

Deep Learning OFDM Receivers for Improved Power Efficiency and Coverage

Jaakko Pihlajasalo¹, Dani Korpi², *Member, IEEE*, Mikko Honkala³, Janne M. J. Huttunen⁴, *Member, IEEE*,
 Taneli Riihonen⁵, *Senior Member, IEEE*, Jukka Talvitie⁶, *Member, IEEE*, Alberto Brihuega⁷,
 Mikko A. Uusitalo⁸, *Senior Member, IEEE*, and Mikko Valkama⁹, *Fellow, IEEE*

Abstract—In this article, we propose multiple machine learning (ML) based physical-layer receiver solutions for demodulating orthogonal frequency-division multiplexing (OFDM) signals that are subject to high level of nonlinear distortion. Specifically, three novel deep learning based convolutional neural network receivers are devised, containing layers in time- and/or frequency-domains, allowing to demodulate and decode the transmitted bits reliably despite the high error vector magnitude (EVM) in the transmit signal. Applicable training procedures are also described, such that the learned layers in the receiver processing properly generalize over different nonlinear distortion and multipath channel characteristics. Extensive set of numerical results is provided, in the context of 5G NR uplink (UL) incorporating also measured terminal power amplifier (PA) characteristics. The obtained results show that the proposed receiver systems are able to clearly outperform the classical linear minimum mean-squared error (LMMSE) receiver as well as the existing ML receiver approaches, especially when the EVM is high compared to modulation order. This is particularly so when the devised ML receiver is of hybrid nature with layers both in time and frequency. The proposed ML receivers can thus facilitate pushing the terminal PA systems deeper into saturation, and thereon improve the terminal power-efficiency, radiated power and network coverage. Through combining the obtained radio link performance results with link budget calculations, all carried out at the 28 GHz mmWave band, it is shown that the proposed ML receivers can enhance the network coverage in terms of maximum UL link distances by close to 100%, when compared to classical LMMSE receiver based networks.

Index Terms—5G NR, 6G, coverage, deep learning, digital post-distortion, EVM, LMMSE, machine learning, mmWave,

nonlinear distortion, OFDM, power amplifier, power-efficiency, receiver.

I. INTRODUCTION

IMPROVING the network coverage and terminal power-efficiency are of fundamental importance in all mobile cellular systems [1], [2]. This is particularly so in wide-area macro deployments as well as in emerging millimeter-wave (mmWave) networks due to the challenges with propagation losses and trade-offs between hardware implementation costs, power consumption and transmit signal quality. Specifically, in the current 4G LTE/LTE-Advanced and 5G NR networks, the uplink coverage is primarily limited by the available user equipment (UE) transmit power while still meeting the unwanted emission and transmit signal inband quality requirements [3].

Interestingly, while the feasible transmit power in below 3 GHz networks is commonly limited by the out-of-band (OOB) emission measures, the role of the passband error vector magnitude (EVM) is becoming more and more critical when the networks evolve towards the mmWave and later even the sub-THz bands [4] in the 6G era. This is primarily because the nonlinear distortion is subject to beamforming [5] as shown through concrete measurements, e.g., in [6].

There are generally many alternative technical approaches to address the network coverage enhancement. These include, e.g., different modulation variants such as $\pi/2$ -BPSK [3], offset QPSK [7] or constrained QPSK [8] that offer reduced envelope variations in terms of peak-to-average-power ratio (PAPR) and thereon larger achievable antenna power with practical nonlinear power amplifier (PA) systems. However, the downside is the limited applicability with orthogonal frequency-division multiplexing (OFDM) physical-layer waveform. 4G LTE/LTE-Advanced and 5G NR networks also support the use of discrete Fourier transform (DFT) spread OFDM (DFT-s-OFDM) in the uplink. This is known to improve coverage, however, at the expense of reduced support for frequency-domain scheduling and link adaptation. The PAPR of OFDM signals can also be explicitly limited, e.g., through the well-known iterative clipping and filtering (ICF) [9] type of processing. Such processing is, however, typically too complex for terminal/UE transmitters

Manuscript received 29 June 2022; revised 13 October 2022 and 22 December 2022; accepted 24 December 2022. Date of publication 13 January 2023; date of current version 14 August 2023. This work was supported in part by the Nokia Bell Labs; in part by the Business Finland; and in part by the Academy of Finland under Grant 319994, Grant 338224, Grant 332361, and Grant 357730. An earlier version of this paper was presented at the IEEE PIMRC 2021 [DOI: 10.1109/PIMRC50174.2021.9569393]. The associate editor coordinating the review of this article and approving it for publication was C.-K. Wen. (*Corresponding author: Mikko Valkama.*)

Jaakko Pihlajasalo, Taneli Riihonen, Jukka Talvitie, and Mikko Valkama are with the Department of Electrical Engineering, Tampere University, 33720 Tampere, Finland (e-mail: mikko.valkama@tuni.fi).

Dani Korpi, Mikko Honkala, Janne M. J. Huttunen, and Mikko A. Uusitalo are with Nokia Bell Labs, 02610 Espoo, Finland.

Alberto Brihuega is with Nokia Mobile Networks, 90620 Oulu, Finland, and also with the Department of Electrical Engineering, Tampere University, 33720 Tampere, Finland.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TWC.2023.3235059>.

Digital Object Identifier 10.1109/TWC.2023.3235059

– particularly when considering mmWave networks with large channel bandwidths.

There are also some prior solutions for dealing with transmitter nonlinear distortion on the receiver side [10], [11], [12]. Especially, the works in [10] and [11] consider a clipping-type nonlinearity and show that their proposed receiver algorithms can efficiently suppress the nonlinear distortion as long as the clipping-function is known. The results in [11] even indicate that moderate clipping might even be beneficial due to the frequency-domain distortion it introduces – in the form of frequency diversity. The work in [12] is applicable also for more general memoryless distortion, albeit the nonlinearity function must still be known.

In this article, we consider an alternative technical approach to improve OFDM-based network coverage and UE power-efficiency, through developing advanced deep-learning aided physical-layer receiver solutions that are capable of decoding the received signal efficiently despite high levels of nonlinear distortion at the transmitter. This has the benefit of learning implicitly to detect distorted signals with high accuracy, without requiring additional processing at the transmitter side, such as digital predistortion (DPD) [6] which is known to be computationally challenging, in general, and particularly for UE transmitters. Our approach does not increase UE complexity while deferring the advanced computations to the gNB side where more processing power is available. Next, we first review the state-of-the-art in machine learning (ML) based receivers available in the literature. After that, the technical contributions of the article are described.

A. State-of-the-Art

Utilizing deep learning in optimizing the receiver performance has already been considered in several works. ML-based channel estimation has been studied in [13], [14], and whereas [15] utilizes convolutional neural networks (CNNs) [16] for equalization. ML-based demapping has been considered in [17], where it was shown to achieve nearly the same accuracy as the optimal demapping rule, albeit with greatly reduced computational cost. Another widely used approach has been to augment the receiver processing flow with deep learning components [18], [19], [20], [21] and thereby achieve improved performance in comparison to conventional benchmark receivers.

Another approach is to consider the task of the receiver as a whole, and train a neural network (NN) to replace the complete physical-layer receiver. Such a solution has been proposed, for instance, in [22] and [23], where a CNN-based receiver, referred to as DeepRx, was shown to achieve high performance especially under sparse pilot configurations. The work in [24] proposes a fully-connected neural network for carrying out joint channel estimation and signal detection. Such an approach is shown to outperform the conventional receiver when there are few channel estimation pilots or when the cyclic prefix is omitted, while also being capable of dealing rather well with clipping noise, a type of hard nonlinearity. The work in [25], on the other hand, applies CNNs to implement a receiver that extracts the bit estimates directly from a linear

time-domain RX signal by learning the DFT operation. The performance boundaries of such fully learned receivers are studied in [26], where data-dependent bit error rate bounds of NN-based receivers are derived.

Fully learned receivers have also been shown capable of dealing with various nonlinear effects or artifacts, such as inter-carrier interference (ICI) stemming from extreme mobility [27]. As for hardware-induced nonlinearities, the impact of various hardware impairments on ML-based receivers has also been analyzed in the literature. The preliminary results in [21], [28], and [29] demonstrate the effectiveness of fully learned receivers in dealing with amplifier-induced nonlinearities. Additionally, in [24] and [30], transmitter-induced clipping effects are considered, with the solution in [30] outperforming a non-ML baseline with similar complexity. In [31], a fully NN-based receiver is shown to be capable of operating efficiently under various impairments, including I/Q imbalance. Similar findings are reported in [32], which also describes a fully learned receiver that can operate under IQ imbalance and carrier frequency offset.

A particularly widely studied impairment in the context of ML-based receivers is phase noise. Especially, [33] proposes an ML-based channel estimator under phase noise and IQ imbalance, demonstrating higher accuracy than conventional methods. The issue of phase noise in sub-THz frequency bands is addressed in [34] by using a deep NN receiver solution. The proposed NN receiver takes in the received signal and a channel estimate, and provides a hard symbol decision as its output, achieving lower bit error rates than the baseline solution. Finally, the work in [35] investigates another type of phase noise resistant ML receiver, consisting of separate NN elements trained to carry out channel estimation and data detection. It is shown that introducing an additional NN for mitigating the effects of phase noise results in higher detection accuracy.

B. Novelty and Contributions

In this work, we focus on developing novel ML-based receiver systems and providing the corresponding new insights for mitigating PA-induced nonlinear distortion at the receiving end. Despite the wide body of work analyzing RF impairments such as phase noise and I/Q imbalance, there is generally considerably less existing work on dealing with nonlinearities using learned receivers. More specifically, the existing works in [21] and [24] propose using a fully-connected NN for receiving nonlinearly distorted signals, showing promising performance compared to conventional baselines. In addition, [29] provides some limited results on learning an end-to-end link as an autoencoder under nonlinear distortion, showing again improved performance over a baseline solution. In all these existing works, however, the receivers are learning only specific PA samples or transmitter clipping level, while the operation under different transmitter realizations is not addressed. Furthermore, the work in [30] focuses on the clipping distortion only, while does not consider realistic responses of practical PAs and their variations from

UE to another. Finally, for completeness it is noted that there is some structural similarity between the receivers proposed in this work and those in [25] – specifically, having ML layers in both time- and frequency-domains. However, in [25], the receiver is allowed to learn the DFT operation, whereas in our approach we include it in the model as a known untrainable layer. Moreover, the work in [25] only considers very mild nonlinear distortion at the transmitter by introducing clipping-based PAPR reduction – opposed to this work, the focus of which is on designing an ML-based receiver for mitigating severe power amplifier-induced nonlinear distortion.

To this end, in this article, we describe three CNN-based receiver architectures and corresponding learning procedures that are capable of accurate signal detection under wide variety of PA samples and even when the level of PA-induced nonlinear distortion is substantially higher than what is allowed by the current 5G NR EVM specifications [36]. The convolutional nature of the proposed solutions lends itself naturally to OFDM-type waveforms, scaling linearly in computational complexity with respect to bandwidth. Especially, we extend our early-stage work reported in [28], where a hybrid time-frequency domain CNN-based receiver was initially proposed. To this end, our contributions can be summarized as follows:

- We develop and present a fully learned physical-layer receiver, referred to as HybridDeepRx, for efficiently demodulating OFDM signals subject to substantial transmitter distortion. Specifically, the HybridDeepRx receiver is equipped with learned convolutional layers in both time- and frequency-domains, such that high-EVM signals can still be demodulated and detected efficiently.
- We also propose a light-weight hybrid receiver, which contains learned convolutional layers only in the time-domain for combating against nonlinear distortion in the received signals. Otherwise, it follows conventional frequency-domain receiver processing for detecting the received symbols. The combination of time-domain CNN (TCNN) and frequency-domain equalization (FEQ) is shown to provide a favorable trade-off between processing complexity and achievable detection accuracy.
- Additionally, we develop and present a purely frequency-domain variant of the HybridDeepRx that operates on post-FFT samples, which facilitates easier hardware implementation and compatibility to ordinary receiver procedures prior to FFT. It utilizes FFT and IFFT pairs within the convolutional layers to achieve the benefits of the time-domain processing without requiring any ML processing before the primary receiver FFT.
- We describe appropriate end-to-end receiver learning procedures such that all the receivers generalize across different PA realizations and channel characteristics.
- Extensive set of numerical results is also provided, in the context of mmWave 5G NR uplink, where measurement-based power amplifier (PA) models are deployed while experimenting with different levels of saturation and corresponding nonlinear distortion in the transmitter system, utilizing the proposed solutions as the base station (BS) receiver. The obtained results show that

all the proposed ML receiver systems outperform the classical linear minimum mean-squared error (LMMSE) receiver, as well as earlier ML-based DeepRx receiver, when there is significant nonlinear distortion in the received signal.

- We also provide corresponding mmWave coverage estimates for the different ML-based receiver architectures, showing as much as $2\times$ improvement in link distance with the proposed frequency-domain HybridDeepRx receiver solution.

For clarity, it is noted that while the presentation focus in this article is on mmWave RF beamforming networks and the corresponding uplink coverage, ML-based receiver schemes can find good use also at lower frequency bands, such as the 3.5 GHz networks. In such cases, covering genuine digital MIMO transmission scenarios is an important ingredient [23].

C. Organization and Notations

The rest of this article is organized as follows. Section II introduces the considered system model and the conventional baseline LMMSE receiver. Section III describes the considered ML-receiver architectures with specific emphasis on the three new receiver solutions with different combinations of time-domain and frequency-domain layers. Section IV discusses the learning procedures and the corresponding data generation for training the receivers. Section V presents the numerical results in 28 GHz 5G NR network context, in terms of radio link BER vs. SNR characteristics for the considered receivers, in different radio channel scenarios and under varying levels of nonlinearity. Additionally, actual coverage results are provided by combining the radio link performance results with 28 GHz pathloss models in different network deployment cases. Finally, the conclusions are drawn in Section VI.

In the forthcoming analysis, matrices are represented with boldface uppercase letters and they can consist of either real- or complex-valued elements, i.e., $\mathbf{X} \in \mathbb{F}^{N \times M}$, where \mathbb{F} stands for either \mathbb{R} or \mathbb{C} .

II. SYSTEM MODEL

Figure 1 depicts the general framework of the considered receiver architectures. The topmost part of Fig. 1 illustrates a conventional OFDM receiver, for reference, while the lower parts show the three proposed receiver systems with varying amounts of learned components. The DeepRx ML receiver from [22] is also shown, serving as another benchmark receiver in addition to the classical LMMSE receiver shown on top.

Let us first describe the basic signal model alongside with the conventional receiver processing. Our primary focus is on mmWave systems where the power-efficiency and coverage challenges are further substantiated [3], [4]. For presentation simplicity, we focus on rank-1 transmission and thus the transmitter and receiver beamforming stages are effectively lumped into the beamformed channel response. To this end, using baseband-equivalent modeling, the received nonlinearly distorted time-domain signal can be expressed as

$$y(n) = h(n) * \phi(x(n)) + w(n), \quad (1)$$

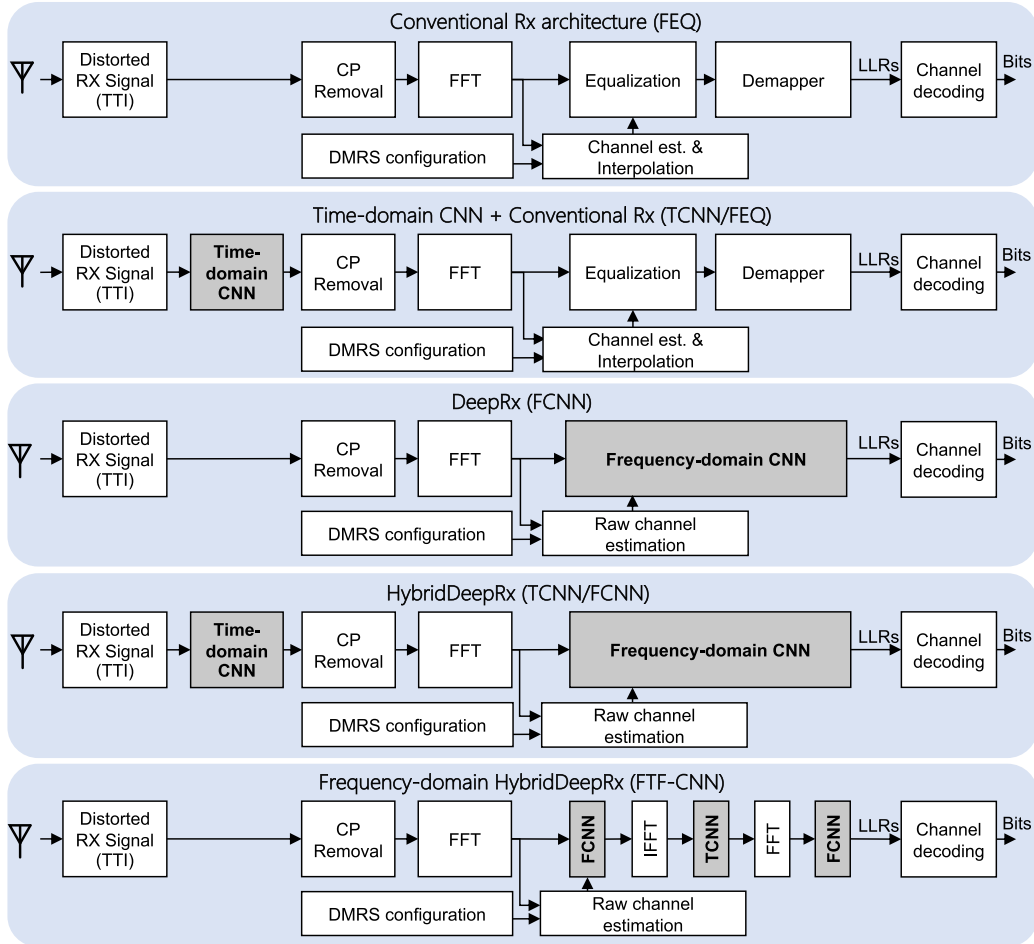


Fig. 1. High-level block diagrams of the proposed and evaluated receiver architectures, compared to a conventional OFDM receiver shown on top. The DeepRx (FCNN) is from [22] while TCNN/FEQ, TCNN/FCNN, and FTF-CNN receivers are proposed in this work.

where $h(n)$ denotes the multipath channel response containing also the effects of the transmitter and receiver linear beamforming, $*$ is the convolution operation, $\phi(\cdot)$ is the effective nonlinear response of the transmitter active array, $x(n)$ is the undistorted transmit waveform, and $w(n)$ is the noise-plus-interference signal. Due to the wide channel bandwidths in mmWave networks, we model the effective nonlinear response of the transmitter active array through the widely-adopted memory polynomial model [37], expressed as

$$\phi(x(n)) = \sum_{\substack{p=1 \\ p \text{ odd}}}^P f_p(n) * (|x(n)|^{p-1} x(n)), \quad (2)$$

where P is the nonlinearity order of the model and $f_p(n)$ denotes the p th-order response of the polynomial model. Importantly, as shown in [5] and [6], such an effective model is able to accurately characterize the *beamformed nonlinear response* of practical mmWave active arrays with multiple parallel PA units, and is thus utilized also here. For clarity, it is also stated that in this work we neglect other RF impairments, such as I/Q imbalance and oscillator phase noise, and consider them in our future work.

Considering next the signal during a single transmission time interval (TTI), the received time-domain signal can be denoted by a matrix $\mathbf{Y}_t \in \mathbb{C}^{(N_{CP}+N) \times N_{\text{symb}}}$, where N_{CP} is

the maximum cyclic prefix (CP) length within the TTI, N is the FFT size and N_{symb} is the number of OFDM symbols. That is, the elements of \mathbf{Y}_t consist simply of the received signal samples, ordered based on their corresponding OFDM symbols. In case the symbols have different CP lengths, zero-padding is used to align the total symbol lengths to $N_{CP} + N$.

Having first removed the CP, the signal is converted to its frequency-domain representation with a fast Fourier transform (FFT), after which it can be expressed as follows:

$$\mathbf{Y}_f = \mathbf{H} \odot \mathbf{X} + \mathbf{N}, \quad (3)$$

where $\mathbf{Y}_f \in \mathbb{C}^{N_D \times N_{\text{symb}}}$ and $\mathbf{X} \in \mathbb{C}^{N_D \times N_{\text{symb}}}$ are the received and transmitted symbols, respectively, $\mathbf{H} \in \mathbb{C}^{N_D \times N_{\text{symb}}}$ is the frequency-domain channel matrix, \odot denotes element wise multiplication, $\mathbf{N} \in \mathbb{C}^{N_D \times N_{\text{symb}}}$ is the noise-plus-interference signal, and N_D denotes the number of data-carrying subcarriers. The noise-plus-interference term incorporates also the effects of nonlinear distortion not captured by the linear part of the signal model.

In a conventional receiver, the demodulation reference signals (DMRSs) are extracted from \mathbf{Y}_f for channel estimation, as illustrated in the upper part of Fig. 1, after which the signal is equalized and the soft bits are extracted. In this work, we consider the widely-used LMMSE receiver as the baseline or reference, due to its broadly-applied nature

TABLE I
MAIN CHARACTERISTICS OF THE CONSIDERED RECEIVER ARCHITECTURES. PRE-FFT PARTS ARE EXECUTED BEFORE THE OFDM FFT TRANSFORMATION, WHILE POST-FFT PARTS ARE EXECUTED AFTER THAT

Name	Pre-FFT	Post-FFT	Other information
FEQ	-	LS channel estimation, LMMSE equalization	Conventional baseline OFDM receiver
TCNN/FEQ	Convolutional ResNet	LS channel estimation, LMMSE equalization	Novel contribution
FCNN	-	Convolutional ResNet	DeepRx, presented originally in [22]
TCNN/FCNN	Convolutional ResNet	Convolutional ResNet	Novel contribution, presented preliminarily in [28]
FTF-CNN	-	Three Convolutional ResNets, IFFT/FFT pairs	Novel contribution, pseudo time-domain processing

in numerous academic and 3GPP standardization studies. For a description of such a receiver, see, e.g., [22]. As a final outcome, the receiver will provide the so-called log-likelihood ratios (LLRs) for each data-carrying resource element (RE). The definition of LLRs is

$$L_{ijl} \triangleq \log \left(\frac{Pr(c_l = 0 | \hat{x}_{ij})}{Pr(c_l = 1 | \hat{x}_{ij})} \right), \quad (4)$$

where $Pr(c_l = b | \hat{x}_{ij})$ is the conditional probability that the transmitted bit c_l is $b \in \{0, 1\}$ given the observed symbol \hat{x}_{ij} , and $l = 0, \dots, B - 1$ where B is the number of bits per symbol. The LLRs represent the receiver's uncertainty about the bits and can be fed, for example, to an LDPC decoder, which then makes decisions regarding the actual information bits.

III. ML-BASED RECEIVER ARCHITECTURES

In this work, we consider four deep learning based receivers, illustrated at high-level already in Fig. 1. For reader's convenience, Table I collects the main characteristics of the considered receiver architectures, while Fig. 2 illustrates their structure in greater detail. These are next described with further rigor.

A. Design Intuition for Proposed Receivers

Let us first take a look at the design intuition we had for the proposed ML receivers, such as the choice of the neural network structure. In this work, all of the proposed neural networks are CNNs as they are a natural choice for OFDM waveforms. This is because they can be naturally represented in 2D space, along the subcarrier and OFDM symbol axes. In such a representation, it suffices that the CNN learns a translationally invariant operation, as the detection task is conceptually the same for all REs. CNNs can also operate under any bandwidth, even if they have not been trained using such data. On the other hand, fully connected layers are tied to certain input and output sizes, meaning that with different signals one would need to have different pre-trained neural networks for each bandwidth configuration. However, convolutional layers function as filters that are slid across the 2D OFDM signal – in case of our receivers, within a single slot – and thus they are considerably more flexible.

This intuition can be expanded to the time-domain. Specifically, applying the time-domain CNN with residual connections on the receiver side is effectively stemming from the fundamental system equations (1) and (2), and the basic well-known fact that neural networks are well capable of approximating almost arbitrary nonlinear functions and transformations. The utilized ResNet structure is beneficial for mitigating the nonlinear distortion of the PA due to the additive nature of the nonlinear response. While the skip connections of the ResNets are generally used to address the vanishing gradient problem, the additive skip connections are also similar to the additive polynomial model of the nonlinear response of the PA or the corresponding active array. Specifically, the output of a ResNet block is of the form $x + F(x)$, where x is the input and $F(x)$ represents the CNN layers and their nonlinear activations, or a linear projection of it. Thus, since the output of the time-domain ResNets should be a signal with the TX induced additive nonlinear distortion mitigated, we believe that the skip connection will help the learning process in achieving this. The nonlinear part of the ResNets only need to learn the additive inverse of the nonlinear part of (2), reflecting a form of digital post-distortion (DPoD) at RX.

Finally, we note that the ML receivers could also be implemented at least partially with complex-valued layers. However, based on our earlier experiments when developing the original DeepRx receiver [22], the use of complex-valued operations did not lead to any significant differences in the number of parameters, complexity, or accuracy of the DeepRx receiver. Since the ML receivers proposed in this work rely on similar ML processing layers, we estimate the same findings to hold also here. For this reason, we have chosen to utilize parallel real-valued signals and the corresponding processing in the proposed ML receiver solutions.

B. Prior DeepRx (FCNN)

DeepRx, first introduced in [22], is a deep learning receiver, that is trained as a single supervised system, instead of training multiple smaller parts of the receiver separately. The benefit of this is that the NN learns directly the task of recovering the transmitted bits, rather than being restricted to learn multiple smaller parts. The goal of the DeepRx is to estimate the received bits from the received frequency-domain signals.

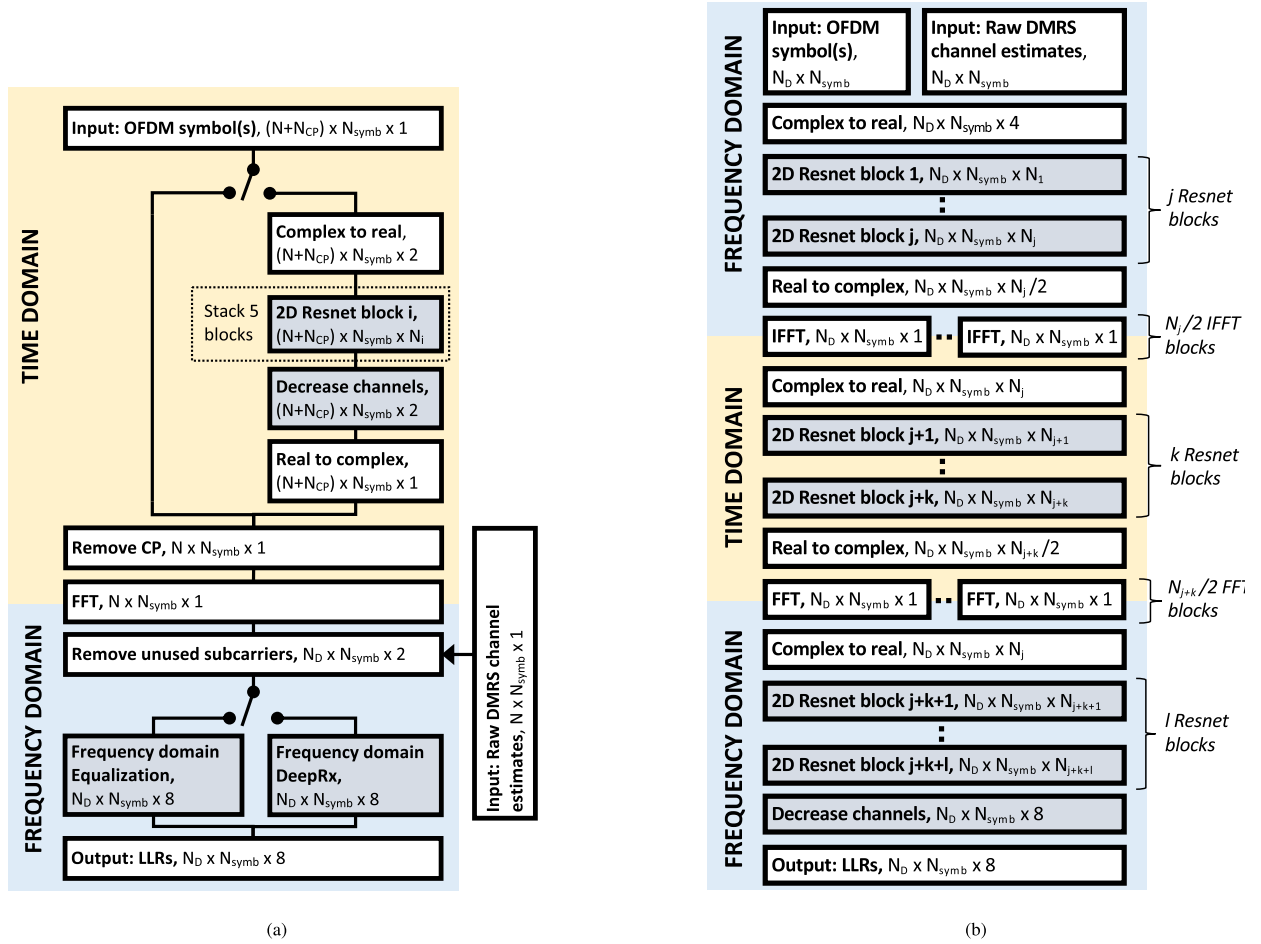


Fig. 2. Detailed illustrations of (a) FEQ, TCNN/FEQ, DeepRx and HybridDeepRx, and (b) frequency-domain HybridDeepRx receiver. The switches in (a) indicate the alternate structures for time- and frequency-domains. The dimensions within each block correspond to its output. The gray blocks represent the components of the learned architecture.

Consequently, the model is constructed in such a manner that the convolutional layers learn to utilize the unknown data symbols and known pilot symbols simultaneously for greater accuracy.

A high-level depiction of the receiver is shown in Fig. 1. The frequency-domain CNN of the DeepRx receives the frequency-domain symbols and the raw least squares DMRS channel estimates as an input. The real and imaginary values of the inputs are concatenated along the channel dimension. For the data-carrying REs, the raw channel estimate array contains zeros. As the channel estimates of the actual data symbols are not calculated or provided to the network, the interpolation is left for the network to learn. Hence, the input to the network is a real valued array $\mathbf{Z}_{\text{post}} \in \mathbb{R}^{N_D \times N_{\text{symb}} \times 4}$. The network architecture follows a residual network structure with 2D preactivation ResNet blocks. With the concatenated inputs, the convolutional filters simultaneously see a neighborhood of the RX signals, pilot symbols and raw channel estimates, facilitating data-aided detection as explained further in [22]. The output is a real valued array $\mathbf{L} \in \mathbb{R}^{N_D \times N_{\text{symb}} \times N_B}$ consisting of the detected LLRs, where N_B is the maximum number of bits per symbol.

DeepRx has not been explicitly designed to operate under PA-induced nonlinear distortion, thus we consider DeepRx as

a reference representing state-of-the-art in general ML-based receivers.

C. Proposed Time-Domain CNN With Conventional Frequency-Domain Receiver (TCNN/FEQ)

The second considered receiver architecture aims at reducing the nonlinear distortion caused by the PA in the time-domain, while simultaneously maintaining the conventional receiver structure in frequency-domain. This is achieved by introducing deep learning processing in the time-domain, before the CP removal, as shown in Fig. 1 and elaborated further in Fig. 2(a). As the nonlinear distortion caused by the PA is inherently a time-domain phenomenon, neural network processing with time-domain inputs is an efficient method for learning to detect such distorted signals. The benefit of this architecture is the more accurate detection of nonlinearly distorted signals, while keeping the computational complexity low. Indeed, the TCNN/FEQ receiver has considerably less parameters than DeepRx, and the other receivers proposed in this paper, and could possibly be used outside the base station scenario for which the other proposed ML receivers are designed for.

Even though the only deep learning processing in the receiver is the time-domain CNN, the receiver is trained

with the transmitted bits as a desired response. This means, that all the parts of the conventional receiver are performed inside the model during training. This is important, and is performed by utilizing untrainable layers within the model. This proposed receiver system has the benefit of keeping the frequency-domain part of the overall receiver the same as in the classical receivers. Such modularity improves the implementation flexibility, even allowing for the CNN block to be by-passed, if so desired. Moreover, the TCNN could be trained separately without the frequency-domain side with synthetic data, but in a real deployment scenario having access to multipath filtered signals with and without PA for training purposes is not feasible. Thus we utilize the end-to-end approach, where the task is to demodulate bits, which are available regardless of the multipath.

The input to the time-domain CNN is constructed using the time-domain RX signals collected during a TTI, represented by the matrix \mathbf{Y}_t . The complex valued signals are split into real and imaginary parts concatenated along the third dimension. To take the varied CP lengths of the 5G specification [38] into account, the OFDM signals with the shorter CP are padded with zeroes, to match the length of the maximum CP length. Thus the input to the time-domain CNN is a real valued array $\mathbf{Z}_{\text{pre}} \in \mathbb{R}^{(N_{\text{CP}}+N) \times N_{\text{symb}} \times 2}$, where N_{CP} again refers to the maximum CP length.

Similar to DeepRx, the network is built with pre-activation ResNet [39] blocks. The network has five ResNet blocks with the number of convolutional filters from 64 up to 256, with filter size 3×3 . The dimensions of the outputs are kept the same size as those of the inputs, to ensure transparency in terms of the ensuing frequency-domain receiver processing. In Fig. 1, the number of filters for ResNet block i is denoted by N_i , such that $N_1 = 64$, $N_2 = 128$, $N_3 = 256$, $N_4 = 128$ and $N_5 = 64$. The last layer of the time-domain CNN is a convolutional layer with two filters of size 1×1 and no activations to match the output size with the input \mathbf{Z}_{pre} , which represents the concatenated real and imaginary parts of the RX signal. This is then converted back to a complex-valued signal, meaning that the rest of the receiver processing can proceed in the usual manner. Indeed, the output of the time-domain CNN is fed to the FFT and consequently to the conventional frequency-domain receiver, which provides the bit estimates.

D. Proposed HybridDeepRx (TCNN/FCNN)

HybridDeepRx, introduced preliminarily in our early-stage work in [28], is a deep learning receiver, which combines the neural network parts of previous two receivers. The receiver is constructed with the nonlinear distortion of the PA in mind. A high-level depiction of the receiver architecture is shown in Fig. 1.

The HybridDeepRx utilizes the same time- and frequency-domain CNNs as the previous two receivers and connects them by including the CP removal and FFT as untrainable layers, which are performed as in a regular receiver. A more detailed depiction of the NN-based receiver is shown in Fig. 2(a). The inputs of the receiver are again the time-domain signals collected during a TTI, denoted by \mathbf{Y}_t , and the raw least

squares DMRS channel estimates, which are directly fed in to the frequency-domain CNN. The DMRS channel estimates are concatenated with the output of the time-domain CNN, similar to the input of the DeepRx. The output of the network is again the detected soft bit estimates.

The time- and frequency-domain CNNs are trained jointly with the untrainable FFT layers between them. The HybridDeepRx is trained end-to-end, such that the non-parametrized operations, such as the FFT, are incorporated to the overall ML model. These operations are differentiable, and thus allow the backpropagation step of the training. In comparison with DeepRx, HybridDeepRx is not a homogeneous neural network as the FFT is performed inside the network. However, this architecture is more efficient as it allows the convolutional layers to focus on learning the detection task, rather than learning the (already known) FFT process. Note that even though HybridDeepRx has more parameters than DeepRx due to the time-domain layers, adding layers with same number of parameters to the frequency side does not significantly affect the performance of the DeepRx.

E. Proposed Frequency-Domain HybridDeepRx (FTF-CNN)

The architecture of the HybridDeepRx is designed to follow a conventional OFDM receiver processing flow with the goal of mitigating the distortion within the RX signal in the time-domain and doing the rest of the detection in the frequency-domain. Essentially, the time-domain network outputs a signal with a generalized solution to the PA distortion. Generally speaking, the time-domain inputs of the HybridDeepRx can be challenging for a practical hardware implementation, meaning that a network that operates fully on post-FFT inputs would be more favourable. However, the effects of the nonlinear distortion of PA can be efficiently compensated for only in the time-domain.

As a solution to this, we propose another ML receiver variant, referred to as frequency-domain HybridDeepRx. Detailed structure of the network is presented in Fig. 2(b). Deferring all the processing to the output domain of the main receiver FFT, allows for keeping the ordinary time-domain receiver functionalities intact, while still benefiting from the rich ML processing structures to efficiently handle the PA distortion. In this architecture, efficient modeling and mitigation of nonlinear distortion is achieved utilizing additional IFFT and FFT conversion pairs inside the frequency-domain NN to carry out consecutive frequency-domain, time-domain and frequency-domain processing.

Another crucial element of this ML receiver architecture is to utilize several parallel IFFT and FFT transformations, in order to avoid introducing a bottleneck inside the ML model. This procedure is illustrated in Fig. 3. If the number of channels of the previous layer N_j is even, the output channels can be paired as real and imaginary parts for $N_j/2$ IFFTs or FFTs. After the IFFT or FFT transformation, the complex numbers can be split to real and imaginary parts to get back to N_j channels for the input of the time- or frequency-domain CNN. In general, this allows the model to be smaller in depth,

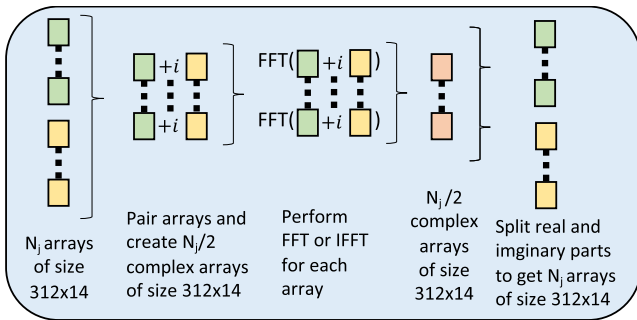


Fig. 3. A detailed illustration of how the parallel IFFTs and FFTs are performed without a bottleneck within the model.

as removing the bottleneck allows the model to maintain the number of channels between the IFFT and FFT conversions.

Altogether, the aforementioned architectural elements allow the network to have the same depth and number of parameters as DeepRx, while mitigating the nonlinear distortion of the PA in the time-domain similar to the original HybridDeepRx. Indeed, the network has the same frequency-domain inputs as DeepRx, denoted by $\mathbf{Z}_{\text{post}} \in \mathbb{R}^{N_D \times N_{\text{sym}} \times 4}$, which are fed into convolutional ResNet blocks. Then the outputs of these ResNet blocks are converted into complex domain, fed into parallel IFFTs, converted back to real values and fed into time-domain ResNet blocks. It should be noted that the transform sizes of the IFFTs are not the same as in the primary FFT transformation. However, this is not an issue, as the goal is not to reconstruct the time-domain waveform, but to transform the signal such that the distortion can be efficiently mitigated. Due to this, the NN processing after the IFFT can be interpreted to operate on a pseudo-time-domain.

After the pseudo-time-domain ResNet blocks, parallel FFT transformations are executed in the same manner as the parallel IFFTs, and the outputs are fed into final frequency-domain ResNet blocks. This structure allows the ML receiver to mitigate the effects of the channel in the first frequency-domain part, then mitigate the PA-induced distortion in the (pseudo-)time-domain and finally map the symbols to bits in the latter frequency-domain part. The structure allows for the nonlinearities in the signal to be mitigated in a logical order relative to how they emerge in the physical transmitting entity (being modelled by (3)). Meanwhile, HybridDeepRx must learn and perform the PA post-distortion on signals that are affected by a multipath channel. Thus frequency-domain HybridDeepRx has performance gains over regular HybridDeepRx in multipath channel scenarios.

IV. TRAINING DATA AND PROCEDURES

In this section, we will discuss the data used for training the receivers and for performance evaluations, as well as the overall training procedure. Altogether, we consider three different channel scenarios to evaluate the performance of the proposed receivers under different conditions, when being affected by the nonlinear distortion of the transmitter system. These scenarios are described below. The training procedure for all of the NN-based receivers is conceptually similar, with the same loss function used for each of them.

A. Considered Channel and PA (Active Array) Models

The channel scenarios considered in this work are as follows:

- Additive white Gaussian noise (AWGN) channel
- Time-invariant multipath channel with a frequency-selective but static channel response
- Time-varying multipath channel with a frequency-selective response that changes over time.

For the latter two scenarios, we utilized 3GPP tapped delay line (TDL) channel models [40]. The AWGN scenario allows to see the isolated benefit of the CNN in compensating for the nonlinearities while the multipath scenarios show the performance under more realistic circumstances. Especially, the latter scenarios show whether the considered receiver architectures are capable of suppressing the nonlinearities also under frequency-selective channels. Moreover, in the multipath channel scenarios, we consider different TDL models for training and validation. Namely, we utilize TDL-B, TDL-C and TDL-D for generating training data, and TDL-A and TDL-E only for validation. This ensures that the reported results represent the generalized performance of the ML-based receivers, not dependent on learning individual channel profiles.

Moreover, to simulate the nonlinear behaviour of the PA, or more generally the effective beamformed nonlinear response of mmWave active array, the response of a real-life mmWave PA module was measured under a high input power. Then, a 17th order polynomial was fitted to the measurements, representing the AM-AM and AM-PM response of the PA. To ensure that the trained receivers do not simply memorize the PA response, we then introduce a dithering term that is applied to the measured PA polynomial to produce several slightly different PA models for training and validation, as also the true PA realizations vary across the transmitting UEs in real systems. Similarly, a change of the network center-frequency or frequency channel may impact the nonlinear response of a given UE. The dithering is performed by adding a normally distributed random number to each polynomial coefficient, with a weight factor that is proportional to the magnitude of the original polynomial coefficient, while also imposing an applicable saturation level to the model such that physical PA behavior is correctly mimicked. The dithered model is then inspected by calculating its first and second derivatives numerically. If the derivative is always positive and second derivative negative within certain error margin, the model is considered valid.

We generated 40 PA models in total, 30 of which are used only for training datasets and 10 only for the validation datasets. As noted earlier, in the context of mmWave active arrays, these represent the effective beamformed nonlinear responses of the UE transmitters. The AM-AM and AM-PM responses of the generated validation models are presented in Fig 4, in addition to their EVM behavior for backoff values from 0dB to 12dB to illustrate how backoff affects the nonlinearity. To take varying levels of nonlinearity into account, we generated datasets for each PA backoff value ranging from -1 dB up to 7 dB, depending on the scenario.

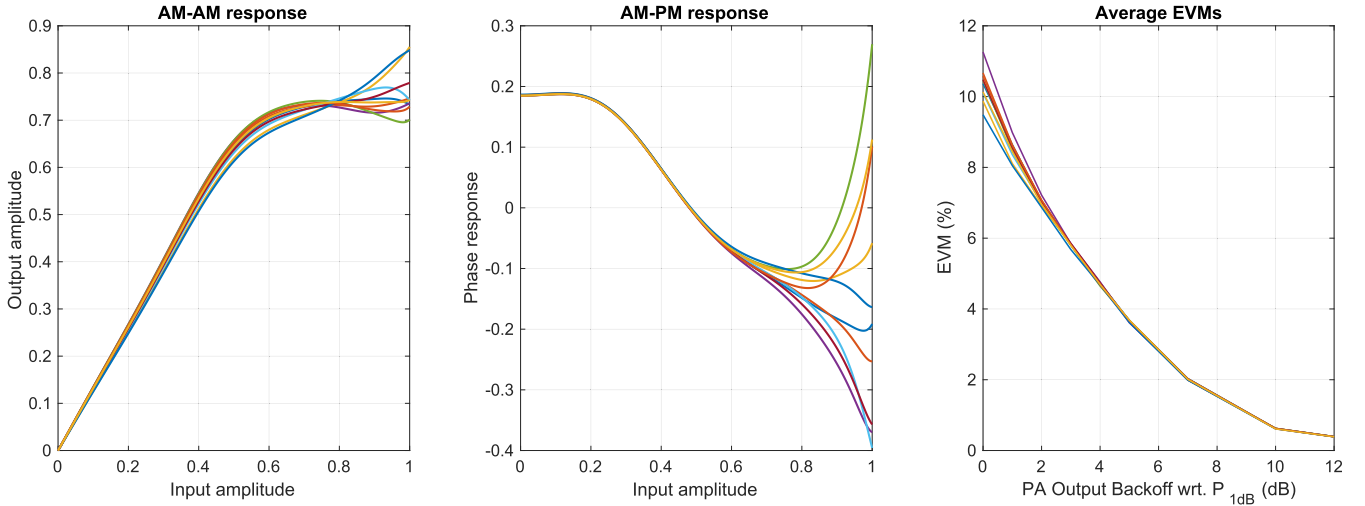


Fig. 4. The AM-AM and AM-PM responses of the generated PA models used for validation. The figure also shows the corresponding PA output waveform EVMs for different backoff values, averaged across multiple OFDM symbols.

TABLE II
EVALUATION PARAMETERS FOR TRAINING AND VALIDATION

Parameter	Value	Randomization
Carrier frequency	28 GHz	N/A
Channel model	AWGN, TDL-A to TDL-E	Noise and channel realizations, delay spread and Doppler characteristics
Doppler shift	0 Hz – 1500 Hz	Uniform distribution or fixed
Delay spread	Up to 100 ns	Uniform distribution
PA model	Measurement based 17th order polynomial	Dithered coefficients
PA backoff value	–1 dB – 7 dB	N/A
SNR	0 dB – 30 dB	Uniform distribution or fixed
Channel bandwidth	50 MHz	N/A
Num. subcarriers (N_D)	792 subcarriers	N/A
FFT size (N)	1024	N/A
Subcarrier spacing	60 kHz	N/A
Max. CP length (N_{CP})	104	N/A
OFDM symbol duration	71.4 μ s	N/A
TTI length (N_{symp})	14 OFDM symbols	N/A
Modulation scheme	64-QAM	N/A

B. Data Generation

In order to generate training data, we simulated a rank-1 5G physical uplink shared channel (PUSCH) link with Matlab's 5G Toolbox [41], using the parameters specified in Table II and the PA and channel models described above.

The SNR values for the datasets were chosen randomly for the training datasets and using a uniform grid for validation. To account for lower number of errors with higher SNR values, more validation data with higher SNR was generated. In the AWGN scenario, the training datasets have a total of 30 000 TTIs and the validation datasets have 15 500 TTIs. For the multipath scenarios, we generated more data to take into account the increased diversity of the channel conditions,

with a total of 105 000 TTIs for training and 48 400 TTIs for validation.

C. Training Procedure

The training for each receiver is performed using the binary cross entropy (CE) as the loss function, training and optimizing the complete receiver algorithm as a whole. Although the actual output of the receivers consists of the LLRs, the training of each neural network receiver is performed using the transmitted bits as the labels. This has the practical benefit that there is no need to define any intermediate desired outputs, such as channel coefficients, or soft information, such as the magnitude of the LLRs. In particular, denoting the set of trainable parameters by θ , the loss function is defined as [22]

$$\text{CE}(\theta) \triangleq -\frac{1}{\#\mathcal{D}B} \sum_{(i,j) \in \mathcal{D}} \sum_{l=0}^{B-1} \left(b_{ijl} \log(\hat{b}_{ijl}) + (1 - b_{ijl}) \log(1 - \hat{b}_{ijl}) \right) \quad (5)$$

where \mathcal{D} denotes the time and frequency indices of data-carrying REs, $\#\mathcal{D}$ is the total number of data-carrying REs, B is the number of bits per resource element, and \hat{b}_{ijl} is the receiver's estimate for the probability that the bit b_{ijl} is one. The bit estimate is obtained by feeding the corresponding LLR through the sigmoid-function as

$$\hat{b}_{ijl} = \text{sigmoid}(L_{ijl}) = \frac{1}{1 + e^{-L_{ijl}}}, \quad (6)$$

where L_{ijl} denotes the LLRs which are the actual output of the receivers. With this procedure, the neural network will learn to implicitly predict the proper soft information, i.e., magnitude of the LLRs. Note that outside training, the output of the network can be extracted before the sigmoid function to receive LLRs instead of the bit probabilities. The chosen stochastic gradient descent (SGD) algorithm in this work is the Adam optimizer, which updates the weights based on the CE loss in (5). We used default parameters of the Adam optimizer. In the case of untrainable layers, such as FFT, the operation

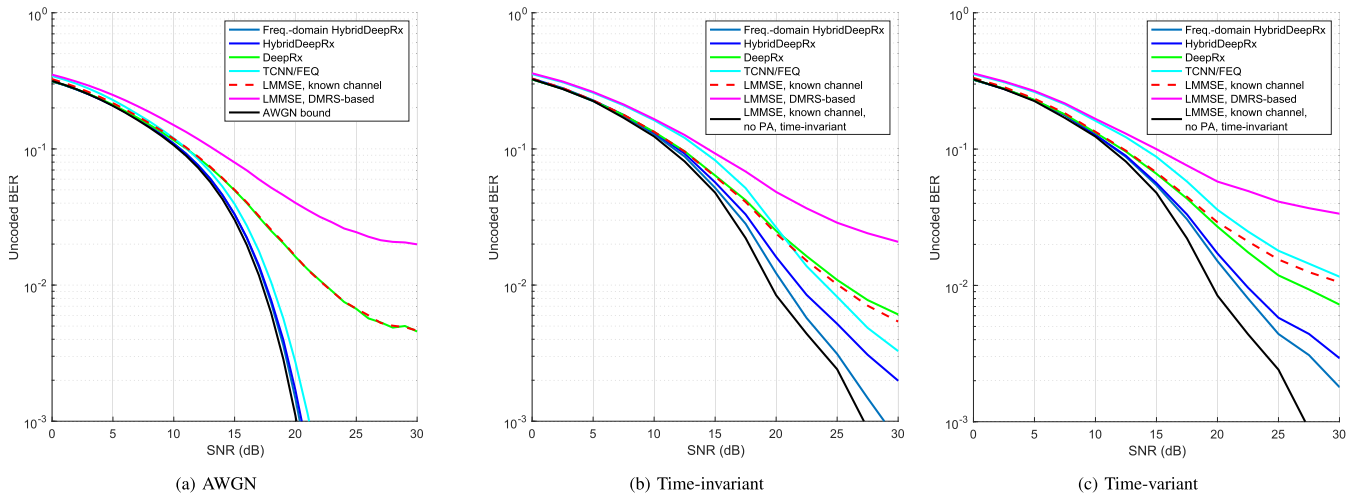


Fig. 5. Uncoded BER performance of the considered receivers, with a PA backoff of 3 dB under (a) AWGN channel, (b) time-invariant multipath TDL channel models and (c) time-variant multipath TDL channel models.

performed in the layer should be differentiable, to facilitate backpropagation over the whole model. Additionally, all the neural networks were trained for 240 to 300 thousand epochs with batch size of 20. We used learning rate of 0.001 with a warmup period of 800 epochs and linear decay after 30% epochs.

V. NUMERICAL RESULTS AND ANALYSIS

The performance of the proposed receivers are evaluated under varying levels of nonlinearity with uncoded bit error rate (BER) as the main performance criteria. Real networks always deploy error control coding, however, as differences in uncoded BER transfer to differences also in coded BER domain, this approach is taken such that the results are independent of the coding scheme. In addition to the four neural network receivers presented in Section III, we evaluate the performance of two LMMSE-based receivers; one with known channel and another performing channel estimation based on two DMRS symbols within the TTI. The DMRS-based channel estimation is performed by first calculating a least squares (LS) channel estimate for the DMRS symbols, and then interpolating it linearly over the whole slot (extrapolation of the channel estimate beyond the last pilot symbols is done with the nearest neighbor rule). The LMMSE based receivers and DeepRx are considered as baseline receivers.

Moreover, for each scenario we show always a reference performance that can be considered as the upper bound for achievable performance. In the AWGN case, the upper bound is the AWGN BER under a completely linear PA. For both multipath channel scenarios, we use an upper bound corresponding to the BER achieved by an LMMSE receiver with known time-invariant channel and a completely linear PA, to ensure a completely ICI-free scenario.

For all three scenarios, we first show the BER performance when the PA backoff value is set to 3 dB, which corresponds to an EVM of roughly 8%. This is the highest allowed EVM in 64-QAM modulation in 3GPP 5G NR specifications. Then we investigate the SNR required to reach 10% and 1% BER

values, with respect to varying levels of nonlinear distortion. This shows the performance gain with respect to PA output backoff in relation to 1 dB compression point, indicating how much the PA can be pushed to reach a target BER with the different receivers. The considered BER values of 1% and 10% of our results correspond to SNR range of 10 dB to 25 dB, which is in line with typical first-transmission coded block-error rate (BLER) of 10% in practical networks [3].

In addition, we show example results for a single-DMRS scenario and receiver trained on varying backoff values to show robustness of the proposed solution. Furthermore, we illustrate the complexity and performance tradeoff of the proposed receivers and discuss the complexity analysis of the solutions. Finally, actual coverage calculations are provided, in terms of the achievable maximum link distances with different receivers, by combining the radio link performance results and applicable mmWave pathloss models.

A. AWGN Channel

Figure 5(a) shows the BER performances with PA backoff value 3 dB for the AWGN scenario. It is evident that the time-domain CNN in the proposed receivers has considerable effect to the performance in comparison with the benchmark receivers. In fact, the nonlinear distortion causes an error floor for the benchmark receivers while both HybridDeepRx variants can almost reach the AWGN bound. Also the TCNN/FEQ receiver achieves a BER close to the AWGN bound. This clearly highlights the benefit of the temporal processing achieved by the trained layers.

Let us next investigate the performance while considering a specific BER value. To this end, Fig. 6 shows the SNR required to achieve uncoded BER values of 10% and 1% with respect to different levels of nonlinear distortion. Lower PA backoff value indicates higher nonlinearity. It can be observed that the proposed HybridDeepRx receivers can achieve the target BER with considerably lower SNR than the benchmark receivers. In fact, in Fig. 6(b), DeepRx and LMMSE receivers can not reach the 1% BER target within the studied SNR range if the PA backoff is less than 3 dB.

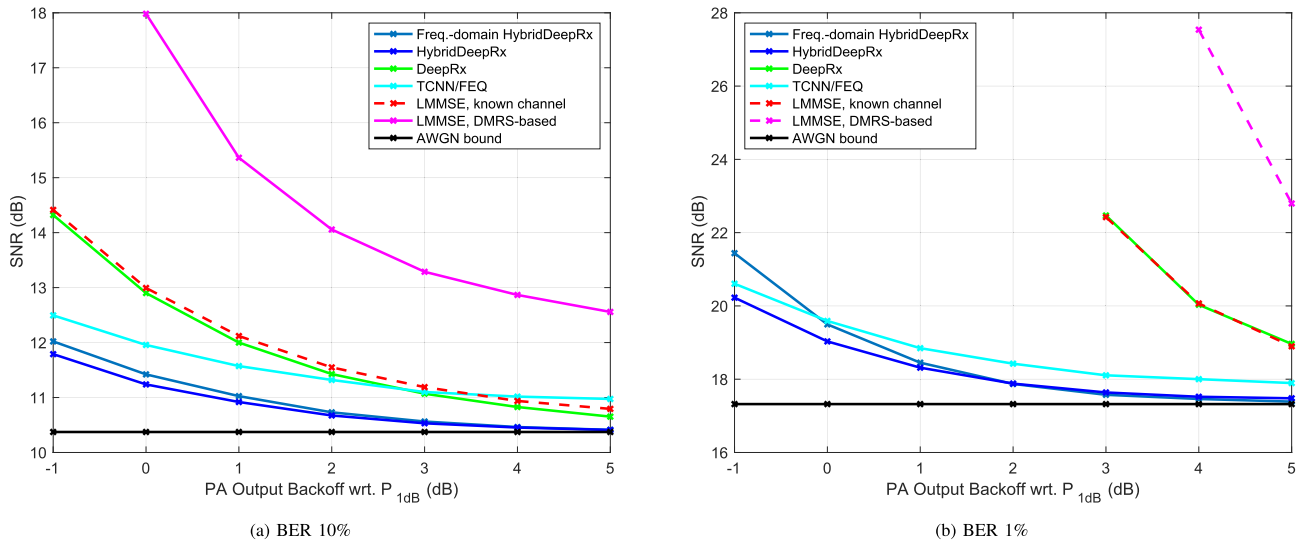


Fig. 6. Performance of the considered receivers under AWGN channel and varying levels of nonlinearity at (a) 10% uncoded BER and (b) 1% uncoded BER for 64-QAM.

As opposed to this, both HybridDeepRx receivers can achieve the BER target even under the most severe nonlinear distortion considered in this work. At low backoff values, the frequency-domain HybridDeepRx is slightly worse than the regular HybridDeepRx likely due to the initial frequency-domain processing not being suitable for the AWGN channel. Furthermore, the TCNN/FEQ receiver performs better than the benchmark receivers for all but the highest backoff values at 10% BER. These findings indicate that the proposed receivers can operate under high levels of nonlinear distortion, thus allowing to push the transmitter PA system towards saturation for improved power-efficiency and coverage. Indeed, the proposed ML receivers reach 1% BER at about 20 dB SNR with some 5 dB lower backoff when compared with the benchmark receivers.

B. Time-Invariant Multipath Channel

Next, let us consider the scenario with time-invariant multipath channel. Figure 5(b) shows the corresponding BER performances with a PA backoff value of 3 dB. Now, the frequency-domain HybridDeepRx has the best performance of the considered receivers, likely due to the fact that it can carry out initial frequency-domain processing before the time-domain processing, unlike the regular HybridDeepRx. Furthermore, the TCNN/FEQ receiver at higher SNR values achieves better results than DeepRx and LMMSE with known channel. These findings indicate that the time-domain CNN in the proposed receivers is effective even under a frequency-selective channel.

Then, in Fig. 7 we see the performance with respect to different levels of nonlinear distortion. In Fig. 7(a), where the BER target is 10%, it can be observed that the difference between the HybridDeepRx variants and other receivers is not as significant as in the AWGN case. This is due to the multipath channel, which makes the mitigation of the nonlinear distortion more difficult. Nevertheless, the benefits of the ML-based receivers are still evident. Moreover,

the frequency-domain HybridDeepRx, which is capable of doing initial frequency-domain processing before the time-domain phase, shows again performance gain over the regular HybridDeepRx.

In Fig. 7(b), where the BER target is 1%, the nonlinear distortion is a more significant bottleneck and it can be observed that the proposed HybridDeepRx receivers can achieve the target BER with considerably lower SNR than the benchmark receivers. As for the TCNN/FEQ receiver, it outperforms the baseline receivers with the lowest backoff values, while falling behind when the PA is more linear. This is likely due to the time-domain CNN having not properly learned to operate under a linear PA, meaning that it might be more favorable to bypass it entirely when the level of distortion is low enough.

Altogether, these results indicate that the nonlinear distortion can be efficiently dealt with also under a frequency-selective channel. Especially, the frequency-domain HybridDeepRx improves the highest allowed PA backoff by about 3 dB at 1% BER compared to classical methods.

C. Time-Varying Multipath Channel

Finally, let us consider the most practical and realistic scenario with a time-varying multipath channel. Fig. 5(c) shows the BER performances with a PA backoff value of 3 dB. Again, both HybridDeepRx variants have the best performance of the considered receivers, indicating that they are capable of dealing with both frequency-selective and time-variant channels while simultaneously mitigating the impact of nonlinear distortion. Moreover, also DeepRx achieves higher performance than LMMSE with known channel. This observation is in line with the findings in [22] and it can be attributed to the fact that DeepRx is capable of mitigating ICI. In this scenario, LMMSE-receiver with known time-invariant channel and without PA achieves the lowest BER among all considered receivers, as it does not suffer from ICI or PA-induced distortion.

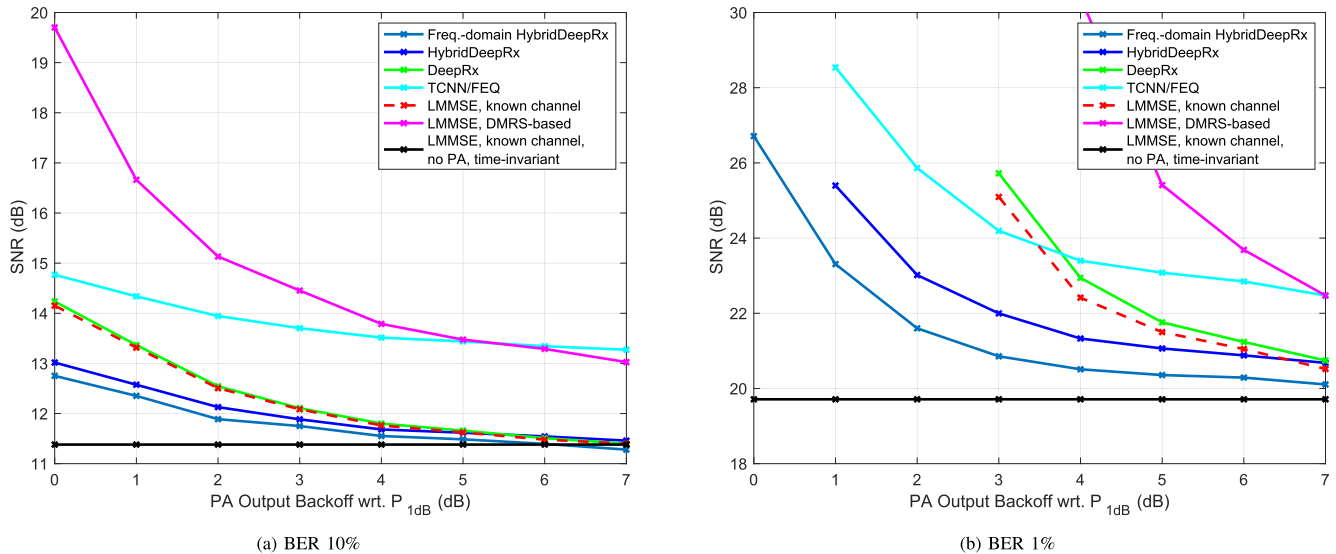


Fig. 7. Performance of the considered receivers under time-invariant multipath channel models and varying levels of nonlinearity at (a) 10% uncoded BER and (b) 1% uncoded BER for 64-QAM.

Furthermore, it can be seen from Fig. 5(c) that the performance of the TCNN/FEQ receiver falls between the other two LMMSE-based receivers. This indicates that, in this scenario, accurate channel knowledge is more beneficial than the ability to mitigate the effects of nonlinear distortion, as the LMMSE receiver with perfect channel knowledge outperforms the TCNN/FEQ receiver.

Finally, Fig. 8 shows the performance with respect to different levels of nonlinear distortion – in terms of varying backoff – for the case of time-variant multipath channel. In this scenario, we also consider 5% BER as the reference point of interest, as the LMMSE with two pilots and the TCNN/FEQ receiver are incapable of reaching the 1% BER target. As in the previous scenario, Fig. 8(c) shows that the proposed HybridDeepRx receivers can reach the 1% BER target with lower backoff than the baselines. This again demonstrates the beneficial effects of pseudo-time-domain ML processing. In fact, the frequency-domain HybridDeepRx is clearly the best of the considered receivers, with a backoff improvement of about 2.5 dB over DeepRx. At 10% and 5% BER targets in Figs. 8(a) and 8(b), the performance gains of the ML-based receivers are also clearly evident. Furthermore, also the TCNN/FEQ receiver outperforms the DMRS-based LMMSE for backoff values below 6 dB.

D. Single-DMRS Results

Let us next consider the proposed receivers under a single DMRS only and output backoff value of 3 dB. In addition to a scenario with the max Doppler shift of 1500 Hz given in Table II, we also consider the performance with lower max Doppler shift of 150 Hz to better show the general performance under these conditions, as the original max Doppler shift affects the single DMRS scenario very significantly. However, the proposed receivers perform well in both scenarios as seen in Fig. 9. Notably the original HybridDeepRx achieves similar results as the frequency-domain variant likely due to more layers in the frequency-domain, which are more useful for

the ICI mitigation. For a more comprehensive analysis on the impact of DMRS patterns for ML-based receivers, we refer to our earlier work on the DeepRx receiver in [22].

E. Complexity and Performance Trade-Offs

To illustrate the trade-off between the complexity and performance of the network, we have tested the performance of frequency-domain HybridDeepRx for varying number of ResNet blocks. In particular, the results in Fig. 10 illustrate the relationship between the model depth/size and the achievable performance, while Table III shows the number of parameters of each variant, with the boldface row representing the baseline HybridDeepRx receiver parametrization. With all the tested model sizes, the ML receiver shows gains over the LMMSE receiver with known channel. Thus, it manages to mitigate the PA-induced distortion even with lower number of ResNets, while increased network depth will result in higher and higher detection accuracy. However, the performance difference between having 11 ResNet blocks, which is also the number of ResNet blocks in DeepRx [22], and having 13 ResNet blocks is not very significant.

Furthermore, the considered ML receivers under default configurations can be compared with their number of trainable parameters. The frequency-domain HybridDeepRx has the same amount of parameters as DeepRx, about 654 thousand, although it is capable of much more accurate detection under a nonlinear PA. It should also be emphasized that the frequency-domain HybridDeepRx also utilizes several IFFT and FFT transformations, which mean that it performs slightly more computational operations than the corresponding DeepRx model. As opposed to this, the original HybridDeepRx, which has about 934 thousand parameters, does not have any extra IFFT or FFT operations as it utilizes additional ResNet blocks before the primary receiver FFT. As for the TCNN/FEQ receiver, it has by far the least number of parameters of about 280 thousand, as it utilizes ResNet blocks only

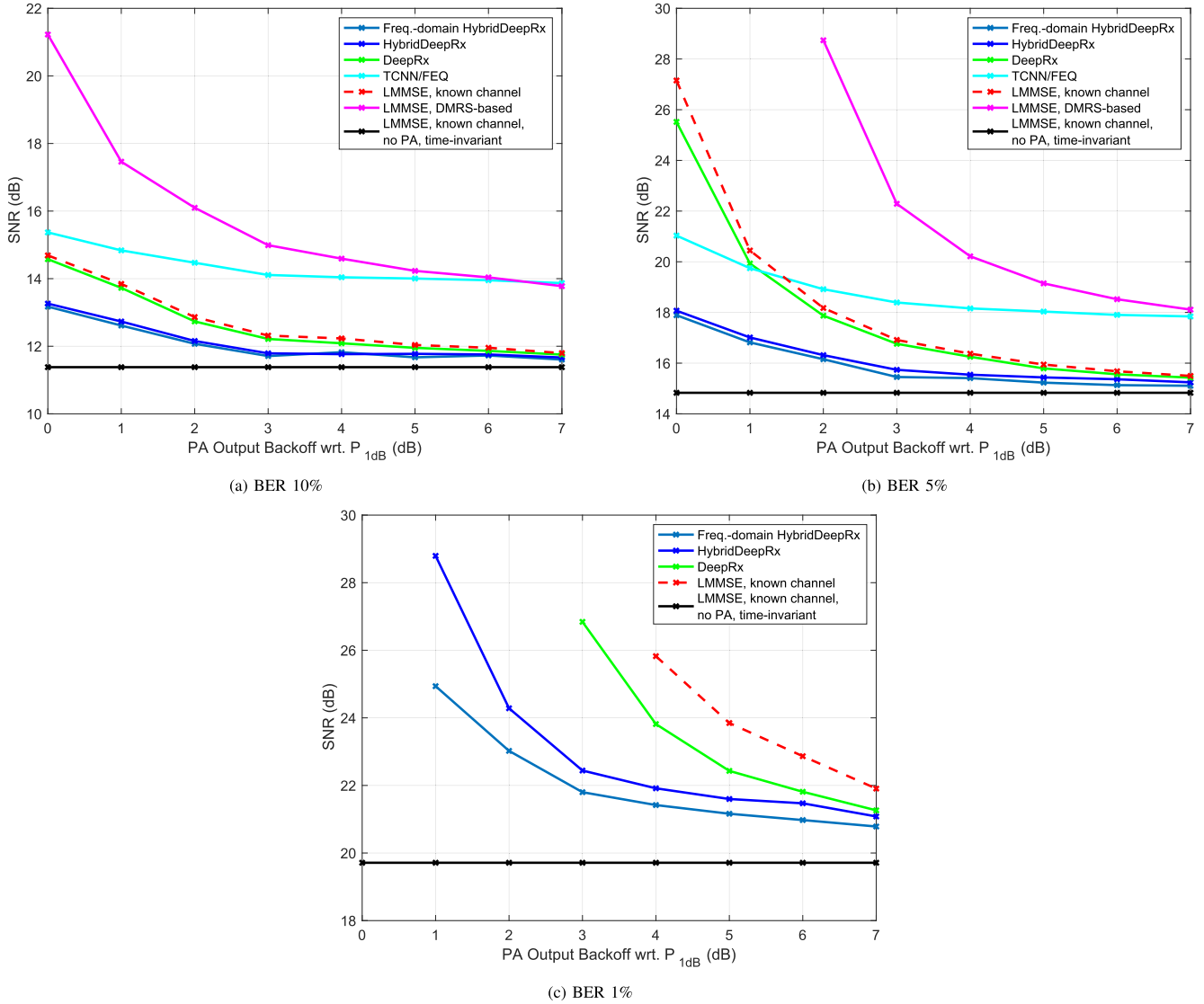


Fig. 8. Performance of the considered receivers under time-varying multipath channel models and varying levels of nonlinearity at (a) 10% uncoded BER, (b) 5% uncoded BER for 64-QAM, and (c) 1% uncoded BER for 64-QAM.

TABLE III
NUMBER OF TRAINABLE PARAMETERS IN FREQUENCY-DOMAIN
HYBRIDDEEPRX WITH VARYING DEPTH

ResNet blocks	Parameters
5	292k
7	311k
9	382k
11	654k
13	725k

in the time-domain, while performing conventional receiver processing after the FFT.

To study the run-time inference complexity, we restrict our analysis to the frequency-domain HybridDeepRx receiver, which is the primary proposed receiver solution of this work. Firstly, we can observe that it has a fully convolutional architecture. Based on this, it is straightforward to deduce that the convolutional layers have an asymptotic complexity

of $\mathcal{O}(N_D N_{\text{symb}})$. Moreover, it is also known that the FFT and IFFT parts have an asymptotic complexity of $\mathcal{O}(N_D \log(N_D) N_{\text{symb}})$, meaning that the overall asymptotic complexity can be written as the addition of the two. With moderate system bandwidths, the number of FFT and IFFT operations is far below the number of operations within the convolutional layers, meaning that the first term is dominant. Hence, we conclude that in most cases the asymptotic complexity of frequency-domain HybridDeepRx scales linearly with respect to bandwidth and slot length as $\mathcal{O}(N_D N_{\text{symb}})$. This means that the complexity of the proposed solution scales similar to conventional receivers, which have also similar asymptotic complexity [22]. However, it should be noted that with extreme bandwidths, the IFFT and FFT blocks will start to dominate the complexity, meaning that in such cases the latter term remains.

Moreover, we also wish to emphasize that in practice the number of operations required for the inference of the

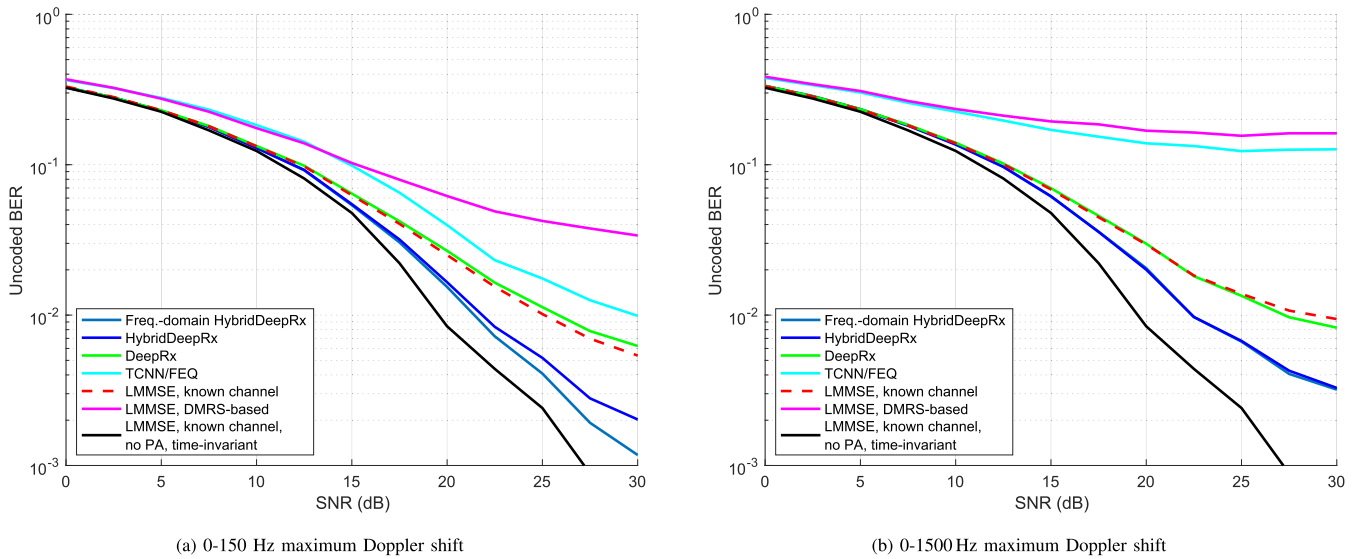


Fig. 9. Single-DMRS results for Doppler shift range of a) 0–150 Hz and b) 0–1500 Hz.

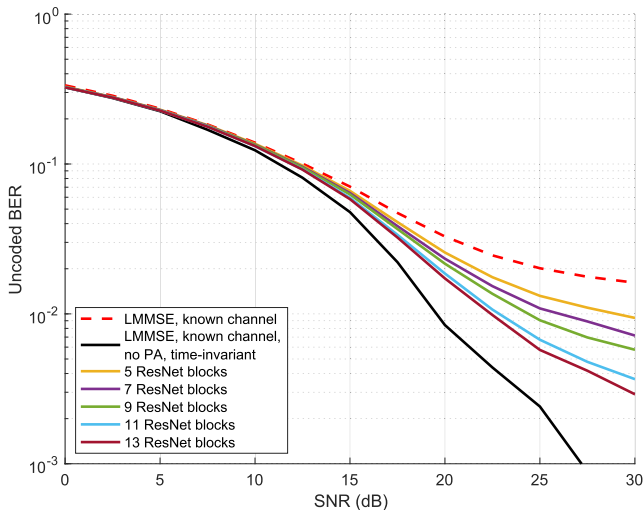


Fig. 10. Comparison of frequency-domain HybridDeepRx performance with different number of total ResNets used.

proposed ML receiver is higher than that of conventional receivers, despite the fact that they scale similarly with respect to bandwidth. In the end, what matters is the power consumption required by each algorithm, and to this end customized ML hardware accelerators have the potential to achieve higher power efficiency for a given number of computational operations (OPS) than conventional digital signal processing (DSP) chips. This is due to the fact that ML algorithms repeat a very small set of simple operations repeatedly, allowing for highly optimized hardware. For instance, the future compute-in-memory chips show very promising power efficiency figures for ML inference [42]. We consider a more detailed hardware-aware complexity analysis an important future work item.

F. Global Network for Different Backoffs

In the previous results, the receivers were initially trained and thus optimized to a certain backoff value. Fig. 11 shows

the performance of frequency-domain HybridDeepRx trained simultaneously on all backoff values from 0 dB to 7 dB. This illustrates that the time-domain layers also perform well with varying levels of nonlinearity. As we can see, the effect of training only on certain backoff value on the results is not significant, which shows that the time-domain layers learn to mitigate varying levels of nonlinearity. Additionally, we tested frequency-domain HybridDeepRx up to backoff values of 12 dB, where the EVM is less than 1%, to see the performance with already essentially linear transmitter. Fig. 12 illustrates that the proposed architecture works well even in case of such practically linear PA. Alternatively, as the network anyway roughly knows the UE transmit power in real deployment, e.g., through pathloss estimates, the base-station could utilize the baseline DeepRx when the UE transmitters are effectively in their linear region.

G. Coverage Analysis

Finally, to illustrate the potential coverage extension enabled by the reduced PA backoff, Table IV shows the maximum expected link coverage in meters for all considered receiver architectures, alongside with the effective isotropic radiated power (EIRP) required for achieving that coverage. The coverages are calculated for a 5G NR uplink scenario, assuming a 50 MHz channel bandwidth at 28 GHz carrier frequency with Urban micro (UMi) and macro (UMa) path loss models (with default parameters), as defined in [40]. Both line-of-sight (LOS) and non-line-of-sight (NLOS) conditions are considered. Moreover, Fig. 13 illustrates the expected coverage with respect to EIRP for all scenarios.

In the underlying link budget analysis, the receiver performance is determined for a BER target of 5% using Fig. 8(b), such that a backoff of 3 dB corresponds to an EIRP of +30 dBm. It is then assumed that each 1 dB increment (or decrement) of backoff results in a similar change in the EIRP. In the base-station receiver side, a total RX beamforming gain of 25 dB is assumed, stemming from the use of a large RX

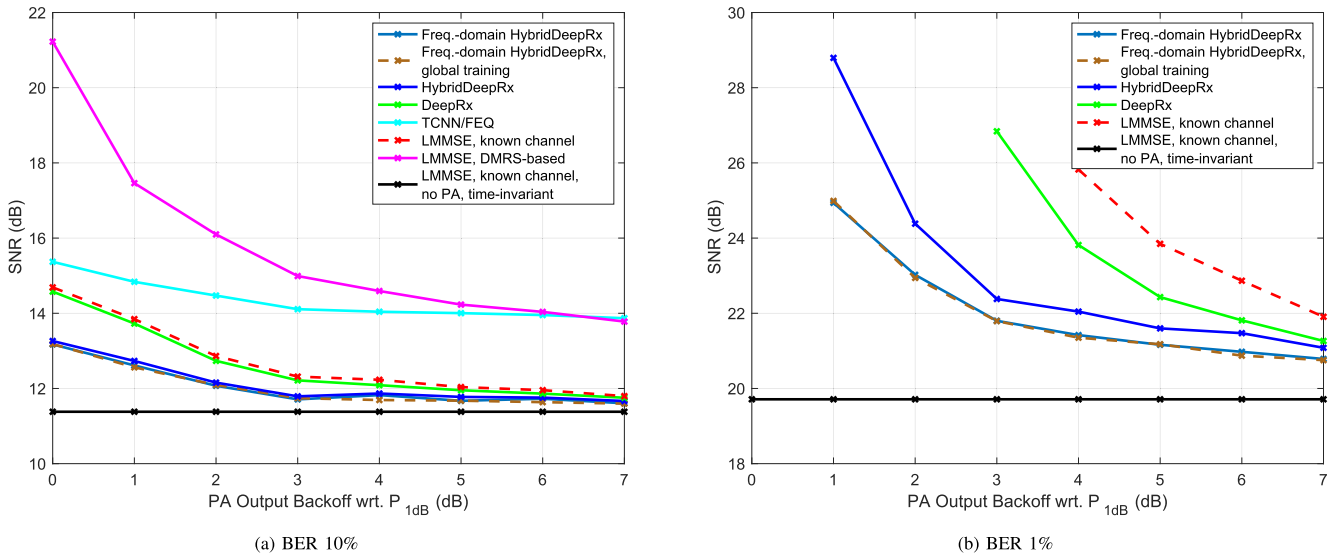


Fig. 11. Performance of frequency-domain HybridDeepRx trained for all backoff values and validated separately at (a) 10% uncoded BER and (b) 1% uncoded BER.

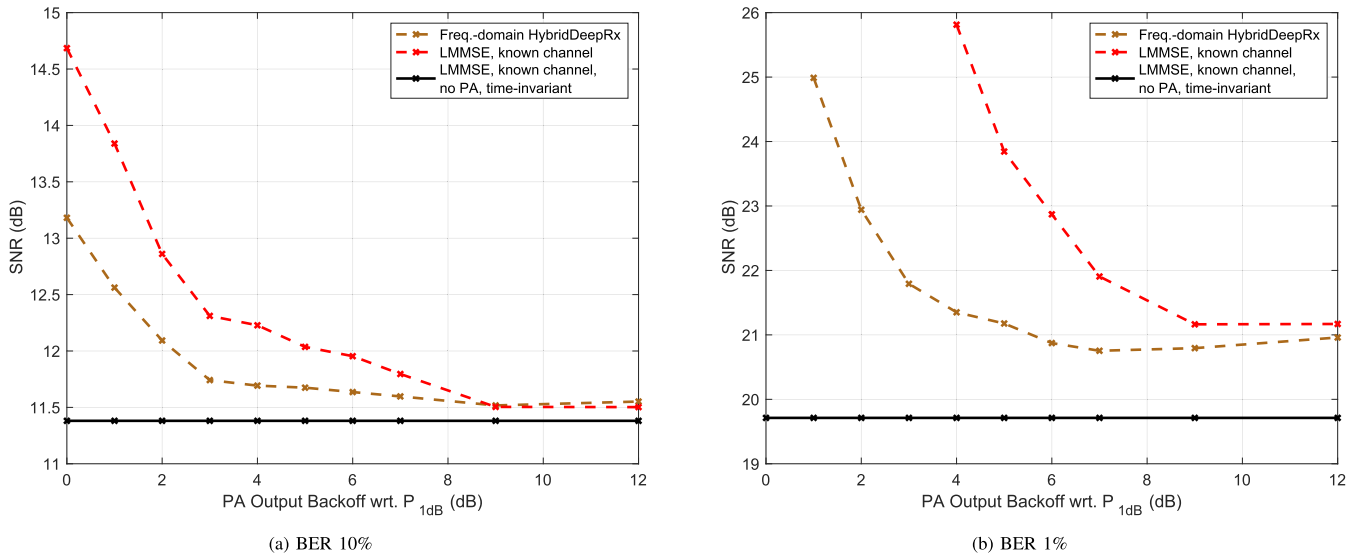


Fig. 12. Performance of frequency-domain HybridDeepRx trained on backoff values 0dB to 12dB at (a) 10% uncoded BER and (b) 1% uncoded BER.

TABLE IV

MAXIMUM ACHIEVED LINK DISTANCES OF THE CONSIDERED RECEIVERS FOR EIRP VALUES BETWEEN +27 AND +32 dBm IN DIFFERENT NETWORK DEPLOYMENT SCENARIOS

Network deployment scenario	UMi LOS		UMi NLOS		UMa LOS		UMa NLOS	
	Distance	EIRP	Distance	EIRP	Distance	EIRP	Distance	EIRP
Freq.-domain HybridDeepRx	1703 m	+32 dBm	143 m	+32 dBm	1945 m	+32 dBm	165 m	+32 dBm
HybridDeepRx	1648 m	+31 dBm	139 m	+31 dBm	1865 m	+31 dBm	161 m	+31 dBm
DeepRx	1383 m	+30 dBm	125 m	+30 dBm	1578 m	+30 dBm	146 m	+30 dBm
TCNN/FEQ	1253 m	+32 dBm	118 m	+31 dBm	1436 m	+31 dBm	138 m	+31 dBm
LMMSE, known channel	1353 m	+30 dBm	124 m	+30 dBm	1545 m	+30 dBm	144 m	+30 dBm
LMMSE, DMRS-based	854 m	+29 dBm	93.8 m	+28 dBm	995 m	+28 dBm	118 m	+29 dBm

antenna array. Moreover, the receiver is assumed to have a noise figure of 8 dB, including also the possible RX losses. The link distance is then determined based on how much path loss can be tolerated with different PA output backoff values in order to achieve the receiver SNR required for a BER of 5%, based on Fig. 8(b).

In general, it can be observed from Table IV that all of the proposed ML-based receiver solutions provide enhanced coverage. The fully-learned receivers (frequency-domain HybridDeepRx, DeepRx and HybridDeepRx) can utilize much higher transmit powers by pushing the transmitter PA deeper into saturation, and thereby achieve higher

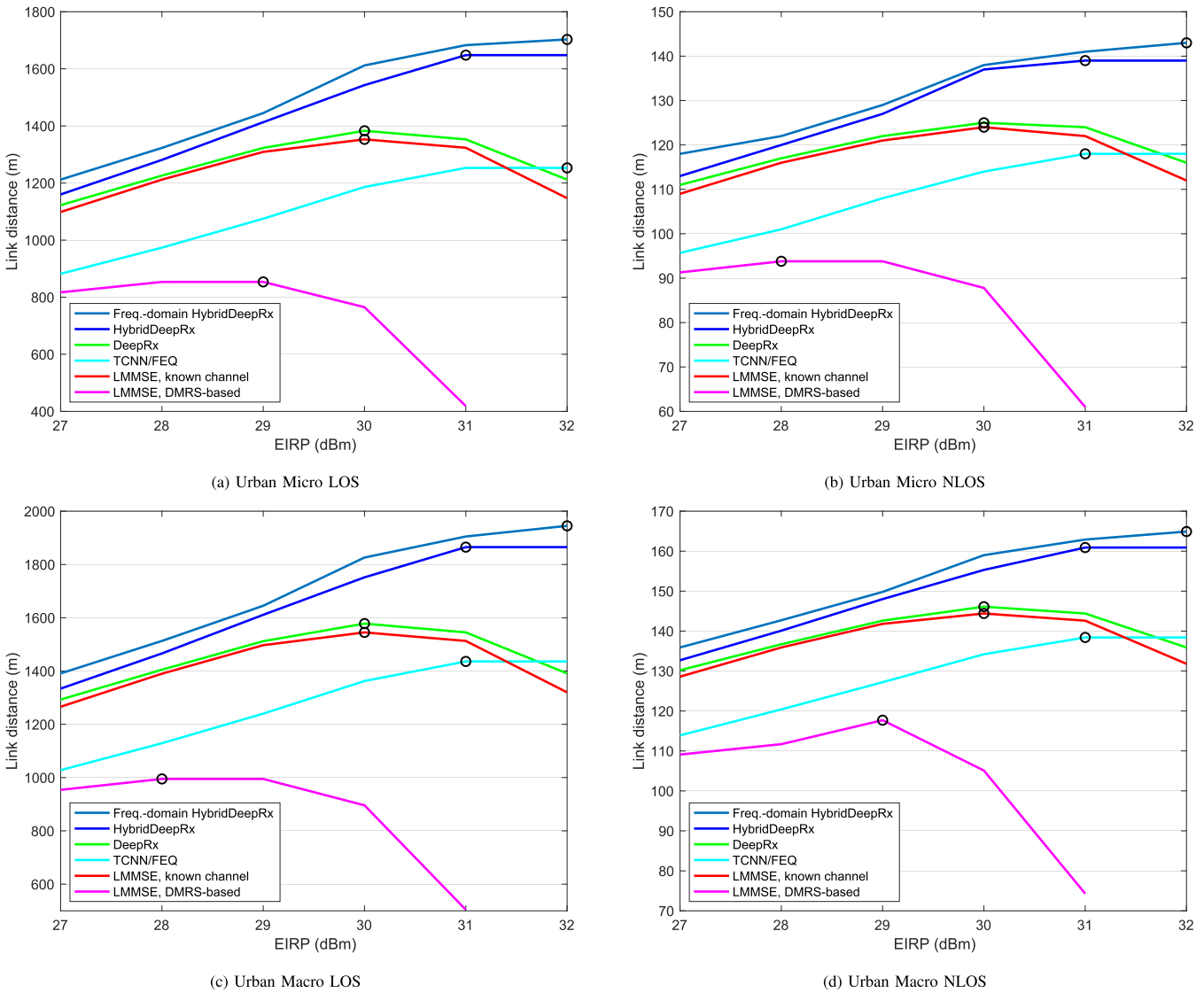


Fig. 13. Link distances with respect to EIRP under different propagation conditions. The maximum coverage of each receiver architecture is denoted by a circle.

coverage. The coverage enhancement achieved by frequency-domain HybridDeepRx ranges from 50% to 100% compared to the DMRS-based LMMSE baseline. While the coverage enhancement of the more light-weight TCNN/FEQ receiver is not quite as substantial, it also outperforms the DMRS-based LMMSE receiver by a clear margin, achieving coverage enhancements in the order of 15%-45%. Therefore, it represents a favorable trade-off between algorithm complexity and link coverage.

Figure 13 illustrates the link coverages with respect to the EIRP, to illustrate the relationship between transmit power and coverage. It is evident that the receivers not able to compensate for nonlinearities have an optimal EIRP which maximizes link coverage, while the frequency-domain HybridDeepRx, TCNN/FEQ, and HybridDeepRx receivers seem to be always able to provide improved coverage when EIRP is increased. This means that the link budget design, when using such ML-based receivers, can focus primarily on the out-of-band

emissions since the higher transmit power is always bound to have a positive impact on the link coverage. This is in contrast to conventional receivers, which do not benefit from higher TX power after the nonlinear distortion becomes the limiting factor.

VI. CONCLUSION

In this article, we presented three novel deep learning based receiver solutions to combat nonlinearly distorted signals due to transmitter PA system. This is achieved by introducing trainable convolutional layers in time-domain, which are particularly suited for dealing with the nonlinear distortion. The proposed receiver architectures are shown to be able to detect even heavily distorted signals, while the benchmark receivers fail to detect such signals reliably. Indeed, the performance gain compared to a conventional linear receiver is several dBs even with reasonable levels of nonlinear distortion.

This translates to as much as a 100% increase in link coverage. In particular, the proposed frequency-domain HybridDeepRx outperforms the other presented ML-based receivers while utilizing only frequency-domain inputs to make the hardware implementation simpler and maintaining the same number of layers as DeepRx. These findings pave the way towards more power efficient radios where the effects of hardware impairments can be dealt with the help of deep learning aided receiver solutions. Moreover, extending the HybridDeepRx to cover actual digital MIMO transmitters and receivers, as well as mapping of the proposed receivers to actual hardware and carrying out full complexity comparisons accordingly are important future work items.

REFERENCES

- [1] G. Liu et al., "Coverage enhancement and fundamental performance of 5G: Analysis and field trial," *IEEE Commun. Mag.*, vol. 57, no. 6, pp. 126–131, Jun. 2019.
- [2] I. P. Nasarre, T. Levanen, K. Pajukoski, A. Lehti, E. Tirola, and M. Valkama, "Enhanced uplink coverage for 5G NR: Frequency-domain spectral shaping with spectral extension," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1188–1204, 2021.
- [3] H. Holma, A. Toskala, and T. Nakamura, *5G Technology: 3GPP New Radio*, 1st ed. Hoboken, NJ, USA: Wiley, 2019.
- [4] U. Gustavsson et al., "Implementation challenges and opportunities in beyond-5G and 6G communication," *IEEE J. Microw.*, vol. 1, no. 1, pp. 86–100, Jan. 2021.
- [5] L. Anttila, A. Brihuega, and M. Valkama, "On antenna array out-of-band emissions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1653–1656, Dec. 2019.
- [6] A. Brihuega et al., "Piecewise digital predistortion for mmWave active antenna arrays: Algorithms and measurements," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 9, pp. 4000–4017, Sep. 2020.
- [7] S. Gronemeyer and A. McBride, "MSK and offset QPSK modulation," *IEEE Trans. Commun.*, vol. C-24, no. 8, pp. 809–820, Aug. 1976.
- [8] I. Peruga Nasarre, T. Levanen, and M. Valkama, "Constrained PSK: Energy-efficient modulation for sub-THz systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–7.
- [9] Y. Rahmatallah and S. Mohan, "Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1567–1592, 4th Quart., 2013.
- [10] J. Guerreiro, R. Dinis, and P. Montezuma, "Optimum and sub-optimum receivers for OFDM signals with strong nonlinear distortion effects," *IEEE Trans. Commun.*, vol. 61, no. 9, pp. 3830–3840, Sep. 2013.
- [11] C. Yang, X. Liu, Y. L. Guan, and R. Liu, "Fast GAMP algorithm for nonlinearly distorted OFDM signals," *IEEE Commun. Lett.*, vol. 25, no. 5, pp. 1682–1686, May 2021.
- [12] S. V. Zhidkov and R. Dinis, "Belief propagation receivers for near-optimal detection of nonlinearly distorted OFDM signals," in *Proc. IEEE 89th Veh. Technol. Conf. (VTC-Spring)*, Apr. 2019, pp. 1–6.
- [13] D. Neumann, T. Wiese, and W. Utschick, "Learning the MMSE channel estimator," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2905–2917, Jun. 2018.
- [14] H. He, C.-K. Wen, S. Jin, and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 852–855, Oct. 2018.
- [15] Z. Chang, Y. Wang, H. Li, and Z. Wang, "Complex CNN-based equalization for communication signal," in *Proc. IEEE 4th Int. Conf. Signal Image Process. (ICSIP)*, Jul. 2019, pp. 513–517.
- [16] Y. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
- [17] O. Shental and J. Hoydis, "'Machine LLRning': Learning to softly demodulate," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–7.
- [18] X. Gao, S. Jin, C.-K. Wen, and G. Y. Li, "ComNet: Combination of deep learning and expert knowledge in OFDM receivers," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2627–2630, Dec. 2018.
- [19] H. He, S. Jin, C. Wen, F. Gao, G. Y. Li, and Z. Xu, "Model-driven deep learning for physical layer communications," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 77–83, Oct. 2019.
- [20] N. Samul, T. Diskin, and A. Wiesel, "Learning to detect," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2554–2564, May 2019.
- [21] L. Xu, F. Gao, W. Zhang, and S. Ma, "Model aided deep learning based MIMO OFDM receiver with nonlinear power amplifiers," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2021, pp. 1–7.
- [22] M. Honkala, D. Korpi, and J. M. J. Huttunen, "DeepRx: Fully convolutional deep learning receiver," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3925–3940, Jun. 2021.
- [23] D. Korpi, M. Honkala, J. M. J. Huttunen, and V. Starck, "DeepRx MIMO: Convolutional MIMO detection with learned multiplicative transformations," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–7.
- [24] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.
- [25] Z. Zhao, M. C. Vuran, F. Guo, and S. D. Scott, "Deep-waveform: A learned OFDM receiver based on deep complex-valued convolutional networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 8, pp. 2407–2420, Aug. 2021.
- [26] M. Careem, A. Dutta, and N. Thawdar, "On equivalence of neural network receivers," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–7.
- [27] J. Pihlajasalo et al., "Deep learning based OFDM physical-layer receiver for extreme mobility," in *Proc. 55th Asilomar Conf. Signals, Syst., Comput.*, Oct. 2021, pp. 395–399.
- [28] J. Pihlajasalo et al., "HybridDeepRx: Deep learning receiver for high-EVM signals," in *Proc. IEEE 32nd Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Helsinki, Finland, 2021, pp. 622–627. [Online]. Available: <https://ieeexplore.ieee.org/document/9569393>, doi: 10.1109/PIMRC50174.2021.9569393.
- [29] A. Felix, S. Cammerer, S. Dörner, J. Hoydis, and S. Ten Brink, "OFDM-autoencoder for end-to-end learning of communications systems," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018, pp. 1–5.
- [30] Y. Xie, X. Liu, K. C. Teh, and Y. L. Guan, "Robust deep learning-based end-to-end receiver for OFDM system with non-linear distortion," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 340–344, Feb. 2022.
- [31] S. Zheng, S. Chen, and X. Yang, "DeepReceiver: A deep learning-based intelligent receiver for wireless communications in the physical layer," *IEEE Trans. Cognit. Commun. Netw.*, vol. 7, no. 1, pp. 5–20, Mar. 2021.
- [32] S. Liu, T. Wang, and S. Wang, "Joint compensation of CFO and IQ imbalance in OFDM receiver: A deep learning based approach," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Jul. 2021, pp. 793–798.
- [33] A. Mohammadian, C. Tellambura, and G. Y. Li, "Deep learning LMMSE joint channel, PN, and IQ imbalance estimator for multicarrier MIMO full-duplex systems," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 111–115, Jan. 2022.
- [34] H. S. Park, E.-Y. Choi, Y. S. Song, S. Noh, and K. Seo, "DNN-based phase noise compensation for sub-THz communications," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2020, pp. 866–868.
- [35] A. Mohammadian, C. Tellambura, and G. Y. Li, "Deep learning-based phase noise compensation in multicarrier systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 10, pp. 2110–2114, Oct. 2021.
- [36] *NR; User Equipment (UE) Radio Transmission and Reception; Part 2: Range 2 Standalone (Release 17)*, document TS 38.101-2, version 17.0.0, 3GPP, Dec. 2020.
- [37] L. Ding et al., "A robust digital baseband predistorter constructed using memory polynomials," *IEEE Trans. Commun.*, vol. 52, no. 1, pp. 159–165, Jan. 2004.
- [38] *NR; Physical channels and modulation*, document TS 38.211, version 16.0.0, 3GPP, Dec. 2019.
- [39] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 630–645.
- [40] *Study on Channel Model for Frequencies From 0.5 to 100 GHz (Release 16)*, document TR 38.901, version 16.0.0, 3GPP, Dec. 2019.
- [41] Mathworks Inc. (2020). *Matlab 5G Toolbox*. [Online]. Available: <https://www.mathworks.com/products/5g.html>
- [42] Q. Dong et al., "A 351 TOPS/W and 372.4 GOPS compute-in-memory SRAM macro in 7 nm FinFET CMOS for machine-learning applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2020, pp. 242–244.



Jaakko Pihlajasalo received the M.Sc. (Tech.) degree in mathematics from the Tampere University of Technology (TUT), Finland, in 2018. He is currently pursuing the Ph.D. degree in mathematics with Tampere University. His research interests include machine learning and statistical modeling, and their applications in wireless communications and positioning.



Jukka Talvitie (Member, IEEE) received the M.Sc. and D.Sc. degrees from the Tampere University of Technology, Finland, in 2008 and 2016, respectively. He is currently a University Lecturer with the Faculty of Information Technology and Communication Sciences, Tampere University (TAU), Finland. His research interests include signal processing for wireless communications, radio-based positioning, radio link waveform design, and radio system design, particularly concerning 5G and beyond mobile technologies.



Dani Korpi (Member, IEEE) received the M.Sc. and D.Sc. degrees (Hons.) in communications engineering and electrical engineering from the Tampere University of Technology (TUT), Finland, in 2014 and 2017, respectively. He is currently a Senior Specialist with Nokia Bell Labs, Espoo, Finland. In particular, he is working on building a native foundation for machine learning in 6G radio networks. His current research interests include applying machine learning to wireless communications, especially on the physical layer. His Ph.D. thesis received the Best Dissertation of the Year Award from TUT and the Finnish Technical Sector's Award for the Best Doctoral Dissertation in 2017.



Alberto Brihuela received the B.Sc. and M.Sc. degrees in telecommunications engineering from the Universidad Politécnica de Madrid, Madrid, Spain, in 2015 and 2017, respectively, and the D.Sc. (Tech.) degree (Hons.) in computing and electrical engineering from Tampere University, Tampere, Finland, in 2022. He is currently an RF System Simulations Engineer with Nokia Mobile Networks, Oulu, Finland. His research interests include statistical and adaptive digital signal processing for compensation of hardware impairments in RF

transceivers, RF algorithms, and RF device modeling.

transceivers, RF algorithms, and RF device modeling.



Mikko Honkala received the D.Sc. (Tech.) degree in computer science from the Helsinki University of Technology (HUT), Finland, in 2007. He worked as a Machine Learning Researcher at the Nokia Research Center. He is currently with Nokia Bell Labs, where he leads the Machine Learning Methods and Applications Group. His research interests include deep learning methods and their applications to new areas, such as wireless communications.



Mikko A. Uusitalo (Senior Member, IEEE) received the M.Sc. (Eng.) and Dr. Tech. degrees and the B.Sc. degree in economics from the Helsinki University of Technology (HUT), in 1993, 1997, and 2003, respectively. He is currently the Head of Research Department of Radio Systems Research Finland at Nokia Bell Labs, Finland. He is leading the European 6G Flagship Projects Hexa-X and Hexa-X-II. Since 2000, he has been at Nokia with various roles, including a Principal Researcher and the Head of International Cooperation at Nokia Research.

He is also a Founding Member of the CELTIC EUREKA and WWRF.



Janne M. J. Huttunen (Member, IEEE) received the M.Sc. degree in medical physics and the Ph.D. degree in physics from the University of Kuopio (UKU), Finland, in 2004 and 2008, respectively. Previously, he worked as an Academic Researcher developing methods for Bayesian inverse problems and applications. He is currently a Senior Researcher with Nokia Bell Labs, Espoo, Finland. His research interests include deep learning in various areas including wireless communications. He received the Best Dissertation Award of the Year from UKU.



Taneli Riihonen (Senior Member, IEEE) received the D.Sc. degree (Hons.) in electrical engineering from Aalto University, Helsinki, Finland, in 2014. He is currently a Tenure-Track Associate Professor with the Faculty of Information Technology and Communication Sciences, Tampere University, Finland. His research interests include communication fundamentals, multicarrier, multiantenna, multihop, and full-duplex wireless techniques with current interest in the evolution of beyond 5G radio systems.



Mikko Valkama (Fellow, IEEE) received the M.Sc. (Tech.) and D.Sc. (Tech.) degrees (Hons.) from the Tampere University of Technology, Finland, in 2000 and 2001, respectively. In 2003, he was with the Communications Systems and Signal Processing Institute, SDSU, San Diego, CA, USA, as a Visiting Research Fellow. He is currently a Full Professor and the Head of the Unit of Electrical Engineering with Tampere University, Finland. His research interests include radio communications, radio localization, and radio-based sensing, with particular emphasis on 5G and 6G mobile radio networks.