IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.20 25.DOI

AI/ML-Driven 6G Network Solutions with Energy Efficiency Considerations

ANDREAS ROESSLER[1], MIGUEL M. LOPEZ[10], ROUAA DIAB[2], RENATO L. G. CAVALCANTE[2], ALI ABDI[4], CHRISTIAN ARENDT[5], CARSTEN BOCKELMANN[6], ALIREZA DANAEE[6], MARCO DANGER[5], ARMIN DEKORSY[6], ALI EL HUSSEINI[10], GERHARD FETTWEIS[8], JOCHEN FINK[2], RICK FRITSCHEK[9], PRAMESH GAUTAM[6], MUHAMMAD QURRATULAIN KHAN[8], SACHINKUMAR B. MALLIKARJUN[3], AHMAD NIMR[8], MOHAMMAD PARVINI[8], ERMA PERENDA[4], ROBERT-JERON REIFERT[7], RAFAEL F. SCHAEFER[9], HANS D. SCHOTTEN[3], PHILIPP SCHULZ[8], AYDIN SEZGIN[7], SLAWOMIR STAŃCZAK[2,11], CHRISTIAN WIETFELD[5], DIRK WÜBBEN[6]

[1]Rohde&Schwarz GmbH and Co. KG, Muehldorfstrasse 15, 81671 Munich, Germany
[2]Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut (HHI), Einsteinufer 37, 10587 Berlin, Germany
[3]University of Kaiserslautern (RPTU), 67663, Kaiserslautern, Germany
[4]Chair for Distributed Signal Processing, RWTH Aachen University, 52056 Aachen, Germany
[5]Communication Networks Institute (CNI), TU Dortmund University, 44227 Dortmund, Germany
[6]Department of Communications Engineering, University of Bremen, 28359 Bremen, Germany
[7]Digital Communication Systems (DCS), Ruhr University Bochum, 44801 Bochum, Germany
[8]Vodafone Chair for Mobile Communications Systems, Technische Universität Dresden, 01069 Dresden, Germany
[9]Chair of Information Theory and Machine Learning, Technische Universität Dresden, 01069 Dresden, Germany
[10]Ericsson Research, Ericsson GmbH, 52134 Herzogenrath, Germany
[11]Technical University of Berlin, Berlin, Germany

Corresponding author: Andreas Roessler (e-mail: andreas@backitapp.com)

This work was funded and supported by Federal Ministry of Education and Research Germany (BMBF) as part of the 6G-ANNA project.

ABSTRACT

This paper highlights key areas in which AI/ML can play a transformative role in 6G, with an emphasis on energy-efficient solutions. Contextualized within Germany's national 6G research initiative, "6G Access, Network of Networks, Automation and Simplification" (6G-ANNA) emerges as the lighthouse project, providing a holistic vision that sets the direction for numerous specialized research efforts. Within this context, this paper aims to present ideas relevant to both academia and industry by identifying key research directions for AI/ML. We begin by reviewing the latest developments in using AI/ML for the Fifth Generation (5G) New Radio (NR) air interface as discussed in 3rd Generation Partnership Project (3GPP) Releases 18 and 19, and examine how these advancements pave the way for a native and energy-efficient AI/ML air interface in 6G. Key results are presented on AI/ML-driven optimization of radio frequency (RF) frontends, along with a strong focus on the role of AI/ML in diverse signal processing tasks and energy saving mechanisms, which demonstrate the potential of AI/ML in improving spectral efficiency and reducing energy consumption. The discussion further introduces methodologies for testing AI/ML-based signal processing tailored for the 6G physical layer, addressing practical challenges relevant to industry stakeholders and standard development organizations. Finally, we discuss the standardization aspects critical for realizing a future AI-native air interface in 6G, aligning our findings with ongoing and upcoming global standardization activities.

INDEX TERMS Artificial intelligence (AI), energy efficiency, machine learning (ML), physical layer (PHY), signal processing, Sixth Generation (6G), standardization, test and measurement

I. INTRODUCTION

The advent of Sixth Generation (6G) wireless communication systems marks a transformative era in connectivity,

driven by unprecedented demands for scalability, reliability, and energy efficiency. As part of the BMBF-funded research project, "6G Access, Network of Networks, Automation and Simplification" (6G-ANNA), this paper explores the role of AI/ML in future wireless networks, with an emphasis on energy-efficient approaches.

The transition from the Fifth Generation (5G) to 6G requires redefining traditional network paradigms, moving toward systems that are not only faster but also inherently capable of leveraging artificial intelligence (AI)/machine learning (ML) for adaptive, real-time decision-making [1], [2]. These capabilities are integral for addressing the growing complexity of communication environments, characterized by heterogeneous user requirements, increased device density, and the need for consistently low latency. While incorporating AI/ML natively within the 6G air interface, it remains a crucial challenge to find a harmonious balance between spectral efficiency, computational demands, and energy consumption.

Many efforts have been made in both academia and industry toward an AI-native 6G design, with an emphasis on energy efficiency. To this end, this paper presents some of the major topics of interest for 6G and aims to present the research directions most relevant for academia and industry. Another aim of this paper is to highlight the necessity to enhance energy efficiency while sustaining high network performance, particularly in light of increasing environmental concerns and the need for sustainable solutions. Current advancements in AI/ML-driven energy optimization, including intelligent power control and adaptive resource allocation, underscore the transformative potential of these technologies. However, several critical challenges remain.

First, there is a need to understand and evaluate the usefulness of AI/ML for radio frequency (RF) frontend optimization, which can significantly impact overall energy consumption and system performance. Second, assessing the applicability and potential performance gains of AI/ML techniques in various signal processing tasks is essential for realizing their full potential in practical deployments. And third, understanding how to test AI/ML-based physical layer (PHY) functionalities is crucial, as these testing methodologies will directly influence the design and requirements of future 6G standards. The collaborative efforts within the 6G-ANNA project aim to address these challenges by establishing a comprehensive framework. This framework will leverage AI/ML to improve signal processing, channel estimation, and resource management while ensuring that energy efficiency remains a core principle of 6G design. Ultimately, these efforts will lay the groundwork for a truly AI-native air interface, shaping future standardization activities and industry practices.

This paper is organized as follows. We continue this section by first introducing what the term AI-native 6G implies, and highlight some of the differences with the current 5G standard. Following this, we discuss energy efficiency and review the current status of 3GPP standardization efforts. Continuing our discussion on standardization, we summarize the current use of AI/ML in 3GPP in Section II. The subsequent sections outline the state of the art and summarize key insights related to different aspects of AI/ML in 6G. In Section III, we present techniques addressing non-linearities in power amplifiers, which are deemed crucial for energy efficiency. Following this, in Section IV, we present the benefits of AI/ML in channel and interference estimation and prediction and in Section V we discuss channel coding. In Sections VI and VII, we present AI/ML techniques for multi-user MIMO and network/subnetwork optimization, respectively. Section VIII covers testing methodologies for AI-based signal processing tasks. In Section IX, future standardization aspects are discussed, and finally Section X concludes our discussion and findings and provides an outlook for future topics of interest.

A. WHAT DOES THE TERM AI-NATIVE AIR INTERFACE FOR 6G IMPLY?

The term "AI-native" implies that AI is a fundamental part of the design and operation of the air interface, rather than just an added enhancement. Unlike 5G, where AI and ML are applied in specific use cases like network optimization, predictive maintenance, and improved user experience, an AI-native approach for 6G integrates AI as an integral part of the network. This means network protocols, signal processing, and system optimization in 6G will inherently use AI capabilities. The shift is from using AI as a performance booster to using AI as a key technology component within the future standard. A good example of this is the channel state information (CSI)-feedback enhancement use case. Currently, it is based on implicit feedback to limit the impact and changes to the existing methodology. However, with a new standard like 6G, explicit feedback could become the foundation for channel feedback. When a system, like 6G, is designed from scratch, it is easier to integrate such changes than with an existing standard, like 5G New Radio (NR).

An AI-native air interface brings several important implications. Firstly, the seamless transfer of AI models between network entities, e.g., base station (BS) and user equipment (UE), or across different layers of the network architecture becomes essential. This allows for dynamic updates and optimizations, ensuring that AI models match current operational conditions and requirements. Secondly, managing the lifecycle of AI models, from development and deployment to maintenance and retirement, becomes critical. This includes version control, performance monitoring, and continuous learning mechanisms to adapt to new data and scenarios, ensuring the models remain effective over time. Lastly, testing aspects present unique challenges, as more than traditional deterministic testing methods may be needed. New strategies that account for the probabilistic nature of AI decisions, the variability of training data, and the potential for model drift over time are needed. This requires a more flexible, adaptive approach to testing and validation, possibly incorporating real-world data in addition to thorough simulations to ensure robustness and reliability.

B. WHAT DOES THE TERM ENERGY EFFICIENCY MEAN IN THE CONTEXT OF AI/ML?

Sustainability and energy efficiency are crucial for future generation systems, yet they often conflict with the current trends of wireless technology [3]. For instance, enhancing the throughput of communication systems has led many research groups to advocate for denser radio networks, utilizing transceivers at high frequencies using advanced signal processing algorithms. While these improvements may boost the transmit energy efficiency, the overall energy expenditure — including hardware and supporting systems such as cooling — often complicates the situation. Typically, the energy radiated by hardware is only a small portion of the total energy budget in wireless systems. This issue is exacerbated by the growing trend of employing power-hungry graphics processing units (GPUs) for ML algorithms in these systems.

Given this context, it is essential to shift the emphasis from merely increasing the transmit energy efficiency to reducing the total hardware energy consumption. One approach to achieve this goal is to tailor the transmission schemes to match user demand, thereby avoiding overprovisioning and the associated energy costs. Furthermore, the introduction of distributed multiple input multiple output (MIMO) systems allows for the simplification of hardware at radio access points (RAPs) while enhancing interference management capabilities and reducing the overall energy consumption by powering down unnecessary network elements. This idea can also be exploited by protocols in conventional systems, where radios should maximize their time spent in sleep modes. However, the resulting optimization problems are typically combinatorial in nature, so provably optimal and fast algorithms for all scenarios are unlikely to exist. ML tools have successfully addressed similar issues in various fields, suggesting their potential also in the wireless domain. For example, in branches of AI research, ML algorithms have been extensively used to determine good positions to place sensors in a network, and similar techniques can be adopted to deactivate RAPs while guaranteeing a necessary level of quality of service (QoS) in the network.

Considering the hardware asymmetry between terminals and RAPs/BSs — where the latter are often equipped with powerful hardware and reliable energy sources — there is also a significant potential for applying learning tools to simplify the terminal hardware (and hence decrease the energy consumption) while mitigating any imperfections introduced by this simplification at the RAPs with learning algorithms. However, challenges exist due to the highly nonlinear distortions caused by simple hardware components, such as simple power amplifiers (PAs). Moreover, the ML algorithms must be sufficiently lightweight to prevent excessive energy consumption at the RAPs.

C. TIMELINE FOR AN AI-NATIVE INTERFACE IN 6G

The development timeline for an AI-native air interface is closely aligned with the definition of a 6G standard and is thus tightly linked to 3rd Generation Partnership Project (3GPP)'s progression. At the time of writing, 3GPP focuses on completing Release 19 by late 2025, with a core specifications freeze completed in September and an Abstract Syntax Notation (ASN).1 syntax freeze planned by the end of December. The first steps toward 6G started with the inaugural 6G workshop on March 10–11, 2025, in Incheon, South Korea. This workshop brought together Technical Specification Group for Service and System Aspects (TSG SA), Technical Specification Group for Core Network and Terminals (TSG CT) and Technical Specification Group for Radio Access Network (TSG RAN) to discuss initial requirements and use cases. Following this, Release 20 will initiate exploratory studies to establish the groundwork for normative 6G standardization starting with 3GPP Release 21. The timeline for Release 21 is expected to be finalized by June 2026, focusing on the preparation of the first formal 6G technical specifications in alignment with ITU-R's IMT-2030 submission requirements [4]. This work will culminate in an ASN.1 freeze by March 2029 (Figure 1). Through this phased approach, 3GPP carefully aligns its milestones with global expectations, paving the way for a comprehensive 6G standard before 2030.

II. THE CURRENT USAGE OF AI/ML IN 3GPP

A. ROLE OF AI/ML IN CURRENT 5G NR NETWORKS

AI and ML have been integrated into the 5G NR standards defined by the 3GPP, which significantly improved network management and operation. At first, AI/ML was mainly used in core network operations, starting from Release 15. This involved gathering lots of data to help manage network functions and simplify operations using advanced data analytics. While the 3GPP standard continues to adopt AI/ML methods for network orchestration and management, their implementation is usually up to the network operators.

An essential first step was the addition of the Network Data Analytics Function (NWDAF) with 3GPP Release 15. The NWDAF is a key part of the 5G core network, offering analysis of network slices that greatly improves operational intelligence across network functions. The NWDAF was further developed in 3GPP Release 17, where it gained two new logical entities: The Analytics logical function (AnLF) and the Model Training Logical Function (MTLF). The MTLF is responsible for creating and improving ML models and supporting the launch of new training services that enhance the network's adaptability. Meanwhile, the AnLF focuses on conducting detailed interference analysis and data analytics, improving network reliability and performance.

Release 17 marked the beginning of a Radio Access Network (RAN)-focused study led by RAN3. The study aimed to outline the functional framework for AI-enabled RAN, exploring several use cases, such as network energy savings, load balancing, and mobility optimization. These use cases demonstrate the potential of AI/ML in improving various aspects of network operations. Through these efforts, 3GPP has paved the way for a smarter, more efficient, and adaptable 5G ecosystem. These advancements indicate a move towards

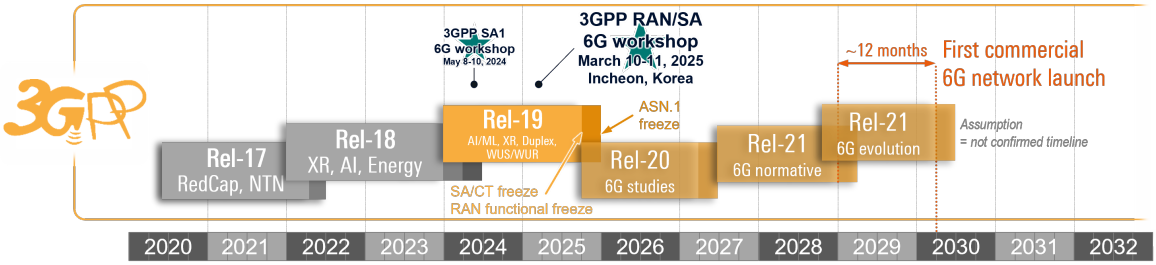


FIGURE 1. 3GPP timeline for 5G-Advanced and 6G

more autonomous and AI-driven network operations, promising major enhancements in operational efficiency and user experience.

B. AI/ML FOR THE 5G NR AIR INTERFACE IN 3GPP RELEASE 18/19

3GPP Release 18 marks the beginning of 5G-Advanced, which includes many new AI/ML-related study items led by various technical specification groups within 3GPP. The System and Service Aspects Working Group 1 (SA1) explores methods for sharing AI/ML models across the 5G network, paving the way for a more interconnected and intelligent network fabric. At the same time, the SA2 group looks at architectural support for AI/ML-based services within the 5G system, laying the groundwork for improved service delivery and operational efficiency. The SA4 group, which focuses on media services, examines potential uses of AI/ML for 5G media codecs, offering significant improvements in media quality and delivery. Meanwhile, SA5 tackles the important issue of AI/ML management, aiming to harmonize AI/ML functions across 5G systems for easier management, orchestration, and charging.

Among these important study items, Technical Report 38.843 on AI/ML for the NR air interface stands out. Many experts from academic institutions and industry believe that 6G will inherently support AI/ML for the air interface. As such, an initial study from a 5G perspective to define a general framework for enhancing air interface functionality to support AI/ML-based algorithms is underway. The goal is to improve performance while reducing complexity and overhead.

To achieve this, 3GPP has agreed to study three pilot use cases. The first scenario, channel state information reference signals (CSI-RS) feedback enhancement, aims to improve the feedback mechanism by applying AI/ML techniques to compress the CSI using an autoencoder, a specific neural network architecture designed to reduce redundancy and thus enable compression. This mechanism is then combined with AI/ML-based prediction of the mobile radio channel's behavior, aiming to increase the interval for sending channel quality reports in the uplink and, therefore, reducing overhead.

The second scenario explores the use of AI/ML to reduce beamforming management overhead and improve beam selection accuracy through spatial and temporal prediction. This could significantly benefit energy consumption and battery life at the UE side.

The third scenario aims to improve positioning estimation accuracy using AI/ML-based techniques, especially in heavy non-line-of-sight (NLOS) conditions like factory halls. Two aspects are investigated. First, a trained AI/ML model is used to directly infer the UEs position, known as 'fingerprinting.' The second aspect is AI/ML-assisted positioning, where the AI/ML model inference output enhances an existing measurement, such as providing an NLOS indicator.

Where the latter two scenarios are a so-called one-sided model use cases, the first scenario is a two-sided model use case. One-sided ML models refer to architectures where the functionality is deployed exclusively on one end of the communication link—either at the transmitter or the receiver. Examples are ML models that are tasked with signal classification or modulation recognition. One-sided models simplify deployment by eliminating the need for coordination between endpoints, making them suitable for scenarios with limited feedback or asymmetric processing capabilities. Two-sided models distribute the ML architecture across the transmitter and receiver—often between a mobile device and a BS. The training of the individual components can be done jointly or individually, each presenting its unique challenges. Joint training offers optimal performance by aligning model parameters across both ends but requires extensive coordination, high data exchange volumes, and unified infrastructure. In contrast, individual training allows each vendor to train their portion independently, reducing complexity but potentially introducing mismatches in learned representations. Ensuring convergence and compatibility when models are trained separately remains a significant challenge, as each side may optimize for different objectives or environments. The answers to these questions are currently controversially discussed among 3GPP standardization experts across the industry.

1) CSI-feedback enhancement for compression and prediction

From all three pilot use cases, the CSI feedback enhancement is most intriguing due to its two-sided aspects, requiring a close interplay between the network and device, as the

encoder portion of the autoencoder resides in the UE and the decoder part in the base station (gNB).

Many industry players believe these AI/ML-based enhancements offer a significant opportunity to stand out from competitors. As a result, many vendors prefer to keep their ML models private. The model's training, inference, and lifecycle management are crucial components that our industry, particularly 3GPP, must ensure compatibility between, even as development and training occur separately. Test and measurement solution providers will play a key role in this process, serving as intermediaries between network equipment vendors, chipset design, and handset developers from a conformance and performance test perspective.

Massive MIMO involves using many antenna elements at the BS to serve multiple users simultaneously using the same physical resources in the time-frequency grid. This technology greatly improves system performance in terms of spectral efficiency and network capacity, but it also presents challenges related to complex signal processing, the need for accurate CSI, and the significant increase in feedback overhead for CSI reporting. Currently, 3GPP has defined compressive sensing- and codebook-based feedback algorithms, the latter being the dominant scenario deployed in today's 5G networks. The UE usually performs signal quality measurements on the downlink signal using the embedded CSI-RS, where it computes several parameters, such as channel quality indicator (CQI), the precoding matrix indicator (PMI), and the rank indicator (RI). These parameters are sent back to the network as part of a measurement report, where the gNB uses the information to select an appropriate precoding matrix from a defined set of codebooks based on the received feedback to maximize signal quality and network performance. This process has been further optimized over the different standards (Fourth Generation (4G) Long Term Evolution (LTE), 5G NR) and actual 3GPP Releases 15 and 16 to improve accuracy and reduce the subsequent overhead due to the high dimension of CSI in such systems.

With the introduction of Release 18 and continuing with Release 19, 3GPP is researching the use of AI to enhance CSI-based feedback, while studying explicit and implicit feedback mechanisms. In the case of the explicit mechanism, the complete channel matrix is sent back to the BS. The implicit mechanism, on the other hand, sends a more compact version of the channel matrix back to the network. This version is in the form of eigenvectors. Compared to the codebook from Release 15/16, the only change in implicit feedback is in the PMI definition. Instead of using the traditional codebook, AI-compressed precoding vectors are transmitted. Most vendors prioritize implicit feedback when establishing their AI framework to enable backward compatibility with the legacy codebook-based approach.

Based on the network configuration, the UE builds the CSI-matrix using the signaled CSI-RS configuration. The size of this matrix depends on the time-frequency grid's configuration (the number of subcarriers, the number of orthogonal frequency division multiplexing (OFDM) symbols), the number of antenna elements for both, transmit and receive, and the number of processed time slots. Vendors might preprocess this data for normalization and noise reduction before it serves as input for the encoder stage of an autoencoder. Inspired by image compression, this type of neural network is used at the UE to compress and quantize the precoding matrix. The resulting bitstream matches the PMI of the original feedback mechanism based on the codebook. This bitstream is sent via the uplink to the BS, where the decoder part of the jointly trained autoencoder reconstructs the original precoding matrix. This AI-based feedback enhancement is a way to reduce reporting overhead with only marginal loss in accuracy compared to conventional methods.

To compare the performance of Release 16 Type II codebook-based feedback with AI-enabled feedback, 3GPP primarily uses (squared generalized cosine similarity (SGCS)), where some vendors also used the (normalized mean squared error (NMSE)) to measure reconstruction accuracy. Most 3GPP contributions focus on implicit feedback, with observed overhead reductions ranging from 5 to 25%. However, these numbers are still a topic of debate among delegates. The scientific literature often investigates explicit feedback compression, which can result in more significant overhead savings due to the higher amount of information to be compressed.

So, how does test and measurement fit into all of this? In a typical deployment, the gNB, the UE, and the modem powering the mobile device, all come from different manufacturers. Generally, it is necessary for these parties to work together, and the encoder and decoder should be trained together to achieve similar performance, which might be a direction the industry prefers to go.

This gap between different vendors could be bridged by employing test and measurement equipment, such as mobile radio tester platforms [e.g., R&S CMX500 5G One-Box Signaling Tester (CMX500)], that can simulate different 5G network configurations, including CSI-RS, and emulate various channel conditions to test the performance of various autoencoder models trained for, e.g., specific locations (site-specific) or channel characteristics. First results are available, which are promising [5]. However, the industry still has a long way to go to ensure full interoperability among vendors when deploying two-sided use cases, not just from a performance standpoint, but also from a model training and model transfer between gNB and UE point of view.

2) AI/ML-based Beam Management

Beamforming via large antenna arrays is indispensable to compensate for the severe pathloss caused by millimeter-wave (mmWave) frequencies [6]. However, to harvest the highest beamforming gain, both the BS and UE should collaborate in beam selection to identify the best beam pair that aligns well with the dominant path of the underlying channel. In 5G communication systems, this is achieved via a beam management (BM) procedure, where beamformed reference signals (RSs), known as synchronization sequence blocks

(SSBs), are periodically transmitted from the BS to the UE. To ensure seamless connectivity, the 3GPP BM procedure has been defined since Release 14 [7]. However, the exhaustive nature of this BM procedure leads to a high RS overhead, latency, and high RF front-end power consumption [8].

To overcome the aforementioned limitations, investigation of an AI/ML based BM procedure is an ongoing study inside the 3GPP. With focus on RS overhead, latency, and power consumption reduction, 3GPP has defined two use cases for beam prediction, i.e., spatial domain and temporal domain beam prediction [9]. The first use case of spatial domain beam prediction aims at exploiting the environment-specific spatial correlation, while the temporal domain beam prediction exploits the correlation arising due to UE mobility patterns. Consequently, an ML model is trained to learn these correlations for efficient beam prediction. Initial investigations have reported more than 90% reduction in overhead with good prediction accuracy over specific environmental conditions [9].

Generalization of an ML model examines how well a model can digest new data and make correct predictions after getting trained on a training set. Since the best beam is mainly dependent on the environmental conditions, ML model generalization is crucial for beam prediction [10]. Considering its importance, the 3GPP study on AI/ML investigates generalization over different beamforming architectures, antenna array dimensions, environmental conditions, UE mobility, and rotation patterns. Initial investigations on model generalization have reported significant performance degradation in terms of prediction accuracy [9]. Alternatively, it has been proposed to train a model on a mixed dataset of various scenarios. However, in such cases, there exists a tradeoff between ML model accuracy performance and its generalization capabilities [11].

To avoid the poor ML generalization, several 3GPP contributors have suggested training scenario-specific models and then to rely on model switching based on the applicable scenario. However, this not only leads to a complex model life cycle management procedure but also requires perfect identification of the applicable scenario. Furthermore, storing, training, switching, and retraining several scenario-specific models may lead to significant memory and power consumption costs, which violates one of the fundamental reasons to use AI/ML for BM.

AI/ML based BM opens several new challenges for test and measurement. As an example, the ML model and its training dataset are vendor and/or device specific, which leads to the challenge of model recreation and validation. Consequently, it is necessary for the test and measurement companies to offer test and validation platforms that can mimic several 3GPP specified channel conditions and network configurations.

3) AI-based position accuracy enhancement

This third pilot use case aims to improve positioning estimation accuracy using AI/ML-based techniques, especially in heavy NLOS conditions, such as in industrial environments, for example, a factory hall [12]. Two aspects were investigated. First, a trained AI/ML-model is used to directly infer the UE's position, known as 'fingerprinting.' The second aspect is AI/ML-assisted positioning, where the AI/ML-model inference output enhances an existing measurement, such as providing an NLOS indicator.

III. ADDRESSING NONLINEAR CHALLENGES THROUGH AI AND MACHINE LEARNING

A. INTRODUCTION AND MOTIVATION

The integration of ML techniques into the physical layer of wireless communication systems represents a pivotal step in addressing the challenges of next-generation 6G networks. On the transmitter side, ML-driven digital pre-distortion (DPD) enhances PA linearity, while ML-based digital post-distortion (DPOD) on the receiver mitigates higher distortion levels, enabling higher transmitter output power. Additionally, ML-powered peak-to-average power ratio (PAPR) reduction techniques are crucial for achieving high power efficiency in OFDM systems without compromising signal fidelity. These advancements, explored in subsequent sections, highlight the transformative role of AI/ML in optimizing spectral efficiency while maintaining energy efficiency.

B. MACHINE LEARNING FOR DPD

PA non-linearities manifest as spectral regrowth and intermodulation distortions (as seen in Fig.2), which compromise signal integrity and energy efficiency, especially at high power levels. This leads to efficiency roll-off and adjacent channel interference which complicate the design for high-performance transmitters. Conventional methods, such as power backoff (BO) or by employing DPD techniques, address these non-linearities at the transmitter level by reducing the PA's output power. For the latter, ML has emerged as a transformative approach to DPD to mitigate the nonlinearity of PA in wireless communication systems. Conventional DPD methods, such as generalized memory polynomial (GMP) and iterative learning algorithms, have provided reliable linearization but are increasingly challenged by the intricate memory effects and dynamic behaviors associated with wideband power amplifiers [13]. Neural network-based solutions have significantly improved our ability to address these challenges, including advanced, recurrent architectures such as long-short term memory (LSTM) and gated recurrent unit (GRU). The use of these models enables accurate prediction of non-linear distortions while accommodating higher-order effects, as evidenced by projects like OpenDPD [14], which achieve notable performance gains in adjacent channel leakage power ratio (ACLR)) and error vector magnitude (EVM) metrics [15]. Unlike conventional methods, such as those employing iterative learning control (ILC) or out-of-band (OOB) linearization, ML-based DPD frameworks provide a more efficient path to linearization under extreme nonlinearity scenarios. For example, while OOB techniques minimize adjacent channel interference by targeting out-of-

band components, they often require iterative processes and significant computational resources, limiting their real-time applicability [13]. In contrast, ML-based models integrate non-linearity correction directly into the signal processing pipeline, reducing both complexity and computational overhead [14]. Furthermore, the ability of ML architectures to adapt through techniques like model pruning and quantization ensures their feasibility in energy-constrained environments, a critical requirement for 6G systems focused on energy efficiency [15].

C. MACHINE LEARNING FOR DPOD

While the techniques outlined in III-B are effective, these solutions involve trade-offs: reduced PA power impacts energy efficiency and coverage, and DPD adds complexity to transmitter design. In 5G, discrete Fourier transform-spread

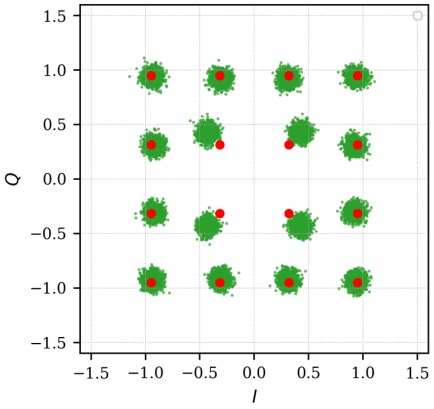


FIGURE 2. The constellations of the received and equalized signals for 16QAM modulated signal and DFT-S-OFDM waveform for a transmitter with non-linear PA with 6dB BO.

orthogonal frequency division multiplexing (DFT-s-OFDM) is employed for uplink transmissions to enhance PA efficiency at the UE side. Similarly, 6G may adopt DFT-s-OFDM for uplinks across FR1/FR2 and potentially in FR3 and sub-THz frequencies. An alternative approach to addressing these challenges is to shift complexity to the network side, thereby relieving the UE from power efficiency constraints. This strategy aims to improve overall energy efficiency, enhance the link budget, and increase throughput by enabling UE operation at higher power levels, closer to the PA's 1 dB compression point. To further address these issues, in this initial work, an AI-enhanced receiver that uses a deep learning architecture, such as residual network (ResNet) (discussed in Section III-C2), is used to mitigate PA non-linearities specifically in DFT-s-OFDM transmitted signals. By analyzing extensive datasets, the AI algorithms learn and adapt to the characteristic behavior of PAs under various conditions, creating an adaptive, ML-driven solution that evolves with environmental changes. This allows the receiver to replace the conventional demapper in the receiver chain by an AI demapper, as illustrated in Fig.3, ultimately maintaining high signal integrity and efficiency. The following content provides an overview of the PA used, outlines the model architecture, and presents the simulation results.

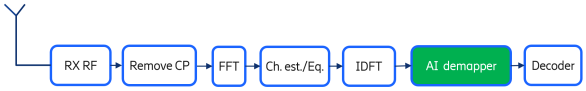


FIGURE 3. AI empowered receiver for DFTs-OFDM.

1) Power Amplifier Model

In this work, a memoryless polynomial model is employed to capture the nonlinear behavior of the PA at the transmitter. The PA nonlinearity is described using a polynomial expression:

$$y_p(n) = \sum_{k \in K_p} a_k \, x(n)|x(n)|^{2k} \tag{1}$$

where n denotes the time index, $x(n)$ is the input to the PA, and $y_p(n)$ represents the corresponding output. The parameter a_k signifies the coefficient of the polynomial term of order k. The PA model applied in this study is a PA-CMOS designed for operation at 28 GHz, with specific polynomial coefficients detailed in [16].

2) Model Architecture

To address the challenges of training deep networks, ResNet leverages "residual learning," a technique that overcomes issues like the vanishing gradient problem by introducing skip connections. These connections enable the training of deeper networks without performance degradation, making ResNet well-suited for tasks requiring increased model depth to enhance learning capacity. In this work, a ResNet variant inspired by ShuffleNet [17] is utilized for soft bit demapping on a per-OFDM-symbol basis, as illustrated in Fig.4. The architecture was selected after numerous experiments with different architectures, aiming to balance computational efficiency and feature extraction for the given problem. The architecture consists of five ResNet blocks, an initial 1D convolution layer (Conv1D) layer for feature extraction, and a final Conv1D layer for output shape adaptation, ensuring compatibility with the regression requirements. Each OFDM symbol is processed with all resource elements in the symbol, and the model outputs log-Likelihood Ratio (LLR) values corresponding to the coded bits. The training dataset comprises observations of equalized symbols in the time domain after the inverse discrete Fourier transform (IDFT), with labels derived from the LDPC-encoded bits in the transmitter (TX) chain. High fidelity between training and deployment is maintained, as some labels can be accurately reproduced at the BS. This model is trained across various backoff (BO) values ranging from 4 to 8 dB, different modulation orders (primarily 64-QAM and 256-QAM), and diverse chipsets.

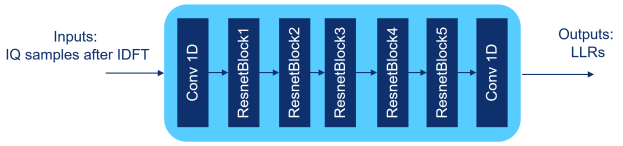


FIGURE 4. AI demapper architecture.

The training also covers a signal-to-noise ratio (SNR) range of -5 to 20 dB and employs a CDL-A channel model, with a UE speed of 3 km/h, enhancing its generalization capability. Furthermore, binary cross-entropy (BCE) is utilized as the loss function, with LLR as the output, making it particularly well-suited for soft-bit regression tasks.

3) Simulation Results

Fig. 5 illustrates the throughput measured in megabits per second (Mbps) as a function of the SNR in dB across multiple PA BO values, demonstrating significant performance differences. In this experimental setup, it is posited that the AI-enhanced receiver feature is integrated on the side of the BS, with the capability to activate or deactivate the ML functionality as required. The ResNetDemapper, evaluated at power BO values of 3 dB and 4 dB, consistently surpasses the performance of configurations when no ML is used at the same SNR levels, and approximates a linear response when no PA is used (the maximum achievable throughput when no PA is used). The ResNetDemapper demonstrates robustness against SNR fluctuations, effectively doubling the throughput to an estimated 500 Mbps at high SNR values in comparison to the conventional receiver when ML is not employed. The results substantiate that the proposed method can mitigate the effects of the non-linearities induced by the transmitter's PA, thereby enhancing throughput. To assess the improvements in

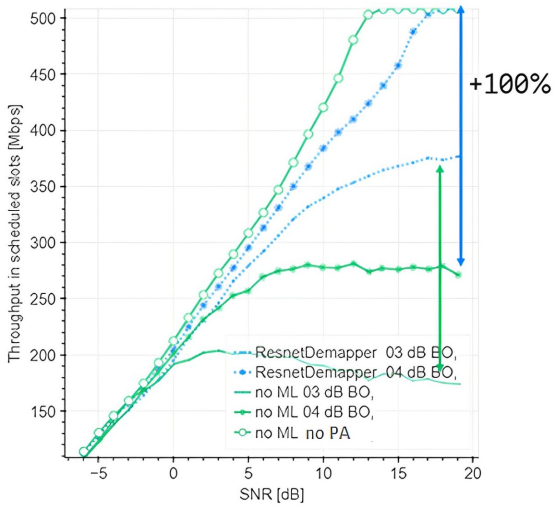


FIGURE 5. Comparison of throughput versus SNR for BO values of 3 dB and 4 dB for an AI-empowered receiver and a conventional receiver where no machine learning is used.

energy efficiency, power efficiency for different BO values is shown in Fig. 6. The analysis is based on a simplified Class A PA model for clarity, where the efficiency follows the relation $\eta = \frac{1}{2 \times BO_{\text{linear}}}$ [18], where BO_{linear} represents the power BO in linear scale. This assumption provides a clear evaluation of efficiency trends while maintaining analytical tractability. The figure shows that increasing the BO value degrades the efficiency of the PA. Since the proposed method enables operation with lower back-off values while achieving similar performance as the legacy receiver, the energy efficiency of the PA can be improved. From this, as illustrated in Fig. 7, it can be deduced that the AI-enhanced receiver, operating with a 5 dB BO at the transmitter, can attain similar performance to that of the conventional receiver with a 9 dB BO. Consequently, the results indicate that the power efficiency has improved by a factor of three. To conclude, this initial work

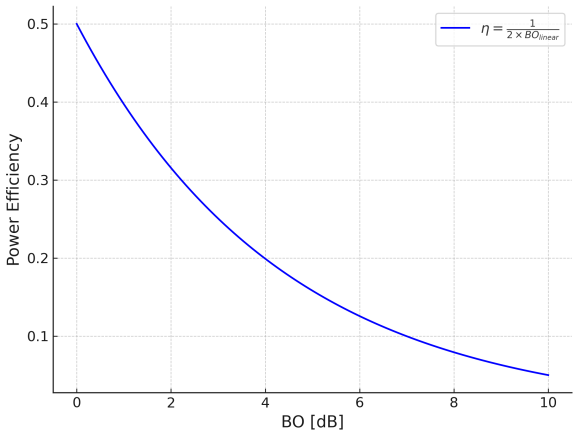


FIGURE 6. Power efficiency of the power amplifier for different backoff values.

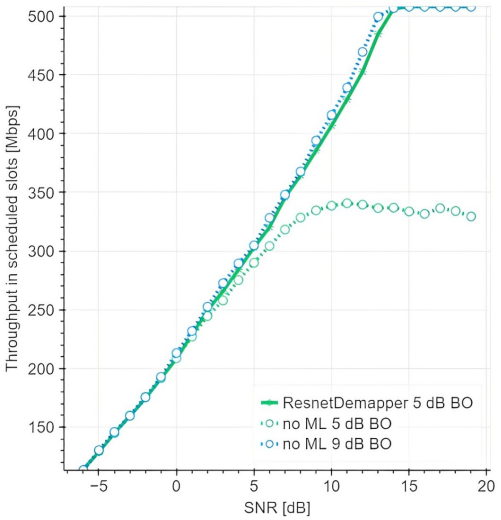


FIGURE 7. Achievable throughput with link adaptation, using an AI-empowered receiver (ResNetDemapper) and a legacy receiver (no ML), for modulation orders up to 256-QAM.

demonstrates the promising potential of machine learning as a viable solution to compensate for nonlinearities generated

by PA when enabling the UE to transmit at higher power levels, thereby improving throughput and power efficiency. For a more comprehensive understanding of the gains, including further enhancements in power efficiency, coverage, and throughput, a potential next step could involve employing a more realistic PA model and conducting further comparisons with additional conventional and novel methods.

D. MACHINE LEARNING FOR PAPR REDUCTION

The demand for higher data rates drives the development of wireless systems with greater bandwidth. In OFDM systems, this translates to a larger number of subcarriers, which typically results in a higher PAPR of the transmit signals. This phenomenon occurs because each time-domain sample of an OFDM waveform constitutes a weighted sum (represented by the discrete Fourier transform) of the constellation points on all subcarriers. Since the constellation points can be seen as independent random variables, the resulting time-domain samples approach a Gaussian distribution as the number of subcarriers increases [19], and thus exhibit a high PAPR. As a result, OFDM signals composed of many subcarriers requires operating power amplifiers with a large power BO to avoid nonlinear signal distortions, which can be prohibitive both in terms of hardware costs and energy consumption (see also Figure 6).

At the same time, OFDM is still the preferred modulation scheme in many settings, due to its low signal processing costs. The PAPR problem can be mitigated with techniques such as DFT-s-OFDM, but this approach can still result in high PAPR, particularly in the presence of multiple streams in MIMO systems. Therefore, PAPR reduction techniques are essential to ensure efficient and affordable operation of wireless systems.

Over the last decades, various PAPR reduction techniques have been proposed. However, their practical applicability is often restricted by strict latency constraints and limited computational resources. Among the most computationally simple techiques is the iterative clipping and filtering (ICF) approach proposed in [20], which repeatedly clips the signal in the time domain and removes the resulting out-of-band radiation via filtering in the frequency domain. In [21]–[23], this approach is generalized into a set-theoretic framework, which allows the derivation of extrapolated projection methods that empirically achieve a considerably lower PAPR at essentially the same computational cost. At the same time, this set-theoretic framework allows for flexibly combining various types of constraints in the frequency-domain, which can improve the detection performance without the need to modify the receiver. In this way, the set-theoretic approach combines the flexibility of PAPR reduction based on convex optimization, as proposed in [24], with the low complexity of ICF. In essence the algorithms proposed in [21]–[23] repeatedly apply a sequence of mappings, which are composed of projections onto convex constraint sets. The extrapolated projection methods in [22] are shown to achieve severe PAPR reduction within only one to four iterations.

ML has the potential to improve further the performance of these techniques. As the iterative algorithms mentioned above typically involve various design parameters that have a considerable impact on their performance, deep unfolding is a promising approach to improve their performance. By interpreting the iterations of an algorithm as the layers of a neural network, and learning the design parameters with standard deep learning techniques such as backpropagation and stochastic gradient descent, deep unfolding gives rise to neural network architectures that incorporate the domain knowledge of the model-based algorithm. This approach has already been applied in the context of multicast beamforming [25] and MIMO detection [26].

IV. CHANNEL ESTIMATION, PREDICTION AND INTERFERENCE ESTIMATION

A. INTRODUCTION AND MOTIVATION

Traditionally, the wireless PHY layer's design has relied on model-based signal processing methods, focusing on accurate stochastic characterization of wireless channels and developing efficient signal models and processing techniques to address various channel and hardware impairments, such as PA non-linearities in Section III. In this section, we discuss the role of AI/ML in estimation and prediction tasks, which are highly challenging tasks due to the random nature of wireless systems and the non-linearities in signal shape introduced by channel and hardware impairments. The developed ML techniques are required to achieve high accuracy while maintaining robustness under a wide range of channel conditions. Furthermore, energy efficiency is a key factor that must be considered in ML-based 6G estimation and prediction solutions. We first address ML-based channel estimation and explore the different methods of learning depending on the type and amount of data available. Following this, we discuss interference estimation in subnetworks and signal-to-interference-plus-noise ratio (SINR) prediction.

B. ML-BASED CHANNEL ESTIMATION

In 6G communication systems, challenges such as the time-varying nature of the channel, RF non-linearities, and the significant overhead in terms of both bandwidth and signaling associated with channel estimation—particularly for massive MIMO systems—can be effectively addressed using data-driven techniques [27]. The existing literature supports adopting ML techniques for channel estimation and proposes various learning algorithms. ML-based channel estimation methods can be grouped into supervised, unsupervised, reinforcement, and hybrid learning approaches, primarily based on how they leverage learning paradigms, dataset availability, and their underlying architectures [28].

1) Supervised ML-based channel estimation

Supervised ML-based methods for channel estimation in 6G systems achieve high accuracy by learning mappings from input data to labeled outputs, minimizing error through loss functions like mean square error (MSE) or cross-entropy.

Fully connected neural networks (FCNNs) perform well in single input single output (SISO) and complex MIMO channels, even handling hardware imperfections in urban scenarios, as shown in [29], [30]. In high-complexity environments like large-scale MIMO and mmWave intelligent reflecting surface (IRS) systems, FCNNs become impractical due to their high computational demands, making convolutional neural networks (CNNs) with optimizations like pruning a more efficient solution [31], [32]. Meanwhile, recurrent neural network (RNN), particularly LSTMs, effectively capture time-varying channel correlations in these dynamic settings [33]. However, these supervised methods require extensive labeled data. Labels in supervised learning refer to the ground-truth CSI values, used to train models, but collecting them is costly because it requires extensive and accurate real-world measurements or simulations, which may not generalize well to diverse and evolving 6G environments; hence, alternative approaches like unsupervised learning are being explored to reduce dependency on labeled data and enhance adaptability.

2) Unsupervised ML-based channel estimation

Unsupervised ML-based methods for channel estimation are used when labels are unavailable due to nonlinearity or nonconvexity problems, optimizing an objective function instead. Autoencoders [34], [35], for instance, have effectively compressed CSI by transforming the channel into the angular delay domain with discrete Fourier transform (DFT) matrices, reducing data volume for efficient transmission across cell zones. While unsupervised and semi-supervised models reduce the need for labeled data, developing a universal model for varied 6G conditions remains challenging.

3) Reinforcement learning-based channel estimation

Reinforcement learning (RL)-based channel estimation uses a trial-and-error approach, where an agent improves estimates through actions guided by rewards and the environment's state. For instance, [36] uses a successive denoising process, leveraging channel curvature to detect unreliable estimates within a Markov decision framework. RL is well-suited for adapting to dynamic channel conditions without needing extensive labeled data, though designing effective reward structures and handling computational demands are key challenges.

4) Hybrid ML-based channel estimation

Beyond the main supervised, unsupervised, and reinforcement learning approaches, other ML-based methods for channel estimation leverage hybrid and specialized techniques to meet complex requirements. Recent efforts focus on multi-task neural networks that integrate channel estimation with transceiver tasks like antenna extrapolation and CSI feedback, leveraging shared prior channel information to enhance operational efficiency and reduce redundant computations, contributing to energy-efficient system design. For instance, a deep plug-and-play prior-based algorithm [37] allows a network to handle these interconnected tasks simultaneously. Additionally, generative adversarial networks (GANs) [38] are being utilized to overcome limitations in traditional deep learning loss functions, generating realistic channel data to enhance estimation accuracy. One notable example of hybrid ML approaches is the use of diffusion models (DMs) for channel modeling [39], [40]. Specifically, diffusion-denoising probabilistic models (DDPMs) learn channel distributions by adding noise to data until a Gaussian distribution is reached, then reversing the process to model and estimate channels effectively. This progressive noise addition and removal allows DDPMs to explore the full range of data points in the channel distribution. Unlike GANs, which often generate high-quality samples from limited modes of the distribution, diffusion models systematically move through all modes, preventing the mode collapse issue common in GANs. By gradually diffusing data and reconstructing it step-by-step, DDPMs capture both the high-probability and low-probability regions of the distribution, enabling more comprehensive coverage of the channel's characteristics. This advantage becomes particularly significant in complex channels, such as specific fading scenarios where GANs may fail to model the full diversity of channel behavior. Therefore, DDPMs provide a flexible, effective alternative for reproducing channel behaviors without repeated physical experiments, outperforming GANs in many cases. One limitation of DMs is their slow sampling speed, which we addressed through techniques like skipped sampling and denoising diffusion implicit models (DDIMs). Additionally, [39] optimized noise scheduling and parameterization using sliced Wasserstein distance (SWD) and end-to-end (E2E) symbol error rate (SER) to balance sampling efficiency and generative performance. The experiments showed that DMs outperform WGANs in scenarios with complex fading, providing more stable and generalizable SER performance, particularly at high Eb/N0. More specific contributions of [40] include: a DM-based channel generation framework compatible with E2E neural network communication systems; pretraining methods for faster, more stable convergence with lower SER; and a detailed analysis of noise scheduling and parameterization to improve trade-offs between performance and sampling speed.

In summary, data-driven methods hold great potential for 6G channel estimation. Supervised approaches offer high accuracy but struggle with labeled data demands, while unsupervised methods address label scarcity but face adaptability challenges. Hybrid techniques, such as diffusion models, excel in modeling diverse channel behaviors and outperform GANs in complex scenarios. Future research should focus on integrating hybrid ML for efficient, adaptable estimation, leveraging multi-task models for system-level optimization, and enhancing the practicality of these methods for 6G deployment.

C. ML-BASED INTERFERENCE ESTIMATION FOR SUB-NETWORKS

Over the past decades, the vision of connecting "anything that may benefit from being connected" has driven an exponential rise in wirelessly connected devices, from small sensors in homes and industries to vehicles and beyond. However, interference has always been a significant challenge, negatively impacting the capacity and performance of wireless communication systems. In ultra-dense environments, where multiple UEs communicate over limited resources such as shared sub-bands, high interference levels degrade the signal-to-interference-plus-noise ratio (SINR) and pose a significant challenge to effective radio resource management. This directly impacts network coverage and makes it difficult to achieve QoS. One such example is hyper-dense deployed In-X subnetwork (SN) which is envisioned to offer short-range communication with extreme requirements such as 0.1-1 ms latency, life-criticality, and high reliability ($99.9999\% - 99.99999\%$) [41]. Examples of In-X SNs include in-robots, in-vehicle, in-body, and in-house environments. The coexistence of the requirement of sub-ms latency which needs high bandwidth of intra-SN communication and limited available bandwidth, inter-SN interference poses a significant challenge. In such scenarios, predictive interference management—leveraging explicit interference predictions through advanced ML techniques—helps allocate resources proactively and prevent performance degradation.

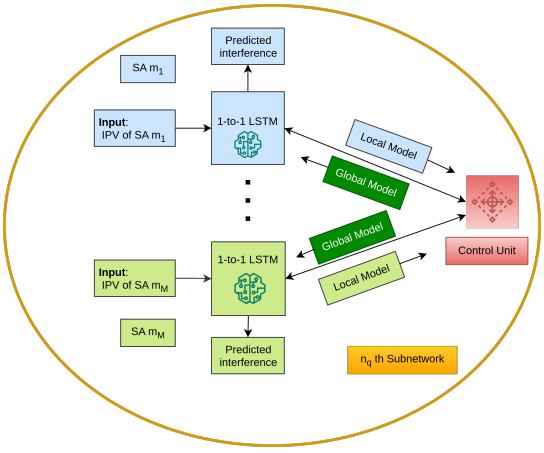


FIGURE 8. FL-based interference predictor proposed in [42].

Various interference estimation techniques have been developed to predict interference, thereby facilitating efficient resource allocation proactively. ML-based approaches, particularly recurrent neural networks such as LSTM-based techniques, are utilized to model the complex, non-linear characteristics of interference by leveraging spatiotemporal dependencies from historical interference power values [43]–[46]. Furthermore, a federated learning (FL)-based interference estimation technique has been introduced, allowing cooperative learning from interference power vectors across multiple SNs. This approach utilizes an existing cellular network for model aggregation, achieving faster convergence and lower estimation errors compared to standalone methods. Detailed insights into this work can be found in [46] and [42]. Due to connectivity limitations or by design, SNs may operate independently from the cellular network. To address this, a prediction technique has been developed in which interference prediction is performed locally at the sensor-actuator (SA) pair using a single-layer LSTM model [46]. Additionally, a many-to-many (M-to-M) LSTM model has been proposed as a centralized approach, where M SA pairs collaboratively learn and predict interference at the SN controller [42]. Incorporating a FL into this framework offers a promising solution for collaboratively learning interference patterns by leveraging the knowledge of participating clients, such as other SA pairs, as illustrated in Fig. 8. The performance of these methods has been compared against the Wiener predictor, which relies on second-order statistics—such as the correlation of interference power values—under the assumption of stationary interference dynamics. Numerical results and analysis demonstrate the effectiveness of the ML-based methods and are detailed in [42]. While these predictors employ point prediction, they assume noise-free interference measurements and deterministic traffic scenarios. To address randomness arising from traffic variations, estimation errors, and other uncertainties, a probabilistic interference prediction framework has been introduced. This framework predicts specific or predefined quantiles instead of the mean, ensuring reliability while incurring minimal additional resource usage. As an alternative approach, a Bayesian framework using sparse Gaussian process regression (SPGPR) has been proposed [47], [48], which can be utilized to reduce the computational complexity in a reasonable factor, has been used for interference prediction [49]. Moreover, to consider the computational cost at low-end SA pairs, this work has been further extended to sparse Gaussian process regression with variational inference (VISPGPR), which selects a subset of salient data with k samples from the entire data with l samples to decrease the computational load. This reduces the complexity from $O(l^3)$ to $O(l \cdot k^2)$, where l is the number of interference power values processed, and k is the number of salient features, with $k \ll l$ [48].

Interference often possesses two properties: (1) heavy-tail distribution and (2) correlation due to small-scale fading. Instead of a Gaussian kernel, the Matern kernel is more suitable to accommodate and enhance these properties. To cope with interference data randomness and outliers due to random traffic, a robust approach using a Student-t distribution instead of Gaussian likelihood has been implemented [49]. A clear downside to Gaussian Process Regression is the requirement to select the covariance kernel and likelihood beforehand, which necessitates pre-knowledge of the distribution and may lead to discrepancies in different scenarios. To overcome this limitation, a probabilistic interference prediction method utilizing quantile bidirectional long-short term memory (QBiLSTM) has been proposed [49]. This method leverages temporal correlations without assumptions about the underlying distribution of target variables. An

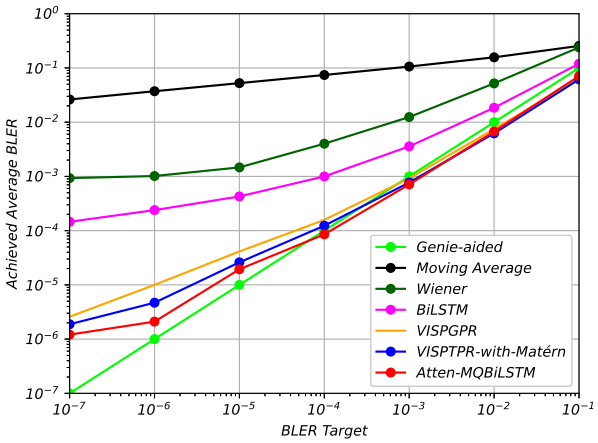


FIGURE 9. Achieved average BLER with isochronous traffic [49].

enhanced version, attention based quantile bidirectional long-short term memory (AttenQBiLSTM), incorporates a modified loss function to further improve prediction accuracy. The efficacy of these predictors has been evaluated using a spatially constant 3GPP channel model with realistic mobility and traffic scenarios. Resource allocation is performed using predicted and true interference, applying finite blocklength theory to determine the number of resource usage. The discrepancy is assessed in terms of the achieved Block Error Rate (BLER) with prediction, considering packets with a finite blocklength; further details can be found in [49]. Results in Fig. 9 indicate that AttenQBiLSTM outperforms other proposed methods. This illustrates that advancements in ML algorithms, along with their efficient utilization, enable the design of data-driven interference predictors with minimal or no reliance on distributional assumptions. Furthermore, accounting for distribution drift and evaluating models across different channels is highly significant. Interference prediction can also be extended to power control using causal inference, as demonstrated in [50], enabling energy-efficient solutions for future 6G systems.

D. ML-BASED SINR PREDICTION

ML-driven SINR prediction enables the PHY layer to proactively adjust its transmission parameters, optimizing resource allocation based on anticipated channel conditions. By accurately predicting SINR, the PHY layer can fine-tune transmission power, modulation schemes, and coding rates in real-time, thereby reducing unnecessary power consumption while maintaining signal quality. This dynamic adaptation minimizes the use of excessive power in favorable conditions. It ensures energy savings in scenarios where high power transmission would otherwise be required to counteract interference or poor channel conditions.

To address one of the key challenges of addressing the optimized private campus network (PCN) operation, we implemented ML models, including bi-directional and vanilla

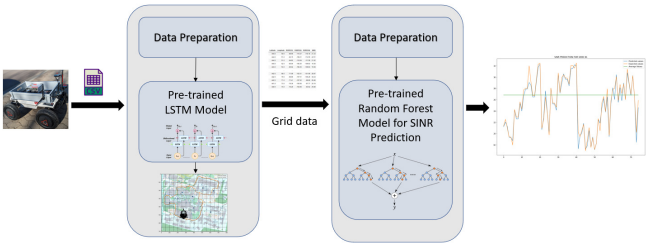


FIGURE 10. Workflow of SINR prediction scheme in PCN [51]

LSTM networks combined with Random Forest algorithms, to accurately predict SINR [51]. By utilizing historical network data and user mobility patterns to predict SINR, these models assist in making decisions in semi-persistent scheduling, smart resource allocation, modulation, and power control. These intelligent actions will lead to lower channel estimations, better energy, lower latency, and overhead. Thus, the efficient and optimal operation of PCN is scheduling radio resources in controlled campus networks.

Our study [51] uses a live PCN at the RPTU-Kaiserslautern campus, where we gathered data to create a grid-based radio environment map. We then applied the mobility attributes of users alongside ML models- vanilla LSTM, bidirectional LSTM, and Random Forest—to predict SINR along users' paths, as shown in Figure 10. The evaluation results for SINR predictions from the random forest model based on the bidirectional-LSTM model's mobility prediction with 0.86 mean absolute error (MAE) and 0.92 R^2 score show that the proposed grid-wise SINR prediction shows reasonable performance in the PCN scenario.

The implications of SINR prediction extend beyond mere energy efficiency, manifesting in support of more nuanced and energy-conscious beamforming techniques and multi-antenna configurations [52]. This predictive control over network parameters aligns with the overarching objectives of sustainable, green communications, ensuring that 6G systems can attain their performance targets without significant escalations in energy consumption. The role of SINR prediction in supporting the energy-conscious techniques in IRS aided Terahertz communication system [53] is of considerable significance, as it highlights the broader implications of this technology and its potential to shape the future of 6G networks.

To demonstrate the influence of SINR prediction in a practical context, one might consider SINR prediction in densely populated 6G networks. In these environments, characterized by a high density of users and devices, the capacity to make real-time adjustments to transmission power and resource allocation is essential for maintaining signal quality and averting interference. In this context, the integration of ML-based SINR prediction represents a crucial element in the network's ability to anticipate future channel conditions and proactively optimize critical parameters, such as transmission power and modulation schemes [54]. By identifying instances where the SINR is likely favorable, the network can

reduce transmission power without compromising quality, leading to tangible reductions in energy consumption.

Furthermore, the foresight inherent in SINR projections facilitates the deployment of more effective beamforming methodologies, wherein energy is precisely allocated toward users likely to experience suboptimal SINR. This targeted energy allocation reduces the overall energy demand in massive MIMO systems and maintains the necessary quality of service. In conjunction with these developments, integrating SINR prediction minimizes the necessity for continuous, energy-intensive real-time channel estimations, further reinforcing the network's energy efficiency and progress toward sustainable operations.

V. CHANNEL CODING

We tackle the key issue of scaling neural networks for error correction to large blocklengths, a problem that arises due to the exponential increase in complexity when using one-hot encoding. While small neural networks perform very well for small blocklengths, by optimizing over symbols rather than bits, scaling them to blocklengths of 100 or more becomes impractical. The number of nodes for the input layer alone grows as 2^k for one-hot encoding, and similar embeddings lose their advantage when drastically downsized, making it computationally prohibitive to scale these methods. On the other hand, bit-based approaches like TurboAE scale better but perform worse for smaller blocklengths due to their bit-level optimization. Our method provides an intermediate solution that allows for practical scaling without sacrificing the advantages of symbol-level optimization. We propose a concatenated classic and neural (CCN) code that combines the flexibility of neural networks with the robustness of Reed-Solomon (RS) codes. Each RS codeword symbol is encoded using a separate instance of the same small neural network, trained with one-hot encoding and optimized via categorical cross-entropy. This structure allows us to extend the effective code dimension while controlling the complexity, providing a scalable solution that maintains high performance at large blocklengths without requiring exponentially larger networks. Our method scales to blocklengths of $n = 255 \cdot 12 = 3060$, while avoiding the training complexity that limits standalone symbol-wise neural codes. In simulations on additive white Gaussian noise (AWGN) channels, a $(255 \cdot 12, 223 \cdot 8)$ CCN code significantly outperforms a standalone $(7, 4)$ reference neural codes, achieving up to 3.1 dB gain at BLER of 10^{-4}. Moreover, an outer RS decoder with an erasure-detection mechanism further improves reliability by correcting both errors and erasures based on a thresholding algorithm applied to the neural decoder's softmax output. CCN codes also demonstrate robustness across different channel conditions, providing resilience to burst errors and fast fading in Rayleigh channels, where they achieve an E_b/N_0 gain of approximately 8.3 dB at a BLER of 10^{-4}. Additionally, CCN codes show an E_b/N_0 gain of 6.2dB as compared to the reference code at a BLER of 10^{-4}, in burst channels.

By leveraging concatenation, we bridge the gap between dense, one-hot encoded neural codes, which perform well for short blocklengths but fail to scale, and bit-based methods like TurboAE, which scale well but underperform at shorter lengths. Our method offers a practical and scalable solution, achieving a balance between performance and complexity [55]. In addition to its demonstrated gains over classical and other neural network-based codes, this hybrid approach aligns with key 6G objectives of reliability, energy efficiency, and adaptability to diverse channels. By supporting very-large blocklengths, CCN codes circumvent the prohibitive training and complexity overhead of standalone symbol-wise neural codes, thus providing a more scalable and practical solution for future communication systems.

VI. MULTI-USER MIMO ASPECTS

A. INTRODUCTION AND MOTIVATION

Multi-user MIMO technology plays a crucial role in enhancing performance of wireless networks due to its ability to increase the spatial degrees of freedom available in the system and take advantage of the diverse nature of wireless channels. For signal processing and resource allocation tasks, such as signal detection, this translates to solving challenging optimization problems in high-dimensional space. To this end, the first part of this section addresses AI/ML-based multi-user MIMO detection to find a lower-complexity alternative to conventional methods without compromising performance. Following this, the focus is shifted to cell-free massive MIMO (CF-mMIMO) systems. Unlike traditional cellular networks, CF-mMIMO eliminates the concept of cell boundaries by allowing distributed RAPs to serve users over a broad area jointly, and all RAPs are connected to a central processing unit (CPU) via a fronthaul network. The straightforward approach of CF-mMIMO encounters scalability challenges with increasing user numbers due to rising computational complexity and fronthaul rates. The User-Centric CF-mMIMO implementation was introduced and adopted in most of the literature where each user is served by only a cluster of RAPs for enhanced scalability. In this context, we tackle the issue of fronthaul compression in user-centric CF-mMIMO for communication overhead reduction. Next, we demonstrate the gains in energy savings obtained by ML-based sleep-mode selection techniques, which involve switching off unnecessary network elements given a target QoS requirement.

B. NEURAL RECEIVER FOR MULTI-USER MIMO

MIMO technology such as multi-user MIMO is a critical ingredient in future wireless systems, as it allows to make efficient use of scarce frequency resources. To this end, modern wireless networks rely on large-scale antenna arrays, to enable high-dimensional signal transmission in the spatial domain, using the same frequency. For instance, in multi-user MIMO systems, several UEs simultaneously transmit data in the uplink, using the same portion of the frequency spectrum. Optimal signal detection at the BS (e.g., based on

a maximum likelihood criterion) is known to be NP-hard [56]. As there is no hope to design a MIMO detector that is both optimal and efficient, a growing body of research has been dedicated to devising suboptimal detectors that strike a balance between performance and complexity, over the past six decades [57]. At one end of this trade-off are high-performance approximation techniques such as sphere decoding [58], which entail high computational costs. At the other end are linear detectors and low-complexity iterative detectors based on approximate message passing (AMP) [59], which can suffer from bad performance on realistic channels [60].

More recently, the trend towards large-scale antenna arrays has driven the need for detectors with extremely low complexity. To meet this goal, various AI/ML-based MIMO detectors have been proposed. Many of them rely on deep unfolding [61] to learn certain design parameters of iterative detectors including projected gradient methods [62]–[64], the alternating direction method of multipliers (ADMM) [65], AMP [60] or orthogonal approximate message passing (OAMP) [66], [67]. These approaches treat the iterations of the respective model-based algorithms as layers of a neural network, which allows the optimization of learnable parameters via standard deep learning techniques such as backpropagation and stochastic gradient descent. Most of the aforementioned AI/ML-based detectors are trained offline based on a synthetic dataset (i.e., simulated channel realizations, transmit- and receive signals). However, realistic conditions in which the data distribution varies over time, pose a severe challenge to such techniques, and it has been shown [60] that many of these techniques suffer severe performance degradation when realistic channel models are considered. The authors of [60] mitigate this problem by proposing an online learning approach, in which the detector is retrained given each individual channel realization, using synthetically generated transmit data. While this online approach is shown to outperform all offline learning methods considered, it requires continuous training at runtime. To avoid the need for online training, and instead improve generalization in the presence of realistic radio environments, a set-theoretic framework for MIMO detection is proposed in [68], [69]. The proposed iterative detector is based on a superiorized [70] version of the adaptive projected subgradient method (APSM) [71]. Compared to other iterative detectors such as the methods in [59], [62], the method in [68], [69] imposes fewer assumptions on the channel matrices, in order to guarantee convergence. A deep-unfolded variant of this set-theoretic MIMO detector has been proposed in [26]. Results in that study show that the proposed detector can outperform other learning-based detectors having similar complexity on realistic channels.

C. USER-CENTRIC CELL-FREE MASSIVE MIMO

1) Relevant Information Processing for Fronthaul Rate Reduction

In a cell-free massive MIMO (CF-mMIMO) system, vector quantization (VQ) can be highly effective in managing and compressing the large volumes of data at the distributed RAPs where each RAP gathers high-dimensional data from several users and needs to forward this data to the central processing unit (CPU) over limited-capacity fronthaul links. VQ significantly reduces the data rate required for these transmissions by representing clusters of observation vectors with single representative points, or centroids, and by reducing the transmission rate, VQ helps lower the energy required for data transfer, a critical factor for large-scale CF-mMIMO systems where energy efficiency is paramount.

The multi-source information bottleneck (IB) method [72] was introduced as a novel distributed noisy source coding scheme for a generic setup wherein, several terminals receive different sets of *noisy* observations from the users and compress their signals using IB-based vector quantization method before transmitting them over multiple rate-limited channels to a CPU. Unlike standard VQ methods, which minimize distortion without considering the specific signal reconstruction needs, the IB-based approach selectively retains information essential for accurately reconstructing a user signal at the CPU. Particularly suited for CF-mMIMO's distributed architecture, the multi-source IB method enables RAPs to collaboratively compress noisy observations, preserving the most critical information needed for user signal recovery. The IB approach achieves an optimized trade-off between information retention and compression rate [73], making it a powerful solution for scalable, energy-efficient CF-mMIMO networks. The IB method, originally introduced in [74], serves as a task-oriented compression framework, and its early applications can be found in the domain of Unsupervised Learning, where it was employed as an information-theoretic strategy for cluster analysis [75].

In CF-mMIMO systems, the multi-source IB method can be adapted by applying linear equalization at each RAP to cancel the spatial interference of different users (which get served by it) and separate their signals [76]. Then the design problem is formulated as a basic trade-off between two mutual information (MI) terms. The first one is the sum of MI terms between each source signal and its sets of the corresponding received signals at the CPU, called the relevant information. The second one called total compression rate is the sum of MI terms between the noisy observations (of the source signals) which are the inputs of the local compressors and their output signals. The goal of IB-based compression is to maximize the relevant information such that the total compression rate does not exceed the capacity of the corresponding fronthaul network which is similar to the concept of data clustering, where the goal is to group observations in a way that retains the most relevant features while reducing the redundancy. The generalized multivariate information bottleneck (GEMIB) algorithm was presented in [72] to address the design problem for multi-source IB-based

compression.

Consider 3 users served by 2 RAPs in a CF-mMIMO system where one of the users is served by both RAPs as illustrated in Fig. 11a. After the linear equalization at RAPs, RAP 1 has the noisy observations $\mathsf{y}_1^{(1)}$ and $\mathsf{y}_2^{(1)}$ of the source signals x_1 and x_2, respectively and RAP 2 has the noisy observations $\mathsf{y}_2^{(2)}$ and $\mathsf{y}_3^{(2)}$ of the source signals x_2 and x_3, respectively. Assume a discrete memoryless channel that approximates a discrete-time, discrete-input, and continuous-output AWGN channel with identical noise variance, σ_{n}^2, for all access channels from the source signals to the corresponding outputs of equalizers at the RAPs. RAPs 1 and 2 compress their noisy observations into z_1 and z_2, respectively, before a forward transmission over ideal (error-free) rate-limited channels (IRCs) to the CPU. To compare the compression schemes, we use the relevant information that is basically the overall transmission rate, i.e., the sum of MI terms between the source signals and the corresponding received signals at the CPU, $I(\mathsf{x}_1;\mathsf{z}_1) + I(\mathsf{x}_2;\mathsf{z}_1\mathsf{z}_2) + I(\mathsf{x}_3;\mathsf{z}_2)$. Fig. 11b illustrates the obtained results comparing GEMIB and K-Means [77] methods for designing the compressors at RAPs in terms of overall transmission rate versus different numbers of output clusters (per RAP). The key takeaway is the clear performance advantage of the GEMIB algorithm over the standard K-Means routine. This superiority is evident in both information preservation and compactness. For example, with $\sigma_{\mathsf{n}}^2 = 1$, the GEMIB algorithm achieves approximately 5 bits of relevant information while requiring only 12 clusters, compared to K-Means, which requires 16 clusters for the same information. Alternatively, if the number of output clusters is fixed at 8, GEMIB supports up to 4.5 bits of relevant information, outperforming K-Means, which can support up to 4 bits.

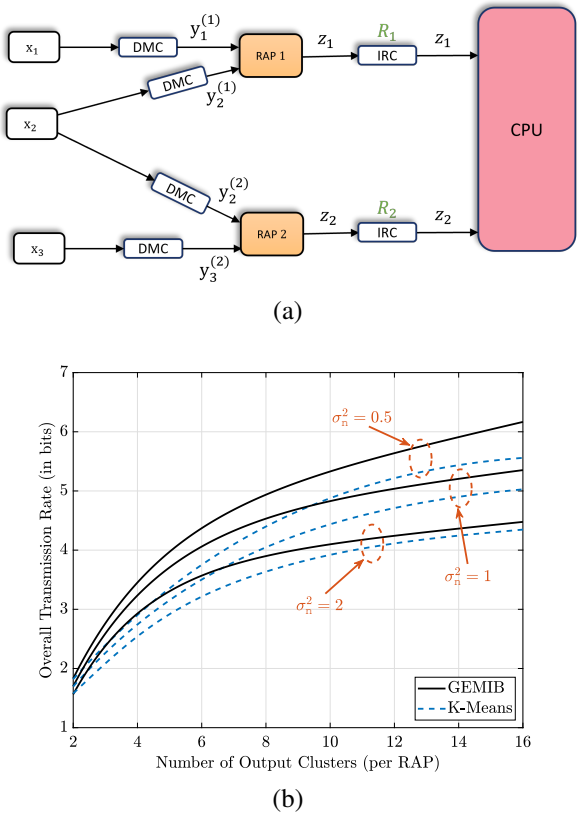


FIGURE 11. **(a)** An example of uplink transmission in a CF-mMIMO network where 3 users employing bipolar 8-ASK source signaling are served by 2 RAPs. **(b)** A performance comparison between GEMIB [72] and K-Means schemes to design the compressors at RAPs in terms of the overall transmission rate vs the number of output clusters (per RAP).

ML-based approaches, such as the forward-aware information bottleneck (FAVIB) method, offer a powerful extension to the traditional IB method by incorporating predictive models into the compression process. These models are particularly useful for learning the underlying statistical relationships between users' signals and optimizing compression decisions based on this knowledge. The deep learning-based FAVIB approach [78], [79], for example, uses variational methods to optimize compression based on a finite set of training samples, making it suitable for scenarios where statistical information about the user signals is not fully available.

2) Sleeping Radio Access Points

Sleep-mode selection is a technique that can be applied to CF-mMIMO systems in order reduce total network power consumption by allowing a subset of serving RAPs to be turned off whenever they are not required to achieve per user QoS requirements.

While turning off a subset of RAPs provides gains in terms of reduced energy usage, further gains can be achieved by combining sleep-mode selection with transmit power control and beamforming design. However, performing efficient resource allocation requires information sharing over fronthaul links between the central unit and distributed RAPs, which, as discussed above, can lead to high signaling overhead. This can be avoided by performing long-term resource allocation, i.e., allocating resources on the basis of channel statistics, rather than instantaneous channel information.

In light of the above discussion, [80] proposes an ML-based framework for the joint optimization of transmitted power, beamforming and the RAP sleep-mode selection to maximize the gains in uplink CF-mMIMO systems while minimizing the network power consumption using channel statistics. In [80], the approach taken condenses the original joint optimization problem into a combinatorial optimization problem with the objective of finding the RAP sleep-mode configuration that minimizes the network power consumption and implicitly finds the optimal long-term transmit power allocation and beamformer design utilizing tools from fixed-point theory, subject to minimum per-user QoS requirements. Typically, efficient methods for solving this type of large-scale combinatorial problems do not exist, and the computational complexity associated with solving a fixed-point algorithm for every possible configuration renders exhaustive search highly impractical.

To this end, surrogate optimization is employed by treating a computationally expensive function as a "black-box" function and approximating it with ML models, such as Gaussian Process Regression or artificial neural networks. The proposed ML-based method in [80] was able to significantly reduce the network power consumption when trained with only a small fraction of the search space of the original combinatorial problem providing an effective solution to the original problem with a much lower computational complexity. This demonstrates the effectiveness and potential of ML-based approaches for handling the large-scale nature of the optimization problems associated with CF-mMIMO systems.

VII. SUB-NETWORKS AND NETWORK OPTIMIZATION

A. INTRODUCTION AND MOTIVATION

The diversity of applications and scenarios envisioned for 6G calls for network optimization methods that are able to adapt quickly to fast-changing environments and serve users with vastly differing requirements. Considering this, AI/ML-based network optimization can improve wireless performance and energy efficiency by serving as lower-complexity and more robust alternative to conventional techniques.In this section, AI/ML is studied in combination with key performance indicator (KPI) monitoring for AI/ML-based network optimization to mitigate performance drops and provide a network planning approach. Then, the problem of ML-based robust resource management with energy efficiency as the main objective is tackled. Finally, the role of AI/ML in vehicular networks is examined and key aspects required for successful deployment in real-world settings are highlighted.

B. PREDICTIVE NETWORK CONTROL BASED ON CROSSBAND KPI-MONITORING

The ability to predict the quality of a wireless channel enables numerous adjustments to be made to network operation in order to increase both performance and energy efficiency. Current and environment-specific network data is required for realistic predictions, which can be obtained, for example, using the spatially distributed traffic and interference generation (STING) approach for real-time network monitoring, as shown in [81], as well as spectrum monitoring, presented in [82]. Based on such information, location- and application-specific performance drops can be detected, which are then remedied by mitigation strategies and predicted in the future due to predictive network control. These strategies are implemented by adjusting context and system parameters that physically intervene in the radio environment and change the network operation and control, respectively.

In [81], the STING-based real-time network monitoring system previously tested in a smaller test environment [83] was transferred to a large-scale industrial manufacturing environment to evaluate the performance of mmWave connectivity in a challenging metallic radio environment. The network performance key performance indicators (KPIs), obtained through static single- and multi-user tests as well as mobile measurements, confirm the robust performance and scalability of mmWave frequencies for future 6G industrial networks. Furthermore, a mitigation strategy consisting of a passive IRS has been validated, whose installation corresponds to a change of context parameters and enables a restoring of line-of-sight (LOS) performance in obstructed locations seamlessly.

As an addition to the STING system, a Software Defined Radio (SDR) has been added to the distributed STING units in order to allow channel monitoring beyond the network itself. This enables an AI based anomaly detection approach, described in detail in [82]. Based on the widely used image-based neural network structure U-Net [84], this approach allows detection and alerting of unwanted channel activity such as in the 3.7 GHz band, with an accuracy of up to 90% with relatively low complexity. These KPI monitoring aspects can be embedded in a holistic network optimization depicted in Fig. 12. The illustration shows that based on given user requirements an AI-driven network planning approach can be used in order to achieve a sufficient initial coverage and network performance through an iterative optimization process of prioritized areas, as in [85]. In addition, continuous KPI and spectrum monitoring enables predictive network control, allowing real-time optimization of network configuration, application characteristics and programmable radio environment components (e.g. IRS), to solve a demand-driven trade-off between energy efficiency and network performance.

C. ROBUST RESOURCE MANAGEMENT VIA ML

Given that energy efficiency is a critical aspect of AI/ML-native 6G, this section focuses on energy-efficient physical layer resource management. The recent work [86] formulates an energy efficiency maximization problem using rate-splitting multiple access in mixed-critical cloud radio access networks. By jointly optimizing beamforming, rate allocation, and decoding groups, [86] proposes a low-complexity iterative algorithm based on fractional programming.

Despite its strong numerical performance, the algorithm lacks real-time feasibility. Many resource allocation problems in communication systems require computationally demanding optimization, where ML provides an efficient alternative, especially in dynamic and uncertain environments. Deep learning has demonstrated its potential for low-complexity and robust 6G solutions [87], such as relay selection [88] and robust optimization [89]. Building on this, [90] proposes a data-driven approach to physical layer resource management, leveraging unsupervised deep learning to handle channel uncertainties and minimize worst-case delays.

In this subsection, we extend the methods from [90] to tackle energy efficiency maximization under channel state uncertainty. Specifically, we consider a system where a single BS with L antennas serves K wireless users, as illustrated in Fig. 13. The channel coefficient between the BS's antenna l and user k is denoted as $h_{l,k} \in \mathbb{C}$, modeled as $h_{l,k} = \hat{h}_{l,k} + h_{l,k}^{\text{err}}$, where $h_{l,k}^{\text{err}} \sim \mathcal{CN}(0, \sigma_h^2)$ represents the channel estimation error. As discussed in Section IV-B,

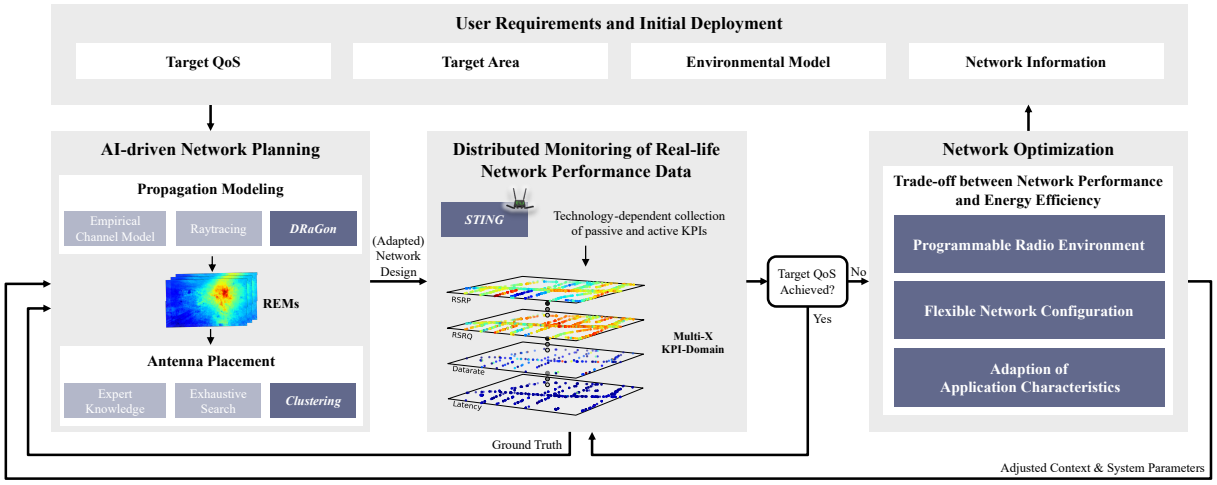


FIGURE 12. Envisioned AI-based network optimization flow.

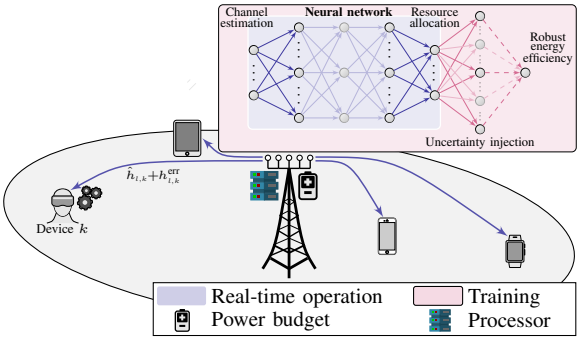


FIGURE 13. System model consisting of one BS performing power control under channel uncertainty. The BS's embedded edge server comprises of a DNN trained for robust energy efficiency maximization.

Communication resources	DNN parameters
$L = 4$ antennas, $K = 4$ users	10 fully connected layers (ReLU)
$p^{\text{fix}} = 30$ dBm, $P^{\max} = 1.1$ Watt	400 neurons per layer
-75 dBm/Hz noise power	150 epochs, 50 minibatches
10 MHz bandwidth	1000 batch size, 2000 test size

TABLE 1. Simulation parameters for Section VII-C.

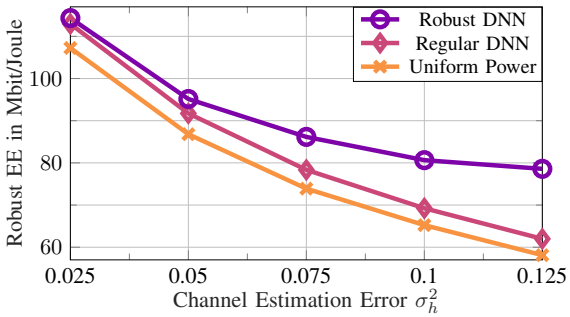


FIGURE 14. Robust energy efficiency in Mbit/Joule as a function of the channel estimation error σ_h^2 comparing robust and regular DNNs and a uniform power allocation scheme.

even ML-based channel estimation techniques introduce errors. In practical systems, the exact distribution of channel uncertainty is often unknown. For simulations, we model the uncertainty as additive and Gaussian, though the framework can accommodate other types of uncertainty [89]. The BS operates under a power constraint $P^{\max}$ and is equipped with embedded computing capabilities. A DNN is employed, taking as input the estimated channel coefficients $\hat{h}_{l,k}$ and $P^{\max}$, and outputting the transmit power allocation p_k for all users. The energy efficiency objective is defined as the ratio of the total achievable rate $R = \sum_{k=1}^{K} r_k$ to the total power consumption $\sum_{k=1}^{K} p_k + p^{\text{fix}}$, where p^{fix} accounts for fixed network power consumption. However, since R depends on uncertain channel conditions, it cannot be deterministically evaluated. Instead, we consider the γ-th of the sum rate distribution, R^{γ}, satisfying $\Pr\left[R < R^{\gamma} \middle| \hat{h}_{l,k} \forall (l,k) \in (\mathcal{L}, \mathcal{K})\right] \leq \gamma$. During training, we leverage the statistical properties of the channel estimation error by generating multiple realizations of $h_{l,k}^{\text{err}}$, computing the energy efficiency for each realization, and extracting the 5%-quantile to obtain an empirical estimate of robust energy efficiency. In practical operation, rather than generating multiple channel error realizations, the scheme *samples* a large number of channel realizations, allowing the DNN to implicitly learn the channel's statistical properties empirically. The DNN is trained to optimize energy efficiency while ensuring robustness against channel uncertainties.

To evaluate the performance of the proposed robust energy efficiency maximization scheme, we compare it against two benchmarks: (i) a standard DNN trained without considering channel uncertainties and (ii) a uniform power allocation scheme, which distributes the available transmit power equally among users. The simulation parameters are set according to Table 1. We additionally utilize regularized zero-forcing beamforming [89].

Fig. 14 presents the robust energy efficiency (in Mbit/Joule) as a function of the channel estimation error variance σ_h^2. As expected, increasing the channel estimation error variance degrades the energy efficiency due to the higher interference, suboptimal power allocation, and general inefficiencies. Notably, the proposed robust DNN

achieves the highest utility and maintains superior energy efficiency under high uncertainty compared to the baselines. While both ML-based approaches outperform uniform power allocation at low estimation errors, the regular DNN (which does not account for uncertainties) loses its advantages as σ_h^2 increases, ultimately approaching the uniform baseline. These results highlight the importance of both ML-based optimization and uncertainty-aware modeling in achieving high energy efficiency in the 6G physical layer.

D. AI/ML IN V2X NETWORKS

One of the key research areas in 3GPP is vehicular networks [91], [92]. Vehicle-to-everything (V2X) communications are the driving force behind vehicular networks' widespread proliferation that enables several intelligent transport systems use cases such as emergency vehicle and collision warning, cooperative lane change, extended sensors, remote driving and vehicular platooning. These use cases have various stringent requirements in terms of throughput, latency, reliability, etc., and require certain levels of QoS. To meet these demands, accurate network behavior prediction is essential. AI/ML is crucial in V2X networks as it facilitates real-time decision-making, optimizes traffic management, and reduces congestion through predictive analytics. It also increases communication efficiency by allocating resources and spectrum as efficiently as possible. By analyzing the data from multiple sensors in vehicular networks, potential failures can be detected. AI/ML algorithms also enable cooperative perception and decision-making by integrating the data gathered by multiple sensors or V2X sources. Cooperative driving (like platooning) is also facilitated through which vehicles communicate and adjust speeds to improve fuel efficiency and safety. All things considered, it may assist with real-time optimization of beamforming, power regulation, spectrum allocation, and link adaption, which enhances the overall use of radio resources, intelligence, and transportation system dependability [93].

Although AI and ML algorithms hold great potential for vehicular networks, their real-world implementation or validation remains elusive due to their high dependency on the collected datasets. The creation of such datasets is resource-intensive task as they need to reflect the whole rage of scenarios that vehicular networks may encounter. Recent studies have acknowledged this challenge and have attempted to generate reference datasets to capture both vehicle and wireless network data and can facilitate the training and evaluation of AI and ML algorithms for V2X applications [94], [95]. However, there is still a significant gap between theoretical models and practical implementations.

Due to the high mobility nature of the vehicular networks, which leads to increased Doppler spread and very short coherence time intervals, acquiring the timely CSI becomes problematic. In this case, ML algorithms can be of interest as they facilitate improved CSI estimation even for high mobilities [96] (Refer to Section IV-B). By learning from past information and continuously adapting to new data, ML algorithms can help maintain network performance even in scenarios where traditional CSI estimation methods struggle. CSI estimation is the preliminary yet crucial step for proper radio resource management (RRM).

In terms of RRM, 3GPP specifies two distinct approaches for V2X networks: centralized and distributed schemes [91], [97]. The centralized scheme (referred to as Mode 1 in NR-V2X, and Mode 3 in LTE-V2X), requires that all the vehicles be under cellular coverage and involves the BS managing the subchannel assignments. This method requires acquiring the CSI from all the vehicles which incurs a lot of signaling overhead on the BS side. This can become a bottleneck, especially in scenarios with a large number of vehicles or high mobility, where frequent CSI updates are required. In the distributed scheme (referred to as Mode 2 in NR-V2X, and Mode 4 in LTE-V2X) the vehicles themselves manage the channel access through sensing-based semi-persistent scheduling (SPS). This scheme offers more flexibility and reduces latency by avoiding constant BS involvement, but it is more prone to errors and packet collisions due to its distributed nature.

In conformity with 3GPP standardization, various ML-based RRM schemes for vehicular networks have been proposed in the literature which are mostly built on the distributed RRM scheme proposed by 3GPP. Among the proposed ML algorithms, RL has been of the highest interest compared to the other ML algorithms due to its capability for online learning and adaptation which makes it suitable for complex environments like vehicular networks.

A common criticism that is often laid on the ML algorithms is that due to their intrinsic complexity, it is often complicated to know to what extent the learned policies and the trained behavior are close to the optimal solution. The black-box nature of many ML algorithms makes it difficult to interpret their decision-making processes, which raises concerns about their reliability and robustness in real-world applications. These kinds of questions that often fall under the umbrella of trustworthiness in AI have been rarely addressed in the recent literature. Recent studies have begun to explore these concerns, highlighting that while ML-derived policies may not always be optimal, they can still be sufficient to meet the specific requirements of certain applications [98].

VIII. TESTING ASPECTS OF AI/ML-BASED SIGNAL PROCESSING FOR 6G PHY LAYER

A. APPLYING TESTING TO AI/ML IN SIGNAL PROCESSING

The integration of AI/ML models for signal processing tasks into traditional communication systems demands robust testing methodologies to ensure reliable performance under real-world conditions. Combining Model-in-the-Loop (MiL), Software-in-the-Loop (SiL), and Hardware-in-the-Loop (HiL) approaches is essential to validate these models across different system layers, from network management to the physical layer. These methodologies provide a structured framework to optimize and verify AI/ML models at each

stage of development.

Model-in-the-Loop. MiL testing evaluates AI/ML models in a simulated environment designed to mimic the physical layer, including RF impairments, noise, and multipath effects. This phase ensures that models can process wireless signals, make real-time decisions, and adapt to simulated channel conditions. For example, a machine learning-based channel estimator can be tested under conditions such as synthetically generated fading and Doppler effects to ensure accurate channel state estimation before transitioning to software and hardware implementation. MiL enables developers to refine model performance and robustness while exploring edge cases that may occur during deployment.

Software-in-the-Loop. SiL testing focuses on running AI/ML algorithms for tasks like signal detection, channel estimation, or spectrum management within a simulated software environment that would later on run on the hardware itself. This method replicates the behavior of digital communication systems without requiring physical hardware, enabling developers to simulate complex scenarios such as varying channel conditions or interference patterns. For instance, an AI/ML-driven Adaptive Modulation and Coding scheme (AMC) scheme for 6G can be tested using SiL by simulating user mobility patterns and fluctuating SNR. SiL is invaluable for evaluating software integration and initial algorithmic robustness.

Hardware-in-the-Loop. HiL testing integrates AI/ML models with hardware such as antennas, RF frontends, and baseband processors. Algorithms are executed in real-time, interacting directly with real-world signals and hardware components. HiL validates the ability of AI/ML models to handle real-time signal processing, manage hardware-induced latencies, and operate within the constraints of an actual communication system. It ensures compatibility and reliability in practical deployments, addressing performance gaps that may not be evident in simulation. The testing workflow typically starts with MiL to refine models in a simulated environment, followed by SiL to validate software integration, and concludes with HiL to ensure seamless operation within the real hardware. HiL is particularly critical for wireless systems with strict real-time requirements, ensuring that AI/ML models meet latency and throughput constraints. Each stage provides iterative feedback, enabling continuous improvements to both models and their integration, thereby reducing the risk of failure in real-world deployments.

B. AN EXAMPLE FOR HIL: A TESTBED FOR NEURAL RECEIVER ARCHITECTURES

Dynamic and rapidly changing environments in wireless systems pose significant challenges to classic signal processing algorithms, prompting research into AI-native air interfaces for 6G. Academic studies have explored replacing conventional blocks, such as channel estimation and equalization, with ML-based approaches. For example, traditional channel estimation methods, which rely on standardized pilot signal patterns, often fail in high-mobility scenarios due to outdated estimates and interpolation errors. Although increasing pilot density can reduce these errors, it comes at the cost of higher overhead and reduced spectral efficiency. A notable innovation is the integration of channel estimation, equalization, and demapping into a single ML-based 'neural receiver.' Neural receivers have emerged as a promising alternative, capable of learning channel time dynamics and mitigating these challenges [99], [100]. However, many assessments are simulation-based and often benchmarked against non-optimized implementations of traditional methods. These systems are typically evaluated within 5G NR environments, use post-Fast Fourier Transform (FFT) data (after synchronization, removal of the cyclic prefix and performing the FFT) as input, and produce LLR for channel decoding. Current research [101] prioritizes uplink scenarios, with neural receivers deployed at the gNB to provide LLR outputs to the Low-Density Parity-Check Code (LDPC) decoder for the 5G NR Physical Uplink Shared Channel (PUSCH). Nvidia's open source framework, SIONNA [102], has become a standard tool for developing and training neural receiver models due to its flexibility and integration capabilities. The neural receiver implemented in the Rohde&Schwarz testbed [104] was designed using SIONNA and trained offline on synthetic data [101], exemplifying the framework's applicability in academia and industry for advancing ML-based wireless communication systems.

C. FIRST STEP TOWARDS E2E LEARNING AND TESTING

It is expected that ML-based signal processing will first be implemented in network infrastructure, e.g., BSs, in the initial version of a future 6G standard. This is due to the complexity and computational effort required, which increases power consumption. However, the transmit side, e.g. the UE, can also contribute to improve the performance, and thus concepts of learning the constellation shape are explored by academia and industry. These so-called "customized constellation" or "non-uniform constellation" are meant to learn the optimal constellation for a chosen modulation scheme during training, aiming to maximize mutual information, and thus the throughput of the respective communication channel [103]. This step considers the impact of the wireless channel and the nonlinearities and impairments caused by the components of the analog transmitter and receiver, distorting the signal and thus degrading performance. Consequently, based on the learned model and thus constellation shape, the neural receiver testbed allows the user of the testbed to define the amplitude and phase for each constellation point in the IQ domain. The learned constellation shape is typically asymmetric compared to classic quadrature modulation and, while learned by the neural receiver. This asymmetry further eliminates the need for pilot signals, as the receiver interpreted the learned shape as a superimposed pilot sequence that can be used for demodulation. Compared to standard 5G NR, those resource elements carrying pilot signals can now be used for data transmission and making the entire

transmission more spectrally efficient. In different literature [100], a potential increase in spectral efficiency of up to, e.g., 14 percent is suggested (Figure 15). However, the observed performance gain largely depends on the chosen configuration and the pilot pattern used in the conventional 5G NR reference system.

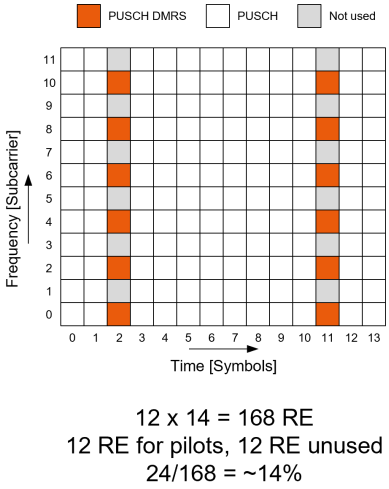


FIGURE 15. Potential performance gains using custom constellation

D. NEURAL RECEIVER ARCHITECTURES HANDLING ANALOG IMPAIRMENTS

Pragmatically, a neural receiver should be trained with slightly impaired data to handle carrier frequency offset (CFO) as tolerated by the 3GPP specification. Figure 16 shows as an example the performance of two neural receiver models with custom constellation, where one was trained without impaired data and the other model with impaired data for a carrier frequency of 2.14 GHz and a CFO of 0.2 parts-per-million (ppm) (428 Hz), where the fine-tuned model still outperforms conventional signal processing algorithms.

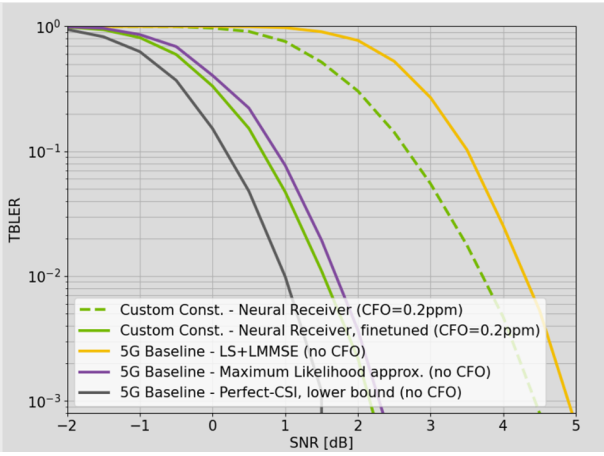


FIGURE 16. Neural receiver trained for CFO of 0.2 ppm

IX. STANDARDIZATION ASPECTS FOR A FUTURE AI-NATIVE AIR INTERFACE IN 6G

The integration of an AI-native air interface in 6G represents a paradigm shift in wireless communication, moving beyond traditional enhancements to rely on AI as a foundational element. Standardization efforts focus on enabling AI-driven functionalities such as autonomous operation, real-time optimization, and continuous learning, emphasizing lifecycle management and testing for interoperability. Strategic approaches include leveraging generative AI for channel modeling, beam management, and semantic communications. These approaches aim to optimize network resource allocation and improve performance in dynamic environments. Key challenges to be addressed include addressing model generalizability, energy efficiency, and ethical concerns. The ongoing transfer from research into standardization highlights the necessity for collaboration among stakeholders to ensure seamless implementation, robust testing frameworks, and adherence to global performance and sustainability goals. These efforts aim to balance the technical, operational, and ethical dimensions of deploying AI in future wireless networks.

X. CONCLUSION AND OUTLOOK

In this paper, we have summarized the findings and proposals of 6G-ANNA concerning an AI/ML-based air interface for 6G, emphasizing energy efficiency. We presented a historical review of developments and applications of AI/ML techniques to physical layer processing and energy efficiency in cellular networks, starting with 5G. We summarized current and upcoming milestones in 3GPP standardization, and clarified how AI/ML is expected to differ between 5G and 6G. We presented AI/ML techniques for RF frontend optimization and showed that these techniques are very promising for digital pre/post-distortion and PAPR reduction, leading to better link performance and energy efficiency gains. We summarized the usefulness and applicability of AI/ML in several signal processing tasks, including channel and interference estimation, SINR prediction, predictive network control, user-centric CF-mMIMO, receiver technologies for MU-MIMO, channel coding and various aspects of V2X networks. Further, we presented testing aspects of AI/ML-based signal processing in the 6G physical layer, which is a subject of practical relevance for the wireless communications industry and which impacts the 3GPP standards. Finally, we described standardization aspects of an upcoming AI native air interface in 6G.

This article highlights research areas that seem very promising for 6G in terms of link performance and energy efficiency gains. The toolbox of AI/ML is versatile and powerful, and its application to the 6G physical layer is promising. Further research is needed on real-time implementation of AI/ML algorithms, and further collaboration among wireless equipment and test equipment vendors is needed to define testing methodologies and to develop robust testing frameworks for AI/ML-based radio technologies.

ACRONYMS

3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
6G	Sixth Generation
6G-ANNA	"6G Access, Network of Networks, Automation and Simplification"
ADMM	alternating direction method of multipliers
ACLR	adjacent channel leakage power ratio
AI	artificial intelligence
AMC	Adaptive Modulation and Coding scheme
AMP	approximate message passing
APSM	adaptive projected subgradient method
AttenQBiLSTM	attention based quantile bidirectional long-short term memory
AWGN	additive white Gaussian noise
BCE	binary cross-entropy
BLER	Block Error Rate
BM	beam management
BMBF	Federal Ministry of Education and Research Germany
BO	backoff
BS	base station
CF-mMIMO	cell-free massive MIMO
CFO	carrier frequency offset
CDL	clustered delay line
CMX500	R&S CMX500 5G One-Box Signaling Tester
CNN	convolutional neural network
CPU	central processing unit
CSI	channel state information
CSI-RS	channel state information reference signals
CQI	channel quality indicator
CMOS	complementary metal–oxide–semiconductor
CCN	concatenated classic and neural
Conv1D	1D convolution layer
DDPM	diffusion-denoising probabilistic model
DDIM	denoising diffusion implicit model
DFT	discrete Fourier transform
DFT-s-OFDM	discrete Fourier transform-spread orthogonal frequency division multiplexing
DM	diffusion model
DNN	deep neural network
DPD	digital pre-distortion
DPOD	digital post-distortion
E2E	end-to-end
EVM	error vector magnitude
FAVIB	forward-aware information bottleneck
FCNN	fully connected neural network
FFT	Fast Fourier Transform
FL	federated learning
FR	frequency range
GAN	generative adversarial network
GMP	generalized memory polynomial
GEMIB	generalized multivariate information bottleneck
gNB	base station
GPU	graphics processing unit
GRU	gated recurrent unit
HiL	Hardware-in-the-Loop
IB	information bottleneck
IDFT	inverse discrete Fourier transform
ILC	iterative learning control
IRC	ideal (error-free) rate-limited channel
IRS	intelligent reflecting surface
KPI	key performance indicator
LLR	log-Likelihood Ratio
LDPC	Low-Density Parity-Check Code
LSTM	long-short term memory
LTE	Long Term Evolution
MAE	mean absolute error
Mbps	megabits per second
MIMO	multiple input multiple output
MI	mutual information
ML	machine learning
mmWave	millimeter-wave
MiL	Model-in-the-Loop
MSE	mean square error
M-to-M	many-to-many
LOS	line-of-sight
NLOS	non-line-of-sight
NR	New Radio
NWDAF	Network Data Analytics Function
NMSE	normalized mean squared error
OAMP	orthogonal approximate message passing
OFDM	orthogonal frequency division multiplexing
OOB	out-of-band
PA	power amplifier
PAPR	peak-to-average power ratio
PCN	private campus network
PHY	physical layer
PMI	precoding matrix indicator
PUSCH	Physical Uplink Shared Channel
ppm	parts-per-million
QBiLSTM	quantile bidirectional long-short term memory
QoS	quality of service
RAP	radio access point
RI	rank indicator
RF	radio frequency
RL	reinforcement learning
RNN	recurrent neural network
ResNet	residual network
RRM	radio resource management
RS	reference signal
SA	sensor-actuator

SDR	Software Defined Radio
SER	symbol error rate
SINR	signal-to-interference-plus-noise ratio
SISO	single input single output
SN	subnetwork
SNR	signal-to-noise ratio
SiL	Software-in-the-Loop
SPS	semi-persistent scheduling
SGCS	squared generalized cosine similarity
SSB	synchronization sequence block
STING	spatially distributed traffic and interference generation
SWD	sliced Wasserstein distance
SPGPR	sparse Gaussian process regression
TSG SA	Technical Specification Group for Service and System Aspects
TSG CT	Technical Specification Group for Core Network and Terminals
TSG RAN	Technical Specification Group for Radio Access Network
UE	user equipment
V2X	vehicle-to-everything
VISPGPR	sparse Gaussian process regression with variational inference
VQ	vector quantization
ICF	iterative clipping and filtering

.

ACKNOWLEDGMENT

The authors would like to thank Torsten Dudda for his support in preparation of this manuscript.

REFERENCES

[1] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6g: Ai empowered wireless networks," IEEE Communications Magazine, vol. 57, no. 8, pp. 84–90, 2019.

[2] H. Yang, A. Alphones, Z. Xiong, D. Niyato, J. Zhao, and K. Wu, "Artificial-intelligence-enabled intelligent 6g networks," IEEE Network, vol. 34, no. 6, pp. 272–280, 2020.

[3] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, "Ai models for green communications towards 6g," IEEE Communications Surveys & Tutorials, vol. 24, no. 1, pp. 210–247, 2022.

[4] International Telecommunication Union (ITU), "Future technology trends of terrestrial imt systems towards 2030 and beyond," Report ITU-R M.2516-2022, ITU Radiocommunication Sector (ITU-R), 2022. Available online: https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-M.2516-2022-PDF-E.pdf.

[5] Q. Rohde&Schwarz, "Title of the video," 2025.

[6] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," IEEE Transactions on Communications, vol. 63, no. 9, pp. 3029–3056, 2015.

[7] 3GPP, "Study on new radio access technology," TR. 38.802, V.14.2.0, 3GPP, Sep. 2017.

[8] M. Q. Khan, A. Gaber, P. Schulz, and G. Fettweis, "Machine learning for millimeter wave and terahertz beam management: A survey and open challenges," IEEE Access, vol. 11, pp. 11880–11902, 2023.

[9] 3GPP, "Study on artificial intelligence (AI)/machine learning (ML) for NR air interface," TR. 38.843, V.2.0.0, 3GPP, Dec. 2023.

[10] M. Q. Khan, A. Gaber, M. Parvini, P. Schulz, and G. Fettweis, "On the generalization of machine learning for mmwave beam prediction," in 2024 19th International Symposium on Wireless Communication Systems (ISWCS), pp. 1–6, 2024.

[11] M. Q. Khan, A. Gaber, M. Parvini, P. Schulz, and G. Fettweis, "A low-complexity machine learning design for mmwave beam prediction," IEEE Wireless Communications Letters, vol. 13, no. 6, pp. 1551–1555, 2024.

[12] M. Alawieh and G. Kontes, "5g positioning advancements with ai/ml," arXiv preprint arXiv:2401.02427, 2023. Fraunhofer Institute for Integrated Circuits IIS, Nuremberg, Germany.

[13] X. Quan, X. Wei, Y. Liu, and Y. Tang, "An out-of-band linearization architecture for pa with strong nonlinearity," IEEE Microwave and Wireless Technology Letters, vol. 34, no. 2, pp. 233–239, 2024.

[14] Y. Wu, G. D. Singh, M. Beikmirza, L. C. N. de Vreede, M. Alavi, and C. Gao, "Opendpd: An open-source end-to-end learning & benchmarking framework for wideband power amplifier modeling and digital pre-distortion," in 2024 IEEE International Symposium on Circuits and Systems (ISCAS), (Singapore), pp. 1–6, IEEE, 2024.

[15] M. A. Bharadwaj, C. Ramesh, and R. V. S. Devi, "Implementation of neural network based digital pre-distorter for power amplifier linearisation," in 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), (Bangalore, India), IEEE, 2024.

[16] Ericsson, "Further elaboration on PA models for NR." R4-165901, 2016. [Online]. Available: http://www.3gpp.org/ftp/tsg_ran/WG4_Radio/TSGR4_80/Docs/R4-165901.zip, 2016.

[17] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6848–6856, 2018.

[18] H. Ochiai, "An analysis of band-limited communication systems from amplifier efficiency and distortion perspective," IEEE Transactions on Communications, vol. 61, no. 4, pp. 1460–1472, 2013.

[19] B. S. Krongold and D. L. Jones, "PAR reduction in OFDM via active constellation extension," IEEE Transactions on broadcasting, vol. 49, no. 3, pp. 258–268, 2003.

[20] J. Armstrong, "Peak-to-average power reduction for OFDM by repeated clipping and frequency domain filtering," Electronics letters, vol. 38, no. 5, pp. 246–247, 2002.

[21] R. L. G. Cavalcante and I. Yamada, "A flexible peak-to-average power ratio reduction scheme for OFDM systems by the adaptive projected subgradient method," IEEE Transactions on Signal Processing, vol. 57, no. 4, pp. 1456–1468, 2009.

[22] J. Fink, R. L. G. Cavalcante, P. Jung, and S. Stańczak, "Extrapolated projection methods for PAPR reduction," in 2018 26th European Signal Processing Conference (EUSIPCO), pp. 1810–1814, IEEE, 2018.

[23] J. Fink, Fixed point algorithms and superiorization in communication systems. PhD thesis, TU Berlin, 2022.

[24] A. Aggarwal and T. H. Meng, "Minimizing the peak-to-average power ratio of OFDM signals via convex optimization," in Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE, vol. 4, pp. 2385–2389, IEEE, 2003.

[25] J. J. Brune, J. Fink, and S. Stańczak, "Deep unfolded multi-group multicast beamforming," in 2022 56th Asilomar Conference on Signals, Systems, and Computers, pp. 1398–1402, IEEE, 2022.

[26] J. Fink, R. L. Cavalcante, Z. Utkovski, and S. Stańczak, "Deep-unfolded adaptive projected subgradient method for mimo detection," in ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5, IEEE, 2023.

[27] I. F. Akyildiz, A. Kak, and S. Nie, "6g and beyond: The future of wireless communications systems," IEEE Access, vol. 8, pp. 133995–134030, 2020.

[28] W. Kim, J. Jung, S. Park, and H. Park, "Towards deep learning-aided wireless channel estimation and channel state information feedback for 6g," Journal of Communications and Networks, vol. 25, no. 1, pp. 61–75, 2023.

[29] S. Joodaki, K. Turbic, A. Sezgin, and H. Gacanin, "Deep learning-based channel estimation in high-speed wireless systems with imperfect frame synchronization," in 2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), (Canada), 2023.

[30] Özlem Tugfe Demir and E. Björnson, "Channel estimation under hardware impairments: Bayesian methods versus deep learning," in 2019 16th International Symposium on Wireless Communication Systems (ISWCS), IEEE, 2019.

[31] S. Liu et al., "Deep denoising neural network assisted compressive channel estimation for mmwave intelligent reflecting surfaces.," IEEE

Transactions on Vehicular Technology, vol. 69, no. 8, pp. 9223–9228, 2020.

[32] S. Gao, J. Li, W. Xu, and X. You, "Lightweight deep learning based channel estimation for extremely large-scale massive mimo systems," IEEE Transactions on Vehicular Technology, 2024.

[33] J. Kim et al., "Parametric sparse channel estimation using long short-term memory for mmwave massive mimo systems.," in ICC 2022-IEEE International Conference on Communications., IEEE, 2022.

[34] Y. Zhang and A. Alkhateeb, "Zone-specific csi feedback for massive mimo: A situation-aware deep learning approach," arXiv preprint arXiv:2401.07451, 2024.

[35] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive mimo csi feedback," IEEE Wireless Communications Letters, vol. 7, no. 5, pp. 748–751, 2018.

[36] M. S. Oh et al., "Channel estimation via successive denoising in mimo ofdm systems: A reinforcement learning approach.," in ICC 2021-IEEE International Conference on Communications., IEEE, 2021.

[37] W. Wan, J. Mo, and O. Simeone, "Deep plug-and-play prior for multitask channel reconstruction in massive mimo systems," IEEE Transactions on Communications, 2024.

[38] Y. Dong, H. Wang, and Y.-D. Yao., "Channel estimation for one-bit multiuser massive mimo using conditional gan.," IEEE Communications Letters, vol. 25, no. 3, pp. 854–858, 2020.

[39] M. Kim, R. Fritschek, and R. F. Schaefer, "Learning end-to-end channel coding with diffusion models," in WSA & SCC 2023; 26th International ITG Workshop on Smart Antennas and 13th Conference on Systems, Communications, and Coding, pp. 1–6, 2023.

[40] M. Kim, R. Fritschek, and R. F. Schaefer, "Robust generation of channel distributions with diffusion models," in ICC 2024 - IEEE International Conference on Communications, pp. 330–335, 2024.

[41] G. Berardinelli, P. Baracca, R. O. Adeogun, S. R. Khosravirad, F. Schaich, K. Upadhya, D. Li, T. Tao, H. Viswanathan, and P. Mogensen, "Extreme communication in 6g: Vision and challenges for 'in-x' subnetworks," IEEE Open Journal of the Communications Society, vol. 2, pp. 2516–2535, 2021.

[42] P. Gautam, C. Bockelmann, and A. Dekorsy, "Interference prediction in unconnected in-x mobile 6g subnetworks using a data-driven approach," in 2024 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 2046–2052, 2024.

[43] C. Jayawardhana, T. Sivalingam, N. H. Mahmood, et al., "Predictive resource allocation for URLLC using empirical mode decomposition," in 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), pp. 174–179, IEEE, 2023.

[44] L. Wei, Z. Liu, C. Zhang, L. Zhang, Y. Wang, and Y. Huang, "Joint model and data-driven two-stage uplink interference prediction in urllc scenarios," in 2024 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6, IEEE, 2024.

[45] M. F. Marzban, W. Nam, T. Luo, A. Kannan, and T. Yoo, "Deep learning for probabilistic interference predictions in mmwave networks," in 2023 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 42–47, IEEE, 2023.

[46] P. Gautam, M. Vakilifard, C. Bockelmann, and A. Dekorsy, "Cooperative interference estimation using lstm-based federated learning for in-x subnetworks," in GLOBECOM 2023 - 2023 IEEE Global Communications Conference, pp. 1338–1344, 2023.

[47] C. K. Williams and C. E. Rasmussen, Gaussian processes for machine learning, vol. 2. MIT press Cambridge, MA, 2006.

[48] M. Titsias, "Variational learning of inducing variables in sparse gaussian processes," in Artificial intelligence and statistics, pp. 567–574, PMLR, 2009.

[49] P. Gautam, C. Bockelman, and A. Dekorsy, "Probabilistic interference prediction for dynamic 6g in-x sub-networks," IEEE Open Journal of the Communications Society, pp. 1–1, 2025.

[50] L. You, "Weighted sum-rate maximization with causal inference for latent interference estimation," in ICC 2023-IEEE International Conference on Communications, pp. 1–7, IEEE, 2023.

[51] S. B. Mallikarjun, S. C. Kusumapani, N. P. Kuruvatti, B. G. Bhat, and H. D. Schotten, "Machine learning based SINR prediction in private campus networks," in 2023 IEEE 97th Vehicular Technology Conference (VTC2023-Spring), pp. 1–6, IEEE, 2023.

[52] I. Orikumhi, J. Kang, H. Jwa, J.-H. Na, and S. Kim, "SINR maximization beam selection for mmwave beamspace MIMO systems," IEEE Access, vol. 8, pp. 185688–185697, 2020.

[53] Q. Wu, Y. Zhang, Z. Yang, and M. R. Shikh-Bahaei, "Deep channel prediction-based energy-efficient intelligent reflecting surface-aided terahertz communications," IEEE Transactions on Wireless Communications, vol. 23, no. 4, pp. 2946–2960, 2024.

[54] Y. Guo, C. Hu, T. Peng, H. Wang, and X. Guo, "Regression-based uplink interference identification and SINR prediction for 5G ultra-dense network," in ICC 2020-2020 IEEE International Conference on Communications (ICC), pp. 1–6, IEEE, 2020.

[55] O. Günlü, R. Fritschek, and R. F. Schaefer, "Concatenated classic and neural (ccn) codes: Concatenatedae," in 2023 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1–6, 2023.

[56] S. Verdú, "Computational complexity of optimum multiuser detection," Algorithmica, vol. 4, no. 1, pp. 303–312, 1989.

[57] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," IEEE Commun. Surv. Tutor., vol. 17, no. 4, pp. 1941–1988, 2015.

[58] E. Viterbo and J. Boutros, "A universal lattice code decoder for fading channels," IEEE Trans. Inf. Theory, vol. 45, no. 5, pp. 1639–1642, 1999.

[59] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in 2015 IEEE Int. Symp. Inf. Theory - Proc. (ISIT), pp. 1227–1231, IEEE, 2015.

[60] M. Khani, M. Alizadeh, J. Hoydis, and P. Fleming, "Adaptive neural signal detection for massive MIMO," IEEE Trans. Wirel. Commun., vol. 19, no. 8, pp. 5635–5648, 2020.

[61] J. R. Hershey, J. L. Roux, and F. Weninger, "Deep unfolding: Model-based inspiration of novel deep architectures," arXiv preprint arXiv:1409.2574, 2014.

[62] H. Liu, M.-C. Yue, A. M.-C. So, and W.-K. Ma, "A discrete first-order method for large-scale MIMO detection with provable guarantees," in 2017 IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), pp. 1–5, IEEE, 2017.

[63] S. Takabe, M. Imanishi, T. Wadayama, R. Hayakawa, and K. Hayashi, "Trainable projected gradient detector for massive overloaded MIMO channels: Data-driven tuning approach," IEEE Access, vol. 7, pp. 93326–93338, 2019.

[64] N. Samuel, T. Diskin, and A. Wiesel, "Learning to detect," EEE Trans. Signal Process., vol. 67, no. 10, pp. 2554–2564, 2019.

[65] M.-W. Un, M. Shao, W.-K. Ma, and P. Ching, "Deep MIMO detection using ADMM unfolding," in 2019 IEEE Data Science Workshop (DSW), pp. 333–337, IEEE, 2019.

[66] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "A model-driven deep learning network for MIMO detection," in 2018 IEEE Glob. Conf. Signal Inf. Process. (GlobalSIP), pp. 584–588, IEEE, 2018.

[67] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Model-driven deep learning for MIMO detection," EEE Trans. Signal Process., vol. 68, pp. 1702–1715, 2020.

[68] J. Fink, R. L. Cavalcante, Z. Utkovski, and S. Stańczak, "A set-theoretic approach to MIMO detection," in Proc. - IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 5328–5332, IEEE, 2022.

[69] J. Fink, R. L. G. Cavalcante, and S. Stańczak, "Superiorized adaptive projected subgradient method with application to mimo detection," IEEE Transactions on Signal Processing, vol. 71, pp. 1350–1362, 2023.

[70] Y. Censor, "Weak and strong superiorization: between feasibility-seeking and minimization," Analele ştiinţifice ale Universităţii" Ovidius" Constanţa. Seria Matematică, vol. 23, no. 3, pp. 41–54, 2015.

[71] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," Numer. Funct. Anal. Optim., 2005.

[72] S. Hassanpour, A. Danaee, D. Wübben, and A. Dekorsy, "Multi-Source Distributed Data Compression Based on Information Bottleneck Principle," IEEE Open Journal of the Communications Society, July 2024.

[73] A. Danaee, S. Hassanpour, D. Wübben, and A. Dekorsy, "Relevance-Based Information Processing for Fronthaul Rate Reduction in Cell-Free MIMO Systems," in 19th International Symposium on Wireless Communication Systems (ISWCS), (Rio de Janeiro, Brazil), July 2024.

[74] N. Tishby, F. C. Pereira, and W. Bialek, "The Information Bottleneck Method," in 37th Annual Allerton Conference on Communication, Control, and Computing, (Monticello, IL, USA), September 1999.

[75] Z. Goldfeld and Y. Polyanskiy, "The Information Bottleneck Problem and Its Applications in Machine Learning," IEEE Journal on Selected Areas in Information Theory, vol. 1, pp. 19–38, April 2020.

[76] A. Danaee, S. Hassanpour, D. Wübben, and A. Dekorsy, "Relevance-Based Multi-User Data Compression for Fronthaul Rate Reduction in Cell-Free Massive MIMO Systems," in Submitted, 2025.

[77] A. K. Jain, "Data Clustering: 50 Years Beyond K-Means," Pattern Recognition Letters, vol. 31, pp. 651–666, June 2010.

[78] M. Hummert, S. Hassanpour, D. Wübben, and A. Dekorsy, "Deep FAVIB: Deep Learning-Based Forward-Aware Quantization via Information Bottleneck Method," in ICC 2024-IEEE International Conference on Communications, (Denver, CO, USA), pp. 1885–1890, August 2024.

[79] M. Hummert, S. Hassanpour, D. Wübben, and A. Dekorsy, "Deep Learning-Based Forward-Aware Quantization for Satellite-Aided Communications via Information Bottleneck Method," in Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), (Antwerp, Belgium), June 2024.

[80] R. Ooi, R. Diab, L. Miretti, R. L. G. Cavalcante, and S. Stańczak, "Joint power control, beamforming, and sleep-mode selection for energy-efficient cell-free networks using surrogate machine learning models," in GLOBECOM 2024 - 2024 IEEE Global Communications Conference, pp. 4956–4962, 2024.

[81] M. Danger et al., "Performance evaluation of IRS-enhanced mmWave connectivity for 6G industrial networks," in 2024 IEEE International Symposium on Measurements and Networking (M&N), 2024.

[82] K.-I. Šabanović, C. Arendt, S. Fricke, M. Geis, S. Böcker, and C. Wietfeld, "AI-based anomaly detection for industrial 5G networks by distributed SDR measurements," in 2024 IEEE International Symposium on Measurements and Networking (M&N), 2024.

[83] M. Danger, S. Häger, K. Heimann, S. Böcker, and C. Wietfeld, "Empowering 6G industrial indoor networks: Hands-on evaluation of IRS-enabled multi-user mmWave connectivity," in 2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), 2024.

[84] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 9351 of LNCS, pp. 234–241, Springer, 2015.

[85] M. Geis, C. Bektas, S. Böcker, and C. Wietfeld, "AI-driven planning of private networks for shared operator models," in IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN), 2024.

[86] R.-J. Reifert, S. Roth, A. A. Ahmad, and A. Sezgin, "Energy efficiency in rate-splitting multiple access with mixed criticality," in Proc. IEEE ICC Workshops, pp. 681–686, 2022.

[87] H. Dahrouj, R. Alghamdi, H. Alwazani, S. Bahanshal, A. A. Ahmad, A. Faisal, R. Shalabi, R. Alhadrami, A. Subasi, M. T. Al-Nory, O. Kittaneh, and J. S. Shamma, "An overview of machine learning-based techniques for solving optimization problems in communications and signal processing," IEEE Access, vol. 9, pp. 74908–74938, 2021.

[88] A. G. Onalan and S. Coleri, "Deep learning based low complexity relay selection for wireless powered cooperative communication networks," in Proc. BalkanCom, pp. 1–6, 2023.

[89] W. Cui and W. Yu, "Uncertainty injection: A deep learning method for robust optimization," IEEE Transactions on Wireless Communications, vol. 22, no. 11, pp. 7201–7213, 2023.

[90] R.-J. Reifert, H. Dahrouj, A. A. Ahmad, H. Gacanin, and A. Sezgin, "Robust communication and computation using deep learning via joint uncertainty injection," in Proc. ISWCS, pp. 1–6, 2024.

[91] "Study on NR vehicle-to-everything (V2X) (Release 16)," TR 38.885, 3GPP, Mar. 2019.

[92] "Overall description of radio access network (RAN) aspects for vehicle-to-everything (V2X) based on LTE and NR (Release 17)," TR 37.985, 3GPP, 2022.

[93] M. H. C. Garcia, A. Molina-Galan, M. Boban, J. Gozalvez, B. Coll-Perales, T. Şahin, and A. Kousaridas, "A tutorial on 5G NR V2X communications," IEEE Communications Surveys & Tutorials, vol. 23, pp. 1972–2026, Thirdquarter 2021.

[94] R. Hernangómez, P. Geuer, A. Palaios, D. Schäufele, C. Watermann, K. Taleb-Bouhemadi, M. Parvini, A. Krause, S. Partani, C. Vielhaus, et al., "Berlin V2X: A machine learning dataset from multiple vehicles and radio access technologies," in Proc. Vehicular Technology Conference (VTC2023-Spring), (Florence, Italy), pp. 1–5, June 2023.

[95] R. Hernangómez, A. Palaios, C. Watermann, D. Schäufele, P. Geuer, R. Ismayilov, M. Parvini, A. Krause, M. Kasparick, T. Neugebauer, et al., "Toward an AI-enabled connected industry: AGV communication and sensor measurement datasets," IEEE Communications Magazine, vol. 62, pp. 90–95, Apr. 2024.

[96] J. Joo, M. C. Park, D. S. Han, and V. Pejovic, "Deep learning-based channel prediction in realistic vehicular communications," IEEE Access, vol. 7, pp. 27846–27858, 2019.

[97] "Overall description of radio access network (RAN) aspects for vehicle-to-everything (V2X) based on LTE and NR (Release 17)," TR 37.985, 3GPP, 2022.

[98] M. Parvini, P. Schulz, and G. Fettweis, "Resource allocation in V2X networks: From classical optimization to machine learning-based solutions," IEEE Open Journal of the Communications Society, 2024.

[99] S. Zheng, S. Chen, and X. Yang, "Deepreceiver: A deep learning-based intelligent receiver for wireless communications in the physical layer," IEEE Transactions on Cognitive Communications and Networking, vol. 7, no. 1, pp. 5–20, 2021.

[100] F. A. Aoudia and J. Hoydis, "End-to-end learning for OFDM: from neural receivers to pilotless communication," CoRR, vol. abs/2009.05261, 2020.

[101] S. Cammerer, F. A. Aoudia, J. Hoydis, A. Oeldemann, A. Roessler, T. Mayer, and A. Keller, "A neural receiver for 5g nr multi-user mimo," arXiv preprint arXiv:2312.02601, 2023. Available online at https://arxiv.org/abs/2312.02601.

[102] NVIDIA Corporation, "Sionna: Nvidia's open-source library for next-generation wireless systems." https://developer.nvidia.com/sionna, 2024. Accessed: 2024-12-15.

[103] M. Stark, F. Ait Aoudia, and J. Hoydis, "Joint learning of geometric and probabilistic constellation shaping," in 2019 IEEE Globecom Workshops (GC Wkshps), pp. 1–6, 2019.

ANDREAS ROESSLER is a technology manager at Rohde & Schwarz. His main focus is on 3GPP's 5G NR standard and advancing 6G research topics. His responsibilities include strategic marketing and product portfolio development for the entire value chain offered by the Rohde & Schwarz test and measurement division. By following industry trends and the standardization process for cellular communication standards very carefully, he gained more than 15 years of experience in the mobile industry and wireless technologies. Andreas holds an MSc from Otto-von-Guericke University in Magdeburg, Germany in electrical engineering, with a focus on wireless communication.

MIGUEL M. LOPEZ received the Ph.D. degree in mathematics from The University of British Columbia, Canada, in 1996. In 1998, he joined Ericsson, in Stockholm, Sweden. He has been involved in signal processing, algorithm development, standardization of cellular technologies and IEEE 802.11 standardization. He currently holds the position of Expert in physical layer design at Ericsson Research, Ericsson GmbH, Herzogenrath, Germany. His research interests include signal processing for wireless communications, transmitter algorithms, receiver algorithms, and physical layer design.

ROUAA DIAB received a B.Sc. degree in Computer Engineering from the American University of Kuwait in 2017, and a M.Sc. degree in Communications and Multimedia Engineering from Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany in 2022. She is currently pursuing a doctoral degree at the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany.

RENATO LUÍS GARRIDO CAVALCANTE received the Electronics Engineering degree from the Instituto Tecnológico de Aeronáutica, São José dos Campos, Brazil, in 2002, and the M.E. and Ph.D. degrees in Communications and Integrated Systems from the Tokyo Institute of Technology, Tokyo, Japan, in 2006 and 2008, respectively. He is currently a Group Leader with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany, and a Lecturer with the Technical University of Berlin. He held appointments as a Research Fellow with the University of Southampton, Southampton, U.K., and as a Research Associate with the University of Edinburgh, Edinburgh, U.K. His interests include signal processing for distributed systems, multiagent systems, convex analysis, machine learning, and wireless communications.

Dr. Cavalcante was the recipient of the Excellent Paper Award from the Institute of Electronics, Information and Communication Engineers of Japan in 2006 and the IEEE Signal Processing Society (Japan Chapter) Student Paper Award in 2008. He is currently serving as a Senior Area Editor for the IEEE Transactions on Signal Processing.

ALI ABDI Ali Abdi completed his Bachelor of Science in Communications Engineering at Shahid Beheshti University, Tehran, in 2015, followed by a Master of Science in Communication Systems Engineering from Sharif University of Technology in 2017. In 2023 he started his PhD at RWTH Aachen University where he also serves as a Research Assistant. Ali's interests span channel estimation for 6G, machine learning in wireless PHY, mmWave and massive MIMO, and Integrated Sensing and Communication (ISAC) for advanced wireless networks.

CHRISTIAN ARENDT received his M.Sc. degree in Electrical Engineering and Information Technology from TU Dortmund University, Germany, in 2018. He is currently pursuing his Dr.-Ing. degree as a research assistant at the Communication Networks Institute (CNI), Faculty of Electrical Engineering and Information Technology, TU Dortmund University. His research focuses on the performance and scalability evaluation of private 5G/6G and Wi-Fi networks, and resource-efficient communication with a focus on industrial environments.

CARSTEN BOCKELMANN (Member, IEEE) received Dipl.-Ing. and Ph.D. degrees in electrical engineering from the University of Bremen, Germany, in 2006 and 2012, respectively. Since 2012, he has been a Senior Research Group Leader with the University of Bremen, coordinating research activities regarding the application of various frameworks like compressive sensing, DMD, Event-based sampling, and machine learning to communication problems. His research interests include 6G, massive machine-type communication, ultra-reliable low latency communications and industry 4.0 as well as health communications.

ALIREZA DANAEE (Member, IEEE) received the bachelor's degree in electronic engineering from the University of Kurdistan, Sanandaj, Iran, in 2008, the master's degree in electrical engineering from Shahid Rajaee Teacher Training University, Tehran, Iran, in 2013, and the Ph.D. degree in electrical engineering, in the area of signal processing, automation and robotics from the Pontifical Catholic University of Rio de Janeiro, Brazil, in 2022. He is currently a Postdoctoral Researcher with the Department of Communications Engineering, University of Bremen, Germany. His current research interests center around statistical signal processing, information theory, and machine learning with applications to wireless communications and distributed data processing.

MARCO DANGER received his M.Sc. degree in industrial engineering from TU Dortmund University, Dortmund, Germany in 2023. He is currently pursuing the Dr.-Ing. degree and is a Research Assistant at the Communication Networks Institute, Faculty of Electrical Engineering and Information Technology, TU Dortmund University. His research interests cover wireless communications in cellular networks of 5G and beyond with a particular focus on the millimeter-wave domain.

ARMIN DEKORSY (Senior Member, IEEE) is a professor at the University of Bremen, where he is the director of the Gauss-Olbers Space Technology Transfer Center and heads the Department of Communications Engineering. With over eleven years of industry experience, including distinguished research positions such as DMTS at Bell Labs and Research Coordinator Europe at Qualcomm, he has actively participated in more than 65 international research projects, with leadership roles in 17 of them. He is a Senior Member of the IEEE Communications and Signal Processing Society and a member of the VDE/ITG Expert Committee on Information and System Theory. He co-authored the textbook 'Nachrichtenübertragung, Release 6, Springer Vieweg,' which is a bestseller in the field of communication technologies in German-speaking countries. His research focuses on signal processing and wireless communications for 5G/6G, industrial radio, and 3D networks.

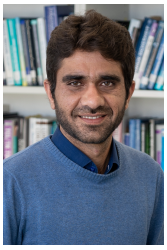
ALI EL HUSSEINI received an Engineering Diploma in Telecommunications from Antonine University, Lebanon, in 2014, followed by an M.Sc. in Signal Processing from the University of Western Brittany (UBO), Brest, France, in 2015. In 2019, he received his Ph.D. in Signal Processing and Telecommunications from the University of Lille, France. He is currently a senior researcher at Ericsson Research, Ericsson GmbH, Herzogenrath, Germany, where he focuses on AI/ML techniques for 5G NR and 6G RAN. His research interests include machine learning, signal processing and wireless communication systems.

GERHARD FETTWEIS (Fellow, IEEE) earned a Ph.D. under H. Meyr at RWTH Aachen in 1990. After postdoctoral work at IBM Research, San Jose, California, he joined TCSI, Berkeley, United States. Since 1994, he has been Vodafone Chair Professor at Technische Universität Dresden. Since 2018 he has also headed the new Barkhausen Institute. In 2019 he was elected into the DFG Senate (German Research Foundation). He researches wireless transmission and chip design, coordinates 5G++Lab Germany, has spun out 17 tech and 3 non-tech startups, and is a member of the German Academy of Sciences (Leopoldina), and German Academy of Engineering (acatech).

JOCHEN FINK received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from the Technical University of Kaiserslautern, Germany, in 2014 and 2016, respectively, and the Dr.-Ing degree (summa cum laude) from the Technical University of Berlin in 2022. He is currently a group leader with the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, Berlin, Germany. His interests include signal processing, wireless communications, fixed point methods, and machine learning.

RICK FRITSCHEK received his B.Sc. degree in Electrical Engineering from Furtwangen University, Germany, in 2010 and his M.Sc. degree in Electrical Engineering from Technische Universität (TU) Berlin, Germany, in 2012. He earned his Ph.D. under the supervision of Prof. Dr.-Ing. Gerhard Wunder at the TU Berlin in 2018. He is currently a Postdoctoral Researcher at the Chair for Information Theory and Machine Learning at the TU Dresden. Previously, he held the position of a Postdoctoral Researcher at the Universität Siegen and as the Principal Investigator of the DFG-funded project "Machine Learning for Physical Layer Security" at the Freie Universität Berlin. His research interests include information theory, machine learning, wireless communications, and physical layer security.

PRAMESH GAUTAM (Graduate Student Member, IEEE) obtained his B.E. in Electrical and Electronics Engineering (Communication) from Kathmandu University and an M.Sc. in Communication and Information Technology from the University of Bremen. Currently, he is pursuing his Ph.D. in Electrical Engineering at the University of Bremen. He actively contributes to the academic community by reviewing papers for prestigious conferences including ICC, Globecom, IEEE ICMLCN, and many more. His research interests include Radio Resource Management, Interference Estimation and Prediction, HRLLC, Machine Learning for Wireless Communication, and low-complexity estimation algorithm design.



MUHAMMAD QURRATULAIN KHAN received the M.Sc. degree in Electrical Engineering from University of Engineering and Technology, Taxila, Pakistan, in 2019. From 2019 to 2022, he was an Algorithm Lead with the Center for Advanced Research in Engineering (CARE) Pvt. Ltd. Islamabad, Pakistan, where he worked on the design of multi-carrier and multi-antenna waveforms suitable for tactical communications. From 2022 to 2024, he was an ESR at National Instruments, Dresden, Germany, in the framework of Marie Skodowska-Curie ITN SEMANTIC project. Since 2024, he is with the Vodafone Chair for Mobile Communications Systems Technische Universität Dresden, Germany.

SACHINKUMAR B. MALLIKARJUN received his B.Eng degree from RV College of Engineering, India, in 2014 and an MSc. degree in Computer Science from the University of Kaiserslautern, Germany, in 2019, where he is currently pursuing a Ph.D. with the Institute for Wireless Communications and Navigation, Kaiserslautern. In 2019, he joined the Institute for Wireless Communications and Navigation as a Researcher. He has been involved in research and industrial projects like 5G Modellregion, Open6GHub, 6G Campus, and 6G ANNA. His research interests include Cellular Network Security, Physical Layer Security in 6G subnetworks, 5G, and O-RAN.

AHMAD NIMR received his Ph.D. degree in Electrical Engineering from TU Dresden, Germany, in 2021, and M.Sc. in Communications and Signal Processing from TU Ilmenau, Germany, in 2014. From 2005 to 2011, he pursued a carrier in software and hardware development. Since then, he has been actively involved in the research, co-authoring several book chapters and papers and presenting at numerous workshops. In 2020, he started leading a research group at the Vodafone Chair Mobile Communications, TU Dresden. His research focuses on radio design and signal processing techniques for communications and sensing, from theory to implementation. He is also a founding member of IEEE ComSoc ETI on Integrated Sensing and Communications. Additionally, he has contributed to several EU and German-funded projects, including the German 6G light house project 6G-ANNA and the European flagship projects Hexa-X and Hexa-X II.

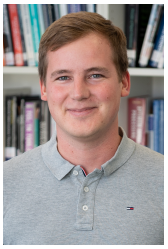
MOHAMMAD PARVINI (Member, IEEE) received the B.Sc. degree in electrical engineering from the Amirkabir University of Technology, Tehran, Iran, in 2019, and the M.Sc. degree in communication systems from Tarbiat Modares University, Tehran, in 2021. He is currently pursuing Ph.D. degree in communication systems at the Technical University of Dresden. His research includes wireless communication systems, signal processing, optimization, and machine learning.

ERMA PERENDA Erma Perenda obtained her Ph.D. with a Networked Systems major from the Department of Electrical Engineering (ESAT), KU Leuven, Belgium, in 2023. Before joining KU Leuven, she worked as a Senior Wi-Fi Software Engineer at the Nokia Digital Home product group at Nokia Shanghai Bell. Since April 2023, she has been working as a Chief Engineer at the Digital Signal Processing chair at RWTH Aachen University. Her research interests include spectrum measurements and management, applied machine learning, wireless network optimization, and architecture design.

ROBERT-JERON REIFERT (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degree in Electrical Engineering and Information Technology from Ruhr University Bochum, Germany, in 2019 and 2021, respectively. He is one of the recipients of the Association for Electrical, Electronic and Information Technologies (VDE) Rhein-Ruhr graduate student award 2021. He is currently pursuing the Ph.D. degree with the Institute of Digital Communication Systems, Ruhr University Bochum, Germany. His research interests include wireless communication systems, mixed criticality, and resilience in 6G communication networks and beyond.

RAFAEL F. SCHAEFER (Senior Member, IEEE) received the Dipl.-Ing. degree in electrical engineering and computer science from the Technische Universität Berlin, Germany, in 2007, and the Dr.-Ing. degree in electrical engineering from the Technische Universität München, Germany, in 2012. He is a Professor and the Head of the Chair of Information Theory and Machine Learning, Technische Universität Dresden, Germany. Since 2023, he is also leading the Wireless Connectivity and Sensing Group at the Barkhausen Institut, Dresden, Germany. From 2013 to 2015, he was a Post-Doctoral Research Fellow with Princeton University. From 2015 to 2020, he was an Assistant Professor with the Technische Universität Berlin. From 2021 to 2022, he was a Professor with the Universität Siegen. Among his publications is the book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017). He was a recipient of the VDE Johann-Philipp-Reis Award in 2013. He received the Joy Thomas Tutorial Paper Award in 2025 and Best Paper Awards from the German Information Technology Society (ITG-Preis) in 2016 and IEEE Global Communications Conference in 2023.

HANS D. SCHOTTEN (S'93-M'97) received the Ph.D. degree from the RWTH Aachen University of Technology, Germany, in 1997. From 1999 to 2003, he worked for Ericsson. From 2003 to 2007, he worked for Qualcomm. He became manager of a R&D group, Research Coordinator for Qualcomm Europe, and Director for Technical Standards. In 2007, he accepted the offer to become a full professor at the University of Kaiserslautern (RPTU). In 2012, he became the scientific director of the German Research Center for Artificial Intelligence (DFKI) and head of the Department for Intelligent Networks. Prof. Schotten served as dean of the Department of Electrical Engineering at the RPTU from 2013 until 2017. Since 2018, he has been chairman of the German Society for Information Technology and a member of the Supervisory Board of VDE.



PHILIPP SCHULZ (Member, IEEE) received the M.Sc. degree in Mathematics and the Ph.D. (Dr.-Ing.) degree in Electrical Engineering from Technische Universität Dresden, Germany, in 2014 and 2020, respectively. He was a Research Assistant with Technische Universität Dresden in the field of numerical mathematics, modeling, and simulation, where he joined the Vodafone Chair for Mobile Communications Systems in 2015 and became a member of the System-Level Group. His research there focused on flow-level modeling and the application of queuing theory on communications systems with respect to ultra-reliable low-latency communications. After more than one year at the Barkhausen Institut, Dresden, Germany, where he studied rateless codes in the context of multi-connectivity, he is currently a research group leader at the Vodafone Chair and focuses on resilience of wireless communications systems.

AYDIN SEZGIN (Senior Member, IEEE) received the Dr. Ing. (Ph.D.) degree in electrical engineering from TU Berlin, in 2005. From 2001 to 2006, he was with the Heinrich-Hertz-Institut, Berlin. From 2006 to 2008, he held a postdoctoral position, and was also a lecturer with the Information Systems Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA, USA. From 2008 to 2009, he held a postdoctoral position with the Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, USA. From 2009 to 2011, he was the Head of the Emmy-Noether-Research Group on Wireless Networks, Ulm University. In 2011, he joined TU Darmstadt, Germany, as a professor. He is currently a professor with the Ruhr University Bochum, Germany. He has published several book chapters, more than 70 journals and 200 conference papers in these topics. Aydin is a winner of the ITG-Sponsorship Award, in 2006. He was a first recipient of the prestigious Emmy-Noether Grant by the German Research Foundation in communication engineering, in 2009. He has coauthored papers that received the Best Poster Award at the IEEE Communication Theory Workshop, in 2011, the Best Paper Award at ICCSPA, in 2015, at ICC, in 2019, and at ISAP, in 2023.

SLAWOMIR STAŃCZAK (Senior Member, IEEE) studied electrical engineering with specialization in control theory from the Wroclaw University of Technology and the Technical University of Berlin (TU Berlin). He received the Dipl.-Ing. and Dr.-Ing. (summa cum laude) degrees in electrical engineering from TU Berlin in 1998 and 2003, respectively, and the Habilitation (Venia Legendi) degree in 2006. Since 2015, he has been a Full Professor of network information theory with TU Berlin and the Head of the Wireless Communications and Networks Department, Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute (HHI). He has been involved in research and development activities in wireless communications since 1997. From 2004 to 2007, he was a Visiting Professor with RWTH Aachen University. In 2008, he was a Visiting Scientist with Stanford University, Stanford, CA, USA. He is the co-author of two books and more than 200 peer-reviewed journal articles and conference papers in the areas of information theory, wireless communications, signal processing, and machine learning. Since 2021, he has been the Coordinator of the 6G Research and Innovation Cluster and the Flagship Project CampusOS. He is also the Project Coordinator of xG-Incubator since November 2023. The project is part of StartUpConnect, which supports start-ups that are working on innovative ideas in the field of 5G/6G communication technologies from radio access to core and fiber optic transport networks. He received research grants from German Research Foundation and the Best Paper Award from the German Communication Engineering Society in 2014. He was the Co-Chair of the 14th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2013). From 2017 to 2020, he was the Chair of the ITU-T Focus Group on Machine Learning for Future Networks, including 5G. Since 2020, he has been the Chairperson of the 5G BERLIN Association and an Editor of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS for the Special Issue on Machine Learning in Communications and Networks. From 2009 to 2011, he was an Associate Editor of European Transactions on Telecommunications (information theory) and an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2012 to 2015.

CHRISTIAN WIETFELD (SM'12) is a Full Professor for Communication Networks at TU Dortmund University. He has been leading and contributing to large-scale research projects on Internet-based mobile communication systems at RWTH Aachen (1992-97, DSRC), Siemens AG (1997-2005, Wireless Internet) and TU Dortmund (since 2005, 4G/5G/6G). Within the 6GEM research hub, in collaboration with the 6G-ANNA project, his current research interests include the design and performance evaluation of AI-based, open 6G networks for cyber-physical systems in energy, transport, robotics, and emergency response, with a particular focus on energy-efficient networking. He has received 15 IEEE Best Paper Awards, is an Associate Editor of the IEEE Wireless Communication Magazine and the recipient of an ITU-T Outstanding Contribution Award for his work on the standardization of next generation mobile network architectures. Christian is an active member of IEEE, VDE/ITG and co-founder of 3 start-ups.

DIRK WÜBBEN (Senior Member, IEEE) is a senior research group leader and lecturer at the Department of Communications Engineering, University of Bremen, Germany. He received the Dipl.-Ing. (FH) degree in electrical engineering from the University of Applied Science Münster, Germany, in 1998, and the Dipl.-Ing. (Uni) degree and the Dr.-Ing. degree in electrical engineering from the University of Bremen, Germany in 2000 and 2005, respectively. He has been active in several national and international projects (e.g., EU FP7 iJOIN, EU FP7 METIS, BMBF TACNET 4.0, FunKI, Open6GHub, 6G-Platform Germany, 6G-TakeOff, 6G-ANNA, ESA AIComS). His research interests include wireless communications, signal processing, multiple antenna systems, cooperative communication systems, channel coding, information theory, and machine learning. He has published more than 160 papers in international journals and conference proceedings. He has been an Editor of IEEE Wireless Communication Letters. He is a board member of the Germany Chapter of the IEEE Information Theory Society and member of VDE/ITG Expert Committee "Information and System Theory". He received the VDE/ITG best paper award 2021 and the VDE/ITG certificate of honor in 2024.