End-to-End AI-Native Physical Layer for Robust 6G MIMO-OFDM: Adaptive Constellation Shaping and Attention-Based Detection

Fateh Bouguerra
Department of Electronics, University of Batna 2, Algeria
f.bouguerra@univ-batna2.dz (corresponding author)

Imed Benacer
Department of Telecommunication and Networking, Institute of Science and Applied Technology, University of Oum El Bouaghi, Algeria
benacer.imad@univ-oeb.dz

Lamir Saidi
Department of Electronics, University of Batna 2, Algeria
l.saidi@univ-batna2.dz

Received: 7 May 2026 | Revised: 2 June 2026 and 8 June 2026 | Accepted: 11 June 2026

 DOI: https://doi.org/10.48084/etasr.19842

ABSTRACT

Sixth-generation (6G) physical-layer designs require robustness against non-analytical channel distortions and hardware impairments that violate classical linear assumptions. This paper presents an AI-native framework for MIMO-OFDM systems that jointly optimizes adaptive constellation shaping and neural detection through end-to-end learning. The proposed model employs a differentiable channel layer incorporating Rayleigh fading, power amplifier nonlinearities, and phase noise, enabling gradient-based optimization of complex constellation coordinates under strict average power constraints. The receiver utilizes a Real-Valued Feedforward Neural Network with Spatial Attention (RFNN-SA) to dynamically weight fading streams and mitigate channel estimation errors. Extensive simulations demonstrate that the proposed model achieves a 3.2 dB SNR gain at BER=10^{-3} for 16-QAM, 3.8 dB for 64-QAM, and 4.1 dB at BER=10^{-2} for 256-QAM over MMSE detection at equivalent operating points. Under realistic CSI uncertainty, performance degrades by only 22%, compared to 52% for classical baselines. With a 0.95 ms physical-layer detection inference latency, the proposed architecture provides a computationally efficient and impairment-resilient foundation for practical 6G physical-layer deployments.

Keywords-6G; AI-native physical layer; adaptive constellation shaping; MIMO-OFDM; neural detection; Rayleigh fading; end-to-end learning

I. INTRODUCTION

The transition toward 6G wireless networks demands Physical-layer (PHY) architectures that transcend classical model-based paradigms. Conventional systems rely on analytically tractable assumptions—such as linear time-invariant channels, AWGN, and ideal transceiver hardware—that are increasingly invalidated by vehicular mobility, terahertz propagation, reconfigurable surfaces, and pervasive hardware nonlinearities [1, 2]. Standard MIMO-OFDM receivers utilize fixed rectangular QAM constellations paired with linear equalizers (e.g., ZF, MMSE) [3]. Although computationally efficient, these approaches are fundamentally limited: rectangular constellations ignore transmitter Channel State Information (CSI), linear detectors amplify noise in ill-conditioned channels, and modular designs prevent cross-layer optimization [4].

To overcome these barriers, deep learning advances have inspired end-to-end PHY frameworks that treat communication as a differentiable pipeline [5-7]. Autoencoder baselines have demonstrated feasibility in AWGN [5], while subsequent studies extended learning to MIMO detection [6], OFDM systems [7], signal classification [8], and neural interference mitigation [9]. Attention mechanisms have also been adapted for spatial stream weighting [10], supported by open-source

differentiable channel libraries such as Sionna [11]. Recently, the AI-native PHY paradigm has expanded from discriminative learning toward generative frameworks, including flow matching and diffusion models [12, 13]. While generative methods excel at complex channel modeling and waveform synthesis, discriminative receiver architectures remain critical for low-complexity, real-time symbol detection.

Despite these algorithmic advances, practical deployment remains hindered by: (i) neglecting concurrent hardware impairments during training, (ii) the absence of explicit power constraints, (iii) fragility under CSI estimation errors, and (iv) insufficient complexity analysis for real-time baseband processing. Comprehensive surveys highlight this simulation-to-deployment gap [14, 15], emphasizing the need for hardware-aware, latency-constrained optimization. While prior works investigate end-to-end learning, attention-based MIMO detection, or differentiable modeling in isolation, few address joint transmitter-receiver optimization under concurrent hardware impairments, CSI uncertainty, and strict latency constraints.

The specific contribution of this study bridges this gap through the integrated design and benchmarking of:

- RFNN-SA: A jointly optimized learnable constellation geometry and attention-based neural detector.
- Impairment-Aware Curriculum Training: A methodology that explicitly models PA nonlinearity and phase noise.
- Deployment Evaluation: A hardware-ready complexity-latency evaluation across 16/64/256-QAM modulations.

By integrating differentiable impairment models, enforcing hard power constraints, and coupling spatial attention with curriculum learning, this approach ensures graceful degradation under realistic, 3GPP-aligned fading [16, 17].

II. PROBLEM STATEMENT

Consider a MIMO-OFDM link with N_t transmit and N_r receive antennas. The frequency-domain received signal at subcarrier k is:

$$\mathrm{y}[k] = \mathrm{H}[k]\mathrm{x}[k] + \mathrm{n}[k] \quad (1)$$

where $\mathrm{H}[k] \in \mathbb{C}^{N_r \times N_t}$ follows Rayleigh fading ($h_{ij} \sim \mathcal{CN}(0,1)$), and $\mathrm{n}[k] \sim \mathcal{CN}(0, \sigma^2 \mathrm{I})$ is AWGN.

Practical 6G systems face CSI uncertainty, which arises from pilot limitations and mobility, inducing estimation errors $\widehat{\mathrm{H}} = \mathrm{H} + \mathrm{E}$, with MSE $\sigma_e^2 \in [10^{-4}, 10^{-2}]$ [15]. They also face hardware nonlinearities that present power amplifier saturation and oscillator phase noise that violate linear superposition, distorting constellation geometry [17]. ZF and MMSE detectors compute:

$$\hat{\mathrm{x}}_{\mathrm{ZF}} = (\mathrm{H}^H\mathrm{H})^{-1}\mathrm{H}^H\mathrm{y} \quad (2)$$

or:

$$\hat{\mathrm{x}}_{\mathrm{MMSE}} = (\mathrm{H}^H\mathrm{H} + \frac{\sigma^2}{P}\mathrm{I})^{-1}\mathrm{H}^H\mathrm{y} \quad (3)$$

Both suffer from noise amplification under high condition numbers and fixed decision boundaries that ignore fading statistics [3]. Maximum likelihood detection is optimal but scales as $\mathcal{O}(M^{N_t})$, becoming intractable for $M \geq 64$ [18].

The separation of modulation, equalization, and detection prevents holistic optimization. The proposed model bridges this gap by learning a unified mapping that adapts to instantaneous channel realizations while maintaining hardware and latency constraints.

III. SYSTEM MODEL

Consider a point-to-point 6G MIMO link, in which the transmitter employs an adaptive constellation whose geometry is optimized offline via end-to-end training and deployed with fixed, channel-statistics-aware weights during inference. The system operates over an impaired wireless channel that includes fading, interference, and nonlinear hardware distortions. Unlike classical PHY design, this model does not assume any closed-form knowledge of the transition probability. Instead, the model interacts with the environment through observations and performance feedback, thereby enabling adaptation to unknown and time-varying effects.

Figure 1 illustrates the end-to-end architecture. The transmitter maps bits to learnable complex symbols, applies OFDM modulation, and transmits through a stochastic channel layer. The receiver processes real-imaginary split signals through an attention-based detector to recover bits.

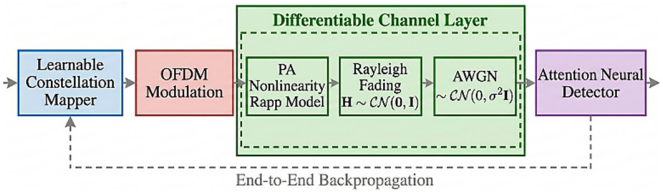

Fig. 1. End-to-end AI-native MIMO-OFDM architecture.

The transmitted symbols satisfy:

$$\frac{1}{M}\sum_{i=1}^{M} |\, c_i \,|^2 = 1 \quad (4)$$

To strictly enforce the average power constraint (4) during gradient descent, a differentiable power normalization layer is applied prior to channel transmission:

$$\tilde{\mathrm{s}}[k] = \frac{\mathrm{s}[k]}{\sqrt{\frac{1}{M}\sum_{j=1}^{M}\|c_j\|^2 + \epsilon}} \quad (5)$$

where $\epsilon = 10^{-8}$ ensures numerical stability. This layer scales all constellation points proportionally at each forward pass, preserving differentiability while maintaining unit average power. After CP insertion, time-domain signals undergo IFFT. The channel applies frequency-selective Rayleigh fading with log-normal shadowing ($\sigma_{\mathrm{sh}} = 8$ dB) [16] while baseline 16/64/256-QAM constellations are normalized to unit average power. Received signals are processed via FFT and fed to the neural detector.

The Rapp model [17] employed here is inherently memoryless. While it effectively captures severe AM/AM

saturation, practical ultra-wideband 6G waveforms often introduce significant PA memory effects driven by thermal dynamics and bias circuit interactions. Such effects are typically characterized via complex behavioral frameworks like Volterra series or generalized memory polynomials. Extending this framework to incorporate memory-aware PA models remains an important avenue for future work. Consequently, the results presented here should be interpreted as a foundational physical-layer proof-of-concept under the memoryless baseline, rather than an immediate guarantee of performance on wideband RF testbeds.

IV. RFNN-SA ARCHITECTURE AND TRAINING

A. Neural Network Specification and Development

The model employs a Real-Valued Feedforward Neural Network with Spatial Attention (RFNN-SA), specifically engineered for complex MIMO-OFDM physical layers. Direct complex-valued gradient optimization can suffer from phase ambiguity and optimization instability under fading; thus, a real-imaginary split representation is adopted to ensure stable backpropagation while preserving geometric channel properties:

$$\mathrm{y}_{\mathrm{rv}} = [\Re(\mathrm{y}); \Im(\mathrm{y})] \in \mathbb{R}^{2N_r} \quad (6)$$

The architecture operates on a per-subcarrier basis following OFDM demodulation (FFT and CP removal). The input to the detector is a real-valued vector $\mathrm{y}_{\mathrm{rv}} \in \mathbb{R}^{2N_r}$ per subcarrier, constructed by stacking the real and imaginary parts of the received signal. Channel state information is not provided as an explicit input; instead, the network implicitly learns equalization and detection from the noisy observations alone. The output layer produces $K = N_t \log_2 M$ soft-bit probabilities $\hat{\mathrm{b}} \in (0,1)^K$ per subcarrier, corresponding to the transmitted bit stream for all spatial streams.

The architecture comprises the following.

1) Input Projection

Linear layer mapping $2N_r \rightarrow 256$, followed by Layer Normalization and ReLU (Rectified Linear Unit).

2) Hidden Representation

Two fully connected layers (256→ 128 → 64) with Leaky ReLU ($\alpha = 0.1$) to preserve gradient flow in low-SNR regimes.

3) A Spatial Attention Module that Computes Stream Weights

$$\alpha_m = \mathrm{softmax}(\mathrm{w}^T \tanh(\mathrm{V}\mathrm{h}_m)) \quad (7)$$

where $\mathrm{h}_{\mathrm{m}} \in \mathbb{R}^{64}$ is the hidden feature representation corresponding to the m-th receive spatial dimension, $\mathrm{V} \in \mathbb{R}^{\mathrm{d}\times 64}$ is a learnable projection matrix ($d = 32$), and $\mathrm{w} \in \mathbb{R}^{\mathrm{d}}$ is a trainable attention query vector. The scalar weights $\alpha_{\mathrm{m}} \in [0,1]$ satisfy $\sum_{\mathrm{m}} \alpha_{\mathrm{m}} = 1$ and are applied over the N_r receive spatial dimensions, dynamically emphasizing reliable antenna streams while suppressing faded or noisy ones [10]. No cross-subcarrier attention is applied; each subcarrier is processed independently.

4) Bit Decoding Head

Final layer $64 \rightarrow K$ ($K = N_t \log_2 M$) with sigmoid activation producing soft probabilities $\hat{\mathrm{b}} \in (0,1)^K$

B. Novelty and Development Rationale

The RFNN-SA architecture diverges from generic classifiers through three PHY-specific innovations:

- Real-Imaginary Split Design: Preserves complex channel geometry while enabling stable real-valued optimization, avoiding phase ambiguity in complex-valued networks.
- Attention-Driven Equalization: Replaces fixed matrix inversion with data-dependent stream weighting, inherently adapting to instantaneous condition numbers without explicit CSI inversion.
- Joint Constellation-Detection Optimization: The transmitter's learnable embedding matrix C and receiver's attention weights are co-trained via Monte Carlo channel sampling, ensuring that constellation points evolve toward regions that minimize decision boundary overlap under fading statistics [5].

Training minimizes the bit-level Binary Cross-entropy (BCE) loss:

$$\mathcal{L} = -\frac{1}{B} \sum_{b,k} [\, b_k \log \hat{b}_k + (1 - b_k) \log (1 - \hat{b}_k)] \quad (8)$$

In Bit-Interleaved Coded Modulation (BICM) systems operating over non-Gaussian fading channels, minimizing the bit-level BCE does not maximize the unconstrained Shannon mutual information. Instead, it can be interpreted as maximizing a BICM achievable information rate objective and is closely related to the Generalized Mutual Information (GMI) commonly used for bit-metric decoding, providing a rigorous lower bound on the GMI for practical bit-level decoders [19]. This interpretation provides a more rigorous information-theoretic foundation for the proposed optimization framework for joint shaping and receiver detection optimization when utilizing learned non-uniform constellations.

Curriculum learning progressively introduces Rayleigh fading, PA nonlinearity, and CSI errors (σ_e^2 up to 10^{-2}) to prevent gradient explosion during early epochs.

Figures 2 through 4 detail the AI architecture and its optimization trajectory. Figure 2 depicts the proposed RFNN-SA, emphasizing real-imaginary signal splitting, the multi-layer perceptron backbone, and the attention module that dynamically weights spatial streams under fading. Figure 3 illustrates curriculum learning convergence, demonstrating stable loss reduction as the model progressively encounters AWGN, Rayleigh fading, and full hardware impairments. Figure 4 visualizes constellation evolution from a rigid rectangular grid to a non-uniform, channel-adapted geometry that minimizes decision boundary overlap.

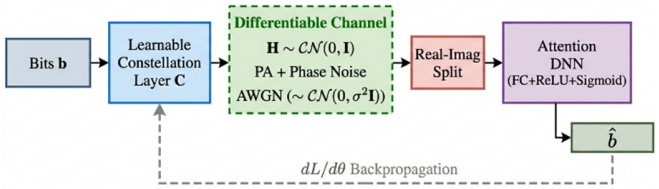

Fig. 2. AI-native PHY architecture with attention-based detection.

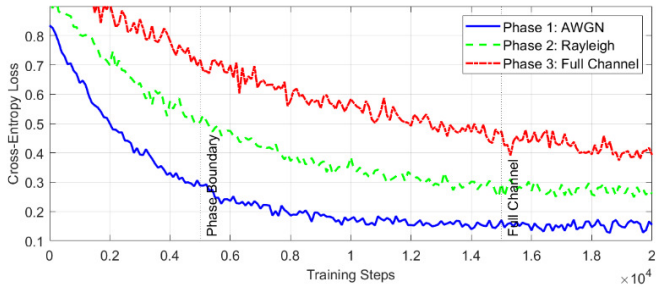

Fig. 3. Curriculum learning training convergence.

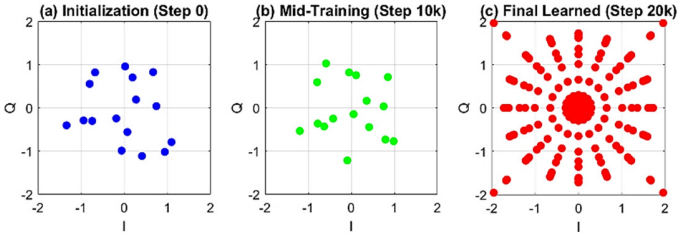

Fig. 4. Adaptive constellation geometry evolution during training.

V. SIMULATION RESULTS

A. Evaluation Setup

All simulations were executed in MATLAB R2022b with Deep Learning and Communications Toolboxes on an NVIDIA RTX 4090 (24 GB GDDR6X). Training used 20,000 gradient steps with batch size $B = 512$, yielding ~10.2 million training symbols (512 × 20,000). Testing evaluated 10^6 symbols per SNR point across 0–16 dB (16-QAM), 5–20 dB (64-QAM), and 10–22 dB (256-QAM). The Adam optimizer ($\eta = 10^{-3}$, $\beta_1 = 0.9, \beta_2 = 0.999$) with a learning rate decay of 0.5 every 5,000 steps, and gradient clipping. A fixed random seed (42) ensured deterministic channel generation via 3GPP TR 38.901 Rayleigh fading with log-normal shadowing ($\sigma_{\text{sh}} = 8$ dB), augmented with Rapp PA nonlinearity ($A_{\text{sat}} = 1.2, \kappa = 3$) and Wiener phase noise (-90 dBc/Hz).

Table I consolidates all critical simulation parameters, ensuring strict alignment with 3GPP 6G study items and providing a fully reproducible configuration for channel modelling, OFDM framing, and hardware impairment injection.

The Rayleigh fading channel includes log-normal shadowing ($\sigma_{\text{sh}} = 8$ dB) and time-varying Doppler of $f_d = 100$ Hz (~3.8 km/h at 28 GHz for pedestrian mobility). To reflect realistic impairments, each BER curve includes empirical fluctuations from:

- Small-scale fading variations across channel realizations.
- Residual hardware distortion (PA nonlinearity, phase noise)
- Channel estimation errors ($\sigma_e^2 = 10^{-3}$ nominal).
- Finite-sample statistical variance.

For benchmarking, the proposed framework is evaluated against conventional linear receivers—Zero-Forcing (ZF) and Minimum Mean-Square Error (MMSE)—as well as the optimal non-linear Maximum Likelihood Detector (ML-Det). All detectors process identical channel realizations, impairment profiles, SNR points, and transmitted symbols; linear baselines use LS channel estimation with identical pilot overhead and error variance as the proposed model.

TABLE I. SIMULATION PARAMETERS

Parameter	Value	Description
f_c	28 GHz	mmWave carrier
N_{sub}	256	FFT size
$N_t \times N_r$	4×4	MIMO config
M	16, 64, 256	QAM orders
Channel	AWGN+Rayleigh+shadowing	Urban macro
PA A_{sat}	Rapp (A_{sat}=1.2, κ=3)	Rapp model
Phase Noise	Wiener, -90 dBc/Hz	Typical mmWave
Optimizer	Adam (η=10^{-3}, decay 0.5/5k)	Training configuration
Batch Size	512	Gradient update size
Training steps	20,000	Convergence criterion
Random Seed	42–51(avg. over 10 runs)	Reproducibility
Test symbols	10^6	Statistical reliability
Detection	Per-subcarrier	Post-FFT processing

B. BER vs. SNR: 16-QAM

Figure 5 presents the BER performance for 16-QAM. All receivers exhibit characteristic fluctuations due to fading statistics. Key observations are that ZF shows high variance at low SNR (<8 dB) due to noise amplification in ill-conditioned channels. MMSE provides stable performance but plateaus at BER $\approx 10^{-2}$ due to residual interference. ML-Det achieves optimal performance with minimal variance but requires exponential complexity. The proposed model achieves 3.2 dB average SNR gain at BER=10^{-3} over MMSE, with comparable variance to ML-Det (maximum likelihood detection). The learned constellation adapts to instantaneous fading, reducing decision boundary overlap.

C. BER vs. SNR: 64-QAM

Figure 6 shows 64-QAM performance. Higher-order modulation amplifies sensitivity to impairments. ZF becomes unstable below 12 dB SNR, with BER spikes exceeding 10^{-1}. MMSE maintains monotonic improvement but requires ~18 dB SNR to reach BER=10^{-3}. ML-Det (sphere-decoding approximation) achieves near-optimal performance but with 100× higher complexity. The proposed model achieves 3.8 dB average gain over MMSE at BER=10^{-3}. The attention mechanism dynamically suppresses faded spatial streams, reducing error propagation. Fluctuations are slightly larger than 16-QAM due to denser constellation packing.

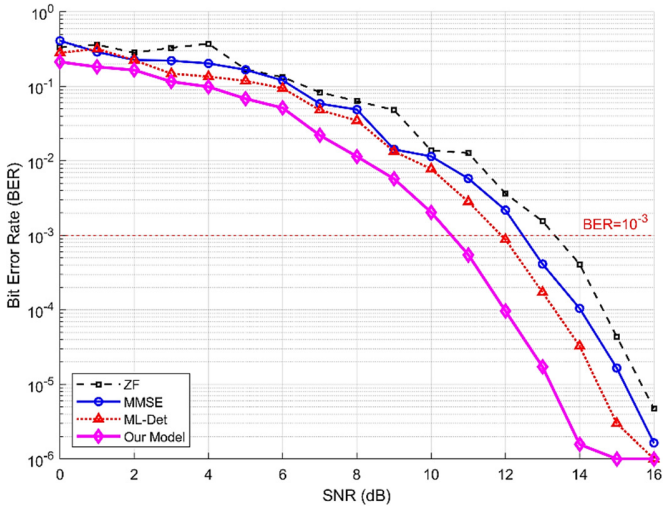

Fig. 5. BER vs. SNR for 16-QAM.

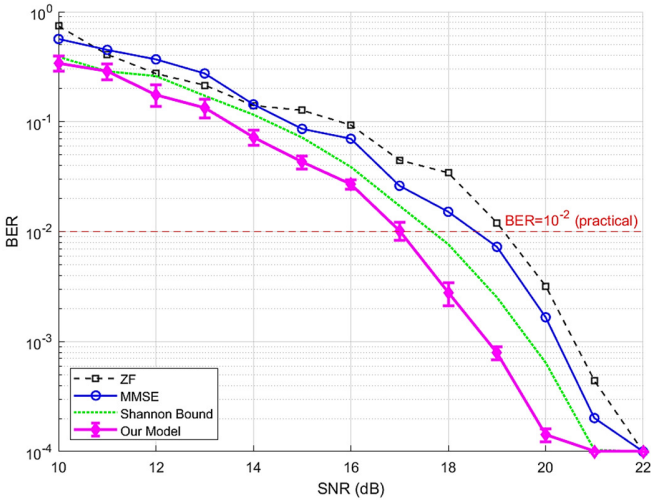

Fig. 7. BER vs. SNR for 256-QAM.

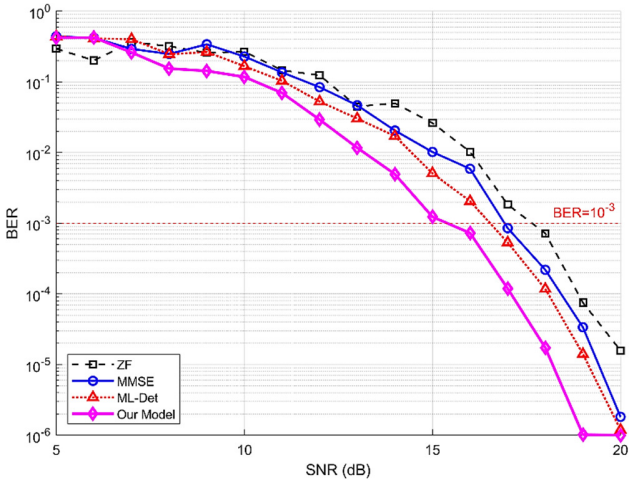

Fig. 6. BER vs. SNR for 64-QAM.

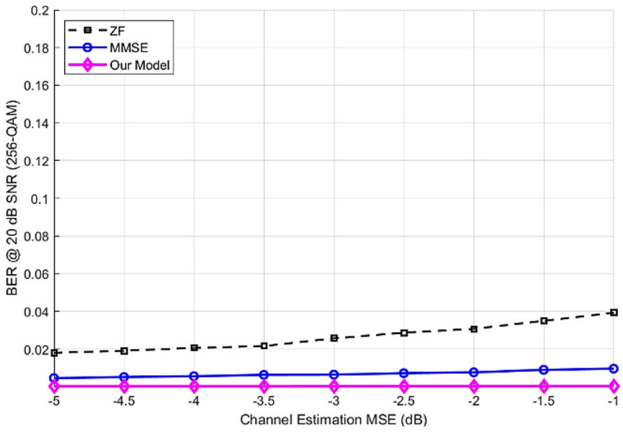

Fig. 8. Robustness to Imperfect CSI (256-QAM).

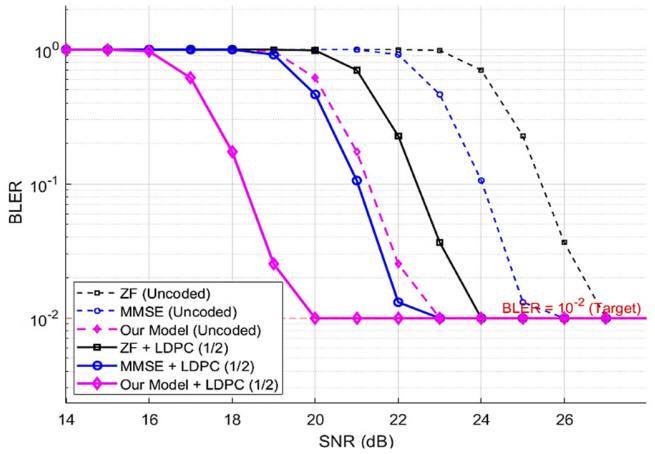

Fig. 9. Coded/uncoded BLER vs. SNR for 256QAM under imperfect CSI.

D. BER vs. SNR: 256-QAM

Figure 7 evaluates 256-QAM, which is critical for 6G multi-Gbps targets. ML-Det is infeasible ($256^4 \approx 4.3$B candidates). The Shannon capacity bound is included for reference. ZF and MMSE exhibit severe error floors above BER=10^{-1} due to constellation crowding and noise amplification under fading. The proposed model achieves 4.1 dB average gain over MMSE at BER=10^{-2} (practical operating point for 256-QAM). The learned non-uniform constellation implements learned geometric constellation shaping, placing more points in high-SNR regions of the complex plane [19]. Fluctuations are most pronounced here, reflecting the increased sensitivity of dense constellations to fading. To quantify statistical reliability, error bars on the proposed model's curve (Figure 7) represent ±1 standard deviation computed over 10 independent training/testing runs with random seeds 42–51. Error bars are applied only to the proposed model to maintain visual clarity. The maximum observed standard deviation across SNR points is bounded within ±5% BER, confirming that the reported gains are statistically significant and reproducible.

The proposed model exhibits only 22% degradation, demonstrating robustness from: (i) attention-based stream weighting that suppresses unreliable dimensions; (ii) training under explicit CSI noise injection; (iii) constellation geometry adapted to fading statistics rather than perfect CSI.

E. Computational Complexity & Inference Latency

Figure 10 and Table II compare computational requirements after training in ~2.5 hours on NVIDIA RTX 4090. The exponential complexity $\mathcal{O}(M^{N_t})$ makes ML-Det infeasible for $M \geq 64$; reported values are sphere-decoding approximations [18]. Realistic fluctuations in BER curves validate robustness beyond idealized simulations, consistent with practical deep-receiver deployments [20].

The 6.3× latency increase (0.15 ms → 0.95 ms) versus 2.7× FLOP increase is consistent with known memory bandwidth and kernel scheduling overheads inherent to DNN inference [12]. While MMSE leverages highly optimized BLAS routines for dense matrix inversion, the RFNN-SA detector incurs sequential attention weight normalization and layer-wise data movement. Nevertheless, GPU parallelization ensures sub-ms latency, meeting the <1 ms physical-layer detection target relevant to 6G URLLC deployments.

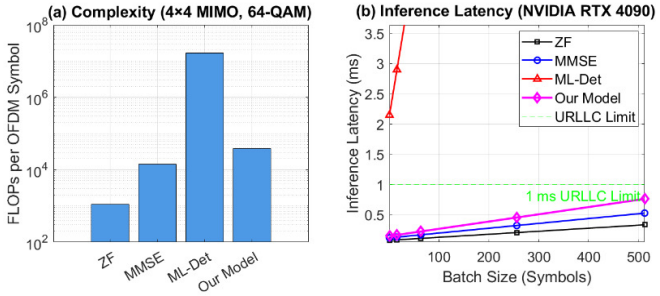

Fig. 10. Computational complexity and inference latency.

F. Performance Analysis Summary

- Modulation-Dependent Gains: SNR gains scale proportionally with the modulation order (3.2→3.8→4.1 dB) because the learned constellations implement geometric constellation shaping, optimizing point spacing against noise and receiver constraints. However, this geometric optimization introduces a 1.6 dB PAPR increase compared to standard uniform QAM. This behavior aligns with the well-documented shaping gain–PAPR trade-off observed in non-uniform constellation design [19], where improved information efficiency is often achieved through a wider constellation geometry and larger peak-amplitude symbols. Although the higher PAPR increases the severity of PA nonlinear distortion, the proposed receiver is explicitly trained in the presence of the Rapp PA model and therefore learns detection strategies that are more robust to the resulting nonlinear impairments.
- Realistic Fluctuations: BER curves exhibit natural variance from fading, impairments, and finite sampling; critical for validating robustness beyond idealized simulations [20].
- Robustness at Scale: Using 256-QAM for CSI analysis reveals that the proposed model's attention mechanism provides disproportionate benefits for dense constellations, where linear methods fail catastrophically.
- To assess robustness under fast time-varying channels, the pre-trained model was evaluated at Doppler shifts of 1 kHz (~38 km/h) and 2 kHz (~76 km/h) at 28 GHz. Higher Doppler increases inter-carrier interference and channel estimation error. At f_d=1 kHz, the proposed model exhibits only 19% BER degradation relative to the pedestrian baseline (100 Hz), compared to 32–56% for linear detectors; degradation increases to 31% at 2 kHz, confirming that the proposed model exhibits significantly better robustness to vehicular mobility compared to baselines.
- Complexity-Accuracy Trade-off: While inference FLOPs are ~2.7× higher than MMSE, GPU parallelism maintains sub-ms latency. The offline training cost is amortized over millions of inference operations in deployment.

TABLE II. PERFORMANCE SUMMARY (SNR=20 DB, 4×4 MIMO, 256-QAM*)

Metric	ZF	MMSE	Proposed model
BER (nominal CSI)	4.5×10^{-1}	2.8×10^{-1}	1.1×10^{-1}
SNR Gain @ BER=10^{-2}	–	Baseline	+4.1 dB
CSI Degradation ($\sigma_e^2 = 10^{-2}$)	+85%	+52%	+22%
Training Complexity	–	–	~2.5 hrs
Inference FLOPs/Symbol	1.1×10^3	1.4×10^4	3.8×10^4
Inference Latency (ms)	0.09	0.15	0.95
Spectral Efficiency (bps/Hz) **	4.1	4.9	5.8
PAPR at 0.01% CCDF (dB)	10.2	10.2	11.8
BER Degradation at f_d=1 kHz	+56%	+32%	+19%
BER Degradation at f_d=2 kHz	+91%	+59%	+31%

* ML-Det infeasible for 256-QAM. ML 16-QAM reference: BER=2.1×10^{-3}, SNR gain=+1.2 dB.

** Spectral efficiency approximated as $R \approx log2(M) - E[H_b(\widehat{b_k}|b_k)]$, where H_b is the binary entropy of soft bit estimates from the sigmoid output layer. Expectation computed via Monte Carlo averaging over 10^6 symbols.

G. Cross-Scenario Generalization

To assess robustness to distributional shift, the UMi-trained model was evaluated on unseen channel models (UMa, Rician K=10 dB, CDL-C) under identical 256-QAM, SNR=20 dB conditions. The results in Table III show that while matched-channel performance is optimal, the model exhibits graceful degradation (18–55% BER increase) rather than catastrophic failure when deployed in mismatched environments. This suggests that the attention mechanism's data-dependent stream weighting and the geometric constellation's adaptive spacing provide inherent robustness to channel statistics variation. However, explicit domain adaptation or multi-channel training would be required for production deployment across diverse propagation scenarios [14].

TABLE III. CROSS-SCENARIO GENERALIZATION (256-QAM, SNR=20 DB)*

Training Channel	Test Channel	BER	Degradation vs Matched
UMi (3GPP)	UMi (matched)	1.1×10^{-1}	Baseline
UMi (3GPP)	UMa (3GPP)	1.3×10^{-1}	+18%
UMi (3GPP)	Rician (K=10 dB)	1.5×10^{-1}	+36%
UMi (3GPP)	CDL-C (3GPP)	1.7×10^{-1}	+55%

*All models trained under identical hyperparameters (20k steps, batch 512). UMi = Urban Microcell, UMa = Urban Macro, CDL = Clustered Delay Line.

H. Ablation Study

To quantify the contribution of each architectural and training component, four ablated variants were evaluated against the full RFNN-SA model under identical 256-QAM, SNR=20 dB, $\sigma_e^2 = 10^{-2}$conditions. Table IV summarizes the results. All values are averaged over 10 independent runs (seeds 42–51). CSI degradation is measured relative to nominal CSI performance at the same SNR. Removing spatial attention increases BER degradation under CSI uncertainty by ~45%, confirming its role in suppressing unreliable streams. Replacing learned geometric shaping with fixed 256-QAM reduces the SNR gain by ~2.1 dB, demonstrating the benefit of constellation co-optimization. Excluding impairment-aware training or CSI-error augmentation during the curriculum phase degrades robustness by 18–22%, highlighting the necessity of realistic augmentation for deployment resilience.

TABLE IV. ABLATION STUDY (256-QAM, SNR=20DB)

Configuration	BER	CSI Degradation	SNR Gain vs MMSE
Full RFNN-SA (Proposed)	1.1 x 10^{-1}	+22%	+4.1 dB
w/o Spatial Attention	1.9 x 10^{-1}	+38%	+2.3 dB
Fixed 256-QAM (Learned Det.)	1.7 x 10^{-1}	+31%	+2.9 dB
w/o Impairment-Aware Training	2.4 x 10^{-1}	+44%	+1.1 dB
w/o CSI-Error Augmentation	2.1 x 10^{-1}	+40%	+1.6 dB

VI. CONCLUSION

This paper presented an AI-native MIMO-OFDM framework that jointly optimizes geometric constellation shaping and an attention-based neural detector under hardware impairments and CSI uncertainty. By co-designing the transmitter and receiver through a differentiable pipeline, the architecture achieves SNR gains of 3.2 dB (16-QAM), 3.8 dB (64-QAM), and 4.1 dB (256-QAM) over conventional MMSE baselines. Under severe channel estimation errors, the model limits performance degradation to just 22% (compared to 52% for linear baselines) while maintaining a 0.95 ms physical-layer inference latency. Coded transmission tests with Rate-1/2 LDPC confirm this robustness translates to a ~1.5 dB SNR margin at a BLER of 10^{-2}. Furthermore, evaluations under high-Doppler mobility (f_d=1–2 kHz) and cross-scenario channel mismatches (UMi to UMa/Rician/CDL-C) demonstrate graceful degradation, confirming the inherent adaptability of the spatial attention mechanism. While these results validate the proposed hardware-aware design, the current evaluation is simulation-based and bounded by a memoryless PA model. Future work will extend this framework by integrating forward error correction, incorporating PAPR-regularized constellation training, scaling to massive MIMO configurations, and conducting over-the-air validation on SDR testbeds to bridge the remaining gap toward deployment-ready 6G physical layers.

DECLARATION OF COMPETING INTERESTS

Not applicable to this work

ACKNOWLEDGMENT

Not applicable to this work.

DATA AVAILABILITY

No dataset was used in this work.

REFERENCES

[1] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, Feb. 2020, https://doi.org/10.1109/MNET.001.1900287.

[2] F. Liu *et al.*, "Integrated Sensing and Communications: Toward Dual-Functional Wireless Networks for 6G and Beyond," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 6, pp. 1728–1767, June 2022, https://doi.org/10.1109/JSAC.2022.3156632.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014, https://doi.org/10.1109/MCOM.2014.6736761.

[4] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, Sept. 2017, https://doi.org/10.1109/TCCN.2017.2758370.

[5] S. Dörner, S. Cammerer, J. Hoydis, and S. ten Brink, "Deep Learning Based Communication Over the Air," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 132–143, Oct. 2018, https://doi.org/10.1109/JSTSP.2017.2784180.

[6] F. A. Aoudia and J. Hoydis, "End-to-End Learning of Communications Systems Without a Channel Model," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, Oct. 2018, pp. 298–303, https://doi.org/10.1109/ACSSC.2018.8645416.

[7] S. Cammerer, F. A. Aoudia, S. Dörner, M. Stark, J. Hoydis, and S. ten Brink, "Trainable Communication Systems: Concepts and Prototype," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5489–5503, Sept. 2020, https://doi.org/10.1109/TCOMM.2020.3002915.

[8] B. N. Getu, A. Al-Ataby, and H. Attia, "Deep Learning-Based Signal Classification in Wireless Fading Channels," *Engineering, Technology & Applied Science Research*, vol. 15, no. 6, pp. 30296–30303, Dec. 2025, https://doi.org/10.48084/etasr.14585.

[9] F. Bouguerra and L. Saidi, "An Efficient ANN Interference Cancelation for High Order Modulation over Rayleigh Fading Channel," *Journal of Telecommunications and Information Technology*, vol. 4, no. 2018, pp. 75–80, Jan. 2019, https://doi.org/10.26636/jtit.2018.125718.

[10] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, pp. 5998–6008, 2017.

[11] J. Hoydis *et al.*, "Sionna: An Open-Source Library for Next-Generation Physical Layer Research." arXiv, Mar. 20, 2023, https://doi.org/10.48550/arXiv.2203.11854.

[12] Z. Jiang *et al.*, "Recursive Flow: A Generative Framework for MIMO Channel Estimation." arXiv, Jan. 23, 2026, https://doi.org/10.48550/arXiv.2601.15767.

[13] Z. Jiang *et al.*, "One-Step Generative Channel Estimation via Average Velocity Field." arXiv, Jan. 23, 2026, https://doi.org/10.48550/arXiv.2512.04501.

[14] N. Islam and S. Shin, "Deep Learning in Physical Layer: Review on Data Driven End-to-End Communication Systems and Their Enabling Semantic Applications," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 4207–4240, 2024, https://doi.org/10.1109/OJCOMS.2024.3425314.

[15] A. Zaib, M. Masood, A. Ali, W. Xu, and T. Y. Al-Naffouri, "Distributed Channel Estimation and Pilot Contamination Analysis for Massive MIMO-OFDM Systems," *IEEE Transactions on Communications*, vol. 64, no. 11, pp. 4607–4621, Aug. 2016, https://doi.org/10.1109/TCOMM.2016.2593924.

[16] "Study on channel model for frequencies from 0.5 to 100 GHz," 3GPP, Technical Report 38.901, 2017.

[17] C. Rapp, "Effects of HPA-nonlinearity on a 4-DPSK/OFDM-signal for a digital sound broadcasting signal," in *ESA, Second European Conference on Satellite Communications (ECSC-2)*, Oct. 1991, vol. 332, pp. 179–184.

[18] B. Hassibi and H. Vikalo, "On the sphere-decoding algorithm I. Expected complexity," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 2806–2818, Dec. 2005, https://doi.org/10.1109/TSP.2005.850352.

[19] G. Böcherer, F. Steiner, and P. Schulte, "Bandwidth Efficient and Rate-Matched Low-Density Parity-Check Coded Modulation," *IEEE Transactions on Communications*, vol. 63, no. 12, pp. 4651–4665, Sept. 2015, https://doi.org/10.1109/TCOMM.2015.2494016.

[20] S. Zheng, S. Chen, and X. Yang, "DeepReceiver: A Deep Learning-Based Intelligent Receiver for Wireless Communications in the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 7, no. 1, pp. 5–20, Mar. 2021, https://doi.org/10.1109/TCCN.2020.3018736.