Horizon Europe Framework Programme

Towards an AI-native, user-centric air interface for 6G networks

---

D5.3 KPIs/KVIs, testing methodologies and benchmarking

Contractual Delivery Date:	31 December 2024
Actual Delivery Date:	28 February 2025
Editor: *(name, organization)*	Carles Navarro Manchón (KEY)
Deliverable nature:	R (Document, report)
Dissemination level:	PU (Public)
Version:	1.0
Keywords: KPIs, Validation, Testing, Benchmarking, AI for Air Interface	

ABSTRACT

This deliverable presents the validation results obtained from performance benchmarking of the technological enablers developed in WP2, WP3, and WP4 of the CENTRIC project. For each enabler, a set of KPIs with corresponding targets are defined and evaluated. In addition, a set of baselines, including AI and non-AI methods, are as well evaluated for benchmarking purposes.

This project has received funding from the European Union's Smart Networks and Services Joint Undertaking (SNS JU) under grant agreement No. 101096379.

Disclaimer

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101096379. This publication reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Impressum

Full project title: Towards an AI-native, user-centric air interface for 6G networks

Short project title: CENTRIC

Number and title of the work package: WP5 – AI-AI Testing and Validation

Number and title of task: T5.1 – AI and Testing, T5.2 – CENTRIC Performance Benchmarking

Document title: KPIs/KVIs, testing methodologies and benchmarking

Editor: Carles Navarro Manchón, Keysight Technologies

Work-package leader: Carles Navarro Manchón, Keysight Technologies

Copyright notice

Executive summary

This deliverable compiles the validation of KPIs of most of the AI-based technological enablers that have been developed in the context of the CENTRIC project. In total, 20 enablers have been validated, and 37 KPIs have been quantified. In addition, baselines for performance reference, including classical and state-of-art AI techniques, have been included for each of the enablers.

The report describes shortly the methodology followed for validation, and then presents an overview of the whole ensemble of considered technologies, their respective KPIs, and the corresponding performance targets.

Subsequently, we delve in detail into the results obtained with each of the methods considered, organized by WP. Snapshots of the results are provided in order to better understand the behaviour of the considered techniques.

While CENTRIC is a low TRL project, and integration of the enablers in higher TRL prototypes or trials are needed to achieve definitive conclusions, the results presented in this report are encouraging with respect to the potential of AI techniques to revolutionize the way the air interface of 6G will be designed. Based on this results, we can confidently conclude that the adoption of AI in the air interface of mobile communication systems has just started, and we foresee a promising future for them in 6G systems and beyond.

List of authors

Company	Author	Contribution
Keysight Technologies	Carles Navarro Manchón	Editor
Keysight Technologies	Issam Maaz	Co-editor
Kings College London	Matteo Zecchin	Enabler developer and results contributor
Kings College London	Osvaldo Simeone	Enabler developer and results contributor
Synthara	Nicolò De Rita	Enabler developer and results contributor
InterDigital	Ahmet Serdar Tan	Enabler developer and results contributor
InterDigital	Anouar Yatribi	Enabler developer and results contributor
InterDigital	Yasser Mestrah	Enabler developer and results contributor
InterDigital	Deepa Gurmukhdas Jagyasi	Enabler developer and results contributor
InterDigital	Muhammad Awais Jadoon	Enabler developer and results contributor
Nokia	Aakash Saini	Enabler developer and results contributor
Nokia	Amir Tehrani	Enabler developer and results contributor
Sequans	Alexander Chiskis	Enabler developer and results contributor
Sequans	Rami Verbin	Enabler developer and results contributor
Sequans	Efstathios Katranaras	Enabler developer and results contributor
Aalborg University	Ramoni Adeogun	Enabler developer and results contributor

Aalborg University	Xiyu Wang	Enabler developer and results contributor
Aalborg University	Saeed Hakimi	Enabler developer and results contributor
NVIDIA	Sebastian Cammerer	Enabler developer and results contributor
Nokia Bell Labs	Alvaro Valcarce	Enabler developer and results contributor
Nokia Bell Labs	Bryan Liu	Enabler developer and results contributor
CNR	Emma Chiaramello	Enabler developer and results contributor
CNR	Marta Parazzini	Enabler developer and results contributor
Oulu University	Abanoub Pipaoy	Enabler developer and results contributor
Oulu University	Septimia Sarbu	

Table of Contents

List of figures and tables

List of figures:

List of tables:

Abbreviations

Abbreviation	Definition	Abbreviation	Definition	Abbreviation	Definition
ICL	In context learning	CSI	Channel State Information	TRP	transmission-reception point
MAML	model-agnostic meta-learning	SGCS	squared generalized cosine similarity	PPO	proximal policy optimization
MMSE	Minimized mean squared error	RU	resource utilization	UE	User equipment
LMMSE	Linear Minimum Mean Squared Error	UCI	uplink control information	mmwave	millimeter-wave
SNR	Signal to noise ratio	MIMO	Multiple input multiple output	PPO	Proximal policy optimization
MSE	Mean squared error	LS	least-squares	AoD	angle of departure
DNN	deep neural networks	UCI	Uplink control information	QAM	quadrature amplitude modulation
SER	Symbol error rate	AoD	angle of departure	BP	best permutation
CDF	cumulative density function	FER	frame error rate	AWGN	additive white Gaussian noise
PDCC	Physical Downlink Control Channel	TS-JEPA	time-series joint embedding predictive architecture	BCE	binary cross-entropy
RNN	recurrent neural network				

1 Introduction

The ambitions of 6G networks are to support low latency in the order of microseconds, massive machine-type connectivity, ultra-high-speed connectivity, and energy efficiency compared to 5G. Furthermore, it will enable immersive communications such as extended reality (XR), high-fidelity holograms, etc. Artificial intelligence will be part of 6G, which will help 6G networks enhance their performance and efficiency. The enhancement brought by AI needs to be validated.

In D5.1, "Early results on KPIs/KVIs testing methodologies and benchmarking," we have identified the KPIs and KVIs relevant to assessing the impact of AI enabled air interface for 6G networks. In this deliverable we present a comprehensive approach to validate Keys performance indicators (KPI) for artificial intelligence (AI) native air interface for 6G networks. For each work package, we provide the enabler technologies that we followed to validate the KPIs. Also, we present the validation assumptions that have been followed along with the validation results. We have provided a benchmarking between the baseline and AI-based algorithms/technologies.

While it was desired to also quantify the KVIs that were identified in D5.1, it was assessed that the technology maturity level of the technologies was not sufficient to approach such quantification. Quantification of KVIs requires integration of the enablers within a full model of the system. Achieving such level of integration was never the goal of CENTRIC, and would have deterred from achieving the impressive amount of enablers developed and validated that the project has produced. Rather than attempting to quantify the KVIs without sufficient basis to do so, which would have been a futile attempt, we refer the reader to CENTRIC deliverable D5.1 where proxy KPIs that can help gauging the contribution of a given enabler to the CENTRIC KVIs were identified.

The rest of the deliverable is structured as follows:

- Section 2 describes the validation methods used in this work and the process used to select KPIs. In addition, an overview of all the enablers for which validation results are provided in this deliverable is provided.
- Section 3 constitutes the main body of this report, and describes for each enabler the KPIs that have been assessed, their target values, and the values attained by the methods in the validation process. In addition, baselines for comparison of the method's performance are identified and assessed, including both traditional and AI-based methods.
- Section 4 draws conclusions from the work.

2 Benchmarking Methodology

2.1 Validation methods and selection of KPIs

The vast majority of enablers reported in this deliverable have been validated using computer simulations, being link-level simulations or system-level simulations. A few of the enablers have been accompanied by mathematical analysis, and this is remarked when applicable. In addition, there are also activities in CENTRIC to validate a small subset of the enablers by means of a proof-of-concept implementation in the lab. However, we do not discuss here validation by experimental means, as there will be a deliverable wholly devoted to it (D5.4) published by end of April 2025, where all details of the experimental validation will be provided.

Regarding the selection of KPIs, careful comparison of this deliverable with the previous D5.1 will reveal that not all the KPIs defined in D5.1 have been validated. Note that D5.1 KPIs were selected as being KPIs of interest for each particular enabler. However, many of those KPIs necessitate integration in a more complete system to be properly evaluated, which is not feasible given the low TRL that is the target of CENTRIC. The KPIs that have been chosen to be validated are those that the developers of the enablers themselves have selected as being feasible to validate with the degree of maturity of each enabler. We believe that this is a sound approach, as attempting to quantify certain KPIs in isolation of the rest of system components may lead to useless or, even worse, misleading results.

2.2 Overview of enablers, KPIs and targets

Table 1 provides an overview of all the enablers whose validation results are reported in this deliverable, organized by WP. For each of the enablers, the validated KPIs are described, as well as their target values.

We emphasize that, while the table is large, containing overall 20 different technological enablers developed in CENTRIC, the list is by no means exhaustive. While it covers most of the work done in the project, there are some enablers in WPs 2, 3, and 4 which are still being developed or are in the process of validation at the time of writing of this deliverable, as the project still has some months before reaching its conclusion.

Table 1: Overview of all enablers validated in this deliverable, their KPIs and the associated targets

WP2 Enablers			
Enabler (Partner)	**Validation approach**	**KPIs**	**Targets**
In-context learning (KCL)	Link-level simulations	KPI#1: Mean-squared error vs varying front-haul capacity	10% reduction with respect to B#1 and B#2
		KPI#2: Mean-squared error vs varying SNR	10% reduction with respect to B#1
Nullhop: Neural Receiver Acceleration (Synthara)	RTL simulations; synthesis and	KPI#1: Inference speed	1ms
		KPI#2: Inference complexity	30% reduction with respect to B#2
		KPI#3: Block-error rate	Improve with respect to B#1

	power estimation		
WP3 Enablers			
Enabler (Partner)	**Validation approach**	**KPIs**	**Targets**
AIML-enabled CSI Compression (NOK)	System-level computer simulations	KPI#1: squared generalized cosine similarity	10% improvement against baselines
		KPI#2: mean and cell-edge user throughput	10% improvement against baselines
Adaptive quantization for AIML-enabled CSI compression	Link-level simulations	KPI#1: SGCS	50% improvement over B#1, loss smaller than 20% over B#2.
		KPI#2: Signalling overhead	50% reduction with respect to B#1.
		KPI#3: Model generalization	Generalization over different payload sizes
CSI Prediction Enhancements (InterDigital)	System-level simulations	KPI#1: Throughput (mean and 5th percentile)	5% increase for mean throughput, 15% increase for 5th percentile.
Linear Coded Multi-TRP CSI Compression (InterDigital)	Computer simulations and mathematical analysis	KPI#1: squared generalized cosine similarity	Less than 0.2 loss with respect to B#1.
UE channel learning and array dimensionality reduction (Sequans)	Link-level simulations	KPI#1: SNR	<1 dB loss with respect to no reduction
		KPI#2: reconstruction accuracy	Within 2dB of B#2
Joint Sensing and Communications (AAU)	Link- and system-level simulations	KPI#1: Communication rate	4 b/s/Hz
		KPI#2: Latency	N/A
ML-enabled Symbol Modulation (InterDigital)	Link-level simulations	KPI#1: Bit/Symbol-error rate	Lower than B#1 under nonlinear impairments
		KPI#2: Model generalization ability	Generalization achieved for multiple modulation orders
Multi-user MIMO Neural Receiver (NVIDIA)	Link-level simulations and proof-of-concept with hardware-in-the-loop	KPI#1: BLER for fixed computational complexity	BLER close to B#2 with lower computational complexity.
		KPI#2: Inference latency	<1ms inference latency on NVIDIA A100 GPU
WP4 Enablers			
DCI Compression (NNF)	System-level and link-level simulations	KPI#1: PDCCH reliability	0.2 dB decoding gain
		KPI#2: Lossless compression ratio	10% improvement versus B#1
Task-oriented Cognitive Wireless Scheduling: Collaborative Navigation (OUL)	Monte Carlo simulations	KPI#1: Parallel task execution (evaluated with Jain's fairness index [1])	As close to 1 as possible
		KPI#2: Latency to destination (in time-steps)	As low as possible

Task-oriented Cognitive Wireless Scheduling: semantic communication and control co-design (OUL)	Link-level simulations	KPI#1: Control performance – Normalized score	In range [0.75—1]
		KPI#2: Communication efficiency – Communication bits and Transmission latency	Minimize while maintaining target for KPI#1
Emerging multiple-access protocols for specialized services (AAU)	System-level simulations in in-factory scenarios	KPI#1: buffer flush rate	0.9 median value
		KPI#2: signalling overhead	50% reduction with respect to B#1
Federated Multi-Agent DRL for Radio Resource Management (AAU)	System-level simulations	KPI#1: User spectral efficiency	6 b/s/Hz at 1st percentile
		KPI#2: average RL reward	12.5
ML-based Sub-band Selection (AAU)	System-level simulations	KPI#1: rate-conforming subnetworks	Median values of 9 for low-rate and 3 for high-rate subnetworks
		KPI#2: Training loss L	L = 1
Joint Sub-band Allocation and Power Control for Outdated CSI Scenarios (AAU)	System-level simulations	KPI#1: Spectral efficiency	average SE = 8.5@median per-user SE = 5.8 @10^{-3}
		KPI#2: Training loss	10^{-3}
Learning-based HARQ (InterDigital)	Mathematical analysis	KPI#1: Retransmission latency	50% decrease versus B#1 and B#2.
		KPI#2: Retransmission overhead	0.2 retransmissions
Probabilistic Time Series Conformal Risk Prediction (KCL)	Link-level simulations	KPI#1: Prediction efficiency	10% improvement against B#1
		KPI#2: Delay, decoding probability, throughput, and energy efficiency	10% improvement over B#2
EMF Reduction via AI-enabled Cell-free Networking (CNR)	System-level simulations	KPI#1: whole body SAR (SARwb)	Not exceed 0.08 W/kg

3 Benchmarking of CENTRIC Enablers

3.1 Benchmarking of WP2 Enablers

We begin by

3.1.1 In-context learning

In-context learning is a learning technique from the family of meta-learning methods. It can be seen as a way to adapt the outcomes of a given model to a particular task without a need for fine-tuning the model parameters. Instead, a set of example input-output pairs are provided to the model as context for the task – hence the name, *in-context learning*. In CENTRIC's WP2, n-context learning techniques have been theoretically developed and evaluated in an example application context corresponding to that of MIMO channel equalization.The outcome and conditions of the validation of this enabler are summarized in Table 2, with the details provided below.

Table 2: Validation Summary for In-context learning

In-context learning (KCL)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Link-level computer Simulation	B#1: (Traditional) Linear Minimum Mean Squared Error (L-MMSE)	KPI#1: Mean-squared error vs varying front-haul capacity	10% reduction with respect to B#1 and B#2	Target exceeded (>50% reduction)
	B#2: (AI) Model Agnostic Meta-Learning (MAML)	KPI#2: Mean-squared error vs varying SNR	10% reduction with respect to B#1	Target exceeded from most SNR values

3.1.1.1 Validation assumptions and measured KPIs

The performance of the in-context learning (ICL)-based MIMO equalizer has been validated via link-level computer simulations using simple single-user MIMO channels as described in [2] as well as under cell-free multi-user MIMO systems [3]. In this report we mainly focus on the latter results A deep receiver that leverages in-context learning and sequence models to adapt to time-varying channel conditions has been implemented and tested. In addition to the ICL-based equalizer, two baselines have been implemented:

- **B#1:** The first baseline is a traditional L-MMSE based MIMO equalizer, such as described in [4].
- **B#2:** The second baseline is also a deep receiver that exploits a concept of meta-learning called model-agnostic meta-learning (MAML) [5].

For all methods, the performance of the equalized channel is evaluated by calculating the mean square error (MSE) between the transmitted and equalized signals. This is measured under two different conditions, leading to the definition of two KPIs:

- **KPI#1:** In the first scenario, the MSE is evaluated against the fronthaul capacity of the assumed cell-free massive MIMO system. With low fronthaul capacity, quantization errors introduced by the fronthaul are expected to impair all methods' performance.
- **KPI#2:** In the second scenario, the MSE is evaluated against different signal-to-noise ratio (SNR) conditions, with the goal of making the proposed equalizer be robust to the relevant SNR range.

For both KPIs, the target is to reduce MSE by 10% with respect to the baselines.

3.1.1.2 Validation results

The validation results for the proposed MIMO equalizer have been reported in [2] and [3], and we summarize the main outcome here.

In the case of KPI#1, the obtained MSE results are presented in Figure 1, where the MSE achieved by the proposed equalizer is presented against that of B#1 (LMMSE) and B#2 (MAML). In addition, the results of B#1 with unconstrained fronthaul are also presented. Two situations are evaluated: one in which the pilot signals used for channel estimation are orthogonal across the different users, and another in which pilot contamination is present. It can be observed that the proposed equalizer exhibits particularly good performance in the latter situation, achieving an MSE that is twice as small as that of LMMSE equalization and several orders of magnitude smaller than the MSE of the MAML-based receiver.

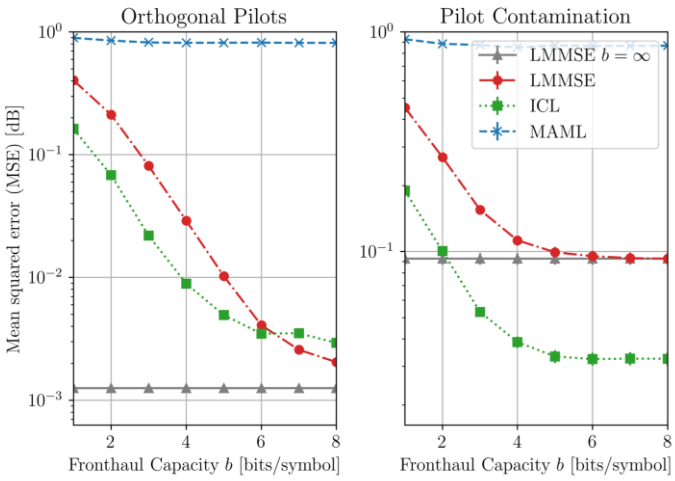


Figure 1: Validation results of In-Context Learning for KPI#1 [3]

The validation of KPI#2 focuses on evaluating the robustness of the proposed ICL-based equalizer against different SNRs, and particularly under pilot contamination conditions. A set of results are depicted in Figure 2, where the MSE of the proposed enabler and that of B#1 are assessed under different levels of pilot contamination. It is worth noting that the plot "ICL without LS tokens" represents the same ICL-based receiver which has not utilized context about large-scale fading conditions of the channel. The results show a significant advantage of the ICL-based equalizer in all scenarios with pilot contamination (Pilot Reuse > 0) for all SNRs above -5 dB. They also show that incorporating large-scale fading information into the receiver is crucial for it to outperform traditional approaches.

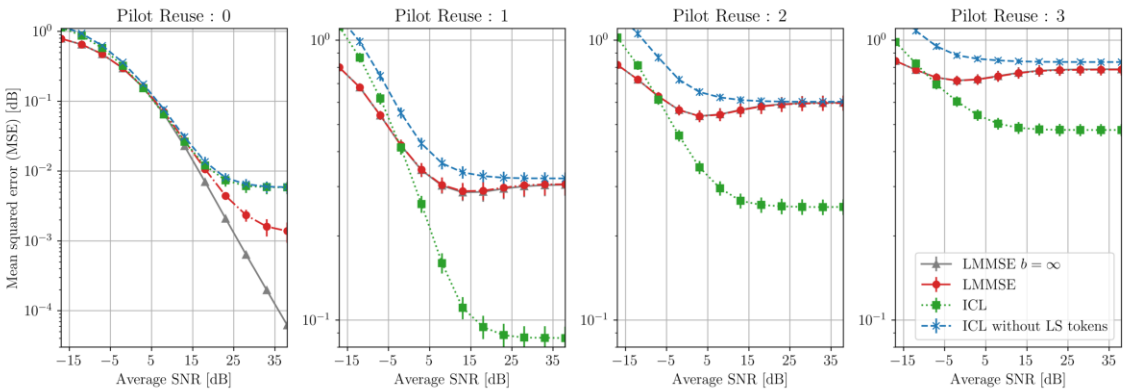


Figure 2: Validation results of In-Context Learning for KPI#2 [3]

3.1.2 Nullhop: Neural Receiver Acceleration

The second enabler that will be evaluated in WP2 is the work done by Synthara?? to accelerate the neural receiver publicly available in NVIDIA's Sionna platform [6].

Table 3: Validation Summary for Nullhop: Neural Receiver Acceleration

Nullhop: Neural Receiver Acceleration (Synthara)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
RTL simulations; synthesis and power estimation	B#1: Receiver based on LS channel estimation [6]	KPI#1: Inference speed	1ms	70ms
	B#2: Original neural receiver model [6]	KPI#2: Inference complexity	30% reduction with respect to B#2	60% sparsity increase
		KPI#3: Block-error rate	Improve with respect to B#1	Achieved

3.1.2.1 Validation assumptions

The receiver's workload has been mapped and deployed onto Nullhop, which is and AI model and convolutional neural network accelerator. The receiver has then been simulated with EDA tools, allowing RTL simulations that reproduce the operations carried out in silicon with high precision. With this, accurate estimates of the latency and throughput achieved by the model can be obtained. Similarly, this setup allows to carry out estimation of the power consumption that the model would incur. Using the aforementioned tools, the receiver in [6] has then been quantized and sparsified in order to reduce its complexity and the inference latency.

Two baselines are used for evaluation of the work:

- **B#1:** the first baseline, used to compare the block error-rate (BLER) of the accelerated receiver, is the traditional (not AI based) receiver available in [6]. It is a receiver based on least-squares (LS) channel estimation, and using minimum mean squared error (MMSE) equalization.

- **B#2:** the second baseline is that of the original neural receiver model in [6]. This baseline is mainly used to evaluate the inference complexity reductions achieved by the accelerated version of the neural receiver model.

As evaluation metrics, two KPIs are used:

- **KPI#1:** the first KPI used to evaluate the accelerated model is the inference speed. The target is to reach 1ms inference time, which would allow running the receiver model in real-time.
- **KPI#2:** the second KPI is the inference complexity, with the target being a 30% decrease with respect to B#2.
- **KPI#3:** the third KPI is the block-error rate (BLER) achieved by the accelerated receiver. The target for it is to improve with respect to B#1.

3.1.2.2 Validation results

At the time of publication of this deliverable, work is still ongoing in the acceleration of the model. Hence, the results reported here will be succinct and should be understood as a temporary snapshot in the development process.

For KPI#1, the current inference speed achieved is 70ms, whereas for KPI#2 a sparsity increase of 60% with respect to the original model has been achieved. These results are still far from the target, in particular for KPI#1, but work is ongoing on further closing the gap.

In term of KPI#3, the BLER achieved by the accelerated receiver and the baselines is depicted in

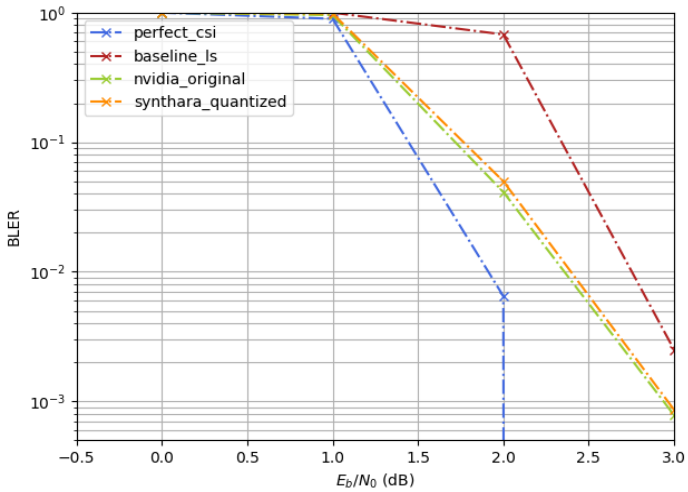


Figure 3: Validation results of Nullhop: Neural Receiver Acceleration for KPI#3

Final results will be reported in the corresponding WP2 deliverable towards the end of the project.

3.2 Benchmarking of WP3 Enablers

3.2.1 AIML-enabled CSI compression

The problem of channel state information (CSI) compression, that is, how to efficiently encode information of on the channel experienced by one user in order to feed the information back to the base station for precoding purposes, has achieved increasing relevance as the dimensions of antenna arrays grow at both network and terminal side. There is a fundamental trade-off between the accuracy of the encoded CSI and the amount of bits it is required to transmit it. In this enabler, deep neural networks (DNNs) are used to perform encoding and decoding of CSI, with the goal of achieving a better compromise between CSI accuracy and the overhead that its transmission entails.

The outcome and conditions of the validation of this enabler are summarized in Table 4, with the details provided below.

Table 4: Validation Summary for AIML-enabled CSI compression

AIML-enabled CSI Compression (NOK)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
System-level computer simulations	B#1: Enhanced Type II Codebook [7]	KPI#1: squared generalized cosine similarity	10% improvement against baselines	12% improvement
	B#3: Transformer based encoder [8]	KPI#2: mean and cell-edge user throughput	10% improvement against baselines	16% and 17% improvement for mean and cell-edge user, respectively.

3.2.1.1 Validation assumptions

The enabler validation has been carried out using system level simulations considering 7 tri-sectorial sites and using 3GPPP standardized channel models. In addition, it is assumed that the encoded CSI feedback channel from user to their base station is ideal. This implies the wireless channel does not introduce further errors in the encoded feedback, but the quantization error including by the CSI encoding process is still present.

In addition to the proposed enabler, two other methods are used as reference baselines, one based on the current standard CSI codebook and one being another DNN based method:

- **B#1:** the first baseline consists of utilizing the 3GPP Release 16 Enhanced Type II codebook [7] to perform the CSI encoding, as currently utilized in 5G NR systems.
- **B#2:** the second baseline consists for a DNN model based on transformers, and is described in [8].

The performances of the different methods are validated using two different KPIs:

- **KPI#1:** the first KPI is the squared generalized cosine similarity, which is a typical measure of the similarity between channel response vectors.

- **KPI#2:** as second KPI, different statistics of the throughput are used, in particular, the mean throughput and the cell-edge user throughput, that is, the 5th percentile of the throughput distribution.

For both KPIs, the target is to achieve a 10% improvement with respect to the baseline.

3.2.1.2 Validation results

A complete description of the validation results is currently being worked on and will be made available in a publication soon. For the moment, the preliminary results show the following improvements with respect to the baselines:

- **KPI#1:** a 12% improvement in squared generalized cosine similarity has been obtained.
- **KPI#2:** in terms of user throughput, 16% improvement has been achieved for mean throughput, while the improvement for cell-edge users is of 17%.

For both evaluated KPIs, the proposed method achieves the initial target.

3.2.2 Adaptive quantization for AIML-enabled CSI compression

Staying within the realm of the CSI compression problem, another related enabler developed in CENTRIC is that of AI-based CSI compression using adaptive quantization. Here, not only the CSI is compressed by an AI model, but the way the compressed information is quantized is adaptive, that is, going beyond uniform quantization.

Results for this enabler have been contributed to 3GPP in [9]. The outcome and conditions of the validation of this enabler are summarized in Table 5, with the details provided below.

Table 5: Validation Summary for Adaptive Quantization for AIML-Enabled Compression

Adaptive Quantization for AIML-Enabled Compression (InterDigital)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Link-level simulations	B#1: uniform quantizer	KPI#1: SGCS	50% improvement over B#1, loss smaller than 20% over B#2.	>100% improvement over B#1, 20% loss over B#2
	B#2: quantization-aware model	Signalling overhead	50% reduction with respect to B#1.	~50%
		Model generalization	Generalization over different payload sizes	Achieved

3.2.2.1 Validation assumptions

The proposed enabler has been evaluated using link-level simulations. Similarly to the previous enabler, the feedback channel between the user and the base station over which the compressed and quantized CSI is exchanged is assumed to be error-free – that is, the CSI itself still contains errors due to compression and quantization, but the transmission of the information over the channel suffers no packet or bit errors.

Two baselines are used to benchmark the performance of the enabler:

- **B#1:** the first baseline is that of using a classical, uniform quantizer [9].
- **B#2:** the second baseline is provided by a compression model that has been trained aware of the type of quantization used.

Using these baselines, three KPIs are assessed:

- **KPI#1:** similarly as the previous enabler, the quality of the compressed CSI is assessed using squared generalized cosine similarity (SGCS). Target for this KPI is to improve 50% over B#1 and have a loss smaller than 20% with respect to B#2.
- **KPI#2:** as a second KPI, the signalling overhead (in bits) incurred by the proposed method is evaluated. The target is to reduce the overhead by 50% with respect to uniform quantization (B#1) with 4 bits.
- **KPI#3:** the third KPI is the model generalization capability, which is evaluated qualitatively.

3.2.2.2 Validation results

Validation results are reported in [9], and the highlights are included in Figure 4. From here, we can evaluate all the above defined KPIs.

Regarding KPI#1, adaptive quantization for CSI compression improves SGCS by more than 100% for low overhead compared to uniform quantization. Adaptive quantization shows around 20% performance loss compared to quantization-aware CSI compression for the lowest quantization (2 bits). For higher quantization (higher overhead) the loss is always below 10% and gap diminishes as overhead increases.

For KPI#2, signaling overhead can also be evaluated based on Figure 4. There, we can observe that adaptive quantization for CSI compression can achieve around 50% overhead reduction (from 256 bits to 128 bits) with similar SGCS performance compared to uniform quantizer. Finally, regarding KPI#3, it is observed that adaptive quantization for CSI compression can achieve generalization to support different payload sizes with minimal performance loss on SGCS. For the quantization-aware model, different models need to be trained for different payload sizes (i.e., for different quantization levels), which complicates its practical implementation.

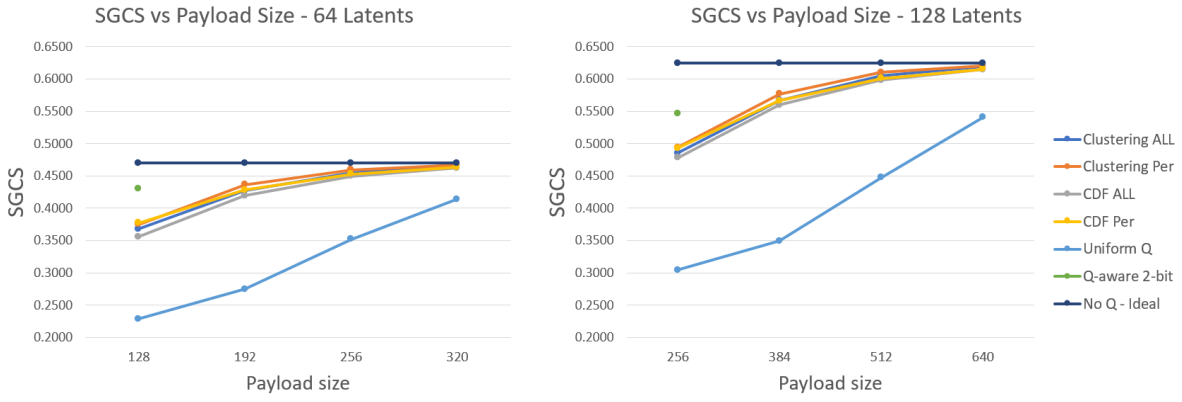


Figure 4: Validation results of Adaptive Quantization for AIML-Enabled Compression [9]

3.2.3 CSI Prediction Enhancements

A third enabler on the topic of CSI compression and prediction is presented here. In this case, the power of AI models is used not just to compress the CSI information, but also to attempt to predict future CSI values. The enabler has been described in a contribution to 3GPP standardization [10].

The outcome and conditions of the validation of this enabler are summarized in Table 6, with the details provided below.

Table 6: Validation Summary for CSI Prediction Enhancements

CSI Prediction Enhancements (InterDigital)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
System-level simulations	B#1: Sample-and-hold B#2: Kalman Filtering	Throughput (mean and 5th percentile)	5% increase for mean throughput, 15% increase for 5th percentile.	Target achieved for certain resource utilization conditions

3.2.3.1 Validation assumptions

For this enabler, system level simulations following the 3GPP simulation assumptions for UE-sided CSI prediction reported in [11].

Two baselines are used to benchmark the performance of the enabler:

- **B#1:** the first baseline, sample-and-hold, simply predicts that the CSI will stay constant until the next CSI measurement [10].
- **B#2:** the second baseline the Kalman Filter, it is a traditional predictor used in statistical signal processing [10].

Using these baselines, a single KPI is assessed for this enabler:

- **KPI#1:** the throughput achieved by the system under the different prediction methods. In particular, two throughput statistics are used: the mean throughput, and the 5th percentile of the throughput distribution, which is typically referred to as the cell-edge user throughput. The target for the KPI is to reach a 5% improvement in mean throughput, and 15% improvement for cell-edge user throughput.

3.2.3.2 Validation results

A sample of the results obtained with the enabler is provided in Figure 5, while a more detailed analysis can be found in [10]. On the left-hand side of the figure, we can find the mean throughput gains of the proposed CSI prediction approach evaluated against B#1 and B#2 under different resource utilization (RU) conditions. The transformer based AIML model CSI prediction performs 10% and 8% better than B#1 and B#2 respectively in terms of mean throughput for 25% resource utilization. The model performs 27% and 25% better than B#1 and B#2 in terms of 5th-percentile throughput for 70% resource utilization.

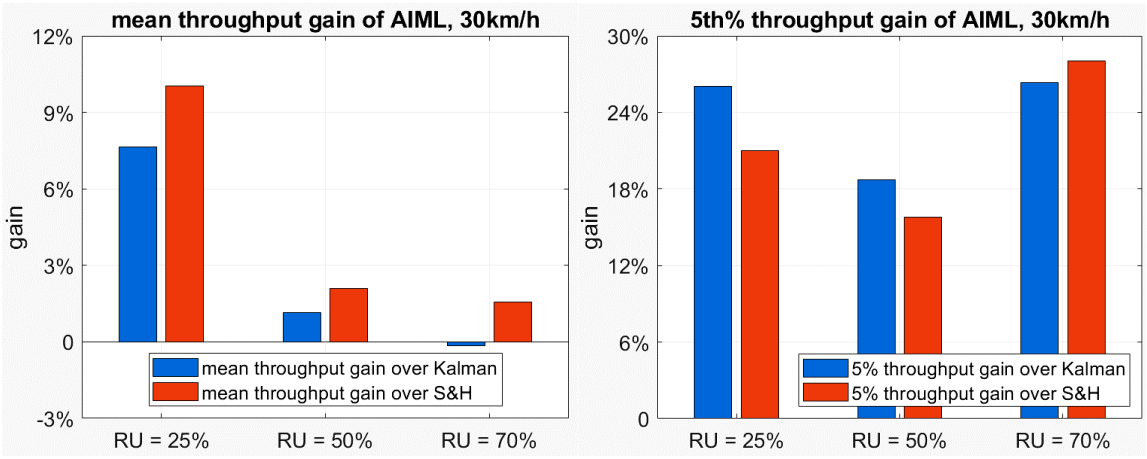


Figure 5: Validation results of CSI Prediction Enhancements

3.2.4 Linear Coded Multi-TRP CSI Compression

As the last enabler in the domain of CSI feedback enhancements, we present now the results of a method for CSI compression in a multi transmission-reception point (TRP) setup. The method relies on linear coding to protect against the loss of uplink control information (UCI). The outcome and conditions of the validation of this enabler are summarized in Table 7, with the details provided below.

Table 7: Validation Summary for Linear Coded Multi-TRP CSI Compression

Linear Coded Multi-TRP CSI Compression (InterDigital)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Computer simulations and mathematical analysis	B#1: lossless compression	KPI#1: squared generalized cosine similarity	Less than 0.2 loss with respect to B#1.	Target achieved with some specific linear coding matrices.

3.2.4.1 Validation assumptions

The validation for this enabler is still in a preliminary state, with simple computer simulations, together with mathematical analysis, have been used.

As baseline for evaluation of the results, a single method is used:

- **B#1:** the baseline is obtained by lossless CSI compression, which represents an ideal case.

Using these baselines, a single KPI is assessed for this enabler:

- **KPI#1:** squared generalized cosine similarity is the KPI used to assess the enabler. The target for the method is to obtain a loss smaller than 0.2 whenever the UCI in a packet is lost.

3.2.4.2 Validation results

A snapshot of validation results obtained in a particular scenario is depicted in Table 8, where the shaded row represents the results obtained for the proposed method. The "Direct decode" method represents B#1, and the numbers indicate the SGCS obtained with the different matrices that can be used to perform linear encoding (A1, A2, and A3).

Table 8: Validation Results for Linear Coded Multi-TRP CSI Compression

Linear coded multi TRP CSI compression				
SGCS	***A1***	***A2***	***A3***	***Direct decode***
$\widehat{H_1}$	0.49	0.76	0.87	0.89
$\widehat{H_2}$	0.90	0.90	0.89	0.89
$\widehat{H_3}$	0.90	0.90	0.89	0.89

The obtained results show that the method's performance depends on the matrix used. We observe that linear coded compressed multi-TRP CSI feedback can achieve 0.76 (0.13 loss) and 0.87 (0.02 loss) for SGCS with specific linear coding matrices (A2 and A3) when the corresponding UCI feedback is lost, hence reaching the target for these two matrices.

3.2.5 UE channel learning and array dimensionality reduction for AI-based MU-MIMO precoding

This enabler evaluates two complementary techniques. The first one consists of an array dimensionality reduction that can lower the complexity of MU-MIMO processing. The second one consists of an approach for channel learning carried out at the UE side. The outcome and conditions of the validation of this enabler are summarized inTable 9, with the details provided below.

Table 9: Validation Summary for UE channel learning and array dimensionality reduction

UE channel learning and array dimensionality reduction (Sequans)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Link-level simulations	B#1: Capacity of MIMO channels [12]	KPI#1: SNR	<1 dB loss with respect to no reduction	0.75 dB loss
	B#2: Channel learning [13]	KPI#2: reconstruction accuracy	Within 2dB of B#2	1.17 dB loss over B#2

3.2.5.1 Validation assumptions

The validation of the enabler is carried out based on link-level simulations performed in MATLAB. The dimensionality reduction method is tested in a system consisting of a UE with a receive array of 8 antenna elements, targeting a reduction to a 4-antenna effective array. An exhaustive search over all possible 4-antenna combinations. For the channel learning technique, on the other hand, a base station with 32 transmit antennas is considered, serving 4 closely spaced single-antenna users.

Two theoretical baselines are used to benchmark the proposed techniques, presented below:

- **B#1:** as first baseline, used for the dimensionality reduction techniques, the capacity bound of MIMO Gaussian channels in [12].
- **B#2:** the second baseline, used to evaluate channel learning, is the method presented in [13].

As KPIs for evaluation of the methods, two are considered:

- **KPI#1:** the first KPI, used to assess the dimensionality reduction technique, is the SNR obtained with the best permutation of antenna reduction. This KPI is evaluated under two conditions: perfect, and imperfect CSI. In the former case, the target is to experience less than 1 dB loss of SNR in the effective, reduced array with respect to the full array. In the case of imperfect CSI, the target is to outperform the full array.
- **KPI#2:** The second KPI, used to evaluate the channel learning method, is defined as the channel reconstruction accuracy per channel matrix element. The target is to perform within 2 dB of B#2.

3.2.5.2 Validation results

Here, we show a snapshot of results for the two KPIs defined above. Note that some results for the dimensionality reduction technique have already been reported in CENTRIC deliverable D3.3. Further results for both array dimensionality reduction and channel learning techniques will be documented in D3.5.

The results for the SNR obtained with the dimensionality reduction technique are reported in Figure 6, where the left-hand side figure represents the perfect CSI case, and the right-hand figure shows the case in which channel estimation errors are present. In the former case, we observe that the best permutation (BP) for array reduction performs within 0.75 dB of the full array when 3 interfering layers are present. Performance degrades though, as expected, when the number of interfering layers grows. In the imperfect CSI case, the different array reduction methods perform closely to the full array, and even surpass its performance for large number of interfering layers.

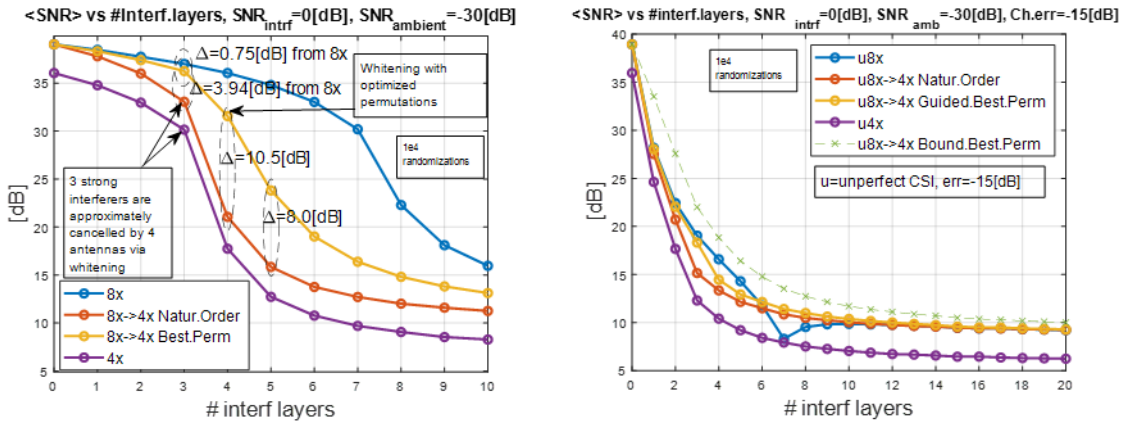


Figure 6: Validation results of array dimensionality reduction for KPI#1

The results for the channel learning part are illustrated in Figure 7, which shows the empirical probability density function of the channel reconstruction error. From this error distribution, it is observed that the channel reconstruction accuracy results is 24.96 dBs, whereas B#2 in the same conditions achieves 26.13 dB. Thus, the proposed method is 1.17 dB from the baseline, fulfilling the target for KPI#2.

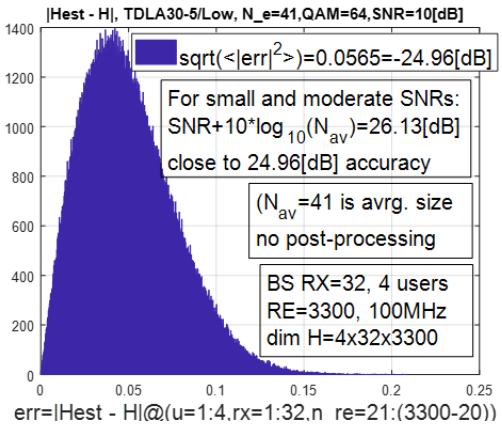


Figure 7: Validation results of channel learning for KPI#2

3.2.6 Joint Sensing and Communications

One of the expected main novelties of 6G systems is that of the inclusion of joint sensing and communication capabilities, where AI-based methods are expected to contribute to achieve the required sensing accuracy in order to be able to exploit the situational information obtained to improve communication performance. In this direction, CENTRIC has developed methods for joint sensing and communication that are applied to the problem of beam management in millimeter-wave (mmWave) systems. The method relies on proximal policy optimization (PPO), a reinforcement learning algorithm.

The outcome and conditions of the validation of this enabler are summarized in Table 10, with the details provided below.

Table 10: Validation Summary for Joint Sensing and Communications

Joint Sensing and Communications (AAU)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Link- and system-level simulations	B#1: X-TDMA	KPI#1: Communication rate	4 b/s/Hz	Target achieved 60% of time
	B#2: Random policy			
	B#3: AoD-based	KPI#2: Latency	N/A	N/A

3.2.6.1 Validation assumptions

The validation of the enabler is performed based on link and system-level simulations, with system parameters as expressed in Table 11.

Table 11: Validation parameters

Parameter	Value
BS antennas N_t	32
Carrier frequency f_c	28 GHz
Tx Power P_t	15 dBm
Noise Power σ_ω^2	-109 dBm
Rician Factor K	10 dB
Time duration of a TTI ΔT	10 ms
Rate - Communication Threshold c	4 b/s/Hz
Radar Cross Section σ_{rcs}	25 m^2
Scenario Dimensions	[100m ×100m]
Number of UEs U	2
Initialized Packet Number $B_{u,0}$	$B_{u,0} \sim Binomial(100,0.6)$
User Speed v	$v \sim Normal(\overline{v}, 4), \overline{v} = 15,20,25,30$

In addition to the proposed PPO-based technique, three different baselines are evaluated:

- **B#1:** The first baseline is based on using a round of periodic beam sweeping every X slots that are used for communication, in a deterministic manner. It is coined *X-TDMA*, due to the way communication and sensing slots are allocated.
- **B#2:** The second baseline corresponds to a random allocation of beam sensing and communication slots, and it is coined *Random.*
- **B#3:** The third baseline corresponds to the base station performing beam selection based on an estimate of the user's channel angle of departure (AoD). A genie-aided version of this baseline, which assumes perfect knowledge of the AoD, is also implemented, representing an ideal, unachievable performance.

The performance of the enabler is benchmarked against two different KPIs:

- **KPI#1:** the first used KPI is communication rate. The simulations assume packets transmitted at a fixed rate corresponding to a spectral efficiency of 4 b/s/Hz, and the target is to minimize the amount of erroneous transmissions together with the transmissions used for sensing.
- **KPI#2:** the second KPI is the communication latency, that is, the time it takes to successfully transmit a packet to the user (accounting for the time required for sensing, buffering, erroneous transmissions). Target for KPI#2 is to achieve a latency no larger than that of B#1.

3.2.6.2 Validation results

Validation results for this enabler have been described in part in [14], while further unpublished results are also included here.

A snapshot of some of the validation results obtained for this enabler in terms of KPI#1 are included in Figure 8. In the left-hand figure, the CDF of the normalized throughput is represented in the right-hand figure, and the average normalized throughput is evaluated as

a function of the user's speed. It is worth noting that throughput is normalized with respect to the data rate target for KPI#1. It can be observed from the CDF that the proposed PPO-based method achieves throughput 1 about 60% of the time, indicating that this percentage of the time the achievable rate of the user is above the target value. It is clearly seen that the proposed method outperforms all other baselines in terms of average throughput and is the method closest to the unachievable performance of the genie-aided version of B#3.

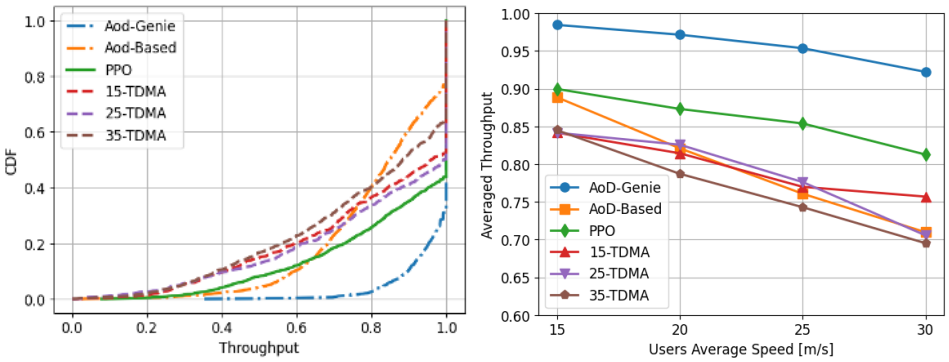


Figure 8: Validation results of Joint Communication and Sensing for KPI#1

In terms of communication latency, while this has not strictly been evaluated, some evidence of the use of slots for transmissions for the different methods can be observed on the left-hand side of Figure 9, whereas in the right-hand side the causes of the erroneous transmissions of each of the methods are analysed. It can be concluded that the PPO-based method allocates fewer resources for sensing than 1-TDMA while getting comparable packet drop performance, which results in lowered latency. Compared to 3/6-TDMA, the proposed method uses more resources for sensing, but leads to significantly lower amount of packet drops. In conclusion, the proposed reinforcement learning algorithm is capable of striking a balance behaviour, smartly allocating sensing and communication slots to achieve a better trade-off and overall increased data rates in comparison to the baselines.

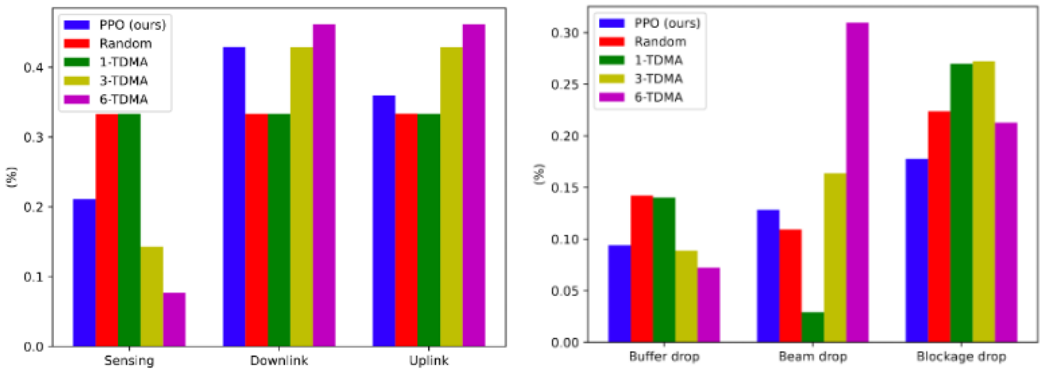


Figure 9: Validation results of Joint Communication and Sensing for KPI#2

3.2.7 ML-enabled Symbol Modulation

We now move on to a different type of enabler developed in CENTRIC's WP3. In this case, an autoencoder neural network is used as a method to automatically learn symbol modulations

specially tailored to specific system conditions. The outcome and conditions of the validation of this enabler are summarized in Table 12, with the details provided below.

Table 12: Validation Summary for ML-enabled Symbol Modulation

ML-enabled Symbol Modulation (InterDigital)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Link-level simulations	B#1: QAM modulations	KPI#1: Bit/Symbol-error rate	Lower than B#1 under nonlinear impairments	>3dB gain at SER = 10%.
		KPI#2: Model generalization ability	Generalization achieved for multiple modulation orders	Generalization to 8-ary and 16-ary modulations.

3.2.7.1 Validation assumptions

In this case, the method is validated by learning modulations for radio channels that include non-linear impairments. Link-level simulations are used to simulate the radio impairments, train an ML model under the impaired channel, and assess the performance of the learned symbol modulation.

The method's performance is evaluated against a single baseline:

- **B#1:** as baseline, performance is compared with that of classical quadrature amplitude modulation (QAM).

As evaluation metrics, we use the following KPIs:

- **KPI#1:** bit and symbol error rate achieved by the method (no channel coding involved). The target is to achieve lower values than those achieved by B#1.
- **KPI#2:** model generalization, in the sense of how well the model can generalize to different number of input bits (i.e., modulation order) is qualitatively assessed.

3.2.7.2 Validation results

The results obtained with the trained autoencoder under nonlinear impairments are presented in Figure 10, where the symbol-error rate (SER) performance of the learned constellations for 8-ary and 16-ary modulations is depicted against the system's signal-to-noise ratio. The SER for the corresponding traditional modulations is also depicted.

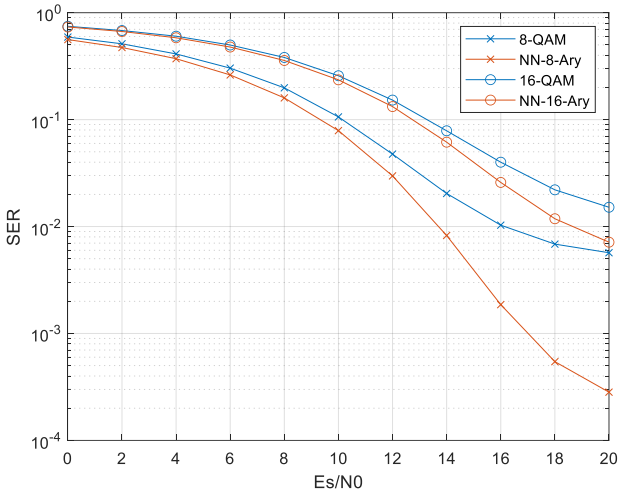


Figure 10: Validation Results for ML-enabled Symbol Modulation

The results show that the target for KPI#1 of improving the SER performance with respect to traditional modulations (B#1) is achieved over a wide range of SNR, obtaining gains of around 3 dBs at a SER of 10%. It is also observed that the target for KPI#2 is as well achieved, as the model is capable of generalizing well to, at least, two different settings of input bits, corresponding to 8-ary and 16-ary modulations.

3.2.8 Multi-user MIMO Neural Receiver

We now present the validation results for the multi-user MIMO neural receiver developed by NVIDIA in the context of WP3. This is one of the enablers in which the validation work has achieved a higher level of maturity. In addition to the work done in WP3 for development of the concept and evaluation of performance via Monte-Carlo simulations, there has also been work on the enabler in WP2 in order to accelerate inference with GPU processing. Furthermore, the receiver has also been evaluated in a proof-of-concept implementation based on hardware-in-the-loop in WP5. In this report, we focus on the validation of the concept's performance via simulations, and the evaluation of the outcome of the acceleration. A detailed analysis of the validation achieved through the proof-of-concept implementation is reserved for D5.4, which will report all the proof-of-concept activities in CENTRIC.

The outcome and conditions of the validation of this enabler are summarized in Table 13, with the details provided below.

Table 13: Validation Summary for Multi-user MIMO Neural Receiver

Multi-user MIMO Neural Receiver (NVIDIA)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Link-level simulations and proof-of-concept with hardware-in-the-loop	B#1: receiver with LS-based channel estimation and LMMSE MIMO detector	KPI#1: BLER for fixed computational complexity	BLER close to B#2 with lower computational complexity.	<1 dB loss with respect to B#2

	B#2: receiver with LMMSE-based channel estimation and K-Best MIMO detector	KPI#2: Inference latency	<1ms inference latency on NVIDIA A100 GPU	1ms latency processing 132 PRBs

3.2.8.1 Validation assumptions

The validation of the concept has been carried out using link-level simulations utilizing NVIDIA's Sionna platform. The receiver has been evaluated in a 5G-compliant uplink configuration, using 3GPP compliant channel models as well as ray-traced channels for specific scenarios.

Two baselines are used for comparison:

- **B#1:** the first baseline consists of a 5G PUSCH MU-MIMO receiver with LS-based channel estimation and LMMSE MIMO detector, as available in open-source implementation in Sionna [15]. This can be considered as a state-of-art commercial receiver for 5G.
- **B#2:** the second baseline is a 5G PUSCH MU-MIMO receiver with LMMSE-based channel estimation and K-Best MIMO detector, with implementation also available in Sionna [15]. This receiver can be considered very close to optimal, but has an infeasible computational complexity for current base stations.

The KPIs used to assess the receiver's performance are:

- **KPI#1:** as first KPI, the block error rate (BLER) performance of the receiver evaluated at a specific computational complexity is used. The target for KPI#1 is to achieve values close to the BLER of B#2 at a lower computational complexity.
- **KPI#2:** the second KPI relates to the inference latency incurred by the accelerated version of the neural receiver. The target for KPI#2 is to achieve an inference latency no larger than 1ms on a specific computational platform (NVIDIA A100 GPU). This would allow running the receiver in real time.

3.2.8.2 Validation results

Starting with the validation results for KPI#1, an illustration of the neural receiver performance for different number of active layers, N_T, is depicted in comparison to the baselines used in Figure 11. The complexity of the neural receiver has been tuned to be lower to that of B#2. As it can be observed, for a number of spatial layers spanning from 1 to 4, the neural receiver performs always within 1dB of B#2, and even achieves better performance for some SNRs. In addition, a very sizable gain (~3dB) is observed in comparison with B#1, which represents a practically achievable baseline. Further performance results for this enabler can be found in CENTRIC's deliverable D3.2.

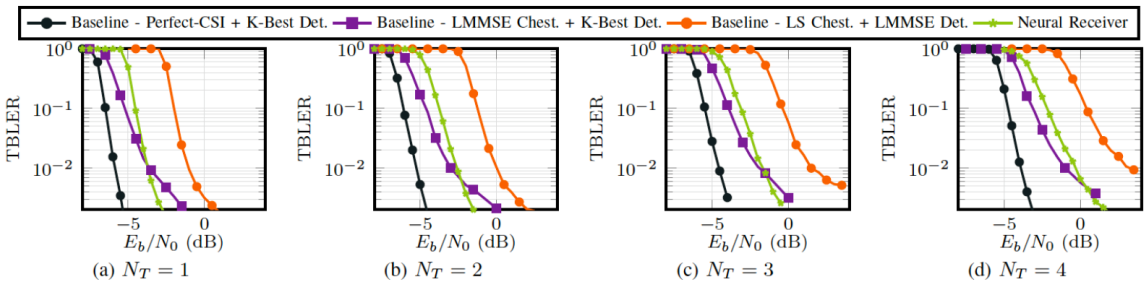


Figure 11: Validation Results of Multi-user MIMO Neural Receiver for KPI#1

Validation results for KPI#2 are presented in Figure 12, where the inference latency (blue) and the SNR required to attain a BLER of 10% (green) are depicted versus the number of iterations run at the neural receiver. As it can be observed, the inference latency grows linearly with he number of iterations run at the receiver, whereas the SNR required to achieve the target BLER decreases. This illustrates the existing tradeoff between inference latency (and computational complexity) and receiver performance. The results show that, while the receiver performance would be slightly degraded, it is possible to limit the number of iterations of the receiver in order to achieve the target of real-time processing.

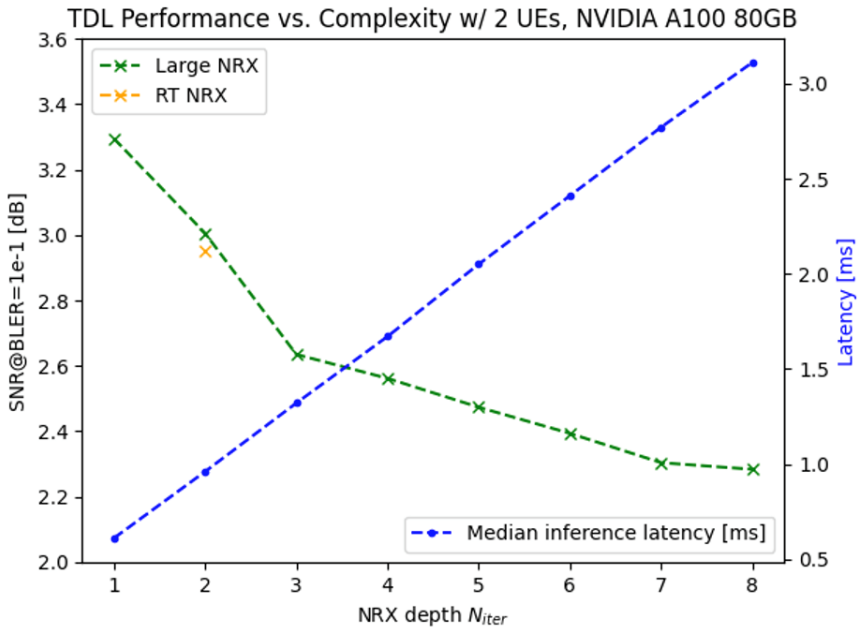


Figure 12: Validation Results of Multi-user MIMO Neural Receiver for KPI#2

3.3 Benchmarking of WP4 Enablers

3.3.1 DCI Compression

The first WP4 enabler that we present is that of DNN-based downlink control channel (DCI) compression. The objective of the enabler is to find way to perform lossless compression of the DCI information that leads to a smaller amount of bits required to transmit the DCI

information. The proposed lossless encoder is based on a transformer DNN architecture, and has been presented in [16].

The outcome and conditions of the validation of this enabler are summarized in Table 14, with the details provided below.

Table 14: Validation Summary for DCI Compression

DCI Compression (NNF)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
System-level and link-level simulations	Huffman Coding [17]	KPI#1: PDCCH reliability	0.2 dB decoding gain	Up to 0.3 dB gain
	Deepzip [18]	KPI#2: Lossless compression ratio	10% improvement versus B#1	18% improvement over B#1.

3.3.1.1 Validation Assumptions

The proposed compressor has been validated using Monte Carlo simulations with two different simulation tools. On the one hand, a system-level simulator with standard-compliant implementation of 5G NR Time Division Duplex symbol-based scheduling has been used to collect datasets to train the designed DNN. As the second tool, a link-level simulator used to model transmission of the Physical Downlink Control Channel (PDCCH) over additive white Gaussian noise (AWGN) channels has been used. This second simulator is used to assess the decoding performance of the PDCCH under the different methods of DCI compression considered.

Two main baselines are used to evaluate the performance of the proposed enabler:

- **B#1:** the first baseline is to use Huffman coding [17], a classical method for lossless compression.
- **B#2:** the second baseline is another AI-based method that relies on recurrent neural networks (RNNs) and is coined Deepzip [18].

The proposed method and baselines are evaluated with respect to two KPIs:

- **KPI#1:** the first KPI is the PDCCH reliability, measured in terms of its frame error rate (FER). The target for this KPI is to achieve 0.2 dB SNR gain in the decoding of the PDCCH compared to the current standard method.
- **KPI#2:** the second KPI is the compression ratio achieved by the proposed lossless compression scheme, defined as the ratio between the original message bitlength and the compressed message bitlength. The target is to improve the compression ratio by 10% compared to B#1.

3.3.1.2 Validation Results

We present now the validation results for the DCI compression enabler. Starting with KPI#1, an illustration of the results can be seen in Figure 13, where the different methods listed in the legend correspond to:

- PDCCH (39,100): The original PDCCH Polar encoding and decoding scheme with an encoded length of 100 bits (current standard)
- HC: PDCCH with a lossless Huffman coding compressor (B#1).
- Transformer-BCE: PDCCH with a lossless transformer-based compressor trained using binary cross-entropy (BCE) loss (proposed enabler)
- Transformer-mFocal: PDCCH with a lossless transformer-based compressor trained using a modified focal loss (proposed enabler)
- RNN-DeepZIP: PDCCH with a lossless recurrent neural network (RNN)–based compressor (B#2).
- Joint Transf. & HC: PDCCH with a lossless compressor that jointly applies a transformer-based model and Huffman coding (proposed enabler)

The results show that a decoding gain of up to 0.3 dB can be obtained with the proposed compression methods.

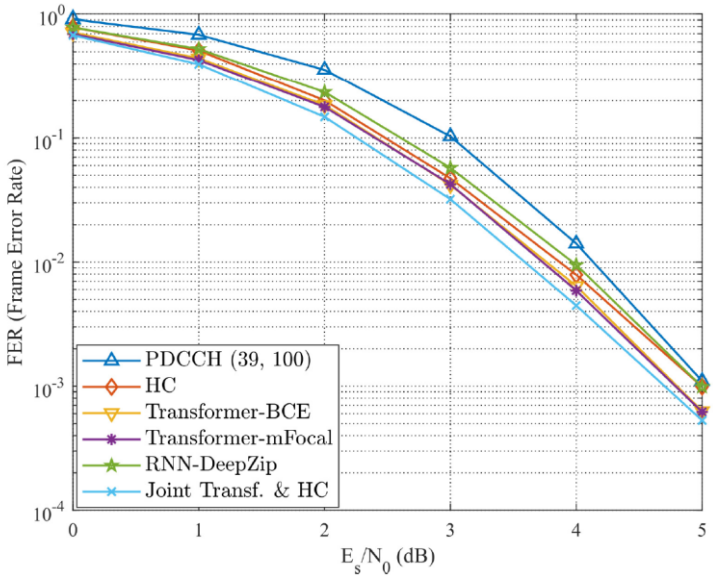
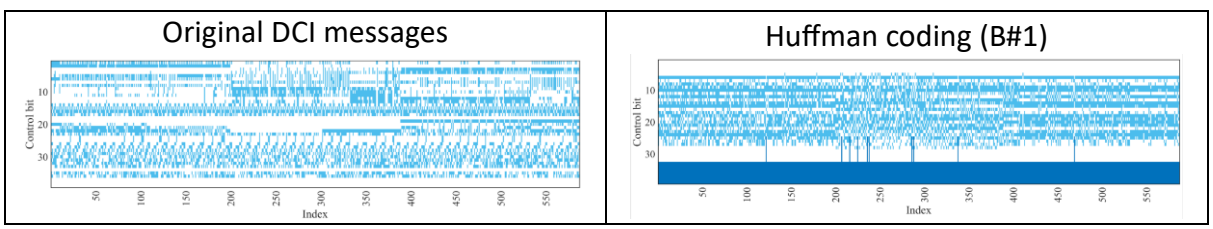


Figure 13: Validation results of DCI Compression for KPI#1

In connection with KPI#2, the compression ratio, a snapshot of the compression levels achieved by the proposed enabler and the baselines is depicted in Figure 14. The figures represent the original DCI messages and their compressed versions with the different compressors. White and light blue colour indicate bit values of 0 and 1, respectively. Dark blue colour indicates bits that remain null after compression, that is, they represent the overhead savings obtained with the compression methods. The outcome of this validation is that the compression ratio obtained by the proposed transformer-based solution showcases a gain of 18% with respect to B#1, thus exceeding the original target of 10% improvement.



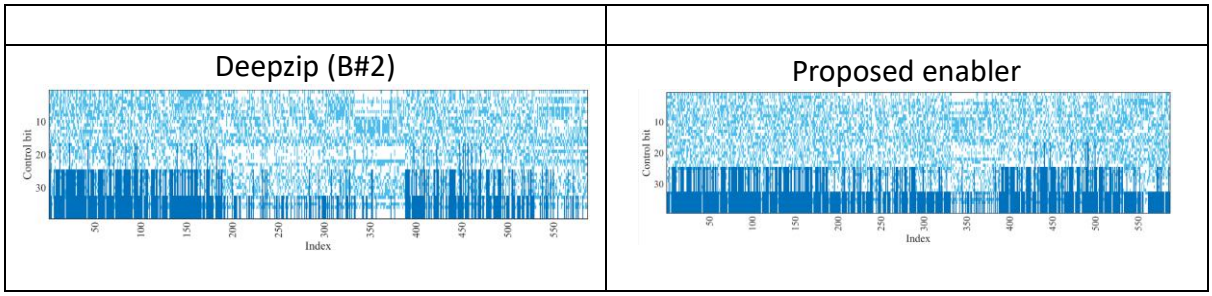


Figure 14: Validation results of DCI Compression for KPI#2

3.3.2 Task-oriented Cognitive Wireless Scheduling: collaborative navigation

The first enabler in the are of task-oriented cognitive wireless scheduling is an AI-based solution to enable collaborative navigation between teams of robots, a scenario pertaining to the social manufacturing use-case of Industry 4.0 shop floor. The problem is illustrated in Figure 15 and Figure 16. Figure 15 shows the system model for the robotic navigation problem. In it, Two teams of 3 robots each need to reach their assigned destination as soon as possible and execute there a task in parallel.

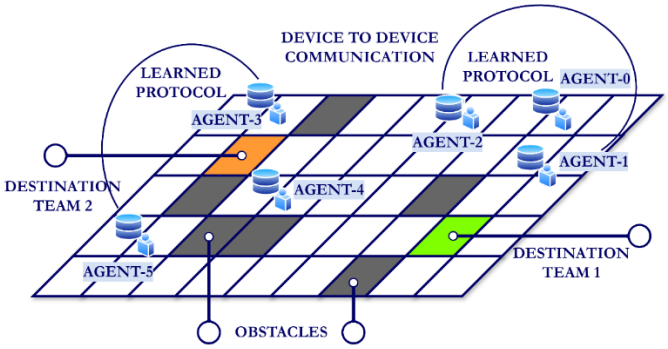


Figure 15: System model of the problem

In Figure 16, the communication model of the problem is depicted. The role of the BS is to allocate the data channel to the team of robots that has less agents at the destination working on their task. The BS needs to ensure that the two task are executed in parallel, that is, that there is an equal number of robots at each of the two destinations. A team of robots can only move on the grid if they exchange their AI-emerged protocol messages. There are 6 robots requiring access to 3 shared uplink data channels and 3 shared downlink channels.

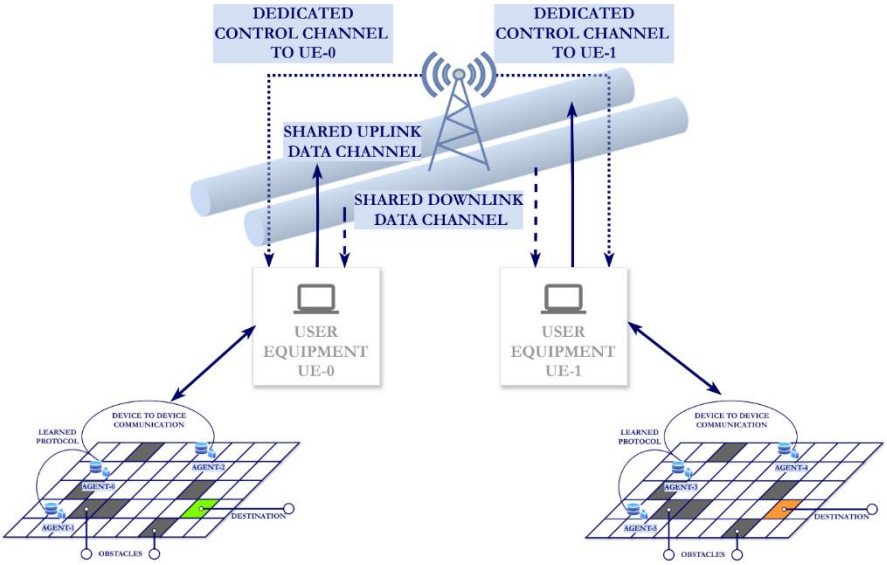


Figure 16: Communication model of the problem

The outcome and conditions of the validation of this enabler are summarized in Table 15, with the details provided below.

Table 15: Validation Summary for Task-oriented Cognitive Wireless Scheduling: collaborative navigation

Task-oriented Cognitive Wireless Scheduling: Collaborative Navigation (OUL)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Monte Carlo simulations	B#1: round robin scheduling	KPI#1: Parallel task execution (evaluated with Jain's fairness index [1])	As close to 1 as possible	Values between 0.9 and 1.
	B#2: optimal navigation policy from [19]	KPI#2: Latency to destination (in time-steps)	As low as possible	103.5 ± 20.2 time steps

3.3.2.1 Validation Assumptions

The validation of the enabler is based on Monte Carlo simulations on a computer implementation of the scenario described above. At the beginning of the simulation, the 6 robots are placed randomly on the grid and they need to navigate to their assigned destination. Each team of 3 robots has an assigned goal position. The simulation is run until all the robots have reached the destination or is terminated after a maximum time-out number of steps. Within a team of robots, and ideal communication channel without packet errors and transmission delays is assumed. Each team of robots navigates on the grid to the destination by exchanging their AI-emerged messages proposed in [19]. The base station coordinates the team of robots via dedicated control channels. The communication channels

are considered ideal, as the focus is on the channel access logic. Hence, no assumptions are used on channel properties, channel estimation, or data encoding and decoding schemes.

Two baselines are used for comparison of the proposed scheme:

- **B#1:** the first baseline is a simple round-robin scheduling scheme for channel access. This baseline does not consider the level of task completion of the different teams of robots, so it is a non-goal oriented scheme.
- **B#2:** the second baseline is the navigation policy proposed in [19]. This method 's goal is to improve the navigation time to destination of a team of robots. The baseline is used to extract the messages given by its optimal navigation policy.

The proposed method is assessed in terms of two KPIs:

- **KPI#1:** the first KPI is the degree of parallel task execution, and is measured using Janin's fairness index [1]. The target is to reach a value as close to 1 as possible in the index, which corresponds to an equal share of resources.
- **KPI#2:** the second KPI is the latency to destination, that is, the number of time-steps that the last robot of a team needs in order to reach the goal position. The target is to decrease it as much as possible.

3.3.2.2 Validation Results

The validation results for KPI#1 are depicted in Figure 17, where Janin's fairness index is plotted for the proposed enabler and for B#1 against the number of robots that have reached the goal. It is clearly seen how the proposed method, in its two variants, produces always fairness much closer to the goal of 1 than a round robin scheduling scheme (B#1). In the proposed scheme, the BS monitors the number of robots at the goal from the current team and switches the channel access as soon as a new agent has reached the goal, leading to the increased fairness depicted in the figure.

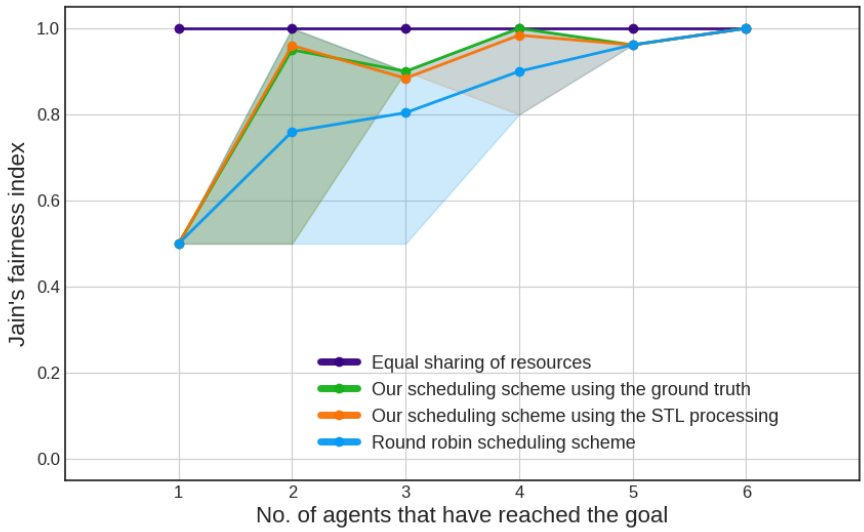


Figure 17: Validation results of collaborative navigation for KPI#1

The values for the latency to destination (KPI#2) are obtained from [19], by evaluating the freely available code provided by the authors. This corresponds to a value of 103.5 ± 20.2 time steps. This shows that enabling communication between robots provides guarantees of convergence of all the agents to their goals, which is something that cannot be guaranteed without communication.

3.3.3 Task-oriented Cognitive Wireless Scheduling: semantic communication and control co-design

As a second enabler in the domain of task-oriented cognitive wireless scheduling, we know present the validation results for a method for semantic communication and control co-design developed within WP4. The proposed method consists of a scheduling approach for multi-device control systems that integrates self-supervised learning and split learning. The method relies on the introduction of time-series joint embedding predictive architecture (TS-JEPA) to model semantic control dynamics within the latent space. To enhance efficiency, a latent space-based scheduling scheme was implemented to optimize wireless resources utilization based on both control performance and channel condition. More details on the method can be found in [20].

The outcome and conditions of the validation of this enabler are summarized in Table 16, with the details provided below.

Table 16: Validation Summary for Task-oriented Cognitive Wireless Scheduling: semantic communication and control co-design

Task-oriented Cognitive Wireless Scheduling: semantic communication and control co-design (OUL)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Link-level simulations	B#1: AI-based methods (autoencoder and supervised learning)	KPI#1: Control performance – Normalized score	In range [0.75—1]	Within the target range
	B#2: classical schedulers (random, greedy, round-robin)	KPI#2: Communication efficiency – Communication bits and Transmission latency	Minimize while maintaining target for KPI#1	10x faster transmissions than control sampling rate

3.3.3.1 Validation Assumptions

The validation was conducted in a simulation environment involving multiple independent non-linear inverted cart-pole systems with different control objectives, representative of industrial internet-of-things (IIoT) use cases such as smart factories. The link-level simulation framework accounts for the application layer (i.e., capturing frames through High frame-rate cameras for real-time sensing), data-link layer (i.e., implementing medium-access control (MAC) policies to prioritize robots for transmission), and physical layer (i.e., dynamically

allocating wireless resources). The BS centralizes transmission scheduling, ensuring that robot with poor control performance but favorable channel conditions are prioritized for communication with the remote controllers. Moreover, while the current validation assumes channel state information (CSIs) availability at the BS, ongoing work explores latent space-based wireless dynamics prediction to predict future channels within the latent space [21]. Moreover, the base station is assumed to have access to the CSI of users for scheduling purposes, the control dynamics are assumed to be unstable on their own, and each control system operates independently but within a shared wireless network.

The evaluation of the proposed semantic communication and control co-design framework is compared against two main categories of baselines. First, AI-based baselines that leverage AI techniques to optimize control performance within the latent space. Second, non-AI-based baselines are the traditional scheduling approaches to allocate wireless resources. These two types of baselines provide a comprehensive performance comparison in balancing communication efficiency and control performance. The selected baselines are:

- **B#1:** as AI baselines, an auto-encoder and a supervised learning approach are used.
- **B#2:** as traditional (non-AI) baselines, classical random, round-robin, and greedy scheduling algorithms are used.

The method's performance is characterized by using two KPIs:

- **KPI#1:** the first KPI relates to the control performance, and the normalized score is used to measure it. The normalized score aims to measure how closely the control systems follow the desired objectives. The target range is [0.75-1.0], ensuring control performance remains high while optimizing wireless resources.
- **KPI#2:** the second KPI characterizes the communication efficiency, by measuring the communication bits used and their transmission latency. The target is to minimize the bit transmission overhead while ensuring good control performance.

3.3.3.2 Validation Results

The validation results for both KPIs are depicted in Figure 18, where the left-hand figure presents the results for KPI#2 and the right-hand figure those of KPI#1. Starting with KPI#1, we observe that the two instances of the proposed TS-JEPA scheme (with and without prediction) stay always above a normalized score of 0.75. The instance using prediction displays a score that tends to the value 1 as the prediction horizon increases.

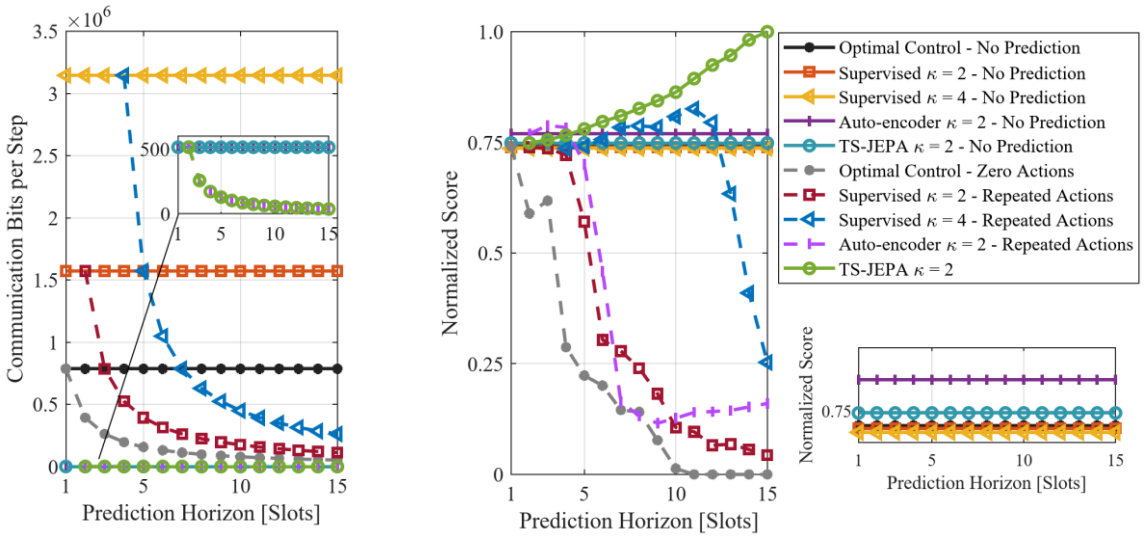


Figure 18: Validation results of Task-oriented Cognitive Wireless Scheduling: semantic communication and control co-design

For KPI#2, we observe that the proposed method obtains much lower communication overhead than most of the evaluated baselines. In particular, the instance that uses prediction exhibits a communication overhead that vanishes as the prediction horizon grows.

3.3.4 Emerging multiple-access protocols for specialized services

Another enabler developed in WP4 is that of a method for emerging multiple-access schemes for specialized communication services. In particular, the use-case employed to validate the method is that of an in-factory scenario made of subnetworks. These subnetworks are installed in mobile robots, and power control of them is necessary in order to control interference while satisfying the communication requirements.

The outcome and conditions of the validation of this enabler are summarized in Table 17, with the details provided below.

Table 17: Validation Summary for Emerging multiple-access protocols for specialized services

Emerging multiple-access protocols for specialized services (AAU)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
System-level simulations in in-factory scenarios	B#1: genie-aided	KPI#1: buffer flush rate	0.9 median value	0.88 median value
	B#2: random protocol	KPI#2: signalling overhead	50% reduction with respect to B#1	80% reduction
	B#3: interference-aware protocol			

3.3.4.1 Validation Assumptions

The enabler is validated by means of system-level simulations of the targeted in-factory subnetworks scenario. It is assumed that a central entity is present and can perform power control based on signalling information received from the subnetworks. The system parameters used for the simulations are detailed in Table 18.

Table 18: Simulation parameters for Emerging multiple-access protocols for specialized services

Parameters	Values
Deployment area	10 m ×10 m
Subnetwork radius	1 m
Operating frequency	6 GHz
Bandwidth	10 MHz
Path loss exponent	2.7
Maximum speed of subnetworks	3 m/s
Maximum transmit power	20 dBm
Minimum transmit power	0 dBm
Number of subnetworks	10
Number of devices per subnetwork	1
Distance between device and and AP	0.5 m
Noise spectral density	−174 dBm/Hz
Payload	64 bytes
Latency	0.001s
Threshold Spectral Efficiency (Rth)	0.05 bps/Hz
Capacity of AP buffer	100

For comparison, three non-AI baselines are implemented as well:

- **B#1:** as ideal baseline, a genie-aided persistent power allocation algorithm is used.
- **B#2:** the second baseline consists of a protocol that performs random actions.
- **B#3:** the third baseline is and interference-aware power control protocol.

The performance of the proposed method is characterized using two KPIs:

- **KPI#1:** as first KPI, the buffer flush rate, that is, the median number of successful packets transmitted per time step, is used. The target for the KPI is to achieve a value of 0.9.
- **KPI#2:** the second KPI relates to the signalling overhead incurred by the protocol, measured in bits. The target for this is to obtain a 50% reduction with respect to B#1.

3.3.4.2 Validation Results

The validation results for KPI#1 are displayed in Figure 19, which depicts the empirical CDF of the buffer flush rate obtained by the different methods. The median value achieved by the proposed enabler (MAPPO MAC) is of 0.88, which slightly falls short of reaching the target of 0.9. Nonetheless, the results show a clear superiority of the proposed AI-based approach with respect to all of the evaluated baselines (except B#1, which represents an ideal unattainable performance).

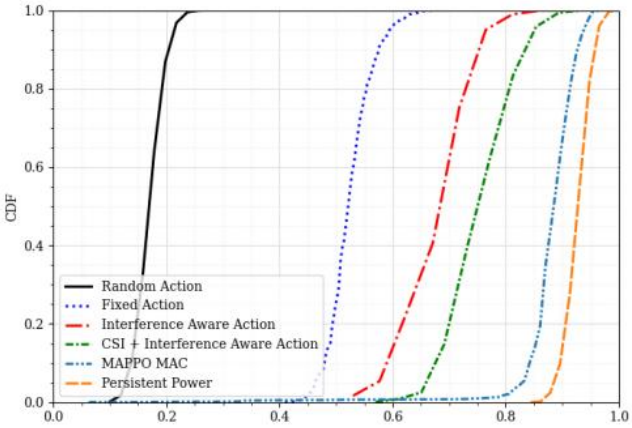


Figure 19: Validation results of Emerging multiple-access protocols for specialized services for KPI#1

The results for KPI#2 are presented in Figure 20, which shows the signaling overhead incurred by the different methods. It is observed that the proposed method (MAPPO) achieves low overhead compared to most of the baselines. Importantly, it exhibits an 80% reduction with respect to B#1, which exceeds the target of 50% reduction. It is important to know that this significant overhead reduction comes at the expense of only a slight degradation in performance with respect to B#1, as can be seen in Figure 19.

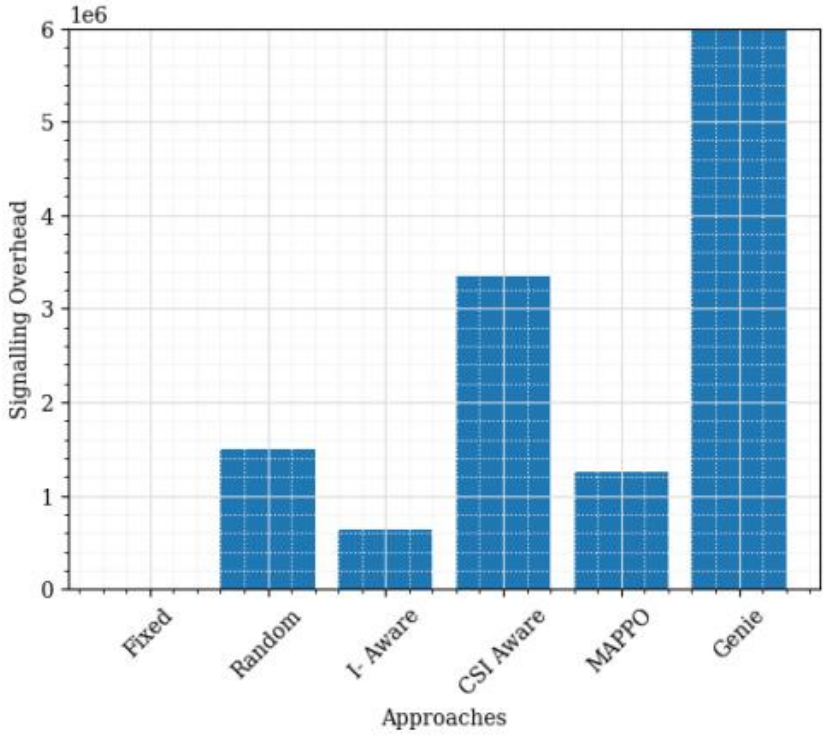


Figure 20: Validation results of Emerging multiple-access protocols for specialized services for KPI#1

3.3.5 Federated Multi-Agent DRL for Radio Resource Management

Continuing within the realm of industrial communication scenarios using subnetworks, we present now the results of a method for radio resource management based on federated multi-agent deep reinforcement learning (DRL). Details on the method can be found in [22].

The outcome and conditions of the validation of this enabler are summarized in Table 19, with the details provided below.

Table 19: Validation Summary for Federated Multi-Agent DRL for Radio Resource Management

Federated Multi-Agent DRL for Radio Resource Management (AAU)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
System-level simulations	B#1: Graph colouring algorithm [23]	KPI#1: User spectral efficiency	6 b/s/Hz at 1st percentile	Target achieved
		KPI#2: average RL reward	12.5	Target achieved

3.3.5.1 Validation Assumptions

The method's performance has been validated by means of system-level simulations, with parameter settings as specified in

Table 20: Simulation parameters for Federated Multi-Agent DRL for Radio Resource Management

Parameter	Value
Total factory area	180 m × 80 m
Clutter type table	Sparse
Number of subnetworks	20
Timestep	0.005 s
Number of episodes	2000
Number of steps per episode	200
Subnetwork separation distance	1 m
Subnetwork radius	1 m
Subnetwork velocity	3 m/s
Transmit power	-10 dBm
Number of frequency channels	4
Carrier frequency	6 GHz
Bandwidth per subnetwork	10 MHz
Noise figure	10 dB
Shadowing decorrelation distance	10 m
Max action switch delay	10

As baseline for performance comparison, one traditional (non-AI) method has been used:

- **B#1:** the selected baseline is the centralized graph colouring algorithm described in [23].

Two KPIs are used to characterize the methods' performance:

- **KPI#1:** User spectral efficiency, evaluated at low percentiles. This allows to characterize the performance of worst-case users. The target is to obtain a rate of 6 bits/s/Hz at the 1st percentile of the spectral efficiency distribution.
- **KPI#2:** The second KPI is the average reward obtained in the operation of the DRL agents. The target is to reach an average value of 12.5.

3.3.5.2 Validation Results

We showcase here a selection of the results that allows to quantify the selected KPIs. Further analysis can be found in [22].

With respect to KPI#1, the results obtained with different instances of multi-agent DRL are depicted in Figure 21, which depicts the cumulative density function (CDF) of the per-user spectral efficiency achieved. It is observed that a couple of the DRL based methods achieve the target of 6 bits/s/Hz at the targeted 1st percentile. All the proposed methods achieve much better performance than a random scheduling algorithm, and similar performance to that of a greedy method. However, they are outperformed by B#1, which is typically considered too complex to be implemented in practice.

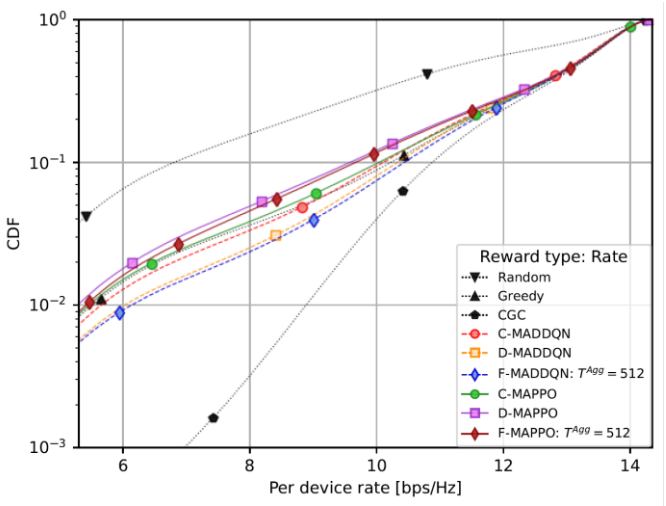


Figure 21: Validation results of Federated Multi-Agent DRL for Radio Resource Management for KPI#1

Regarding KPI#2, results are presented in Figure 22, which depicts the reward over episodes of the operation of the reinforcement learning agents. Results are presented separately for the two DRL frameworks utilized: MADDQN is presented on the left-hand side, whereas different variants of MAPPO are depicted in the right-hand side. The results show that the MAPPO agents converge much faster to the target reward value than the MADDQN counterparts do. The achieved rewards after convergence tend to oscillate around the target of 12.5.

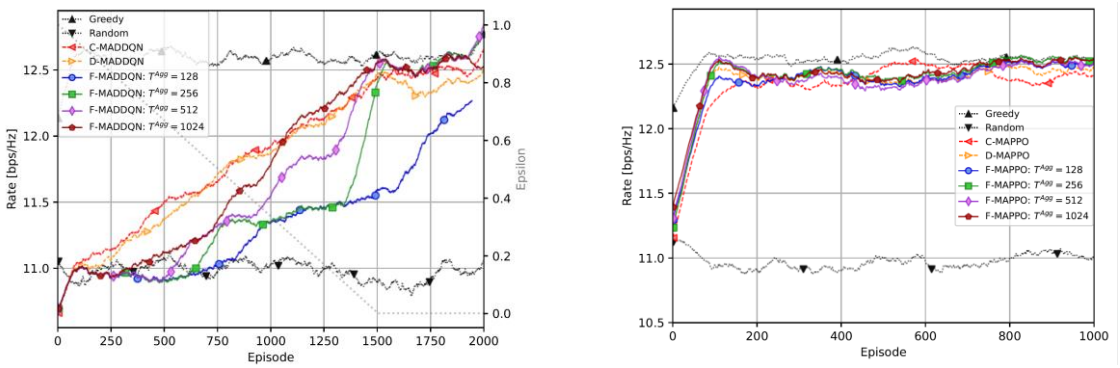


Figure 22: Validation results of Federated Multi-Agent DRL for Radio Resource Management for KPI#2

3.3.6 ML-based Sub-band Selection

Another enabler that deals with radio resource management in 6G In-factory subnetwork scenarios that has been developed in WP4 is a ML-based method for sub-band allocation between the different subnetworks. The method relies on a deep, fully-connected neural network architecture that is trained to maximize the number of subnetworks that fulfill certain rate criteria. The details of the approach are extensively described in [24].

The outcome and conditions of the validation of this enabler are summarized in Table 21, with the details provided below.

Table 21: Validation Summary for ML-based Sub-band Selection

ML-based Sub-band Selection (AAU)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
System-level simulations	B#1: SISA algorithm [25]	KPI#1: rate-conforming subnetworks	Median values of 9 for low-rate and 3 for high-rate subnetworks	9 for low-rate and 2.6 for high-rate networks.
		KPI#2: Training loss L	L = 1	L = 1.1

3.3.6.1 Validation Assumptions

The performance of the proposed allocation method has been evaluated using system-level simulations, with parameters set as in Table 22.

Table 22: Simulation parameters for ML-based Sub-band Selection

Parameter	Value
Factory area, $L \times L$	20 m×20 m
Number of subnetworks, N	20
Number of sub-bands, K	4
Subnetwork radius, R	1 m
Number of devices per subnetwork, J	1
Minimum distance between APs	2 m
device to AP minimum distance, d_{min}	0.8

Shadowing standard deviation, λ	7.2 dB
DL clutter density, r, clutter size, d_s	0.6, 2
De-correlation distance, d_c	5 m
Transmit power, $\mathrm{P_m}$	0 dBm
Bandwidth, B	40 MHz
Carrier frequency, f_c	10 GHz
Noise figure, NF	5 dB
Low-rate subnetwork required SE, $\mathrm{SE_L^{req}}$	0.4
High-rate subnetwork required SE, $\mathrm{SE_H^{req}}$	8

A state-of-the-art algorithm relying on traditional model-based processing is used as baseline:

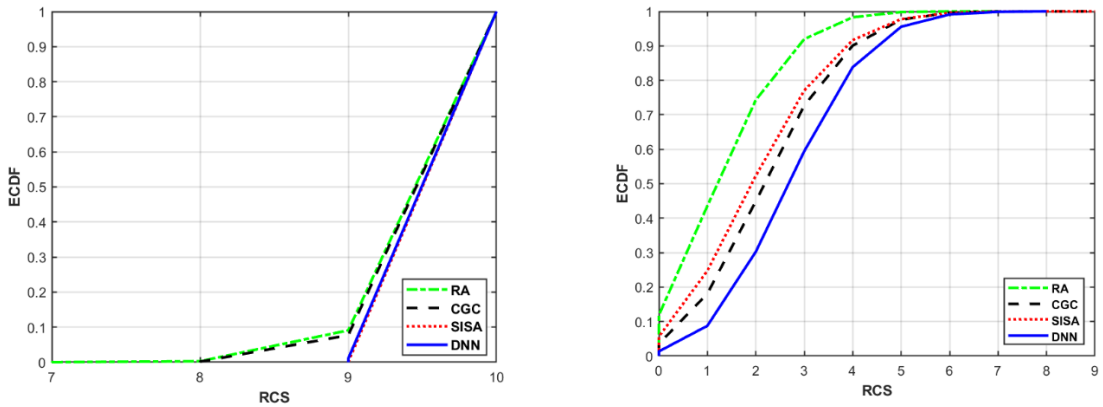
- **B#1:** the sequential iterative sub-band allocation (SISA) algorithm [25].

Two KPIs are used to characterize the enabler's performance:

- **KPI#1:** the first KPI consists of the count of subnetworks that reach their required rate, which is coined rate-conforming subnetworks (RCS). To ensure fairness, the KPI is evaluated separately for high- and low-rate subnetworks, with their targets being a median value of 9 for low-rate subnetworks and a median value of 3 for high-rate ones.
- **KPI#2:** The second KPI is is the loss value achieved by the neural network during training. The loss function is defined as $L = \frac{\sigma(SE_n^{req} - SE_n)}{SE_n^{req}}$, where σ denotes the sigmoid function, SE_n denotes the rate achieved by the nth subnetwork, and SE_n^{req} is its rate requirement. The target value for the loss function is $L = 1$.

3.3.6.2 Validation Results

Starting with KPI#1, the validation results for it are depicted in Figure 23, where the left-hand figure shows results for the low-rate networks and the right-hand one depicts the high-rate subnetworks. It can be observed that the propose DNN method performs equally to B#1 for the low-rate case, whereas it clearly outperforms it in the high-rate count. The target of RCS reaching 9 for the low-rate networks is achieved, but the method is slightly below the target for high-rate ones, reaching only a value of 2.6 instead of the targeted value of 3. Nonetheless, the proposed method is clearly better than B#1 and outperforms all other methods evaluated.



(a): Low-rate subnetworks (b): High-rate subnetworks

Figure 23: Validation results of ML-based Sub-band Selection for KPI#1

With respect to KPI#2, the training convergence curve of the model is depicted in Figure 24, which shows the value of the loss function over training epochs. The loss function achieves, with sufficient training iterations, a value slightly below 1.1, which is close but not reaching the target value of 1.

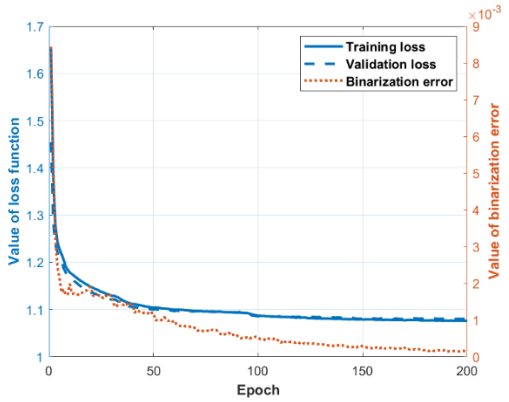


Figure 24: Validation results of ML-based Sub-band Selection for KPI#2

3.3.7 Joint Sub-band Allocation and Power Control for Outdated CSI Scenarios

As a last enabler developed for the problem of radio resource allocation in In-factory subnetworks, we present now the results of an extension of the previous enabler which, in addition to performing sub-band allocation, also carries out power control. In addition, the method operates with outdated CSI.

The outcome and conditions of the validation of this enabler are summarized in Table 23, with the details provided below.

Table 23: Validation Summary for Joint Sub-band Allocation and Power Control for Outdated CSI Scenarios

Joint Sub-band Allocation and Power Control for Outdated CSI Scenarios (AAU)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
System-level simulations	B#1: SISA [25] combined with WMMSE [26]	KPI#1: Spectral efficiency	average SE = 8.5@median per-user SE = 5.8 @10^{-3}	average SE = 8.7@median per-user SE = 5.8 @10^{-3}
		KPI#2: Training loss	10^{-3}	$0.6 \cdot 10^{-3}$

3.3.7.1 Validation Assumptions

The proposed method is evaluated via system-level simulations, with parameters detailed in

Table 24: Simulation parameters for Joint Sub-band Allocation and Power Control for Outdated CSI Scenarios

Parameter	Value
Factory area, $L \times L$	20 m×20 m

Number of subnetworks, N	10
Number of sub-bands, K	3
Subnetwork radius, R	1 m
Number of devices per subnetwork, J	1
Minimum distance between APs	2 m
device to AP minimum distance, d_{min}	0.8
Shadowing standard deviation, λ	4 dB
DL clutter density, r, clutter size, d_s	0.7, 1
De-correlation distance, d_c	5 m
Maximum transmit power, $\mathrm{P_m}$	0 dBm
Subband Bandwidth, B	40 MHz
Carrier frequency, f_c	10 GHz
Noise figure, NF	5 dB
Sounding reference signal period, Δt	100 ms
CSI buffer length, T	5
Prediction length (delay) τ	4

For comparison of the method's performance, a baseline has been defined as follows:

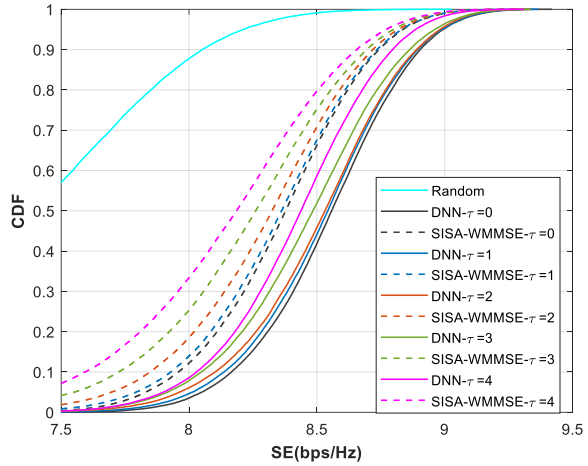
- **B#1:** the baseline consists of the combination of two non-AI algorithms, one for sub-band allocation and one for power control. Sub-band allocation is done by the already mentioned SISA algorithm [25], whereas power allocation is carried out using the weighted sum mean-squre error minimization (WMMSE) method [26].

As evaluation metrics, two KPIs are used:

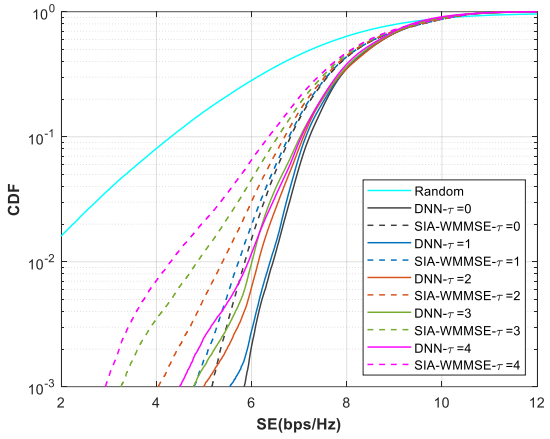
- **KPI#1:** as first KPI, the spectral efficiency (SE) achieved by the method is evaluated. In particular, two statistics of the SE are in focus: the median of the average SE, with a target value of 8.5 b/s/Hz, and the CDF value at 10^{-3} of the per-user SEs, with a target of 5.8 b/s/Hz.
- **KPI#2:** the second KPI used is the training loss achieved during the training process, where the loss function used to train the method is the mean squared error (MSE). The target for KPI#2 is a value of 10^{-3}.

3.3.7.2 Validation Results

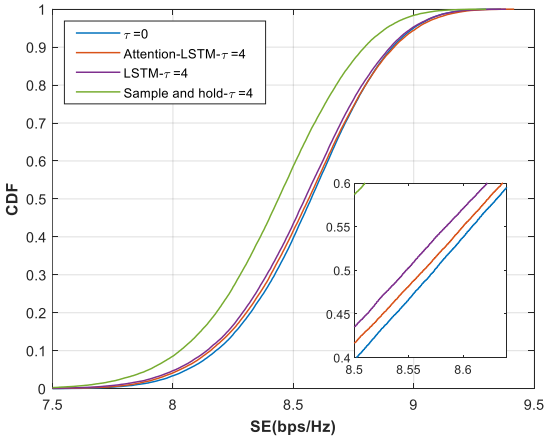
The validation results for KPI#1, that is, the obtained CDF of the SE, is depicted in Figure 25. Subfigures (a) and (b) show the CDFs of the average and per-user SEs respectively under different levels of outdating of the CSI for the proposed enabler and B#1. It can be observed that the proposed method (DNN) is quite more robust to CSI outdating than the baseline. Subfigures (c) and (d) evaluate different predictors that can be used in combination with the proposed method. Fron these results, we can see that the average SE has a median value of 8.7 b/s/Hz, slightly above target, whereas the per-user SE CDF has a value of 5.8 at 10^{-3}, which fall slightly short from the target.



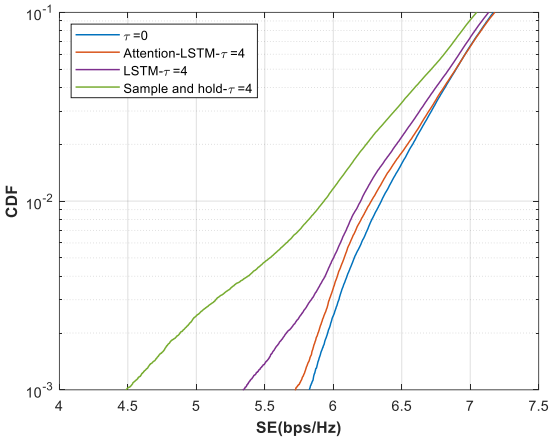
(a): CDF of the average SE across all subnetworks, comparing the proposed DNN-based RRM and the benchmark under varying delay conditions.



(b): CDF of the individual SE across all subnetworks, comparing the proposed DNN-based RRM and the benchmark under varying delay conditions.



(c): CDF of the average SE across all subnetworks for DNN-based RRM with a $(\tau = 4)$-sample delay, comparing sample-and-hold, Attention-LSTM, and LSTM predictors.



(d): CDF of the SE for all subnetworks for DNN-based RRM with a $(\tau = 4)$-sample delay, comparing sample-and-hold, Attention-LSTM, and LSTM predictors.

Figure 25: Validation results of Joint Sub-band Allocation and Power Control for Outdated CSI Scenarios for KPI#1

To validate KPI#2, the training dynamics of the proposed DNNs are represented in Figure 26. Two different DNN architectures are evaluated: a plain LSTM network, and an LSTM network with dual attention. Of these, the network with attention achieves the lower loss, exceeding the target value of 10^{-3} and reaching $0.6 \cdot 10^{-3}$.

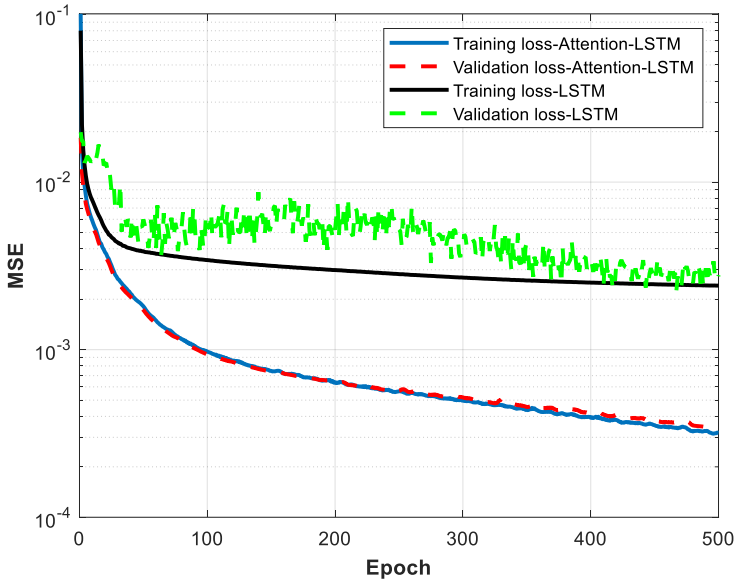


Figure 26: Validation results of Joint Sub-band Allocation and Power Control for Outdated CSI Scenarios for KPI#2

3.3.8 Learning-based HARQ

Moving on to a different topic, we now present the validation results for a learning-based scheme developed to predict the result of the decoding process of a packet, hence being able to request early retransmissions before the channel decoding has finalized.

The outcome and conditions of the validation of this enabler are summarized in Table 25, with the details provided below.

Table 25: Validation Summary for Learning-based HARQ

Learning-based HARQ (InterDigital)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Mathematical analysis	B#1: Legacy HARQ	KPI#1: Retransmission latency	50% decrease versus B#1 and B#2.	Achieved when prediction accuracy larger than 0.8
	B#2: Proactive HARQ	KPI#2: Retransmission overhead	0.2 retransmissions	Achieved when prediction accuracy larger than 0.85

3.3.8.1 Validation Assumptions

The method is validated by using mathematical analysis. For the analysis, 5G NR numerology and frame structure are assumed, with a subcarrier spacing of 60 KHz (i.e., $\mu = 3$). In addition, it is assumed that the packet-error rate of initial transmissions is 10%.

As baselines, two methods detailed in 3GPP standard are utilized:

- **B#1:** as first baseline, the legacy HARQ process in [27] is used.
- **B#2:** the second baseline is the proactive HARQ method in [27].

The enabler is evaluated using two KPIs:

- **KPI#1:** the first KPI is the retransmission latency in round-trip time (RTT), measured in slots. The target is to reach at least 50% decrease with respect to the baselines.
- **KPI#2:** the second KPI used is the retransmission overhead, measured as the average number of retransmissions per original packet. The target for KPI#2 is to reach an average number of retransmissions of 0.2 for high prediction accuracies.

3.3.8.2 Validation Results

The validation results of the method are presented in Figure 27, with the left-hand figure relating to KPI#1 and the right-hand figure dealing with KPI#2. Starting with KPI#1, we see that the proposed method (eHARQ) always reduces latency in comparison with respect to B#1 and B#2. The reduction amounts to 55% and 50% respectively for a prediction accuracy of 0.8.

For KPI#2, we see that the average number of retransmissions ranges from around 0.5 to 0.1 as the prediction accuracy increases towards perfect prediction. The target value is reached whenever the prediction accuracy is, at least, 0.85.

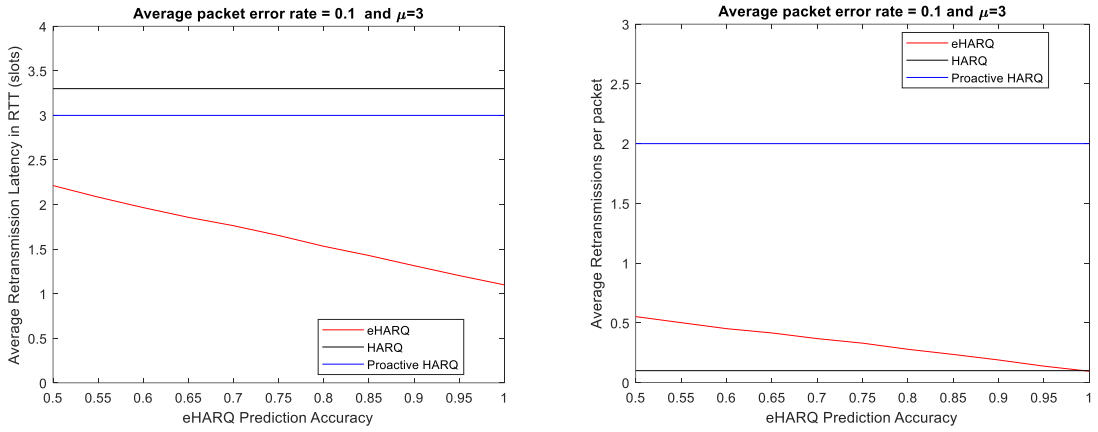


Figure 27: Validation results for Learning-based HARQ

3.3.9 Probabilistic Time Series Conformal Risk Prediction

This enabler uses probabilistic time-series conformal risk prediction techniques tested under channel prediction tasks, applied to the problems of power control under interference constraints and energy-efficiency hybrid automatic repeat request with incremental redundancy under decoding constraints. The method has been documented and analysed in [28].

The outcome and conditions of the validation of this enabler are summarized in Table 26, with the details provided below.

Table 26: Validation Summary for Probabilistic Time Series Conformal Risk Prediction

Probabilistic Time Series Conformal Risk Prediction (KCL)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
Link-level simulations	B#1: Time-Serie Conformal Prediction [29]	KPI#1: Prediction efficiency	10% improvement against B#1	Up to 25% improvement

	B#2: Model-predictive control based on Time-Serie Conformal Prediction [30]	KPI#2: Delay, decoding probability, throughput, and energy efficiency	10% improvement over B#2	25%, 15%, 20% and 25%, respectively, for each of the metrics in KPI#2.

3.3.9.1 Validation Assumptions

Validation is performed through simulations of propagation conditions in Marienhof Square, Munich, using the Sionna Ray-Tracing software [31]. The system is assumed to comprise a fixed single-antenna transmitter, multiple moving receivers with pre-defined mobility patterns and random blockages. Communication occurs at a center frequency of 2.14 GHz with a bandwidth of 120 kHz.

Two AI-based baselines are used to benchmark the method's performance:

- **B#1:** Time-Serie Conformal Prediction, a method proposed in [29]. This baseline is used to validate the prediction accuracy of the method.
- **B#2:** Model-predictive control based on Time-Serie Conformal Prediction, as proposed in [30]. This baseline is used to benchmark the performance of the control tasks based on the prediction method.

As evaluation metrics, two KPIs are considered:

- **KPI#1:** the first KPI is the prediction efficiency, measured using the prediction set size. The target for this KPI is to improve 10% over B#1.
- **KPI#2:** the second KPI is, in fact, a set of KPIs that should fulfil a joint target. The KPIs involved are: delay, decoding probability, throughput, and energy efficiency. The target is to improve (decrease for the first one, increase for all the rest) each of them 10% with respect to B#2.

3.3.9.2 Validation Results

The validation results for KPI#1 are represented in Figure 28. There, the proposed schemes, PTS-CRC and E-PTS-CRC, are benchmarked against B#1 for the task of channel prediction. All methods return prediction regions that cover the true channel evolution with a probability larger than 1-α. However, the proposed methods returns prediction sets that are in average more efficient (smaller). Across various miscoverage requirements α, the proposed method PTS-CRC reduces the inefficiency up to 25% while ensuring the target miscoverage rate.

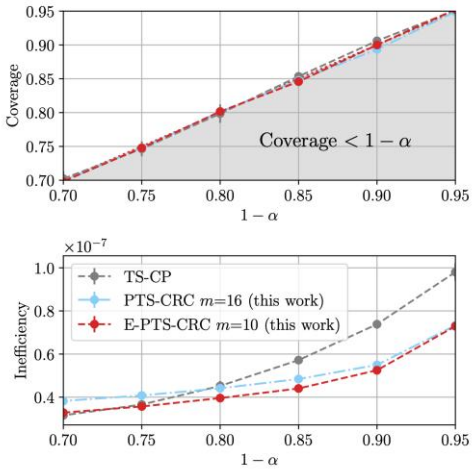
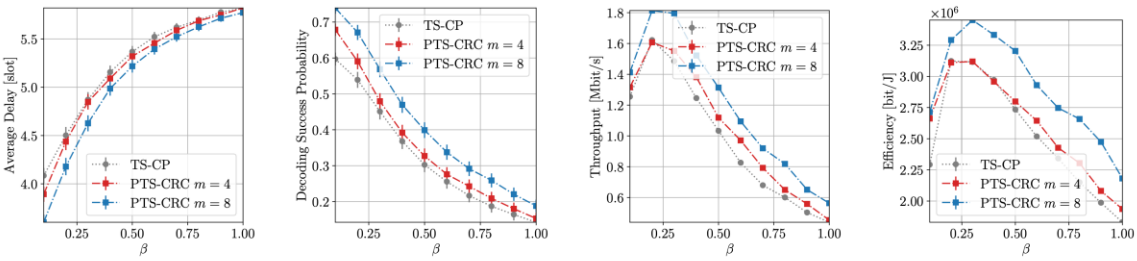


Figure 28: Validation results of Probabilistic Time Series Conformal Risk Prediction for KPI#1

The results for KPI#2 are showcased in Figure 29. The HARQ-IR schemes are tested based on closed-loop model predictive control (MPC). The MPC is solved using predictions from the baseline and the proposed method. The performance metrics of the HARQ-IR algorithm obtained by solving the MPC problem using TS-CP (B#1) and the proposed scheme, PTS-CRC, are depicted. The parameter β determines the target rate of the HARQ-IR scheme, where a larger β corresponds to a higher target rate. We observe a significant increase in the energy efficiency of the HARQ-IR protocol when the MPC algorithm uses predictions from the proposed model. Specifically, we achieve up to a 1-slot delay reduction, a 15% increase in decoding probability, a 20% throughput gain, and a 25% improvement in energy efficiency.



(a) Average Delay (b) Decoding Probability (c) Throughput (d) Energy Efficiency

Figure 29: Validation results of Probabilistic Time Series Conformal Risk Prediction for KPI#2

3.3.10 EMF Reduction via AI-enabled Cell-free Networking

To finalize the presentation of validation results, we now treat the case of the enabler for electromagnetic field (EMF) exposure reduction by exploitation of the cell-free paradigm. Specifically, a method called Cluster-then-Match (CtM), an efficient algorithm making joint decisions about user assignment and power levels of access points, is evaluated. The method has been described in detail in [32].

The outcome and conditions of the validation of this enabler are summarized in Table 27, with the details provided below.

Table 27: Validation Summary for EMF Reduction via AI-enabled Cell-free Networking

EMF Reduction via AI-enabled Cell-free Networking (CNR)				
Validation approach	**Baselines**	**KPIs**	**Targets**	**Achieved Result**
System-level simulations	B#1: Max Rate	KPI#1: whole body SAR (SARwb)	Not exceed 0.08 W/kg	Orders of magnitude below target

3.3.10.1 Validation Assumptions

The EMF exposure was assessed by estimating the specific absorption rate (SAR). More specifically, SAR was evaluated over the whole body, (SARwb), which is the power of the EMF absorbed over the entire body mass, by making use of computational approaches applied on different human models. SAR was chosen since, for high frequencies environments, it is listed as one of the main basic restrictions that are the actual exposure limits in the guidelines. Basic restrictions, indeed, are based on verified health effects and provided as internal physical quantities (as SAR in case of high frequencies). Moreover, since SAR is related to the body mass, it allows to account for the anatomical variability of human bodies.

The performance of the proposed cell-free networking method was validated in indoors factory scenarios as standardized in 3GPP TR 38.901, which includes non-human network users and non-user humans. In this context, high quality decisions are needed about (i) which access point to use when serving an end user and (ii) how to manage access points, e.g., how to set their power levels. Following the human-centric networking paradigm, such decisions account not only for the performance of the network, but also for the level of electromagnetic field exposure to which human bodies incur and energy consumption.

One baseline is defined to validate the method:

- **B#1:** the selected baseline consists of a traditional (non-AI) optimization method seeking to maximize the sum-rate of the system.

A single KPI is used to validate the enabler:

- **KPI#1:** as KPI, the specific absoroption rate (SAR) is employed. The target for it is the value set in the guidelines for the ICNIRP, which corresponds to a SARwb of 0.08 W/kg for general public.

3.3.10.2 Validation Results

The validation results for the proposed CtM method are depicted in Figure 30 and Figure 31. It is observed in Figure 30 that the distribution of SARwb obtained by CtM is significantly lower than that of B#1 and several orders of magnitude below the target (ICNIRP limit). In Figure 31, The black marker on the color bar corresponds to the ICNIRP limit. Square and triangle markers correspond (resp.) to humans associated with the Duke and Ella models. Red stars represent access points.

As conclusion of these results, we remark that in all the conditions tested and for all the human model considered, SARwb levels resulted significantly below the limit values recommended (i.e. the target one), with at least one order of magnitude of difference.

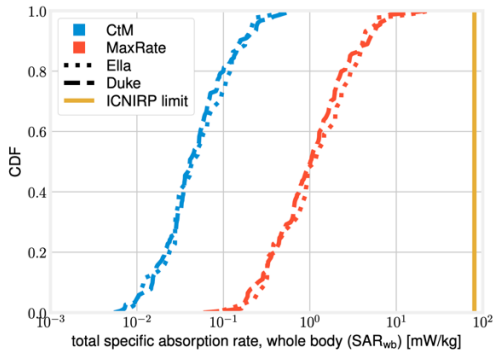


Figure 30: CtM and the MaxRate benchmark: SARwb values

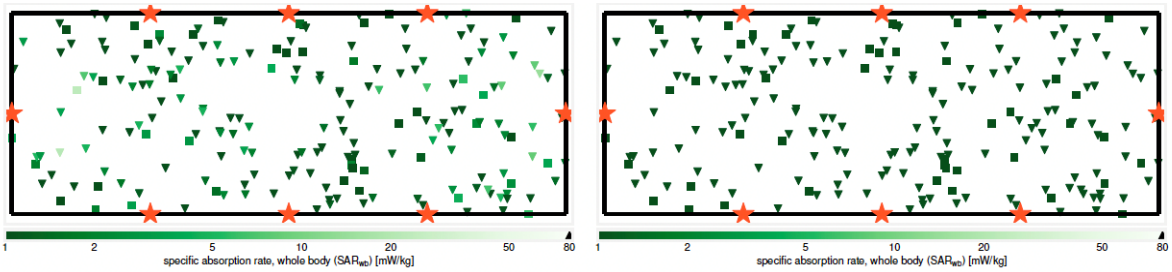


Figure 31: SARwb experienced by different humans under the MaxRate (left) and CtM (right) strategies.

4 Conclusions

This deliverable has presented the validation results of the enablers developed in CENTRIC WPs 2, 3 and 4. We have restricted the presentation to include only validation by simulation results and theoretical analysis, as validation via proof-of-concept implementation will be reported in D5.4, due end of April 2025.

Overall, 20 AI-based technological enablers have been validated, comprising the evaluation of a total of 37 KPIs. When evaluating the assessment of the KPIs, we observe a large proportion of enablers that have achieved the set targets. While the validation of the enablers has been done in isolation of each other, and integration of multiple components is required in order to faithfully assess the potential of AI to the air interface of 6G, these results constitute an encouraging prospect for the disruptive capabilities of AI techniques. We hope that follow-up projects in the upcoming phases of SNS-JU, with a higher target in terms of TRL, will be able to overtake the task of integrating some of the developed enablers in a complete prototype of 6G systems.

We conclude by remarking that, while extensive, the list of enablers included in this deliverable is not exhaustive of all the work done in the project. There are still several months to go until the conclusion of CENTRIC, and we direct the attention of the interested reader towards the publication of final deliverables of each of the WPs in the project, where further techniques that were not ready at the point of publication of this deliverable will be included.

References

[1] R. Jain, D.-M. Chiu and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems, DEC Research Report TR-301," 1984.

[2] M. Zecchin, K. Yu and O. Simeone, "In-Context Learning for MIMO Equalization Using Transformer-Based Sequence Models," in *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, Denver, CO, USA, 2024.

[3] M. Zecchin, K. Yu and O. Simeone, "Cell-Free Multi-User MIMO Equalization via In-Context Learning," in *2024 IEEE 25th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Lucca, Italy, 2024, 2024.

[4] E. Björnson and L. Sanguinetti, "Making Cell-Free Massive MIMO Competitive With MMSE Processing and Centralized Implementation," *IEEE Transactions on Wireless Communications,* vol. 19, no. 1, pp. 77-90, 2020.

[5] C. Finn, P. Abbeel and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *34th International Conference on Machine Learning - Volume 70 (ICML'17)*, 2017.

[6] [Online]. Available: https://nvlabs.github.io/sionna/examples/Neural_Receiver.html.

[7] 3GPP, "TS 138 214 - V16.2.0 - 5G; NR; Physical layer procedures for data (3GPP TS 38.2014 version 16.2.0 Release 16)," 2020.

[8] H. Xiao, Z. Wang, W. Tian, X. Liu, W. Liu, S. Jin, J. Shen, Z. Zhang and N. Yang, "AI Enlightens Wireless Communication: Analyses, Solutions and Opportunities on CSI Feedback," *China Communications,* vol. 18, no. 11, pp. 243-256, 2021.

[9] InterDigital, Inc., "R1-2307251, Evaluation on AI/ML for CSI feedback enhancement, 3GPP TSG-RAN WG1 #114".

[10] InterDigital, Inc., "R1-2406501, Discussion on AI/ML-based CSI prediction, 3GPP TSG-RAN WG1 #118".

[11] 3GPP, "TR38.843 V18.0.0 (2023-12) Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface (Release 18)," 2023.

[12] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European transactions on telecommunications,* vol. 10, no. 6, pp. 585-595, 1999.

[13] S. Yang and L. Hanzo, "Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs," *IEEE Communications Surveys & Tutorials,* vol. 17, no. 4, pp. 1941-1988, 2015.

[14] C. J. Vaca-Rubio, C. Navarro Manchón, R. Adeogun and P. Popovski, "Proximal Policy Optimization for Integrated Sensing and Communication in mmWave Systems," *arXiv preprint arXiv:2306.15429.,* 2023.

[15] [Online]. Available: https://nvlabs.github.io/sionna/examples/5G_NR_PUSCH.html#.

[16] B. Liu, A. Valcarce and K. Srinath, "A lossless compression technique for the downlink control information message," in *IEEE Int. Workshop Signal Proc. Adv. Wireless Commun. (SPAWC)*, 2024.

[17] D. Huffman, "A Method for the Construction of Minimum-Redundancy Codes," in *Proceedings of the IRE*, 1952.

[18] M. Goyal, K. Tatwawadi, S. Chandak and I. Ochoa, "Deepzip: Lossless data compression using recurrent neural networks," in *Data Compression Conf. (DCC)*, 2019.

[19] T. Lin, M. Huh, C. Stauffer, S.-N. Lim and P. Isola, "`Learning to ground multi-agent communication with autoencoders," in *Advances in Neural Information Processing Systems 35 (NeurIPS 2021)*, 2021.

[20] A. M. Girgis, A. Valcarce and M. Bennis., *Time-Series JEPA for Predictive Remote Control under Capacity-Limited Networks,* arXiv preprint arXiv:2406.04853, 2024.

[21] C. B. Chaaya, A. M. Girgis and M. Bennis., "Learning Latent Wireless Dynamics from Channel State Information," *IEEE Wireless Communications Letters,* 2024.

[22] B. Madsen and R. Adeogun, "Federated Multi-Agent DRL for Radio Resource Management in Industrial 6G in-X subnetworks," in *2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Valencia, Spain, 2024.

[23] R. Adeogun, G. Berardinelli and P. Mogensen, "Learning to Dynamically Allocate Radio Resources in Mobile 6G in-X Subnetworks," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Helsinki, Finland, 2021.

[24] S. Hakimi, R. Adeogun and G. Berardinelli, "Rate-conforming Sub-band Allocation for In-factory Subnetworks: A Deep Neural Network Approach," in *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, Antwerp, Belgium, 2024.

[25] D. Li, S. R. Khosravirad, T. Tao and P. Baracca, "Advanced Frequency Resource Allocation for Industrial Wireless Control in 6G subnetworks," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, Glasgow, United Kingdom, 2023.

[26] Q. Shi, M. Razaviyayn, Z. -Q. Luo and C. He, "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Transactions on Signal Processing,* vol. 59, no. 9, pp. 4331-4340, 2011.

[27] 3GPP, "3GPP TS 38.321, Medium Access Control (MAC) protocol specification".

[28] M. Zecchin, S. Park and O. Simeone, "Forking Uncertainties: Reliable Prediction and Model Predictive Control with Sequence Models via Conformal Risk Control," *IEEE Journal on Selected Areas in Information Theory,* vol. 5, pp. 44-61, 2024.

[29] K. Stankeviciute, A. M. Alaa and M. v. d. Schaar, "Conformal time-series forecasting," *Advances in neural information processing systems,* vol. 34, pp. 6216-6228, 2021 .

[30] L. Lindemann, M. Cleaveland, G. Shim and G. Pappas, "Safe planning in dynamic environments using conformal prediction"," *IEEE Robotics and Automation Letters,* vol. 4, 2023 .

[31] [Online]. Available: https://nvlabs.github.io/sionna/api/rt.html.

[32] E. Chiaramello, C. Chiasserini, F. Malandrino, A. Nordio, M. Parazzini and A. Valcarce, "Cluster-then-Match: Efficient Management of Human-Centric, Cell-Less 6G Networks," in *25th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMOM 2024*, 2024.