This deliverable has not yet been externally reviewed and approved by the European Commission.

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**

D3.1 – PRELIMINARY RESULTS ON PHY AND MAC ENABLERS FOR IN-X SUBNETWORKS

Due date	29/02/2024	Delivery date	29/02/2024
Work package	WP3		
Responsible Author(s)	Fotis Foukalas (COGN)		
Contributor(s)	Baldomero Coll-Perales (UMH), Javier Gozalvez (UMH), Miguel Sepulcre (UMH), Hamid Reza Hashempour (AAU), Gilberto Berardinelli (AAU), Ramoni Adeogun (AAU), Davide Dardari (CNIT), Marina Lotti (CNIT), Alessandro Mirri (CNIT), Enrico Paolini (CNIT), Shin Horng Wong (SONY), Filipe Conceicao (IDE), Ognen Ognenoski (IDE), Thomas Jacobsen (Nokia), Meng Li (IMEC), Wei Liu (IMEC), Spilios Giannoulis (IMEC), Baiheng Chen (IMEC), Christos Tsakos (COGN), Jürgen Hupp (FHG)		
Version	V1.0		
Reviewer(s)	Renato Abreu (Nokia), Henrik Klessig (BOSCH)		
Dissemination level	Public		

Co-funded by the European Union

Horizon Europe Grant Agreement No. 101095738. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or SNS JU. Neither the European Union nor the granting authority can be held responsible for them.

VERSION AND AMENDMENT HISTORY

Version	Date (DD/MM/YYYY)	Created/Amended by	Changes
0.1	01/12/2023	Fotis Foukalas	First template
0.2	15/01/2024	Fotis Foukalas	First draft ready
0.7	31/01/2024	Fotis Foukalas	Version completed
0.8	15/02/2024	Renato Abreu, Henrik Klessig	Review completed
0.9	22/02/2024	Fotis Foukalas	Second round of review completed
1.0	28/02/2024	Fotis Foukalas	Final version

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS

Acronym	Description
3GPP	3rd Generation Partnership Project
AAL	antenna array-like
AD	autonomous driving
AI	artificial intelligence
AO	alternate optimization
AOA	angle of arrival
AOD	angle of departure
AP	access point
ATS	asynchronous traffic shaper
AWGN	additive white gaussian noise
BACKCOM	backscatter communication
BCD	block coordinate descent
BS	base station
BW	Bandwidth
BWP	bandwidth part
CBS	credit based shaper
CDF	cumulative distribution function
CG	configured grant
COT	channel occupancy time
CPE	customer premises equipment
CQI	channel quality indicator
CR	constraint relaxation
CSA	coded slotted ALOHA
CSI	channel state information
CSI-RS	channel state information – reference signal
DAC	digital-to-analog converter
DC	dual connectivity
DCI	downlink channel information
DCF	distributed coordinated function
DetNet	deterministic network
DF	decode and forward
DL	Downlink
DMA	dynamic metasurface antenna
DMRS	demodulation reference signal
DRX	discontinuous reception

DSSS	direct-sequence spread spectrum
E/E	Electrical/Electronic
ECU	Electronic Control Unit
EIRP	equivalent isotropic radiated power
EM	Electromagnetic
eMBB	enhanced mobile broadband
FBE	frame based equipment
FD	fully digital
FDD	frequency division duplexing
FDMA	frequency division multiple access
FEC	forward error correction
FFT	Fast Fourier Transform
FHSS	frequency hopping spread spectrum
FPGA	field programmable gate arrays
GNSS	global navigation satellite system
GOPS	giga operations per second
HARQ	hybrid automatic repeat request
HC	high-capacity device
HFC	hybrid fully connected
HPC	hybrid partially connected
HPCU	high-performance computing unit
IFFT	inverse Fast Fourier Transform
IMU	inertial measurement unit
ISM	industrial, scientific, and medical
IOT	internet-of-things
IO	indoor office
IVN	In-vehicle network
KKT	Karush-Kuhn-Tucker
LA-MAC	latency aware medium access control
LAD	latency aware deconfliction
LBE	load based equipment
LBT	listen before talk
LC	low-capability device
LDPC	low density parity check
LO	local oscillator
LOS	line-of-sight
LUT	look-up table
MAC	medium Access Control

MCOT	maximum channel occupancy time
MCS	modulation and coding scheme
MDS	maximum distance separable
MIMO	multiple-input multiple-output
MISO	multiple input single output
ML	machine learning
MLO	multi-link operation
mMTC	massive machine-type communication
MSE	mean square error
NACK	negative acknowledgment
NCC	network coded cooperation
NLOS	non-line-of-sight
NOMA	non orthogonal multiple access
NR-U	New radio unlicensed
NS-3	Network simulator 3
N-RIS	non-reconfigurable intelligent surface
OFDM	orthogonal frequency division multiplexing
PA	power amplifier
PAPR	peak-to-average power ratio
PDCCH	physical downlink control channel
PER	packet-error rate
PHY	physical layer
PLC	programmable logic controller
PMI	precoding matrix indicator
PSR	packet success ratio
PUCCH	physical uplink control channel
QAM	Quadrature amplitude modulation
QoS	quality of service
RAN	Radio Access Network
RB	resource block
RF	radio frequency
RI	rank indicator
RIS	reconfigurable intelligent surface
RRM	radio resource management
RRC	Radio resource control
RV	random variable
SCM	self-conjugating metasurface
SCS	sub-carrier spacing

SDG	sustainability development goal
SDP	semidefinite programming
SDR	semidefinite relaxation
SIC	successive interference cancellation
SPCA	sequential parametric convex approximation
SR	scheduling request
SINR	signal-to-interference-plus-noise ratio
SNE	subnetwork element
SNR	signal-to-noise ratio
SPS	Semi-persistent scheduling
SRE	smart radio environment
SSB	synchronization signal block
TA-CCA	time aware clear channel assessment
TAS	time aware shaper
TC	technology component
TDD	time division duplexing
TDMA	time division multiple access
TE	transversal electric
TM	transversal magnetic
TRP	transmission/reception points
TSN	time-sensitive network
UAV	unmanned aerial vehicles
UE	user equipment
UL	uplink
URLLC	ultra-reliable low latency communication
UTD	uniform theory of diffraction
UWIN	Ultra Reliable Wireless Industrial Network
WA	weight adaptation
WMMSE	weighted minimum mean square error
XR	extended reality

EXECUTIVE SUMMARY

Deliverable D3.1 presents the preliminary studies carried out in WP3 on physical layer and medium access control enablers for in-X subnetworks. Connection of the presented methods with the original technology components defined in the project proposal is first presented, along with an initial mapping with the use cases defined in WP2, the targeted KPIs/KVIs, and a discussion on their standardization potential.

Identified physical layer enhancements leverage the opportunities offered by operations at high carrier frequencies (i.e., mmWave, sub-THz) for improving throughput or reducing latencies. In particular, the potential of different antenna arrays is evaluated, and a novel solution for ultra-fast joint communication and beamforming based on self-conjugating metasurfaces is proposed. For the case of sub-THz spectrum, an analysis of the trade-off between energy consumption and throughput for different transceiver architectures is presented.

Medium access control solutions exploit characteristics of the services or deployments targeted by in-X subnetworks for reducing the overhead associated to traditional communication protocols; enhancements based on proximity grouping, configured grant and coded random access are proposed. Also, methods that leverage specific operational conditions (e.g., predefined device movements) or characteristics of data traffic for proactive and predictive resource allocation are introduced. Flexible duplexing is foreseen as a promising enabler for multiplexing services with different requirements by supporting different uplink/downlink switching points on a subband basis. Since the presence of blockage can jeopardize the possibility of achieving demanding communication requirements in a subnetwork, multi-link enhancements based on secondary nodes or network coded cooperation are proposed and evaluated. Also, enhancements to the access procedures for subnetworks operating in the unlicensed spectrum are presented.

This deliverable reflects the work in progress of the WP3 partners in Task 3.1 and Task 3.2. For each activity, an outlook to the coming work towards the final deliverable D3.3 is presented.

1 INTRODUCTION

This deliverable presents our preliminary studies related to the physical layer and medium access control solutions for in-X subnetworks. It is our hypothesis that the peculiar characteristics of in-X subnetworks (e.g., very short-range, devices in static or predictive motion) allow to cope with demanding communication requirements while limiting complexity and energy consumption.

Concepts and results presented in this deliverable are based on works in progress, and further findings and recommendations will be provided in the D3.3 deliverable. We remark that it is not in the scope of the project to provide a full and coherent PHY and MAC design for future in-X subnetworks, since this would require an effort which is beyond what the project has committed to. We rather focus on the individual technology solutions that are identified as the most promising "elements" for achieving the project objectives. A coherent system design for in-X subnetworks as part of the 6G 'network of networks' will rather be object of the phase 2 of the SNS work programme.

In line with the 6G-SHINE project proposal, we present how the methods described in this deliverable cope with the originally defined TCs. Also, we present an initial mapping of the methods to the use cases of the project, whose preliminary definition is presented in deliverable D2.1 [1], and whose refined definition is presented in deliverable D2.2 [2] (submitted at the same time as this deliverables). Since KVIs are at the core of the design of the 6G-SHINE technologies, we show how the designed technologies can address environmental, economic and -when relevant- social sustainability.

The initial portfolio of technology solutions presented in this deliverable aims at addressing objective 3 and objective 4 of the 6G-SHINE proposal, though their perspective in terms of applicability is diverse. Some of the presented technologies follow indeed an evolutionary path with respect to current 3GPP specifications and can be implemented as an incremental update, tailored to the specific characteristics and requirements of in-X subnetworks; other solutions have instead a longer timespan as they require novel hardware and/or processing capabilities, and can be seen as potential candidates for 3GPP Release 20 and beyond. An outlook of the standardization potential of the proposed methods is presented.

Section 2 of the deliverable focuses on the physical layer enhancements for in-X subnetworks. We first discuss the potential of having antenna panels at both transmitter and receiver ends for in-X subnetworks, for the sake of reducing energy consumption while guaranteeing high spectral efficiency. The novel concept of self-conjugating meta surfaces is then introduced, providing an efficient solution for MIMO communication using ultra-low complexity and power as well as fast setup time. Since communication within in-X subnetworks is predominantly short-range, this makes high carrier frequencies, such as in the sub-THz region, a promising option to explore. In particular, throughput and energy efficiency characteristics are explored for sub-THz transceivers with different architectures. As jamming is identified as a major threat for achieving high reliability, we present our vision for improving robustness to such attacks by a native system design and detector enhancements.

Section 3 focuses on the novel medium access protocols and algorithms for in-X subnetworks. We first introduce protocol enhancements tailored to the achievement of lower latencies than what currently supported by the 3GPP specifications, based on proximity grouping and configured grant extensions. We then present enhancements to the paradigm of coded random access, as a fully grant-free solution that can reduce latency in in-X subnetworks serving many devices. In-X subnetworks are expected to multiplex services with diverse requirements (e.g., video feeds together with periodic low latency traffic), via a tailored solution design. Approaches based on flexible duplexing are introduced. Since latency and signalling overhead can also be reduced via prediction, we present solutions for predictive scheduling either based on context information or knowledge of underline application and services. We present different approaches for leveraging cooperative communication for the sake of improving

reliability in a subnetwork, based on the presence of secondary nodes and network coded cooperation among devices. Finally, we introduce novel approaches for reducing communication latency in subnetworks operating over the unlicensed spectrum, meant at overcoming the limitations of 'listen-before-talk' medium access.

Finally, **Section 4** presents a concise summary of the initial studies.

Research presented in this deliverable cover Task 3.1 and Task 3.2 described in the project proposal. The mapping of the presented methods with the specific subtasks described in the proposal is the following (note that some of the methods can be mapped to more than one subtask):

- Subtask 3.1a: Radio interface enablers for ultra-reliable, ultra-short transmissions in subnetwork scenarios.
    - Metasurfaces for short-range joint beamforming and communication
    - Proximity Grouping for Pre-emptive retransmissions & Preventive scheduling
    - Configured grant enhancements
- Subtask 3.1b: Intra-subnetwork macro-diversity and cooperative schemes.
    - Cooperative in-X communication
    - Network coded cooperation
- Subtask 3.1c: Jamming-robust native PHY design.
    - Native-robust design
    - Baseband receiver improvement
- Subtask 3.1d: Native support of low complexity beam forming and beam aware scheduling.
    - Antenna arrays/panels for in-X subnetworks
    - Metasurfaces for short-range joint beamforming and communication
    - AI/ML-based prediction for beam management
- Subtask 3.1e: Sub-THz PHY System Models and Abstractions
    - Sub-THz transceiver architecture and analysis for subnetworks
- Subtask 3.2a: Enhanced multiple access procedures for service multiplexing and reduced latencies.
    - Flexible/full duplex – capable nodes
    - Deterministic and Predictive Traffic Scheduling for In-Vehicle Networks
- Subtask 3.2b: Proactive and predictive interference-robust scheduling algorithms.
    - Mobility-aware predictive scheduling in industrial subnetworks
    - AI/ML-based prediction for beam management
    - Coded random access
    - Deterministic and Predictive Traffic Scheduling for In-Vehicle Networks
- Subtask 3.2c: Novel access schemes in unlicensed spectrum for in-X subnetworks.
    - Latency-aware access in unlicensed bands
    - Adaptive Low latency and reliable unlicensed spectrum access

1.1 Relevance of the presented methods in the 6G-SHINE context

Research conducted in WP3 addresses objective 3 and 4 of the project, reported below:

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.*

In the proposal, we have identified 16 technology components (TCs) to be investigated in 6G-SHINE. The list of TCs is the following:

TC1. In-X data traffic models

TC2. Channel models for in-X scenarios

TC3. Sub-THz system model

TC4. Ultra-short transmissions with extreme reliability

TC5. Analog/hybrid beamforming/beamfocusing

TC6. Jamming-aware native PHY design

TC7. RIS enhancements

TC8. Intra-subnetwork macro-diversity

TC9. Flexible/full duplex scheduler

TC10. Predictive scheduler

TC11. Latency-aware access in the unlicensed spectrum

TC12. Centralized radio resource management

TC13. Distributed/hybrid radio resource management.

TC14. Jamming detection and mitigation

TC15. Hybrid management of traffic, spectrum, and computational resources

TC16. Coordination of operations among subnetworks in the same entity

Research in WP3 covers TC3-TC11, and the preliminary work presented in D3.1 covers all these TCs excluding TC7, which is addressed by the dedicated deliverable D3.2. In Table 1, we present a list of the technology/methods studied in this deliverable, and the connection with the original TCs.

TABLE 1. CONNECTION OF THE STUDIED METHODS WITH THE 6G-SHINE TCS.

Technology/method	6G-SHINE TCs
Antenna arrays/panels for in-x subnetworks	TC5
Metasurfaces for short-range joint beamforming and communication	TC4, TC5
Sub-THz transceiver architecture and analysis for subnetworks	TC3
Jamming-robust PHY design – native robustness	TC6
Jamming-robust PHY design – baseband receiver improvement	TC6
Proximity Grouping for Pre-emptive retransmissions & Preventive scheduling	TC4, TC10
Configured grant enhancement	TC4
Coded Random Access	TC4, TC11

Flexible/full duplex – capable nodes	TC9, TC4
AI/ML-based prediction for beam management	TC5, TC10
Mobility-aware predictive scheduling	TC4, TC10
Deterministic and Predictive Traffic Scheduling	TC10
Cooperative in-X communication	TC8
Network coded operations	TC8
Latency-aware access in unlicensed bands	TC11
Adaptive Low latency and reliable unlicensed spectrum access	TC11

PHY and MAC methods in 6G-SHINE are meant to provide support for achieving unprecedented radio performance requirements leveraging the advantageous short-range radio propagation characteristics.

Table 2 describes the main KPIs and KVIs targeted by the presented methods. It is worth mentioning that, since 6G-SHINE is a low technology readiness level (TRL) project, we do not aim at directly measuring the impact of the designed solutions in terms of KVIs, as such assessment will only be possible once the designed solutions are implemented and integrated in a coherent system design, which is beyond the scope of the project. Still, KVIs are at the center of our technology design, and we speculate how our solutions can be the basic "bricks" for addressing relevant KVIs. A description on how 6G-SHINE research addresses environmental, economic and social sustainability is included in deliverable D2.2. Given the nature of the research in physical layer and medium access control, the methods presented in D3.1 are prevalently addressing environmental and economic sustainability.

TABLE 2. KPIS AND KVIS TARGETED BY THE PRESENTED METHODS.

Technology/method	Main target KPIs	Targeted KVIs
Antenna arrays/panels for in-X subnetworks	High data rate (e.g. > 80Mbps), short beam misalignment time (e.g. <1ms with $1\text{-}10^{-3}$ probability of success), low energy consumption	Improved environmental sustainability due to lower energy consumption on the devices.
Metasurfaces for short-range joint beamforming and communication	Latency <500 µs, devices (e.g., sensors) with ultra-low complexity (no RF chains, no DAC), extremely low signalling overhead, sensors with extremely low energy consumption, ultra-short packet transmission (down to 32 bytes).	Improved environmental sustainability thanks to lower energy consumption enabled by the high-energy efficient sensors. Potentially, sensors might be power supplied using energy harvesting techniques (no battery)
Sub-THz transceiver architecture and analysis for subnetworks	No specific target KPIs. Targeting more accurate power estimation for different transceiver architecture.	Improved environmental sustainability thanks to a lower power consumption enabled by optimized transceiver architectures.
Jamming-robust PHY design – native robustness	Error probability in the order of 10^{-6} and below	Improved economic sustainability thanks to the jamming-resilient

Jamming-robust PHY design – baseband receiver improvement	Error probability in the order of 10^{-6} when SNR > 5 dB	operation in e.g., industrial subnetworks, reducing operation downtime.
Proximity Grouping for Pre-emptive retransmissions & Preventive scheduling	Low latency (< 1 ms) and high reliability, with error probability $<10^{-6}$.	Improved economic sustainability due to highly efficient transmission, allowing to reduce operational delays in e.g., industrial manufacturing.
Configured grant enhancement	Error probabilitiies < 10^{-6}.	Improved environmental sustainability due to lower energy consumption of the nodes.
Coded Random Access	Low-latency (<500 µs) and 10^{-6} probability of two consecutive errors. Up to 50 sensors and nodes supported.	Improved environmental sustainability due to lower energy consumption of the nodes.
Flexible/full duplex – capable nodes	High data rates (> 80 Mbps) flows multiplexed with traffic with short cycle time (sub-ms), and 10^{-6} error probability	Improved economic sustainability thanks to the capability of scheduling device with very diverse requirements by a same node, without additional infrastructure.
AI/ML-based prediction for beam management	Beam misalignment time, low complexity for the algorithm and short-time intervals for measurements, and low probability of prediction error (method is generic, and no quantitative benchmarks)	Social sustainability by facilitating full automation of systems, environment sustainability via low complexity improving energy consumption.
Mobility-aware predictive scheduling	Low latency (< 1 ms) and high reliability, with error probabilities $<10^{-6}$.	Improved economic sustainability due to highly efficient transmission. Improved environmental sustainability due to extreme reliability and low latency enabling it to replace wired connections, thereby reducing cabling, weight, easier to install and manufacture.
Deterministic and Predictive Traffic Scheduling	KPIs emerging from the supported automotive systems (e.g. ADAS, autonomous driving, etc) whose requirements include: latency < 1ms, bandwidth from Mbps to Gpbs, reliability > 99%, determinism, and control loop times in the order of ms.	Improved economic and environmental sustainability thanks to the gained flexibility in service provisioning and configuration by replacing rigid cables with configurable predictive scheduler.
Cooperative in-X communication	Sub-ms cycle times with error probability < 10^{-6} for at least 20 devices	Improved environmental sustainability due to lower energy consumption enabled by the cooperative nodes

Network Coded Operation	Sub-ms cycle times with packet loss probability < 10^{-6} for at least 20 devices	Improved environmental sustainability due to higher efficiency of error control and thus lower energy consumption enabled by the network coded cooperative nodes
Latency-aware access in the unlicensed spectrum	Lower Medium access delay for priority traffic by 50%, packet success ratio same or better than SoTA baseline, end-to-end latency reduction by 50%	Improved economic sustainability as unlicensed band networks will be able to offer also near-real-time network services, dismissing the need for specialized duplicate deployments. Also using an unlicensed spectrum should be more affordable in terms of equipment cost and no fees for using the spectrum, as long as regulatory constraints are respected.
Adaptive Low latency and reliable unlicensed spectrum access	Error probabilitiies < 10^{-6}.	Improved environmental sustainability due to lower energy consumption of the nodes.

Table 3 presents the mapping of the presented methods to the use case categories (and specific use cases) as defined in deliverable D2.1. We remark that, we do not aim at evaluating each presented method for all the mapped use cases. In our performance evaluation, we rather highlight the main use case or use case category of interest, as we believe the extension to a different use case with similar KPIs is straightforward. This deliverable is being submitted at the same time as D2.2 deliverable, where a more detailed description of use cases and requirement is provided. Also, radio propagation models for in-X subnetworks are being developed in WP2 and will be presented in deliverable D3.3, with expected submission in June 2024. Therefore, some of the methods have still been evaluated according to prior assumptions in terms of use case scenarios and radio propagation characteristics. Future work in WP3 will further address the capabilities of the proposed methods of addressing the use case scenarios and requirements as defined in deliverable D2.2 and will also leverage the channel models presented in D3.3.

TABLE 3. MAPPING BETWEEN PRESENTED METHODS AND USE CASES AS DEFINED IN D2.1 [1]

Technology/method	**Main use case category(ies)**	**Relevant use cases**
Antenna arrays/panels for in-x subnetworks	Industrial	Robot control
	Consumer	Immersive education
Metasurfaces for short-range joint beamforming and communication	Industrial	Robot control, Unit test cell
Sub-THz transceiver architecture and analysis for subnetworks	Consumer	XR indoor gaming

Jamming-robust PHY design – native robust design	Industrial	Robot control
	In-vehicle	Wireless Zone ECU, Collaborative Wireless Zone ECU, Inter-subnetwork Coordination, Virtual ECU
	Consumer	AR navigation
Jamming-robust PHY design – baseband receiver improvement	Industrial	Robot control
	In-vehicle	Wireless Zone ECU, Collaborative Wireless Zone ECU, Inter-subnetwork Coordination, Virtual ECU
	Consumer	AR navigation
Proximity Grouping for Pre-emptive retransmissions & Preventive scheduling	Industrial	Robot control
	Consumer	Immersive education, Indoor interactive gaming
Configured grant enhancement	Industrial	Robot control
Coded Random Access	Industrial	Robot control, Robot swarms, Unit test cell
Flexible/full duplex capable nodes	Industrial	Robot control, Unit test cell, Visual inspection cell
	In-vehicle	Wireless Zone ECU, Collaborative Wireless Zone ECU
	Consumer	All use cases
AI/ML-based prediction for beam management	Consumer and industrial	Immersive education, Augmented Reality Navigation, Robot control, Co-existence in factory hall
Mobility-aware predictive scheduling	Industrial	Robot control
	Consumer	Immersive education, Indoor interactive gaming
Deterministic and Predictive Traffic Scheduling	In-vehicle	Wireless Zone ECU, Collaborative Wireless Zone ECU, Virtual ECU
Flexible/full duplex–capable nodes	Industrial	Robot control, Unit test cell, Visual inspection cell
	In-vehicle	Wireless Zone ECU, Collaborative Wireless Zone ECU
	Consumer	All use cases
Cooperative in-X communication	Industrial	Robot control, Unit test cell, Visual inspection cell, Robot swarms
	In-vehicle	Cooperative wireless zone ECU
	Consumer	Immersive education
Network coded operations	Industrial	Robot control, Unit test cell, Visual inspection cell, Coexistence in factory hall

Latency-aware access in unlicensed spectrum	Industrial	Robot control, Robot swarms
Adaptive low latency and reliable unlicensed spectrum access	Industrial	Robot control, Unit test cell, Visual inspection cell, Coexistence in factory hall

An outlook of the standardization potential of the proposed methods is presented in Table 4. Some of the envisioned methods can be seen in an evolutionary path with respect to current 3GPP methods, while others have not yet a direct connection with coming 3GPP work but may have a longer-term impact (Release 20 and beyond) as they require fundamental changes on the operational procedures of wireless systems.

TABLE 4. STANDARDIZATION POTENTIAL OF THE PRESENTED METHODS

Technology/method	Standardization potential
Antenna arrays/panels for in-X subnetworks	The concepts developed within this technology component has potentially 3GPP RAN1 and RAN2 impacts when considering beam management and beam forming for device-to-device technologies.
Metasurfaces for short-range joint beamforming and communication	This method has potential relevance for 6G standardization, 3GPP RAN 1 Release 20 and beyond, with particular reference to backscatter-based communications for IoT scenarios.
Sub-THz transceiver architecture and analysis for subnetworks	This is not relevant for standardization since it only concerns transceiver design and analysis; however, this work can contribute to ETSI THz group.
Jamming-robust PHY design – native robust design	This concept has potential relevance for 3GPP RAN 1 and RAN2, Release 20 and beyond. Considering a jamming-robust design can be a native condition for 6G physical layer design, at least for critical services. Also, the hopping protocol with repetition must be standardized.
Jamming-robust PHY design – baseband receiver improvement	This technology component does not have obvious link with standardization, since it concerns receiver implementation.
Proximity Grouping for Pre-emptive retransmissions & Preventive scheduling	Potential 3GPP RAN1 and RAN2 impacts, notably Layer 1 & RRC signaling to perform proximity grouping, maintaining the group and providing scheduling for pre-empting retransmissions.
Configured grant enhancement	URLLC enhancements have standardization potential for 3GPP Release 19, in particular RAN1 solutions.
Coded random access	This method can potentially be integrated already into 5G NR where the transmission protocol involves the organization of time into frames and slots. Since 5G NR allows the transmission to start at any symbol in the frame ('mini-slot' transmission), the proposed sliding window decoding technique fits the standard and can be further exploited to enhance performance. Still, this solution will rather have relevance for 6G standardization, as a solution for improving latency and reliability of grant-free transmissions.
Flexible/full duplex–capable nodes	Since in-band flexible duplexing is already discussed in the 3GPP Release 18 NR study item on duplexing evolution, this method can have potential already for Release 19 standardization. Still, in the context of subnetworks, the method has higher potential in Release 20 and beyond, RAN1, as new

	frame formats and resource grids might need to be defined to support the enhanced duplexing operations.
AI/ML-based prediction for beam management	This method targets at standardisation for 3GPP Release 19+, in particular for 3GPP RAN1 and RAN2, topics related to coverage and capacity optimisation. It relates to current AIML studies and work items such as: AIML for air interface (SI and WI), AIML for NG-RAN (WI). The method is based on monitoring cross layer criteria and measurements, evaluation of AIML model performance, and identifying optimal datasets for coverage and capacity optimisation.
Mobility-aware predictive scheduling	Potential 3GPP RAN1 and RAN2 impacts for new Layer 1 & RRC signalling to providing new trigger and positions of interest, e.g., UE/AP to indicate that UE is within T_{move} ms from a position of interest.
Deterministic and Predictive Traffic Scheduling	Potential 3GPP RAN1 and RAN2 impact for Release 20 and beyond, including integration of TSN and wireless for the end-to-end support of time sensitive traffic, full duplexing (see next entry in this table), efficient configure grant schemes, and more flexible frame formats and resource grids.
Cooperative in-X communication	This method has potential relevance for 3GPP RAN1 and RAN2, Release 20 and beyond. The method may require indeed tailored reference sequences in order to acquire the channel state information for all the involved links in the training phase. Also, it requires a tailored protocol definition for exploiting the presence of the relay nodes.
Network coded cooperation	This method has potential relevance for 3GPP RAN1 and RAN2, Release 20 and beyond.
Latency-aware access in unlicensed spectrum	This method can be submitted as a possible enhancement in the next iteration of update of the ETSI ISM access rules standards like ETSI EN 300.328 for 2GHz ISM band or ETSI EN 301.893 for 5GHz ISM band
Adaptive Low latency and reliable unlicensed spectrum access	This method has potential relevance as NR-U enhancement for URLLC in 3GPP RAN1 and RAN2.

Figure 1 clusters the presented methods in terms of their expected short-, medium- and long-term impact. Note that some of the methods are pure PHY or MAC enhancements, while others resides in between as they require joint PHY and MAC design.

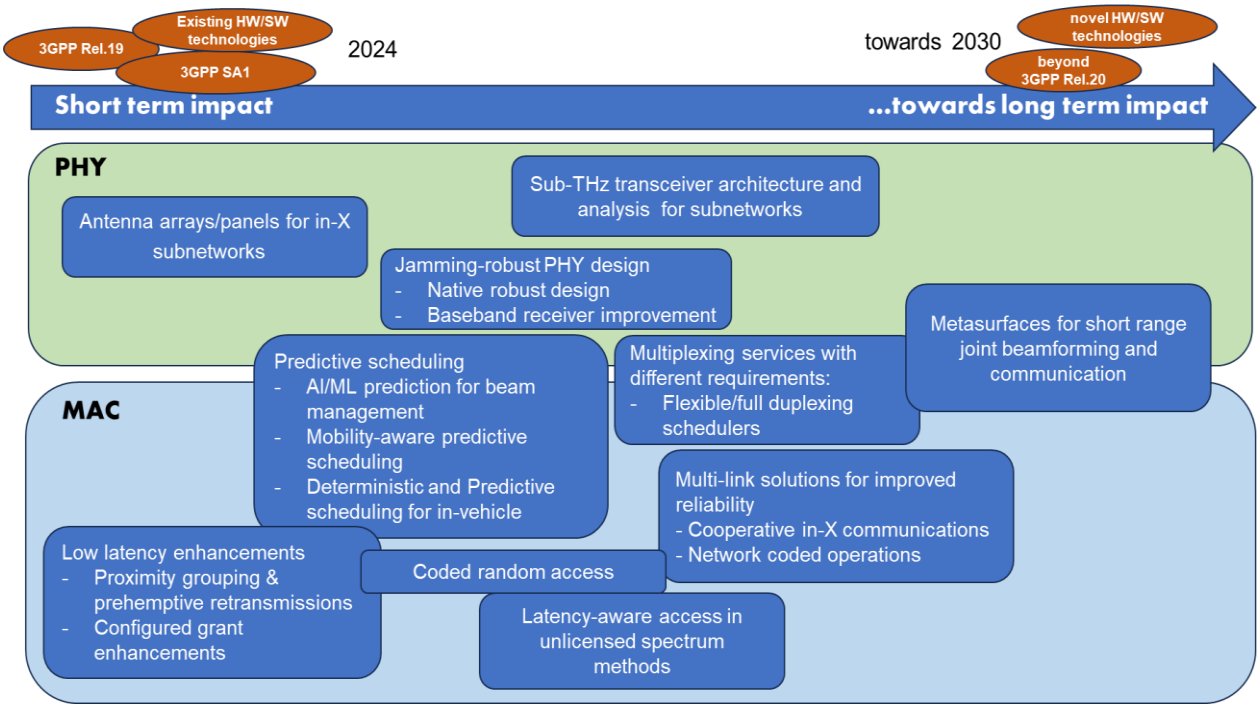


FIGURE 1. PERSPECTIVE ON THE POTENTIAL IMPACT OF THE PRESENTED METHODS.

We foresee that enhancements for optimizing antenna arrays and panels for relevant in-X subnetworks use cases are feasible in the short term, as they rely on existing technologies and are already well defined in current 3GPP standards. Design of a jamming-robust physical layer has also solid roots in current technology methods (e.g., packet repetitions, channel hopping) but they may require a longer timespan to be extensively discussed in the 3GPP. The adoption of sub-THz transceiver is also foreseen in a longer term, along with the usage of metasurfaces for joint communication and beamforming, as they require novel hardware and subsume a major change in the way communication protocols are designed.

The low latency enhancements based on proximity grouping and configured grant, are incremental with respect to the current 3GPP specification. Predictive schedulers may also require incremental updates in terms of enabling signalling, but significant effort in terms of algorithm development. A longer timespan is also foreseen for the eventual adoption of coded random-access techniques in subnetworks, though its basic concept can in principle be supported by current 5G numerology and frame structure. Solutions exploiting multi-link enhancements may require a novel definition of subnetwork roles and can be seen feasible for Release 20 and beyond. Proposed methods for latency-aware unlicensed access may require regulation updates for what concerns access rules.

In the description of the methods, we will adopt the nomenclature currently being defined in WP2, referred to the relevant subnetwork components. For the work carried out in D3.1, the following elements are relevant:

- Element with High Capabilities (HC). An element with high capabilities is a device/node with increased capabilities in terms of networking and computation. Such a node might act as the central communication node in a subnetwork and also might offer compute resources to other devices in the subnetwork. Multiple such HCs can be installed in a single subnetwork. An HC device can be a user equipment as defined by 3GPP or a non-3GPP device.

- Element with Low Capabilities (LC). An element with low capabilities is similar to an HC but has limited capabilities in terms of networking and computation. This can reduce the functionalities this device provides to the subnetwork and even there might be no connection to the 6G base station. In a hierarchical or nested subnetwork, the LC might act as an aggregator. An LC device can be a user equipment as defined by 3GPP or a non-3GPP device.
- Subnetwork Element (SNE). Subnetwork elements are computationally constrained devices that have limited form factor, cost footprint, and include devices such as sensors/actuators. A SNE device can be a user equipment as defined by 3GPP or a non-3GPP device.

We refer to deliverable D2.2 for a thorough description of all subnetwork elements [2].

2 ROBUST AND LOW COMPLEXITY PHY ENHANCEMENTS

This section presents the initial work done in the context of Task 3.1, Robust and low complexity PHY enhancements. Short-range communication makes high carrier frequencies -mmWave and sub-THz- a promising solution for achieving demanding communication requirements. Most of the presented methods in this section refer to the usage of antenna arrays operating at these frequencies. In particular, the potential of antenna array and panels for in-X subnetworks operating at mmWave bands are presented, along with the novel concept of self-conjugating metasurfaces for short range joint communication and beamforming, that capitalizes from electromagnetic processing for a low cost and low signalling overhead radio design. An evaluation of sub-THz transceiver architectures, considering throughput and energy efficiency trade-offs, is also presented. Moreover, initial discussions and findings related to a jamming robust native PHY design and related receiver enhancements are included.

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

Beamforming/ beam focusing is identified as one of the TCs of 6G-SHINE (specifically, TC5) and is a potential solution for increasing the radio resource spectral efficiency. In this section, the focus is on beamforming and its application in-X subnetworks. Specifically, we discuss how and when beamforming can be considered a promising solution for in-X subnetworks, and what are the related challenges.

2.1.1 Beamforming for in-X subnetworks

Beamforming is traditionally considered a coverage enhancement technique, that in wide-area cellular networks is a prerequisite at high carrier frequencies (e.g. mmWave) as it can increase signal power in a particular direction, thereby compensate for the increased path loss. For In-X subnetworks the motivation for its use in is rather to increase spectral efficiency, reduce power consumption, reduce signal power to keep interference at a minimum, improve resilience to jamming and strengthening privacy by making it harder to intercept and eavesdrop the subnetwork communication links. High frequency bands such as mmWave spectrum region are promising for in-X subnetworks given the short range and the availability of a large spectrum.

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 analog beamformer, digital beamformer and/or hybrid beamformer. Analog beamformers are relatively simple by design, as they work by applying a phase-shifter in front of the radio frequency chain, giving a finite set of pre-configured phase-shifts, which may only be active

one at a time. This makes the task of beam alignment a matter of finding the best phase-shift among the pre-configured shifts for the sake of providing the best link quality. In a digital beamformer, which is envisioned to be the more frequently used in the 6G era, a communication link may comprise of multiple radio frequency chains that enable the formation of the beams in digital domain, hence providing additional freedom in forming the beams to the direction(s) with the best link quality. Further, they are not constrained to a single beam at a time. This additional freedom comes with a cost in terms of hardware cost, power consumption and signalling processing complexity. A middle ground between the two is a hybrid beamformer, where one or more radio frequency chain(s) can be associated to one or more analog phase-shifter(s), hence reducing hardware cost compared to a fully digital beamformer and processing complexity, but still increased compared to the analog beamformer.

For in-X subnetworks, at least two types of devices are considered; the first type is a subnetwork device which can take the role as a gateway or scheduling entity for an entire subnetwork and the second type is a subnetwork device which is served by the first type of devices. The first type can also be called an access point or high capability (HC) subnetwork device, following the nomenclature defined in D2.2 [2]. The second device type can be called a subnetwork entity (SNE) or low capability (LC) subnetwork device. Both device types are an enabler of each other, i.e., 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 constrains. Regarding antennas, it is straightforward to consider that a LC device can be equipped with relatively few 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, 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 high capability subnetwork entity can be a 6G smartphone, a pair of eXtended Reality (XR) glasses, or a 6G customer premises equipment (CPE), whereas examples of a low-capability subnetwork device could be a sensor/actuator.

There are plenty of 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 beamforming 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 [3] and up to 8 dB for a 4 element antenna array [3][4]. 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 [4]. A reference model of a typical smartphone placement with four antenna panels from [5] is shown in Figure 2, and in Figure 3 antenna panel placement for XR glasses are 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

4) 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. This is visualized in Figure 4 where the antenna radiation pattern without and with a human hand grip is shown [3]. 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 [3][5], 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 wires, robotic arms, human hands or human arms) 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.

2.1.2 Performance evaluation

From the above description of factors that may impact the beamforming gains, we learn that evaluation of beamforming gains would need to account for at least the antenna design and implementation, potential interactions to the devices and potential interactions with the environment.

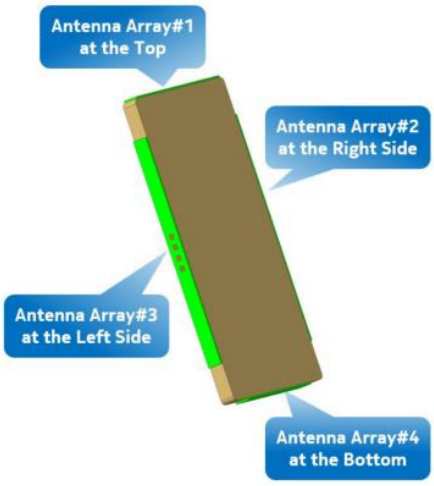


FIGURE 2. ANTENNA PANEL PLACEMENT EXAMPLES FOR A SMARTPHONE. PICTURE FROM [5].

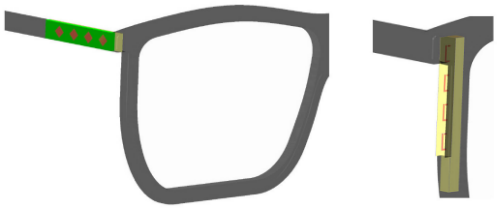


FIGURE 3. POSSIBLE PLACEMENT OF ANTENNA ARRAYS ON XR GLASSES. PICTURE FROM [5].

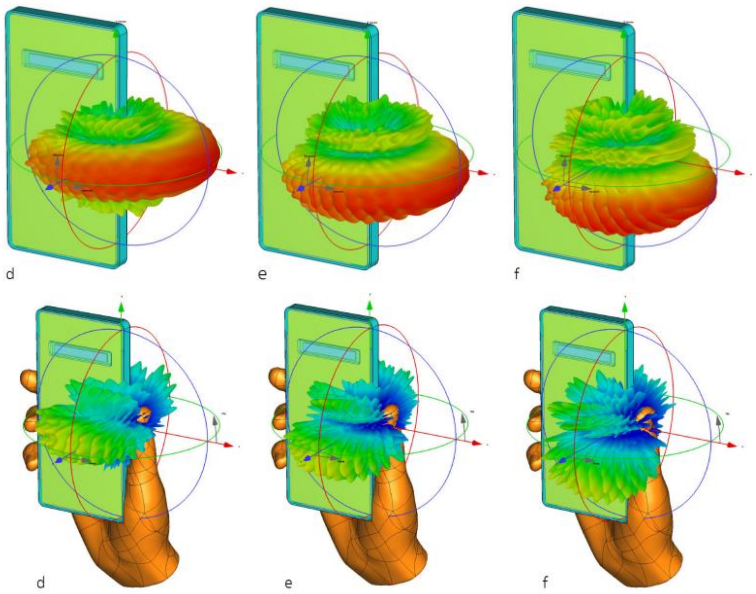


FIGURE 4. PROPAGATION PATTERN WITH THE REFERENCE SMARTPHONE, WITHOUT (D,E,F IN THE TOP) AND WITH (D, E, F IN THE BOTTOM) A HUMAN HAND INVOLVED AND FOR BEAMS STEERED TOWARDS 0, 15, 30 DEG (D,E,F RESPECTIVELY). PICTURE FROM [5].

In an attempt to shed light on the question of potential benefits in equipping subnetwork devices with beamforming capabilities, a two-stage approach has been taken. Firstly, evaluating the signal-to-interference power ratio improvements in a static deployment and considering ideal beam selection, and secondly, considering a beam alignment procedure with mobile subnetworks and subnetwork entities within it. The reason is that the first stage will give an idea of the magnitude of gains to be expected and serves as input to the configurations to be considered in the second stage, which will then introduce the overhead of beam management. This section focuses on the first stage and leaves the second stage for further study.

Two scenarios have been chosen to represent a selection of the in-X subnetwork use case scenarios [1], an indoor factory scenario for the robot control use case, and an indoor classroom scenario for the immersive education use case. In the indoor factory, a set of robot units are deployed, and each is equipped with a high capability subnetwork device and multiple subnetwork entities acting as sensors and actuators within the robot. The sensors and actuators connected via the high capability subnetwork device are assumed to be statically mounted within the robot. The other scenario, the indoor classroom, consists of multiple pupils, each assumed to operate its own subnetwork which is equipped with at least one HC and at least one LC or SNE. The subnetwork device, which can

be either an LC or an SNE, could be a pair of XR glasses. Additional SNEs can be considered as well, which could be sensors mounted around the body providing measurements to be used when compiling the XR domain content. The indoor classroom scenario used for immersive education can be assumed to be very dynamic, meaning that both the pupils and the devices on the pupils will be moving around. Further, it can be assumed that there are far less obstacles in the indoor classroom scenario than in an indoor factory scenario particularly impacting the link quality between subnetworks (inter-subnetworks).

The evaluation assumptions are summarized in Table 5. Note that the modelling assumptions for traffic and radio channels are taken from existing sources; in the future studies, they will be aligned with the assumptions provided in deliverables D2.2 (submitted at the same time as this deliverable) and D2.3 (with expected submission in June 2024).

TABLE 5. EVALUATION PARAMETERS USED TO STUDY THE POTENTIAL OF BEAM FORMING.

Parameter	**Value**
Scenario	A single room of 20 m x 20 m x 3 m (length x width x height) for both indoor classroom and indoor factory. 2 or 5 or 10 subnetworks are deployed in the scenarios. Each subnetwork consists of: - Indoor classroom: 1 high capability entity and 1 subnetwork entity (could be a pair of XR glasses) - Indoor factory: 1 high capability entity and 3 subnetwork entities (could be sensors/actuators)
Subnetwork deployment	The HC is always deployed in the center of the subnetwork, other subnetwork entities (LC and/or SNEs) are deployed at a random location within a circle around the HC. - Indoor factory: Circle diameter of 2.5 m. Minimum distance between subnetworks are 0.5 m (i.e. no overlap of subnetworks). Subnetwork entities deployed in a height between 0.5 m to 1.5 m, with the HC at a height of 1 m. - Indoor classroom: Circle diameter of 1 m. Subnetwork entities are deployed in a height between 1.4 m and 1.8 m and the height of the HC is between 0.7 m and 1.2 m.
Traffic modelling	- Indoor classroom: XR traffic following 3GPP TR 38.835. From XR glasses to the high capability entity a periodicity of 4 ms and a size of 100 B. From the high capability entity to the XR glasses a video stream modelled as an interarrival rate between 4 ms and 16.67 ms with a packet size between 32 kB and 93 kB following a truncated Gaussian. - Indoor factory: Periodic transmissions of 32 B packet with a periodicity of 1 ms both to and from the HC and the other subnetwork entities.
Channel model	- Indoor classroom: Indoor office (InH) from 38.900. - Indoor factory: Indoor factory model (InF) from 38.901 in the dense clutter low height (DL) configuration.
Antenna panel locations	2 panels: Panel#1-#2 placed as per Figure 2 4 panels: Panel#1-#4 placed as per Figure 2
Panel configuration	Configuration 1: Theoretical reference configuration with a monopole antenna (referred to as omni) with 0dB antenna gain. Configuration 2: 1x1x2 (MxNxP): 1 beam per panel at 0deg. Configuration 3: 1x4x2 (MxNxP): 4 beams configured and equally spaced from -56deg to +56deg

	Configuration 4: 1x16x2 (MxNxP): 16 beams configured equally spaced from -60deg to +60deg Note 1: All antenna elements are having a per antenna element gain of 5dB. Note: 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.
Antenna implementation loss scenarios	Scenario 0: Ideal antenna implementation Scenario 1: 4 dB antenna implementation loss added to all antenna configurations. Scenario 2: 4 dB antenna implementation loss added to antenna configuration 1 and 6dB antenna implementation loss added to antenna configuration 2, 3 and 4. Scenario 3: 4 dB antenna implementation loss added to antenna configuration 1 and 8dB antenna implementation loss added to antenna configuration 2, 3 and 4. *Note: The values of 4 dB and 8 dB are chosen from [5] to represent an implementation loss from an FR2 antenna panel on a smartphone on a pair of XR glasses respectively.*
Beam selection	Ideal (always selecting the best beam pair)
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	Round robin scheduling strategy in both uplink (UL) and downlink (DL). No inter-subnetwork scheduling coordination.
Link adaptation	Fixed modulation and coding scheme (MCS): 1/3 QPSK, so no link adaptation applied.
Power control	Fixed transmit power of 0dBm for both HC and subnetwork entities.

Figure 5 shows the cumulative distribution function (CDF) of the per sub-carrier signal-to-interference plus noise ratio (SINR) of the downlink transmissions in the indoor factory scenario (on the left) and the indoor classroom (on the right) for a sub-set of different antenna configurations on the HC (the transmitter (TX)) and the subnetwork entity (the receiver (RX)). Here, only 2 subnetworks are deployed and either 2 or 4 antenna panels, where each antenna panel can be equipped with either 1, 4 or 16 antenna elements. Ideal antenna implementation gains are assumed. The figure gives a starting point for identifying how many panels should be utilized at the subnetwork entities, by comparing the solid lines (either theoretical omni or 4 panels at the transmitter and 4 panels at the receiver) with dashed lines (2 panels receiver) and the dotted line which is a scenario of 2 panels at both transmitter and receiver. Overall, and in particular applicable for the indoor classroom scenario, all options of two panels performs on par or worse than a the theoretical omni antenna, which is particularly clear at the tails of the distribution, e.g. the 10% SINR percentile. There does seem to be a difference between the indoor factory and indoor classroom scenario where in the indoor factory, the average SINR of a configuration with 4 panels at the transmitter and only 2 panels at the receiver is sufficient to perform on par on average with the theoretical omni (and worse in some cases), whereas for the indoor classroom also 4 panels at the receiver is needed to perform on par with the theoretical omni antenna configuration. This can be due to the relatively close placement of the transmitter and receiver and its height separation used in this evaluation which reveals that these antenna panel configurations and placements both give positive directivity gains in some directions, as well as negative directivity gains in others when compared to a theoretical omni antenna. In summary, it seems to be clear that when panels are used, it should be either such that there are enough panels (to close the areas with negative directivity gains) or that they are properly placed in the considered scenario.

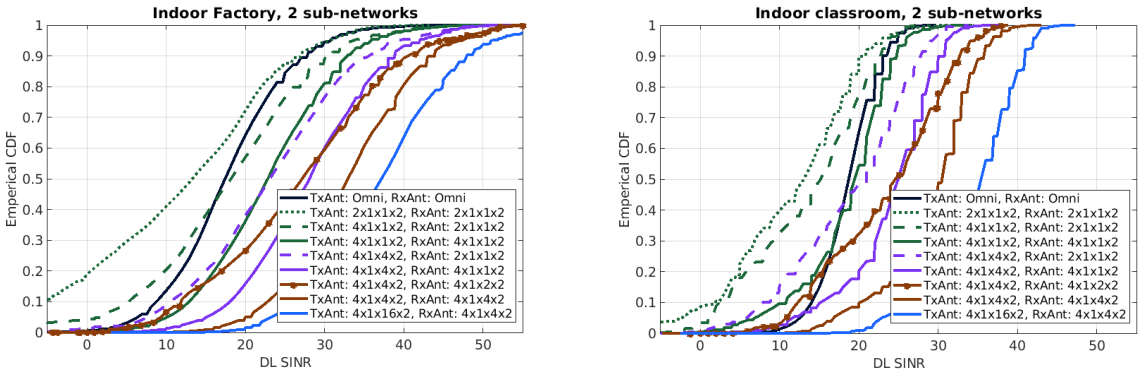


FIGURE 5. EMPIRICAL CDF OF DL SINR WITH 2 SUBNETWORKS IN INDOOR SETTINGS, CONSIDERING ANTENNA LOSS (SCENARIO 0, TABLE 5)."

Another relevant aspect is the size of the antenna panels for in-X subnetworks. From Figure 5 it can be observed that a gain can be achieved using antenna panels which scale proportionally with the number of antenna elements in the antenna panels. The ideal gains of about 6 dB on average may be expected when introducing four panels compared to theoretical omni, and again when introducing 4 antenna elements per panel; this is also what is seen when increasing the number of antenna elements at the transmitter from 1 to 4 (green to purple), when increasing the number of antenna elements from 1 to 4 at the receiver (purple to red), and once again when increasing the number of antenna elements at the transmitter from 4 to 16 (red to blue). This seems to hold for both scenarios.

However, Figure 5 does not give the complete picture, and for that reason, evaluations with antenna implementation losses and a higher quantity of subnetwork devices have also been considered. Figure 6 and Figure 7 shows a bar plot of the average DL SINR considering three different number of subnetworks deployed in the indoor factory and indoor classroom scenario, respectively.

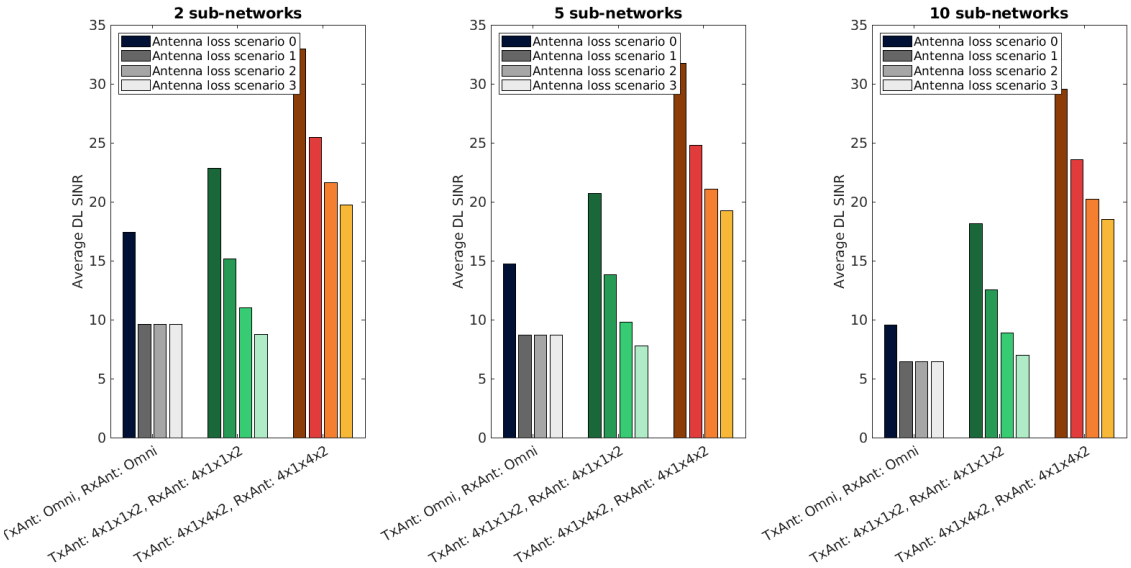


FIGURE 6. AVERAGE SINR IN INDOOR FACTORY SCENARIO, VARYING SUBNETWORKS, ANTENNA SETUPS, AND LOSS SCENARIOS (TABLE 5)."

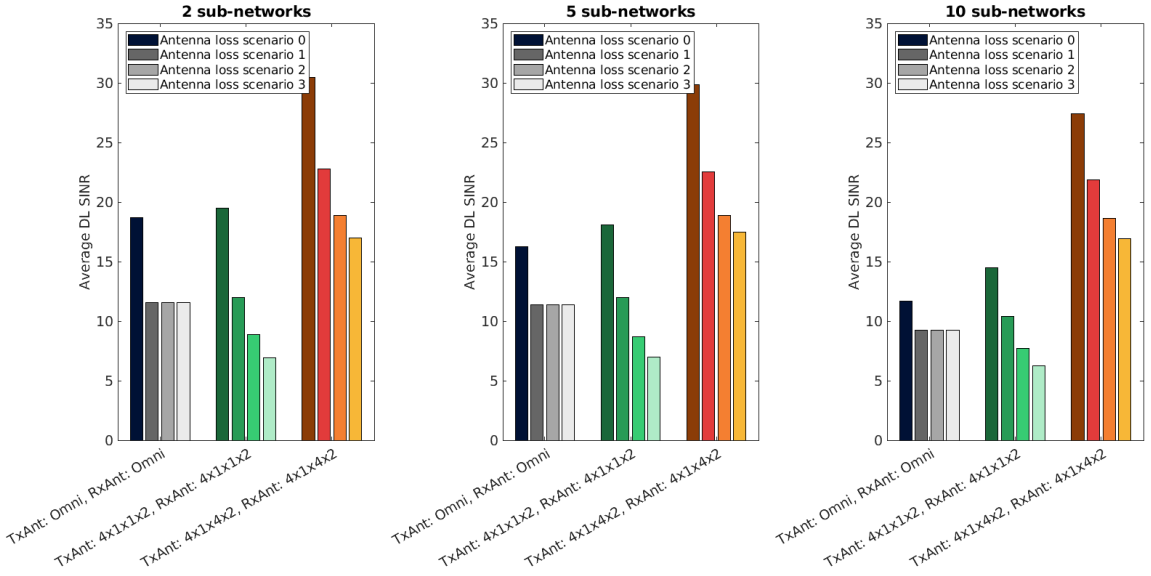


FIGURE 7. AVERAGE SINR IN THE INDOOR CLASSROOM SCENARIO, ACHIEVED FOR DIFFERENT NUMBER OF SUBNETWORKS, DIFFERENT ANTENNA CONFIGURATIONS, AND UNDER DIFFERENT ANTENNA IMPLEMENTATION LOSS SCENARIOS DEFINED IN TABLE 5.

Three antenna configuration combinations are considered in these figures: theoretical omni antennas at TX and RX (black group), 4 panels with 1 antenna element at TX and RX (green groups) and 4 panels with 4 antenna elements at TX and RX (red/orange group). For each antenna configuration, the four antenna implementation loss scenarios are also represented, where scenario 0 is the ideal scenario with no additional implementation losses added, scenario 1 adds 4 dB loss to all antennas, and scenario 2 and 3 adds 6 dB and 8 dB implementation loss to antenna panels respectively.

It can be observed that as the number of deployed subnetworks grows, the average SINR decreases for all scenarios, as would also be expected. It can also be observed that under ideal antenna losses (scenario 0) there are gains by utilizing antenna panels (with both 1 and 4 antenna elements per panel), particularly in the indoor factory scenario where gains are in the order of 10 dB to 20 dB for 2 subnetworks deployed and 8 dB to 20 dB for 10 subnetworks. In the indoor classroom scenario these gains are in the order of 2 dB to 12 dB for 2 subnetworks and 4 dB to 15 dB for 10 subnetworks. These gains are smaller when considering the other antenna implementation loss scenarios. For example, with scenario 1 (where a 4 dB loss is added), the gains are in the order of 6 dB to 16 dB for the indoor factory scenario and 1 dB to 12 dB for the indoor classroom scenario. This reason in this case is that, while there is a directivity gain, the antennas are also emitting less energy due to the antenna implementation losses. For antenna scenario 2 (where 6 dB implementation loss is added to all panels), these gains decrease by another 3 dB to 4 dB which also makes the single element per panel case, performing almost on par with the theoretical omni (for the indoor factory scenario and worse for the indoor classroom scenario). It would be expected that a larger antenna panel size (i.e. 16 antenna elements) at the high capability device would further improve the performance, as was indicated in Figure 5.

The take-aways from this study 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.

And while these take-aways applies for both scenarios, it would be expected that, especially in the indoor classroom scenario, other factors such as subnetwork device designs might make the antenna performance in the worse end of the antenna implementation losses. Moreover, when recalling that the indoor classroom scenario can be assumed to be a highly dynamic scenario with human interaction, there might be challenges in providing gains with larger antenna panels compared to the indoor factory scenario. The subnetwork entities in the indoor factory scenario could even be designed with antenna performance in mind and may even be assumed to be statically deployed; this scenario seems more promising for deploying devices with antenna panels and beamforming capabilities.

2.1.1 Next steps

It should be noted that there are other factors mentioned earlier that have not been covered in detail in this simulation-based study, such as a channel model calibrated for in-X subnetworks which are being worked on in 6G-SHINE WP2. For example, none of the channel models used in the study have been tuned for short ranges, and both have a concave increasing LOS probability as the distance between transmitter and receiver is reduced, and this might not always be the case for in-X subnetworks if NLOS paths exists e.g. due to a blockage. For example, if the updated channel model would show a higher likelihood of non-line-of-sight in short ranges, and an updated path-loss for shorter distances, much less significant beamforming gains would be expected.

It should also be noted that this is the take-away from the first stage of the study (the static deployments), and the conclusions of the benefits of antenna panels for in-X subnetworks needs the second stage of the study as well, which considers the beam management aspects.

2.2 EXPLOITING META SURFACES FOR SHORT-RANGE JOINT BEAMFORMING AND COMMUNICATION

As mentioned in the previous section, the current trend in exploiting higher frequency bands allows for more available bandwidth, higher data rate, and lower latency. Nevertheless, the increase in the frequency results in an increase in the path loss that could limit the operating range, even for short-range communication.

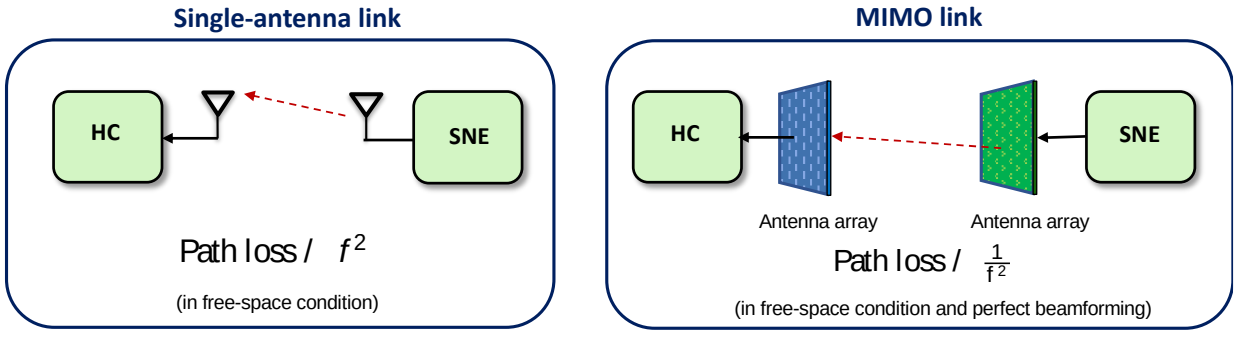


FIGURE 8. PATH LOSS USING SINGLE-ANTENNA AND MULTI-ANTENNA DEVICES IN FREE SPACE CONDITION.

With reference to the industrial subnetwork category, we consider the use case of wireless robot control operations where the operational robot is expected to be equipped with numerous SNEs (e.g., sensors) connected wirelessly to an HC acting as the central communication node of the specific subnetwork. The primary way to address the potentially significant path loss at high frequencies is to implement an antenna array not only at the HC but also at the SNE side. In fact, it is well-known that the path loss in free space using single-antenna devices grows with the square of the working frequency f and distance d (see Figure 8-left). On the contrary, assuming an array of aperture A_{AP} at the HC and an array of aperture A_s at the SNE, the path loss becomes proportional to $d^2/(f^2 A_{AP} A_s)$. This suggests that, in this case, the loss due to the distance can be compensated by either increasing the size of the arrays (within physical constraints) or increasing the frequency (see Figure 8-right), thus making advantageously working at high frequencies. Nevertheless, it is essential to note that achieving such a result requires beam steering at both the HC and the SNE. While this does not represent an issue at least for use cases in which the SNEs are static relative to the HC, for example, test cells and in in-vehicle subnetworks, for use cases in which the SNEs may move relative to the HC, such as in-classroom and on-body subnetworks, achieving fast beamforming is very challenging. In the results shown in section 2.1, ideal beam alignment was assumed, and therefore neglecting the problem of achieving proper beamforming. Unfortunately, while this is manageable for the HC, implementing beam steering algorithms and an antenna array (involving RF phase shifters) in the SNE might not be compliant for the case of extremely low complexity, small, and energy-efficient devices. Beamforming at both transmitter (TX) and receiver (RX) introduces high power consumption, signaling overhead (e.g., for channel estimation), high hardware complexity, and slow beamforming procedures.

Therefore, it is important to conceive a PHY technique to allow an efficient uplink MIMO communication between SNEs and an HC with ultra-low set up time, using low-complexity and low-power devices at millimeter waves and beyond. The following main KPIs are targeted:

- Almost zero signaling overhead to allow efficient short packets transmission (down to 32 bytes);
- Joint communication and beamforming to allow the application to dynamic environments, e.g., sensor on a rotating part of a machine;
- Sensors/devices with extremely low complexity and low energy consumption (no RF chains, no DAC);
- Initial beamforming and access with < 100 µs latency

Sensors might eventually be power supplied using energy harvesting techniques (no battery).

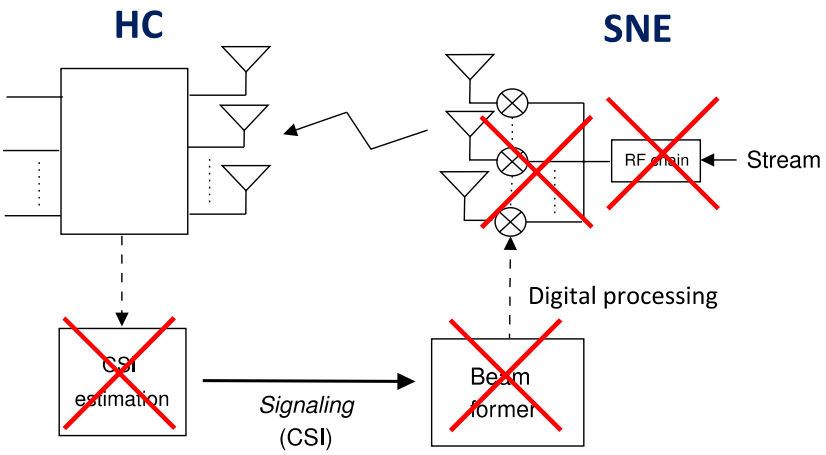


FIGURE 9. CLASSICAL SINGLE-LAYER MIMO LINK AND THE COMPONENTS THAT CAN BE AVOIDED THROUGH THE PROPOSED APPROACH.

Note that the requirements above will be further tuned in the second phase of the project, according to the input from WP2.

To tackle the above challenging requirements, it has been proposed 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 channel state information (CSI) estimation and signaling overhead (see Figure 9) [6].

2.2.1 Self-conjugating arrays/metasurfaces

A schematic representation of a self-conjugating array is reported in Figure 10, where we consider *M* antenna elements organized, for simplicity of explanation, as a uniform linear array (ULA) and a plane incident wave with angle ϕ. Each antenna is supposed to be loaded in such a way the load reflects a conjugated version of the impinging signal.

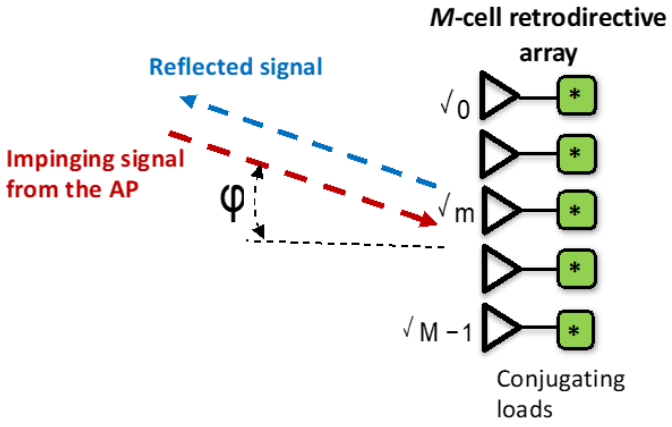


FIGURE 10. SCHEMATIC REPRESENTATION OF A RETRODIRECTIVE ARRAY/METASURFACE.

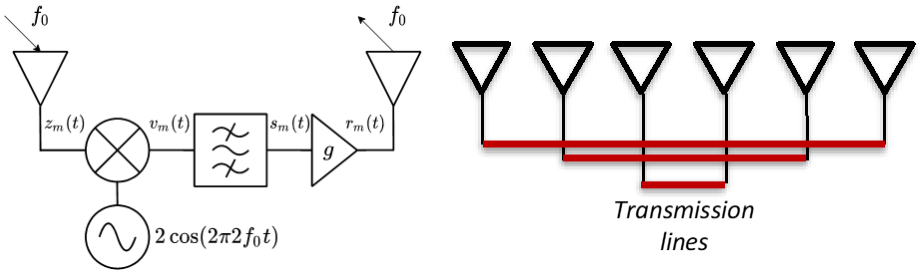


FIGURE 11. (LEFT) HETERODYNE TECHNIQUE FOR PHASE CONJUGATION AT THE MTH ANTENNA. (RIGHT) VAN ATTA PASSIVE RETRODIRECTIVE ARRAY.

In particular at the *m*th antenna, the impinging wave accumulates a phase shift θ_m, with respect to the first antenna, given by

$$\theta_m = \frac{2\pi}{\lambda} m\Delta \, sin(\phi)$$

for *m=0,1,.., M-1*, where $\Delta = \lambda/2$ is the inter-antenna separation and λ the wavelength. At each antenna element, the phase of the backscattered signal is conjugated with respect to that of the impinging signal so that the beam direction vector of the reflected signal is

$$\left[1, e^{-\frac{2\pi}{\lambda} m\Delta \, sin(\phi)}, \ldots, , e^{-\frac{2\pi}{\lambda}(M-1)\Delta \, sin(\phi)}\right]^T$$

which is required to steer the array beam back to ϕ, that is, to have the impinging wave reflected in the same direction of arrival (retrodirectivity).

From a technological point of view, the possibility of realizing ULAs capable of the retrodirective effect has been investigated for many years [7]. The main methods to achieve retrodirectivity are phase conjugating arrays (as explained before), exploiting heterodyne mixing (see Figure 11- left), and Van Atta arrays. While phase-conjugating arrays necessitate active components, Van Atta arrays realize passive retrodirectivity [8], making them more suitable for implementation in low-cost SNEs (e.g., low-cost sensors). As depicted in Figure 11- right, a Van Atta array comprises antenna elements arranged in symmetrical pairs and connected through transmission lines. Each element of the pair serves as a receiving and transmitting antenna. Specifically, the signal received by an element is re-irradiated by its pair after traveling through the transmission line, designed to avoid introducing phase shifts. The elements are deployed in a mirror-symmetric manner to induce an equivalent phase-conjugation effect for the reflected wave compared to the incident wave. Recent developments have demonstrated that metasurfaces can also be designed to provide passive retrodirectivity. This property is achieved by engineering the surface impedance based on a supercell design periodicity greater than the wavelength [9].

2.2.2 Joint communication and beamforming

The initial considered scenario consists of an HC equipped with an antenna array of *N* elements, and a SNE made of a SCM that backscatters the impinging signal from the HC by conjugating and modulating it according to the data to be transmitted. The SCM is composed of *M* cells and it does not require neither DAC nor digital processing (low complexity and energy consumption). The system is supposed to be narrowband, and the HC has full-duplex capabilities.

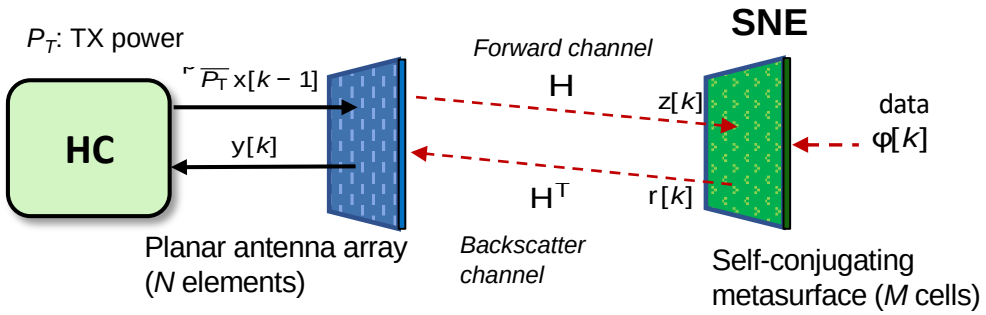


FIGURE 12. JOINT COMMUNICATION AND BEAMFORMING USING AN SCM-BASED SNE.

Differently from conventional retrodirective arrays, we propose that the SCM not only performs the conjugation and possibly amplification of the signal received from the HC, thus retro-directing the impinging signal, but also introduces, in the *k*th symbol interval, a phase shift $\phi[k]$ (the same for all the antennas) that incorporates the information to be transmitted by the HC in that interval. Note that the phase $\phi[k]$ associated with the information affects all the antennas of the array, thus not compromising the retro directivity behavior. Most importantly, the implementation of the modulated SCM does not require DAC chains as data directly modulates the phase sequence $\{\phi[k]\}$, thus allowing a low-cost, low-complexity, low-energy consumption multiantenna device. Thanks to the conjugating properties of the SCM, the following iterative joint communication and beamforming scheme working at the HC has been developed and analysed in the single-SNE scenario.

With reference to Figure 12, let $\sqrt{P_T}\boldsymbol{x}[k]$ be the vector of size *N* containing the signal transmitted by the *N* elements of the HC's antenna array, where P_T is the transmitted power and **x**[k] is a unit norm beamforming vector, i.e., the precoding vector, at the generic time interval *k*. At the startup, i.e., at *k* = 0, the optimal beamforming vector for the link with the given SNE is not known by the HC, which therefore randomly generates a unit norm beamforming vector **x**[0]. At the end of each time interval *k*, with *k* ≥ 1, the beamforming vector **x**[k] will be iteratively updated, as described in algorithm below, where

$$\mathbf{A} = \sqrt{P_{\mathrm{T}}}\, g\, \mathbf{H}^{\dagger}\mathbf{H}$$

is the modified round-trip channel matrix and *g* is the SCM gain (*g<1* is passive).

Initialization: generate a guess unitary norm precoding vector $\mathbf{x}[0]$;
for $k = 1, \ldots, K$ **do**
  transmit: $\sqrt{P_{\mathrm{T}}}\,\mathbf{x}[k-1]$
  $\mathbf{z}[k] = \sqrt{P_{\mathrm{T}}}\,\mathbf{H}\,\mathbf{x}[k-1] + \boldsymbol{\eta}[k]$
  $\mathbf{r}[k] = g e^{J\phi[k]}\mathbf{z}^{*}[k]$
  receive: $\mathbf{y}[k] = e^{J\phi[k]}\,\mathbf{A}^{*}\,\mathbf{x}^{*}[k-1] + \mathbf{n}^{*}[k]$
  $\mathbf{x}[k] = \mathbf{y}^{*}[k]/\|\mathbf{y}[k]\|$
  $u[k] = \mathbf{x}^{\dagger}[k-1]\,\mathbf{x}[k]$
  $\widehat{\phi}[k] = \mathrm{detection}\,(-\arg\{u[k]\})$
end

Specifically, as appears in the pseudo-code of this procedure shown above, the process starts (step 0 of the pseudo-code) with the random generation of a guess unitary norm beamforming vector $\boldsymbol{x}[0]$. Obviously, any prior information (e.g., past transmission, SNE position, etc.) can be exploited to speed up the process, therefore here we are implicitly considering a worst-case scenario. At the $(k-1)$th iteration, with $k \geq 1$, the HC transmits the current version of the beamforming vector, $\boldsymbol{x}[k-1]$ (step 2), which is then received by the SNE (step 3). The latter reflects the received signal along the direction of arrival (thanks to the conjugation operation), modulating its phase based on the data intended for the HC (step 4). Specifically, at the SNE, information data is associated with the phase sequence $\{\phi[k]\}$, forming a packet of length *K* symbols (*K* = ∞ in case of continuous transmission), according to any phase-based signaling scheme (e.g., BPSK, QPSK, ...). The HC receives back the response at time instant *k* from the SNE and the environment (step 5). Then, a normalized and conjugated version of the received vector $y[k]$ is computed (step 6) and used as the updated beamforming vector $x[k]$ in the subsequent iteration. Information data is extracted by the HC at each iteration by correlating the current received vector with the previous beamforming vector $x[k-1]$, thus forming the decision variable $u[k]$ (step 7); the decision on the modulation symbol conveyed by $\phi[k]$ at the *k*th time instant is obtained by means of the function *detection*(·),

according to the adopted modulation scheme (step 8). It is worth noticing that data demodulation is operated while the HC is transmitting its precoding vector, thanks to the full-duplex BS assumption. More details can be found in [6].

The proposed algorithm is a modified version of the well-known power method which is a simple iterative scheme for estimating the strongest eigenvector of a square matrix, in our case representing the optimal beamforming vector. Differently from the conventional power method, here at each iteration both data (phase shift) and noise are added (see Figure 13) thus making the convergence to the strongest eigenvector no longer granted.

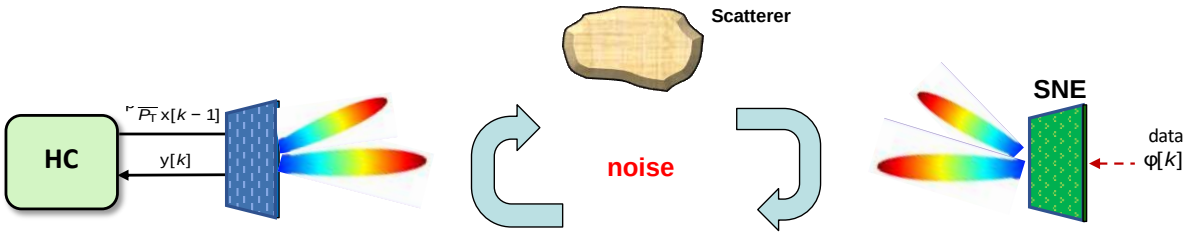


FIGURE 13. OPTIMAL SINGLE-LAYER BEAMFORMING AT THE CONVERGENCE. AT EACH ITERATION, ADDITIONAL NOISE GENERATED BY THE SNE AND HC ENTERS IN THE LOOP.

The convergence behavior of the iterative beamforming scheme has been characterized assuming an initial random guess of the precoding vector. In particular, the signal-to-noise ratio (SNR) evolution is given by the following formula

$$\mathrm{SNR}[k] = \mathrm{SNR}^{(\max)} \frac{\mathrm{SNR}[k-1]+1}{N+\mathrm{SNR}[k-1]}$$

where

$$\mathrm{SNR}[1] \simeq \mathrm{SNR}^{(\max)}/N$$

and

$$\mathrm{SNR}^{(\max)} = \frac{\lambda_1^2}{\sigma^2}$$

is the maximum SNR along the strongest left eigenvector (direction) v_1 of the channel **H,** i.e., the SNR one would obtain if all the power was concentrated to direction v_1 (optimal beamforming), where λ_1 is the corresponding strongest eigenvalue and σ^2 is the noise power. The scheme convergences to the optimal single-layer MIMO performance (i.e., optimal beamforming) when the following condition on the SNR is satisfied, in particular when $\mathrm{SNR}^{(\mathrm{boot})} > 1$, where

$$\mathrm{SNR}^{(\mathrm{boot})} = \mathrm{SNR}^{(\max)}/N$$

is denoted as *boostrap SNR*.

In Figure 14, examples of SNR evolution are provided for different values of bootstrap SNR for rank-1 and rank-3 channels. If condition $\mathrm{SNR}^{(\mathrm{boot})} > 1$ is not satisfied, the convergence is towards an extremely low SNR and the communication does not take place. On the contrary, the convergence value is given by $\mathrm{SNR}^{(\max)}$ and hence it does not depend on the initial random guess ***x****[0]*, but only on the bootstrap SNR which, in turn, depends on the link budget and increases with *N* (array size at the HC) and M^2 (number of cells of the SCM). The same results hold also for higher rank channels.

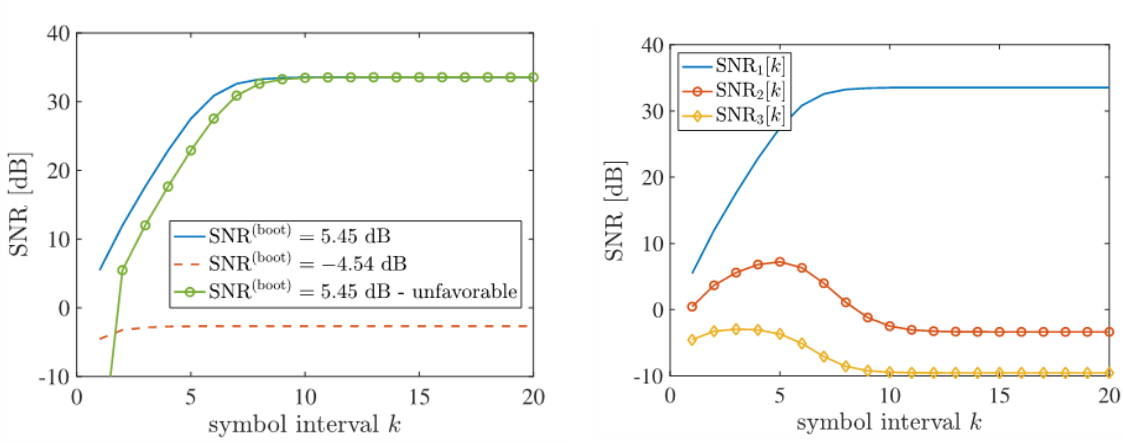


FIGURE 14. SNR EVOLUTION FOR DIFFERENT CONDITIONS. LEFT: RANK-1 CHANNEL. RIGHT: RANK-3 CHANNEL.

2.2.3 Performance evaluation

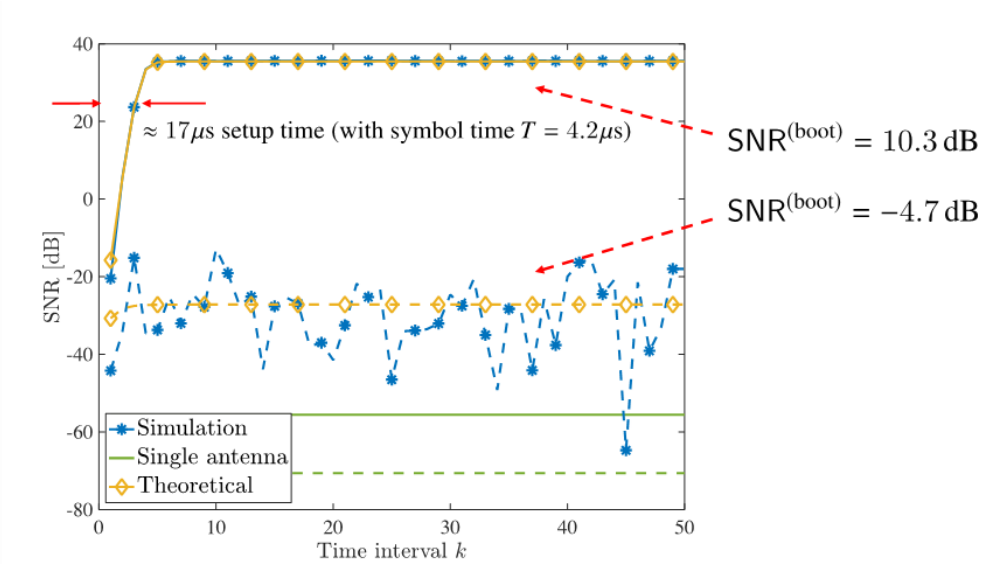
In the following, some simulation results are shown. An HC equipped with a planar array deployed along the xy-plane located at [0m,0m,0m] and a single SNE with a SCM deployed along the xy-plane located at [0m,0m,10m] are considered. The main simulation parameters are listed in Table 6.

TABLE 6. SIMULATION PARAMETERS FOR EVALUATION OF SELF-CONJUGATING METASURFACES

Parameter	**Value**
Carrier frequency	28 GHz
Bandwidth (single carrier)	240 KHz
Symbol time	4.2 us
Total TX power	-30 dBm
SCM gain	20 dB
HC antennas	20 x 20 (10 x 10 cm)
SCM antennas	10 x 10 (5 x 5 cm)
HC noise figure	3 dB
Path-loss exponent	2 (free space/LOS) - 2.5 (NLOS)

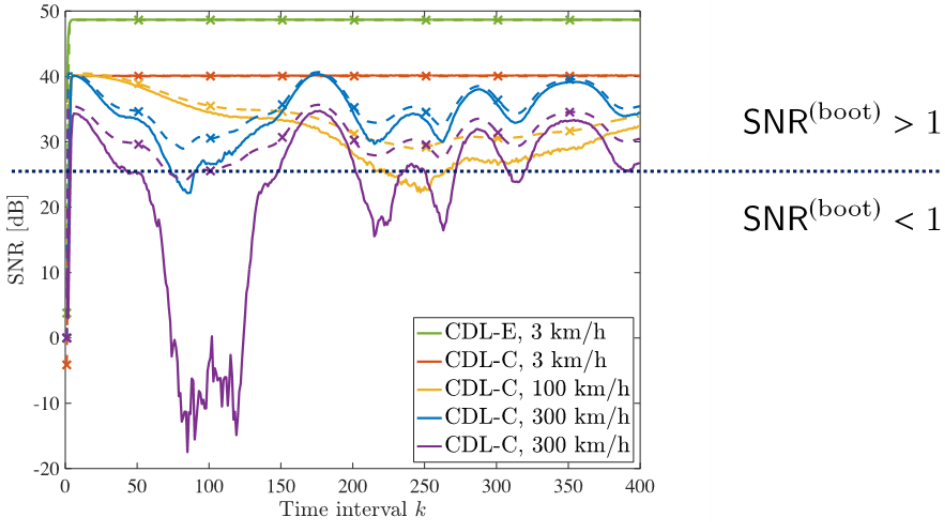
In Figure 15, the evolution of the SNR is shown in static free-space condition for two values of the transmitted power, $P_T = -30$ dBm and $P_T = -45$ dBm, corresponding to $\mathrm{SNR}^{(\mathrm{boot})} = 10.3$ dB, and $\mathrm{SNR}^{(\mathrm{boot})} = -4.7$ dB, respectively. Simulation results are compared with the theoretical ones. From the plots, it can be clearly noticed that when $\mathrm{SNR}^{(\mathrm{boot})} > 0$ dB the SNR converges to $\mathrm{SNR}^{(\mathrm{max})}$ within 4-5 time intervals (i.e., 4-5 iterations of the algorithm), whereas it converges to very low values when $\mathrm{SNR}^{(\mathrm{boot})} < 0$ dB, as well predicted by the theoretical analysis. The SNRs that would be obtained with a single antenna backscattering SNE is also given as a reference (the bottommost

flat curves) in order to emphasize the large gain introduced by the SCM and the antenna array at the HC, despite the large path loss experienced at the considered carrier frequency (28 GHz) and the two-way backscatter channel.



Continuous lines are for $P_T = -30$ dBm; dashed lines are for $P_T = -45$ dBm.

FIGURE 15. TRANSIENTS OF THE SNR FOR SINGLE SNE (FREE SPACE).



LOS/NLOS 3GPP channel models. Continuous lines refer to simulation results; dashed lines with markers refer to theoretical curves

FIGURE 16. TRANSIENTS OF THE SNR: DYNAMIC LOS/NLOS MULTIPATH CHANNELS. FREE-SPACE AND 3GPP LOS/NLOS CHANNEL MODELS (CDL-C/E) CONSIDERED.

Results with more realistic, time-variant, multipath 3GPP channel models, namely the Clustered Delay Line (CDL) models, are shown in Figure 16 for $P_T = -10$ dBm and $P_T = -15$ dBm. In particular, the line-of-sight (LOS) CDL-E and

the non-line-of-sight (NLOS) CDL-C channel models have been considered for different speeds from 3 km/h up to 300 km/h. The behavior of the SNR assuming perfect estimation of the top eigenvector, i.e., $SNR^{(max)}$, is reported for comparison (see the theoretical curves plotted with a dashed style). The results demonstrate the capability of the communication scheme to track the fading evolution at very high speed as long as $SNR^{(boot)} = SNR^{(max)} /N \gg 1$. When $SNR^{(boot)} < 1$, i.e., SNR < 26 dB (note that 10 log10 N = 26), because of fading, the algorithm stops tracking until SNR > 26 dB again, showing a fast recovery. It can also be observed that the range of fading variation during tracking is relatively narrow, approximately 10 dB. This limited range is a consequence of the diversity gain intrinsic in the communication associated with the strongest eigenvector. Moreover, it can be noticed that the initial transient to achieve the convergence in NLOS conditions (CDL-C), characterized by significant multipath components, is slightly larger than that of the LOS condition (CDL-E). However, it falls below 8 time intervals (i.e., 36 μs), confirming that the proposed scheme achieves initial beamforming and link setup with extremely low latency.

Summarizing, the proposed scheme is characterized by the following characteristics:

- In the presence of noise, the convergence value does not depend on the initial random guess of the precoding vector, but only on the bootstrap SNR.
- The bootstrap SNR increases with N (array size at the HC) and M^2 (number of cells of the SCM).
- Despite the SCM can be composed of hundreds of elements, no RF chains, signal processing, beam alignment schemes are needed at SNE side (very simple!).
- The scheme works independently of the channel (far-field, near-field, free-space, multipath), number of antenna elements, and is completely blind.
- It allows channel tracking provided that the channel dynamics is not faster than the convergence speed.
- Optimal single-layer beamforming in the SNR maximization sense is obtained. Consequently, maximum diversity gain is reached contributing to reduce the fluctuations of the SNR.
- The sampling time depends on the propagation round-trip time and HC's computational power. Each iteration could be very fast (<1us). It has been verified that the convergence is reached within a few symbol times, in this setup within 20 μs.
- Extremely low-complexity MIMO communication with SNEs.

The analysis carried out considered narrowband (low data rate) transmission that might correspond to a symbol or resource block of an OFDM link. In principle the proposed scheme can be applied to wideband transmission as well provided that some frequency filtering solution is applied in the SCM. This aspect is currently subject of investigation. The main drawback of the proposed scheme lies in its two-way communication characteristics. In fact, due to the backscattering nature of the communication, the path loss increases with the distance to the power of four. On the other hand, such large path loss can be compensated by increasing the number of antenna elements N and M at the HC and SNE, respectively. In fact, increasing N and M is beneficial also for the bootstrap SNR, even though M has a higher impact than N. Another way to compensate for the path loss is to fix the areas of the HC and SNE and increase the frequency. More details and further numerical investigations can be found in [6].

2.2.4 Next steps

The next research activity aims at:

- Analysing the effect of static clutter;
- Analysing the impact of hardware impairments such as non-reciprocity effects;
- Extending the scheme to a multi-SNE synchronous and asynchronous scenarios.

- Improving the reliability using successive interference cancellation (SIC) and/or Coded MAC approaches (see section. 3.1.3);
- Making it full duplex (e.g., for fast loop control);
- Extend to multicarrier transmission;
- Consider channel models/measurements from WP2;
- Evaluate possible integration with higher layer methods in WP4.

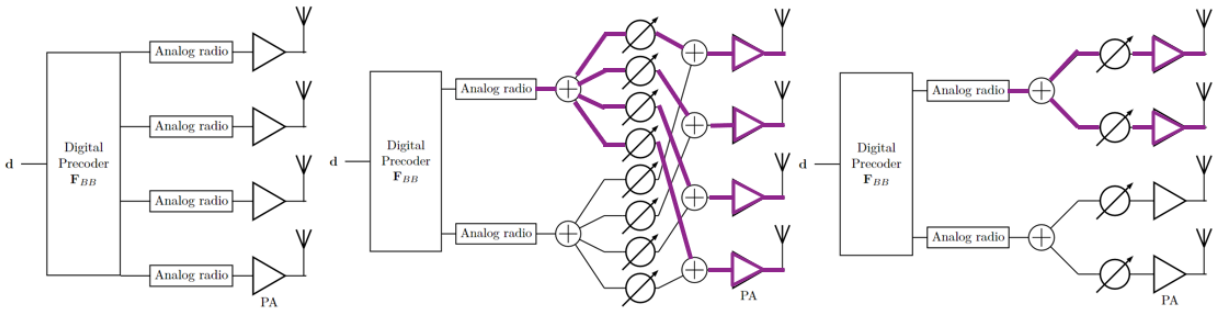
2.3 SUB-THZ TRANSCEIVER ARCHITECTURES & ANALYSIS FOR SUBNETWORKS

Previous sections have highlighted the benefits of mmWave spectrum for in-X subnetworks, evaluating the performance of different antenna panels, and proposing a new joint communication and beamforming method for low cost SNEs. Here, the focus is on even higher frequency range, in the sub-THz region (100 - 300 GHz), which can be seen as a promising option for subnetworks given the availability of large spectrum (several GHz). Operations in the sub-THz spectrum raise concern in terms of energy consumption given the large bandwidth and greater numbers of antenna elements with respect to operations in the mmWave spectrum region. On the other hand, large bandwidth can be exploited to reduce the power consumption of radios, i.e., only activate the radio hardware to achieve the instantaneous data throughput demand and switch it off otherwise. In this section, the focus is on the throughput and power analysis for sub-THz transceiver with different beamforming architectures at HC devices. Reference use case is the indoor gaming use-case defined in WP2.

2.3.1 Transceiver architectures

Bands in the mmWave to sub-THz range are good candidates for the indoor gaming use cas as it requires high data rate and low latency short range communication. The smaller wavelengths associated with sub-THz frequencies enable the use of compact antenna designs. Antenna arrays operating in the sub-THz range can create more focused and precise directional beams, allowing for improved spatial control over the transmitted signals and less interference between different beams. The large antenna array and the beamforming technique increase the link budget to ensure high data rate transmission. Hence, sub-THz communication is attractive for short distance, LOS channel, high throughput, and high directional communication scenarios. However, the sub-THz communication faces hardware impairment compared to lower frequency communication. The most critical issue is the insufficient transmission power from a single power amplifier (PA). The trend of maximum power delivery capability of all different technologies of PA is declining as a function of frequency and has a steeper slope practically in all technologies above 100 GHz. The decreasing slope rate differs between semiconductor technologies. The CMOS technology is not capable of providing sufficient power while compound III-V semiconductor devices like GaN, InP are more suitable for sub-THz communication [10]. To compensate for the output power (Pout) reduction and higher path loss for sub-THz, a phase array is a must to ensure sufficient equivalent isotropic radiated power (EIRP) to build reliable directional high data rate communication. The size of an antenna for a given frequency is chosen such that it is an integer multiple of half-wavelength (λ/2) or full wavelength (λ). For a frequency of 140 GHz (with wavelength of 2.1 mm), a phase array with up to a hundred antennas can be practical in HC devices, and tens of antennas still practical in LC or SNE devices. Three different beamforming architectures: fully digital (FD), hybrid fully connected (HFC) beamforming and hybrid partially connected (HPC) beamforming have been introduced in the section 2.1.1 and are depicted in Figure 17. The different beamforming architectures result in differences in throughput, power consumption and energy efficiency, especially under different antenna panels with different number of antennas. It is worth exploring the throughput and energy efficiency for sub-THz transceivers with

different beamforming architectures under sub-THz channel characteristics to meet ultra-high speed and design power efficient HC and LC devices for in-X subnetworks.



a. Full digital (FD) b. Hybrid fully connected (HFC) c. Hybrid partially connected (HPC)

FIGURE 17. BLOCK DIAGRAM OF THE FULLY DIGITAL, HYBRID FULLY CONNECTED AND PARTIALLY CONNECTED ARCHITECTURES.

To address this problem, the system is modelled as following. We consider a HC device with N_{HC} antennas serving K devices having N_{UE} antennas each. Let s_k denote the data symbol transmitted to user k. Prior to transmission, data symbols are precoded by the precoding matrix $F = [f_1, f_2, \cdots f_k] \in \mathbb{C}^{N_{BS} \times K}$. The transmitted signal is $x = Fs$. The variable $H_k = \sqrt{\frac{N_{HC}N_{UE}}{L}} \sum_{l=1}^{L} a_l a_R(\theta_l^R, \emptyset_l^R)\, a_T^H(\theta_l^T, \emptyset_l^T)$ denotes the channel between the HC and user k, which is based on the geometric sub-THz channel model [11], where L is the number of LOS path, a_l is the complex path gain for the l-th path, $\theta_l^R, \emptyset_l^R$ are the angle of arrival (AOA) for l-th path in azimuth and elevation, respectively, and $\theta_l^T, \emptyset_l^T$ are the angle of departure (AOD) for l-th path in azimuth and elevation, respectively. $a_R(\theta_l^R, \emptyset_l^R)$ and $a_T^H(\theta_l^T, \emptyset_l^T)$ are steering vectors, which are described in [11]. The received signal at user antennas can be expressed as $r_k = H_k x + n$ where $n \sim \mathcal{N}(0, \sigma^2 I_{N_{UE}}) \in \mathbb{C}^{N_{UE} \times 1}$ is the additive noise. The received signal can be separated into desired signal, interference signal and noise components as follows: $r_k = H_k f_k s_k + \sum_{n=1|n\neq k}^{K} H_k f_k s_k + n$.Device k further processes the received signal by combining with its own combining vector $w_k \in \mathbb{C}^{N_{UE} \times 1}$, $y_k = w_k^H r_k$.

The EIRP takes the different user position realizations and different beamforming architectures into account. The EIRP for user k can be expressed as $EIRP(\theta, \emptyset)_k = \frac{N_{HC}}{K} P_{PA}^{(out)} |a_T^H(\theta_l^T, \emptyset_l^T) f_k|^2$, where $P_{PA}^{(out)}$ denotes the output power of the PA. The total EIRP P_k in $(\theta, \emptyset)$ direction is the sum of $EIRP(\theta, \emptyset)_k$. The average received desired signal power of device k after combining is: $P_k = \frac{N_{HC}}{K} P_{PA}^{(out)} \mathrm{E}[|w_k^H H_k f_k|^2]$ and the interference signal power of device k is: $I_k = \frac{N_{HC}}{K} P_{PA}^{(out)} \sum_{n=1|n\neq k}^{K} \mathrm{E}[|w_k^H H_k f_k|^2]$. The SINR for device k is calculated by $\beta_k = P_k/(I_k + P_N)$. A suitable MCS is selected, depending on the SINR per user β_k, and the achieved throughput of device k can then be calculated by: $\Gamma(\beta_k, B) = B \sum_i t_i \mu(\beta_k - \overline{\beta_k})$, where $\mu(\cdot)$ is the unit step function which is equal to unity when the input argument is non-negative, and zero otherwise, t_i denotes the incremental spectral efficiency corresponding to $\overline{\beta_k}$, B is the bandwidth and i is the MCS index. The average throughput over all devices is $\frac{1}{K} \sum_{k=1}^{K} \Gamma(\beta_k, B)$.

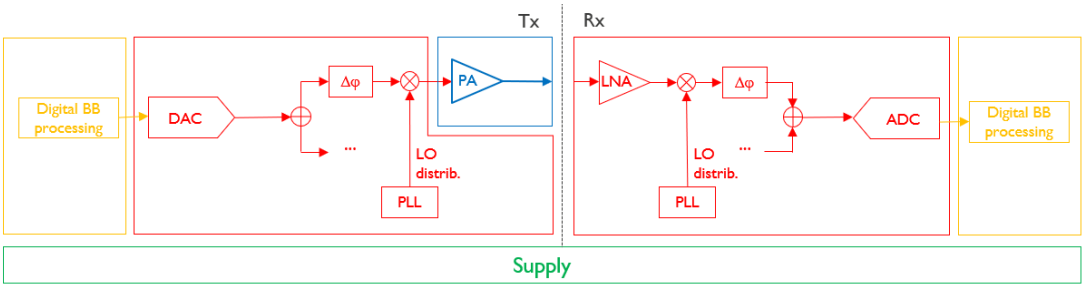


FIGURE 18. BLOCK DIAGRAM OF TRANSCEIVER AND THE CORE COMPONENTS TAKEN INTO ACCOUNT FOR POWER ANALYSIS

The power model analysis simulation mainly analyses 4 components shown in Figure 18. Block diagram of transceiver and the core components taken into account for power analysis, i.e. PA, analog front-end, digital baseband, and power supply for both TX and RX at the HC or SNE side; more details could be found in [12]. For this work, we focus on differences between beamforming architectures with different number of antennas. The properties of power amplifier with different technologies for sub-THz are integrated, such as $P_{PA}^{(out)}$ for III-V semiconductor devices like GaN, InP. The power model for the three different beamforming architectures at HC device for downlink is presented in Table 7. The data rate and power consumption of uplink compared with downlink is relatively small [12] ,hence in this analysis only the power consumption for downlink is considered. The transmit power consumption is analysed as analog and digital circuits. The analog part contains the following components: PA power ($P_{PA,DC}^{(out)} = P_{PA}^{(out)}/\eta$), where η is the efficiency of PA, power of the mixer (P_{mix}), power of the local oscillator (LO) (P_{LO}) with its distribution power $P_{LO,d}$, power of the phase shifter (P_{PS}), and power for digital-to-analog converter (DAC) (P_{DAC}). For the HC device, the digital part contains channel encoding, constellation mapping, inverse fast Fourier transform (IFFT), MIMO precoding, and upsampling and filtering operations for downlink and filtering, down sampling, MIMO processing, channel estimation, equalization, demapping and decoding for uplink. $P_{BB,dig}$ denotes the total baseband power for the FD beamforming architecture with $N_{HC} = K$ digital chains, while $P_{BB,hyb}$ denotes the total baseband (BB) power for HFC and HPC beamforming architecture with only K digital chains. Single carrier modulation and Single Carrier Frequency Domain Equalization (SC-FDE) is considered for this study for its superior performance to mitigate the effects of frequency-selective fading, its low implementation complexity, low Peak-to-Average Power Ratio (PAPR). The main block diagram of digital baseband architecture is shown in Figure 19. Q represents the number of quantization bits, and the power consumption is scaled with Q. Numerical value of each analog components is based on lab experiments, more details can be found in [13][14]. The power modelling is based on rough estimation of the complexity: giga operations per second (GOPS) and quantization of different functionalities, then those two factors are translated to power consumption in Watts as function of selected technology [12]. The state-of-art 7nm and up-coming 2nm technology nodes are considered to highlight the impact of circuit technology on energy efficiency.

2.3.2 Performance evaluation

Indoor gaming is an envisioned use case for In-X subnetworks, where the sub-THz communication is a good candidate to build the wireless link between the HC and LC. To ensure an immersive gaming experience, the high data rate and low latency transmission are the two crucial KPIs. Since the signal processing latency in PHY is very small compared with the network latency and application-level latency, the over-the-air transmission latency is not considered for this work. In the following, we apply the methodology mentioned above to explore the throughput and energy efficiency for different beamforming architectures under different technologies nodes.

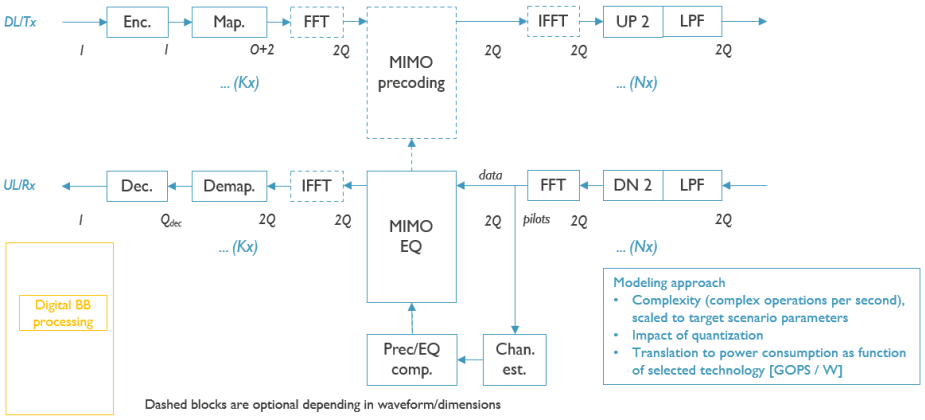


FIGURE 19 BLOCK DIAGRAM OF BASEBAND PROCESSING AND THE CORE FUNCTIONALITY TAKEN INTO ACCOUNT FOR BB POWER ANALYSIS.

TABLE 7. TRANSMIT POWER CONSUMPTION BREAK DOWN OF DIFFERENT BEAMFORMING ARCHITECTURES FOR DOWNLINK.

Architecture	$\boldsymbol{P_{TX}}$
FD	$N_{HC}\left(P_{PA,DC}^{(out)} + P_{DAC} + P_{mix} + P_{LO,d}\right) + P_{LO} + P_{BB,dig}$
HFC	$N_{HC}\left(P_{PA,DC}^{(out)} + KP_{PS}\right) + K\left(P_{mix} + P_{LO,d} + P_{DAC} + P_{amp}\right) + P_{LO} + P_{BB,hyb}$
HPC	$N_{HC}\left(P_{PA,DC}^{(out)} + P_{PS}\right) + K\left(P_{mix} + P_{LO,d} + P_{DAC} + P_{amp}\right) + P_{LO} + P_{BB,hyb}$

The parameters for power model analysis simulation for the VR indoor gaming scenario have been listed in Table 8. The simulator assumes that the number of users is equal to the number of spatial data streams and digital chains, and the number of PAs is equal to the number of antennas. Due to the small indoor area, only 2 devices are considered, otherwise the interferences from other beams will be large. The HC is assumed to be fixed at the centre of the ceiling, and the two users' positions are characterized by their height and the elevation and azimuth angles regards to the perpendicular axis to the ceiling. Hundreds of different channel realizations are simulated, and the average performance is reported.

TABLE 8. PARAMETERS FOR SUB-THZ TRANSCEIVER POWER ANALYSIS SIMULATION FOR XR INDOOR GAMING SCENARIOS.

Parameter	**Specification**	**Value**
Geometry (indoor gaming)	AP height [m]	5
	user height [m]	U[1.5,1.8]
	User locations in elevation	$\{^{-\pi}/_{3}, ^{-\pi}/_{6}, ^{\pi}/_{6}, ^{\pi}/_{3}\}$
	User locations in azimuth	$\{^{-\pi}/_{3}, ^{-\pi}/_{6}, ^{\pi}/_{6}, ^{\pi}/_{3}\}$
Carrier frequency		140 GHz
RF bandwidth		{1,2,4} GHz
PHY performance	Active (minimal and peak) throughput per UE (DL, UL)	1Gbps DL
	Target BER (DL, UL)	0.00001
	Number of simultaneously active users in same band	2
#streams data = #user = # digital chains		2

Possible waveforms, constellations, coding rate/options		802.11ad SC/OFDM
Propagation channel for each user	Type of propagation (LOS, NLOS)	LOS
BS architecture constraints	Max PA power (techno-dependent) CMOS /GaN [dBm]	14
	Max number of antennas and array size	{4,..,32}
	Antenna element gain, interconnect loss, radiation diagram	+3 dBi
	Nosie figure	8 dB
	Optimization criteria (energy eff. or other)	energy eff.
UE architecture constraints	number of antennas and array size	2X2

Figure 20 depicts the average throughput for 500 different channel realizations versus the number of Tx antennas comparison of FD, HFC, and HPC architectures for different systems with different bandwidth. The transmitter antenna configurations used in this experiment are 2×2, 4×2, 6×2, 4×4, 6×4, and 8×4 in vertical and horizontal axis, respectively. The EIRP is linearly scaled with bandwidth, hence the higher bandwidth supports the higher throughput. For fixed bandwidth, the FD and HFC architectures support the same throughput, which both benefit from spatial diversity. The HPC cannot achieve the same EIRP as HFC and FD, hence leading to a lower throughput. A digital stream is distributed to the whole antenna array in HFC, while to the partial antenna array in HPC, as illustrated in Figure 17.

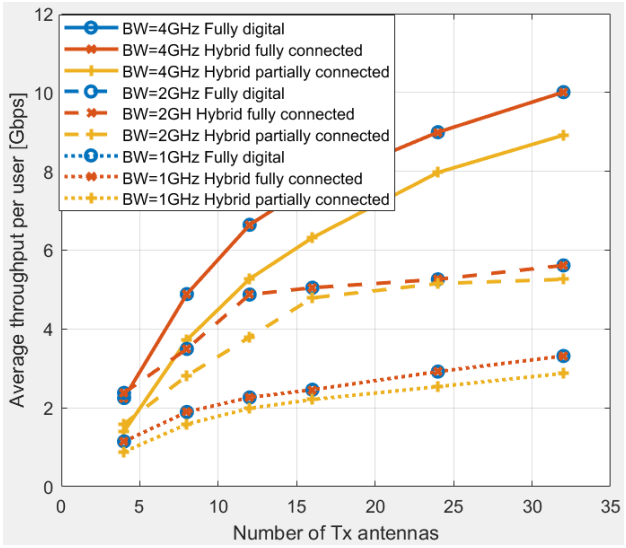


FIGURE 20. AVERAGE THROUGHPUT PER USER VERSUS THE NUMBER OF TX ANTENNAS COMPARISON OF FULLY DIGITAL, HYBRID FULLY CONNECTED, AND HYBRID PARTIALLY CONNECTED ARCHITECTURES CONSIDERING 1 GHZ, 2GHZ AND 4GHZ OF BANDWIDTH.

For energy efficiency optimization, we assume that duty-cycling can be exploited. This means that if the system requires, e.g., 1 Gbps while maximum allowed throughput is 10 Gbps, short bursts of data can be transmitted 10% of the time at 10 Gbps while the system can be put in sleep mode the rest of the time. This is generally more efficient in terms of system energy efficiency, and if designed properly, all electronic components are able to fast enough switch on/off such that the impact of this strategy on application latency is negligible.

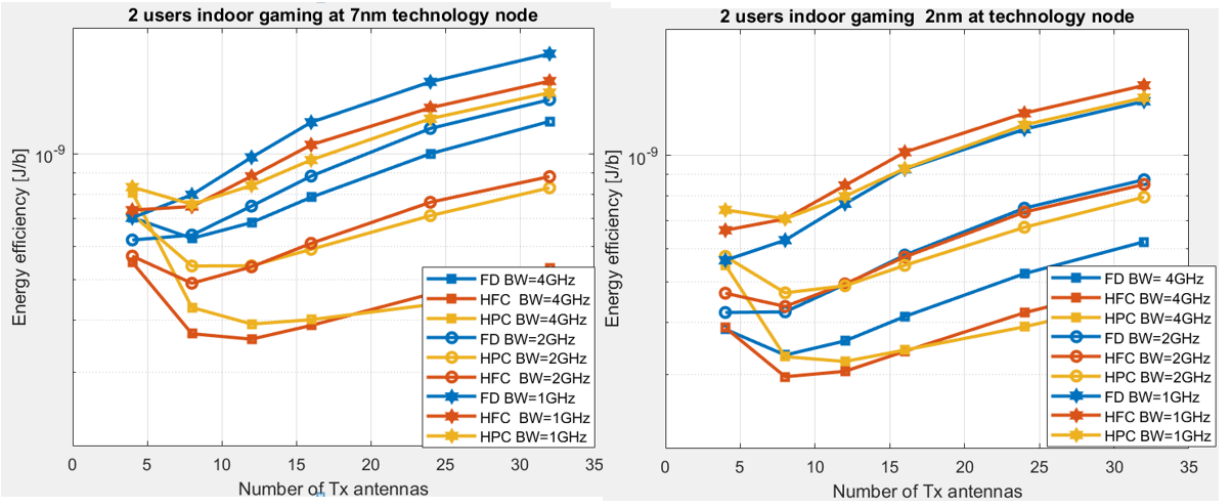


FIGURE 21. ENERGY EFFICIENCY VERSUS THE NUMBER OF TX ANTENNAS COMPARISON OF FULLY DIGITAL, HYBRID FULLY CONNECTED, AND HYBRID PARTIALLY CONNECTED ARCHITECTURES FOR 7NM AND 2NM TECHNOLOGY NODE AND WITH DIFFERENT BANDWIDTH.

Figure 21 depicts the energy efficiency versus the number of Tx antennas comparison of the three architectures for 7nm and 2nm technology node for different bandwidth: 1 GHz, 2 GHz and 4 GHz, respectively. In general, for fixed architecture and fixed technology node, the wider bandwidth system provides better energy efficiency. For 7nm technology node, for a fixed bandwidth, the FD architecture has the highest energy efficiency value when the antenna size is larger than 4, where the full digital and analog chain dominates the power consumption. For a fixed bandwidth, the energy efficiency value of HFC is higher than HPC when the antenna array is small, and the HPC becomes the most energy-efficient architecture when the number of transmit antennas is greater than a certain value at different bandwidths. The decrease in energy-efficiency of HFC is due to its increasing splitter losses. For 2nm technology node, different from 7nm, FD outperforms the HFC and HPC architectures when the number of transmit antennas is smaller than certain value, due to reduced baseband power consumption. We observe that FD architecture benefits from technology scaling more since the power is dominated by digital components. HPC architecture becomes the most energy-efficient compared to FD and HFC architectures when there is no limit on the antenna array size due to good enough spectral efficiency with many antennas while saving on power compared to other architectures.

We identified that the most energy-efficient architecture changes with respect to bandwidth, technology, and antenna array size constraints. HFC architecture is the most energy-efficient architecture considering the 7nm technology node and size-constrained antenna arrays, while HPC architecture becomes more energy-efficient for large antenna arrays. FD architecture becomes the most energy-efficient for the 2nm technology node and 2,4 GHz bandwidth with size-constrained antenna arrays. For 6G in-X subnetworks, it is recommended to use wider bandwidth and only activate the radio hardware to achieve the instantaneous data throughput demand, which in general provides the best energy efficiency. For fixed bandwidth, the HPC and HFC architecture normally provide better energy efficiency compared to FD architecture. HFC architecture is more energy efficient than HPC when the number of antennas is smaller and HPC architecture is more energy efficient than HFC architecture when the number of antennas is large.

2.3.3 Next steps

For future work, we will further continue to analyse the power consumption of baseband transceiver of the head mounted devices (HMD) and perform HC and HMD energy efficient co-optimization for the indoor gaming use case. In the meantime, we will perform D-Band physical layer simulation to verify the communication performance for different transceiver architectures.

2.4 JAMMING-ROBUST PHY DESIGN

Jamming and malicious interferers can pose a threat to communication reliability, especially for those use cases demanding very low latencies. Physical layer in 6G should be designed to provide a tier of robustness to jamming and malicious attacks. This is relevant for all use cases demanding high reliable communication, such as robot control (industrial), Wireless Zone ECU, Collaborative Wireless Zone ECU, Inter-subnetwork Coordination, Virtual ECU (in-vehicle), and AR navigation (consumer). As mentioned in section 2.3.1, receive beamforming is already providing a tier of protection to jamming attacks due to spatial filtering of signals from unwanted directions. In this section, we present two foreseen approaches for improving native robustness to external interference foreseen in 6G-SHINE. It is worth remarking that, while these approaches can reduce the impact of jammers, PHY enhancements might not suffice for ensuring ultra-reliable communication and should eventually be combined with detection and mitigation techniques. Such solutions are studied in WP4.

2.4.1 Native robust design

It is our assumption that, services that require ultra-reliable communication, should use a frame structure and a data mapping approach able to provide inner robustness to jamming and malicious attacks, as well as from interference generated by other systems eventually operating over the same spectrum (especially in the case of unlicensed spectrum).

Jammers can be broadly classified as wideband and narrowband jammers [15]. Wideband jammers transmit their signal over a broad frequency range. While their effect can be instantaneously disruptive, they can be easily detected, and reactive mitigation techniques can be promptly set in place. On the other side, narrowband jammers are operating over a limited portion of the operational spectrum; their impact is more subtle, and they might be hard to detect (or distinguished from legitimate interferers). A smart jammer might be able to identify the operational channel or sub-channel and jeopardize the communication.

The effects of narrowband jammers can be then counteracted by spread spectrum, i.e. by transmitting over a band which is significantly larger than the bandwidth potentially used by the jammers [16]. Nonetheless, instantaneous transmission over a large bandwidth may be costly from a device perspective as it demands a wideband transmitter with a high baseband sampling rate. Channel hopping techniques, where each transmission happens over a different frequency channel, is an interference robust technique that allows maintaining a low device complexity thanks to the small instantaneous bandwidth [17]. By hopping operational channel at each transmission, one may avoid persistent interference by a jammer. We therefore believe that, for services demanding high reliability, physical layer should follow the following criteria:

- Consecutive transmissions towards and from a certain device, should happen over different frequency channels/subbands, to the largest extent in a non-predictable manner. The hopping pattern can be

generated by using pseudorandom sequences known only at the transmitter and receiver. Transmitter and receiver entities should share the seeds used for the pseudorandom sequence generation in e.g., the authentication phase prior operation, and ensure that transmit and receive pattern are aligned. It is worth remarking that the presented approach is similar to the one adopted by ad hoc technologies operating in the unlicensed spectrum, such as Bluetooth, but is not used in cellular type of technologies [18]. We also believe that channel hopping should happen not only for data packets but also for control information, including broadcast channels. This is because jamming broadcast channels may have a disruptive effect on the communication in the entire subnetwork, and not only for individual links.

- Reliability can be enhanced by encoding the packet across multiple hops or repeating the packet. This obviously translates to a lower spectral efficiency and therefore this approach should only be used when ultra-reliable communication is required. Also, the system should be dimensioned such that all repetitions can be accommodated within a tolerable latency budget. This may require the utilization of extremely short transmission units, combined with a large bandwidth. In case an OFDM air interface is used, a packet can be mapped over a single OFDM symbol, having a large subcarrier spacing. For example, the numerology in [18] recommends subcarrier spacings in the order of 480 kHz. This may allow accommodating more than 20 symbols in a 0.1 ms communication interval.

The usage of pseudo-random channel hopping combined with packet repetitions can offer a robust tier of protection to jamming attacks, but also increases the risk of collisions with transmissions of neighbor subnetworks, especially in the case of uncoordinated subnetworks without possibility of aligning transmission patterns. In other terms, improving resilience to external attacks may come at the price of an increase of occurrence of legitimate interference.

2.4.1.1 Dimensioning of radio resources

Figure 22 shows, a generic radio resource grid comprising of q frequency channels and p time slots, where each channel/slot combination represents a resource unit where a transmission packet is mapped. Devices belonging to different subnetworks are using different transmission patterns, each with n repetitions, and they may collide over one or more resource units. This can hold for both uplink and downlink; for the sake of simplicity, we will refer to the uplink case next.

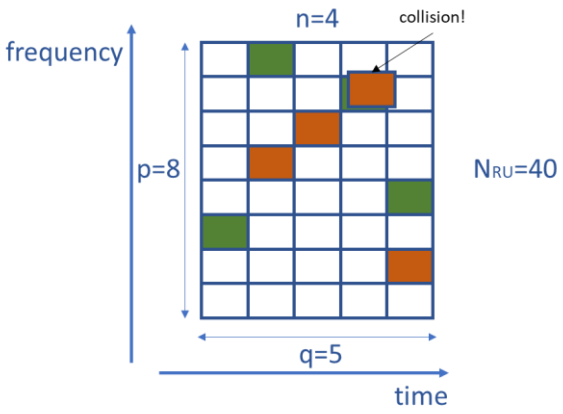


FIGURE 22. EXAMPLE OF TRANSMISSION PATTERNS BY 2 DEVICES, DENOTED BY DIFFERENT COLORS, INCLUDING ONE COLLISION EVENT.

It is therefore important to derive the minimum number of resource units needed to support a given reliability level in terms of packet error rate. Let us consider first a classic packet erasure channel model, where collisions are disruptive events. A transmission by a given device is therefore successful if at least one of the packet repetitions is interference-free. In our paper [19], we have derived a closed form expression of the minimum number of packet repetitions and minimum number of resource units required for supporting a given failure probability, where a failure event corresponds to the situation where none of the packet repetitions is interference-free.

For a high number of interfering devices d, the minimum amount of resource units for achieving a failure probability target P_f can be approximated by [19]

$$N_{min} \approx \left\lceil \frac{\ln(P_f)}{\left[\left(\frac{1}{2}\right)^{\frac{1}{d}} - 1\right]\ln(2)} \right\rceil$$

where $\lceil \cdot \rceil$ denotes the ceiling operator. For large d, we can resort to a Maclaurin expansion $\left(\frac{1}{2}\right)^{\frac{1}{d}} \approx 1 + \frac{1}{d} ln\left(\frac{1}{2}\right)$, and therefore $N_{min} \approx \left\lceil \frac{d}{ln^2(2)} \ln(P_f) \right\rceil = -\lceil 2.0814 \cdot \ln(P_f) \rceil$. The number of packet repetitions for achieving the error probability target with such number of resource units is given by $n = -\left\lceil \frac{\ln(P_f)}{\ln(2)} \right\rceil$. Interestingly, for a high number of interfering devices such value is independent from the number of devices itself.

Let us consider now the case in which the receiver is able to resolve some of the collisions, for example based on the usage of successive interference cancellation receivers, or multi-antenna processing. In this case, we can soften the assumptions of the basic packet erasure channel model, and consider a failure event the case when all packet repetitions of the device experience more than N_c^{max} collisions, where N_c^{max} is the maximum number of resolvable collisions. Such failure probability is given by

$$P_{f,N_c^{max}} = \left(1 - \sum_{c=0}^{N_c^{max}} P(N_c = c)\right)^n$$

where $P(N_c = c) = \binom{d}{c}\left(\frac{n}{N_{RU}}\right)^c \left(1 - \frac{n}{N_{RU}}\right)^{d-c}$. The minimum number of resource units for achieving a given failure probability, can be derived numerically from the expression above. For the specific case of receivers able to resolve a single collision ($N_c^{max} = 1$), we have shown in [19] that the minimum number of resource units able to cope with a given error probability target can be approximated as $N_{min} \approx -\lceil 0.7502 \cdot \ln(P_f) \rceil$, and the number of repetitions is equal to $n = -\lceil 0.6905 \cdot \ln(P_f) \rceil$. By comparing this result with the case of no collision resolution capability, it turns out that the capability of resolving one collision leads to a reduction of the amount of resource units of around ~x2.77.

Figure 23 shows the number of resource units in a scenario with d=100 potential interfering subnetwork links, and a target failure probability of 10^{-6}, as a function of the number of packet repetitions. The number of packet repetitions ensuring the minimum resource units, is also highlighted. For $N_c^{max} = 0$ (i.e., no collision resolution capability) and $N_c^{max} = 1$ (capability of resolving one collision), analytical results are used, while for the other cases results are based on numerical inversion of the expression for failure probability. In all cases, simulation

results show a perfect match with the analytical or numerical values. Results highlight the trade-off between diversity and generated interference, i.e. a certain number of packet repetitions is needed to provide sufficient diversity for having interference-free transmissions; however, if the number of repetitions becomes too high, the generated interference leads to an increase of the amount of resource units needed for achieving the given error probability target. In general, the usage of receiver with collision solving capabilities leads to a significant reduction on the number of resource units for supporting ultra-reliable communication.

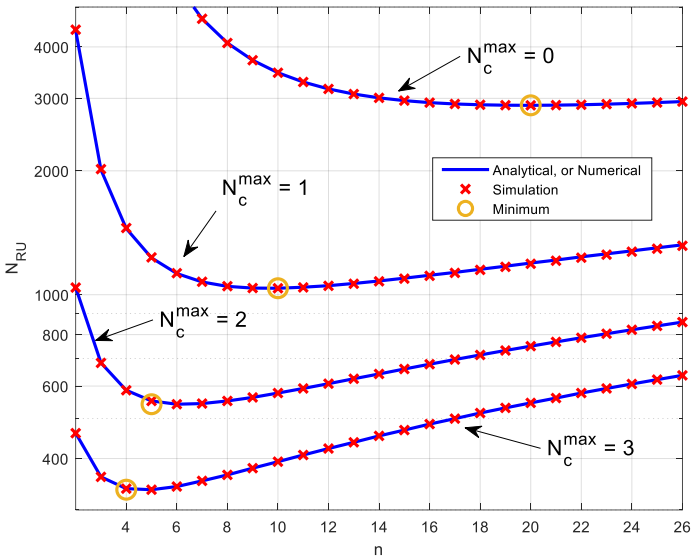


FIGURE 23. NUMBER OF RESOURCE UNITS FOR ACHIEVING AN ERROR PROBABILITY TARGET OF 10^{-6}, ASSUMING 100 INTERFERING DEVICES.

It is worth to remark that, the wireless system designer has the freedom of selecting the number of frequency channels p and time slots q such that $pq = N_{RU}$, with the constraint that the minimum number of frequency channels or time slots should be at least equal to the number of packet repetitions. Such choice depends on the spectrum availability and latency requirements. In case of a large available spectrum, the designer can opt for a large number of frequency channels, and a low number of time units, therefore, be able of supporting low latency. Conversely, in case of limited spectrum resources, the designer must opt for a large number of time slots, leading to higher latencies. For example, results above show that 1040 resource units and 10 packet repetitions are needed in case of receivers capable of resolving a single collision; by assuming a maximum time slot of 1 ms and 1 GHz available spectrum, the system designer can opt for a resource grid of 104 frequency channels of 9,61 MHz and 10 time units of 100 µs duration, where devices can transmit their 10 packet repetitions using a pseudorandom hopping pattern.

Note that main scope of the analysis carried out in [19] was to obtain general insights on the relationship between number of required resource units and failure probability. Results are based on a packet erasure channel model where interference is the only performance-limiting factor. Still, we argue that, in spite of its simplicity, the packet erasure channel model can be seen as a good approximation for short-range subnetworks where the receive signal strength is expected to be high, and the HC can use techniques such as power control and link adaptation to counteract fading and ensure correct reception for interference-free transmissions. In this respect, the number of

interfering devices refers to those devices having the possibility of clearly jeopardizing the performance. Also, collisions that cannot be resolved by a successive interference cancellation receiver are not necessarily leading to disruptive events. In the next section, we present a receiver enhancement meant at improving robustness to colliding transmissions.

In summary, a physical layer featuring inner robustness to jamming, should include channel hopping with pseudo-random hopping pattern, possibly combined with packet repetitions in case ultra-reliable communication is demanded. However, such pseudo-random channel hopping inevitably causes interference in the case of dense subnetwork deployments and is therefore important to properly allocate the number of resources needed for ensuring a given target error probability. We have provided here guidelines for dimensioning such number of resources.

2.4.1.2 Next steps

In the initial study presented here, though the robustness to jamming is the starting point for the proposed design, performance is analysed with focus on the legitimate interference only. Also, we have assumed the same number of repetitions for all devices. In coming research, we will be dimensioning the number of resources and packet repetitions based on the actual signal-to-interference plus noise ratio of each device. Also, we will discuss the actual impact on jamming attacks on the performance, and how to take this into account in the resource dimensioning.

2.4.2 Baseband Receiver Improvement

In the previous section, we investigated how to reduce the probability of collision with interferers, or transmissions generated by another uncoordinated system. Though collision is indeed undesired, packet errors may not necessarily occur in case of collisions. In modern communication systems, Forward Error Correction (FEC) coding techniques, such as convolutional codes, turbo codes [20], low density parity check (LDPC) codes [21] are commonly being applied, that introduce some level of redundancy in exchange for robustness, and can resolve colliding transmissions. The performance of the baseband receiver plays indeed an important role when the signal of interests is overlapped with interference. Apart from the coding mechanism, the demapping stage, which happens before removing FEC coding, also plays a critical role in the final performance. In this section we present a demapper based on Approximated Likelihood Ratio (ALLR) to improve robustness against jamming. We focus here on its algorithm and operational principle, and we also illustrate the gained performance using simulation in Matlab. The simulation is done using the bit true model of the baseband module, which has its corresponding realization on the field programmable gate arrays (FPGA) hardware, studied in WP5.

2.4.2.1 ALLR demapper

Demapper is a critical module in the baseband receiver and translates complex numbers in a constellation into soft bits, which will then be further processed by various decoders depending on the coding techniques applied.

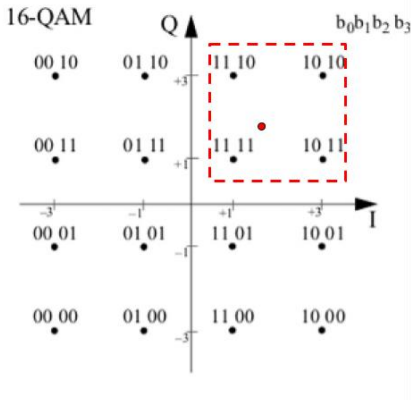


FIGURE 24. EXAMPLE OF DEMAPPER FOR 16-QAM

As an example, Figure 24 shows the basic operation of demapper for 16-QAM modulation, where the symbol in the red dot will be mapped into 4 "hard" bits. When using soft bits each of the 4 bits actually receives a "score" to indicate the confidence in the decision. There are many ways to perform demapping, where the most straightforward approach is to determine the soft bits based on distance between constellation points. If the received symbol is very close to one of the reference points (i.e., distance falls into a predefined threshold), then it will receive a high score for the corresponding bits. On the other hand, if the received symbol is far away (i.e., distance to reference points bigger than certain threshold), the soft bits will have a lower confidence. A demapper that determines soft bits purely by distance to the reference symbol, is referred to as threshold-based demapper hereafter. However, with this approach a received symbol could be entirely moved towards a wrong reference point due to noise or interference; this may eventually lead to packet decoding errors.

Therefore, a more reliable approach is to consider not only the distance of a symbol to its reference point, but also the distribution of noise or interference on a particular subcarrier. For OFDM modulation, received symbols will be equalized using channel estimation. Channel is typically estimated by known preambles, and the estimate is referred to as Channel State Information (CSI) hereafter. When narrowband interference is present on certain subcarriers, or part of the subcarriers are heavily shaded by objects, then the CSI on those subcarriers will be affected.

The metric to determine the most likely value of a bit given the received symbol is the measurement of likelihood ratio. For a given bit (*b*) and received symbol *r(x,y),* where *x y* denotes its corresponding location in the complex plane, the likelihood ratio of the bit being '0' rather than being '1' is calculated as such:

$$L(b) = \log\left(\frac{\Pr\,(b=0|r(x,y))}{\Pr\,(b=1|r(x,y))}\right).$$

This is the definition of log likelihood ratio (LLR) [22], that can be used as a way to calculate the "score" for soft bits.

When assuming equally distributed symbols, the LLR for transmissions over an Added White Gaussian Noise (AWGN) channel is expressed as follows:

$$L(b) = \log\left(\frac{\sum_{s\in S0} e^{-\frac{1}{\sigma^2}((x-s_x)^2+(y-s_y)^2)}}{\sum_{s\in S1} e^{-\frac{1}{\sigma^2}((x-s_x)^2+(y-s_y)^2)}}\right),$$

where *S0* and *S1* are the collection of reference constellations that make the bit 0 or 1 respectively, s_x and s_y are the real and imaginary coordinates of the reference constellation point on a complex plain, and $\sigma^2$is the noise variance of the AWGN channel.

As shown in the equation above, the exact likelihood requires to calculate the distance of a symbol to all reference symbols in the constellation. Although this is trivial to achieve in software, the physical layer processing of a packet usually needs to happen in hardware. Hence another level of simplification is made, i.e. instead of the exact probability, we simply compare the likelihood between the most nearby reference constellation that makes the bit '0' and the most nearby reference constellation that makes the bit '1'. This is referred to as the ALLR [22], as described in the equation below:

$$L(b) = -\frac{1}{\sigma^2}\left(\min_{s\in S_0}((x-s_x)^2+(y-s_y)^2) - \min_{s\in S_1}((x-s_x)^2+(y-s_y)^2)\right).$$

For each symbol, the ALLR only requires calculation of distance to the most nearby reference constellation belonging to either S_0 *or* S_1. Although its complexity is significantly reduced in comparison with the exact LLR, for high order QAM constellations (e.g, 64 QAM, 1024 QAM), however the process of finding the right reference constellation points on communication hardware in real time is not obvious. There is not much related work in this aspect in the open scientific literature. In [23], authors further simplified the ALLR calculation by relying on a hard coded look-up table (LUT) specific to 16QAM. The solution consists of 2 substructions and 2 multiplications in the calculation; however it is not discussed how this type of simplification can be done for other modulation types and how the LUT size needs to scale. Also, authors only discussed polarity of LLR (hard bit) for performance evaluation, disregarding soft bit. Due to these limitations, it is hard to directly apply the method of [23] for our baseband receiver, which needs soft bit and the support for different QAM types.

We start from introducing another level of QAM agnostic simplification. We observe that for QAM modulation, e.g, 64-QAM, it is noticed that each row and column in Gray-coding format has at least 1 bit in common, and the first half of bits (first 3 bits in the 6-bit 64-QAM symbols in this case) are always associated with I-direction magnitude, while the second half of bits are always associated with Q-direction magnitude. It is hence logical that for a bit determined by I-direction, its $S_0$and S_1have the same imaginary part. Viceversa, for a bit determined by Q-direction, its $S_0$and S_1have the same real part. Then the identical parts would cancel out. By combining these facts, we can further simplify the equation as below:

$$L(b) = \frac{1}{\sigma^2}({S_0}' - {S_1}')(2r' - ({S_0}' + {S_1}'))$$

where r' is the real or imaginary value of the received symbol, depending on the concerned bit position. For the example of QPSK, we let r'=*Real*(r) when calculating the ALLR for first bit, and r'=*Imag*(r) when calculating the ALLR for the second bit. The same applies to S_0 and S_1. This expression contains 2 multiplications and 2 subtractions; however, since multiply by 2 is easily achieved in hardware by bit shifting by 1 towards the left, we can neglect this multiplication, which leads to a lower resource consumption than [23].

In order to search for the nearest S_0 and S_1, a dedicated hardware module is to be designed. In the first step, for 16-QAM the constellation plain is fragmented into cells that consists of only 4 points. The cell where the symbol belongs (as indicated by red dashed line in Figure 24) is found by comparing its *x* and *y* coordinates. In the second step, if among the 4 points in a cell both subsets of S_0 and S_1 exist, then we select 2 out of the 4 points in the cell as the most nearby constellation to make the bit 0 and 1 respectively. A comparison is made based on calculation of distance square and a LUT containing the bit map of reference constellation points. However, in case all points

in a cell belong to the same set (either S_0 or S_1), then the selection needs to carry on further. Intrinsically, constellation points making the bit into opposite value exist in cells that are in the same row or column as the current cell. Hence the search does not need to happen over all reference points. In most cases, it is simply the neighboring cell, or with one adjacent cell in between.

After finding S_0 and S_1, we also need the estimation of noise variance σ^2. In most wireless standards, communication includes the usage of reference sequences, and noise or interference can be then estimated based on those. The advantage of this approach is that the variance is estimated at the same time when packets are being received, no additional samples need to be collected. Though this estimation is in time domain, for OFDM modulation, the variance of each sub carrier is the overall noise variance in time domain divided by the corresponding CSI of a subcarrier, i.e.

$$\sigma_k^2 = \frac{\sigma^2}{CSI_k}$$

where *k* denotes the subcarrier index. In this way, the subcarrier with less noise and interference has higher ALLR than otherwise, and soft bits based on ALLR makes the subsequent decoding process much more robust. Here we however have assumed that interference or jammer's signal power also follows Gaussian distribution, which might not always be the case. Hence further research is foreseen to investigate the performance when jammer's signal does not comply to this assumption.

Finally, after ALLR is computed, it should be delivered to decoder FEC capability, which expects the input to be delivered in a finite amount of bits. This means the ALLR needs to be quantized before it can be used by the next module. For example, in the case of OpenWiFi platform, the FEC decoder is the Viterbi decoder, and it is configured to accept 3 bit input. Typically, FEC decoders accept no more than 3 bit input, which allows the value between 0 to 7. How the ALLR with large dynamic range is quantized into these values greatly affects the final performance of the decoder.

The LLR quantization is a well-studied problem. In [24], the authors aim to maximize mutual information between LLR and its corresponding bits to optimize the LLR quantization scheme. For implementation, the authors suggest that the optimal quantization level is to be obtained by curve fitting, depending on the LLR's variance. This is however challenging for implementation in hardware. Although it is possible to store some discrete value of the curve in a LUT, the finite resolution of the LUT will cause errors; hence, some additional mechanism for interpolation between the LUT entries is needed, which may affect the performance and the decoding latency. Secondly, to use the LUT, the LLR variance must be known, but there is no indication how this can be estimated. Of course, statistics of LLR can be first collected after a packet is completely received, and then start the quantization process of LLR. However, this can significantly increase the decoding latency. As real-life baseband decoder usually works in a pipelining approach, i.e when LLR is computed, it should be quantized right afterwards to be fed to the FEC decoder. Due to these considerations, we did not apply the optimal LLR quantization levels directly from work such as [24].

Another type of work is using modern AI approach for LLR quantization. For instance, [25] presents a deep learning-based method for LLR quantization. The quantizer is essentially composed by a deep autoencoder and a decoder, which reconstructs the LLR in a lower number of bits, based on the intermediate output, referred to as latent representation. The latent representation is obtained during training. However, in this work the autoencoder is only trained for 256 QAM, LPDC coding with a Rayleigh fading channel for each OFDM symbol. Generalization to other modulation and coding schemes is not addressed, neither memory aspects are discussed. Even after the

deep encoder's compression, the resulting LLR still needs at least 9 bits, which might be a significantly high value. Hence, due to the overhead of training, the complexity of the autoencoder, as well as the insufficient compression rate for the output, this kind of solution is also very hard to be applied on real-time communication hardware.

Therefore, an ALLR quantizer is designed from scratch, taking into account the need to support different QAM types and the varying amplitude of the input signal. The quantization is linear, and the step is proportional to the sum of CSI magnitude square of all sub carriers, and inversely proportional to the noise variance and a factor determined by the type of QAM. The exact formula to determine quantization step is shown below:

$$Q_{step} = \frac{\eta \sum_k CSI_k^2}{{\sigma_k}^2 \theta_{qam}}$$

where the sum is taken over the set of subcarriers where the reference signal is mapped. In the above equation, the θ_{qam} is a coefficient determined by the QAM type. The coefficient essentially reflects the distance between two adjacent constellation points for a specific QAM, which is predefined by a communication protocol. For example, in Wi-Fi it is equal to 1 for BPSK, 2 for QPSK, 10 for 16 QAM and 42 for 64 QAM. The other coefficient η however is a constant coefficient we determined heuristically to scale the quantization step according to the input signal strength, represented by the CSI. For every packet, a new quantization step is calculated using the above equation. The process to calculate CSI squared sum happens in parallel when CSI for each subcarrier is obtained, and it only consumes one instance to obtain CSI magnitude, and an accumulator to calculate sum sequentially. The coefficient η is 0.1875, equal to 1/8 plus 1/16, the multiplication of this coefficient is achieved through bit shifting. The multiplication with θ_{qam} for QAM 16 and QAM 64 however is achieved through a real multiplier. Last, a dedicated divider is needed to get the final value of Q_{step}.

2.4.2.2 Performance evaluation

In order to verify the performance of the proposed solution, we have implemented two types of demapper: the threshold demapper which determines soft bits by comparing symbol's distance against predefined threshold; and the ALLR based demapper, which determines soft bits using quantized ALLR. The remaining part of the baseband receiver [26] uses OFDM modulation of 64 sub carriers, with BPSK, QPSK, 16 QAM and 64 QAM. In terms of coding, convolutional codes with the following polynomials are applied (i.e., $G_0 = 133_8, G_1 = 171_8$).

Bit true model of the baseband receiver is built correspondingly in MATLAB environment. Figure 25 shows the packet error rate (PER) results at various SNRs, when using ALLR based demapper (in red line) and threshold based demapper (in blue line). For each SNR, the simulation consists of 2000 packets filtered with different channel response, packet payload is 4000 Byte, and 10 kHz carrier frequency offset are applied.

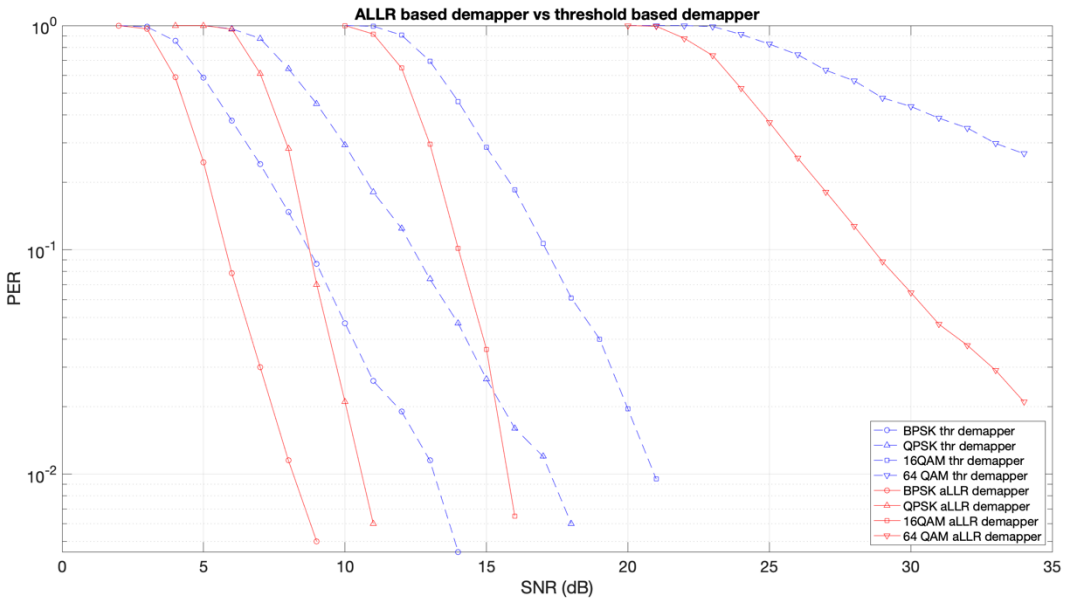


FIGURE 25. OPENWIFI BIT TRUE PER SIMULATION: THRESHOLD VS ALLR BASED DEMAPPER

We observe significant improvement when comparing the blue (threshold based demapper) and red (LLR base demapper) curve, on average an improvement of 4 dB is observed for different types of modulation at 10% PER. This is because in comparison to ALLR based demapper, the threshold based demapper does not consider the property of noise or interference. The generated soft bits could look like a confident 1 or 0, but in reality, it is actually an error caused by noise or interference. The Viterbi decoder however will be misled by this confidence indicated in soft bit. In the ALLR based demapper, the soft bit would correctly indicate the confidence of the concerned bit by considering the input signal strength, noise/interference on a specific sub carrier, hence producing much better decoding performance. The baseband design is also prototyped on an FPGA of software-defined radio, which is studied in WP5.

Based on the research result of this subsection, we recommend that receivers for 6G in-X subnetwork use a ALLR based demapper rather than the standard threshold based demapper, to make the baseband packet decoding process more robust against noise and interference.

2.4.2.3 Next steps

As a future activity, it is the plan to first verify the performance ALLR based demapper in non-Gaussian type of interference and extend the design if necessary.

3 MEDIUM ACCESS ENHANCEMENTS AND SERVICE MULTIPLEXING

The demanding communication requirements of in-X subnetworks call for efficient MAC mechanisms, that are presented in this section. First, low-latency MAC enhancements are presented. They include the use of proximity grouping for pre-emptive retransmissions and scheduling, enhancements to the 5G configured grant mechanisms, and coded random-access schemes. In-X subnetworks are also more suitable for the implementation of flexible/full duplexing node given their low-power characteristics compared with wide area networks; we discuss the opportunity of leveraging such capabilities for multiplexing services with diverse requirements. The controlled operation of subnetwork in entities like robots or vehicles offers opportunities for predictive MAC scheduler; we present the advancements towards achieving efficient context information and coverage assessment using AI/ML techniques, mobility-aware scheduling in industrial scenarios, and deterministic traffic handling in in-vehicle networks. Multi-link enhancements for improved reliability of in-X subnetworks are also discussed, including the use of cooperative communication and network coded operations. Finally, this section introduces innovative solutions for improved channel access and reliability for subnetworks operating in the unlicensed spectrum.

3.1 LOW LATENCY MAC ENHANCEMENTS

Three different approaches for reducing latencies in subnetwork supporting time-critical services are presented in this section: proximity grouping for pre-emptive retransmissions, configured grant enhancement, and coded random access. The proximity grouping solution can be implemented as an enhancement to traditional dynamic grant-based procedures, while the configured grant enhancement solution is tailored to periodic traffic with pre-allocated resources. The proposed coded random-access solution is instead intended to improve latency of fully grant-free solutions.

3.1.1 Proximity grouping for pre-emptive retransmissions & preventive scheduling

We propose here a novel scheme for reducing the delays associated to scheduling, by leveraging the characteristics of close proximity of devices served by in-X subnetworks. In order to highlight the benefits of the proposed proximity grouping schemes, in the next subsection we will first recall how transmission mechanisms are defined in 3GPP systems, considering communication between a 5G base station (called a “gNB”) and the User Equipment (UE).

3.1.1.1 5G HARQ Transmissions

In the current 3GPP 5G systems, Hybrid Automatic Repeat Request (HARQ) retransmissions are used for physical channels carrying data such as physical downlink shared channel (PDSCH) in the downlink and physical uplink shared channel (PUSCH) in the uplink. The HARQ protocol is based on feedback indications from the receiver and retransmissions by the transmitter. Here, an initial transmission of a physical channel is transmitted to a receiver and the receiver will feedback an ACK (positive acknowledgement) if it successfully decodes the physical channel, otherwise it feedbacks a NACK (negative acknowledgement). In case of a NACK reception, a retransmission of the physical channel may take place; in this case, the receiver may soft combine the LLRs (soft bits) after demodulation at the receiver, of the retransmitted physical channel with all previous transmissions of the same physical channel,

thereby increasing its effective SNR and increasing the chance of correct decoding. There is typically a maximum number of retransmissions of a physical channel before the transmission is abandoned.

An example of the legacy HARQ retransmission mechanism is shown in Figure 26, where frequency division duplex (FDD) mode is assumed, i.e. UL and DL transmissions happen over paired bands. The time labels t_1, t_2, …, etc, indicates events of interest, e.g., a packet transmission started or ended, a beginning of a slot. It should be noted that in Figure 26, the time label t_7 and t_8 are added so that it is easier to compare with the timing in Figure 27, since Figure 27 shows the same scenario as that in Figure 26 but using an enhanced HARQ transmission method. The downlink grants in the Dowlink Control Channel (DCI) #1 and DCI#2 schedule downlink channels PDSCH#1 and PDSCH#2 in physical resources in Slot *n* and Slot *n*+1 respectively, and their HARQ feedbacks are multiplexed and transmitted in the physical uplink control channel (PUCCH)#1. In this example, the UE fails to decode PDSCH#1 and feedbacks a NACK in PUCCH#1. Consequently, the gNB sends another downlink grant in DCI#3 to schedule a retransmission for PDSCH#1, where the UE soft combines this PDSCH#1 with the previous PDSCH#1. Here the UE successfully decodes it and feedbacks an ACK in PUCCH#2. The time for the UE to successfully receive PDSCH#1 is $t_{14} - t_2$ ms, in Figure 26. By assuming an OFDM numerology with a 240 kHz subcarrier spacing, this corresponds to a 2.5 ms time.

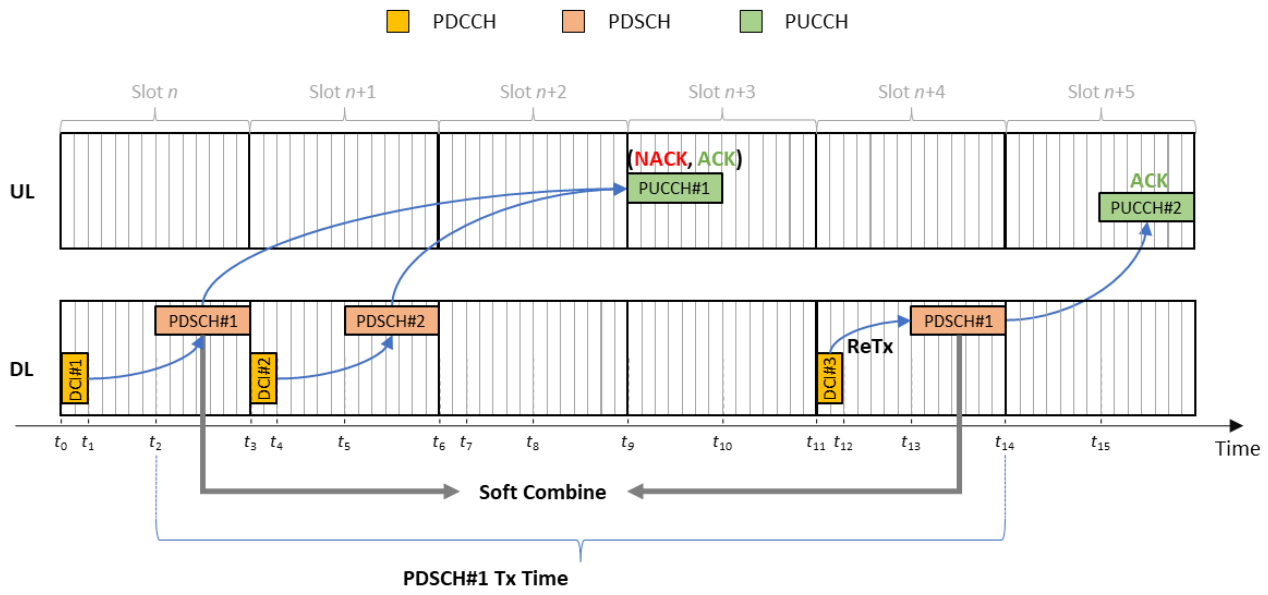


FIGURE 26. 5G HARQ RETRANSMISSIONS

3.1.1.2 Fast HARQ feedback

Since the transmitter needs to receive a NACK from the receiver before it will retransmit the packet, retransmission incurs delay. Whilst such delay is tolerable in some services, such as enhanced Mobile Broadband (eMBB), it may not meet the low latency requirement of URLLC services. Recognising this, methods using fast HARQ feedback have been proposed for 5G but not implemented.

A method prosed in [27] is to feedback an estimated HARQ feedback based on the LLRs (soft bits) of the received packet, in addition to a HARQ feedback based on the fully decoded packet. Since the soft bits can be processed faster and hence earlier than decoding the entire packet and checking the cyclic redundancy code (CRC), the estimated HARQ feedback can be transmitted earlier. Figure 27 shows an example implementation of fast estimated HARQ feedback using the same scenario as that in Figure 26. Similar to the previous scenario, the gNB transmits DCI#1 and DCI#2 to schedule PDSCH#1 and PDSCH#2, respectively, where their HARQ feedback messages are scheduled to be transmitted in PUCCH#2. Here, instead of decoding the entire PDSCH#1 channel, the UE estimates using LLR that the UE will likely fail to decode PDSCH#1 and sends out a fast HARQ feedback in PUCCH#1. The gNB then transmits a downlink grant in DCI#3 to schedule a retransmission for PDSCH#1 in Slot n+2, prior to receiving PUCCH#2, which contains the actual decoding outcome of PDSCH#1. The UE soft combines the retransmitted PDSCH#1 with its previously received PDSCH#1, which leads to a successful decoding of PDSCH#1, and transmits the HARQ feedback in PUCCH#3. After estimating the decoding outcome of the first (initial) PDSCH#1 channel, the UE continues to decode the channel completely and fails to decode it, which would also indicate to the gNB that the estimation of the unsuccessful decoding of PDSCH#1 was correct. The UE feedbacks the actual decoding outcome of the first PDSCH#1 in PUCCH#2, to confirm that its estimation was correct. It can be observed that using the estimate of the decoding outcome of a channel enables a faster HARQ feedback which results in a reduced latency in scheduling a retransmission. In this example, the latency is reduced from $t_{14} - t_2$ to $t_9 - t_2$,. By assuming an OFDM numerology with a 240 kHz subcarrier spacing, this would correspond to a reduction from 2.5 ms to 1.25 ms.

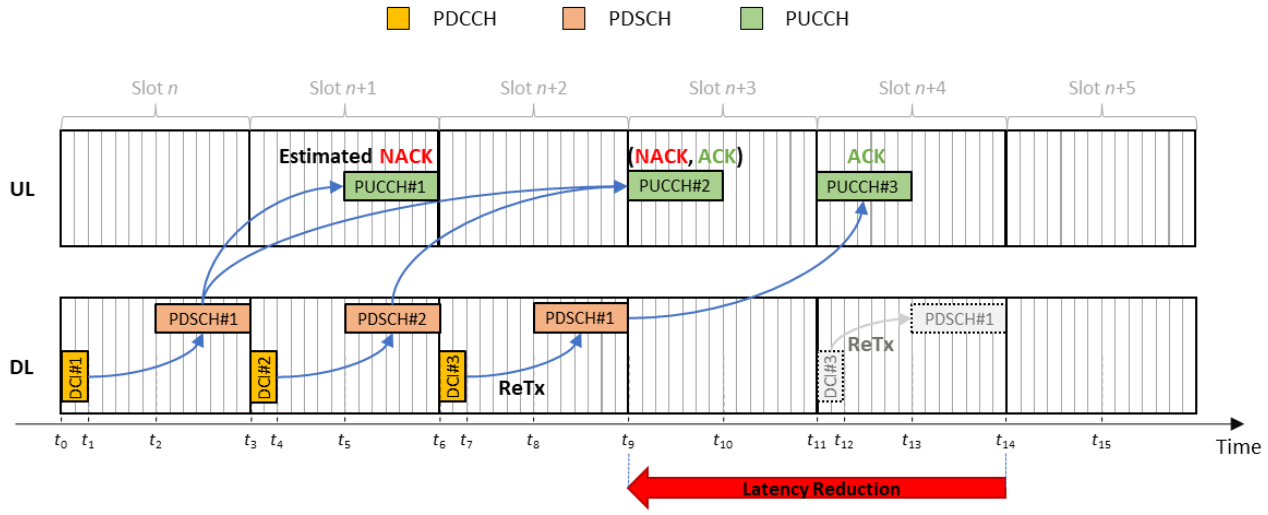


FIGURE 27. FAST ESTIMATED HARQ FEEDBACK

Methods employing fast HARQ feedback require uplink resources to be available for the UE to transmit the estimated ACK or NACK, which is feasible in an FDD operation such as in the example in Figure 26 and Figure 27. However, in time division duplexing (TDD) operations, uplink resource may not be available when the UE needs to transmit an estimated NACK or ACK, and therefore fast HARQ feedback is not practical. Figure 28 shows a similar scenario as the one in Figure 27, but Figure 28 operates in a TDD mode. Similarly, downlink grants DCI#1 and DCI#2 schedule PDSCH#1 and PDSCH#2 respectively. The UE estimates that it is likely to fail the decoding of PDSCH#1. However, unlike in FDD, there is no uplink resource immediately available for the UE to transmit the estimated NACK (e.g., in Slot n+1), since the earliest uplink resource is in Slot n+4.

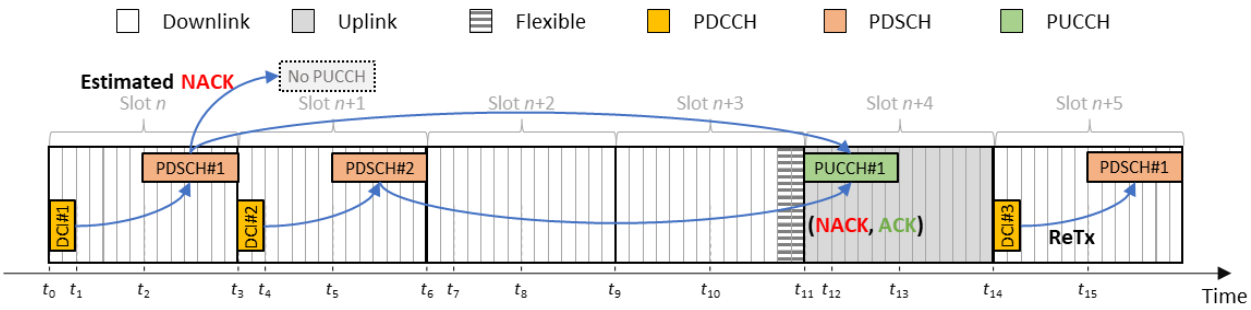


FIGURE 28. NO UPLINK RESOURCE FOR ESTIMATED NACK

In-X subnetworks are mainly expected to operate at high frequency carriers such as in the mmWave or Sub-THz region, where TDD is likely to be used. Given the extreme low latency requirement of 6G-SHINE, there is the need for a method that provides fast retransmission in TDD operation. In the following, we introduce a method to address the latency issue in providing HARQ feedback for in-X subnetworks operating in TDD mode.

3.1.1.3 Proposed Proximity Grouping scheme

Proximity grouping is proposed here to reduce the latency in retransmissions and to provide pre-emptive scheduling. A proximity group contains two or more devices that experience similar radio channel conditions. 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. An example of proximity grouping in a subnetwork is shown in Figure 29, consisting of a HC element that is connected to a core network and serves five SNEs. Here we will use the term HC to refers to an entity with functionality of an Access Point. In this example, two proximity groups are formed, where proximity group 1 contains SNE#1 and SNE#2, whilst proximity group 2 contains SNE#3, SNE#4 and SNE#5. The proximity grouping can be based on the SNEs' locations and/or the measured channel estimates relative to the HC.

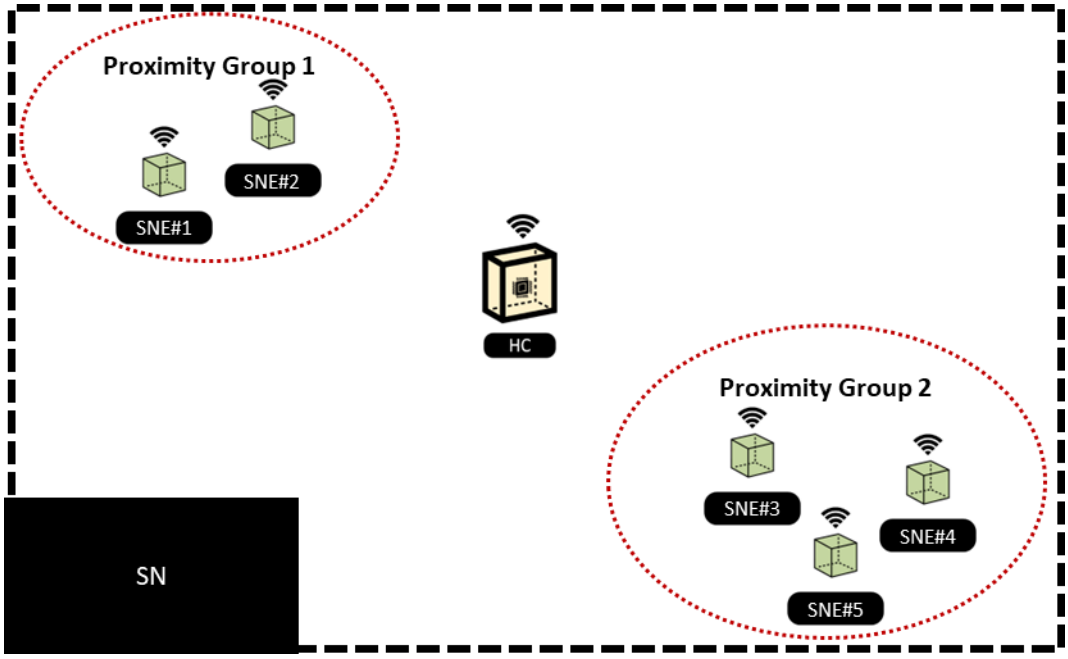


FIGURE 29. PROXIMITY GROUPS.

In pre-emptive retransmissions, the network schedules a retransmission to a UE for a downlink packet such as PDSCH, before it receives the HARQ feedback from that UE. Pre-emptive retransmission is highly beneficial in reducing the transmission latency in a TDD network. The network can utilise the proximity grouping property to provide pre-emptive retransmissions. An example is shown in Figure 30, where SNE#1 and SNE#2 are in a proximity group. The network transmits via the HC, DCI#1 and DCI#2 to schedule PDSCH#1 and PDSCH#2 to SNE#1 and SNE#2 respectively. PDSCH#1 is scheduled in Slot *n* whilst PDSCH#2 is scheduled in Slot *n*+2, after the uplink Slot *n*+1. SNE#1 fails to decode PDSCH#1 and consequently, SNE#1 feedbacks a NACK in PUCCH#1 in Slot *n*+1. Since SNE#2 is in the same proximity group as SNE#1, the network transmits DCI#3 in Slot *n*+2 to pre-emptively schedule a retransmission for PDSCH#2 for Slot *n*+3 before it could receive a HARQ feedback from SNE#2. SNE#2 soft combines the retransmitted PDSCH#2; in this example, it successfully decodes it and feedbacks an ACK in PUCCH#2 in Slot *n*+6. The network transmits DCI#4 in Slot *n*+3 to schedule a retransmission for PDSCH#1 in Slot *n*+4, where SNE#1 soft combines the two PDSCH#1 and successfully decodes it. SNE#1 then feedbacks an ACK in PUCCH#3 in Slot *n*+6. This method can at the minimum reduce the latency by a TDD cycle (consisting of a downlink to uplink cycle); for example, in a 3.125 ms TDD cycle of a 240 kHz subcarrier spacing system, where a TDD cycle consists of 5 slots (4 downlink slots and 1 uplink slot), this would reduce latency by 3.125 ms. It is worth to mention that, in a TDD cycle, the HARQ feedback for a PDSCH may not be transmitted in the uplink slot of that cycle, for example, due to lack of uplink resources in the uplink slot, and therefore an SNE may not transmits its HARQ feedback within the same TDD cycle. In this case, SNE#2 may not be able to transmit its HARQ feedback in Slot *n*+6 (if the resources in uplink Slot n+6 are congested), and therefore the reduction in latency will be a multiple of TDD cycles.

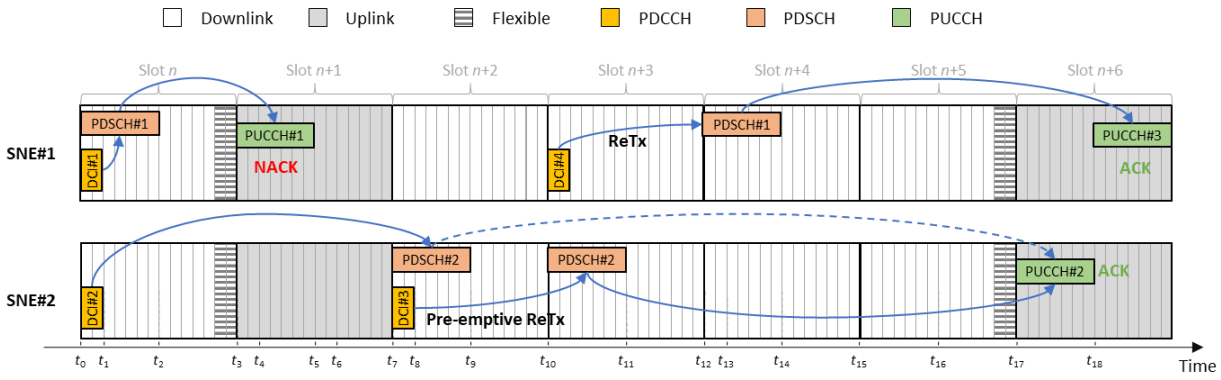


FIGURE 30. PRE-EMPTIVE RETRANSMISSION USING PROXIMITY GROUPING

In addition to pre-emptive retransmissions, proximity grouping also aids the network such as the HC, in performing preventive scheduling. If the network that wishes to schedule to a SNE, and another SNE in a proximity group fails to decode a packet, the network may schedule a more robust packet to that SNE. An example is shown in Figure 31, where SNE#1 and SNE#2 belong to the same proximity group. The network transmits DCI#1 to schedule SNE#1 with PDSCH#1, but SNE#1 fails to decode PDSCH#1 and consequently, feedbacks a NACK in PUCCH#1 in uplink Slot *n*+1. In Slot *n*+2, the network wishes to schedule SNE#2, and learning from the HARQ feedback from SNE#1, the network schedules a more robustly coded physical channel to SNE#2. Here, the network transmits DCI#3 to schedule PDSCH#2 with 2× repetitions for SNE#2 to increase the reliability of PDSCH#2. In this example, both SNE#1 and SNE#2 managed to decode their respective PDSCH after receiving two of their transmissions and they feedback an ACK each in uplink Slot *n*+6. It should be appreciated that this method will effectively eliminate any latency due to retransmission. This means, it can at least reduce the latency by a TDD cycle; for example, if the scenario in Figure 31 has a TDD downlink to uplink cycle of 3.125 ms, this method will at the minimum reduce the latency by 3.125 ms.

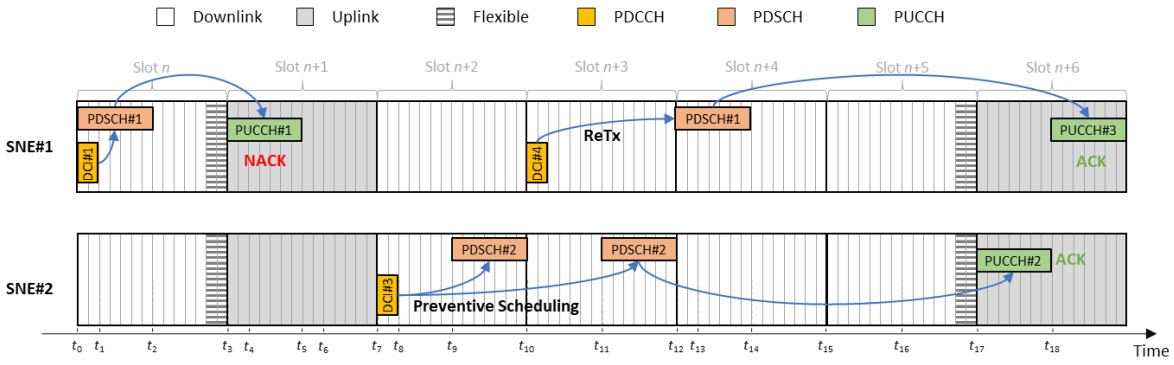


FIGURE 31. PREVENTIVE SCHEDULING USING PROXIMITY GROUPING

Proximity grouping can also be used for uplink transmissions. For example, if the HC, fails to decode an uplink transmission from an SNE, the HC can group broadcast to the SNEs in the proximity group such outcome, so that

other SNEs in the group may take preventive steps to avoid sending uplink transmissions that may fail. The SNE wishing to transmit to the HC may for example increase its transmission power or use a more robust MCS.

To implement proximity grouping, the HC in the subnetwork needs measurements from SNEs such as their locations and/or their channel estimations. For example, the HC can group the SNEs based on similar channel state information. Additionally, or alternatively, the HC may form a proximity group with the SNEs that fall within a predefined area in LOS conditions. Hence, new types of measurements and reporting may be needed to implement proximity grouping.

It may also be beneficial that a SNE transmits the maximum MCS that it can tolerate; in this way if the HC schedules an MCS higher than the maximum MCS, it is likely to fail the decoding under the current radio condition. This is beneficial for the HC as it might be able to schedule another SNE in the same proximity group.

3.1.1.4 Next steps

We will further investigate the details in signalling required between SNE and the HC to implement proximity grouping, especially in determining which SNEs to group. The maintenance of a proximity group also needs to be investigated, as SNEs may be mobile (e.g., sensors mounted over a robot arm) and may move out of a proximity group and into another proximity group.

Enhancements utilising the proximity group will be studied and the required signalling will be determined to support this enhancement. Enhancement may include SNEs in the proximity group collaborating in decoding downlink packets to improve reliability.

Quantitative evaluation will also be performed to evaluate the latency savings this method offered with respect to existing solutions.

3.1.2 Configured grant enhancements

The proximity group enhancements presented in the previous sub-section are based on the assumption that a typical scheduling procedure is done. In the uplink, this is based on scheduling request by the device and DCI message by the HC. Such procedure can be used for low latency non-periodic packet transmissions. An updated mechanism for periodic traffic is proposed in this subsection. The proposed method is based on Configured grant (CG), an uplink scheduling mechanism which allows a UE with periodic (deterministic) traffic to transmit its data by bypassing the traditional control messaging.

3.1.2.1 Configured grant in 3GPP

CG was introduced in 3GPP Release 15 being an enabler for URLLC, especially for industrial use cases. Through Release 16 and 17 there have been various enhancements, making the mechanism more flexible and providing better latency and reliability. Specifically, these enhancements include:

- **Enhanced transmission repetition for reliability:** Improving reliability through increased transmission repetitions.
- **Multiple CG configurations:** Enabling support for diverse traffic patterns with multiple CG configurations.

- **Dynamic CG configuration Activation/Release:** Adopting adaptive management of CG configurations to changing network demands.
- **PHY layer prioritization:** Ensuring timely delivery of critical payloads by prioritizing CG transmissions at the PHY layer.

By using CG, the resources for transmission can be granted through radio resource control (RRC) signaling just after the connection between the HC and the SNE has been established. These resources are repeated in a periodic basis so the SNE can use them again every time it has data to transmit.

To apply enhancements for industrial control use case, the 3GPP has introduced two types of CG: type 1 and type 2. For type 1 the resource allocation or the grant is provided directly by RRC parameters, while for type 2, similarly to semi-persistent scheduling (SPS), only the periodicity is provided by RRC signaling while the rest of the grant is activated by L1/L2 control messages, such as DCI. The obvious advantage of type 1 is that it takes effect immediately after receiving the configuration without waiting for a DCI and our aim is to enhance the Type 1 with the obvious latency advantage it offers. For type 1 CG in particular, the resources for transmission are provided by RRC *ConfiguredGrantConfig IE,* including the periodicity, symbols, Physical Resource Blocks (PRBs) for transmission but also the MCS and Demodulation Reference Signal (DMRS) configuration. After configured, the SNE can, based on the predefined periodicity, calculate a transmission opportunity along with the HARQ ID to be associated with this transmission. Figure 32 highlights the major steps of CG type 1.

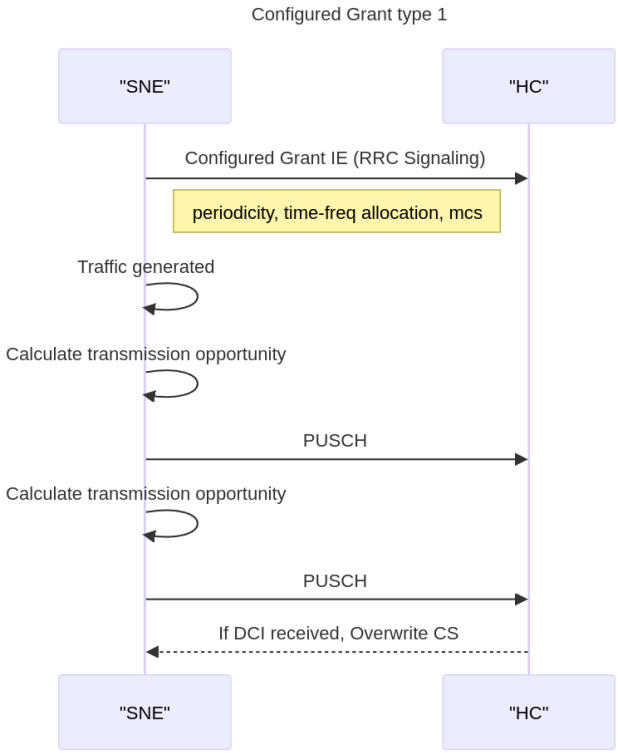


FIGURE 32. TYPE 1 CG SCHEDULING MECHANISMS.

Previous research in GF scheduling can be categorized into two classes: GF scheduling using (i) shared resources, and (ii) using dedicated resources. For sporadic traffic, periodic allocation of resources can lead to poor resource utilization because SNEs may have not data to transmit in the reserved resources. For these scenarios, shared resources can be allocated to many SNEs as proposed in many studies [28][29]. In case of periodic traffic patterns, the best choice is to allocate dedicated resources to the UEs matching their traffic demands. However, collisions can still occur as firstly described in [30], where the collision between periodicities is first addressed.

The effective management of uplink configured grant resources is crucial in order to follow the strict reliability and latency requirements. A significant issue can arise when SNEs with different periodicities are allocated the same resource blocks, leading to potential overlaps. In an optimized scenario a scheduler within a subnetwork would allocate distinct resources to each SNE to prevent any possibility of collisions. However, several practical factors could lead to situations where the scheduler could end up allocating a part of overlapping resources to different SNEs, e.g for resource optimization, or to support a large number of SNEs. For example, in a scenario where one CG SNE is configured with a 5 ms transmission periodicity and another one with a 10 ms periodicity a periodic overlapping of some resources could arise. These SNEs will inevitably reach a point of collision every 10 ms, where both require access to the same resources for data transmission. While this collision is harmful for both users, it is especially damaging for the SNE with the longer periodicity since this has to wait for longer time for a transmission opportunity. Addressing this conflict in configured grant scheduling is thus imperative to keep the needed reliability levels. Figure 33 highlights the event where 2 SNEs with different periodicities would still experience some conflict in their transmissions due to the overlap of their pre-scheduled transmissions.

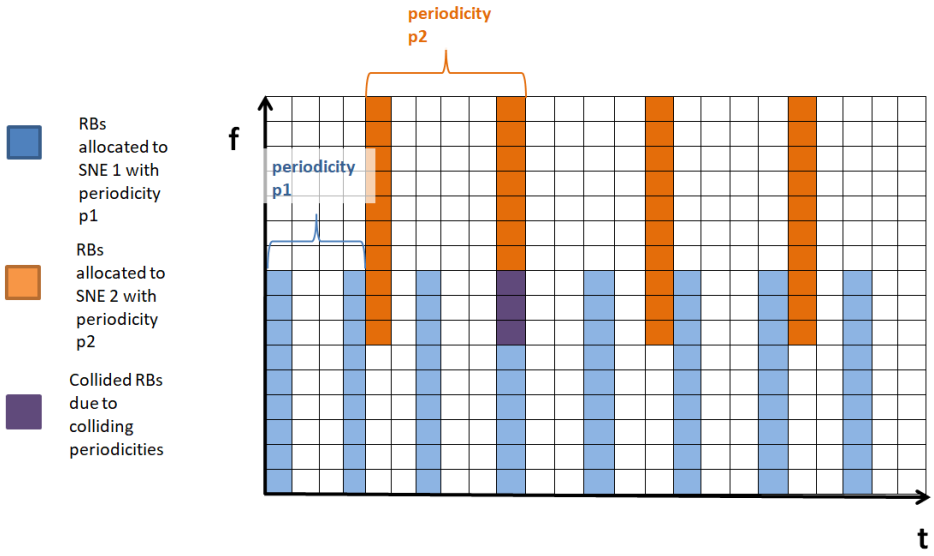


FIGURE 33. CONFLICT EXAMPLE - RESOURCE ALLOCATION OVERLAP IN SNE 1 AND SNE 2 [28]

3.1.2.2 Proposed enhancement

To mitigate the aforementioned challenge, a novel approach can be implemented at the SNE side. This involves the development of an intelligent algorithm at each SNE that constantly monitors for transmission success rates and identifies patterns which could mean periodic collisions. When a SNE perceives a regular disruption in its transmissions (for example a SNE with a 1 ms transmission periodicity experiences collisions every 10 ms) it can

imply a conflict with a SNE with a longer periodicity. The algorithm then follows a process where it must decide its action based on the comparison of its own periodicity with the detected collision pattern.

For SNEs with shorter periodicities, the algorithm can initiate an adjustment to their transmission, by skipping their transmissions to avoid the collision. These calculated skips ensure that collisions do not happen and as a result the reliability is not harmed.

On the other hand, for SNEs with longer or equal periodicity schedules, the algorithm follows a different strategy. These SNEs, upon detecting a collision pattern that aligns with their transmission periodicity, maintain their scheduled transmission times. This decision is based on the fact that SNEs with shorter periodicity have greater flexibility to adjust their transmission timings since they will soon have a transmission opportunity than the SNEs with higher periodicities. The main process of the proposed idea is depicted in Figure 34.

To detect collision patterns, which should occur at the observation window of the algorithm, a systematic approach to monitoring and analyzing the sequence of acknowledgments (ACKs) post-transmission is followed. When a SNE successfully transmits data, it is acknowledged by the base station (ACK). In the case of CG, HARQ is implicit meaning the SNE assumes ACK if not DCI with NACK is received (indicating a request for retransmission) in a predefined time period. An absence of such an acknowledgement, suggests a failure in transmission, which could be due to various factors including deep fading, interference, or a collision. The algorithm is designed to differentiate between these causes by detecting the regularity and timing of these failures. For instance, sporadic failures may indicate fading issues, while consistent failures at regular times, such as every 10 milliseconds, may suggest a collision due to overlapping transmission schedules with another SNE. By statistically analyzing the timing of these unsuccessful transmissions and correlating them with the SNE's own transmission periodicity, the algorithm can detect a collision pattern. In implementing these decisions, the algorithm may utilize an ML technique or even the Fast Fourier Transform (FFT), as a tool for analyzing the frequency domain representation of the transmission feedback. Peaks at the spectrum of FFT can reveal the regular intervals at which collisions occur.

For instance, if NACKs are consistently received at intervals that match or are multiples of the SNE's periodicity, the algorithm interprets this as a strong indicator of periodic collisions rather than random occurrences or failures due to other factors like fading or interference. If a pattern is recognized, the algorithm confirms this through continuous monitoring over several cycles, and, upon validation, it proactively adjusts the SNE’s transmission schedule to skip certain transmissions in order to avoid future collisions.

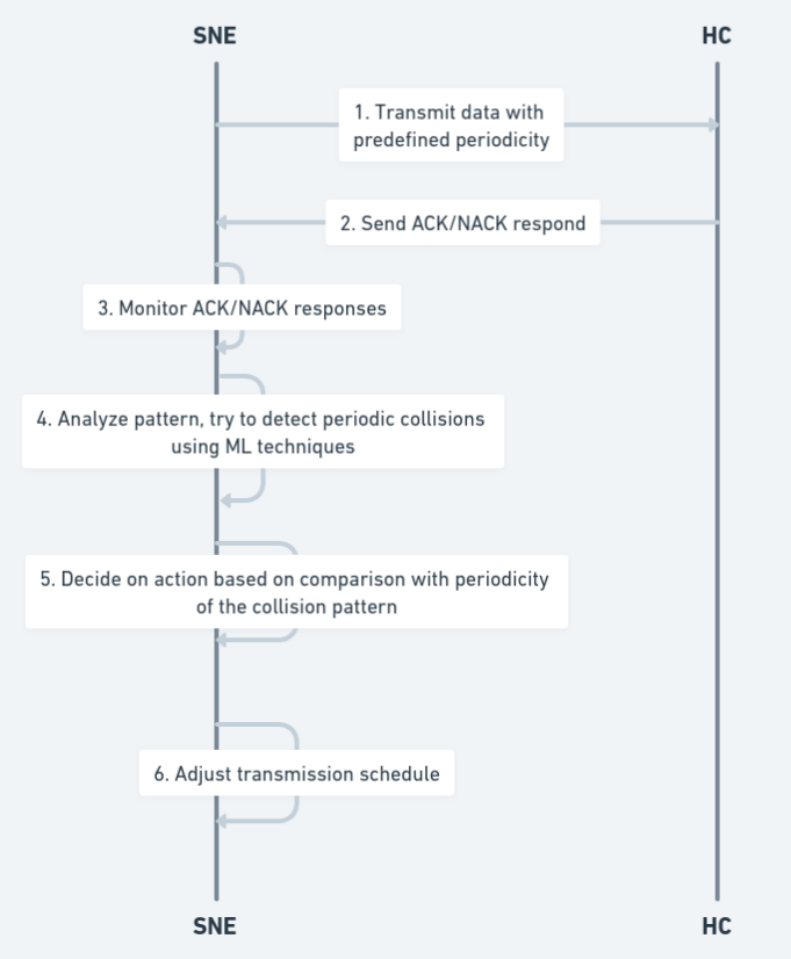


FIGURE 34. ALGORITHM TO ENHANCE CG FOR PERIODIC COMMUNICATIONS

The fairness of the proposed algorithm is one of its strengths. When a collision is detected, the algorithm prioritizes adjustments for SNEs with shorter periodicities, which implies skipping transmissions on the instances where a collision is expected to happen. The approach of skipping transmissions for the shorter periodicities SNEs, is preferred because these SNEs transmit more often, and their schedules can typically accommodate adjustments without a significant impact on their performance. Alternatively, for SNEs with longer periodicity schedules, maintaining their scheduled transmission times is generally preferred, as their less frequent transmissions mean that each opportunity to transmit is more critical and losing one would lead to a greater performance degradation.

3.1.2.3 Next steps

In the upcoming phase towards D3.3, we will focus on testing the enhanced configured grant mechanisms, specifically evaluating its reliability enhancements and effectiveness in reducing collisions in a multi-SNE environment. Performance results will be gathered, using srsRAN open-source testing environment [31]. Results will provide insights into the effectiveness of the proposed algorithm and the challenge to manage the periodicity of several SNE connections to HC in subnetworks.

3.1.3 Coded Random Access

The solutions presented above are intended as enhancements for improving latencies in in-X subnetworks either based on traditional grant approaches or configured grant for periodic type of traffic. Here, we present a solution for improving latency for SNEs operating in a truly grant-free manner, i.e. the SNEs can initiate their transmission at any time without receiving any grant from the serving HC. While the schemes presented above are intended as enhancements to current 3GPP specification, or rely on 3GPP concepts such as configured grant, the method here presented is in principle agnostic to 3GPP specifications. Still, it can in principle be integrated into the current 5G air interface, where critical transmissions can eventually start at any OFDM symbol in the frame ("mini-slot" transmission).

The presented solution is based on the idea of coded slotted ALOHA (CSA) [32], and is strategically designed to reduce latency and enhance the overall communication reliability in a subnetwork. The main use cases of interest are industrial, and in particular robot control, where each operational robot is expected to be equipped with numerous SNEs (e.g., sensors and actuators). Each of these components operates with very low communication cycles, possibly as low as 50 μs, with the primary goal of ensuring fast and accurate movement commands. The control information, typically encapsulated into short packets of 200-300 bytes, must exhibit an extremely high level of reliability. The system can be designed to guarantee a probability of encountering two consecutive errors over the same packet of less than 10^{-6}.

3.1.3.1 Coded Slotted ALOHA (CSA)

CSA is a multiple access protocol based on erasure coding and successive interference cancellation, able to bridge multiple access with iteratively decodable modern codes on sparse graphs and therefore achieving high reliabilities. Packet erasure correcting codes (or, simply, packet erasure codes) are building blocks for the construction of CSA protocols.

A packet erasure encoder receives in input a sequence of k packets, hereafter called "information packets" and all having the same size (say m bits) and produces a sequence of $n > k$ packets, called "encoded packets" and again of size m bits each. The number of input packets, k, is called the code dimension. The number of output packets n, instead, is called the code length. The rate of the erasure code is k/n. A packet erasure code with dimension k and length n is referred to as an (n, k) code. Note that if less than k packets (of the total n) are decoded there is no hope of reconstructing the original information packets. The encoded packets are generated in such a way that, if some of them are erased, the known ones allow retrieving the original information packets. This latter operation is called erasure decoding. Optimum erasure correcting codes are called maximum distance separable (MDS) codes [33], holding the property that the original k information packets can be retrieved from any subset of the encoded packets. Several classes of MDS codes are known, for example algebraic-based ones (e.g., Reed-Solomon codes and their shortened versions) or graph-based ones. Usually, the packet erasure encoder is "systematic" meaning that the first k encoded packets coincide with the information packets. In this case, the last $n - k$ encoded packets are called redundancy. Packet erasure encoders in practical applications are usually systematic and linear. Here, linear means that every encoded packet can be obtained as a linear combination of a subset of the information packets. The coefficients of the linear combination belong to a finite field (binary or, most often, non-binary). The simplest examples of packet-level MDS codes are repetition codes, where $k = 1$ and each of the $n - 1$ encoded packets is a copy of the unique information packet, and by a packet parity-check code,

where instead $k = n - 1$ and the unique redundant packet is obtained as a bit-wise XOR of the information packets.

The proposed scheme is a modification of the well-known uncoordinated access protocol called CSA. In conventional CSA, the time is slotted and framed (with N_s slots per frame), and the active SNEs are slot- and frame-synchronous. Each slot has a time duration of T_s and is partitioned into k sub-slots (or "slices"), each of duration T_s/k . The device-side access protocol of conventional CSA, from wake up to transmission, is exemplified in the following and pictorially represented in Figure 35.

Each SNE, being active over a frame, splits its data packet (of length M bits) into k data fragments (or segments), each of length $m = M/k$ bits. These data segments are encoded via a packet-level linear erasure correcting code of dimension k and length n , yielding n encoded fragments (hereafter simply fragments), each one again of length m bits. The (n, k) erasure correcting code, employed by the active SNE to generate the n segments, is drawn randomly from an ensemble $C = \{C_1, C_2, \dots, C_r\}$ of r "component codes", this set being known to all SNEs and to the HC (e.g., an access point). The h -th code in the ensemble, C_h has code rate k/n . Any time a SNE is active, it draws its erasure code C_h from the set C without any coordination with the other active SNEs and independently of its choices in past activation events. Fragments are indexed by $s = \{1, \dots, n\}$. The SNE then draws randomly and without replacement n integers between 1 and $k \cdot N_s$, representing the indexes of the sub-slots where the n segments are transmitted by the SNE. The generic segment is equipped with its index s , the index h of the erasure code used to generate it, and with pointers to the positions of the other $n - 1$ segments. The segment bits are further encoded via a physical layer channel code and then mapped onto complex constellation symbols; a synchronization preamble is prepended to the complex payload, and the so obtained burst is transmitted in the corresponding sub-slot. Neglecting guard times, the time duration of a segment is T_s/k and coincides with that of a slice (i.e., sub-slot). In the example depicted in the figure, the user MAC payload is fragmented into $k = 3$ segments and $(5,3)$ packet erasure code with linear and systematic encoding is applied to generate $n = 5$ encoded segments, of which $n - k = 2$ are redundancy segments.

The HC data recovery procedure is described hereafter. It is based on combining segment erasure decoding and successive interference cancellation across slices. The frame is first processed by the HC in a slice-by-slice fashion. All segments that are successfully recovered in a sub-slot (after physical layer processing that includes segment detection, channel estimation, demapping, and physical layer decoding) are stored in memory by the HC. For each such segment, the HC extracts the erasure code index h , the position index s , and the pointers to the positions of the other $n - 1$ segments. This way, the HC can perform an association between the current segment and those segments, transmitted by the same SNE, that have been recovered in the previously processed sub-slots. All of these segments are characterized by the same erasure code index h and by different position indexes s . The HC then attempts packet erasure decoding for code h using the so far available segments as input. If successful, packet erasure decoding allows recovering all segments, including the ones unsuccessfully received in previous sub-slots and the ones to be received in future sub-slots; when this is the case, the whole SNE data packet is recovered and immediately forwarded to the upper layers. At the end of the first pass on the frame, some segments have been recovered by the HC. The complex samples associated with each of these segments are then subtracted from the slices in which they have been received: note that this operation, called interference subtraction, requires accurate physical layer processing. Moreover, all segments corresponding to a SNE packet that has been sent to the upper layers are removed from memory. A new iteration is then triggered, in which new segments may be successfully recovered from slices where interference subtraction was performed, possibly yielding the recovery of new SNE data packets after erasure decoding. This process is iterated until the memory is empty.

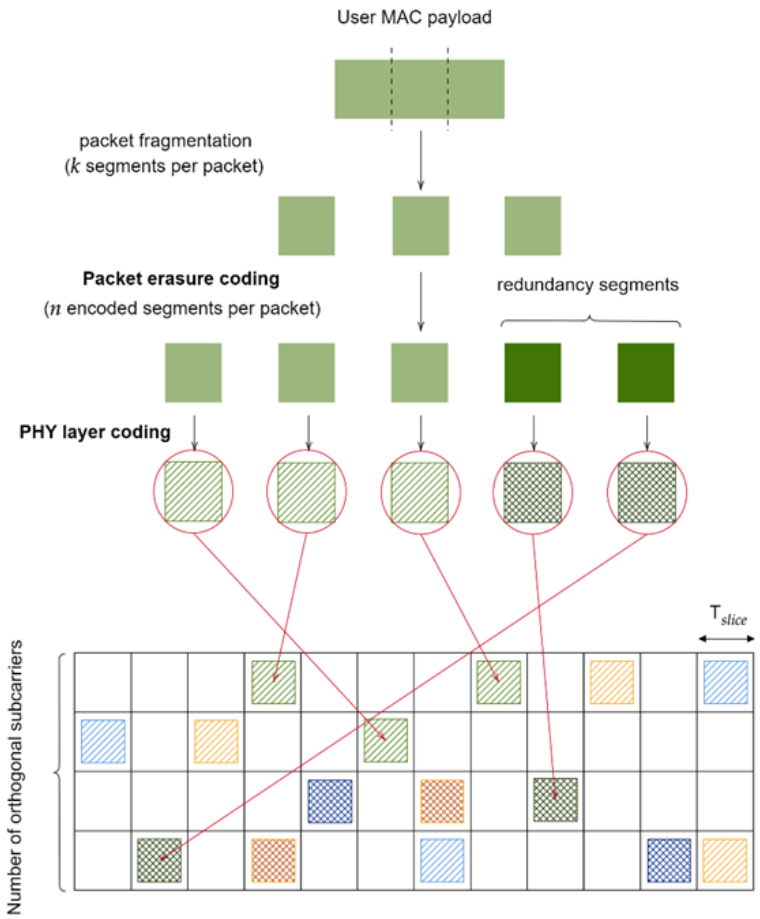


FIGURE 35. ENCODING PROCEDURE FOR CONVENTIONAL CSA.

3.1.3.2 CSA for Ultra-Reliable and Low-Latency Communications

The iterative recovery procedure characterizing CSA has been well-investigated in literature and has been shown to exhibit a thresholding phenomenon similar to the one known for modern codes with iterative decoding (such as LDPC codes). Actually, in fact, the successive interference cancellation process of CSA admits a graphical interpretation that resembles the so called “peeling decoder” of LDPC codes. This thresholding phenomenon may be roughly summarized as follows: if the load on the CSA frame, defined as the ratio between the number of contending users and the number of slots, is larger than a certain threshold, then the packet loss probability is high (equivalently, reliability is low); in contrast, if the load becomes smaller that the threshold, then the SIC process achieves very low packet loss probability values. The thresholding load value depends on several factors, among which the choice of the erasure codes employed by the users. Despite the very promising reliability levels it can ensure, ordinary CSA relies on successive interference cancellation across the slots of a frame, so that some of the packets are recovered with a delay that is not compliant with low-latency communication requirements. To cope with this issue, a modified CSA scheme is introduced. This is addressed in the following.

To meet the latency and reliability requirements of the wireless robot control use case, we propose a modified version of the CSA protocol specifically thought to address the associated constraints of communication cycle duration (e.g., 50 μs) and packet loss probability (two consecutive errors over the same packet with less than 10^{-6}). In terms of the device-side access mechanism, a key modification focuses on the packet erasure coding step, featuring a "normal-priority" erasure code ensemble, C_L, consisting of a set of (n_L, k) erasure correcting codes, together with a "high-priority" one, C_H, comprising a set of (n_H, k) erasure correcting codes, where $n_H > n_L$. As the reliability constraint translates into a requirement on the probability of encountering two consecutive errors, each active SNE follows a specific operational sequence. When the active SNE transmits a packet for the first time, it randomly selects a packet erasure correcting code from the ensemble C_L, thus transmitting n_L encoded fragments (in the same number of sub-slots), following PHY layer encoding. The probability distribution according to which one erasure code is drawn from C_L represents a degree of freedom of the system and can be designed to improve the system performance. At the end of each transmission, the HC acknowledges SNEs whose packets were correctly received. If no acknowledgment is received, indicating packet loss, the SNE attempts packet transmission again, employing this time an erasure code belonging to the "high priority" subset C_H; again, the probability distribution from which one erasure code is drawn from C_H can be appropriately designed. Assuming that all erasure codes are maximum distance separable (MDS) codes, the packet is correctly received whenever at least k fragments out of n_L (in the first transmission) or n_H (in the second transmissions) are correctly received. The described approach increases the probability of decoding at least k fragments (out of n_H), thereby reducing the likelihood of losing the consecutive packet.

The second modification with respect to standard CSA pertains to the fragment transmission procedure, as shown in Figure 36. Unlike the conventional protocol, where fragments are randomly transmitted across all possible slices of a frame (from 1 to $k \cdot N_s$), in the modified protocol each active SNE transmits its information message over a "virtual frame" of a predefined length $W \leq k \cdot N_s$ that is started in the first available sub-slot after activation. Upon waking up, the SNE opens a virtual frame and randomly selects, without replacement, n_L (or n_H, depending on the situation) integers between N_{slice} and $N_{slice} + W$, where N_{slice} is the index of the current slice and W is the virtual frame size. The size W of the virtual frame is selected to align with the communication cycle of the SNEs, ensuring compliance with the latency constraint. The introduction of the virtual frame presents challenges for the decoding procedure at the HC. To meet the latency constraint, successive interference cancellation decoding must be performed in a sliding window fashion, over a window of length equal to the frame size W . This sliding window decoding procedure allows the HC to perform interference subtraction over the window, with the processing conducted in a slice-by-slice fashion. As the window slides through the virtual frame, interference subtraction is carried out in real-time, enhancing the robustness of the decoding process.

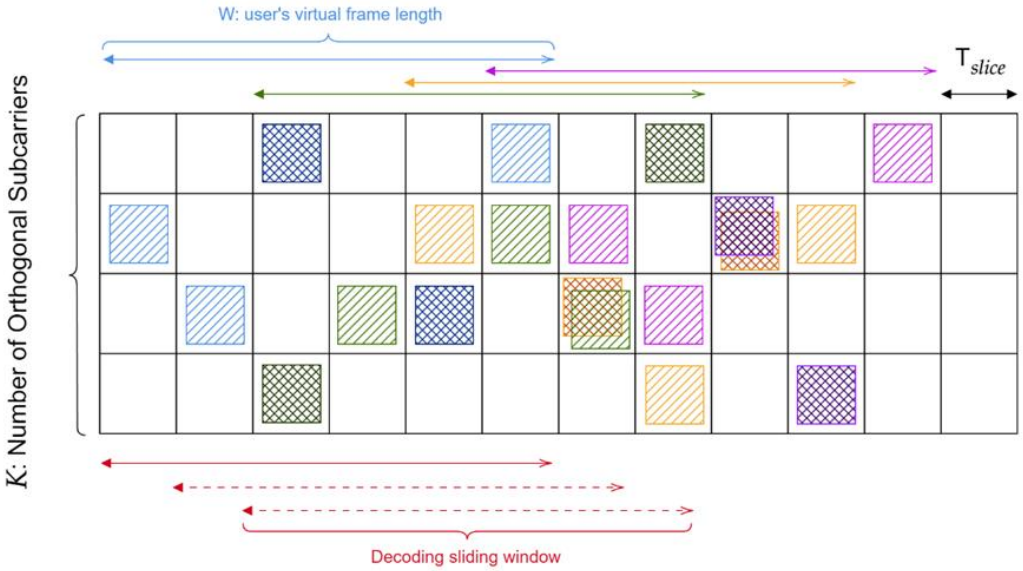


FIGURE 36. ENHANCED CSA FOR ULTRA-RELIABLE AND LOW-LATENCY COMMUNICATION: DECODING PROCEDURE.

3.1.3.3 Next steps

Next steps will include the performance evaluation of the proposed CSA protocol with sliding window decoding. This evaluation will encompass both numerical and theoretical analyses. Firstly, we will study the effect of several system parameters through realistic Monte-Carlo simulations. The tuning phase will be focused on key protocol parameters, such as the ensembles C_H and C_L representing "high-priority" and "normal-priority" users, and the virtual frame size, denoted as W. The objective is to discern the optimal configuration that fits the specified reliability and latency requirements. Moreover, our contribution will extend to the development of an analytical tool that will enable us to directly identify the most favourable trade-off between the system parameters and the performance requirements.

3.2 MULTIPLEXING SERVICES WITH DIVERSE REQUIREMENTS IN IN-X SUBNETWORKS

Subnetworks are expected to support services having different requirements, such as high data rates with sporadic traffic with low latency requirements. For example, in the visual inspection cell industrial use case introduced in D2.1 [1], video feeds from control cameras need to coexist with control traffic with potentially tight inter-arrival times. Also, subnetworks may need to support control loops having different cycle times.

Multiplexing such traffic types with diverse requirements may pose significant challenges to the resource scheduling in a subnetwork. A significant effort in this direction has been spent in the 3GPP for 5G NR standardization, where features such as pre-emptive scheduling [34], enhanced HARQ retransmission mechanisms [35] and dynamic TDD [36] have been introduced. In particular, pre-emptive scheduling combines eMBB and ultra-reliable low latency communication traffic by opportunistically blanking some of the eMBB resources when a critical packet arrives. Dynamic TDD, where the link direction can change at each slot based on the traffic demand, was also motivated by the need of supporting heterogeneous traffic. The usage of dynamic TDD in macro-cell deployment can be challenged by the high cross-link interference between neighbour high power base stations,

i.e. DL-to-UL interference, while it can be considered feasible in small cells deployment where a similar power level is used in UL and DL [37].

Dynamic TDD offers the possibility of rapidly coping with the variable traffic demand but is still constrained by the necessity of accommodating the same UL/DL switching point for all served devices. This can represent a significant limitation for those networks where high data rate traffic is to be multiplexed with services requiring tight communication cycles.

3.2.1 Full/flexible duplexing for enhanced scheduling flexibility

In-band full duplexing, where a same node transmits and receive simultaneously over the same band, and flexible duplexing, where transmission and reception happen simultaneously but over different subbands, have also been advocated as solutions for improving the support of heterogeneous services, by overcoming the limitation of requiring the same UL/DL switching point for all users [38][39].

Historically considered unfeasible given the overwhelming self-interference from transmitter to receiver ports, full duplex and flexible duplexing solutions have been benefitting from the advances in analog and digital self-interference cancellation techniques in the last decade, and they can be considered viable at an affordable cost, at least for low power nodes. In flexible duplexing transceivers, the implementation of self-interference cancellation is eased by the insertion of a guard band to reduce the energy leakage between the transmit and receive port [40].

Flexible duplexing has also been discussed in the 3GPP Rel-18 NR study item on duplexing evolution, as part of 5G-Advanced standardization, for the sake of assessing the performance and feasibility of such solution for different bands and network deployments [41].

We believe that the usage of flexible duplexing is a promising approach for efficiently multiplexing different traffic types in a subnetwork; in particular, for services that can benefit from different UL/DL switching points, such as periodic traffic with tight periodicity multiplexed with high data rate traffic. We foresee that, in a subnetwork, only the HC may be flexible-duplexing capable, while the LC and the SNEs are only half-duplex capable, being low-cost devices.

Figure 37 highlights the difference between traditional half-duplexing mode, with fixed UL/DL switching point for all subbands, and flexible duplexing, where the UL/DL switching is subband specific. When half-duplexing mode is used, the UL/DL switching point for all devices should be adapted to the device requesting a tighter UL/DL temporization, possibly affecting their throughput and latency.

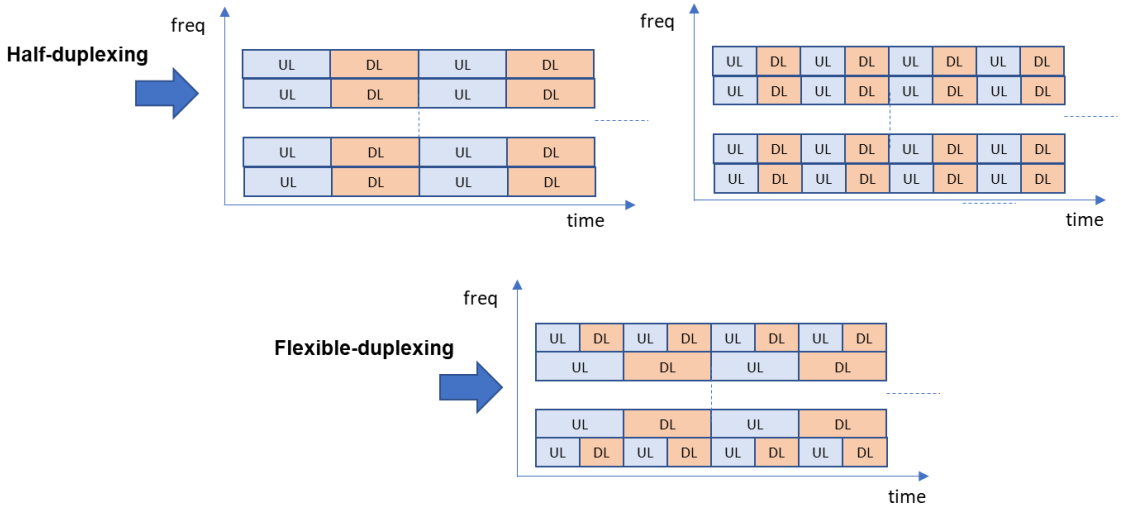


FIGURE 37. HALF-DUPLEXING VS. FLEXIBLE DUPLEXING.

3.2.2 Challenges of flexible duplexing

Flexible duplexing offers the possibility of overcoming such restriction, introducing though additional sources of interference with respect to co-channel interference among neighbours' subnetworks. Such interference sources are the following:

- Self-interference at the HC, generated by the leakage of DL transmissions on the neighbour subbands where the HC is receiving transmissions from devices. This can be handled via analog and digital cancellation techniques.
- UL-to-DL interference, generated by devices transmitting in adjacent subbands to devices receiving data from the HC.
- DL-to-UL interference, generated by the HC transmissions to the UL transmissions of devices in neighbour subnetworks transmitting over adjacent subbands.

In a traditional macro-network setup, DL-to-UL interference is the most critical component given the large power imbalance between base station and UEs [34]. In short-range subnetworks though, we can assume though a similar transmit power for SNEs, LCs and HCs, in the order of 10 dBm or below. Therefore, DL-to-UL interference is not a major concern in subnetworks, as UL performance is dominated by co-channel interference in dense deployments.

Given the limited area/volume of subnetworks, with devices potentially located at a short mutual distance, UL-to-DL interference can instead significantly affect the performance. The HC can take care of scheduling devices with significantly different ratios of UL and DL data demands over different subbands, and their position in the spectrum has to be selected such that the residual inter-subband interference does not affect the communication performance.

3.2.3 Next steps

In Deliverable D3.3, solutions for flexible duplexing schedulers aiming at enhancing the support of diverse traffic types in a subnetwork will be presented, and their performance evaluated via simulation studies. In our analysis, we will be adopting models for in-band emissions defined by the 3GPP [42].

3.3 PREDICTIVE SCHEDULING

Solutions presented in this subsection are addressing the possibility of leveraging predictive mechanisms for enhancing the transmission reliability in subnetworks. Predictive solutions can address beam management, mobility of potential signal obstructor, or characteristics of the underline service and application.

3.3.1 AI/ML-based predictions for beam management and coverage assessment in subnetworks

AI/ML based methods can be utilised for improved protocol engineering and enhancing functionalities across the entire subnetworks' architecture. A specific challenge arises when these AI/ML methods are applied for subnetworks, primarily because of the need to support low complexity. As extensively described in Section 2, subnetworks operating at high carrier frequencies are expected to use beamforming for improving spectral efficiency and reduce energy consumption. Several aspects around meeting the low complexity requirements for beam management and coverage assessment for subnetworks are discussed in the following subsections.

3.3.1.1 Background: 5G-NR used as a baseline

Devices being served by 5G RANs are required to perform measurements. The main goal of measurement functionalities are coverage and radio link assessment. The latter is more important for active devices (i.e., with an active service connection), to make sure the radio link is suitable for the active connection. This measurement functionality focuses on the active beam pair being used for communications. The former however, is more important to guarantee service continuity, as coverage assessment is fundamental to eliminate coverage holes, weak coverage areas, and forms the basis for mobility procedures.

In 5G (and, in general for cellular) deployments, the coverage area of a cell is formed by the coverage of a set of beams. In RRC_CONNECTED mode, a device measures one or more beams of a cell and the measurements results, which can relate to power, signal quality, noise or interference, are averaged to derive one or more cell quantities. Devices are therefore configured to consider a subset of the possibly detected beams. Synchronization Signal Block (SSB) type of beams provide wider coverage and synchronisation resources, whereas Channel State Information Reference Signal (CSI-RS) type of beams provide reference for measurements, channel feedback, and beam management procedures. Filtering procedure takes place at two different levels. First, at the physical layer, a beam consolidation procedure takes place to derive beam qualities. Secondly, at RRC level (or Layer 3), physical layer measurements are averaged to derive either beam and/or cell quantities from multiple beams. Cell quantities from beam measurements are derived in the same way for serving cell(s) and for non-serving or neighbour cell(s).

While performing measurements is necessary, it comes at the cost of energy consumption (due to both measurement and measurement reporting) and service interruption during measurement gaps. The cell coverage area will typically be larger in a cellular deployment, when compared to the coverage area for subnetworks deployment. In 5G, there are mechanisms in place to avoid devices to perform measurements e.g., when the device is at the centre of a cell, where coverage is expected to be good and satisfactory for the service. In certain type of subnetworks, however, devices might be expected to assess coverage much more often. This can be the case of immersive education use cases, with dynamic interaction among pupils, or industrial use cases such as robot control and subnetwork coexistence in the factory all, where robots -or their parts- can be mobile. Due to the resource constrained nature of subnetwork devices, the mentioned problems may become even more evident.

3.3.1.2 Preliminary method design for beam management and coverage assessment

Subnetworks radio deployments may leverage on AI/ML methods that rely on supervised learning, then measurement reporting becomes a significant problem due to the signalling overhead resulting from the amount of data typically required for good model generalization capabilities. Moreover, supervised learning AI/ML models that can anticipate beam related quantities are fundamentally difficult to train due to the smaller coverage area they provide. This makes the window of opportunity to collect useful beam related data and with that, to train models, smaller when compared to cell level related AI/ML models.

One way to tackle this problem would be to include lightweight AI/ML functionalities in devices themselves. Devices could then execute AI/ML operations such as inference, instead of measuring and reporting measurements. A few examples of AI/ML operations that can form the basis for the method design, may include anticipating the below metrics/characteristics:

- Radio measurement quantity regression, where the models may output one or several predicted values
- Classification of beams, which can include a basic metric such as a *good beam* (in 5G specification, a *good beam* would be above a *absThreshSS-BlocksConsolidation* value), or a more complete classification system, e.g., threshold-based with more levels
- Clustering of a set of beams, as an example, in the above-mentioned classification categories
- Counting the number of expected good beams, e.g., a short time in the future
- Beam failure detection, e.g., detecting exactly when this would be expected to happen

Having subnetwork devices anticipating any of the above example metrics can aid enhancing beam management procedures at a network level. Any of these useful beam related predictions could be a valuable input for beam scheduling techniques (for example, cluster of beams *X* serves these set of subnetwork users *Y* because it can provide resources for the users service), or the input can serve as a handover trigger (for example, cluster of beams *X* does not provide coverage for subnetwork users *Y*, hence these users may need to execute handover towards another subnetwork or towards the umbrella network) by considering the assessment of results of coverage from neighbours (e.g., as a result of regular measurement configuration). Such methods facilitate resource allocation and would reduce measurement related signalling overhead. In some cases, however, the overhead is difficult to be reduced, as it may be important to continue to transmit measurements and/or predicted measurements, so the serving network can assess the performance of a model. In some other cases, the overhead of measurements, predictions and training information reporting is too great for a certain radio service link, e.g., due to battery constrains and/or air interface overload. In these cases, the action taken by the device can be associated with other system performance aspects.

To address the mentioned scenarios, devices may also be configured with cross-layer related criteria, including monitoring of L1, L2 and L3 events, where upon determination that the criteria is met, the UE will either perform inference, and/or will train an AI/ML model, reporting on both the predictions and/or the training process. As an example, a device may decide to report on predictions only, measurements only, or initiating a batch of training to a certain model and not report at all, based on MAC layer buffer status values. If these are high, a possibility is to report on measurements and predictions only. If, however the air interface experiences overload, then a device might report on the predictions only, in an effort to reduce signalling overhead.

In order to achieve the example methods described above, it is also important to consider aspects such as dataset size, due to the resource constrained nature of subnetwork devices, and the fact that keeping a dataset that is able to satisfy training requirements is fundamental for good AI/ML model performance and generalisation properties.

There are various factors that can affect the performance of a model and the accuracy of its predicted values. One of them is, if the model has been exposed with particular input data characteristics. In these cases, more accurate predictions are obtained with e.g., less data redundant points as input in training. This would, however, depend as well on the data features of the dataset.

One way to address the dataset size optimality problem, is to make the devices determine optimal dataset size (e.g., in terms of number of data points and data features) based on identification of sub-optimal performance of an AI/ML model. A device may test different settings for model performance, with a focus on the input data. For example, a model may be trained with an *X* number of data points as training inputs. Training of that model is then executed using the current dataset. After training, the model is then ready for testing. Inference is executed, where inference results can be any of the example outputs listed above. The inference results are then compared to ground truth results that will be available a short time after inference. The dataset used for training can then be assessed by a device. Information related to the input data that best describes the conditions the device experienced when it produced the inference results can be extracted.

To summarise, the preliminary design of the low-complexity method for beam management and coverage assessment for subnetworks would comprise of the following stages:

- Stage 1: Pre-emptive inference for beam-related prediction, particular for low coverage assessment. For example, on detection of low coverage from beam (or cluster) *X*, infer next best beam (or cluster) *Y* based on AI/ML model and filtering of L1 data which has high degree of variability and unpredictability.
- Stage 2: AI/ML model activation strategy in order to save computational resources and increase the chance of good model generalisation. For example, using a carefully tailored mix of radio channel conditions and training data characteristics.
- Stage 3: Dataset design for subnetworks, in particular considering the dataset size and feature engineering including optimal selection of features based on the available dataset.

In the end, it is important to highlight that the output of the discussed method may be used as a relevant input for predictive scheduling mechanisms.

3.3.1.3 Next steps

The preliminary design of the method will be finalised and evaluated and used as a prescription for building predictive scheduling schemes.

3.3.2 Mobility-aware Predictive Scheduling in Industrial Subnetworks

In this section, we introduce a novel approach for improving latency and reliability in industrial subnetworks by leveraging context information related to position and movement of devices of interest.

For the sake of highlighting the novelty of the presented approach, we first recall the basic link adaptation procedures in modern cellular systems.

In the current 3GPP 5G NR system, the UE is configured by the network to perform measurements on the radio channel condition and reports it to the network. The radio channel condition is reported in the form of Channel Quality Indicator (CQI) which represents the estimated MCS required for a hypothetical PDSCH to reach a target PER. The CQI is reported in the Channel State Information (CSI) indicator, which may also additionally contain the

Precoding Matrix Indicator (PMI) and Rank Indicator (RI) for MIMO operations. The CSI can be reported in an aperiodic (one off) or periodic manner and the gNB uses the CSI reports and, the reported HARQ feedbacks to perform link adaptation for the UE, to determine a suitable MCS for a downlink packet transmission on a downlink physical channel, such as PDSCH.

The CQI received by the gNB is based on the past information, i.e., the gNB relies on past radio conditions in scheduling downlink channels to the UE. This mechanism is unable to react to sudden changes to the radio condition. For example, if the gNB receives CQI indicating a good radio condition for the past 100 ms, the gNB would schedule a PDSCH with high order MCS to the UE. If there is an object that blocks the LOS between that UE and the gNB, then the PDSCH transmissions to the UE may likely fail. This may be acceptable in legacy cases such as in the current 5G NR system, where latency and reliability are not very stringent, and the UE can rely on retransmission to recover the transmissions. However, for 6G-SHINE with extreme latency and reliability, this may not be acceptable. Hence, a new method is required to support the extreme requirement faced in use cases considered in 6G-SHINE.

3.3.2.1 General idea

Ideally, to mitigate against sudden changes to the radio condition in a wireless network, the network scheduler in the gNB or AP would have the capability to anticipate such changes, i.e., the network would have a predictive scheduler. In a cellular network deployed in an urban area, it is very challenging for the gNB scheduler to know when there will be a sudden change in radio condition from CSI reports due to the presence of many moving objects in that environment. As described previously, sporadic performance degradation and packet loss due to sudden changes in the radio environment is acceptable for a legacy cellular network, such as 5G, as the latency and reliability requirements are not very stringent, and hence there is not much motivation to pursue a predictive scheduler, especially in an urban environment with many moving objects.

Sporadic packet losses may not be acceptable in 6G-SHINE use cases, where a reliability with PER as low as 10^{-7}, and extreme low latency are targeted. Fortunately, unlike in a cellular network, the characteristics of some subnetworks are expected to be highly predictable. For example, a robotic arm forming a subnetwork as shown in Figure 38, has a highly predictable movement where the arm moves from Position 1 to Position 2. An SNE can be located at the tip of the robotic arm and a HC can be located at the base. When the SNE moves from Position 1 to Position 2, it will cause a sudden change to the radio environment between the SNE and the HC; however, this change is highly predictable and such characteristics of the subnetwork can be used to enable predictive scheduling. In the example in Figure 38, the SNE can change its served beam from Beam 1 to Beam 2 when it anticipates that its position will change from Position 1 to Position 2, and/or, the HC can schedule a more robust (i.e. lower order) MCS when it anticipates that the SNE is going to move from Position 1 which is closer to the AP to Position 2 which is further from the HC.

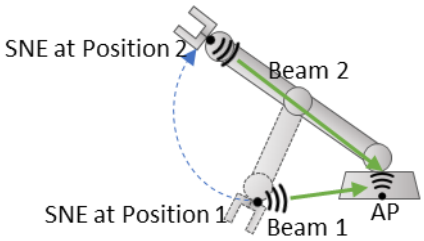


FIGURE 38. SUBNETWORK CONSISTING OF A ROBOTIC ARM

3.3.2.2 Implementation signalling

To implement a predictive scheduling solution, the scheduler in the HC needs to know which movements are of interest and when these movements occur. The movements can be predetermined prior to network operation or during pauses in operation, e.g., maintenance, and they are also determined during network operation. The movements that are of interest are movements that causes significant changes to the radio condition, for example due to shadowing leading to non-LOS conditions or for the example in Figure 38, movements that cause a beam direction to be obsolete. It should be appreciated that changes in radio condition due to fast fading is not considered in these movements of interest as it is not feasible to predict ahead of time the impact of fast fading.

The movements can be predetermined; for example, in the simplified robotic arm subnetwork in Figure 38, the movements can be characterised as Position 1 and Position 2. Measurements can be performed in these two positions to determine the optimum transmission parameters when the SNE is in these positions.

The movements can also be determined during operation of the subnetwork. For example, the HC may experience a poor signal from an SNE and the position of the subnetwork or the SNE can be noted. The HC may then configure the SNE to perform measurements when the SNE is in that position again and determine optimum transmission parameters for that position.

Once these movements and the associated positions are known, the times when they occur need to be signalled to the scheduler in the HC device.

Signalling should be used by the SNE to inform the HC or other SNEs in the subnetwork, that a movement that causes significant changes to the radio condition is going to occur. For example, the SNE with indoor positioning capability, can inform the other elements in the subnetwork such as the HC or other SNEs, that it is T_{move} ms towards a position of interest. It should be mentioned that positioning is only one way of triggering the SNE to send T_{move} to the HC, and other trigger can be a timer. For example, a robot with a known path will start a timer when it starts its process and will send the T_{move} countdown when its timer hits a pre-determined time. The position or timer information may be determined at the higher layers, or system they may be implemented at lower layers of the SNE and pass down to the lower layers such as Layer 2 or Layer 1 of the SNE, to trigger the SNE to send the T_{move} countdown to the HC. It is beneficial that the SNE informs the other elements early, e.g., by a time T_{move}, that a movement is to occur to enable the scheduler to have sufficient time to mitigate any changes due to the radio condition. For example, Figure 39 shows a subnetwork with an HC acting as an AP, and three SNE's, SNE#1, SNE#2 and SNE#3. SNE#2 and SNE#3 are in fixed locations whilst SNE#1 is a robot that moves around the subnetwork with a known path. The positions of interest to the scheduler are Position 2 and Position 3 of SNE#1, where in these positions, SNE#1 blocks the LOS between SNE#2 and the HC, and between SNE#3 and the

HC. SNE#1 can indicate to the HC or other SNEs in the subnetwork when it is T_{move} ms away from Position 2. The HC may also broadcast this information to the SNEs or other HCs in the subnetwork. If the HC has downlink transmission to SNE#2 when SNE#1 is in Position 2, the HC can redirect the transmission via a RIS to mitigate against SNE#1 blocking the LOS between SNE#2 and the HC. Similarly, when SNE#1 is T_{move} ms away from Position 3, it indicates this to the HC and the HC can broadcast this information to other SNEs in the subnetwork, so that other SNE may take this into account in their uplink transmissions. For example, SNE#3 may transmit an uplink packet to the HC when SNE#1 is in Position 3 and being aware that SNE#1 blocks its LOS with the HC, SNE#3 can increase its transmission power or use a more robust (i.e., lower) MCS for its uplink transmission. In-X subnetworks SNEs may be enhanced to have more autonomy on their transmission parameters compared to the current 5G UEs, i.e., a SNE can be preconfigured with multiple transmission parameters, and it can use one of these transmission parameters based on its own triggers; for example, the trigger can be information received regarding an impending loss in LOS conditions.

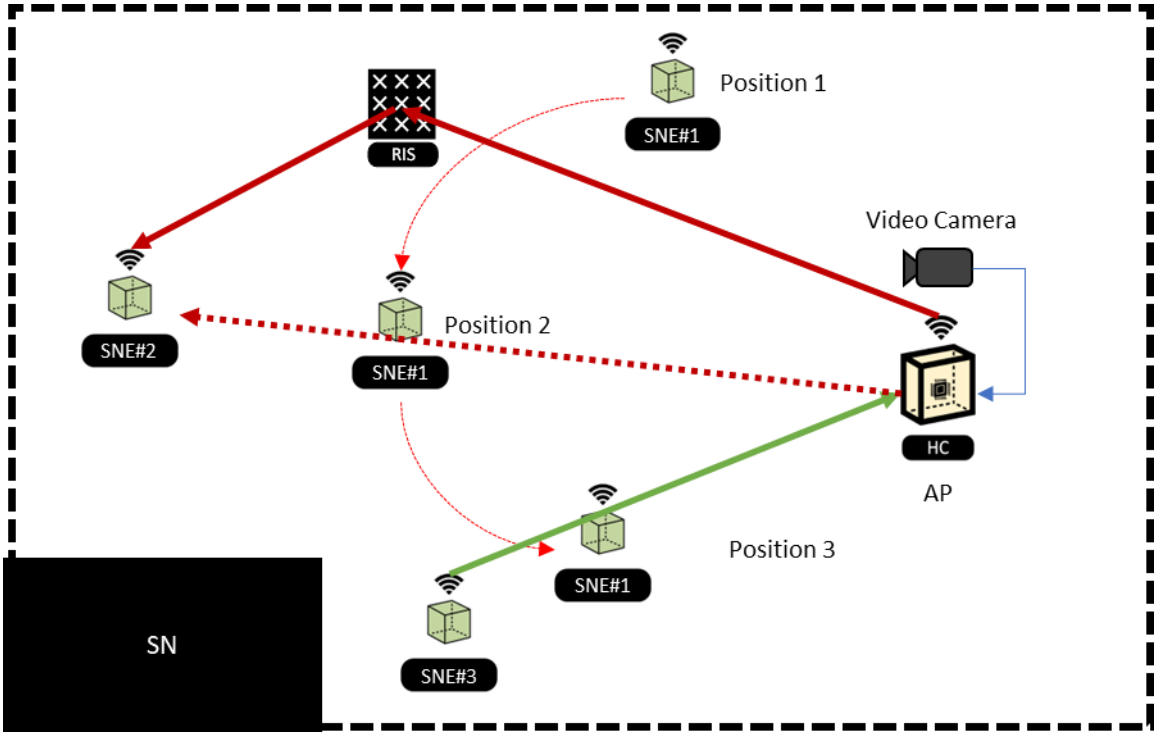


FIGURE 39. INDICATING TIME OF ARRIVAL AT POSITIONS OF INTEREST FOR PREDICTIVE SCHEDULING

Another signalling method is for the application layer to inform the HC or SNE that a movement is going to occur, i.e., the application layer can provide control information to the lower layer, such as the physical layer of the HC or SNE. For example, in the case of the robotic arm in Figure 38, when the application layer determines that it needs to move from Position 1 to Position 2, it indicates to the SNE or HC that a position change from Position 1 to Position 2 will occur in the next T_{move} ms. The signalling from the application layer can be separated from the SNE. For example in the subnetwork in Figure 39, a video camera can be used to monitor the movements of SNE#1 and when SNE#1 is approaching Position 2, the video camera can inform the scheduler in the HC when SNE#1 is T_{move} ms away from Position 2, and if the HC has a downlink packet for SNE#2 when SNE#1 is in Position 2, it can

redirect the transmission via a RIS to SNE#2 since the LOS is blocked by SNE#1. Similarly, when SNE#1 is T_{move} ms away from Position 3, the video camera informs the scheduler. Here the scheduler may broadcast this information to the SNEs within the subnetwork, which is beneficial for the SNEs transmitting in the uplink. For example, when SNE#1 is T_{move} ms away from Position 3 the HC informs SNE#3 (which still has LOS with the HC) and when SNE#1 is in Position 3 and SNE#3 has an uplink transmission to the HC, it can increase its transmission power or use a more robust MCS to overcome the non-LOS caused by SNE#1.

3.3.2.3 Next step

Further investigation will be done on methods to determine the described position of interests and the signalling required to update these positions. This is the mechanism used to trigger a SNE to send the T_{move} countdown to the HC, or the HC to takes mitigative steps to avoid a predicted bad radio condition. The SNE and HC may require effective signalling to inform each other of the predicted bad radio condition and the preventive scheduling taken to mitigate against it. Such a method will also require giving the UE some autonomy in scheduling its own resources, which is not available in current 5G system. Hence, the level of autonomy given to the SNE will be further considered.

Enhancements to this method will be explored, such as the determination of beams at position of interests that will improve reliability and reduce latency for packet transmissions.

Finally, some quantitative evaluation will be performed to determine the benefits of this method.

3.3.3 Deterministic and Predictive Traffic Scheduling

Supporting traffic flows with diverse requirements in subnetworks, including the simultaneous support of deterministic time-sensitive traffic and best-effort traffic, requires highly intelligent and predictive scheduling solutions that efficiently and effectively handle their proper multiplexing. 6G-SHINE seeks exploiting predictability of subnetworks in multiple dimensions, including channel conditions, traffic patterns, communication context, etc., to anticipate resource requirements and program scheduling ahead. The designs must consider the stochastic nature of the wireless links that introduce varying link capacities, potential interference and higher transmissions errors compared to their wired counterparts [43]. The proposed deterministic and predictive MAC designs rely on a common framework that can be utilized for the support of mixed traffic flows in subnetworks in general. We will demonstrate the effectiveness of the proposed framework in the in-vehicle networks (IVN) leveraging the traffic characterization presented in deliverable D2.2 [2] to predict the demand of resources and schedule IVN subnetwork traffic flows. Such traffic flows exhibit certain correlation patterns that can be exploited to anticipate the traffic demand and predictively schedule wireless transmissions. Our objective is to exploit this predictability to efficiently support deterministic services with high-priority time-sensitive traffic as well as bounded worst-case performance under saturated conditions without over dimensioning or underutilizing resources or starving lower priority traffic.

3.3.3.1 Scheduling in 5G networks

Supporting mixed traffic flows with diverse requirements, including deterministic time-sensitive traffic and best-effort traffic, goes well beyond the current state of the art [44]. The challenge augments when certain traffic flows simultaneously demand high data rate and reliability levels as well as extremely low latency as we see with

increasing levels of automation and service reliance in critical verticals such as autonomous driving (AD). As mentioned in Section 3.1, 3GPP Rel. 15 5G NR introduced semi-static scheduling scheme (i.e., configured grant - CG in the UL and semi-persistent in DL) to enable shorter UL and DL transmission delays. These scheduling schemes perform periodic reservations of radio resources. However, if the traffic's inter-packet period and the resource allocation periodicity do not coincide, the service delay requirement might be compromised. To address some of these challenges, 3GPP Rel. 16 in TS 38.331 allows configuring and activating multiple CG for uplink traffic with different resource allocation periods. This might lead to overprovisioning of resources to satisfy the most demanding QoS requirements, or it may result in high signalling overhead to dynamically activate the most suitable CG. The importance of integrating time-sensitive capabilities in 5G and beyond wireless networks is acknowledged with continuous specifications, as observed since Rel. 16 of 3GPP, that aim at delivering end-to-end deterministic latencies utilizing Time Sensitive Network (TSN) and Deterministic Network (DetNet) technologies that have been designed for wired networks. The fundamental differences between wired and wireless networks challenge the integration of time-sensitive capabilities in (5G and beyond) wireless networks.

Exploiting knowledge about traffic patterns can contribute towards a more efficient scheduling of deterministic time-sensitive communication flows [45] [46]. Predictive and proactive resource allocation has shown to be a promising solution to effectively support massive machine-type communication (mMTC) applications, for example, by identifying correlated traffic patterns between related devices of a group or using deep reinforcement learning to predict users' trajectories [47]. However, current scheduling solutions cannot support the requirements of mixed traffic flows where high-priority time-sensitive traffic is multiplexed with low-priority traffic, as well as high-data rate and low-latency services foreseen in 6G. Time-sensitive capable schedulers pre-empt or down-prioritize the transmission of low-priority traffic in favour of time-sensitive traffic. Depending on the volume or frequency of time-sensitive traffic, a low-priority traffic might be pre-empted several times which might compromise the on-time scheduling of low priority traffic [44]. The effective support of diverse traffic flows with stringent requirements becomes even more challenging for reduced capabilities (REDCAP) devices, like we could encounter in subnetworks. A light version of 5G NR using REDCAP (a.k.a. 5G NR REDCAP or NR Light) is being designed since Rel-17 in TR 38.875, but it is limited to address the communication needs of Industrial IoT (IIoT) deployments and mMTC communications that are characterized by low data rate requirements and best-effort traffic. To this aim, 5G NR REDCAP utilizes reduced number of Tx/Rx antennas and bandwidth, relaxed processing times, MIMO layers and modulation orders, limits to half-duplex FDD operation (full-duplex FDD is optional for legacy 5G NR UE) and extends the discontinuous reception (DRX) operation.

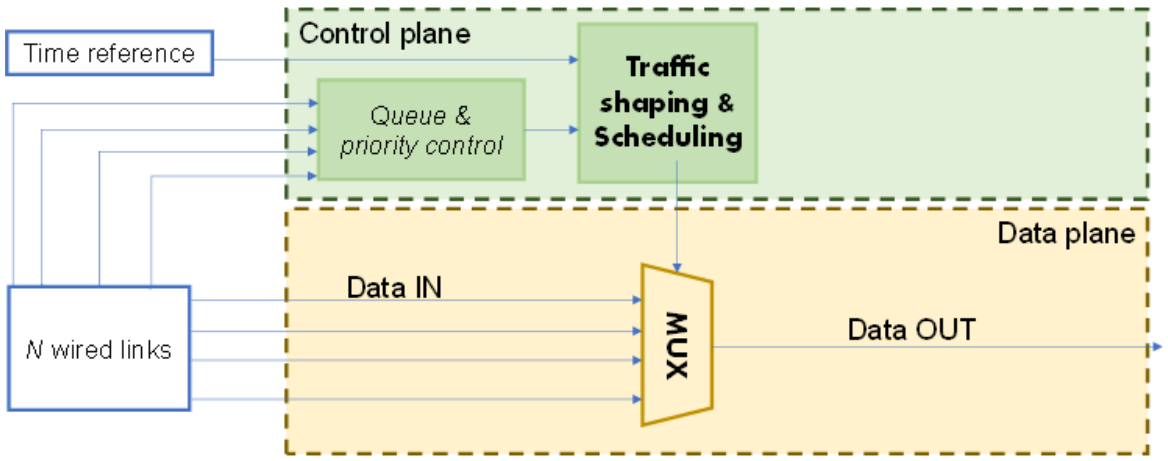
3.3.3.2 Time-sensitive networking in hybrid wireless and wired automotive in-vehicle networks

In-vehicle networks (IVN) and Electrical/Electronic (E/E) architectures are evolving towards zonal-based deployments supported by a high-speed automotive Ethernet backbone that is used to connect zonal Electronic Control Units (ECUs) together, as well as to the vehicle's central computing unit or High-Performance Computing Unit (HPCU) [1][2]. This evolution is motivated by the increasing critical bounded-timing, reliability, and safety requirements of automotive functions and services, especially with the advent of AD. Autonomous vehicles are complex real-time systems that require deterministic communication networks to guarantee that messages are exchanged between the in-vehicle elements (e.g., sensors, actuators, control units, etc.) within a period or bounded delay. The in-vehicle deterministic traffic mixes with other traffic with varying requirements, such as data rate, latency, and reliability. This challenges the design of an IVN's wireless subnetwork – and in particular its MAC-capable to support all services efficiently without high priority traffic compromising the rest of traffic. 6G-SHINE targets the support of mixed traffic flows in in-vehicle subnetworks with data-driven and AI/ML-powered predictive scheduling mechanisms. IVN schedulers should be implemented at subnetwork elements with higher

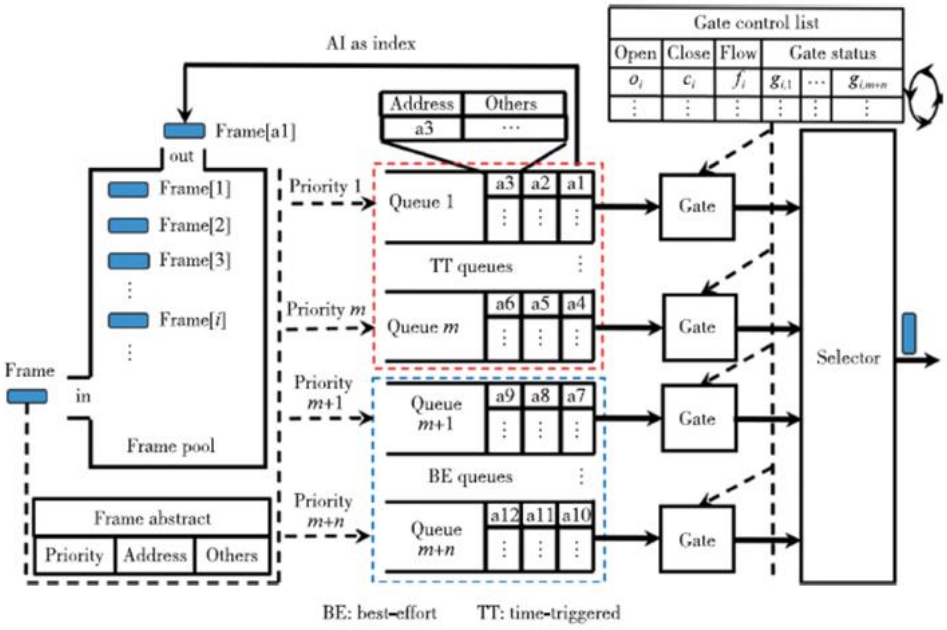
capabilities, i.e., zone ECU and HPCU. Zone ECUs and HPCUs will natively exploit AI/ML for their service operation (e.g., data fusion and object identification) and ensure the necessary processing power.

The definition and specification of time-sensitive networking in automotive in-vehicle ethernet communications is being addressed by the IEEE 802.1DG working group. The goal of this specification is to support "*the entire range of in-vehicle applications including those requiring security, high availability and reliability, maintainability, and bounded latency*". The on-time and deterministic management of the heterogeneous in-vehicle traffic represents one of the main challenges towards this end. This is why the vast majority of the IEEE 802.1DG specification focuses on defining the processing order of the received and transmitted messages at the different IVN elements. Figure 40.a) shows a high-level block diagram for the deployment of TSN in automotive ethernet-based IVN. In Figure 40.a), we can observe control plane and data plane operations executed by intermediate bridge elements on (zone) ECUs. Additionally, it depicts data received through the ingress ports (data IN) and data transmitted through egress ports (data OUT) over wired links connected to end stations, such as sensors and actuators. Scheduling and traffic shaping represent key building blocks in Figure 40.a) for the management of TSN networks. The main purpose of traffic shaping is to control the rate at which traffic is sent or received to ensure that it conforms to a desired traffic profile. It typically involves the use of buffers to store and regulate the flow of traffic. Scheduling, on the other hand, manages the access to shared resources and is focused on determining the order in which packets are transmitted or processed. It involves prioritizing packets based on various criteria (priority, deadline, etc.) to meet specific quality of service (QoS) requirements. Both concepts play crucial roles in optimizing network performance and ensuring QoS. It should be noted that since traffic shaping is a bandwidth management technique, if the resource assigned by a scheduler is the bandwidth, then a scheduler will implicitly perform traffic shaping.

In the automotive domain, the Asynchronous Traffic Shaper (ATS), Credit Based Shaper (CBS), and Time Aware Shaper (TAS) are common approaches to controlling and regulating the IVN data flows. The TAS (see its block diagram in Figure 40.b)) is specifically designed to support time-sensitive applications. To this aim, the TAS defines intervals in which the gates of queues open/close, and it cyclically repeats the open/close sequence providing timely synchronized transmissions [48].



a) High-level building blocks



b) Time-aware shaper operation [49]

FIGURE 40. REFERENCE DEPLOYMENT OF TSN IN AUTOMOTIVE ETHERNET IN-VEHICLE NETWORKS

6G-SHINE aims to replace wired IVN connections with wireless links while supporting heterogeneous automotive traffic, including deterministic and time-sensitive applications. In Figure 41, the resulting building block for the 6G-SHINE deployment of hybrid wireless and wired connections in IVN is depicted. Figure 41 highlights in red the three main blocks added compared to the reference TSN deployment in automotive ethernet IVN shown in Figure 40. We foresee a gradual evolution for reaching this hybrid wireless and wired deployment. First efforts are to be focused on the design of a wireless scheduler that supports the traffic generated/addressed to/from sensors/actuators with stringent and deterministic requirements. This scheduler is represented with a red trapezium in Figure 41. Then, the hybrid wireless and wired scenario involves replacing a subset (N-X) of the wired links with wireless connections, seamlessly integrating them into the data and control planes for effective flow management. Additionally, Figure 41 introduces a new block labelled 'Predictive,' which plays a crucial role in managing and scheduling incoming (and possible outgoing) wireless transmissions. The 'Predictive' module is closely integrated with the 'Traffic Shaping & Scheduling' module to ensure that the integration of wireless links does not interfere with operations supporting data flows, including those for deterministic and time-sensitive applications. This implies, for example, that the scheduled wireless transmissions must arrive at the TAS's queues (see Figure 40.b)) before the gates open. To enhance predictive decisions, 6G-SHINE envisions feeding the 'Predictive' module with context control information derived from processed traffic (e.g., closed feedback on channel state, BLER statistics, traffic-flow characteristics, and their correlation) and environmental factors (e.g., vehicle speed, road type). Additionally, AI/ML models could contribute to the predictive capabilities. In the context of 6G-SHINE, the "Predictive" module is envisioned to play a crucial role in the joint management and coordination of both wired traffic shaping and wireless link scheduling mechanisms. For instance, the "Predictive" module could coordinate the opening timing of gates for wired traffic (e.g., advancing when the queue opens if the traffic is predicted to be waiting on the queue), while simultaneously relaxing the time requirements for the wireless link.

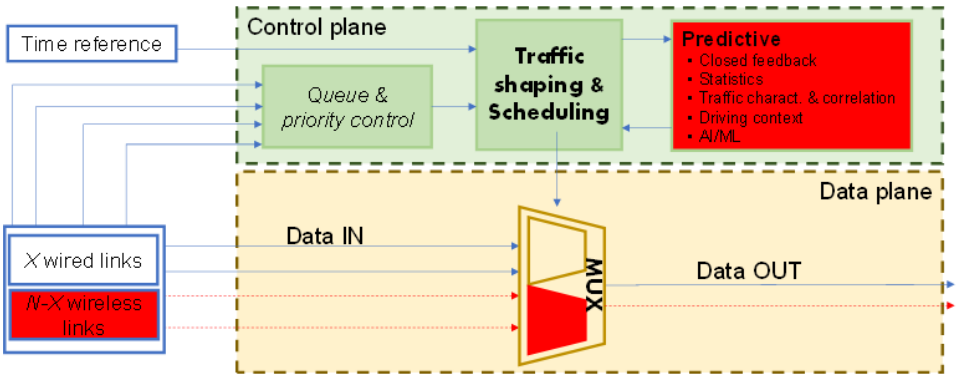


FIGURE 41. PROPOSED FRAMEWORK FOR PREDICTIVE-BASED AND TSN-CAPABLE HYBRID WIRELESS AND WIRED IN-VEHICLE NETWORKS

3.3.3.3 Proposed framework

Sustaining determinism over wireless links has been identified as a challenge due to [50]: 1) wireless channel variations, 2) channel access latency, and 3) interference and coexistence. These three challenges impact most wireless systems, but their impact strongly depends on the operating conditions and the sensitiveness of service to QoS/QoE variations. The analysis of these drawbacks for wireless TSN support in the context of the 6G-SHINE's in-vehicle subnetworks is as follows:

1) Wireless channel variations are more likely to appear in mobile and dynamic environments. In the in-vehicle subnetworks, wireless links are established between SNE, e.g., sensors and actuators, and higher capabilities elements, e.g., zone ECU and HPCU. These elements are static and their wireless channel conditions can be well studied and characterized beforehand. In addition, 6G-SHINE is performing measurement campaigns to derive in-vehicle channel models that characterize the wireless channel behaviour of in-vehicle subnetworks.
2) The wireless medium shared by multiple devices involves some form of channel access procedure in most wireless systems. Grant-based, random access and listen-before-talk channel access procedures are well known sources of (uncontrolled) delay. 6G-SHINE seeks exploiting the characterization and predictability of the in-vehicle traffic and the proactive reservation of resources to reduce the channel access delay. TDD and FDD duplexing schemes can also impact the time to access wireless shared resources. TDD is often utilized in time-sensitive systems due to the possibility of programming transmissions. In TDD systems, the guaranteed bounded latency is very dependent on the utilized frame format. Most commonly deployed TDD frames allocate more frequently slots for downlink transmissions. This results in increasing delays to access uplink resources [51]. On the other hand, both uplink and downlink resources are always available in FDD systems. This comes at the expense of having rigid dedicated spectrum for uplink and downlink, which might be inefficient. 6G-SHINE can investigate the utilization of discontinuous allocation schemes [52] that seek to take advantages of both duplexing schemes.
3) Interference and coexistence need be addressed in order to enable wired equivalent TSN performance over wireless links. The in-vehicle environment can be thought of as well-planned and characterized environment. The co-existence between in-vehicle subnetworks can be controlled due to the static nature

of the in-vehicle subnetwork elements. A more challenging scenario is controlling potential interference between subnetworks embedded in different neighbouring vehicles. 6G-SHINE is also investigating radio resource management schemes for managing interference from known and potentially unknown systems (e.g., jammers).

In line with the vision of the SNS JU's funded DETERMINISTIC6G project shared in [53], 6G-SHINE foresees the potential of predictability on achieving deterministic communications in 6G networks. Predictability could be exploited for different and complementary usages, while 6G-SHINE pursues initially exploiting data-driven and AI/ML-based approaches for accurate characterization and forecasting of in-vehicle traffic demand and patterns that allow optimizing the scheduling of resources and meeting time-sensitive requirements in mixed traffic conditions over the wireless links.

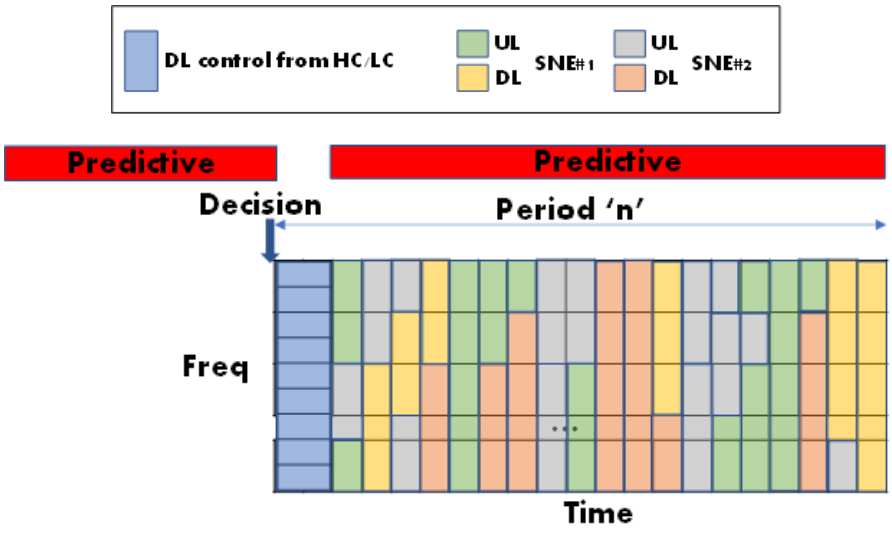


FIGURE 42. PREDICTIVE SCHEDULING OPERATION FRAMEWORK

In the following, the main features of the framework for the operation of the proposed predictive scheduling scheme illustrated in Figure 42 are summarized. Although the framework could be applied to any subnetwork in general, examples of application are provided for the in-vehicle subnetwork:

1) Centralized coordination and scheduling

Meeting stringent Quality of Service (QoS) requirements, such as ensuring bounded latency, poses a significant challenge when relying on distributed random-access mechanisms [50]. For the specific context of IVN, the proposed scheduling scheme aligns with the trend toward centralized and zonal E/E in-vehicle architectures, leveraging the higher capabilities available at zone ECUs and HPCUs. This approach facilitates central coordination of wireless medium access, enabling efficient execution of predictive scheduling. This centralized approach also favours the deployment of IVN subnetwork that are used to serve multiple low-cost and lower-capabilities SNEs, e.g., sensors and actuators. Implementing the predictive scheduling at the zone ECUs and HPCU helps keeping low the complexity of SNE.

2) Matrix-map allocation of resources with flexible period duration

Operating in a centralized scheduling mode requires efficient ways to distribute the scheduling information to the SNEs in the subnetwork. Current 5G systems rely on persistent scheduling to assign deterministic access to resources to the end devices. Although this approach minimizes the signalling and control overhead, persistent scheduling might require overprovisioning of resources to satisfy mixed traffic requirements. In 6G-SHINE, we would explore that the central entity deployed in the zone ECU or HPCU (depending on the scenario) evaluates, adapts and allocates the resources regularly. The duration of the allocation period is to be set based on the achievable predictability horizon of the 'Predictive' module (see Figure 41). Without loss of generality, 6G-SHINE proposal assumes radio resources are divided in the time and frequency domains forming a matrix-like grid of resource blocks. Resource blocks are then the minimal 'matrix's element-like' allocation unit. As it is depicted in Figure 42, the central entity at the zone ECU or HPCU utilizes the RBs in the first time slot of the period '*n*' as control information to signal the allocated resources to the SNEs in the downlink. Resources are allocated to the SNEs for the following period, and they are signalled to the SNEs using a matrix-format which indicates whether the resource is allocated for uplink or downlink transmissions. Pre-configured matrix-format patterns could be used to minimize the amount of signalled data in the downlink. The decisions about the duration of the allocation period, the amount of the resources allocated to each SNE, and how they are distributed in the period's allocation matrix are based on the outputs of the 'Predictive' module. It hence potentially accounts for the requirements of the predictive traffic, the channel status, feedback statistics, etc., and considers trade-offs between flexibility/reactiveness to traffic changes and control overhead (i.e., frequency with which allocated resources are signalled).

3) Discontinuous allocation of resources

The proposed matrix-map allocation of resources leverages the possibility for discontinuous allocation that is being investigated for future 6G systems [52]. Discontinuous allocation of resources allows for a more flexible and efficient time-frequency scheduling and utilization of the resources, compared to conventional 5G-based OFDMA schemes that impose spectrum contiguity constraints in assigning the RBs. It also brings the possibility for additional freedom for channel-state aware RB allocation.

4) Dynamic configuration of BW parts

The flexibility in the radio resource management (i.e., 2) and 3)) could also be extended to dynamically configure the RB format or bandwidth part (BWP). As part of centralized coordination and scheduling (1)), additional decision-based predictions – for example, regarding traffic needs or requirements, radio channel conditions – could include the configuration of the duration in time and frequency of the RB (i.e., numerology) for the following period. This decision must consider the switch delay associated with changing the BW part configuration as detailed in 3GPP TS 38.133. In line with the flexible grid envisioned, we would also explore the case in which the available spectrum is configured with different BW parts to support traffic with different requirements in each BW part.

5) In-band flexible duplexing

Finally, we initially consider that the predictive scheduling mechanism operates over a single-band (see Figure 42). With the utilization of the same band for UL and DL transmissions, in-band flexible duplexing, introduced in Section 3.2, targets the utilization of the minimum bandwidth to help reduce the SNE's power consumption. Additionally, the coordinated flexible duplexing-predictive scheduling operation also targets maintaining the low complexity of SNEs by only allocating uplink or downlink resources in a time slot (i.e., half duplexing). On the other hand, flexible duplexing provides opportunities for the higher capabilities SNEs (LC or HC) implemented at the zone ECU or HPCU to transmit and receive simultaneously on non-overlapping frequency resources. The concurrent uplink and downlink transmissions at the LC or HC might create different types of interference, including self-interference

from power leakage from the downlink to the uplink transmissions and vice versa [39], which challenges the resource allocation process.

3.3.3.4 Problem formulation

Following the gradual approach for the deployment of hybrid wireless and wired technologies in IVN, we first consider the problem formulation for the design of a wireless scheduler that support mix traffic flows with diverse requirements, including deterministic time-sensitive traffic. We consider that the matrix-like resource pool P_i is divided into a number W_i of RBs in the period interval r_i. As it is illustrated in Figure 43, the size of each RB is f_i (Hz) $\times$ τ_i(s) which depends on the numerology μ_i employed in the period interval r_i. In this two-dimensional frequency-time resource pool P_i, $w_i(f,t)$ denotes the RB associated with the *fth* RB in the frequency domain during the time slot *t*. Note that f v(0, …, F_i-1) and t v(0, …, T_i-1), where F_i and T_i are the total number of RBs and slots in the frequency and time domains, respectively. Considering *BW* is the total single-band bandwidth for the operation of the subnetwork, F_i can be calculated as *BW*/f, and T_i is equal to r_i/τ_i.

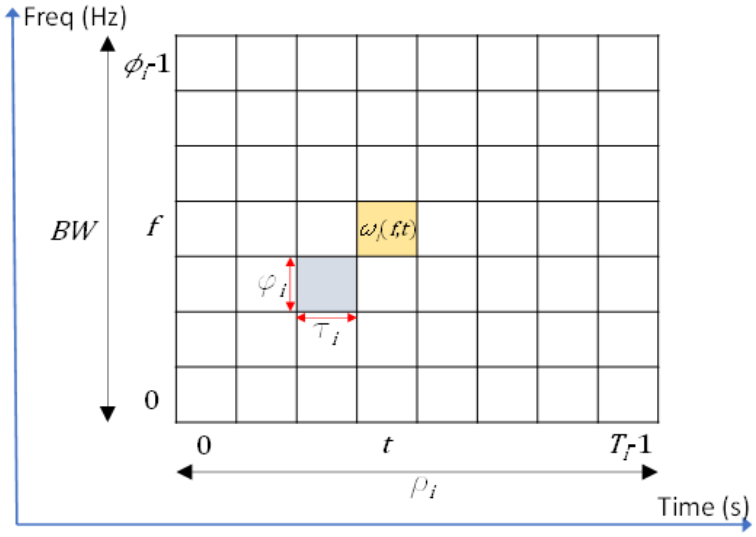


FIGURE 43. TIME-FREQUENCY GRID

We consider that a number *N* of SNE (i.e., sensors and actuators) are configured in the IVN subnetwork to communicate through a wireless connection to the LC or HC (i.e., zone ECU or HPCU). We define the variable $\eta_{j,i}^{UL/DL}$ as the number of RBs needed to accommodate the traffic generated (UL, $\eta_{j,i}^{UL}$) or addressed (DL, $\eta_{j,i}^{DL}$) towards the SNE *j*, *j* v(0, …, *N*-1), in the period interval r_i. In this period interval, a SNE *j* may generate (UL) or receive (DL) several packets $P_{j,i}^{UL/DL}$. The number of RBs needed for the transmission of each of the packets $p^{UL/DL}$, $p^{UL/DL}$ v [1, …], can be expressed as $\delta_{j,i}^{p,UL/DL}$. Note that $\delta_{j,i}^{p,UL/DL}$ accounts for possible retransmissions (k-repetitions mechanism is assumed, i.e., no ACK) needed to satisfy reliability requirements. Then, $\sum_p \delta_{j,i}^{p,UL/DL} = \eta_{j,i}^{UL/DL}$. It should be noted that $\delta_{j,i}^{p,UL/DL}$ depends on the coding rate and modulation order used to transmit the packet $p^{UL/DL}$. Let denote $a_{j,i}^{p,UL/DL}$ as the arrival time of the packet $p^{UL/DL}$, and $l_{j,i}^{p,UL/DL}$ be its latency requirement from/to the SNE *j* to the LC or HC (i.e. Zone ECU or HPCU) in the period interval r_i.

To mathematically formulate the scheduling assignment problem, we declare the variable $w_i(f,t)$ as binary as follows:

$$w_i(f,t) = \begin{cases} 1, & if\ the\ fth\ RB\ at\ time\ slot\ t\ is\ assigned, \\ 0, & otherwise \end{cases}$$

and the new binary variables *UL* and *DL* as follows:

$$UL_{j,i}^{(f,t)} = \begin{cases} 1, & if\ the\ RB\ w_i(f,t) \text{ is assigned for UL transmissions to the SNE } j, \\ 0, & otherwise \end{cases}$$

$$DL_{j,i}^{(f,t)} = \begin{cases} 1, & if\ the\ RB\ w_i(f,t) \text{ is assigned for DL transmissions to the SNE } j, \\ 0, & otherwise \end{cases}$$

Objective function: Several objectives' functions could defined for this scheduling assignment problem. As a preliminary problem-solving approach, we consider as initial target the minimization of the bandwidth (BW) utilized to accommodate the wireless traffic generated in the IVN zone. By setting this objective function, we seek maintaining the low-complexity and low-power of SNE in IVN subnetworks.

$$\mathbf{min}\ BW$$

Constraints:

C1) *Exclusive resource occupancy:* RBs cannot be assigned to more than one SNE simultaneously

$$w_i(f,t) \leq 1, \qquad \forall f, t$$

C2) *Half-duplex allocation for SNE:* RBs allocated to a SNE at a given time slot can only be used for UL or DL

$$\sum_f UL_{j,i}^{(f,t)} \cdot \sum_f DL_{j,i}^{(f,t)} = 0, \qquad \forall t, j$$

C3) *Scheduling limit in every slot:* the total number of RBs assigned to all SNEs (either for UL or DL) at a given time slot is limited by the overall number of RBs in the frequency domain

$$\sum_j \sum_f UL_{j,i}^{(f,t)} + \sum_j \sum_f DL_{j,i}^{(f,t)} \leq \Phi i, \qquad \forall t$$

C4) *RB allocation limit reached*: the traffic generated in the subnetwork can be supported with the number of available RB

$$\sum_j \eta_{j,i}^{UL} + \sum_j \eta_{j,i}^{DL} \leq \Omega_i$$

C5) *Traffic latency requirements:* sufficient resources are allocated to the transmission of the packets generated by and addressed towards the SNEs.

$$\sum_{p^{UL}} \sum_{t=0}^{t=l_{j,i}^{p,UL}} \sum_f UL_{j,i}^{(f,t)} \geq \delta_{j,i}^{p,UL}, \qquad \forall\, j$$

$$\sum_{p^{DL}} \sum_{t=a_{j,i}^{p,DL}}^{t=l_{j,i}^{p,DL}} \sum_{f} DL_{j,i}^{(f,t)} \geq \delta_{j,i}^{p,UL}, \qquad \forall\, j$$

Other constraints, for example imposed by the power leakage generated due to the full-duplexing operation at the LC or HC (i.e., zone ECU or HPCU) that might require a "guard band" or leaving some RB empties are yet to be defined.

3.3.3.5 Description of the scenario and initial results

Following the problem formulation described above, we perform an initial study to derive a first approach about the potential required bandwidth for the wireless transmission of the data generated by the vehicle's sensors to the zone ECUs. We consider the reference sensor deployment for L3 autonomous driving described in 6G-SHINE D2.2 [2] and depicted in Figure 44. Without loss of generality, sensors have been grouped into zones following proximity relations and functional safety principles. Functional safety principles aim to ensure that a vehicle continues to function properly despite any unexpected malfunctions in its components. In this context, the defined areas are as follows:

- Zone front left: cameras C2 and C1, and radars R2 and R1
- Zone front right: cameras C4 and C1, and radars R3 and R1
- Zone rear left: radar R4 and camera C5
- Zone rear right: radar R5 and camera C5
- Zone center: camera C3, Lidar L1, Inertial Measurement Unit (IMU) and Global navigation satellite system (GNSS)

For example, note that data from camera C1 and radar R1 are transmitted to the ECUs located in both the front left and front right zones. This ensures that the data generated by these sensors can be transferred to the HPCU even if one of the two front zone ECUs experiences an issue.

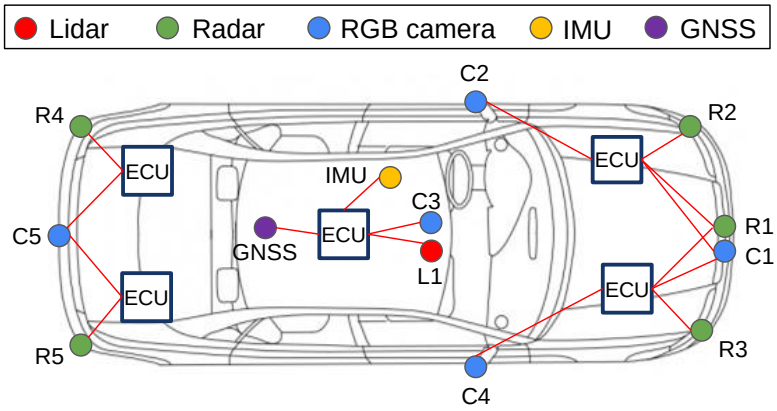


FIGURE 44. REFERENCE VEHICLE SENSORS DEPLOYMENT

We consider only uplink traffic from the sensors (or SNE) to the zone ECU (or LC) for these initial evaluations. An example of the amount of data generated by each of the sensors is summarized in Table 9. These values are derived

from the IVN traces reported in 6G-SHINE D2.2 [2]. The number of RBs required for the transmission of the sensor data has been estimated following the 3GPP transport block size determination procedure defined in 3GPP TS 38.214, considering different MCS configurations. The lowest (i.e. QPSK) and highest (i.e. 256QAM) order of modulation scheme supported in 5G NR are utilized. Table 9 also indicates the target code rate (x 1024) utilized together with each modulation scheme among the ones defined in 3GPP TS 38.214, i.e., QPSK-602, 16QAM-490, 64QAM-616, 256QAM-797. A sub-carrier spacing of 30 KHz is considered, and the number of MIMO layers (v) is set to 1 and 2. Table 9 also shows the delay budget for the transmission of the data from the sensor to the zone ECU. This delay budget has been set to match the rate at which the sensor generates new data. IMU, GNSS, radars and cameras operate at 20 Hz, and Lidar at 10 Hz, according to the reference sensor deployment identified in 6G-SHINE D2.2 [2].

TABLE 9. IN-VEHICLE SENSOR DATA AND RB TRANSMISSION REQUIREMENTS (ASSUMING 30 KHZ SUB-CARRIER SPACING)

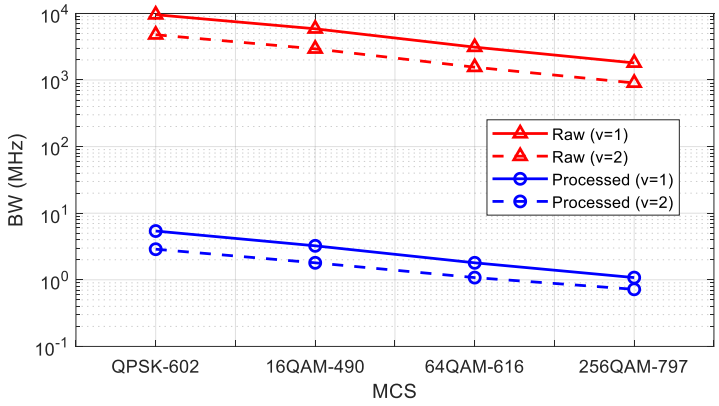
Data	Size (Bytes)	#RBs								Delay Budget (ms)
		QPSK-602		16QAM-490		64QAM-616		256QAM-797		
		v=1	*v*=2	*v*=1	*v*=2	*v*=1	*v*=2	*v*=1	*v*=2	
IMU raw	365	17	9	10	5	6	3	4	2	50
IMU proc.	357	16	8	10	5	6	3	3	2	50
GNSS raw	169	8	4	5	3	3	2	2	1	50
Lidar raw	5769	261	131	161	81	85	43	50	25	100
Lidar proc.	1725	77	39	48	24	26	13	15	8	100
Radar raw	16853	745	373	458	229	243	122	141	71	50
Radar proc.	3013	137	69	84	42	45	23	26	13	50
Camera raw	14.7MB	662993	331497	407267	203634	215975	107988	125195	62598	50
Camera proc.	4921	221	111	136	68	72	36	42	21	50

Figure 45 shows the resulting bandwidth needs for the transmission of the raw or processed data from the sensors to the zone ECU for each of the considered zones: front left and front right (Figure 45.a), rear left and rear right (Figure 45.b) and center (Figure 45.c). The obtained BW results consider the utilization of a fixed MCS (i.e. no MCS adaptation) for transmissions among the considered ones (QPSK-602, 16QAM-490, 64QAM-616, and 256QAM-797) during the period interval r_i. The period has been set to allow an integer number of transmissions from each sensor in the zone. For example, in the case of the front left zone where there are two cameras and two radars operating at 20Hz (delay budget = 50 ms), r is set to 50ms. During this period, each sensor of the left zone performs 1 transmission. For the center zone r is set to 100ms to allow an integer number of transmissions from the camera,

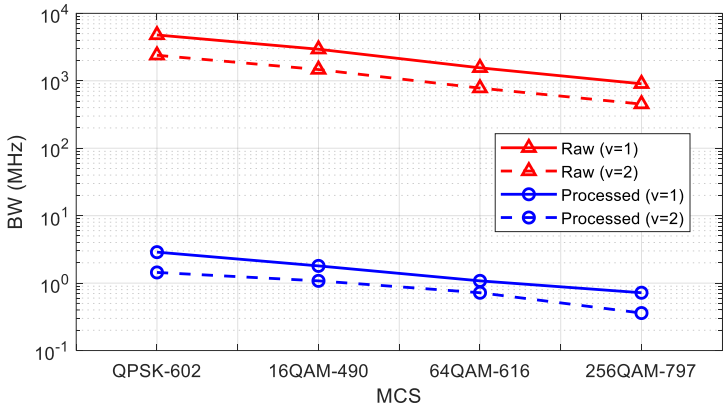
IMU, and GNSS operating at 20Hz and Lidar operating at 10Hz. During this period, each sensor of the center zone performs 2 transmissions but the lidar that performs 2 transmissions.

For the different zones considered, Figure 45 highlights the decrease in BW needs when higher order modulation schemes are employed for the transmission. For example, Figure 45.a shows that the required BW decreases from 9.5 GHz to 1.8 GHz when QPSK-602 and 256QAM-797 are utilized, respectively. Figure 45 also shows the decrease in BW needs with the number of MIMO layer. In this case, increasing *v* from 1 to 2 reduces roughly by half the BW requirements. Figure 45 also shows the significant impact on the BW needs of transmitting raw or processed sensor data. It should be noted that the higher share of the BW needs is required for the transmission of raw camera data (Table 9). The obtained results also evidence the high BW reduction that would be achieved if sensors process raw data and transmit the processed information (e.g. detected objects, region of interest, objects classification, etc.) to the zone ECU. For example, Figure 45.a shows that the required BW decreases from 9.5 GHz (*v=1*) to 5.4 MHz (*v=1*) and from 1.8 GHz (*v*=1) to 1.1 MHz (*v*=1) when QPSK-602 and 256QAM-797 are utilized.

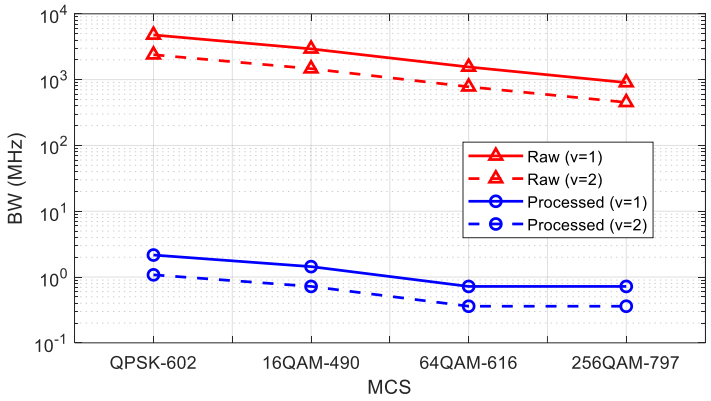
The obtained results shed light on the challenges of transmitting raw sensor data in in-vehicle subnetworks due to the high BW requirements. Advanced MCS and MIMO techniques can contribute to increasing the throughput and thus reducing BW requirements. However, this also increases the complexity of SNE which should ideally be kept low. Using high order MCS schemes also requires good and stable channel conditions. RIS elements strategically located within the vehicle could be exploited towards this aim. On the other hand, transmitting processed data has shown to significantly reduce BW requirements. The two approaches (i.e. transmitting raw and processed data from sensors to zone ECU) align with the trends in IVN and E/E architecture, which either consider increased centralization with simpler sensors (requiring sensors to transmit raw data to zone ECUs) or preprocessing within the sensor (where sensors transmit processed data to zone ECUs) [54].



a) Zone: front let & front right



b) Zone: rear left & rear right



c) Zone: center

FIGURE 45. BW ESTIMATION TO SUPPORT THE SENSORS DATA TRANSMISSION

3.3.3.6 Next steps

Future work will focus on exploiting predictability for the simultaneous support of deterministic time-sensitive traffic and best-effort traffic in subnetworks. On one hand, data-driven and AI-powered scheduling mechanisms will be designed leveraging for example the traffic characterization performed in the 6G-SHINE D2.2 for in-vehicle networks [2]. Correlations between subnetworks' traffic patterns will be exploited to anticipate resource demands, program scheduling request, and allocate resources when and where necessary. In addition, we will continue exploring the possibilities of integrating predictive scheduling and flexible duplexing for the advanced support and multiplexing of mixed traffic flows. The utilization of RIS elements to further control the propagation conditions and avoid potential blockage between transmitters and receivers will also be explored. The introduction and strategic location of RIS elements in subnetworks can also be considered as a complementary tool for the enhanced predictability of the subnetwork operation and hence for the improved traffic scheduling.

3.4 Multi-Link Solutions for Improved Reliability

Despite the opportunities posed by limited range, establishing highly reliable, low-latency communication within industrial subnetworks—particularly those governing robot control systems—remains a formidable task, especially in scenarios where the desired signal might be shadowed by, e.g. metallic machinery parts. The methods presented in the previous sections have addressed the problem of reducing latency and reliability via enhanced transmission procedures, and predictive schedulers. In this section, we explore the possibility of leveraging multilink protocols -either based on cooperative communication and network coded cooperation- for further improving communication reliability.

3.4.1 Cooperative in-X Communication

The focus of the cooperative communication method proposed here is on industrial subnetworks, though the method can be adapted to other subnetwork types operating with tight communication cycles. Our hypothesis asserts that facilitating ultra-short communication cycles within factory subnetworks demands the integration of link diversity to counteract blockage effects, due to e.g., metallic obstructors [55]. Extensively documented in existing literature [56][57], relaying emerges as a well-established and effective strategy for mitigating fading by leveraging spatial diversity. Notably, studies have evidenced that two-hop relaying protocols substantially improve both capacity and the quality of service [58]. The efficient integration of relays into the subnetwork architecture necessitates tailored communication protocols.

A multitude of investigations have delved into employing relays in URLLC networks, as evidenced in [55]-[60]. The concept explored in [57] regarding cooperative transmission seeks to maximize the potential of wireless networks through collaborative device-to-device relaying for improved reliability. The central emphasis within [57] revolves around creating transmission methods that function independently of transmitter CSI. Although these CSI-agnostic strategies eliminate the necessity for transmitter CSI, they lack adaptability to fluctuating instantaneous channel conditions. As a result, they typically opt for a suboptimal and cautious strategy when determining transmission rates, often based on the weakest user's conditions, leading to diminished spectral efficiency. On the other hand, [55] introduces a specialized wireless communication protocol tailored explicitly for industrial control systems. This protocol employs cooperative transmission to assist nodes dealing with poor channel conditions and prioritizes direct transmission for stronger links. However, criteria for selecting between single-hop and two-hop devices are not explicitly addressed. Additionally, the protocol's reliance on fixed power levels leads to inefficiencies in power emission. Conversely, [59] presents an algorithm aimed at minimizing transmit power for cooperative communication within smart factories. This algorithm focuses solely on throughput and reliability constraints, disregarding timing constraints introduced by short control cycles. In another study [60], the focus is on utilizing unmanned aerial vehicles (UAVs) as decode-and-forward (DF) relays to facilitate the transmission of short URLLC control packets between a controller and multiple mobile robots. Notably, the authors focus on the optimization of time-varying optimal block lengths rather than the communication cycle time.

Furthermore, while the aforementioned research has explored relay-assisted URLLC transmission, it is crucial to note that none have specifically addressed the unique context of in-factory subnetworks. The solution presented here aims to enhance support for ultra-short, highly reliable communication while simultaneously reducing overall transmitted power in a subnetwork. This reduction not only aids in energy saving but also decreases interference generated to neighbouring subnetworks. Central to our approach is the utilization of LC elements, which relay packets from a selected pool of SNEs facing challenging propagation conditions to a HC with embedded controller capabilities. We address the critical issue of minimizing total transmit power while meeting stringent low latency criteria. In order to address the inherent nonconvex nature of the optimization problem, we utilize a promising

strategy called sequential parametric convex approximation (SPCA). This approach entails simplifying the problem through the use of successive convex approximations, ultimately transforming it into a conventional convex semidefinite programming (SDP) format. This transformation enables efficient solving through numerical solvers like the SDP tool available in convex optimization toolbox (CVX) [61].

3.4.1.1 Considered scenario

We consider an industrial subnetwork where N SNEs connect wirelessly to K LCs and one HC. The HC command the actuators, while the K LCs equipped solely with radio capabilities.

The set of all LCs is labelled as $\mathcal{K}$, while $\mathcal{N}$ represents the set of all SNEs within the subnetwork. SNEs can communicate directly with the HC through single-hop transmission or communicate with both the HC and LC. In the latter case, the LC forwards the received message to the HC, facilitating a two-hop cooperative transmission. $\mathcal{N}_{1h}$ indicates the set of SNEs scheduled for single-hop direct transmission, while $\mathcal{N}_{2h}$ denotes the set of SNEs selected for two-hop cooperative transmission.

Figure 46 depicts an example subnetwork with 5 SNEs and 2 LCs, where SNEs A, B, and C operate in two-hop cooperative mode, while SNEs D and E function in single-hop mode. Notably, signals transmitted by SNEs in cooperative mode are also received by the HC, allowing for potential energy combination between the signals received by the SNEs and the LC.

All LCs are assumed to be synchronized in time and frequency, while SNEs are randomly distributed across the subnetwork area. SNEs are allocated time resources in a time division multiple access (TDMA) manner to prevent intracell interference. In this study, our focus lies solely on the UL, where a transmission consists of a packet of B_n bits, for $n = 1, \cdots, N$, from a SNE to the HCs/ LCs across a bandwidth of W Hz. It is crucial that the HC receives the packets transmitted by all SNEs within a time slot of T seconds.

We assume that the HC possesses knowledge of the channel responses across all communication links between the SNEs and the HC/LCs, as well as between LCs and the HC. This information is obtained in a training phase before operation, where channel responses are estimated at the HC and LCs through the transmission of reference sequences. In the latter case, this data is reported to the HC. We assume here that channel responses are expected to remain relatively stable within a subnetwork, i.e. we are considering what is known as a quasi-static case. This means that although the channels might undergo changes due to movement, it is possible to update the CSI by repeating the training phase. To accommodate the dynamic nature of robots with fast-moving parts, predictive mechanisms as the ones introduced in section 3.3.2 can be incorporated. This prediction will guide the timing for repeating the CSI acquisition process, ensuring the system remains accurate and responsive to movement. The HC scheduler then utilizes complete CSI from all subnetwork links to determine the subset of SNEs for single-hop and two-hop transmission. This determination also includes their transmission rate (thus, the allocation of time resources) and transmit power.

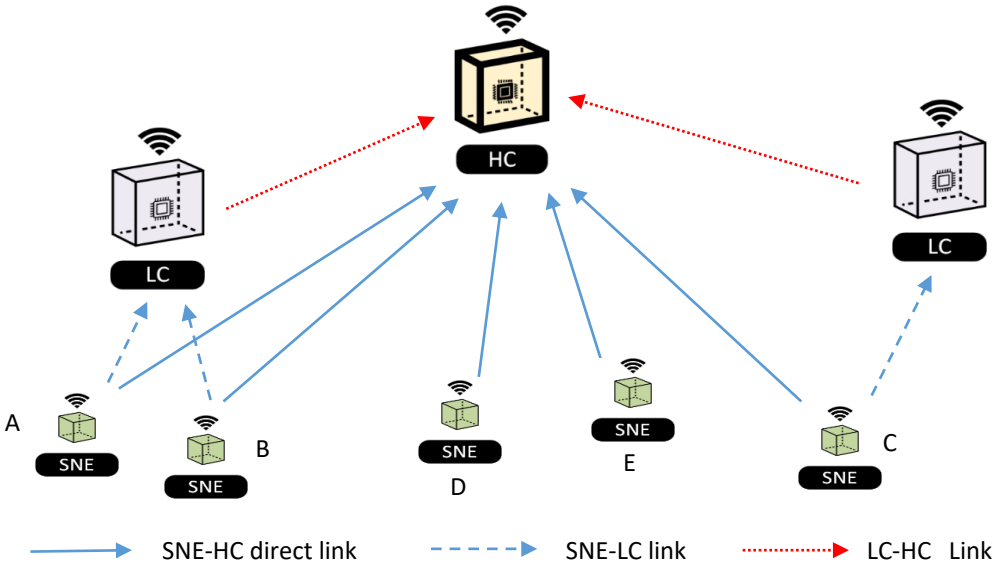


FIGURE 46. SYSTEM MODEL FOR MULTIPLE LCS SUBNETWORK TRANSMISSION.

3.4.1.2 Proposed communication protocol

We assume that the presence of LCs and their forwarding capabilities can significantly increase communication reliability within a subnetwork. Leveraging the potential of LCs necessitates a custom-tailored communication protocol.

Our proposed protocol, depicted in Figure 47 and compared with the basic TDMA protocol, divides the UL time slot into three sub-slots of variable duration. In this protocol, the first phase sub-slot is dedicated to SNEs belonging to the $\mathcal{N}_{2\mathcal{h}}$ set, facilitating their transmission. These transmissions are received by both the HC and LC, leveraging the most favorable channel conditions offered by a specific LC for each SNE within $\mathcal{N}_{2\mathcal{h}}$. The subsequent phase is allocated to the LC for relaying the received messages from SNEs to the HC, acting as a DF relay. It is worth noting that the HC combines energy received from the SNE during the first phase with the energy obtained from LC in the second phase.

In contrast, SNEs scheduled for single-hop transmissions utilize the second sub-slot, transmitting solely to the HC. It is essential to note that accommodating the two phases of cooperative operation within the UL slot necessitates shorter transmission intervals for SNEs in $\mathcal{N}_{2\mathcal{h}}$ compared to the basic TDMA protocol. Consequently, a higher transmission rate is required for their packets. Additionally, the proposed protocol incorporates single-hop transmission in the second sub-slot to accommodate the processing time required for DF relaying.

For a comprehensive description of the signal model, please refer to [62]. 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 [62].

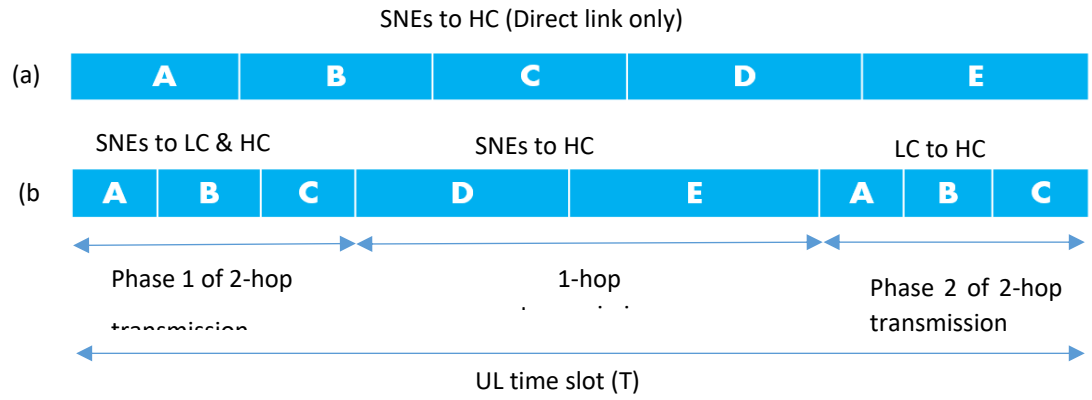


FIGURE 47. COMPARISON OF COMMUNICATION PROTOCOL IN SCHEME (A) SINGLE-HOP TRANSMISSION AND (B) MULTIPLE LCS TRANSMISSION.

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 [62].

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 $0 < \theta < 1$. 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.

3.4.1.3 Proposed method for transmit power optimization

Within industrial subnetworks, minimizing the total emitted power not only extends SNE battery life but also diminishes interference in neighboring 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. The optimization problem can be formulated as:

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

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

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

$$P_{D_n}^{S} \le P_{max},\ \forall n \in \mathcal{N}_{2\hbar}$$

where P_n and $P_{D_n}^{S}$ represent the transmit power of each SNE and LC, respectively. $\hat{g}_n^d$, $\hat{g}_{n,D_n}$, 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 [62] for detailed insights.

3.4.1.4 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. The deployment of SNEs and HC/LCs is uniformly spread across the subnetwork area. Our model entails each SNE transmitting a 32-byte packet 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 established at -174 dBm/Hz.

The wireless channels between SNEs and HC/LCs, along with those between LCs and the HC, are subjected to independent frequency flat Rayleigh fading. The path loss model adopted aligns with the factory and open-plan building channel model used in [55]. To incorporate shadow fading, consistent with [63], our scenario incorporates a shadow fading model with a standard deviation of 7 dB. These assumptions will be revised in the second phase of the project, based on the input from WP2.

We explore cooperative schemes employing different numbers of selectable LCs, including '1h' and '1 of K' in P-CSI case. In Figure 48, we present a comparative analysis of the empirical CDF showcasing the transmission power per SNE and LCs across various schemes, assuming a maximum power P_{max} of 20 dBm.

Our simulations are framed around a half-duplex UL scenario employing TDMA, where ideally, there is no interference within this setup. Nonetheless, potential interference from adjacent subnetworks could be further explored in future. Incorporating the modelling of a non-ideal transceiver represents an additional dimension that could be considered in future to enhance realism in practical scenarios.

Our findings indicate a reduction in transmit power ranging from 4 dB to 7.5 dB for the two-hop schemes compared to the single-hop scheme. Specifically, the '1 of 1' scheme achieves a median saving of 4 dB, while the '1 of 4' scheme can save up to 7.5 dB in transmit power. 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 [62]), we analyse the CDF of transmit power in the P-CSI case, as depicted in Figure 48, 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.

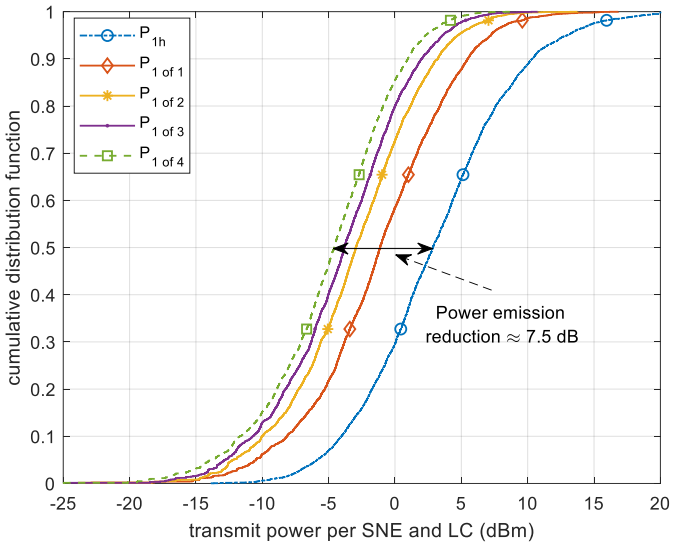


FIGURE 48. CDF OF TRANSMIT POWER ACROSS VARIOUS SCHEMES WITH FOR P_{MAX} = 20 DBM IN P-CSI CASE.

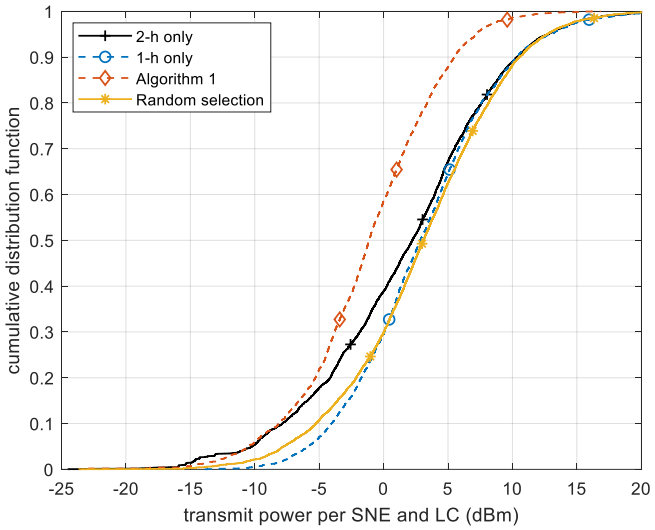


FIGURE 49. CDF OF TRANSMIT POWER FOR VARIOUS SNE CLASSIFICATION METHODS IN P-CSI CASE.

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 50, 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 fulfill time constraints, defaulting to maximum power in worst-case scenarios. However, if this level proves inadequate, overflow instances occur.

In the context of the I-CSI scenario, we set *L = 4*, dedicating less than 1% of time resources to the training phase. This allocation is based on empirical selection and adheres to the rule of thumb that a contribution of 1% can be considered negligible in the context of total power emission. To minimize channel estimation error, maximum power is applied during training, followed by the power control method for the remaining time slots. Given the marginal allocation (less than 1%) to training, its influence on average power transmission can be considered negligible.

The outcomes presented compare θ = 0.5 and θ = 0.9 in the I-CSI case with the ideal P-CSI case. The trend observed indicates that as θ increases, the overflow rate decreases, converging toward the P-CSI scenario. For P_{max} = 30 dBm, all two-hop transmissions adhere to the minimum overflow rate requirement of 10^{-6} even in the I-CSI case with θ = 0.5. In contrast, single-hop transmission fails to achieve better than 2×10^{-4}. A P_{max} in the order of ~11 dBm and a '1 of 4' scheme is instead needed for supporting the 10^{-6} minimum overflow rate requirement, assuming θ = 0.5.

In Figure 51, we illustrate the impact of the discount factor on error probability, comparing results across θ values ranging from 0.5 to 0.9. Particularly, the overflow rate and error probability represent conflicting requirements: as the discount factor approaches 1, the overflow rate decreases because SNE packets might be allocated onto a smaller number of resources. However, this reduction leads to a higher error probability. In contrast, decreasing the value of θ results in a higher power demand to meet stringent low-latency requirements.

To maintain an acceptable overflow rate, P_{max} is set to 25 dBm in this analysis, as shown in Figure 51. The '1 of 3' and '1 of 4' schemes successfully meet the specified constraint of $P_e < 10^{-6}$ for θ values below 0.7 and 0.8, respectively. Conversely, other schemes fail to meet this constraint. The feasibility of employing a limited discount factor (and thereby a restricted power increase) is attributed to the generally advantageous propagation conditions within a short-range subnetwork.

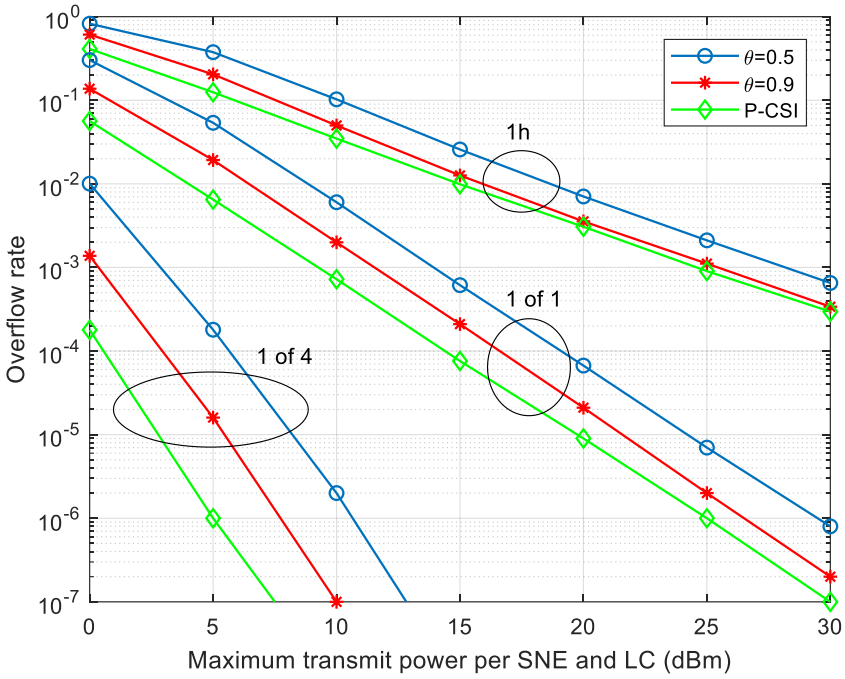


FIGURE 50. OVERFLOW RATE AGAINST MAXIMUM TRANSMIT POWER OF DIFFERENT SCHEMES FOR P-CSI CASE AND I-CSI CASE WITH L = 4.

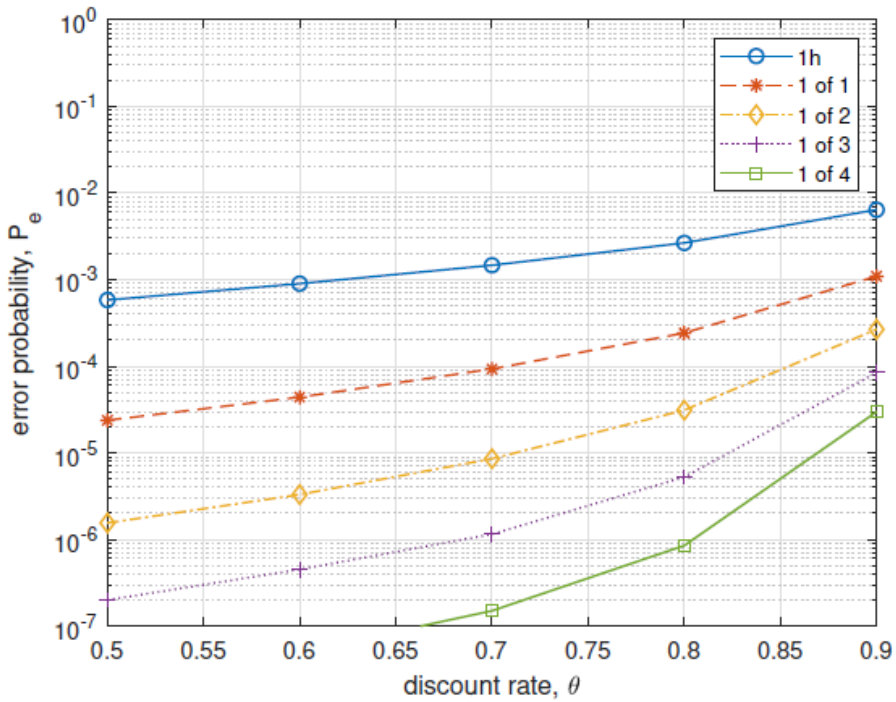


FIGURE 51. P_E AGAINST Θ FOR I-CSI CASE WITH PMAX = 25 DBM AND L = 4.

3.4.1.5 Summary of the initial findings

We have tackled the challenge of ensuring successful communication within an industrial wireless subnetwork while adhering to tight cycle time constraints for numerous SNEs. Our approach centers on the presence of multiple LCs equipped with relaying capabilities, coupled with a novel communication protocol accommodating SNEs operating in either a single-hop or a two-hop manner. Additionally, we have addressed the crucial task of minimizing emitted power while navigating timing constraints, employing sequential parametric convex approximation techniques. Through simulation, we have validated the efficacy of the proposed protocol.

Specifically, employing a '1 of 4' scheme yields a remarkable 7.5 dB reduction in power compared to optimized single-hop transmission. Furthermore, cooperative transmission dramatically decreases error probability and overflow rate in contrast to single-hop scenarios. For instance, with a maximum transmit power of around ~11 dBm, the overflow rate diminishes from 0.1 in single-hop transmission to an impressive 10^{-6} in the case of two-hop transmission.

3.4.1.6 Next steps

Our future work aims to broaden the analysis to encompass amplify-and-forward relaying within subnetworks. We will also conduct a comparative analysis of relay-based methods and multiple RIS systems, evaluating the latter as a potential alternative to traditional relays. Moreover, we intend to tackle the support of services with heterogeneous requirements within the cooperative framework, leveraging the capabilities of flexible/full duplexing-capable LCs.

3.4.2 Network Coded Operations

As mentioned above, 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 machines and 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. Since programmable logic controllers (PLC) typically work in deterministic cycle times, not only the latency of a transmission but also the round-trip time of a communication becomes very important.

Besides the cooperative approach presented above, 6G-SHINE is exploring combined approaches for multilink transmission and cooperative methods using network coding, especially for low cycle times and extremely high reliability. The challenge is to design a transmission and coding scheme optimized in terms of challenging KPIs for subnetworks with multiple devices. Innovative Multi-Link Operation (MLO) based network coded cooperation (NCC) concepts are to be designed and evaluated with the parameters and properties of available radio technologies.

Similarly, to the previous study, we consider a TDMA protocol where transmission resources are allocated to each SNE according to a predefined schedule. TDMA protocols mostly specify a deterministic cycle time where several

SNEs are transmitting to one BS in a time multiplexed fashion, and bidirectional communication on the application layer takes place between one BS and SNEs (Figure 52).

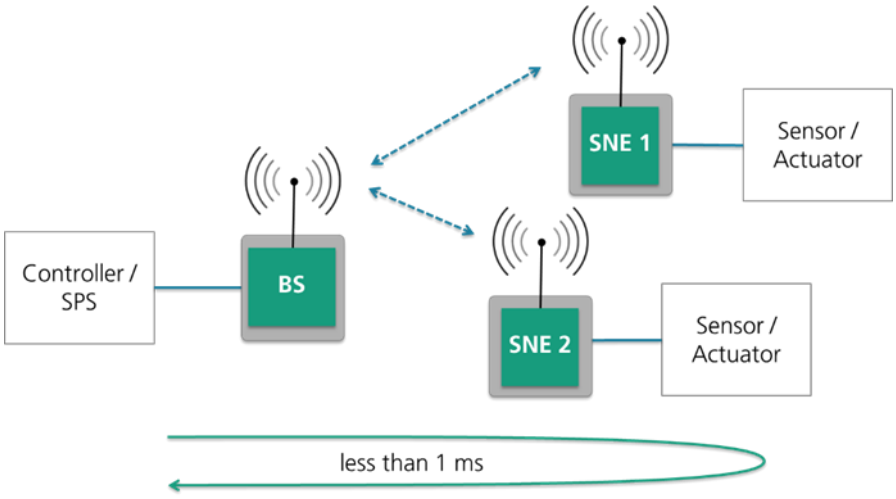


FIGURE 52. SCHEMATIC ILLUSTRATION ULTRA RELIABLE WIRELESS INDUSTRIAL NETWORK

In contrast to typical single link physical layer FEC, cooperative communication and network coding describe procedures that can be employed at the link level and benefit from the involvement of several radio participants [64]. The combination of scheduling techniques (e.g. TDMA) and network coded cooperation is promising [65], as the determinism of medium access allows the precise planning of combined packet encoding opportunities.

The MLO approach was introduced by IEEE with WiFi 7 (IEEE 802.11be), while in former versions, even WiFi 6 enhancements, a multi-link operation was not possible. A packet-level link aggregation in the MAC layer across different PHY links is provided. Hence it is possible to transmit data in parallel in 2.4, 5 and 6 GHz frequency bands, with a bandwidth of up to 320 MHz and a 4K QAM modulation. MLO is used to improve the link resiliency by packet duplication over multiple links or to increase the throughput by using the links independently.

There are also approaches which investigate the benefit of applying network coding to a dynamic multi-link transmission technology. In [66], network coding is only employed at the application layer and is used to increase the reliability. However, the usage of cooperative communication and the increase of spatial diversity is outside the scope of the cited paper. Therefore, the full potential for an increase in reliability is not utilized.

3.4.2.1 Network Coded Cooperation in TDMA/TDD-Systems

We here recall the basic principle of network coded cooperation. Let us consider a bidirectional data exchange taking place between one BS and SNEs, resulting in a star topology. In a network with two SNEs, the BS sends a data packet “A” to the SNE1 (downlink) and the SNE1 sends a data packet “a” to the HC (uplink). Accordingly, the HC sends data packet “B” to SNE2 and SNE2 sends data packet “b” to the BS, as illustrated in Figure 53. Due to the

overhearing effect, however, the actually uninvolved radio node within radio range can still receive a data packet that is not intended for it. Cooperative procedures use this effect by retransmitting received packets that are intended for other participants and thus forwarding them. In the example, SNE1 retransmits the received data packet “B”.

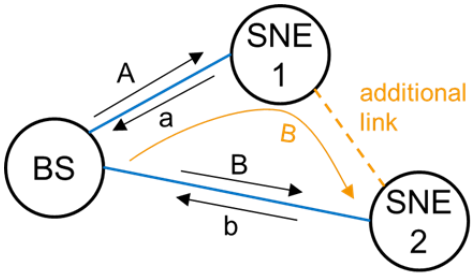


FIGURE 53. DATA PACKET FLOW WITH COOPERATIVE COMMUNICATIONS

As the transmission at link layer level takes place via several, spatially separated paths, cooperative communication increases the diversity order and the communication robustness. A second method is now used for forwarding the data packets, which is network coding. Here, data packets are coded with each other at the symbol level. This coding is based on arithmetic in finite fields and, in the simplest case, can be implemented in binary form with a bitwise XOR operation (encoding process). The encoded packet is thus no longer but carries both pieces of information. According to a coding scheme, packets are transmitted either natively (e.g. “A”) or combined coded (e.g. “A⊕B”). When evaluating, it should be noted that the receiver of a combined packet can only reconstruct the contained native packets if it receives other suitable combined coded or native packets (decoding process).

Past research has investigated the applicability and benefits of such network coded cooperation for a TDMA/TDD systems like Ultra Reliable Wireless Industrial Network (UWIN) by Fraunhofer IIS. Two error control methods were investigated and compared with a single transmission: blind retransmission (which is a simple repetition of the packet) and network coded cooperation with a first systematic coding scheme approach. Figure 54 shows the transmission and coding schedule for the described packet data flow from Figure 53. Note that the considered TDMA frame, consisting of 12 slots, has a duration of 125 µs.

Single Transmission
No error correction
4 TX slots

	TDMA Frame												
Slot	0	1	2	3	4	5	6	7	8	9	10	11	12
BS	TX "A"	TX "B"				RX	RX						
SNE1	RX					TX "a"							
SNE2		RX					TX "b"						

Blind Retransmission
8 TX slots

	TDMA Frame												
Slot	0	1	2	3	4	5	6	7	8	9	10	11	12
BS	TX "A"	TX "B"	TX "A"	TX "B"		RX	RX		RX	RX			
SNE1	RX		RX			TX "a"			TX "a"				
SNE2		RX		RX			TX "b"			TX "b"			

Network Coded Cooperation
7 TX slots

	TDMA Frame												
Slot	0	1	2	3	4	5	6	7	8	9	10	11	12
BS	TX "A"	TX "B"	TX"A⊕B"			RX	RX		RX	RX			
SNE1	RX	RX	RX			TX"B⊕a	RX		TX"a⊕b"	RX			
SNE2	RX	RX	RX			RX	TX"A⊕b"		RX	TX"a⊕b"			

FIGURE 54. TRANSMISSION AND CODING SCHEMES TO COMPARE ERROR CONTROL MECHANISMS

The packet loss rate, i.e. the probability that a packet cannot be received correctly at application level, was determined for these three transmission schemes. For this purpose, an equally distributed packet error rate of 1% at link level was assumed. As we consider dense subnetworks, we also assume that all nodes are within coverage area, i.e. a link between any two nodes exists. Table 10 shows the packet loss rates resulting from a stochastic simulation. In the case of a simple transmission, each packet error also results in a packet loss. With blind retransmissions, there is a quadratic improvement in the packet loss rate (assuming all retransmissions are experiencing uncorrelated channel conditions) compared to the single transmission. NCC results in a significant further improvement in reliability and a reduction in the packet loss rate with fewer used time slots compared to blind retransmission. This results from the additional diversity path via the respective other SNE.

TABLE 10. COMPARISON OF PACKET LOSS RATES (NETWORK CODED COOPERATION)

Transmission Scheme	Number of used slots	Packet loss rate			
		A BS->SNE1	a SNE1->BS	B BS->SNE2	b SNE2->BS
Single Transmission	4	1.0×10^{-02}	1.0×10^{-02}	1.0×10^{-02}	1.0×10^{-02}
Blind Retransmission	8	1.0×10^{-04}	1.0×10^{-04}	1.0×10^{-04}	1.0×10^{-04}
Network Coded Cooperation	7	3.8×10^{-06}	2.9×10^{-06}	4.1×10^{-06}	4.2×10^{-06}

The major challenge in NCC is the joint definition of a transmission strategy with a coding strategy, especially in subnetworks with many devices. With a high number of nodes, the protocol overhead of the signaling required for network coding must be taken into account.

As a planned innovation, MLO will add transport capacity per unit of time and together with NCC this additional capacity can be used to flexibly add redundancy to improve reliability or to support a larger number of SNEs in subnetworks.

3.4.2.2 Simulation Model and Initial Results

In 6G-SHINE, the simple stochastic simulation is currently being extended to a network simulation, assuming certain spatial SNE constellations or network sizes. For this purpose, a model of the protocol stack is being developed, with the focus on the MAC and data link layer since this is where the network-coded cooperation is to be handled. It is assumed that due to the short spatial distance of the devices within the subnetwork, all SNEs are generally within radio range, which means that a channel between all nodes exists, and the overhearing effect can be used. This results in an additional link between any two SNEs at the data link level as illustrated in Figure 55. These links can be used to make the connection more reliable at application level.

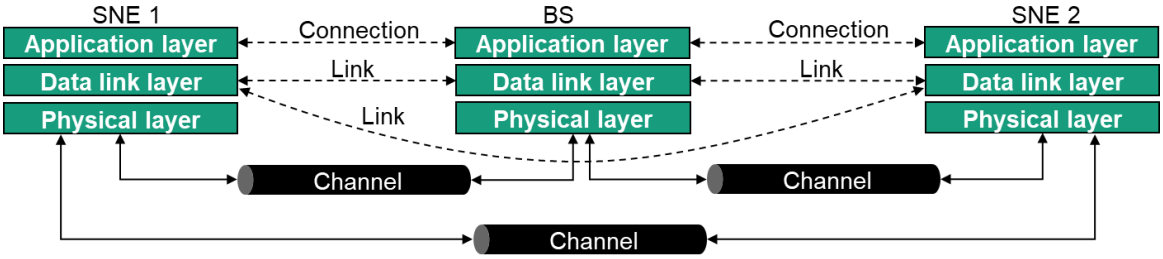


FIGURE 55. DATA LINKS VS. CONNECTIONS

In order to investigate innovative NCC with MLO, the simulation model is being developed in the first project phase of 6G-SHINE for modelling the transmission and coding schemes and for developing allocation strategies for subnetworks with multiple devices. The developed model supports:

- TDMA multiframes, which comprise several TDMA frames. This results in multiples of the transmission capacity of a single frame for longer cycle times. For example, a cycle time of 1 ms results in 8 frames of 125 µs per multiframe for UWIN.
- Multiple frequency channels that can be used simultaneously in parallel and thus form the basis of the MLO.

The activity of every node in a dense subnetwork is defined by one global resource allocation matrix (see Figure 56). An activity takes place in a slot on a logical channel. The logical channel is assigned to the frequency channel in the PHY. For the shown example, it schedules a transmit activity for SNE1 in slot 5 of frame 0 on the logical channel 1 and RX activities for all other devices. The global resource allocation matrix also defines the concrete network coding scheme which is employed. For example, it specifies for every transmit activity either a native (uncoded) packet to transmit or a rule for the assembly of one network encoded packet. In the case of Figure 56, for SNE1 the TX packet has to be encoded by using Packet 1, Packet 2 and Packet 3. A defined network coding scheme follows the concept of a systematic network coding method, which we consider as superior to random linear network coding for deterministic data traffic patterns as it requires a lower signalling overhead.

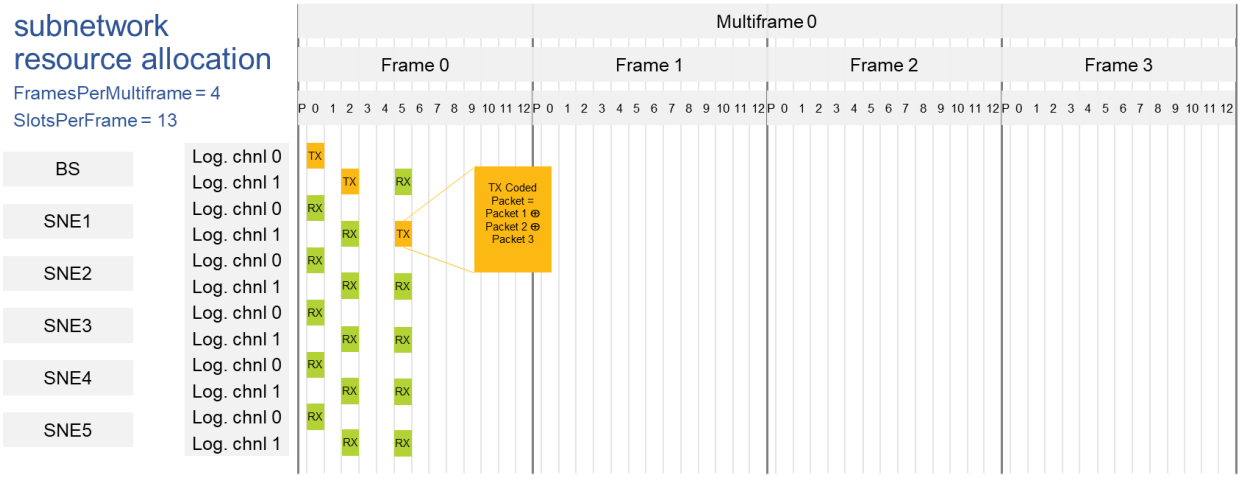


FIGURE 56. NETWORK SCHEDULE MATRIX.

The results achieved currently include modelling concepts and implementations for the 6G-SHINE subnetwork macro diversity simulation model using a network simulation framework.

3.4.2.3 Next steps

Regarding the multi-link MAC designs using network coded cooperation, the implementation work to realize the model in a network simulation framework has begun and will be followed by work on transmission and coding schemes. The KPIs determined in simulation can then be used to optimize the generation of the transmission and coding schemes. In order to reduce the protocol overhead for NCC, clustering procedures are to be investigated. This involves forming groups of nodes that work together using NCC.

In order to base the simulation results on realistic industrial scenarios, it is planned to include PHY Layer effects, including radio propagation, into the network simulation.

3.5 Solutions For Improved Latency in The Unlicensed Spectrum

Subnetworks are expected to operate over different spectrum bands, some of which have been explored in previous sections (28 GHz, sub-THz, etc.). Also, both options of licensed and unlicensed spectrum are to be considered for subnetworks. Access to the unlicensed spectrum can be subject to tight regulations, such as a listen before talk policy, which may affect latency and therefore time-critical operations in subnetworks. In this section, we explore enhancements for reducing latency for operations in the unlicensed spectrum.

3.5.1 Latency Aware Access in unlicensed Bands

We introduce here an innovative medium access control mechanism for unlicensed spectrum operations. We consider the exemplary case of ISM bands, though the proposed solution can be applied to other systems subject to similar regulations in terms of channel access. The proposed solution can coexist with the current standardized approach enforced today by ETSI as far as best-effort traffic is concerned, as it introduces a non-standard way to support real-time traffic prioritization in the MAC for unlicensed bands, making sure it is granted directly higher priority in the medium access mechanism against the best-effort approach enforced today based on clear channel assessment (CCA). We examine the impact of this new mechanism on the standard best effort CCA based channel access mechanism regarding medium access latency and possible collisions as well as its performance against several related parameters.

3.5.1.1 General channel access mechanisms to existing standardized bands

Various ETSI standards are organizing and defining the minimum requirements for accessing unlicensed bands, including ISM bands. For this work, we focus on the 2.4 GHz ISM band but the approach is easily extendable and applicable to all other unlicensed bands where similar rules apply. The standard that describes the general rules for accessing the 2.4 GHz ISM band is the ETSI EN 300 328 V2.2.2 [67]. The general access rule for 2.4 GHz ISM band is depicted in Figure 57. According to the standard, the minimum requirement for non-load adaptive devices is to perform a CCA check for at least 18 µsec (for non-adaptive frequency-hopping spread spectrum (FHSS) adaptive devices, i.e. devices that are able to hop over multiple channels) or 0.2% of the Channel Occupancy time in case of Adaptive FHSS devices using LBT, whichever is bigger.

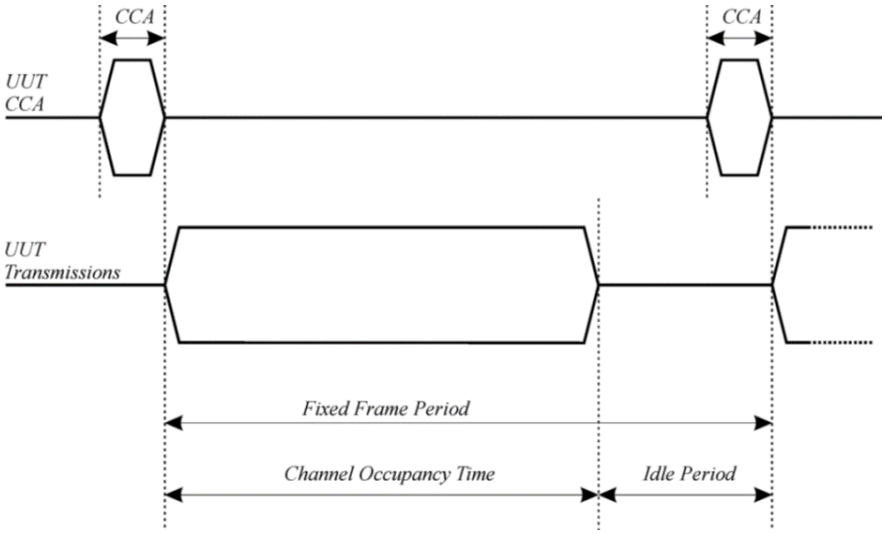


FIGURE 57.GENERAL ACCESS RULES FOR 2.4 GHZ ISM BAND BASED ON CCA AS DEFINED BY ETSI EN 300 3298 [67].

These minimum requirements were defined based on the capabilities of devices to sense the medium and react in a fast manner taking into account the nominal turnaround time of most devices, which at the time being is in the order of few µsec. Families of standards like IEEE 802.15 [68] and IEEE 802.11 [69] but also 3GPP TS 37.213 [70] are built on top of these general requirements and define further rules on the sensing time needed to access the medium. For instance, IEEE 802.11n [71] defines the minimum sensing time that the medium needs to be non-busy as *DIFS = SIFS + 2·Slot_Time* which for a direct sequence spread spectrum (DSSS) based PHY is calculated as: DIFS = 10 + 2 * 20 = 50 µsec. Hence, a IEEE 802.11n device employing a DSSS PHY layer is actually sensing for a longer time than the minimum 18 µsec defined for the 2.4GHz ISM band regulation from ETSI. Of course IEEE 802.11 Distributed Coordination Function (DCF)- based devices have built on top a backoff mechanism that make it even more complicated to access the medium in favour of supporting higher load or networks of higher density. The backoff mechanism can make a device wait for as long as *aCWmax* (1023) * *aSlotTime* (20) = 20460 µsec of medium idle time in worst case before accessing the medium, where *aCWmax* is the maximum duration of the contention window measured in number of slots and *aSlotTime* is the defined slot time duration.

We will focus on the basic requirements coming from the ETSI standard, as this is technology agnostic and is setting the basis for all ISM based technologies general rule of operation.

3.5.1.2 Proposed latency aware channel access scheme

Based on the presented general access rules set by ETSI for ISM bands, we will present here an innovative approach aiming to push forward the standardized capabilities, adding time aware medium access capabilities to all ISM enabled devices.

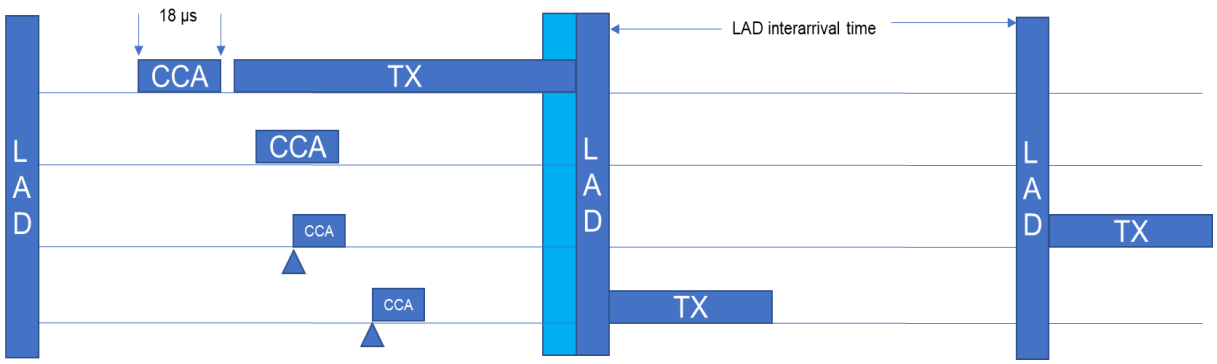


FIGURE 58. LATENCY AWARE DECONFLICTION PERIOD STRUCTURE AND TIMINGS

We propose the introduction of a time period every specified interarrival time that will support real time traffic only, called latency aware deconfliction (LAD). By real time traffic, we refer to traffic flows that have a hard or soft deadline requirement for end-to-end latency. If the differentiation of incoming traffic is made between best effort and real-time (or priority classes, as this is the most typical differentiation), then best effort traffic/low priority class can continue to access the medium as being done today (supporting backwards compatibility) while the real time/high priority traffic can access the medium in a deterministic manner employing the LAD periods. As it can be seen in Figure 58, LAD periods will be present in the medium access scheme, placed in a specific LAD interarrival time (LAD-IAT) taking precedence over the best effort access scheme by using a CCA time shorter than 18 µs. The exact timing will be determined after simulation tests for the Time Aware CCA (TA-CCA) as well as for the LAD-IAT. The integration of the LAD assumes that a time synchronization service is present for all active nodes, making sure there is a µs level synchronization between active nodes. Every LAD is defined to start at a specific time, but it will shift to wait for any active transmission during the best effort period like it can be seen in Figure 59, on the middle LAD instance. As there is a transmission active, LAD is shifted until that transmission is finished and only then, LAD procedure takes over. How this shifting happens, and the overall architecture of the LAD period, is depicted in Figure 59.

Once the LAD start time arrives, all nodes start performing CCA continuously, and only when they find the medium free for TA-CCA (let us assume a time of 10 µs for TA-CCA), they assume LAD to start, and the internal LAD operation commences. Once medium is found free for the TA-CCA time, all nodes having high priority traffic to send (in our example the 2 upper nodes only as shown in Figure 59) select a random step from S=1 to X (X to be defined, for this example set to 10) and start performing TA-CCA again. Each step S1 to S10 can be from 1 to 3 µs, as long as the max duration of *MAX_step + TA-CCA* is below 18 µsec to maintain priority over best effort traffic. If the channel is found to be free for TA-CCA time, then they transmit their packet. If at any point in time of TA-CCA a transmission is detected, all other nodes abort their own medium access effort and defer. By doing so, there is deconfliction between active nodes with priority traffic to send, equal to the number of steps. Statistically, with 10 steps available, there is a 10% chance of collision between active nodes. The only case where there will be a collision is if two or more nodes select the fastest active step (for example S3 is selected by 2 nodes and all the rest select S4 and below), hence not been able to detect each other transmission, as they will start transmitting at the same time. To further minimize the probability of collision, each node will randomize the CCA time done within each step to be (50% to 100% TA-CCA) with a randomization resolution of 1 µs (the optimal value is still under investigation), allowing to deconflict and avoid collisions even between nodes selecting the same fastest step randomly.

Lastly, specific enhancements like adaptive step size vs fixed step size are going to be evaluated. The adaptive step design refers to the ability of a node to halve the max step size in order to randomly select a step every time an

LAD period is passed without being able to access the medium. Hence, if for example the max step size is 16, then it will select a random step between 1 to 16 during the first LAD; and if it fails to access the medium, then it will select a random step between 1 to 8; in the next LAD, it will use 1 to 4 ending up in the minimum which is 1 to 2. This proposed adaptive step max size selection scheme enhances further the time awareness of the proposed MAC but could potentially lead to more collisions in the LAD period. To get a proper insight on the matter at hand, this will be simulated and evaluated before moving forward to the general evaluation of the proposed LA-MAC, in order to make a proper selection between the two possible maximum step calculation mechanisms and employ the selected algorithm in the simulation campaign.

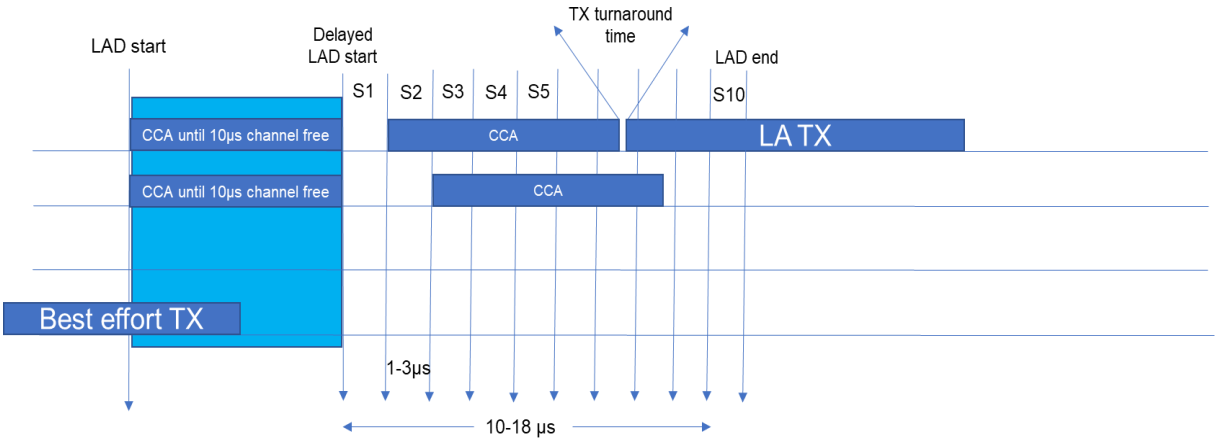


FIGURE 59. LAD INTERNAL OPERATION AND TIMINGS TO MINIMIZE PRIORITY PACKET COLLISIONS

The proposed method is implemented in ns-3 [72] and simulated in order to evaluate its capabilities and possible pitfalls. Crucial parameters are identified and evaluated against general MAC KPIs like Medium Access Delay and application layer KPIs like Packet Success Ratio and Throughput/Latency. In the next subsection the details of the ns-3 model implementation are presented and simulation results are discussed.

3.5.1.3 Simulation model description

In the ns-3 simulator, each new MAC layer is part of its corresponding net device, which is used to attach the MAC to the rest of the OSI stack, i.e. connecting new MAC to the network layers and PHY layers. Therefore, in the ns-3 a new net device called LANetDevice was introduced, which consists of two medium access schemes, namely the proposed Latency-Aware MAC (LA-MAC) scheme and typical CSMA-MAC, which is the standardized CSMA/CA medium access scheme available in the ns-3 libraries. LANetDevice also consists of 3 supporting modules LA-MAC modules: LAMacQueue, LAMacController and LAMacSchedule as it can be seen in Figure 60.

LA-MAC and CSMA-MAC are both responsible for processing the outgoing application layer packets, generation and embedding of MAC headers, trying to access the medium and sending the packets to the PHY layer for transmission. Further, they are responsible for the generation of acknowledgements, ready-to-send (RTS)/ clear-to-send (CTS) packets and retransmission of the data packets when needed. While one of these two MAC schemes

is active, the other one is disabled. Which one is going to be active is decided by the LAMacController, which consults LAMacSchedule and based on the predefined periodicity of the LAD zone and current time decides which MAC variant is active at any given time.

LAMacSchedule module stores all the LA-MAC related information, such as time information (time position in the scheme) of best effort medium access intervals and LAD zones, number of steps within LAD zone, step duration, CCA duration etc. This schedule can be accessed or modified by other modules in the LAMac net device. LAMacController accesses it to determine if it is time to use LA-MAC or CSMA-MAC and make an appropriate switch, while LA-MAC consults the LAMacSchedule module to acquire LA-MAC related parameters, which are employed for regulating access to the medium during LAD slots.

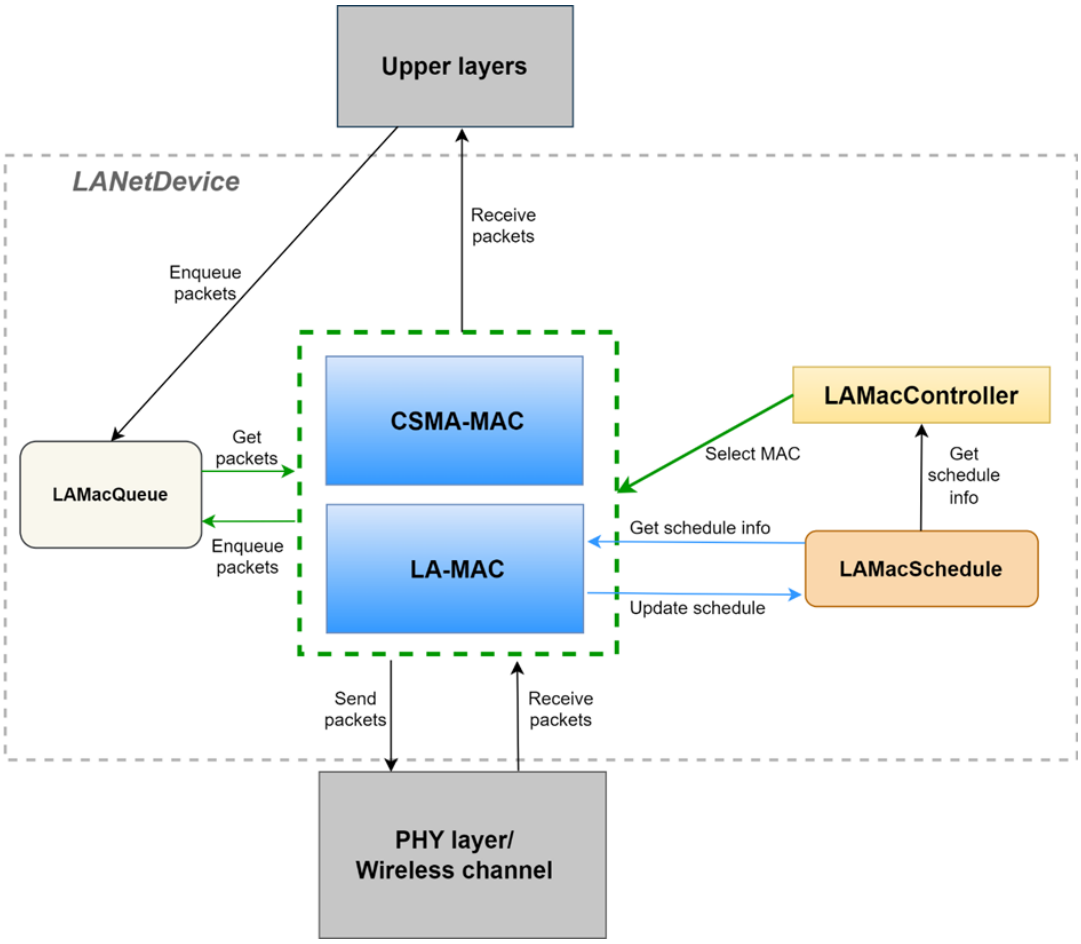


FIGURE 60. LAD SIMULATION MODEL ARCHITECTURE.

The purpose of LAMacQueue is to store the data packets, as they arrive from the upper layers of the OSI stack, and for retransmissions in case they were not successfully sent by the MAC layer. While for the data packets from the upper layers, the operation of adding packets to the back of the queue is performed, packets that are not successfully sent are added to the front of the queue. Packets waiting in the queue will be acquired by one of the available MAC schemes for further processing.

Finally, based on the timings, the appropriate MAC access scheme will be selected and the procedure to get access to the physical medium will be executed. Therefore, the implementation is a dual MAC scheme layer architecture, offering controllable time sharing between the standardized CSMA/CA and the LA-MAC access schemes while efficiently solving all transition issues during the switch of the MAC scheme.

3.5.1.4 Simulation results

3.5.1.4.1 Adaptive vs fixed # of maximum steps algorithm evaluation

Initially, an adaptive max step size adaptation algorithm was tested in order to evaluate its effect in latency and throughput on the proposed MAC scheme. We executed simulation runs for the maximum density of nodes that was planned to be evaluated, which is 32 nodes in the same collision domain. As the adaptive step enhancement will limit the possible selectable steps over time, it makes sense to evaluate its impact in the most high-density scenario. The tested algorithm defines that after a failed access attempt in an LAD, the relevant node will limit its possible range of step selection in the next LAD by half. As an example, in the first try the range would be 1 to *step_max* and if that access attempt would fail (i.e., it does not achieve to get access to the medium) then in the next LAD it would try 1 to *step_max*/2, in the next one 1 to *step_max*/4 and so on.

The setup includes 32 nodes in the same collision domain trying to establish and maintain 16 active data flows, 8 of high priority and 8 of best effort priority. Each data flow is a typical periodical traffic of ~8.38ms interarrival time with a 500 bytes packet size, resulting in a per data flow maximum throughput of ~476.8 kbps. The PHY layer rate used throughout in the entire simulation campaign was set to 11 Mbps.

The results between the static and adaptive adaptation scheme are depicted in Table 11.

TABLE 11. ADAPTIVE VS FIXED ADAPTATION OF MAX STEP SELECTION

	Maximum number of steps in LAD	
Metric	**Adaptive**	**Static**
Minimum access delay (μs)	15	15
Maximum access delay (μs)	39361.90	61026.90
Average access delay (μs)	6662.67	7369.85
SD of access delay – jitter (μs)	4789.81	6325.03
Average packet latency (μs)	24609.62	30968.14
Packet Success Rate (PSR) (%)	58.18	64.04
Total data send successfully (Mb)	106.5304	117.2611

It can be seen that the adaptive adaptation offers significantly better results in latency and access delay (maximum is lowered by almost 35%), while it mildly affects maximum throughput and PSR. Such a result can be explained by realizing that lowering the max selectable step increases chances of collision between competing nodes within the LAD in high density scenarios like this one. Hence the lower PSR of the adaptive scheme and the lower overall throughput. As we focus on latency aware characteristics though and the impact on Packet success rate (PSR) and Throughput are only lowered by around ~9.4% for both cases in the worst possible case, we will go ahead and execute the evaluation simulation campaign selecting as default the adaptive maximum step adaptation scheme as part of the LA-MAC.

3.5.1.4.2 LA-MAC evaluation and results

The simulation campaign for the LA-MAC performance evaluation focuses on 3 KPIs: network density, offered load and the LAD interarrival time. The general simulation parameters are depicted in Table 12, followed by the assigned parameters for the CSMA/CA based MAC model in Table 13 and lastly the LA-MAC model parameters in Table 14.

All simulation runs were executed 20 times using a random seed for the random generator of ns-3 and the results were averaged across runs. The LAD periodicity was set to 3 ms, except for the runs where LAD periodicity was the examined parameter. This means that a single priority packet can be send every 3 ms (with a typical transmission duration of ~541μs , the LAD airtime is calculated to be around 18% of total airtime), while the rest of the time MAC access is driven by the standard CSMA/CA. With a worst-case scenario of 46μs of sensing/waiting before accessing the medium in the LAD operation, LAD has hard priority over any legacy node running simple CSMA/CA only, as DIFS is 50μs. The CSMA/CA MAC parameter values are in line with the IEEE 802.11n DSSS PHY settings. The values used in the LA-MAC settings are an initial approach and are of course to be updated based on these initial results and further simulation studies.

Every simulation set generated a wide set of measurements, i.e bits per second, packets per second, average/min/max values of MAC access delay, application layer latency, access jitter, PSR and throughput. For the sake of space we will only present the most important results as a function of the number of deployed nodes, that being medium access delay and application layer latency.

TABLE 12. SIMULATION GENERAL PARAMETERS

Simulation duration	50 s
Number of simulations per use case	20
Number of nodes (N)	4, 8, 16 or 32
Number of active data flows	N/2 => 2, 4, 8 or 16
Nodes mobility	fixed positions in the same collision domain
PHY supported data rate	11 Mbps
Propagation delay	Simulated
Propagation loss	Simulated

TABLE 13. 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 14. LA-MAC PARAMETERS

LAD period	3 ms
LAD number of steps (step_max)	12
Step duration	3 μs

Initial CCA duration (TA-CCA)	10 μs
CCA duration during step	Random from CCAmin to CCAmax
CCAmin	5 μs
CCAmax	10 μs

In the case of four deployed nodes where 2 data flows are active (1 best effort and 1 priority), the baseline scenario was based on standard CSMA/CA MAC. The results are depicted in the table below:

TABLE 15. FOUR NODES - TWO ACTIVE DATA FLOWS BASELINE SCENARIO DETAILED RESULTS

Metric	CSMA MAC
Minimum access delay (μs)	50
Maximum access delay (μs)	3260.75
Average access delay (μs)	506.61
SD of access delay – jitter (μs)	419.64
Average packet latency (μs)	972.98
PSR (%)	99.99
Throughput (Mb)	45.7756

TABLE 16. FOUR NODES - TWO ACTIVE DATA FLOWS LA-MAC ENABLED SCENARIO DETAILED RESULTS

	MAC type	
Metric	**LA MAC**	**CSMA MAC**
Minimum access delay (μs)	15	50
Maximum access delay (μs)	1462.25	1821.75
Average access delay (μs)	37.79	763.51
Standard Deviation of access delay – jitter (μs)	57.10	90.36
Average packet latency (μs)	482.74	1210.68
PSR (%)	100	100
Throughput (Mb)	22.8881	22.8881

LOOKING AT THESE RESULTS, THE IMPACT OF THE PRESENCE OF THE LAD PERIOD IN THE LA-MAC ENABLED IS CLEAR AS FAR AS LATENCY AND MEDIUM ACCESS DELAY IS CONCERNED. AS THIS IS A LOW LOAD SCENARIO, BOTH VARIANTS ACHIEVE ~100% PSR, HENCE WE CAN FOCUS DIRECTLY ON THE ACHIEVED LATENCY. IN ALL LATENCY RELATED METRICS, THE LA-MAC ACHIEVES SIGNIFICANTLY LOWER LATENCIES COMPARED TO THE BASELINE, RANGING FROM 55% LOWER MAXIMUM MAC ACCESS DELAY TO 92% LOWER AVERAGE ACCESS DELAY. EVEN JITTER WAS ENHANCED TO AN ASTOUNDING 86% REDUCTION COMPARED TO THE BASELINE. LOOKING AT THE IMPACT OF THE EXISTENCE OF THE LAD PERIOD ON THE BEST EFFORT TRAFFIC (CSMA/CA LINE COMPARED TO THE BASELINE AT FIGURE 61 AND FIGURE 62. APPLICATION AVERAGE DELAY VS NODE DENSITY AND LOAD WITH AND WITHOUT THE 32 NODES RESULTS FOR BETTER CLARITY OF PRESENTATION

), still trying to access the medium in the standardized way of IEEE 802.11, we see that the impact is mostly within the limits of the expected. Around 18% of airtime is now allocated to the LA MAC, meaning that the CSMA MAC has only 82% of airtime compared to the baseline and therefore a decline in QoS characteristics should be observable. Indeed, average delay increased by around 250 μs on the CSMA/CA period, but the maximum medium access delay was 44% lower, dropping from 3260 μs to just 1821 μs. This is expected as the LAD period deconflicts a significant amount of priority traffic, allowing the load on the CSMA part to be less, reducing collisions and backoff

times. Jitter was also significantly decreased by ~78%. Lastly the application layer delay was increased in a similar manner to the MAC access delay for approximately 240 μs, an increase of ~24%. This increase is again in line with the 18% less airtime that CSMA/CA had in this scenario because of the existence of the LA-MAC.

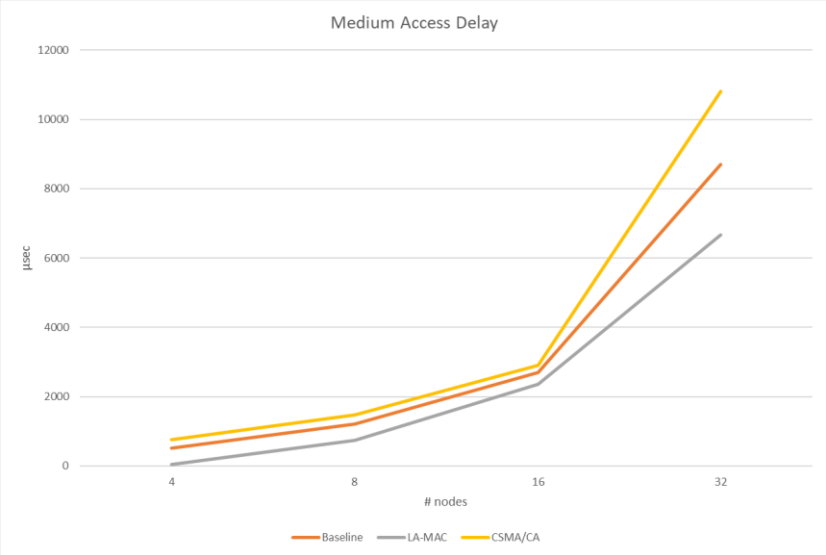
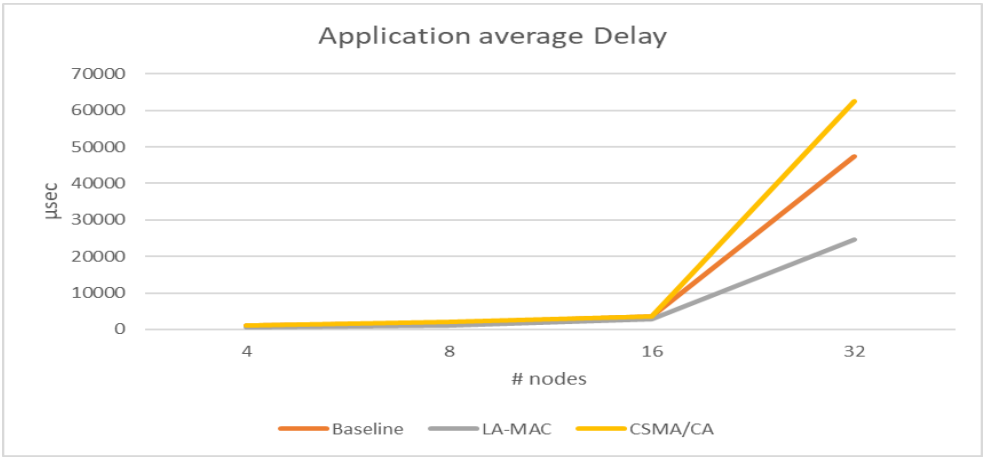


FIGURE 61. MEDIUM ACCESS DELAY VS NODE DENSITY AND LOAD.

Looking at the bigger picture, we now present the results of application layer average latency and medium access average delay over the whole series of simulation runs vs number of nodes (and load increasing analogous to number of nodes from 4 to 32 nodes, depicted in the X axis of the figures).



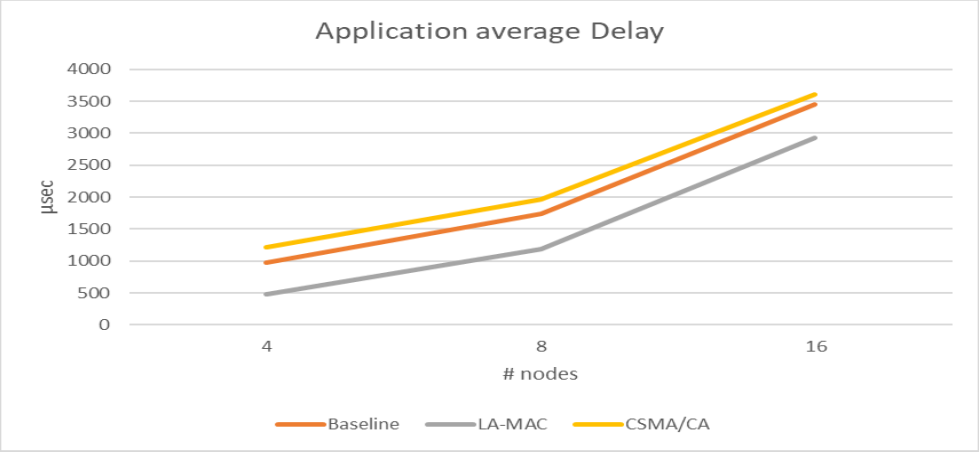


FIGURE 62. APPLICATION AVERAGE DELAY VS NODE DENSITY AND LOAD WITH AND WITHOUT THE 32 NODES RESULTS FOR BETTER CLARITY OF PRESENTATION

It can be seen that increasing load density has as expected an impact in the best effort traffic served by the CSMA/CA MAC, but the gains in the LA-MAC are far greater in all cases. We see LA-MAC achieving significantly lower latency in all scenarios compared to the baseline, being able to reduce application layer latency from 23% to 90% and medium average access latency from 24% to 50%. The most important thing to notice is that the impact in the best effort traffic is not significant compared to the gains of priority traffic, proving that the proposed LA-MAC can offer true priority services on the MAC layer.

To better understand the results, the application load and the capacity % of the medium is presented in Table 17.

TABLE 17. PHY AND LA-MAC LOAD PERCENTAGE BASED ON ACTIVE DATA FLOWS

Number of nodes	4	8	16	32
Number of active data flows	2	4	8	16
Total load (Mbps)	0,95	1,90	3,81	7,62
PHY capacity %	9%	17%	35%	69%
LA-MAC load %	24,0%	48,0%	96,1%	192,1%

We observe that for most scenarios CSMA/CA has to serve a load that is no more than 35% of the PHY capacity, but on the 32 nodes scenario the load is 69%, CSMA/CA is saturated and its QoS performance declined. The priority traffic on the other hand is not as severely impacted even when the LAD load % is almost more than 2 times the LAD capacity, meaning almost 1 out of 2 priority packets are not sent during the LA-MAC time, but are forced to be transmitted also during the CSMA/CA period. Let it be noted that all packets can use the CSMA/CA period but only priority packet flows can use the LA-MAC period by design. Hence, if the LA-MAC period is saturated, priority traffic is trying to access the medium also during the CSMA/CA period. That explains why the LA-MAC latency is very stable and significantly lower than the CSMA/CA period on all scenarios except the 32 nodes scenario, where LA-MAC is saturated and hence many of priority packets are sent using the standard CSMA/CA MAC scheme.

On the last part of our analysis, the proposed MAC scheme is evaluated as a function of the active number of priority data flows, in order to examine the priority load impact on the LA-MAC performance, as well as the impact

of varying LAD interarrival times (effectively varying the LA-MAC airtime) to QoS. For this part of the analysis, we focus on the 2 most demanding node density setups, that of 16 and 32 active nodes.

The results of varying priority aware traffic and their impact on the average access delay are presented in the figures below:

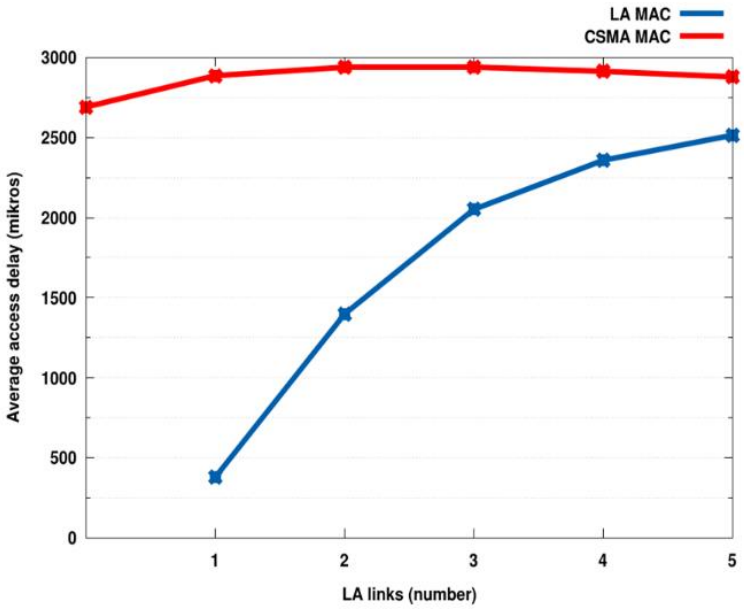


FIGURE 63. MAC DELAY VS PRIORITY LOAD FOR 16 NODES DENSITY

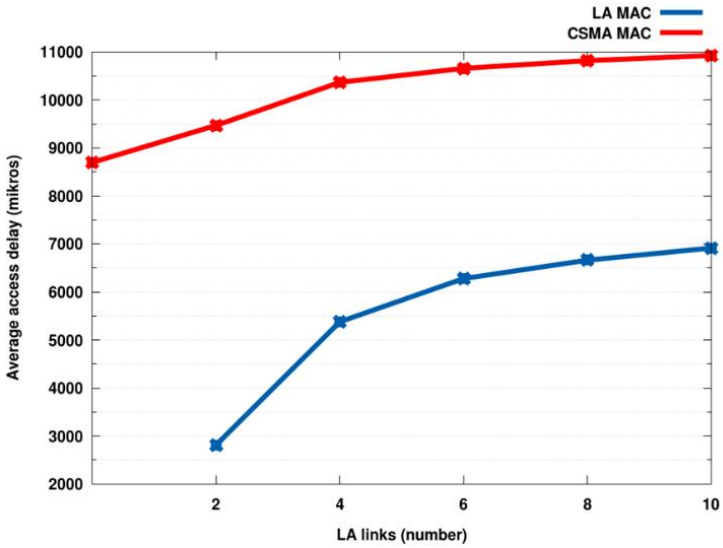


FIGURE 64. MAC DELAY VS PRIORITY LOAD FOR 32 NODES DENSITY

The impact of the number of LA data flows (priority flows) is clear in both the CSMA/CA part as well as on the LA-MAC part. However, the improvement of average access delay in all cases with the introduction of the LAD period is impressive. Especially in the 32 active nodes scenario, where CSMA/CA reaches its limits and cannot offer a good latency aware service, the LA-MAC offers significantly lower medium access delay in all cases ranging from 38% to 71% reduction.

Lastly, the results of the varying LAD interarrival time (varying LA-MAC airtime) to the average MAC delay are presented in Figure 65, again focusing on the 16 and 32 nodes density scenarios.

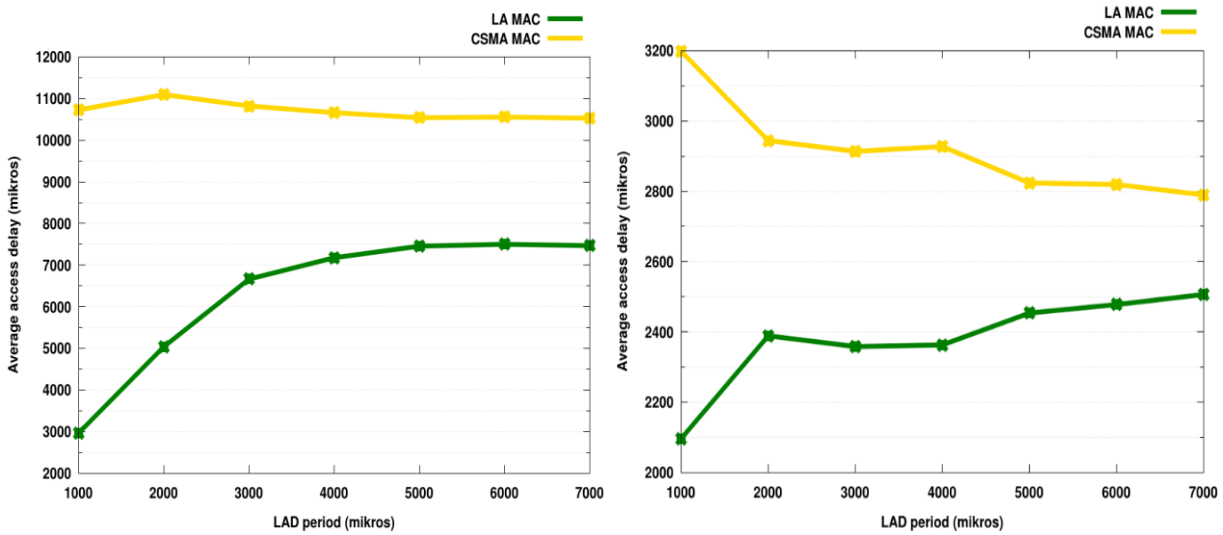


FIGURE 65. IMPACT OF VARYING LA-MAC PERIOD (LAD-IAT) ON OFFERED AVERAGE ACCESS DELAY FOR 16 AND 32 NODES

As it can be clearly seen from Figure 65, increasing the capacity of the LA-MAC (smaller LAD-IAT or LAD period as depicted in the graph) has a very positive impact on the offer medium access latency, far exceeding the performance of the baseline CSMA/CA MAC algorithm. As LAD-IAT increases though, latency also increases but appears to hit a stable performance state above 6 to 7 msec IAT. To make clearer the connection between the LAD-IAT and the actual capacity of LAD in Table 18 shows the calculated LAD airtime vs the LAD IAT:

TABLE 18. LAD AIRTIME VS LAD-IAT

LAD-IAT (ms)	1	2	3	4	5	6	7
LAD airtime %	54,15%	27,07%	18,05%	13,54%	10,83%	9,02%	7,74%

On the other hand, we can observe that when LAD-IAT is very small, the proposed MAC scheme is able to offer very low average MAC access delay compared to the standard CSMA/CA. The presented gains appear to be as big as a 70-75% reduction in MAC delay for the 1 ms LAD-IAT case.

To conclude, the presented initial results show a very promising performance enhancement on priority traffic when the proposed medium access scheme is present, providing clear and significant enhancements in packet latency for high priority packets. On the other hand, the best effort traffic appears not to be significantly impacted; in fact, in some cases, it is even also experiencing a higher level of QoS for KPIs like 44% less maximum MAC delay and

78% less jitter, as a direct impact of the LAD presence. Hence a regulation update could be initiated based on this work, focusing on supporting at least 2 different CCA timings, one for best effort as it is today (at least 18 µsec) and another for real-time priority traffic in the range of 11-16 µsec (12 µsec was used in this evaluation for the TA-CCA). Then schemes like LAD zones and the whole LA-MAC can be built on top that would indeed enable real time access to priority traffic even in unlicensed multi-technology populated frequency bands.

3.5.1.4.3 Next steps

The presented initial results clearly show that the proposed MAC scheme has high potential and further research is needed to examine its performance compared to other possible control parameters of the proposed MAC scheme. To that end, our next steps will involve design of enhancements based on the knowledge gained up until this time, to provide, if possible, even deterministic medium access in unlicensed bands for low throughput high priority flows. The capacity of the LAD enhancement is a major parameter impacting performance, so more research is needed to determine how the LAD capacity should be optimized to benefit the most priority traffic without imposing a high airtime loss in best effort traffic in more complex scenarios. Furthermore, the ability to support multiple packets per LAD might also be examined based on the findings of the capacity control parameter evaluation and its impact in priority traffic offered QoS.

3.5.2 Adaptive Low Latency and reliable unlicensed spectrum access

We present here a different approach for improving latencies and reliability in in-X subnetworks operating in the unlicensed spectrum. While the previously presented solution proposes the introduction of a dedicated time period for deconfliction, the approach presented here works on the adjustment of relevant access parameters for the sake of reducing the latency in channel access according to the experienced radio conditions. Main use cases of interest are in the industrial domain, such as robot control.

3.5.2.1 Adaptive semi-static access mechanism

ETSI has specified the following two channel access strategies in the unlicensed spectrum [73]:

- the semi-static frame-based equipment (FBE) channel access mechanism.
- the dynamic load-based equipment (LBE) channel access mechanisms.

It has been noted that while LBE is helpful in a non-congested setting where the wireless channel is not expected to be busy, FBE is advantageous in a crowded setting where the wireless channel is likely to be busy. Still, in the case of NR-U systems FBE is expected to be perform well in scenarios where there is no other incumbent technology, e.g. WiFi. Due to the fact that the frame intervals can be set to match the boundaries of the slots used by the NR licensed carrier, FBE is particularly advantageous for NR-U networks. According to the NR-U work item, NR-U systems can only use FBE LBT in environments where the absence of Wi-Fi networks is guaranteed. A transmitter in FBE can only sense the channel at specific times. In contrast to LBE, there is no random back-off before attempting to reach the channel, even in the event that the transmitter is unable to do so due to LBT [74]. The fraction of time during which collisions can occur is limited to the beginning of each frame interval because all possible transmitters sense the channel simultaneously, that is, at the beginning of the synchronized frame intervals. Moreover, coordinated interference mitigation and suppression schemes can be used to synchronize

transmissions on the channel, allowing multiple devices belonging to the same operator to transmit simultaneously and boosting spectral efficiency.

Given all the above radio access conditions in the unlicensed spectrum that are related to the nature of in-X subnetworks (i.e. high number of devices looking for fast opportunities), we would like to consider a semi-static (i.e. FBE) type of LBT that can be synchronized with the slot boundary as depicted in Figure 66 below [75]. There are several elements in this implementation that need to be specified in here as follows:

- **Backoff adaptation:** the backoff period speed can be adapted according to the interference state information received from the SNE to the HC [76].
- **Optimal Channel Occupancy Time (OCOT):** having multiple numerology URLLC implementations, an optimal COT can be chosen [77].
- **Centralized or distributed energy detection:** a centralized or federated machine learning can be used to guide an optimal channel selection [78].

We expect that with all three mechanisms will enhance the reliability of the accessing the unlicensed spectrum by avoiding increases to the latency, as explained for each mechanism in detail below.

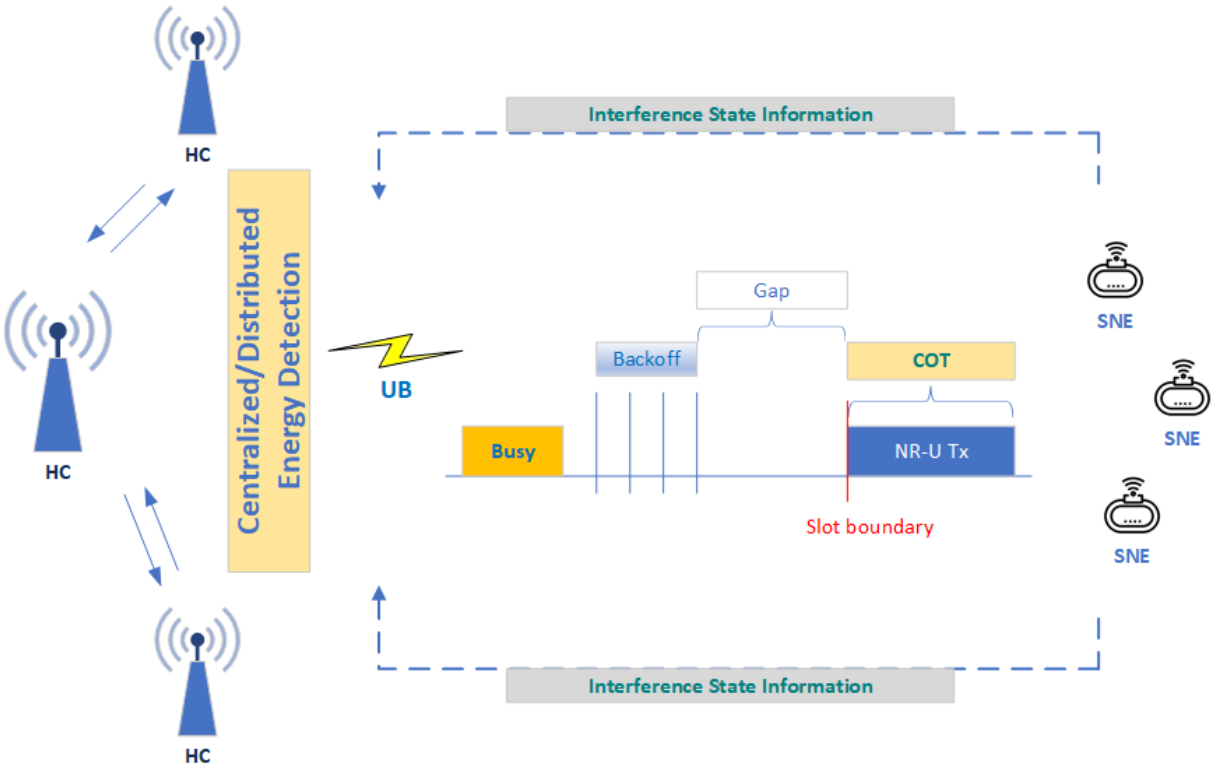


FIGURE 66. ADAPTIVE SEMI-STATIC CHANNEL ACCESS FOR URLLC IN UNLICENSED SPECTRUM.

3.5.2.1.1 Backoff speed adaptation

In unlicensed spectrum bands, a myriad of wireless technologies must coexist and compete for limited bandwidth. This competition often leads to congestion and interference, which can significantly degrade the reliability, as well as the performance of the system. Current LBT access scheme uses a fixed energy detection threshold and backoff

speed, making it unable to adapt to varying interference levels. Against this, an innovative solution has been introduced to enhance the functionality of LBT, which are foundational to modern wireless communication systems [76]. This approach adjusts the back-off mechanism speed based on interference measured by the SNEs and informed to the HC side. Backoff speed adaptation is a technique designed to dynamically adjust the waiting period before a device attempts to access the channel again after a failed attempt, due to the presence of interference. By adjusting the backoff speed to the current interference environment, this method allows for efficient channel access attempts. In scenarios where interference levels are high, the backoff speed can be increased, reducing the probability of collision by delaying the next access attempt. Respectively, in low-interference conditions, the backoff speed can be decreased to accelerate channel access, in order to optimize throughput and reduce latency. As depicted in Figure 67, SNEs could quantify instantaneous interference and inform the HC side with the goal of adjusting their back-off speed.

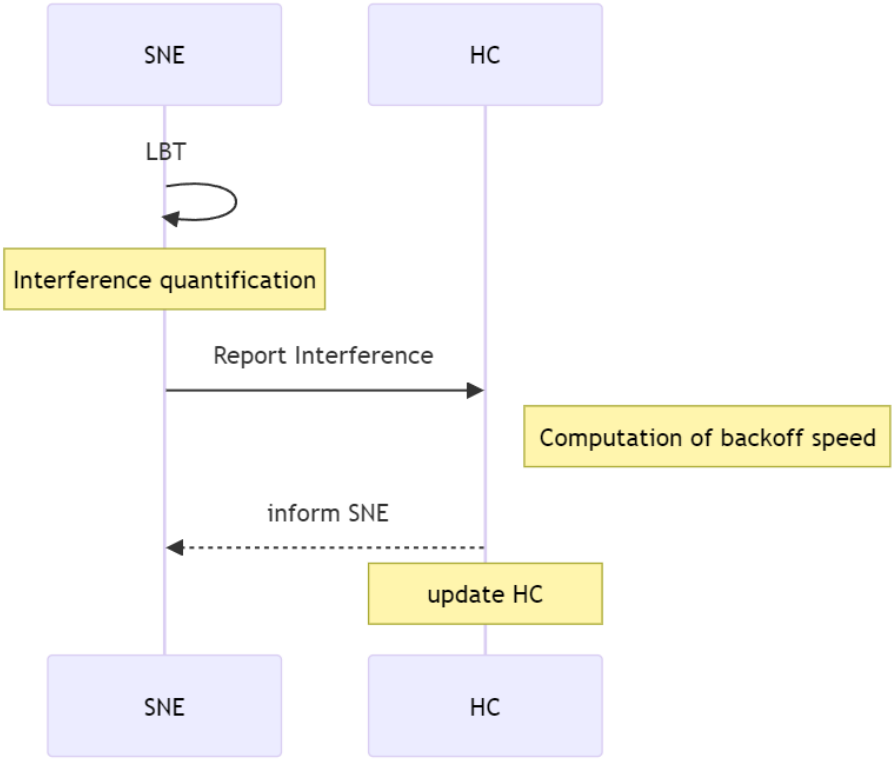


FIGURE 67. BACKOFF ADAPTATION MECHANISM

3.5.2.1.2 Optimal Channel Occupancy Time (OCOT)

In NR-U, the gap duration is the time between the end of the LBT procedure and the next available transmission slot boundary. This gap duration inherently varies, influenced both by the end time of LBT and the synchronization with slot boundaries dictated by the HC's numerology. One of the core issues arising from this scenario is the

difference in gap durations experienced by HC nodes operating with different numerologies. HCs with higher numerologies, have smaller slot durations and as a result, more frequent slot boundaries. This leads to shorter gaps, and subsequently, these HC nodes access the channel more frequently. On the other hand, HC nodes with longer slots, which means longer gaps, have a lower chance to access the channel.

This variation in gap duration creates an imbalance in channel access opportunities across the network. Particularly in dense network environments, like the ones envisioned in in-X subnetworks, HC nodes with shorter slots can dominate the channel, leading to a significant fairness issue. HC nodes with longer slots are at a disadvantage, potentially impacting the Quality of Service (QoS) for end-users connected to these nodes. For the sake of addressing this fairness issue, an algorithm can be implemented that adjusts the maximum channel occupancy time (MCOT) based on slot duration and number of competing nodes in each numerology. The proposed algorithm adjusts MCOT to account for the varying slot duration, and by doing so it aims to enable a fairer channel access. The approach for design of a MCOT adaptation procedure is depicted in Figure 68; it includes a process of development, implementation and evaluation with the goal of ensuring fair channel occupancy.

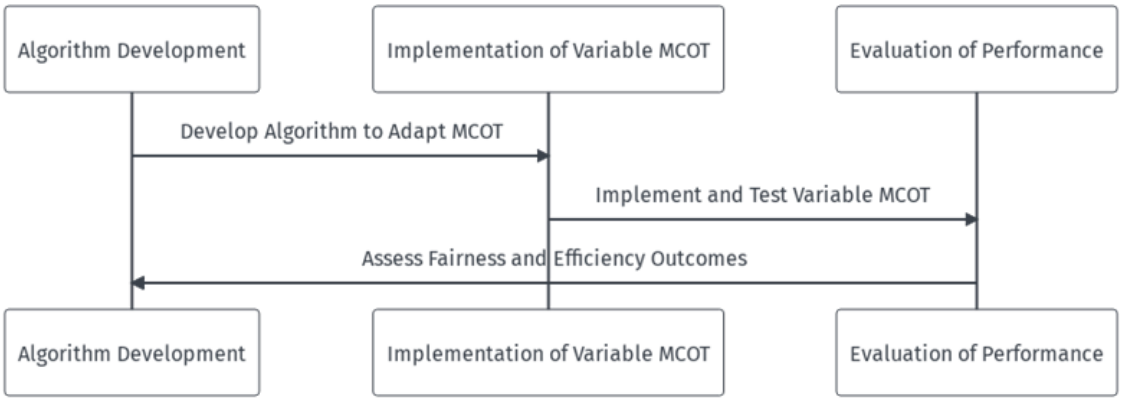


FIGURE 68. MCOT ADAPTATION PROCEDURE.

3.5.2.1.3 Centralized or distributed energy detection

The Energy Detection (ED) threshold impacts the channel access and can affect system performance. For instance, when set to a high value, multiple nodes may perceive the channel as idle and attempt to transmit simultaneously, resulting in transmission failures and a decline in reliability. Conversely, in a low ED threshold setting, nodes may erroneously interpret the channel as occupied, even when no transmission is present, leading to underutilization of the channel and increased latency. In [78], the use of Deep Reinforcement Learning is explored for the optimization of channel access where NR-U and Wi-Fi systems coexist. The focus is on ensuring URLLC and with this goal it is proposed to dynamically adjust the ED threshold in order to improve the reliability of NR-U while securing the fair coexistence with Wi-Fi. The adjustment of ED can be implemented in both centralized and distributed manners as depicted in Figure 69, offering flexibility and adaptability. In a centralized approach, a central controller or coordinator is responsible for managing and regulating the ED process across all nodes within the network. This centralized control allows for coordinated adjustments of ED thresholds based on a global view of the network state and objectives. The learning component of the model is driven by a reward function, which is designed to maximize the successful transmission of packets. On the other hand, the distributed approach

decentralizes the ED adjustment process, allowing individual nodes to autonomously determine their ED thresholds based on local observations and interactions with the environment. If Federated Deep Reinforcement Learning (FDRL) is applied, local models from different nodes are aggregated to update global parameters, enhancing the learning process through shared experiences while maintaining local autonomy.

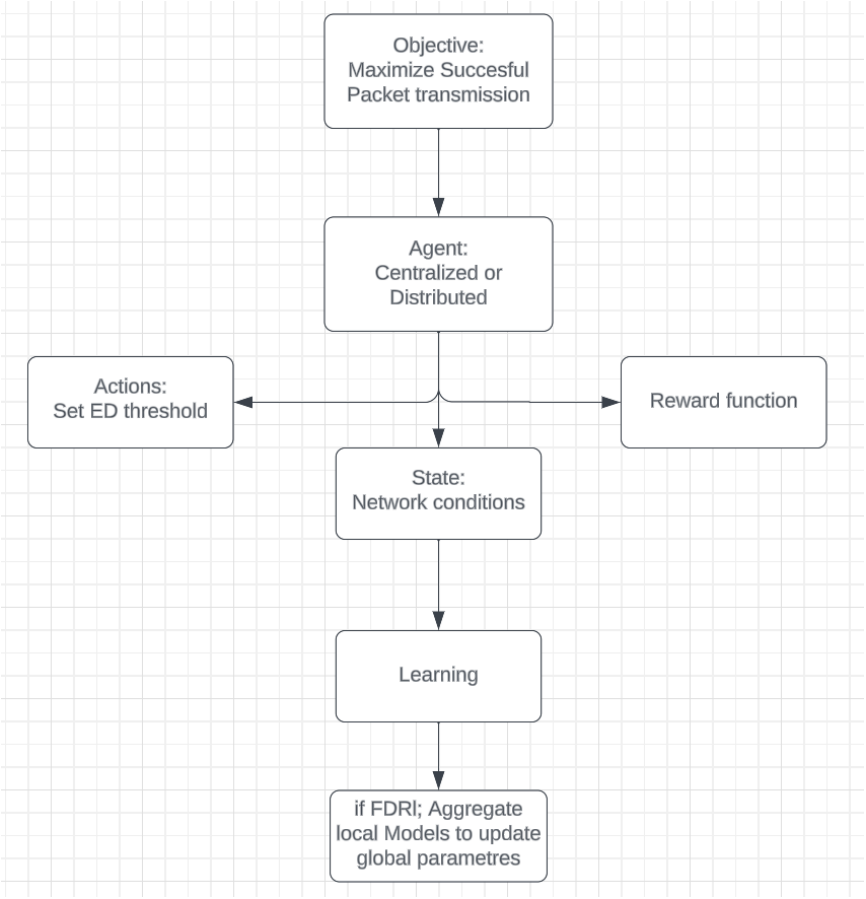


FIGURE 69. CENTRALIZED/DISTRIBUTED ENERGY DETECTION.

3.5.2.2 Next steps

The plan is to develop the proposed mechanisms that will enhance the latency and reliability in the in-X subnetwork operating in the unlicensed band. Our solution will optimize both the channel access and occupancy time that can be enhanced with adaptive backoff and cooperative spectrum sensing using artificial neural networks. The implementation of this solution will also be tailored to 3GPP systems. We expect that that the results will demonstrate the advanced unlicensed spectrum access solution that will be presented in deliverable D3.3.

4 CONCLUSION AND DISCUSSION

In this deliverable, we have presented the initial work carried out in 6G-SHINE for the design of physical layer and medium access control enablers for in-X subnetworks. This deliverable is reflecting the work in progress by the project partners in WP3. For some of the presented methods, initial results have been included. Since the work presented in D3.1 has been carried out while the work in WP2 (use cases definition, scenarios and requirements, and radio channel modelling) had still to be finalized, results here presented have mainly been based on prior assumptions (on scenarios, simulation parameters, channel models) made by the partners based on their research experience.

We have first discussed how the methods presented in the deliverable are addressing the original technology components defined in the 6G-SHINE proposal, as well as their KPIs and potential KVIs, possible connection with the use cases defined in WP2, and standardization perspective.

Section 2 has presented our initial studies on robust and low complexity physical layer enhancements.

Installation of antenna panels in use cases such as industrial and in-classroom is expected to be beneficial for enhancing short-range communication performance, especially in case of dense subnetworks; and pave the way for further enhancements based on scheduling coordination and beam-aware radio resource management.

In order to reduce the device complexity while achieving high performance at high carrier frequencies, we have explored the adoption of self-conjugating metasurfaces, able to automatically adjust to the direction of impinging signal; the proposed blind self-configuring iterative scheme for joint communication and beamforming is able to converge with a very limited number of iterations (4-5), without acquiring channel state information and without signalling overhead.

The usage of sub-THz spectrum can also be promising for in-X subnetworks given their short-range characteristics and the availability of large bands in that region. We have then investigated throughput and energy efficiency of sub-THz transceivers; a hybrid partially connected architecture results to be the most energy efficient solution when compared to a fully digital and to an hybrid fully connected architecture, in case there are no strict limits on the antenna array size.

For the sake of achieving robustness to potential interferers and jamming attacks that can harm critical services in subnetworks, we propose the usage of fast frequency hopping with pseudo-random pattern and packet repetitions. We have derived the minimum number of resources needed for reducing interference between multiple devices. Additionally, we have proposed a bit demapping approach based on approximated log-likelihood ratios to further increase robustness to interference, showing SNR gains of ~4dB with respect to traditional demappers.

Section 3 presented our studies on medium access control and transmission protocol enhancements for subnetworks, meant at improving latency and reliability and to enable the multiplex of diverse services in a same air interface, as envisioned by many of the in-X subnetwork use cases defined in D2.1.

Since latency can be negatively impacted by signalling overhead, we propose a fast retransmission scheme based on proximity grouping, that relies on close vicinity of subnetwork elements for preventive scheduling and retransmissions. Also, configured grant enhancements can be beneficial for scheduling devices having different periodicities. Since subnetworks may need to support uncoordinated access by multiple concurrent elements

while ensuring low latencies and high reliability, we have proposed a novel coded random access scheme based on the assignment of erasure codes with different priorities.

In order to multiplex services with different requirements, we foresee flexible and full duplexing at the HC to be a promising approach as it removes the constraint of a fixed TDD switching point for all devices in a subnetworks. Flexible and full duplexing multiplexing schemes can benefit from the low power conditions in subnetworks, though cross-link interference is a concern to be addressed.

We have then introduced possible solutions for predictive scheduling, for the sake of proactively allocating resources and reduce measurement and signalling overhead. Proactive beam alignment can be based on AI/ML models able of performing pre-emptive inference. In the case of industrial subnetworks, the scheduler can leverage knowledge of the motion of industrial parts (e.g., robot arms) for optimizing transmission procedures. In the case of in-vehicle subnetworks, featuring static devices, data driven AI/ML approaches can leverage accurate characterization and forecasting of in-vehicle traffic conditions (e.g., video streaming from the ADAS sensors) for optimizing the resources and meeting the time-sensitive requirements; an objective function of the schedulers and service-specific constraints have been formulated, along with preliminary results.

Communication reliability within a subnetwork can be enhanced via the introduction of LC devices with relaying capability; specifically, we have shown how the introduction of LCs can reduce the overall power emitted in a subnetwork of above 7 dB while keeping the reliability constraint. Also, network coded cooperation can be used for further improving reliability.

Subnetworks are also expected to operate over unlicensed spectra, subject to possible listen-before-talk constraints that may jeopardize the possibility of achieving low latencies. We have proposed a novel latency-aware access scheme where critical traffic can access the medium in a deterministic manner via predefined deconfliction periods. The proposed solution is able to reduce the medium access delay up to ~75% compared to standard CSMA/CA. Other approaches for reducing latencies in the unlicensed spectrum can be based on the adjustment of relevant access parameters such as back-off time and energy detection.

Deliverable D3.3 will present the consolidated effort in the design of PHY and MAC enablers of subnetworks. The coming activities will leverage the refined scenarios, assumptions and use cases defined in D2.2 (submitted at the same time as D3.1). Synergies among the presented technologies will be explored, along with their capabilities of addressing the defined KPIs and the project objectives. Also, performance analysis will be aligned -when relevant- to the radio channel models developed in WP2, that will be presented in Deliverable D2.3 (with expected submission in June 2024).

REFERENCES

[1] 6GSHINE_Deliverable D2.1 “Initial definition of scenarios, use cases and service requirements for in-X subnetworks”, submitted in August 2023.

[2] 6GSHINE_Deliverable D2.2 “Refined definition of scenarios use cases and service requirements for in-X subnetworks”, submitted in February 2023.

[3] 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.

[4] 3GPP TR 38.817 V16.4.0, Table 7.2.1.4.1-1.

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

[6] 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.

[7] R. Miyamoto and T. Itoh, “Retrodirective arrays for wireless communications,” IEEE Microwave Magazine, vol. 3, no. 1, pp. 71–79, 2002.

[8] E. Sharp and M. Diab, “Van Atta reflector array,” IRE Transactions on Antennas and Propagation, vol. 8, no. 4, pp. 436–438, 1960.

[9] M. Kalaagi and D. Seetharamdoo, “Fano resonance based multiple angle retrodirective metasuface,” in 2020 14th European Conference on Antennas and Propagation (EuCAP), 2020.

[10] Deliverable D2.1 “Towards Tbps Communications in 6G: Use Cases and Gap Analysis”, Hexa-X, 2020.

[11] S. Ju and T. S. Rappaport, "Sub-Terahertz Spatial Statistical MIMO Channel Model for Urban Microcells at 142 GHz," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021, pp. 1-6.

[12] C. Desset, P. Wambacq, Y. Zhang, M. Ingels and A. Bourdoux, "A flexible power model for mm-wave and THz high-throughput communication systems," 2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, UK, 2020, pp. 1-6, doi: 10.1109/PIMRC48278.2020.9217264.

[13] C. Desset, N. Collaert, S. Sinha and G. Gramegna, "InP / CMOS co-integration for energy efficient sub-THz communication systems," 2021 IEEE Globecom Workshops (GC Wkshps), Madrid, Spain, 2021, pp. 1-6.

[14] 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, pp. 1-5.

[15] W. Xu, K. Ma, W. Trappe, and Y. Zhang, (2006). Jamming sensor networks: attack and defense strategies. *IEEE network*, *20*(3), 41-47.

[16] C. Popper, M. Strasser, and S. Capkun,. (2010). Anti-jamming broadcast communication using uncoordinated spread spectrum techniques. IEEE Journal on Selected Areas in Communications, 28(5), 703-715.

[17] R. Adeogun, G. Berardinelli, P. E. Mogensen, I. Rodriguez, &: Razzaghpour (2020). Towards 6G in-X subnetworks with sub-millisecond communication cycles and extreme reliability. *IEEE Access*, *8*, 110172-110188.

[18] N. Golmie, O. Rebala, and N. Chevrollier (2003, October). Bluetooth adaptive frequency hopping and scheduling. In *IEEE Military Communications Conference, 2003. MILCOM 2003.* (Vol. 2, pp. 1138-1142). IEEE.

[19] G. Berardinelli, R. Adeogun (2023). On the required radio resources for ultra-reliable communication in highly interfered scenarios. IEEE Communications Letters, vol.17, no.6, June 2023.

[20] S. Benedetto, Sergio, and G. Montorsi. "Role of recursive convolutional codes in turbo codes." Electronics Letters 31.11 (1995): 858-859.
[21] A. Shokrollahi. "LDPC codes: An introduction." Coding, cryptography and combinatorics. Birkhäuser Basel, 2004.
[22] Digital Baseband Modulation https://nl.mathworks.com/help/comm/ug/digital-baseband-modulation.html, last accessed 20/12/2023.
[23] X. Cao, Y. Liu, and D. Hu. "Simplified LLR algorithm for m-QAM demodulation." the Journal of Engineering 2019.21 (2019): 7370-7375.
[24] W. Rave, "Quantization of Log-Likelihood Ratios to Maximize Mutual Information," in IEEE Signal Processing Letters, vol. 16, no. 4, pp. 283-286, April 2009, doi: 10.1109/LSP.2009.2014094.
[25] M. Arvinte, A. H. Tewfik, and S. Vishwanath. "Deep log-likelihood ratio quantization." 2019 27th European Signal Processing Conference (EUSIPCO). IEEE, 2019.
[26] X.Jiao, W. Liu, M. Mehari, M. Aslam, and I. Moerman, (2020, May). 0penwifi: a free and open-source IEEE802. 11 SDR implementation on SoC. In 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring) (pp. 1-2).
[27] R1-1808256, "Prediction-Based early feedback," TCL Communication, 3GPP RAN1#93.
[28] G. Berardinelli *et al*., "Reliability Analysis of Uplink Grant-Free Transmission Over Shared Resources," in *IEEE Access*, vol. 6, pp. 23602-23611, 2018, doi: 10.1109/ACCESS.2018.2827567.
[29] B. Singh, O. Tirkkonen, Z. Li and M. A. Uusitalo, "Contention-Based Access for Ultra-Reliable Low Latency Uplink Transmissions," in IEEE Wireless Communications Letters, vol. 7, no. 2, pp. 182-185, April 2018, doi: 10.1109/LWC.2017.2763594.
[30] A. Larrañaga, M. C. Lucas-Estañ, I. Martinez and J. Gozalvez, "5G Configured Grant Scheduling for 5G-TSN Integration for the Support of Industry 4.0," *2023 18th Wireless On-Demand Network Systems and Services Conference (WONS)*, Madonna di Campiglio, Italy, 2023, pp. 72-79, doi: 10.23919/WONS57325.2023.10062219.
[31] Open source 5G protocol stack, srsRAN, Web: https://github.com/srsran.
[32] E. Paolini, G. Liva, and M. Chiani. "Coded slotted ALOHA: A graph-based method for uncoordinated multiple access." IEEE Transactions on Information Theory 61.12 (2015): 6815-6832.
[33] S. Zhuo, et al. "Coded slotted ALOHA for erasure channels: Design and throughput analysis." IEEE Transactions on Communications 65.11 (2017): 4817-4830.
[34] A.A. Esswie, and K.I. Pedersen, K. I. (2018). Opportunistic spatial preemptive scheduling for URLLC and eMBB coexistence in multi-user 5G networks. IEEE Access, 6, 38451-38463.
[35] S.R. Khosravirad, G. Berardinelli, K. I. Pedersen, and F. Frederiksen, (2016, May). Enhanced HARQ design for 5G wide area technology. In *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)* (pp. 1-5). IEEE.
[36] Kim, H., Kim, J., & Hong, D. (2020). Dynamic TDD systems for 5G and beyond: A survey of cross-link interference mitigation. *IEEE Communications Surveys & Tutorials*, *22*(4), 2315-2348.
[37] Catania, D., Sarret, M. G., Cattoni, A. F., Frederiksen, F., Berardinelli, G., & Mogensen, P. (2015, May). Flexible UL/DL in small cell TDD systems: A performance study with TCP traffic. In 2015 IEEE 81st Vehicular Technology Conference (VTC Spring) (pp. 1-6). IEEE.
[38] Kolodziej, K. E., Perry, B. T., & Herd, J. S. (2019). In-band full-duplex technology: Techniques and systems survey. IEEE Transactions on Microwave Theory and Techniques, 67(7), 3025-3041.
[39] M. Mokhtari, G. Pocovi, R. Maldonado and K. I. Pedersen, "Modeling and System-Level Performance Evaluation of Sub-Band Full Duplexing for 5G-Advanced," in IEEE Access, vol. 11, pp. 71503-71516, 2023, doi: 10.1109/ACCESS.2023.3295121.

[40] Ji, H., Kim, Y., Muhammad, K., Tarver, C., Tonnemacher, M., Kim, T., ... & Lee, J. (2021). Extending 5G TDD coverage with XDD: Cross division duplex. *IEEE Access*, *9*, 51380-51392.

[41] Study on Evolution of NR Duplex Operation (Release 18), document 3GPP RP-213591, Dec. 2022.

[42] 3GPP Technical Specification TS 38.101-1, "NR; User equipment (UE) radio transmission and reception; Part 1: range 1 standalone", Release-18.

[43] D. Cavalcanti, J. Perez-Ramirez, M. M. Rashid, J. Fang, M. Galeev and K. B. Stanton, "Extending Accurate Time Distribution and Timeliness Capabilities Over the Air to Enable Future Wireless Industrial Automation Systems," in Proceedings of the IEEE, vol. 107, no. 6, pp. 1132-1152, June 2019, doi: 10.1109/JPROC.2019.2903414.

[44] A. Nasrallah et al., "Ultra-Low Latency (ULL) Networks: The IEEE TSN and IETF DetNet Standards and Related 5G ULL Research," in IEEE Communications Surveys & Tutorials, vol. 21, no. 1, pp. 88-145, Firstquarter 2019, doi: 10.1109/COMST.2018.2869350.

[45] 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.

[46] D. Ginthör, J. von Hoyningen-Huene, R. Guillaume and H. Schotten, "Analysis of Multi-user Scheduling in a TSN-enabled 5G System for Industrial Applications," 2019 IEEE International Conference on Industrial Internet (ICII), 2019, pp. 190-199.

[47] J. Li, X. Zhang, J. Zhang, J. Wu, Q. Sun and Y. Xie, "Deep Reinforcement Learning-Based Mobility-Aware Robust Proactive Resource Allocation in Heterogeneous Networks," in IEEE Transactions on Cognitive Communications and Networking, vol. 6, no. 1, pp. 408-421, March 2020, doi: 10.1109/TCCN.2019.2954396.

[48] A. G. Marino, F. Fons, Z. Haigang and J. M. M. Arostegui, "Traffic Shaping Engine for Time Sensitive Networking Integration within In-Vehicle Networks," 2021 IEEE Vehicular Networking Conference (VNC), Ulm, Germany, 2021, pp. 182-189, doi: 10.1109/VNC52810.2021.9644668.

[49] Z. H. Li, S. Q. Yang, J. H. Yu, et al.," Predictive scheme for mixed transmission in time-sensitive networking," ZTE Communications, vol. 20, no. 4, pp. 78–88, Dec. 2022.

[50] D. Cavalcanti, J. Perez-Ramirez, M. M. Rashid, J. Fang, M. Galeev and K. B. Stanton, "Extending Accurate Time Distribution and Timeliness Capabilities Over the Air to Enable Future Wireless Industrial Automation Systems," in Proceedings of the IEEE, vol. 107, no. 6, pp. 1132-1152, June 2019, doi: 10.1109/JPROC.2019.2903414.

[51] M.C. Lucas-Estañ, B. Coll-Perales, M. I. Khan, S. S. Avedisov, O. Altintas, J. Gozalvez, M. Sepulcre, "Support of Teleoperated Driving with 5G Networks", Proceedings of the 2023 IEEE 98th Vehicular Technology Conference (VTC2023-Fall), 10-13 October 2023, Hong Kong.

[52] H.B. Salameh, H. Al-Obiedollah, R. Mahasees, Y. Jararweh, "Opportunistic non-contiguous OFDMA scheduling framework for future B5G/6G cellular networks", Simulation Modelling Practice and Theory, Volume 119, 2022. https://doi.org/10.1016/j.simpat.2022.102563.

[53] G. P. Sharma et al., "Toward Deterministic Communications in 6G Networks: State of the Art, Open Challenges and the Way Forward," in IEEE Access, vol. 11, pp. 106898-106923, 2023, doi: 10.1109/ACCESS.2023.3316605.

[54] T. Huck, A. Achtzehn, A. Deberling, E. Goerres, and K. Wehefritz, "The next step in E/E architectures", Bosch's WhitePaper, 2024.

[55] S. R. Khosravirad, H. Viswanathan, and W. Yu, "Exploiting Diversity for Ultra-Reliable and Low-Latency Wireless Control," IEEE Trans. Wireless Commun., vol. 20, no. 1, pp. 316-331, Jan. 2021.

[56] S. Karmakar and M. K. Varanasi, “The Diversity-Multiplexing Tradeoff of the Dynamic Decode-and-Forward Protocol on a MIMO Half-Duplex Relay Channel,” IEEE Transactions on Information Theory, vol. 57, no. 10, pp. 6569-6590, Oct. 2011.

[57] V. N. Swamy, S. Suri, P. Rigge, M. Weiner, G. Ranade, A. Sahai, and B. Nikolic, “Cooperative communication for high-reliability low-latency wireless control,” in IEEE Intl. Conf. on Commun. (ICC’15), June 2015, pp. 4380–4386.

[58] Y. Hu, M. C. Gursoy, and A. Schmeink, “Relaying-Enabled Ultra- Reliable Low-Latency Communications in 5G,” IEEE Network, vol. 32, no. 2, pp. 62-68, March-April 2018.

[59] J. Cheng and C. Shen, “Relay-Assisted Uplink Transmission Design of URLLC Packets,” IEEE Internet of Things Journal, vol. 9, no. 19, pp. 18839-18853, Oct. 2022.

[60] A. Ranjha, G. Kaddoum, and K. 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, vol. 18, no. 7, pp. 4954-4965, July 2022.

[61] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” http://cvxr.com/cvx, Mar. 2014.

[62] H.R. Hashempour, G. Berardinelli, and R. Adeogun, “A Power Efficient Cooperative Communication Protocol for 6G in-Factory Subnetworks,” to be submitted to EuCNC 2024.

[63] 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.

[64] S. Tedik Başaran, G. Karabulut Kurt, M. Uysal and İ. Altunbaş, "A tutorial on network coded cooperation," in IEEE Communications Surveys & Tutorials, vol. 18, no. 4, pp. 2970-2990, Fourthquarter 2016, doi: 10.1109/COMST.2016.2562027

[65] A. Lee and M. Gerla, "Performance evaluation of wireless network coding in TDMA networks," MILCOM 2012 - 2012 IEEE Military Communications Conference, Orlando, FL, USA, 2012, pp. 1-6, doi: 10.1109/MILCOM.2012.6415880.

[66] H. Nie, S. Li and Y. Liu, "Dynamic Multi-link Transmission Technology with High Reliability based on Link State Awareness and Network Coding," 2021 International Wireless Communications and Mobile Computing (IWCMC), Harbin City, China, 2021, pp. 1884-1889, doi: 10.1109/IWCMC51323.2021.9498620.

[67] ETSI Standard EN 300 328 V2.2.2, HTTPS://WWW.ETSI.ORG/DELIVER/ETSI_EN/300300_300399/300328/02.02.02_60/EN_300328V020202P.PDF

[68] IEEE 802.15 Working group , https://www.ieee802.org/15/

[69] IEEE 802.11 Working group, https://www.ieee802.org/11/

[70] 3GPP TS 37.213 Standard, Physical layer procedures for shared spectrum channel access, https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3435

[71] IEEE 802.11n wireless standard, https://standards.ieee.org/ieee/802.11n/3952/

[72] Ns-3 network simulator, https://www.nsnam.org/

[73] S. Park, H. Ryu, Y. Kim and J-K. Han, Performance of Channel Access Mechanisms for 5G Industrial-IoT over Unlicensed Bands, 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall), Sept. 2021.

[74] T-K. Ley, 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), Dec. 2021.

[75] M. Zajac and S. Szott, Resolving 5G NR-U Contention for Gap-Based Channel Access in Shared Sub-7 GHz Bands, IEEE Access, vol. 10, pp. 4031-4047, Jan. 2022.

[76] H. Zhou, Y. Deng and A. Nallanathan, Novel Listen-Before-Talk Access Scheme with Adaptive Backoff Procedure for Uplink Centric Broadband Communication, IEEE Internet of Things Journal, vol. 10, no. 22, Nov. 23.

[77] A. B. K. M. and A. M. Baswade, "Numerology-Aware MCOT Selection for Improving Fairness in 5G NR-U Networks," in IEEE Communications Letters, vol. 27, no. 3, pp. 1045-1049, March 2023, doi: 10.1109/LCOMM.2023.3241401.

[78] Y. Liu, H. Zhou, Y. Deng and A. Nallanathan, Channel Access Optimization in Unlicensed Spectrum for Downlink URLLC: Centralized and Federated DRL Approaches, IEEE Journal on Selected Areas in Communications, pp. 2208-2222.