Deep-Learning–Aided Joint Detection and Channel Estimation for Massive MIMO–OFDM Systems

Célio A. de Souza Junior and Jaime L. Jacob and Taufik Abrão

***Abstract*— Accurate channel state information (CSI) and efficient multi-user detection are critical for massive MIMO-OFDM. We present a frame-level deep-neural-network (DNN) receiver that *jointly* estimates the channel and detects symbols. Monolithic and spectrum-partitioned architectures are trained with Monte-Carlo data under EPA-7 frequency-selective fading channel model, AWGN, and QAM. Compared with least-squares (LS) and time-domain MMSE (TD-MMSE) followed by zero-forcing, the partitioned DNN achieves up to 10 dB bit-error-rate (BER) gain at 10^{-3} in a 64×64 uplink antenna while keeping latency comparable. The model is robust to cyclic-prefix (CP) suppression and to a $4\times$ pilot reduction, indicating suitability for 5G/6G deployments.**

***Keywords*— Massive MIMO, OFDM, Channel estimation, Deep learning, 5G, 6G.**

I. Introduction

The increasing demands for ultra-high throughput, minimal latency, and stringent energy efficiency in next-generation wireless networks, such as 5G and beyond, require innovative physical layer solutions [1]. Massive Multiple-Input Multiple-Output (MIMO) combined with Orthogonal Frequency Division Multiplexing (OFDM) has emerged as a cornerstone technology to meet these ambitious targets. However, the full realization of Massive MIMO-OFDM benefits critically hinges on the availability of accurate channel state information (CSI) and the deployment of computationally efficient multi-user detection algorithms. Indeed, next-generation wireless networks must deliver terabits per second per square kilometer throughput with sub-millisecond latency and stringent energy budgets [2]. In this sense, massive MIMO-OFDM is the key physical-layer technology that enables those targets, yet its advantages rely on precise CSI and computationally affordable detectors.

The seminal monograph [2] establishes the theoretical and practical foundations necessary for understanding the context of joint detection and channel estimation in massive MIMO-OFDM systems. The work is regarded as a cornerstone in the field, offering insights into scalable multi-antenna technologies. In [3], an adaptive detection scheme using deep learning (DL) to improve detection accuracy in massive MIMO systems is introduced. Moreover, training a neural network (NN) to learn detection patterns addresses the computational complexity of traditional methods, making it highly relevant for joint detection tasks in massive MIMO-OFDM systems.

Traditional approaches to channel estimation, such as Least Squares (LS) and Time-Domain Minimum Mean Square Error (TD-MMSE), when coupled with linear detectors like Zero-Forcing (ZF) or MMSE, often struggle in the massive MIMO context. These methods typically exhibit high computational complexity, usually scaling cubically with the number of antennas, which can be prohibitive for large-scale deployments. Furthermore, their performance is notably susceptible to practical impairments such as imperfect cyclic prefix (CP) utilization and scarcity of pilot resources, which are common in realistic wireless environments.

To overcome these limitations, machine learning (ML), particularly deep learning (DL), has garnered significant research interest as a powerful paradigm for enhancing various aspects of wireless communication systems. The ability of deep neural networks (DNNs) to learn complex patterns and relationships from data without explicit mathematical models makes them highly attractive for tackling the intricate challenges of joint channel estimation and data detection in dynamic and complex Massive MIMO-OFDM systems.

Early investigations of DL for wireless communications demonstrated promising results for OFDM systems. For example, He et al. [4] proposed a DL-based framework for joint channel estimation and signal detection, which showcases notable performance gains, especially under low signal-to-noise ratio (SNR) conditions. Similarly, Ye et al. [5] introduced a DNN architecture that achieved superior performance over conventional methods in OFDM systems, particularly in scenarios characterized by high mobility and rich multipath fading. Building upon such foundational work, the focus has increasingly shifted towards the more complex Massive MIMO regime. Recent studies, such as the work presented in [6] and [7], explore advanced NN architectures for joint channel estimation and symbol detection specifically tailored for MIMO-OFDM systems, while also highlighting challenges like the robustness of pre-trained models in dynamic channel conditions. Deep Expectation-Maximization, as introduced in [8], bridges the gap between classical expectation–maximization and modern DNNs, enabling efficient joint detection and channel estimation in MIMO settings. A critical consideration in these DL-based approaches is the trade-off between achievable performance and computational complexity. Although sophisticated DNNs can offer significant accuracy improvements, their inference latency and resource requirements must be carefully managed for practical deployment. The research by Zhang et al. [9] provided early analyses on the performance and complexity of various DL architectures for these tasks, and this remains an active area of investigation, with ongoing efforts to develop lightweight yet powerful NN solutions that effectively balance these competing factors.

Our work extends the application of DL to joint chan-

Department of Electrical Engineering, State University of Londrina (UEL), PR, Brazil. (e-mails: {celio.junior, jaime.jacob,taufik}@uel.br). This work was partially supported by CAPES (1724/2023), Fundação Araucária PR (CONFAP 2022) e CNPq (314618/2023-6).

nel estimation and data detection specifically for the massive MIMO-OFDM regime. We aim to address the critical performance-complexity trade-off by proposing and evaluating novel DNN-based receiver architectures. The primary contributions of this work are fourfold: ***i***) a comprehensive computational feasibility study of our proposed DL-based joint detection and channel estimation approach, benchmarking its complexity and performance against established classical schemes like LS+ZF and TD-MMSE+ZF; ***ii***) we introduce a scalable convolutional DNN architecture with sub-carrier partitioning; ***iii***) We evaluate their efficiency and efficacy in systems with 64×64 antennas; ***iv***) We provide a detailed comparative analysis against conventional methods, particularly under challenging conditions such as significant pilot reduction, the absence of cyclic prefix, thereby demonstrating the robustness and suitability of the proposed techniques for future wireless deployments.

Notice that recent studies apply NNs either to signal equalization [3] or to joint detection–channel estimation [5], [10], but only for SISO scenarios. By contrast, current wireless standards rely mainly on (massive) M-MIMO architectures, whose much larger channel dimensionality makes both training and real-time inference substantially more computationally demanding.

In our work, we extend *joint* learning to the massive regime and provide specific analyses: **a**) Computational feasibility study of DL-based joint detection and channel estimation in massive MIMO–OFDM, benchmarking its complexity and performance against classical LS+ZF and TD-MMSE schemes. **b**) The scalable DNN architecture (partitioned) evaluated in 64×64 antennas; **c**) Compare LS+ZF and TD-MMSE+ZF under pilot reduction and CP removal.

II. System Model

A. OFDM Signal Model

Let $\mathcal{U} = \{1, \ldots, N_{\text{users}}\}$ denote the set of single-antenna users served by a base-station (BS) with N_{rx} antennas, typically $N_{\text{rx}} \gg N_{\text{users}}$ in the massive-MIMO regime. Every OFDM symbol employs N_{sub} orthogonal sub-carriers and a cyclic prefix (CP) of length $L_{\text{cp}} > L_h$, where L_h is the maximum delay spread in samples.

User $u \in \mathcal{U}$ maps its coded bits onto an M-QAM constellation, producing

$$\mathbf{x}_u \;=\; \big[x_u(0),\, x_u(1), \ldots, x_u(N_{\text{sub}}{-}1)\big]^{\mathsf{T}} \in \mathbb{C}^{N_{\text{sub}} \times 1}.$$

An N_{sub}-point inverse fast fourier transform (IFFT) converts the frequency-domain vector into the time-domain block that is transmitted over the channel. After CP removal and an N_{sub}-point FFT at the BS, the observation on sub-carrier k ($0 \le k < N_{\text{sub}}$) is

$$\mathbf{y}(k) \;=\; \mathbf{H}(k)\,\mathbf{x}(k) \;+\; \mathbf{z}(k), \tag{1}$$

where

$$\mathbf{x}(k) = \begin{bmatrix} x_1(k) \\ \vdots \\ x_{N_{\text{users}}}(k) \end{bmatrix} \in \mathbb{C}^{N_{\text{users}} \times 1}, \quad \mathbf{y}(k) \in \mathbb{C}^{N_{\text{rx}} \times 1},$$

$\mathbf{H}(k) \in \mathbb{C}^{N_{\text{rx}} \times N_{\text{users}}}$ is the MIMO channel frequency response on sub-carrier k, and $\mathbf{z}(k) \sim \mathcal{CN}\big(\mathbf{0}, \sigma^2 \mathbf{I}_{N_{\text{rx}}}\big)$ denotes additive white Gaussian noise.

The stacked vector forms $\mathbf{x} = [\mathbf{x}(0)^{\mathsf{T}} \cdots \mathbf{x}(N_{\text{sub}}{-}1)^{\mathsf{T}}]^{\mathsf{T}}$ and $\mathbf{y}$ are used later for channel estimation and symbol detection.

B. Channel Statistics

The wireless channel is modeled according to the 3GPP Extended-Pedestrian-A (EPA-7) tapped-delay line, whose maximum excess delay is $\tau_{\max} = 410$ ns and rms delay spread $\tau_{\text{rms}} \approx 44$ ns. Each tap $\ell \in \{0, \ldots, 6\}$ follows a zero-mean circularly symmetric complex Gaussian distribution

$$h_{m,u}(\ell) \;\sim\; \mathcal{CN}\big(0,\, p_\ell\big), \tag{2}$$

where m indexes the receive antennas and u the single-antenna users. The tap powers in decibels are

$$\big[p_\ell\big]_{\text{dB}} = \{0,\, -1,\, -2,\, -3,\, -8,\, -17.2,\, -20.8\}\ \text{dB},$$

which are converted to linear scale and normalized such that $\sum_{\ell=0}^{6} p_\ell = 1$. Unless otherwise stated, taps are assumed i.i.d. across antenna pairs (favorable propagation). All simulations adopt block fading over one OFDM frame.

C. Classical Estimation and Detection Methods

We compare the proposed deep learning scheme with classical baselines comprising least-squares (LS) and time-domain minimum mean square error (TD-MMSE) estimators followed by zero-forcing (ZF) detection.

1) Least-Squares (LS) Estimation: Assuming orthogonal pilot symbols, the LS channel estimate for subcarrier k is obtained by

$$\widehat{\mathbf{H}}_{\text{LS}}(k) = \mathbf{Y}_{\text{pilots}}(k)\, \mathbf{X}_{\text{pilots}}^{-1}(k), \tag{3}$$

where $\mathbf{Y}_{\text{pilots}}(k)$ contains the received pilot signals and $\mathbf{X}_{\text{pilots}}(k)$ is a known diagonal pilot matrix.

2) Time-Domain MMSE (TD-MMSE) Estimation: This method exploits the channel's delay-domain correlation. Let $\mathbf{h}$ denote the vectorized channel impulse response. The MMSE estimate is given by

$$\widehat{\mathbf{h}} = \mathbf{R}_h\, \mathbf{S}^H \big(\mathbf{S}\mathbf{R}_h\mathbf{S}^H + \sigma^2\mathbf{I}\big)^{-1} \mathbf{y}_{\text{pil}}, \tag{4}$$

where $\mathbf{R}_h$ is the channel tap covariance matrix, $\mathbf{S}$ is the pilot spreading matrix, and $\mathbf{y}_{\text{pil}}$ is the received pilot vector in the time domain. The estimate $\widehat{\mathbf{h}}$ is then transformed into the frequency domain to reconstruct $\widehat{\mathbf{H}}(k)$.

3) Zero-Forcing (ZF) Detection: Once $\widehat{\mathbf{H}}(k)$ is available from LS or TD-MMSE, the data symbols are detected via the ZF equalizer:

$$\widehat{\mathbf{x}}(k) = \widehat{\mathbf{H}}^{\dagger}(k)\, \mathbf{y}(k), \tag{5}$$

where $\widehat{\mathbf{H}}^{\dagger}(k)$ denotes the Moore–Penrose pseudo-inverse of the estimated channel matrix.

These classical approaches form the benchmark against which our DNN-based joint detector-estimator is evaluated in terms of BER performance and computational efficiency.

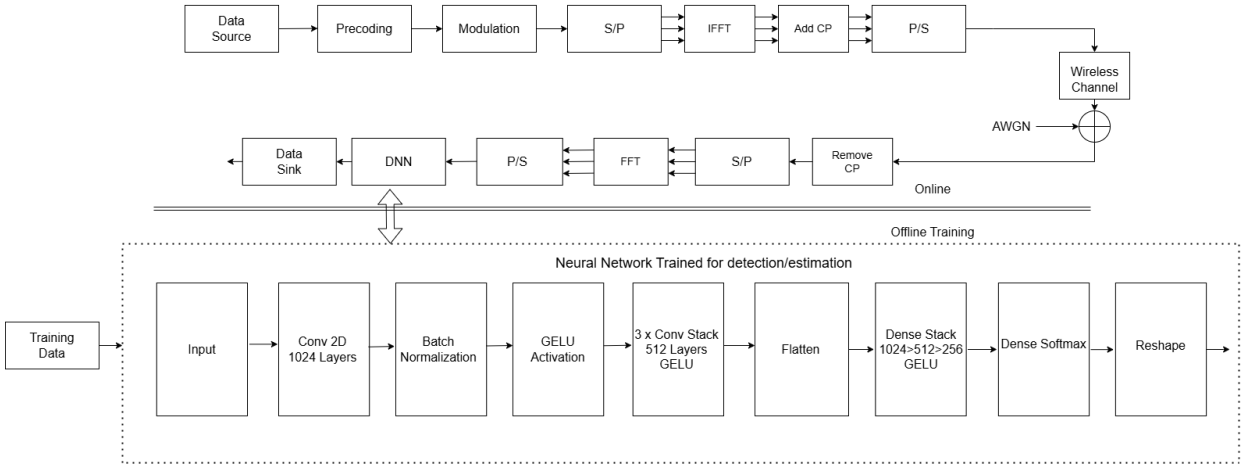


Fig. 1: Proposed MIMO-OFDM receiver: pilot insertion, FFT, feature stacking and the sub-carrier-partitioned DNN.

III. Proposed DNN Architecture

The network replaces the conventional *channel-estimation → linear-detection* cascade with a single data-driven module as illustrated by 1. Real and imaginary parts of the data grid $\mathbf{Y}_{\mathrm{data}}$ and the pilot grid $\mathbf{Y}_{\mathrm{pil}}$ are stacked, producing input feature maps.

Sub-carrier partitioning: To keep the parameter count tractable for massive MIMO we partition the N_{sub} sub-carriers into G disjoint groups of equal size, processed by G identical *sub-networks* in parallel; here we set $G = 16$, so that each sub-network handles $N_{\mathrm{sub}}/G = 4$ adjacent sub-carriers.

Each sub-network receives an input tensor with dimensions $N_{\mathrm{rx}} \times (N_{\mathrm{sub}}/G) \times 2N_{\mathrm{blocks}}$, corresponding, respectively, to the receive antennas, the adjacent sub-carriers in the current group, and the real/imaginary parts of both data and pilot grids across N_{blocks} consecutive OFDM symbols.

The input tensor is first processed by a 2-D convolution with 1 024 filters and a 1×1 kernel; the output is normalised by batch normalisation, activated by GELU, and regularised with a dropout of 0.2. A stack of three additional convolutional layers, each with 512 filters and GELU activation, refines these features and preserves spatial locality. The resulting feature maps are then flattened and passed to a fully connected multilayer perceptron $1024 \to 512 \to 256$, all layers again using GELU. Finally, a soft-max layer of width $N_{\mathrm{users}} \times (N_{\mathrm{sub}}/G)$ produces the posterior probabilities of the QPSK symbols conveyed by the sub-carriers in the current group for every user, completing the joint estimation–detection task.

All G outputs are concatenated to recover the full block decision. This design reduces the total parameter count compared with a monolithic dense model and brings the inference time down.

The network minimises sparse categorical cross-entropy with RMSprop (learning rate 4×10^{-4}). Early stopping with a patience of eight epochs prevents over-fitting; the best weights are selected on a validation set spanning 0–20 dB SNR.

IV. Numerical Results and Implementation

The parameter values adopted are summarized in Table I. Training and test datasets were generated in MATLAB using these parameters. Model training was carried out in Python/TensorFlow on a Google Cloud server equipped with an NVIDIA A100. Performance was assessed via Monte-Carlo simulations: the frozen network was evaluated on 10^6 additional OFDM frames per SNR point, and its BER and latency were compared against classical ZF, LS, and TD-MMSE baselines under identical channel realisations on local CPU.

TABLE I: Simulation parameters

Parameter	**Value**
MIMO–OFDM	
N_{rx}	64
N_{users}	64
N_{sub}	64 (fixed)
Modulation (M-QAM)	4 (fixed)
SNR range (dB)	0 – 20
CP length	$N_{\mathrm{cp}} = 0.25\, N_{\mathrm{sub}}$
Channel taps L	7
Channel model	
3GPP EPA-7, Rayleigh	$\tau_{\mathrm{rms}} = 44$ ns
N_{pil} (number of pilots)	64,32,16
Estimator	LS, TD-MMSE
Equaliser	ZF
Neural network (CNN + FC)	
Conv layers	1×1/1024, 3×3/512
Dense head	$1024 \to 512 \to 256 \to 256$
Hidden act.	GELU
Output act.	Softmax
Loss	Categorical cross-entropy
Optimiser	RMSprop, $\alpha = 4 \times 10^{-4}$
Training frames	10^7
Batch size	32
Epochs	50
Hardware	Google Colab GPU

A. Pilot Reduction

Three separate DNNs were trained—one for each pilot density of 64, 32 and 16 pilots—using the same 64 × 64 MIMO-OFDM configuration, $N_{\mathrm{sub}} = 64$ and 4-QAM modulation. Fig. 2 compares their BER curves with the classical LS and TD-MMSE baselines that operate under the *same* pilot budgets. Reducing the pilot count from 64 to 32 moves TD-MMSE

curve to the right by roughly **4 dB** in the low-SNR region, whereas the corresponding DNN experiences a loss of **2 dB** and still outperforms TD-MMSE that keeps all 64 pilots. Reducing further to 16 pilots adds only $\approx$ **2.5 dB** of penalty for the DNN; even with half the pilots, the DNN attains BER $< 10^{-5}$ at 12 dB, still surpassing the TD-MMSE solution that uses twice as many pilots. With the full set of 64 pilots, the DNN exceeds the ZF bound under perfect CSI by nearly 1 dB, highlighting its superior pilot efficiency in the massive-MIMO regime and confirming the trends depicted in Fig. 2.

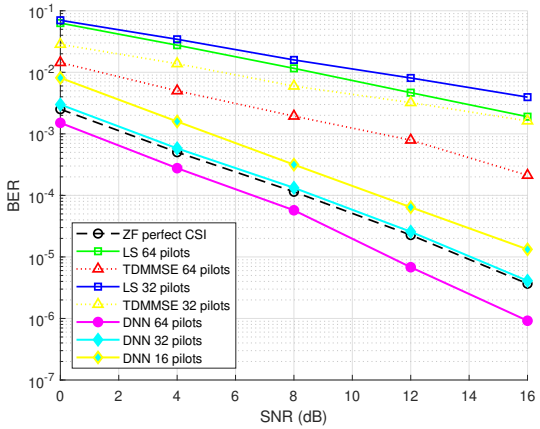


Fig. 2: BER vs. SNR for a 64×64 massive-MIMO link (64 subcarriers, 4-QAM) under pilot budgets of 64, 32, and 16 symbols; in each case the DNN equalizer is tested with the *same* pilot count used during its training.

To investigate whether a single network can be generalized to pilot budgets different from those seen during training, we have trained two independent DNNs under the same conditions as Fig. 2. In Fig. 3a, the DNN was trained exclusively with 64 pilots, whereas Fig. 3b shows the same architecture trained with 32 pilots. *Key observations*:

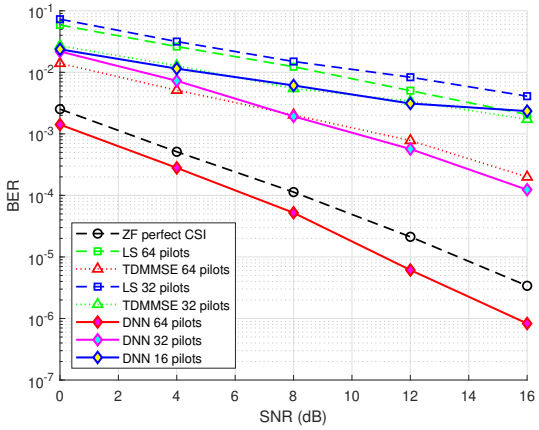
Adaptation to under-sampling: When DNN trained over 64 pilots is forced to operate with 32 or 16 pilots, the curves shift by more than **10 dB** in the BER range 10^{-3}.

Intrinsic robustness of low-pilot training: The DNN-32, learned with a sparser pilot pattern, shows a noticeably smaller BER penalty when pilots are further removed, indicating that networks trained with less information internalize features that generalize better to even harsher pilot-scarce scenarios.

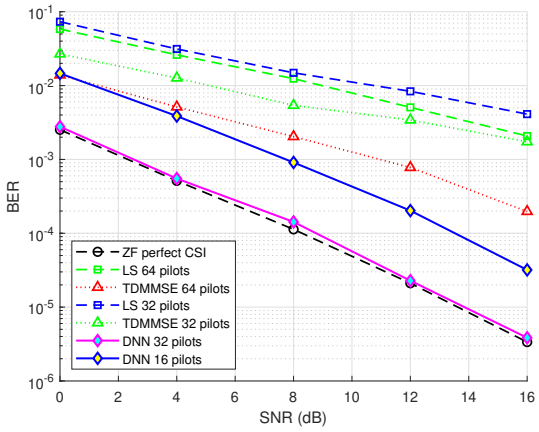
B. Absence of Cyclic Prefix

When the CP is omitted, as illustrated in Fig. 4, each OFDM symbol experiences frequency-selective inter-symbol interference (ISI) that the classical linear estimators are unable to cancel. Both **LS** and **TD-MMSE** develop a clear error floor near $\mathrm{BER} \approx 10^{-2}$: By contrast, the proposed **DNN** remains far more resilient. Its curve shifts by only $\sim$ 2.1 **dB**. At $\mathrm{SNR} = 12$ dB, for instance, the DNN without CP still attains $\mathrm{BER} \approx 7 \times 10^{-5}$, whereas LS and TD-MMSE stay two orders of magnitude higher.

It is noteworthy that the network *was trained* with *a CP* using the same pilot budget employed for inference. The fact that its performance degrades only modestly when the CP is *removed* at test time indicates that the learned non-linear equaliser has internalised a channel-invariant strategy that mitigates the ISI introduced by the missing guard interval, rather than merely memorising CP-specific statistics.



(a) DNN Trained over 64 pilots.



(b) DNN Trained over 32 pilots.

Fig. 3: BER vs. SNR—64×64 M-MIMO, 64-subcarriers 4-QAM; DNN trained on N_{pil} pilots, and tested at $\leq N_{\text{pil}}$.

V. Complexity Analysis

Table II reports the average inference time per OFDM frame on a 3.0 GHz quad-core CPU for a 64×64 M-MIMO link with 64 subcarriers and 4-QAM modulation. The neural equaliser is split into 16 parallel subnetworks, each with about 1.6×10^7 trainable parameters. Although the DNN is roughly five times slower than TD-MMSE and ten times slower than the The LS+ZF baseline remains in the same latency scale, suggesting practical viability for specific applications.

TABLE II: Average execution time per OFDM frame

Method	Time (s)
LS + ZF	0.0877
TD-MMSE	0.1613
DNN	0.8617

From a learning-theoretic viewpoint, the actual bottleneck is *training*: the weight count grows almost linearly with the array size and pilot grouping. In ultra-large arrays, a single subnetwork, if huge blocks of subcarriers are assigned, the number of parameters can exceed 5×10^8 parameters, demanding GPUs with extensive memory usage, motivating future work on pruning, quantization, and distributed training.

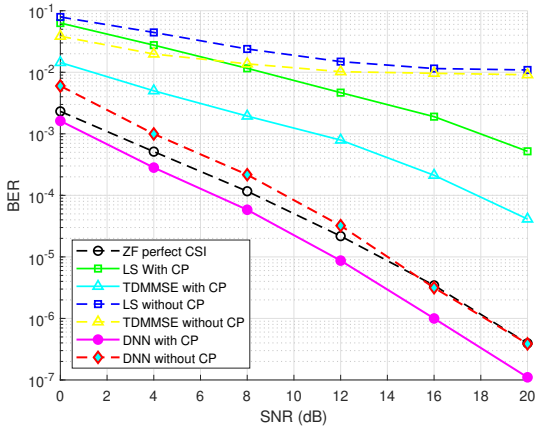


Fig. 4: Impact of removing the CP on BER for a M-MIMO link 64×64, 4-QAM, with 64 subcarriers.

VI. Conclusions

The proposed convolutional neural network architecture, especially in its subcarrier-partitioned configuration, proves certain robustness in challenging scenarios that are critical for practical implementations in 5G and 6G systems.

The numerical results acquired demonstrate that the DNN-based receiver maintains reliable performance even under conditions of aggressive pilot reduction, showing a degradation of only 2 dB when the number of pilots is reduced from 64 to 32, and only an additional 2.5 dB when reduced to 16 pilots. This behavior significantly contrasts with classical LS and TD-MMSE methods, which suffer substantially greater degradations. Notably, even with half the pilots (32), the proposed DNN still outperforms the TD-MMSE solution that uses the complete set of 64 pilots, achieving BER $\leq 10^{-5}$ at 12 dB. With the full set of pilots, the DNN exceeds the ZF bound under perfect CSI by nearly 1 dB, highlighting its superior efficiency in the Massive MIMO regime.

The model demonstrated resilience to the complete removal of the cyclic prefix (CP), a condition that normally introduces severe inter-symbol interference (ISI). While classical linear estimators develop an error floor near BER $\approx 10^{-2}$, the proposed DNN approach experiences a shift of only approximately 2.1 dB in its performance curve. At SNR = 12 dB, for example, the DNN without CP still achieves BER $\approx 7\times10^{-5}$, while LS and TD-MMSE remain two orders of magnitude higher. This result is particularly significant considering that the network was trained with CP, indicating that the learned non-linear equalizer internalized a channel-invariant strategy that efficiently mitigates the ISI introduced by the absence of the guard interval.

The computational complexity analysis revealed that, despite the significant BER improvement, the DNN exhibits latency on the same order of magnitude as the TD-MMSE and LS methods. This difference, although present, is not prohibitive for many practical applications, especially given the substantial performance gains achieved.

From a computational learning perspective, we identified that the main bottleneck lies in training, as the number of weights grows almost linearly with the dimensions that define the wireless scenario. Training these M-MIMO DNNs demands high-memory GPUs, often impractical in real-time operation. The strategy of partitioning the model into subnetworks partially mitigates the scalability problem, but the memory footprint and the number of floating-point operations remain substantial.

Experiments with alternative setups show that, when the DNN is trained on the harsher i.i.d. Rayleigh channel with sparser pilots, its BER rises only modestly after further pilot cuts—evidence that scarcity forces the model to learn features that transfer well. The same mechanism is expected to hold for milder Rician links, and future work will extend the study to vehicular, high-Doppler channels. Although the achieved performance is encouraging, the overall balance between spectral efficiency and computational burden does not yet fully meet the optimization targets envisioned for Massive MIMO-OFDM applications in real-world scenarios. Future work could explore more efficient network architectures, model quantization and pruning techniques, or hybrid approaches that combine the power of deep learning with traditional analytical methods to achieve a better balance between performance and complexity.

References

[1] F. Boccardi, R. W. H. Jr., A. Lozano, T. L. Marzetta, and P. Popovski, "Five Disruptive Technology Directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[2] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[3] X. Zhang, K. Huang, and E. G. Larsson, "Adaptive Massive MIMO Detection with Deep Learning," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5437–5451, 2020.

[4] H. He, C. Wen, S. Jin, and G. Y. Li, "Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems," *IEEE Wireless Communications Letters*, vol. 7, no. 5, pp. 852–855, 2018.

[5] H. Ye, G. Y. Li, and B. H. Juang, "Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems," *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 114–117, 2018.

[6] C. Silpa, A. Vani, and K. R. Naidu, "Deep Learning Based Channel Estimation for MIMO–OFDM System with Modified ResNet Model," *Indian Journal of Science and Technology*, vol. 16, no. 2, pp. 97–108, 2023.

[7] Y. Sun, H. Shen, B. Li, W. Xu, P. Zhu, N. Hu, and C. Zhao, "Trainable Joint Channel Estimation, Detection, and Decoding for MIMO URLLC Systems," *IEEE Transactions on Wireless Communications*, vol. 23, no. 9, pp. 12 172–12 188, 2024.

[8] Y. Zhang, J. Sun, J. Xue, G. Y. Li, and Z. Xu, "Deep Expectation–Maximization for Joint MIMO Channel Estimation and Signal Detection," *IEEE Transactions on Signal Processing*, vol. 70, pp. 4483–4497, 2022.

[9] C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.

[10] W. de Souza Junior and T. Abrão, "Machine Learning-Based Methods for Joint Detection and Channel Estimation in OFDM Systems," *Internet Technology Letters*, vol. 6, no. 2, 2023.