n+7$. SNE1 then takes another N_1 = 20 OFDM symbols to successfully decode the PDSCH and transmits an ACK in Slot $n+9$.
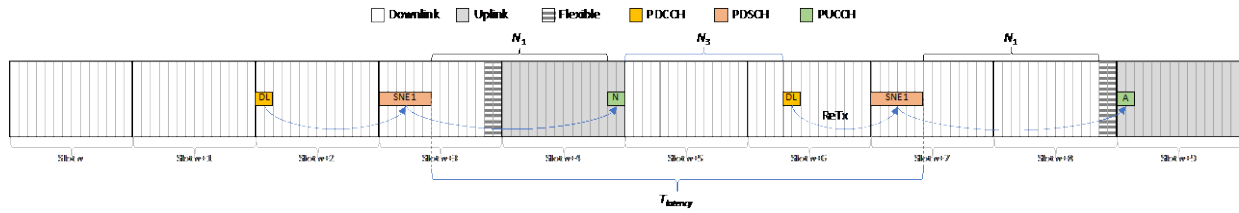
FIGURE 37: SNE PDSCH AND HC PUCCH PROCESSING TIMES

To evaluate the latency, we consider the case where the SNE requires 1 retransmission to successfully decode the PDSCH, that is an initial PDSCH and a retransmission of that PDSCH. The latency in receiving these two PDSCH (initial + retransmission) is shown in Figure 37 as $T_{latency}$. We calculate the latency for PDSCH in each downlink slot when pre-emptive retransmission is used, and when it is not used. We then determine the reduction in latency using pre-emptive retransmission.

We first determine the latencies for the cases where pre-emptive retransmission is not available. These cases are shown in Figure 38, where SNE1, SNE2, SNE3 and SNE4 belong to the same proximity group. The HC schedules PDSCHs to SNE1, SNE2 and SNE3 in Slot n, Slot $n+1$ and Slot $n+2$ respectively, where there is no PDSCH being scheduled prior to Slot n for at least a period longer than multiples of the coherence time. Since there is no HARQ feedback at the HC, prior to Slot n from any SNE in the proximity group, the HC is unaware of the channel condition for these SNEs, and therefore pre-emptive retransmission cannot be utilised for these SNEs. In another scenario, SNE4 is being scheduled a PDSCH in Slot m, where $m \gg n$, i.e. Slot m is several multiples of coherence time after Slot $n+9$. Similarly, there is no PDSCH prior to Slot m for at least a period longer than multiples of the coherence time, and hence pre-emptive retransmission cannot be used for the PDSCH for SNE4 in Slot m. For the PDSCH scheduled for SNE1, SNE2, SNE3 and SNE4 in these cases, the HC needs to rely on the HARQ feedback from these SNEs in order to determine whether retransmissions are required or not, and in this example, they all send a NACK in their scheduled PUCCHs in uplink Slot $n+4$. The HC then schedule PDSCH retransmissions for SNE1, SNE2 and SNE3 in Slot $n+6$, where here they are successfully decoded and these SNEs feedback ACKs in their scheduled PUCCH in uplink Slot $n+9$. Similarly, for SNE8, the HC schedules a PDSCH retransmission in Slot $m+8$ after receiving its NACK in Slot $m+6$, where SNE8 then successfully decodes the PDSCH after combining the retransmitted PDSCH with the initial PDSCH. The latency $T_{latency}$ for each of these cases, i.e. for each of these SNEs to receive their PDSCH successfully are calculated.
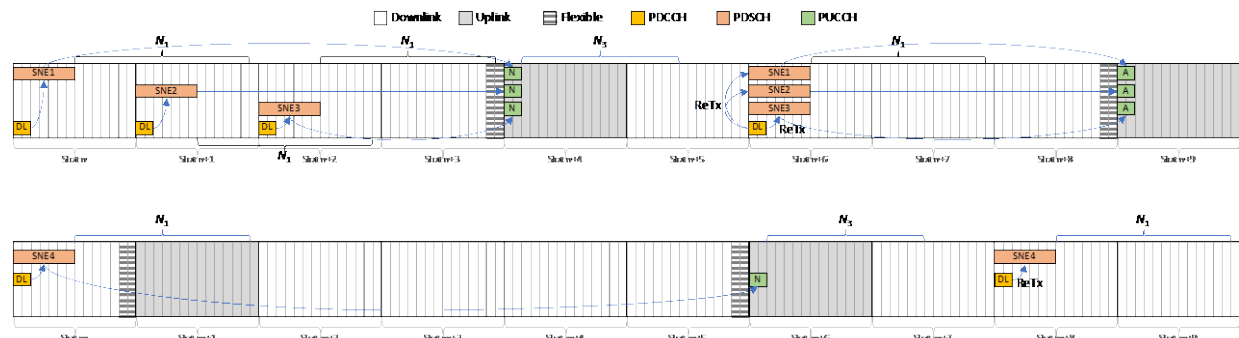
FIGURE 38: CASES WITHOUT PRE-EMPTIVE RETRANSMISSIONS

The cases where PDSCH pre-emptive retransmission is utilised is shown in Figure 39, for SNE5, SNE6, SNE7 and SNE8. The HC schedules SNE5 with a PDSCH in Slot n+3, but SNE5 is unable to provide a HARQ feedback in time in the UL Slot n+4 due to its PDSCH processing time. Since the HC managed to get HARQ feedbacks (which are NACKs) from SNE1, SNE2 and SNE3 in Slot n+4, and that these SNEs belong in the same proximity group as SNE5, the HC, when it finishes decoding these HARQ feedbacks in Slot n+5, can then assume that SNE5 did not manage to decode the PDSCH successfully. The HC then schedules a PDSCH pre-emptive retransmission for SNE5 in Slot n+6. The HC schedules a PDSCH for SNE6 in Slot n+5 and since the next available UL resource is in Slot n+9, SNE6 cannot provide a HARQ feedback until Slot n+9. Similar with SNE5, SNE6 is also in the same proximity group as SNE1, SNE2 and SNE3, and so the HC can utilise their (SNE1, SNE2 and SNE3) NACK feedback to predict that SNE6 requires a retransmission and consequently schedules a PDSCH pre-emptive retransmission for SNE6 in Slot n+6. Since the HC finishes decoding the HARQ feedbacks from SNE1, SNE2 and SNE3, in Slot n+5, it knows that the radio channel is poor from Slot n+6 onwards. Hence for SNE7 and SNE8, which also belong to the same proximity group, the HC schedules 2× repetitions for these SNEs, i.e., SNE7's 2× PDSCH repetitions are in Slot n+6 and Slot n+7, whilst SNE8's 2× PDSCH repetitions are in Slot n+7 and Slot n+8. In this examples all the SNEs feedback positive HARQ acknowledgements in uplink Slot n+9 after receiving two PDSCH transmissions. Similarly, the latencies $T_{latency}$ for SNE5, SNE6, SNE7 and SNE8 that benefited from pre-emptive retransmissions are calculated.
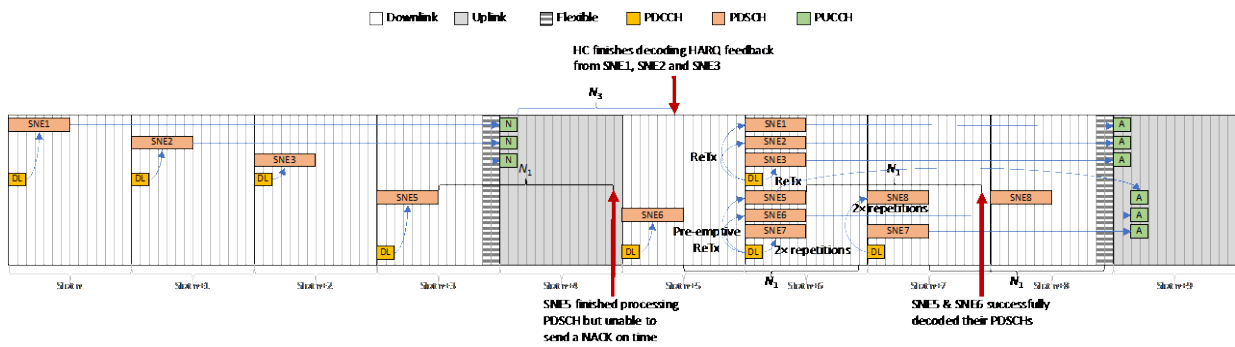
FIGURE 39. CASES WITH PRE-EMPTIVE RETRANSMISSIONS

The latencies $T_{latency}$ for the cases without pre-emptive retransmission and with pre-emptive retransmission are summarised in Table 9. It can be observed that on average, the latency using pre-emptive retransmissions in a proximity group can be reduced by 50%, which is highly beneficial for the requirement of subneworks.

TABLE 9: LATENCIES $T_{LATENCY}$ WITHOUT AND WITH PRE-EMPTIVE RETRANSMISISONS

Parameter	Latency $T_{latency}$ (ms)		Reduction	
	No pre-emptive ReTx SNE1, SNE2, SNE3, SNE4	Pre-emptive ReTx SNE5, SNE6, SNE7, SNE8	ms	Percentage
Average (ms)	0.96	0.43	0.53	55.3%
Min (ms)	0.74	0.37	0.38	50.6%
Max (ms)	1.24	0.62	0.63	50.4%

4.1.1.2   Eavesdropping in proximity group

In this section we explore another benefit of proximity grouping. Since SNEs within the same proximity group shares the same radio channel profile, they can decode each other's PDSCH. That is, one or more SNE in the proximity group can eavesdrop another SNE's downlink channel. An example is shown in Figure 40, where a subnetwork contains a HC and 5 SNEs. SNE1, SNE2 and SNE3 forms a proximity group (Proximity Group 1), and SNE4 and SNE5 forms another proximity group (Proximity Group 2). In this example, the HC transmits a PDSCH to SNE1, but since SNE3 shares the same proximity group, it also receives the PDSCH targeted at SNE1, that is SNE3 eavesdrops the downlink transmission to SNE1. By having more than one SNE receive a downlink channel, the reliability can be improved, for example, SNE3 can send the eavesdropped PDSCH to SNE1, that is, SNE3 acts like a receiver diversity for SNE1, which would be beneficial if SNE1 fails to decode the PDSCH. There can be more than one eavesdropping SNE, for example SNE2 may also eavesdrop the PDSCH targeted at SNE1.
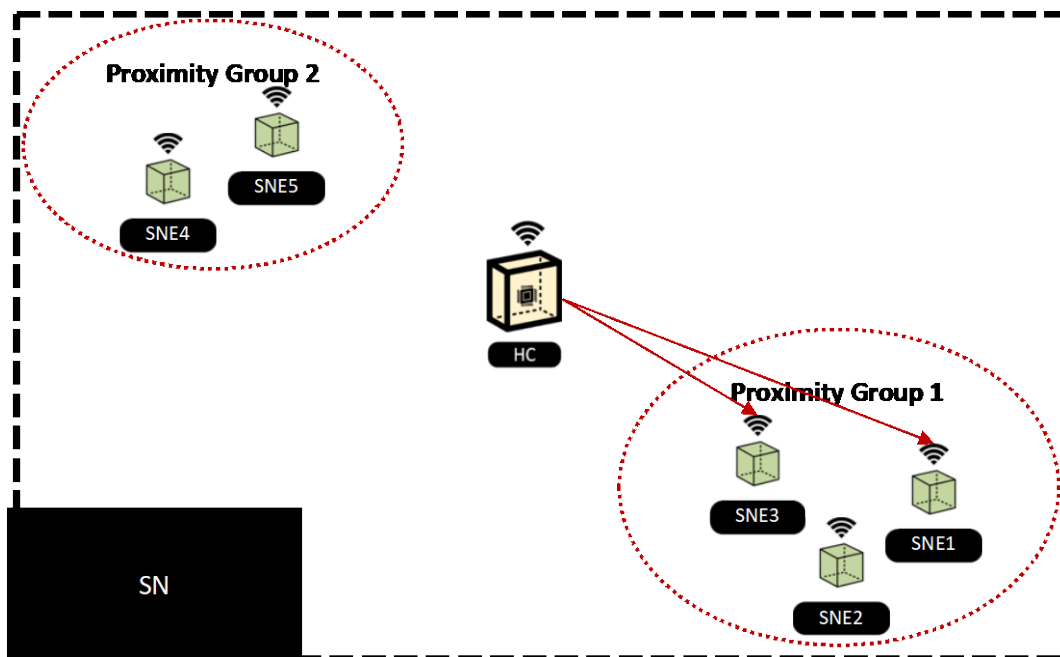
FIGURE 40: EAVESDROPPING IN A PROXIMITY GROUP

Another benefit of eavesdropping in a proximity group is that it can aid the HC in determining pre-emptive retransmission. Here the SNEs can provide HARQ feedback on eavesdropped PDSCH, which would provide the HC with information of the radio channel condition. An example of this is shown in Figure 41, where SNE1, SNE2 and SNE3 belong to a proximity group. The HC schedules PDSCH1 to SNE1 in Slot n, and SNE2 and SNE3 eavesdrop this PDSCH and attempt to decode it. SNE1 and SNE2 fail to decode PDSCH#1 and they feedback a NACK in PUCCH#1 and PUCCH#2 respectively in uplink Slot $n+1$. Although SNE3 has similar radio channel profile as SNE1 and SNE2, it is slightly closer to the HC and it managed to decode PDSCH#1, and feedbacks an ACK in PUCCH#3 in uplink Slot $n+1$. The HC retransmits PDSCH#1 to SNE1 in Slot $n+4$ after receiving the NACK from SNE1. In Slot n, the HC also schedules SNE2 with PDSCH#2 in Slot $n+2$, in which case SNE2 can only provide a HARQ feedback in the next available UL slot that is two slots after Slot $n+4$. However, the HC has feedback from SNE2 in Slot $n+1$ for PDSCH#1 (which was targeted at SNE1) and in addition to the HARQ feedback from SNE1, the HC can make a judgement that SNE2 requires a retransmission and hence it schedules a PDSCH pre-emptive retransmission for PDSCH#2 in Slot $n+3$. In Slot n+3, the HC schedules PDSCH#3 to SNE3, and since it learned that SNE3 managed to decode PDSCH#1 successfully, and in this example PDSCH#3 has the same

MCS as PDSCH#1, the HC schedules a single PDSCH#2 instance to SNE3 instead of providing additional transmissions as for SNE2.
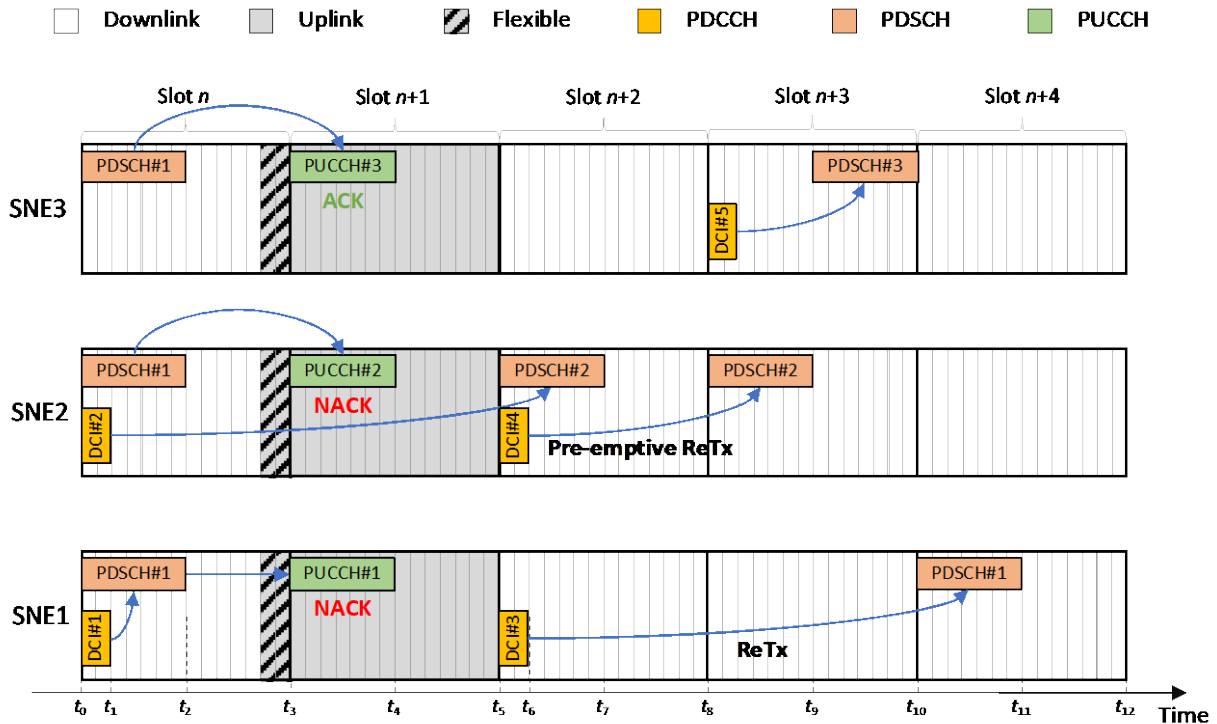
FIGURE 41: PRE-EMPTIVE RETRANSMISSION USING EAVESDROPPING HARQ FEEDBACKS

Providing HARQ feedback for eavesdropped PDSCH can also help the scheduler in performing link adaption, since link adaption typically uses these HARQ feedback to determine whether the SNE is meeting the BLER target. For extreme high reliability requirement such as BLER of 10^{-9}, relying on a single SNE to provide sufficient HARQ feedback to determine its BLER may take a long time, but if multiple SNEs in a proximity group can provide HARQ feedbacks, these can reduce the time for the HC to gauge the BLER the SNE are experiencing.

To implement eavesdropping in a subnetwork, the downlink grant carried by the Physical Downlink Control Channel (PDCCH) can use a RNTI (Radio Network Temporary Identifier) that is common among the SNEs that participate in eavesdropping in a proximity group. This enables the eavesdropping UE to read the scheduling information in the downlink grant and decode the corresponding PDSCH. The HC can also indicate which SNE the PDSCH is targeted at, so that the target SNE will pass the PDSCH to higher layers whilst the eavesdropping SNEs does not need to decode further after the physical layer processing.

In conclusion, the eavesdropping in a proximity group where one or more SNE receive downlink packet targeted at another SNE belonging to the same proximity group can provide reliability via receiving diversity and provide more information to the HC to aid in pre-emptive retransmissions and in link adaptation at the HC.

4.1.1.3   Piggybacking in pre-emptive retransmissions

The PDSCH pre-emptive retransmission scheduled to an SNE is based on HARQ feedback from other SNEs in a proximity group, that is, it is based on prediction that the SNE could not decode the previous PDSCH, prior to receiving a HARQ feedback from that SNE. Since pre-emptive retransmission is a prediction, it may not be needed by the SNE and the resources it occupies may be wasted. To reduce wastage in resources in unwanted pre-emptive retransmission, additional data for the same SNE or other SNE can be piggybacked onto the pre-emption retransmission. The HC can decide which SNE may receive this piggybacked data, for example, based on whether it has feedbacked an ACK for a previous PDSCH, which can be used as a gauge on the level of Modulation & Coding Scheme (MCS) that SNE can tolerate. An example is shown in Figure 42, where SNE1, SNE2 and SNE3 are in the same proximity group. In Slot *n*, the HC schedules PDSCH#1 in the same slot for SNE1, where PDSCH#1 uses MCS=5. SNE2 and SNE3 eavesdrop on SNE1's PDSCH#1. SNE1 and SNE2 fail to decode PDSCH#1 and they feedback a NACK in PUCCH#1 and PUCCH#2 respectively in uplink Slot *n*+1, whilst SNE3 successfully decodes PDSCH#1 and feedbacks an ACK in PUCCH#3. In Slot *n*, the HC schedules SNE2 with PDSCH#2 with MCS=6 to be received in Slot *n*+2. Since SNE2 feedbacks a NACK for PDSCH#1 which has a lower MCS, the HC determines that SNE2 is likely going to fail to decode PDSCH#2 and schedules a pre-emptive retransmission for PDSCH#2 in Slot *n*+3. The HC schedules SNE3 in Slot *n*+2 with PDSCH#3 with MCS=3, to be received in Slot *n*+3 where PDSCH#3 is piggybacked onto the pre-emptive retransmission for PDSCH#2. From SNE3's ACK feedback for PDSCH#1, the HC determines that it can receive a PDSCH with MCS=5 and hence it estimates that it may receive a PDSCH with a lower MCS (i.e. more robust), which is piggybacked onto another PDSCH. In this example, SNE3 also eavesdrop PDSCH#2 targeted at SNE2, which can be beneficial in receiving PDSCH#3 that is piggybacked onto PDSCH#2, since if it successfully decodes PDSCH#2, it can employ successive interference cancellation to remove PDSCH#2 from the downlink transmission containing PDSCH#2 with the piggybacked PDSCH#3.
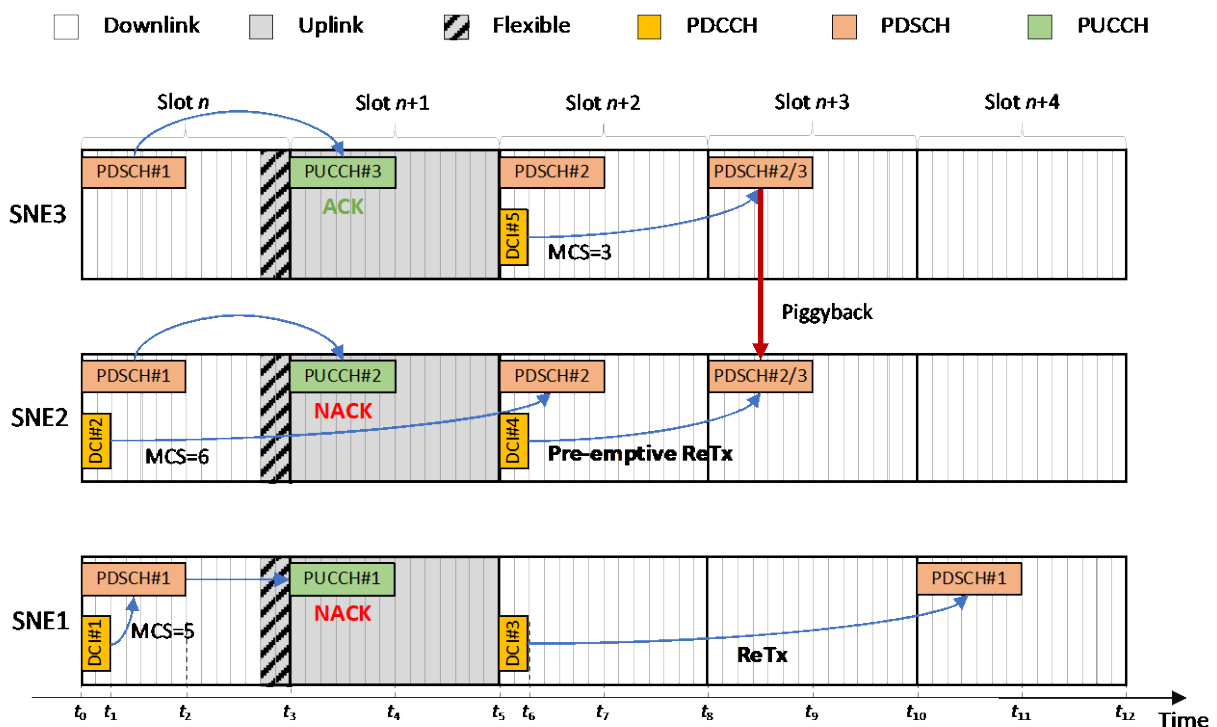
FIGURE 42: PIGGYBACK DATA ON PRE-EMPTIVE RETRANSMISSION

Although, the HC can use MU-MIMO to transmit two data to different SNEs using the same time and frequency resource, this may be challenging for SNEs in the same proximity group since they have similar radio channel profile. Another method is to punctures the pre-emptive retransmission modulated symbols with the modulated symbols from the piggybacked data, similar to the multiplexing UCI onto PUSCH in 5G. However, puncturing can carry limited amount of data, since puncturing would increase the code rate of the pre-emptive PDSCH retransmission. For the piggybacking of data onto pre-emptive retransmission, we propose to use overlaying modulation. For example, the piggyback data and pre-emptive retransmission use QPSK, and we can rotate the QPSK constellation by 45 degrees for the piggyback data so that we can overlay the modulated symbols of the piggyback data and pre-emptive retransmission together. In the next section we will discuss this scheme in further detail.

4.1.1.4   A trellis coded pre-emptive transparent retransmission-like mechanism

We next consider an alternative pre-emptive mechanism that mitigates the need for retransmissions. The outline of the mechanism is as follows:

1. Consider the transmission of a coded sequence of data blocks $s_1, s_2, s_3, \dots$ where each s_k contains multiple data symbols (e.g., QAM), and each s_k carrying a complete codeword.
2. The rate and modulation order are determined such that a target BLER is achieved, e.g., 10%.
3. Instead of transmitting the data blocks $s_1, s_2, s_3, \dots$ we transmit a modified version, $\tilde{s}_1, \tilde{s}_2, \tilde{s}_3, \dots$ derived from the original sequnce.
4. For data block $\tilde{s}_k$, and if there is no need for any HARQ transmission for any preceeding block $\tilde{s}_{k-\ell}$, $\ell = 1,2,3 \dots$, then the operations at the receiver are virtually identical to those if the original s_k had been transmitted. The BLER remains at the target level, e.g., 10%. It is, more-or-less, transparent to the receiver that $\tilde{s}_k$ was transmitted instead of s_k.
5. But if any $\tilde{s}_{k-\ell}$, $\ell = 1,2,3 \dots$ is in error, then the effect of the modification from s_k to $\tilde{s}_k$ kicks in, and the block in error may be decoded without any need of HARQ/retransmission. As soon as this is done, the current block $\tilde{s}_k$ can be decoded at the target BLER without taking the modification into (much) account.
6. The blocks $\tilde{s}_1, \tilde{s}_2, \tilde{s}_3 \dots$ can be transmitted without any additional need of time, frequency, or spatial resources. Spectral efficiency remains identical, no matter how it is computed. The only effect is that the number of HARQ/retransmission drastically reduces (down to null for all practical purposes in some cases). The price to pay for the receiver is caching, as it need to store previously received blocks and/or Log-likelihood ratios.

What is described above may be claimed to be a form of pre-emptive HARQ/retransmission. The retransmission is there for the receiver to make use, but if not needed, then it plays no role and can be fully discarded by the receiver; this, despite not requiring any additional time-frequency resources.

We consider a sequence of transport blocks to be transmitted. All blocks comprise K in bits (including CRC), and the k:th bit of the t:th block is denoted u_t^k. Each block is protected by an error correcting code of rate K/N which implies that to each transport block, there is a corresponding codeword of length N; the n:th bit of the t:th codeword is denoted v_t^n. Each codeword is modulated into a sequence of transmit symbols by a b-ary modulator producing a length L symbol sequence s_t^ℓ. Here "b-ary modulator" should be understood as a modulator the encodes b bits from each codeword into each transmit symbol. Consequently, $N = bL$. See Figure 43 for an illustration.

$$\boxed{u_1^1\ u_1^2\ \cdots\ u_1^K}\ \boxed{u_2^1\ u_2^2\ \cdots\ u_2^K}\quad\cdots\quad\boxed{u_t^1\ u_t^2\ \cdots\ u_t^K}$$

Error correction encoding (LDPC in examples) $N > K$

$$\boxed{v_1^1\ v_1^2\ \cdots\ v_1^N}\ \boxed{v_2^1\ v_2^2\ \cdots\ v_2^N}\quad\cdots\quad\boxed{v_t^1\ v_t^2\ \cdots\ v_t^N}$$

Modulation $N = bL$

$$\boxed{s_1^1\ s_1^2\ \cdots\ s_1^L}\ \boxed{s_2^1\ s_2^2\ \cdots\ s_2^L}\quad\cdots\quad\boxed{s_t^1\ s_t^2\ \cdots\ s_t^L}$$
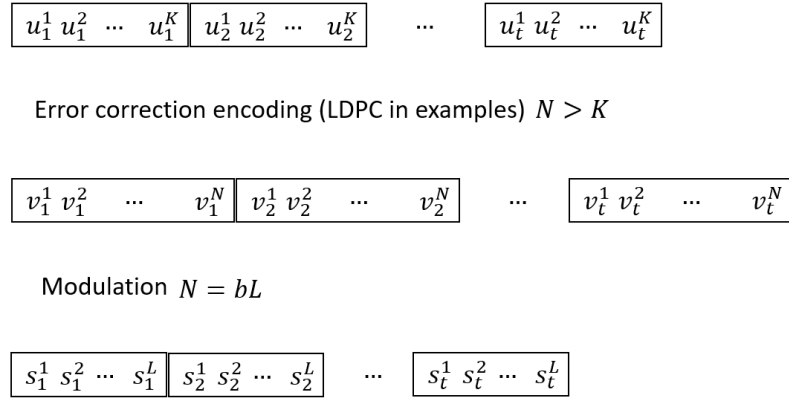
FIGURE 43. ILLUSTRATION OF ERROR CORRECTION ENCODING AND MODULATION

The novelty of our method lies in the specific way in which the modulator operates. A general overview of the modulator, which does not illustrate any novelties, is provided in Figure 44.

The figure illustrates the modulator for the t:th codeword. This codeword is first partitioned into L groups of b bits. Each group is then sent to a modulator producing a transmit symbol. Finally, all L transmit symbols are collected and sent over the channel (by means of, e.g., OFDM). It is worthwhile to mention that there are L modulators that operate in parallel, and what shall be described next applies to any of the L modulators.

At this point, the reader may be surprised about why the particular wording "modulator" was used, and not the more commonplace wording "constellation map". The standard way of forming the transmit symbols is to map each group of b bits to a 2^b-ary constellation. For example, if $b = 4$, then the symbols s_t^k could belong to a 16QAM constellation. This is, however, *not* what shall be done in the proposed method.

Instead, we use a modulator with memory which maps each group of b bits to a 2^n-ary constellation where $n > b$. An illustration of the modulator is provided in Figure 3 for the case of $b = 2$.

$$\boxed{v_t^1\ v_t^2\quad\cdots\quad v_t^N}$$

Partition into L groups of b bits ($N = bL$)
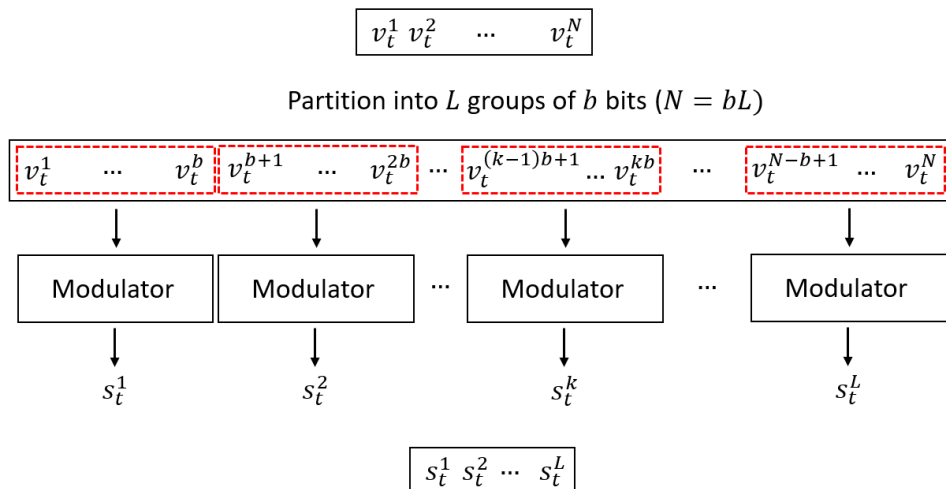
FIGURE 44. ILLUSTRATION OF MODULATOR FROM FIGURE 43

FIGURE 45. AN OVERVIEW OF THE MODULATOR FOR $b = 2$, $n = 3$, AND $v = 2$.
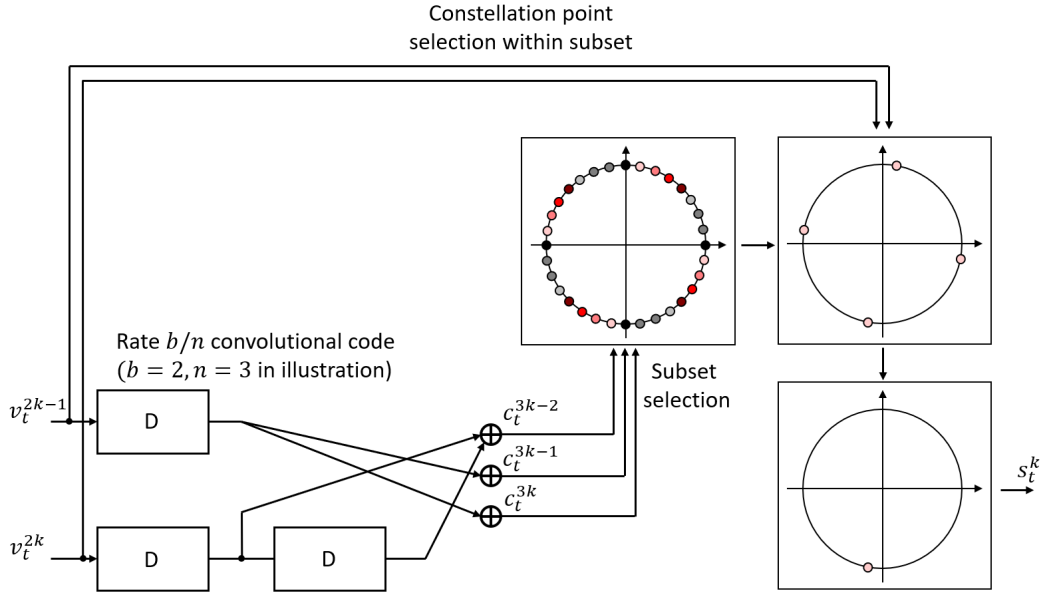
At time t, the k:th modulator takes the group $v_t^{(k-1)b+1}, \ldots, v_t^{kb}$ as input; for $b = 2$ this reduces to the tuple v_t^{2k-1}, v_t^{2k}. The modulator has previous pairs of groups in memory (the D-elements in the lower left corner of the figure), i.e., it has stored $v_{t-m}^{(k-1)b+1}, \ldots, v_{t-m}^{kb}$, $m = 1 \ldots v$, where v is the memory-depth of the modulator. For the case shown in Figure 3, $v = 2$ and this reduces to $v_{t-1}^{2k-1}, v_{t-1}^{2k}, v_{t-2}^{2k}$. Based on the content of said memory, the k:th modulator produces n bits $c_t^{(k-1)n+1}, \ldots, c_t^{nk}$; for the case $n = 3$ this is $c_t^{3k-2}, c_t^{3k-1}, c_t^{3k}$. These n bits select one out of 2^n subsets from a 2^{n+b} constellation (the one with colors). Each subset has precisely 2^b constellation points. Which point that is to be transmitted within the selected subset is determined by the b bits in the group $v_t^{(k-1)b+1}, \ldots, v_t^{kb}$, which is the input to the modulator at time t.

For given values of b, n, the design of the colorful constellation of Figure 3 is made systematically as follows.

1. Select a standard constellation with 2^b points. We refer to this as the base constellation. This may be any 2^b-ary constellation such as QAM/PSK.
2. Construct an expanded constellation with 2^{n+b} points as the collection of 2^n rotated versions of the base constellation. The rotation angles are up to design, but herein we shall exemplify with uniformly spaced rotations.

See Figure 46 for an illustration with $b = 2$ and $n = 3$. The base constellation has been chosen as QPSK.

FIGURE 46. DESIGN OF EXPANDED CONSTELLATION FROM A BASE CONSTELLATION BY MEANS OF ROTATIONS.
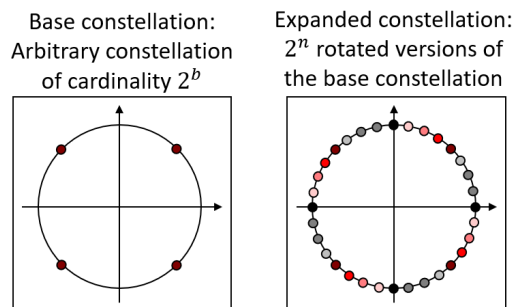
Benefits of proposed method

The rationale of the method is that it almost eliminates the need of HARQ transmissions. Further, in cases where HARQ would not be needed for a legacy transmission of a transport block (i.e., directly using the base constellation and not the expanded one), then it is trivial to remove the changes to the received signal stemming from our method. Thus, as long as HARQ is not needed, our method is not changing the situation and workload for the receiver. However, as soon as HARQ is needed, the receiver can utilize the modified signal structure and in almost all cases avoid a HARQ request. We describe the above made statements with examples.

Example 1. Without our method (legacy), the "modulator" of Figure 44 would be replaced with a "constellation map" to the 2^b-ary base constellation. So, decoding of a transport block using the base constellation can be regarded as a baseline complexity level. Now, assume that all transport blocks prior to the t:th one has been successfully decoded. The received signal is the received version of $s_t^1, \ldots, s_t^L$, and this signal has indeed been generated by our invention according to Figure 3. However, as all previous transport blocks have been correctly decoded, the receiver has v_{t-m}^k at hand for $m = 1,2 \ldots$. This implies that the state of the state machine in Figure 45 is fully known. As a direct consequence, the selected subsets used in the transmission of $s_t^1, \ldots, s_t^L$ are known to the receiver. Now recall that each subset is merely a rotation of the base constellation. Therefore, the receiver can de-rotate the received signals accordingly. As pure rotations are not impacting the SNR we reach the conclusion that, after de-rotation, the task of the receiver is identical to that of a legacy decoding.

Example 2 (Example 1 cont.). Assume now that the decoding of transport block t fails; the UE may determine if this occurs by inspection of the CRC. In this situation the effect of the method kicks in. The receiver next obtains the block $s_{t+1}^1, \ldots, s_{t+1}^L$, and this block is not independent of the codeword $v_t^1 \ldots, v_t^N$ due to the memory of the modulator. Based on the L groups s_t^k, s_{t+1}^k, we may specify L parallel soft-output algorithms to produce LLRs of the bits $v_t^1, \ldots, v_t^N$ from which a second decoding attempt can be made. It is well known how to implement said soft-output receiver algorithms, and we refrain from providing full details. However, we remark that the optimal BCJR algorithm does not require its forward recursion as the backward recursion suffices for our case. In case the CRC is correct, then the t:th transport block is known, and the receiver may now attempt decoding of the $t + 1$:th. This, however, puts us in the same situation is in Example 1.

Example 3 (Example 2 cont.). In case the CRC this fails, the receiver may await block $t + 2$, run the L parallel soft-output algorithms and attempt a third attempt. If successful, the receiver goes back to to block $t + 1$ and attempts a decoding. For each block the receiver obtains there is more and more information about the t:th codeword. This is true even for blocks further away than the memory length of the state machine in Figure 45 (commonly referred to as decision depth in coding literature which is known to be a multiple of the memory length). However, at some point there is no further information present in subsequent blocks, and in those cases the receiver must request formal HARQ. Once that has been received, the t:th block can be decoded, and the receiver can now attempt a decoding of block $t + 1$. This also highlights the need of buffering whenever the CRC does is not correct.

Numerical results.

We consider the same setup as the one shown in Figure 46, namely: a QPSK base constellation, a rate 2/3 convolutional code identical to the one drawn in the figure, resulting in an expanded constellation

with 32 points. Moreover, a rate 3/4 LDPC code with block length 648 has been used. In Figure 47 we provide simulation results. The different curves correspond to the block error rate (BLER) at different decoding attempts. That is, the blue curve corresponds to the BLER resulting from the initial attempt – which coincides with the legacy system. The red curve is the BLER when the receiver uses blocks t and $t + 1$ for decoding of block t, and similar for the remaining curves. As a 10% BLER is the target, we may observe that using 6 blocks at 10% legacy BLER results in a BLER of 1.25%. Thus, about 85% of all HARQ transmissions can be avoided using our method. We remark that the convolutional code used is very weak, and much better performance is obtained by using a stronger code. To that end, we also show the results for a stronger code for the sixth attempt (green curve). As can be seen at 10% legacy BLER, the resulting BLER is about 0.8%, which implies that 92% of the HARQ transmissions can be avoided by our method. The particular code used in shown in Figure 48.
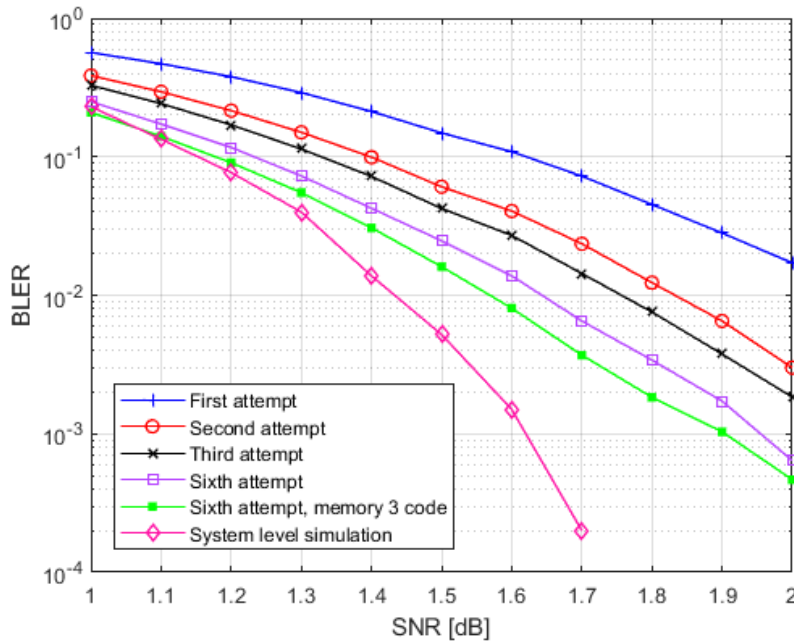
FIGURE 47. NUMERICAL RESULTS. AT THE TIME OF FIRST REFERENCING THIS FIGURE, THE READER SHOULD IGNORE THE BOTTOM CURVE (SYSTEM LEVEL SIMULATION).
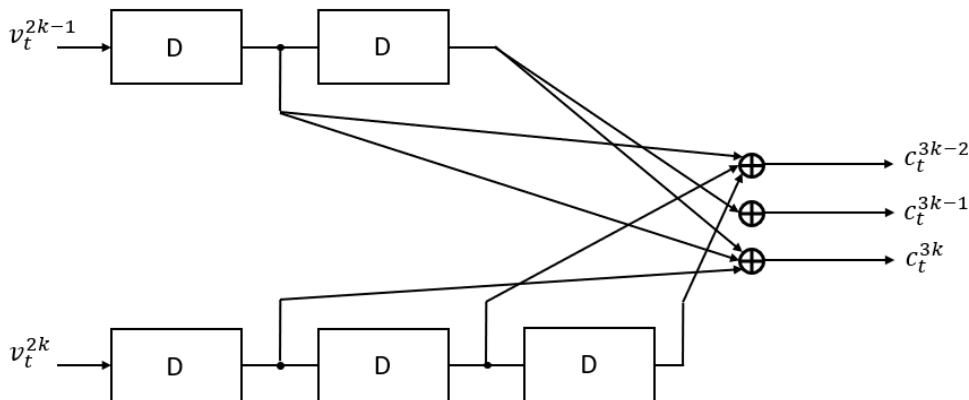
FIGURE 48. A STRONGER CODE THAN THE ONE SHOWN IN FIGURE 45. THIS PRODUCES THE GREEN CURVE OF FIGURE 47.

We next provide a simulation result for an LDPC with block length 1368 (the LDPC code is in all examples a rate 3/4 quasi-cyclic code with the same prototype matrix taken from the WLAN standard). The results

are shown in Figure 49. As can be seen, the benefit of our method remains even for (stronger) codes with larger block lengths.
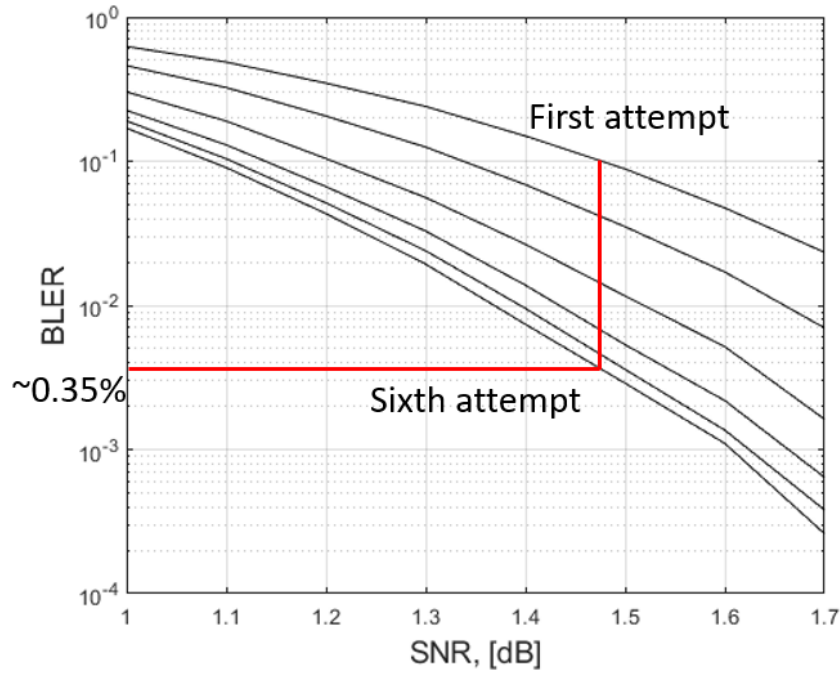
FIGURE 49. NUMERICAL RESULTS FOR A CODE WITH BLOCK LENGTH 1368.

We next consider the impact of having a decoding failure. If codeword u_t was erroneously decoded at time t, then at time $t+1$, the situation changes for the decoding of codeword u_{t+1}. So far in our exposition, it has been implicitly assumed that the decoding of u_{t+1} does not commence until the decoding of u_t is correct. This may be achieved already at time $t+1$ as previously discussed, but it could also occur at a later time; possibly even a HARQ transmission is needed. However, although the decoding of u_t is not successful at time $t+1$, it is still possible to attempt a decoding of u_{t+1}. If this fails, then we may observe also s_{t+2} and attempt a second decoding of u_{t+1}. The results of this are shown in Figure 50 for the (very weak but low complexity) code in Figure 52. As may be seen, there is a high chance of getting the decoding correct despite that u_t is not decoded, especially at later decoding attempts. This helps to decrease latency and the need of buffering as the decoding process may continue even when some blocks are in error.

For the case that both u_t and u_{t+1} are not decoded at time $t+2$, we may attempt a decoding of u_{t+2} nonetheless. The performance of this is shown in Figure 51. Again, we observe that at later stages the probability that the decoding is successful remains reasonable.

Finally, we mention that the code used, and the bitmapping are very important for performance. Herein, we have made no attempts to optimize performance whatsoever, and major improvements are likely by doing so.
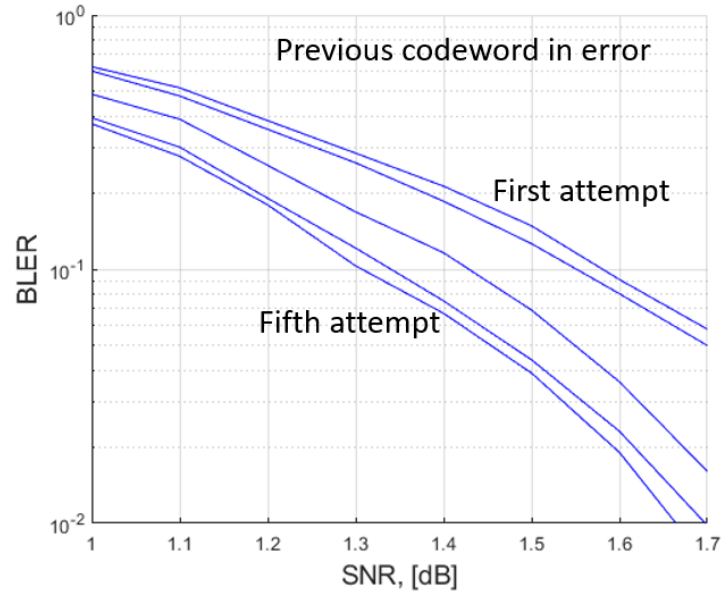
FIGURE 50. NUMERICAL RESULTS FOR THE CASE WHEN THE PREVIOUS CODEWORD WAS ERRONEOUSLY DECODED
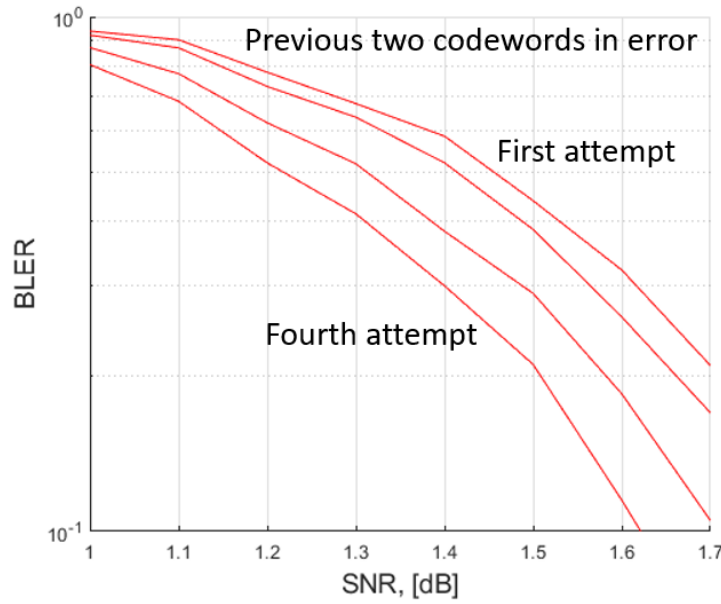
FIGURE 51. NUMERICAL RESULTS FOR THE CASE WHEN THE PREVIOUS TWO CODEWORDS WERE ERRONEOUSLY DECODED
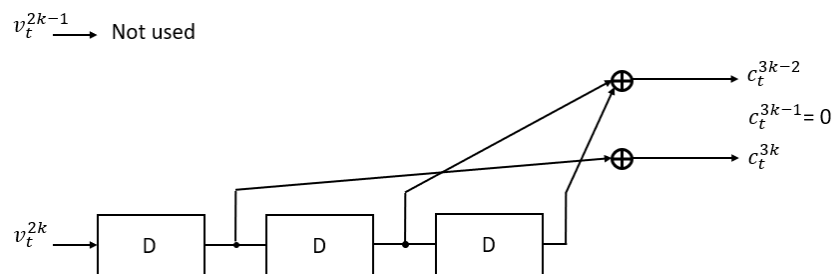
FIGURE 52. CODE USED FOR FIGURE 50 AND FIGURE 51. THIS CODE HAS ONLY 8 STATES.

System level simulation

In our next experiment, we consider a full system in which 10000 codewords have been transmitted. At each time t, the receiver attempts to decode previous codewords that have not been decoded

successfully at earlier time points, i.e., $s_t, s_{t-1}, s_{t-2}, \ldots s_{t-L}$, where L is the length of the decoding window before declaring an error. If the codeword s_{t-L} has not been decoded successfully at time t, then a HARQ request is sent for s_{t-L}. Once the HARQ is received, we consider the codeword s_{t-L} to be decoded with probability one. The receiver now backtracks to time $t - L + 1$, (i.e., it needs to buffer signals) and attempts to decode s_{t-L+1} (if it is not already decoded). This may incur a delay of $L + L_{\mathrm{HARQ}}$ where the latter parameter is the time between request of HARQ and receiving the HARQ. The results of this scheme are shown by the bottom curve in Figure 47. As can be seen, drastic improvements are obtained. At 10% BLER of the legacy system (i.e., the top curve at SNR ~1.62 dB), the BLER of the entire system is about .07%. This implies that essentially all HARQ transmissions can be avoided.

4.1.1.5   Summary and recommendations

The requirements for 6G-SHINE are very stringent notably a block error rate (BLER) below 10^{-6}, which is a significant jump from current 5G URLLC. Together with the stringent latency requirement in 100 µs, this requires every single transmission to be highly reliable and efficiently transmitted.

Proximity grouping makes use of some unique characteristics of subnetworks, since the SNEs are within a small area, by grouping SNEs that have similar radio propagation channel together so that they can collectively provide information to the HC that enables it to perform pre-emptive retransmissions and predictive scheduling. For some scenarios with known or highly predictable movements, this information can further be used to enhance predictive scheduling. It is shown that using pre-emptive retransmission on a typical TDD slot format provides at least 50% reduction in latency.

Proximity grouping also enables SNEs of the same group to eavesdrop each other's reception or transmissions thereby providing receive or transmit diversity for each packet, to improve the reliability of the transmissions between HC and SNEs.

In order to fully utilise the available resources, we introduced piggybacking on pre-emptive retransmission packet, so that even if the pre-emptive retransmission is not needed, e.g. the SNE could decode the previous transmission, the resource is not wasted as other information for a different SNE or the same SNE can be piggybacked on top of it. An overlay modulation method is introduced as one way to implement piggybacking.

For further work, the methods presented here can be recommended to standardisation bodies such as 3GPP for 6G wireless communications.

4.1.2   Configured grant enhancements

4.1.2.1   Introduction and Background

In Deliverable 3.1 [2], Configured Scheduling (CS) and its potential applications within subnetworks use cases in the industrial domain were explored. Operations such as robot control can heavily rely on deterministic, periodic communication, marking them ideal candidates for configured grant (CG)-based solutions. However, for CS to be able to serve such traffic needs it should maintain collision free resource allocation, taking into account the periodic patterns that arise and carefully split the available time-frequency grid between users; as highlighted in the previous deliverable D3.1 [2]. An equally important

challenge is that CS, due to its semi-static nature, lacks the inherent link adaptation present in its dynamic counterpart [45]. This limitation may prove especially critical when serving traffic with strict timing requirements [46]. In such cases, potential retransmissions that might come from the absence of fast link adaptation and hence decoding failures, can compromise the timely reception of packets. Alternative methods of increasing the successful reception, such as the repeated transmission of same packets might indeed increase reliability but at the expense of increased latency, which eventually makes them unsuitable for such strict demands. From the above it becomes clear that conventional CS falls short of meeting the stringent timing requirements of certain subnetwork use cases. In the following section, a Link Adaptation Aware (LAA) - Multiple Configured Grant (MCG) scheme is introduced to take advantage of the extra flexibility multiple CG configurations can provide. Differently to their conventional use [47], multiple configurations are harnessed to cover multiple channel conditions through the assignment of different MCS levels with the goal of enhancing the reception of single traffic flow SNEs. Overall, the goal is to shorten the gap from the slow LA single CG solutions provide.

4.1.2.2   Proposed LAA-MCG scheme

This subsection describes how MCGs are leveraged to address the lack of fast link adaptation in conventional configured scheduling. In standard implementations, multiple configured grants are assigned typically to serve separate traffic flows or enhance the latency of one flow under uncertain packet generation time [47]. The proposed scheme, however, builds upon the idea that in case each SNE has a unique traffic flow with fixed periodic packet generation multiple CGs can have the role of providing diverse configurations in terms of MCS assignment. Should those configurations be constructed in a way that they cover a wide range of MCS levels, which correspond to different possible channel cases, they can provide a higher level of flexibility to the otherwise semi-static CG. This flexibility can allow the SNEs to use in scenarios with unfavorable channel conditions, configurations with more conservative MCS levels, whereas more optimistic ones can be used in situations the wireless channel allows it. The HC node should construct those configurations ensuring collision-free operation regardless of which configuration each SNE may operate and with the goal of covering diverse channel conditions. Following the construction phase the configurations are distributed to the SNEs with one of them serving as the active one, which the SNE will use for transmission. Based on the channel conditions observed, the HC node is responsible for deciding the active configuration each SNE should operate under. The framework for LAA-MCG is treated in the following subsection.

4.1.2.3   LAA-MCG Scheduling

The construction of the MCGs has the overall goal of producing configuration sets that are both conflict-free and cover a wide range of channel conditions through diverse MCS assignments. To accomplish that an optimization problem is defined, which takes into account possible collisions between different configurations, from different SNEs and at the same time tries to spread the configurations each user is assigned regarding the MCS assignment. For the above reason an ideal MCS is defined which spreads the available MCS levels between configurations:

$$MCS_{\text{ideal},k} = \left\lceil MCS_{min} + \frac{k-1}{K-1} \times (MCS_{max} - MCS_{min}) \right\rceil$$

where K is the maximum number of configurations assigned and MCS_{max}, MCS_{min} are the max and min possible MCS values, typically 0 and 28. However, spreading the configurations as $MCS_{\text{ideal},k}$ dictates might be not feasible because of practical issues. The main challenge comes from the fact that lower MCS levels require more RBs to accommodate the same traffic payload. For this reason, we are looking for the MCS assignment that creates also a viable resource allocation. To accomplish this, the goal is to

minimize the distance from those ideal values in a way also that also feasible allocations are produced. More specifically, constructing the configurations involves solving the following optimization problem (P1):

$$\min_{x_{i,m,k}, RB_{\text{start},i,k}, RB_{\text{end},i,k}} \sum_{i=1}^{N} \sum_{k=1}^{K} \sum_{m \in \mathcal{M}} |m - MCS_{\text{ideal},k}| \cdot x_{i,m,k}$$

where $x_{i,m,k}$ is a binary variable representing if SNE i is assigned MCS m in configuration k.

The objective is paired to the following constraints:

- $PRB_m = \left\lceil \frac{D}{(S_m * N_s * N_{sym})} \right\rceil$, which calculates number of RBs needed for a given MCS to transmit the fixed payload.

- $RB_{\text{end},i,k} - RB_{\text{start},i,k} + 1 = PRB_{\text{mcs}_{i,k}}, \quad \forall i, \forall k. RB_{\text{start},i,k} = RB_{\text{start},i}$, which) ensures the RB allocation matches the MCS requirement.

- $\sum_{i=1}^{N} \max_{k} \{ PRB_{\text{mcs}_{i,k}} \} \leq B$, ensures that the overall RB usage does not exceed the available bandwidth B.

- $[RB_{\text{start},i,k}, RB_{\text{end},i,k}] \cap [RB_{\text{start},j,l}, RB_{\text{end},j,l}] = \emptyset, \quad \forall i \neq j$. This equation is to avoid RB overlap between different Ues.

- $\sum_{m=1}^{M} \sum_{k=1}^{K} x_{i,m,k} = K, \quad \forall i \in \{1,2,\dots,N\}$, ensures that each UE is assigned exactly K configurations.

- $\sum_{m=1}^{M} x_{i,m,k} = 1, \quad \forall i \in \{1,2,\dots,N\}, \quad \forall k \in \{1,2,\dots,K\}$, assigns an MCS value to all selections.

- $\sum_{k=1}^{K} x_{i,m,k} \leq 1, \quad \forall i \in \{1,2,\dots,N\}, \quad \forall m \in \mathcal{M}$, enforces each distinct MCS value to be present once in each UE selections to avoid redundancy and waste of selections.

P1 is classified as a Mixed-Integer Nonlinear Programming (MINLP) problem. It involves optimizing over both binary decision variables x_i, m, k, which determine the MCS assignment, and continuous variables ($RB_{\text{start},i,k}$) and ($RB_{\text{end},i,k}$) which represent the resource block allocations.

To make the problem tractable, it is transformed into a Mixed-Integer Linear Program (MILP) by introducing auxiliary variables and precomputing constant values. More specifically, the absolute deviation term $|m - MCS_{ideal,k}|$ is precomputed for all m and k as a constant $c_{m,k}$, allowing the objective function to become linear. Additionally, auxiliary variables y_i are introduced to represent the worst-case RB requirement per UE. These variables enforce the total PRB usage constraint, ensuring it remains within the available bandwidth B. With these transformations, the original nonlinear problem is reformulated into Problem P2 — a MILP — which can be efficiently solved using standard solvers such as Gurobi [48]. Since the MILP formulation remains computationally intensive for large-scale problems, the optimization is performed offline during the network planning phase, where the number of SNEs and their traffic characteristics are known in advance. Once the MCGs per SNE are constructed they are distributed to them during initial connection establishment.

In practice, the HC node keeps track through measuring the uplink channel conditions for each SNE which of the pre-assigned CGs is most appropriate at any given time. If an SNE's active configuration is no longer suitable (for example, because the channel improved and can accommodate a higher rate), the HC node sends a short control message to indicate the new configuration index to be used. The decision is based on which of the current configuration is closer to what would the HC node use if it was to schedule a new transmission.

This short message only carries the information required to switch from one set of transmission parameters to another, which is significantly less overhead than a full dynamic grant, containing frequency/time resources etc. As a result, devices can adapt to the variations in wireless medium without excessive signaling while still benefiting from the low-latency nature of semi-persistent resource allocation, in cases the channel has no significant variation and no switching is needed. In cases none of the existing configuration is appropriate for the observed channel conditions, the HC node can override the active configuration. This can happen for example in cases of extreme channel degradation or improvement, where the HC node may intervene by issuing a one-time dynamic grant as Figure 53 indicates.
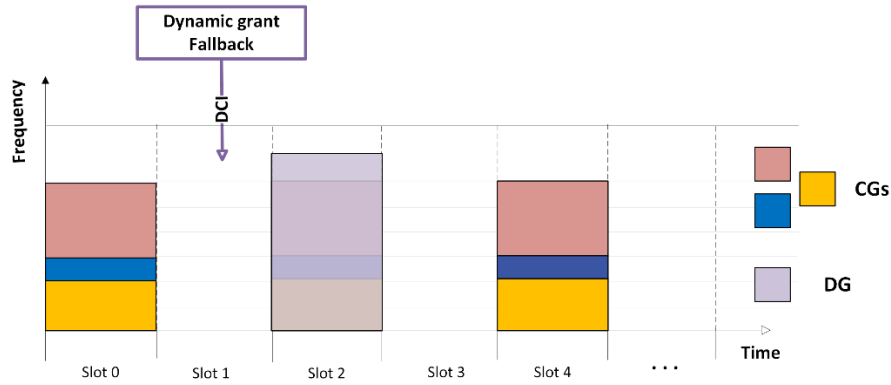
FIGURE 53. FULL DCI OVERRIDE OF CG OPPORTUNITY

In summary, the LAA-MCG scheme operates in two distinct phases: an initial configuration phase, where collision-free MCGs are assigned to each SNE during setup, and the data transmission, where the HC node monitors channel conditions for all SNEs and dynamically selects the appropriate configuration for each one.

4.1.2.4   Experimental Set-up and Performance Evaluation

To assess the proposed scheme the open source srsRAN platform was used, which provides a flexible software implementation for both the HC node (gNB) and the SNEs (UEs) [48]. In our experiments, the HC and SNEs communicate over a simulated time-frequency grid, representing an industrial wireless environment with a 10 MHz bandwidth and a carrier frequency of 1.8 GHz. Each SNE generates periodic, deterministic traffic with a fixed payload size. To accurately model the industrial environment, the uplink channel incorporates both large-scale and small-scale effects. The large-scale path loss is modeled using the 3GPP TR 38.901 Indoor Factory Dense Layout path loss model. This choice is supported by the measurements conducted in deliverable D2.3 [4],which confirmed that 3GPP InF-DL model provides a good fit for large-scale path loss in short-range industrial subnetwork scenarios. Based on this, the path loss is given by:

$$PL(d) = 18.6 + 35.7 \log_{10}(d) + 20 \log_{10}(fc) + SF ,$$

where is d the transmitter-receiver distance in meters, fc is the carrier frequency in GHz, and SF is the shadow fading component. Small-scale fading is captured using a first-order autoregressive (AR(1)) process:

$$h_{ss}(t + 1) = \rho h_{ss}(t) + \eta(t),$$

where $\eta(t) \sim N(0, 1 - \rho^2)$.

Signal processing is handled in a GNU Radio environment, and ZeroMQ is used to facilitate efficient message exchange between the HC and SNEs. The experimental setup is designed to capture key performance metrics such as goodput, latency, and signaling overhead, providing a comprehensive evaluation of the proposed scheme under realistic conditions. A summary of the experimental set-up is given by Table 10. The configuration reflects some key characteristics of industrial robot control use case, as defined in D2.2 [3]. While the selected payload size is slightly smaller than the specified for closed-loop robot control, it serves as a representative value for feedback messages in industrial environments. The carrier frequency of 1.8GHz falls within the commonly sub-6 GHz bands used in industrial set-ups. Furthermore, the choice of 52 RBs comes as a result of the 10MHz channel bandwidth combined with a 15kHz subcarrier spacing.

TABLE 10. EXPERIMENTAL SETUP PARAMETERS

SNEs N	8
Available RBs	52
Configurations K	3
Payload Size D	40
Carrier Frequency (fc)	1.8GHz
Bandwidth	10 MHz
Shadow Fading Std Dev (sSF)	7.2db
P	0.8-0.99
Noise Amplitude	100 n
Update Interval	0.01s - 0.1s

Our experimental evaluation compares the performance of the proposed LAA-MCG scheme against several benchmark schemes based on single configured grants. For clarity, the benchmark schemes are defined as follows:

- the Conservative MCS scheme employs a single configuration using the lowest MCS level, prioritizing reliability but potentially sacrificing throughput;
- the Moderate MCS scheme uses a mid-range MCS level in an attempt to balance reliability and throughput;
- and the Optimistic MCS scheme utilizes the highest MCS level to maximize throughput, though it is more susceptible to retransmissions when channel conditions degrade.

By employing LAA-MCG and solving the optimization problem (P1), all SNEs receive conflict-free configurations during the initial connection. Due to the optimization constraints needed to avoid resource allocation conflicts, the specific configurations assigned to each SNE may differ slightly. Nevertheless, the relative performance improvements provided by the LAA-MCG scheme over the single-CG schemes are consistent across all SNEs. To simplify the analysis without loss of generality, we present the performance results for a single SNE as a representative case. The assessment focuses on three key metrics: goodput, retransmission latency, and signaling overhead. Goodput is measured as the rate of successfully delivered data—excluding retransmissions—since it provides a more accurate indication of effective network performance under varying channel conditions. Figure 54 demonstrates the goodput performance as a function of average SNR. As depicted, the LAA-MCG scheme consistently outperforms single CG configurations. This enhanced performance is due to the mechanism's flexibility, which minimizes retransmissions under low SNR conditions while capitalizing on favorable channel conditions to increase throughput.
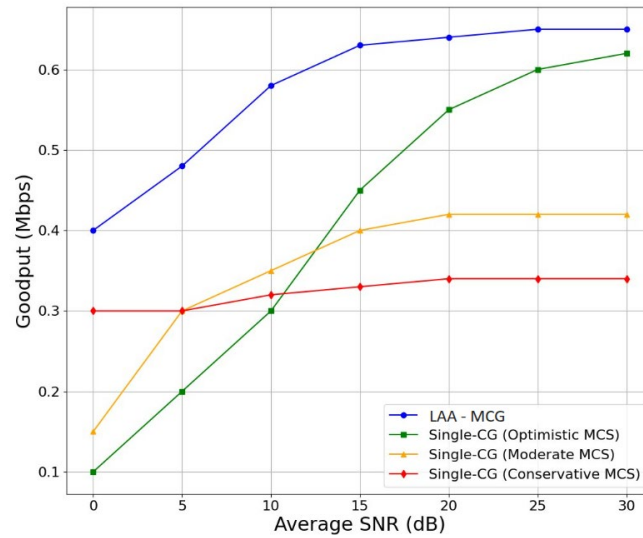
FIGURE 54. GOODPUT (MBPS) VS. AVERAGE SNR (DB) FOR MCG-RRM AND SINGLE-CG

Retransmission latency is evaluated by recording the number of transmission slots required for a packet to be successfully received. This metric captures the delay incurred due to retransmissions, which can significantly impact overall latency. Figure 55 displays the cumulative distribution of retransmission latency for the LAA-MCG scheme compared to the single-CG benchmarks across various SNR levels. In Figure 55 (LEFT), it can be seen that the LAA-MCG scheme exhibits low latency, achieved through its adaptive switching strategy—which includes full DCI overrides in challenging conditions. When the channel deteriorates significantly, this override feature enables the system to quickly switch to a much lower MCS, thereby reducing retransmissions. However, occasional retransmissions are still observed, particularly under intermediate SNR conditions. This is because the switching strategy—selecting the MCS level closest to the optimal based on absolute difference—can sometimes be overly optimistic, choosing an MCS that is slightly too high for the prevailing channel quality, leading to an initial transmission failure. Figure 55 (RIGHT) further illustrates that under these conditions, the Conservative MCS scheme, which opts for a lower MCS, tends to yield more stable latency performance. To mitigate these issues, a more sophisticated switching strategy or increasing the number of CG configurations per SNE could provide finer granularity in MCS selection, allowing LAA-MCG to more precisely track channel variations and further reduce retransmissions.
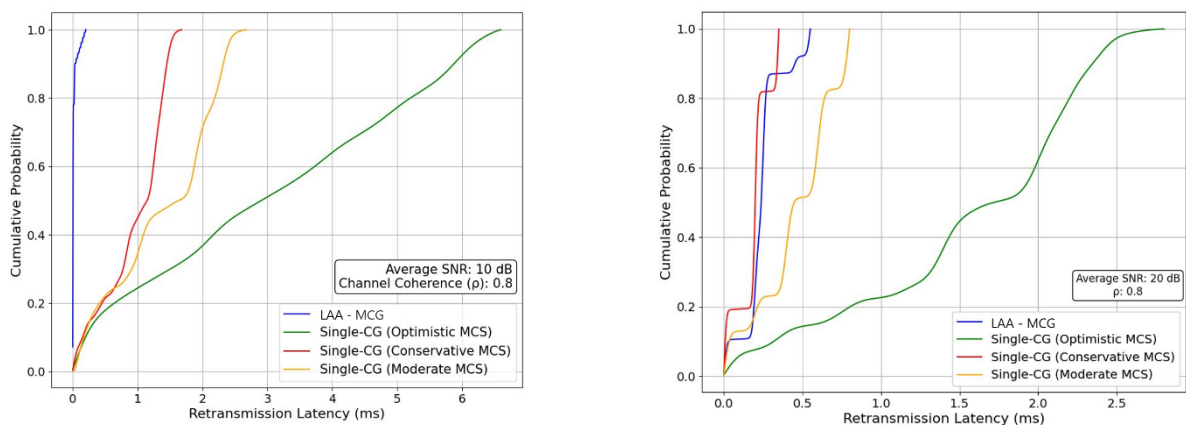
FIGURE 55. CDF OF RETRANSMISSION LATENCY FOR LOWER (LEFT) AND MIDDLE SNR (RIGHT) FOR LAA-MCG AND SINGLE-CG.

Signaling overhead is quantified by counting the control messages required for configuration updates. In the LAA-MCG scheme, overhead depends on channel fluctuations—specifically, on the channel correlation coefficient (ρ) and the channel change frequency (t). More frequent channel variations (i.e., lower ρ or higher t values) lead to more configuration switches or dynamic grants, thereby increasing the number of initial DCIs sent by the HC node. For comparison, we also evaluate a Proactive Scheduling approach, where a full DCI is transmitted at every transmission occasion regardless of channel variability. Proactive Scheduling is the version of dynamic scheduling, that still uses a full grant but without a Scheduling Request (SR) since the traffic is periodic of fixed payload. Figure 56 (a) and (b) illustrate that as channel conditions become more stable, the signaling overhead for LAA-MCG decreases significantly—thanks to fewer configuration updates and the use of shortened DCIs which uses less bits. In contrast, Proactive Scheduling maintains a constant overhead since it sends a full DCI for every transmission occasion.
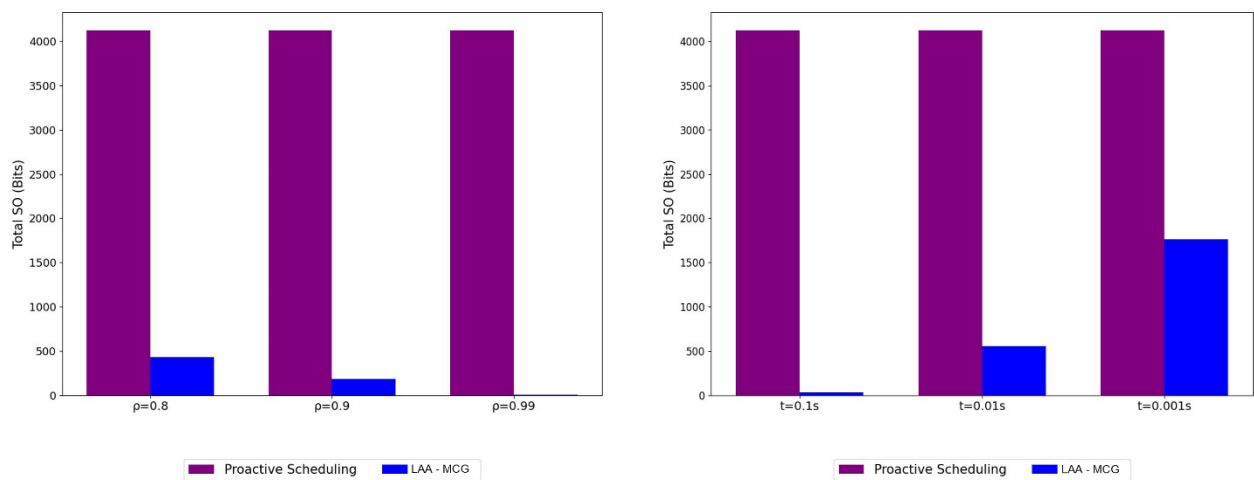
FIGURE 56. SIGNALING OVERHEAD (BITS) VS. (LEFT) CHANNEL CORRELATION COEFFICIENT P AND (RIGHT) UPDATE INTERVAL T, COMPARING LAA-MCG AND PROACTIVE SCHEDULING.

4.1.2.5   Summary and recommendations

In conclusion, LAA-MCG combines the strengths of fixed single-CG schemes and Proactive Scheduling. Its design, which leverages pre-configured grants and adaptive real-time switching, reduces retransmissions and resource waste while maintaining low signaling overhead. This flexible strategy not only enhances throughput and minimizes delays but also effectively addresses the limitations of both static and fully dynamic scheduling approaches acting as an effective middle ground. Experimental results show that LAA-MCG can increase spectral efficiency by approximately 1.5–2× compared to single-CG schemes, while also reducing worst-case retransmission latency by up to 2–3×, all with limited control overhead. These improvements are especially relevant for the robot control use case [3] which requires low latency and retransmission-free communication within strict periodic cycles. In such settings, packet loss due to MCS mismatch or channel variability can disrupt the timely delivery of packets and LAA-MCG addresses this by providing a more flexible scheduling framework.

The current implementation does not yet target the most stringent latency requirements of robot control use case since it still relies on slot-based scheduling and fixed TTI duration. However, integrating LAA-MCG with higher subcarrier spacing and mini-slot scheduling could bring it closer to the required latencies for the applicable use case. We therefore consider LAA-MCG to be a practical solution for

industrial robot control, where retransmission avoidance is critical and future enhancements can further close the gap to fully ultra-low latency operation.

4.1.3   Coded Random Access

As previously mentioned in Deliverable D3.1 [2], Coded Random Access (CRA) is a class of grant-free protocols designed to support massive access from uncoordinated SNEs. It leverages packet diversity and successive interference cancellation (SIC) to enable multi-packet reception (MPR) capabilities at the HC. In industrial subnetworks, such as those used for wireless robot control and automated unit test cells, low-latency communication is essential to ensure real-time responsiveness and operational efficiency. Meeting the stringent latency and reliability demands of these applications requires access protocols capable of supporting ultra-low communication cycle durations (e.g., 50 µs) while maintaining extremely low packet loss probabilities (e.g., preventing two consecutive errors within the same packet with a probability below 10^{-6}). To achieve this, efficient contention resolution mechanisms must be designed to rapidly adapt to varying traffic conditions without introducing excessive signalling overhead. Synchronous CRA protocols typically require SNEs to synchronize at both the frame and slot levels, using a synchronization beacon broadcast by the base station at the start of each frame. However, frame-level synchronization introduces additional signalling overhead and increases system latency, as SNEs must wait for the beginning of the next frame before initiating contention. This delay is particularly problematic in industrial applications where timely data transmission is essential. To mitigate these issues, asynchronous CRA protocols have been explored. For example, Asynchronous Contention Resolution Diversity ALOHA (ACRDA) [49] is a fully asynchronous scheme where users are misaligned at both the slot and frame levels. This approach was extended to irregular user repetition rates to enhance flexibility. Another variation, Frame-Asynchronous Coded Slotted ALOHA (FA-CSA) [50], maintains slot-level synchronization while allowing user frames to remain asynchronous. This system has been analysed for its asymptotic load threshold and packet loss rate (PLR) error floor, drawing connections with spatially coupled LDPC codes.

In this Deliverable, we extend FA-CSA protocols by incorporating packet fragmentation and erasure coding instead of multiple packet repetitions, as outlined in Deliverable D3.1. This method enhances transmission efficiency beyond what is achievable through simple packet repetition. Maximum Distance Separable (MDS) erasure codes are considered for this purpose. Additionally, we extend the system model to include orthogonal subcarriers, accounting for physical layer decoding errors in non-colliding fragments. Our main contributions include an analysis of FA-CSA's asymptotic load threshold and PLR error floor in this extended setting [51]. The results demonstrate that under sliding window (SW) decoding, the proposed scheme approaches capacity limits at high efficiencies, and the error floor analysis provides accurate performance predictions. By reducing latency and improving spectral efficiency, these enhancements to FA-CSA protocols contribute to more reliable and responsive communication in industrial subnetworks, enabling precise real-time control and monitoring of robotic systems and automated unit test cells. Furthermore, we extend our study by introducing numerical results for two additional mechanisms. The first considers an extension of the proposed scheme with retransmissions, allowing SNEs to resend unsuccessfully decoded packets to improve reliability at a cost of a slightly degraded latency. The second mechanism introduces a scheduled access approach, where coordinated transmission scheduling is performed by the HC to further optimize resource allocation and minimize contention-induced delays. These additional evaluations provide a comprehensive understanding of how different access strategies impact the performance of multiple access protocols in industrial subnetworks.

4.1.3.1   System Model

Recalling what we have already described in Deliverable D3.1, a packet erasure encoder takes k information packets, all with the same size b [bits], and generates $n > k$ encoded packets, also having size b [bits]. Encoding is performed so that the k original information packets can be retrieved from subsets of the encoded packets, any such subset having a size of at least k. MDS codes are optimum erasure codes as they hold the property that the original k packets can be retrieved from any subset of the n encoded ones. We consider a slotted system where a sufficiently high number of SNEs (hereafter also referred to as "users") wake up unpredictably to transmit data packets to the HC. In this first analysis, we assume that the channel access is contention-based, as no grant is provided by the HC, and SNE activity is fully uncoordinated. The contention phase occurs over a contention window consisting of N_S slots. Each slot is further divided into mini-slots, or "slices," forming a contention window with a total number of $N_s = N_S \cdot k$ slices. SNEs become active in specific slices and immediately contend for transmission over a defined time window, referred to as the virtual frame (VF), of length N_ℓ slices. While SNEs are synchronized at the slice level, their VF are generally not aligned in time. The number of new SNEs becoming active in a slice, independent from slice to slice, is modelled as a Poisson random variable whose mean G_s [users/slice] represents the system load. We further assume that no SNEs are active in the slice immediately preceding the contention window start and no SNEs are allowed to transmit in the last $N_\ell - 1$ slices of the contention window. SNEs employ OFDM for transmission of their packets, with a total of K subcarriers. We call "resource" a slice-subcarrier pair in the time-frequency grid. The device-side access protocol and the receiver processing are addressed in the next section.

4.1.3.2   Packet Encoding and Transmission

Each active SNE partitions its MAC layer payload, which consists of $k \cdot b$ bits, into k individual information fragments, each containing b bits. These fragments are then encoded using a (n_h, k) packet erasure component code, denoted as C_h, resulting in n_h encoded fragments. The encoding process involves randomly selecting C_h from a predefined set of component codes $C = \{C_1, C_2, \ldots, C_\theta\}$, which is shared among all SNEs and known to the HC. The rate of a given component code C_h, where $h \in \theta$, is defined as $R_h = \frac{k}{n_h}$. The code C_h is picked in C according to a probability mass function $\{\Lambda_h\}_{h=1}^{\theta}$ that is the same for all SNEs. The scheme's efficiency is $R_c = \frac{k}{\Lambda'(1)}$, where $\Lambda'(1)$ represents the expected component code length.

Each of the n_h encoded fragments is further encoded via a PHY layer code before transmission. We assume that, after PHY encoding, each fragment fits exactly one slice. In its VF, each SNE draws n_h resources uniformly at random (without replacement) from the available $k \cdot N_\ell$ ones and transmits its fragments using those resources. The transmission scheme is illustrated in Figure 57.
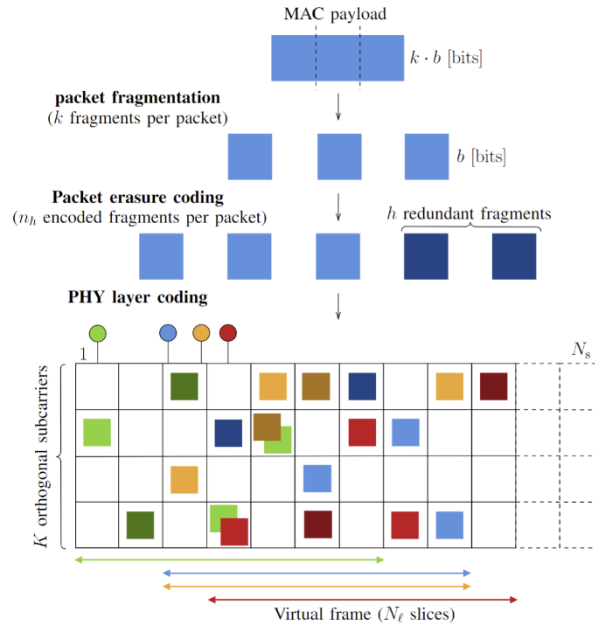
FIGURE 57. PICTORIAL REPRESENTATION OF THE PROPOSED FA-CSA TRANSMISSION SCHEME. IN THE EXAMPLE, EACH SNE SPLITS ITS PAYLOAD INTO $k = 3$ INFORMATION FRAGMENTS, WHICH ARE THEN ENCODED USING A $(5,3)$ PACKET ERASURE CODE. THE ENCODED FRAGMENTS ARE TRANSMITTED OVER A VF OF LENGTH $N_\ell = 7$ SLICES. CONTENTION WINDOW SIZE: $N_s = 12$ SLICES; NUMBER OF ORTHOGONAL SUBCARRIERS: $K = 4$.

4.1.3.3   Windowed Decoder

Let $P_{e,t}$ be the fragment decoding error probability at PHY layer, given t fragments in the resource. It depends on the PHY code protecting each fragment, on the signal-to-noise ratio, and on the PHY channel model. We assume no MPR capabilities at PHY layer, i.e., $0 \leq P_{e,1} \leq 1$ and $P_{e,t} = 1$ for all $t > 1$. The case $P_{e,1} = 0$ corresponds to the collision channel, while $P_{e,1} > 0$ allows us incorporating in the analysis PHY decoding errors on fragments that are alone in a resource.

The data recovery procedure features a combination of fragment erasure decoding and SIC across slices. The HC runs an SW decoder with size $W \geq N_\ell$ slices. Within the current window, resources are first processed in a slice-by-slice manner to recover as many fragments as possible (out of those not experiencing collisions). For $P_{e,1} > 0$, a fragment decoding error at PHY layer is assumed to be always detected, to prevent erroneous processing during the SIC phase. For each decoded fragment, the HC extracts the erasure code index h , the time-frequency position, and the pointers to the positions of the other $n_h - 1$ fragments; then it stores this information, together with the fragment samples, in a list. This way, the receiver can perform an association between fragments transmitted by the same user. The BS then attempts maximum a-posteriori (MAP) packet erasure decoding for each discovered user, exploiting the so far available fragments as input. If successful, erasure decoding recovers all fragments including, possibly, the ones to be received in slices not yet covered by the SW. The complex samples associated with each of these fragments are then subtracted through SIC from the resources in which they were received. A new iteration is then triggered, in which new fragments may be recovered by PHY decoding in resources where interference subtraction was performed, possibly yielding recovery of new packets after erasure decoding. When no more packets are recovered, the window is slid forward by one slice, and the process restarts.

4.1.3.4   Numerical Results: FA-CSA Grant-Free Protocol

Figure 58 presents the PLR as a function of the system load, denoted as G_S[users/slot], for a single-subcarrier FA-CSA system with $K = 1$, a coding rate of $R_c = \frac{2}{3}$ and a PHY layer decoding error probability of $P_{e,1} = 0$. The dashed vertical lines represent the asymptotic load thresholds obtained from density evolution, while the dashed curves illustrate the analytical error floor [51]. The results indicate that a (9,6) MDS erasure code ($k = 6$) outperforms a (6,4) MDS erasure code ($k = 4$) in both the waterfall and error floor regions. Moreover, increasing the VF size significantly reduces the error floor. For example, at a system load of $G_S^* = 0.1$, the scheme with $k = 6$ achieves a PLR of approximately $P_L^* \approx 10^{-5}$ for $N_L = 100$ and $P_L^* \approx 10^{-6}$ for $N_L = 250$. Figure 59 illustrates the packet loss rate P_L versus G_S for different values of K and $P_{e,1}$. The curves assume $k = 6$, $R_c = \frac{2}{3}$, and $N_L = 250$ slots. As expected, increasing K results in a performance boost in both the waterfall and error floor regions. Regarding the load threshold, it increases by a factor of K compared to the single-subcarrier scenario. Additionally, we highlight that the PLR can be significantly impaired by a high PHY decoding error probability $P_{e,1}$. For example, the system performance is notably affected by $P_{e,1} = 0.1$, as also predicted by the error floor analytical estimate, while already for $P_{e,1} = 0.01$ it approaches that of the $P_{e,1} = 0$ case.
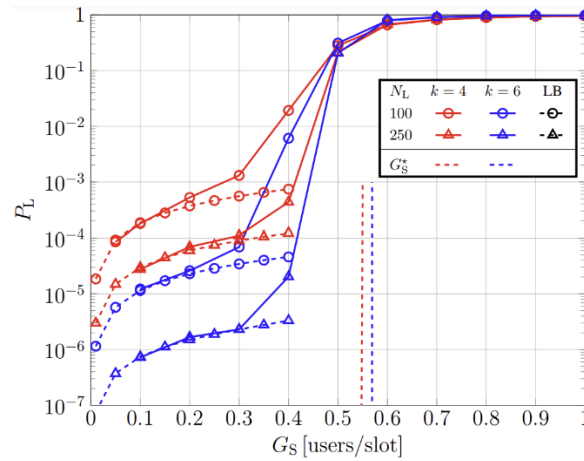
FIGURE 58. ACHIEVED PACKET LOSS RATE P_L VERSUS THE SYSTEM LOAD G_S FOR DIFFERENT VALUES OF k AND N_L. DASHED-MARKED: ERROR FLOOR PREDICTION [51].
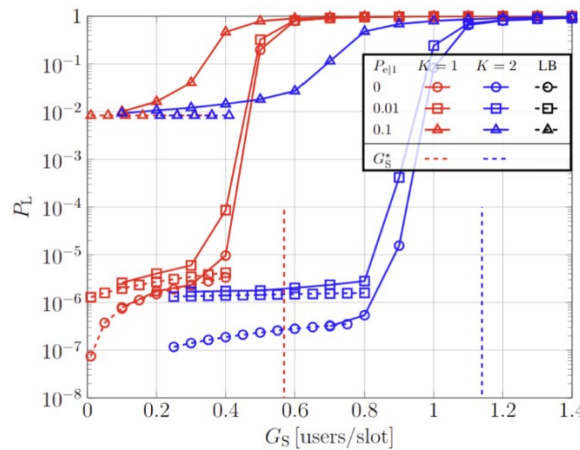
FIGURE 59. ACHIEVED PACKET LOSS RATE P_L VERSUS THE SYSTEM LOAD G_S FOR DIFFERENT VALUES OF K AND $P_{e,1}$. DASHED-MARKED: ERROR FLOOR PREDICTION [51].

4.1.3.5　Numerical Results: FA-CSA with Retransmissions Protocol

Figure 60 illustrates the performance curve of an FA-CSA system implementing a retransmission strategy. This strategy imposes that when a user packet is lost (i.e., the packet is not decoded and the SW decoder has moved beyond the user's VF), the user immediately retransmits the packet after the first loss. To enhance the recovery of retransmitted packets, given that the PLR target concerns the loss of two consecutive user packets, multiple simulations are conducted, gradually increasing the number of redundancy packets during retransmission h_{Rtx}. The performance curve is evaluated under the assumption of $k = 4$ information fragments, $N_L = 100$ slots, $K = 1$ subcarrier, and a fragment decoding error probability of $P_{e,1} = 0$. The performance improvement due to retransmission is evident in both the waterfall region (medium-to-high PLR values) and the error floor region (low-to-extremely-low PLR values). Compared to the baseline scenario, the error floor is significantly reduced. This is because the retransmission strategy accounts for packet loss events occurring due to two consecutive decoding failures of the same user packet, ensuring higher reliability.
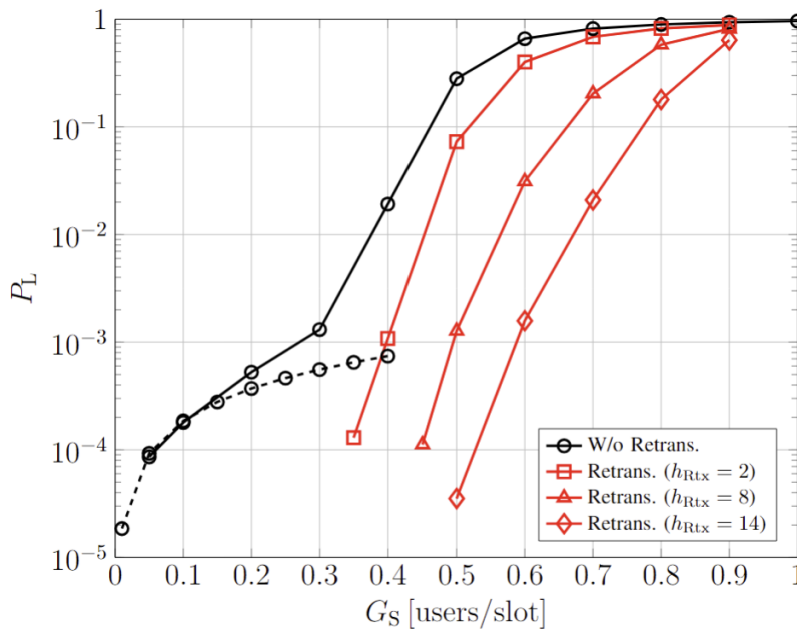
FIGURE 60. ACHIEVED PACKET LOSS RATE P_L VERSUS THE SYSTEM LOAD G_S FOR THE SCHEME WITH AND WITHOUT RETRANSMISSIONS.

4.1.3.6　Numerical Results: FA-CSA Scheduled Protocol

In the FA-CSA scheduled algorithm, the HC dynamically allocates resources to user fragments following the procedure outlined in Figure 61. The HC operates with precise knowledge of each user's activation time within the contention window and the real-time state of that window. Resource allocation begins from the activation index slice of each user, selected within the allowed range $\{1, ..., N_s - N_\ell\}$. The algorithm ensures that each user transmits the required number of encoded fragments n, while allowing controlled overlap with other users to maximize resource efficiency. Allocation is performed on a resource-by-resource basis, where each resource is defined as a slice-subcarrier pair. Within each slice, the algorithm iterates over all subcarriers, verifying their availability. A fragment is allocated if:

1. The subcarrier is unoccupied (fragPerRes=0), or

2. The remaining available resources in the frame satisfy the latency constraint imposed on transmitting users.

To maintain fairness with the FA-CSA grant-free counterpart, the maximum latency is set equal to the VF length N_ℓ, ensuring alignment with system constraints. If a subcarrier is already occupied, the algorithm allows allocation only if the number of overlapping fragments remains below the predefined threshold h. This controlled overlap prevents excessive interference while maintaining high spectral efficiency. The allocation process iterates until all user fragments are assigned, moving to the next slice when necessary. This adaptive strategy ensures efficient resource utilization, low-latency communication, and robust contention resolution, making it well-suited for high-reliability applications.

Algorithm FA-CSA Scheduled

1: **for** $u = 1$ to N_{users} **do**
2:     idxSlice $\leftarrow$ idxStartSlice$(u) = s \in \{1, \ldots, N_\text{s} - N_\ell\}$;
3:     nFragmentsAllocated $\leftarrow 0$;   nFragmentsOverlapped $\leftarrow 0$;
4:     **while** nFragmentsAllocated $< n$ **do**
5:         **for** idxSubcarrier $= 1$ to K **do**
6:             **if** fragPerRes(idxSlice, idxSubcarrier) $= 0$ **or**
7:     $(N_\ell + \text{idxStartSlice} - \text{idxSlice} - 1) \cdot K + (K - \text{idxSubcarrier} + 1) \le (n - \text{nFragmentsAllocated})$ **then**
8:                 nFragmentsAllocated++;
9:                 fragPerRes(idxSlice, idxSubcarrier)++;
10:             **else if** fragPerRes(idxSlice, idxSubcarrier) $= 0$ **and** nFragmentsOverlapped $< h$ **then**
11:                 nFragmentsAllocated++;   nFragmentsOverlapped++;
12:                 fragPerRes(idxSlice, idxSubcarrier)++;
13:             **end if**
14:         **end for**
15:         idxSlice++;
16:     **end while**
17: **end for**

FIGURE 61. FA-CSA SCHEDULED ALGORITHM.

Figure 62 presents the PLR performance of the FA-CSA scheme, comparing its grant-free and scheduled variants. The red curves represent the baseline FA-CSA grant-free scheme, where each user transmits $n = 6$ encoded fragments, with $k = 4$ being information fragments and $h = 2$ being redundant fragments. The system operates with a single subcarrier ($K = 1$), and the PHY layer fragment error probability is set to one, implying that decoding relies entirely on the protocol's redundancy and collision resolution mechanisms to recover lost fragments. The blue curves illustrate the performance of the FA-CSA scheduled variant, where the HC dynamically allocates resources following the algorithm described above. Two separate simulations are conducted, each considering different values of the VF length N_ℓ, which directly impacts the system's latency and scalability. As observed in the simulation results, the scheduled FA-CSA protocol achieves superior PLR performance compared to its grant-free counterpart, particularly in the waterfall region, where the error rate sharply decreases after a certain load threshold. This improvement is attributed to the coordinated scheduling mechanism, which effectively reduces contention and enhances fragment allocation efficiency. The scheduled FA-CSA variant exhibits a threshold phenomenon, meaning that the system performs optimally up to a specific load limit, beyond which performance rapidly degrades, leading to a sharp increase in the PLR. The VF length N_ℓ plays a crucial role in determining system performance, as it affects both latency requirements and scalability.

If low latency is a priority (i.e., a smaller N_ℓ), the FA-CSA scheduled variant experiences reduced scalability, meaning the system can support fewer users before reaching the threshold. Conversely, if the system allows for a larger N_ℓ (i.e., more transmission slices per SNE fragments transmission), scalability improves, enabling the system to support higher loads while maintaining low PLR. The results confirm that the FA-CSA scheduled protocol outperforms its grant-free counterpart under moderate loads, making it a viable solution for latency-sensitive and high-reliability applications. However, its performance is constrained by the VF length, which introduces a trade-off between latency and scalability. System design must carefully consider N_ℓ based on the application's latency constraints and user load requirements to achieve optimal performance.
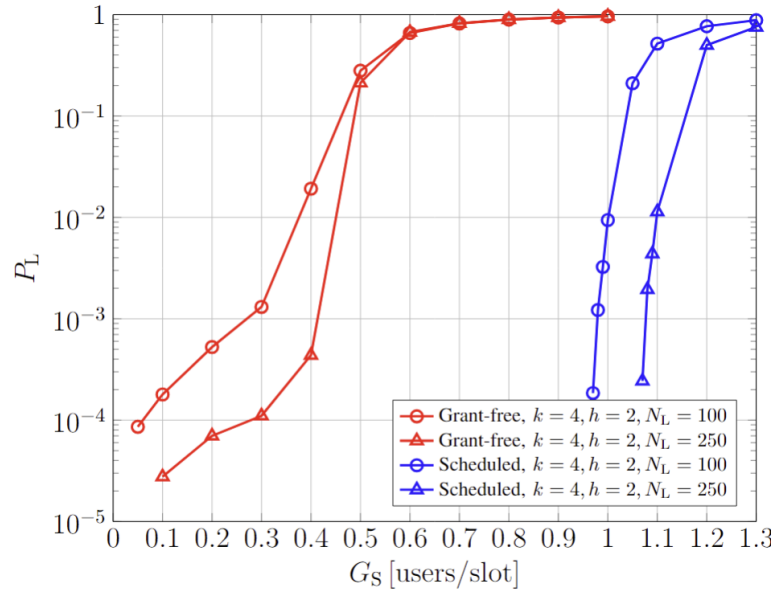
FIGURE 62. ACHIEVED PACKET LOSS RATE P_L VERSUS THE SYSTEM LOAD G_S FOR THE GRANT-FREE AND SCHEDULED PROTOCOLS.

4.1.3.7 Summary and recommendations

The proposed FA-CSA protocol enables low-latency, ultra-reliable communication in industrial subnetwork scenarios by incorporating packet fragmentation with MDS erasure coding and orthogonal subcarrier support. Analytical and numerical results demonstrate that the scheme supports communication cycles as short as 50 µs and meets reliability target by keeping the probability of two consecutive packet errors below 10^{-6}. Additional evaluations of retransmission strategies and scheduled access mechanisms highlight further trade-offs between latency, reliability and coordination overhead. Adoption of FA-CSA with erasure coding and sliding window decoding is recommended for industrial subnetwork applications with stringent latency and reliability requirements, with retransmissions or scheduled access employed as needed to meet specific performance targets.

4.2 MULTIPLEXING SERVICES WITH DIVERSE REQUIREMENTS IN IN-X SUBNETWORKS

4.2.1 Full/flexible duplexing for enhanced scheduling flexibility

In deliverable D3.1, the opportunity of using flexible duplexing -enabled HCs as a solution for supporting services with diverse characteristics was discussed. Flexible duplexing HCs are able to simultaneously transmit and receive radio signals over different frequency subbands, thus enabling the support of

different UL/DL switching point in a same subnetwork. Flexible duplexing enabled HCs suffer from different types of interference, i.e. self-interference due to the energy leakage between the DL transmission over the subbands where UL transmissions are allocated; and the UL-to-DL interference, generated by the transmissions of SNEs transmitting over adjacent bands to the DL SNEs served by the HC. For a more detailed description of the different interference sources in flexible-duplexing enabled subnetworks, we refer to Deliverable D3.1.

Self-interference can be counteracted via analog and digital interference cancellation at the HC; transmission and reception over different subbands, as well as the low transmit power, make such process easier compared to full-duplexing transceivers, and high-power macrocell base stations, respectively.

On the other hand, the UL-to-DL interference, also known as cross-link interference, might be significantly affecting the performance, especially in case of SNEs simultaneously active in UL and DL and located in very close proximity within the already small subnetwork area. This is pictorially illustrated in Figure 63. There is therefore the risk that the in-band emission (IBE) generated by the UL SNEs affects significantly the communication quality of the DL SNEs.
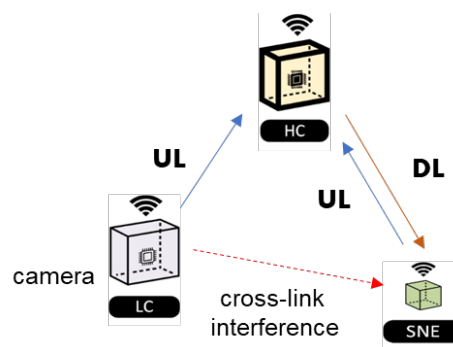
FIGURE 63. CROSS-LINK INTERFERENCE IN FLEXIBLE DUPLEXING ENABLED SUBNETWORKS

We consider here two potential solutions for reducing the impact of IBE:

- Ensuring a sufficient frequency separation between the UL SNEs and DL SNEs. The scheduler should therefore avoid to allocate over neighbor subbands such UL and DL transmission, in case of critical communication in the DL.
- UL and/or DL power control. Adjusting the transmit power can help mitigating the cross-link interference.

The 3GPP has defined specific requirements for the IBE (see Table 6.4.3.2-1 in TS 36.521 [52]), that includes requirements for general leakage outside the allocated resource block, IQ image, and carrier leakage, with respect to the power of the UL allocation. The general leakage component is a function of the EVM requirements. Higher order modulations (e.g., 64QAM) have tighter EVM requirements than low-order modulations (e.g., QPSK). Figure 64 shows the IBE limit per RB as a function of the frequency separation from the UL allocation, measured in terms of number of RBs (ΔRB), considering different frequency allocations (50 and 100 RBs), and different used modulation order in the UL. Higher order modulations have tight EVM requirements, translating to lower spurious and tighter IBE limits in the adjacent RBs. This is because the usage of high order modulations requires higher linearity receivers to be correctly detected, and this turns out to also be beneficial for the devices operating in the neighbor subbands thanks to the reduced spurious, especially in the case of low number of allocated RBs.
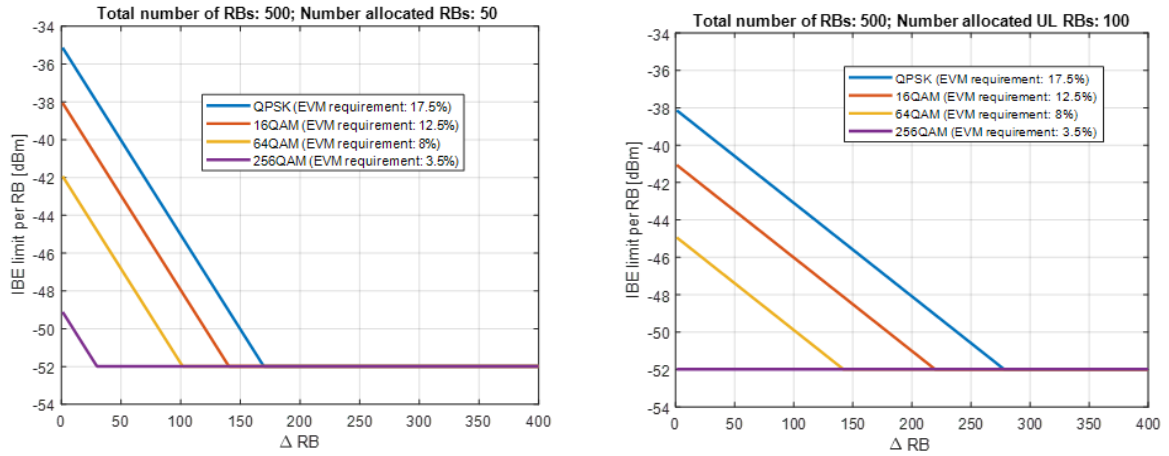
FIGURE 64. IBE LIMITS ACCORDING TO 3GPP TS 36.521

Such result suggests that, in case the device transmitting in the UL is operating with a high order MCS, the benefit of frequency separation with the DL allocation are likely to be limited, as the device has to fulfill tighter requirements.

Let us consider the visual inspection cell use case, where a LC can transmit a video feed in the UL while the HC can communicate in the DL to its sensors/SNEs. Given the high data rate requirements of the video transmission, and the advantageous propagation condition in a short-range subnetwork, the LC will likely be operating with a high order MCS; this will enforce tight IBE requirement and therefore the impact of frequency separation of the DL transmissions will be lower.

We therefore propose to address the problem of mitigation of cross-link interference due to IBE via power control. In the following, different power control strategies are proposed, respectively for UL, DL, and a joint UL/DL solution.

DL power control
The rationale behind the usage of DL power control, is to increase the transmit power at the HC, such that the energy leakage from the UL transmissions in neighbor subbands can be reduced. Our proposed DL power control strategy consists of the following steps:

- The victim SNE estimates the average energy leakage $\delta_{leakage}$ experienced in his allocated subband, and report it to the HC
- The HC **increases** its transmit power of an amount equal to $\Delta P_{DL} = \min(\frac{\sigma_n^2 + \delta_{leakage}}{P_{R,DL}} \gamma_t \rho_\gamma, \ \Delta_{max,DL})$,

where σ_n^2 is the noise power, $P_{R,DL}$ is the receive power in the DL (before correction), γ_t is the SINR target, $\Delta_{max,DL}$ is the maximum accepted power increase, and ρ_γ is a system margin which is empirically set to take into account transceiver non-idealities.

UL power control

UL power control aims at reducing the energy leakage generated by the SNEs active in the UL towards the SNEs receiving data from the HC, by reducing the UL transmit power. Our proposed UL power control consists of the following steps:

- The victim SNE estimates the average energy leakage $\delta_{leakage}$ experienced in his allocated subband, and report it to the HC
- The HC instructs the LC to **reduce** its transmit power of an amount equal to

  $$\Delta P_{UL} = \min\left(\frac{P_{R,DL} - \sigma_n^2 \gamma_t \rho_\gamma}{\delta_{leakage} \gamma_t \rho_\gamma}, \Delta_{max,UL}\right)$$

where σ_n^2 is the noise power, $P_{R,DL}$ is the receive power in the DL, γ_t is the SINR target in the DL, $\Delta_{max,UL}$ is the maximum accepted power decrease.

Joint UL/DL power control

A joint proposal for UL/DL power control consists of the following steps:

- The victim SNE estimates the average energy leakage $\delta_{leakage}$ experienced in his allocated subband, and report it to the HC
- The HC increase its transmit power of an amount equal to $\Delta P_{DL} = \min\left(\frac{\sigma_n^2 + \delta_{leakage}}{P_{R,DL}} \gamma_t \rho_\gamma, \Delta_{max,DL}\right)$, where σ_n^2 is the noise power, $P_{R,DL}$ is the receive power in the DL (before correction), γ_t is the SINR target and Δ_{max} is the maximum accepted power increase.
- If $\Delta P_{DL} = \Delta_{max,DL}$ (i.e., correcting the power in the DL was not sufficient since the maximum accepted power increase is already reached), the HC instructs the LC to reduce its transmit power of an amount equal to $\Delta P_{UL} = \min\left(\frac{P_{R,DL} \Delta_{max,DL} - \sigma_n^2 \gamma_t \rho_\gamma}{\delta_{leakage} \gamma_t \rho_\gamma}, \Delta_{max,UL}\right)$

4.2.1.1 Performance evaluation

We evaluate the performance of the proposed power control solutions by considering a 5x5 m^2 subnetwork, where HC, LC and SNE are randomly deployed with a minimum mutual distance of 0.5 m. We refer to a typical visual inspection cell use case as defined in D2.2, with a combination of high throughput video feeds with low data rate communication links with sensors. Specifically, we consider an LC device (e.g., camera) transmitting a video feed in the UL to an HC, which is also involved in a DL communication with one or more sensors. In order to avoid a single UL/DL switching point which will force a drastic reduction of the UL throughput, the HC is flexible-duplexing enabled and multiplex its UL and DL transmissions over neighbour subbands, as shown in Figure 65.
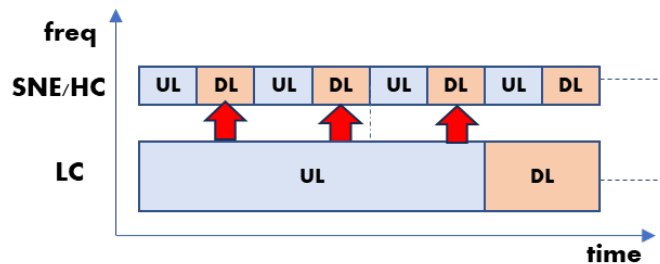
FIGURE 65. MULTIPLEXING OF PERIODIC TRAFFIC WITH UL-SKEWED TRAFFIC IN A FLEXIBLE DUPLEXING-EQUIPPED SUBNETWORK

We also consider the following simulation parameters:

- Overall transmission bandwidth: 90 MHz (500 RBs)
- Carrier frequency: 10 GHz (FR3)
- Lognormal shadowing with 7 dB standard deviation (aligned with findings in D2.3)
- HC transmission (DL)
    - Packet size: 100 bytes (from visual inspection cell use case)
    - DL slot duration: 142 µs (minislot with 2 OFDM symbols)
    - Bandwidth: 9 MHz (50 RBs with 15 kHz subcarrier spacing)
    - 20% pilot/control overhead
    - Ideal spectral efficiency target: 0.8333 bps/Hz
    - Transmit power (before control): 0 dBm,
    - $\Delta_{max,DL}$ = 23 dB
    - ρ_γ =3
- LC transmission (UL)
    - Data rate: 50 Mbps (from visual inspection cell use case)
    - Bandwidth: 18 MHz (100 RBs with 15 kHz subcarrier spacing), unless differently specified
    - 20% pilot/control overhead
    - Ideal spectral efficiency target: 6.9 bps/Hz
    - Transmit power (before control): 0 dBm
    - $\Delta_{max,UL}$ = 23dB
- In modelling IBE, carrier leakage at centre frequency and IQ image are disregarded → it is assumed that DL resources are allocated such that carrier leakage and IQ image are avoided

- Self-interference at the HC also modelled as in [53], assuming that a combination of antenna separation and analog/digital cancellation techniques ensure a reduction value of above 100 dB

In-band emission is modelled as a Gaussian random variable with zero-mean and standard deviation equal to the IBE limit; note that this may represent a worst-case scenario since devices might be designed with a given margin with respect to such limits.

Figure 66 displays the CDF of the DL spectral efficiency for the proposed power control strategies, assuming different values for the spacing ΔRB between the simultaneous DL and UL transmissions, measured in number of RBs. Power control strategies bring significant benefits in terms of achievable spectral efficiency. DL power control clearly outperforms UL power control; this is due to the higher benefits of controlling the DL receive power at the SNE with respect to reducing the leakage from the neighbor band. The joint UL/DL power control solution further improves the achievable spectral efficiency at the very low percentile, highlighting that the additional benefits of UL power control are more visible when combined with an increase in DL power for the SNE of interest. Interestingly, ensuring a large frequency separation between UL and DL transmissions seems to be beneficial only for the case of DL power control when looking at the very low percentiles of the distribution. In the case of UL power control, benefits are instead visible at the median values of the distribution. Since the focus for the DL transmissions to the sensors is to ensure high reliability at the low percentiles, frequency separation appears then to have a negligible impact in case a joint UL/DL power control strategy is adopted. However, the spectral efficiency target can be achieved only at a probability above 10[-4], suggesting the need of a retransmission to cope with a 10[-6] target. The figure also shows the case where the UL transmission allocation is 50 RBs rather than 100 RBs as in the other cases; improvements in spectral efficiency are visible in this case. When 50 RBs are used, the spectral efficiency target is now 6.9 bits/s/Hz, translating to the usage of high order MCSs that require tighter EVM restrictions on the UL, translating to lower IBE in neighbour subbands, as also shown in Figure 66. This suggests that the usage

of a smaller bandwidth combined with a high MCS can improve performance of the DL low-data rate traffic in a neighbour subband.

Figure 67 displays the CDF of the UL spectral efficiency. Here, the scale is linear since for best effort high data rate type of traffic, median and percentiles not lower than 5% are of interest for assessing performance. While UL power control inevitably leads to a spectral efficiency loss, joint UL/DL power control allows to fully compensate for such loss, leading to the same performance as when no power control strategy is used. In any case, performance at the median and ~5% of the distribution are well beyond the spectral efficiency targets. Note that, in a practical system, maximum data rates can be capped by the highest order modulation and coding scheme adopted, e.g. 10 bits/s/Hz in case of 1024QAM.
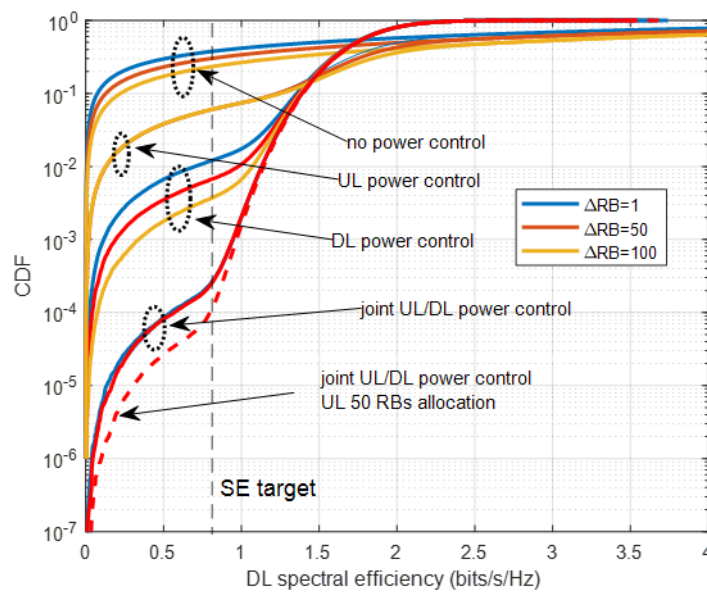
FIGURE 66. DL SPECTRAL EFFICIENCY FOR FLEXIBLE-DUPLEXING ENABLED OPERATIONS.
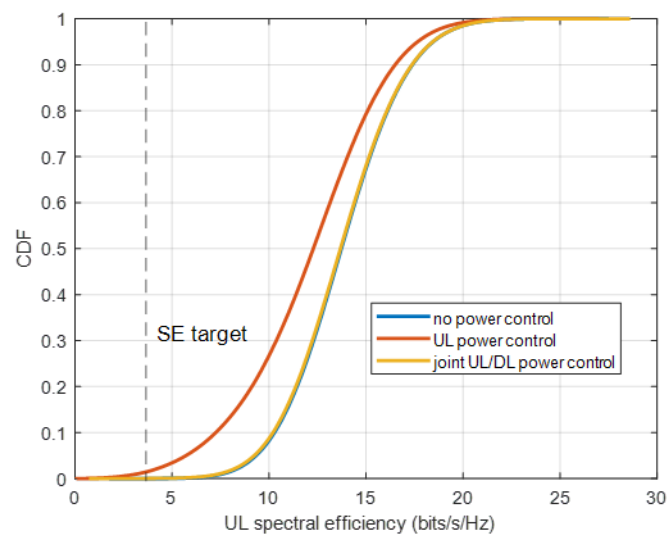
FIGURE 67. UL SPECTRAL EFFICIENCY FOR FLEXIBLE-DUPLEXING ENABLED SUBNETWORKS

4.2.1.2 Summary and recommendations

In this subsection, we have studied the performance of flexible duplexing as a solution for enabling simultaneous support of UL and DL traffic with different characteristics by avoiding the usage of a same UL/DL switching point. IBE generated by the energy leakage of the UL transmissions to the DL subbands is a major performance-limiting factor. We have proposed the usage of different power control techniques for reducing the impact of such interference, with a joint UL/DL solution being the most effective. Frequency separation appears to have a negligible impact on the performance in case a joint UL/DL power control is in place. Also, DL spectral efficiency can improve when a high order MCS is used in the UL transmissions, since this translates to tighter EVM requirements and lower IBE.

It is therefore recommended to apply a joint UL/DL power control for flexible-duplexing equipped in-X subnetworks, with the usage of as smaller as possible UL transmit bandwidth combined with a high order MCS.

4.3   PREDICTIVE SCHEDULING

The predictable nature of industrial systems (such as robotic movements) and the data traffic generated within vehicles presents an opportunity for subnetworks to utilize predictive strategies for more efficient resource allocation. The solutions presented in this section apply channel state information prediction to support robotic control, and utilize camera-generated data in vehicular contexts, incorporating AI/ML techniques to improve context awareness and mobility management.

4.3.1   Predictive scheduling and pre-allocation mechanisms

In dynamic industrial environments, and use cases such as robot control, scheduling communication between moving SNEs within a subnetwork that communicates with a HC poses significant challenges. One of the primary challenges is the potential for communication blockage caused by the movement of SNEs, which can lead to disruptions and decreased reliability. To mitigate these issues, employing diversity techniques to avoid collision becomes a practical solution as shown in Figure 68. Building on the concept of cooperative communication introduced in Deliverable D3.1, particularly in subsection 3.4.1 on Cooperative in-X Communication and [54], we previously demonstrated how utilizing multiple LCs enhances reliability in static scenarios.

In this section, we extend this model to account for dynamic scenarios where SNEs are mobile and follow predefined trajectories. In such dynamic environments, efficient resource allocation depends on the timely and accurate knowledge of Channel State Information (CSI). Since the CSI is time-varying due to the mobility of SNEs, it is crucial to proactively manage resources by predicting the CSI for upcoming time slots. This foresight allows for more informed decision-making in relay selection and SNE association, which are vital for maintaining communication reliability and efficiency.

To achieve this, we propose a Long Short-Term Memory (LSTM) based approach for predicting future CSI values, leveraging the temporal correlations present in the channel data [55][56][57]. The predicted CSI is then used as input to the resource allocation algorithm previously developed in Deliverable D3.1, enabling proactive scheduling and pre-allocation of resources. By anticipating CSI variations, the system can adaptively select relays, optimize SNE associations, and allocate the time/frequency resources thereby enhancing communication performance. Our motivation for the usage of LSTM stems from the fact that CSI can be modelled as a time-correlated stochastic process due to user mobility, scattering

environments, and Doppler effects. LSTM networks, with their ability to learn both short-term and long-term dependencies in sequential data, are well-suited for capturing such temporal correlations.

We evaluate the effectiveness of our predictive approach by comparing the overflow rate and outage probability when using predicted CSI versus historical (outdated) CSI. Our hypothesis is that leveraging predicted CSI allows for more precise resource allocation, leading to improved network performance compared to the conventional approach that relies on outdated CSI over the prediction horizon. This analysis highlights the advantages of integrating LSTM-based CSI prediction into predictive scheduling and pre-allocation mechanisms for dynamic industrial communication subnetworks.

In addition to performance gains, predictive CSI reduces channel overhead by increasing the sampling interval for channel estimation. By sending pilot signals less frequently to the SNE, the HC can allocate more resources to data transmission, enhancing overall efficiency. The general workflow of our predictive scheduling framework is illustrated in Figure 69. Initially, the HC collects historical CSI from all links. This data is then used to predict future CSI samples. Once the CSI predictions are updated, a relay selection and resource allocation algorithm schedules transmissions by assigning resource blocks to each SNE. Data transmission then proceeds according to the predefined protocol.

If real-time channel estimation is required in each time frame, pilot sequences are transmitted to update the historical CSI. However, the frequency of pilot transmission depends on the prediction accuracy, as more reliable CSI predictions allow for less frequent channel estimation. The dashed box representing pilot transmission and channel estimation in Figure 69 indicates that these steps are not always necessary and can be performed less frequently due to the availability of predicted CSI.
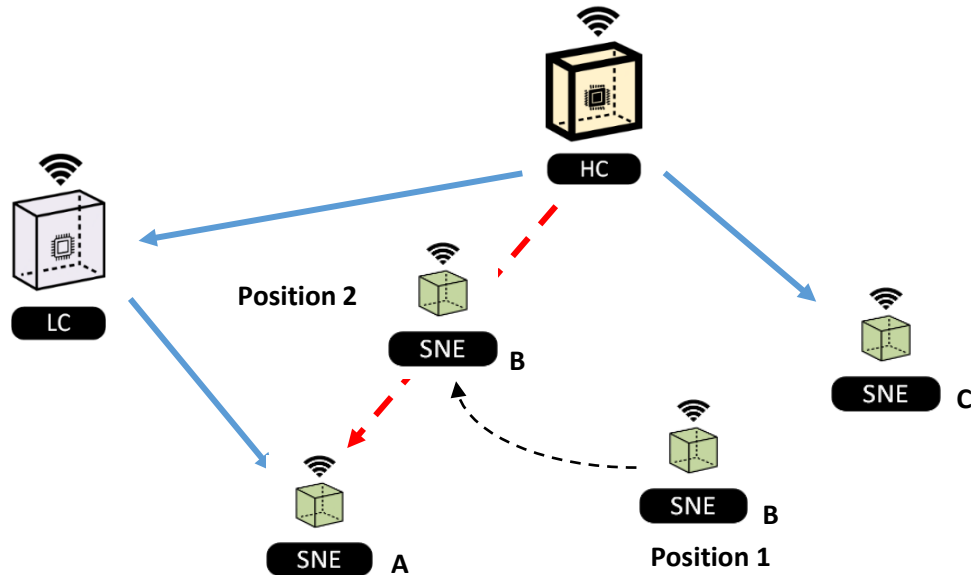
FIGURE 68. SYSTEM MODEL DEPICTING BLOCKAGE IN A DYNAMIC SUBNETWORK OF MULTIPLE MOBILE SNEs
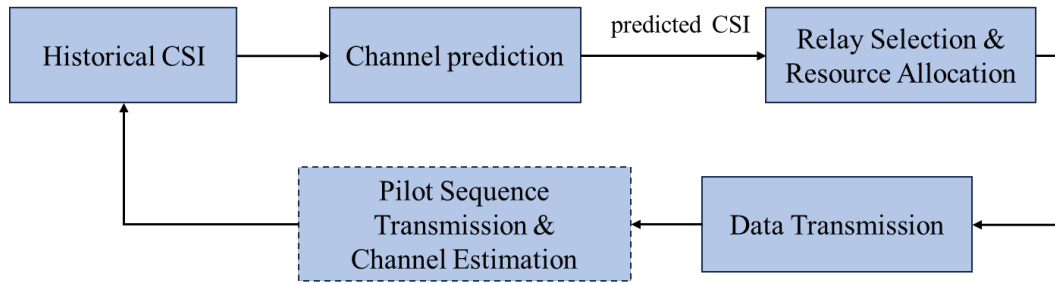
FIGURE 69. PREDICTIVE CSI-DRIVEN RESOURCE ALLOCATION

4.3.1.1 LSTM-Based CSI Prediction

This section explores the use of LSTM networks to predict future CSI values, with the objective of evaluating the model's performance against a benchmark that relies on outdated CSI values. The motivation for using LSTM stems from its ability to capture temporal dependencies, which are inherent in wireless communication channels due to their time-varying nature.

In this study, we make two key assumptions. First, the channel is sampled at intervals of 0.5 ms, allowing us to capture rapid fluctuations in the channel state, assuming velocity ranges as defined for devices in the robot control use case presented in D2.2. Second, the temporal correlation of the channel is governed by its coherence time, which defines the time duration over which the channel can be considered approximately constant. These assumptions provide a realistic basis for predicting future CSI values, as the temporal correlation directly influences the predictability of the channel.

To leverage the temporal dependencies in the CSI data, we use an input sequence length of 20, corresponding to the past 10 ms of channel observations. This sequence length is chosen to provide sufficient historical context for the LSTM to learn temporal patterns and dependencies. The model is tasked with predicting future CSI values for horizons ranging from 1 ms to 4 ms, which corresponds to 2 to 8 samples ahead, with increments of 0.5 ms. This approach allows us to assess the model's predictive accuracy over short-term intervals, which is crucial for applications requiring real-time channel estimation and adaptation.

The channel environment is modelled using a Rician fading profile, reflecting a scenario where a dominant line-of-sight component is present alongside scattered multipath components. This model is particularly relevant for in-factory communication scenarios aligned with 3GPP standards [57] and deliverable D2.3. Additionally, the Doppler frequency, which influences temporal correlation, is determined by the relative velocity between the transmitter and receiver. Different SNE velocities, such as 5-20 m/s, are considered to evaluate the model performance under varying mobility conditions. By modelling temporal correlation, we account for the impact of coherence time and Doppler frequency on the predictability of the channel, thus enabling the LSTM to learn complex temporal dependencies.

To train and evaluate the LSTM model, the CSI dataset is divided into training, validation, and testing sets, comprising 15,000, 2,000, and 3,000 samples, respectively. Input sequences are constructed using sliding windows of 20 samples, which represent the past 10 ms of CSI data. Corresponding output sequences are generated for prediction horizons ranging from 2 to 8 samples, allowing the model to learn mappings from historical data to future CSI values. This data preparation strategy ensures that the model is trained on diverse temporal patterns, enhancing its generalization capability.

For benchmarking, we compare the LSTM model's performance against an outdated CSI approach, which simply uses the last observed CSI value as the prediction for all future time steps. This benchmark assumes channel invariance over the prediction horizon and serves as a baseline to evaluate the predictive advantage of the LSTM model. By contrasting the LSTM's performance with this naive approach, we can quantify the benefits of leveraging temporal correlations for CSI prediction.

LSTM Model Architecture

The LSTM model architecture is designed to effectively capture temporal dependencies in the input sequences. It consists of the following key layers:

- Sequence Input Layer: Accepts time-series CSI data as input, with a shape of (20 time steps, 1 feature). We selected 20 time steps based on the coherence time of the channel in our simulation setup. This length offers a balanced trade-off between capturing sufficient temporal dynamics and maintaining manageable computational complexity. Longer sequences were also tested, but did not yield significant improvements in prediction accuracy.
- LSTM Layer (100 units): Extracts temporal patterns from input sequences using 100 hidden units. The number of units was chosen through grid search experiments where we varied the number of units (e.g., 32, 64, 100, 128) and evaluated performance on a validation set. 100 units consistently provided better accuracy without overfitting. This size allows the model to learn rich temporal features while remaining computationally efficient.
- Dropout Layer (0.2): Introduces a dropout rate of 20% to prevent overfitting. Dropout randomly sets a fraction of the LSTM unit outputs to zero during training, forcing the model to rely on multiple pathways rather than memorizing specific sequences. This improves generalization and enhances robustness to variations in channel conditions. A dropout rate of 0.2 is commonly used in literature as a moderate regularization value. In our case, it helped reduce overfitting, especially when training on limited CSI data, by encouraging the network to learn more generalizable features.
- Fully Connected Layer: Maps the LSTM outputs to the prediction vector. This layer ensures that the extracted temporal features are translated into meaningful CSI predictions.
- Regression Layer: Optimizes the model by minimizing the Mean Squared Error (MSE) loss function, which measures the discrepancy between predicted and actual CSI values. MSE penalizes larger errors more heavily, ensuring that the model prioritizes minimizing significant deviations. This layer with Mean Squared Error (MSE) is suitable for continuous-value predictions such as CSI, where the goal is to minimize the average squared deviation from the true values.

This architecture allows the LSTM model to generate multi-step predictions, capturing the short-term dynamics of the wireless channel. The proposed method is illustrated in Figure 70.
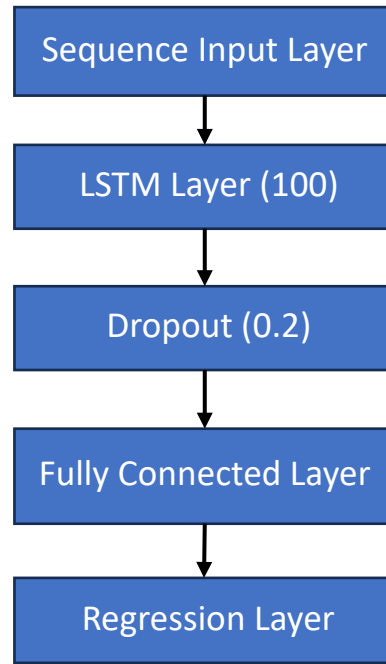
FIGURE 70. PROPOSED LSTM-BASED CSI PREDICTION ALGORITHM

4.3.1.2   Training workflow and evaluation settings

To effectively train the LSTM model for CSI prediction, we follow a structured workflow designed to optimize learning and generalization. The input sequences and target outputs are normalized to the range [0,1] using MinMaxScaler, ensuring that all features contribute equally to the learning process. This normalization technique prevents any single feature from dominating due to larger numerical values and enhances the convergence speed of the model. We employ the Adam optimizer for training, chosen for its adaptive learning rate capability that accelerates convergence while maintaining stability. Mean Squared Error (MSE) is used as the loss function, as it accurately measures the average squared difference between the actual and predicted CSI values, penalizing larger errors more heavily. The model is trained for 10 epochs, balancing training time and performance, although additional epochs can be utilized to improve accuracy if overfitting is not observed. A batch size of 64 is selected to ensure efficient computation while maintaining generalization, and 10% of the training data is reserved for validation. This validation strategy allows continuous monitoring of the model's performance and facilitates early stopping if signs of overfitting are detected. This training configuration enables the LSTM model to effectively capture temporal dependencies in the CSI data, optimizing its predictive accuracy over short-term intervals.

To benchmark the performance of the LSTM-based CSI prediction model, we compare it against the outdated CSI approach, which serves as a baseline. The outdated CSI method assumes that the last observed CSI value remains constant across all future time steps, relying on the assumption of channel invariance over the prediction horizon. This approach provides a straightforward baseline to evaluate the predictive advantage of the LSTM model. To quantify the prediction accuracy, we use Normalized MSE (NMSE), which measures the prediction error relative to the magnitude of the actual CSI values. NMSE is calculated as:

$$\text{NMSE} = \frac{E\left[|\boldsymbol{S} - \widehat{\boldsymbol{S}}|_2^2\right]}{E\left[|\boldsymbol{S}|_2^2\right]}$$

where S represents the true CSI values, and $\widehat{\boldsymbol{S}}$ denotes the predicted values.

4.3.1.3   Simulation results

Building on the discussions in the previous sections, we now evaluate the performance of our approach through a series of simulations. For simplicity, we first assess our model in a single-link scenario before extending it to a more realistic setting with multiple SNEs and LCs, as described in Subsection 3.4.1 of Deliverable D3.1. The channel settings are based on the results presented in D2.3 [4]. Figure 71 presents a comparison between the proposed LSTM prediction method and outdated CSI. The LSTM model consistently outperforms the outdated CSI method across all prediction horizons. However, as the prediction length increases, the NMSE rises for both methods due to reduced temporal correlation, which diminishes the predictability of future CSI values. Despite this, the LSTM model's ability to capture temporal dependencies allows it to maintain lower NMSE values compared to the outdated CSI baseline, highlighting its effectiveness in short-term CSI prediction. For instance, the NMSE for both methods is under 0.04 for a 1 ms prediction, while for a 4 ms prediction, the NMSE for the LSTM remains below 0.05, whereas for the outdated CSI, it is around 0.13. This significant difference underscores the necessity of using prediction models for larger prediction steps.
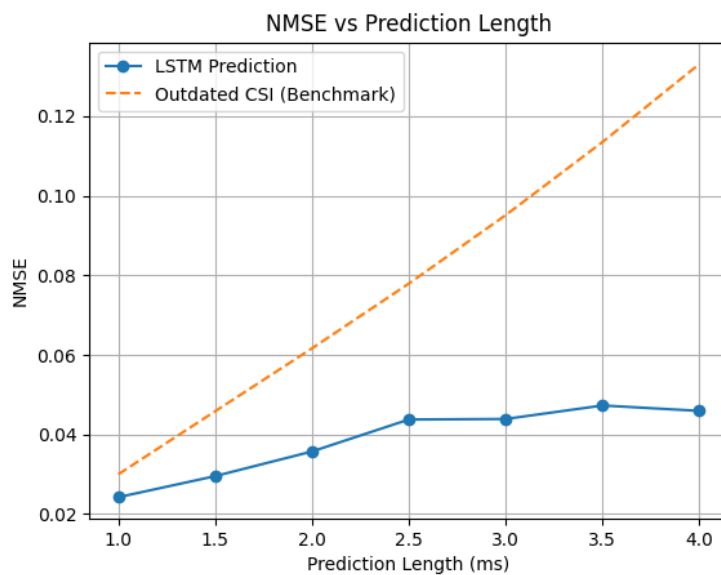
FIGURE 71. PROPOSED LSTM PREDICTION VS OUTDATED CSI

In In the second simulation, we compare the NMSE against velocities for two methods, as shown in Figure 72. The objective is to evaluate and compare the performance of the LSTM-based CSI prediction and the outdated CSI method under varying mobility scenarios. Specifically, we assess the NMSE for different SNE velocities ranging from 1 m/s to 25 m/s, in 5 m/s increments, to account for the impact of Doppler frequency on temporal correlation. This range reflects realistic mobility conditions in wireless communication environments. The model is tasked with predicting the next two CSI samples (1 ms) based on the past 20 CSI samples (10 ms), enabling real-time channel estimation and adaptation. This comparative analysis provides insights into the robustness of the LSTM model across diverse mobility scenarios, highlighting its adaptability to dynamic wireless channels and its predictive advantage over conventional outdated CSI methods. As shown in in Figure 72, the performance of the historical CSI method dramatically degrades at higher velocities due to reduced temporal correlation, which is caused by increased Doppler effects at higher speeds.
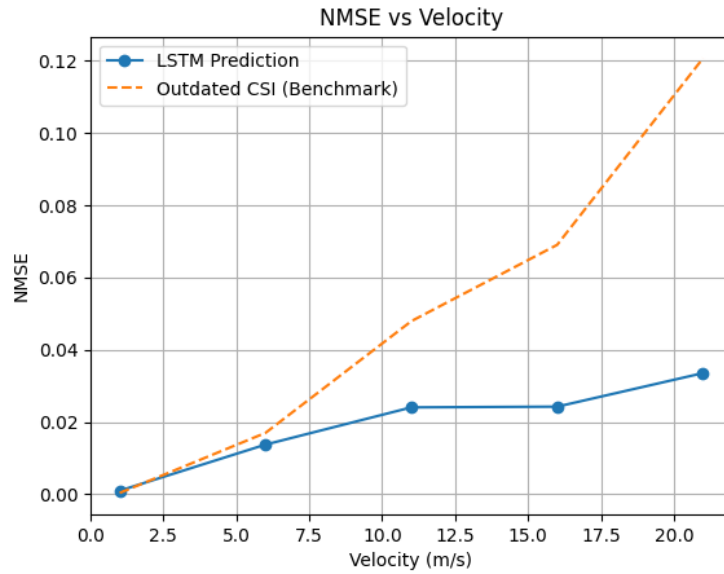
FIGURE 72. COMPARISON OF NMSE BETWEEN TWO METHODS FOR DIFFERENT VELOCITIES

In the next simulation, we evaluate the effect of channel prediction on outage probability. Figure 73 shows the outage probability versus discount rate for predicted and delayed CSI for a single link. We use the system model described in subsection 3.4.1 of Deliverable D3.1 and omit the details for brevity. To achieve an outage probability of less than 10^{-6}, which is crucial for reliability, a maximum discount rate (θ) of 0.92 is required for predicted CSI, and 0.83 for one-sample delay. As the delay in historical CSI increases, the outage probability significantly degrades. This result underscores the necessity of prediction to achieve reliability while optimizing system resources. In Figure 74 we present the results for the same scenario, but with a zoomed-in version of θ ranging from 0.9 to 1, highlighting the gap between predicted and historical CSIs.
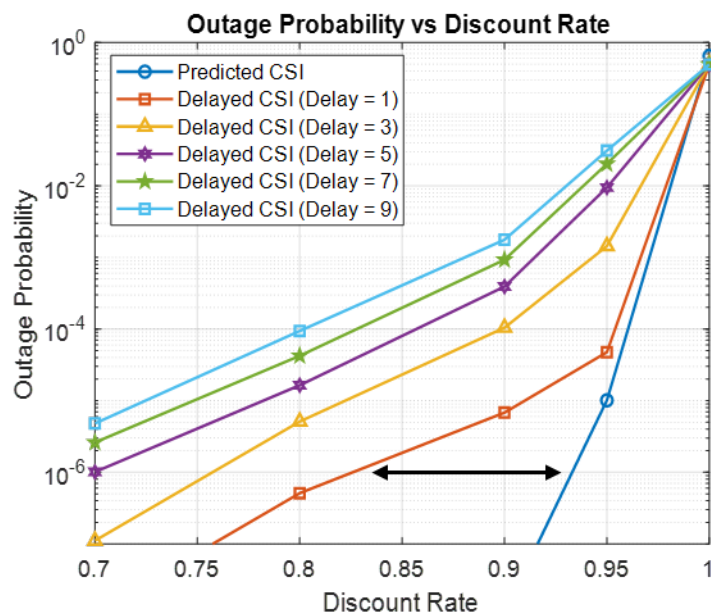
FIGURE 73. OUTAGE PROBABILITY VS DISCOUNT RATE FROM 0.7 TO 1 FOR PREDICTED AND DELAYED CSI
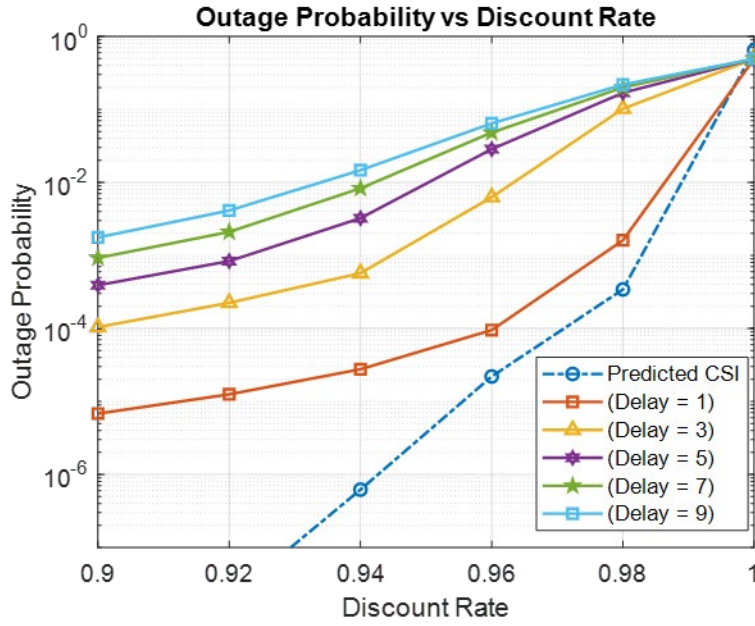
FIGURE 74. OUTAGE PROBABILITY VS DISCOUNT RATE FROM 0.9 TO 1 FOR PREDICTED AND DELAYED CSI

We extend the simulation to the case of 4 LCs and 20 SNEs. The channel settings follow the results outlined in D2.3. The results for both outage and overflow are plotted in Figure 75 and Figure 76, respectively. The outage of predicted CSI remains below 10^{-6} for discount rates under 0.92, which is acceptable. However, for historical CSI, θ must be below 0.82 to meet this threshold. This indicates that more resources are required for outdated CSI to compensate for the lower rate caused by the discount factor.

To provide further insight, we use the overflow rate as a metric to assess the impact of the discount rate on time constraints. If the overflow exceeds $100\ \mu s$, which is the threshold based on the Robot Control use case, a violation occurs. Figure 76 shows that for a K-factor of 6, θ should be greater than 0.87. Therefore, an acceptable bound for selecting the discount rate in this case is:

$$0.87 \leq \theta \leq 0.92$$

Furthermore, to extend our analysis to other use cases and environments, we examine the results for different Rician K-factors. On the other hand, the primary difference between the channel measurements from D2.3 and the 3GPP [58] model lies in the K-factor, which is lower than expected based on 3GPP predictions. To address this discrepancy, we conduct an additional simulation. Figure 76 shows that increasing the K-factor from 0 to 6 significantly reduces overflow. This is due to the strengthened line-of-sight link, which leads to higher acceptable rates. For instance, when $\theta = 0.9$, the overflow in Rayleigh fading is on the order of 5×10^{-2}, whereas for Rician fading with a K-factor of 6, the overflow drops below 10^{-6}.

In the final experiment, we consider different sampling periods and evaluate our method with outdated CSI. A lower CSI sampling frequency reduces overhead and frees up resources for data transmission. However, it is crucial to determine how much we can extend this time interval while maintaining system reliability. According to Figure 77, increasing T_s from 0.5 ms to 4 ms allows the predicted CSI to achieve the same reliability of 10^{-6} by reducing the discount factor from 0.96 to 0.82. Interestingly, the required $\theta = 0.82$ is higher than that of delayed CSI, even when $T_s = 0.5$ ms. In other words, using prediction enables an eightfold reduction in pilot overhead compared to relying solely on delayed CSI. However,

lowering the discount factor in this case reduces available time resources by approximately 20%. In other words, an 20% improvement in spectral efficiency is achieved with the predictive scheduling scheme compared to relying on outdated data for $T_s = 0.5$ ms under the same reliability conditions. Thus, finding an appropriate tradeoff between T_s and θ is crucial, as it depends on the accuracy of the prediction method in designing a proactive resource allocation scheme.
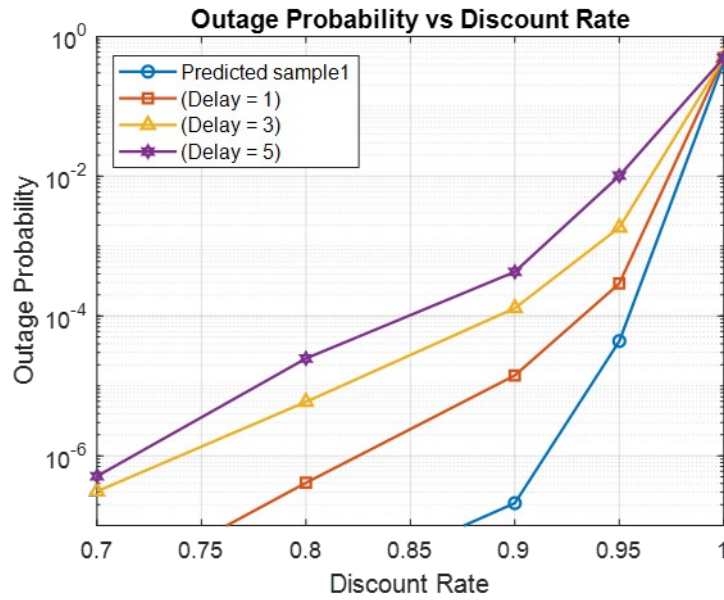
FIGURE 75. OUTAGE PROBABILITY VS DISCOUNT RATE FOR PREDICTED AND DELAYED CSI WITH 4 LCS AND 20 SNES.
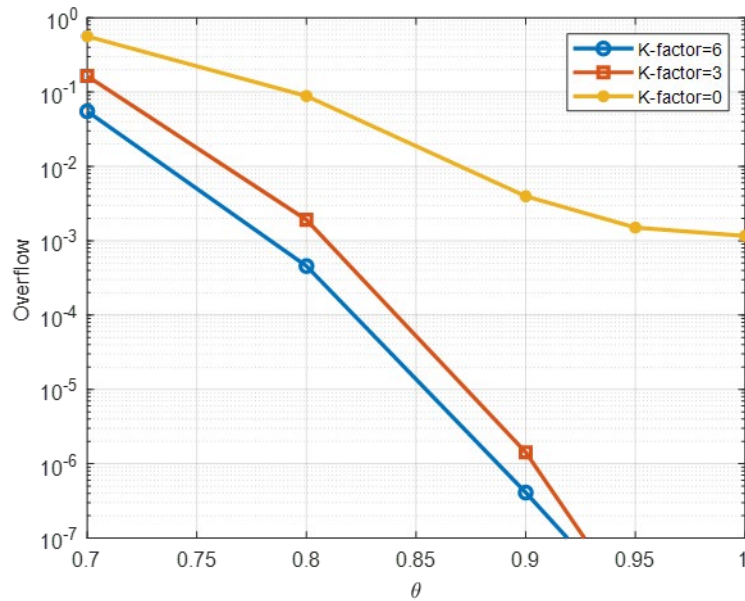
FIGURE 76. OVERFLOW RATE VS DISCOUNT RATE FOR PREDICTED AND DELAYED CSI WITH 4 LCS AND 20 SNES.
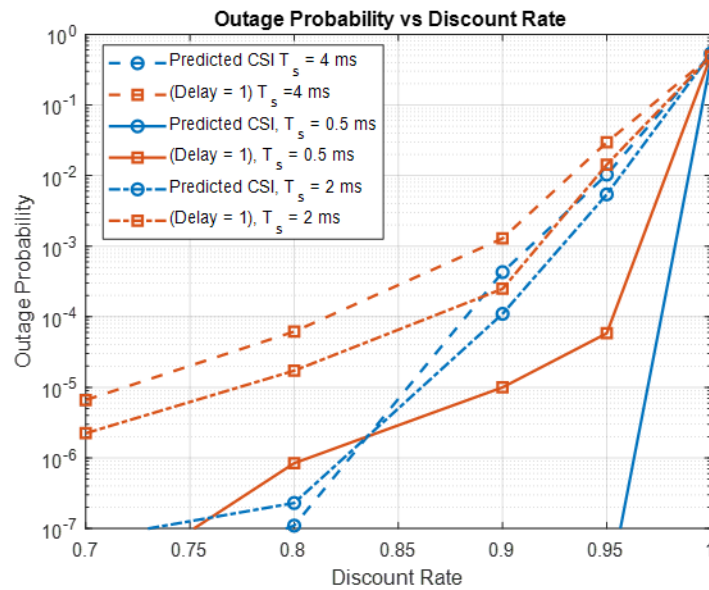
FIGURE 77. OUTAGE PROBABILITY VS DISCOUNT RATE FOR PREDICTED AND DELAYED CSI WITH DIFFERENT SAMPLING PERIODS.

4.3.1.4   Summary and recommendations

In this subsection, we investigated predictive scheduling in an industrial subnetwork (robot control use case), focusing on the impact of CSI prediction in proactive resource allocation. We leveraged LSTM as a powerful machine learning method for forecasting temporally correlated CSI, highlighting it as a promising approach. The NMSE metric was used to assess prediction accuracy, comparing it with outdated CSI. We then applied a practical scenario with multiple LCs and SNEs, as discussed in D3.1 for in-X cooperative communication. Our results demonstrated that leveraging LSTM-based CSI prediction improves both outage and overflow rates for a fixed discount factor compared to relying solely on historical CSI data. Specifically, an 20% improvement in spectral efficiency is achieved with the predictive scheduling scheme compared to relying on outdated data.

We recommend selecting an appropriate CSI sampling frequency based on system requirements. For instance, using our LSTM-based method, the sampling period can be extended from 0.5 ms to 4 ms, reducing pilot overhead by a factor of eight while maintaining the required reliability. However, enhancing prediction accuracy through more advanced methods could further improve resource utilization and system reliability. Therefore, we suggest further exploration of state-of-the-art CSI prediction techniques to minimize prediction errors. Another promising direction is the development of a comprehensive algorithm that jointly optimizes CSI prediction and proactive resource allocation, improving performance in highly dynamic scenarios.

4.3.2   Link quality assessment for multi-path transmission schemes in subnetworks

The work carried out in D3.1 [2] addressed predictive mechanisms for coverage within subnetworks, and from the parent 6G network to and from subnetworks. It proposed AI/ML based mechanisms to effectively predict beam related metrics such as radio quantities or measurements (like SNR, RSRP, etc.), classification of beams (e.g., good beam vs bad beam), clustering of beams according to their signal quality and beam failures. The quality of the radio link is a fundamental component to be assessed for a transmitter to choose the best serving radio link.

The nodes in the subnetwork are constrained and may have difficulties assessing the radio link based on high number of physical layer samples. Hence, this is typically assessed by higher layers metrics, like SNR or RSRP, that result from averaging of physical layer measurement sampling, which are collected at much higher rate.

Given this restriction in subnetwork nodes, it becomes more important to evaluate the radio link with higher layer metrics. With these, due to the distributed nature of the subnetworks, their heterogeneity in terms of measurement capabilities, and cascaded channels introduced by coverage extenders like the RIS, the continuation of this work moves to D3.4 [59].

4.3.3 Deterministic and Predictive Traffic Scheduling

Future wireless networks must support emerging time-sensitive and deterministic services driven by the increasing digitalization and automation of key verticals such as connected and autonomous mobility and Industry 4.0. These applications demand ultra-reliable low-latency communications (URLLC) with stringent bounded latency and high reliability. However, achieving deterministic wireless communications in beyond-5G networks remains a challenge due to the inherently stochastic nature of wireless channels and systems. This challenge becomes even more complex where mixed traffic flows with diverse Quality of Service (QoS) requirements coexist, including both time-sensitive and best-effort traffic.

The ability to support deterministic service levels in mixed traffic environments heavily depends on efficient radio resource management and advanced scheduling mechanisms capable of anticipating traffic demands and dynamically allocating resources. In 5G New Radio (NR), scheduling can be either dynamic or semi-static [60]. Dynamic scheduling allocates radio resources on a per-packet basis upon receiving a scheduling request, whereas semi-static scheduling, such as Configured Grant (CG) for uplink (UL) and Semi-Persistent Scheduling (SPS) for downlink (DL), pre-allocates radio resources periodically to users, eliminating the need for scheduling requests and reducing latency. While semi-static scheduling has proven effective in ensuring low transmission latencies for periodic traffic with fixed packet sizes, its efficiency decreases when packet arrival frequencies fluctuate, packet sizes vary, or multiple traffic flows with different periodicities coexist [61][62]. In such cases, semi-static scheduling may result in inefficient resource utilization or failure to meet deterministic latency requirements.

To mitigate these inefficiencies, existing research has explored predictive scheduling mechanisms that leverage traffic forecasts to enhance resource allocation [63]. Some studies propose semi-static scheduling schemes that pre-allocate resources based on traffic predictions to improve support for deterministic services. However, such pre-allocations are susceptible to inefficiencies due to prediction inaccuracies and the stochastic nature of wireless traffic. This unpredictability can reduce the system's capacity to guarantee deterministic communication requirements and lead to resource allocation inefficiencies [64].

In this context, we proposed a novel predictive scheduling approach that advances the state-of-the-art by integrating predictive capabilities into both dynamic and semi-static scheduling frameworks. Unlike traditional semi-static scheduling approaches that rely on pre-allocations, our proposed predictive dynamic scheduling scheme allocates resources upon receiving a scheduling request while leveraging traffic predictions to prioritize allocations that optimize resource utilization and meet deterministic

latency requirements. Additionally, by delving deeper into the concepts of scheduling mechanisms, we proposed a predictive Configured Grant (CG) scheduling scheme that accounts for inaccuracies in traffic predictions when pre-allocating radio resources based on anticipated traffic demands. By intelligently utilizing traffic predictions and adapting to mixed traffic flows with diverse QoS requirements, our approach significantly enhances deterministic service levels, maximizing the percentage of transmissions that meet bounded latency requirements and improving overall resource efficiency in beyond-5G networks.

4.3.3.1   State of the Art for Deterministic and Predictive Traffic Scheduling

Predictive solutions will play a key role in next-generation networks by enhancing radio resource utilization efficiency and reducing excessive over-dimensioning, particularly for traffic with stringent QoS requirements, including deterministic services [65]. Predictive schedulers have emerged as a promising approach to optimize resource management in Time-Sensitive Networks (TSN). For instance, an online traffic scheduler leveraging deep reinforcement learning (DRL) and convolutional neural networks (CNN) is introduced to extract flow features and optimize scheduling and resource allocation in TSN environments [66]. Similarly, a deterministic federated learning framework and a DRL-based resource scheduling algorithm are proposed for managing time-sensitive industrial IoT services [67]. Furthermore, a reinforcement learning-based scheduler is designed to handle traffic flows with different QoS profiles (5G QoS Identifiers, 5QIs) in an asynchronous deterministic backhaul network, utilizing predictive data analytics to optimize resource allocation [68].

Several studies also propose predictive schedulers to efficiently manage radio resources at the RAN (Radio Access Networks) and better support services using traffic demand forecasts. For example, a semi-persistent downlink scheduler integrated with a short-term traffic predictor demonstrated that predictive scheduling could achieve comparable performance to proportional-fair dynamic scheduling while reducing computational complexity in bursty video traffic scenarios [69]. Another study proposed a DRL-based grant allocator combining offline and online learning to support URLLC traffic using semi-static grant-free scheduling, which minimizes signaling overhead and communication latency [63]. Additionally, multi-objective DRL techniques have been explored for priority-enabled grant-free (GF) or configured grant (CG) scheduling in a massive Machine-Type Communication (mMTC) setting, demonstrating that predictive approaches improve successful transmission probabilities by reducing collisions in shared resource access [70]. An alternative approach analyzed spatio-temporal traffic correlations in industrial environments, using them to optimize semi-static scheduling configurations in beyond-5G networks [71].

Moreover, the 3GPP standard defines two types of Configured Grant (CG) scheduling. In Type 1, the uplink grant is configured via RRC signaling, whereas, in Type 2, the periodicity is predefined but activation and deactivation occur dynamically through the PDCCH control channel, similar to Semi-Persistent Scheduling (SPS) [60]. The adaptability of semi-static scheduling has been explored extensively, with studies proposing dynamic adjustments to resource allocations based on network conditions. For instance, offline and online learning techniques have been applied to optimize CG scheduling, where parameters such as node buffer status and unused resources are continuously monitored to determine optimal allocations [63]. Furthermore, reinforcement learning has been used to enable mMTC nodes to autonomously select radio resources in a grant-free scheduling framework while avoiding collisions with eMBB transmissions [72]. To enhance predictive scheduling, proactive schemes have been developed to forecast traffic demand and optimize resource allocation. One such

approach assigns radio resources to high-priority nodes based on buffered data, expected data generation in the next scheduling period, and predicted average data rate [69]. Another study proposed a proactive grant to allocate resources based on predictions of the data expected to be generated by different nodes. The first approach predicts data sizes and allocates the necessary radio resources for each user in every slot. While this minimizes latency, it leads to over-reserving radio resources that may go unused. The second approach enhances efficiency by predicting both data sizes and traffic generation times [73]. While these solutions have demonstrated significant improvements, inaccuracies in traffic predictions can still lead to inefficiencies in resource utilization and cause some packets to experience increased latency.

To address the challenges mentioned above, we advance the state-of-the-art by proposing a predictive dynamic scheduler that leverages traffic predictions to select resource allocations that can satisfy the demand of current transmissions and increase the likelihood of satisfying future transmissions. In addition, current approaches that focus on semi-static scheduling, where resources are reserved based on traffic predictions, remain susceptible to resource management inefficiencies due to inaccuracies in predictions or the stochastic nature of wireless systems and traffic sources. These inaccuracies can cause some packets to experience increased latency. In this context, we aim to handle these inaccuracies to reduce the impact of incorrect predictions. We extend the state-of-the-art with a novel predictive CG scheduling scheme that pre-allocates resources, considering both traffic predictions and prediction inaccuracies. By accounting for these inaccuracies, the scheme improves the ability to meet bounded latency requirements, enhances the satisfaction rate of transmissions, and thus supports deterministic services.

4.3.3.2   Proposed Deterministic and Predictive Traffic Scheduling Mechanisms

4.3.3.2.1   Predictive Dynamic Scheduling

In this subsection, we introduce our first proposed Predictive Dynamic Scheduling (PDS) scheme for Beyond 5G, aimed at achieving deterministic and predictive scheduling in future networks. PDS leverages traffic predictions to prioritize resource allocations that meet the requirements for transmissions that sent a Scheduling Request (SR) while avoiding resources likely to be needed by future predicted packets. Before presenting our proposed scheme, we first highlight the potential of PDS in scenarios with mixed traffic flows and varying QoS requirements. In 5G NR, dynamic scheduling allocates radio resources on a per-packet basis upon receiving a SR. The scheduler can implement various policies to select among the available radio resources based on the SR. Figure 78 illustrates an example where a new SR for the transmission of packet pkt_0 is generated at t_0. This packet must be transmitted within a maximum latency deadline d_0 and requires R_0=9 radio resources. In NR, a radio resource is defined by a Resource Block (RB) in the frequency domain and a slot in the time domain. The RBs must be selected within the transmission window to meet the latency deadline. The transmission window includes all available RBs between t_0 and d_0. Figure 78-left shows that two allocation options within this window would meet the latency deadline. A common policy is to minimize latency, selecting option 1 in Figure 78-left. Now, suppose that a new SR for packet pkt_1, requiring R_1=6 RBs with a latency deadline d_1, arrives at t_1 (Figure 78-middle). The figure shows that it is not possible to meet pkt_1's latency deadline as the necessary RBs within its transmission window were previously assigned to pkt_0. However, both packets could meet their latency requirements if the scheduler had anticipated at t_0 that a new SR for pkt_1 (demanding R_1=6 RBs) would be received at t_1. In this case, the scheduler could have assigned to pkt_0 the RBs corresponding to option 2 in Figure 78-left, allowing pkt_1 to receive the necessary RBs and meet

its latency deadline d_1 (Figure 78-right). This example highlights the potential of predictive dynamic scheduling, and the performance gains it can achieve in scenarios with mixed traffic flows and varying QoS requirements.
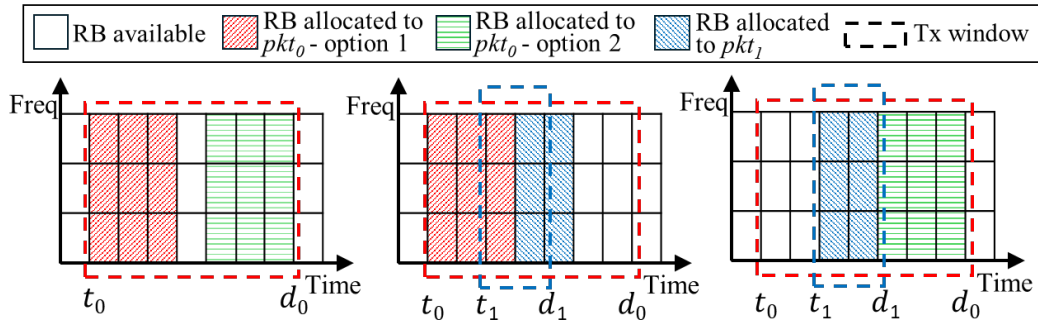
FIGURE 78. ON THE POTENTIAL OF PREDICTIVE DYNAMIC SCHEDULING

Let's consider a 5G NR cell supporting N nodes generating data for uplink transmission. Let pkt_0 be a packet of size s_0 generated by a node n at time t_0 with a latency requirement L_0. The node n sends an SR to the gNB to request RBs for transmitting this packet. The gNB determines the number of RBs as $R_0 = f(s_0, mcs_0)$, where $f(\cdot)$ is a function that determines the number of RBs needed to transmit a packet of size s_0 using a Modulation and Coding Scheme (MCS) mcs_0 following [60]. A standard dynamic scheduler searches for R_0 RBs available in consecutive slots within its transmission window, i.e., between the packet's generation time t_i and its deadline $d_i = t_i + L_i$. In contrast, the PDS proposal searches for the R_0 RBs considering not only the requirements of pkt_0 but also those of future predicted packets $\widehat{pkt_i}$. To this aim, a predictor forecasts the generation time $\hat{t}_i$ and size $\hat{s}_i$ of the next P packets after pkt_0. The scheduler then estimates the number of RBs required for each of these packets as $\widehat{R}_i = f(\hat{s}_i, mcs_i)$, with each packet to be transmitted before $\hat{d}_i = \hat{t}_i + L_i$.

PDS avoids assigning pkt_0 RBs that may be needed for future packets, ensuring that latency requirements are met for both current and upcoming packets. To this end, PDS identifies first the transmission window of pkt_0 and the available RBs within the window (Figure 79.a). PDS then identifies the set Φ of predicted packets whose transmission window may overlap with that of pkt_0. Two packets pkt_i and pkt_j have overlapping transmission windows if $t_i < d_j$ and $t_j \leq d_i$. In Figure 79.b, the transmission windows of $\widehat{pkt}_1$, $\widehat{pkt}_2$, and $\widehat{pkt}_3$ overlap with that of pkt_0, while non-overlapping windows (e.g. $\widehat{pkt}_4$) do not affect the transmission of pkt_0 and are not considered by PDS. PDS temporarily marks as *RBs to avoid* the available RBs between $\hat{t}_i$ and $\hat{d}_i$ for each predicted packet $\widehat{pkt}_i \subset \Phi$ as shown in Figure 79.c. Even if the number of RBs within $\hat{t}_i$ and $\hat{d}_i$ exceeds the $\widehat{R}_i$ RBs required by $\widehat{pkt}_i$ (computed based on $\hat{s}_i$), PDS reserves them to be able to accommodate upcoming packets even with inaccurate $\hat{t}_i$ and $\hat{s}_i$ predictions. PDS tries to avoid allocating pkt_0 any RBs between $\hat{t}_i$ and $\hat{d}_i$ for all $\widehat{pkt}_i$ (i = 1, 2,…, P) to ensure pkt_i will have sufficient RBs available to meet its latency deadline.

PDS then searches for R_0 available RBs in consecutive slots for the transmission of pkt_0. These RBs must be within the transmission window of pkt_0 to meet its latency deadline (Figure 79.d). If multiple options are available, PDS selects the RBs that offer the lowest latency for pkt_0. If no viable options exist, PDS must allocate to pkt_0 RBs initially marked as *RB to avoid*. This situation is illustrated in Figure 79.e where the number of unavailable RBs (allocated to previous transmissions) has been modified. PDS identifies the predicted packet $\widehat{pkt}_i \subset \Phi$ with the largest latency requirement ($\widehat{pkt}_3$ in Figure 79) and checks whether the number of available RBs within its transmission window $\widehat{U}_i$ (between $\hat{t}_i$ and $\hat{d}_i$) is equal to

or greater than the sum of RBs required to transmit pkt_0 and $\widehat{pkt_i}$, i.e. $\widehat{U_i} \geq R_0 + \widehat{R_i}$. If this condition is satisfied, PDS allocates R_0 RBs for pkt_0 in the first available slots after $\hat{t_i}$. Otherwise, PDS evaluates the same condition for $\widehat{pkt_j} \subset \Phi$ with the next largest latency requirement L_i. In the example in Figure 79.e, we consider $R_0=6$, $\widehat{R_2}=3$, and $\widehat{R_3}=3$. The transmission window of $\widehat{pkt_3}$ has only 6 available RBs, which is insufficient to accommodate both pkt_0 and $\widehat{pkt_3}$. PDS then checks whether the available RBs within the transmission window of $\widehat{pkt_2}$ are sufficient to accommodate pkt_0 and $\widehat{pkt_2}$. Since $\widehat{U_2}= 9$ in Figure 79.e, which satisfies the condition $\widehat{U_2} \geq R_0+\widehat{R_2}$, PDS allocates RBs to pkt_0 within the transmission window of $\widehat{pkt_2}$. If the condition $\widehat{U_i} \geq R_0 + \widehat{R_i}$ cannot be met for any predicted packets in Φ, PDS allocates R_0 RBs to pkt_0 from the available RBs within the transmission window $\widehat{U_i}$ of the predicted packet $\widehat{pkt_i} \subset \Phi$ with the largest latency requirement ($\widehat{pkt_3}$ in Figure 79).
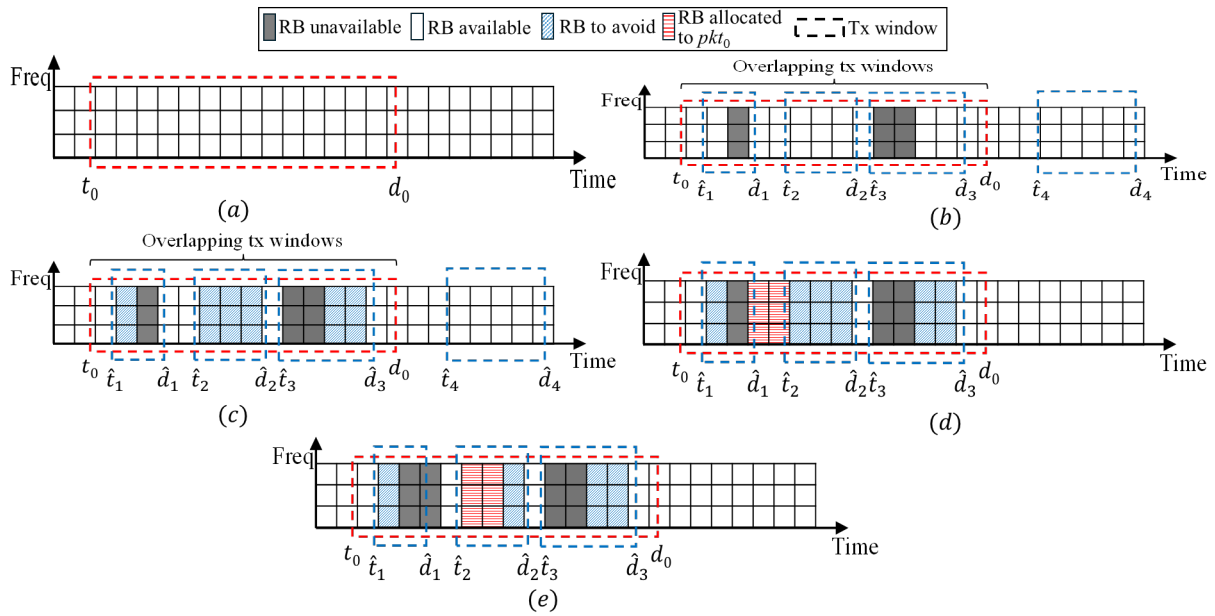
FIGURE 79. ILLUSTRATION OF THE OPERATION OF THE PROPOSED PREDICTIVE DYNAMIC SCHEDULING.

4.3.3.2.2 *Predictive Configured Grant Scheduling*

To proactively allocate resources and minimize the impact of scheduling requests in dynamic scheduling, we extend our proposed scheme with Configured Grant (CG) scheduling. In Section 4.1.2, we introduced the LAA-MCG scheme, which enhances CG by leveraging multiple CG configurations to adapt to varying channel conditions, enabling collision-free allocation with a focus on periodic and deterministic traffic, such as in industrial subnetwork use cases. Building on this, we now extend CG to support aperiodic and mixed traffic flows, particularly in vehicular use cases where traffic patterns are highly dynamic but still require deterministic service levels. To address this variability in traffic patterns, we leverage traffic predictions to anticipate transmission demands and use the flexibility of CG to proactively pre-allocate resources. This approach enables reliable communication under stochastic conditions, reduces collisions, and meets bounded latency requirements in dynamic scenarios. Furthermore, this approach not only leverages predictive knowledge but also accounts for inaccuracies in predictions. By incorporating these inaccuracies, the scheduler dynamically adapts resource allocation to mitigate the effects of incorrect predictions and enhance overall resource utilization.

Let us consider a scenario with N nodes, where each node i ($i=0,1,…,N-1$) generates data with an inter-packet generation time characterized by a mean value p_i and a standard deviation σ_i^t. The packet size is

also characterized by a mean value s_i and a standard deviation σ_i^s. Each packet has a latency requirement L_i. Without loss of generality, we consider $p_i = p$ ms for all N nodes. The proposed predictive CG scheduler determines the radio resources to be allocated for all nodes within a scheduling interval of p ms, and scheduling decisions can be updated at every scheduling interval p. Resources are proactively pre-allocated to maximize the likelihood of satisfying the requirements of each transmission. To achieve this, a predictor forecasts the generation time $\hat{t}_i$ and size $\hat{s}_i$ of the next packet pkt_i for each node within the upcoming scheduling interval p. The scheduler then calculates the number of resources required for transmitting each packet pkt_i, considering not only the predicted size $\hat{s}_i$ but also a margin Δs_i to account for prediction inaccuracies; Δs_i can be set equal to the standard deviation σ_i^s or other value related to the prediction inaccuracy. We assume a NR radio interface where a radio resource is defined by a Resource Block (RB) in the frequency domain and a slot in the time domain. The scheduler estimates the number of RBs needed for pkt_i as $R_i = f(\hat{s}_i + \Delta s_i, mcs_i)$, where $f(\cdot)$ is a function that determines the number of RBs required to transmit a packet of size $\hat{s}_i + \Delta s_i$ using a Modulation and Coding Scheme (MCS) mcs_i following [60]. The proposed scheduler seeks to allocate R_i RBs for pkt_i within the allocation window $w_i = w(t_i^{ini}, t_i^{end})$, where $t_i^{ini} = \hat{t}_i + \Delta t_i$ and $t_i^{end} = \hat{t}_i - \Delta t_i + L_i$, $\hat{t}_i$ is the predicted generation time of pkt_i and Δt_i is a time margin to account for prediction inaccuracies of the generation time. The w_i ensures that a packet generated between $\hat{t}_i - \Delta t_i$ and $\hat{t}_i + \Delta t_i$ can be transmitted within its latency requirement L_i, as shown Figure 80.
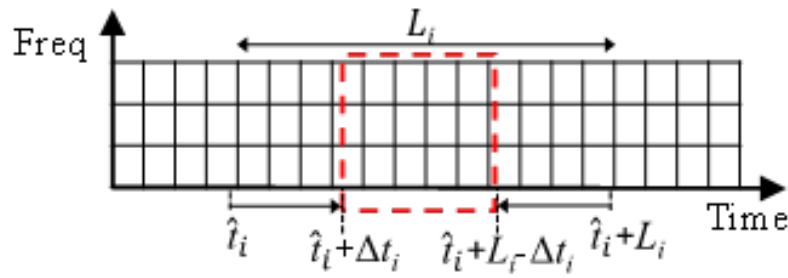
FIGURE 80. ALLOCATION WINDOW w_i FOR PACKET PKT_i.

The scheduler identifies w_i for all packets pkt_i within the scheduling interval p, and searches for a scheduling solution that meets the latency requirements of the maximum possible number of packets. We assume a mixed traffic flow scenario where packets may have different latency requirements. The operation of the proposed predictive CG scheduling scheme is presented in Figure 81. The scheme defines the set Φ, which is initialized with all the packets pkt_i scheduled for the next scheduling interval p. The scheduler starts processing packets with the shortest latency requirements (line 2 in Figure 81). For each packet pkt_i, the scheduler first searches for available RBs within its allocation window w_i that do not overlap with the allocation window w_j of other packets pkt_j ($\forall j \neq i$) (line 3 of Figure 81).

```
Algorithm I: Predictive CG Scheduling
Input: set Φ of packets to schedule, t̂ᵢ, Δtᵢ, ŝᵢ, Δsᵢ, Lᵢ ∀i
1.  Define for each pktᵢ in Φ: Rᵢ= f(ŝᵢ+Δsᵢ, mcsᵢ),
    wᵢ=w(tᵢⁱⁿⁱ, tᵢᵉⁿᵈ), tᵢⁱⁿⁱ= t̂ᵢ+ Δtᵢ, tᵢᵉⁿᵈ= t̂ᵢ + Lᵢ - Δtᵢ
2.  Sort packets in Φ by Lᵢ
3.  Call Φ=Search_in_non-overlapping_area(Φ)
4.  If Φ ≠ Ø (there are packets without RBs)
5.     Call Search_in_overlapping_area(Φ)
6.  End procedure
Procedure I: Search_in_non-overlapping_area(Φ)
1.  Set Θ = Ø (Θ is a set with pending packets)
2.  Repeat
3.     N = size(Φ)
4.     For each pktᵢ in Φ
5.        Identify non-overlapping area wᵢⁿᵒ = w(tᵢⁿᵒ⁻ⁱ, tᵢⁿᵒ⁻ᵉ)
          between wᵢ and wⱼ ∀j≠i and pktⱼ in Φ
6.        Rᵢⁿᵒ = available RBs within wᵢⁿᵒ
7.        If Rᵢⁿᵒ ≥ Rᵢ
8.           Allocate Rᵢ RBs in wᵢⁿᵒ to pktᵢ: call RB_allocation
9.        Else → pktᵢ is included in Θ
10.       Removed pktᵢ from Φ
11.    End For
12.    Φ = Θ, Θ = Ø
13. Until size(Φ) == N
14. Return Φ and end procedure

Procedure II: allocate_RB(pktᵢ, wᵢⁿᵒ=w(tᵢⁿᵒ⁻ⁱ,tᵢⁿᵒ⁻ᵉ),
Rᵢ, tᵢⁱⁿⁱ,tᵢᵉⁿᵈ)
1.  If tᵢⁿᵒ⁻ⁱ== tᵢⁱⁿⁱ & tⁿᵒ⁻ᵉ== tᵉⁿᵈ
2.     Allocate Rᵢ RBs in the center of wᵢⁿᵒ
3.  Elseif tᵢⁿᵒ⁻ⁱ-tᵢⁱⁿⁱ > tᵢᵉⁿᵈ - tᵢⁿᵒ⁻ᵉ
4.     Allocate Rᵢ RBs in wᵢⁿᵒ starting from tᵢⁿᵒ⁻ⁱ
5.  Else
6.     Allocate Rᵢ RBs in wᵢⁿᵒ ending at tᵢⁿᵒ⁻ᵉ
7.  End procedure
Procedure III: Search_in_overlapping_area(Φ)
1.  For each pktᵢ in Φ
2.     Repeat
3.        R = available RBs within wᵢ= w(tᵢⁱⁿⁱ, tᵢᵉⁿᵈ)
4.        If R ≥ Rᵢ → Allocate Rᵢ RBs in wᵢ to pktᵢ
5.        Elseif odd iteration → tᵢᵉⁿᵈ=tᵢᵉⁿᵈ+Δtᵢ
6.        If even iteration → tᵢⁱⁿⁱ= tᵢⁱⁿⁱ-Δtᵢ
7.        Until pktᵢ receives RBs or wᵢ==w(t̂ᵢ-Δtᵢ,
          t̂ᵢ+Lᵢ+Δtᵢ)
8.        If pktᵢ has not received RBs & Rᵢ > f(ŝᵢ, mcsᵢ)
9.           Rᵢ = f(ŝᵢ, mcsᵢ), tᵢⁱⁿⁱ= t̂ᵢ+ Δtᵢ, tᵢᵉⁿᵈ= t̂ᵢ+ Lᵢ - Δtᵢ
10.          Goto line 2
11.    End For
12. End procedure
```

FIGURE 81. ALGORITHM FOR PREDICTIVE CG SCHEDULING

We define this non-overlapping area of w_i as $w_i^{no} = w(t_i^{no-i}, t_i^{no-e})$. If the number of available RBs R_i^{no} within w_i^{no} is smaller than R_i ($R_i = f(\hat{s}_i + \Delta s_i, mcs_i)$), then pkt_i is added to an auxiliar set Θ as a pending packet awaiting RB allocation (lines 9 of *Search_in_non-overlapping_area*). Conversely, if $R_i^{no} \geq R_i$, the scheduler allocates R_i RBs to pkt_i within w_i^{no} (lines 4-8 of *Search_in_non-overlapping_area*). If multiple allocation options exist, the scheduler follows the policy outlined in *allocate_RB*. In particular, if w_i does not overlap with any other w_j ($\forall j \neq i$), the scheduler allocates RBs for pkt_i at the center of w_i. This increases robustness against inaccuracies in the estimated generation time of pkt_i. If w_i partially overlaps with the allocation window w_j of other packets, the scheduler prioritizes allocating resources from the boundary of w_i^{no} that is furthest from the respective limits t_i^{ini} or t_i^{end} of w_i. This ensures that packets generated with a deviation larger than Δt_i from the estimated time $\hat{t}_i$ can still meet their latency requirements. Once pkt_i receives RBs, it is removed from Φ (line 10 in *Search_in_non-overlapping_area*), and w_i is no longer considered when allocating RBs for other packets pkt_j ($j \neq i$). After completing the allocation process for all packets in Φ, the process in lines 2-13 in *Search_in_non-overlapping_area* is repeated for the remaining packets in Θ. We should note that a packet pkt_j that previously had $R_i^{no} < R_i$ might now have $R_i^{no} \geq R_i$ after removing from Φ the packets pkt_j that successfully received RBs. This iterative process continues until no further packets can be allocated the required R_i resources within non-overlapping areas.

For packets that cannot be assigned the required resources R_i within non-overlapping areas, the scheduler follows *Search_in_overlapping_area* (line 5 in Figure 81). The process begins with packets that have the lowest latency requirements. For each packet pkt_i remaining in Φ after *Search_in_non-overlapping_area*, the scheduler checks whether there are sufficient RBs within w_i to satisfy R_i ($R_i = f(\hat{s}_i + \Delta s_i, mcs_i)$). If so, the scheduler allocates R_i RBs within w_i to pkt_i (lines 3-4 in *Search_in_overlapping_area*). If there are insufficient RBs, the size of w_i is iteratively augmented by Δt_i until there are sufficient RBs available within w_i to satisfy R_i, or w_i reaches its maximum possible size $L_i + 2 \cdot \Delta t_i$ (lines 5-7 in *Search_in_overlapping_area*). The maximum size is defined so that it is still possible to meet the latency requirement L_i if the packet is generated at $\hat{t}_i + \Delta t_i$ or $\hat{t}_i - \Delta t_i$. If w_i reaches its maximum allowed size, the scheduler reduces the number of RBs to allocate for pkt_i to match only the predicted

size (i.e., R_i is updated as $R_i = f(\hat{s}_i, mcs_i)$), and searches again for R_i RBs in w_i for pkt_i (lines 8-10 in *Search_in_overlapping_area*). Increasing w_i reduces the likelihood of meeting the latency requirement for packets that deviate significantly from their expected generation time $\hat{t}_i$, but the requirement can still be satisfied in some cases. For example, consider the case where t_i^{end} is equal to $t_i^{end} = \hat{t}_i + L_i$, and the scheduler finds RBs for pkt_i at the end of w_i as illustrated in Figure 82. In this case, the latency requirement L_i cannot be met if pkt_i is actually generated before $\hat{t}_i$ as illustrated in Figure 82, but it can still be satisfied if pkt_i is generated after $\hat{t}_i$.
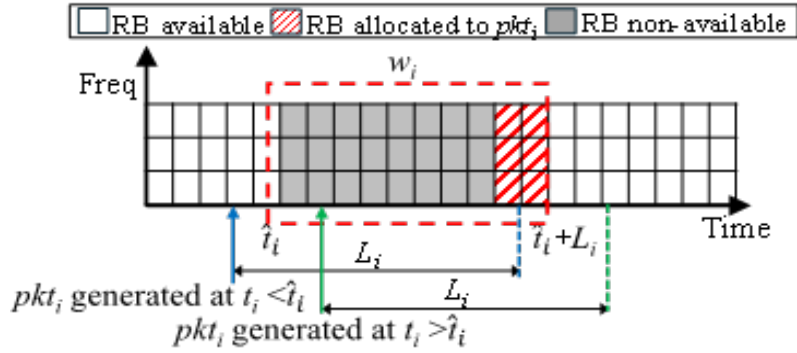
FIGURE 82. EXAMPLE OF PACKET PKT_I GENERATED AT T_I FOR WHICH ITS LATENCY REQUIREMENT L_I IS SATISFIED WHEN $T_I < \hat{T}_I$ BUT NOT WHEN $T_I < \hat{T}_I$.

4.3.3.2.3   *Joint Predictive Scheduling and Flexible Duplexing for Mixed Traffic Support*

Building upon the work presented in Section 4.2.1, we further extend our proposal presented in Section 4.3.3.2.1 and proposed a joint radio resource management scheme for handling both UL and DL communications. This approach leverages predictive scheduling and the flexibility offered by full/flexible duplexing. While flexible duplexing introduces opportunities for improved spectral efficiency, it also raises concerns regarding cross-link interference, particularly between UL and DL transmissions from different SNEs. To address this, we extend the predictive scheduling algorithm from Section 4.3.3.2.1 by integrating flexible duplexing to intelligently manage resource allocation while minimizing interference risks, particularly in terms of in-band emission (IBE) as discussed in Section 4.2.1, thereby enhancing scheduling efficiency in mixed traffic flows. Consider the example scenario illustrated in Figure 83, where four SNEs (SNE0–SNE3) communicate with an HC, transmitting UL packets and receiving corresponding DL transmissions. Each UL transmission from an SNE to the HC may cause interference to others SNEs, depending on spatial proximity. For instance, when SNE0 transmits to HC in UL, it may have higher cross link interference with SNE2's DL reception from HC than with SNE3, due to the relative distances between the sensors, as depicted in Figure 83. To mitigate such interference, the proposed scheme leverages both predictive knowledge of upcoming traffic demands and spatial diversity among SNEs. It prioritizes resource allocation in regions where nodes are spatially distant, reducing interference and increasing the likelihood of successful transmissions. In cases where conflict-free allocation is not possible, the algorithm strategically selects regions with minimal IBE impact, favoring transmissions between SNEs that are farther apart.
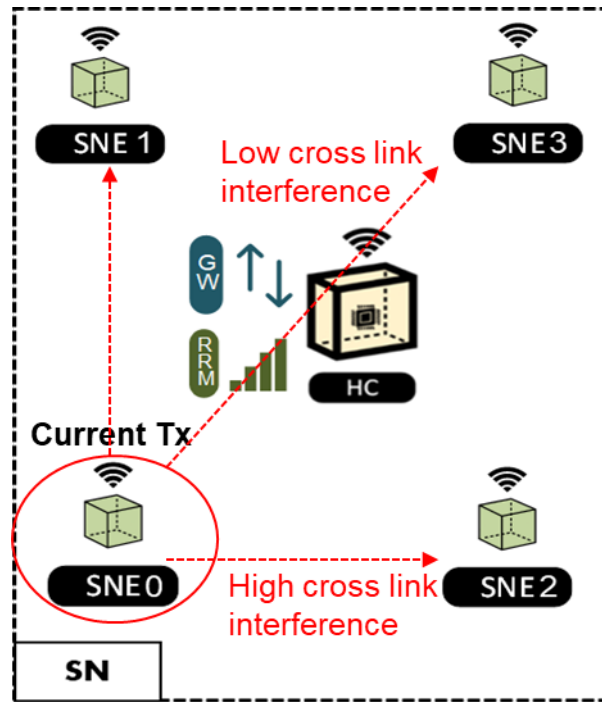
FIGURE 83. ILLUSTRATION OF CROSS-LINK INTERFERENCE IN A WIRELESS SENSOR NETWORK

Considering SNE0 sends a scheduling request (SR) at time t_o to transmit packet pkt_o. The HC must allocate resources for this transmission while also considering the predicted traffic from other SNEs (SNE1, SNE2, and SNE3), which are expected to generate packets pkt_1, pkt_2, and pkt_3, respectively. The predicted regions for these upcoming packets are marked in advance, as detailed in Section 4.3.3.2.1. In addition, all previous constraints also remain same, particularly that each packet must be transmitted within a maximum latency deadline d_o. However, the problem now includes an additional constraint i.e., the HC must allocate DL resources following UL reception. Once the UL transmission from SNE0 is received, the HC must process the data and send a corresponding DL response, all within the same strict deadline. Efficient DL resource allocation within the transmission window is critical to meeting latency requirements, which underscores the importance of flexible full duplexing.

The proposed algorithm schedules both UL and DL resource blocks for a real-time transmission, referred to as pkt_o, which arrives at time t_o and must be completed by its deadline d_o. The goal is to allocate the required UL and DL RBs within the valid transmission window, ensuring the DL follows the UL without conflicting with other scheduled or predicted transmissions. The process begins by determining the transmission window using the t_o, d_o, and the latest predicted deadline $\widehat{d_i}$ to ensure a safe allocation margin. The algorithm then scans the resource grid to identify a contiguous region that can accommodate the required UL RBs within this window. Once a suitable UL region is found, these RBs are tentatively reserved. The algorithm then searches for a corresponding DL allocation that follows the UL, and ensures it fits within the remaining time before the deadline d_o. If both UL and DL resources are successfully identified, the allocations for both UL and DL are finalized, and the scheduling for the current transmission is marked as complete, as illustrated in the Figure 84.a).

However, if DL RBs cannot be found after a tentative UL reservation (as shown in Figure 84.b), the algorithm releases the UL allocation and either searches for alternative UL regions or retries DL allocation in a broader or more flexible search space. This may involve considering RBs overlapping with predicted regions that exhibit high spatial diversity. In such cases, when no free RBs are available for

either DL or both UL and DL, the algorithm selects resources from the predicted regions, again prioritizing spatial diversity to minimize cross-link interference. This approach ensures the current transmission is satisfied while minimizing disruption to future predicted traffic. Figure 84.c) illustrates this scenario, showing how resource allocation may extend into predicted regions based on spatial diversity.

When there are insufficient resources to support UL or DL traffic simultaneously in conflict free region, the proposed scheme opportunistically allocates resources within predicted traffic regions prioritized by spatial diversity. It identifies predicted transmissions whose transmission windows partially overlap with the current transmission, following the strategy detailed in Section 4.3.3.2.1. The scheduler then attempts to schedule both UL and DL, aiming for low interference and high transmission success probability. This process involves evaluating the compatibility of predicted regions, assessing resource availability, and applying flexible duplexing by first allocating UL RBs, then searching for suitable DL slots. If both UL and DL allocations succeed, resources are finalized in regions with minimal interference (i.e., high spatial diversity). If DL allocation fails, the UL reservation is rolled back, and the scheduler considers the next most suitable prediction, while avoiding high-interference regions as shown in Figure 84.c).

When initial strategies fail due to the unavailability of suitable UL or DL resources within the designated transmission window, the proposed scheme initiates a fallback mechanism that performs a full reallocation over the interval (t_o, d_o). It reassesses the entire transmission window, including overlaps with predicted traffic regions, to identify the earliest feasible UL slots with minimal interference and no scheduling conflicts. If such slots are unavailable, the algorithm relaxes its constraints and selects the next best available RBs (as shown in Figure 84.d)), even within higher-interference or overlapping regions, to guarantee that the current transmission is not dropped. Once tentative UL slots are reserved, the scheduler attempts to allocate corresponding DL slots. If DL allocation fails, the reserved UL slots are released, and the search continues with alternative configurations. This iterative reassessment continues until a viable UL-DL pair is successfully allocated within the deadline, ensuring reliable real-time transmission even under resource contention. The scheme dynamically adapts to interference levels and resource availability, prioritizing low-interference regions but allowing higher interference regions when necessary, maximizing the likelihood of successful real-time transmission. Ultimately, the approach ensures that both UL and DL requirements are met within deadline bounds, maintaining service continuity under varying network conditions.
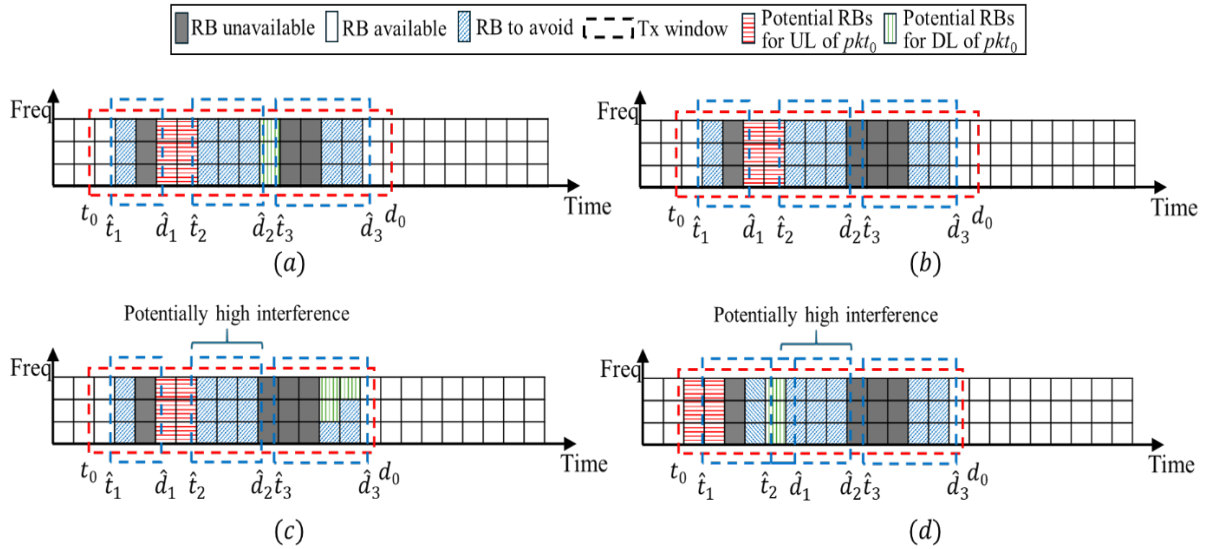
FIGURE 84. ILLUSTRATION OF THE *JOINT PREDICTIVE SCHEDULING AND FLEXIBLE DUPLEXING*

4.3.3.3 Traffic Characterization and Prediction

To evaluate our proposal, we consider a 6G-envisioned autonomous driving scenario in which data collected or generated by an autonomous vehicle is sent to the network for processing at the edge [74]. Communications must meet a bounded latency deadline to ensure that offloading processing workloads to the network does not disrupt the vehicle's operation. For our evaluation, we use realistic sensor data generated by an autonomous vehicle through the Connected and Automated Mobility (CAM) platform presented in [75] and in Deliverable 2.2 [3]. This platform integrates realistic sensing and autonomous driving (AD) capabilities using the open-source CARLA and AUTOWARE software's, interconnected via a Robot Operative System (ROS) bridge. The platform generates sensor data from an autonomous vehicle with a full suite of Level 3 (L3) AD sensors, including five cameras and five radars mounted on the top, front, rear and sides of the vehicle. These sensors detect objects such as vehicles, obstacles and pedestrians, which are then processed by the AD software to control the vehicle. The sensors generate raw data at periodic intervals, which is then processed by a perception module to extract detected objects. The processed sensor data is transmitted to the network for processing at the edge. We have collected extensive datasets of the processed sensor packets -including size, timestamp and sensor ID- from realistic urban environments. Figure 85 shows an example of processed sensor data packets generated with the CAM platform. The packet size varies over time based on the number of objects detected by each sensor as the vehicle navigates an urban scenario.
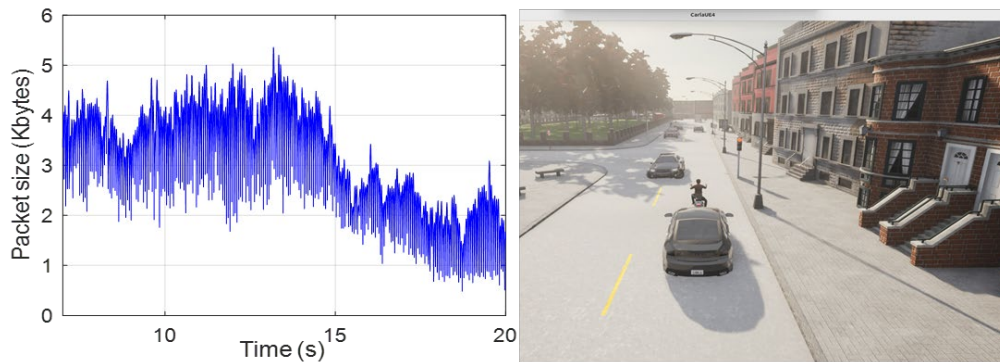
FIGURE 85. SAMPLE TRACE OF PROCESSED SENSOR DATA PACKETS.

Sensors (cameras and radars) generate packets at 50 ms intervals, but the object detection algorithm introduces a processing delay that depends on the driving scenario and the number of detected objects. The sensor configuration follows a cyclic pattern repeated every 50 ms, with packets from the five cameras generated first, followed by packets from the five radars. Table 11 reports the packet inter-arrival times between the ten packets generated every 50 ms, where Δt_{ij} represents the time interval between packets i and j. The first row shows the time interval between packets sequentially generated by the five cameras, while the second row reports the interval between the last camera packet and the first radar packet. Similarly, the third row details the intervals between packets from the five radars, and the fourth row shows the interval between the last radar packet and the first camera packet in the next 50 ms cycle. The table shows that the packet inter-arrival times remain nearly constant within each category, as reflected by the small standard deviation, with variations primarily due to the object detection algorithm's processing time. Given this consistency, the timing of the next ten packets can be directly estimated from the traffic characterization in Table 11 using the mean inter-arrival time. Importantly, any small inaccuracies in these estimations, as indicated by the standard deviation, will not impact on our predictive scheduler since it only allocates resources to the current packet. PDS uses the estimated future packet arrivals to select the radio resources for the current packet that satisfy its latency requirements and increase the likelihood of meeting the latency constraints of future packets upon their arrival.

TABLE 11: PACKET INTER-ARRIVAL TIME

Parameter	Mean (ms)	Std. dev (ms)
$\Delta t_{1\text{-}2}$, $\Delta t_{2\text{-}3}$, $\Delta t_{3\text{-}4}$, $\Delta t_{4\text{-}5}$	8.25	0.58
$\Delta t_{5\text{-}6}$	3.91	0.62
$\Delta t_{6\text{-}7}$, $\Delta t_{7\text{-}8}$, $\Delta t_{8\text{-}9}$, $\Delta t_{9\text{-}10}$	0.25	0.37
$\Delta t_{10\text{-}11}$	13.62	0.89

The size of processed sensor packets varies significantly as shown in Figure 85, since the size depends on the number of objects detected in the driving scenario. To predict the size of future packets, we implemented a predictor based on a Long Short-Term Memory (LSTM) network. LSTM networks are particularly suited for tasks involving sequential or time-series data, where the order and context of input elements are relevant [71]. This includes tasks where dependencies between data points over time need to be learned, making it suitable for predicting autonomous vehicle sensor data since the size of each packet is tied to past packets and sensor behaviors. The hyper-parameters used for the configuration of the LSTM are summarized in Table 12 and are selected to balance the learning and training stability. A sequence length of 150 is selected to capture sufficient historical context for predicting the next packet sizes. Three LSTM layers with 256, 128, and 64 units, respectively, are employed for feature extraction and to capture complex temporal patterns, with a dropout rate of 0.1 applied to prevent overfitting. Each input sample includes features such as sensor ID, packet size, arrival time of packet, inter-arrival time of packets, as well as the mean and standard deviation of previous packet sizes, to capture key characteristics of the sensor data. Min-Max scaling is used for data normalization to ensure that all features contribute proportionally during training. The number of output parameter is set to ten, corresponds to the predicted sizes of the next ten packets. Further, these hyperparameters were optimized to minimize the Mean Absolute Error (MAE), which quantifies the average magnitude of prediction errors. MAE is defined in equation below, where P represents the number of predicted packets, s_i is the actual packet size, and $\hat{s}_i$ is the predicted packet size. We used a

dataset containing 9780 samples, with 60% allocated for training the LSTM, 20% for validation, and the remaining 20% for testing.

$$MAE = \frac{1}{P}\sum_{i=1}^{P} |s_i - \widehat{s_i}|$$

The LSTM network employs a regression-based supervised learning method and the predicted packet sizes can take any real numerical value. However, our analysis of the dataset revealed that packet sizes are limited to a discrete set of values that depend on the number of objects detected by the sensors and the specific sensor type. The output of the LSTM is then discretized by adjusting the predicted packet size to the nearest possible discrete packet size value.

TABLE 12: LSTM HYPER-PARAMETERS

Parameter	Value	Parameter	Value
Sequence Length	150	Number of layers	3 (units: 256,128,64)
Batch Size	32	Features Per Sample	6
Dropout	0.1	Epochs	100
Optimizer	Adam	Scaling Method	Min-Max Scaling
Learning rate	0.001	Number of Outputs	10

4.3.3.3.1 Predictive Dynamic Scheduling

The proposed PDS scheme leverages predictive knowledge of future traffic to schedule current transmissions while accounting for upcoming traffic demands. The LSTM network is configured to predict the next ten packets, considering the previously described cyclic pattern generation. Figure 86 illustrates the predictor's performance using the CDF of the relative absolute error for the prediction of the first, fifth and tenth packet. As expected, the figure shows that prediction accuracy decreases for later packets but remains sufficient for the purpose of the PDS. The predictor accurately forecasts 85% of the next packets without error, with this percentage decreasing to 78% and 70% for the fifth and tenth packets, respectively. The average relative error for the first prediction is only 9%, increasing to 13% and 24% for the prediction of the fifth and tenth packets. It is important to highlight that the lower prediction accuracy for later packets has a limited impact on the PDS. This is because the likelihood that the transmission windows of these distant packets overlap with the transmission window for the currently scheduled packet is low. In contrast, the higher accuracy of predictions for first packets has a more significant influence on the scheduler's performance, as their transmission windows are more likely to overlap with that of the packet being scheduled in dynamic scheduling.
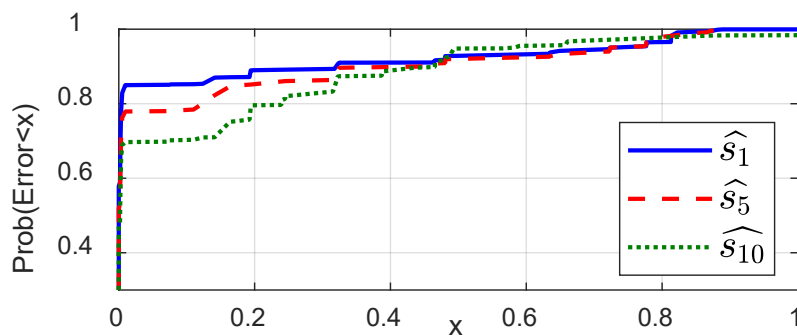
FIGURE 86. CDF OF THE RELATIVE ABSOLUTE ERROR $\left(|s_i - \widehat{s_i}|/s_i \right)$

4.3.3.3.2   Predictive Configured Grant Scheduling

The proposed predictive CG scheduler pre-allocates resources for future traffic demands by leveraging predicted traffic information while accounting for potential prediction inaccuracies. Different sensors have a sampling rate of 20 Hz, collecting data every 50 ms. Then, the object detection algorithm introduces a processing delay that depends on the driving scenario and the number of detected objects. We have characterized this processing delay for each sensor, showing a standard deviation ranging from 1.47 to 1.80 ms. The proposed predictive CG scheme considers 50 ms as the inter-packet generation time p, $\hat{t}_i$ is estimated as $t_i^0 + p$ where t_i^0 is the generation time of the first packet for each sensor i of the N=10 sensors, and Δt_i is set to the characterized standard deviation value σ_i^t. However, following the earlier explanation of the LSTM, it is now configured to predict the size of the next ten packets, i.e., the processed sensor packets generated by each of the N=10 sensors within the period p=50 ms. Figure 87 illustrates the predictor's performance using the CDF of the absolute error for the prediction of the first, fifth and tenth packets. The figure illustrates the magnitude of possible prediction inaccuracies and reinforces the need to account for such inaccuracies when reserving resources based on predicted traffic demands. requirement. The proposed predicted CG scheduling scheme accounts for the inaccuracy of the predicted packet size $\hat{s}_i$ when determining the resources that should be pre-allocated to each pkt_i to minimize the impact of prediction errors on the capacity to meet the latency requirements.
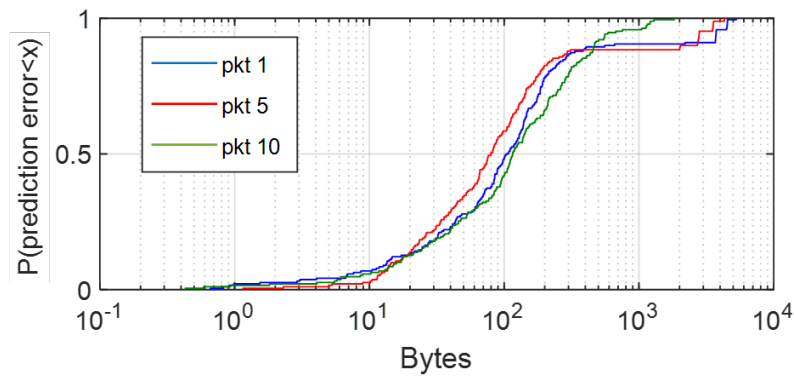
FIGURE 87. CDF OF THE ABSOLUTE PREDICTION ERROR ($|s_i - \hat{s}_i|$)

4.3.3.4   Evaluation and discussion

4.3.3.4.1   Evaluation of the proposed Predictive Dynamic Scheduling scheme

This section compares the performance of the proposed scheme against a reference 5G dynamic scheduling scheme. Following [76], the reference scheme is configured to minimize the latency of each packet requesting resources, and allocates radio resources in the earliest available slot that meets the packet's transmission requirements. Unlike the PDS scheme, the reference scheduler considers only the latency requirements of the current packet and does not account for future traffic demands in its resource allocation decisions. Since each sensor generates packets every 50 ms, a packet is dropped if it cannot be scheduled for transmission before the next packet is generated 50 ms later.

Performance is evaluated in a 5G NR cell with 5 MHz bandwidth and a subcarrier spacing (SCS) of 30 kHz with a slot duration of 0.5 ms in accordance with [60]. Uplink sensor data packets are generated based on the pattern described in Section 4.3.3.2.2 and transmitted using MCS6 (Modulation and Coding Scheme) to ensure their robust and reliable transmission. Each sensor data packet has a latency requirement derived from 3GPP specifications for enhanced V2X scenarios [77]. Camera-generated packets must be transmitted within a maximum latency of 50 ms. To assess different mixed traffic flow

conditions, two configurations are tested for radar packets. In the first configuration (C1), the maximum latency for radar packets is randomly set to either 10 ms or 5 ms. In the second one (C2), the maximum latency for radar packets is randomly set to either 10 ms or 3 ms. We evaluate performance under different network loads, considering a single autonomous vehicle per cell (scenario S1) and two autonomous vehicles per cell (scenario S2).

Table 13 presents the percentage of transmissions that meet latency requirements for both the PDS scheme and the reference (Ref.) scheme across all scenarios and all latency requirement configurations. The results show that in the low-load scenario (S1) with configuration C1, the PDS scheme can support deterministic service levels as it successfully supports all transmissions within their latency deadline, whereas the reference scheme meets latency requirements for only 85.7% of them. The performance gains of the PDS scheme over standard 5G dynamic scheduling stem from its ability to anticipate short-term traffic demands and allocate current transmissions the radio resources, from all possible options, that satisfy their demand and are less likely needed by upcoming packets to satisfy their demand. By leveraging predictive knowledge of future traffic, the PDS scheme flexibly utilizes the latency budgets of mixed traffic flows with different QoS requirements to better support deterministic communications and increase the percentage of satisfied transmissions. This is visible in Figure 88.a that plots the cumulative distribution function (CDF) of the latency experienced by packets from cameras and radars in scenario S1 under configuration C1. Table 13 shows that both the PDS and reference schemes satisfy all camera transmissions in S1-C1. However, Figure 88.a reveals that while the PDS scheme ensures 100% of camera packets meet their 50 ms latency requirement, it slightly increases the latency of approximately 10% of camera packets beyond 5 ms compared to the reference scheme. This adjustment is intentional, as PDS strategically utilizes the 50 ms latency budget of camera packets to prioritize predicted radar transmissions with stricter latency constraints. As a result, Figure 88.a shows that the PDS scheme reduces the latency of radar transmissions compared to the reference scheme, with greater improvements observed for packets with stricter latency requirements. The reference scheme lacks this predictive knowledge and meets the latency requirements for only 95% of radar transmissions with a 10 ms deadline and just 50% of radar transmissions with a 5 ms deadline (Table 13), whereas PDS ensures deterministic service levels for 100% of radar transmissions.

TABLE 13: PERCENTAGE (%) OF TRANSMISSIONS MEETING THEIR LATENCY REQUIREMENTS

Latency req.	Transmissions	Scenarios			
		S1		S2	
		Ref.	PDS	Ref.	PDS
C1	Total	85.70	100	55.02	97.00
	Cameras (50 ms)	100	100	99.84	97.61
	Radars (10 ms)	95.35	100	12.37	99.89
	Radars (5 ms)	47.95	100	7.89	95.90
C2	Total	78.60	92.90	53.95	87.80
	Cameras (50 ms)	100	100	99.84	96.00
	Radars (10 ms)	95.35	100	12.37	93.66
	Radars (3 ms)	20.08	71.93	3.69	68.95

The benefits of predictive knowledge in dynamic scheduling augments as latency requirements become stricter within mixed traffic flows. Under the C2 configuration, where certain radar packets have a latency requirement of 3 ms, Table 13 shows that the reference scheme satisfies only 78.6% of

transmissions in S1, with this percentage dropping to just 20% for radar packets with the strictest 3 ms deadline. This is compared to 48% of satisfied radar transmissions with a 5 ms latency deadline under the C1 configuration. In contrast, the proposed PDS scheme can better support deterministic service levels as it successfully satisfies 92.9% of all transmissions and nearly 72% of radar packets with a 3 ms latency requirement. The remaining 7.1% of transmissions that fail to meet their latency constraints with PDS correspond to radar packets with large sizes that exceed the transmission capacity within the 3 ms deadline, given the available radio resources in the cell. For instance, in a best-case scenario, 66 RBs are available within a 3 ms window. A radar packet containing data from five detected objects has a size of 3780 bytes and requires 67 RBs when transmitted using MCS6, making it impossible to meet the 3 ms deadline regardless of the scheduling scheme. It is worth noting here that with 5 MHz bandwidth and 30 kHz SCS, there are 11 radio resources or RB per slot, totalling 66 RB within 3 ms. This is a best-case analysis assuming that packet generation is synchronized with the start of the slot and that the scheduling requests and response do not add any additional latency. This finding highlights that PSD effectively guarantees latency requirements for all feasible transmissions under S1-C2 considering the cell resources.

When the traffic load increases (scenario S2), more packets have overlapping transmission windows (as illustrated in Figure 79), leading to greater competition for the same radio resources. In this case, careful allocation of radio resources becomes even more critical. The gains achieved with the PDS scheme, which leverages predicted traffic demands to optimize resource allocation while accounting for both current and future (potentially conflicting) transmissions, become even more significant compared to standard 5G dynamic scheduling. Table 13 shows that the percentage of transmissions meeting latency requirements drops significantly for the reference scheme in S2, with only 55.02% and 53.95% of all transmissions satisfied under the C1 and C2 configurations, respectively. The performance deteriorates further for packets with the most stringent latency requirements. Specifically, Table 13 shows that only 7.89% of radar transmissions with a 5 ms deadline (C1) and just 3.69% of those with a 3 ms deadline (C2) are successfully transmitted with the reference scheme. In contrast, the proposed PDS scheme successfully meets the latency requirements for 97% and 87.8% of all transmissions under the higher-load scenario with mixed traffic flows, underscoring its capacity to better support deterministic service levels. Additionally, it ensures that 95.9% and 68.9% of radar transmissions with 5 ms and 3 ms latency requirements under C1 and C2, respectively, meet their deadlines. It is important to note that the PDS scheme meets the latency requirements for 88.57% of transmissions in S2-C2 using 89.8% of the available radio resources. In contrast, the reference scheme, which consumes 97.6% of the resources, manages to satisfy the latency requirements for only 54.95% of transmissions. This clearly demonstrates the most efficient resource utilization achieved by the PDS scheme by leveraging predictive knowledge of future traffic.

The superior performance of the proposed scheme stems from its ability to anticipate short-term traffic demands and allocate radio resources to current transmissions in a way that minimizes conflicts with upcoming packets. To achieve this, the PDS scheme strategically utilizes the available latency budget of camera packets to allocate resources that ensure the timely delivery of radar packets with stricter latency constraints without compromising camera traffic with more relaxed requirements. This effect is evident in Figure 88.b, where 10% of camera packets experience latencies exceeding 8 ms, allowing most radar transmissions to meet their deadlines. Compared to the lower-load scenario (Figure 88.a), Figure 88.b shows that a wider portion of the 50 ms latency budget for camera packets is leveraged to accommodate radar transmissions when the load increases. These results highlight the effectiveness of

the predictive scheduler in optimizing resource allocation for mixed traffic flows with diverse QoS requirements.
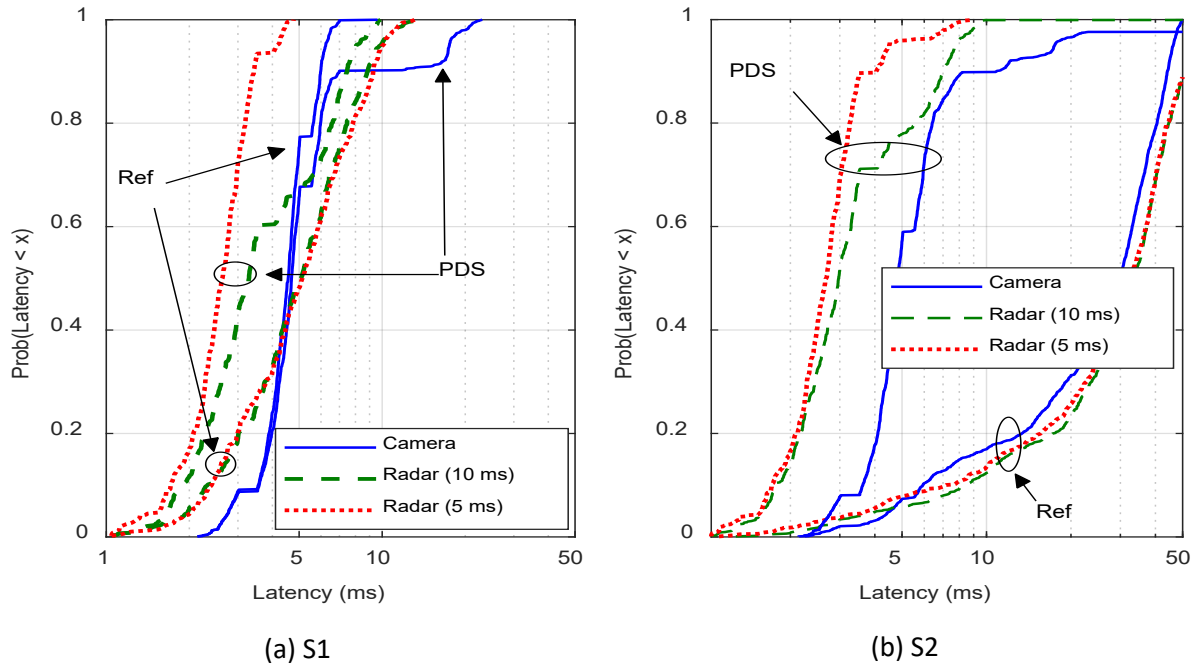
FIGURE 88. CDF OF THE LATENCY UNDER C1 CONFIGURATION.

4.3.3.4.2   Evaluation of the Proposed Predictive Configured Grant Scheduling Scheme

This section compares the performance of the proposed predictive CG scheduler (PCG) against a reference (non-predictive) 5G CG scheduling scheme. The reference scheme periodically allocates RBs for transmitting packets generated by each node. Following [76], the reference scheme allocates RBs to minimize the latency of each packet. For a fair comparison with PCG, the reference scheme uses an allocation period p=50 ms and seeks to allocate RBs to pkt_i within w_i = w(t_i^{ini}, t_i^{end}) with t_i^{ini}= $\hat{t}_i$+Δt_i, t_i^{end}= $\hat{t}_i$-Δt_i+L_i, and Δt_i=σ_i^t, which ensures that a packet generated between $\hat{t}_i$-Δt_i and $\hat{t}_i$+Δt_i can be transmitted within its latency requirement L_i. The reference scheme allocates the amount of RBs necessary to transmit packets of size S_i. We test different configurations of the reference scheme where S_i is established based on the most frequent value of the packet sizes, i.e., the mode of the packet size - mode(s_i)- (C1 configuration), or on the 80th, 90th, 95th, and 99th percentiles of the packet size (C2, C3, C4 and C5 configurations for the reference scheme), respectively. The percentile value is represented by the function $P_x(s_i)$, where x represents the desired percentile. We also evaluate different configurations of the proposed PCG scheduler to assess the impact of the selection of the prediction inaccuracy Δs_i. PCG is evaluated with values of Δs_i equal to the standard deviation of the predicted size σ_i^s (C1 configuration for PCG), or equal to the 80th, 90th, 95th, or 99th percentiles of the absolute error of the prediction, i.e., $P_x(AE)$ with x=80, 90, 95 and 99, and AE=$|s_i-\hat{s}_i|$ (C2, C3, C4 and C5 configurations for PCG, respectively). Table 14 shows the evaluated configurations for PCG and the reference schemes. In addition, we also evaluate a predictive CG scheduler that allocates RBs for packets transmissions following a scheduling policy similar to PCG but only considering the predicted packet size $\hat{s}_i$ and not taking into account prediction inaccuracies.

TABLE 14. EVALUATED CONFIGURATIONS FOR PCG AND REF.

Scheduler	Configurations				
	C1	C2	C3	C4	C5
PCG	$\Delta s_i = \sigma_i^s$	$\Delta s_i = P_{80}(AE)$	$\Delta s_i = P_{90}(AE)$	$\Delta s_i = P_{95}(AE)$	$\Delta s_i = P_{99}(AE)$
Reference	$S_i = \text{mode}(s_i)$	$S_i = P_{80}(s_i)$	$S_i = P_{90}(s_i)$	$S_i = P_{95}(s_i)$	$S_i = P_{99}(s_i)$

The performance of the proposed PCG scheme and the reference scheme is evaluated considering a 5 MHz cell bandwidth, a 30 kHz NR subcarrier spacing (SCS) [60], and a MCS11 (Modulation and Coding Scheme) to balance between robustness and spectral efficiency; similar trends have been observed for other configurations. We consider a mixed traffic flow scenario with latency requirements derived from 3GPP specifications for enhanced V2X scenarios [77]. We consider that 50%, 25% and 25% of the packets must be transmitted within a maximum latency of 50, 20 and 10 ms, respectively; other scenarios with mixed traffic flows and different latency requirements have also been evaluated, and similar trends have been observed.

Figure 89 presents the percentage of transmissions that meet their latency requirements using the proposed PCG scheme and the reference (Ref.) scheme. The performance is shown for the different configurations evaluated for each scheme following Table 14. Figure 89 shows that PCG significantly increases the percentage of packets that meet their latency requirements compared to the reference scheme. PCG can reach satisfaction levels higher than 87% with all the evaluated configurations, while the reference scheme cannot surpass a 64% satisfaction level regardless of the configuration used. Figure 89 also reports the performance achieved with a predictive configured grant scheme that does not consider prediction inaccuracies and pre-allocates resources based only on the predicted packet size $\widehat{s}_i$. In this case, only about 64% of packets meet their latency requirement (similar to the reference scheme) compared with a satisfaction level of 92.4% and 98.9% with the PCG when $\Delta s_i = \sigma_i^s$ (C1) and $\Delta s_i = P_{99}(AE)$ (C5), respectively. These results strongly highlight the importance of accounting for prediction inaccuracies in the design of predictive scheduling schemes capable of supporting deterministic communications. It is worth noting here that the transmissions that PCG cannot satisfy correspond to packets with a size larger than $\widehat{s}_i + \Delta s_i$. In this case, the number of allocated RBs is insufficient for transmitting these packets in time to meet their latency requirement.
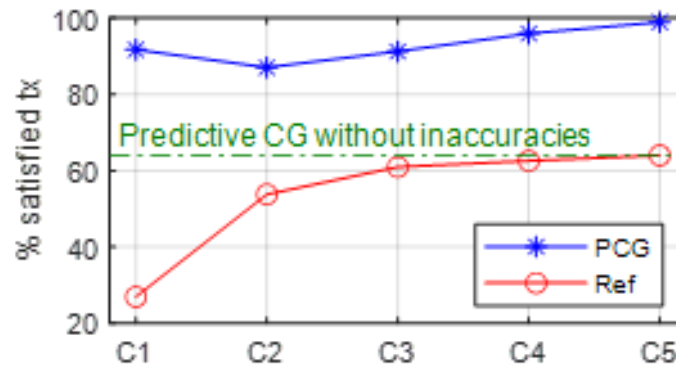
FIGURE 89. PERCENTAGE (%) OF TRANSMISSIONS MEETING THEIR LATENCY REQUIREMENTS

Figure 89 shows that the proposed PCG increases its performance as the prediction inaccuracy Δs_i is set equal to higher percentile values of the prediction's absolute error. Similarly, the reference scheme increases its satisfaction level when it allocates RBs based on the highest percentiles of the packet size.

However, these improvements come at the expense of pre-allocating a larger number of RBs per packet than actually needed, thus reducing resource efficiency. This is illustrated in Figure 90, where Figure 90.a represents the percentage of RBs allocated for the transmission of packets and Figure 90.b represents the percentage of pre-allocated RBs that are actually unutilized. Figure 90.a shows that the percentage of allocated RBs increases for PCG when Δs_i is set equal to higher percentile values of the prediction's absolute error, and increases for the reference scheme when allocating RBs based on the highest size percentiles. However, PCG pre-allocates in general a lowest number of RBs than the reference scheme while achieving higher satisfaction levels (Figure 89). In addition, PCG results in a significantly lower percentage of non-utilized RBs compared to the reference scheme, which highlights its highest resource efficiency. higher percentile values of the prediction's absolute error. Similarly, the reference scheme increases its satisfaction level when it allocates RBs based on the highest percentiles of the packet size. However, these improvements come at the expense of pre-allocating a larger number of RBs per packet than actually needed, thus reducing resource efficiency. This is illustrated in Figure 90, where Figure 90.a represents the percentage of RBs allocated for the transmission of packets and Figure 90.b represents the percentage of pre-allocated RBs that are actually unutilized. Figure 90.a shows that the percentage of allocated RBs increases for PCG when Δs_i is set equal to higher percentile values of the prediction's absolute error, and increases for the reference scheme when allocating RBs based on the highest size percentiles. However, PCG pre-allocates in general a lowest number of RBs than the reference scheme while achieving higher satisfaction levels (Figure 89). In addition, PCG results in a significantly lower percentage of non-utilized RBs compared to the reference scheme, which highlights its highest resource efficiency.

(a) Allocated RBs   (b) Allocated but non-used RBs

FIGURE 90. PERCENTAGE OF ALLOCATED RBS AND PERCENTAGE OF ALLOCATED RBS THAT ARE NON-USED FOR PACKET TRANSMISSIONS

Following the URLLC principles in 5G, the reference scheme focuses on minimizing the transmission latency. On the other hand, PCG has been designed to support deterministic communications in beyond 5G networks, and hence prioritizes maximizing the number of transmissions that meet their latency requirements. Figure 91 plots the CDF of the latency experienced by packets with 10, 20, and 50 ms latency requirements, respectively. The figure shows that packets with the most stringent latency requirement (10 ms) experience the lowest latency values when using the PCG scheme. On the other hand, PCG slightly increases the latency of packets with more relaxed latency constraints. This is particularly the case for packets with a 20 ms requirement as these packets often compete for the same RBs as the packets that require 10 ms latency. PCG intentionally delays the transmission of packets with more relaxed latency requirements, while still meeting their latency deadline, to ensure that packets with tighter deadlines can be transmitted in time. In contrast, a scheduler that seeks minimizing the

transmission latency (Ref.) does not leverage varying latency requirements in mixed traffic flows to increase the percentage of satisfied transmissions, and the latency experienced increases uniformly up to the maximum allowed latency.
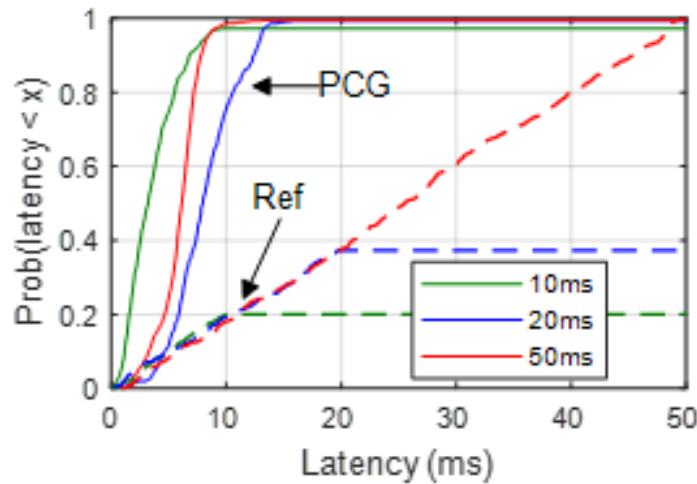
FIGURE 91. CDF OF THE LATENCY EXPERIENCED BY PACKETS WITH **10, 20** AND **50** MS LATENCY REQUIREMENTS WITH **PCG** AND THE REFERENCE SCHEME

4.3.3.4.3 Evaluation of the Proposed Predictive Scheduling and Flexible Duplexing scheme

To evaluate the joint management of radio resources using flexible duplexing, our proposed scheme incorporates traffic characterization and prediction, as detailed in Section 4.3.3.3. The scenario considers an in-vehicular environment with 10 sensors including 5 cameras and 5 radars, mounted on the top, front, rear, and sides of the vehicle (deployment is illustrated in Figure 92), all transmitting packets to a HC, centrally located within the vehicle.
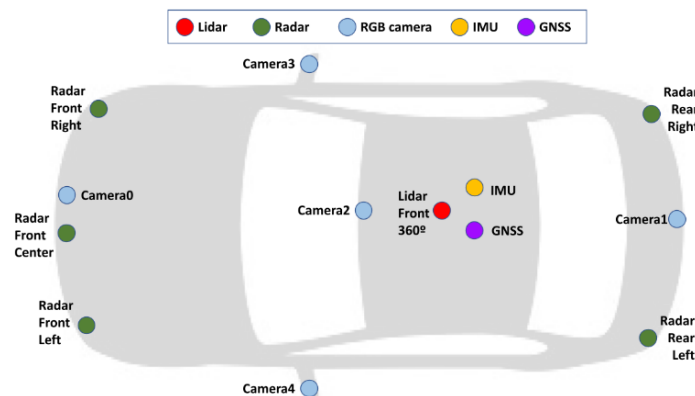
FIGURE 92. ILLUSTRATION OF SENSOR DEPLOYMENT IN A VEHICLE

Figure 93 illustrates the spatial diversity matrix of sensors mounted on a vehicle, where each cell represents the relative spatial separation between a pair of sensors, ranging from 0 (closely located) to 1 (far apart). Sensors located in close proximity (e.g., Camera0 and Radar Front Center) exhibit low spatial diversity (blue areas in Figure 93), which means that if their transmissions are scheduled simultaneously, they are more likely to experience higher levels of interference, potentially resulting in degraded performance or unsatisfied transmission. On the other hand, sensors located at opposite ends of the vehicle, such as front sensors (Camera0, Radar Front Left/Right) and rear sensors (Camera1, Radar Rear Left/Right), exhibit high spatial diversity (red areas in Figure 93), indicating a lower likelihood of

interference when scheduled at the same time. This spatial separation can be exploited by the scheduler to allocate uplink and downlink resources to sensor pairs with high spatial diversity, thereby reducing the impact of IBE. For example, assigning resources to the rear-right radar while front sensors are actively transmitting can significantly minimize interference and enhance communication reliability. While we do not explicitly predict DL traffic, we assume it constitutes a fixed percentage (X%) of the corresponding UL traffic, allowing us to estimate DL requirements for scheduling purposes. The DL traffic represents the processed counterpart of the UL traffic.
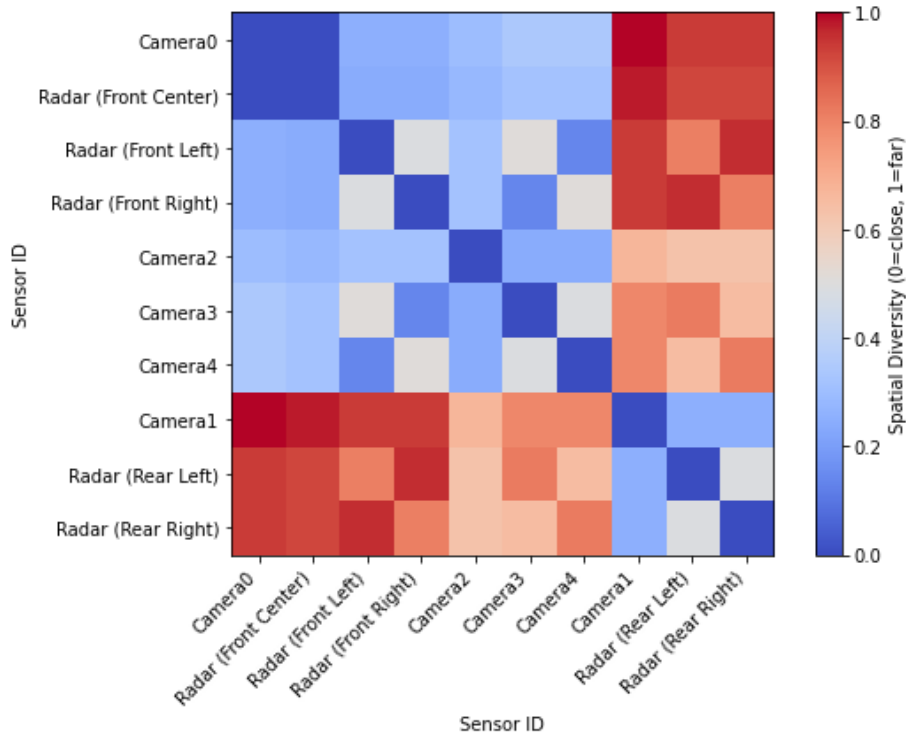
FIGURE 93. SPATIAL DIVERSITY CONSIDERING IN VEHICULAR SENSORS

The performance of the proposed scheme is evaluated against a baseline implemented using the same predictive dynamic scheduling framework described in Section 4.3.3.2.1 and compliant with 3GPP standard [60]. However, the baseline does not consider spatial diversity during resource allocation. The objective of this comparison is to highlight the benefits of the proposed scheme in jointly allocating radio resources while accounting for spatial diversity and its effect on IBE. In both schemes, resources are always allocated to ensure transmission occurs, but satisfaction of a transmission is evaluated using the SINR value.

$$SINR_i = \frac{P_{DL[i]}}{IBE_{UL[j] \rightarrow DL[i]} + P_{noise}}$$

Where $P_{DL[i]}$ represents the DL power of *SNE_i* from *HC*, $IBE_{UL[j] \rightarrow DL[i]}$ denotes the interference caused by the UL transmission of *SNE_j* affecting the DL transmission of victim *SNE_i*. This IBE interference is calculated as explained in Section 4.2.1 and 3GPP guidelines provided in TS 36.521 [52] (see Table 6.4.3.2-1), using path loss (PL) values from Deliverable D2.3 [4] based on vehicular scenario models. P_{noise} represents additive white Gaussian noise (AWGN). By comparing SINR values, we evaluate how

effectively the proposed method maintains the transmission quality through spatially-aware resource allocation.

Figure 94 represents the CDF of SINR values for both the predictive and benchmark schemes, clearly illustrating the superiority of the predictive method. The predictive scheme consistently achieves higher SINR values, as shown by its right shifted CDF curve. This rightward shift in the CDF highlights the efficient utilization of spatial diversity. By prioritizing transmissions in spatially diverse regions, the proposed scheme effectively enhances SINR performance. At the median (CDF = 0.5), the predictive method achieves an SINR of approximately 24 dB, compared to about 16 dB for the benchmark, with an improvement of around 8 dB. Similarly, at the lower tail (CDF = 0.2), the predictive method reaches approximately 12 dB, versus 7 dB for the benchmark, indicating better performance even under less favorable conditions. This indicates that the predictive method successfully allocates radio resources in spatially diverse regions, leading to a more consistent and balanced distribution of resources across sensors. By effectively leveraging the spatial diversity between sensors, it minimizes the impact of interference caused by IBE, thereby improving the achieved SINR.
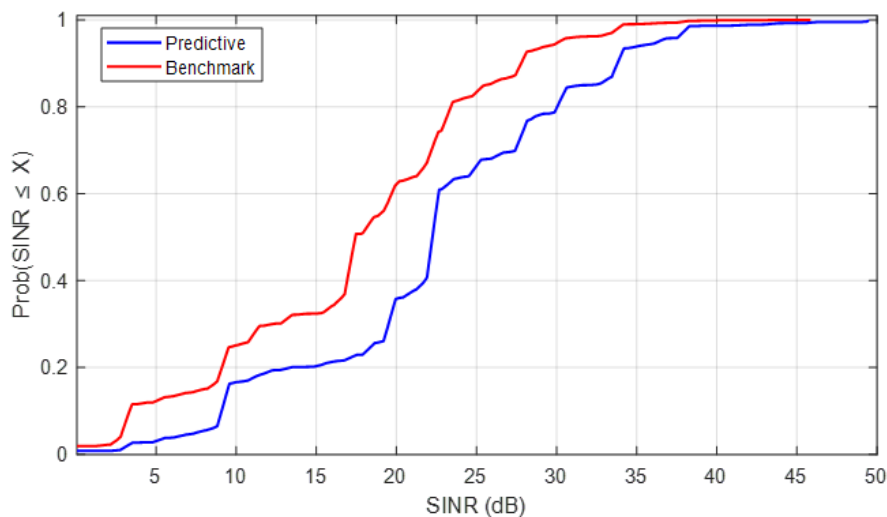
FIGURE 94. CDF OF SINR VALUES FOR PREDICTIVE AND BENCHMARK METHODS ACROSS ALLOCATED TRANSMISSIONS

Figure 95 represents the percentage of the satisfied transmissions as a function of transmit power under two different cases, each comparing the predictive and benchmark methods. Case 1 (Figure 95-left) considers SNEs transmitting UL packets with power ranging from 0 to 23 dB, while the HC's DL transmit power is fixed at 10 dB. Case 2 (Figure 95-right) is considers varying HC DL transmit power from 0 to 23 dB, while SNEs transmit with a fixed power of 10 dB.

In Case 1 (Figure 95-left), as SNE transmit power increases, interference (IBE) to neighbouring SNEs also increases, significantly degrading SINR. This leads to a noticeable decline in the satisfaction percentage, particularly for the benchmark method, which drops sharply from nearly 100% to below 50% at 23 dB. In contrast, the predictive method maintains a much higher satisfaction rate, remaining above 80% even at high SNE power. This demonstrates the predictive method's superior interference awareness and its ability to allocate resources effectively by considering distant regions where the impact of IBE is minimized. In Case 2 (Figure 95-right), increasing HC transmit power improves SINR at the SNEs by overcoming UL interference. Both methods show rising satisfaction rates, but the predictive method consistently outperforms the benchmark across all power levels. For instance, at 6 dB HC power, the predictive method achieves about 93% satisfaction versus 80% for the benchmark. The difference

narrows at higher powers (≥14 dB), where both reach nearly 100%, but the predictive method's advantage is most notable under low-power conditions which is a critical aspect for sub-network communications.

Figure 95 also highlights the predictive method's clear advantage in supporting low-power transmissions, which is essential for sub-networks where energy efficiency and minimal interference are key requirements. In Case 1, the predictive method maintains over 95% satisfaction in low-power conditions (0–10 dB), while the benchmark's performance drops significantly. Similarly, in Case 2, even at just 2 dB HC power, the predictive method achieves over 80% satisfaction, whereas the benchmark falls below 70%, emphasizing the predictive method's ability to ensure reliable communication even with minimal power.
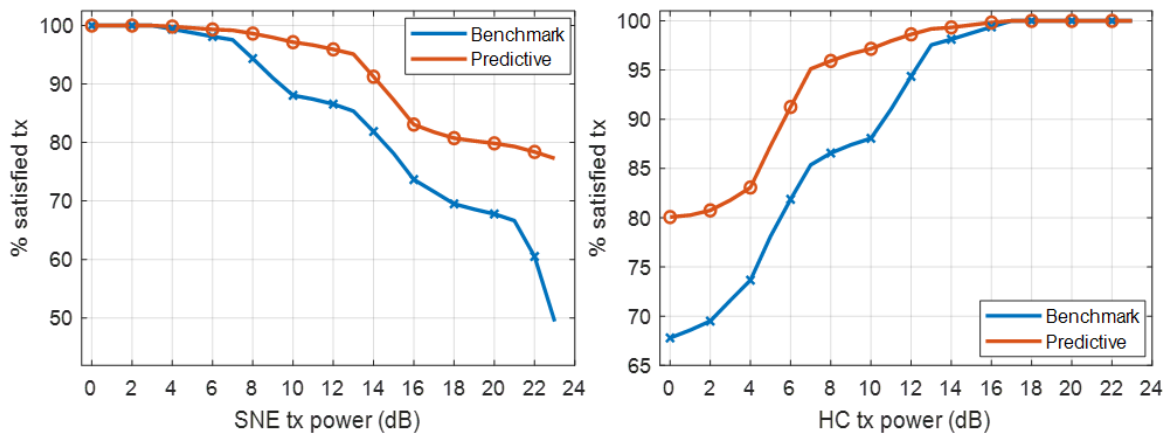
FIGURE 95. TRANSMISSION SATISFACTION UNDER VARYING SNE (LEFT) AND HC (RIGHT) TRANSMIT POWER FOR PREDICTIVE AND BENCHMARK METHODS

This capability is particularly valuable in subnetwork environments, where maintaining low transmission power helps reduce interference without compromising performance. The predictive method thus offers a robust and efficient solution for future wireless systems operating under constrained power and dense deployments. Moreover, the predictive approach enables more reliable and efficient resource allocation, clearly showcasing its advantage in enhancing both network performance and overall transmission satisfaction.

4.3.3.5 Summary and recommendations

The proposed research focuses on enhancing scheduling mechanisms by leveraging predictive knowledge to proactively allocate resources, aiming to improve resource efficiency and minimize latency for time-sensitive services. The primary goal is to achieve deterministic service provisioning in beyond-5G systems while considering prediction inaccuracies, which is critical for meeting QoS requirements like bounded latency and reliable communication. In the first part, we investigate how predictive knowledge can enhance resource allocation by introducing a predictive dynamic scheduler, which leverages traffic predictions to allocate resources efficiently to incoming scheduling requests, ensuring they meet their latency requirements while avoiding resources likely to be needed by future predicted packets. Our evaluation demonstrates that the proposed scheme outperforms standard 5G dynamic scheduling and significantly enhances the ability to provide deterministic service levels with bounded latency deadlines in scenarios with mixed traffic flows and varying QoS requirements. Our work further extends existing methods by incorporating prediction inaccuracies into the scheduling process, thereby mitigating the impact of incorrect predictions and enhancing resource allocation efficiency. This is achieved through the introduction of the second proposal, the predictive configured grant scheme,

which improves grant-free scheduling. The PCG scheduling scheme leverages traffic predictions to pre-allocate resources and adjusts the scheduler's decision-making process by accounting for prediction inaccuracies. This approach effectively manages stochastic traffic and ensures deterministic service provisioning, even in unpredictable conditions. By considering prediction uncertainties, the scheme minimizes the effects of incorrect predictions, especially in sensitive environments (such as in vehicular scenarios) with mixed traffic and varying demands. Further extending the proposals, we presented a joint radio resource allocation of UL and DL utilizing flexible duplexing and predictive scheduling. The proposed method dynamically adapts to varying network conditions by jointly managing uplink and downlink resources with spatial awareness. It effectively minimizes the impact of in-band emissions (IBE) and enhances network performance by leveraging spatial diversity and by using flexible frequency band for allocation of radio resources. This approach ensures efficient and reliable transmissions, particularly in sub-network scenarios where interference sensitivity and power limitations are critical.

It is therefore recommended to incorporate predictive knowledge into scheduling decisions, as it significantly enhances deterministic service provisioning. Additionally, leveraging spatial diversity in combination with predictions enables more effective radio resource management by identifying and exploiting low-interference regions dynamically. However, to mitigate the impact of incorrect predictions, especially in sensitive environments, it is essential to integrate mechanisms that account for prediction uncertainties. In addition, advanced error correction techniques could also be developed to handle inaccuracies, which will improve network performance, help meet bounded latency requirements and enable more efficient and reliable communication. Furthermore, refinement of prediction models is essential, particularly in handling unpredictable traffic patterns. While LSTM-based networks have shown promising results, further exploration of hybrid models that combine reinforcement learning and predictive analytics could offer even more accurate predictions in complex, real-time environments.

In conclusion, this research contributes to predictive scheduling by improving resource utilization and ensuring bounded latency to achieve deterministic communication services. The proposed PDS and PCG schemes have significant potential for optimizing resource allocation and meeting QoS requirements, particularly in high-demand, low-latency, and mixed traffic scenarios. In addition, the proposed joint predictive scheduling with flexible duplexing method further enhances performance through joint uplink-downlink resource management. It achieves high transmission satisfaction for both UL and DL, by mitigating the impact of interference via spatial separation considering predictive knowledge. It ensures reliable communication even in low-power or high-interference scenarios, which is crucial for sub-network environments and aligns well with the goals of reliable, energy-efficient future wireless systems. Further development will pave the way for more intelligent and adaptive scheduling mechanisms for future communication networks.

4.4   MULTI-LINK SOLUTIONS FOR IMPROVED RELIABILITY

Reliable communication in in-X subnetworks can be jeopardized by the presence of obstacles, that leads to reduced received signal power and risk of transmission errors. Here, we present our findings on enhancing communication reliability through multi-link strategies in environments characterized by obstacles and non-line-of-sight communication. Specifically, we explore solutions leveraging cooperative communication and network coding techniques.

4.4.1 Cooperative in-X Communication

In the deliverable D3.1, we elaborated on the problem of cooperative communication within industrial subnetworks, though the method can be adapted to other subnetwork types operating with tight communication cycles.

The cooperative communication method proposed focuses on enhancing ultra-reliable, low-latency communication (URLLC) by leveraging link diversity to mitigate blockage effects, which are common in factory environments due to metallic obstructors. As extensively documented in the literature, relaying has been recognized as an effective strategy for combating fading through spatial diversity. Studies indicate that two-hop relaying protocols significantly improve both capacity and quality of service. However, integrating relays efficiently into industrial subnetworks requires tailored communication protocols.

Several investigations have explored relay-assisted URLLC transmission, emphasizing cooperative transmission strategies, CSI-agnostic methods, and power minimization algorithms (e.g., [79]-[82]). While cooperative transmission methods enhance network reliability, existing approaches often lack adaptability to instantaneous channel variations, leading to suboptimal spectral efficiency. Moreover, some solutions prioritize direct transmission for stronger links but fail to provide explicit criteria for selecting between single-hop and two-hop SNEs. Others optimize power consumption while neglecting timing constraints critical for short control cycles.

The solution presented here specifically addresses these challenges in in-factory subnetworks. It aims to support ultra-short, highly reliable communication while minimizing overall transmitted power. This reduction not only conserves energy but also mitigates interference with neighbouring subnetworks. Our approach employs LC elements to relay packets from SNEs experiencing challenging propagation conditions to an HC with embedded controller capabilities. The optimization problem, formulated to minimize total transmit power while meeting stringent low-latency criteria, is inherently nonconvex. To address this, we employ the sequential parametric convex approximation (SPCA) technique, which reformulates the problem into a convex semidefinite programming (SDP) framework. This transformation allows for efficient numerical solutions using optimization tools like CVX.

By integrating cooperative relaying within industrial subnetworks, our method enhances communication reliability, improves resource efficiency, and ensures seamless operation under strict latency requirements as targeted by the robot control use case presented in D2.2.

4.4.1.1 Setting Up the Scene

We examine cooperative transmission in an industrial subnetwork with N SNEs wirelessly connected to K LCs and one HC. The HC issues control commands, while LCs provide radio communication support. The sets of LCs and SNEs are denoted by $\mathcal{K}$ and $\mathcal{N}$, respectively. Each SNE and LC has a single antenna. Figure 96 illustrates the uplink (UL) transmission model with multiple LCs.

SNEs transmit either directly to the HC (single-hop, $\mathcal{N}_{1h}$) or via an LC (two-hop, $\mathcal{N}_{2h}$). In Figure 96, SNEs D and E use single-hop transmission, while A, B, and C employ two-hop cooperative mode. In the latter, SNEs first transmit to the HC, then relay via an LC, allowing the HC to combine direct and relayed signals. We consider DF and AF relaying, each with two configurations: (1) TDMA, where SNEs are assigned time slots to prevent interference, and (2) FDMA, where they use orthogonal subchannels within a shared

slot. In both cases, UL transmissions occur over W Hz bandwidth, with SNEs sending B_n bits that must reach the HC within T seconds.
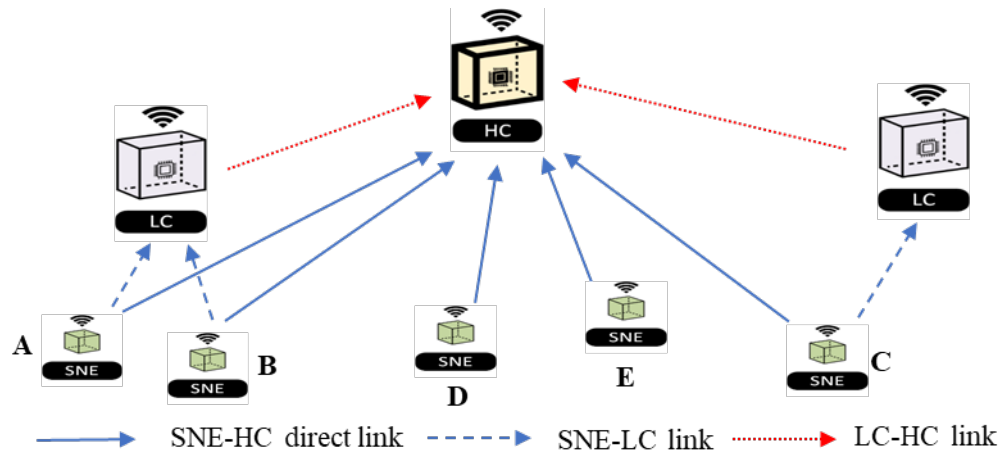
FIGURE 96. SYSTEM MODEL FOR MULTIPLE LCS SUBNETWORK TRANSMISSION.

To enhance clarity in the communication process and define the role of each component, Figure 97 presents a step-by-step breakdown:

- Channel Estimation: SNEs and LCs transmit pilot sequences, allowing the HC to evaluate channel conditions. Based on these assessments, the HC determines whether SNEs should operate in single-hop or two-hop transmission modes to optimize both efficiency and reliability.
- Resource Management: The HC oversees the allocation of resources such as time slots, bandwidth, and power for UL transmission. It dynamically adjusts resource distribution for single-hop and two-hop SNEs to enhance power efficiency and minimize latency.
- UL Transmission: SNEs initiate data transmission according to the HC's scheduling plan. Single-hop SNEs send data directly to the HC, whereas two-hop SNEs first transmit data to both an LC and the HC.
- Relaying: Designated LCs relay signals using either AF or DF techniques. During the second phase of cooperative transmission, LCs forward the received signals to the HC, thereby completing the two-hop transmission.
-
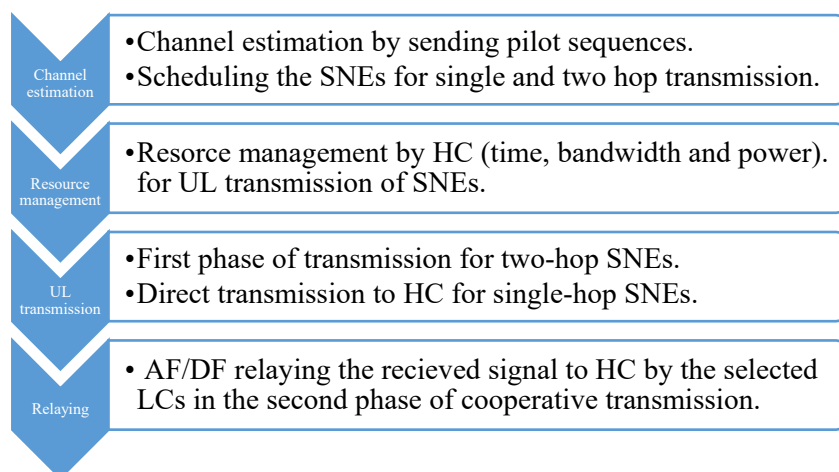
FIGURE 97. OVERVIEW OF THE COMMUNICATION PROCESS.

4.4.1.2   Proposed Communication Protocols for Relay-assisted Network

We now elaborate on our communication protocols for the TDMA and FDMA cases.

Figure 98 illustrates our proposed TDMA protocol. The UL timeslot is divided into three variable-duration sub-slots: the first phase of two-hop transmissions, single-hop transmissions, and the second phase of two-hop transmissions. In the first sub-slot, SNEs in the set $\mathcal{N}_{2\hbar}$ transmit their signals, which are received by both the HC and the LC. Each SNE in $\mathcal{N}_{2\hbar}$ is served by the LC with the best channel conditions.

In the second phase, the LC acts as a DF/AF relay, forwarding the received messages to the HC. The HC then combines the energy from the first-phase signal with that forwarded by the LC before decoding. Single-hop transmissions are received only by the HC.

To allow sufficient processing time at the LCs, SNEs in two-hop mode are scheduled with a gap between receiving and retransmitting packets. This ensures efficient data processing without extending the timeslot duration.

In the second phase sub-slot, the LC forwards the received messages to the HC, acting as a DF/AF relay. The HC combines the energy from the first-phase signal with that forwarded by the LC before decoding. Single-hop transmissions are only received by the HC.

To allow sufficient processing time at the LCs, SNEs in two-hop mode are scheduled with a gap between receiving and retransmitting packets. This ensures efficient data processing without extending the timeslot duration.

In the FDMA case, all SNEs share the same time slot but use separate bandwidths to avoid interference. Figure 99 shows the proposed cooperative FDMA protocol. For two-hop transmissions, a processing time is included to account for DF processing at the LCs.

In Figure 99, β_n and β_n^s represent the bandwidth fractions allocated to SNE n during the first and second phases of cooperative transmission, respectively. Due to limited bandwidth,

$$\sum_{n \in \mathcal{N}} \beta_n = 1, \quad \sum_{n \in \mathcal{N}_{2\hbar}} (\beta_n^s - \beta_n) \le 0$$

In TDMA, since all SNEs use the full bandwidth, we set $\beta_n = 1$ and $\beta_n^s = 1$.

For a comprehensive description of the signal model, please refer to [83]. The HC employs an algorithm to determine whether a SNE should be scheduled for single-hop or two-hop mode, opting for the shortest expected transmission time between the two modes. Further details about this classification algorithm can be found in [83].
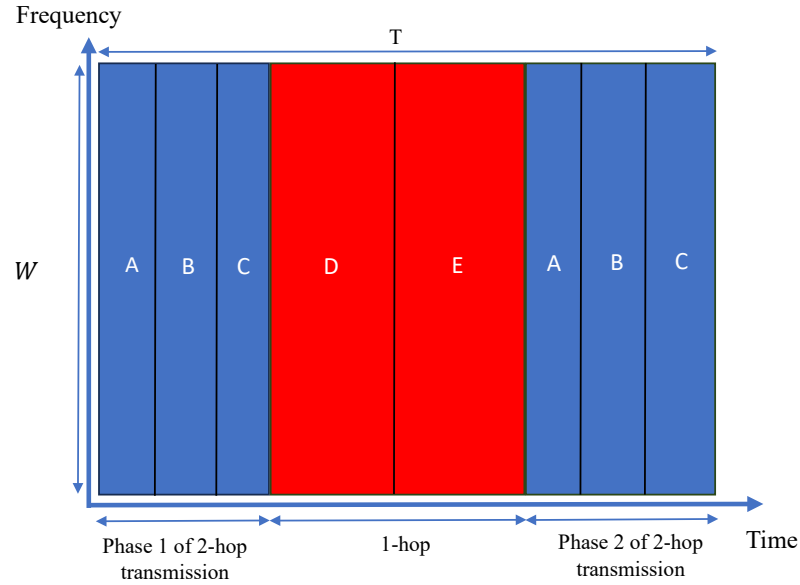
FIGURE 98. PROPOSED COOPERATIVE TDMA PROTOCOL FOR RELAY-ASSISTED NETWORK
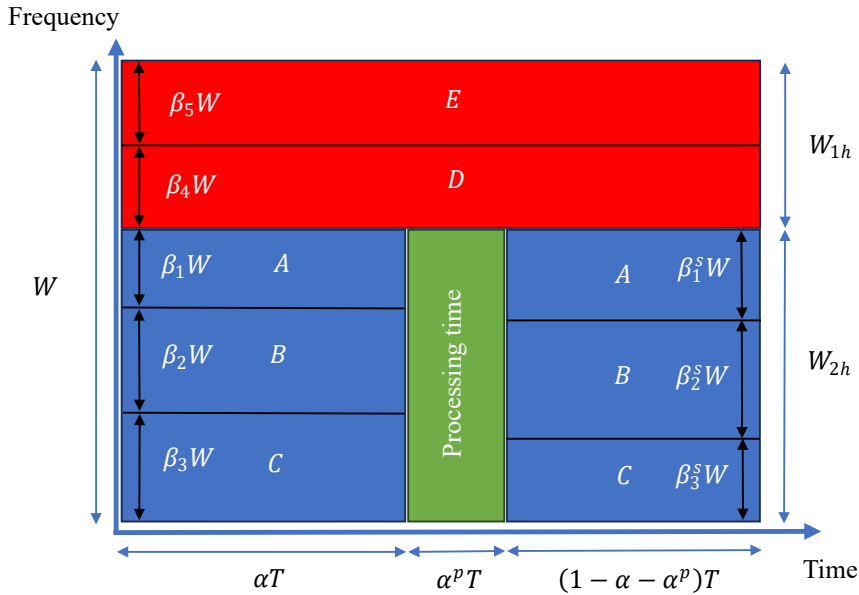
FIGURE 99. PROPOSED COOPERATIVE FDMA PROTOCOL FOR RELAY-ASSISTED NETWORK

We generally consider the imperfect CSI (I-CSI) case, where channel responses are estimated using pilot training sequences for each SNE. The estimation of HC-LC and LC/HC-SNE channels utilizes L training symbols. A more extensive use of pilot sequences results in reduced channel errors. For a detailed mathematical description of channel error and its dependency on L, please refer to [83].

In addressing the impact of outage errors stemming from inaccuracies in channel estimation, the transmitter has the flexibility to mitigate these errors by reducing the data transmission rate through a rate discount factor, denoted as θ, where $0 < θ < 1$. However, opting for smaller values of θ to achieve lower error probability entails a trade-off: while it minimizes errors, it also reduces the transmission rate. A low transmission rate may lead to resource overflow, as it might not be possible to successfully accommodate all transmissions within the time slot. On the other side, using a high transmission rate enabled by a high θ, may increase the failure probability. Hence, the optimal selection of the discount factor becomes crucial in balancing these trade-offs.

4.4.1.3   Proposed Communication Protocol for Multiple RIS-assisted Network

In the previous section, we explored cooperative communication with multiple LCs to enhance URLLC transmission. Now, we replace each LC with an RIS while maintaining the same system configuration and examine the UL transmission scenario. Figure 100 illustrates an RIS-assisted UL channel comprising one HC, a set $\mathcal{N}$ of N SNEs, and a set $\mathcal{K}$ of K RISs. Both the HC and each SNE are equipped with a single antenna. Each RIS, indexed by $k \in \mathcal{K}$, consists of J_k reflecting elements and is optimized to support communication between the HC and the SNEs.

SNE-HC Link ⟶ **RIS-HC Link** ┈┈┈▶ **SNE-RIS Link** ╌╌╌▶
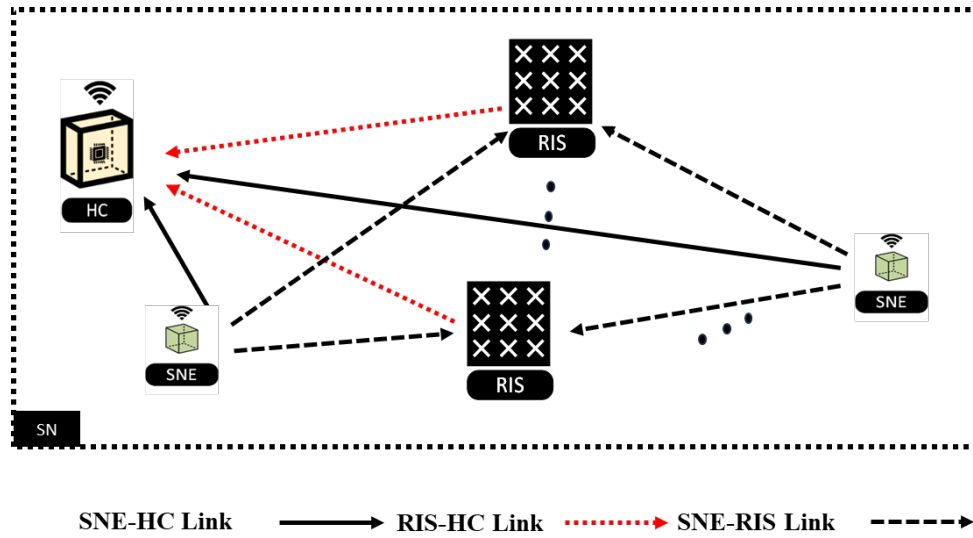
FIGURE 100. SYSTEM MODEL FOR MULTIPLE RISS IN THE SUBNETWORK.

We consider each RIS as an ideal passive element, which implies no frequency selectivity. As a result, our model focuses solely on TDMA transmission. It is worth noting that FDMA leads to significant performance degradation compared to TDMA, as demonstrated in [84].

The transmitted signal from SNE n is $\sqrt{P_n}x_n$. The phase shift matrix of RIS k can be optimized through a diagonal matrix $\Phi_k = \mathrm{diag}\left(e^{j\phi_{k1}}, \cdots, e^{j\phi_{kJ_k}}\right) \in C^{J_k \times J_k}$ with $\phi_{kj} \in [0,2\pi], k \in \mathcal{K}, and j \in \mathcal{J}_k = \{1,2,\cdots,J_k\}$, where Φ_k captures the effective phase shifts applied by all reflecting elements of RIS k. Thus, the received signal of n-th SNE at the destination through the direct link and multiple RISs is denoted as:

$$y_n = \sqrt{P_n}\left(\left(h_n^d\right)^* + \sum_{k=1}^{K} h_{n,k}^H \Phi_k h_k^r\right)x_n + z_0, \ \forall n \in \mathcal{N}$$

where $h_n^d, \boldsymbol{h_{n,k}} \in C^{J_k}$, and $\boldsymbol{h_k^r} \in C^{J_k}$ represent the channel responses from SNE n to the HC, from SNE n to RIS k and from RIS k to the HC respectively. The term z_0 denotes the additive white Gaussian noise. Since the signal transmitted through multiple RISs is weak, we neglect multi-hop transmission among RISs when deriving the received signal [85].

4.4.1.4   Channel estimation in RIS scenario

We assume quasi-static fading for all channels, meaning the channel coefficients stay constant within one coherence interval. In RIS-aided systems, a key challenge in channel estimation is the scaling ambiguity, which prevents separate identification of the RIS-HC and SNE-RIS channels [4].

By rewriting the cascaded channel, we get: $h_{n,k}^H \Phi_k h_k^r = \phi_k^T u_{n,k}$ where $u_{n,k} = \mathrm{diag}\left(h_{n,k}^H\right)h_k^r \in C^{J_k}$ and $\phi_k = \left[e^{j\phi_{k1}}, \cdots, e^{j\phi_{kJ_k}}\right]^T$. As shown in [86], the cascaded channel $\boldsymbol{u_{n,k}}$ and direct channel h_n^d are

sufficient for RIS-aided communication design. Therefore, most existing work focuses on estimating these channels separately [86]. Using TDMA, channel estimation is performed for each SNE individually. Following a DFT-based training scheme from [87], we estimate the cascaded channel for each RIS by activating the k-th RIS while deactivating others, inspired by the on-off scheme in [88]. Figure 101 illustrates the channel estimation process. Under this approach, we rewrite RIS equation to estimate the channel for the k-th RIS as follows:

$$y_{n,k} = \sqrt{P_n}[1, \phi_k^T]\tilde{u}_{n,k}x_n + z_0, \ \forall n \in \mathcal{N},$$

where $\tilde{u}_{n,k} = [h_n^d, u_{n,k}]$. Our goal is to estimate $\tilde{u}_{n,k}$ using the known pilot vector ϕ_k, which depends on the pilot sequence L_k used for channel estimation. To uniquely estimate $\tilde{u}_{n,k}$, the number of time slots for channel training of the k-th RIS must satisfy $L_k \geq (J_k + 1)$.
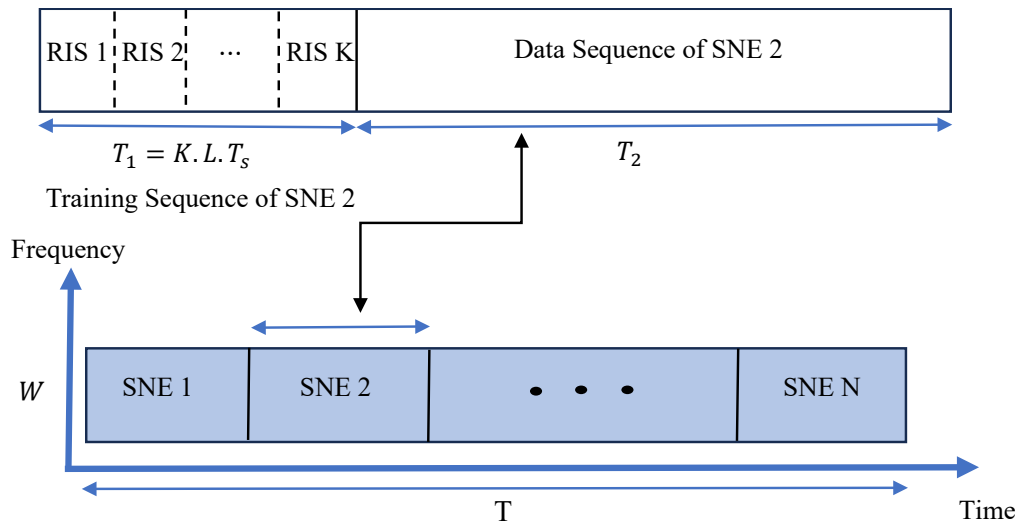
FIGURE 101. PROTOCOL FOR UL TRANSMISSION AND CHANNEL ESTIMATION WITH MULTIPLE RISs.

According to [86], two widely used methods for channel estimation are: 1) Least Squares (LS) Estimator and 2) Linear Minimum Mean-Squared-Error (LMMSE). The LMMSE approach generally achieves lower estimation error than LS because it utilizes prior knowledge of the channel distributions. Using LMMSE, the error variance for the channels, assuming all channels experience uncorrelated Rician fading, is given by [86]:

$$\text{var}\left(\widehat{h_n^d}\right) = \text{var}(\hat{u}_{n,k}) = \frac{1}{1 + L_k \cdot P_n/\sigma_0}, \ \forall k \in \mathcal{K}$$

where σ_0 is the noise power. Therefore, in special case that $J_k = J$ and $L_k = L, \forall k \in \mathcal{K}$ the total training sequence time for SNE n as shown in Figure 101 is obtained by: $T_1 = K \cdot L \cdot T_s$, which requires a large amount of pilot training overhead, by increasing K and J_K compared to relay. In practice, channel estimation is not required in every time frame, as it depends on the channel conditions. In many cases, this leads to high training overhead, especially when the number of reflecting elements J_K is large, significantly reducing the available time slots for data transmission. To minimize pilot overhead, the element grouping (EG) method is applied, as proposed in [87], [88]. In RIS-aided communication systems, EG groups adjacent elements and assigns them a common reflection pattern, which is particularly effective when the elements are closely spaced.

For the rest of discussion, we assume the LMMSE estimators with $L_k = (J_k + 1)$.

4.4.1.5   Proposed method for transmit power optimization with multiple LCs

Within industrial subnetworks, minimizing the total emitted power not only extends SNE battery life but also diminishes interference in neighbouring subnetworks. In this section, we introduce an approach aimed at minimizing the total transmit power across all SNEs and LCs while effectively managing the timing constraints.

To manage the complexity of jointly selecting relays and minimizing power, we first determine the optimal transmission link before optimizing transmit power. Assuming constant power and uniform bandwidth allocation, we reformulate the objective to maximize the transmission rate, thereby minimizing total delay. This involves calculating the maximum channel gain for all LCs and the HC, then selecting the link with the highest gain, as described in [54] . Algorithm 1 in [54] is then applied to choose the strongest transmission link. With the sets $\mathcal{N}_{1\hbar}$, $\mathcal{N}_{2\hbar}$, and $\mathcal{D}$ identified, the optimization problem in TDMA case can be formulated as:

$$\min\left(\sum_{n\in\mathcal{N}} P_n + \sum_{n\in\mathcal{N}_{2\hbar}} P^s_{D_n}\right),$$

$$\text{s.t.} \sum_{n\in\mathcal{N}_{1\hbar}} \frac{B_n}{W\theta\log_2(1+\hat{g}^d_n)} + \sum_{n\in\mathcal{N}_{2\hbar}} \frac{B_n}{W\theta\log_2(1+\hat{g}_{n,D_n})} + \sum_{n\in\mathcal{N}_{2\hbar}} \frac{B_n}{W\theta\log_2(1+\hat{g}^d_n+\hat{g}_{D_n,p})} \leq T_U,$$

$$P_n \leq P_{max}, \forall n \in \mathcal{N}$$

$$P^s_{D_n} \leq P_{max}, \forall n \in \mathcal{N}_{2\hbar}$$

where P_n and $P^s_{D_n}$ represent the transmit power of each SNE and LC, respectively. $\hat{g}^d_n$, $\hat{g}_{n,D_n}$, and $\hat{g}_{D_n,p}$ are the SNR of nth SNE- HC, SNE- LC, and LC- HC links, respectively.

The first constraint concerns the total time limitation for transmitting a packet from each of the SNEs, where T_U is obtained by subtracting the total training sequence time from the time slot i.e., $T_U = T - NLT_s$, with $T_s = 1/W$ denoting the symbol period. It is important to note that the problem is formulated for the general I-CSI case. For the perfect CSI (P-CSI) case, $\theta = 1$ and $T_U = T$. Additionally, the second and third constraints are related to the maximum allowable power. We solve this problem utilizing the SPCA method. For a comprehensive understanding, please refer to [54] for detailed insights.

With a similar approach the problem formulation in FDMA case can be obtained as follows:

$$\min_{\mathbf{P},\beta,\alpha}\left(\sum_{n\in\mathcal{N}} P_n + \sum_{k\in\mathcal{K}} P^s_k\right),$$

$$\text{s.t.} \frac{B_n}{\beta_n W\theta\log_2(1+\hat{g}^d_n)} \leq T', \forall n \in \mathcal{N}_{1\hbar}$$

$$\frac{B_n}{\beta_n W\theta\log_2(1+\hat{g}_{n,D_n})} \leq \alpha T', \forall n \in \mathcal{N}_{2\hbar}$$

$$\frac{B_n}{\beta^s_n W\theta\log_2(1+\hat{g}_{D_n,p})} \leq (1-\alpha-\alpha^p)T', \forall n \in \mathcal{N}_{2\hbar}$$

$$\sum_{n\in\mathcal{N}} \beta_n = 1$$

$$\sum_{n\in\mathcal{N}_{2\hbar}} (\beta^s_n - \beta_n) \leq 0$$

$$0 \leq \alpha \leq 1$$

$$P_n \leq P_{max}, \forall n \in \mathcal{N}$$

$$P^s_{D_n} \leq P_{max}, \forall n \in \mathcal{N}_{2\hbar}$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \cdots, \beta_N, \beta_1^s, \cdots, \beta_K^s]$. Due to coupling of the variables and nonconvexity of constraints the global optimum solution is not straightforward. Please refer to [54] for the detailed solution.

4.4.1.6   Performance evaluation

We evaluate our proposed solution through Monte Carlo simulations within a $3 \times 3m^2$ subnetwork, catering to 20 SNEs, resembling a production module in a factory setting for robot control operations. The deployment of SNEs and HC/LCs is uniformly spread across the subnetwork area. Our model entails each SNE transmitting a 64-256 bytes packet (aligned with the assumptions for industrial traffic presented in deliverable D2.2) to the HC within a total time frame of 0.1 ms, utilizing a bandwidth of 100 MHz. Additionally, the power spectral density of the additive white Gaussian noise is set at -174 dBm/Hz. In our simulations, we adopt the channel settings based on the results outlined in WP2 D2.3 and the 3GPP channel model [89]. Most results in D2.3 follow the 3GPP model for in-factory environments, except for the K-factor, which rarely exceeds 10 dB. Building on this model, we apply Rician fading for all SNE-to-HC/LC and LC-to-HC links, reflecting the high likelihood of LoS conditions in small subnetworks. The path loss model is derived from the 3GPP channel model for indoor factory scenarios [89] aligning with the results of D2.3. Additionally, we incorporate a shadow fading model with a standard deviation of 7 dB.

In the following, we evaluate the performance of the relay and RIS scenario, respectively and compare the results with the cooperative communication method.

Relay performance:
We explore cooperative schemes employing different numbers of selectable LCs, including '1h' and '1 of K' in P-CSI case. In Figure 102, we present a comparative analysis of the empirical CDF showcasing the transmission power per SNE and LCs across various schemes for TDMA with the DF method, assuming a maximum power P$_{max}$ of 10 dBm and $B_n = 64$.
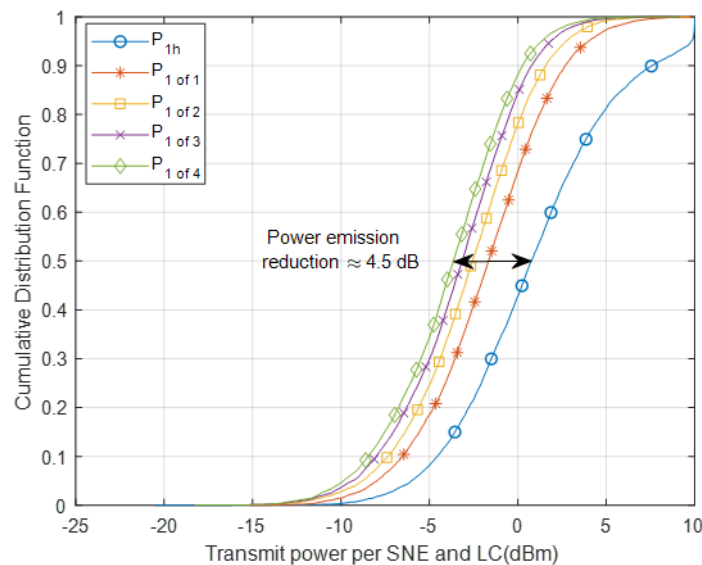
FIGURE 102. COMPARISON OF THE **CDF** OF TRANSMIT POWER ACROSS VARYING NUMBERS OF **LC**s FOR **TDMA** WITH THE **DF** METHOD, GIVEN $B_n = 64$ AND $P_{max} = 10$ DBM.

Our findings indicate a reduction in transmit power of up to 4.5 dB for the two-hop schemes compared to the single-hop scheme. This reduction is primarily due to the ability to select relays experiencing more favorable propagation conditions compared to the direct link with the HC.

The results clearly illustrate that increasing the number of LCs leads to greater power savings. However, it is noteworthy that the disparity in this power gain diminishes as the value of K grows.

To evaluate the performance of Algorithm 1 (please refer to [83]), we analyse the CDF of transmit power as depicted in Figure 103, across multiple scenarios:
1. Cooperative transmission employed by all SNEs.
2. All SNEs directly transmitting to the HC.
3. SNEs classified for single-hop or two-hop operations based on Algorithm 1.
4. Random selection of SNEs categorized as single-hop or two-hop.

Throughout this simulation only one LC ('1 of 1') is available. It is apparent from the results that all classification methods necessitate higher power transmission compared to Algorithm 1. This substantiates the effectiveness of Algorithm 1 in significantly reducing power consumption.
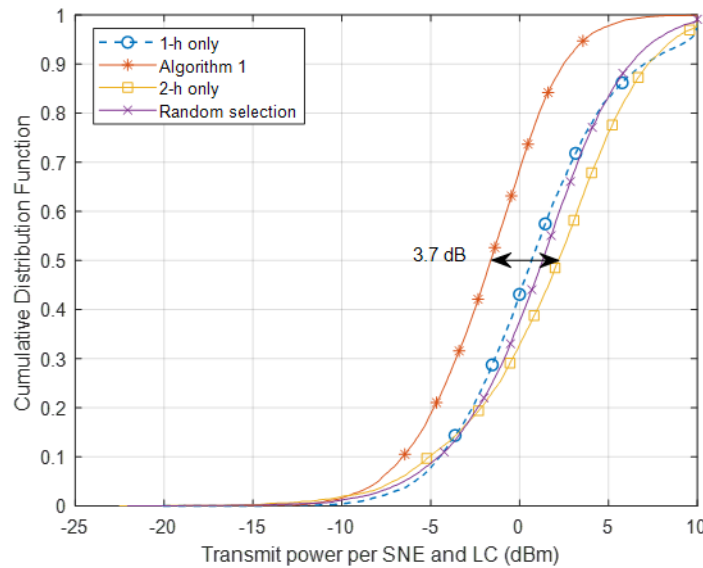
FIGURE 103. CDF OF TRANSMIT POWER FOR VARIOUS SNE CLASSIFICATION METHODS.

Figure 104 compares the AF and DF relaying methods for TDMA and FDMA transmissions, focusing on the case of $B_n = 64$ bytes. The results demonstrate that FDMA is more power-efficient than TDMA because reducing the bandwidth lowers the noise power for each device, leading to a decrease in the required transmit power. Additionally, the comparison between AF and DF relaying reveals the superior performance of DF, owing to its enhanced signal decoding capability. Specifically, for $B_n = 64$, the DF method achieves power savings of 1.8 dB and 0.8 dB for TDMA and FDMA, respectively, compared to AF relaying.

The total transmission time must not exceed the predefined time T, i.e., $T^{DF} \leq T$. Otherwise, a time overflow occurs. We utilize the $overflow\ rate$ to quantify the occurrence of events where $T^{DF} > T$. In Figure 106 the time overflow rate is depicted across different power levels for various schemes. These results, derived using maximum power, remain unaffected by the actual power optimization process.

This is because the power optimization method adjusts power to fulfil time constraints, defaulting to maximum power in worst-case scenarios. However, if this level proves inadequate, overflow instances occur.

In the I-CSI scenario, we set $L = 4$, allocating less than 1% of time resources to training, which is considered negligible for total power emission. To minimize channel estimation error, maximum power is used during training, followed by power control for the remaining time slots. Due to the minimal training allocation, its impact on average power transmission is negligible.
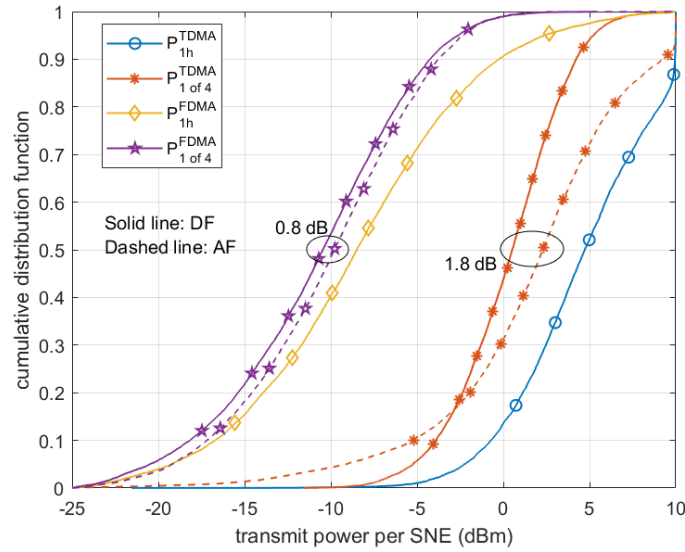
FIGURE 104. COMPARISON OF TRANSMIT POWER CDFS FOR DF AND AF METHODS USING TDMA OR FDMA TRANSMISSION WITH $B_n = 64$ BYTES.
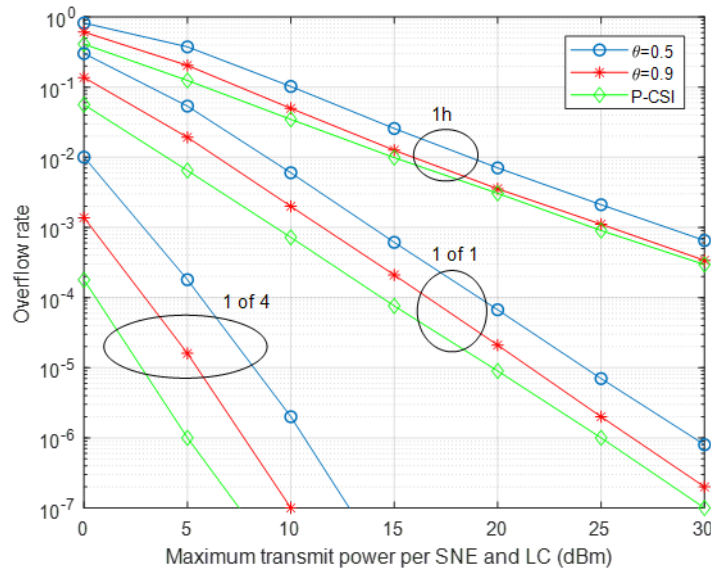
FIGURE 105. OVERFLOW RATE AGAINST MAXIMUM TRANSMIT POWER OF DIFFERENT SCHEMES FOR P-CSI CASE AND I-CSI CASE WITH L = 4.

The results compare $\theta = 0.5$ and $\theta = 0.9$ in the I-CSI case with the ideal P-CSI case. As θ increases, the overflow rate decreases, approaching the P-CSI scenario. For $P_{\max} = 30$ dBm, all two-hop transmissions

satisfy the minimum overflow rate of 10^{-6} even with θ = 0.5, while single-hop transmission exceeds 2×10^{-4}. Achieving the 10^{-6} requirement with θ = 0.5 requires $P_{max} \approx 11$ dBm and a '1 of 4' scheme. Figure 106 shows the impact of the discount factor on error probability for θ values from 0.5 to 0.9. As θ approaches 1, the overflow rate decreases because SNE packets are allocated fewer resources, increasing error probability. Conversely, lower θ\thetaθ values require more power to meet low-latency demands.

To maintain an acceptable overflow rate, P_{max} is set to 25 dBm in this analysis (Figure 106). The '1 of 3' and '1 of 4' schemes meet the $P_e < 10^{-6}$ constraint for θ < 0.7 and $\theta <$ 0.8, respectively, while other schemes fail. Specifically, for a '1 of 2' configuration, the discount rate must be below 0.5. This indicates that spectral efficiency can improve by 60% with the addition of two relays. The '1 of 1' and '1 of 4' configurations achieve approximately ~x2 and ~x4 **spectral efficiency** compared to single-hop transmission under the same reliability conditions, demonstrating the effectiveness of our method. This limited discount factor remains feasible due to favourable propagation conditions within the short-range subnetwork.
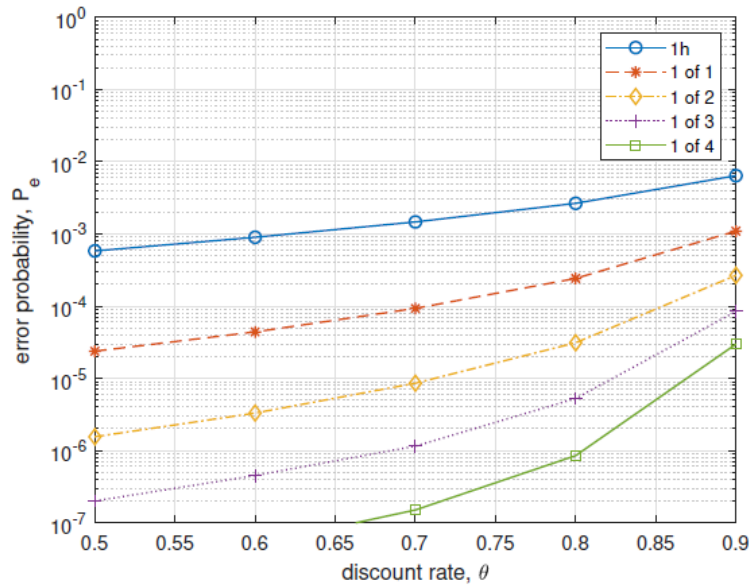
FIGURE 106. P_E AGAINST Θ FOR I-CSI CASE WITH PMAX = 25 DBM AND L = 4.

RIS Performance

Figure 107 shows the results for RISs with 16 elements. To evaluate the effectiveness of our phase-setting approach, we compare the results with those obtained using random phase settings, which lead to significant performance degradation.

We also compare the single RIS (1-RIS) deployment with the four RISs (4-RIS) scheme. Using four RISs increases the degrees of freedom, achieving power savings of up to 3 dB. Additionally, we examine the impact of the discount factor θ in the I-CSI case, presenting results for $\theta = 0.5$ and $\theta = 0.9$. With $\theta = 0.5$, the data rate is halved, requiring more power to compensate for the loss.
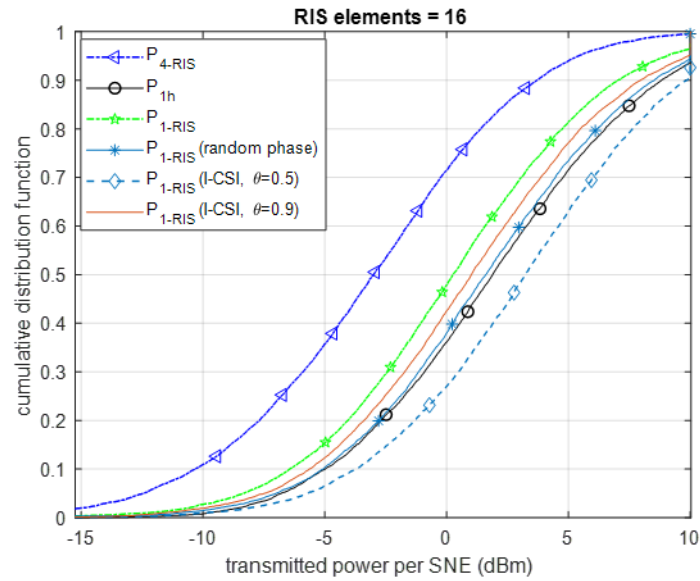
FIGURE 107. CDF OF TRANSMIT POWER FOR THE RIS SCENARIO WITH DIFFERENT CONFIGURATIONS USING 16 ELEMENTS.

Figure 108 illustrates the effect of the number of RIS elements on transmit power. Deploying four RISs with 64 elements significantly reduces transmit power, achieving nearly a 10 dB reduction compared to single-hop transmission without RIS.
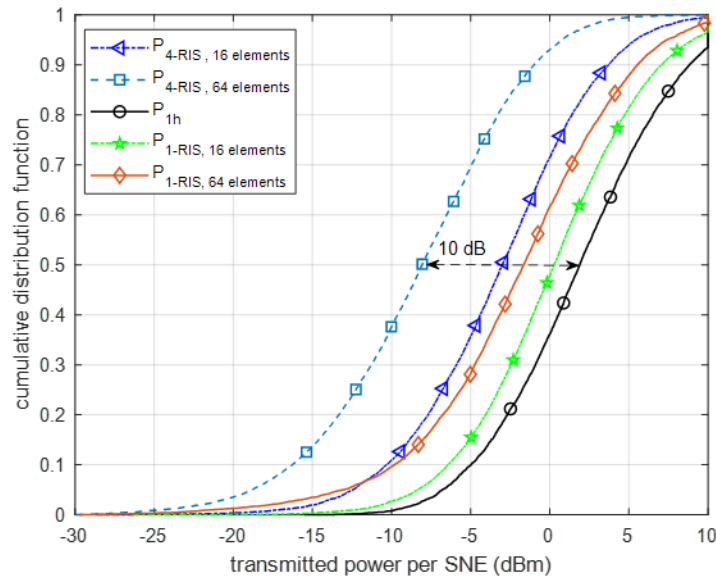
FIGURE 108. CDF COMPARISON OF TRANSMIT POWER FOR MULTIPLE RIS CONFIGURATIONS WITH 16 AND 64 ELEMENTS.

Figure 109 compares the performance of the RIS scenario with relaying in a TDMA setup, as RIS is compatible only with this configuration. The CDF curve shows that the '1 of 1' relaying scheme performs similarly to a RIS with 64 elements. However, using 4 RISs with 64 elements outperforms the '1 of 4' relaying scheme, reducing transmit power by up to 3 dB. This highlights the potential of RIS to achieve lower transmit power than relays.

It is important to note that this analysis focuses solely on minimizing transmit power, excluding the power required to operate the RIS. To accurately compare total power consumption, operational power must be considered, which will be explored in future work.
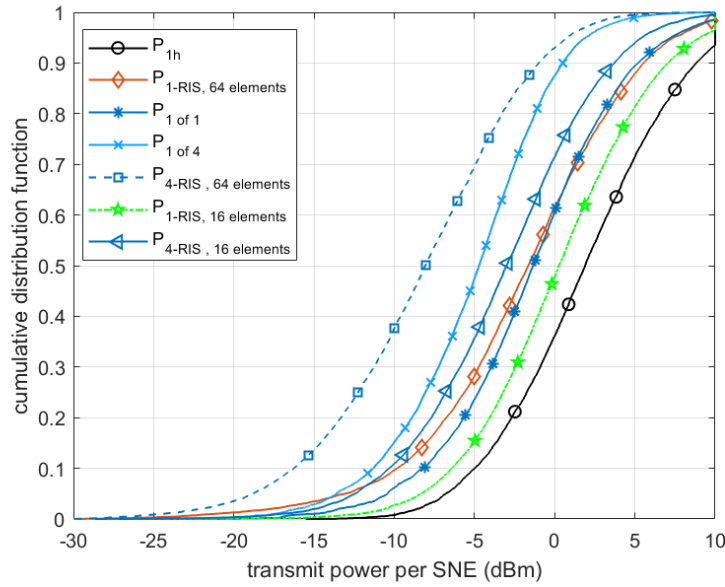
FIGURE 109. CDF COMPARISON OF TRANSMIT POWER FOR RIS VERSUS RELAY.

Overflow rate and P_{out} are depicted in Figure 110 and Figure 111 as functions of θ, respectively. Obviously, relays outperform RIS in terms of reliability. Specifically, using two relays results in both P_{out} and overflow rates dropping below 10^{-6}, whereas even with 4 RISs with 16 elements, the overflow rates do not reach 10^{-6}. This is mainly because RIS is passive and cannot provide the same level of reliability as an active relay.
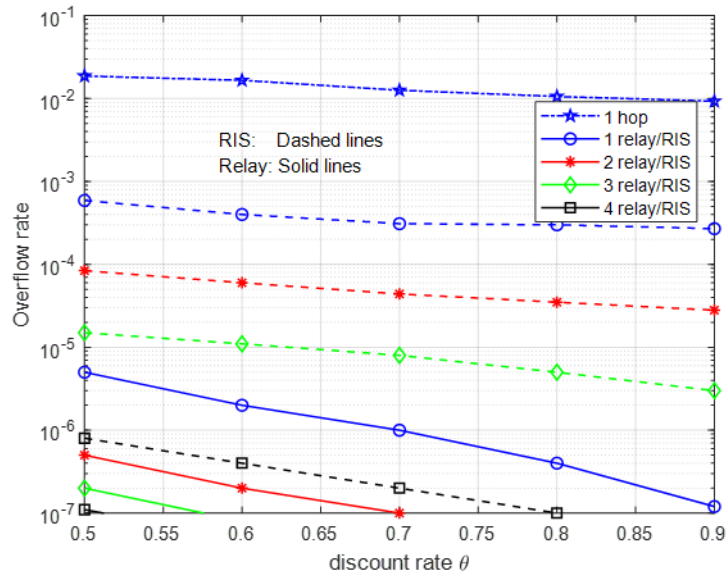
FIGURE 110. OVERFLOW RATE AGAINST θ FOR I-CSI CASE WITH $P_{max} = 23$M, $J_k = 16$ AND $L = 17$.
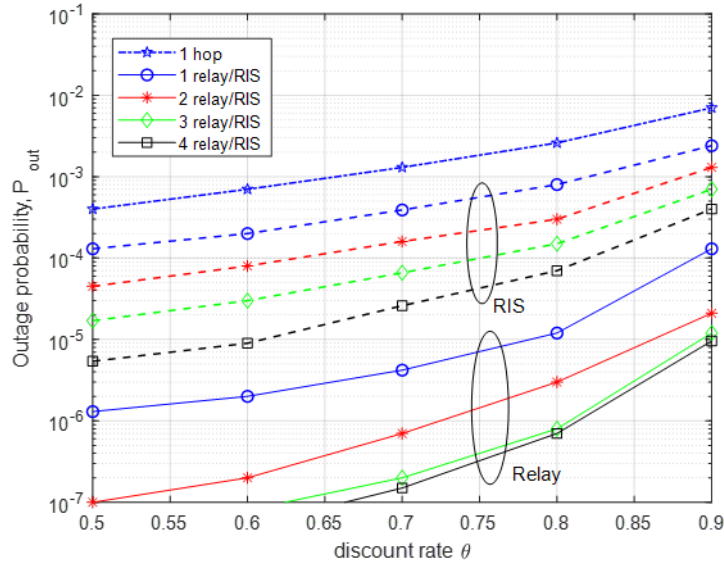
FIGURE 111. P_{OUT} AGAINST θ FOR I-CSI CASE WITH $P_{max} = 23$ DBM, $J_k = 16$ AND $L = 17$.

Increasing the number of RIS elements and training pilots improves performance. Figure 110 and Figure 111 show that configurations with 3 or 4 RISs achieve an outage rate below 10^{-6}, meeting the reliability constraint. In contrast, a single relay meets this requirement with $\theta < 0.7$ using only 65 pilots.

In conclusion, deploying RISs effectively reduces transmit power compared to relay-based systems. Four RISs with 64 elements achieve up to 3 dB more power savings than a comparable relay setup while satisfying URLLC requirements.

However, reducing the number of RISs or their elements may compromise reliability, which is crucial for IIoT subnetworks. For example, four RISs with 16 elements cannot match the reliability of a single relay in terms of outage probability.

4.4.1.7   Summary and recommendations

This subsection addressed the challenge of enabling communication with strict cycle times for numerous SNEs within an IIoT subnetwork. We proposed using multiple LCs with relaying capabilities and introduced new TDMA and FDMA protocols supporting both single-hop and two-hop communication. We also compared relay-aided and RIS-assisted scenarios to evaluate their strengths and limitations, focusing on minimizing power consumption while meeting timing constraints.

Simulation results demonstrate that RIS-assisted approaches offer significant power savings, while relay-based methods provide better reliability, lower outage probability, and less overhead. Based on the results, we recommend selecting the configuration according to the application's priority on power efficiency, reliability, or training overhead. For example, to meet URLLC requirements, using multiple LCs as proposed is the most suitable option.

4.4.2   Network Coded Cooperation

In the previous section, novel solutions for supporting fast and reliable control cycles for industrial subnetworks leveraging LCs acting as relays, or RISs, have been presented. Here, we present further methods for exploiting spatial diversity via the PHY multi-links and via subnetwork devices in range by using cooperative mechanisms based on network coded cooperation (NCC).

Such macro-diversity solutions can be used in industrial subnetworks for achieving robustness to blockage effects that may hinder the efficient support of critical services. It is a very likely scenario that installations in production cells, machines or robots may lead to situations where the desired link between a base station (BS) which is a HC device and multiple SNEs such as sensors, actuators or robots is obstructed by obstacles such as metallic items.

Deliverable 3.1 presented the preliminary considerations for Network Coded Cooperation (NCC) in a subnetwork and for modeling in a network simulator. In the further course of the project, the communication parameters were specified for the industrial use case in accordance with Deliverable 2.2, various network topologies were modeled, and various error control mechanisms were simulated on this basis. The simulation runs generated KPI figures on the required radio resources and statistical results on packet loss rates. These generated KPIs were used to evaluate NCC. Various cooperation and coding strategies were developed for NCC and compared with a simple transmission or a method with packet retransmission (so-called blind retransmission).

Network coding (NC) addresses, in communication systems, information that is to be transmitted between two end points, and that is subject to transformation through several processing functions above a channel coding on the physical layer that take input packets and produce output packets wherein each of the output data packets has distinct importance, significance or meaning in terms of contribution to the recovery of the information to be transmitted or the quality of user experience, wherein not all the output packets are required at the receiver for a successful recovery of the input packets or targeted quality of experience [97].

We use NC as a packet level processing function, or more specifically a packet coding function that transforms input packets, also denoted as source packets or Service Data Units (SDU) or NC SDUs, into output packets, also denoted as coded packets or Protocol Data Units (PDU) or NC PDUs. The input packets being coded together form a NC generation, also denoted as a generation. An NC PDU output packet of the network coding processing is a linear combination on a symbol level of one or more of the NC SDU, i.e. a first NC PDU is generated based on a linear combination of one or more of the NC SDUs using a first set of network coding coefficients, and a second NC PDU is generated based on a second linear combination of one or more of the NC SDUs using a second set of network coding coefficients that is independent from the first set of network coding coefficients, i.e., the first NC PDU and the second NC PDU are linearly independent.

In order to successfully recover n NC SDUs at the receiver, the receiver needs to correctly decode at least n linearly independent NC PDUs out of the transmitted NC PDUs. Therefore, even if the receiver fails to correctly decode some of the received NC PDUs, the receiver can still recover the NC SDUs. As mentioned earlier, network coding is used to describe packet coding techniques wherein coding may be

    a)   end-to-end coding where coding is performed at the source (i.e. originating node of the packet), decoding is performed at the destination. There is no recording or recoding at any intermediary

node. By producing more linearly independent NC PDUs than the number of NC SDUs of a generation results in a redundancy of packets, i.e. some of the NC PDU transmissions can fail as long as enough NC PDUs are received to fully decode the NC SDUs. Also, NC PDUs can be distributed over multiple paths.

b) coding is performed at the source and/or intermediate forwarding node, and decoding is performed at the destination. Forwarding of coded / recoded NC PDUs allows cooperation of subnetwork nodes and is denoted as **Network Coded Cooperation (NCC)**.

Deliverable 3.1 presented preliminary considerations for NCC in a subnetwork and for modeling such an approach in a network simulator. In the further course of the project, the communication parameters were specified for the industrial use case in accordance with Deliverable 2.2. Various network topologies were modeled, and various error control mechanisms were simulated on this basis. The simulation runs generated KPIs for evaluating the error control procedures in terms of resource requirements and reliability.

Various cooperation and coding strategies were developed for NCC and compared to simple transmission methods like single transmission or static packet retransmission (so-called blind retransmission). In the following, the requirements for communication are described first based on the underlying scenarios. Next, simulation model and the KPIs for a single or repeated packet transmission are presented as a baseline. Various static and opportunistic NCC procedures are then described and the simulated KPIs are presented. Finally, a recommendation for action is given based on the results.

4.4.2.1 Scenario and Requirements

The focus is on the industrial use case, with the specific scenarios of robot control, unit test and - to a limited extent due to the limited data rate assumed in the simulation - visual inspection. The considerations can be transferred to other use cases with similar requirements and assumptions, such as automotive.

The requirements for this use case from D2.2 were implemented as follows:

TABLE 15 COMPARISON OF THE REQUIREMENTS WITH THE IMPLEMENTATION IN 6G-SHINE

Requirements from D2.2	Configuration of the communication parameters used in simulation
Cycle time < 100 µs - 1 ms for communication in loops other communication cycles < 10 - 100 ms	Cycle time is 250 µs, which corresponds to a TDMA frame duration of 250 µs, comprising 26 time slots. We assume symmetric bidirectional communication between HC and SNEs. Data packets should be delivered within a single frame (deterministic cycle time!) **This makes the approach in 6G-SHINE faster by a factor of approx.. 10 compared to 5G URLLC!**
Data packet size 100..300 Byte	28 Byte of higher layer data The data packet size could be extended by using fragmentation or higher modulation schemes, but it was considered as being sufficient to evaluate the error control mechanisms.
Data rate 1..2 Mbit/s (in case of video up to 80 Mbps)	With the selected data packet size the resulting data rate for higher layer data is 896 kbit/s.

20..50 SNE	Predefined and stationary set of up to 6 SNE The number of 6 SNE is sufficient to evaluate NCC strategies. More SNE can be supported depending on the available resources.
Movement speed of nodes none or up to 20 m/s	As we simulate on a packet level, movement speed is not relevant. Changing transmission conditions based on topology and distances were simulated by a periodic permutation of all SNEs. In particular, this allows all possible cooperation relationships between the SNEs to be investigated.
No requirement defined	Multilink operation: two parallel links between devices (operation on second channel is synchronous to first channel (same slot and frame alignment)) This operation mode allows shorter transmission phases und thus better NCC strategies.
No requirement defined	Star topology: no connection on application level between SNEs, only between HC and SNEs
probability of having two consecutive errors < 10^{-6}	This given requirement seems to be too weak for fast industrial control applications, as 2 consecutive errors will lead to an error condition. With a given cycle time of 250 µs a probability of 10^{-6} would statistically lead to an error condition every 250 seconds! We recommend to consider as probability of having two consecutive errors at least < 10^{-9}, which means $3,2*10^{-5}$ for a single packet loss in case of uncorrelated packet errors. **The packet loss result is a KPI will be the outcome of the simulations.**

The industrial subnetwork consists of an HC and 6 randomly arranged SNEs. All SNE are within the coverage area of the HC, but the different distances between HC and SNE lead to different link properties between HC and SNE. In addition to the HC-SNE link, the SNE_x-SNE_y link properties are particularly important for NCC, as cooperative communication is based on the forwarding of packets via other SNEs. To simulate difficult transmission conditions, such as those encountered in a robot cell, 2 obstacles can be added to the environment, which can hinder the line of sight between HC and SNE or between SNE_x and SNE_y (Figure 112. Subnetwork topology with 2 obstacles). For a complete evaluation and the generation of statistical KPIs, all possible cooperation relationships between the SNEs were simulated by a periodic permutation of all SNE positions.

FIGURE 112. SUBNETWORK TOPOLOGY WITH 2 OBSTACLES
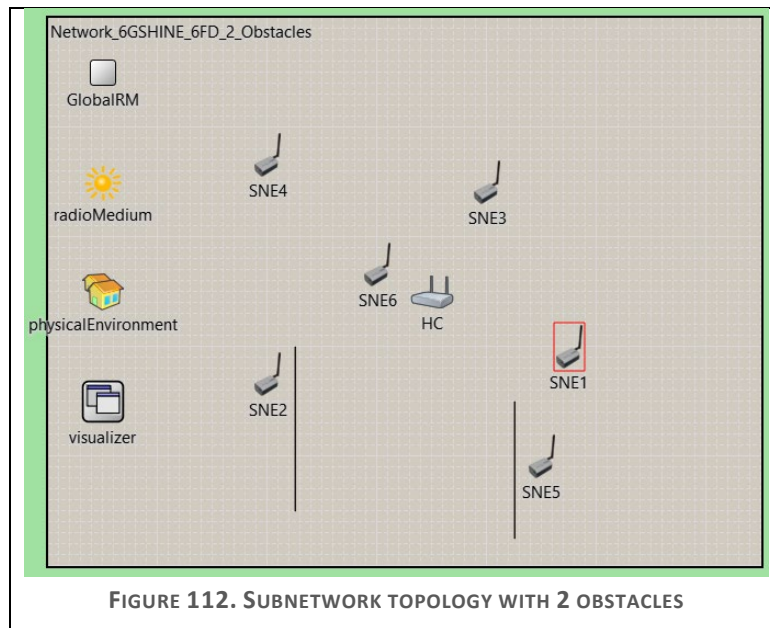
Based on this network constellation we examined how spatial diversity via the PHY multi-links and via other subnetwork devices in range by using network coded cooperative mechanisms improve reliability. The central questions were:

- Which nodes can and should cooperate depending on distance / link quality?
- What influence do obstacles have on the transmission quality in the network?
- How can cooperation strategies be adapted during operation, especially within the cycle time?

It is assumed that each cycle one native packet is transmitted from the HC to an SNE and one native packet is transmitted from an SNE to the HC (Time Division Duplex TDD). A packet is transmitted within a certain timeslot on a certain frequency channel, denoted as Resource Block (RB).
The following resource allocation serves as a baseline:

FIGURE 113. RESOURCE ALLOCATION EXAMPLE FOR 1-RETRANSMISSION
(GREEN = TX ACTIVITY; GRAY = RX ACTIVITY)
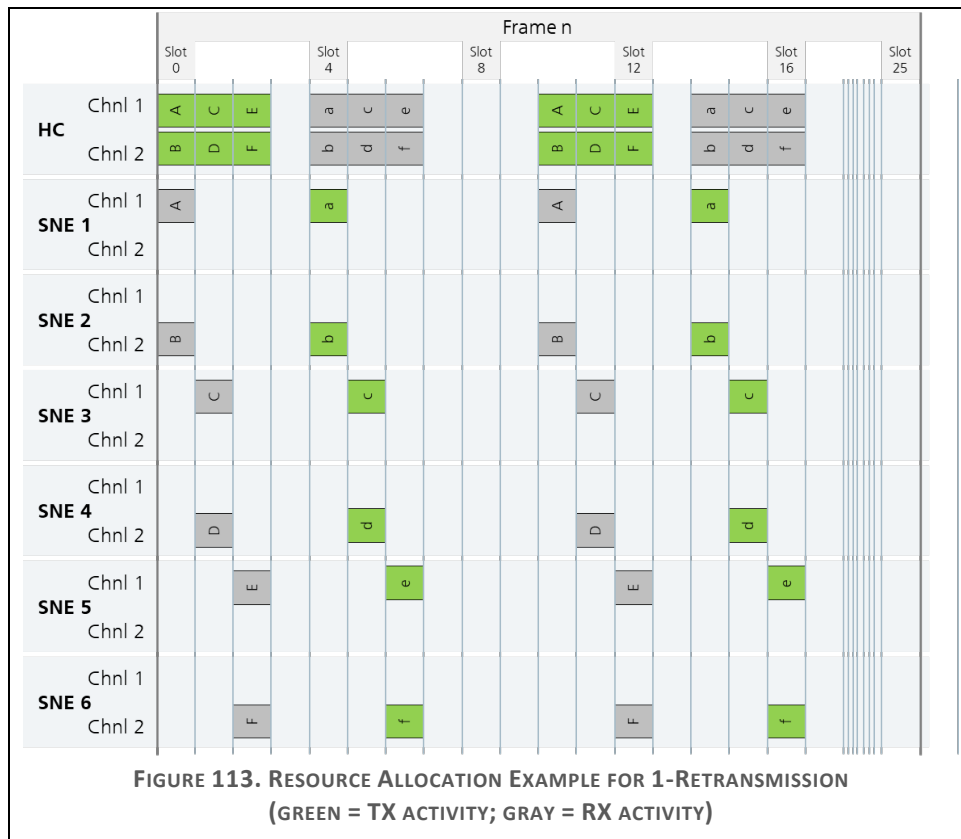
In a downlink phase, for example, a packet "A" is transmitted from the HC to the SNE_1 (shown here in slot 0), which sends back a packet "a" in an uplink phase (slot 4). In the case of a single transmission, this completes the communication cycle. In the case of blind retransmission, the packets are sent repeatedly (slot 10 and slot 14). With given assumptions for transmission power, frequency band, modulation bandwidth and sensitivity we get the following baseline results (simulation run 200000 cycles (= 50 s) with position permutations every 1 ms):
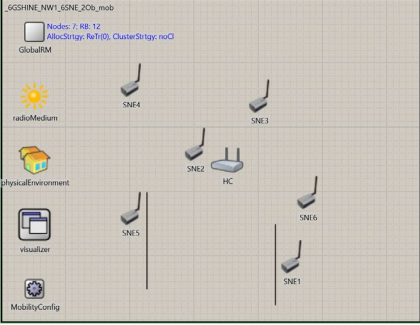
TABLE 16 KPI AND SIMULATION RESULTS FOR PACKET LOSS RATE FOR n-RETRANSMISSION AND NO OBSTACLES

0 Obstacle	Single Transmission	Blind Retransmission
KPI		
Total number of used resource blocks per frame	12	24
Number of TX slot activities at SNE	1	2
Number of RX slot activities at SNE	1	2
Paket Loss Rates downlink (to SNE)	0,053250%	0,000000%
Paket Loss Rates uplink (from SNE)	0,056667%	0,000083%

The assumptions made show that a simple transmission would not be sufficiently statistically reliable overall. But already a simple blind retransmission would suffice to produce a sufficiently reliable transmission under line-of-sight conditions.

In practice, problems do not or rarely occur under such ideal line-of-sight conditions but occur in situations where direct communication is disrupted by obstacles. The following values result in case of 2 obstacles:

TABLE 17 KPI AND SIMULATION RESULTS FOR PACKET LOSS RATE FOR N-RETRANSMISSION AND 2 OBSTACLES

2 Obstacles	Single Transmission	Blind Retransmission
KPI		
Total number of used resource blocks per frame	12	24
Number of TX slot activities at SNE	1	2
Number of RX slot activities at SNE	1	2
Paket Loss Rates downlink to SNE	33,374333%	33,333333%
Paket Loss Rates uplink from SNE	33,375917%	33,333333%

In the event of obstacles, direct transmission between the HC and the SNE behind an obstacle fails.

4.4.2.2   Network Coded Cooperation Strategies

Cooperative communication is based on the ability of SNEs to forward packets on behalf of other SNEs or HC. This results in additional propagation paths, which increase the diversity order of the communication. This means that blocked direct links can be compensated.

NCC extends cooperative communications with the option of forwarding several native packets together as a single coded packet by adding them at symbol level.

The following scheduling was used to process an NCC strategy within a communication cycle:
- Downlink phase, in which the HC sends native packets to all SNEs once (slot 0 to slot 2 in Figure 114)
- Uplink phase, in which each SNE sends native packets to the HC (slot 5 to slot 7 in Figure 114)
- NCC phase, in which each SNE forwards packets for other SNEs, whereby downlink and uplink packets are encoded together as a coded packet. The composition of the coded packet is described and signaled via a coding vector. (Slot 11 to slot 13 in Figure 114)
- Optional further NCC phases with other coding vectors.

FIGURE 114. RESOURCE ALLOCATION EXAMPLE FOR 1 NCC PHASE (GREEN = TX ACTIVITY; GRAY = RX ACTIVITY)
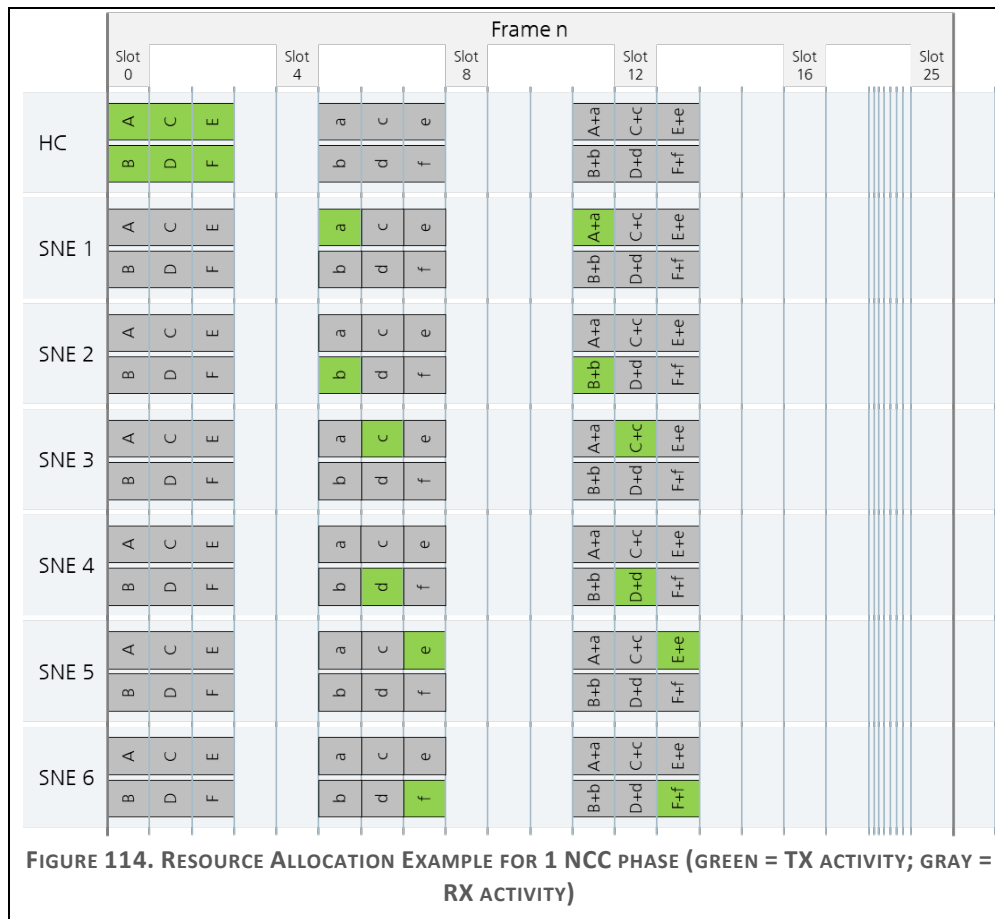
Coded packets have the same payload length as the native packets from which they are encoded. This means that the transmission of the coded packet requires the same slot size at the same data rate as the transmission of the native packet. This leads to a significant increase in efficiency compared to cooperative communication, where all native packets have to be repeated individually.

Scheduling and coding strategies for NCC:

Strategy	Description
NCC 1 Everyone helps everyone (max. coding degree = 2)	All SNEs receive all packets in the downlink phase of the HC (slots 0-2) and in the uplink phase (slots 5-7) of the other SNEs by so called overhearing. In a first NCC phase, SNEs each send the downlink and uplink packets of the SNE with the next highest SNE ID in coded form. In a second NCC phase, the packets of the SNE with the next but one SNE ID are forwarded, and so on. In this way, each SNE forwards the packets for all other SNEs. Each SNE generates a coded packet of grade 2 from the native downlink and uplink packets. If only one native packet is available for coding, this is forwarded natively. If both native packets are missing, no packet is sent. This procedure has the greatest potential for mutual support, but requires the most resource blocks of all NCC procedures. With 6 SNEs, 6 downlink RBs + 6 uplink RBs + 5*6 NCC RBs = 42 RBs would be required! Taking processing latencies into account,
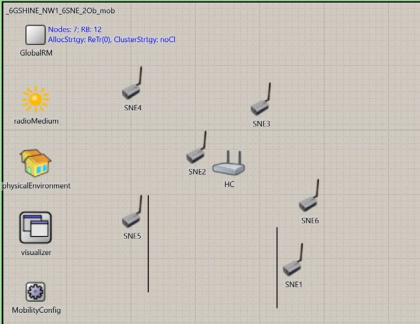
	only a maximum of 3 NCC phases can be mapped in the selected frame length of 26 slots, which results in a number of 30 RBs. In addition, the overhearing effort of a SNE is very high, i.e. a SNE requires many RX slot activities.
NCC 2 2 Cluster of 3 cooperating SNE (1-2-3, 4-5-6)	The number of RBs and RX slot activities required can be drastically reduced by forming clusters. In this case, 3 SNEs form a cluster. Within the cluster, each SNE supports the other two SNEs in 2 NCC phases. With 6 SNEs and 2 clusters, 24 RBs are required, which corresponds to the number of blind retransmission.

TABLE 18 KPI AND SIMULATION RESULTS FOR PACKET LOSS RATE FOR STATIC NCC AND NO OBSTACLES

0 Obstacle	Blind Retransmission	NCC 1 everyone helps everyone limited to 3 NCC phases because of the limited number of slots per frame	NCC 2 2 clusters (1-2-3), (4-5-6) 2 NCC phases
KPI			
Total number of used resource blocks per frame	24	30	24
Number of TX slot activities at SNE	2	4	3
Number of RX slot activities at SNE	2	26	9
Paket Loss Rates downlink to SNE	0,000000%	0,000000%	0,003583%
Paket Loss Rates uplink from SNE	0,000083%	0,001167%	0,003500%

Under line-of-sight conditions, NCC methods proved to be suitable, although for smaller cluster sizes there are topology permutations in which the SNEs of the cluster have an unfavorable constellation, making cooperation more difficult. In the next step, more difficult conditions were investigated in which 2 obstacles block the propagation to some SNE.

TABLE 19 KPI AND SIMULATION RESULTS FOR PACKET LOSS RATE FOR STATIC NCC AND 2 OBSTACLES

2 Obstacles	Blind Retransmission	NCC 1 everyone helps everyone limited to 3 NCC phases	NCC 2 2 clusters (1-2-3), (4-5-6) 2 NCC phases
KPI			

Total number of used resource blocks per frame	24	30	24
Number of TX slot activities at SNE	2	4	3
Number of RX slot activities at SNE	2	26	9
Paket Loss Rates downlink (to SNE)	33,333333%	4,049417%	10,735833%
Paket Loss Rates uplink (from SNE)	33,333333%	6,260917%	10,744500%

Even if the average packet loss rates of all permutations do not reach the desired 10^{-5}, the results clearly show the potential of NCC to increase the reliability of the network. This is especially true in the case of well-chosen cooperation clusters. It is evident that even connections blocked by obstacles can be fully maintained due to the additional path diversity of NCC.

Each NCC phase increases the number of RBs as well as the number of TX and RX slot activities. Clustering can reduce the number of RBs and the number of TX and RX slot activities.

However, smaller clusters mean an increase in packet loss rate due to the smaller number of cooperation partners. If the cluster is chosen incorrectly and there are obstacles between the SNEs in the cluster, NCC can also fail.
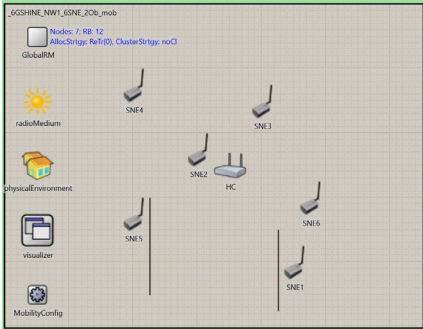
4.4.2.3 Proposed Opportunistic Network Coded Cooperation

So far, methods have been investigated in which the cooperating SNEs and thus the coding vectors of the NCC phases were statically predefined. NCC can become even more efficient if feedback about the downlink reception is provided with the uplink packet. For example, the uplink packet can be encoded with the downlink packet.

An SNE can analyze the coded packets received in the uplink phase via overhearing and deduce which SNE in the same cluster needs the most support. In order to support the SNE determined in the analysis, the coding vector for the NCC phase is dynamically adapted. This decision is made decentrally in the SNE and does not require any interaction with the HC. The basic scheduling in the frame also remains the same. All cycle time requirements are met, as the dynamic, flexible support takes place in the same frame.

Strategy	Description
NCC 3 No Cluster with dynamic coding vectors	There is only one NCC phase defined, in which a SNE determines autonomously which other SNE needs support. With 6 SNEs and no clusters, only 18 RBs are required, which is far less compared to the number of blind retransmissions.
NCC 4 2 Cluster of 3 cooperating SNE (1-2-3, 4-5-6) with dynamic coding vectors	Again 3 SNEs form a cluster. There is only one NCC phase defined, in which a SNE determines itself which other SNE within the same cluster needs help. With 6 SNEs and 2 clusters, also only 18 RBs are required, which is far less compared to the number of blind retransmission.
NCC 5 3 Cluster of 2 cooperating SNE (1-5), (3-4), (2-6) with dynamic coding vectors	With 2 cooperating SNEs, the smallest possible cluster size was also examined. This results in 3 clusters in the example. With 18 RBs, 2 TX activities and 4 RX activities, the minimum resource requirement is achieved.

2 Obstacles	Blind Retransmission	NCC 3 no cluster	NCC 4 2 cluster (1-2-3), (4-5-6)	NCC 5 3 cluster (1-5), (3-4), (2-6)
		Opportunistic Network Coded Cooperation	Opportunistic Network Coded Cooperation	Opportunistic Network Coded Cooperation
KPI				
Total number of used resource blocks per frame	24	18	18	18
Number of TX slot activities at SNE	2	2	2	2
Number of RX slot activities at SNE	2	16	7	4
Paket Loss Rates downlink to SNE	33,333333%	0,032167%	12,916250%	20,503417%
Paket Loss Rates uplink from SNE	33,333333%	0,000500%	12,912500%	20,508917%

Dynamic adaptation of the coding vector by analyzing and evaluating the actual topology reduces the packet loss rate by a factor of over 1000. While a network operating with retransmission has completely failed connections and an average packet loss rate of 33%, a network operating with NCC can maintain full operation.

At the same time, the number of RBs can be reduced by 25%! The freed-up RBs can either be used for further redundancy in the network (e.g. 2nd NCC phase) or for additional SNEs.

4.4.2.4   Summary and recommendations

Network Coded Cooperation shows excellent results in 6G subnetworks to increase the availability of SNEs and the reliability of the transmission. NCC can also be used in TDMA/TDD systems with extremely short cycle times.

The right choice of cooperating SNEs in a cooperation cluster is important. The larger the clusters are selected, the more redundancy is available in the network and thus the possibility of reducing the packet loss rate. On the other hand, larger clusters increase the resource requirements.

With opportunistic Network Coded Cooperation, a new method was investigated in which an evaluation of the connection quality is carried out in the SNE to identify which other SNE can / should be helped in the NCC phase.

The method presented

- works with cycle times of 250 μs and is therefore around 10 times faster than 5G-URLLC systems,
- reduces the packet loss rate by a factor of 1000 compared to systems with packet retransmission,
- requires 25% fewer resource blocks for transmission compared to systems with packet retransmission.

It is therefore strongly recommended to include opportunistic Network Coded Cooperation in future 6G subnetworks and to create the conditions for overhearing in cooperation clusters.

4.5   SOLUTIONS FOR IMPROVED LATENCY IN THE UNLICENSED SPECTRUM

Subnetworks are anticipated to operate across various spectrum bands, including those previously discussed (e.g., 28 GHz, sub-THz). Both licensed and unlicensed spectrum options are under consideration for these subnetworks. However, accessing unlicensed spectrum may be subject to strict regulations—such as listen-before-talk protocols—which can impact latency and thus hinder time-sensitive operations. This section focuses on enhancements aimed at minimizing latency for operations conducted within unlicensed spectrum bands, leveraging the initial findings presented in deliverable D3.1.

4.5.1   Latency Aware Access in unlicensed Bands

This work focuses on providing solutions for priority differentiation when accessing unlicensed bands (e.g., ISM bands), focusing on high priority and bounded latency traffic flows. It is building on top of the initial design of the LAD (Latency Aware Deconfliction) mac access scheme presented in D3.1 within the 6G-SHINE context. It will play a significant role in the case where subnetworks need to access ISM bands for additional spectrum resources but want to avoid the bottleneck of current LBT + backoff rules, which will surely impact severely the offered latency and overall QoS.

The shortcomings of the current ETSI standard were presented previously in D3.1, not allowing for any differentiation in the basic access mechanism of ISM bands, defining a flat minimum CCA time for each ISM band. We already showed in our previous work in D3.1 that the adoption of a two level CCA time (one for best access as it is today and a smaller CCA timing for priority access), if employed correctly, would provide significant performance enhancements for priority traffic with minimum impact on the best effort traffic even when multiple technologies are present in the same spectrum band/channel. But also shortcomings in the design were identified, with the main one being the inability to scale up on supporting a high number of priority flows with packet interarrival time similar or even lower than the periodicity of the LAD. More information on the pinpointed disadvantage and the proposed solution will be given in the next subchapter where the enhanced design of LAD will be presented in detail.

In this work, we identify the design shortcomings, defined design enhancements that can solve the pinpointed issues and evaluated the proposed solution using an updated NS3 simulation model of the LAD MAC function.

4.5.1.1   Description of enhanced design

The initial design of the LAD is shown for reference in Figure 115.
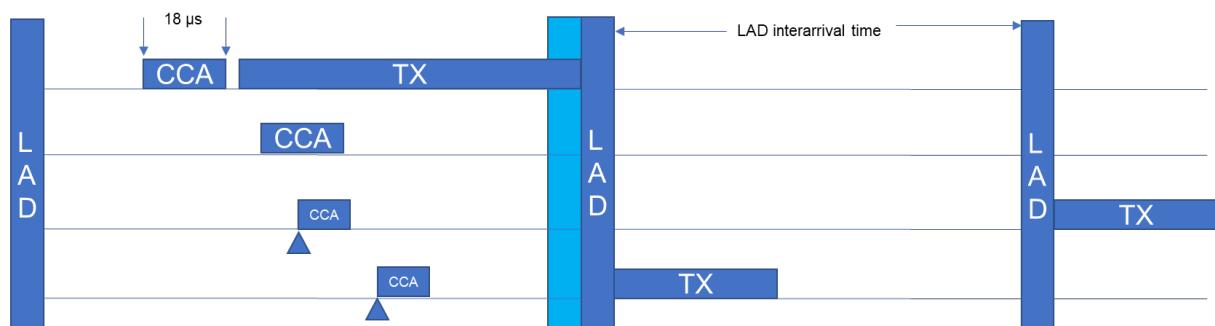
FIGURE 115. VERSION 1.0 OF LATENCY AWARE DECONFLICTION PERIOD STRUCTURE AND TIMINGS

Since LAD access timing is fixed based on a predetermined interarrival time, it was found to struggle to support traffic flows that generate multiple packets per LAD period. The main issue identified is that, if the total high priority packets generated per LAD period is more than 2, then the LAD capacity is saturated as there is no alternative way to forward the packets in time. As a result, packets are queued on the high priority queue, more nodes are trying to access each LAD period every time and finally more collisions are probable in the LAD period (as more packets are competing during the LAD , the higher the chance of a LAD collision) and packets could end up waiting multiple LAD periods before being served. This would have an impact on MAC access latency as well as packet success ratio as more packets could reach the maximum retransmissions setting of 2 and get discarded.

To solve the aforementioned bottleneck and make the LAD design able to cope with high priority traffic capacity issues in a scalable and abstract way without making the LAD interarrival time being decided in a dynamic and decentralized way (that would require a decision protocol to be put in place, making the design unnecessarily complex), we ended up deciding to allow multiple LAD packets being strobed next to each other once an LAD start time came. The proposed update to the LAD design is presented in Figure 116 and is named LAD V2.0.
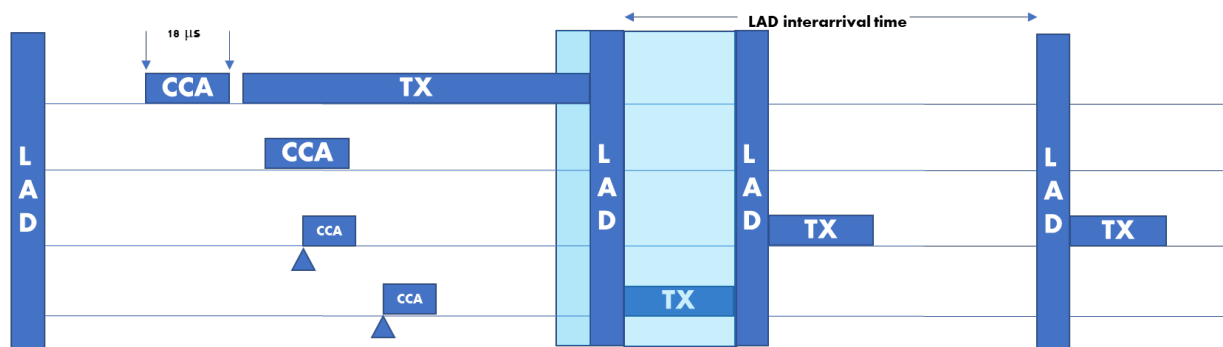
FIGURE 116. LAD V2.0 DESIGN

As it can be seen in the figure, now when 2 or more nodes have a packet to send when an LAD period starts, they are able to compete in the first LAD, one will win and transmit its packet and immediately after the transmission of the first packet, a new LAD gets initiated from the nodes still having high priority traffic to send. In this way a bottleneck is not created and in each LAD time all high priority traffic is transmitted. The impact of this strobing approach on the legacy best effort access scheme will be analysed and if possible a rule on how far can the LAD strobing go will be defined versus the size of the LAD interarrival time. The internals of the LAD function remain the same, taking the best performing settings identified in D3.1 as the default values for overall operation.

On a second axis, two different designs will be evaluated, one focused on a pure LAD operation (called simple LAD from now on) where high priority packets are being transmitted only in LAD periods, and one where LAD packets can use both the CSMA period using typical CSMA but then also use the LAD period for deterministic operation (called hybrid LAD or H-LAD). Comparison between the 2 modes will reveal possible advantages existing in each approach and when they are best applicable to offer the optimum performance latency wise.

4.5.1.2   Updated simulation model

The NS3 model of LAD that was implemented in our previous work in D3.1 was now updated in order to support 2 major changes. Such model is depicted in Figure 117. First of all, the choice of enabling an LAD period is no longer controlled only from LADMacSchedule module based on a starting time and the LAD interarrival time, but now strobing support is also required when more packets are present in the high priority queue. Hence, the LAD MAC packet queue module is also now feeding queue utilization information to the LADMacController in order to be able to trigger a new LAD period if a previous LAD just ended and there are still packets in the high priority queue.
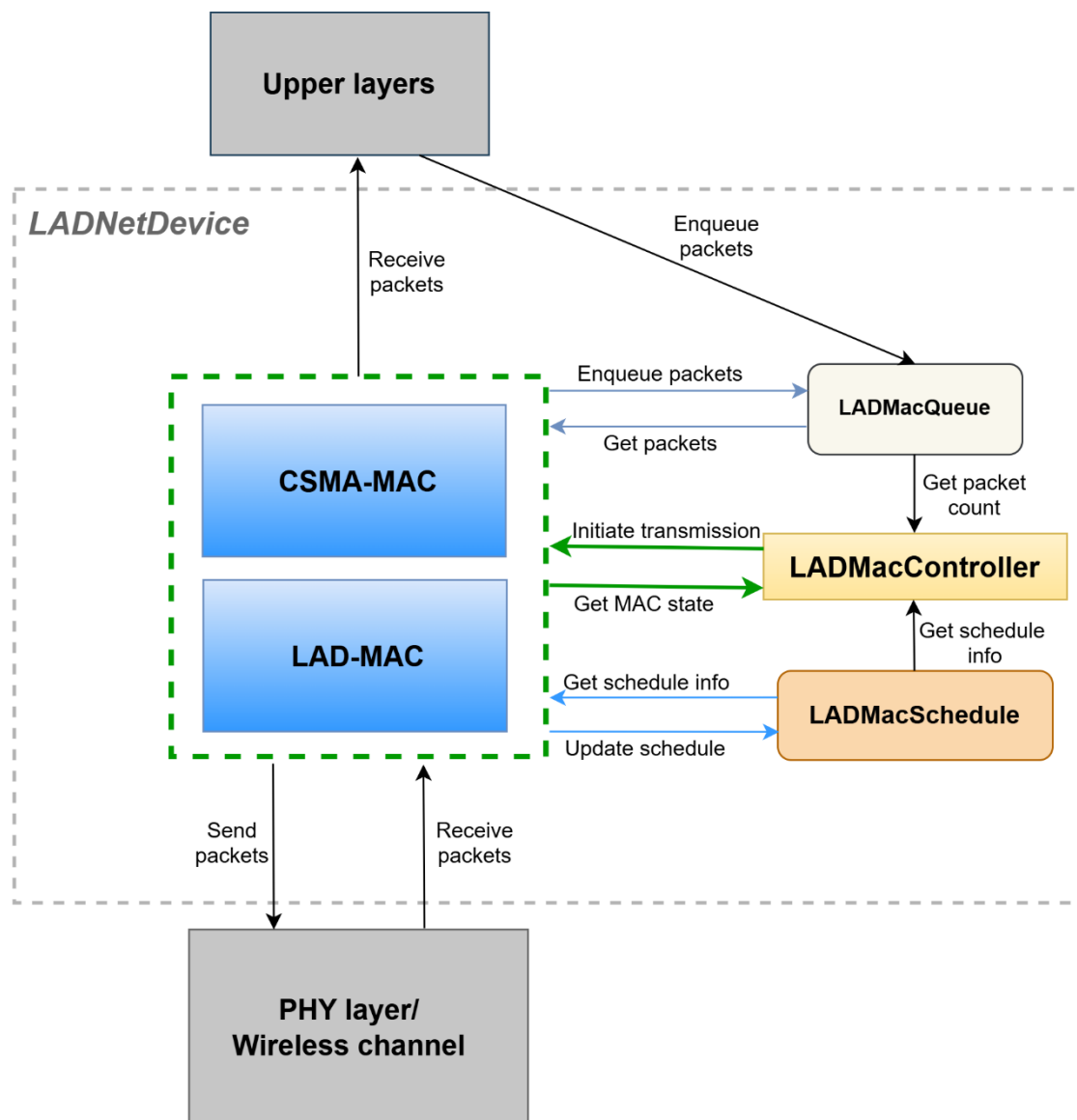
FIGURE 117. LAD V2.0 NS3 UPDATED MODEL

Furthermore, for the LAD to be able to be re-triggered once the previous is ended and all nodes to be still synchronized, a LAD binary variable on waiting for ACK was added in the LAD trigger logic. If there is ACK support for any heterogeneous technology node in the vicinity of the LAD nodes (WiFi in our case here), then LAD needs to be re-initiated after the ACK is sent or after ACK is timed out, hence the waiting time to declare the medium as empty is 12 μsec once the RSSI drops to noise floor. If there is no ACK

support, then the medium is declared as idle as soon as the RSSI drops to noise floor without any delay and LAD is triggered immediately.

Lastly, support for high priority packets to use the CSMA time is added and enabled in the relevant simulation scenarios to evaluate performance of H-LAD against the baseline but also against pure bursty LAD.

4.5.1.3   Simulation setup

Two major simulation series were executed with the updated model, one general performance evaluation series of LAD and on top of that an added scenario coming from WP2 , depicting the operation of a Robotic Arm control loop over the proposed LAD MAC access scheme versus the basic CSMA based access used today from most wireless technologies in the ISM bands.

The simulation series settings used to evaluate the performance of the update LAD scheme are depicted in the 3 tables below:

TABLE 21. SIMULATION GENERAL PARAMETERS

Simulation duration	50 s
Number of simulations per use case	5
Number of nodes (N)	4 to 32
Number of communication links	N/2 in total, equally divided
Nodes mobility	fixed positions in the same collision domain
PHY supported data rate	11 Mbps
Propagation delay	Simulated
Propagation loss	Simulated

TABLE 22. CSMA/CA MAC PARAMETERS

RTS/CTS	Disabled
ACKs	Enabled
Number of retransmissions	2
CWmin	15
CWmax	1023
Slot time	20 µs
SIFS	10 µs
DIFS	50 µs

TABLE 23. LA-MAC PARAMETERS

LAD period	1, 2, 3 ms
LAD number of steps (step_max)	12
Step duration	2 µs
Initial CCA duration	10 µs
CCA duration during step	rand(CCAmin, CCAmax)
CCAmin	3 µs
CCAmax	6 µs

The main variable axis for the general evaluation scenarios was the number of nodes present (from 4 and 32 nodes. It is well known that CSMA offers very low MAC access delay when network load is below

40%, but we still decided to check in detail the performance of all modes also in the lower network load part of the simulation scenarios.

Each simulation scenario was executed 10 times with random seed generation in order to avoid simulation artifacts and results of each run were then averaged over the 10 runs.

4.5.1.4   Results

We executed both simulation modes of pure LAD and H-LAD from 4 nodes to 32 nodes density to evaluate the offered network performance.

In the following graphs the Application layer latency as well as the MAC access delay is presented in the form of Min, Max, Average and Standard Deviation to present in detail the overall performance offered from each mode and setting in the simulation scenario plane. Each mode is labelled as either Baseline (CSMA only), LAD (pure LAD mode) and H-LAD accompanied by the number of nodes present in the scenario. Five scenarios in total were executed in each simulation run based on the number of nodes (4-8-16-24-32 nodes) and the results are combined in a single graph for easy comparison.

In Figure 118, the application layer latency is presented for the 2 LAD modes vs the relevant CSMA baseline but also the coexisting CSMA based best effort traffic. In order to realise what the graph presents easier, allow us to give a short explanation. 5 experiment cycles in total were executed from 4 to 32 nodes (4,8,16,24,32 nodes) adding network load as the total node count increases. In each experiment cycle we got the baseline where all nodes were running basic CSMA, and 2 experiments running a variant of LAD and standard CSMA to implement a robotic arm control priority traffic. Hence each experiment cycle generated 5 flows performance graphs (Baseline, CSMA, H-CSMA, LAD and H-LAD). Direct comparison can be made between Hybrid or non-Hybrid variants of the CSMA and LAD access schemes, as well as all of them against the relevant baseline. Also, all variants performance profile can be observed as the network load increases. For each flow, Min, Max, Average and deviation of the KPI is given in order to have a full view of the profile of performance of each flow.

At a first glance it is obvious that the performance latency wise for both LAD variants largely outperform all CSMA traffic in most cases, with only exception the 4 nodes case where pure LAD is offering larger latencies. That is expected as it is well known that CSMA offers very low access times in low capacity and node density scenarios. The interesting part here is that H-LAD outperforms the baseline and CSMA variants even in the 4 nodes density scenarios. The pattern in almost all scenarios is the same, CSMA offers better performance when coexisting with LAD except the 32 nodes scenario, which indicated that offloading traffic to the LAD MAC access scheme benefits also CSMA, even in the H-LAD variant where CSMA is shared for priority and not priority traffic. LAD offers always worse latency than H-LAD, but also has less impact on the relevant CSMA links.

Maximum values remain quite stable for the LAD variants while the CSMA based variants all exhibit significant degradation as the network load and density increases. A very interesting result is also the very low jitter of all LAD variants vs the CSMA variants in all cases except the 4 nodes scenario.

Looking at the MAC layer access latency in Figure 119, a similar picture is painted, revealing that the experienced Application layer latency is based on the offered MAC access delay of each tested MAC scheme. The MAC access offered from LAD outperforms the CSMA variants with a factor of 2-10,

especially in mid to high load and density scenarios the difference is obvious even in the logarithmic axis used in both figures. For sake of clarity, also the value points are offered for each measurement so the viewer can dive even deeper in the performance gains comparing absolute values.

To get a more complete picture of the offered performance, we need to look at the PSR of each scenario and see the packet delivery performance of each scheme that offered the presented latency performance. As it is obvious, having very good latency performance is only as good as achieving it with also very high PSR, else it is just a trade-off between latency and reliability.

Looking at the PSR of the execute simulation runs, presented in Figure 120,we can see that in most scenarios the PSR is >99.8%. But this pattern changes for CSMA variants in the 24 and 32 nodes runs. There CSMA variants including the baseline exhibit again degrading performance losing a significant number of packets. All LAD variants perform admirably even in the 32 nodes density, being able to offer >99% PSR, showcasing the power of the LAD MAC scheme even when load of the network is more than 70% and the density of the network is 32 nodes.

LATENCY LAD performance vs #nodes

	AVG	MIN	MAX	DEV
Baseline-4	0,97	0,50	3,74	0,42
CSMA-4	0,72	0,50	1,51	0,26
H-CSMA-4	1,21	0,50	2,15	0,18
LAD-4	1,51	0,46	2,38	0,50
H-LAD-4	0,48	0,46	1,12	0,21
Baseline-8	1,73	0,50	5,38	0,91
CSMA-8	1,59	0,50	4,91	0,93
H-CSMA-8	2,16	0,50	4,92	0,44
LAD-8	1,49	0,46	3,62	0,50
H-LAD-8	0,79	0,46	3,26	0,34
Baseline-16	3,46	0,50	13,60	1,90
CSMA-16	3,56	0,50	8,11	1,78
H-CSMA-16	4,16	0,50	7,97	0,97
LAD-16	2,13	0,46	5,61	0,81
H-LAD-16	1,45	0,46	4,61	0,76
Baseline-24	27,26	0,50	129,75	18,29
CSMA-24	20,05	0,50	111,28	15,01
H-CSMA-24	23,86	0,50	123,44	16,95
LAD-24	2,76	0,46	6,78	1,14
H-LAD-24	2,54	0,46	6,75	1,19
Baseline-32	47,87	0,50	166,02	24,10
CSMA-32	75,71	0,55	285,92	37,65
H-CSMA-32	83,95	2,12	304,36	40,77
LAD-32	3,42	0,46	8,58	1,50
H-LAD-32	3,24	0,46	8,22	1,59

Experiments and Flows

● AVG   MIN   MAX   ● DEV

FIGURE 118. LAD AND H-LAD APPLICATION LAYER LATENCY VS BASELINE AS NETWORK DENSITY AND LOAD INCREASES
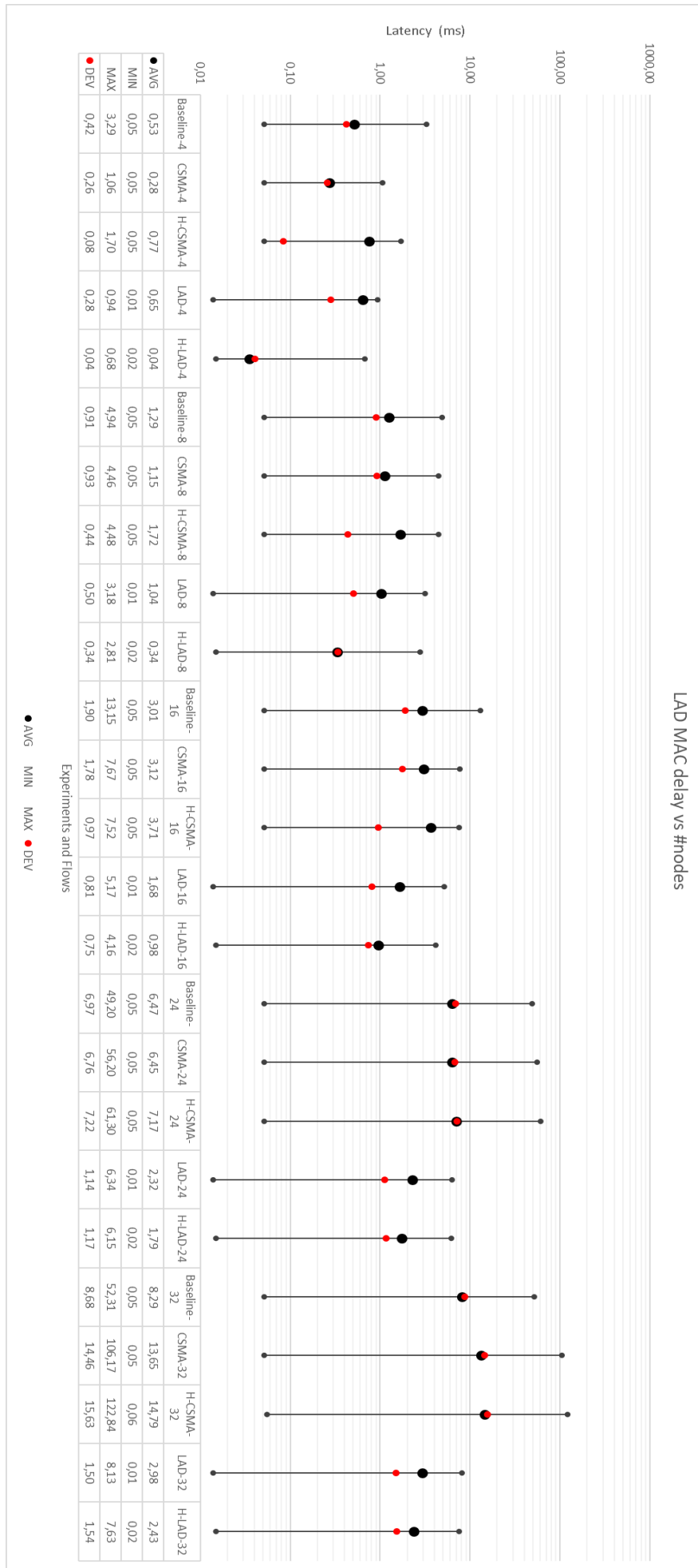
FIGURE 119. MAC ACCESS DELAY FOR LAD VARIANTS VS BASELINE AS DENSITY AND LOAD INCREASES

	AVG	MIN	MAX	DEV
Baseline-4	0,53	0,05	3,29	0,42
CSMA-4	0,28	0,05	1,06	0,26
H-CSMA-4	0,77	0,05	1,70	0,08
LAD-4	0,65	0,01	0,94	0,28
H-LAD-4	0,04	0,02	0,68	0,04
Baseline-8	1,29	0,05	4,94	0,91
CSMA-8	1,15	0,05	4,46	0,93
H-CSMA-8	1,72	0,05	4,48	0,44
LAD-8	1,04	0,01	3,18	0,50
H-LAD-8	0,34	0,02	2,81	0,34
Baseline-16	3,01	0,05	13,15	1,90
CSMA-16	3,12	0,05	7,67	1,78
H-CSMA-16	3,71	0,05	7,52	0,97
LAD-16	1,68	0,01	5,17	0,81
H-LAD-16	0,98	0,02	4,16	0,75
Baseline-24	6,47	0,05	49,20	6,97
CSMA-24	6,45	0,05	56,20	6,76
H-CSMA-24	7,17	0,05	61,30	7,22
LAD-24	2,32	0,01	6,34	1,14
H-LAD-24	1,79	0,02	6,15	1,17
Baseline-32	8,29	0,05	52,31	8,68
CSMA-32	13,65	0,05	106,17	14,46
H-CSMA-32	14,79	0,06	122,84	15,63
LAD-32	2,98	0,01	8,13	1,50
H-LAD-32	2,43	0,02	7,63	1,54

Experiments and Flows

● AVG   MIN   MAX ●   DEV

LAD MAC delay vs #nodes

Latency (ms)

FIGURE 120. PACKET SUCCESS RATIO ACROSS ALL SCENARIOS AND MAC MODES

Another very interesting observation is that the CSMA variants perform better than the baseline in the 24 nodes scenario, showcasing that the existence of LAD and offloading of 50% of packets through it has a positive impact in the CSMA flows performance as well. Such an impact proves that there is still space for win-win situations in the MAC design and not all is simply a trade-off between different approaches. Finally, as expected the hybrid variance of LAD has a slightly higher impact in PSR of the coexisting CSMA flows, since it tries to reuse also the CSMA time and not only the LAD periods.
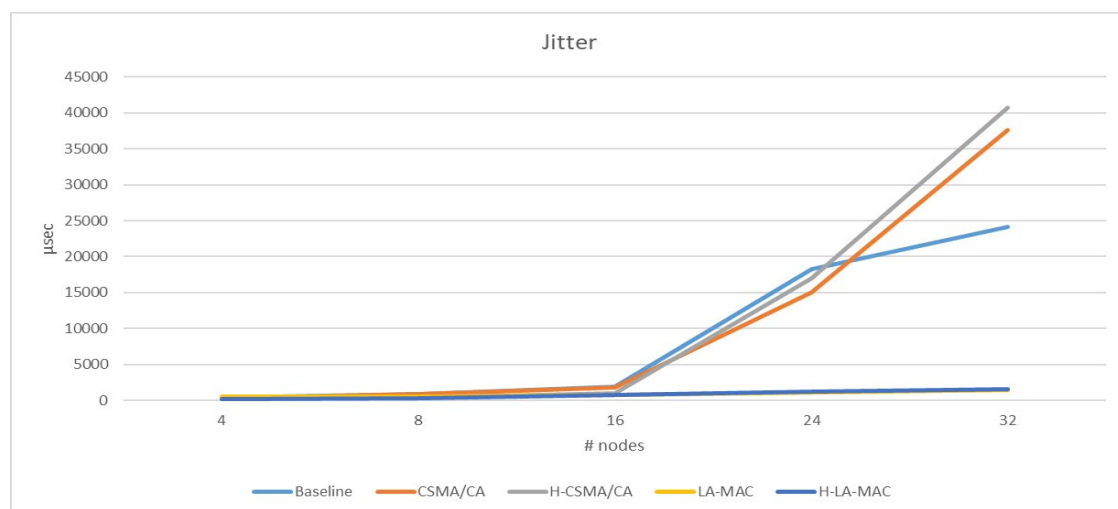
FIGURE 121. JITTER VS # NODES FOR LAD VS WIFI AND BASELINE

Regarding jitter, the results from all scenario runs are presented below in Figure 121. The perceived jitter in the application layer is presented and evaluated vs the baseline. As it is obvious the LAD and H-LAD offer almost stable (from 200 to 1400 μsec) jitter while the Wi-Fi variants exhibit highly increasing jitter as the network load increases, reaching values of 25 to 45 msec compared to the 1.4 msec

maximum jitter presented in LAD. Hence, jitter wise, we achieved x17 to x32 times lower jitter than the baseline in high load conditions and similar or better jitter in the low load scenarios, especially for the H-LAD variant. Since the graph is scaling up to high numbers due to the Wi-Fi achieved jitter, we also present a standalone graph to compare easier between the 2 LAD variants in Figure 122.
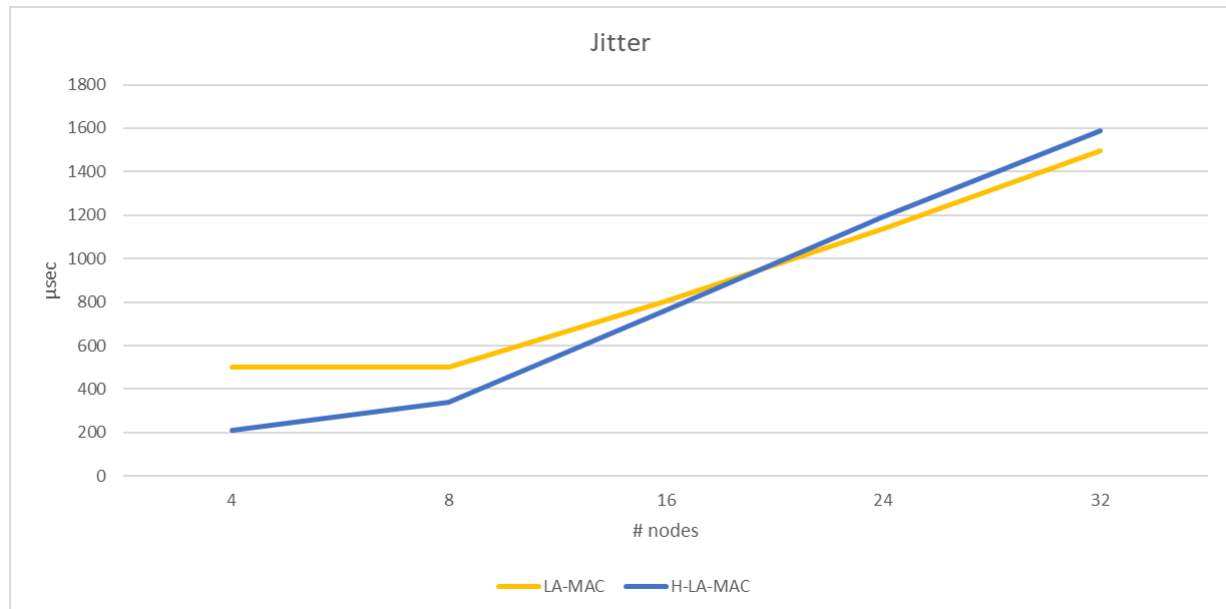
FIGURE 122. LAD VARIANTS JITTER VS # NODES

The H-LAD is performing better in low to medium number of nodes (linearly equal to network load) and in high network load conditions the pure LA-MAC is better. This result makes sense as the H-LA uses also the CSMA time for medium access and therefore increasing network load should impact it more. On the other hand, the pure LA-MAC is less impacted from load, as the only chance of a packet loss or collisions is if there is a failed deconfliction of the packets trying to get access in the medium during the randomization part of the LA. Overall we can state that H-LA is more than 50% better jitter-wise in low network load conditions, but only 10% worse in high network load conditions. Therefore, we consider it to be a better candidate for final selection between the two, seeing that also the impact between LA and H-LA to the Wi-Fi priority flows are not very different.

4.5.1.5 WP2 usecase scenario results

To test the proposed MAC access scheme for a real deployment based on WP2 identified usecases, we chose the robotic arm control usecase and setup a simulation environment similar to the characteristics of the PHY identified in WP2. The detailed characteristics of the simulation environment can be seen in Table 24 :

TABLE 24. ROBOTIC ARM CONTROL USE CASE ENVIRONMENT SIMULATION SETTINGS

Simulation duration	10 s
Number of simulations per use case	1
Number of nodes (N)	4
Number of communication links	N/2 => 2 -> 1 LAD and 1 CSMA
Nodes mobility	fixed positions in the same collision domain
PHY supported data rate	165 Mbps

The setup now includes 4 nodes forming 2 links, one link will be treated by the typical CSMA algorithm and the second with the H-LA MAC. We will compare the results based on Application layer latency, MAC access delay and observed jitter for both data links. Based on the envisioned wider PHY channel bandwidth for such applications within the definition of the usecase in WP2, we choose to scale the available capacity of the PHY to 165 Mbps compared to the basic 11 Mbps used in the general simulations. The CSMA link is acting as a best effort link to saturate the PHY and help pinpoint the ability of the LA-MAC to find gaps and send packets through in a deterministic way even when the PHY is saturated. The PHY rate of the CSMA link is around 56.1 Mbps, using around 34% of the PHY capacity while the PHY rate of the robotic arm control is around 9.24 Mbps or 5.6% of the available PHY capacity. We consider this to be a medium network load environment that can be representative of most real-life scenarios.

In Figure 123 we present the achieved Application layer latency, the MAC access delay and the jitter of the LA robotic arm control link vs the CSMA background traffic link.
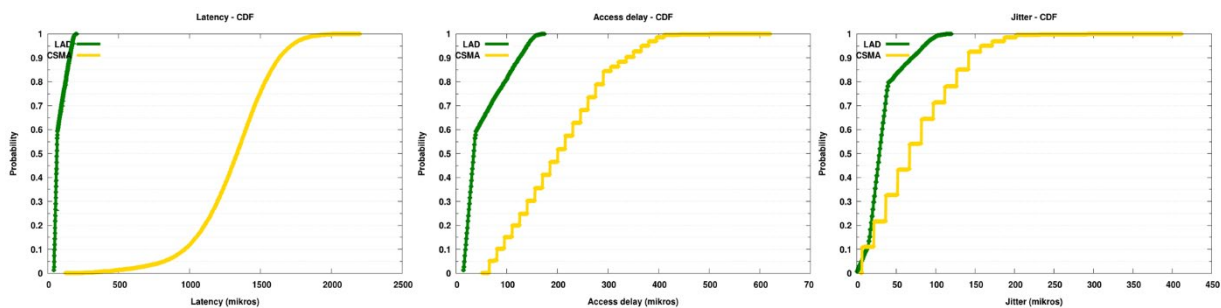

FIGURE 123. LATENCY, ACCESS DELAY AND JITTER OF LA VS CSMA DURING ROBOTIC ARM SIMULATION SCENARIO.

As it can be clearly seen, the LA is outperforming the CSMA MAC scheme in all aspects, providing bounded latency around 200μsec, bounded access delay capped at 175 μsec and a jitter no mor than 115 μsec in all cases. This is achieved while the background CSMA link is using 34% of the PHY capacity, clearly demonstrating that the LA access scheme can decouple low priority and high priority flows and offer deterministic network operation in terms of latency and access delay with extremely low jitter with minimum impact of other traffic flows that follow the legacy CCA timings.

4.5.1.6 Summary and recommendations

After analysing the general results but also the WP2 use case simulation scenario we can safely state that having a hard prioritization capability based on the CCA timings can significantly decouple low and high priority flows on the MAC layer and offer deterministic networking performance in low to medium network load environments. The proposed LA mechanism in either of each two forms (pure or hybrid) achieved significantly better and stable networking performance compared to the legacy CSMA approaches that are based on a static CCA duration before accessing the medium as it is currently the case in all related ETSI standards for ISM bands access. This can be a true enabler for deterministic networking in ISM bands that any wireless technology can take advantage of to offer stable QoS independently from the actual load of the network. If ETSI standards are revised to allow differentiated CCA timings for priority data flows, radio technologies operating in the unlicensed spectrum can take advantage of this enhancement and offer hard priority for the highest priority flows or for specific application domains like industrial automation.

Specifically, we achieved similar or better MAC access delay in low network conditions, ranging from a 93% to a 71% average access delay reduction, while the maximum reduction was from 36% to 88% in high network load conditions. Jitter was also reduced from 33% to 83% compared to the baseline.

The achieved average Application layer delay improvements ranged from 14% (low load) to 93% (high load) b lower latency, always compared to the baseline. The maximum delay improvements were from 33% (low network load) to the astounding 95% (high network load), establishing clearly a picture that depicts that the proposed LA scheme can indeed offer deterministic operation independent of best effort traffic presence and load.

We plan to move forward and propose the dual priority-based support, that is the basis of the proposed LA MAC scheme, to be integrated in the ETSI standards for ISM 2.4 and 5GHz bands, as it will provide the capability of true priority support when sharing the medium with best effort technologies like Wi-Fi.

4.5.2 Adaptive Low Latency and reliable unlicensed spectrum access

We propose here an alternative approach to enhancing latency and reliability in in-X subnetworks operating within the unlicensed spectrum. Unlike the previously discussed solution—which introduces dedicated time slots for deconfliction—this method focuses on adjusting key access parameters to minimize channel access latency based on current radio conditions. The primary use cases of interest lie in industrial settings, such as unit test control.

4.5.2.1 Introduction & Background

In Deliverable 3.1 [2], the challenges of subnetworks operating in the unlicensed bands were explored, highlighting critical issues such as unpredictable channel access and LBT-induced delays. 3GPP defines two unlicensed channel access schemes: dynamic channel access—also known as Load-Based Equipment (LBE), which employs random backoff procedures—and semi-static channel access, or Frame-Based Equipment (FBE), which restricts channel access attempts to fixed, periodic intervals [90]. The absence of Wi-Fi in certain industrial environments enables controlled spectrum conditions, under which FBE can serve as a reliable access mechanism—especially when traffic follows predictable, periodic patterns. This access strategy aligns with the Unit Test Cell use case described in Deliverable D2.2 [3], where sensor nodes transmit measurements individually to a central controller or aggregator. In FBE mode the channel sensing intervals can be independently defined for each device. This flexibility makes it possible to support a variety of traffic patterns across different SNEs, while still preserving coordinated and predictable medium access. The HC node acts as the subnetwork's access point, while the SNEs communicate over the unlicensed band, each following its own access periodicity within the FBE framework.

To reduce latency once access is granted, FBE can be paired with configured grant (CG) scheduling and UE-initiated channel occupancy [91]. This enables SNEs to perform LBT at fixed intervals and, upon success, immediately transmit using pre-assigned uplink resources—thereby avoiding dynamic scheduling delays and supporting timely packet delivery. However, while this configuration suits periodic traffic well, the Unit Test Cell use case may also involve aperiodic transmissions [3], where some sensors send data irregularly. Such cases do not align well with the fixed timing structure of FBE and is more naturally handled using LBE access, where a device contends for the channel as soon as data becomes available.

Even in Wi-Fi-free environments, the coexistence of FBE and LBE SNEs within the same NR-U system introduces contention dynamics [92]. The unpredictable channel access behaviour of LBE devices can interfere with the scheduled transmissions of FBE SNEs, leading to repeated LBT failures and reducing the reliability. This challenge highlights the need for more flexible mechanisms that enable FBE-based SNEs to react to congestion and channel variability, without discarding their periodic access structure. The following section examines the interaction between FBE and CG in NR-U systems and identifies key limitations under congestion, motivating the need for adaptive use of the granted Channel Occupancy Time (COT).

4.5.2.2   FBE and Configured Grant

While FBE regulates when a SNE is allowed to attempt channel access, CG scheduling determines when uplink resources are available for transmission once access has been granted. CG provides each SNE with transmission opportunities that repeat periodically, based on parameters configured by the HC node. Since both FBE and CG operate on periodic schedules, it is beneficial to align them [93]. By configuring CG transmissions to coincide with the expected start of a COT after a successful LBT attempt, the SNE can immediately utilize its granted uplink resources. Without this synchronization, an SNE might gain access to the channel only to find that no CG transmission is currently valid—or conversely, a CG opportunity might expire while the SNE waits for access. This not only minimizes latency but also makes delay more predictable—an essential requirement for low latency services operating in unlicensed spectrum. To further enhance responsiveness, NR-U introduces the concept of UE-initiated COT. Traditionally, the initiation of a COT is controlled by the gNB (or HC node in a subnetwork), which may introduce additional signaling delays. With UE-initiated COT, an SNE that completes LBT can begin its COT immediately, enabling prompt use of its aligned CG opportunity and avoiding unnecessary idle time. This allows the SNE to make full use of its aligned CG opportunity as soon as the channel becomes available, further reducing access latency and avoiding unnecessary idle periods. However, even with careful FBE–CG alignment and the responsiveness offered by UE-initiated COT, contention can still arise—not only from external technologies like Wi-Fi, but also from other SNEs operating under LBE or FBE access. This kind of intra-subnetwork contention can result in repeated missed opportunities despite pre-scheduled grants. Figure 124 illustrates this coordinated yet contention-prone access model.
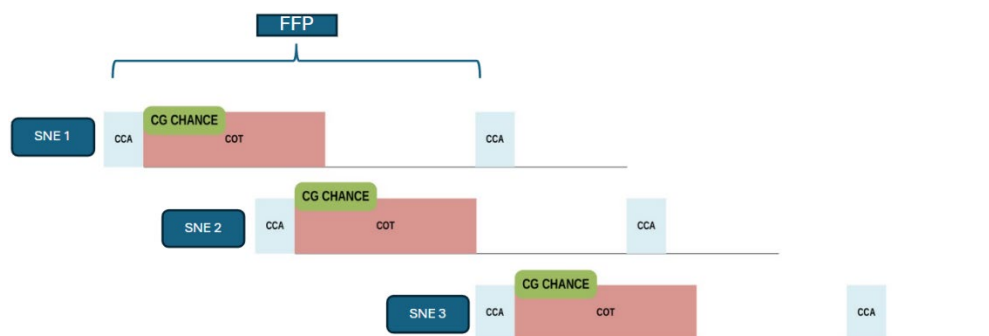
FIGURE 124: CONFIGURED GRANT WITH FBE

4.5.2.3   COT strategy Trade-Offs

Once an SNE successfully gains access to the channel, it transmits its data within the granted COT, a time window whose structure and duration are constrained by NR-U regulations and system design. In FBE-based systems, this COT is bounded by the Fixed Frame Period (FFP) and subject to regulatory constraints—typically capped at 90–95% of the FFP duration—to ensure fair coexistence in the

unlicensed band. Additionally, not all of the granted COT is available for data transmission: portions must be reserved for HARQ feedback and guard intervals. As such, only a subset of the COT can be effectively used for uplink data. In this work, we model that usable portion as a fixed number of full transmission slots, each carrying one PUSCH.

Under FBE-CG alignment, the transmission behavior within the granted COT is predefined. Specifically, the strategy for filling the available slots is determined through RRC signaling and remains fixed at data delivery time. Two strategies are currently supported. In repetition-based transmission, the same transport block (TB) is sent across multiple slots within the COT to improve decoding reliability. This is what is considered as default mode for many URLLC applications over unlicensed [93]. An alternative mode supported in NR-U is multi-PUSCH transmission [94], where each slot in the granted COT carries a different TB. This approach is more suited to high-throughput services, as it increases spectral efficiency by maximizing data delivery per grant. These two strategies represent distinct trade-offs between reliability and throughput and are illustrated in Figure 125, which shows how the available uplink slots within a COT are used with each configuration.
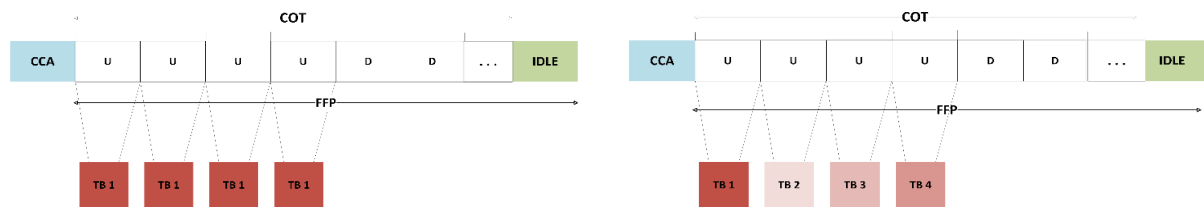

FIGURE 125: CONFIGURED TRANSMISSION STRATEGIES WITHIN A COT, (A) REPETITION-BASED AND (B) MULTI-PUSCH TRANSMISSION SCHEME.

While repetition and multi-PUSCH each offer benefits under specific traffic and channel conditions, the key limitation arises from the semi-static nature of CG scheduling when paired with FBE. In this configuration, the strategy for using the COT—whether to transmit repetitions or multiple distinct transport blocks—is determined through RRC signaling and remains fixed during operation. For instance, an SNE configured for repetitions may continue to resend the same TB even when its uplink buffer is building up due to prior LBT failures. Conversely, an SNE configured for multi-PUSCH may transmit distinct packets without redundancy, even when channel conditions are poor after LBT success—leading to packet loss. To demonstrate the limitations of current static transmission schemes, we simulate a typical scenario with periodic traffic, constrained latency, and varying channel access opportunities. Upon each successful LBT attempt, the SNE transmits either repeated or distinct transport blocks within the granted COT, depending on its preconfigured mode. A packet is considered dropped if it is not delivered within its latency budget. The key simulation parameters are summarized in Table 25.

TABLE 25: SIMULATION PARAMETERS

Parameter	Setting
Traffic Pattern	Periodic; uplink generation aligned with CG periodicity
CG / Packet Generation Period	8ms (example value; representative of typical URLLC configurations)
Wireless Channel	3GPP industrial [95]
Modulation and Coding Scheme (MCS)	14
FFP	Aligned with CG periodicity
COT Duration	7 ms (in line with 3GPP-compliant limits)
Packet Latency Budget	Varied (10–20 ms)
Channel Access Probability	Varied (p ∈ [0.25, 1])

Number of UL slots within COT (M)	4

Figure 126 and Figure 127 show the packet drop ratio as a function of average SNR for both repetition-based and multi-PUSCH configurations, accordingly, evaluated under different channel access probabilities. As expected, repetition delivers superior reliability under low-SNR conditions due to redundancy, but it quickly becomes inefficient in access-limited situations—since repeated transmissions consume valuable COT time that could be used to transmit new packets that wait in the transmission buffer of the SNE. Conversely, multi-PUSCH is more efficient in terms of buffer clearance under congestion but suffers at low SNR, where the lack of redundancy results in higher packet loss. While packets are forwarded for transmission more frequently their deadlines are violated by staying in the HARQ too long to either reach maximum retransmission counter or expire before successful reception. These results reveal that neither static approach alone can meet the deadline latency demands across diverse conditions, highlighting the need for a different transmission strategy.
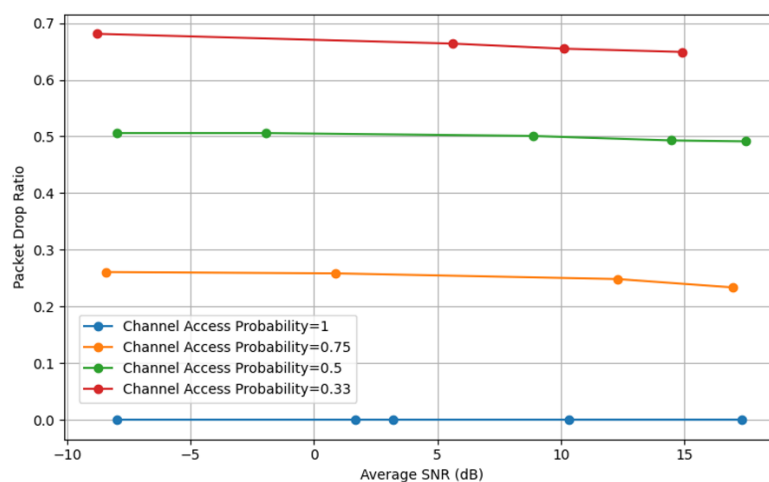
FIGURE 126: PACKET DROP RATIO VS AVERAGE SNR FOR REPETITION MODE COT UNDER DIFFERENT CHANNEL ACCESS PROBABILITIES
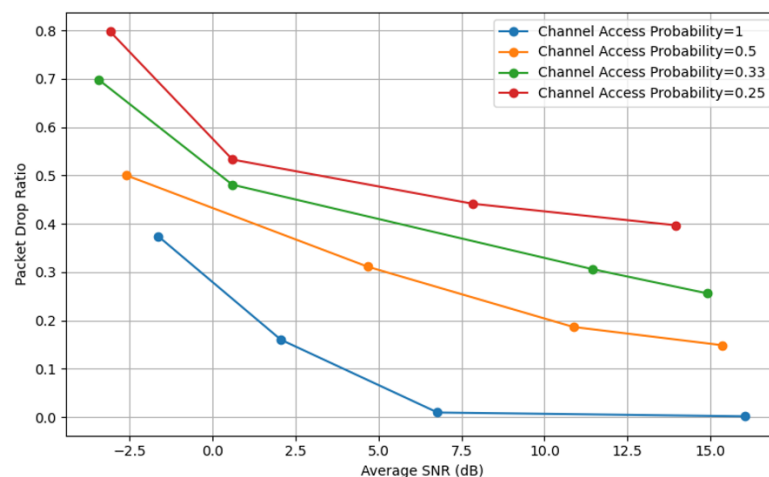
FIGURE 127: PACKET DROP RATIO VS AVERAGE SNR FOR MULTI-PUSCH COT MODE UNDER DIFFERENT CHANNEL ACCESS PROBABILITIES

4.5.2.4 Adaptive COT mode utilization

These opposing failure reasons—repetition's vulnerability to channel-access congestion, and multi-PUSCH's vulnerability to link-quality —motivate a solution that combines strengths from both modes. A hybrid mode, where inside the COT we could have a portion of TBs repeated partially, could be a

promising approach, and it has indeed been floated in 3GPP discussions; however, its realization is blocked by the fundamental HARQ-process mapping ambiguity it would introduce [96]. As an alternative, a per-COT adaptation is proposed that, without changing the CG parameters, dynamically selects either repetition or multi-PUSCH at each channel access opportunity. To do so, both the SNEs and HC node maintain locally updated counters for consecutive LBT failures and recent HARQ NACK indications. Because CG occasions occur at strictly periodic instants, the absence of an expected PUSCH transmission directly implies an LBT failure at the HC node side. Both ends therefore maintain an identical view of channel-access outcomes and can synchronously switch modes. In this scheme, every granted COT begins in repetition mode to exploit soft-combining gain. Once the locally tracked LBT-failure counter reaches a threshold value N, the very next successful channel occupation automatically becomes a multi-PUSCH burst, with each slot in the COT carrying a distinct TB to rapidly clear any backlog and minimize queuing delay. This threshold can be configured by the network and signalled via RRC — just like other FBE and CG parameters—using, for example, a new RRC field. After such a burst is transmitted, the mechanism evaluates the outcome to decide how to proceed. If any block in that burst is NACK-ed the COT mode reverts to repetition mode. In the rare case that the LBT-failure counter again reaches N before the forced repetition has been sent, the LBT-failure counter takes priority —ensuring we don't endlessly repeat a TB that might already be expired. The basic flow of this algorithm is illustrated in Figure 128. Notably, the entire mechanism operates solely via locally maintained counters and HARQ outcomes, imposing zero additional signaling overhead—an important advantage in unlicensed bands, where signaling would be subject to additional LBT procedures, increasing access delays.
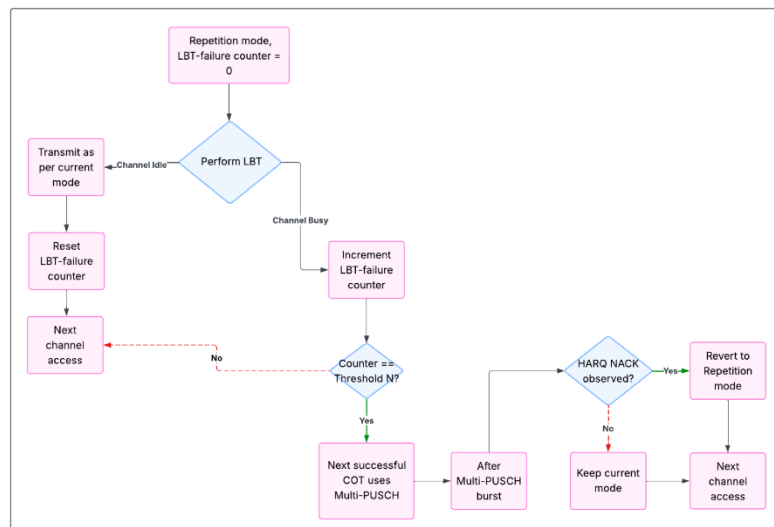

FIGURE 128: FLOWCHART OF MODE SWITCHING ALGORITHM

The selection of the threshold N is subject to fundamental system constraints. First, the backlog constraint ensures that after N consecutive LBT failures, the number of accumulated packets, $N+1$, does not exceed the maximum number of distinct packets, M, that can be transmitted in one *multi(M) - PUSCH* burst:

$$N \leq M - 1$$

Second, the latency constraint ensures compliance with the end-to-end latency budget D measured in slots, given a CG – FFP periodicity of T_{CG} slots. Thus, the maximum allowable number of failures N must satisfy:

$$N \leq \left\lceil \frac{D}{T_{CG}} \right\rceil - 1$$

Combining these the maximum feasible threshold is:

$$N_{max} = \min\left\{ M - 1, \ \left\lceil \frac{D}{T_{CG}} \right\rceil - 1 \right\}$$

For example, if we keep each multi-PUSCH to carry 4 UL slots and the CG-FFP period is equal to 8 slots, with a packet deadline of 15 slots we get a $N_{max} = 1$. Importantly, this bound on N does not drive the packet-drop rate to zero. By limiting the SNEs to at most one missed CG before a multi-PUSCH burst, we try to guarantee that no packet will expire simply because it spent too long in repetition mode, and that every queued packet can be carried in a single burst without overflow. However, packets can still be lost to link-layer effects—HARQ retransmissions—or congestion due to consecutive LBT failures. Figure 129 illustrates exactly this point under the same parameter set. As predicted by our bound $N_{max} = 1$, the adaptive curve lies below all other schemes over the entire SNR range. The static curves consistently underperform, reflecting the penalty of exceeding the backlog and latency limits. For reference, the figure also includes "LBT-only" variants of the same per-COT logic: the UE simply counts up LBT failures until it reaches N, fires exactly one multi-PUSCH burst, then reverts to repetition on the next CG—without ever using any HARQ feedback to adjust its behaviour. Because it has no link-quality indicator, this "LBT-only" scheme cannot tell when the channel has recovered and harvest the multi-PUSCH at higher SNRs. Overall, the adaptive scheme with $N = 1$ and HARQ achieves on average 2–3× lower packet-drop ratio compared to static schemes across the SNR range.
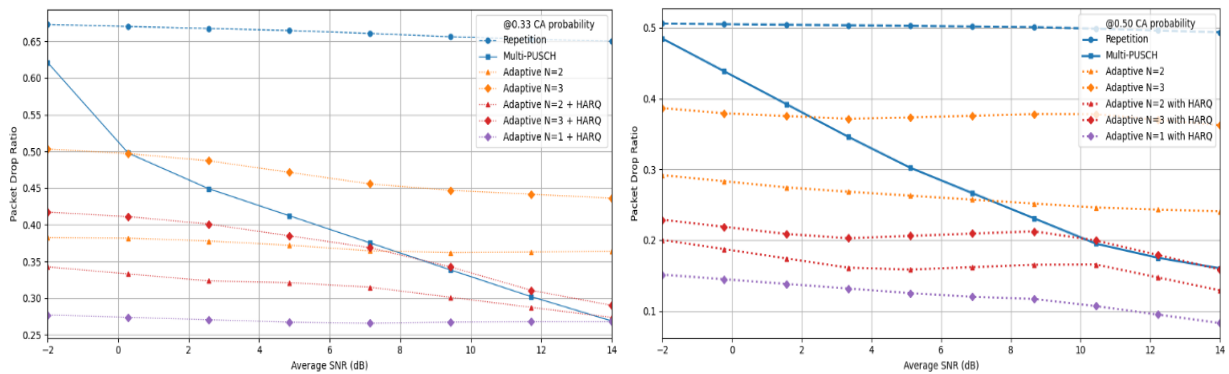
FIGURE 129: PACKET DROP RATIO VS SNR FOR PROPOSED AND STATIC SCHEMES ACROSS (LEFT) CHANNEL ACCESS PROBABILITY =0.33 AND CHANNEL ACCESS PROBABILITY =0.5

In more general cases—where $N_{max} > 1$—each choice of N in $\{1, \ldots, N_{max}\}$ trades off two different effects: larger N means the SNE stays longer in repetition mode (improving soft-combining reliability) but allows more packets to queue up (increasing backlog and risk of deadline-driven drops), smaller N forces earlier multi-PUSCH bursts (quickly clearing buffer under contention) at the expense of losing repetition gain when the link is weak. In that case the choice of optimal in N is the value in $\{1, \ldots, N_{max}\}$

that minimizes the long-run packet-drop probability for the expected operating conditions. While the instantaneous channel-access probability $p_c(t)$ and SNR vary over time due to traffic load and interference, to derive a strategy to choose the optimal N, N_{opt}, we assume they fluctuate around slowly changing averages $\overline{p_c} = E[p_{c(t)}]$ and $\overline{SNR} = E[SNR(t)]$. These average values can be estimated through periodic measurements of the expected operation environment. Under this model, the optimal threshold is given by:

$$N_{opt} = \arg \min_{1 \le N \le N_{max}} P_{drop(N; \bar{p}_c, \overline{SNR})}$$

Where P_{drop} denotes the long-run drop probability under the operating conditions. To illustrate, we have a representative example where $D = 25$, $T_{CG} = 8$, $M = 4$, so that $N_{max} = min\left(M - 1, \left\lfloor \frac{D}{T_{CG}} \right\rfloor\right) = 3$. Fixing $\overline{SNR} \approx 10$ dB, we sweep $\overline{p_c}$ across a range of typical values and record the resulting packet-drop probabilities for each threshold N ∈ {1, 2, 3} as shown in Figure 130.
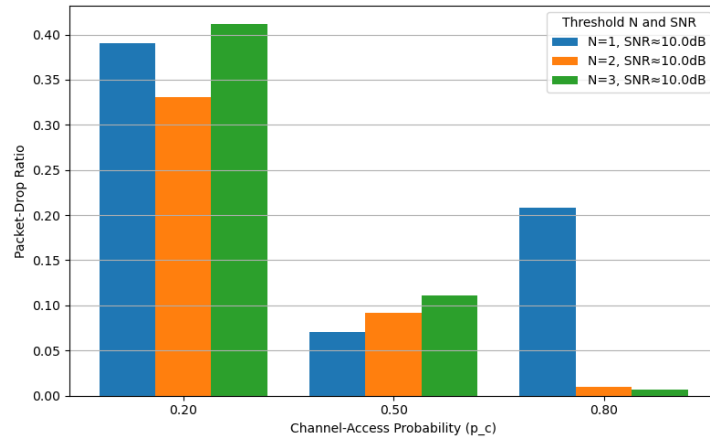

FIGURE 130: PACKET-DROP RATIO VS $\overline{p_c}$ FOR DIFFERENT THRESHOLD VALUES N UNDER FIXED $\overline{SNR} \approx 10$ DB.

As we can see, N_{opt} varies with $\overline{p_c}$ as expected. For each operating point, the threshold that minimizes the packet-drop probability changes. This demonstrates that the proposed approach allows N to be tuned to match congestion levels in the network, thereby improving packet delivery under different conditions. Since N can be reconfigured via RRC signalling, the HC node may periodically re-estimate $\overline{SNR}$ and $\overline{p_c}$ to update the threshold accordingly.

4.5.2.5 Summary and recommendations

This subsection addressed the challenges that follow the operation of subnetworks in unlicensed bands and the delays that come for the uncertain outcomes of LBT procedures along with the ones also present in licensed channels like interference and fading. Solutions for channel access and data transmission which rely on semi-static periodic behaviour, although help with reducing the signalling, suffer themselves from inflexibility. A dynamic transmission scheme was proposed that switched COT modes between repetition and multi-PUSCH based on real time conditions such as LBT failures and HARQ feedback. Rather than attempting to enhance the channel access procedure itself, the proposed algorithm focuses on what happens after the channel access is won and how to best utilize the granted COT once it becomes available. This adaptive behaviour enables improved packet delivery ratio without

additional signalling overhead. Experimental results indicate a ~x2–3 reduction in packet drop ratio due to missed deadlines across the full SNR range compared to static repetition or multi-PUSCH strategies. These improvements are particularly beneficial in unit-test cell deployments [3], where periodic and aperiodic traffic coexist, and timely delivery is critical despite contention and interference in unlicensed environments. The scheme primarily targets periodic nodes using semi-static channel access, as they are most affected when coexisting with dynamically accessing SNEs [92].

5   CONCLUSIONS AND DISCUSSION

The 6G-SHINE project has presented in this Deliverable a comprehensive and technically grounded set of PHY and MAC solutions tailored for in-X subnetworks. By exploring concrete subnetwork configurations in consumer, industrial, and vehicular domains, this deliverable has shown how the developed PHY and MAC solutions can meet demanding KPIs in throughput, latency, reliability, scalability, determinism, and synchronization—beyond what current solutions can typically support.

As a representative of the consumer use cases, the indoor interactive gaming scenario highlights the need for massive bandwidth and ultra-low-latency synchronization to deliver immersive XR experiences. The combination of outside-in tracking, local rendering, and actuator control within a confined subnetwork has stringent KPI targets. Through the use of sub-THz communication, 6G-SHINE achieves the high throughput needed for 8K–24K video streaming while keeping power consumption below 1W at the headset level, all this while achieving ~x2 lower complexity at device side, and high reliability (BLER < 10^{-6}).

In the industrial context, wireless robot control imposes some of the most stringent requirements: communication cycles below 100 µs, jitter below 1 µs, and BLER below 10^{-6}. 6G-SHINE proposes a suite of novel technologies, such as metasurfaces (SCM and DSA), jamming-resilient PHY, proximity grouping, predictive scheduling, coded random access, and network-coded cooperation, to not only replicate but in many cases surpass the capabilities of wired systems like EtherCAT. Results show ~x10 latency reduction, almost a ~x20 improvement in average and maximum latency in unlicensed bands, packet error rate <10^{-6} for communication cycles <100 us (up to 10^{-7} in less strict scenarios), up to ~x4 improvement in spectral efficiency, and significant power savings through EM-level signal processing, beamforming and cooperative communication. Through these schemes, the subnetworks can maintain ultra-reliable operation even under co-channel interference and adversarial jamming, proving high resilience and robustness.

In the vehicular domain, the Virtual (Wireless) ECU use case represents a transformative step toward fully wireless in-vehicle networks, replacing traditional cable harnesses with flexible, high-capacity 6G subnetwork wireless zones. Typical KPIs for this scenario include latencies below 1 ms, data rates from 2 Mbps to Gbps accommodating diverse sensor data streams, reliability greater than 99.9%, deterministic bounded latencies and tight control loop times. To meet these demanding KPIs, 6G-SHINE has developed innovative solutions grounded in deterministic and predictive traffic scheduling as well as flexible duplexing mechanisms. Compared to state-of-the-art schemes, the proposed predictive scheduling schemes achieve up to an 18-fold improvement in communication reliability, improve scalability performance by factors of up to ~7x, and boost resource utilization efficiency by up to 2x, ensuring that network resources are optimally used to meet latency and reliability constraints without unnecessary overprovisioning.

Several other PHY and MAC solutions are presented in this Deliverable which, even though they have been configured to address a primary use case, they can be applied across multiple use cases. This includes:

- Antenna arrays and reciprocity-based beam alignment procedure can increase the throughput in the order of +50% to +100% compared to a single wide-beam transmit antenna configuration.

- Full/flexible duplexing for enhanced scheduling flexibility achieving a factor of ~x10 reduction of latency with respect to 5G and a spectral efficiency gain of up to ~x8.
- Adaptive low-latency and reliable unlicensed spectrum access schemes showing a 2-3x decreased packet drop ratio across the full SNR range compared to state-of-the-art schemes.
- Interference-resilient and scalable access schemes allow for dense subnetwork deployments even in spectrum-constrained or unlicensed environments.
- Predictive and adaptive mechanisms, including LSTM-based scheduling and proximity-aware retransmissions, increase responsiveness while optimizing resource usage.
- Hardware-native innovations, such as metasurfaces and sub-THz transceivers, unlock performance levels not achievable with conventional digital techniques.

In conclusion, this deliverable contributes with the recommended PHY and MAC solutions to the 6G-SHINE vision that subnetworks, built around context-specific design principles and empowered by 6G-era technologies, can meet the diverse and extreme demands of future applications and use cases, and can be a foundational 6G component.

REFERENCES

[1] SNS JU Architecture WG, "Towards 6G Architecture: Key Concepts, Challenges, and Building Blocks", White Paper, May 2025. DOI: 10.5281/zenodo.15001377

[2] 6G-SHINE D3.1, "Preliminary results on PHY and MAC enablers for in-X subnetworks", Feb. 2024. Available online at: https://6gshine.eu/wp-content/uploads/2024/11/D3.1-Preliminary-results-on-PHY-and-MAC-enablers-for-in-X-subnetworks_v1.0.pdf

[3] 6G-SHINE D2.2, "Refined definition of scenarios, use cases and service requirements for in- X subnetworks", Feb. 2024. Available online at: https://6gshine.eu/wp-content/uploads/2024/11/D2.2-Refined-definition-of-scenarios-use-cases-and-service-requirements-for-in-X-subnetworks-_v1.0.pdf

[4] 6G-SHINE D2.3, "Radio propagation characteristics for in-X subnetworks", Nov. 2024. Available online at: https://6gshine.eu/wp-content/uploads/2024/12/6GSHINE_Deliverable_2.3_v1.0_disclaimer.pdf

[5] F. Fernandes, C. Rom, J. Harrebek, S. Svendsen and C. N. Manchón, "Hand Blockage Impact on 5G mmWave Beam Management Performance," in IEEE Access, 2022.

[6] 3GPP TR 38.817-01, "General aspects for Base Station (BS) Radio Frequency (RF) for NR" V16.4.0, Table 7.2.1.4.1-1, Sept. 2022.

[7] "Nokia White Paper" How 5G is making us rethink UE antenna design".

[8] 3GPP TS 38.300, "NR; NR and NG-RAN Overall description; Stage-2", V18.5.0, March 2025.

[9] 6G-SHINE D4.1, "Preliminary results on the management of radio resources in subnetworks in the presence of legitimate and malicious interferers", June 2024. Available online at: https://6gshine.eu/wp-content/uploads/2024/11/D4.1_Preliminary-results-on-the-management-of-radio-resources-in-subnetworks_v1.0.pdf

[10] 3GPP RAN1#104, August 2023, R1-2308605 "FL summary #4 for AI 9.4.3 Enhanced sidelink operation on FR2 licensed spectrum"

[11] A. Dang. S., B. O. Shihada, and et al., "What should 6G be?" Nature Electronics 3, vol. 20-29, 2020.

[12] H. Thang, N. Shlezinger, F. Guidi, D. Dardari, and Y. C. Eldar, "6G wireless communications: From far-field beam steering to near-field beam focusing." IEEE Communications Magazine, vol. 61, no. 4, pp. 72-77. April 2023.

[13] N. Gonzlez-Prelcic, M. Furkan Keskin, O. Kaltiokallio, M. Valkama, D. Dardari, X. Shen, Y. Shen, M. Bayraktar, and H. Wymeersch, "The integrated sensing and communication revolution for 6G: Vision, techniques, and applications," Proceedings of the IEEE, vol. 112, no. 7. pp. 676-723, July 2024.

[14] T. Gong, P. Gavriilidis, R. Ji, C. Huang, G. C. Alexandropoulos, L. Wei, Z. Zhang. M. Debbah, H. V. Poor, and C. Yeen, "Holographic MIMO communications: Theoretical foundations, enabling technologies, and future directions," IEEE Communications Surveys & Tutorials, vol. 26, no. 1, pp. 196-257, 2024.

[15] E. Björnson, Y. C. Eldar, E. G. Larsson, A. Lozano, and H. V. Poor, "Twenty-five years of signal processing advances for multiantenna communications: From theory to mainstream technology," IEEE Signal Processing Magazine, vol. 40, no. 4, pp. 107-117, 2023.

[16] J. Zhu, Z. Wan, L. Dai, M. Debbah, and H. V. Poor, "Electromagnetic Information Theory: Fundamentals, Modeling. Applications, and Open Problems," arXiv e-prints, p. arXiv:2212.02882, Dec. 2022.

[17] M. D. Renzo and M. D. Mighiore, "Electromagnetic signal and information theory," IEEE BITS the Information Theory Magazine, pp. 1-13, 2024.

[18] D. Dardari, G. Torcolacci, G. Pasolini, and N. Decarli, "An overview on over-the-air electromagnetic signal processing," 2024. [Online]. Submitted to IEEE Signal Processing Magazine. Available: https://arxiv.org/abs/2412.14968

[19] M. Di Renzo, F. H. Danufane, and S. Tretyakov, "Communication models for reconfigurable intelligent surfaces: From surface electromagnetics to wireless networks optimization," Proceedings of the IEEE, vol. 110, no. 9, pp. 1164–1209, 2022.

[20] D. Dardari, "Reconfigurable electromagnetic environments: A general framework," IEEE Journal on Selected Areas in Communications, vol. 42, no. 6, pp. 1479-1493, June 2024.

[21] D. Dardari, M. Lotti, N. Decarli and G. Pasolini, "Establishing Multi-User MIMO Communications Automatically Using Retrodirective Arrays," in IEEE Open Journal of the Communications Society, vol. 4, pp. 1396-1416, 2023, doi: 10.1109/OJCOMS.2023.3289326.

[22] J. An, C. Xu, D. W. K. Ng, G. C. Alexandropoulos, C. Huang, C. Yuen, and L. Hanzo, "Stacked intelligent metasurfaces for efficient holographic MIMO communications in 6G," IEEE Journal on Selected Areas in Communications, vol. 41, no. 8, pp. 2380-2396, 2023ç

[23] D. Dardari, "3D electromagnetic signal processing," in Proc. Asilomar Conf. on Signals, Systems and Computers, Oct 2024.

[24] D. Dardari, M. Lotti, N. Decarli and G. Pasolini, "Grant-Free Random Access With Backscattering Self-Conjugating Metasurfaces," in IEEE Transactions on Cognitive Communications and Networking, vol. 10, no. 5, pp. 1620-1634, Oct. 2024.

[25] M. Lotti, N. Decarli, G. Pasolini, and D. Dardari, "Localization Based on MIMO Backscattering from Retro-Directive Antenna Arrays," arXiv e-prints, p. arXiv:2404.14206, Apr. 2024. Accepted for publication on IEEE Trans. on Vehicular Techn., 2025.

[26] Q. Li, M. Wen, and M. Di Renzo, "Single-RF MIMO: From spatial modulation to metasurface-based modulation," IEEE Wireless Commun., vol. 28, no. 4, pp. 88–95, 2021.

[27] Y.-C. Liang, Q. Zhang, J. Wang, R. Long, H. Zhou, and G. Yang, "Backscatter communication assisted by reconfigurable intelligent surfaces," Proc. IEEE, vol. 110, no. 9, pp. 1339–1357, 2022.

[28] N. Decarli, F. Guidi, and D. Dardari, "Passive UWB RFID for tag localization: architectures and design," IEEE Sensors J., vol. 16, no. 5, pp. 1385–1397, March 2016.

[29] J. C. Bucheli Garcia, M. Kamoun, and A. Sibille, "Low-complexity adaptive spatial processing of ESPAR antenna systems," IEEE Trans. Wireless Commun.", vol. 19, no. 6, pp. 3700–3711, Feb. 2020.

[30] D. Dardari "Frequency-selective Dynamic Scattering Arrays for Over-the-air EM Processing", Submitted to URSI-B EMTS 2025, Bologna, 2025, arXiv, https://arxiv.org/abs/2502.07336

[31] G. -H. Park, C. W. Byeon and C. S. Park, "60 GHz 7-Bit Passive Vector-Sum Phase Shifter With an X-Type Attenuator," in IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 70, no. 7, pp. 2355-2359, July 2023.

[32] J. Zhou, H. J. Qian and X. Luo, "A 9-Bit Vector-Sum Digital Phase Shifter Using High Resolution VGAs and Compensated Quadrature Signal Generator," 2019 IEEE MTT-S International Microwave Conference on Hardware and Systems for 5G and Beyond (IMC-5G), Atlanta, GA, USA, 2019.

[33] Y. Ertuğrul, C. Desset and S. Pollin, "Revisiting energy-efficient hybrid and digital beamforming architectures above 100 GHz," 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), Florence, Italy, 2023.

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

[35] IEEE, "IEEE Standard for Information technology–Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements-Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3:

Enhancements for Very High Throughput in the 60 GHz Band," IEEE Std 802.11ad-2012 (Amendment to IEEE Std 802.11-2012, as amended by IEEE Std 802.11ae-2012 and IEEE Std 802.11aa-2012), pp. 1–628, 2012.

[36] G. Berardinelli, "A centralized packet repetition scheme for dense in-X subnetworks", in preparation for submission to IEEE Internet of Things journal.

[37] 6G-SHINE D2.4, "In-X subnetwork architectures and integration into 6G 'networks of networks'", Feb. 2025. Available online at: https://6gshine.eu/wp-content/uploads/2025/02/6GSHINE_D2.4_Deliverable_v1.0_Disclaimer.pdf

[38] 6G-SHINE D5.2, "First implementation and evaluation of PoCs", Dec. 2024. Available online at: https://6gshine.eu/wp-content/uploads/2025/02/D5.2-First-implementation-and-evaluation-of-PoCs_v1.0_Disclaimer.pdf

[39] Thijs Havinga et al, Wi-Fi 6 Cross-Technology Interference Detection and Mitigation by OFDMA : an Experimental Study, https://arxiv.org/abs/2503.05429

[40] IEEE 802 LAN/MAN Standards Committee. "IEEE Standard for Information technology-Telecommunication and information exchange between systems-Local and metropolitan area networks-Specific requirements Part11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) ", Revision of IEEE Std 802.11-2020

[41] J. Liu, R. Porat et al., "TGax Channel Model Document," IEEE 802.11-14/0882r4, September 2014.

[42] M. Aslam, X. Jiao, W. Liu, M. Mehari, T. Havinga, and I. Moerman, "A novel hardware efficient design for IEEE 802.11ax compliant OFDMA transceiver," Computer Communications, vol. 219, pp. 173–181, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366424000926

[43] 3GPP TR38.824, "Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC) (Release 16)," v2.0.0, March 2019

[44] 3GPP TS38.214, "Physical layer procedures for data (Release 18)," v18.5.0, December 2024

[45] Aleksei Shashin, Andrey Belogaev, Artem Krasilov, Evgeny Khorov, Adaptive parameters selection for uplink grant-free URLLC transmission in 5G systems, Computer Networks, Volume 222, 2023, 109527, ISSN 1389-1286, https://doi.org/10.1016/j.comnet.2022.109527.

[46] R. B. Abreu, G. Pocovi, T. H. Jacobsen, M. Centenaro, K. I. Pedersen and T. E. Kolding, "Scheduling Enhancements and Performance Evaluation of Downlink 5G Time-Sensitive Communications," in IEEE Access, vol. 8, pp. 128106-128115, 2020, doi: 10.1109/ACCESS.2020.3008598.

[47] 3GPP TSG RAN WG1, "Enhanced UL transmission with configured grant for URLLC," 3GPP, Gothenburg, Sweden, Tech. Rep. R1-1809165, 2018, 3rd Generation Partnership Project (3GPP)

[48] Gurobi Optimization, LLC, "Gurobi Optimizer Reference Manual," 2024. [Online]. Available: https://www.gurobi.com

[49] R. De Gaudenzi, O. del Rìo Herrero, G. Acar, and E. Garrido Barrabés, "Asynchronous contention resolution diversity ALOHA: Making CRDSA truly asynchronous," IEEE Trans. Wireless Commun., vol. 13, no. 11, pp. 6193–6206, Nov. 2014.

[50] E. Sandgren, A. Graell i Amat, and F. Brännström, "On frame asynchronous coded slotted ALOHA: Asymptotic, finite length, and delay analysis," IEEE Trans. Commun., vol. 65, no. 2, pp. 691–704, Feb. 2017.

[51] Alessandro Mirri and Enrico Paolini, "Frame-Asynchronous Coded Slotted ALOHA with MDS Component Codes," IEEE WCNC 2025, Mar. 2025, accepted.

[52] 3GPP TS 36.521-1 V18.7.0 , "User Equipment (UE) conformance specification; Radio transmission and reception; Part 1: Conformance Testing, Release 18", December 2024.

[53] Reply LS on Interference Modelling for Duplex Evolution (Release 18), document 3GPP R4-2214376, Aug. 2022

[54] H. R. Hashempour, G. Berardinelli, R. Adeogun and E. A. Jorswieck, "Power Efficient Cooperative Communication Within IIoT Subnetworks: Relay or RIS?," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2024.3521001.

[55] W. Jiang and H. D. Schotten, "Neural network-based fading channel prediction: A comprehensive overview," IEEE Access, vol. 7, pp. 118112–118124, 2019.

[56] T. Peng, R. Zhang, X. Cheng, and L. Yang, "LSTM-based channel prediction for secure massive MIMO communications under imperfect CSI," in Proc. IEEE Int. Conf. Commun. (ICC), Dublin, Ireland, Jun. 2020, pp. 1–6.

[57] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in Proc. Adv. Neural Inf. Process. Syst., Cambridge, MA, USA, Dec. 2015, pp. 802–810.

[58] 3GPP TR 38.901, v17.0.0, "Technical Specification Group Radio Access Network; Study on channel model for frequencies from 0.5 to 100 GHz," 2022.

[59] 6G-SHINE D3.4, "RIS Management strategies for in-X subnetworks", to be released in June 2025. It will be available at: https://6gshine.eu/deliverables-ii/

[60] 3GPP, TS 38.300. V17.11.0, "NR; NR and NG-RAN Overall Description; Stage 2 (Release 18))", Release 17, 2024.

[61] Yungang Pan, et al., "Resource Optimization with 5G Configured Grant Scheduling for Real-Time Applications", in Proc. Design, Automation & Test in Europe Conference & Exhibition (DATE), Antwerp, Belgium, 2023, pp. 1-2.

[62] A. Larrañaga-Zumeta, M.C. Lucas-Estañ, J.Gozálvez, A. Arriola, "5G configured grant scheduling for seamless integration with TSN industrial networks", Computer Communications, 2024, 107930..

[63] M. Elsayem, et al., "Intelligent Resource Allocation for Grant-Free Access: A Reinforcement Learning Approach", IEEE Networking Letters, vol. 5, no. 3, pp. 154-158, Sept. 2023.

[64] A. Valcarce, et al., "The Role of AI in 6G MAC", in Proc. 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Antwerp, Belgium, 2024, pp. 723-728.

[65] European Technology Platform Networld Europe, "Smart Networks in the context of NGI", Technical Annex to Strategic Research and Innovation Agenda 2022-27, 2024.

[66] X. Hong, Y. Xi, and P. Liu, "Resource-aware online traffic scheduling for time-sensitive networking", IEEE Transactions on Industrial Informatics, vol. 20, no. 12, pp. 14267-14276, Dec. 2024.

[67] D. Yang, et al., "DetFed: Dynamic resource scheduling for deterministic federated learning over time-sensitive networks", IEEE Transactions on Mobile Computing, vol. 23, no. 5, pp. 5162-5178, May 2024.

[68] P. Prados-Garzon, et al., "LEARNET: Reinforcement learning based flow scheduling for asynchronous deterministic networks", in Proc. 2020 IEEE International Conference on Communications (ICC 2020), Dublin, Ireland, Jun. 7–11, 2020.

[69] Q. He, G. Dán and G. P. Koudouridis, "Semi-Persistent Scheduling for 5G Downlink based on Short-Term Traffic Prediction", in Proc. of the IEEE Global Communications Conference (GLOBECOM 2020), Taipei, Taiwan, 2020, pp. 1-6

[70] Y. Kaura, et al., "Adaptive Scheduling of Shared Grant-Free Resources for Heterogeneous Massive Machine type Communication in 5G and Beyond Networks", IEEE Trans. on Network and Service Management, Early Access, Nov. 2024.

[71] S. Cavallero, et al., "A new scheduler for URLLC in 5G NR IIoT networks with spatio-temporal traffic correlations", in Proc. IEEE International Conference on Communications (ICC 2023), Rome, Italy, May 28–Jun. 1, 2023.

[72] G. Di Gennaro, et al., "Decentralized Grant-Free mMTC Traffic Multiplexing With eMBB Data Through Deep Reinforcement Learning", IEEE Transactions on Machine Learning in Communications and Networking, vol. 2, pp. 1440-1455, Sept. 2024.

[73] D. Overbeck, N. A. Wagner, F. Kurtz and C. Wietfeld, "Proactive Resource Management for Predictive 5G Uplink Slicing", in Proc. of the 2022 IEEE Global Communications Conference (GLOBECOM 2022), Rio de Janeiro, Brazil, 2022, pp. 1000-1005.

[74] 6G IA, European Vision for the 6G Network Ecosystem, v2, Nov. 2024.

[75] L. Lusvarghi, et al., "Characterization of In-Vehicle Network Sensor Data Traffic in Autonomous Vehicles", in Proc. 2024 IEEE Vehicular Networking Conference (VNC), Kobe, Japan, May, 2024.

[76] A. Omer, et al., "Performance Evaluation of 5G Delay-Sensitive Single-Carrier Multi-User Downlink Scheduling", in Proc. IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Toronto, Canada, 2023, pp. 1-6.

[77] 3GPP, TS 22.186. V17.0.0, "Enhancement of 3GPP support for V2X scenarios", Release 17, 2022.

[78] M. Mokhtari, et al., "Modeling and system-level performance evaluation of sub-band full duplexing for 5G-advanced," *IEEE Access*, vol. 11, pp. 71503–71516, 2023.

[79] Cheng, Jing, and Chao Shen. "Relay-assisted uplink transmission design of URLLC packets." IEEE Internet of Things Journal 9.19 (2022): 18839-18853.

[80] Van Huynh, Dang, et al. "Multiple relay robots-assisted URLLC for industrial automation with deep neural networks." 2021 IEEE Global Communications Conference (GLOBECOM). IEEE, 2021.

[81] Ranjha, Ali, Georges Kaddoum, and Kapal Dev. "Facilitating URLLC in UAV-assisted relay systems with multiple-mobile robots for 6G networks: A prospective of agriculture 4.0." IEEE Transactions on Industrial Informatics 18.7 (2021): 4954-4965.

[82] Ren, Hong, et al. "Joint transmit power and placement optimization for URLLC-enabled UAV relay systems." IEEE Transactions on vehicular technology 69.7 (2020): 8003-8007.

[83] H. R. Hashempour, G. Berardinelli, R. Adeogun and E. A. Jorswieck, "Power Efficient Cooperative Communication Within IIoT Subnetworks: Relay or RIS?," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2024.3521001.

[84] W. Jiang and H. D. Schotten, "User Selection for Simple Passive Beamforming in Multi-RIS-Aided Multi-User Communications," in Proc. IEEE 97th Veh. Technol. Conf. (VTC2023-Spring), Florence, Italy, 2023, pp. 1–5.

[85] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," IEEE Trans. Veh. Technol., vol. 69, no. 12, pp. 14960–14973, Dec. 2020.

[86] C. Pan et al., "An Overview of Signal Processing Techniques for RIS/IRS-Aided Wireless Systems," IEEE J. Sel. Topics Signal Process., vol. 16, no. 5, pp. 883–917, Aug. 2022.

[87] B. Zheng and R. Zhang, "Intelligent reflecting surface-enhanced OFDM: Channel estimation and reflection optimization," IEEE Wireless Commun. Lett., vol. 9, no. 4, pp. 518–522, Jul. 2020.

[88] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," IEEE Trans. Commun., vol. 68, no. 7, pp. 4522–4535, Jul. 2020.

[89] 3GPP TR 38.901, v17.0.0, "Technical Specification Group Radio Access Network; Study on channel model for frequencies from 0.5 to 100 GHz," 2022.

[90] 3GPP, "Study on NR-based access to unlicensed spectrum," 3rd Generation Partnership Project (3GPP), TR 38.889, Release 15, May 2017

[91] 3GPP, "Physical layer procedures for shared spectrum channel access," 3rd Generation Partnership Project (3GPP), TS 37.213, Release 18, 2025

[92] T. -K. Le, F. Kaltenberger and U. Salim, "Dynamic switch between load based and frame based channel access mechanisms in unlicensed spectrum," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021, pp. 1-6, doi: 10.1109/GLOBECOM46510.2021.9685120.

[93] T. -K. Le, U. Salim and F. Kaltenberger, "Enhancements of PUSCH repetitions for URLLC in licensed and unlicensed spectrum," 2021 17th International Symposium on Wireless Communication Systems (ISWCS), Berlin, Germany, 2021, pp. 1-6, doi: 10.1109/ISWCS49558.2021.9562142.

[94] 3GPP, "NR; Physical layer procedures for data," 3rd Generation Partnership Project (3GPP), TS 38.214, Release 18, 2025

[95] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), TR 38.901, version 16.1.0, Release 16, 2020.

[96] 3GPP RAN1, "R1-23xxxxx FL Summary2 – XR Specific Enhancements – v026_ZTE_Mod," 3rd Generation Partnership Project (3GPP), RAN1#114, 2023.

[97] Başaran, Semiha Tedik, Güneş Karabulut Kurt, Murat Uysal, and İbrahim Altunbaş. "A tutorial on network coded cooperation." IEEE Communications Surveys & Tutorials 18, no. 4 (2016): 2970-2990.