ultra-massive MIMO for future cell-free heterogeneous networks - **MiFuture**

AI procedures and protocols

Deliverable 3.2

Grant Agreement Number: 101119643
Project Acronym: MiFuture
Project Title: ultra-massive MIMO for future cell-free heterogeneous networks
Call: HORIZON-MSCA-2022-DN-01
Type of action: HORIZON TMA MSCA Doctoral Networks - Industrial Doctorates
Granting authority: European Research Executive Agency
Project start date: 01/01/2024

VERSION DETAILS	
Original Delivery Date:	31/12/2025
Actual Delivery Date:	31/12/2025
Version:	V2.0
Number of Pages:	21
Status:	Draft

CONTRIBUTIONS	
Participant	Name
Universidad Carlos III de Madrid (UC3M)	Prof. Ana García Armada
Vodafone Intelligent Solutions (VOIS)	Lucia De Miguel Albertos
Vodafone Intelligent Solutions (VOIS)	Carlos Úbeda Castellanos
Universidad Carlos III de Madrid (UC3M)	Masoumeh Jamshidi
Instituto de Telecomunicações (IT)	Omid Abbasi
Tampere University (TAU)	Mehmet Fatih Ayten

DOCUMENT HISTORY			
Version	Date	Responsible	Description
1.0	14.12.2025	Masoumeh Jamshidi	Initial Version
1.1	18.12.2025	Prof. Ana García Armada	Revised
2.0	19.12.2025	Masoumeh Jamshidi	Revised
2.0	19.12.2025	Prof. Ana García Armada	Revised
2.1	23.12.2025	Masoumeh Jamshidi	Revised

DELIVERY REVIEWS			
Version	Date	Reviewed by	Conclusion*
2.1	29.12.2025	Supervisory Board	Accepted

*e.g. Accepted, Develop, Modify, Re-edit, Update

Contents

Beneficiaries and Partners 4
.1 Introduction 5
.2 AI-Based Channel Estimation 6
.3 AI-Driven Localization and Sensing with Deep Learning 9
3.1 Deep Learning for Integer Ambiguity Resolution 9
3.2 Failure-Tolerant Phase-Only Indoor Positioning 10
3.3 Phase Synchronization Errors and AP Selection 11
3.4 Neural Network Integrated Multistatic Sensing for CF JCAS 12
3.5 Unified AI Framework for Sensing and Localization in 6G 13
4. AI Procedures and Protocols in Open RAN 14
4.1 Open RAN Architecture and Intelligent Control 14
4.2 Non-Real-Time and Near-Real-Time RIC Control Loops 14
4.3 xApps and rApps as AI Execution Units 15
4.4 Distributed AI Execution at the DU/CU Level (dApps) 15
4.5 AI-Based Physical-Layer and MAC Procedures in Open RAN 16
4.6 AI-Driven Positioning and Localization Support in Open RAN 17
4.7 AI-Integrated Multistatic Sensing and CF JCAS 18
4.8 Toward Autonomous and AI-Native Open RAN: Challenges and Outlook 18
5. Conclusions and Outlook 19
References 21

Beneficiaries and Partners

Beneficiaries:

1. UNIVERSIDAD CARLOS III DE MADRID (UC3)
2. INSTITUTO DE TELECOMUNICACOES (IT)
3. TAMPEREEN KORKEAKOULUSAATIO SR (TAU)
4. LUNDS UNIVERSITET (ULUND)
5. UNIVERSITAT AUTONOMA DE BARCELONA (UAB)
6. VODAFONE INTELLIGENT SOLUTIONS ESPANA SL (VOIS)
7. OY L M ERICSSON AB (ERF)
8. ERICSSON AB (ERS)
9. NOKIA SOLUTIONS AND NETWORKS SP ZOO (NOP)
10. NOKIA SOLUTIONS AND NETWORKS OY (NOF)
11. ALTYS TECHNOLOGIES (ALT)
12. KEYSIGHT TECHNOLOGIES BELGIUM (KEY)

Associated Partners:

1. UNIVERSIDADE NOVA DE LISBOA (UNL)
2. UNIVERSIDADE DE AVEIRO (UAv)
3. UNIVERSIDADE DE COIMBRA (UCo)
4. KATHOLIEKE UNIVERSITEIT LEUVEN (KUL)
5. KEYSIGHT TECHNOLOGIES SPAIN (KEY-SP)

Chapter 1

1. Introduction

Future wireless communication systems are expected to support unprecedented levels of connectivity, reliability, and service diversity, driven by emerging applications such as high-accuracy localization, integrated sensing and communication, and cell-free network architectures. Achieving these requirements using conventional rule-based signal processing and protocol design has become increasingly challenging in highly dynamic, dense, and heterogeneous radio environments. In this context, artificial intelligence (AI) has emerged as a key enabler for redesigning wireless system procedures and protocols toward adaptive and learning-driven operation.

At the physical layer, AI-based channel estimation represents a fundamental procedure that directly impacts beamforming, localization, sensing, and resource allocation. By leveraging data-driven inference, AI-based channel estimation can reduce pilot overhead, improve robustness under mobility, and mitigate model mismatch in complex propagation environments, providing a reliable foundation for higher-level wireless functionalities.

Building upon accurate channel knowledge, phase-only positioning with deep learning enables high-precision localization using phase information as a compact and hardware-efficient feature set. Learning-based phase processing relaxes calibration requirements and enhances robustness against noise and environmental dynamics, extending the role of AI from communication performance optimization to context-aware wireless services.

In parallel, the convergence of communication and sensing has led to growing interest in joint communication and sensing (JCAS) architectures. Neural-network-integrated multistatic sensing for cell-free JCAS systems enables cooperative sensing, distributed data fusion, and intelligent interpretation of multistatic measurements across geographically separated nodes. Such AI-driven sensing procedures are particularly relevant in dense and cell-free deployments, where coordination among multiple access points is essential.

At the network level, the deployment and coordination of these AI-driven procedures are enabled by AI procedures and protocols in Open Radio Access Network (Open RAN) architectures. Open RAN provides disaggregated, programmable, and standardized interfaces that support the integration, orchestration, and lifecycle management of AI models across real-time and non-real-time control loops, enabling scalable, interoperable, and AI-native wireless network operation.

This document presents the initial work of MiFuture Doctoral Candidates and their supervisors in AI-aided channel estimation, localization and sensing, and their feasibility of integration with the ORAN architecture.

Chapter 2

2. AI-Based Channel Estimation

As the number of users and connected devices in the Internet of Things (IoT) continues to rise, and as many applications demand higher data rates and lower latency, the motivation for integrating artificial intelligence (AI) into 6th Generation (6G) networks becomes increasingly compelling. AI can assist in handling highly complex communication tasks, including signal detection, channel estimation, beamforming, sensing, and other sophisticated processes [1]. These capabilities are particularly critical as 6G aims to support ultra-reliable, low-latency, and high-capacity services. One of the central components we aim to enhance in 6G is sensing and channel estimation, especially in applications such as autonomous driving and wearable devices [2]. By leveraging AI, these tasks can be performed more efficiently and accurately, enabling better overall system performance and user experience.

Traditional channel estimation methods, such as Least Squares (LS) and Linear Minimum Mean Square Error (LMMSE), face clear limitations in these scenarios. For example, LS often requires transmitting a larger number of pilot symbols, which reduces the effective data rate. LMMSE, on the other hand, requires that the receiver has prior knowledge of the channel and noise statistics and introduces significant computational complexity, making it less suitable for real-time or large scale IoT systems [2]. These constraints highlight the need for more adaptive and intelligent approaches capable of coping with dynamic wireless environments.

Studies show that even a lightweight one-dimensional convolutional neural network (1-D CNN) used as a post-processing unit after conventional channel estimation can yield significant improvements, such as a 30% reduction in residual carrier-frequency-offset (CFO) error. This evidence motivates the use of AI in channel estimation through several possible approaches [3]:

Deep-learning-based algorithms can dynamically allocate pilot symbols online based on the channel condition and available system resources, thereby reducing overhead while maintaining high performance.

Learning-based methods can adaptively correct offsets by capturing and modeling the noise pattern over time.

End-to-end models can jointly perform channel estimation, offset correction, and angle-of-arrival (AoA) estimation within a single processing block.

AI-based approaches can enable scaling of the system to massive multiple-input multiple-output (MIMO) configurations for multi-user scenarios.

Data-driven models can be trained on real-world measurements to ensure that the proposed methods remain effective and robust in practical deployment scenarios. To this end, many studies have explored the use of machine-learning (ML) and deep-learning (DL) methods for channel estimation. One approach treats the channel matrix as a two-dimensional (2D) image and formulates the estimation task as an image-processing problem [4]. Another line of research applies conventional pilot-based channel estimation and performs interpolation using a Super-Resolution Convolutional Neural Network (SRCNN) [5, 6]. Other methods combine CNNs for initial channel estimation with recurrent neural networks (RNNs) to leverage temporal correlations and predict channel evolution over time [7]. These methods demonstrate that deep learning can provide substantial improvements over traditional techniques, particularly in dynamically varying wireless channels.

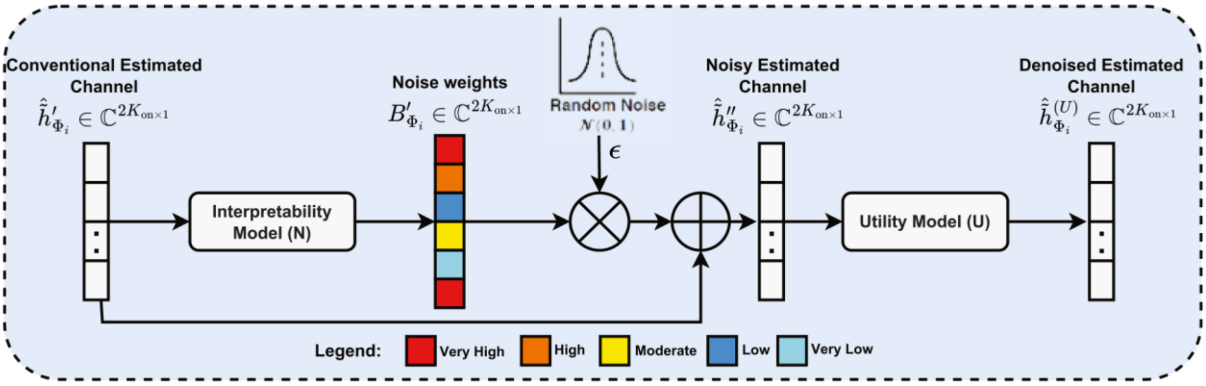

Fig. 1: XAI-based channel estimation method. Model N identifies the relevant and irrelevant subcarriers for channel estimation, guiding the DL-based utility model [8].

However, the blind application of deep-learning models is not fully reliable in the context of wireless communication, as it is crucial to understand how a model makes its decisions. In [2], the authors raise three critical questions that challenge the current use of DL models in wireless systems:

”Is there a way to better select the DL black-box high-dimensional model inputs?”

”Is there a real need for such high-complex architectures?”

”Is there a way to provide interpretability to the decision-making strategy employed DL black box mod- els?”

This is where the use of explainable artificial intelligence (XAI) becomes particularly relevant. The fundamental idea behind XAI is to provide a clear explanation of why a DL model produces a particular prediction. It is not enough to justify the decisions of the model; we must also be able to interpret them reliably. This capability is essential to build trust in AI-based systems, especially in safety-critical and sensitive applications such as AI-assisted remote surgery [2].

Two primary methods are commonly employed to design XAI models: perturbation-based and gradient-based approaches. In the perturbation-based approach, the model is analyzed by systematically modifying and perturbing input features and observing how these perturbations affect the model’s output. In the gradient-based approach, the gradient of the output with respect to the input is backpropagated to assess the importance of each input feature, providing insight into the decision-making process of the network [2].

XAI-trained models can also be classified into two categories: model-agnostic and model-specific. In the model-agnostic case, the XAI training process does not rely on the internal structure or weights of the original network. In contrast, model-specific approaches build the interpretability mechanism based on the particular architecture, parameters, and learned weights of the already trained model [2].

In [8], the authors introduced XAI into wireless communications for the first time in the context of channel estimation. They employed a model-agnostic, perturbation-based XAI method to estimate the wireless channel in an orthogonal frequency-division multiplexing (OFDM) system. Their proposed channel-estimation algorithm is illustrated in Fig. 1. The core concept is to train a model, referred to as Model N, to identify which subcarriers are relevant or irrelevant for channel estimation. This model operates alongside a conventional DL-based utility model built using a feedforward neural network (FNN). During training, intentional noise is added to the subcarriers. By minimizing the loss function, the model learns which subcarriers do not contribute to accurate estimation and should therefore be ignored. This approach enhances system performance in terms of bit-error-rate (BER) while also reducing computational complexity, as fewer subcarriers need to be considered for channel estimation.

Chapter 3

3. AI-Driven Localization and Sensing with Deep Learning

This chapter summarizes a set of artificial intelligence (AI) contributions to high-precision positioning and joint communication and sensing (JCAS) in next-generation cell-free/distributed MIMO systems. The attached works focus on (i) carrier phase positioning (CPP) using phase-only measurements, avoiding time-of-arrival requirements, and (ii) neural network (NN)-based multistatic sensing for joint angle-of-arrival (AoA) and angle-of-departure (AoD) estimation in cell-free JCAS. Across the four papers, deep neural networks are used to:

- resolve integer ambiguities inherent to phase-only positioning with much lower complexity than maximum-likelihood estimation (MLE),
- provide failure-tolerant positioning in the presence of antenna faults,
- maintain centimeter-level accuracy under phase synchronization errors via NN-based access point (AP) selection,
- enable scalable multistatic sensing for joint AoA/AoD estimation from communication signals.

3.1 Deep Learning for Integer Ambiguity Resolution

In distributed CPP, APs observe only carrier phases of the user equipment (UE) signal (Fig. 2). The corresponding range information is ambiguous by integer multiples of the wavelength. Classical MLE resolves these ambiguities by a dense grid search over the deployment region, which is computationally prohibitive. To address this challenge, two deep learning approaches are proposed. A multi-layer perceptron (MLP) directly maps differential phase measurements to the 2D UE position. A second, hybrid architecture first estimates integer ambiguity labels using a neural classifier and then refines the position using a convolutional neural network (CNN) driven by both phase measurements and the predicted ambiguities.

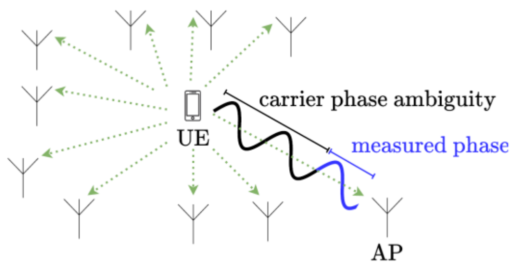

Fig. 2: Uplink positioning with distributed APs where only the carrier phase measurements at different APs are used to estimate the UE position.

Both models are trained in supervised fashion using synthetically generated measurements over random UE positions. Mean-squared error loss is used for regression, and sparse categorical cross-entropy is used for ambiguity classification. Network pruning is applied after initial training to remove low-magnitude weights and further reduce inference complexity.

In a scenario with 20 APs and carrier frequencies of 800 MHz and 1.8 GHz, the CNN-based approach achieves centimeter-level accuracy while reducing complexity by two to three orders of magnitude compared with MLE. As shown in Fig. 3, typical 95th-percentile positioning errors are on the order of 2–3 cm at a transmit power of 0 dBm, illustrating that phase-only CPP with deep learning can match or surpass the accuracy of more traditional time-based methods, without requiring clock synchronization.

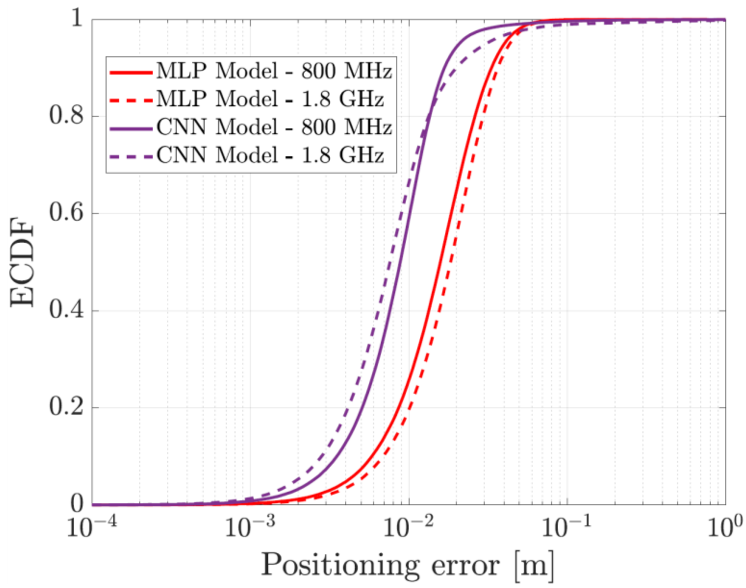

Fig. 3: ECDF of the proposed NN at a transmit power of 0 dBm.

3.2 *Failure-Tolerant Phase-Only Indoor Positioning*

A second contribution extends phase-only CPP to indoor deployments with hardware impairments, in particular antenna element failures at the APs. In this case, differential phase measurements are affected both by noise and by the failure pattern, which can significantly degrade position estimates if not handled explicitly. The proposed architecture combines a neural differential ambiguity estimator with a gradient descent based geometric solver. The estimator is implemented as an MLP with a shared trunk and multiple parallel output branches, each branch predicting the probability distribution over integer ambiguities associated with a particular AP pair. The most likely ambiguity labels are then fed into a hyperbola-intersection based solver that refines the UE position by minimizing a quadratic cost function between estimated and geometry-induced differential distances. Training data covers both fault-free and failure scenarios, with different numbers of failing APs and multiple transmit power levels. This protocol allows the network to learn robust features that generalize across hardware conditions. As shown in Fig. 4, the resulting scheme achieves sub-centimeter to low-centimeter accuracy in a realistic indoor layout, even when several APs are in failure. Reported 95th-percentile positioning errors remain below approximately 2 cm over a wide range of transmit powers and AP failure probabilities. At the same time, the overall FLOP count is reduced by roughly 20–30% compared with earlier NN-based baselines, while being orders of magnitude lower than MLE. A threshold test on the final cost value also enables reliable AP-failure detection, providing a simple integrity monitoring mechanism.

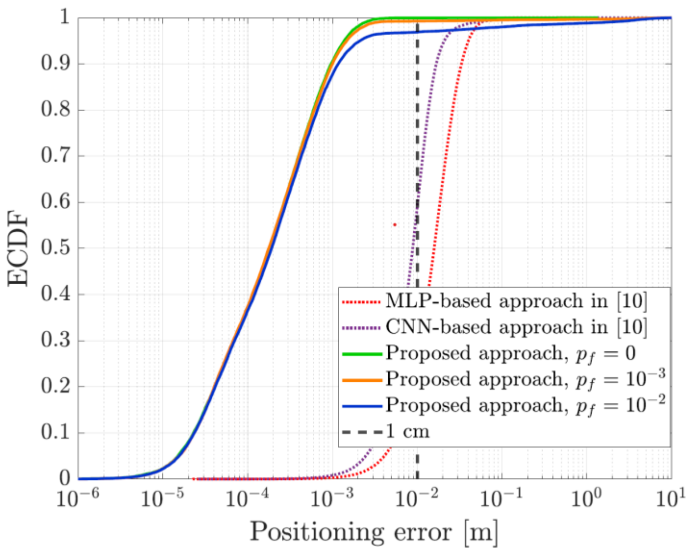

Fig. 4: ECDF of the proposed approach and NN approaches in previous work at a transmit power of 0 dBm. The previous work benchmark performance assumes failure-free data (pf = 0).

3.3 Phase Synchronization Errors and AP Selection

A third study investigates phase-only positioning under phase synchronization errors between the APs and a central location management function (Fig. 5). Random phase perturbations are added to the ideal measurements, representing practical synchronization imperfections.

The underlying hyperbola-intersection method remains the core positioning engine, but an additional MLP is introduced to perform intelligent AP selection. Instead of using all ambiguity combinations, the network selects an ambiguity pair that is predicted to yield the lowest positioning error. The input features include both measurement-based quantities (differential phases and per-AP SNRs) and geometry-based quantities (inter-AP distances and angular relations). The output of the MLP is an estimated error metric for each candidate pair, and the pair with minimum predicted error is chosen.

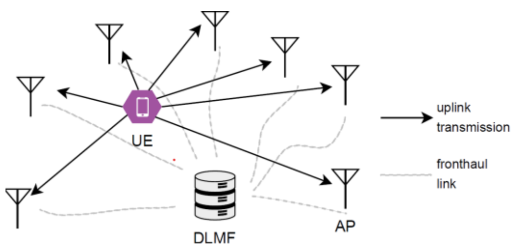

Fig. 5: Uplink UE positioning in a distributed AP deployment, where only phase measurements are taken into account for positioning.

The AP selection network is trained using labeled data where ground-truth errors are computed offline for all ambiguity pairs by a reference solver. Once trained, the selection can be performed with a single forward pass.

Simulation results show that this AI-based AP selection yields 95th-percentile errors in the sub centimeter to sub-half-centimeter range, even in the presence of significant phase perturbations. Compared to a baseline that uses all ambiguities, the proposed selection strategy reduces FLOP count by around 20% while improving high-percentile error performance. In contrast, heuristic selection rules such as random choice or maximum-SNR choice suffer from large error floors and occasional catastrophic failures (Fig. 6).

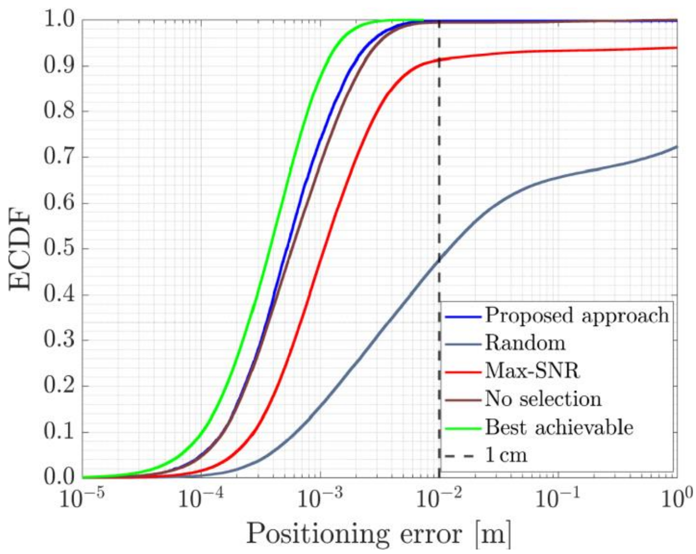

Fig. 6: ECDF of the proposed approach and benchmarks.

3.4 Neural Network Integrated Multistatic Sensing for CF JCAS

The fourth work moves from pure positioning to multistatic radar sensing in a cell-free JCAS setting. A set of distributed transmit APs and receive APs, each equipped with uniform linear arrays, simultaneously provide communication and sensing functionalities. Communication waveforms, such as OFDM signals, are reused for sensing stationary targets via the reflections captured at the receive APs.

The sensing pipeline proceeds as follows. First, least-squares channel estimation is performed on the received communication signals to obtain frequency-domain channel matrices over subcarriers and time. These matrices are then rearranged and reduced in dimension using a coarse timing estimation step, which focuses on delay taps where target echoes are expected. This reduction is important to keep the NN input size manageable.

An MLP is then trained to map the reduced channel representation to the concatenated AoA and AoD values of all targets in the scene. The network is trained with mean-squared error loss over a dataset that spans a range of SNR values, numbers of APs, and target positions.

The NN-based estimator is benchmarked against an MLE solution that jointly searches over AoA, AoD, and delay parameters. While MLE provides a performance upper bound, it is computationally expensive and scales poorly with the number of targets and APs. In contrast, the NN approach achieves similar root-mean-squared error across a wide SNR range for single target scenarios, with a single forward pass. For multi-target cases, both methods see an increase in error due to overlapping echoes, but the NN retains competitive performance and substantially lower complexity. These results demonstrate that NN-based multistatic sensing is a promising building block for practical CF JCAS implementations.

3.5 Unified AI Framework for Sensing and Localization in 6G

The four works collectively establish a consistent set of AI procedures and protocols for sensing and localization in future 6G networks:

- Deep architectures (MLPs, CNNs, and hybrid NN plus optimization schemes) are tailored to the structure of phase and channel measurements, enabling direct regression of positions and angles, as well as classification of integer ambiguities.
- Training protocols explicitly include hardware and synchronization impairments, as well as a wide range of SNRs, to ensure robustness beyond idealized conditions.
- Complexity is treated as a primary design constraint. Network pruning, feature reduction (via coarse timing), and AP selection are used to reduce FLOP counts by factors ranging from tens of percent (against prior NN baselines) to several orders of magnitude (against exhaustive MLE), while keeping centimeter-level accuracy.
- The learned models are not only estimators but also decision tools, supporting AP selection and AP-failure detection, which are essential for reliable and efficient operation in largescale cell-free deployments.

Overall, the body of work demonstrates that carefully designed AI/ML solutions can unlock high-precision, phase-only positioning and scalable multistatic sensing in realistic 6G-oriented cell-free and distributed MIMO systems.

Chapter 4

4. AI Procedures and Protocols in Open RAN

4.1 Open RAN Architecture and Intelligent Control

Open Radio Access Network (Open RAN) is based on the principles of functional disaggregation, open interfaces, and software-defined control, enabling interoperability and multi-vendor deployments. The architecture decomposes the base station into Radio Units (RU), Distributed Units (DU), and Central Units (CU), while abstracting control and monitoring functionalities into the RAN Intelligent Controller (RIC). This separation provides the structural foundation for embedding artificial intelligence into radio access network procedures.

A defining feature of Open RAN is the introduction of native data-driven intelligence through the RIC concept, which allows AI-based procedures to operate independently of vendor-specific baseband implementations. This architectural openness directly supports the applicability of machine learning across multiple layers of the RAN, from physical-layer processing to network-level automation.

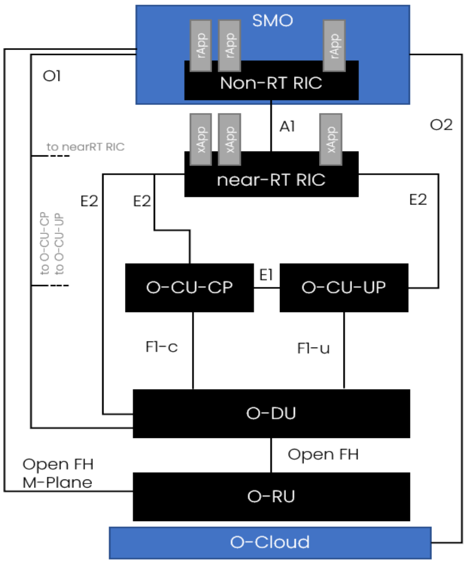

Fig. 7: Open RAN Architecture

4.2 Non-Real-Time and Near-Real-Time RIC Control Loops

The RIC is logically split into the Non-Real-Time (Non-RT) RIC and the Near-Real-Time (Near-RT) RIC, each supporting AI-driven procedures at different time scales. The Non-RT RIC operates on time scales larger than one second and focuses on long-term optimization, policy management, analytics, and AI/ML model training. It hosts rApps that generate policies and enrichment information based on aggregated RAN data and external inputs, and distributes trained models to the Near-RT RIC via the A1 interface.

The Near-RT RIC operates on time scales between tens of milliseconds and one second and enables near-real-time control of RAN elements through xApps. These xApps consume measurements exposed by RAN nodes and execute AI-based inference to drive control actions via the E2 interface, enabling adaptive radio resource management under dynamic channel and traffic conditions.

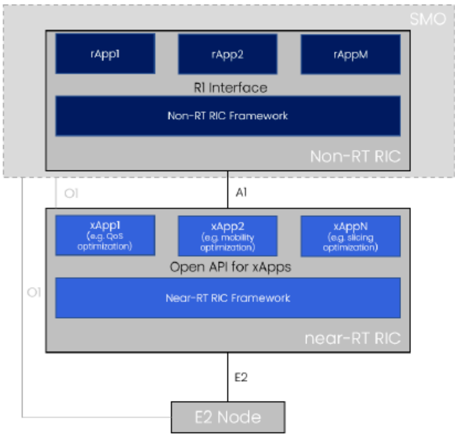

Fig. 8: Near-RT RIC and Non-RT RIC

4.3 xApps and rApps as AI Execution Units

AI procedures in Open RAN are implemented through modular applications known as xApps and rApps. xApps reside within the Near-RT RIC and execute latency-sensitive AI inference tasks, such as adaptive resource allocation, interference mitigation, and real-time parameter tuning. These applications consume RAN measurements exposed through standardized interfaces and translate AI outputs into actionable control commands.

rApps operate within the non-RT RIC and are responsible for computationally intensive tasks such as model training, policy optimization, and performance analytics. rApps leverage large-scale RAN data and external information sources to continuously improve AI models and system policies. The interaction between rApps and xApps establishes a closed loop learning and control framework that supports continuous adaptation and optimization.

4.4 Distributed *AI Execution at the DU/CU Level (dApps)*

While the Non-Real-Time and Near-Real-Time RAN Intelligent Controllers (RICs) enable AI-driven control at time scales above several milliseconds, a growing class of AI-based radio functions requires ultra-low-latency execution and direct access to user-plane data. Such functions include real-time scheduling, link adaptation, beamforming and beam management, spectrum sensing, and AI-assisted localization, whose performance is fundamentally constrained when implemented exclusively through RIC-based control loops.

To address these requirements, the concept of distributed applications (dApps) has been proposed as an architectural extension to Open RAN. dApps are lightweight, cloud-native microservices deployed directly at the Distributed Unit (DU) and Central Unit (CU), where strict latency constraints and data locality prevent effective execution through the RIC alone. By operating close to the physical and MAC layers, dApps enable closed-loop AI-based control with sub-10 ms and even sub-millisecond response times.

In contrast to xApps and rApps, which primarily rely on control-plane information exchanged via the E2 interface, dApps can directly access fine-grained PHY/MAC measurements and user-plane data. Transferring such data to the RIC is often impractical due to bandwidth, latency, and privacy considerations. Local execution of dApps at the DU/CU therefore provides a scalable and efficient mechanism for real-time AI inference and control.

From a functional perspective, dApps extend AI applicability to the lowest layers of the RAN protocol stack. Representative use cases include AI-driven scheduling and link adaptation, beamforming and beam management in massive MIMO systems, real-time spectrum sensing and sharing, and low-latency positioning and sensing for integrated communication and sensing (ISAC). In this sense, dApps complement xApps and rApps within a hierarchical AI control framework, where intelligence is distributed across multiple time scales and architectural layers.

Architecturally, dApps interact with RAN functions through high-speed, service-based interfaces that enable subscription to telemetry streams, access to user-plane data, and delivery of control actions without disrupting ongoing RAN operation. Furthermore, cooperation between dApps and xApps enables coordinated AI-driven decisions across real-time and near-real-time domains, supporting the evolution toward fully AI-native Open RAN architectures.

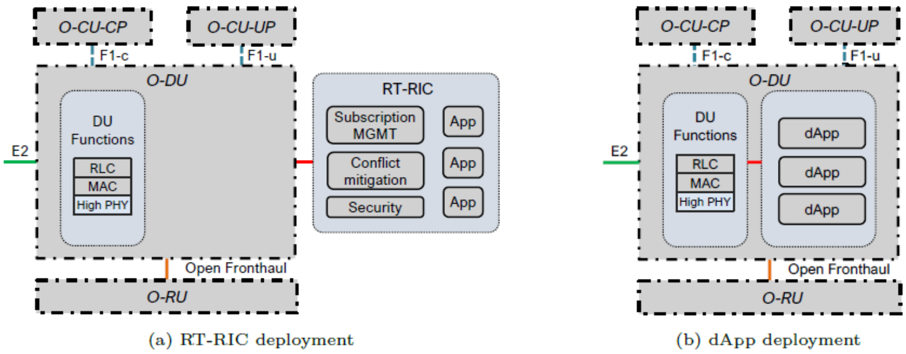

Fig. 9: Comparison between RT RIC and dApp architectures.

4.5 AI-Based Physical-Layer and MAC Procedures in Open RAN

Channel estimation is a core physical-layer function and one of the primary applicability domains of machine learning in next-generation radio access networks. Learning-based techniques can significantly enhance channel estimation performance by improving robustness under user mobility, reducing pilot overhead, and mitigating model mismatch effects in complex propagation environments. Within the Open RAN framework, AI-based channel estimation procedures can be flexibly deployed at different architectural points depending on latency constraints and computational requirements.

Latency-critical inference tasks, such as channel estimation refinement and fast adaptation to short-term channel variations, may be executed close to the Distributed Unit (DU) or Radio Unit (RU), where direct access to low-level physical-layer measurements is available. In contrast, model training, retraining, and long-term adaptation naturally reside in the Non-Real-Time RIC, where large-scale data analytics and cross-cell information can be exploited. Open RAN enables standardized exposure of channel-related measurements and performance indicators, allowing AI models to be continuously trained, validated, and refined using aggregated RAN data.

While AI-based channel estimation represents a fundamental enabling function, the role of machine learning in Open RAN is not limited to estimation tasks alone. In massive MIMO systems, accurate and timely channel knowledge directly feeds into a range of higher-layer and cross-layer decisions, including scheduling, link adaptation, power control, and beam management. As the dimensionality of the channel state increases and multi-user interactions become more complex, rule-based optimization approaches become increasingly inefficient and difficult to scale.

Consequently, similar AI-driven principles can be extended beyond channel estimation to support data-driven scheduling decisions, adaptive modulation and coding selection, and dynamic resource allocation. These AI-assisted PHY- and MAC-layer procedures are particularly relevant in MaMIMO deployments, where real-time coordination among users, beams, and resources is essential to fully exploit spatial multiplexing gains. Within the Open RAN architecture, such functions can be distributed across different execution points, ranging from DU/CU-level real-time inference to RIC-based coordination, policy optimization, and longer-term learning.

By framing channel estimation as part of a broader class of AI-enabled physical-layer and MAC procedures, Open RAN provides a unified and flexible platform for integrating intelligence across multiple layers and time scales. This holistic view is essential for evolving toward AI-native RAN operation, where estimation, adaptation, and control are jointly optimized to meet the performance and scalability requirements of next-generation massive MIMO systems.

4.6 AI-Driven Positioning and Localization Support in Open RAN

Beyond communication-centric optimization, Open RAN enables advanced AI-driven localization and sensing procedures by supporting coordinated operation across multiple RAN nodes. Phase-based positioning and Multistatic sensing rely on joint processing of measurements collected from geographically distributed RUs, requiring tight coordination and synchronization across the network.

AI models deployed within the RIC can perform data fusion, feature extraction, and inference to jointly estimate user locations, environmental characteristics, and target parameters. In cell-free and JCAS architectures, Open RAN provides the necessary orchestration mechanisms to manage sensing resources, coordinate measurement collection, and adapt sensing strategies dynamically based on network conditions.

4.7 AI-Integrated Multistatic Sensing and CF JCAS

The convergence of communication and sensing is strongly supported by the Open RAN paradigm. Neural-network-integrated Multistatic sensing in cell-free joint communication and sensing (CF JCAS) systems relies on cooperative operation across geographically distributed RUs and coordinated processing of sensing and communication data.

Open RAN enables such coordination by exposing sensing-related measurements to the RIC, where AI-based data fusion and interpretation can be performed. AI-driven Multistatic sensing procedures benefit from flexible model placement across Near-RT and Non-RT RICs, enabling adaptive sensing–communication trade-offs and scalable deployment in dense and cell-free environments.

4.8 Toward Autonomous and AI-Native Open RAN: Challenges and Outlook

By combining disaggregated architectures, standardized interfaces, and multi-timescale control loops, Open RAN provides a solid foundation for autonomous RAN operation. Within this framework, AI-driven procedures for channel estimation, positioning, sensing, and radio resource management can be embedded into closed-loop control mechanisms that continuously adapt to time-varying network conditions, traffic dynamics, and propagation environments. This architectural and procedural alignment enables a shift from static configuration toward data-driven, self-optimizing RAN behavior.

At the same time, realizing fully autonomous and AI-native Open RAN operation introduces several open challenges. Strict latency constraints limit the complexity and placement of AI inference, particularly for real-time PHY and control-loop operations. Scalability in ultra-dense and cell-free deployments remains a critical concern due to increased data exchange, coordination overhead, and model management complexity. In addition, data heterogeneity across vendors and deployments, along with privacy and security considerations, complicates the training and deployment of robust AI models.

Ensuring stability, robustness, and interpretability of AI-driven procedures is therefore essential for the practical adoption of autonomous Open RAN. Addressing these challenges will be a key step toward enabling intelligent radio access networks capable of supporting future wireless services with high reliability, adaptability, and operational efficiency.

Chapter 5

5. Conclusions and Outlook

This document presents a set of initial research results on AI-aided physical-layer signal processing, localization and sensing, and AI-enabled network procedures, with a focus on their feasibility and integration in next-generation wireless systems and Open RAN architectures.

At the physical layer, the presented studies demonstrate that AI-based channel estimation can effectively overcome the limitations of conventional rule-based estimators in highly dynamic and heterogeneous radio environments. By leveraging data-driven inference, learning-based approaches reduce pilot overhead, improve robustness under mobility and hardware impairments, and mitigate model mismatch effects. The integration of explainable artificial intelligence (XAI) further strengthens interpretability and trust, which are essential for reliable deployment in large-scale and safety-critical wireless systems.
Beyond channel estimation, these AI-based physical-layer capabilities also act as key enablers for higher-layer and cross-layer decisions in massive MIMO systems, including scheduling, link adaptation, power control, and beam management, where rule-based optimization becomes increasingly inefficient due to high-dimensional channel dynamics.

Building on accurate channel knowledge, deep-learning-based phase-only positioning enables high-precision localization without relying on time-of-arrival measurements or strict clock synchronization. Learning-driven resolution of integer ambiguities, failure-tolerant positioning strategies, and intelligent access point selection allow centimeter- and sub-centimeter-level accuracy to be maintained even in the presence of antenna faults and phase synchronization errors, while significantly reducing computational complexity compared to exhaustive model-based methods.

In parallel, the integration of neural networks into multistatic sensing for cell-free joint communication and sensing (CF JCAS) demonstrates that communication signals can be efficiently reused for sensing purposes in distributed deployments. AI-based multistatic sensing enables joint estimation of angles of arrival and departure with competitive accuracy and substantially reduced complexity, making it a promising building block for scalable and practical CF JCAS implementations.

At the network level, Open RAN architectures provide a flexible and standardized foundation for deploying, orchestrating, and managing AI-driven procedures across multiple time scales. The separation of Non-Real-Time and Near-Real-Time control loops, together with the use of rApps and xApps, enables closed-loop, data-driven RAN operation and supports the practical integration of AI-based physical-layer and sensing functionalities into interoperable and vendor-agnostic networks.
In addition, emerging concepts such as distributed AI execution at the DU/CU level (dApps) highlight the need to complement RIC-based intelligence with ultra-low-latency AI functions operating close to the PHY and MAC layers, enabling real-time scheduling, beam management, and link adaptation in massive MIMO and cell-free deployments.

Looking forward, advancing these results toward operational systems requires continued research on model generalization, robustness under non-stationary environments, and computational efficiency for real-time deployment on resource-constrained devices. Further integration of online learning, continual adaptation, and cross-layer optimization will be critical to fully exploit the potential of AI-native physical-layer processing, localization, and sensing. At the network level, future efforts should focus on scalable model lifecycle management, efficient data collection and distribution, and robust control-loop design under strict latency constraints, including the coordinated operation of AI functions across RIC-based control loops and distributed execution points, while addressing challenges related to interpretability, stability, privacy, and security.

Overall, the combined results and outlook presented in this deliverable indicate that AI-native approaches constitute a fundamental pathway toward autonomous, scalable, and context-aware wireless networks, enabling reliable communication, high-precision localization, and intelligent sensing in next-generation wireless systems, with AI-driven PHY and MAC intelligence jointly supporting both real-time and long-term network optimization.

References

[1] B. Sun, Z.Wang, A. Yang, X. Liu, S. Jin, P. Sun, R. Tamrakar, and D. Jiang, "Ai enlightens wireless communication: Analyses and solutions for dmrs channel estimation," China Communications, vol. 20, no. 5, pp. 275–287, 2023.

[2] A. K. Gizzini, Y. Medjahdi, A. J. Ghandour, and L. Clavier, "Explainable ai for enhancing efficiency of dl-based channel estimation," IEEE Transactions on Machine Learning in Communications and Networking, vol. 3, pp. 976–996, 2025.

[3] A. Aldelemy, A. Al-Dulaimi, G. T. A. El Sanousi, P. O. Siaw, N. T. Ali, V. Doychinov, Y. Dama, and R. A. Abd-Alhameed, "Ai-enhanced pilot-assisted angle-of-arrival estimation for wearable devices in rician fading channels: A low-snr focused method," IEEE Access, vol. 13, pp. 197381–197413, 2025.

[4] J. Yang, C.-K. Wen, S. Jin, and F. Gao, "Beamspace channel estimation in mmwave systems via cosparse image reconstruction technique," IEEE Transactions on Communications, vol. 66, no. 10, pp. 4767–4782, 2018.

[5] M. Soltani, V. Pourahmadi, A. Mirzaei, and H. Sheikhzadeh, "Deep learning-based channel estimation,"IEEE Communications Letters, vol. 23, no. 4, pp. 652–655, 2019.

[6] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295– 307, 2016.

[7] Y. Liao, Y. Hua, X. Dai, H. Yao, and X. Yang, "Chanestnet: A deep learning based channel estimation for high-speed scenarios," in ICC 2019 - 2019 IEEE International Conference on Communications (ICC), pp. 1–6, 2019.

[8] A. K. Gizzini, Y. Medjahdi, A. J. Ghandour, and L. Clavier, "Towards explainable ai for channel estimation in wireless communications," IEEE Transactions on Vehicular Technology, vol. 73, no. 5, pp. 7389–7394, 2024.

[9] Ferrús, R., Sallent, O., Pérez-Romero, J., & Agustí, R. (2019, December). Applicability domains of machine learning in next generation radio access networks. In 2019 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1066-1073). IEEE.

[10] Kliks, A., Dryjanski, M., Ram, V., Wong, L., & Harvey, P. (2023). Towards autonomous open radio access networks. ITU Journal on Future and Evolving Technologies , 4 (2), 251-268.

[11] Hoffmann, M., Janji, S., Samorzewski, A., Kułacz, Ł., Adamczyk, C., Dryjański, M., ... & Bogucka, H. (2023). Open RAN xApps design and evaluation: Lessons learnt and identified challenges. IEEE Journal on Selected Areas in Communications, 42(2), 473-486.

[12] Lacava, A., Bonati, L., Mohamadi, N., Gangula, R., Kaltenberger, F., Johari, P., ... & Melodia, T. (2025). dApps: Enabling real-time AI-based Open RAN control. *Computer Networks*, 111342.