6G AI-driven air interface – Hexa-X-II view

Hamed Farhadi, Bitan Banerjee, Rafael Berkvens, Nabeel Nisar Bhat, Emmanuelle Bodji, Dilin Dampahalage, Eslam Eldeeb, Jeroen Famaey, Gerhard P. Fettweis, Jaeseong Jeong, Dani Korpi, Siddhartha Kumar, Yann Lebrun, Rodolphe Legouable, José Miguel Mateos-Ramos, Nurul Huda Mahmood, Mohammad Hossein Moghaddam, Ahmad Nimr, Nandana Rajatheva, Nuwanthika Rajapaksha, Athanasios Stavridis, Tommy Svensson, Han Yu, Leif Wilhelmsson, Henk Wymeersch

***Abstract*— This paper presents the European 6G Flagship project Hexa-X-II's view on 6G AI-driven air interface. It outlines motivations for AI in the physical layer, selected applications of AI for communication and sensing, their achieved performance, and the challenges to be addressed. The paper also provides an overview of the relevant standardization activities.**

***Index Terms*—6G, artificial intelligence (AI), integrated sensing and communications (ISAC), machine learning (ML).**

I. Introduction

The IMT-2030 framework serving as the sixth generation (6G) wireless system roadmap is being prepared by the International Telecommunication Union – Radio Sector (ITU-R) [1]. Europe's 6G flagship project, Hexa-X, laid the foundation for 6G towards this framework [2]; and its successor, Hexa-X-II, addresses a holistic end-to-end system design for 6G.

Considering the recent advances in artificial intelligence (AI) and machine learning (ML) there is potential for adoption of an *AI-driven air interface (AI-AI)* in 6G [2], [3], [4], [5], [6]. An AI-AI refers to a radio air interface for communication and/or sensing, wherein one or multiple functionalities in the lower layers across the transmitter and/or receiver are replaced by AI-based methods. These methods learn functionalities to enhance performance and adaptability in wireless networks. The vision of a new air interface, partially designed by AI to optimize communication, is outlined in [6]. Exemplary enablers leveraging AI, e.g., AI-assisted mobility, traffic, and radio channel prediction, and their applications are presented in [4], [5]. Hexa-X vision on AI integration in 6G networks with selected examples in the physical layer (PHY) was presented in [3]. Given the vision that 6G radio networks will provide services beyond communications, e.g., sensing [7], the role of AI in PHY should be extended. An overview of deep learning techniques for radar-based sensing and for integrated sensing and communication (ISAC) is provided in [8], and [9], respectively.

Unlike previous works that focus on specific AI applications or narrow use cases in the physical layer [2], [3], [4], [5], [6], [7], this paper presents a comprehensive analysis of the role of AI in the 6G PHY, representing views from major representative stakeholders in the 6G ecosystem, such as network and user equipment (UE) vendors, chipset manufacturers, service providers, and academia. This paper presents motivations for AI-AI, bridges the gap by addressing both communication and ISAC scenarios, offering a dual perspective absent in prior studies Furthermore, this paper incorporates quantitative evaluations of AI methods in PHY, outlines associated challenges, and highlights opportunities for standardization. By integrating Hexa-X-II Key Value Indicators (KVIs), this paper also extends the analysis to include the societal impact of AI, presenting a comprehensive view that aligns with 6G's broader objectives.

II. Motivations for AI-driven air interface

The key drivers towards AI-AI are as follows.

AI-friendly hardware: The recent advancements on dedicated hardware for AI processing for either *inference* or *training* facilitates AI integration in PHY. *Inference-dedicated chipsets* are optimized for real-time and low-power processing and commonly operate with low-precision, e. g. 8-bit or 4-bit [10]. These are optimized to be cost effective and to support high throughput and outperform general purpose hardware by one to two orders of magnitude in terms of energy efficiency (EE). *Training-dedicated chipsets* have high memory bandwidth for data manipulation and high precision, e.g., 32-bit.

***Performance enhancements*:** Employing AI at PHY can enhance performance in terms of reliability and throughput by mitigating non-linear distortions that are not captured by classical model-based techniques. Additionally, an AI-driven approach can perform end-to-end optimization, where multiple functionalities at the PHY across the transmitter and receiver, as well as cross-layer functionalities, can be jointly optimized. This contrasts with classical solutions that focus on the optimization of individual components. The potential performance improvements for communication and sensing are discussed in Section III and Section IV, respectively.

***Flexibility*:** AI capability provides flexibility by autonomous operation, e.g., adaptive resource allocation and signal processing, under changing conditions for traffic, radio propagation, user's behavior or speed [3], [4]. Moreover, the deployed AI-driven methods can be refined by uploading new models, without modifying the hardware or algorithms.

***Scalability*:** The capability to re-train models with new data as

This work was supported by the Hexa-X-II project which has received funding from the Smart Networks and Services Joint Undertaking (SNS JU) under the European Union's Horizon Europe research and innovation programme under Grant Agreement No 101095759.

Hamed Farhadi is with Ericsson Research, Gerhard P. Fettweis, Ahmad Nimr, and Bitan Banerjee are with TU Dresden, Jeroen Famaey, Rafael Berkvens, and Nabeel Nisar Bhat are with Univ. Antwerp and imec, Rodolphe Legouable, and Emmanuelle Bodji are with Orange, Nurul Huda Mahmood, Nandana Rajatheva,, Dilin Dampahalage, Eslam Eldeeb, and Nuwanthika Rajapaksha are with Univ. Oulu, Jaeseong Jeong, Athanasios Stavridis, Leif Wilhelmsson are with Ericsson Research, Dani Korpi is with Nokia Bell Labs, Mohammad Hossein Moghaddam, and Siddhartha Kumar are with Qamcom, Yann Lebrun is with Qualcomm, Henk Wymeersch, Tommy Svensson, José Miguel Mateos-Ramos, and Han Yu are with Chalmers Univ. Hamed Farhadi is the first author and the corresponding author. The other co-authors' names ordered alphabetically.

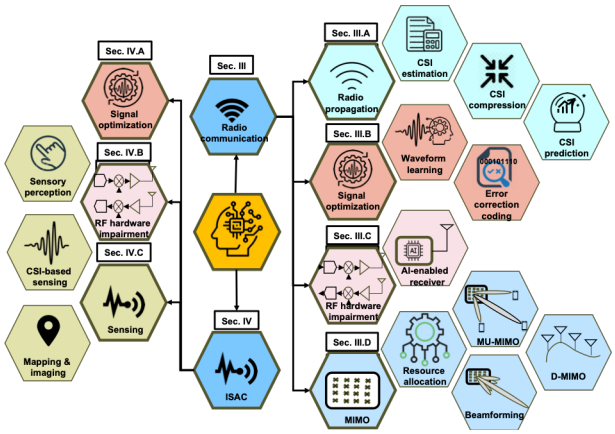


Figure 1 AI-driven air interface techniques for radio communication.

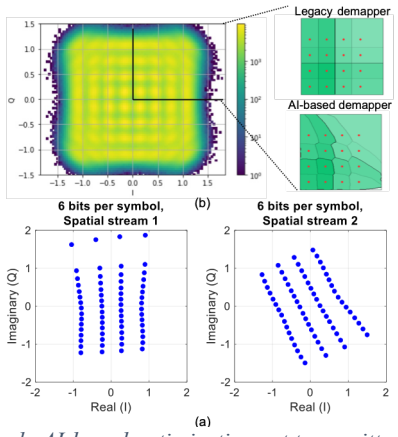


Figure 2 Example AI-based optimizations at transmitter and receiver (a) AI-optimized constellation for pilotless MIMO transmission. (b)AI-based demapper in the presence of hardware impairments.

scenarios change supports scalable solutions. This allows the network to remain scalable, regardless of the increasing number and variety of devices and services it must support. To ensure the scalability of AI-driven methods, model lifecycle management (LCM), i.e., the process for model development, deployment, and management during the entire lifecycle, needs to be supported by the standard. Hence, the 3rd Generation Partnership Project (3GPP) initiated a study on AI/ML for new radio (NR) air interface [13].

III AI-DRIVEN AIR INTERFACE FOR RADIO COMMUNICATION

This section introduces AI-driven techniques for radio communications, as illustrated in Figure 1. The integration of AI-driven methods into the end-to-end radio system is depicted in Figure 3, with methods discussed in detail below. For each proposed method, one or few of the blocks in Figure 3 are performed by the AI/ML-based methods while the other blocks are performed using legacy (i.e. non-AI/ML) methods.

A. CSI acquisition

AI improves channel state information (CSI) acquisition, including estimation, compression, and prediction as follows.

1) CSI estimation

Accurate CSI estimation in massive multi-input multi-output (MIMO) systems presents challenges including a) limited sounding reference signal (SRS) availability, b) high computational complexity of traditional precoding due to matrix inversion [15], and c) hardware impairments degrading CSI estimation. These challenges can be conceptualized as superimposed noise on the ideal CSI. Generative adversarial networks (GANs) can learn the underlying distribution of the data and generate high-quality CSI estimates, even in the presence of noise. A low-complexity least squares estimation with applying conditional GAN on top of that leads to accurate CSI estimation (See Table 1).

2) CSI compression

CSI feedback is a key enabler of multi-antenna techniques. However, as the number of antenna elements increases, CSI overhead becomes significant. AI can learn to efficiently compress and decompress CSI. Hence, reduce the overhead and enhance the accuracy of CSI feedback compared to the existing SRS and codebook-based (e.g., eType2) schemes. However, the AI-based CSI feedback requires training of the UE-side encoder and the base station (BS)-side decoder in a *sequential learning* [14], where BS and UE share training data without sharing the AI model. The encoders and the decoder models are trained, separately. Our results show that a) sequential training enables AI-based CSI feedback without disclosing proprietary AI models and b) a common decoder sequentially trained for multiple UE vendor encoders obtains good performance (see Table 1). Autoencoder (AE)-based CSI compression has been studied for AI-based CSI feedback in 3GPP Release-18 [13]. However, this approach may encounter degradations due to channel aging as AE model struggles to capture wireless channel's time-varying nature.

3) CSI prediction

We introduce an evolutionary CSI neural network (NN), a predictive model designed to learn radio channel dynamics in latent space and predict CSI to addresses the issue of channel aging. This model consists of two components: an AE and a dynamicNet. The AE processes CSI to produce a latent representation, and the dynamicNet learns the temporal evolution of latent vectors over time. This enables recursive multi-step CSI predictions within the latent space which a) minimizes performance degradation due to channel aging, and b) reduces feedback overhead (see Table 1).

B. Signal optimization

AI-driven methods can optimize transmit signal as follows.

1) Waveform learning

We propose a technique to train a MIMO transmitter and receiver jointly to communicate without channel estimation pilots. This enables to perform spatial multiplexing without consuming resources for pilots, thereby improving spectral efficiency (SE). It can be shown that by learning jointly the constellation shapes used by individual spatial streams, illustrated in **Error! Reference source not found.** (a), and a convolutional NN (CNN)-based receiver, it is possible to transmit several spatial MIMO streams without any pilots. This leads to 20% SE gain compared to a legacy pilot-based MIMO system (see Table 1).

2) LDPC code learning

In 5G networks, low-density parity-check (LDPC) codes are selected for handling user data. These codes are expected to be retained and utilized in 6G. While LDPC codes demonstrate excellent performance for large data sizes, their performance

degrades with smaller data sizes. To address this, we introduce

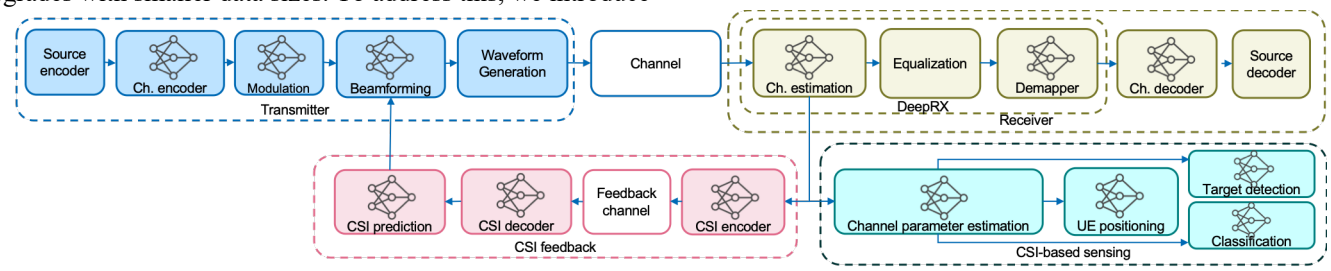


Figure 3 AI-driven air interface techniques for radio transmitters and receiver.

AI techniques to enhance 5G LDPC parity check matrices while preserving their original structure. The system is composed of an NN decoder model, where the weights of the NN represent the coefficients of the original parity check matrix. At the end of the decoding process, we process the results of gradients and losses to preserve the quasi-cyclic structure of LDPC codes. The results confirm gain for small codewords (see Table 1), however, extended work is needed to generalize this method to matrices of other sizes.

3) Semantic communications

Intelligent semantic communication proposes transmission of the meaning of a message instead of the information bits. This reduces the needed resources and lowers traffic. For example, consider a multi-task-oriented semantic communication for connected and autonomous vehicles (CAV), where traffic signs are transmitted from one CAV to another one. At the transmitter, the message can be encoded using a CNN-based semantic encoder. The receiver performs two tasks: a) classification, and b) image reconstruction, where the semantic decoder is composed of CNN layers for image reconstruction followed by a classifier (see Table 1).

C. RF hardware impairment compensation

Radio frequency (RF) hardware is subject to impairments, such as power amplifier (PA) nonlinearity, and oscillator phase noise, which degrade performance. The impact of the RF hardware is expected to be more severe on 6G systems as the need for enhanced EE would require the PAs to operate in nonlinear regime, and transmission at higher bands leads to higher phase noise. We propose an *AI-based digital post distortion (AI-DPoD)* method to compensate distortions at the receiver side as far as the requirements on the out of band emissions are fulfilled. The proposed method relies on an *AI-based demapper*, illustrated in **Error! Reference source not found.** (b), to generate soft bits based on the equalized received signals. Performance evaluations for uplink scenario (see Table 1) confirm 15% throughput gain, 10-45% UE EE enhancement, and 3 dB coverage extension for PA nonlinearity compensation, and 2.6 dB gain in reliability for 64 QAM for phase noise compensation compared to the legacy receivers (see Table 1).

D. MIMO transmissions

Leveraging MIMO capability imposes stringent requirements, e.g., in terms of CSI quality and computational complexity. This is even more demanding in 6G considering the increasing number of antennas. AI-based methods are envisioned to lower the complexity as outlined in the following examples.

1) Beamforming with imperfect CSI

Multi-user MIMO can improve SE; however, its performance would be compromised by CSI inaccuracies at the transmitter. These inaccuracies stem from various sources such as channel aging, estimation error, and the CSI feedback quantization. To mitigate the imperfect CSI's impact, we propose a transformer-based method to compute precoders for a group of users based on CSI estimate and associated uncertainties which leads to enhanced SE (see Table 1).

2) Scalable D-MIMO and radio resource allocation

Distributed MIMO (D-MIMO) leverage on the spatial diversity provided by distributed antennas to enhance reliability and improve coverage. However, varying channel conditions makes pilot assignment for D-MIMO transmissions challenging when the number of antennas scales. We represent it as a graph coloring problem and propose coloring algorithm based on graph NN (GNN) and RL that learns to code on graphs and benefits from parallel computing [14]. Furthermore, we propose a scalable method for access point (AP) selection and power control [14]. This is performed by unsupervised learning using a DNN. The results (see Table 1) confirm that it improves sum rate compared to the full power transmission scenario with significantly lower computational complexity compared to the optimization-based baseline.

IV. AI-DRIVEN AIR INTERFACE FOR ISAC

This section presents the AI-AI techniques, illustrated in Figure 1, for various ISAC scenarios. The motivation for AI-based approaches stems from three distinct sources. First, the fact the ISAC systems are inherently more difficult to design than communication systems, due to the integration of two functions in the same system, i.e., sensing and communication, leading to complex multi-objective optimization. Second, the presence of hardware impairments leads to performance degradations that are more severe for ISAC than for pure communication, which motivates the use of AI-driven mitigation or exploitation strategies. Third, sensing data captured by receivers are a complex nonlinear mapping of the physical environment, people, objects, and their movements. Inverting such mappings for detecting, localizing, or classifying objects or object's status is often beyond the capabilities of model-based signal processing.

A. Signal optimization

The requirements of sensing and communication signal design may be different in terms of modulation format, pulse shaping, or beamforming. Joint optimization approaches are usually non-convex or computationally demanding. AI can be harnessed to optimize existing signals and to design new

signals in space-time-frequency domains under various constraints, e.g., sparse signal design. Given the different objectives of communications and sensing, the loss function

Table 1 AI-driven air interface applications for communications, model specifications, and assessments.

Application	Evaluation assumptions: Carrier freq./BW/# of TX antennas/# of RX antennas/UL or DL/Waveform	Training type	Model type	Training features	Training labels	Training loss function	Model deployment			Model size (Number of parameters)	KPI [gain]	KVI		
							Transmitter	Receiver	CPU			Environmental	Inclusiveness	Trustworthiness
CSI estimation	2GHz/20MHz/128/1/DL/-	S	CGAN	CSI measurements	UL CSI, DL CSI	BCE	✓	×	×	1.6M	SE [~50% / 96%] without / with hardware impairments BM: MMSE precoding with incomplete CSI	✓	×	✓
CSI compression	4GHz/20MHz/32/4/UL & DL/OFDM	S	TR	SVD of subband of CSI	CSI	SGCL	✓	✓	×	100K	CSI feedback overhead reduction [30-70%] BM: eType2	✓	×	✓
CSI prediction	4.9GHz/10MHz/8/1/DL/-	S	CNN	CSI measurements	Future CSI measurements	MSE	✓	✓	×	1M	10%-outage capacity [11%] at 9 slots CSI aging BM: Autoencoder (AE)-based CSI compression in 3GPP Release-18 [13]	✓	×	✓
Waveform learning	3.5 GHz/ Narrowband/4/2/UL/CP-OFDM	S	CNN	Autoencoder	Transmitted bits	BCE	✓	✓	×	-	SE/TP [15–20%] BM: Conventional 5G link with DMRS for channel estimation [12]	✓	×	×
PA post distortion compensation	30GHz/400MHz/64/1/UL/DFT-s-OFDM	S	FNN	Equalized symbols	Transmitted bits	BCE	×	✓	×	50K	TP [~15%] EE [~10-45%] Coverage [3 dB] BM: Max-LogMAP demapper [11]	✓	✓	✓
Phase noise compensation	100GHz/1.5GHz/1/1/UL/DFT-s-OFDM	S	FNN	Equalized symbols	Transmitted bits	BCE	×	✓	×	3K	BLER [~2.6 dB] for 64 QAM BM: Max-LogMAP demapper	×	✓	✓
Beamforming with imperfect CSI	3.5GHz/20MHz/8/1/DL/CP-OFDM	S	TR	Estimated CSI, estimated CSI uncertainties	True CSI	SE	✓	×	×	-	SE [~85%] BM: SLNR maximization precoding	✓	×	✓
Learning LDPC codes	-	S	RNN	FER Measurement	Transmitted bits	BCE	×	✓	×	-	FER [~1 dB] BM: 5G NR LDPC codes with layered normalized min-sum decoding	✓	×	×
Pilot assignment for D-MIMO	28GHz/100M/1/1/UL/single sub-carrier	RL	GCNN	Random generated ER graph	-	BCE	×	×	✓	500K	Rate [~20-40%] BM: Semi-random and MDS-based pilot assignment	✓	×	×
Access point selection and power control in D-MIMO	1.9 GHz/20 MHz/ 1/1-64/UL/single sub-carrier	U	DNN	Large scale fading channel coefficients	No labels	NSR	×	×	✓	500K	Rate [~20-30 %] and computational complexity reduction BM: Optimization-based with full power	✓	×	×
Semantic communication	2GHz/20MHz/-/1/1/UL/QAM	S	CAE	Transmitted bits	Transmitted bits	MSE	✓	✓	×	2.5M	SE [~50% / 90%] and SSIM [~40%] in noisy channels BM: Classical encoding	✓	×	×

Model types: Transformer (TR), Conditional generative adversarial network (CGAN), Convolutional neural network (CNN), fully connected neural network (FNN), recurrent neural network (RNN), Graph convolutional neural network (GCNN), Deep neural network (DNN), **Training types:** supervised (S), unsupervised (U), reinforcement learning (RL), **Training loss functions:** Binary cross-entropy (BCE), Squared Generalized Cosine Loss (SGCL), Mean square error (MSE), Spectral efficiency (SE), Negated sum rate (NSR)**, KPIs:** Spectral efficiency (SE), Throughput (TP), Energy Efficiency (EE), Normalized mean square error (NMSE), Frame error rate (FER), Benchmark method (BM), **Note:** All trainings are offline, so not mentioned in the table.

should be designed based on the use-case goals. We design an end-to-end learning approach to perform joint beam design and AoA estimation of a single target in a monostatic scenario (see Table 2). We replaced conventional beamforming and estimation functions with fully connected NNs (FNNs) achieving 8% improvement in detection probability with respect to a conventional approach.

B. RF hardware impairment compensation

Sensing performance is more severely degraded than communication performance under hardware impairments. This emphasizes the importance of compensating for hardware impairments with the assumed models do not match the real hardware models in sensing, and hence, in ISAC. AI has been shown to enhance performance of hardware impairment systems. Moreover, to enhance the explainability of AI and reduce the number of parameters to learn, we use a model-based ML approach ISAC signal processing algorithms with learnable parameters to compensate for impairments. We consider the sensing of multiple passive objects in a scene while performing communication with an intended device. To accomplish both goals, we parameterize the orthogonal matching pursuit (OMP) algorithm, to detect and estimate the position of a random number of targets.

C. Detection/sensing/classification

1) Gesture and pose classification

ISAC based on CSI can rely on ubiquitously deployed network infrastructure which already estimates and collects CSI for communication purposes. The presence of living and non-living entities, such as humans and objects, respectively, in the environment have a detectable influence on the

estimated CSI. With the use of AI, this influence can be translated to presence, location, and a more elaborate state information of the present objects. Due to the similarities between CSI data and images, CNNs, which are widely used in computer vision, have shown accurate results for CSI-based

Table 2 AI-driven air interface applications for integrated sensing and communications, model specifications, and assessments.

Application	Radio assumptions : Carrier freq./BW/# of TX antennas/# of RX antennas/ Waveform	Training type	Model type	Training mode	Training features	Tanning labels	Training loss function	Model deployment		Model size [number of parameters]	KPI [gain]	KVI		
								TX	RX			Environmental Sustainability	Inclusiveness	Trustworthiness
Signal optimization	60 GHz/ N/A / 16 / 16 / Single-carrier	S	FNN	Offline	Target area, comm. messages, received signals	Target presence and position, comm. Messages.	BCE, NLL, and CCE.	✓	✓	TX: 2386 RX: 9751	8% better detection probability BM: least-square beamforming & maximum a posteriori ratio test at receiver	✓	×	✓
Hardware impairment compensation	60GHz/ 60MHz/ 64/ 64/ OFDM	S	Model-based	offline	TX: target area RX: angle-delay map	Target positions	GOSPA	✓	✓	64	67 m better GOSPA loss BM: least-squares beamforming & orthogonal matching pursuit	✓	×	×
CSI-based sensing for presence detection and localization	5 GHz/ 80 MHz/ 1/ 1/ OFDM	S	FNN	offline	Amplitude of the channel frequency response	Labeled 2D space	MSE	✓	✓	16K	60% less miss-classification BM: logistic regression	✓	×	✓
CSI-based human sensing	60 GHz/2.16 GHz/36/36/ SC	S	CNN	offline	RAW CSI	Target gestures and poses from MOCAP	CE	✓	✓	107K	26% higher classification accuracy BM: FNN	✓	×	✓
Human sensing	60 GHz.//800 MHz16/16/ OFDM	S	CNN	offline	Power per beam pair	Target gestures	CE	✓	✓	180K	17% higher classification accuracy BM: FNN	✓	×	✓

Training types: Supervised (S); **Model types:** Fully connected neural network (FNN), Convolutional neural network (CNN); **Loss functions:** Cross entropy (CE), Binary cross entropy (BCE), Mean square error (MSE), Generalized optimal sub-pattern assignment (GOSPA), Negative log-likelihood (NLL), Categorical cross-entropy (CCE), Benchmark method (BM),

sensing tasks (e.g., obstacle detection, human presence detection, gesture recognition). In our work, a CNN is trained in a supervised manner by combining CSI measurements with ground truth gesture data collected from motion capture cameras, resulting in over 93% gesture classification accuracy (see Table 2). Also, more classical AI methods such as FNN have sufficiently good performance for presence detection and localization (see Table 2).

Alternately, we leverage power per beam pair to capture fine-grained information of human gestures. We consider a similar system as above, comprising two millimeter-wave antennas set up in a bi-static configuration that facilitates the capture of different gestures/poses of the users. The two antennas serve the role of a transmitter and a receiver, respectively, as shown in Figure 4. The TX continuously transmitted a 5G-like signal, comprising one synchronization block and random data, across 50 beams, whereas the RX continuously received the reflected signal from the closed environment across 56 beams. For each measurement, we computed a detailed grid, of shape (50 x 56), of beam power per TX and RX beams. This was then processed using a CNN, designed to handle spatial and temporal data variations across TX and RX beams. The network was trained on a dataset of gestures/poses from 8 participants, where each participant performed each gesture/pose for ten seconds. The trained network classified the gestures across 8 users (see Figure 4), with an average accuracy of ≈ 96%.

V. AI-DRIVEN AIR INTERFACE: STANDARDIZATION

Standards development organizations (SDOs) have already started the introduction of AI/ML for PHY.

A. AI/ML for air interface in 3GPP

3GPP initiated a study item on AI/ML for 5G NR air interface in Release-18 to establish a framework for enhancing the air interface through the application of AI/ML. This study considers different aspects including AI/ML model's LCM. The study investigated the potential benefits of AI/ML through evaluations and comparisons with legacy methods for three use cases, i.e. a) *CSI feedback* to reduce overhead, improve feedback accuracy and enable CSI prediction, b) *beam management* to reduce overhead, minimize latency, and enhance beam selection accuracy, and c) and *positioning* to enhance positioning accuracy. Following the insights from this study, Release 19 is set to define a framework for employing AI/ML for the air interface, focusing on use cases such as positioning and beam management. Release 19 will also expand on the exploration of AI/ML for additional applications. The developments by 3GPP aim to accelerate the AI/ML integration in telecom industry, setting the stage for its extensive adoption as the industry progresses towards 6G.

B. AI/ML for air interface in O-RAN

O-RAN ALLIANCE aims at evolving RANs with two core principles: *intelligence* and *openness*. O-RAN proposes two

ways to manage the network infrastructure. The first one is done via service management and orchestration (SMO) framework that contains non-real-time RAN intelligent controller (RIC) and rApps for intelligent RAN optimization in non-real-time (>1 s) that uses data analytics and AI/ML. The second one considers near-real-time RIC and xApps that enable control and optimization of centralized unit (CU) and distributed unit (DU) with near-real-time control loops (10ms - 1s). The implementation of AI/ML processing via rApps and xApps enables prediction and adaptation of the network parameters in real time, especially at the PHY, for different optimizations. This would allow more flexibility in the RAN by enabling AI-driven methods for MIMO beamforming, traffic steering, resource allocation, power saving (extinction of radio antenna during period of the day according to the traffic analysis), etc., that are among potential 6G use-cases currently discussed at O-RAN Alliance.

VI. Challenges of AI-driven air interface

The challenges for the adoption of AI in PHY include:

Complexity: AI/ML methods for the physical layer in 6G introduce varying computational complexities. The complexity of these methods largely depends on model size as detailed in Tables I and Table II. Several methods such as *pruning, quantization,* and *distillation* can be applied to lower the complexity of the models. While training is typically performed offline and infrequently, inference complexity significantly impacts real-time operations within the networks, especially for energy-constrained devices. Additionally, model LCM—encompassing data collection, retraining, and monitoring—introduces further complexity, necessitating additional protocols and signaling to support the integration of AI/ML into the 6G physical layer.

***Training data*:** High quality datasets are essential for both training and assessment of AI/ML models during the development of the AI-driven methods. However, *lack of standardized datasets* is one of the challenges towards development, assessment, and benchmarking the AI-driven methods. In addition, obtaining datasets which accurately represent the *wide variety of scenarios* that the system might encounter is challenging. *Collection of training data from the network* is another challenge which requires *observability* of certain data in the PHY and *signaling* between the BS and the UE to coordinate data collection. In many physical-layer applications, at least part of the training can be carried out with *simulated data* to reduce the overall need of measured training data. Leveraging on *generative AI (GenAI)* provide opportunity for creating synthetic training dataset.

***Real-time processing*:** There are stringent delay constraints for PHY signal processing. On one hand, AI-driven processing may impose higher computational complexity compared to the non-AI alternatives. On the other hand, AI-driven methods can leverage parallel processing by running on AI accelerators and benefit from enhanced efficiency. Real-time responsiveness requires a balance of optimized hardware and model efficiency. Parallel processing and hardware acceleration via GPUs, TPUs, or dedicated ASICs can be used to handle intense computation. Additionally, model compression techniques can be utilized to further enhance real-time processing. AI-driven methods need to be adopted deliberately for specific functionalities to improve performance while benefiting from AI accelerators.

***Energy efficiency*:** The energy consumption of AI-driven methods can be divided into two main components: training and inference. Training occurs once, with potential re-training occurring less frequently, hence, the energy consumption associated with model training should be evaluated over its entire lifecycle. In contrast, inference occurs every time the model is used, hence, energy-efficient inference is critical for minimizing the overall energy consumption of the system. It is anticipated that model training will primarily take place in cloud environments, where computational resources are more abundant, while inference will be executed on UE or BS, which face stricter energy consumption constraints. To meet these constraints, AI accelerators should be integrated into hardware platforms dedicated to PHY processing, optimizing energy efficiency during inference. An end-to-end assessment of energy consumption, considering both training and inference throughout the model's lifetime, is essential for a comprehensive comparison with legacy methods.

***New design paradigm*:** AI/ML integration in PHY requires new design processes to be adopted for the model LCM, e.g., for data collection, data validation, model training, and model monitoring. This can be even more challenging in the case of distributed models; hence, it is essential to have standard procedures for AI model LCM when applicable to ensure multi-vendor coordination. The integration of AI poses new challenges from user's perspective. Preserving user privacy and handling sensitive information during data collection for model training emerges as a crucial topic. For ensuring trustworthiness and maintaining data privacy, differential privacy and federated training methods ensure that sensitive data remains on local devices rather than centralized servers.

***Generalization*:** Considering the complexity and the cost associated with re-training of AI models in each new environment, e.g., for sensing applications, it is essential for the AI models to generalize to new environments to ensure cost-efficient solutions. This requires considering generalization capability of the model as one of the key design factors. Techniques such as *data augmentation*, *domain adaptation*, and *transfer learning* expand the effective training set and enable models to quickly adapt to new scenarios, while *masked encoder architectures* and *self-supervised learning* approaches leverage unstructured or unlabeled data to learn more robust representations. Incorporating *regularization*, *meta-learning*, and *few-shot learning* strategies further builds resilience against overfitting, ensuring that models remain accurate across diverse conditions and evolving environments.

***Contributions towards 6G key values*:** AI is envisioned to contribute towards 6G key values, i.e., *environmental sustainability*, *trustworthiness*, and *inclusiveness*. However, aligning with these values presents several challenges. While AI-driven methods can enhance network EE, training AI models is energy-intensive, and overall energy consumption must be considered. For sustainability, *model compression* techniques can be highly beneficial. AI can enhance PHY security, such as through AI-based jamming detection that boosts resilience and trustworthiness, but it also faces security risks from other AI-powered entities. Extensive data collection

for AI/ML introduces risks like unauthorized access and regulatory non-compliance, which necessitate privacy-preserving techniques such as *federated learning (FL)* and *differential privacy*. However, FL is also vulnerable to attacks like model inversion and data poisoning. Robust frameworks for secure data management, explainable AI, and advanced intrusion detection are critical to ensuring trustworthiness and resilience in AI/ML-driven 6G systems.

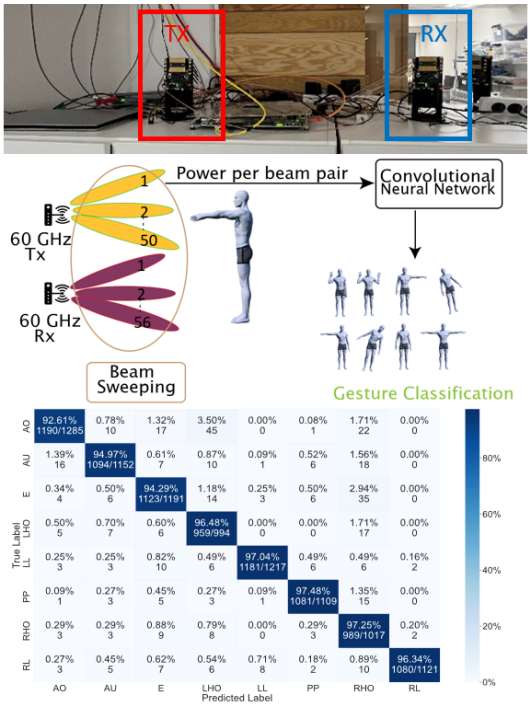


Figure 4: The measurement setup for gesture recognition and results.

REFERENCES

[1] R. Liu, H. Lin, H. Lee, F. Chaves, H. Lim and J. Sköld, "Beginning of the journey toward 6G: Vision and framework," *IEEE Comm. Mag.*, vol. 61, no. 10, pp. 8-9, October 2023.

[2] M. A. Uusitalo, *et al.*, "6G vision, value, use cases and technologies from European 6G flagship project Hexa-X," *IEEE Access*, vol. 9, pp. 160004-160020, 2021.

[3] M. Merluzzi *et al.*, "The Hexa-X project vision on artificial intelligence and machine learning-driven communication and computation co-design for 6G," *IEEE Access*, vol. 11, pp. 65620-65648, 2023.

[4] H. Rydén, H. Farhadi, A. Palaios, L. Hévizi, D. Sandberg and T. Kvernvik, "Next generation mobile networks' enablers: Machine learning-assisted mobility, traffic, and radio channel prediction," *IEEE Comm. Mag.*, vol. 61, no. 10, pp. 94-98, October 2023.

[5] H. Kim, S. Oh and P. Viswanath, "Physical layer communication via deep learning," *IEEE J. Selected Areas in Info. Theory*, vol. 1, no. 1, pp. 5-18, May 2020.

[6] J. Hoydis, F. A. Aoudia, A. Valcarce and H. Viswanathan, "Toward a 6G AI-native air interface," *IEEE Comm. Mag.*, vol. 59, no. 5, pp. 76-81, May 2021.

[7] Hexa-X, "Deliverable D3.3: Final models and measurements for localisation and sensing," [available online], https://hexa-x.eu/deliverables/.

[8] Z. Geng, H. Yan, J. Zhang and D. Zhu, "Deep learning for radar: A survey," *IEEE Access*, vol. 9, pp. 141800-141818, 2021.

[9] U. Demirhan and A. Alkhateeb, "Integrated sensing and communication for 6G: Ten key machine learning roles," *IEEE Comm. Mag.,* vol. 61, no. 5, pp. 113-119, May 2023.

[10] J. Deng, *et al.*, "5G and AI integrated high performance mobile SoC process-design co-development and production with 7nm EUV FinFET technology," *IEEE Symp. VLSI Tech.*, Honolulu, HI, USA, 2020.

[11] H. Farhadi, J. Haraldson and M. Sundberg, "A deep learning receiver for non-linear transmitter," *IEEE Access*, vol. 11, pp. 2796-2803, 2023.

[12] D. Korpi, M. Honkala, J.M.J. Huttunen, "Deep learning-based pilotless spatial multiplexing," in *Proc. 57th Asilomar Conference on Signals, Systems, and Computers*, Oct. 2023.

[13] 3GPP TR 38.843, "Study on artificial intelligence (AI)/machine learning (ML) for NR air interface," V0.1.0., June 2023.

[14] Hexa-X-II, "Deliverable D4.3: Early results of 6G radio key enablers," [available online] https://hexa-x-ii.eu/results/.

[15] M. Ataeeshojai, *et al.*, "Iterative matrix inversion methods for precoding in cell-free massive MIMO systems," *IEEE Trans. Vehicular Tech.*, vol. 71, no. 11, pp. 11972-11987, Nov. 2022.

BIBLIOGRAPHIES

Hamed Farhadi is a Senior Researcher at Ericsson Research, Sweden. He was the Technical Manager of Hexa-X project and is leading intelligent radio access design in Hexa-X-II.

Bitan Banerjee is a postdoctoral researcher at the Vodafone Chair for Mobile Commun. System, TU Dresden, Germany.

Rafael Berkvens is a Professor with the Department of Electronics-ICT at the Univ. of Antwerp, and a principal investigator with imec, Belgium.

Nabeel Nisar Bhat is a PhD student at the Univ. of Antwerp and imec, Belgium.

Emmanuelle Bodji is a PhD student with the Department of Connectivity Technologies at Orange Innovation, France.

Dilin Dampahalage is a PhD student with Centre for Wireless Communications (CWC), Univ. of Oulu, Finland.

Eslam Eldeeb is PhD student with CWC, Univ. of Oulu, Finland.

Jeroen Famaey is a Professor with the Department of Computer Science at the Univ. of Antwerp and a principal investigator with imec, Belgium.

Gerhard P. Fettweis is Vodafone Chair Professor at TU Dresden, Germany, CEO of the Barkhausen Institute, and coordinator of the 5G++ Lab Germany.

Jaeseong Jeong is a Senior Specialist in AI at Ericsson Research, Stockholm Sweden.

Dani Korpi is a Senior Specialist with Nokia Bell Labs, Espoo, Finland.

Siddhartha Kumar: is a Senior Systems Engineer at Qamcom Research and Technology AB, Gothenburg, Sweden.

Yann Lebrun is a Staff Engineer at Qualcomm, Paris, France.

Rodolphe Legouable is the Manager of the Connectivity Technologies team at Orange Innovation, France.

José Miguel Mateos Ramos is a PhD student with the Department of Electrical Eng. at Chalmers Univ. of Technology, Sweden.

Nurul Huda Mahmood is a Sr. researcher and Adjunct Professor with CWC, Univ. of Oulu, Finland and a co-lead of Hexa-X-II work on radio innovations for 6G.

Mohammad Hossein Moghaddam is a senior system engineer, and joint communication and radar sensing expert at Qamcom Research and Technology AB, Gothenburg, Sweden.

Ahmad Nimr is a research group leader at Vodafone Chair, TU Dresden, Germany. He is a co-lead of Hexa-X-II work on radio innovations for 6G.

Nandana Rajatheva is a Professor with CWC, Univ. of Oulu, Finland.

Nuwanthika Rajapaksha is a PhD student with CWC, Univ. of Oulu, Finland.

Athanasios Stavridis is a system engineer at Business Area

Networks, Ericsson, Lund, Sweden.
Tommy Svensson is a Professor and leader of Wireless Systems within the Department of Electrical Eng. at Chalmers Univ. of Technology, Sweden.
Han Yu is a postdoctoral researcher with the Department of Electrical Eng. at Chalmers Univ. of Technology, Sweden.
Leif Wilhelmsson is a Principal Researcher at Ericsson Research, Lund, Sweden.
Henk Wymeersch is a Professor with the Department of Electrical Eng. at Chalmers Univ. of Technology, Sweden.