



TG#1: State-of-the-Art AI/ML Applications for Improvements in RAN Performance

AI-RAN Alliance

WG1 AI-for-RAN

ai-ran.org

AI-RANTM
ALLIANCE

Contents

Executive Summary.....	4
1. Introduction and scope.....	6
2. AI-based air interface.....	7
2.1. AI-based receiver.....	8
2.2. AI-based transmitter	9
2.3. Learned end-to-end air interface	10
3. AI-based CSI and feedback.....	11
3.1. Channel modeling.....	11
3.2. AI-based channel estimation.....	14
3.3. CSI compression and feedback.....	15
3.4. CSI prediction.....	17
3.5. 3GPP work on CSI feedback.....	18
4. Radio resource management.....	19
4.1. Beamforming.....	19
4.2. Beam management.....	22
4.3. Scheduling	25
4.4. Spectrum sensing and sharing.....	28
4.5. Power control.....	30
4.6. Mobility management	33
5. Novel network architectures.....	34
5.1. Distributed MIMO (D-MIMO)	35
5.2. Reconfigurable Intelligent Surfaces (RIS).....	37
5.3. Extremely large-scale MIMO (XL-MIMO)	38
6. Sustainability and energy efficiency enhancements via AI-RAN.....	40
6.1. Cell energy saving.....	41
6.2. Transmit antenna muting.....	43
7. AI-RAN network management.....	44
7.1. Load balancing and traffic steering	44
7.2. Anomaly detection, fault prediction, and mitigation.....	46
7.3. Slice service level agreement management.....	48
8. Integrated sensing and communication	50
8.1. Integrated sensing and communication	50
8.2. AI-based positioning and localization	52

9. Emerging topics	54
9.1. Semantic communications	54
9.2. Digital twins for RAN	56
10. Security in AI-for-RAN	57
11. Definitions of terms and abbreviations	59
12. References	64
13. Document History	100

Executive Summary

This report from the **AI-RAN Alliance** provides a comprehensive overview of how **artificial intelligence (AI) and machine learning (ML)** are transforming the **radio access network (RAN)**, summarising the latest techniques and models leveraging AI-for-RAN technologies and outlining their role as foundational elements in **5G and 6G** wireless networks.

AI-for-RAN focuses on enhancing RAN with AI from physical layer to management and operations layers. To do this, AI is embedded across the RAN hardware and software stack. The report highlights how advanced neural architectures and learning strategies, including convolutional, transformer, and reinforcement learning models, enhance key RAN functions, such as **channel estimation, beamforming, scheduling, spectrum sensing, energy saving, and power control**. These advances are part of **3GPP Releases 18–20** discussions, which mark the transition from AI-enhanced to **AI-native networks**.

By approaching this subject through a thorough, detailed review of the existing academic and industrial literature on AI in advanced wireless networks, the authors detail multiple techniques, experiments, tests, architectures, and models. Starting with the **AI-native air interface**, the document charts the RAN's transition from traditional transceivers to AI-integrated physical layer. It describes a three-stage evolution: first, AI-assisted parameter optimization enhances existing components; second, AI begins to replace key blocks such as channel estimation and decoding; and finally, complete end-to-end learning enables self-adaptive, AI-native transceivers. Demonstrated systems like AI-based transceiver already show up significant bit error rate performance gains and throughput improvements in real-time environments, marking tangible progress toward fully learned air interfaces.

The report also details AI-driven approaches to **channel state information (CSI)** processing. In particular, the report details how these approaches can potentially outperform legacy estimation methods, especially in complex massive multiple-input multiple-output (mMIMO) channels. These advances are part of the 3GPP Releases 18–20 standardization discussions, which define interoperable frameworks for AI/ML-based CSI management across vendors.

This document further examines **radio resource management (RRM)**, as AI/ML now underpins the intelligent control of RAN resources, including beamforming and scheduling, as well as spectrum and power optimization. Transformer and deep reinforcement learning (DRL) methods are investigated in the literature and promise scalable, resilient beamforming even under imperfect CSI. Reinforcement learning-based schedulers could also potentially deliver expert-level throughput and fairness with an order-of-magnitude reduction in complexity. AI-based

spectrum sensing and sharing is investigated to improve interference management, while ML-driven power control is studied as a promising scheme to boost energy efficiency in mMIMO systems.

An overview of novel or emerging network architectures is also provided, as next-generation architectures increasingly rely on AI for coordination and adaptability. **Distributed MIMO** benefits from AI/ML techniques for resource allocation, while reconfigurable intelligent surfaces (RIS) use graph and convolutional networks for channel shaping. Furthermore, **AI-RAN technologies** can directly enhance sustainability through predictive cell-sleep control, transmit-antenna muting, and energy-aware scheduling. On the other hand, advanced anomaly detection and fault prediction can reduce downtime and operational waste where AI/ML techniques are widely investigated in the literature. Looking ahead, **semantic communications, digital-twin RANs, integrated sensing and communication (ISAC), and AI-based security** define the frontier of the AI-enabled network era, laying the groundwork for self-optimizing, self-healing, and energy-efficient 6G systems.

In summary, the report details how AI-for-RAN presents a foundational shift toward **AI-native wireless networks in future**. As global standardization efforts and AI capabilities continue to develop, these elements will underpin AI-native 6G systems that unify communication, sensing, and computation, transforming the efficiency and intelligence of future mobile infrastructures.

1.Introduction and scope

With the rapid advancement of computational capabilities for training and deploying artificial intelligence (AI)-driven solutions, there is a tremendous opportunity to enhance radio access network (RAN) functionalities through advanced AI. This focus has become central to the development of next-generation wireless networks, where neural networks and various machine learning (ML) techniques are being widely explored to enhance RAN performance across three primary paradigms: AI-for-RAN, AI-and-RAN, and AI-on-RAN.

This document aims to outline the state of the art (SOTA) of the AI-for-RAN use case, where sophisticated AI/ML methods can achieve higher spectral efficiency, higher data rates, and improved coverage while promoting energy efficiency. Research efforts in developing AI-for-RAN functionalities span the entire RAN stack. This includes addressing core physical (PHY) layer functionalities such as modulation and demapping, equalization, channel estimation, adaptive beamforming, and medium access control (MAC) layer functionalities such as resource allocation, scheduling, adaptive modulation and coding (AMC), power control, protocol learning, dynamic spectrum allocation, and multi-user interference management—all of these are essential capabilities for high-density, high-capacity networks. As mentioned earlier, energy-efficient networking is another focus in AI-for-RAN, where AI-driven optimization protocols help achieve significant power savings, a major focus towards sustainable, greener networks.

We start reviewing the SOTA from the PHY layer, where significant progress has been made in recent years, gradually moving to higher layers through MAC and network-level functionalities. Furthermore, we discuss novel network architectures such as distributed multiple-input multiple-output (D-MIMO), reconfigurable intelligent surfaces (RIS), and extremely large-scale MIMO (XL MIMO) with a focus on AI/ML to improve their performance. We also briefly discuss emerging topics such as integrated sensing and communication (ISAC) and semantic communications, which are concepts readily harnessing the power of AI/ML for RAN enhancements. An overview of different AI-RAN functionalities discussed is given in Fig. 1.

When considering the standardization aspects for AI-RAN, the completion of the 3rd Generation Partnership Project (3GPP) Release 18 for 5G-Advanced (5G-A) initiated a study to integrate AI and ML into mobile communications. Its study item on AI/ML for New Radio air interface focused on three key applications to enhance the PHY layer: channel state information (CSI) feedback enhancement, beam management, and improvements in positioning accuracy [Lin23]. Additionally, Release 18 investigated critical aspects related to AI/ML life cycle management, including performance monitoring and testing [Lin23]. With these advancements, it is anticipated that Release 19 will further incorporate AI-based features, contributing to the development of an AI-native PHY layer in future network releases. Moving on, 3GPP Release 20/Release 21 will focus on advancing 5G-A toward 6G, implementing AI-native networks for

automation, optimization, and improved efficiency in network management and resource allocation [3GPP25]. To this end, integrating an extensible AI/ML framework built on 5G-A with native support for AI/ML lifecycle management (e.g., configuration, performance monitoring, deactivation, and seamless transition to conventional algorithms) will be considered, along with exploring new use cases. We will explore these standardization efforts in more detail in their respective sections.

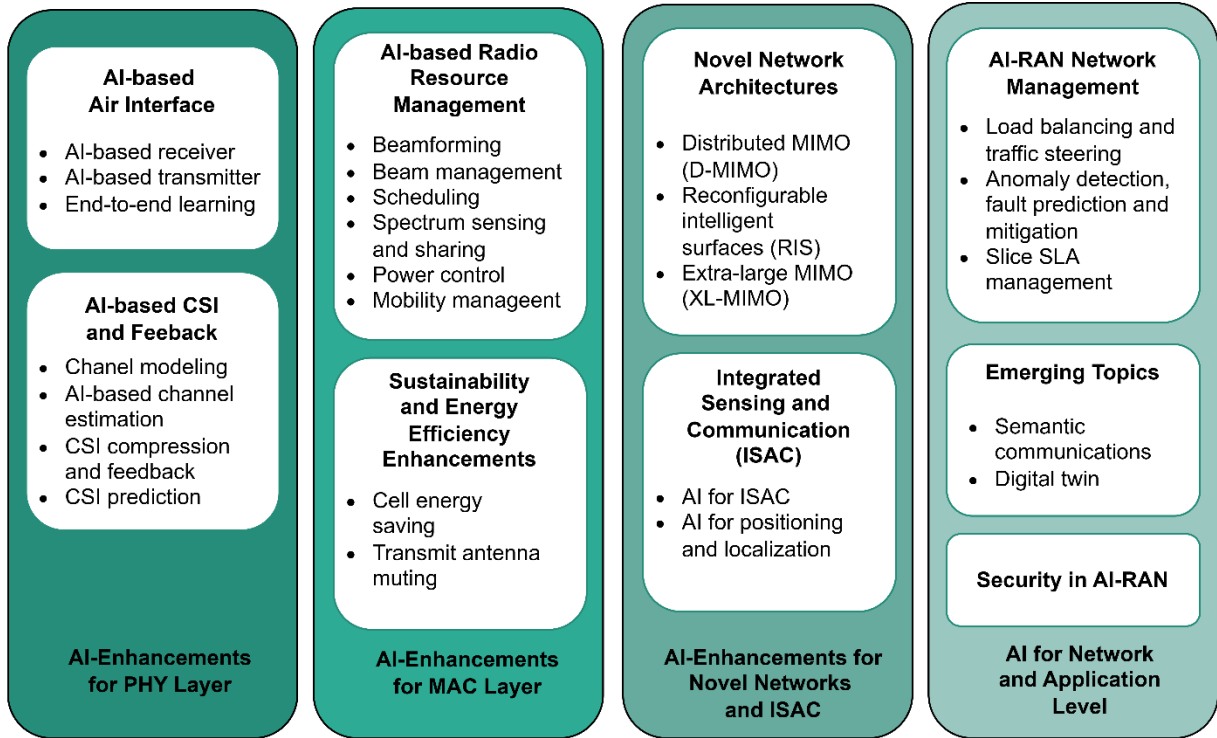


Fig. 1: Overview of different AI-RAN functionalities examined in this paper.

2. AI-based air interface

The seminal work on applying deep learning to the physical layer was presented in [OH17], which demonstrated the potential of modeling a transceiver as an autoencoder. Since then, the application of AI/ML to various functional blocks of the transceiver has been extensively studied in the literature. Depending on the blocks being assisted or replaced by AI/ML, they can be categorized into three application stages, representing an evolutionary path from the conventional transceiver, as illustrated in Fig. 2.

- Baseline: Conventional transceiver in RAN.
- Stage 1: The baseline transceiver is optimized using AI/ML-based parameters and controls.
- Stage 2: Some conventional transceiver blocks are replaced by AI/ML-based blocks.

- Stage 3: The transceiver is integrated with AI/ML across almost all blocks.

In Stage 1, the conventional functional blocks in the transceiver are primarily maintained while AI/ML-based parameter optimization and adjustments enhance performance. For example, AI/ML can infer additional channel state information and control a conventional channel estimator by channel categorization based on factors such as mobility speed or line-of-sight status. In Stage 2, some major transceiver blocks are replaced by AI/ML-based solutions. Any block from the front-end to channel coding is a potential target for AI/ML applications. The long-term vision, as outlined in [Hoy+21], is to integrate AI/ML into nearly all blocks, as exemplified in Stage 3. All these stages contribute to the development of key technologies for improving the AI-native air interface.

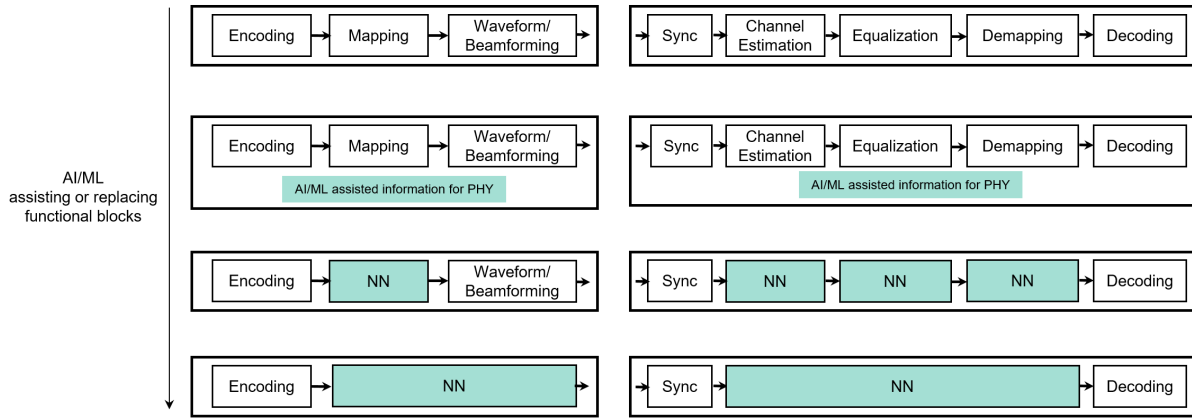


Fig.2: Evolution towards the AI-native air interface.

2.1. AI-based receiver

Building on the idea of residual networks, a ResNet-based convolutional neural network (CNN) was developed as an orthogonal frequency-division multiplexing (OFDM) receiver, achieving remarkable success in [HKH21]. Termed "DeepRX", this architecture takes in the received signals in the frequency domain after the fast Fourier transform (FFT) operation. It predicts the log-likelihood ratios for each bit, fed to the channel decoder using the bit-metric decoding technique, while being compatible with the 5G New Radio (5G NR) resource grid structure. This architecture marked a significant step in the realization of AI-based receivers.

Further research introduced variants such as a hybrid architecture [Pih+23], where convolutional layers operating in both the time and frequency domains were incorporated to address distortions in each domain, mitigating the impact of power amplifier nonlinearity, and a fully learned multiple-input multiple-output (MIMO) receiver in [Kor+20]. Recently, a significant step in designing a practical neural receiver for 5G NR with real-time performance in a graphics processing unit (GPU)-accelerated platform was presented in [Wie+24]. There, the authors

proposed an adaptive neural receiver architecture capable of dynamic modulation and coding scheme (MCS) configurations without the need for re-training or additional inference cost.

In contrast to deep neural receivers, another approach is to use neural networks in some blocks, such as channel estimation and equalization, which will be discussed in subsequent chapters. Replacing only the demapper while keeping other blocks similar to the conventional architecture, a deep learning receiver has been introduced in [Sko+23], which accounts for oscillator phase noise in high-band transmissions, and extended in [FHS23] to account for nonlinear transmitter effects, e.g., power amplifier (PA) nonlinearity. The reported evaluations in [FHS23] and [Sko+23] have shown major performance improvements over legacy methods in terms of block error rate, throughput, coverage extension, and energy efficiency, with significantly lower complexity than the fully AI receiver approach.

Targeting energy efficiency, in [ML24], a neuromorphic AI prototype receiver demonstrated energy-efficient radio channel estimation by activating only neurons that detect changes, unlike traditional neural networks that require constant computation. This dynamic approach reduced energy consumption by adapting to variations in channel activity and suggests broader applications for low-power AI. Extending the neuromorphic AI to the CNN-based ResNet architecture, “SpikingRx” in [Gup+24] shows comparable performance to the aforementioned neural network (NN)-based architectures while achieving significant power savings. A demonstration of SpikingRx at the user equipment (UE) combined with AI/ML-optimized higher-order modulations [Mad+23] at the base station (BS) was presented at Mobile World Congress 2025 (MWC 2025) as one of the AI-RAN Alliance demos.

One challenge with existing AI-based receiver studies is the widespread use of simulated radio channel data for both training and validation, raising concerns about real-world generalization. The authors in [Luo+25] studied the generalization of DeepRx across various simulated channel scenarios and validated its performance over-the-air (OTA), highlighting its strong generalization. The study emphasizes that relying on individual 3GPP channel models can hinder both training and validation, and shows the importance of tailored channel models and OTA testing in learned receiver development.

2.2. AI-based transmitter

Moving to the transmitter side, “DeepTx” [HKH23] proposed a ResNet-based CNN model that enhances beamforming by predicting downlink (DL) channel information from uplink (UL) channel estimates. Trained in a supervised manner, the CNN model minimizes a loss function tied to UE receiver performance. The network not only predicts channel evolution between UL and DL but also compensates for inefficiencies in the processing chain, including beamforming. Numerical results indicate improved beamforming efficiency and accuracy. Many other learned

transmitter approaches usually combine the deep neural receivers at the receive side and are covered under the end-to-end learning section. Besides the neural transmitter approaches, AI/ML has been explored for predicting beam codebook indices and channel predictions [KS24]. Furthermore, NN-based digital pre-distortion (DPD) is being explored at the radio frequency (RF) level.

AI/ML-based beamforming with imperfect CSI for multi-user MIMO transmission is proposed in [FGS25]. The proposed method based on transformer neural networks (TNNs) architecture has shown robust performance in the presence of imperfect CSI, due to the channel, while scaling with the number of users.

AMC schemes enable data transmission in accordance with channel conditions, thereby improving signal quality and reliability. It selects the best MCS, considering channel estimates, data size, and system constraints. Challenges include balancing aggressive transmissions for spectral efficiency and conservative transmissions for latency and block error rate (BLER). ML-based solutions for AMC are reported in the literature, including function approximation using NNs [MWB+21] and deep reinforcement learning (DRL) [GWH+22].

A recent work in [KK24] proposed interference perceptual multi-modulation, a UL data transmission scheme for full duplex systems in which different MCS levels are assigned to individual UL sub-bands, achieved through a deep CNN that predicts the appropriate MCS for each sub-band by assessing potential residual self-interference based on the DL resource allocation pattern. Several other studies in [Bal18; Guo+19; Mur+21] also investigated AI/ML-based approaches for self-interference estimation and cancellation in full-duplex systems. Furthermore, the authors in [MTL22] investigated joint estimation of the channel, phase noise (PN), and in-phase and quadrature-phase (I/Q) imbalance in multicarrier MIMO full-duplex wireless systems, combining a deep learning-based approach with linear minimum mean-squared error (LMMSE) solution. These studies align with the 3GPP Release 19 standardization efforts on sub-band full duplex (SBFD) to enhance network capacity and reduce latency, particularly for UL communications, by enabling simultaneous transmission and reception across different sub-bands within the same frequency band. Furthermore, the authors in [Hua+20] developed an ML framework for non-orthogonal multiple access (NOMA) by modeling the whole NOMA system comprising the BS, channels, and the UEs as a blackbox and utilizing a deep neural network (DNN) to approximate it to learn the encoding and decoding, power control, and data detection.

2.3. Learned end-to-end air interface

Further developments to [OH17] were explored in [Cam+20] towards more practical bit-metric decoding (BMD) with constellation shaping, along with a fully differentiable neural iterative

demapping and decoding (IDD) structure. One of the major problems in training such a transceiver in an end-to-end fashion as an autoencoder is the missing channel gradient problem, which was solved by approximating the channel gradient in [AH19]. Authors in [AH22a] demonstrated that pilotless communications achieve a significant gain in spectral efficiency by employing a deep neural receiver while adapting the transmitter's constellation for OFDM.

Further progress was shown by completely learning a waveform for both single and multicarrier transmissions in [AH22] and [Gou+21]. Improving hardware-impaired communications by leveraging the concept of waveform learning was shown in [Kor+22], while more enhanced deep receivers to decode power amplifier (PA) non-linearity impaired signals were presented in [Pih+23]. Moving to higher frequencies towards sub-THz, waveform learning was explored under the increasing phase noise impairment in [Mar+24]. Exploring the channel coding gains using NNs has also been presented in [Jia+19; Gru+17]

Learned air interface with online learning was recently demonstrated as one of the AI-RAN Alliance demos at MWC 2025, showcasing adaptation to different channel conditions, allowing the modulation scheme to specialize under specific channel conditions, such as high mobility dynamic channels or more stable fixed wireless access channels, opportunistically using a fixed AI updating scheme OTA.

3. AI-based CSI and feedback

Channel state information (CSI) is essential for ensuring reliable communication, efficient resource allocation, and interference management. Accurate channel estimation is crucial for precoding, interference suppression, and signal detection, especially in fast-fading and high-mobility scenarios. Conventional channel estimation methods rely on mathematical and statistical techniques to estimate CSI using pilot signals. Widely used approaches include least squares (LS), which is simple but noise-sensitive, and minimum mean square error (MMSE), which improves accuracy by incorporating noise statistics. Compressed sensing (CS), on the other hand, exploits channel sparsity to reduce pilot overhead, while Kalman filtering tracks temporal variations in CSI in fast-changing environments. However, these traditional channel estimation methods struggle with complex, time-varying, and high-dimensional channels, making ML a powerful tool for denoising, prediction, and adaptation to dynamic channel conditions. In this section, we discuss AI/ML-based SOTA techniques for channel estimation, CSI compression and feedback, and CSI prediction.

3.1. Channel modeling

When considering SOTA channel modeling techniques used in literature and industry, different statistical, deterministic, or hybrid approaches are used that vary in complexity and accuracy,

catering to different wireless communication needs. Global standardization efforts have largely favored statistical channel models (SCMs) due to their low computational complexity, enabling faster simulation implementations. SCMs characterize the stochastic nature of wireless channels in specific environments, such as urban streets, by reconstructing the channel using key parameters derived from statistical distributions [Sha+24]. Geometry-based stochastic models (GBSMs), on the other hand, are derived from a predefined stochastic distribution of effective scatterers by applying the fundamental laws of wave propagation [MED16; Wan+22a]. Ray tracing-based channel models fall under the deterministic modeling approach and employ physics-based simulations for highly accurate propagation analysis, albeit with high computational costs. Recent advancements in high-performance computing and AI-assisted acceleration have further enabled more efficient ray tracing simulations, reducing computational overhead and making large-scale, site-specific channel modeling more practical for next-generation wireless systems. Ray tracing techniques are usually applied to high-frequency channel models, which consider ray propagation mechanisms such as direct, reflection, and scattering rays, traces ray propagation paths, and then analyzes channel characteristics. The DeepMIMO dataset introduced in [Alk19] is a publicly available dataset generated using ray-tracing-based channel models to provide realistic channel characteristics, with customization capabilities for different use cases and scenarios, supporting AI/ML solution development in massive MIMO (mMIMO) and millimeter wave (mmWave) networks. In addition to these channel modeling approaches, emerging AI/ML-based generative models leverage deep learning to generate adaptable channel representations, balancing accuracy and efficiency. Table 2 below summarizes the comparison of these channel modeling approaches in terms of accuracy, complexity, generalization, and flexibility.

Table 2: Summary of different channel modeling approaches.

	Statistical Channel Models	Ray Tracing-Based Models	Generative Channel Models
Approach	Uses empirical/statistical distributions to model radio channel behavior.	Uses a physics-based simulation of wave propagation that considers obstacles and reflections.	Uses AI/ML techniques to generate realistic channel data based on learned distributions. E.g., generative adversarial networks (GANs), variational autoencoders, diffusion models
Accuracy	Moderate (depends on the accuracy of statistical parameters).	High (captures physical propagation effects accurately).	High (if trained on high-quality data, can generalize well).
Computational complexity	Low (fast simulations).	High (requires complex 3D models and computations). With accelerated computing, such as GPUs, simulations can be done fast.	Moderate to high (depends on model architecture and inference complexity).
Generalization	Suitable for large-scale system simulations but lacks spatial/temporal detail.	Limited to the specific environment modeled. Ideal for site-specific channel modeling.	Can generalize across environments if trained with a diversity of data.
Flexibility	Easy to modify parameters for different scenarios.	Requires a 3D model of the environment for each scenario, including material characteristics.	Flexible, can learn from diverse datasets and adapt.
Example models	3GPP spatial channel model, WINNER, QuaDRiGa.	Shooting and bouncing ray (SBR), METIS.	GAN-based CSI prediction, variational autoencoder-based models.
Use cases	System-level simulations, theoretical studies.	Link and system-level simulations.	AI-driven channel prediction, dataset augmentation for simulations.

3.2. AI-based channel estimation

Compared to widely used traditional channel estimation methods, ML can learn nonlinear channel characteristics directly from data, making it robust in high-mobility scenarios and sparse channels such as the mmWave and THz bands. Furthermore, AI/ML algorithms reduce computational complexity by replacing iterative estimations with fast feedforward computations. Moreover, ML methods can be designed to minimize pilot overhead and enable efficient channel modeling for mMIMO systems [MG21].

When considering channel estimation in doubly-dispersive environments, the deep learning algorithms are utilized in two different manners: (i) feedforward neural networks with different architectures and configurations are employed on top of the conventional symbol-by-symbol channel estimators and (ii) CNNs are used where the estimated channel for the entire frame is modeled as a two-dimensional (2D) low-resolution noisy image [Sol+19; Kar+21]. One of the initial works under this second approach is [Sol+19], which considers the time-frequency response of a fast-fading communication channel as a 2D image and reconstructs and enhances it using image processing techniques. Pilot signals are used, and the channel is restored, considering it as a low-resolution image using a cascade of neural networks for image super-resolution and image restoration. This was demonstrated for 5G NR physical uplink shared channel (PUSCH) on a real-time OTA setup at MWC 2025. Furthermore, [He+18] incorporated a special CNN as a denoiser to the approximate message passing algorithm for beamspace channels, and [Don+19] adapted deep CNNs for three-dimensional (3D) mmWave mMIMO channels to exploit the correlations in frequency, time, and space.

A similar approach based on denoising for 5G NR PUSCH leveraging AI to handle high-dimensional wireless channel data in the joint frequency-temporal-spatial domain, achieving better channel estimation results and leading to improved coverage, was demonstrated as an AI-RAN Alliance demo at MWC 2025. The work in [BA19; BDA20] proposed an untrained NN to prevent the training complexity of CNNs used for channel estimation. Specifically, in [BDA20], a DNN based on a deep image prior network was first used to denoise the received signal, followed by conventional least-squares estimation.

Furthermore, combining the least square estimator with NNs is proposed in [Yan+19a; RWX20]. In addition, on sounding reference signals (SRS) channel estimation performed within a sparse periodicity, the work in [Yoo+25] applied AI denoising to SRS channel estimation and showed SRS-based beamforming performance improvement in lab tests. While AI denoising directly improves estimation accuracy, AI can also be used to infer channel parameters indirectly thereby improving its performance. Channel conditions and estimation behaviors vary with parameters such as line-of-sight (LoS)/non-line-of-sight (NLoS), Doppler spread, delay spread and signal-to-noise ratio (SNR), which significantly impact conventional channel estimation performance.

AI can improve channel estimation performance by assisting the channel parameter estimation or control as introduced in [Oh+25].

ML-based channel estimation techniques investigated in the literature for Terahertz (THz) communications systems include deep CNN-based models [CH20], GANs [BA21], and deep kernel learning [NA19]. They are shown to reduce the complexity involved in channel inference and that arising from ultra-massive antenna elements. Furthermore, federated learning-based wideband THz channel estimation is studied in [Elb+23a; Elb+23b], where the labels of the model are obtained via beamspace support alignment and sparse Bayesian learning algorithms, respectively. A real-time CSI feedback platform is developed in [CH24] using an autoencoder-based deep learning model integrated into a 5G Open RAN (O-RAN) protocol stack, demonstrating accurate CSI reconstruction and reduced feedback overhead on both OTA and emulated testbeds. As discussed, these advancements highlight the potential of ML-based approaches to tackle the challenges of channel estimation across diverse communication environments. By leveraging the unique capabilities of NNs, they offer robust, efficient, and scalable solutions.

3.3. CSI compression and feedback

The performance of mobile communications systems heavily relies on the systems' accurate downlink CSI at the BS, typically obtained via estimation at the UE and fed back to the transmitter. However, in mMIMO systems, the large number of antennas at the BS creates a high CSI dimension and substantial feedback overhead. This raises the need for an alternative approach for the UE to locally compress the downlink CSI before sending a limited version back to the BS to reduce feedback overhead. Conventional CSI feedback methods based on codebook and compressive sensing do not meet the requirements of low complexity and high accuracy [Guo+22]. Recently, deep learning techniques have been used for CSI estimation in frequency division duplex (FDD) systems, following two main research directions: (i) learning the uplink-to-downlink channel mapping to eliminate the feedback by using uplink channel knowledge at the BS to gain downlink channel knowledge and (ii) compressing the estimated or available channel matrix with a deep learning architecture to minimize the feedback [Ner+23]. Such a feedback compression method using an autoencoder is illustrated in Fig. 3.

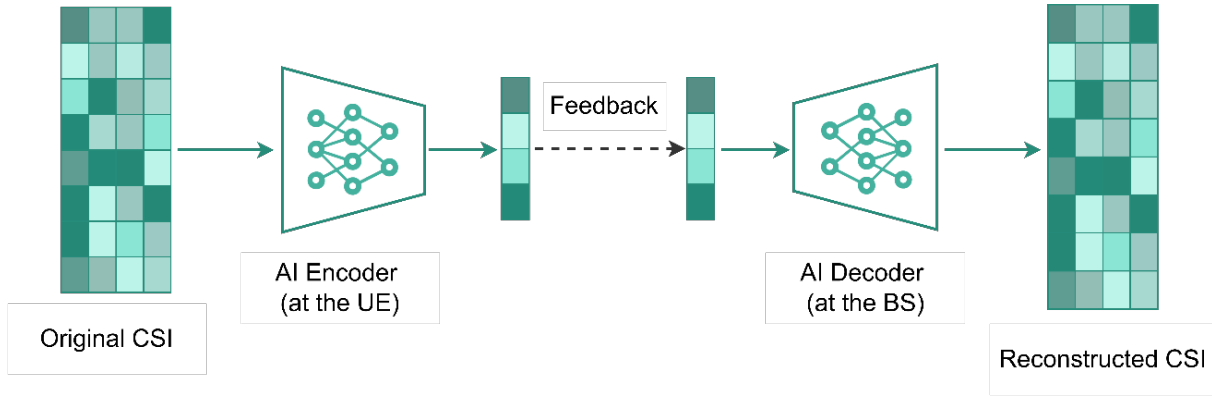


Fig 3: CSI feedback using autoencoder (adapted from [Guo+22]).

The authors in [Guo+22] presented an overview of different deep learning-based CSI feedback techniques. The existence of uplink-to-downlink mapping was first investigated in [AA19]. Different deep learning techniques have since been used, such as fully connected NNs [AA19; Arn+19], sparse complex-valued NNs [Yan+19a], and CNNs where the space-frequency channel matrix is treated as an image [Don+19; Han+20a]. The authors in [Riz+20] proposed an autoencoder-based dimensionality reduction for channel compression followed by uplink-to-downlink mapping via random forests.

The authors in [WSJ18] first applied deep learning to CSI-limited feedback and built a novel CSI sensing and recovery neural network (CsiNet) that fits a transformation from CSI to codewords and an inverse transformation from codewords to CSI. CsiNet is shown to outperform the conventional CS-based CSI feedback method in terms of reconstruction quality. The authors in [LZD20; MYG21; Cao+21; Sun+21] also investigated deep autoencoder architectures to embed the DL channel matrix at the UE into a latent space with reduced dimensionality. Other deep learning techniques used for CSI compression and feedback include long short-term memory (LSTM) network and attention mechanism in encoder and decoder network, respectively, proposed in [CDN19], joint convolutional residual network MIMO channel feature extraction and recovery in terms of bit-level quantization performance in [Lu+20], multi-resolution convolutional network to extract CSI features on multiple resolutions as proposed in [LWS20], etc.

Another technique recently used for AI-based CSI compression and feedback is the transformer-based approach, which can more effectively exploit correlations among elements in the channel matrix, leading to improved CSI reconstruction quality. The authors in [XYP21] introduced CsiTransformer, an effective and robust transformer-based CSI feedback scheme, and MixedCsiNet, a lower-complexity variant designed for power-limited devices by using CNNs as the encoder and a transformer as the decoder, which outperformed existing CSI feedback schemes across various compression rates. Furthermore, the authors in [Ju+23] proposed

transformer-aided parametric CSI feedback (T-PCF) to compress and feed back key mmWave mMIMO channel parameters such as path angles, time delays, and complex gains into a codeword for efficient feedback to the BS. By leveraging parameter correlations, T-PCF reconstructs the CSI matrix, thereby enhancing feedback efficiency and channel recovery in mmWave systems. Given that channel conditions vary and data distributions change accordingly, AI-based CSI reporting must generalize to these evolving distributions. In this context, a causality-aware CSI encoding system proposed in [Ban+23] adapts to changes in the input data distribution, and evaluations demonstrated its effectiveness in domain adaptability, out-of-distribution generalization, and power savings. The authors in [Bai+25] proposed a transformer-based temporal-domain multi-length optimized decoder that effectively leverages temporal correlations in MIMO channels to enhance CSI reconstruction. Unlike schemes that require different models or additional signaling overhead for re-synchronization between the encoder and decoder under varying conditions, the transformer-based temporal-domain multi-length-optimized decoder can robustly recover CSI by leveraging feedback on the precoding matrix and temporal information with a single fixed model for inputs with varying sequential time steps.

3.4. CSI prediction

Outdated CSI due to delays between BS and UE, or rapid channel variations resulting from UE mobility, known as channel aging, can be challenging to achieve timely and precise CSI. Channel prediction techniques are utilized to overcome this challenge without spending additional channel training resources. Radio channel prediction is considered one of the enablers of next-generation mobile networks, and is envisioned to enhance beamforming for multi-user MIMO systems and serve as a cornerstone for emerging use cases such as wireless energy transfer [Ryd+23]. An effective method for modeling an aging channel is an autoregressive (AR) stochastic model whose parameters are computed based on the channel correlation-matching property across adjacent coherence intervals [BB05]. In addition to this AR process and the use of a Kalman filter to predict, AI techniques capable of time-series prediction have also been investigated in recent literature.

When considering suitable AI/ML approaches for channel prediction, since CSI forecasting is a typical time-series learning problem, a recurrent neural network (RNN) is well-suited for exploring the hidden patterns in CSI variations [YNM19]. Such an approach was first proposed in [LYH06], which is extended to MIMO channels in [PVK08; DH14; JS18]. A performance and complexity comparison between AI-based and Kalman filter-based predictors is presented in [JS19]. A hybrid approach combining deep learning and statistical models is presented in [She+22a]. There, NeuralProphet, a novel time-series model composed of statistical components, is combined with RNNs to achieve better channel prediction performance than individual models.

Furthermore, the performance evaluation of an ML-based channel predictor using real-world channel realizations from a testbed is presented in, [She+22b] demonstrating its predictive capability. The authors in [YNM19] proposed a CNN-based channel prediction scheme by using a CNN to identify the channel aging pattern and adopting an AR predictor or an autoregressive network with exogenous inputs recurrent neural network (NARX-RNN) to forecast CSI. A recent work on channel prediction also includes a federated learning-based LSTM model for real-time online CSI prediction to avoid the computational, storage, and communication challenges associated with centralized approaches [LLD24].

Recently, transformer-based channel prediction approaches have been proposed to address the limitations of CNN or RNN-based approaches, such as restricted receptive fields and error propagation. To this end, the authors in [LT23] proposed Channelformer, an encoder-decoder neural architecture for OFDM downlink channel estimation that leverages multi-head self-attention in the encoder, a residual CNN decoder, and customized weight pruning to achieve low complexity and low latency while reducing parameters by up to 70% without performance loss. Furthermore, [Kan+24] proposed CS3T-UNet, a Transformer-based model for mMIMO channel prediction that integrates cross-shaped spatial attention and group-wise temporal attention to capture spatial and temporal dependencies.

3.5. 3GPP work on CSI feedback

In 3GPP Release 18, significant efforts were made to explore the integration of AI/ML techniques into CSI feedback processes for 5G and beyond. The main goal was to examine how AI/ML models can replace traditional methods, like codebook-based feedback, to enhance performance and reduce overhead. This included studying the use of both one-sided models (e.g., UE-sided CSI prediction) and two-sided models involving collaboration between the UE and gNodeB (gNB) for adaptive learning [Lin23] and a scalable codebook for the CSI framework using a location-specific optimized codebook basis [Lim+25]. The 3GPP work focused on defining the requirements for model architectures, data collection procedures, and training methodologies for inter-vendor collaboration, applied to CSI feedback. The main challenge in AI-based CSI compression and feedback, as identified in the 3GPP work, is that AI/ML deployments spanning multiple companies differ from traditional approaches to standardization, notably, since the model itself defines the air interface. Thus, a large part of the effort was devoted to ensuring that a potential specification would allow interoperability. This remains a challenging task for RAN1 and RAN4, as parts of the models are deployed across UE and gNB. The work continued in Release 19 and culminated in the standardization of CSI prediction. The standardization of spatial-frequency domain compression of CSI feedback has now begun in Release 20.

4. Radio resource management

Energy-efficient and intelligent radio resource management (RRM) is crucial for future 6G networks, optimizing resource allocation to reduce energy consumption while maintaining high network performance and real-time adaptability. Dynamic and complex networks, along with the increasing number of devices requesting access and data-intensive applications, make it imperative to allocate spectrum resources to ensure reliable network performance and meet user quality-of-service (QoS) requirements. Resource allocation problems are formulated as optimization problems subject to total power, user QoS, and other relevant constraints, and solved using optimization-based techniques to find optimal solutions.

With their ability to analyze complex data and uncover intricate associations using DNNs and CNNs, deep learning methods are applied in RRM tasks such as resource scheduling across multiple users or applications, beam management, subcarrier and power allocation, spectrum sharing, etc. The multiparameter nature of these RRM problems and the complexity of associations between channel characteristics make deep learning approaches an efficient way to address the overall RRM problem [Bar+22]. DRL techniques such as soft actor-critic (SAC), deep deterministic policy gradient (DDPG), and deep Q-network (DQN) enable joint optimization of different radio resources and adaptively optimize complex, dynamic environments, efficiently handling large state-action spaces and achieving real-time decision-making. Unsupervised learning techniques such as K-nearest neighbors (KNN), K-means clustering, F-test, etc., focus on clustering based on the underlying patterns in data and, therefore, are more suitable for RRM tasks such as BS selection, user grouping, clustering, and handover management tasks. The following subsections discuss SOTA AI/ML approaches for different RRM problem areas in detail.

4.1. Beamforming

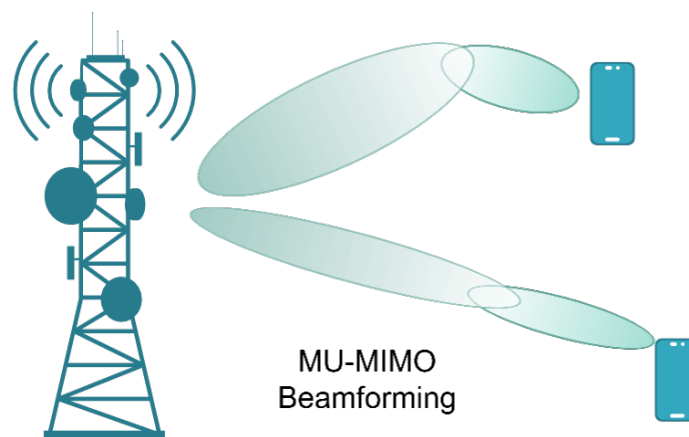


Fig. 4: MU-MIMO beamforming.

Multi-user multiple-input multiple-output (MU-MIMO) enables spatial multiplexing, allowing multiple users to share the same time-frequency resources to improve spectral efficiency and user capacity, which is crucial for supporting the massive access scenario in networks. To enable this, proper precoding design is essential to manage and shape multiple data streams to minimize interference among users while maximizing data throughput. The focused energy beams provided by beamforming (as shown in Fig. 4) and improved reliability achieved by interference reduction through spatial separation enable the support of dense networks, especially in high-frequency communications, and maintaining robust connections in challenging environments. When considering advanced beamforming techniques, hybrid beamforming (HBF) combines digital and analog beamforming, leveraging digital processing for flexible signal management while using analog components to efficiently direct beams, optimizing power and complexity. On the other hand, intelligent beamforming uses ML to dynamically adjust and optimize beam patterns based on real-time network conditions, user locations, and environmental factors. Several recent studies that investigated HBF and/or intelligent beamforming approaches to improve the system performance and reduce the complexity/signaling overhead are summarized in the survey paper [Gko23].

Due to channel aging, MU-MIMO is prone to imperfect CSI. In [FGS25], an AI/ML-based beamforming method is proposed to address imperfect CSI. Traditional ML-based beamformers, particularly those using fully connected layers, face scalability issues: they scale poorly with increasing numbers of users and require extensive training across numerous user permutations. To overcome these challenges, [FGS25] employs permutation-equivariant neural network architectures, focusing on TNNs, which inherently handle user permutations more efficiently. The evaluation results demonstrate that TNNs maintain strong beamforming performance under imperfect CSI while scaling better with varying user counts. As such, TNNs present a promising solution for scalable and efficient beamforming in MU-MIMO systems with imperfect CSI.

Deep unfolding is a popular model-driven approach for AI/ML-based precoding algorithm design [Hu+21; PBJ22], which unfolds an existing iterative algorithm and embeds trainable parameters to improve performance. The authors in [Hu+21] proposed an iterative algorithm-induced deep unfolding neural network based on the iterative weighted minimum mean square error (WMMSE) algorithm to solve the sum rate maximization problem for precoding design in MU-MIMO systems. [Jin+23] further proposed a model-driven deep learning approach to maximize the spectral efficiency of mmWave MU-MIMO systems through hybrid precoding. Specifically, the WMMSE optimization problem is combined with manifold optimization and deep learning to improve performance and simplify the process. They have also shown robustness to changing environments and imperfect CSI.

Another deep learning-based framework for low-complexity MU-MIMO precoding design is proposed in [ZGZ22], which transforms the MIMO precoding problem into a multiple-input single-output (MISO) precoding problem, enabling the optimal precoding structure to be obtained in closed form. There, a DNN is designed to fit the mapping from the channels to the precoding matrix. Furthermore, the authors have used input dimensionality reduction, network pruning, and recovery module compression to improve computational efficiency. CNN-based models are proposed in [Elb19; Pek+20; Liu+22a] for precoding/beamforming designs to extract spatial correlations from channel matrices or to decrease trainable parameters. Specifically, the authors in [Liu+22a] proposed a joint antenna selection and hybrid precoding design via unsupervised learning by employing ResNet to extract features from the channel matrices.

A DNN for efficient and distributed channel estimation, quantization, feedback, and downlink multiuser precoding in an FDD mMIMO system with multiple served users was proposed in [SAY21]. It was shown that the proposed DNN-based algorithm with short pilot sequences and very limited feedback overhead can already approach the performance of conventional linear precoding schemes with full CSI. In [Hoj+21], a low-complexity unsupervised learning approach for fully digital precoder and HBF design was proposed, which does not rely on CSI but is based on the received signal strength indicator (RSSI), demonstrating that despite not having access to full CSI, the BS can be trained to robustly design the HBF. Furthermore, the authors in [LMD21] proposed a DRL approach to design HBF matrices, also minimizing the associated signaling overhead.

A transformer-based HBF scheme consisting of three transformer encoder modules is proposed in [Wan+23]. Motivated by the fact that the analog RF beamformer and digital baseband beamformer can be jointly optimized to approximate the performance of a fully digital beamformer, each transformer encoder is designed to handle a specific part of the HBF optimization. By combining the structural priors of traditional optimization methods with the feature-extraction capabilities of self-attention, they have shown superior performance compared to both conventional and existing deep learning-based methods in both complete and limited CSI feedback scenarios. Additionally, the authors in [LL25] considered a frequency-selective TDD mmWave massive MIMO-OFDM system with a hybrid architecture and proposed a transformer-based approach that bypasses high-dimensional channel estimation by learning channel sensing from uplink pilots and simplifying hybrid precoding into neural network optimization. Their solution employs downlink hybrid precoding based on an interleaved-polymerization-transformer-based analog precoding network (IPTAP-Net), which decomposes on a per-user basis and a conventional linear digital precoding algorithm to reduce computational complexity in multi-user systems.

Recently, [Liu+25] proposed a low-rank adaptation blocked LSTM transformer with message-passing GNN, an ML-assisted beamforming design for ISAC in ultra-dense device-to-device mmWave networks, named LoRA_BLT-MPGNN. There, the LoRA_BLT module captures temporal and spatial patterns with low complexity, while the MPGNN models capture inter-node relationships to adapt across different network sizes. Using historical channel data, the method predicts the next-slot beamforming matrix, reducing signaling overhead and achieving near-optimal sum rates, while maintaining robust sensing under imperfect CSI. Finally, another recent study [Eme+25] proposed a transformer-based foundation model for mMIMO precoding that minimizes transmitter energy while adapting to per-user rate requirements. The results show that with the same energy consumption, the zero-shot deployment of the proposed foundation model substantially outperforms zero-forcing and approaches weighted minimum mean squared error performance with 8 times less complexity. They also proposed a cosine-similarity-based data augmentation method to handle data-scarce scenarios, enabling implementation of DL-based solutions in practice by addressing challenges of data availability and training complexity.

4.2. Beam management

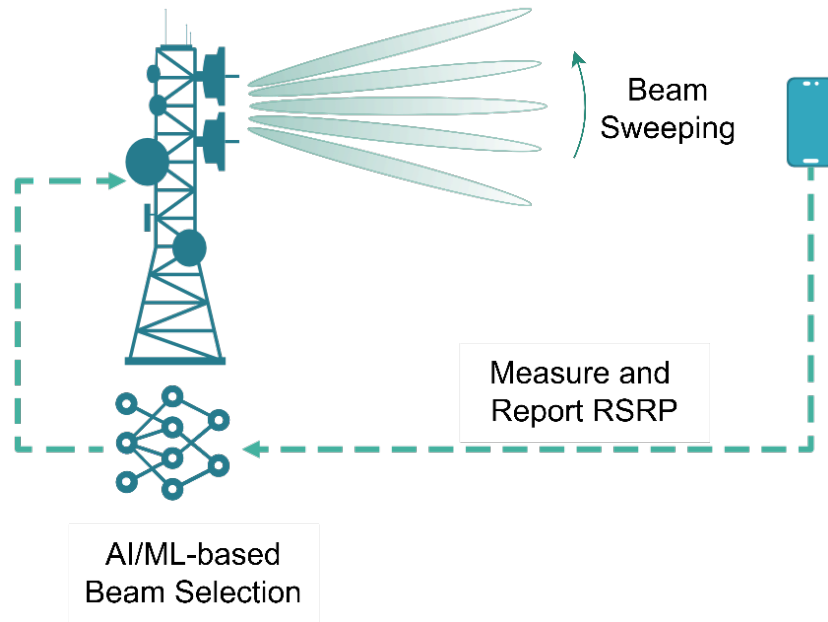


Fig. 5: AI/ML-based MU-MIMO beam management.

Beam management (BM) refers to a set of procedures that enable the establishment, maintenance, and recovery of directional radio links in different frequency bands such as FR1 and FR2. It is essential in 5G NR and beyond, ensuring reliable high-frequency transmissions despite challenges like high path loss and blockages. The procedures include periodic beam sweeping, beam measurement, beam selection and refinement, and beam recovery.

Conventional BM procedures face several challenges, such as high overhead, higher latency, and slow adaptation, being vulnerable to blockages, mobility challenges, and computational complexity and scalability issues when the number of beams increases due to the large number of antennas in mMIMO. AI/ML-based BM techniques can be helpful in such instances, as also identified by 3GPP Release 18 BM use cases [Lin23]. An example AI/ML-based BM approach is illustrated in Fig. 5.

An overview of the standardized MIMO framework supporting BM and CSI acquisition in 5G NR is presented in [Li+20], including future research directions for AI/ML-based BM. The studies in [Qur+23; Ma+23] also explored ML-based approaches for mmWave BM management, where [Qur+23] provided a comprehensive overview of different AI/ML techniques for BM in mmWave/THz networks and discussed open challenges. ML-based beam training that accounts for the sparse nature of mmWave channels is advantageous for massive beam-forming training, as the number of dominant beam pairs is far fewer than the total number of candidate pairs [DSJ19]. In addition, BM for efficient RRM with reduced complexity is investigated in [Zho+19a], where a deep learning-based BM and interference coordination method in a dense mmWave network is presented to reduce the complexity of the RRM problem via a DNN-based approximation. The authors in [Zuo+22] proposed an AI-based BM approach for beam selection with reduced beam measurement overhead. A well-known VGG-based network is proposed to predict the quality of all beam pairs using the measured quality of partial beam pairs. When it comes to mmWave analog beam management and selection, one effective approach to reduce signaling overhead is to leverage reciprocity between UL and DL channels, where applicable. In [Isa+23], the authors proposed a personalized federated learning technique that exploits the correlation between different frequency bands in a heterogeneous scenario with non-IID characteristics. The idea is to predict the downlink precoder codebook based on channel estimation obtained in the sub-6 GHz uplink, which the gNB can acquire without additional overhead. The proposed algorithm not only reduces the signaling overhead but also improves beam selection accuracy by 33%.

ML-based time domain mmWave BM for 5G-A and beyond is proposed in [Li+23a], including reference signal receive power (RSRP) prediction and beam change prediction. The proposed AI/ML algorithms are first evaluated through simulations with new UE mobility models based on recent standard evolutions in 3GPP. The study [Van+21] explored reinforcement learning (RL) for optimizing beam allocation and relay selection in vehicular networks, focusing on serving multiple vehicles simultaneously while mitigating blockage effects. Furthermore, [BAM19] utilized an LSTM model to track the angle of arrival (AoA) based on channel information, thereby facilitating BM while the UE is in motion along a straight trajectory. The authors in [PRM21] adopted a CNN to extract angle information from in-phase-quadrature signal samples, reducing the time required to complete the 5G NR initial access process.

On ML-based beam management in FR3, the authors in [DH24] proposed an ML-refined codebook design for XL-MIMO, using a neural network and beam selection strategy to design the initial access and refinement codebooks using end-to-end learning from beamspace representations. It is extended in [DH25], to encapsulate the achievable spectral efficiency to produce codebooks that support full-rank MIMO connections, while also mitigating inter-cell interference. The proposed network beamspace learning (NBL) end-to-end codebook design algorithm requires minimal shared information and achieves over 10 dB improvement in beam alignment and more than 25% gain in network spectral efficiency compared to traditional codebooks.

When considering AI/ML model management for BM, a recent study [Bai+24] investigated AI-based BM across model training, inference, and monitoring, focusing on optimizing the data collection time window for temporal beam prediction. Furthermore, the authors in [Jay+24] investigated AI/ML for BM, focusing on spatial-domain and time-domain beam prediction in terms of beam prediction accuracy, model generalization, and system-level performance evaluated using a 3GPP-compliant simulator, with results showing the impact on user throughput and provided design guidelines for deploying lightweight, low-complexity AI/ML models for practical BM. The authors in [Sun+24] also explored various aspects of AI model management, including model selection, monitoring, and activation/deactivation operations, from a 3GPP perspective. They demonstrated effective management of model switching between AI-based BM and traditional BM, with simulation results showing a reduction in unnecessary switches.

In addition to performing beam prediction solely based on reference signal measurements, assisted information such as location data can also be used to improve AI/ML prediction accuracy. To this end, the authors in [KC21] proposed an RNN-based beam prediction for mmWave systems using sensor data and beam measurements as inputs to improve prediction accuracy. [EMC21] investigated model-based and model-free hybrid beamforming approaches for joint ultra-massive MIMO radar-communications in the THz band, and proposed DeepBF, a CNN-based approach to learn hybrid beamformer weights.

The authors in [Liu+22e] considered ISAC in vehicular networks. They proposed a convolutional LSTM RNN, denoted as CLRNet, to predict the angle of vehicles for the design of predictive beamforming, where the CNN and the LSTM module are adopted to exploit the spatial features and the temporal dependency from the estimated historical angles of vehicles to facilitate the angle prediction. Another notable work on predictive beamforming in ISAC for vehicle-to-infrastructure networks in [Liu+22f] proposed a deep learning approach that learns from historical channels to directly predict the next-slot beamforming matrix to maximize the achievable sum rate. A recent study in [Zha+24] proposed a Transformer-based predictive

beamforming scheme for ISAC vehicular networks, using a novel protocol that maps signal echoes directly to the beamforming matrix, avoiding CSI and historical channel requirements. The resulting optimization problem is solved via the echoes-based convolution Transformer network (ECT-Net), which combines convolution and attention to capture spatial dependencies and improve predictive beamforming performance. Considering non-RF sensing modalities, an infrastructure-mounted light detection and ranging (LiDAR) sensor system operating in parallel with the wireless system is used to predict future beam decisions in [Mar+22]. There, an RNN-based learning model is proposed for beam prediction, employing user-tracking information obtained from LiDARs and beam-sequence information from the wireless system. In contrast, the authors of [HK23a] proposed using a radio environment map that maps localization to the statistics of RSRP and UE mobility trajectories to train an RL agent to predict beam sequences along the UE motion path.

From the standardization efforts in 3GPP, AI/ML for BM was considered in Release 18 of 5G-A. The aim is to explore AI/ML-based approaches to improve beam prediction accuracy while reducing associated overhead. Two primary sub-use cases have been identified: spatial-domain downlink beam prediction and temporal downlink beam prediction. Spatial-domain prediction utilizes current measurements from a specific set of beams (Set B) to predict the optimal beam from another set (Set A), while temporal prediction leverages historical data to forecast the best beam for future instances. These AI/ML models can be implemented on either the gNB or UE side, with corresponding signaling mechanisms established for reporting and inference. The successful completion of these studies in Release 18 has paved the way for normative work in Release 19, focusing on further enhancements in AI/ML support for beam management and other use cases.

Extensive research within the O-RAN Alliance has explored the application of AI/ML to beam management. These studies aim to enhance cluster-level performance metrics, such as interference mitigation, handover efficiency, overall power consumption, and transmission overhead. Notably a few use cases are well studied and standardized in O-RAN Alliance, including grid-of-beams beamforming optimization and beam-based mobility robustness optimization (bMRO).

4.3. Scheduling

Scheduling is essential in wireless networks to optimize resource allocation (e.g., time slots, frequency bands, and antennas) among multiple users or applications, thereby enhancing overall network efficiency, capacity, and user experience. With effective scheduling, the network can balance throughput optimization and fairness, ensuring all users experience an acceptable QoS and that resources are not monopolized by users with the best channels. It also optimizes network power usage by aligning resource allocation with users' channel states and system load,

reducing unnecessary energy consumption — a crucial objective for green communication in mMIMO. These reasons make scheduling a fundamental feature in modern communications networks. Scheduling algorithms like round robin (RR), proportional fairness (PF), and best channel quality indicator (best CQI)-based scheduling have been fundamental in earlier wireless networks but present limitations in the context of 5G and beyond [Mam+22]. While 5G already supports a limited version of MU-MIMO, scaling the number of co-scheduled users in future networks will be increasingly challenging due to the need for even faster scheduling within smaller time slots [Nok24]. In 6G, the scheduler will face additional challenges, including managing network slicing with increased carrier bandwidth (100–400 MHz), accommodating diverse QoS requirements, and co-scheduling 5G and 6G devices on shared carriers [Nok24]. The complex dependencies among these responsibilities challenge and require AI/ML techniques to design novel scheduling algorithms that cater to the increasing demands of future networks.

Different types of DRL models proposed in the literature for decision-making in discrete action space that fit the resource scheduling problem include DQN [Mni+13; AAK17], double DQN (DDQN) [HGS16], advantage actor-critic (A2C) [Xu+20], asynchronous advantage actor-critic (A3C) [Mni+16; Wan+22b], actor-critic with experience replay (ACER) [Wan+17a], and proximal policy optimization (PPO) [Sch+17; RCT24], to name a few. Specifically, the work in [AAK17] focused on safety and QoS in a vehicle-to-infrastructure scenario using a DQN to learn energy-efficient scheduling policy from high-dimensional inputs corresponding to the characteristics and requirements of vehicles residing within a roadside unit's communication range. On the other hand, [Wan+22b] focused on DRL-based no-wait flow scheduling for time-sensitive networks (such as industrial control networks and in-vehicle networks) using an A3C-based algorithm. A recent work in [RCT24] proposed a PPO-based DRL scheduling algorithm to solve the NOMA uplink ultra-reliable low-latency communication (URLLC) scheduling problem involving strict deadlines. The performance of a downlink MU-MIMO scheduler using a DDQN is presented in [Nok24], which used imitation learning to avoid performance-damaging random exploration, rewarding the DDQN agent when its decisions matched those of an expert proportional fair scheduler. Results show that the deep scheduler learned to leverage frequency selectivity and dynamic frequency-hopping patterns. Such DRL methods are particularly advantageous, as an agent can learn from a vast number of radio challenges during offline training and make fast, real-time scheduling decisions with a constant number of operations during the forward pass. Similarly, the authors in [Nér+25] proposed a novel DRL approach for the global resource block packet scheduler in 5G networks, featuring an encoder-only Transformer with action branching to handle large action spaces in resource block allocation. Trained on different varying numbers of users, the proposed approach generalizes across configurations and demonstrates superior packet-loss performance compared to heuristic methods.

DDPG is another DRL algorithm capable of handling continuous, high-dimensional action spaces to find the optimal allocation strategy. The authors in [Guo+20] investigated user-channel correlation to select users in a MU-MIMO system using DDPG. By modeling MU-MIMO scheduling as a Markov decision process (MDP), the proposed real-time DDPG algorithm is shown to be well-trained to simultaneously select users and their frequency-domain resources across all sub bands, satisfying the minimum correlation among selected users. The authors in [Che+21a] also proposed a DDPG-based method for joint user scheduling and transmit precoding selection for MU-MIMO to maximize the system sum rate. The method considers not only the channel correlation between different users but also the antenna correlation of each user.

The problem of frequency scheduling with MU-MIMO has also been addressed in [SKC22] and [Gio+24] to maximize a fairness throughput metric among the users. The action space of all possible user associations across all available frequency sub-bands is combinatorial and explodes even for small problem sizes. AlphaGo-inspired approach in [SKC22] searches for the best performing solution using Monte Carlo tree search and RL. The authors in [Gio+24] addressed the same MU-MIMO frequency scheduling problem using DQN with action-branching, a DRL technique that learns a joint representation of the user channels and buffers and then feeds them to separate parallel DQN branches, each responsible for a single sub-band. The approach yields results comparable to [SKC22] and offers the added benefit of enabling fast inference of scheduling actions through parallelized decision-making. It also allows online fine-tuning after deployment.

A DRL-based algorithm selection framework for cross-layer QoS-aware scheduling and antenna allocation in mMIMO systems is proposed in [Hua+23a]. Rather than directly assigning execution parameters, the proposed framework utilizes DRL to select combinations of algorithms based on current traffic conditions. The DRL model is trained using a specialized MDP model with a componentized action structure and is evaluated in realistic traffic scenarios. More recent work proposed a novel dynamic scheduler, SMART, based on the SAC DRL model and the KNN algorithm [An+23]. They have shown that the proposed model performs very closely to the optimal proportionally fair scheduler in terms of spectral efficiency and fairness while having more than an order of magnitude lower computational complexity in medium network sizes. A comprehensive evaluation of DRL for slice-based resource allocation is conducted in [Tsa+24], where the authors introduce PandORA. This framework automates the design, training, and testing of DRL agents, and analyze how architectural choices such as reward formulation, action space, and decision timescale impact performance across enhanced mobile broadband (eMBB), URLLC, and massive machine-type communications (mMTC) slices. Furthermore, a DQN-based distributed framework is proposed in [Cas+21] to enable self-optimization in software-based 5G

networks, where each cell independently learns optimal slicing and scheduling policies by interacting with its environment and adapting to dynamic traffic demands in real time.

When considering scheduling different applications, there is a vast literature on DRL-based scheduling of IoT applications, including centralized DRL techniques such as DQN, DDQN, and SAC, and distributed approaches such as A3C, to name a few [Che+21b]. A recent work [WGB25] proposed a transformer-enhanced distributed DRL (TF-DDRL) framework for IoT application scheduling in edge and cloud environments, overcoming the limitations of classical DRL-based approaches. It is built on an actor-critic architecture with scalable multi-server support and off-policy correction for stable training. Furthermore, the authors in [Zha+25] proposed a Decision Transformer (DT)-based adaptive decision architecture for resource management in RIS-aided communications and UAV-aided mobile edge computing (MEC) leveraging cloud pre-training and edge fine-tuning to personalize models. By utilizing offline-trained DT models, the architecture achieves faster convergence and superior performance in new scenarios compared with DRL.

Furthermore, building on the latest LLM technologies, [Elk+25] proposed ALLSTaR, an LLM-driven framework for automated, intent-based RAN scheduler design, implementation, and evaluation. By interpreting natural-language intents, generating scheduler code from research literature, and enabling on-the-fly deployment, ALLSTaR supports large-scale OTA testing as demonstrated by the automatic synthesis and comparison of 18 scheduling algorithms on a multi-vendor 5G testbed.

While there is extensive research on AI/ML-based schedulers, a feasible 3GPP-compliant scheduler capable of delivering fair throughput across users, while keeping a low computational complexity for 5G and beyond, is still missing. To address this, SOTA DRL algorithms, such as PPO and distributed SAC discrete algorithm, can be enhanced and adapted to train a deep scheduler using novel training techniques as proposed in [KLV24]. Also, leveraging generative AI in intent-based RAN scheduler design frameworks, such as the ALLSTaR proposed in [Elk+25], can significantly enhance advanced RAN synthesis and optimization. Furthermore, multi-transmission and reception point (multi-TRP) joint scheduling is being explored in 5G-A to enhance network performance. Key aspects of 3GPP's work in this area include coordinated scheduling and beamforming, enhanced carrier aggregation, and dual connectivity enhancements to support simultaneous connections with multiple TRPs thereby improving mobility robustness and seamless handovers.

4.4. Spectrum sensing and sharing

Spectrum sensing and sharing include the processes that detect and utilize underused spectrum and allocate it efficiently among users. Effective spectrum sharing relies on a robust sensing

infrastructure capable of detecting and monitoring spectrum usage, authorizing access, and managing arbitration between multiple legitimate users sharing the same frequency band. Spectrum sensing not only supports spectrum sharing but also enables interference management. Additionally, the infrastructure must sense and detect interference in co-channels and adjacent channels to maintain reliable communication. These techniques are essential to address spectrum scarcity, improve the utilization of underused resources, and support the growing demand for wireless services in 5G, 6G, Internet of Things (IoT), cognitive radio networks, and beyond.

They enable dynamic and fair allocation of the spectrum in real-time, ensuring efficient and reliable communication. AI-based techniques for radio spectrum allocation and management have been developed to intelligently and adaptively handle spectrum sensing and allocation, including related channel allocation, user association, cell selection, and dynamic network access. Spectrum sharing can leverage (i) unused or underutilized spectrum (e.g., due to network traffic load is lower than the planned capacity) as opportunistic access of spectrum without using 5G spectrum coexistence techniques, (ii) 5G spectrum coexistence techniques to optimize the RANs to avoid co-channel interference or to minimize the impact of co-channel interference to each other while still maintaining adequate QoS as well as (iii) combinations of (i) and (ii).

Accurate and effective spectrum sensing has attracted significant attention in the literature, especially AI/ML-based solutions for spectrum sensing and dynamic spectrum management, which have been widely investigated in cognitive radio communication to increase spectral efficiency. To this end, supervised and unsupervised learning techniques such as support vector machines (SVMs), K-means clustering algorithms, and Gaussian mixture models have been proposed in the literature [Thi+13; Li+18; Zho+19b; LKC19; Mah+24]. The authors in [LKC19] proposed deep cooperative sensing (DCS), the first cooperative spectrum sensing (CSS)-based algorithm using a CNN, in which multiple secondary users cooperate to detect a primary user. Furthermore, RL approaches are investigated [SK20; Son+21a]. In [SK20], a DRL-based CSS algorithm is proposed for cognitive radio networks and shown to outperform existing DCS, K-out-of-N, and SVM-based CSS algorithms. In [Son+21a], a novel Q-learning-based spectrum management framework is proposed to enable dynamic spectrum allocation users to perform effective, individualized, and intelligent spectrum management, thereby addressing interference coordination among secondary users and interference suppression for primary users. The efficiency of the Q-learning process is further improved by utilizing a DQN.

Additionally, [Uva+24] presents a deep learning-based semantic segmentation framework for real-time multi-label, multi-class spectrum sensing, where a U-Net architecture with non-local blocks classifies overlapping signals at I/Q-level granularity—achieving high accuracy and enabling robust generalization across devices, sampling rates, and RF environments. In addition,

learning-based spectrum sensing framework development for RIS-assisted communications is investigated in [Yan+21a; Yan+21b]. A recent survey of AI-empowered multiple access technologies, focusing on spectrum sensing, protocol design, and optimization, in [Cao+24] presents detailed SOTA spectrum sensing and sharing studies. The recent AI-RAN demo on AI-based spectrum sensing presented at MWC 2025 demonstrated the path towards the practical implementation of AI-driven spectrum sensing, which continuously and programmatically monitors the uplink spectrum and dynamically adjusts radio resource allocation based on AI-driven signal classification. The demo is based on recent work on spectrum sensing based on I/Q samples processing discussed in [Uva+21; Gan+24; Uva+24, Rob+24a].

AI/ML-based spectrum sensing techniques discussed above pave the way for the development of intelligent spectrum-sharing solutions enabled by reduced prior information requirements and higher sensing accuracy. Intelligent spectrum sharing, which improves spectral efficiency, is required to satisfy the following rules: (i) interference is reduced or mitigated, or (ii) sharing is enabled in time, frequency/band, or spatial domain [Cao+24]. The research spans three main areas: (i) spectrum sharing between secondary and primary users in cognitive radio networks, (ii) spectrum sharing across licensed and unlicensed bands for coexistence networks (such as Wi-Fi and cellular coexistence), and (iii) spectrum sharing between active and passive users [Cao+24].

In addition to the above-mentioned studies on spectrum sensing and sharing, RL-based investigations of spectrum sharing and coexistence between long-term evolution licensed assisted access (LTE-LAA) and Wi-Fi systems, and between active and passive users in high-frequency communications are reported in [GFT16; Han+20b; Bos+21]. Finally, multi-agent dueling DDQN-based DRL model for cellular heterogeneous networks (HetNets) for joint BS and channel selection [Zha+19], single-agent actor-critic DRL-based channel assignment algorithm for hybrid NOMA in 5G networks [Zhe+21], DQN-based dynamic spectrum sharing between 4G long-term evolution (LTE) and 5G NR networks in multi-RAT HetNets [CS21], are some of the other relevant studies along similar research problems which are based on DRL techniques. Finally, the ChARM framework is presented in [Bal+22], enabling real-time spectrum sharing in O-RAN by using deep residual neural networks trained on I/Q samples to classify wireless technologies and dynamically adjust RAN parameters such as frequency and coexistence mode for interference mitigation and spectrum optimization.

4.5. Power control

Power control is a critical component of RRM in modern wireless networks, as it enhances communication quality, reduces interference, and improves overall network performance by effectively allocating power to network entities like BSs, gateways, and edge servers [AA24]. It is a longstanding challenge in cellular networks due to the need to balance coverage, capacity,

and interference. However, in advanced networks such as mMIMO, MU-MIMO, and NOMA, the impact of power control is even more significant, since multiple users share the same time-frequency resources, making inter-user interference management and fairness highly dependent on optimal power allocation. Therefore, in this section, we mainly focus on power control in advanced communication networks, and Table 3 summarizes notable works on mMIMO and NOMA systems.

Table 3: Summary of AI/ML-based power control methods.

System Type	Objective	AI/ML Technique	Methodology Details	Key References
mMIMO Downlink	Max-min and max-product power allocation	DNN	Trained based on UE positions for SINR and SLNR criteria	[SZD18], [PNA21]
mMIMO Uplink	Maximize the sum spectral efficiency	CNN	Joint pilot and data power control from large-scale fading coefficients	[VBL19]
NOMA mmWave Networks	Energy efficiency maximization	DNN	User association, subchannel allocation, and power control under QoS and interference constraints	[Zha+20]
Multi-cell Multi-carrier NOMA	Energy efficiency maximization	ADCNN	Dual-pipeline network for spatial feature extraction and refinement	[Ada+22]
HetNets with NOMA	Improved sum rate with minimum data rates	DNN (unsupervised)	Lagrangian decomposition for subchannel allocation and DNN for power control	[Kim+23]
MU-MIMO NOMA	Sum-rate maximization	DRL (DQN, DDQN)	Joint beamforming and power control	[LZL21]
Multi-cell Cellular Networks	Downlink sum-rate maximization	DRL (actor-critic)	Single-agent and multi-agent actor-critic models outperform WMMSE	[KA20]
6G Connectivity with QoS Constraints	Maximize the number of links meeting QoS	GNN (GCN, GAT)	Graph-based admission and power control using Q-learning and attention networks	[Yan+24]
Cell-free Massive MIMO/D-MIMO	Power control optimization	Supervised, Unsupervised, and DRL	Various methods to handle scalability and complexity	[DAn+19], [ZND20], [Raj+21], [Luo+22], [Rah+22]

In the context of power control in legacy cellular networks, the authors in [Fal+18] proposed ML-based uplink transmission power prediction for LTE based on available passive network quality indicators and application-level information such as RSRP, velocity, and data rate. Trace data from extensive drive tests with periodic uplink transmissions over a public mobile network have been used to evaluate and compare three techniques: random forests, DNNs, and ridge regression. Furthermore, [ME18] proposed an RL-based closed-loop power control algorithm for downlink voice over LTE (VoLTE) in indoor small-cell networks. The method addresses performance tuning and leverages a derived lower bound on effective signal-to-interference-plus-noise ratio (SINR) loss from neighboring cell failures for practical applicability, and simulation results showed improvements in voice retainability and mean opinion score over existing standards.

One of the earliest works on deep learning-based power control for mMIMO was proposed in [SZD18], where a DNN was trained to perform max-min and max-product power allocation in the DL of a mMIMO network. Similarly, [PNA21] introduced a DNN-based method to solve max-min and max-product power allocation problems for mMIMO systems, using SINR and signal-to-leak-plus-noise ratio (SLNR) criteria for linear precoder design, with the DNN trained based on UE positions.

In [VBL19], the authors investigated joint pilot and data power control to maximize the sum spectral efficiency in uplink mMIMO systems. They applied deep learning using a CNN to learn the optimal data and pilot powers from large-scale fading channel coefficients. The work in [Zha+20] further expanded on energy efficiency maximization under QoS constraints, interference, and power limitations by proposing a DNN-based framework for user association, subchannel allocation, and power control in NOMA mmWave networks. This approach achieved higher energy efficiency with reduced complexity.

Other studies on ML-based power control for NOMA systems include [Ada+22; Kim+23; LZL21; Gou+23]. Specifically, [Ada+22] proposed an energy-efficiency maximization framework for multi-cell, multi-carrier NOMA using a dual-pipeline augmented deep convolutional neural network (ADCNN) for real-time power allocation. The first pipeline extracts high-quality features, which are spatially connected and refined using an attention-based network, while the second pipeline handles low-quality spatial features. In [Kim+23], the authors addressed downlink power-domain NOMA for HetNets and introduced resource allocation algorithms for subchannels and transmit powers to improve sum rate performance while satisfying minimum data-rate requirements. The subchannel allocation was performed through Lagrangian dual decomposition-based user association, and a DNN-based unsupervised learning algorithm was used for power control.

For sum-rate maximization in a single-cell MU-MIMO NOMA network, [LZL21] proposed a DRL-based joint beamforming and power control algorithm using DQN and DDQN. Additionally, DRL methods for power control, both single-agent and multi-agent actor-critic DRL approaches, were explored in [KA20], where the goal was to maximize the downlink sum-rate in multi-cell, multi-user cellular networks. These DRL approaches outperformed conventional fractional programming and WMMSE methods with reduced complexity.

In [Yan+24], the authors investigated graph neural network (GNN)-based algorithms to support massive connectivity in 6G. They aimed to maximize the number of links that meet QoS requirements in a given area through joint admission and power control, using the power intensity of the received signal or SINR as a communication quality measure. Two models were proposed: one based on a graph convolution network (GCN) with the Q-learning approach and another using a graph attention network (GAT). Both models were shown to preserve computation time and deliver significant performance gains even with large network scales.

ML-based power control in cell-free mMIMO or D-MIMO systems has also been widely explored, with supervised learning, unsupervised learning, and DRL techniques employed [DAn+19; ZND20; Raj+21; Raj+23; ZND21; Luo+22; Rah+22]. A detailed discussion on D-MIMO power control follows in the next section.

This collection of studies highlights the growing application of deep learning and RL techniques to optimize power control in modern wireless networks, addressing various performance objectives such as energy efficiency, sum-rate maximization, and QoS compliance.

4.6. Mobility management

Mobility management is essential in 5G and 6G to ensure seamless connectivity, low-latency communication, and efficient resource utilization as users move across network cells. With higher frequencies such as mmWave, signals are more susceptible to blockages and rapid degradation, making intelligent handover strategies crucial. High-speed applications such as vehicle-to-everything (V2X), smart cities, unmanned aerial vehicles (UAVs), and industrial automation demand ultra-reliable, low-latency connections, requiring advanced mobility prediction and optimization. Without effective mobility management, users experience dropped connections, increased latency, and degraded QoS, impacting the performance of real-time applications such as augmented reality (AR)/virtual reality (VR), autonomous driving, and mission-critical communications. 5G relies on enhanced handover techniques, beam tracking, and multi-connectivity, while 6G aims to integrate AI/ML-driven predictive mobility management for proactive handovers, reducing latency and failures. AI/ML models analyze user mobility patterns, signal quality, and network conditions to dynamically optimize handover decisions. This

intelligent approach enhances energy efficiency, network reliability, and user experience, making mobility management more adaptive in next-generation wireless networks.

3GPP introduced the conditional handover (CHO) in 5G to improve handover, which enhances mobility by allowing UEs to execute handovers when predefined conditions are met, reducing failures and ping-pong effects. However, CHO may lead to increased radio resource usage due to multiple handover preparations. AI/ML-based proactive handover mechanisms are studied in recent literature to overcome this challenge, exploiting user trajectory prediction [Zai+23; Ryd+23], measurement prediction [Lee+20], handover prediction [Lim+23], throughput prediction [Mas+21], etc. For example, the authors in [Lee+20] proposed a prediction-based CHO (PCHO) scheme that leverages deep learning to overcome the limitations of CHO and enable more intelligent preparation decisions. PCHO uses former blockage information, based on changes in BS signal patterns, to predict the best next BS to which to perform handover. The authors in [Par+24] also proposed a proactive handover approach for 6G networks, leveraging AI/ML-based time-series forecasting to predict optimal handover timing and target cells. Simulations show that the proposed method achieves nearly zero handover failures, effectively balancing resource efficiency and user throughput. In [Jan+25], a user-level mobility optimization mechanism is proposed to address recurring service interruptions by leveraging AI/ML to extract patterns of user behavior and paths where the problem occurs.

In the context of mobility management, a recent demonstration by the AI-RAN Alliance at MWC 2025 demonstrated the potential of beamforming for mobility-aware interference mitigation and power savings. To address the challenge of imperfect CSI due to mobility, a mobility-aware AI-based RAN intelligent controller (AI-RIC) that jointly optimizes beamforming and power allocation for interference mitigation and energy saving is developed by leveraging realistic RAN scenario data generations for both training and testing. The AI-RIC comprises a lightweight neural network and is trained using RL to take beamforming and beam-on/off actions.

5. Novel network architectures

Emerging network architectures such as D-MIMO, RIS, XL-MIMO, and 3D aerial platforms, including UAVs and high-altitude platforms (HAPS), are essential for realizing the ambitious goals of 6G and beyond, particularly in higher-frequency communications. Upper mid-band (FR3), mmWave, and sub-THz frequencies offer higher capacity than sub-6 GHz bands but suffer more path loss, limiting range. D-MIMO can overcome these challenges by utilizing geographically distributed transmission and reception points (TRPs), also known as access points (APs) to enhance coverage, improve reliability and provide spatial diversity, and improve interference handling, thereby enhancing the high capacity and spectrum efficiency of higher

frequency ranges. RIS can dynamically reflect and shape signals to optimize coverage and reduce energy consumption, which is useful in obstructed or complex urban environments.

Additionally, 3D aerial platforms provide flexible, on-demand coverage, enhancing network resilience and adaptability by adjusting their position based on user density and demand. However, managing these complex architectures efficiently requires real-time, intelligent decision-making. AI/ML solutions are essential for signal processing and resource allocation tasks, such as beam management and interference management. ML techniques, particularly RL and federated learning, enable distributed, scalable approaches to resource allocation, that adapt to rapid changes in the environment and user behavior. These AI-driven approaches optimize network performance, ensuring seamless connectivity and efficient resource utilization in the dynamic, high-frequency environment of 6G. This section provides an overview of novel network architectures that will be useful for AI-RAN and for AI/ML-based SOTA studies in the D-MIMO, RIS, and XL-MIMO research areas.

5.1. Distributed MIMO (D-MIMO)

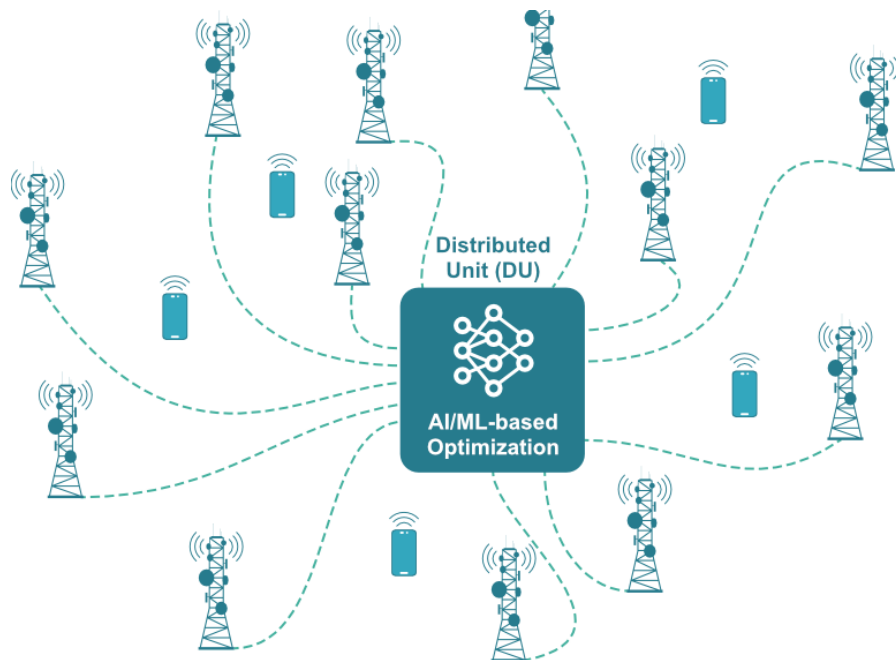


Fig. 6: D-MIMO with AI/ML-based optimization performed in the DU. This could be a central processing unit (CPU), a GPU, or any other computing block.

Cell-free massive MIMO, or D-MIMO, introduced in [Ngo+15], is a novel network paradigm to overcome the challenges associated with traditional cellular MIMO, such as poor QoS for cell-edge users and inter-cell interference. It can provide uniformly good throughputs for all the users by combining the advantages of both mMIMO and ultra-densification [DBS21; Ngo+15; Ngo+17;

Nay+17]. In D-MIMO, the TRPs are connected to a distributed unit (DU) or multiple coordinating DUs via fronthaul links, where the TRPs coherently serve all the UEs using the same time-frequency resource [Ngo+17]. In addition to the beamforming and spatial multiplexing gains, this signal co-processing across multiple distributed TRPs provides several additional benefits: macro-diversity gains, coordinated interference management, and more uniform QoS among users [Int20]. With these gains, the D-MIMO architecture is a promising candidate for many emerging applications that require stringent requirements such as massive connections, low latency, high reliability and resilience, and high-speed mobility. PHY and MAC layer optimization for D-MIMO has been widely studied over the years, developing an extensive line of research and development parallel to conventional mMIMO technology. The signal processing and radio resource allocation concepts used in co-located mMIMO are adapted to D-MIMO, with modifications and improvements to suit the cell-free architecture.

Effective and efficient radio resource allocation is critical for proper D-MIMO implementation to reap the full benefits of this novel network architecture, including user association and TRP selection, pilot assignment, power control, TRP switch on/off, etc. In recent literature, different AI/ML approaches have been utilized to overcome the computational complexity and performance limitations in conventional techniques for solving these resource allocation tasks. A D-MIMO network with AI/ML optimization performed in the DU is illustrated in Fig. 6. Depending on the scenario, AI/ML-based optimization could also be implemented at TRPs. Various supervised learning [DAn+19; ZND20], RL [ZND21; Luo+22; Rah+22], and unsupervised learning [Raj+21; Raj+23] techniques were proposed for D-MIMO power control to achieve different optimization objectives, such as sum-rate and min-rate maximization. Supervised learning-based approaches performed worse than the baseline in shadow fading and mobile scenarios [DAn+19; ZND20].

On the other hand, unsupervised learning-based techniques directly train a DNN using the optimization objective as the custom loss function to learn the resource allocation outputs, eliminating the need for labeled data for model training, and producing results similar to the baseline [Raj+23]. Different DRL techniques used for power control include DQN [ZND21], DDPG [Luo+22], and twin delayed DDPG [Rah+22]. Furthermore, distributed power control is investigated to reduce the signaling overhead and complexity [CBS19; Zah+23; LKC18]. The work in [Zah+23] investigated downlink power control for sum-rate maximization with proportional fairness using a distributed DNN approach based on local channel statistics between the base station and different UEs, reducing fronthaul signaling. The performance is further improved by providing side information to the DNN. Furthermore, [LKC18] proposed targeted power control for sum-rate and energy-efficiency maximization in a device-to-device network using a CNN.

The increased number of antennas in the D-MIMO system significantly increases the complexity and overhead of the optimization problems to be addressed, underscoring the need for efficient, scalable solutions. For example, the overhead for beam selection in analog beamforming systems becomes severe in D-MIMO networks with a large number of beamforming APs. A novel beam selection method is proposed in [Mer+23] based on the CS concept to reduce beam selection overhead, which is further optimized by using learned beam dictionary and neural sparse decoders. The proposed method reduces overhead and, consequently, the latency of beam selection.

DRL techniques for TRP selection and user clustering in cell-free mMIMO/D-MIMO networks include [Ghi+23; MSR23; TAK22; CHC21]. The authors in [Ghi+23] used DRL for TRP selection to perform user association in a cell-free mMIMO network, maximizing energy efficiency while satisfying minimum rate constraints for all UEs. The challenge proposed by a large discrete action space in the optimization problem is overcome by approximating it with either a continuous set or a smaller discrete set, and by modifying existing DRL methods accordingly. A DDPG-based approach to the joint cooperation clustering and content caching problem in cell-free mMIMO is proposed in [CHC21]. The proposed approach is shown to achieve close performance to the optimal exhaustive solution in small networks and to outperform SNR-based clustering and popularity-based caching in large networks thanks to its joint design. A flexible DRL-based user-centric clustering approach, where each user is served by only a subset of TRPs to achieve sum-rate maximization or user QoS requirements is proposed in [MSR23] and allows different-sized cluster formation.

5.2. Reconfigurable Intelligent Surfaces (RIS)

RIS, also known as intelligent reflecting surfaces, are passive planar arrays formed by many passive elements. By dynamically adjusting the phase, amplitude, or polarization of incoming signals, RIS can improve signal strength, reduce interference, and optimize propagation paths without requiring additional active power. RIS-assisted communications are useful in higher-frequency bands, such as mmWave, to compensate for propagation loss and overcome blockages in NLoS scenarios.

Although RIS offers significant advantages, its signal processing and implementation present several challenges that need to be addressed to fully utilize its potential. Obtaining accurate CSI of the cascaded channel in real time is challenging due to the large number of propagation paths caused by the large number of antennas in the RIS in addition to the direct path. The cascaded channel makes it inherently difficult to estimate the individual channels (transmitter-to-RIS and RIS-to-receiver), and the large number of passive elements further complicates RIS channel estimation compared to traditional MIMO. Scalability is another issue, as increased

computational complexity makes real-time RIS optimization and control challenging. Hardware limitations and integration with existing networks are other concerns.

Recently, the literature has shown some AI/ML techniques that have been used to overcome the above challenges to a certain extent. A comprehensive survey of AI/ML-based RIS techniques is presented in [FC22] under different areas such as channel estimation, beamforming, energy efficiency, resource management, detection, and security. RIS channel estimation studies in the literature include CNN-based deep learning techniques [Liu+20; Elb+20; Zha+21] and residual learning techniques [Jin+21; Liu+21; Liu+22b]. Federated learning techniques are also studied extensively, for example in [EC22; Qiu+24; SQN22; Das+24], due to their ability to minimize transmission overhead [FC22]. Considering RIS phase shift/reflection coefficient optimization, the work in [Hua+19; Ale+20] utilized supervised learning techniques while [Per+23; Ale+22; Has+22] proposed deep RL approaches. Studies [Gao+20; Son+21b] proposed unsupervised learning approaches to overcome the labeling overhead of supervised learning.

A recent study proposed RIS phase optimization using channel-chart-based generative flow networks to overcome the high training dataset requirements and generalization challenges in supervised learning and the slow convergence and unstable performance arising from large action spaces in RL [BB24]. [Li+23b] proposed a double deep learning network for joint optimal phase shift and beamforming in the RIS-assisted MIMO network that is superior to reinforcement learning in the RIS joint optimization. A recent survey paper [ZHL24] discusses RIS-aided 6G network optimization in detail, with a focus on different RL techniques such as deep Q-learning, multi-agent RL, transfer RL, hierarchical RL, and also how large language models (LLMs) can be combined with RL to handle network optimization problems. The integration of RIS into O-RAN architectures is discussed in [Tsa+25; TM24], where AI-powered control of the BS's scheduling policy, along with spectrum allocation strategies and RIS deployments, is used to enhance network performance across eMBB and URLLC slices.

5.3. Extremely large-scale MIMO (XL-MIMO)

Extremely large-scale MIMO (XL-MIMO) refers to a massive extension of traditional MIMO technology consisting of thousands of antenna elements with extremely large antenna array dimensions often spanning physical scales that cover a substantial area. To fully leverage the system's spatial resolution and performance benefits, users must also be within the near-field of the physical aperture of the array. This concept arises in the context of next-generation wireless networks, such as 6G, and is envisioned to meet ultra-high data rate, reliability, and connectivity demands. Besides the very large antenna dimensions, the main differences between traditional MIMO and XL-MIMO lie in signal propagation and the channel model. XL-MIMO operates in the near-field region, requiring spherical wave models due to user proximity to parts of the array instead of the far-field planar model [Lu+24]. This near-field effect results from the

large array aperture in XL-MIMO, due to the extremely large number of elements in the antenna array, and is further enhanced by the shorter wavelengths of high-frequency communications, such as in FR2 and FR3. This, along with hardware and signal processing complexity challenges, requires novel techniques to enable XL-MIMO. The authors in [Yu+25] discussed principles and techniques for using deep learning for near-field XL-MIMO transceiver design, focusing on developing scalable, low-complexity, and robust algorithms that can efficiently characterize and leverage the properties of the near-field channel.

The benefit of near-field beam focusing relies on accurate channel estimation, utilizing the sparsity in both angle and distance dimensions. To this end, several AI/ML-based approaches for estimating the near-field channel of XL-MIMO systems are investigated in the literature [Zha+23a; Lee+22; Yu+22; Lei+24]. The authors in [Zha+23a] proposed a model-based deep learning approach to enhance the polar-domain estimation accuracy. CNN-based and fixed-point network-based deep learning approaches have been applied to the near-field channel estimation in THz XL-MIMO systems in [Lee+22; Yu+22]. A recent work in [Gao+24] proposed the XL-MIMO channel network (XLCNet) to estimate high-dimensional channel, serving as a universal solution for both near-field and far-field users with different channel statistics. Weight pruning and quantization are further used to design a compressed XLCNet (C-XLCNet) to accelerate the model inference and facilitate the model storage and transmission. In addition to the channel estimation complexity arising from the spherical wave characteristics in the near-field channel, the larger antenna array scale also significantly increases the size of the CSI matrix, making CSI feedback in the near-field channel highly challenging. To this end, the authors in [Pen+24] proposed a deep learning-based ExtendNLNet that can compress the CSI and further reduce the overhead of CSI feedback in XL-MIMO.

Codebook-based beam training is a technique with potential to improve beamforming gains in XL-MIMO systems. However, the near-field beam training requires information on both angle and distance, which significantly increases the beam training overhead. To this end, the authors in [Liu+23a] proposed a deep learning-based beam training scheme that uses the received signals corresponding to the far-field wide beams to estimate the optimal near-field beam. The authors in [Nie+25] also proposed a near-field beam training method based on deep learning, where a CNN is trained using the negative of the user-average achievable rate as the loss function to optimize the beamformer, thereby maximizing the rate in multi-user networks without relying on predefined beam codebooks. Finally, [Liu+24a; Liu+24b] presented investigations on combined cell-free and XL-MIMO architectures. Specifically, [Liu+24a] proposed the enhanced cell-free XL-MIMO paradigm, which incorporates multi-agent reinforcement learning (MARL) to provide a distributed strategy for addressing the challenges of high-dimensional signal processing and costly energy consumption. [Liu+24b], on the other hand, proposed a MARL framework for power control in a cell-free XL-MIMO network.

Finally, novel Transformer-based techniques have also been applied to XL-MIMO since self-attention effectively captures long-range spatial dependencies across large antenna arrays and models complex spatial-temporal correlations, including near-field effects. Moreover, transformers can handle high-dimensional, irregular channel data and enable parallel computation for faster training and real-time inference. To this end, the authors in [Li+25] proposed NEFT, a unified Transformer framework for efficient near-field CSI feedback in XL-MIMO systems. Built on a hierarchical Vision Transformer backbone, NEFT is extended with lightweight variants to suit different deployment constraints: utilizing multi-level knowledge distillation (KD) to reduce complexity and via encoder- and edge-constrained scenarios through attention-free encoding combined with KD. A recent study [LD24] proposed MAT-CENet, a Mixed Attention Transformer-based neural network for XL-MIMO channel estimation that combines feature-map and spatial attention mechanisms to capture XL-MIMO channel characteristics. Using multi-head attention, the network effectively identifies and leverages important features for improved estimation. Finally, the authors in [Cao+25] utilized last-bounce cluster-based channel model to capture spatial-temporal characteristics in XL-MIMO systems for channel prediction and proposed a CNN-Transformer-based channel prediction method to improve spatial-temporal information extraction.

6. Sustainability and energy efficiency enhancements via AI-RAN

Given the design criteria of 6G, sustainability plays a more prominent role than in any previous generation [Mat+20; Ahm+23]. Since about 95% of mobile radio network product life-cycle emissions occur during use, minimizing energy consumption during operation is key to reducing both emissions and operational expenditure [LL23]. In current mobile networks, most energy is consumed by RAN, underscoring the importance of RAN energy-efficiency enablers. To this end, developing low-power hardware components as well as designing efficient signal processing and resource allocation mechanisms to reduce power consumption while achieving the user QoS requirements is important. Moreover, low-complexity algorithms that can be implemented with minimal resources are beneficial for further improving energy efficiency. Some of the energy-saving RAN software features proposed by the industry include cell switch on/off, dynamic baseband power management, and mMIMO muting [Nok23]. To this end, AI/ML-based approaches are promising candidates for providing low-complexity, energy-efficient solutions to radio resource allocation problems [Hus+20].

6.1. Cell energy saving

BS sleeping and switching off are widely adopted mechanisms to reduce energy consumption in RANs. These approaches rely on putting BSs into various sleep modes (SMs), each characterized by the degree of hardware dormancy, the corresponding energy savings, and the latency needed to restore full functionality [3GPP23]. Depending on the selected SM, BSs can deactivate specific hardware components, such as power amplifiers, mixers, digital-to-analog converters, or the entire RF chain, substantially reducing energy consumption. However, reactivating these components incurs latency, which varies by component. For instance, waking up an RF frontend may take several hundred milliseconds, while temperature-compensated crystal oscillators can require additional time to reheat.

Table 4 summarizes the energy-saving potential and recovery latencies of different BS sleep modes. In general, BS sleeping techniques can be categorized into two timescales:

- **Short-term sleeping:** Short-term sleeping typically operates on a near-real-time timescale, with reactivation delays ranging from a few microseconds (shallow modes) to hundreds of milliseconds (deep modes). These decisions are typically made locally by the BS. For example, a BS in a particular sleep mode will immediately reactivate upon receiving traffic from its associated UEs, without requiring coordination with neighboring BSs.
- **Long-term sleeping:** Long-term sleeping involves delays in the order of seconds (up to minutes). These approaches are typically managed at the network level and involve macro BSs or edge clouds.

Table 4: Descriptions of various BS SMs.
The numbers can vary depending on the specific BS technology and 3GPP standards.

Short-Term Sleeping		
Sleep Mode	Switch Off Components	Description
Micro sleep	Power amplifier	This is the shallowest but fastest SM. The duration of this SM is at the level of the symbol time. Once a BS has no traffic on some subcarriers, it can switch off its power amplifier for a symbol time.
Light sleep	Analog frontend such as a mixer, a digital-to-analog converter (DAC), and a frequency synthesizer	A slightly deeper sleep level, with a delay of the order of milliseconds.
Deep sleep	All digital baseband processing as well as analog frontend components	This mode has a deeper sleep level, requiring a minimum duration of tens of milliseconds.
Long-Term Sleeping		
Sleep Mode	Switch Off Components	Description
Carrier shutdown	Power amplifier, TRX, DPD chips, and some digital basebands.	Reactivation completes within seconds, but network-level coordination typically imposes guard times of minutes.
Hibernate	All Radio components, including the power amplifier, baseband processing, and analog frontend.	This mode is the deepest SM, requiring a duration of minutes.

On short-term sleeping, [Jan+20] introduced an ML approach combining LSTM-based traffic prediction with a Lyapunov optimization-driven BS switching algorithm. The paper reports around a 15% reduction in total BS power consumption via SM optimizations while maintaining queue stability. However, the switching and SM selection decisions are obtained via heuristic optimization. The authors in [Mas+23a] adopted a different approach, modeling multiple short-term SMs as a sequential Markov decision process. They developed an online, traffic-aware deep Q-learning technique to maximize long-term energy savings and introduced a digital-twin-based framework to start model retraining in response to dynamic network conditions. In the context of long-term sleeping, in [YZ20], the authors propose DRAG, a DRL framework for activating and deactivating small base stations in heterogeneous networks. By formulating the activation problem as an MDP and employing an actor-critic algorithm alongside a dedicated action-refinement scheme, DRAG incorporates arrival-rate predictions, switching costs, and service latency to minimize cumulative energy consumption. Additionally, various approaches have sought to optimize RAN configuration to reduce energy consumption without compromising key performance indicators (KPIs). For example, [Gan+23a] proposed a branching soft-actor critic (BSAC) algorithm aimed at meeting reliability requirements through optimized

power allocation across physical resource blocks and the orchestration of diversity transmissions, and notably reduced transmit power consumption relative to baseline methods that offer similar levels of reliability. Building on this work, [Shi+23] proposed a hierarchical SAC framework that further reduces energy consumption by reducing the number of signal exchanges required during training. [Lop+22] provided a comprehensive overview of key technologies that enable energy efficiency in RAN.

6.2. Transmit antenna muting

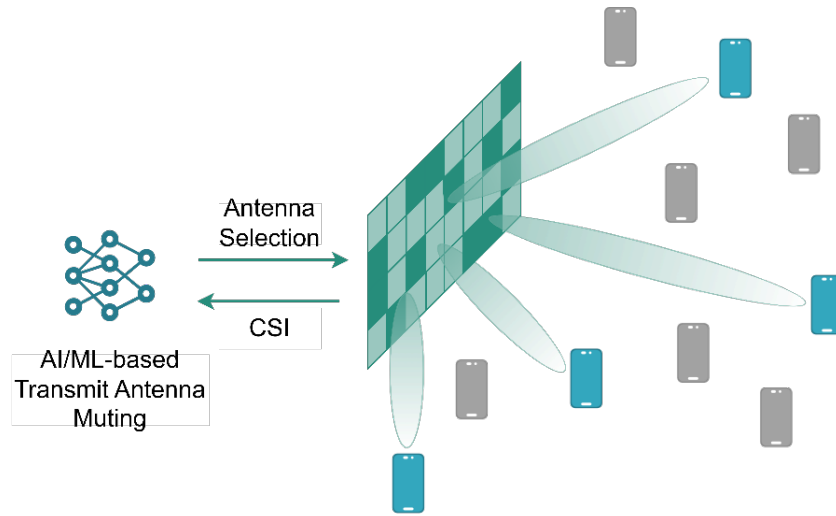


Fig 7: AI/ML-based transmit antenna muting in downlink MU-MIMO with user scheduling (adapted from [Raj+24]).

While the advantages of mMIMO systems are rooted in the use of a large number of antennas [HBD11], this large number of antennas, their corresponding power amplifiers, and RF chains lead to significant power consumption, system complexity, and cost. Transmit antenna muting (TAM) is a technique that reduces the BS power consumption by carefully deactivating a subset of antennas while satisfying the desired system performance. In addition to improving energy efficiency, TAM helps mitigate interference, improve the SNR, reduce signal processing complexity, and reduce overall system complexity. ML-based TAM approaches have been investigated to achieve a trade-off between the computational complexity and the system performance. An example AI/ML-based TAM approach is illustrated in Fig. 7.

The authors in [Jou16; Yan+19b] proposed TAM as a multi-class classification approach using SVMs and the KNN method for a single-user MIMO system. A joint TAM and hybrid beamforming design for single-user mmWave MIMO is proposed in [EM20], which uses two serial CNNs. The authors in [lbr+18] proposed a joint multicast beamforming and TAM approach that learns the mapping between the CSI and a subset of antennas that maximizes the minimum SNR at the receivers in a multicast setup. Joint antenna selection and precoding design algorithm to maximize the system sum rate of the multiuser MIMO downlink, subjected to a transmit power

constraint and QoS requirements, is proposed in [Vu+21]. Furthermore, in [YWW21], the authors proposed a low-complexity TAM algorithm for MU-MIMO based on multi-label learning, where a DNN is employed to determine the set of selected antennas for a given channel matrix. The proposed multi-label learning network combined deep canonical correlation analysis with an autoencoder to extract low-dimensional features from the channel matrix and the interdependencies among selected antennas thereby accurately predicting the antenna set with a relatively small-scale learning mode. The authors in [Raj+24] considered a different optimization problem for TAM. There, an ML-based TAM algorithm for MU-MIMO downlink is proposed to minimize the number of active antennas at the BS while guaranteeing per-user throughput requirements, where supervised learning and a novel custom loss function, are used to ensure user QoS requirements. In [HK21], the authors proposed using UE location information to train an RL agent to configure the antenna array in a MU-MIMO system. The solution was then adapted to the grid of beams beamforming in [HK24].

7. AI-RAN network management

AI-driven solutions have arguably become fundamental in optimizing key network operations and management. With the advent of 5G and the emergence of 6G networks, AI/ML models are being employed to enhance decision-making processes, automate resource allocation, and improve overall network performance. These intelligent systems can reduce human intervention, allowing for self-optimizing and self-healing networks that can dynamically adapt to evolving conditions.

7.1. Load balancing and traffic steering

The rapid evolution of wireless communication technologies has emphasized the critical role of load balancing and traffic steering in ensuring network efficiency and QoS. As networks grow increasingly dense and heterogeneous with the advent of 5G and 6G, leveraging AI and ML for dynamic traffic management has become a cornerstone of modern network optimization. The integration of AI-driven solutions into load-balancing and traffic-steering processes has demonstrated remarkable potential for addressing the complexities of heterogeneous networks. In [NA+24], the authors introduced an ML-based traffic steering model for 6G networks that uses historical data to classify traffic and allocate resources optimally, leveraging mixed-integer linear programming and multi-output classification models. Authors in [Hab+24] proposed a hierarchical deep Q-learning (h-DQN) framework for traffic steering. This bi-level architecture decomposes decision-making into high-level policies and low-level actions, significantly enhancing adaptability and convergence efficiency in dense 5G networks. Furthermore, [Max+24] presents ScalO-RAN, a framework for energy-aware scaling and placement of AI-based applications in O-RAN, leveraging models like CNNs, LSTMs, and DRL agents to optimize

resource usage while meeting latency constraints across heterogeneous RAN nodes. A DRL-based traffic steering mechanism is proposed in [She+22c] for per-user mobility management in ORAN, leveraging CNN-powered Conservative Q-learning and Random Ensemble Mixture models to improve throughput and spectral efficiency by up to 50% in large-scale ns-3 simulations. The paper [Bon+21b] presents a slice-aware scheduling use case in O-RAN, where a DRL agent based on the PPO algorithm is deployed in the near-real-time RIC to autonomously optimize resource allocation across network slices.

Small cells have emerged as pivotal components for managing traffic load in ultra-dense networks. The authors in [Cho+24] introduced a practical load-balancing algorithm for 5G small cell networks to dynamically adjust small cell coverage based on real-world traffic data, ensuring fairness and improving resource utilization. Unlike conventional static biasing methods, their approach integrates traffic prediction to dynamically adjust the bias of small cells based on predicted loads. By leveraging daily traffic patterns, this algorithm has the potential to significantly reduce load imbalance and enhances resource utilization while maintaining computational efficiency and, therefore, could achieve superior performance compared to conventional techniques like cell range expansion.

Traffic steering in heterogeneous ultra-dense networks has also been enhanced by advanced ML models. Authors in [Cha+23] developed a bidirectional LSTM model that predicts QoS based on channel state information. This model enables proactive resource allocation, preserving QoS across multiple heterogeneous links, including 3GPP and non-3GPP technologies, even under varying network conditions. In addition to these network-centric approaches, cloud environments can provide valuable insights into load balancing.

Authors in [Kav+24] introduced a DRL-based framework combined with multi-agent learning to optimize traffic flow across dynamic environments, achieving significant improvements in eMBB throughput and URLLC latency. Similarly, [Gan+25] developed a DRL-based solution that steers distributed learning traffic via device selection, aiming to reduce convergence time while minimizing disruptions to the coexisting URLLC service. The authors in [Tra+24] proposed a computing-aware traffic steering architecture that leverages dynamic data from edge servers to steer traffic efficiently, resulting in reduced latency compared to traditional methods. GNNs have also been employed to optimize traffic steering by recognizing intricate relationships among network topologies, traffic features, and routing patterns. In [Raf+23], the authors demonstrated the utility of GNNs in service function chaining, achieving a 20% improvement in resource utilization and a 5% increase in throughput. For predictive traffic management, LSTM networks have proven effective in forecasting dynamic traffic demands. Studies by [Kav+24] integrated LSTM models into RAN slicing frameworks to predict traffic patterns and allocate resources dynamically, thereby ensuring seamless service amid competing demands.

Fair traffic steering remains a challenge in 5G systems due to network densification and heterogeneous traffic types. The authors in [Tha+22] addressed this by proposing Balanced5G, a data-driven framework that integrates supervised learning techniques to distribute traffic equitably across cells, thereby reducing network congestion and improving QoS compared to conventional models.

Additionally, hierarchical learning approaches have emerged as a robust methodology for multi-layered decision-making in traffic steering. [Hab+24] introduced an h-DQN model for O-RAN, which decomposes complex traffic management tasks into high-level policies and low-level actions, achieving faster convergence and better network adaptability. The hierarchical traffic steering in O-RAN for the V2X scenarios was also demonstrated in [Sro+24]. Finally, experimental evaluations of ML models, such as those by [Cha+23], underscore the importance of selecting suitable algorithms for specific traffic scenarios. Their work evaluated 22 ML models in a real-world 5G environment, providing critical insights into model performance for traffic prediction and steering.

7.2. Anomaly detection, fault prediction, and mitigation

Anomaly detection has become a critical component for enhancing the performance and reliability of the RAN, particularly as 5G and Beyond 5G systems introduce greater complexity. A significant body of research has focused on leveraging AI and ML techniques to address these challenges. For instance, federated learning frameworks have gained prominence for enabling collaborative anomaly detection while preserving data privacy. The TenaxDoS framework, which combines federated learning with continual learning, has demonstrated remarkable success in mitigating catastrophic forgetting, maintaining over 98.8% accuracy in detecting distributed denial-of-service (DDoS) attacks in RAN environments [BHT+24]. Similarly, research integrating LSTM auto-encoders within network data analytics function (NWDAF) systems has shown promising results in identifying traffic anomalies using real-world network data [Mek+23]. GAN-based models have also emerged as a robust method for detecting anomalies in dynamic RAN environments by adapting to model drift and reducing false positives [Gud+24].

Furthermore, bi-level federated GNNs could effectively address the interconnected nature of telecom systems, capturing hierarchical relationships between software execution and node interactions [Bou+23]. As another good enhancement of anomaly detection, AI-based classifiers play a crucial role in detecting abnormalities such as discontinuous transmission (DTX) and unintended transmission (UTX) [Jan+24]. These advancements highlight the potential of ML-based methods in enhancing the adaptability and robustness of RANs.

MEC has emerged as a pivotal enabler for real-time anomaly detection in wireless RANs. By processing data closer to the source, MEC architectures effectively reduce latency and improve

responsiveness. Studies leveraging MEC for distributed anomaly detection have introduced innovative systems that combine flow collection, symptom detection, and decision-making modules at the network edge [Sha+24]. The SpotLight system, which integrates deep generative models across edge and cloud, provides efficient and explainable anomaly detection for RAN environments, utilizing methods like JVGAN and MRPI for distribution learning and anomaly confirmation [Sun+24]. Similarly, platforms like Deep RAN employ recurrent convolutional neural networks to identify anomalies in live cellular networks, enabling real-time anomaly management across multi-technology systems [Tan+20]. The abnormal load detection in the network control plane, i.e., the detection and mitigation of signaling storms, is shown in [HK23b]. Explainable AI methods also contribute to transparency, as shown in unsupervised approaches for anomaly localization in virtualized environments [Joh+22; ElB+23].

Building on these foundations, a recent analysis [YSS24] underscores how the disaggregated nature of O-RAN, combined with software-defined networks (SDN), network functions virtualization (NFV), and AI/ML-driven RAN intelligent controller (RIC) components, heightens the risk of configuration conflicts and unintended behavior across xApps and rApps. This study highlights the potential of AI/ML not only to optimize RAN management but also to detect and proactively resolve misconfigurations, thereby preventing service degradation. A case study within the same work illustrates how policy conflicts among xApps can adversely affect user experience and proposes AI-based mechanisms for conflict mitigation—marking a critical step toward safer, more autonomous RAN environments.

In the domain of runtime threat analysis, the 6G-XSEC framework [Wen+24] introduces an innovative approach to anomaly detection at the cellular edge. By fusing lightweight unsupervised deep learning models with LLMs, it enhances the O-RAN control plane with real-time analytics and explainability. This methodology demonstrated 100% detection accuracy across five documented cellular attacks and validates the use of LLMs in real-time, interpretable network security, advancing the vision for trustworthy AIOps in future 6G infrastructures. Complementing these efforts, the FALCON framework [Kum+25] offers proactive fault prediction in virtualized O-RAN deployments. By collecting and analyzing telemetry from infrastructure, platform, and RAN layers, and applying a random forest-based classifier, FALCON anticipates faults with over 98% accuracy and F1-score. This multi-level AI/ML strategy empowers operators to pre-emptively address system vulnerabilities, significantly enhancing network resilience and operational efficiency.

To address the scarcity and spatio-temporal modeling challenges in fault management, the Simba framework [Has+24] combines GNNs and transformers for robust anomaly detection and root-cause analysis in 5G RANs. Using the Simu5G simulator, Simba creates synthetic but realistic datasets for both normal and failure scenarios, facilitating comprehensive training. This

hybrid approach substantially outperforms LTE-focused solutions, enhancing fault diagnosis accuracy and supporting the development of self-healing networks.

Finally, managing conflicts among independently operating xApps in O-RAN's near-real-time RIC remains a significant challenge. To this end, recent research [Zol+25] introduces a data-driven method for dynamic conflict graph reconstruction using GraphSAGE, effectively uncovering hidden dependencies between xApps, parameters, and KPIs. In tandem, the GRAPHICA framework [SYF25], built on GNNs, predicts and analyzes three types of conflicts—direct, indirect, and implicit—while identifying responsible xApps. Achieving over 98% F1-score in imbalanced datasets, GRAPHICA exemplifies the potential of scalable, AI-powered conflict management for stable and intelligent multi-agent RAN systems.

Despite these significant strides, challenges remain in scaling these solutions to accommodate the diverse and dynamic nature of RANs. Issues such as the computational overhead of ML models, ensuring privacy in distributed environments, and optimizing resource utilization at the edge continue to hinder widespread adoption. Unsupervised learning techniques for throughput enhancement demonstrate the ability to identify performance bottlenecks and improve service quality [ElB+23].

In addition, RL-based frameworks for scalable RAN optimization, which incorporate anomaly detection for task grouping, highlight the potential for intelligent load balancing and resource allocation [Li+24a]. Future research must focus on overcoming these obstacles while exploring novel technologies like blockchain for secure data sharing and adaptive MEC infrastructure to support dynamic wireless networks. By integrating these approaches, wireless RANs can achieve improved reliability, scalability, and performance in increasingly complex network environments. The authors in [Gro+24a] proposed TIMESAFE, a security monitoring framework for O-RAN fronthaul synchronization, leveraging Transformer- and CNN-based machine learning models—achieving over 97.5% detection accuracy—and demonstrating significant improvements over rule-based methods in detecting spoofing and replay attacks on the precision time protocol.

7.3. Slice service level agreement management

ML methods for network slice management can enhance the efficiency and adaptability of 5G and beyond networks. A comparative study in [Tsa+24] evaluates 12 DRL-based algorithms for slice-based resource allocation in RAN, showing how reward shaping, action spaces, and hierarchical control strategies affect QoS trade-offs across eMBB, URLLC, and mMTC slices. RL has been a prominent approach, particularly in dynamic resource allocation for slice reconfiguration. For example, the AI-driven network slice management framework proposed in [Tuf+23] integrates a multi-agent deep deterministic policy gradient RL algorithm. This approach

predicts traffic patterns and allocates resources adaptively to improve QoS, reduce latency, and optimize resource utilization. By incorporating RL, the framework enables real-time adjustments to network conditions and traffic demands, ensuring seamless slice management. In [Cam+23], ML models are used for traffic prediction and capacity forecasting, leveraging CNNs and RNNs. These models enable proactive slice reconfiguration by predicting resource demands with up to 91% accuracy. The application of hysteresis rules further refines these predictions, reducing unnecessary resource adjustments and ensuring compliance with service level agreement (SLA) requirements. In [Pan+25], the authors present a proof-of-concept resource management scheme for augmented reality in 5G MEC that adapts network and computing resources to meet strict delay constraints. Their approach splits the round-trip delay budget into uplink, downlink, and edge segments and uses a multi-armed bandit algorithm to optimize resource allocation, reducing both bandwidth usage and power consumption while maintaining high QoS. A DRL model is also employed in [Raf+24] to manage latency-aware network slicing under time-varying service level agreements, dynamically allocating PRBs to minimize SLA violations and improve resource efficiency. Such predictive capabilities allow mobile network operators to allocate computational resources more effectively, optimizing both performance and cost-efficiency.

The MonB5G framework, as described in [Kuk+21], utilizes distributed AI control loops to enable intent-based slice reconfiguration. This framework integrates unsupervised learning methods for anomaly detection and supervised learning for SLA compliance monitoring, ensuring scalability and fault tolerance. Similarly, the integration of transfer learning into ML-based slicing frameworks, as discussed in [Yeh+23], improves traffic load predictions and resource distribution, supporting dynamic network conditions and enhancing multi-tenant slice deployments. [DO24] introduces OrchestRAN, a framework for orchestrating AI/ML-driven functionalities in disaggregated RANs by automatically selecting, placing, and instantiating models across network nodes based on timescale, resource, and data availability constraints, enabling scalable and latency-aware network intelligence. An explainability framework is also proposed in [Rob+24b] to enhance the interpretability of DRL-based control in slice-aware RAN environments, using attributed graphs to analyze and improve decision-making transparency under dynamic conditions. A data-driven SLA control framework is presented in [Gem+25], enabling dynamic adaptation of RAN configurations through real-time monitoring and optimization to enforce soft and strict bitrate guarantees across network slices.

Latency is a critical parameter in SLAs, especially in industrial automation, where guarantees such as 10 ms latency at 0.9999 reliability are required. Probabilistic latency prediction has gained attention to address these stringent demands. For instance, in [FYJ20], the authors employed random forests and a histogram-based approach to condition delay distribution predictions on contextual variables such as position, time, radio channel, and signal power in IoT

networks. In [MSG23], mixture models were applied to real 5G delay measurements, using MCS indices as conditioning variables, demonstrating that mixtures of Gaussians and extreme-value theory models with neural networks can effectively capture rare latency events. This emerging trend in temporal probabilistic prediction is a new direction toward more confident predictions, as exemplified by [Sko+23], which leveraged RNN and LSTM models on mobile network operator data to derive conditional delay probability density functions that account for network and traffic conditions. These dynamic predictions can be integrated into network optimization frameworks to efficiently achieve SLA targets.

These methods collectively demonstrate the power of ML to transform network slice management by enabling predictive analytics, dynamic resource allocation, and enhanced operational efficiency. Through innovative ML techniques, these frameworks address the challenges of scalability, SLA adherence, and real-time adaptability in complex network environments.

8. Integrated sensing and communication

Integrated sensing and communication (ISAC) and high-precision localization are essential for 6G, enabling context-aware, ultra-reliable, and low-latency applications such as autonomous vehicles, smart cities, industrial automation, extended reality (XR), and advanced IoT. ISAC enables networks to simultaneously transmit data and sense the environment, enhancing situational awareness and enabling new capabilities such as wireless imaging, gesture recognition, and environmental mapping. The idea of ISAC focusing on joint communication and sensing lets us reuse the existing waveforms, hardware, sites, and spectrum. Location is a key enabler for a plethora of new use cases of interest to telecom. As 6G integrates location-based services as a core feature, achieving centimeter-level accuracy becomes crucial for seamless operation. However, traditional signal processing struggles with complex, dynamic environments, making ML a potential enabler for enhancing sensing accuracy, multipath mitigation, adaptive beamforming, and real-time localization. ML-driven models can improve positioning accuracy, resource allocation, and robustness in challenging conditions, ensuring efficient and intelligent 6G networks. In this section, we briefly discuss state-of-the-art studies under AI/ML-based ISAC and positioning/localization.

8.1. Integrated sensing and communication

ISAC refers to the convergence of wireless communication and sensing functionalities within a single system, sharing spectrum, hardware, and infrastructure [Wan+22c]. It has been identified as a key pillar of 6G due to its potential to maximize spectrum usage and reduce hardware costs and power consumption [WBV21; Liu+22c; Fan+23]. One of the primary differentiators of 6G

compared to 5G will be its emphasis on radio-based positioning and sensing. These capabilities are expected to support not only conventional location-aware services but also new use cases with extreme performance requirements [Beh+23].

Furthermore, the sensing and communication functions can benefit from each other. This integration is essential for applications such as autonomous vehicles, smart cities, and next-generation wireless networks, where communication and environment perception must work seamlessly together.

The potential of AI/ML techniques for extracting and leveraging temporal/spatial patterns, approximating complex models, and solving difficult optimization problems has also been shown to be beneficial for ISAC systems. AI/ML techniques can intelligently manage the integration of communication and sensing, ensuring efficient, adaptive, and high-performance operation in diverse and dynamic environments [DA23]. These methods can learn directly from the data, achieve better performance, and be useful even under model deficits. Specifically, the authors in [DA23] discussed roles and methods for utilizing AI/ML for joint sensing and communication systems under three main directions: joint sensing and communication, sensing-aided communication, and communication-aided sensing. For ISAC, the identified potential areas for AI/ML include waveform design optimization, learning site-specific spatial sensing and communication beam patterns, self-interference cancellation at ISAC transceivers, wireless and system resource allocation optimization, ISAC system security enhancement, and enabling network-level operation of ISAC systems [DA23]. A comprehensive survey of ML-enhanced ISAC systems, covering architectures, algorithms, and applications, is presented in [AL24].

The studies in [Liu+22d; Che+24; Liu+23b; Liu+23c] proposed different AI/ML techniques for channel estimation in RIS-assisted ISAC systems. Specifically, [Liu+22d] presented a DL framework consisting of two DNNs deployed at the ISAC BS and assigned to each downlink UE to estimate sensing and communication channels, respectively. It is extended to a multi-user scenario in [Liu+23c] using a neural network framework consisting of two extreme learning machines in the downlink user and the ISAC BS to estimate the sensing and communication channels, showing performance improvements compared to least squares, CNN, and feedforward neural network approaches, with reduced training complexity and faster training time. In [Liu+23b], a CNN-based three-stage approach is proposed to decouple the estimation problem into sub-problems, including the estimation of direct sensing and communication channels in the first stage, reflected communication channel in the second stage, and the reflected sensing channel in the third stage.

In [Qi+23; Qi+24], a deep learning solution “ISACNN,” based on unsupervised learning is proposed by jointly optimizing the sensing signal waveform and the communication receive beamforming from the perspective of weighted-sum sensing rate and communication rate

maximization. A DRL approach to designing an automotive radar system with ISAC is proposed in [XZS22], where sparse transmit arrays with quantized phase shifters are used for transmit beamforming to enhance both radar sensing and communication performance. Several recent studies also explored DRL approaches for optimizing radio resources for radar services in ISAC systems [Sme+24], reconfigurable beamforming for automotive radar sensing and communication [Xu+24], and related applications. The work in [Mat+22; MS23] jointly learned beamformers, target-detection mapping, and receiver processing within a deep end-to-end autoencoder model, focusing on fully digital MIMO with a single receiver. Hardware impairments can degrade the performance of ISAC systems. In [Riv+23; Mat+22], the authors provide an end-to-end beamforming design using AI/ML and demonstrate the robustness of the proposed methods over model-based benchmarks. Finally, deep unfolding methods are also investigated for waveform design and beamforming in ISAC systems [KNJ24; Ngu+24], combining the interpretability of traditional iterative algorithms with the flexibility and learning capacity of DL approaches.

Unlike traditional sensing, which relies on backscattered signals from the target, sensing can also be accomplished by exploiting perturbations in the CSI. This methodology for sensing is useful in multi-path rich channels with deficient LoS. AI/ML has been shown to provide superior performance over traditional methods in these scenarios. In [YW23, YG24, YGD25], the authors use CSI perturbations in indoor deployments for target detection and estimation of target parameters, such as position, velocity, and range.

Some other AI/ML approaches for signal processing and optimization in ISAC systems and related applications include deep learning-based MIMO communication for integrated sensing, communication, and computing systems [DZ23], deep learning for ICAS in vehicular networks [Zha+23b], ISAC system for fall detection and recognition using ultrawideband signals [Li+24b], daily activity monitoring using commodity Wi-Fi [Zho+23].

All 6G technologies, including ISAC, align well with the Agenda 2030 for Sustainable Development by the United Nations. ISAC technologies are not only focusing on cost and performance trade-offs, but also on broader values such as sustainability, inclusiveness, and trustworthiness [Wym+25]. AI/ML can efficiently solve the key value indicator (KVI)-induced KPI optimization problem discussed in [Wym+25] to ensure a sustainable approach for joint sensing and communication.

8.2. AI-based positioning and localization

To support the advanced use cases in 6G, positioning and localization technologies must meet several key requirements: centimeter-level to millimeter-level ultra-high positioning accuracy, sub-millisecond latency for real-time tracking, extremely high reliability and robustness, secure

and privacy-preserving localization, energy efficiency and scalability, and context awareness. Outdoor positioning and localization technologies primarily rely on Global Navigation Satellite Systems (GNSS) such as GPS, Galileo, GLONASS, and BeiDou, which provide global coverage using satellite signals. Complementary technologies like cellular network positioning (5G, LTE), Wi-Fi, ultra-wideband (UWB), and AI-based sensor fusion enhance accuracy in urban and indoor environments where GNSS signals may be weak or obstructed. Indoor localization, which is more challenging than outdoor localization, is classified into fingerprinting and geometric categories based on the underlying technology.

ML is increasingly vital for localization, especially in GNSS-denied environments such as indoor and urban areas, where traditional signal processing struggles with noisy, indirect, and complex observations. A hybrid localization approach is proposed in [Gan+23b], combining a physics-based joint angle and delay estimation technique with a deep neural network to accurately predict user location in wideband single-input multiple-output systems, achieving high accuracy in multipath environments. Recently, CSI-based fingerprinting approaches have gained more popularity to overcome the limitations in RSSI-based techniques due to signal fluctuations. To this end, [Wan+17b] proposed DeepFi, a CSI-based fingerprinting framework for indoor localization. The weights of a DNN are trained in the offline stage, and a greedy algorithm is used to optimize the weights and reduce complexity. In the online localization phase, a radial basis function-based probabilistic method is used to estimate the location.

Furthermore, [Zha+22a] proposed a transformer-based indoor positioning system (TIPS) to address the multipath fading effects associated with CSI-based fingerprinting methods in NLoS environments. There, direction of arrival (DoA) information is also used in addition to CSI to learn the relationship between the positions and the CSI and DoA. [MCK25] introduced an advanced transformer architecture designed to improve indoor localization in NLoS-dominated wireless environments. The authors address the limitations of traditional approaches and existing deep learning models, particularly focusing on reducing computational complexity while maintaining high accuracy. A lightweight swish-gated linear unit transformer (L-SwiGLU Transformer) architecture is proposed that reduces computational complexity without sacrificing localization accuracy. It replaces the standard feed-forward network with an L-SwiGLU-based network and incorporates other modifications to enhance efficiency. The sensor snapshot tokenization, a novel approach, introduces a method that maintains variable-specific representations of power delay profiles, enhancing the transformer model's ability to capture multivariate correlations. This method improves attention mechanisms and reduces reliance on large datasets.

Different ML techniques and deep learning models used in fingerprinting techniques include CNNs [ITE18; JH18; WWM20], LSTM networks [DAY22], DQNs [Dou+18], denoising autoencoders [Wan+19; Lee+21], autoencoder and GAN-based unsupervised learning [SS20],

RL techniques [Dou+24], federated learning techniques [Che+22], etc. A complex-valued NN model is proposed to enhance radio frequency positioning accuracy in multi-antenna systems. The model jointly performs line-of-sight (LoS) classification and first-path time-of-arrival (ToA) estimation. This research indicates that complex-valued NNs hold significant promise for improving RF positioning, with potential applications in navigation, industrial automation, and other areas where precise location data is critical [CKB22].

CSI-based fingerprints in MIMO OFDM systems can capture multipath characteristics, which is challenging in Wi-Fi systems due to their limited bandwidth and number of antennas. To this end, a DL-based mMIMO indoor localization is investigated in [Arn+18], where a two-step training procedure is proposed: first, training on LoS data, and second, finetuning on measured NLoS positions. For outdoor positioning, CNNs were utilized in [Vie+17; Sun+19] to map the CSI into the 2D position coordinates, whereas CNN-based indoor positioning is proposed in [Cer+21]. A novel 3D CNN-enabled learning method to localize the 3D positions of the mobile terminals in a MIMO-OFDM system is proposed in [Wu+21]. Furthermore, recent 3GPP studies focused on two key AI/ML approaches: AI/ML direct positioning and AI/ML-assisted positioning. In direct positioning, raw channel measurements are directly fed into an AI/ML model, allowing the system to calculate the device's position without relying on conventional positioning methods. This approach is particularly beneficial in environments where GPS or other traditional systems are inadequate.

In assisted positioning, AI/ML models enhance conventional positioning methods by refining raw channel measurements before passing them to traditional algorithms for final position estimation. 3GPP has specifically explored UE-based Positioning where AI/ML models are implemented on the UE side, using measurements from positioning reference units (PRUs) to generate labels and estimate position. Additionally, NG-RAN node-assisted positioning involves network-side models that help in determining the device's position by integrating AI/ML insights with traditional positioning techniques. These efforts in Release 18 have laid the groundwork for further AI/ML positioning enhancements in Release 19, which will aim to standardize these methods for improved accuracy and performance across diverse 5G networks.

9. Emerging topics

9.1. Semantic communications

Semantic communication is an emerging paradigm designed to optimize communication resource usage in future wireless networks, which leverages AI/ML techniques to enable efficient semantic encoding and decoding, reducing spectrum consumption while preserving the intended meaning of transmitted information. By incorporating shared background knowledge at both the

transmitter and receiver, semantic communications enable accurate content recovery with minimal semantic errors, even in challenging channel conditions. While semantic communication is particularly beneficial for AI-driven learning and prediction tasks that rely on large-scale data, achieving these advantages requires extensive data collection and pre-training of deep learning-based semantic models, posing challenges in acquiring and managing background knowledge. AI-enabled techniques investigated in literature under semantic communications include semantic-aware sensing, coding, modulation, and integration with emerging applications. The survey paper [Qin+24] on AI-empowered wireless communications presented a comprehensive overview of AI-based semantic communications.

A semantic-aware CS technique enables the generation and compression of meaningful content, thereby facilitating the optimization of wireless resources for enhanced efficiency. Semantic-aware image CS architectures were studied in [CZ22; ZQL23]. The authors in [Lok+23] also proposed a semantic communication-based end-to-end image transmission system. A recent study [ZQL24] proposed an RL-based video CS architecture. In addition, a joint semantic sensing and communication framework for XR was proposed in [Zha+22b]. On the other hand, age-of-information-based semantic sampling is more suitable for real-time scenarios involving correlated data sequences generated by sensing devices.

The rapid advancement of natural language processing serves as the foundation for analyzing and understanding the semantics of information, enabling the development of joint semantic-channel coding for text, improving communication efficiency by encoding meaning rather than raw data. To this end, the authors in [FRG18; Xie+21; Zho+22] proposed different LSTM and Transformer model-based techniques for joint semantic-channel coding for text, an approach that aims to extract and encode the semantic information of sentences rather than raw data such as a sequence of bits or a word, thereby improving communication efficiency. Extending deep learning-based joint semantic-channel coding to speech and images is also a widely studied topic [WQ21; Wen+23; Zha+23b; Hua+23b]. Integrating semantic communications with emerging applications and technologies is also of interest in recent literature, especially with the development of advanced deep learning techniques such as LLMs [Xie+25].

These novel AI/ML-based semantic communications technologies improve efficiency, reduce bandwidth usage, and enhance reliability, especially in resource-constrained or noisy environments, which is particularly useful in applications like human-machine interaction, IoT, and multimedia transmission, where understanding the context and intent is more important than exact data replication.

9.2. Digital twins for RAN

Digital twins for RAN (DT-RAN) create real-time replicas of physical networks, facilitating design, diagnosis, simulation, and AI/ML-based control of 6G wireless networks [Lin+23]. These models leverage vast amounts of data, including channel measurements, mobile sensor data, positioning, and spectrum awareness, to improve network optimization, coordination, and resilience. The shift towards DT-RANs is driven by the increasing complexity and demands of modern mobile networks, particularly as we move from 5G to 6G. Traditional network optimization methods were effective in simpler networks but are no longer sufficient in the era of complex, dynamic, and high-performance networks. The growth of mobile devices, the proliferation of IoT, the need for URLLC, and the rise of data-intensive applications such as AR/VR and autonomous vehicles all require more sophisticated network management. Legacy approaches, such as self-organizing networks, struggle with the complexity of managing thousands of parameters in dynamic network environments. The introduction of neural receivers, neural beamforming, and AI-driven models further complicates optimization. DT-RAN serves as a key solution by providing realistic network simulations that enable accurate tuning and optimization, particularly through real-time DTs that capture evolving environmental dynamics.

DT-RAN improves network performance by enabling precise simulation and optimization of network conditions. The integration of DTs with O-RAN principles is expected to enhance intelligence, autonomy, and openness in 6G RAN deployments [Mas+23b]. The O-RAN Alliance next-Generation Research Group (nGRG) recently published a comprehensive research report focused on DT-RAN, highlighting key enablers and use cases that are expected to improve performance, operational efficiency, and utility in future O-RAN deployments [Nvi+24]. The authors in [Wan+25] presented detailed insights into the requirements and possible architecture of a complete digital twin channel (DTC) for 6G. The feasibility of real-time DTC is experimentally validated, and the potential applications and open issues in future DTC research are explored, drawing on typical 6G use cases. In [Rob+24c], the authors present TwiNet, a real-time digital twin framework for wireless networks that uses CNN-based deep learning models to dynamically detect pilot jamming and adaptively retrain models within the DT environment, achieving up to 99.9% training accuracy and significantly improving threat resilience in resource-constrained systems. Tools like Colosseum wireless network emulator with hardware-in-the-loop serve as a DT platform, enabling AI-RAN development through large-scale emulation, data-driven experimentation, and real-time training and validation of AI/ML-based control and optimization policies across diverse RF and protocol environments [Vil+24, Pol+24, Bon+21a]. In [BO24], the authors also develop a methodology to twin commercial LTE traffic traces on the Colosseum O-RAN testbed, using the Toeplitz inverse-covariance-based clustering algorithm and generating a large cross-layer dataset designed to train and evaluate AI/ML-based control policies, including applications in forecasting, classification, and explainable AI. A U-Net-based

deep learning model is introduced in [Sae+24] to enable real-time, high-fidelity propagation modeling for digital twin wireless networks, leveraging elevation maps and rough propagation estimates to achieve sub-0.035 dB error and drastically reduce latency compared to traditional ray tracing methods. Moreover, the BostonTwin framework introduces a city-scale digital twin of Boston integrating high-fidelity 3D building and antenna data into ray-tracing tools, enabling accurate electromagnetic propagation modeling for 6G wireless network planning and immersive communication use cases [Tes+24]. The use of digital twins for sensing applications is demonstrated in [Vil+23], where a high-fidelity digital twin of a radar-cellular coexistence scenario is implemented on Colosseum to train a CNN-based AI agent that detects radar transmissions in real time and mitigates interference by dynamically halting cellular transmissions.

10. Security in AI-for-RAN

As AI and ML become core components of next-generation RAN, they enable advanced automation and optimization—but also introduce new security vulnerabilities that must be carefully addressed. AI/ML models are inherently data-driven and can be susceptible to manipulation through adversarial inputs, data poisoning, model inversion, or backdoor attacks. In addition, the closed-loop nature of AI-RAN decision-making amplifies the risk of cascading failures if malicious data or models go undetected. Protecting AI pipelines from training to inference, ensuring the integrity and confidentiality of input data, and defending against runtime attacks are essential for the safe deployment of AI-enabled RAN systems. Thus, AI-RAN security must go beyond traditional network defenses and embrace robust, explainable, and attack-resilient learning frameworks that are secure by design.

The work in [Gro+24b] provides a comprehensive framework for securing AI-enabled RAN systems by addressing vulnerabilities across open interfaces, virtualized platforms, and data-driven intelligence. It highlights threats such as data poisoning, adversarial attacks, and backdoors targeting AI/ML models, and demonstrates the use of autoencoders to mitigate these threats in deep reinforcement learning-based control loops, reducing erroneous decisions by up to 70%. A real-time AI-based framework is also proposed in [Sca+24] to enhance 5G RAN security by leveraging cross-layer features and deep learning models (including an attention-augmented Xception) to detect radio resource control-layer attacks, such as identity spoofing and denial-of-service (DoS), achieving over 85% accuracy across diverse mobility scenarios.

In parallel, recent research [YSS24] highlights that beyond conventional cyber threats, misconfigurations caused by AI/ML-enabled xApps and rApps themselves pose a critical and underexplored security risk in O-RAN environments. Due to the disaggregated and heterogeneous nature of O-RAN—coupled with AI-driven automation—independent third-party

applications may enact conflicting policies that degrade performance or destabilize the network. This type of configuration drift, especially when propagated autonomously across layers, can lead to cascading failures and service disruptions. The study provides a first in-depth analysis of such misconfiguration scenarios, specifically those arising from interactions among SDN, NFV, and AI/ML within the RIC. A case study illustrates how conflicting xApp actions can directly impact the user experience and demonstrates how AI/ML techniques could be harnessed to proactively detect and resolve misconfigurations before they manifest as faults. This reinforces the need to treat AI/ML misconfiguration not only as an operational challenge but also as a strategic component of AI-RAN security, requiring robust detection, verification, and conflict-resolution mechanisms.

11. Definitions of terms and abbreviations

2D	Two-Dimensional
3D	Three-Dimensional
3GPP	3rd Generation Partnership Project
4G	Fourth-Generation
5G	Fifth-Generation
5G NR	5G New Radio
5G-A	5G-Advanced
6G	Sixth-Generation
A2C	Advantage Actor-Critic
A3C	Asynchronous Advantage Actor-Critic
ACER	Actor-Critic with Experience Replay
ADCNN	Augmented Deep Convolutional Neural Network
AI	Artificial Intelligence
AI-RIC	AI-Based RAN Intelligent Controller
AMC	Adaptive Modulation and Coding
AoA	Angle of Arrival
AP	Access Point
AR	Autoregressive
AR	Augmented Reality
BER	Bit Error Rate
best CQI	Best Channel Quality Indicator
BLER	Block Error Rate
BM	Beam Management
BMD	Bit-Metric Decoding
bMRO	Beam-Based Mobility Robustness Optimization
BS	Base Station
BSAC	Branching Soft-Actor Critic
CHO	Conditional Handover
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CSI	Channel State Information
CsiNet	CSI Sensing and Recovery Neural Network
CSS	Cooperative Spectrum Sensing

DAC	Digital to Analog Converter
DCS	Deep Cooperative Sensing
DDoS	Distributed Denial-of-Service
DDPG	Deep Deterministic Policy Gradient
DDQN	Double Deep Q-Network
DL	Downlink
D-MIMO	Distributed Multiple-Input Multiple-Output
DNN	Deep Neural Network
DoA	Direction of Arrival
DoS	Denial-of-Service
DPD	Digital Pre-Distortion
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
DT	Digital Twin
DTC	Digital Twin Channel
DT-RAN	RAN Digital Twin
DTX	Discontinuous Transmission
DU	Distributed Unit
eMBB	Enhanced Mobile Broadband
FDD	Frequency Division Duplex
FFT	Fast Fourier Transform
GAN	Generative Adversarial Network
GAT	Graph Attention Network
GBSM	Geometry-Based Stochastic Models
GCN	Graph Convolution Network
gNB	gNodeB
GNN	Graph Neural Network
GNSS	Global Navigation Satellite Systems
GPU	Graphics Processing Unit
HAP	High Altitude Platform
HBF	Hybrid Beamforming
h-DQN	Hierarchical Deep Q-Learning
HetNet	Heterogeneous Network
IoT	Internet of Things
I/Q	In-Phase and Quadrature-Phase

ISAC	Integrated Sensing and Communication
KD	Knowledge Distillation
KNN	K-Nearest Neighbors
KPI	Key Performance Indicator
KVI	Key Value Indicator
LiDAR	Light Detection and Ranging
LLM	Large Language Model
LMMSE	Linear Minimum Mean-Squared Error
LoS	Line-of-Sight
LS	Least Squares
LSTM	Long Short-Term Memory
LTE	Long-Term Evolution
LTE-LAA	Long-Term Evolution Licensed Assisted Access
MAC	Medium Access Control Layer
MARL	Multi-Agent Reinforcement Learning
MDP	Markov Decision Process
MEC	Mobile Edge Computing
MIMO	Multiple-Input Multiple-Output
MISO	Multiple-Input Single-Output
ML	Machine Learning
mMIMO	Massive MIMO
MMSE	Minimum Mean Square Error
mMTC	Massive Machine-Type Communications
mmWave	Millimeter Wave
MSC	Modulation and Coding Scheme
multi-TRP	Multi-Transmission and Reception Point
MU-MIMO	Multi-User MIMO
MWC 2025	Mobile World Congress 2025
NBL	Network Beamspace Learning
NFV	Network Function Virtualization
NLoS	Non-Line-of-Sight
NN	Neural Network
NOMA	Non-Orthogonal Multiple Access
NWDAF	Network Data Analytics Function
OFDM	Orthogonal Frequency Division Multiplexing

OFDMA	Orthogonal Frequency Division Multiple Access
O-RAN	Open RAN
OTA	Over-the-Air
PA	Power Amplifier
PCHO	Prediction-Based Conditional Handover
PF	Proportional Fairness
PHY	Physical Layer
PN	Phase Noise
PPO	Proximal Policy Optimization
PRU	Positioning Reference Unit
PUSCH	Physical Uplink Shared Channel
QoS	Quality-of-Service
RAN	Radio Access Network
RF	Radio Frequency
RIC	RAN Intelligent Controller
RIS	Reconfigurable Intelligent Surface
RL	Reinforcement Learning
RNN	Recurrent Neural Network
RR	Round Robin
RRM	Radio Resource Management
RSRP	Reference Signal Receive Power
RSSI	Received Signal Strength Indicator
SAC	Soft Actor-Critic
SBFD	Sub-Band Full Duplex
SBR	Shooting and Bouncing Ray
SCM	Statistical Channel Model
SDN	Software-Defined Networks
SINR	Signal-to-Interference-Plus-Noise Ratio
SLA	Service Level Agreement
SLNR	Signal-to-Leak-Plus-Noise Ratio
SM	Sleep Mode
SNR	Signal-to-Noise Ratio
SOTA	State-of-the-Art
SRS	Sounding Reference Signals
SVM	Support Vector Machine

TAM	Transmit Antenna Muting
THz	Terahertz
TNN	Transformer Neural Network
TOA	Time-of-Arrival
T-PCF	Transformer-Aided Parametric CSI Feedback
TRP	Transmission and Reception Point
UAV	Unmanned Aerial Vehicle
UE	User Equipment
UL	Uplink
URLLC	Ultra-Reliable Low-Latency Communication
UTX	Unintended Transmission
UWB	Ultra-Wideband
V2X	Vehicle-to-Everything
VoLTE	Voice over LTE
VR	Virtual Reality
WMMSE	Weighted Minimum Mean Square Error
XLCNet	XL-MIMO Channel Network
XL-MIMO	Extremely Large-Scale MIMO
XR	Extended Reality

12. References

- [3GPP23] 3rd Generation Partnership Project (3GPP). “Study on network energy savings for NR,” 3GPP TR 38.864 v18.1.0, 2023.
- [3GPP25] 3rd Generation Partnership Project (3GPP). “Chair’s summary of the GPP workshop on 6G”, Accessed on 18/03/2025 [Online.] Available: https://www.3gpp.org/ftp/workshop/2025-03-10_3GPP_6G_WS/Docs/6GWS250243.zip
- [AA19] Muhammad Alrabeiah and Ahmed Alkhateeb. “Deep Learning for TDD and FDD Massive MIMO: Mapping Channels in Space and Frequency.” In: 2019 53rd Asilomar Conference on Signals, Systems, and Computers. 2019, pp. 1465–1470. doi: 10.1109/IEEECONF44664.2019.9048929.
- [AAK17] Ribal Atallah, Chadi Assi, and Maurice Khabbaz. “Deep reinforcement learning-based scheduling for roadside communication networks,” 2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Paris, France, 2017, pp. 1-8, doi: 10.23919/WIOPT.2017.7959912.
- [Ada+22] Abuzar B. M. Adam et al. “Energy-Efficient Power Allocation in Downlink Multi-Cell Multi-Carrier NOMA: Special Deep Neural Network Framework.” In: IEEE Transactions on Cognitive Communications and Networking 8.4 (2022), pp. 1770–1783.
- [Adi+23] Muhammad Adil et al. “CA-MLBS: Content-Aware Machine Learning Based Load Balancing Scheduler in the Cloud Environment.” In: Expert Systems 40.4 (2023), e13150.
- [AH19] Fayccal Ait Aoudia and Jakob Hoydis. “Model-Free Training of End-to-End Communication Systems.” In: IEEE Journal on Selected Areas in Communications 37.11 (2019), pp. 2503–2516. doi: [10.1109/JSAC.2019.2933891](https://doi.org/10.1109/JSAC.2019.2933891).
- [AH22] Fayccal Ait Aoudia and Jakob Hoydis. “Waveform Learning for Next-Generation Wireless Communication Systems.” In: IEEE Transactions on Communications 70.6 (2022), pp. 3804–3817. doi: [10.1109/TCOMM.2022.3164060](https://doi.org/10.1109/TCOMM.2022.3164060).
- [Ahm+23] Ijaz Ahmad et al. “Sustainability in 6G Networks: Vision and Directions.” In: 2023 IEEE Conference on Standards for Communications and Networking (CSCN). 2023, pp. 202– 208. doi: [10.1109/CSCN60443.2023.10453200](https://doi.org/10.1109/CSCN60443.2023.10453200).
- [AL24] Mikael Ade Krisna Respati and Byung Moo Lee. “A Survey on Machine Learning Enhanced Integrated Sensing and Communication Systems: Architectures, Algorithms, and Applications.” In: IEEE Access 12 (2024), pp. 170946–170964. doi: [10.1109/ACCESS.2024.3501363](https://doi.org/10.1109/ACCESS.2024.3501363).

- [Ale+20] George C. Alexandropoulos et al. "Phase Configuration Learning in Wireless Networks with Multiple Reconfigurable Intelligent Surfaces." In: 2020 IEEE Globecom Workshops (GC Wkshps. 2020, pp. 1–6. doi: 10.1109/GCWkshps50303.2020.9367575.
- [Ale+22] George C. Alexandropoulos et al. "Pervasive Machine Learning for Smart Radio Environments Enabled by Reconfigurable Intelligent Surfaces." In: Proceedings of the IEEE 110.9 (2022), pp. 1494–1525. doi: 10.1109/JPROC.2022.3174030.
- [Alk19] Ahmed Alkhateeb. "DeepMIMO: A Generic Deep Learning Dataset for Millimeter Wave and Massive MIMO Applications." 2019. arXiv: 1902.06435 [cs.IT]. url: <https://arxiv.org/abs/1902.06435>.
- [An+23] Qing An et al. "A Deep Reinforcement Learning-Based Resource Scheduler for Massive MIMO Networks." In: IEEE Transactions on Machine Learning in Communications and Networking 1 (2023), pp. 242–257. doi: [10.1109/TMLCN.2023.3313988](https://doi.org/10.1109/TMLCN.2023.3313988).
- [Arn+18] Maximilian Arnold et al. "On Deep Learning-Based Massive MIMO Indoor User Localization." In: 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). 2018, pp. 1–5. doi: [10.1109/SPAWC.2018.8446013](https://doi.org/10.1109/SPAWC.2018.8446013).
- [Arn+19] Maximilian Arnold et al. "Enabling FDD Massive MIMO through Deep Learning-based Channel Prediction." 2019. arXiv: 1901.03664 [cs.IT]. url: <https://arxiv.org/abs/1901.03664>.
- [BA19] Eren Balevi and Jeffrey G. Andrews. "Deep Learning-Based Channel Estimation for High-Dimensional Signals." 2019. arXiv: 1904.09346 [eess.SP]. url: <https://arxiv.org/abs/1904.09346>.
- [BA21] Eren Balevi and Jeffrey G. Andrews. "Wideband Channel Estimation With a Generative Adversarial Network." In: IEEE Transactions on Wireless Communications 20.5 (2021), pp. 3049–3060. doi: 10.1109/TWC.2020.3047100.
- [Bai+24] Yingshuang Bai et al. "AI-Based Beam Management in 3GPP: Optimizing Data Collection Time Window for Temporal Beam Prediction." In: IEEE Open Journal of Vehicular Technology 5 (2024), pp. 48–55. doi: [10.1109/OJVT.2023.3337357](https://doi.org/10.1109/OJVT.2023.3337357).
- [Bai+25] Jungsuk Baik et al. "Transformer-Driven Robust Recovery of Massive MIMO CSI Feedback with Temporal Information." In: ICC 2025 - 2025 IEEE International Conference on Communications (ICC), 2025.
- [Bal+22] Lorenzo Baldesi et al. "ChARM: NextG Spectrum Sharing Through Data-Driven Real-Time O-RAN Dynamic Control." In: Proc. of IEEE International Conference on Computer Communications (INFOCOM). London, UK, May 2022, pp. 91–100.

- [Bal18] A. Balatsoukas-Stimming, "Non-Linear Digital Self-Interference Cancellation for In-Band Full-Duplex Radios Using Neural Networks," 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Kalamata, Greece, 2018, pp. 1-5, doi: 10.1109/SPAWC.2018.8445987.
- [BAM19] Daoud Burghal, Naveed A. Abbasi, and Andreas F. Molisch. "A Machine Learning Solution for Beam Tracking in mmWave Systems." In: 2019 53rd Asilomar Conference on Signals, Systems, and Computers. 2019, pp. 173–177.
doi: [10.1109/IEEECONF44664.2019.9048770](https://doi.org/10.1109/IEEECONF44664.2019.9048770).
- [Ban+23] Serene Banerjee et al. "Causality-Aware Channel State Information Encoding." In: 2023 15th International Conference on COMmunication Systems & NETworkS (COMSNETS). 2023, pp. 617-625. doi: 10.1109/COMSNETS56262.2023.10041353.
- [Bar+22] Ioannis A. Bartsiokas et al. "ML-Based Radio Resource Management in 5G and Beyond Networks: A Survey." In: IEEE Access 10 (2022), pp. 83507–83528.
doi: [10.1109/ACCESS.2022.3196657](https://doi.org/10.1109/ACCESS.2022.3196657).
- [BB05] K.E. Baddour and N.C. Beaulieu. "Autoregressive modeling for fading channel simulation." In: IEEE Transactions on Wireless Communications 4.4 (2005), pp. 1650–1662. doi: [10.1109/TWC.2005.850327](https://doi.org/10.1109/TWC.2005.850327).
- [BB24] Charbel Bou Chaaya and Mehdi Bennis. "RIS Phase Optimization via Generative Flow Networks." In: IEEE Wireless Communications Letters 13.7 (2024), pp. 1988–1992.
doi:10.1109/LWC.2024.3400127.
- [BDA20] Eren Balevi, Akash Doshi, and Jeffrey G. Andrews. "Massive MIMO Channel Estimation With an Untrained Deep Neural Network." In: IEEE Transactions on Wireless Communications 19.3 (2020), pp. 2079–2090. doi: 10.1109/TWC.2019.2962474.
- [Beh+23] A. Behravan et al. "Positioning and Sensing in 6G: Gaps, Challenges, and Opportunities," In: IEEE Vehicular Technology Magazine, vol. 18, no. 1, pp. 40-48, March 2023.
- [BHT+24] C. Benzaïd, F. M. Hossain, T. Taleb, P. M. Gómez, and M. Dieudonne, "A Federated Continual Learning Framework for Sustainable Network Anomaly Detection in O-RAN," 2024 IEEE Wireless Communications and Networking Conference (WCNC), Dubai, United Arab Emirates, 2024, pp. 1-6,
doi: 10.1109/WCNC57260.2024.10570951.
- [BO24] Leonardo Bonati, Ravis Shirkhani, Claudio Fiandrino, Stefano Maxenti, Salvatore D'Oro, Michele Polese, and Tommaso Melodia. "Twinning Commercial Network Traces on Experimental Open RAN Platforms." In: Proceedings of the 30th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '24). 2024, pp. 1914–1921.

- [Bon+21a] Leonardo Bonati et al. "Colosseum: Large-Scale Wireless Experimentation Through Hardware-in-the-Loop Network Emulation." In: Proc. of IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN). Virtual Conference, Dec. 2021, pp. 1–6. doi: 10.1109/DySPAN52821.2021.9616557.
- [Bon+21b] Leonardo Bonati et al. "Intelligence and Learning in O-RAN for Data-Driven NextG Cellular Networks." In: IEEE Communications Magazine, vol. 59, no. 10, Oct. 2021, pp. 98–104. doi: 10.1109/MCOM.101.2001120.
- [Bos+21] Christopher Bosso et al. "Ultrabroadband Spread Spectrum Techniques for Secure Dynamic Spectrum Sharing Above 100 GHz Between Active and Passive Users." In: 2021 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN). 2021, pp. 45–52. doi: [10.1109/DySPAN53946.2021.9677346](https://doi.org/10.1109/DySPAN53946.2021.9677346).
- [Bou+23] R. Bourgerie and T. Zanoluda. "Fault Detection in Telecom Networks Using Bi-Level Federated Graph Neural Networks." In: 2023 IEEE International Conference on Data Mining Workshops (ICDMW), Shanghai, China, 2023, pp. 1608–1617. doi: 10.1109/ICDMW60847.2023.10449399.
- [Cam+20] Sebastian Cammerer et al. "Trainable Communication Systems: Concepts and Prototype." In: IEEE Transactions on Communications 68.9 (2020), pp. 5489–5503. doi: [10.1109/TCOMM.2020.3002915](https://doi.org/10.1109/TCOMM.2020.3002915).
- [Cam+23] Juan Sebastian Camargo et al. "Dynamic Slicing Reconfiguration for Virtualized 5G Networks Using ML Forecasting of Computing Capacity." In: Computer Networks 236 (2023), p. 110001.
- [Cao+21] Zheng Cao et al. "Lightweight Convolutional Neural Networks for CSI Feedback in Massive MIMO." In: IEEE Communications Letters 25.8 (2021), pp. 2624–2628. doi: 10.1109/LCOMM.2021.3076504.
- [Cao+24] Xuelin Cao et al. "AI-Empowered Multiple Access for 6G: A Survey of Spectrum Sensing, Protocol Designs, and Optimizations." In: Proceedings of the IEEE 112.9 (2024), pp. 1264–1302. doi: [10.1109/JPROC.2024.3417332](https://doi.org/10.1109/JPROC.2024.3417332).
- [Cao+25] [Chaoqun Cao](https://orcid.org/0009-0001-9210-4460) et al. "Deep Learning-Enhanced Channel Prediction for XL-MIMO Systems," in IEEE Communications Letters, vol. 29, no. 6, pp. 1260–1264, June 2025, doi: 10.1109/LCOMM.2025.3558838.
- [Cas+21] Luca Casasole et al. "QCell: Self-optimization of Softwarized 5G Networks through Deep Q-learning." In: Proc. of IEEE Conference on Network Softwarization (NetSoft). Tokyo, Japan, June 2021, pp. 1–6.
- [CBS19] Sucharita Chakraborty, Emil Björnson, and Luca Sanguinetti. "Centralized and Distributed Power Allocation for Max-Min Fairness in Cell-Free Massive MIMO." In: 2019 53rd Asilomar Conference on Signals, Systems, and Computers. 2019, pp. 576–580. doi: <https://doi.org/10.1109/IEEECONF44664.2019.9048903>

- [CDN19] Qiuyu Cai, Chao Dong, and Kai Niu. "Attention Model for Massive MIMO CSI Compression Feedback and Recovery." In: 2019 IEEE Wireless Communications and Networking Conference (WCNC). 2019, pp. 1–5. doi: 10.1109/WCNC.2019.8885897.
- [Cer+21] Gregor Cerar et al. "Improving CSI-based Massive MIMO Indoor Positioning using Convolutional Neural Network." In: 2021 Joint European Conference on Networks and Communications 6G Summit (EuCNC/6G Summit). 2021, pp. 276–281. doi: [10.1109/EuCNC/ 6GSummit51104.2021.9482604](https://doi.org/10.1109/EuCNC/6GSummit51104.2021.9482604).
- [CH20] Yuhang Chen and Chong Han. "Deep CNN-Based Spherical-Wave Channel Estimation for Terahertz Ultra-Massive MIMO Systems." In: GLOBECOM 2020 - 2020 IEEE Global Communications Conference. 2020, pp. 1–6. doi: 10.1109/GLOBECOM42002.2020.9322174.
- [CH24] Hai Cheng, Pedram Johari, Mohamed Amine Arfaoui, Francois Periard, Philip Pietraski, Guodong Zhang, and Tommaso Melodia. "Real-Time AI-Enabled CSI Feedback Experimentation with Open RAN." In: Proc. of IEEE/IFIP Wireless On-demand Network Systems and Services Conference (WONS). Chamonix, France, Jan. 2024, pp. 1–6.
- [Cha+23] Ilias Chatzistefanidis et al. "ML-based Traffic Steering for Heterogeneous Ultra-Dense Beyond-5G Networks." In: 2023 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2023.
- [CHC21] Ronald Y. Chang, Sung-Fu Han, and Feng-Tsun Chien. "Reinforcement Learning-Based Joint Cooperation Clustering and Content Caching in Cell-Free Massive MIMO Networks." In: 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall). 2021, pp. 1–7. doi:[10.1109/VTC2021-Fall52928.2021.9625449](https://doi.org/10.1109/VTC2021-Fall52928.2021.9625449).
- [Che+21a] Hongchao Chen et al. "Joint User Scheduling and Transmit Precoder Selection Based on DDPG for Uplink Multi-User MIMO Systems." In: 2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall). 2021, pp. 1–5. doi: <https://doi.org/10.1109/VTC2021-Fall52928.2021.9625046>
- [Che+21b] [Wuhui Chen](#) et al. "Deep Reinforcement Learning for Internet of Things: A Comprehensive Survey," in IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1659-1692, third quarter 2021, doi: 10.1109/COMST.2021.3073036.
- [Che+22] Shuaifei Chen et al. "A survey on user-centric cell-free massive MIMO systems." In: Digital Communications and Networks 8.5 (2022), pp. 695–719. issn: 2352-8648. doi: [//doi.org/10.1016/j.dcan.2021.12.005](https://doi.org/10.1016/j.dcan.2021.12.005).
- [Che+24] Zijian Chen et al. "Joint Target Sensing and Channel Estimation for IRS-Aided mmWave ISAC Systems." In: 2024 IEEE Wireless Communications and Networking Conference (WCNC). 2024, pp. 1–6. doi: [10.1109/WCNC57260.2024.10570612](https://doi.org/10.1109/WCNC57260.2024.10570612).

- [Cho+24] Young-Jun Cho, Hyeon-Min Yoo, Kyung-Sook Kim, Jeehyeon Na, and Een-Kee Hong. "Practical Load Balancing Algorithm for 5G Small Cell Networks Based on Real-World 5G Traffic and O-RAN Architecture." In: IEEE Access (2024).
- [CKB24] J. -F. T. Cheng, A. R. Khamesi, and Y. Blankenship, "Complex-Valued Neural Networks for Distributed Machine Learning Assisted RF Positioning," In: IEEE MILCOM, Washington, DC, USA, 2024, pp. 451-456, doi: [10.1109/MILCOM61039.2024.10773773](https://doi.org/10.1109/MILCOM61039.2024.10773773).
- [CS21] Ursula Challita and David Sandberg. "Deep Reinforcement Learning for Dynamic Spectrum Sharing of LTE and NR." In: ICC 2021 - IEEE International Conference on Communications. 2021, pp. 1–6. doi: [10.1109/ICC42927.2021.9500325](https://doi.org/10.1109/ICC42927.2021.9500325).
- [CZ22] Bin Chen and Jian Zhang. "Content-Aware Scalable Deep Compressed Sensing." In: IEEE Transactions on Image Processing 31 (2022), pp. 5412–5426. doi: [10.1109/TIP.2022.3195319](https://doi.org/10.1109/TIP.2022.3195319).
- [DA23] Umut Demirhan and Ahmed Alkhateeb. "Integrated Sensing and Communication for 6G: Ten Key Machine Learning Roles." In: IEEE Communications Magazine 61.5 (2023), pp. 113–119. doi: [10.1109/MCOM.006.2200480](https://doi.org/10.1109/MCOM.006.2200480).
- [DAn+19] C. D'Andrea et al. "Uplink Power Control in Cell-Free Massive MIMO via Deep Learning." In: 2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). 2019, pp. 554–558.
- [Das+24] Sree Krishna Das et al. "A Survey on Federated Learning for Reconfigurable Intelligent Metasurfaces-Aided Wireless Networks." In: IEEE Open Journal of the Communications Society 5 (2024), pp. 1846–1879. doi: [10.1109/OJCOMS.2024.3378266](https://doi.org/10.1109/OJCOMS.2024.3378266).
- [DAY22] Yinhuan Dong, Tughrul Arslan, and Yunjie Yang. "An Encoded LSTM Network Model for Wi-Fi-based Indoor Positioning." In: 2022 IEEE 12th International Conference on Indoor Positioning and Indoor Navigation (IPIN). 2022, pp. 1–6. doi: [10.1109/IPIN54987.2022.9918116](https://doi.org/10.1109/IPIN54987.2022.9918116).
- [DBS21] Özlem Tufe Demir, Emil Björnson, and Luca Sanguinetti. "Foundations of User-Centric Cell-Free Massive MIMO." In: Foundations and Trends in Signal Processing 14.3-4 (2021), pp. 162–472. issn: 1932-8346. doi: [10.1561/2000000109](https://doi.org/10.1561/2000000109). url: <http://dx.doi.org/10.1561/2000000109>.
- [DH14] Tianben Ding and Akira Hirose. "Fading Channel Prediction Based on Combination of Complex-Valued Neural Networks and Chirp Z-Transform." In: IEEE Transactions on Neural Networks and Learning Systems 25.9 (2014), pp. 1686–1695. doi: [10.1109/TNNLS.2014.2306420](https://doi.org/10.1109/TNNLS.2014.2306420).

- [DH24] Ryan M. Dreifuerst and Robert W. Heath, "Hierarchical ML Codebook Design for Extreme MIMO Beam Management," in *IEEE Transactions on Machine Learning in Communications and Networking*, vol. 2, pp. 688-702, 2024, doi: 10.1109/TMLCN.2024.3402178.
- [DH25] Ryan M. Dreifuerst and Robert W. Heath, "Neural Codebook Design for MIMO Network Beam Management," in *IEEE Transactions on Wireless Communications*, vol. 24, no. 5, pp. 3909-3922, May 2025, doi: 10.1109/TWC.2025.3536290.
- [DO24] Salvatore D'Oro, Leonardo Bonati, Michele Polese, and Tommaso Melodia. "OrchestRAN: Orchestrating Network Intelligence in the Open RAN." In: *IEEE Transactions on Mobile Computing*. In press, 2024, pp. 1–13.
- [Don+19] Peihao Dong et al. "Deep CNN-Based Channel Estimation for mmWave Massive MIMO Systems." In: *IEEE Journal of Selected Topics in Signal Processing* 13.5 (2019), pp. 989–1000. doi: 10.1109/JSTSP.2019.2925975.
- [Dou+18] Fei Dou et al. "Top-Down Indoor Localization with Wi-Fi Fingerprints Using Deep Q-Network." In: 2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS). 2018, pp. 166–174. doi: [10.1109/MASS.2018.00037](https://doi.org/10.1109/MASS.2018.00037).
- [Dou+24] Fei Dou et al. "On-Device Indoor Positioning: A Federated Reinforcement Learning Approach with Heterogeneous Devices." In: *IEEE Internet of Things Journal* 11.3 (2024), pp. 3909–3926. doi: [10.1109/IJOT.2023.3299262](https://doi.org/10.1109/IJOT.2023.3299262).
- [DSJ19] Yu Ding, Li Su, and Depeng Jin. "New fast multi-user beam training scheme based on compressed sensing theory for millimetre-wave communication." In: *IET Communications* 13.6 (2019), pp. 642–648. doi: <https://doi.org/10.1049/iet-com.2018.5483>.
- [DZ23] Chaowei Duan and Jian Zhang. "A modified deep learning-based MIMO communication for integrated sensing, communication and computing systems." In: *Digital Signal Processing* 142 (2023), p. 104196. issn: 1051-2004. doi: <https://doi.org/10.1016/j.dsp.2023.104196>
- [EC22] Ahmet M. Elbir and Sinem Coleri. "Federated Learning for Channel Estimation in Conventional and RIS-Assisted Massive MIMO." In: *IEEE Transactions on Wireless Communications* 21.6 (2022), pp. 4255–4268. doi: 10.1109/TWC.2021.3128392.
- [Elb+20] Ahmet M. Elbir et al. "Deep Channel Learning for Large Intelligent Surfaces Aided mm-Wave Massive MIMO Systems." In: *IEEE Wireless Communications Letters* 9.9 (2020), pp. 1447–1451. doi: 10.1109/LWC.2020.2993699.
- [ElB+23] M. A. ElBahaay, T. Arafa, and N. AbdelBaki. "Unsupervised Anomaly Detection for Throughput Enhancement in Mobile Networks." In: 2023 International Conference on Computer and Applications (ICCA), Cairo, Egypt, 2023, pp. 1–6. doi: 10.1109/ICCA59364.2023.10401508

- [Elb+23a] Ahmet M. Elbir et al. "Federated Multi-Task Learning For THZ Wideband Channel And DOA Estimation." In: 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW). 2023, pp. 1–5.
doi: 10.1109/ICASSPW59220.2023.10193511.
- [Elb+23b] Ahmet M. Elbir et al. "Terahertz-Band Channel and Beam Split Estimation via Array Perturbation Model." In: IEEE Open Journal of the Communications Society 4 (2023), pp. 892–907. doi: 10.1109/OJCOMS.2023.3263625.
- [Elb19] Ahmet M. Elbir. "CNN-Based Precoder and Combiner Design in mmWave MIMO Systems." In: IEEE Communications Letters 23.7 (2019), pp. 1240–1243. doi: [10.1109/LCOMM.2019.2915977](https://doi.org/10.1109/LCOMM.2019.2915977).
- [Elk+25] Maxime Elkael et al. "ALLSTaR: Automated LLM-Driven Scheduler Generation and Testing for Intent-Based RAN," arXiv preprint arXiv:2505.18389, 2025.
- [EM20] Ahmet M. Elbir and Kumar Vijay Mishra. "Joint Antenna Selection and Hybrid Beamformer Design Using Unquantized and Quantized Deep Learning Networks." In: IEEE Transactions on Wireless Communications 19.3 (2020), pp. 1677–1688.
doi: [10.1109/TWC.2019.2956146](https://doi.org/10.1109/TWC.2019.2956146).
- [EMC21] Ahmet M. Elbir, Kumar Vijay Mishra, and Symeon Chatzinotas. "Terahertz-Band Joint Ultra-Massive MIMO Radar-Communications: Model-Based and Model-Free Hybrid Beamforming," in IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 6, pp. 1468-1483, Nov. 2021, doi: 10.1109/JSTSP.2021.3117410.
- [Eme+25] Jérôme Emery et al. "A Foundation Model for Massive MIMO Precoding with an Adaptive per-User Rate-Power Tradeoff." 2025. arXiv: 2507.18587 [eess.SP].
<https://arxiv.org/abs/2507.18587>
- [Fal+18] Robert Falkenberg et al, "Machine Learning Based Uplink Transmission Power Prediction for LTE and Upcoming 5G Networks Using Passive Downlink Indicators," 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), Chicago, IL, USA, 2018, pp. 1-7, doi: 10.1109/VTCFall.2018.8690629.
- [Fan+23] Xinran Fang et al. "Joint Communication and Sensing Toward 6G: Models and Potential of Using MIMO." In: IEEE Internet of Things Journal 10.5 (2023), pp. 4093–4116. doi:[10.1109/JIOT.2022.3227215](https://doi.org/10.1109/JIOT.2022.3227215).
- [FC22] K. M. Faisal and Wooyeol Choi. "Machine Learning Approaches for Reconfigurable Intelligent Surfaces: A Survey." In: IEEE Access 10 (2022), pp. 27343–27367.
doi: 10.1109/ACCESS.2022.3157651.
- [FGS25] Hamed Farhadi, Prayag Gowgi, and David Sandberg, "AI/ML-based beamforming with imperfect channel state information for multi-user MIMO transmission." In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2025.

- [FHS23] Hamed Farhadi, Johan Haraldson, and Mrarten Sundberg. "A Deep Learning Receiver for Non-Linear Transmitter." In: IEEE Access 11 (2023), pp. 2796–2803. doi: 10.1109/ACCESS.2023.3234501.
- [FRG18] Nariman Farsad, Milind Rao, and Andrea Goldsmith. "Deep Learning for Joint SourceChannel Coding of Text." In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2018, pp. 2326–2330. doi: [10.1109/ICASSP.2018.8461983](https://doi.org/10.1109/ICASSP.2018.8461983).
- [FYJ20] Christofer Flinta, Wenqing Yan, and Andreas Johnsson. "Predicting round-trip time distributions in IoT systems using histogram estimators." In: IEEE/IFIP Network Operations and Management Symposium (NOMS). 2020, pp. 1-9. doi: 10.1109/NOMS47738.2020.9110315.
- [Gan+23a] Milad Ganjalizadeh et al. "Saving energy and spectrum in enabling URLLC services: A scalable RL solution." In: IEEE Transactions on Industrial Informatics 19.10 (2023), pp. 10265–10276.
- [Gan+23b] Rajeev Gangula et al. "Localization in Wideband SIMO Systems with Spatial-Wideband and Pulse-Shaping Effects." In: Proc. of IEEE Global Communications Conference (GLOBECOM). Kuala Lumpur, Malaysia, Dec. 2023, pp. 1–6.
- [Gan+24] R. Gangula et al., "Listen-While-Talking: Toward dApp-based Real-Time Spectrum Sharing in O-RAN," MILCOM 2024 - 2024 IEEE Military Communications Conference (MILCOM), Washington, DC, USA, 2024, pp. 651-652, doi: 10.1109/MILCOM61039.2024.10773685.
- [Gan+25] Milad Ganjalizadeh et al. BSAC-CoEx: Coexistence of URLLC and Distributed Learning Services via Device Selection. 2025. arXiv: 2212.11805 [cs.NI]. url: <https://arxiv.org/abs/2212.11805>.
- [Gao+20] Jiabao Gao et al. "Unsupervised Learning for Passive Beamforming." In: IEEE Communications Letters 24.5 (2020), pp. 1052–1056. doi: 10.1109/LCOMM.2020.2965532.
- [Gao+24] Shen Gao et al. "Lightweight Deep Learning Based Channel Estimation for Extremely Large-Scale Massive MIMO Systems." In: IEEE Transactions on Vehicular Technology 73.7 (2024), pp. 10750–10754. doi: 10.1109/TVT.2024.3364510.
- [Gem+25] Gabriele Gemmi et al. "An Open RAN Framework for the Dynamic Control of 5G Service Level Agreements." In: arXiv preprint arXiv:2309.07508. 2023, pp. 1–10.
- [GFT16] Apostolos Galanopoulos, Fotis Foukalas, and Theodoros A. Tsiftsis. "Efficient Coexistence of LTE With Wi-Fi in the Licensed and Unlicensed Spectrum Aggregation." In: IEEE Transactions on Cognitive Communications and Networking 2.2 (2016), pp. 129–140. doi: [10.1109/TCCN.2016.2594780](https://doi.org/10.1109/TCCN.2016.2594780).

- [Ghi+23] Niyousha Ghiasi et al. "Energy Efficient AP Selection for Cell-Free Massive MIMO Systems: Deep Reinforcement Learning Approach." In: IEEE Transactions on Green Communications and Networking 7.1 (2023), pp. 29–41.
doi: [10.1109/TGCN.2022.3196013](https://doi.org/10.1109/TGCN.2022.3196013).
- [Gio+24] A. Giovanidis et al. "Online Frequency Scheduling by Learning Parallel Actions," In: 3rd International Conference on 6G Networking (6GNet), Paris, France, 2024, pp. 153-160, doi: 10.1109/6GNet63182.2024.10765674.
- [Gko23] Panagiotis K. Gkonis. "A Survey on Machine Learning Techniques for Massive MIMO Configurations: Application Areas, Performance Limitations and Future Challenges." In: IEEE Access 11 (2023), pp. 67–88. doi: [10.1109/ACCESS.2022.3232855](https://doi.org/10.1109/ACCESS.2022.3232855).
- [Gou+21] Mathieu Goutay et al. "End-to-End Learning of OFDM Waveforms with PAPR and ACLR Constraints." In: 2021 IEEE Globecom Workshops (GC Wkshps). 2021, pp. 1–6. doi: [10.1109/GCWkshps52748.2021.9682132](https://doi.org/10.1109/GCWkshps52748.2021.9682132).
- [Gou+23] Sotirios K. Goudos et al. "Joint QoS Aware Admission Control and Power Allocation in NOMA Downlink Networks." In: IEEE Access 11 (2023), pp. 30873–30887. doi: [10.1109/ACCESS.2023.3262117](https://doi.org/10.1109/ACCESS.2023.3262117).
- [Gro+24a] Joshua Groen et al. "TIMESAFE: Timing Interruption Monitoring and Security Assessment for Fronthaul Environments." In: arXiv preprint arXiv:2412.13049. 2024, pp. 1–13.
- [Gro+24b] Joshua Groen et al. "Implementing and Evaluating Security in O-RAN: Interfaces, Intelligence, and Platforms." In: IEEE Network. 2024, pp. 1–13.
- [Gru+17] Tobias Gruber et al. "On deep learning-based channel decoding." In: 2017 51st Annual Conference on Information Sciences and Systems (CISS). 2017, pp. 1–6. doi: [10.1109/CISS.2017.7926071](https://doi.org/10.1109/CISS.2017.7926071).
- [Gud+24] Venkateswarlu Gudepu et al. "GAN-Based Drift and Anomaly Detection for Open Radio Access Networks." In: 2024 IEEE 25th International Conference on High Performance Switching and Routing (HPSR). IEEE, 2024.
- [Guo+19] Hanqing Guo et al. "DSIC: Deep Learning Based Self-Interference Cancellation for In-Band Full Duplex Wireless," 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9013521.
- [Guo+20] Xiaojun Guo et al. "A Novel User Selection Massive MIMO Scheduling Algorithm via Real Time DDPG." In: GLOBECOM 2020 - 2020 IEEE Global Communications Conference. 2020, pp. 1–6. doi: [10.1109/GLOBECOM42002.2020.9322383](https://doi.org/10.1109/GLOBECOM42002.2020.9322383).
- [Guo+22] Jiajia Guo et al. "Overview of Deep Learning-Based CSI Feedback in Massive MIMO Systems." In: IEEE Transactions on Communications 70.12 (2022), pp. 8017–8045. doi:10.1109/TCOMM.2022.3217777.

- [Gup+24] Ankit Gupta et al. SpikingRx: From Neural to Spiking Receiver. 2024. arXiv: 2409.05610 [cs.IT]. url: <https://arxiv.org/abs/2409.05610>.
- [Hab+24] Md Arafat Habib et al. "Machine Learning-Enabled Traffic Steering in O-RAN: A Case Study on Hierarchical Learning Approach." In: IEEE Communications Magazine (2024).
- [Han+20a] Yu Han et al. "Deep Learning-Based FDD Non-Stationary Massive MIMO Downlink Channel Reconstruction." In: IEEE Journal on Selected Areas in Communications 38.9 (2020), pp. 1980–1993. doi: 10.1109/JSAC.2020.3000836.
- [Han+20b] Mengqi Han et al. "Reinforcement Learning for Efficient and Fair Coexistence Between LTE-LAA and Wi-Fi." In: IEEE Transactions on Vehicular Technology 69.8 (2020), pp. 8764–8776. doi: [10.1109/TVT.2020.2994525](https://doi.org/10.1109/TVT.2020.2994525).
- [Has+22] Ramin Hashemi et al. "Deep Reinforcement Learning for Practical Phase Shift Optimization in RIS-assisted Networks over Short Packet Communications." In: 2022 Joint European Conference on Networks and Communications 6G Summit (EuCNC/6G Summit). 2022, pp. 518–523. doi: 10.1109/EuCNC/6GSummit54941.2022.9815804.
- [Has+24] Antor Hasan et al. "Root Cause Analysis of Anomalies in 5G RAN Using Graph Neural Network and Transformer." arXiv preprint arXiv:2406.15638 (2024).
- [HBD11] Jakob Hoydis, Stephan ten Brink, and Mérouane Debbah. "Massive MIMO: How Many Antennas Do We Need?" In: 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton). 2011, pp. 545–550. doi: [10.1109/Allerton.2011.6120214](https://doi.org/10.1109/Allerton.2011.6120214).
- [He+18] Hengtao He et al. "Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems." In: IEEE Wireless Communications Letters 7.5 (2018), pp. 852–855. doi: 10.1109/LWC.2018.2832128.
- [HGS16] Hado van Hasselt, Arthur Guez, and David Silver. "Deep Reinforcement Learning with Double Q-Learning." In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. AAAI'16. Phoenix, Arizona: AAAI Press, 2016, pp. 2094–2100.
- [HK21] M. Hoffmann and P. Kryszkiewicz, "Reinforcement Learning for Energy-Efficient 5G Massive MIMO: Intelligent Antenna Switching," in IEEE Access, vol. 9, pp. 130329–130339, 2021, doi: 10.1109/ACCESS.2021.3113461
- [HK23a] M. Hoffmann and P. Kryszkiewicz, "Beam Management Driven by Radio Environment Maps in O-RAN Architecture," 2023 IEEE International Conference on Communications Workshops (ICC Workshops), Rome, Italy, 2023, pp. 54–59, doi: 10.1109/ICCWorkshops57953.2023.10283527.
- [HK23b] M. Hoffmann and P. Kryszkiewicz, "Signaling Storm Detection in IIoT Network based on the Open RAN Architecture," IEEE INFOCOM 2023 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Hoboken, NJ, USA, 2023, pp. 1–2, doi: 10.1109/INFOCOMWKSHPS57453.2023.10226043.

- [HK24] M. Hoffmann and M. Dryjański, "Energy Efficiency in Open RAN: RF Channel Reconfiguration Use Case," in IEEE Access, vol. 12, pp. 118493–118501, 2024, doi: [10.1109/ACCESS.2024.3449700](https://doi.org/10.1109/ACCESS.2024.3449700).
- [HKH21] Mikko Honkala, Dani Korpi, and Janne M. J. Huttunen. "DeepRx: Fully Convolutional Deep Learning Receiver." In: IEEE Transactions on Wireless Communications 20.6 (2021), pp. 3925–3940. doi: [10.1109/TWC.2021.3054520](https://doi.org/10.1109/TWC.2021.3054520).
- [HKH23] Janne M. J. Huttunen, Dani Korpi, and Mikko Honkala. "DeepTx: Deep Learning Beamforming With Channel Prediction." In: IEEE Transactions on Wireless Communications 22.3 (2023), pp. 1855–1867. doi: [10.1109/TWC.2022.3207055](https://doi.org/10.1109/TWC.2022.3207055).
- [Hoj+21] Hamed Hojatian et al. "Unsupervised Deep Learning for Massive MIMO Hybrid Beamforming." In: IEEE Transactions on Wireless Communications 20.11 (2021), pp. 7086– 7099. doi: [10.1109/TWC.2021.3080672](https://doi.org/10.1109/TWC.2021.3080672).
- [Hoy+21] Jakob Hoydis et al. "Toward a 6G AI-Native Air Interface." In: IEEE Communications Magazine 59.5 (2021), pp. 76–81. doi: [10.1109/MCOM.001.2001187](https://doi.org/10.1109/MCOM.001.2001187).
- [Hu+21] Qiyu Hu et al. "Iterative Algorithm Induced Deep-Unfolding Neural Networks: Precoding Design for Multiuser MIMO Systems." In: IEEE Transactions on Wireless Communications 20.2 (2021), pp. 1394–1410. doi: [10.1109/TWC.2020.3033334](https://doi.org/10.1109/TWC.2020.3033334).
- [Hua+19] Chongwen Huang et al. "Indoor Signal Focusing with Deep Learning Designed Reconfigurable Intelligent Surfaces." In: 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). 2019, pp. 1–5. doi: [10.1109/SPAWC.2019.8815412](https://doi.org/10.1109/SPAWC.2019.8815412).
- [Hua+20] Hongji Huang et al., "Deep Learning for Physical-Layer 5G Wireless Techniques: Opportunities, Challenges and Solutions." in IEEE Wireless Communications, vol. 27, no. 1, pp. 214–222, February 2020, doi: [10.1109/MWC.2019.1900027](https://doi.org/10.1109/MWC.2019.1900027).
- [Hua+23a] Chih-Wei Huang et al. "A DRL-Based Automated Algorithm Selection Framework for Cross-Layer QoS-Aware Scheduling and Antenna Allocation in Massive MIMO Systems." In: IEEE Access 11 (2023), pp. 13243–13256. doi: [10.1109/ACCESS.2023.3243068](https://doi.org/10.1109/ACCESS.2023.3243068).
- [Hua+23b] Danlan Huang et al. "Toward Semantic Communications: Deep Learning-Based Image Semantic Coding." In: IEEE Journal on Selected Areas in Communications 41.1 (2023), pp. 55–71. doi: [10.1109/JSAC.2022.3221999](https://doi.org/10.1109/JSAC.2022.3221999).
- [Hus+20] Fatima Hussain et al. "Machine Learning for Resource Management in Cellular and IoT Networks: Potentials, Current Solutions, and Open Challenges." In: IEEE Communications Surveys Tutorials 22.2 (2020), pp. 1251–1275. doi: [10.1109/COMST.2020.2964534](https://doi.org/10.1109/COMST.2020.2964534).

- [lbr+18] Mohamed S. Ibrahim et al. "Learning-Based Antenna Selection for Multicasting." In: 2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). 2018, pp. 1–5. doi: [10.1109/SPAWC.2018.8446042](https://doi.org/10.1109/SPAWC.2018.8446042).
- [Int20] Giovanni Interdonato. "Cell-Free Massive MIMO: Scalability, Signal Processing and Power Control." PhD thesis. Linköping University Electronic Press, 2020.
- [Isa+23] Martin Isaksson et al. "mmWave Beam Selection in Analog Beamforming using Personalized Federated Learning", In IEEE Future Networks World Forum (FNWF), Nov. 2023.
- [ITE18] Mai Ibrahim, Marwan Torki, and Mustafa ElNainay. "CNN-based Indoor Localization using RSS Time-Series." In: 2018 IEEE Symposium on Computers and Communications (ISCC). 2018, pp. 01044–01049. doi: [10.1109/ISCC.2018.8538530](https://doi.org/10.1109/ISCC.2018.8538530).
- [Jan+20] Gunhee Jang et al. "Base station switching and sleep mode optimization with LSTM-based user prediction." In: IEEE Access 8 (2020), pp. 222711–222723.
- [Jan+24] Min Jang et al. "Learning of Discrepancy to Validate Decoding Results with Error-Correcting Codes," In: IEEE International Conference on Communications (ICC), 2024, pp. 2931–2936, doi: 10.1109/ICC51166.2024.10622266
- [Jan+25] Seowoo Jang et al. "AI-powered Network-level and User-level Optimization: Challenges, Real-World Results, and the Path Forward" In: Samsung Research Blog [Online]. Available: <https://research.samsung.com/blog/AI-powered-Network-level-and-User-level-Optimization-Challenges-Real-World-Results-and-the-Path-Forward>
- [Jay+24] Nalin Jayweera et al. "5G-Advanced AI/ML Beam Management: Performance Evaluation with Integrated ML Models." 2024. ArXiv: 2404.15326 [eess.SP]. <https://arxiv.org/abs/2404.15326>
- [JDB25] Bhushan Joshi, Ayodele Damola, and Shruti Bothe "AI-driven sustainability in telecom: Ericsson's vision for a greener 6G future." In: Ericsson Blog [Online]. Available: <https://www.ericsson.com/en/blog/2025/3/ai-driven-sustainability>.
- [JH18] Jin-Woo Jang and Song-Nam Hong. "Indoor Localization with Wi-Fi Fingerprinting Using Convolutional Neural Network." In: 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN). 2018, pp. 753–758. doi: [10.1109/ICUFN.2018.8436598](https://doi.org/10.1109/ICUFN.2018.8436598).
- [Jia+19] Yihan Jiang et al. "Turbo Autoencoder: Deep Learning Based Channel Codes for Point-to-Point Communication Channels." In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2019.
- [Jin+21] Yu Jin et al. "Channel Estimation for Semi-Passive Reconfigurable Intelligent Surfaces With Enhanced Deep Residual Networks." In: IEEE Transactions on Vehicular Technology 70.10 (2021), pp. 11083–11088. doi: 10.1109/TVT.2021.3109937.

- [Jin+23] Weijie Jin et al. "Model-Driven Deep Learning for Hybrid Precoding in Millimeter Wave MU-MIMO System." In: IEEE Transactions on Communications 71.10 (2023), pp. 5862–5876. doi: [10.1109/TCOMM.2023.3296622](https://doi.org/10.1109/TCOMM.2023.3296622).
- [Joh+22] S. S. Johari et al. "Anomaly Detection and Localization in NFV Systems: An Unsupervised Learning Approach." In: NOMS 2022 - 2022 IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, 2022, pp. 1–9. doi: 10.1109/NOMS54207.2022.9789938.
- [Jou16] Jingon Joung. "Machine Learning-Based Antenna Selection in Wireless Communications." In: IEEE Communications Letters 20.11 (2016), pp. 2241–2244. doi: [10.1109/LCOMM.2016.2594776](https://doi.org/10.1109/LCOMM.2016.2594776).
- [JS18] Wei Jiang and Hans Dieter Schotten. "Multi-Antenna Fading Channel Prediction Empowered by Artificial Intelligence." In: 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall). 2018, pp. 1–6. doi: [10.1109/VTCFall.2018.8690550](https://doi.org/10.1109/VTCFall.2018.8690550).
- [JS19] Wei Jiang and Hans Dieter Schotten. "A Comparison of Wireless Channel Predictors: Artificial Intelligence Versus Kalman Filter." In: ICC 2019 - 2019 IEEE International Conference on Communications (ICC). 2019, pp. 1–6. doi: [10.1109/ICC.2019.8761308](https://doi.org/10.1109/ICC.2019.8761308).
- [Ju+23] H. Ju et al. "Transformer-Aided Parametric CSI Feedback for mmWave Massive MIMO Systems," ICC 2023 - IEEE International Conference on Communications, Rome, Italy, 2023, pp. 3596-3601, doi: 10.1109/ICC45041.2023.10279055.
- [KA20] Ahmad Ali Khan and Raviraj S. Adve. "Centralized and Distributed Deep Reinforcement Learning Methods for Downlink Sum-Rate Optimization." In: IEEE Transactions on Wireless Communications 19.12 (2020), pp. 8410–8426. doi: [10.1109/TWC.2020.3022705](https://doi.org/10.1109/TWC.2020.3022705).
- [Kan+24] Hua Kang, [Qingyong Hu](#), [Huangxun Chen](#), [Qianyi Huang](#), Qian Zhang, and Min Cheng. "Cross-shaped Separated Spatial-Temporal UNet Transformer for Accurate Channel Prediction," IEEE INFOCOM 2024 - IEEE Conference on Computer Communications, Vancouver, BC, Canada, 2024, pp. 2079-2088, doi: 10.1109/INFOCOM52122.2024.10621187.
- [Kar+21] Abdul Karim Gizzini et al. "CNN Aided Weighted Interpolation for Channel Estimation in Vehicular Communications." In: IEEE Transactions on Vehicular Technology 70.12 (2021), pp. 12796–12811. doi: 10.1109/TVT.2021.3120267.
- [Kav+24] Fatemeh Kavehmadavani et al. "Empowering Traffic Steering in 6G Open RAN with Deep Reinforcement Learning." In: IEEE Transactions on Wireless Communications (2024).

- [KC21] Shubham Khunteta and Ashok Kumar Reddy Chavva. "Recurrent Neural Network Based Beam Prediction for Millimeter-Wave 5G Systems." In: 2021 IEEE Wireless Communications and Networking Conference (WCNC). 2021, pp. 1–6.
doi: [10.1109/WCNC49053.2021.9417509](https://doi.org/10.1109/WCNC49053.2021.9417509).
- [Kim+23] Donghyeon Kim et al. "Deep Learning-Based Resource Allocation Scheme for Heterogeneous NOMA Networks." In: IEEE Access 11 (2023), pp. 89423–89432.
doi: [10.1109/ACCESS.2023.3307407](https://doi.org/10.1109/ACCESS.2023.3307407).
- [KK24] Taehyoung Kim and Gyuyeol Kong. "Deep Learning-Driven Interference Perceptual Multi-Modulation for Full-Duplex Systems." Mathematics. 2024. 12, no. 10: 1542.
<https://doi.org/10.3390/math12101542>
- [KLV24] Petteri Kela, Bryan Liu, and Alvaro Valcarce. "Towards Practical Deep Schedulers for Allocating Cellular Radio Resources." 2024. arXiv: 2411.08529 [eess.SP]. url: <https://arxiv.org/abs/2411.08529>.
- [KNJ24] Prashanth Krishnananthalingam, Nhan Thanh Nguyen, and Markku Juntti. "Constant Modulus Waveform Design for Wideband Multicarrier Joint Communications and Sensing via Deep Unfolding." In: 2024 IEEE Wireless Communications and Networking Conference (WCNC). 2024, pp. 1–6.
doi: [10.1109/WCNC57260.2024.10571059](https://doi.org/10.1109/WCNC57260.2024.10571059).
- [Kor+20] Dani Korpi et al. "DeepRx MIMO: Convolutional MIMO Detection with Learned Multiplicative Transformations." 2020. arXiv: 2010.16283 [eess.SP]. url: <https://arxiv.org/abs/2010.16283>.
- [Kor+22] Dani Korpi et al. "Waveform Learning for Reduced Out-of-Band Emissions Under a Nonlinear Power Amplifier." 2022. doi: [10.48550/ARXIV.2201.05524](https://doi.org/10.48550/ARXIV.2201.05524). url: <https://arxiv.org/abs/2201.05524>.
- [KS24] Naveed Ali Khan and Stefan Schmid. "AI-RAN in 6G Networks: State-of-the-Art and Challenges." In: IEEE Open Journal of the Communications Society 5 (2024), pp. 294–311. doi: [10.1109/OJCOMS.2023.3343069](https://doi.org/10.1109/OJCOMS.2023.3343069).
- [Kum+25] LS, Yaswanth Kumar, et al. "FALCON: A Framework for Fault Prediction in Open RAN Using Multi-Level Telemetry." arXiv preprint arXiv:2503.06197 (2025).
- [LD24] Shuangshuang Li and Peihao Dong, "Mixed Attention Transformer Enhanced Channel Estimation for Extremely Large-Scale MIMO Systems," 2024 16th International Conference on Wireless Communications and Signal Processing (WCSP), Hefei, China, 2024, pp. 394–399, doi: [10.1109/WCSP62071.2024.10827075](https://doi.org/10.1109/WCSP62071.2024.10827075).
- [Lee+20] C. Lee et al., "Prediction-Based Conditional Handover for 5G mm-Wave Networks: A Deep-Learning Approach", IEEE Veh. Technol. Mag., vol. 15, no. 1, pp. 54-62, Mar. 2020.

- [Lee+21] Woong-Hee Lee et al. "Noise Learning-Based Denoising Autoencoder." In: IEEE Communications Letters 25.9 (2021), pp. 2983–2987.
doi: [10.1109/LCOMM.2021.3091800](https://doi.org/10.1109/LCOMM.2021.3091800).
- [Lee+22] Anho Lee et al. "Intelligent Near-Field Channel Estimation for Terahertz Ultra-Massive MIMO Systems." In: GLOBECOM 2022 - 2022 IEEE Global Communications Conference. 2022, pp. 5390–5395. doi: [10.1109/GLOBECOM48099.2022.10001612](https://doi.org/10.1109/GLOBECOM48099.2022.10001612).
- [Lei+24] Hao Lei et al. "Channel Estimation for XL-MIMO Systems with Polar-Domain Multi-Scale Residual Dense Network." In: IEEE Transactions on Vehicular Technology 73.1 (2024), pp. 1479–1484. doi: [10.1109/TVT.2023.3311010](https://doi.org/10.1109/TVT.2023.3311010).
- [Li+18] Zan Li et al. "Improved cooperative spectrum sensing model based on machine learning for cognitive radio networks." In: IET Communications 12 (19 2018), pp. 2485–2492. doi:[10.1049/iet-com.2018.5245](https://doi.org/10.1049/iet-com.2018.5245).
- [Li+20] Yu-Ngok Ruyue Li et al. "Beam Management in Millimeter-Wave Communications for 5G and Beyond." In: IEEE Access 8 (2020), pp. 13282–13293.
doi: [10.1109/ACCESS.2019.2963514](https://doi.org/10.1109/ACCESS.2019.2963514).
- [Li+23a] Qiaoyu Li et al. "Machine Learning Based Time Domain Millimeter-Wave Beam Prediction for 5G-Advanced and Beyond: Design, Analysis, and Over-The-Air Experiments." In: IEEE Journal on Selected Areas in Communications 41.6 (2023), pp. 1787–1809. doi:[10.1109/JSAC.2023.3275613](https://doi.org/10.1109/JSAC.2023.3275613).
- [Li+23b] Kaiyue Li et al. "Double Deep Learning for Joint Phase-Shift and Beamforming Based on Cascaded Channels in RIS-Assisted MIMO Networks." In: IEEE Wireless Communications Letters 12.4 (2023), pp. 659–663. doi: [10.1109/LWC.2023.3238073](https://doi.org/10.1109/LWC.2023.3238073).
- [Li+24a] J. Li et al. "Anomaly Detection for Scalable Task Grouping in Reinforcement Learning-Based RAN Optimization." In: 2024 IEEE International Conference on Communications Workshops (ICC Workshops), Denver, CO, USA, 2024, pp. 1395–1400.
doi: [10.1109/ICCWorkshops59551.2024.10615617](https://doi.org/10.1109/ICCWorkshops59551.2024.10615617).
- [Li+24b] Anna Li et al. "An Integrated Sensing and Communication System for Fall Detection and Recognition Using Ultrawideband Signals." In: IEEE Internet of Things Journal 11.1 (2024), pp. 1509–1521. doi: [10.1109/IJOT.2023.3290421](https://doi.org/10.1109/IJOT.2023.3290421).
- [Li+25] [Haiyang Li](https://arxiv.org/abs/2509.12748) et al. "NEFT: A Unified Transformer Framework for Efficient Near-Field CSI Feedback in XL-MIMO Systems." 2025. arXiv: 2509.12748 [eess.SP]. URL: <https://arxiv.org/abs/2509.12748>
- [Lim+23] J. P. S. H. Lima et al., "Deep Learning-Based Handover Prediction for 5G and Beyond Networks," ICC 2023 - IEEE International Conference on Communications, Rome, Italy, 2023, pp. 3468–3473, doi: [10.1109/ICC45041.2023.10279195](https://doi.org/10.1109/ICC45041.2023.10279195).

- [Lin+23] Xingqin Lin et al. "6G Digital Twin Networks: From Theory to Practice." In: IEEE Communications Magazine 61.11 (2023), pp. 72–78.
doi: 10.1109/MCOM.001.2200830.
- [Lin23] Xingqin Lin. "An Overview of the 3GPP Study on Artificial Intelligence for 5G New Radio." 2023. arXiv: 2308.05315 [cs.NI]. url: <https://arxiv.org/abs/2308.05315>.
- [Lim+25] Seongmok Lim et al. "6G MIMO: Toward Intelligent and Energy-Efficient Radio" In: SR Blog [Online]. Available: <https://research.samsung.com/blog/6G-MIMO-Toward-Intelligent-and-Energy-Efficient-Radio>
- [Liu+20] Shicong Liu et al. "Deep Denoising Neural Network Assisted Compressive Channel Estimation for mmWave Intelligent Reflecting Surfaces." In: IEEE Transactions on Vehicular Technology 69.8 (2020), pp. 9223–9228. doi: 10.1109/TVT.2020.3005402.
- [Liu+21] Chang Liu et al. "Deep Residual Network Empowered Channel Estimation for IRS-Assisted Multi-User Communication Systems." In: ICC 2021 - IEEE International Conference on Communications. 2021, pp. 1–7.
doi: 10.1109/ICC42927.2021.9500708.
- [Liu+22a] Zhiyan Liu et al. "Deep Unsupervised Learning for Joint Antenna Selection and Hybrid Beamforming." In: IEEE Transactions on Communications 70.3 (2022), pp. 1697–1710.
doi: [10.1109/TCOMM.2022.3143122](https://doi.org/10.1109/TCOMM.2022.3143122).
- [Liu+22b] Chang Liu et al. "Deep Residual Learning for Channel Estimation in Intelligent Reflecting Surface-Assisted Multi-User Communications." In: IEEE Transactions on Wireless Communications 21.2 (2022), pp. 898–912.
doi: 10.1109/TWC.2021.3100148.
- [Liu+22c] Fan Liu et al. "Integrated Sensing and Communications: Toward Dual-Functional Wireless Networks for 6G and Beyond." In: IEEE Journal on Selected Areas in Communications 40.6 (2022), pp. 1728–1767. doi: [10.1109/JSAC.2022.3156632](https://doi.org/10.1109/JSAC.2022.3156632).
- [Liu+22d] Yu Liu et al. "Deep-Learning-Based Channel Estimation for IRS-Assisted ISAC System." In: GLOBECOM 2022 - 2022 IEEE Global Communications Conference. 2022, pp. 4220– 4225. doi: [10.1109/GLOBECOM48099.2022.10001672](https://doi.org/10.1109/GLOBECOM48099.2022.10001672).
- [Liu+22e] Chang Liu et al. "Deep CLSTM for Predictive Beamforming in Integrated Sensing and Communication-Enabled Vehicular Networks," in Journal of Communications and Information Networks, vol. 7, no. 3, pp. 269–277, Sept. 2022,
doi: 10.23919/JCIN.2022.9906941.
- [Liu+22f] Chang Liu et al., "Learning-Based Predictive Beamforming for Integrated Sensing and Communication in Vehicular Networks," in IEEE Journal on Selected Areas in Communications, vol. 40, no. 8, pp. 2317–2334, Aug. 2022,
doi: 10.1109/JSAC.2022.3180803.

- [Liu+23a] Wang Liu et al. "Deep Learning Based Beam Training for Extremely Large-Scale MassiveMIMO in Near-Field Domain." In: IEEE Communications Letters 27.1 (2023), pp. 170–174. doi: [10.1109/LCOMM.2022.3210042](https://doi.org/10.1109/LCOMM.2022.3210042).
- [Liu+23b] Yu Liu et al. "Deep-Learning Channel Estimation for IRS-Assisted Integrated Sensing and Communication System." In: IEEE Transactions on Vehicular Technology 72.5 (2023), pp. 6181–6193. doi: [10.1109/TVT.2022.3231727](https://doi.org/10.1109/TVT.2022.3231727).
- [Liu+23c] Yu Liu et al. "Extreme Learning Machine-Based Channel Estimation in IRS-Assisted Multi-User ISAC System." In: IEEE Transactions on Communications 71.12 (2023), pp. 6993–7007. doi: [10.1109/TCOMM.2023.3308150](https://doi.org/10.1109/TCOMM.2023.3308150).
- [Liu+24a] Zhilong Liu et al. "Cell-Free XL-MIMO Meets Multi-Agent Reinforcement Learning: Architectures, Challenges, and Future Directions." In: IEEE Wireless Communications 31.4 (2024), pp. 155–162. doi: [10.1109/MWC.007.2300176](https://doi.org/10.1109/MWC.007.2300176).
- [Liu+24b] Ziheng Liu et al. "Double-Layer Power Control for Mobile Cell-Free XL-MIMO With Multi-Agent Reinforcement Learning." In: IEEE Transactions on Wireless Communications 23.5 (2024), pp. 4658–4674. doi: [10.1109/TWC.2023.3321334](https://doi.org/10.1109/TWC.2023.3321334).
- [Liu+25] Haochen Liu et al. "Predictive ISAC Beamforming Using LoRA-Transformer LSTM Integrated with GNN in Ultra-Dense D2D mmWave Networks," in IEEE Transactions on Vehicular Technology, vol. 74, no. 9, pp. 13938-13952, Sept. 2025, doi: [10.1109/TVT.2025.3559445](https://doi.org/10.1109/TVT.2025.3559445).
- [LKC18] Woongsup Lee, Minhoe Kim, and Dong-Ho Cho. "Deep Power Control: Transmit Power Control Scheme Based on Convolutional Neural Network." In: IEEE Communications Letters 22.6 (2018), pp. 1276–1279. doi: [10.1109/LCOMM.2018.2825444](https://doi.org/10.1109/LCOMM.2018.2825444).
- [LKC19] Woongsup Lee, Minhoe Kim, and Dong-Ho Cho. "Deep Cooperative Sensing: Cooperative Spectrum Sensing Based on Convolutional Neural Networks." In: IEEE Transactions on Vehicular Technology 68.3 (2019), pp. 3005–3009. doi: [10.1109/TVT.2019.2891291](https://doi.org/10.1109/TVT.2019.2891291).
- [LL23] Daniela Laselva and Mads Lauridsen. "Reducing energy use with 5G-Advanced." In: Nokia White Paper [Online]. Available: <https://onestore.nokia.com/asset/213599> (2023).
- [LL25] Ken Long and Hongjun Liu. "Transformer-Enabled Hybrid Precoding for TDD Large-Scale Antenna Arrays Systems with Channel Sensing," in Journal of Communications and Networks, vol. 27, no. 1, pp. 23-31, Feb. 2025, doi: [10.23919/JCN.2025.0000002](https://doi.org/10.23919/JCN.2025.0000002).
- [LLD24] Siying Liu, Congduan Li, and Jieyi Deng. "Real-Time Wireless Channel Prediction Based on Online Federated Learning." In: 2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring). 2024, pp. 1–6. doi: [10.1109/VTC2024-Spring62846.2024.10683453](https://doi.org/10.1109/VTC2024-Spring62846.2024.10683453).

- [LMD21] Enrique M. Lizarraga, Gabriel N. Maggio, and Alexis A. Dowhuszko. "Deep reinforcement learning for hybrid beamforming in multi-user millimeter wave wireless systems." In: 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring). 2021, pp. 1–5. doi: [10.1109/VTC2021-Spring51267.2021.9449053](https://doi.org/10.1109/VTC2021-Spring51267.2021.9449053).
- [Lok+23] Maheshi U. Lokumarambage et al. "Wireless End-to-End Image Transmission System Using Semantic Communications." In: IEEE Access 11 (2023), pp. 37149–37163. doi: 10.1109/ACCESS.2023.3266656.
- [Lop+22] D. López-Pérez et al. "A Survey on 5G Radio Access Network Energy Efficiency: Massive MIMO, Lean Carrier Design, Sleep Modes, and Machine Learning," In: IEEE Communications Surveys & Tutorials 24.1, pp. 653–697.
- [LT23] [Dianxin](#) Luan and John S. Thompson, "Channelformer: Attention Based Neural Solution for Wireless Channel Estimation and Effective Online Training," in IEEE Transactions on Wireless Communications, vol. 22, no. 10, pp. 6562-6577, Oct. 2023, doi: 10.1109/TWC.2023.3244484.
- [Lu+20] Chao Lu et al. "Bit-Level Optimized Neural Network for Multi-Antenna Channel Quantization." In: IEEE Wireless Communications Letters 9.1 (2020), pp. 87–90. doi: 10.1109/LWC.2019.2942908.
- [Lu+24] [Haiquan](#) Lu et al. "A Tutorial on Near-Field XL-MIMO Communications Toward 6G," in IEEE Communications Surveys & Tutorials, vol. 26, no. 4, pp. 2213-2257, Fourth quarter 2024, doi: 10.1109/COMST.2024.3387749.
- [Luo+22] Lirui Luo et al. "Downlink Power Control for Cell-Free Massive MIMO With Deep Reinforcement Learning." In: IEEE Transactions on Vehicular Technology 71.6 (2022), pp. 6772– 6777. doi: [10.1109/TVT.2022.3162585](https://doi.org/10.1109/TVT.2022.3162585).
- [Luo+25] Riku Luostari et al. "Adapting to Reality: Over-the-Air Validation of AI-Based Receivers Trained with Simulated Channels." 2025. arXiv: 2408.04182 [eess.SP]. url: <https://arxiv.org/abs/2408.04182>.
- [LWS20] Zhilin Lu, Jintao Wang, and Jian Song. "Multi-resolution CSI Feedback with Deep Learning in Massive MIMO System." In: ICC 2020 - 2020 IEEE International Conference on Communications (ICC). 2020, pp. 1–6. doi: 10.1109/ICC40277.2020.9149229.
- [LYH06] Wei Liu, Lie-Liang Yang, and L. Hanzo. "Recurrent Neural Network Based Narrowband Channel Prediction." In: 2006 IEEE 63rd Vehicular Technology Conference. Vol. 5. 2006, pp. 2173–2177. doi: [10.1109/VETECS.2006.1683241](https://doi.org/10.1109/VETECS.2006.1683241).
- [LZD20] Zhenyu Liu, Lin Zhang, and Zhi Ding. "An Efficient Deep Learning Framework for Low Rate Massive MIMO CSI Reporting." In: IEEE Transactions on Communications 68.8 (2020), pp. 4761–4772. doi: 10.1109/TCOMM.2020.2993626.

- [LZL21] Tongwei Lu, Haijun Zhang, and Keping Long. "Joint Beamforming and Power Control for MIMO-NOMA with Deep Reinforcement Learning." In: ICC 2021 - IEEE International Conference on Communications. 2021, pp. 1–5.
doi: [10.1109/ICC42927.2021.9500713](https://doi.org/10.1109/ICC42927.2021.9500713).
- [Ma+23] Ke Ma et al. "Deep Learning for mmWave Beam-Management: State-of-the-Art, Opportunities and Challenges." In: IEEE Wireless Communications 30.4 (2023), pp. 108–114. doi: [10.1109/MWC.018.2100713](https://doi.org/10.1109/MWC.018.2100713).
- [Mad+23] P. Madadi et al. "AI/ML Optimized High-Order Modulations," in ICC 2023 - IEEE International Conference on Communications, Rome, Italy, 2023, pp. 6373-6378, doi: 10.1109/ICC45041.2023.10279091.
- [Mah+24] Haitham Mahmoud et al. "Machine Learning-based Spectrum Allocation using Cognitive Radio Networks." In: 2024 International Joint Conference on Neural Networks (IJCNN). 2024, pp. 1–6. doi: [10.1109/IJCNN60899.2024.10649989](https://doi.org/10.1109/IJCNN60899.2024.10649989).
- [Mam+22] A. Mamane et al. "Scheduling Algorithms for 5G Networks and Beyond: Classification and Survey," in IEEE Access, vol. 10, pp. 51643-51661, 2022, doi: 10.1109/ACCESS.2022.3174579.
- [Mar+22] Dileepa Marasinghe et al. "LiDAR aided Wireless Networks - Beam Prediction for 5G." In: 2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall). 2022, pp. 1–7. doi: [10.1109/VTC2022-Fall57202.2022.10012751](https://doi.org/10.1109/VTC2022-Fall57202.2022.10012751).
- [Mar+24] Dileepa Marasinghe et al. "Waveform Learning under Phase Noise Impairment for SubTHz Communications." In: IEEE Transactions on Communications (2024), pp. 1–1. doi: [10.1109/TCOMM.2024.3430982](https://doi.org/10.1109/TCOMM.2024.3430982).
- [Mas+21] A. Masri et al., "Machine-learning-based predictive handover." Proc. IFIP/IEEE Int. Symp. Integr. Netw. Manage. (IM), pp. 648-652, May 2021.
- [Mas+23a] Meysam Masoudi et al. "Digital Twin Assisted Risk-Aware Sleep Mode Management Using Deep Q-Networks." In: IEEE Transactions on Vehicular Technology 72.1 (2022), pp. 1224–1239.
- [Mas+23b] Antonino Masaracchia et al. "Digital Twin for Open RAN: Toward Intelligent and Resilient 6G Radio Access Networks." In: IEEE Communications Magazine 61.11 (2023), pp. 112– 118. doi: 10.1109/MCOM.003.2200879.
- [Mas+25] Saad Masrur et al. "Transforming Indoor Localization: Advanced Transformer Architecture for NLOS Dominated Wireless Environments with Distributed Sensors." 2025. arXiv: 2501.07774 [cs.LG]. url : <https://arxiv.org/pdf/2501.07774>.
- [Mat+20] Marja Matinmikko-Blue et al. "White paper on 6G drivers and the UN SDGs." In: 6G Research Visions, No. 2. University of Oulu. (2020). url: <http://urn.fi/urn:isbn:9789526226699>.

- [Mat+22] José Miguel Mateos-Ramos et al. "End-to-End Learning for Integrated Sensing and Communication." In: ICC 2022 - IEEE International Conference on Communications. 2022, pp. 1942–1947. doi: [10.1109/ICC45855.2022.9838308](https://doi.org/10.1109/ICC45855.2022.9838308).
- [Max+24] Stefano Maxenti et al. "ScalO-RAN: Energy-aware Network Intelligence Scaling in Open RAN." In: Proc. of IEEE International Conference on Computer Communications (INFOCOM). May 2024, pp. 1–6.
- [ME18] Faris B. Mismar; Brian L. Evans. "Q-Learning Algorithm for VoLTE Closed Loop Power Control in Indoor Small Cells," 2018 52nd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 2018, pp. 1485-1489, doi: 10.1109/ACSSC.2018.8645168.
- [MED16] À. O. Martínez, P. Eggers, and E. De Carvalho, "Geometry-based stochastic channel models for 5G: Extending key features for massive MIMO," 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Valencia, Spain, 2016, pp. 1-6, doi: 10.1109/PIMRC.2016.7794648.
- [Mek+23] A. Mekrache, K. Boutiba, and A. Ksentini. "Combining Network Data Analytics Function and Machine Learning for Abnormal Traffic Detection in Beyond 5G." In: GLOBECOM 2023 - 2023 IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 2023, pp. 1204–1209. doi: 10.1109/GLOBECOM54140.2023.10436766.
- [Mer+23] Mattia Merluzzi et al., "The Hexa-X Project Vision on Artificial Intelligence and Machine Learning-Driven Communication and Computation Co-Design for 6G." In: IEEE Access 11(2023), pp. 65620-65648. doi: 10.1109/ACCESS.2023.3287939.
- [MG21] Mahdi Boloursaz Mashhadi and Deniz Gündüz. "Pruning the Pilots: Deep Learning-Based Pilot Design and Channel Estimation for MIMO-OFDM Systems." In: IEEE Transactions on Wireless Communications 20.10 (2021), pp. 6315–6328. doi: 10.1109/TWC.2021.3073309.
- [ML24] Marie Hogan and Johan Lundsjö "6G Straight from the Ericsson Labs." In: Ericsson Blog [Online]. Available: <https://www.ericsson.com/en/blog/2024/4/6g-straight-from-the-ericsson-labs> (2024).
- [Mni+13] Volodymyr Mnih et al. "Playing Atari with Deep Reinforcement Learning." 2013. arXiv: [1312.5602](https://arxiv.org/abs/1312.5602)[cs.LG]. url: <https://arxiv.org/abs/1312.5602>.
- [Mni+16] Volodymyr Mnih et al. "Asynchronous Methods for Deep Reinforcement Learning." 2016. arXiv: [1602.01783](https://arxiv.org/abs/1602.01783) [cs.LG]. url: <https://arxiv.org/abs/1602.01783>.
- [MS23] Charlotte Muth and Laurent Schmalen. "Autoencoder-based Joint Communication and Sensing of Multiple Targets." In: WSA SCC 2023; 26th International ITG Workshop on Smart Antennas and 13th Conference on Systems, Communications, and Coding. 2023, pp. 1–6.

- [MSG23] Samie Mostafavi, Gourav Prateek Sharma, and James Gross. "Data-driven latency probability prediction for wireless networks: Focusing on tail probabilities." In: GLOBECOM 2023 – 2023 IEEE Global Communications Conference. 2023, pp. 4338–4344. doi: 10.1109/GLOBECOM54140.2023.10437281.
- [MSR23] Charmae Franchesca Mendoza, Stefan Schwarz, and Markus Rupp. "User-Centric Clustering in Cell-Free MIMO Networks using Deep Reinforcement Learning." In: ICC 2023 IEEE International Conference on Communications. 2023, pp. 1036–1041. doi: [10.1109/ICC45041.2023.10279626](https://doi.org/10.1109/ICC45041.2023.10279626).
- [MTL22] Amirhossein Mohammadian, Chintha Tellambura, and Geoffrey Ye Li. "Deep Learning LMMSE Joint Channel, PN, and IQ Imbalance Estimator for Multicarrier MIMO Full-Duplex Systems," in IEEE Wireless Communications Letters, vol. 11, no. 1, pp. 111–115, Jan. 2022, doi: 10.1109/LWC.2021.3121662.
- [Mur+21] Konstantin Muranov et al. "On Deep Learning Assisted Self-Interference Estimation in a Full-Duplex Relay Link," in IEEE Wireless Communications Letters, vol. 10, no. 12, pp. 2762–2766, Dec. 2021, doi: 10.1109/LWC.2021.3116188.
- [MYG21] Mahdi Boloursaz Mashhadi, Qianqian Yang, and Deniz Gündüz. "Distributed Deep Convolutional Compression for Massive MIMO CSI Feedback." In: IEEE Transactions on Wireless Communications 20.4 (2021), pp. 2621–2633. doi: 10.1109/TWC.2020.3043502.
- [NA+24] Hisyam Ng, Hibatul Azizi, and Toktam Mahmoodi. "Machine Learning-Driven Dynamic Traffic Steering in 6G: A Novel Path Selection Scheme." In: Big Data and Cognitive Computing 8.12 (2024), p. 172.
- [NA19] Shuai Nie and Ian F. Akyildiz. "Deep Kernel Learning-Based Channel Estimation in Ultra-Massive MIMO Communications at 0.06–10 THz." In: 2019 IEEE Globecom Workshops (GC Wkshps). 2019, pp. 1–6. doi: 10.1109/GCWkshps45667.2019.9024624.
- [Nay+17] E. Nayebi et al. "Precoding and Power Optimization in Cell-Free Massive MIMO Systems." In: IEEE Transactions on Wireless Communications 16.7 (2017), pp. 4445–4459. doi: [10.1109/TWC.2017.2698449](https://doi.org/10.1109/TWC.2017.2698449).
- [Ner+23] Matteo Nerini et al. "Machine Learning-Based CSI Feedback with Variable Length in FDD Massive MIMO." In: IEEE Transactions on Wireless Communications 22.5 (2023), pp. 2886–2900. doi: 10.1109/TWC.2022.3215104.
- [Nér+25] [Sylvain](#) Néronat et al. "Efficient 5G Resource Block Scheduling Using Action Branching and Transformer Networks," *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, Barcelona, Spain, 2025, pp. 1–6, doi: 10.1109/ICMLCN64995.2025.11140453.

- [Ngo+15] H. Q. Ngo et al. “Cell-Free Massive MIMO: Uniformly Great Service for Everyone.” In: 2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). 2015, pp. 201–205. doi: [10.1109/SPAWC.2015.7227028](https://doi.org/10.1109/SPAWC.2015.7227028).
- [Ngo+17] H. Q. Ngo et al. “Cell-Free Massive MIMO Versus Small Cells.” In: IEEE Transactions on Wireless Communications 16.3 (2017), pp. 1834–1850.
- [Ngu+24] Nhan Thanh Nguyen et al. “Joint Communications and Sensing Hybrid Beamforming Design via Deep Unfolding.” In: IEEE Journal of Selected Topics in Signal Processing 18.5 (2024), pp. 901–916. doi: [10.1109/JSTSP.2024.3463403](https://doi.org/10.1109/JSTSP.2024.3463403).
- [Nie+25] Jiali Nie et al. “Near-Field Beam Training for Extremely Large-Scale MIMO Based on Deep Learning.” In: IEEE Transactions on Mobile Computing 24.1 (2025), pp. 352–362. doi: [10.1109/TMC.2024.3462960](https://doi.org/10.1109/TMC.2024.3462960).
- [Nok23] Nokia. “Taking 5G-Advanced to the Next Level and Bridging into the 6G Era.” In: Nokia White Paper [Online]. Available: <https://onestore.nokia.com/asset/213705> (2023).
- [Nok24] Nokia. “AI opportunities in 6G Layer 2.” In: Nokia White Paper [Online]. Available: <https://onestore.nokia.com/asset/213703> (2024).
- [Nvi+24] NVIDIA, Bell CA, CMCC, DeepSig, Dell, Ericsson, Nokia, Qualcomm, Rakuten, Verizon, VIAVI, “O-RAN Next Generation Research Group (nGRG). “Digital Twin RAN: Key Enablers.” 2024, O-RAN Contributed Research Report, Report ID: RR-2024-09 (2024).
- [OH17] Timothy O’Shea and Jakob Hoydis. “An Introduction to Deep Learning for the Physical Layer.” In: IEEE Transactions on Cognitive Communications and Networking 3.4 (2017), pp. 563–575. doi: [10.1109/TCCN.2017.2758370](https://doi.org/10.1109/TCCN.2017.2758370).
- [Oh+25] Jongho Oh et al. “AI-Driven Modem Innovations for Shaping the Future RAN” In: Samsung Research Blog [Online]. Available: <https://research.samsung.com/blog/-Blog-AI-Driven-Modem-Innovations-for-Shaping-the-Future-RAN>
- [Pan+25] Panagiotis Nikolaidis, et al. “A Proof of Concept Resource Management Scheme for Augmented Reality Applications in 5G Systems.” 2025. arXiv:2501.01398 [cs.NI]. url: <https://arxiv.org/abs/2501.01398>.
- [Par+24] H. -S. Park et al. “Mobility Management Paradigm Shift: From Reactive to Proactive Handover Using AI/ML,” in IEEE Network, vol. 38, no. 2, pp. 18-25, March 2024, doi: [10.1109/MNET.2024.3357108](https://doi.org/10.1109/MNET.2024.3357108)
- [PBJ22] Lissy Pellaco, Mats Bengtsson, and Joakim Jaldén. “Matrix-Inverse-Free Deep Unfolding of the Weighted MMSE Beamforming Algorithm.” In: IEEE Open Journal of the Communications Society 3 (2022), pp. 65–81. doi: [10.1109/OJCOMS.2021.3139858](https://doi.org/10.1109/OJCOMS.2021.3139858).

- [Pek+20] Ture Peken et al. “Deep Learning for SVD and Hybrid Beamforming.” In: IEEE Transactions on Wireless Communications 19.10 (2020), pp. 6621–6642. doi: [10.1109/TWC.2020.3004386](https://doi.org/10.1109/TWC.2020.3004386).
- [Pen+24] Zhangjie Peng et al. “Deep Learning-Based CSI Feedback for XL-MIMO Systems in the Near-Field Domain.” 2024. arXiv: 2405.09053 [eess.SP]. url: <https://arxiv.org/abs/2405.09053>.
- [Per+23] Dariel Pereira-Ruisánchez et al. “Deep Contextual Bandit and Reinforcement Learning for IRS-Assisted MU-MIMO Systems.” In: IEEE Transactions on Vehicular Technology 72.7 (2023), pp. 9099–9114. doi: 10.1109/TVT.2023.3249353.
- [Pih+23] Jaakko Pihlajasalo et al. “Deep Learning OFDM Receivers for Improved Power Efficiency and Coverage.” In: IEEE Transactions on Wireless Communications 22.8 (2023), pp. 5518– 5535. doi: [10.1109/TWC.2023.3235059](https://doi.org/10.1109/TWC.2023.3235059).
- [PNA21] Ridho Hendra Yoga Perdana, Toan-Van Nguyen, and Beongku An. “Deep Learning-based Power Allocation in Massive MIMO Systems with SLNR and SINR Criteria.” In: 2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN). 2021, pp. 87–92. doi: [10.1109/ICUFN49451.2021.9528565](https://doi.org/10.1109/ICUFN49451.2021.9528565).
- [Pol+24] Michele Polese et al. “Colosseum: The Open RAN Digital Twin.” In: IEEE Open Journal of the Communications Society. August 2024, pp. 1–13.
- [PRM21] Michele Polese, Francesco Restuccia, and Tommaso Melodia. “DeepBeam: Deep Waveform Learning for Coordination-Free Beam Management in mmWave Networks.” In: MobiHoc '21. Shanghai, China: Association for Computing Machinery, 2021, pp. 61–70. isbn: 9781450385589. doi: [10.1145/3466772.3467035](https://doi.org/10.1145/3466772.3467035). url: <https://doi.org/10.1145/3466772.3467035>.
- [PVK08] Chris Potter, Ganesh K. Venayagamoorthy, and Kurt Kosbar. “MIMO Beam-Forming with Neural Network Channel Prediction Trained by a Novel PSO-EA-DEPSO Algorithm.” In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). 2008, pp. 3338–3344. doi: [10.1109/IJCNN.2008.4634272](https://doi.org/10.1109/IJCNN.2008.4634272).
- [Qi+23] Qiao Qi et al. “DL-based Joint Waveform and Beamforming Design for Integrated Sensing and Communication.” In: 2023 IEEE Globecom Workshops (GC Wkshps). 2023, pp. 110– 116. doi: [10.1109/GCWkshps58843.2023.10464912](https://doi.org/10.1109/GCWkshps58843.2023.10464912).
- [Qi+24] Qiao Qi et al. “Deep Learning-Based Design of Uplink Integrated Sensing and Communication.” In: IEEE Transactions on Wireless Communications 23.9 (2024), pp. 10639– 10652. doi: [10.1109/TWC.2024.3373797](https://doi.org/10.1109/TWC.2024.3373797).
- [Qin+24] Zhijin Qin et al. “AI Empowered Wireless Communications: From Bits to Semantics.” In: Proceedings of the IEEE 112.7 (2024), pp. 621–652. doi: [10.1109/JPROC.2024.3437730](https://doi.org/10.1109/JPROC.2024.3437730).

- [Qiu+24] Bin Qiu et al. "Federated Learning-Based Channel Estimation for RIS-Aided Communication Systems." In: IEEE Wireless Communications Letters 13.8 (2024), pp. 2130–2134. doi: 10.1109/LWC.2024.3402968.
- [Qur+23] M. Qurratulain Khan et al. "Machine Learning for Millimeter Wave and Terahertz Beam Management: A Survey and Open Challenges." In: IEEE Access 11 (2023), pp. 11880–11902. doi: [10.1109/ACCESS.2023.3242582](https://doi.org/10.1109/ACCESS.2023.3242582).
- [Raf+23] Adeel Rafiq et al. "Knowledge Defined Networks on the Edge for Service Function Chaining and Reactive Traffic Steering." In: Cluster Computing 26.1 (2023), pp. 613–634.
- [Raf+24] Rafail Raftopoulos et al. "DRL-based Latency-Aware Network Slicing in O-RAN with Time-Varying SLAs." In: Proc. of International Conference on Computing, Networking, and Communication (ICNC). Big Island, Hawaii, Feb. 2024, pp. 1–6.
- [Rah+22] Mostafa Rahmani et al. "Deep Reinforcement Learning-based Power Allocation in Uplink Cell-Free Massive MIMO." In: 2022 IEEE Wireless Communications and Networking Conference (WCNC). 2022, pp. 459–464. doi: [10.1109/WCNC51071.2022.9771964](https://doi.org/10.1109/WCNC51071.2022.9771964).
- [Raj+21] Nuwanthika Rajapaksha et al. "Deep Learning-based Power Control for Cell-Free Massive MIMO Networks." In: ICC 2021 - IEEE International Conference on Communications. 2021, pp. 1–7. doi: 10.1109/ICC42927.2021.9500734.
- [Raj+23] Nuwanthika Rajapaksha et al. "Unsupervised Learning-Based Joint Power Control and Fronthaul Capacity Allocation in Cell-Free Massive MIMO With Hardware Impairments." In: IEEE Wireless Communications Letters 12.7 (2023), pp. 1159–1163. doi: 10.1109/LWC.2023.3265348.
- [Raj+24] Nuwanthika Rajapaksha et al. "Minimizing Energy Consumption in MU-MIMO via Antenna Muting by Neural Networks with Asymmetric Loss." In: IEEE Transactions on Vehicular Technology 73.5 (2024), pp. 6600–6613. doi: 10.1109/TVT.2023.3339340.
- [RCT24] B. -M. Robaglia, M. Coupechoux and D. Tsilimantos, "Deep Reinforcement Learning for Uplink Scheduling in NOMA-URLLC Networks," in IEEE Transactions on Machine Learning in Communications and Networking, vol. 2, pp. 1142-1158, 2024, doi: 10.1109/TMLCN.2024.3437351.
- [Riv+23] Steven Rivetti et al. "Spatial Signal Design for Positioning via End-to-End Learning," In: IEEE Wireless Communications Letters, vol. 12, no. 3, pp. 525-529, March 2023. doi: 10.1109/LWC.2022.3233475.
- [Riz+20] Valentina Rizzello et al. "Learning the Channel State Information Across the Frequency Division Gap in Wireless Communications." In: WSA 2020; 24th International ITG Work-shop on Smart Antennas. 2020, pp. 1–6.

- [Rob+24a] Clifton Paul Robinson et al. "DeepSweep: Parallel and Scalable Spectrum Sensing via Convolutional Neural Networks," 2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN), Stockholm, Sweden, 2024, pp. 505-510, doi: 10.1109/ICMLCN59089.2024.10624803.
- [Rob+24b] Clifton Paul Robinson et al. "EXPLORA: Explainability for Deep Reinforcement Learning-based Control of Open RAN." In: Proc. of IEEE International Conference on Computer Communications (INFOCOM). 2024, pp. 1–6.
- [Rob+24c] Clifton Paul Robinson et al. "TwiNet: Connecting Real World Networks to their Digital Twins Through a Live Bidirectional Link." In: arXiv preprint arXiv:2411.03503. 2024, pp. 1–13.
- [RWX20] Xin Ru, Li Wei, and Youyun Xu. "Model-Driven Channel Estimation for OFDM Systems Based on Image Super-Resolution Network." In: 2020 IEEE 5th International Conference on Signal and Image Processing (ICSIP). 2020, pp. 804–808. doi: 10.1109/ICSIP49896.2020.9339375.
- [Ryd+23] Henrik Rydén et al. "Next Generation Mobile Networks' Enablers: Machine Learning-Assisted Mobility, Traffic, and Radio Channel Prediction," In: IEEE Communications Magazine 61.10 (2023), pp. 94-98. doi: 10.1109/MCOM.001.2200592.
- [Sae+24] Ali Saeizadeh et al. "AI-assisted Agile Propagation Modeling for Real-time Digital Twin Wireless Networks." In: arXiv preprint arXiv:2410.22437. 2024, pp. 1–13.
- [SAY21] F. Sahrabi, K. M. Attiah and W. Yu, "Deep Learning for Distributed Channel Feedback and Multiuser Precoding in FDD Massive MIMO," in IEEE Transactions on Wireless Communications, vol. 20, no. 7, pp. 4044-4057, July 2021, doi: 10.1109/TWC.2021.3055202.
- [Sca+24] Alessio Scalingi et al. "Det-RAN: Data-Driven Cross-Layer Real-Time Attack Detection in 5G Open RANs." In: IEEE INFOCOM 2024 – IEEE Conference on Computer Communications. May 2024, pp. 1–6.
- [Sch+17] John Schulman et al. "Proximal Policy Optimization Algorithms." 2017. arXiv: [1707.06347 \[cs.LG\]](https://arxiv.org/abs/1707.06347). url: <https://arxiv.org/abs/1707.06347>.
- Shuang shuang Li and Peihao Dong, "Mixed Attention Transformer Enhanced Channel Estimation for Extremely Large- },
- [Sha+24] Riaz Shaik et al. "Real-Time Anomaly Detection in 5G Networks Through Edge Computing." In: 2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS). IEEE, 2024.
- [Sha+24] D. Shakya et al. "Radio Propagation Measurements and Statistical Channel Models for Outdoor Urban Microcells in Open Squares and Streets at 142, 73, and 28 GHz." in IEEE Transactions on Antennas and Propagation, vol. 72, no. 4, pp. 3580-3595, April 2024, doi: 10.1109/TAP.2024.3366581.

- [She+22a] M. Karam Shehzad et al. "Real-Time Massive MIMO Channel Prediction: A Combination of Deep Learning and NeuralProphet." In: GLOBECOM 2022 - 2022 IEEE Global Communications Conference. 2022, pp. 1423–1428.
doi: [10.1109/GLOBECOM48099.2022.10001362](https://doi.org/10.1109/GLOBECOM48099.2022.10001362).
- [She+22b] Muhammad K. Shehzad et al. "ML-Based Massive MIMO Channel Prediction: Does It Work on Real-World Data?" In: IEEE Wireless Communications Letters 11.4 (2022), pp. 811–815. doi: [10.1109/LWC.2022.3146230](https://doi.org/10.1109/LWC.2022.3146230).
- [She+22c] Wei Shen et al. "AI for Per-User Mobility Management and Traffic Steering in Open RAN." In: Proc. of ACM International Conference on emerging Networking EXperiments and Technologies (CoNEXT). Rome, Italy, Dec. 2022, pp. 446–459.
- [Shi+11] Qingjiang Shi et al. "An Iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel." In: IEEE Transactions on Signal Processing 59.9 (2011), pp. 4331–4340. doi: [10.1109/TSP.2011.2147784](https://doi.org/10.1109/TSP.2011.2147784).
- [Shi+23] Wei Shi et al. "Communication-Efficient Orchestrations for URLLC Service via Hierarchical Reinforcement Learning." In: 2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). 2023, pp. 1-6, doi: [10.1109/PIMRC56721.2023.10293856](https://doi.org/10.1109/PIMRC56721.2023.10293856).
- [SK20] Rahil Sarikhani and Farshid Keynia. "Cooperative Spectrum Sensing Meets Machine Learning: Deep Reinforcement Learning Approach." In: IEEE Communications Letters 24.7 (2020), pp. 1459–1462. doi: [10.1109/LCOMM.2020.2984430](https://doi.org/10.1109/LCOMM.2020.2984430).
- [SKC22] D. Sandberg, T. Kvernvik, and F. D. Calabrese, "Learning Robust Scheduling with Search and Attention," In: ICC 2022 - IEEE International Conference on Communications, Seoul, Korea, Republic of, 2022, pp. 1549–1555, doi: [10.1109/ICC45855.2022.9838584](https://doi.org/10.1109/ICC45855.2022.9838584).
- [Sko+23] Marco Skocaj et al. "Data-driven predictive latency for 5G: A theoretical and experimental analysis using network measurements." In: 2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). 2023, pp. 1-6. doi: [10.1109/PIMRC56721.2023.10293861](https://doi.org/10.1109/PIMRC56721.2023.10293861).
- [Sme+24] Carsten Smeenk et al. "Optimizing Radio Resources for Radar Services in ISAC Systems by Deep Reinforcement Learning." In: 2024 IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). 2024, pp. 1–6. doi: [10.1109/PIMRC59610.2024.10817257](https://doi.org/10.1109/PIMRC59610.2024.10817257).
- [Sol+19] Mehran Soltani et al. "Deep Learning-Based Channel Estimation." In: IEEE Communications Letters 23.4 (2019), pp. 652–655.
doi: [10.1109/LCOMM.2019.2898944](https://doi.org/10.1109/LCOMM.2019.2898944).

- [Son+21a] Hao Song et al. "A Deep Reinforcement Learning Framework for Spectrum Management in Dynamic Spectrum Access." In: IEEE Internet of Things Journal 8.14 (2021), pp. 11208– 11218. doi: [10.1109/JIOT.2021.3052691](https://doi.org/10.1109/JIOT.2021.3052691).
- [Son+21b] Haobo Song et al. "Unsupervised Learning-Based Joint Active and Passive Beamforming Design for Reconfigurable Intelligent Surfaces Aided Wireless Networks." In: IEEE Communications Letters 25.3 (2021), pp. 892–896. doi: 10.1109/LCOMM.2020.3041510.
- [SQN22] Wenhan Shen, Zhijin Qin, and Arumugam Nallanathan. "Federated Learning Enabled Channel Estimation for RIS-Aided Multi-User Wireless Systems." In: 2022 IEEE International Conference on Communications Workshops (ICC Workshops). 2022, pp. 812–817. doi: 10.1109/ICCWorkshops53468.2022.9814694.
- [Sro+24] P. Sroka et al. "Policy-Based Traffic Steering and Load Balancing in O-RAN-Based Vehicle-to-Network Communications," in IEEE Transactions on Vehicular Technology, vol. 73, no. 7, pp. 9356–9369, July 2024, doi: 10.1109/TVT.2024.3399924.
- [SS20] J. H. Seong and D. H. Seo. "Selective Unsupervised Learning-Based Wi-Fi Fingerprint System Using Autoencoder and GAN." In: IEEE Internet of Things Journal 7.3 (2020), pp. 1898–1909. doi: [10.1109/JIOT.2019.2956986](https://doi.org/10.1109/JIOT.2019.2956986).
- [Sun+19] Xiaoyu Sun et al. "Fingerprint-Based Localization for Massive MIMO-OFDM System with Deep Convolutional Neural Networks." In: IEEE Transactions on Vehicular Technology 68.11 (2019), pp. 10846–10857. doi: [10.1109/TVT.2019.2939209](https://doi.org/10.1109/TVT.2019.2939209).
- [Sun+21] Yuyao Sun et al. "A Lightweight Deep Network for Efficient CSI Feedback in Massive MIMO Systems." In: IEEE Wireless Communications Letters 10.8 (2021), pp. 1840–1844. doi: 10.1109/LWC.2021.3083331.
- [Sun+24] Chuanhao Sun et al. "SpotLight: Accurate, Explainable and Efficient Anomaly Detection for Open RAN." In: Proceedings of the 30th Annual International Conference on Mobile Computing and Networking, 2024.
- [SYF25] Maryam Al Shami, Jun Yan, and Emmanuel Thepie Fapi. "O-RAN xApps conflict management using graph convolutional networks." arXiv preprint arXiv:2503.03523 (2025).
- [SZD18] Luca Sanguinetti, Alessio Zappone, and Merouane Debbah. "Deep Learning Power Allocation in Massive MIMO." In: 2018 52nd Asilomar Conference on Signals, Systems, and Computers. 2018, pp. 1257–1261. doi: [10.1109/ACSSC.2018.8645343](https://doi.org/10.1109/ACSSC.2018.8645343).
- [TAK22] Fitsum Debebe Tilahun, Ameha Tsegaye Abebe, and Chung G. Kang. "DRL-based Distributed Resource Allocation for Edge Computing in Cell-Free Massive MIMO Network." In: GLOBECOM 2022 - 2022 IEEE Global Communications Conference. 2022, pp. 3845– 3850. doi: [10.1109/GLOBECOM48099.2022.10000910](https://doi.org/10.1109/GLOBECOM48099.2022.10000910).

- [Tan+20] M. R. Tanhatalab et al. "Deep RAN: A Scalable Data-Driven Platform to Detect Anomalies in Live Cellular Network Using Recurrent Convolutional Neural Network." In: 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, Slovakia, 2020, pp. 269–274. doi: 10.1109/SAMI48414.2020.9108742.
- [Tes+24] Paolo Testolina et al. "BostonTwin: The Boston Digital Twin for Ray-Tracing in 6G Networks." In: Proceedings of the ACM Multimedia Systems Conference 2024 (MMSys '24), Bari, Italy. Apr. 2024.
- [Tha+22] Anurag Thantharate et al. "Balanced5G-Fair Traffic Steering Technique Using Data-Driven Learning in Beyond 5G Systems." In: 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2022.
- [Thi+13] Karaputugala Madushan Thilina et al. "Machine Learning Techniques for Cooperative Spectrum Sensing in Cognitive Radio Networks." In: IEEE Journal on Selected Areas in Communications 31.11 (2013), pp. 2209–2221. doi: [10.1109/JSAC.2013.1311120](https://doi.org/10.1109/JSAC.2013.1311120).
- [TM24] Maria Tsampazi and Tommaso Melodia. "System-Level Experimental Evaluation of Reconfigurable Intelligent Surfaces for NextG Communication Systems." In: arXiv preprint arXiv:2412.12969. 2024, pp. 1–16.
- [Tra+24] Minh-Ngoc Tran, Van-Binh Duong, and Younghan Kim. "Design of Computing-Aware Traffic Steering Architecture for 5G Mobile User Plane." In: IEEE Access (2024).
- [Tsa+24] Maria Tsampazi et al. "PandORA: Automated Design and Comprehensive Evaluation of Deep Reinforcement Learning Agents for Open RAN." In: IEEE Transactions on Mobile Computing. 2024, pp. 1–13.
- [Tsa+24] Maria Tsampazi et al. "A Comparative Analysis of Deep Reinforcement Learning-based xApps in O-RAN." In: Proc. of IEEE International Conference on Computer Communications (INFOCOM). 2024, pp. 1–6.
- [Tsa+25] Maria Tsampazi et al. "O-RIS-ing: Evaluating RIS-Assisted NextG Open RAN." In: Proc. of the 3rd IEEE Workshop on Next-generation Open and Programmable Radio Access Networks (NG-OPERA). 2025, pp. 1–12. arXiv: 2502.18753.
- [Tuf+23] L.-M. Tufeanu et al. "Enabling Reinforcement Learning for Network Slice Management in Multi-Agent 5G Networks." In: 2023 IEEE 9th World Forum on Internet of Things (WF-IoT), Aveiro, Portugal, 2023, pp. 1–6. doi: 10.1109/WF-IoT58464.2023.10539464.
- [Uva+21] D. Uvaydov et al. "DeepSense: Fast Wideband Spectrum Sensing Through Real-Time In-the-Loop Deep Learning," IEEE INFOCOM 2021 - IEEE Conference on Computer Communications, Vancouver, BC, Canada, 2021, pp. 1-10, doi: 10.1109/INFOCOM42981.2021.9488764.

- [Uva+24] D. Uvaydov et al. "Stitching the Spectrum: Semantic Spectrum Segmentation with Wideband Signal Stitching," IEEE INFOCOM 2024 - IEEE Conference on Computer Communications, Vancouver, BC, Canada, 2024, pp. 2219-2228, doi: [10.1109/INFOCOM52122.2024.10621332](https://doi.org/10.1109/INFOCOM52122.2024.10621332).
- [Van+21] Nguyen Van Huynh et al. "Optimal Beam Association for High Mobility mmWave Vehicular Networks: Lightweight Parallel Reinforcement Learning Approach." In: IEEE Transactions on Communications 69.9 (2021), pp. 5948–5961. doi: [10.1109/TCOMM.2021.3088305](https://doi.org/10.1109/TCOMM.2021.3088305).
- [VBL19] Trinh Van Chien, Emil Bjornson, and Erik G. Larsson. "Sum Spectral Efficiency Maximization in Massive MIMO Systems: Benefits from Deep Learning." In: ICC 2019 - 2019 IEEE International Conference on Communications (ICC). 2019, pp. 1–6. doi: [10.1109/ICC.2019.8761234](https://doi.org/10.1109/ICC.2019.8761234).
- [Vie+17] Joao Vieira et al. "Deep convolutional neural networks for massive MIMO fingerprint-based positioning." In: 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC). 2017, pp. 1–6. doi: [10.1109/PIMRC.2017.8292280](https://doi.org/10.1109/PIMRC.2017.8292280).
- [Vil+23] Davide Villa et al. "AI-based Radar Sensing and Coexistence Using a Wireless Digital Twin." In: Proc. of ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization (WiNTECH). Madrid, Spain, Oct. 2023, pp. 1–6.
- [Vil+24] Davide Villa et al. "Colosseum as a Digital Twin: Bridging Real-World Experimentation and Wireless Network Emulation." In: IEEE Transactions on Mobile Computing, vol. 23, no. 10. 2024, pp. 9150–9166.
- [Vu+21] Thang X. Vu et al. "Machine Learning-Enabled Joint Antenna Selection and Precoding Design: From Offline Complexity to Online Performance." In: IEEE Transactions on Wireless Communications 20.6 (2021), pp. 3710–3722. doi: [10.1109/TWC.2021.3052973](https://doi.org/10.1109/TWC.2021.3052973).
- [Wan+17a] Ziyu Wang et al. "Sample Efficient Actor-Critic with Experience Replay." 2017. arXiv: [1611.01224\[cs.LG\]](https://arxiv.org/abs/1611.01224). url: <https://arxiv.org/abs/1611.01224>.
- [Wan+17b] Xuyu Wang et al. "CSI-Based Fingerprinting for Indoor Localization: A Deep Learning Approach." In: IEEE Transactions on Vehicular Technology 66.1 (2017), pp. 763–776. doi: [10.1109/TVT.2016.2545523](https://doi.org/10.1109/TVT.2016.2545523).
- [Wan+19] Rongrong Wang et al. "A Robust Wi-Fi Fingerprint Positioning Algorithm Using Stacked Denoising Autoencoder and Multi-Layer Perceptron." In: Remote Sensing 11.11 (2019). issn: 2072-4292. doi: [10.3390/rs11111293](https://doi.org/10.3390/rs11111293). url: <https://www.mdpi.com/20724292/11/11/1293>.

- [Wan+22a] C. -X. Wang et al., "Pervasive Wireless Channel Modeling Theory and Applications to 6G GBSMs for All Frequency Bands and All Scenarios," in IEEE Transactions on Vehicular Technology, vol. 71, no. 9, pp. 9159-9173, Sept. 2022, doi: 10.1109/TVT.2022.3179695.
- [Wan+22b] X. Wang et al. "Deep Reinforcement Learning Aided No-wait Flow Scheduling in Time-Sensitive Networks," 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 2022, pp. 812-817, doi: 10.1109/WCNC51071.2022.9771665.
- [Wan+22c] Jian Wang et al. "Integrated Sensing and Communication: Enabling Techniques, Applications, Tools and Data Sets, Standardization, and Future Directions." In: IEEE Internet of Things Journal 9.23 (2022), pp. 23416–23440. doi: [10.1109/IJOT.2022.3190845](https://doi.org/10.1109/IJOT.2022.3190845).
- [Wan+23] Yang Wang et al. "Transformer-Empowered 6G Intelligent Networks: From Massive MIMO Processing to Semantic Communication," in IEEE Wireless Communications, vol. 30, no. 6, pp. 127-135, December 2023, doi: 10.1109/MWC.008.2200157.
- [Wan+25] H. Wang et al., "Digital Twin Channel for 6G: Concepts, Architectures and Potential Applications," in IEEE Communications Magazine, vol. 63, no. 3, pp. 24-30, March 2025, doi: 10.1109/MCOM.001.2400213.
- [WBV21] Thorsten Wild, Volker Braun, and Harish Viswanathan. "Joint Design of Communication and Sensing for Beyond 5G and 6G Systems." In: IEEE Access 9 (2021), pp. 30845–30857. doi: [10.1109/ACCESS.2021.3059488](https://doi.org/10.1109/ACCESS.2021.3059488).
- [Wen+23] Zhenzi Weng et al. "Deep Learning Enabled Semantic Communications with Speech Recognition and Synthesis." In: IEEE Transactions on Wireless Communications 22.9 (2023), pp. 6227–6240. doi: [10.1109/TWC.2023.3240969](https://doi.org/10.1109/TWC.2023.3240969).
- [Wen+24] Haohuang Wen et al. "6G-XSec: Explainable Edge Security for Emerging OpenRAN Architectures." Proceedings of the 23rd ACM Workshop on Hot Topics in Networks. 2024.
- [WGB25] [Zhiyu Wang](#), [Mohammad Goudarzi](#), and [Rajkumar Buyya](#). "TF-DDRL: A Transformer-Enhanced Distributed DRL Technique for Scheduling IoT Applications in Edge and Cloud Computing Environments," in IEEE Transactions on Services Computing, vol. 18, no. 2, pp. 1039-1053, March-April 2025, doi: 10.1109/TSC.2025.3528346.
- [Wie+24] Reinhard Wiesmayr et al. "Design of a Standard-Compliant Real-Time Neural Receiver for 5G NR." 2024. arXiv: [2409.02912\[cs.IT\]](https://arxiv.org/abs/2409.02912). url: <https://arxiv.org/abs/2409.02912>.
- [WQ21] Zhenzi Weng and Zhijin Qin. "Semantic Communication Systems for Speech Transmission." In: IEEE Journal on Selected Areas in Communications 39.8 (2021), pp. 2434–2444. doi: [10.1109/JSAC.2021.3087240](https://doi.org/10.1109/JSAC.2021.3087240).

- [WSJ18] Chao-Kai Wen, Wan-Ting Shih, and Shi Jin. "Deep Learning for Massive MIMO CSI Feedback." In: IEEE Wireless Communications Letters 7.5 (2018), pp. 748–751. doi: [10.1109/LWC.2018.2818160](https://doi.org/10.1109/LWC.2018.2818160).
- [Wu+21] Chi Wu et al. "Learning to Localize: A 3D CNN Approach to User Positioning in Massive MIMO-OFDM Systems." In: IEEE Transactions on Wireless Communications 20.7 (2021), pp. 4556–4570. doi: [10.1109/TWC.2021.3060482](https://doi.org/10.1109/TWC.2021.3060482).
- [WWM20] Xuyu Wang, Xiangyu Wang, and Shiwen Mao. "Deep Convolutional Neural Networks for Indoor Localization with CSI Images." In: IEEE Transactions on Network Science and Engineering 7.1 (2020), pp. 316–327. doi: [10.1109/TNSE.2018.2871165](https://doi.org/10.1109/TNSE.2018.2871165).
- [Wym+25] Henk Wymeersch, et al. "6G Positioning and Sensing Through the Lens of Sustainability, Inclusiveness, and Trustworthiness", In: IEEE Wireless Communications, vol. 32, no. 1, pp. 68–75, February 2025. doi: [10.1109/MWC.011.2400055](https://doi.org/10.1109/MWC.011.2400055).
- [Xie+21] Huiqiang Xie et al. "Deep Learning Enabled Semantic Communication Systems." In: IEEE Transactions on Signal Processing 69 (2021), pp. 2663–2675. doi: [10.1109/TSP.2021.3071210](https://doi.org/10.1109/TSP.2021.3071210).
- [Xie+25] Huiqiang Xie et al. "Toward Intelligent Communications: Large Model Empowered Semantic Communications." In: IEEE Communications Magazine 63.1 (2025), pp. 69–75. doi: [10.1109/MCOM.001.2300807](https://doi.org/10.1109/MCOM.001.2300807).
- [Xu+20] C. Xu et al. "Buffer-aware Wireless Scheduling based on Deep Reinforcement Learning," 2020 IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea (South), 2020, pp. 1-6, doi: [10.1109/WCNC45663.2020.9120729](https://doi.org/10.1109/WCNC45663.2020.9120729).
- [Xu+24] Lifan Xu et al. "Reconfigurable Beamforming for Automotive Radar Sensing and Communication: A Deep Reinforcement Learning Approach." In: IEEE Journal of Selected Areas in Sensors 1 (2024), pp. 124–138. doi: [10.1109/JSAS.2024.3431462](https://doi.org/10.1109/JSAS.2024.3431462).
- [Xue+24] Qing Xue et al. "AI/ML for Beam Management in 5G-Advanced: A Standardization Perspective." In: IEEE Vehicular Technology Magazine 19.4 (2024), pp. 64–72. doi: [10.1109/MVT.2024.3431790](https://doi.org/10.1109/MVT.2024.3431790).
- [XYP21] Y. Xu, M. Yuan and M. -O. Pun, "Transformer Empowered CSI Feedback for Massive MIMO Systems," 2021 30th Wireless and Optical Communications Conference (WOCC), Taipei, Taiwan, 2021, pp. 157-161, doi: [10.1109/WOCC53213.2021.9602863](https://doi.org/10.1109/WOCC53213.2021.9602863).
- [XZS22] Lifan Xu, Ruxin Zheng, and Shunqiao Sun. "A Deep Reinforcement Learning Approach for Integrated Automotive Radar Sensing and Communication." In: 2022 IEEE 12th Sensor Array and Multichannel Signal Processing Workshop (SAM). 2022, pp. 316–320. doi: [10.1109/SAM53842.2022.9827815](https://doi.org/10.1109/SAM53842.2022.9827815).

- [Yan+19a] Yuwen Yang et al. “Deep Learning-Based Channel Estimation for Doubly Selective Fading Channels.” In: IEEE Access 7 (2019), pp. 36579–36589. doi: 10.1109/ACCESS.2019.2901066.
- [Yan+19b] Ping Yang et al. “Antenna selection for MIMO system based on pattern recognition.” In: Digital Communications and Networks 5.1 (2019). Artificial Intelligence for Future Wireless Communications and Networking, pp. 34–39. issn: 2352-8648. doi: <https://doi.org/10.1016/j.dcan.2018.10.001>. url: <https://www.sciencedirect.com/science/article/pii/S2352864818301512>.
- [Yan+21a] Bo Yang et al. “Intelligent Spectrum Learning for Wireless Networks with Reconfigurable Intelligent Surfaces.” In: IEEE Transactions on Vehicular Technology 70.4 (2021), pp. 3920–3925. doi: [10.1109/TVT.2021.3064042](https://doi.org/10.1109/TVT.2021.3064042).
- [Yan+21b] Bo Yang et al. “Spectrum-Learning-Aided Reconfigurable Intelligent Surfaces for “Green” 6G Networks.” In: IEEE Network 35.6 (2021), pp. 20–26. doi: [10.1109/MNET.110.2100301](https://doi.org/10.1109/MNET.110.2100301).
- [Yan+24] Mengke Yang et al. “Joint Admission and Power Control for Massive Connections via Graph Neural Network.” In: IEEE Transactions on Vehicular Technology 73.8 (2024), pp. 11806–11820. doi: [10.1109/TVT.2024.3371019](https://doi.org/10.1109/TVT.2024.3371019).
- [Yeh+23] Shu-ping Yeh et al. “Deep Learning for Intelligent and Automated Network Slicing in 5G Open RAN (O-RAN) Deployment.” In: IEEE Open Journal of the Communications Society (2023).
- [YG24] Vijaya Yajnanarayana and Philipp Geuer “Bi-Static Sensing in OFDM Wireless Systems for Indoor Scenarios”, In: 3rd International Conference on 6G Networking (6GNet), Paris, France, 2024, pp. 227-233. doi: 10.1109/6GNet63182.2024.10765701.
- [YGD25] Vijaya Yajnanarayana, Philipp Geuer, and Satyam Dwivedi. “Indoor Sensing using Measurements.” 2024. arXiv:2409.00634 [cs.IT]. url: <https://arxiv.org/abs/2409.00634>.
- [YNM19] Jide Yuan, Hien Quoc Ngo, and Michail Matthaiou. “Machine Learning-Based Channel Estimation in Massive MIMO with Channel Aging.” In: 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). 2019, pp. 1–5. doi: [10.1109/SPAWC.2019.8815557](https://doi.org/10.1109/SPAWC.2019.8815557).
- [Yoo+25] Daejung Yoon et al. “From AI Concept to Real World: AI-powered Modem for Next Samsung Cellular Networks”, 2025, In: SR Blog [Online]. Available: <https://research.samsung.com/blog/From-AI-Concept-to-Real-World-AI-powered-Modem-for-Next-Samsung-Cellular-Networks>
- [YSS24] Noe M. Yungaicela-Naula, Vishal Sharma, and Sandra Scott-Hayward. “Misconfiguration in O-RAN: Analysis of the impact of AI/ML.” Computer Networks (2024): 110455.

- [Yu+22] Wentao Yu et al. “Hybrid Far- and Near-Field Channel Estimation for THz Ultra-Massive MIMO via Fixed Point Networks.” In: GLOBECOM 2022 - 2022 IEEE Global Communications Conference. 2022, pp. 5384–5389. doi: 10.1109/GLOBECOM48099.2022.10001564.
- [Yu+25] Wentao Yu et al. “Deep Learning for Near-Field XL-MIMO Transceiver Design: Principles and Techniques.” In: IEEE Communications Magazine 63.1 (2025), pp. 52–58. doi: 10.1109/MCOM.001.2300829.
- [YW23] Vijaya Yajnanarayana and Henk Wymeersch, “Multistatic Sensing of Passive Targets Using 6G Cellular Infrastructure”, In: EuCNC/6G Summit, Gothenburg, Sweden, 2023, pp. 132-137. doi: 10.1109/EuCNC/6GSummit58263.2023.10188243.
- [YWW21] Wentao Yu, Tianyu Wang, and Shaowei Wang. “Multi-Label Learning Based Antenna Selection in Massive MIMO Systems.” In: IEEE Transactions on Vehicular Technology 70.7 (2021), pp. 7255–7260. doi: [10.1109/TVT.2021.3087132](https://doi.org/10.1109/TVT.2021.3087132).
- [YZ20] Junhong Ye, and Ying-Jun Angela Zhang. “DRAG: Deep reinforcement learning based base station activation in heterogeneous networks.” In: IEEE Transactions on Mobile Computing 19.9 (2019), pp. 2076–2087.
- [Zah+23] Mahmoud Zaher et al. “Learning-Based Downlink Power Allocation in Cell-Free Massive MIMO Systems.” In: IEEE Transactions on Wireless Communications 22.1 (2023), pp. 174–188. doi: [10.1109/TWC.2022.3192203](https://doi.org/10.1109/TWC.2022.3192203).
- [Zai+23] A. Zaidi et al. “A framework to address mobility management challenges in emerging networks.” IEEE Wireless Commun., vol. 30, no. 4, pp. 90-97, Aug. 2023.
- [ZGZ22] Maojun Zhang, Jiabao Gao, and Caijun Zhong. “A Deep Learning-Based Framework for Low Complexity Multiuser MIMO Precoding Design.” In: IEEE Transactions on Wireless Communications 21.12 (2022), pp. 11193–11206. doi: [10.1109/TWC.2022.3190435](https://doi.org/10.1109/TWC.2022.3190435).
- [Zha+19] Nan Zhao et al. “Deep Reinforcement Learning for User Association and Resource Allocation in Heterogeneous Cellular Networks.” In: IEEE Transactions on Wireless Communications 18.11 (2019), pp. 5141–5152. doi: [10.1109/TWC.2019.2933417](https://doi.org/10.1109/TWC.2019.2933417).
- [Zha+20] Haijun Zhang et al. “Deep Learning Based Radio Resource Management in NOMA Networks: User Association, Subchannel and Power Allocation.” In: IEEE Transactions on Network Science and Engineering 7.4 (2020), pp. 2406–2415. doi: [10.1109/TNSE.2020.3004333](https://doi.org/10.1109/TNSE.2020.3004333).
- [Zha+21] Shunbo Zhang et al. “Deep Learning Optimized Sparse Antenna Activation for Reconfigurable Intelligent Surface Assisted Communication.” In: IEEE Transactions on Communications 69.10 (2021), pp. 6691–6705. doi: 10.1109/TCOMM.2021.3097726.
- [Zha+22a] Zhongfeng Zhang et al. “TIPS: Transformer-Based Indoor Positioning System Using Both CSI and DoA of Wi-Fi Signal.” In: IEEE Access 10 (2022), pp. 111363–111376. doi: [10.1109/ACCESS.2022.3215504](https://doi.org/10.1109/ACCESS.2022.3215504).

- [Zha+22b] Bowen Zhang et al. Semantic Sensing and Communications for Ultimate Extended Reality. 2022. arXiv: [2212.08533\[eeess.SP\]](https://arxiv.org/abs/2212.08533). url: <https://arxiv.org/abs/2212.08533>.
- [Zha+23a] Xiangyu Zhang et al. "Near-Field Channel Estimation for Extremely Large-Scale Array Communications: A Model-Based Deep Learning Approach." In: IEEE Communications Letters 27.4 (2023), pp. 1155–1159. doi: 10.1109/LCOMM.2023.3245084.
- [Zha+23b] Zhibo Zhang et al. "Deep-learning methods for integrated sensing and communication in vehicular networks." In: Vehicular Communications 40 (2023), p. 100574. issn: 22142096. doi: <https://doi.org/10.1016/j.vehcom.2023.100574>. url: <https://www.sciencedirect.com/science/article/pii/S2214209623000049>.
- [Zha+24] Yunwu Zhang et al. "Transformer-Based Predictive Beamforming for Integrated Sensing and Communication in Vehicular Networks," in IEEE Internet of Things Journal, vol. 11, no. 11, pp. 20690-20705, 1 June1, 2024, doi: 10.1109/IJOT.2024.3372060.
- [Zha+25] Jie Zhang et al. "Decision Transformers for Wireless Communications: A New Paradigm of Resource Management," in IEEE Wireless Communications, vol. 32, no. 2, pp. 180-186, April 2025, doi: 10.1109/MWC.007.2400124.
- [Zhe+21] Jianzhang Zheng et al. "Channel Assignment for Hybrid NOMA Systems With Deep Reinforcement Learning." In: IEEE Wireless Communications Letters 10.7 (2021), pp. 1370– 1374. doi: [10.1109/LWC.2021.3058922](https://doi.org/10.1109/LWC.2021.3058922).
- [ZHL24] Hao Zhou, Chengming Hu, and Xue Liu. "An Overview of Machine Learning-Enabled Optimization for Reconfigurable Intelligent Surfaces-Aided 6G Networks: From Reinforcement Learning to Large Language Models." 2024. arXiv: 2405.17439 [cs.NI]. url: <https://arxiv.org/abs/2405.17439>.
- [Zho+19a] Pei Zhou et al. "Deep Learning-Based Beam Management and Interference Coordination in Dense mmWave Networks." In: IEEE Transactions on Vehicular Technology 68.1 (2019), pp. 592–603. doi: [10.1109/TVT.2018.2882635](https://doi.org/10.1109/TVT.2018.2882635).
- [Zho+19b] Fuhui Zhou et al. "Dynamic Spectrum Management via Machine Learning: State of the Art, Taxonomy, Challenges, and Open Research Issues." In: IEEE Network 33.4 (2019), pp. 54–62. doi: [10.1109/MNET.2019.1800439](https://doi.org/10.1109/MNET.2019.1800439).
- [Zho+22] Qingyang Zhou et al. "Semantic Communication with Adaptive Universal Transformer." In: IEEE Wireless Communications Letters 11.3 (2022), pp. 453–457. doi: [10.1109/LWC.2021.3132067](https://doi.org/10.1109/LWC.2021.3132067).
- [Zho+23] Shuang Zhou et al. "Wi-Monitor: Daily Activity Monitoring Using Commodity Wi-Fi." In: IEEE Internet of Things Journal 10.2 (2023), pp. 1588–1604. doi: [10.1109/IJOT.2022.3210378](https://doi.org/10.1109/IJOT.2022.3210378).

- [ZND20] Y. Zhao, I. G. Niemegeers, and S. H. De Groot. “Power Allocation in Cell-Free Massive MIMO: A Deep Learning Method.” In: IEEE Access 8 (2020), pp. 87185–87200. doi: [10.1109/ACCESS.2020.2992629](https://doi.org/10.1109/ACCESS.2020.2992629).
- [ZND21] Yu Zhao, Ignas G. Niemegeers, and Sonia M. Heemstra De Groot. “Dynamic Power Allocation for Cell-Free Massive MIMO: Deep Reinforcement Learning Methods.” In: IEEE Access 9 (2021), pp. 102953–102965. doi: [10.1109/ACCESS.2021.3097243](https://doi.org/10.1109/ACCESS.2021.3097243).
- [Zol+25] Arshia Zolghadr et al. “Learning and reconstructing conflicts in o-ran: A graph neural network approach.” 2025 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2025.
- [ZQL23] Bowen Zhang, Zhijin Qin, and Geoffrey Ye Li. “Semantic-Aware Image Compressed Sensing.” In: 2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP). 2023, pp. 1–6. doi: [10.1109/MLSP55844.2023.10285864](https://doi.org/10.1109/MLSP55844.2023.10285864).
- [ZQL24] Bowen Zhang, Zhijin Qin, and Geoffrey Ye Li. “Compression Ratio Learning and Semantic Communications for Video Imaging.” In: IEEE Journal of Selected Topics in Signal Processing 18.3 (2024), pp. 312–324. doi: [10.1109/JSTSP.2024.3405853](https://doi.org/10.1109/JSTSP.2024.3405853).
- [Zuo+22] Jun Zuo et al. “Artificial Intelligence-Based Spatial Domain Beam Prediction for 5G Beyond.” In: 2022 IEEE Globecom Workshops (GC Wkshps). 2022, pp. 1460–1465. doi: [10.1109/GCWkshps56602.2022.10008651](https://doi.org/10.1109/GCWkshps56602.2022.10008651).