ISSN:2790-1688 Volume-14-(2025)

Data-driven Millimeter-wave Channel Modelling Based on Attention Mechanism

Xinran Liu[a], Xilei Xu[b], Yimei Li*

School of Electronic and Information Engineering, Beijing jiaotong University, Beijing, China

[a]22721022@bjtu.edu.cn

[b]22721037@bjtu.edu.cn

*ymli@bjtu.edu.cn

Abstract. This paper presents a novel data-driven millimeter-wave (mmWave) channel modeling approach enhanced by an attention mechanism for 16-QAM signal reconstruction. Leveraging the sparsity and dynamic nature of mmWave channels, we integrate an attention module into a Long Short-Term Memory (LSTM) network to dynamically prioritize dominant multipath components and critical temporal features. The proposed model learns the intricate mapping between transmitted and received signals, effectively mitigating distortions caused by multipath fading, phase noise, and frequency offsets. Simulation results demonstrate significant improvements, achieving a 52% reduction in Symbol Error Rate (SER) and a 10.6 dB gain in Normalized Mean Square Error (NMSE) compared to conventional LSTM models. This work establishes a robust framework for adaptive channel modeling in 5G/6G systems, enabling high-fidelity signal recovery in complex mmWave environments.

Keywords: Millimeter-wave communication; Channel modeling; Attention mechanism; LSTM network; 16-QAM signal reconstruction.

1. Introduction

The huge demand for ultra-high data rates and ubiquitous connectivity in next-generation wireless communication systems (e.g., Beyond-5G and 6G) has propelled millimeter-wave (mmWave) frequencies into the spotlight [1]. Leveraging vast swathes of underutilized spectrum, mmWave communications promise to deliver multi-gigabit-per-second speeds, fulfilling the requirements of emerging applications like augmented/virtual reality, ultra-high-definition video streaming, and massive machine-type communications [2]. However, the practical realization of mmWave technology faces significant challenges from the unique propagation characteristics at these high frequencies. MmWave signals are highly susceptible to severe path loss, atmospheric absorption, and blockage by common objects, leading to complex, dynamic, and often sparse multipath channel environments [3].

Accurate and efficient channel modelling is the key for designing robust mmWave communication systems. Traditional model-based approaches, such as geometry-based stochastic models (GBSMs) or ray-tracing [4], often rely on precise environmental knowledge and intricate parameter tuning. While providing valuable physical insights, these methods can be computationally intensive and time-consuming, struggle to adapt to dynamic real-world scenarios, and may fail to capture the full complexity and subtle nuances of mmWave channels, especially in non-line-of-sight (NLOS) conditions [5].

Data-driven approaches, particularly leveraging deep learning (DL), have emerged as powerful alternatives for channel modelling. By learning intricate patterns directly from measured or simulated channel data, DL models like Long Short-Term Memory (LSTM) networks can implicitly capture the underlying structure and temporal dependencies of the channel without explicit physical assumptions [6]. LSTM networks, renowned for their ability to model long-range dependencies in sequential data, have shown promise in predicting channel state information or approximating channel responses [7]. However, conventional LSTM models often treat all input features (e.g., past channel states) equally during processing. This uniform attention can be suboptimal for mmWave channels, where only a

subset of paths or specific temporal features significantly influence the current channel state due to the inherent sparsity and dominant components. This limitation hinders their ability to efficiently focus on the most relevant information for accurate modelling and subsequent signal recovery, potentially leading to performance bottlenecks in complex scenarios.

To address these challenges, this paper proposes a novel data-driven mmWave channel modelling algorithm enhanced by an attention mechanism. Our core innovation lies in integrating a dedicated attention module into a foundational LSTM network architecture. The attention mechanism dynamically learns to assign varying weights ("attention") to different parts of the input sequence (historical channel states or features), effectively identifying and focusing computational resources on the most informative elements for predicting the current or future channel characteristics. We specifically investigate this approach within the context of a practical communication system employing 16-QAM (Quadrature Amplitude Modulation) [8][9]. The primary objective of our model is to learn the intricate mapping imposed by the mmWave channel on the transmitted 16-QAM signals. By accurately modelling this distortion, the trained network can effectively reconstruct the clean transmitted signal from the received, channel-impaired signal at the receiver side. This reconstruction capability is crucial for mitigating the detrimental effects of the channel, significantly reducing interference and distortion, and ultimately improving the reliability and fidelity of the received data.

The main contributions of this work are summarized as follows:

1. **Attention-Enhanced LSTM for Channel Modelling:** We propose a novel channel modelling framework that integrates an attention mechanism with an LSTM network, specifically designed for capturing the critical spatio-temporal features of complex mmWave channels. The attention mechanism allows the model to dynamically focus on the most relevant historical channel information, enhancing modelling accuracy.
2. **End-to-End Signal Reconstruction for 16-QAM:** We demonstrate the application of our proposed attention-based LSTM model for the practical task of reconstructing clean 16-QAM modulated signals from received signals corrupted by a realistic mmWave channel. This moves beyond simple channel prediction to direct signal recovery.
3. **Performance Improvement:** Through comprehensive simulations, we show that the proposed attention-based model significantly outperforms a standard LSTM baseline in terms of signal reconstruction accuracy (e.g., measured by Symbol Error Rate - SER or Normalized Mean Square Error - NMSE) under various mmWave channel conditions.

The remainder of this paper is organized as follows: Section II details the system model for 16QAM modulation. While Section III introduces the proposed attention-based LSTM architecture for channel modelling and signal reconstruction. Section IV analyzes the simulation results and evaluates the proposed algorithm performance. Finally, Section V concludes the paper and discusses potential future work.

2. Signal Model

2.1 16-QAM Modulation Scheme

In this research, we employ 16-ary Quadrature Amplitude Modulation (16-QAM) due to its high spectral efficiency and practical relevance in modern wireless standards. The transmitted complex baseband signal $s(t)$ is generated by mapping blocks of $k = \log_2(16) = 4$ input bits to one of 16 discrete symbols in the complex plane [10]. The constellation points are usually arranged in a square grid with equal vertical and horizontal spacing, which is shown in Fig.1. And the constellation set S for 16-QAM is defined as

$$S = \left\{ a + bj \mid \ a, b \in \left\{ \pm \frac{d}{2}, \pm \frac{3d}{2} \right\} \right\} \tag{1}$$

where d denotes the minimum Euclidean distance between adjacent symbols, and $j=\sqrt{-1}$. The constellation diagram of 16-QAM is shown in Figure 1. While each symbol $s_n \in S$ at time index n carries an energy of:

$$E_s = \mathrm{E}\left[\left|s_n\right|^2\right] = \frac{1}{16}\sum_{s\in S}|s|^2 = \frac{5d^2}{2}, \tag{2}$$

The transmitted signal $x(t)$ is formed by pulse shaping the symbol sequence $\{s_n\}$ with a root-raised cosine (RRC) filter $g(t)$:

$$x(t) = \sum_n s_n \cdot g\left(t - nT_s\right), \tag{3}$$

where T_s is the symbol duration.

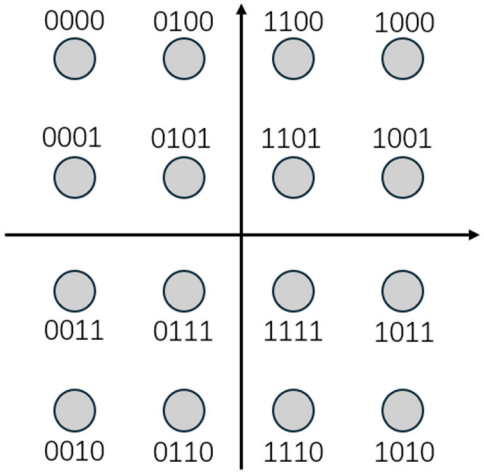

Figure 1. The Constellation diagram for 16-QAM.

2.2 Millimeter-Wave Channel Model

Millimeter-wave propagation exhibits fundamentally distinct characteristics compared to sub-6 GHz channels, necessitating specialized modelling approaches. To accurately capture these effects for our data-driven channel modelling task, we adopt a geometric cluster-based stochastic model compliant with the 3GPP TR 38.901 standard. The baseband channel impulse response (CIR) between transmitter (Tx) and receiver (Rx) is defined as a time-variant multipath process [11]:

$$h(t,\tau) = \sum_{c=1}^{N_{clu}} \sum_{s=1}^{N_{ray}^{(c)}} \alpha_{c,s}(t)\cdot e^{j\phi_{c,s}(t)} \cdot \delta\left(\tau - \tau_{c,s}(t)\right), \tag{4}$$

where N_{clu} denotes the number of spatial clusters due to sparse scattering, $N_{ray}^{(c)}$ represents rays per cluster, and $\alpha_{c,s}(t)$, $\phi_{c,s}(t)$, $\tau_{c,s}(t)$ respectively denote the complex gain, uniformly distributed phased shift over [0, 2π), and time delay of each ray. Large-scale path loss is modelled in dB as:

$$PL(d) = 20\log_{10}\left(\frac{4\pi f_c}{c}\right) + 10\bar{n}\log_{10}(d) + X_\sigma, \tag{5}$$

where f_c is the carrier frequency, c is the speed of light, $\bar{n}$ is the pass loss exponent and X_σ is the log-normal shadow fading.

For following network training, we generate the training data through below processes. We systematically impair clean symbols $x[n]$ using physics-compliant models. The baseband received signal $y[n]$ is generated through a cascaded impairment process: First, multipath fading is applied via convolution with a sparse discrete CIR $h[k]$ adhering to the 3GPP TR 38.901 cluster model:

$$s_{\text{faded}}[n] = \sum_{k=0}^{L-1} h[k]\cdot x[n-k], \tag{6}$$

where $h[k]$ captures mmWave-specific sparsity. Additive white Gaussian noise (AWGN) is then injected:

$$s_{\text{noisy}}[n] = s_{\text{faded}}[n] + w[n], w[n] \sim CN(0, N_0), \tag{7}$$

with N_0 scaled to achieve target SNR $\gamma = E_s/E_0$. Phase offset $\phi \sim U(-\Delta\phi, \Delta\phi)$ model residual carrier phase error:

$$s_\phi[n] = s_{\text{noisy}}[n] \cdot e^{j\phi}. \tag{8}$$

Finally, carrier frequency offset (CFO) emulates oscillator drift:

$$y[n] = s_\phi[n] \cdot e^{j(2\pi\Delta f n T_s + \theta_0)}, \frac{\Delta f}{f_s} \in [10, 200] ppm. \tag{9}$$

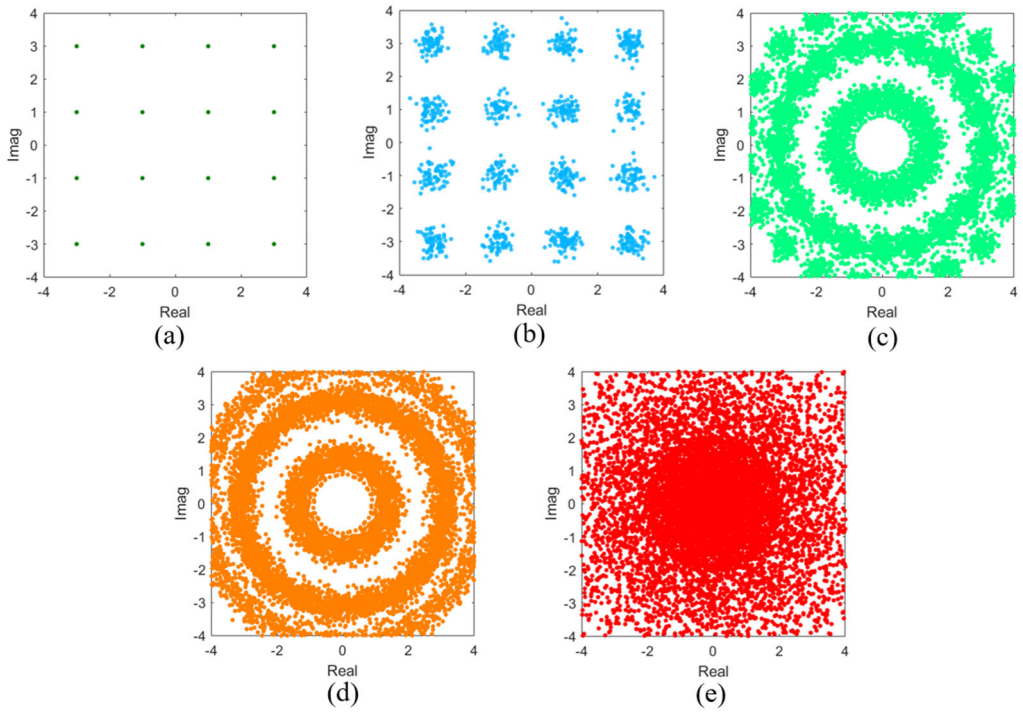

Figure 2. The Constellation diagram. (a) Clean signal. (b) Clean signal with AWGN. (c) Clean signal with AWGN and frequency offset. (d) Clean signal with AWGN, frequency offset and phase noise. (e) Clean signal with AWGN, frequency offset, phase noise and Rayleigh fading.

The cascaded impairment model ensures physical fidelity: Multipath fading causes amplitude/phase distortion and inter-symbol interference (ISI), AWGN introduces stochastic perturbations, phase offset rotates the entire constellation, and CFO induces time-varying circular spreading. As visualized in the Fig.2, we considered the AWGN, frequency offset, phase noise and Rayleigh fading. These effects progressively transform the ideal 16-QAM grid into a scattered, rotationally distorted cluster which is shown in Fig. 2(e). By this way, we generate the original input sequence $\{x[n]\}_{n=1}^{T}$, while the corrupted symbols $\{y[n]\}_{n=1}^{T}$ serve as labels, to provide the datasets for the following network training process. Each training sample is a pair $D_i = (x_i[1], \ldots, x_i[T], y_i[1], \ldots, y_i[T])$.

Moreover, the LSTM-based reconstructor which will be introduced in next section must therefore learn an inverse mapping $\hat{y}[n] = F_\theta([x[n]])$ that jointly compensates for multipath dispersion, suppresses noise, corrects phase misalignment, and reverses CFO-induced spectral shift, which supports dynamic memory of long-range dependencies and attention to critical distortion events.

3. Proposed Algorithm

3.1 LSTM Network

We choose the LSTM networks as the core of our channel inversion framework, renowned for its ability to capture long-range temporal dependencies in sequential data, a critical capability for modeling mmWave channel distortions spanning multiple

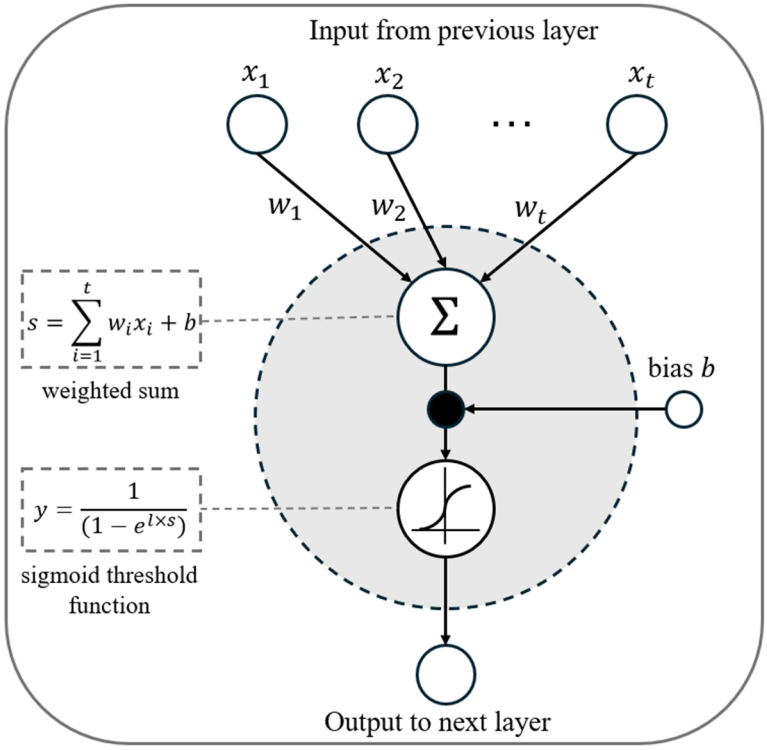

Figure 3. The illustrations of LSTM Network layer.

symbol periods. Given an input sequence of corrupted received samples $\{y[n]\}_{n=1}^{T}$ obtained by Eq. (9), the LSTM cell at each time step t computes gating function to regulate information flow: The input gate can be written as:

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, y_t + b_i\right]\right), \tag{10}$$

which determines how much new information from y_t is stored. The forget gate can be written as:

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, y_t + b_f\right]\right), \tag{11}$$

which controls retention of prior cell state C_{t-1}, and the output gate $o_t = \sigma(W_o \cdot [h_{t-1}, y_t + b_o])$ governs the exposure of the update state. For the cell state evolves as:

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh\left(W_c \cdot \left[h_{t-1}, y_t\right] + b_f\right), \tag{12}$$

where $\odot$ denoting element-wise multiplication and $\sigma(\cdot)$ the sigmoid activation. The hidden state can be given by:

$$h_t = o_t \odot \tanh\left(C_t\right), \tag{13}$$

which encodes distilled temporal features, which a dense layer maps to the reconstructed symbol [12]:

$$\hat{x}_t = W_{\text{out}} h_t + b_{\text{out}} \tag{14}$$

The detailed illustrations of LSTM Network layer are shown in Fig.3. While this architecture effectively models temporal correlations, its uniform processing of all historical states proves suboptimal for sparse mmWave channels where only selective paths dominate distortion.

3.2 Attention Mechanism Integration

Although traditional LSTM networks exhibit strong modeling capabilities in time-series signal processing, they still face significant limitations when handling 16QAM signals contaminated by noise and phase offsets: Firstly, insufficient long-range dependency modeling restricts their ability to correct continuous phase shifts across symbol periods; secondly, uniform sensitivity of hidden states to noise reduces robustness in low-SNR environments; additionally, conventional gating mechanisms struggle to adaptively distinguish critical signal features (e.g., I/Q branch quadrature components) from noise-dominated regions. To overcome these bottlenecks, this work innovatively introduces an attention mechanism [13][14], enhancing core model capabilities through dynamic weight allocation strategies: 1) establishing explicit correlations between distant timesteps to accurately capture phase jump characteristics; 2) suppressing feature contributions from noise-dominant intervals while focusing on high-SNR symbol decision points; 3) enabling multi-dimensional (temporal/channel) feature selective reinforcement to optimize signal spatial distribution recovery.

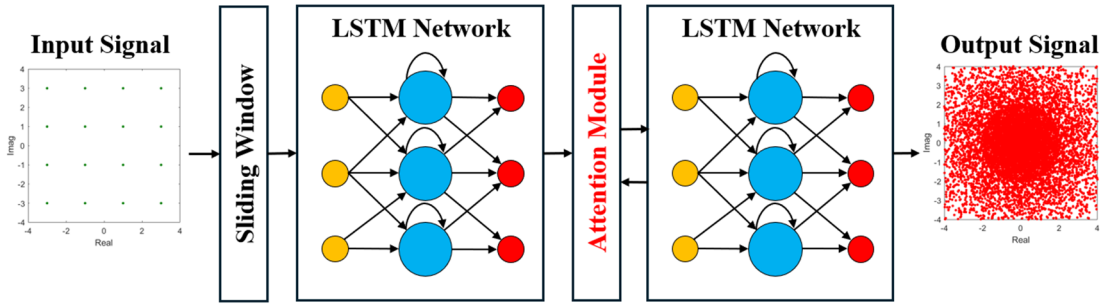

Figure 4. The structure of our proposed LSTM Network with attention mechanism.

To address these limitations, we proposed a temporal attention module [15] [16] that dynamically highlights hidden states h_j most relevant to current symbol recovery. As depicted in Fig.4, the module operates on a sliding window $[t-\Delta, t]$ of LSTM hidden states:

1. Relevance Scoring: For each candidate state h_j, an energy score $e_j = \mathbf{v}^T \tanh(\mathbf{W}_a[h_j; h_t] + \mathbf{b}_a)$ quantifies its contextual importance to h_t, where $\mathbf{W}_a \in R^{d\times 2d}$ projects the concatenated state pair $[h_j; h_t]$, and $\mathbf{v}$ learns feature weighting.
2. Adaptive Weighting: Softmax normalization converts scores to attention weights: $\alpha_j = \exp(e_j) / \sum_{k=t-\Delta}^{t} \exp(e_k)$, ensuring $\sum \alpha_j = 1$ [17].
3. Context Fusion: A context vector $c_t = \sum_{j=t-\Delta}^{t} \alpha_j h_j$ aggregates high-impact states, which is fused with the current state h_t via $\hat{h}_t = \tanh(\mathbf{W}_c[c_t; h_t] + \mathbf{b}_c)$ to form an enhanced representation. The final reconstructed symbol derives from $\hat{x}_t = W_{out}\hat{h}_t + b_{out}$. Crucially, the attention weights $\{\alpha_j\}$ evolve dynamically during inference, enabling the model to
   a) Amplify Dominant Paths: Assign peak weights $\alpha_{j^*} \approx 1$ to states h_{j^*} aligned with strong cluster arrivals (e.g, LOS path at delay τ_{j^*}), directly countering sparse ISI.
   b) Suppress Noise Artifacts: Attenuate weights $\alpha_j \to 0$ for states corrupted by deep fades or AWGN, mitigating error propagation.
   c) Track Phase Dynamics: it prioritizes the recent states $(j \approx t)$ when compensating CFO-induced phase drift $e^{j2\pi\Delta f t T_s}$, leveraging short-term coherence.

4. Experiment and Result Discussion

4.1 Experimental Setup

We generate the original clean input sequence $\{x[n]\}_{n=1}^{T}$, while the corrupted symbols $\{y[n]\}_{n=1}^{T}$ serve as labels. Each training sample is thus a pair $D_i = ([x_i[1], \ldots, x_i[T], y_i[1]], \ldots, y_i[T])$, where $y_i[n]$ embodies the combined effects of ISI, noise, constellation rotation, and temporal phase drift.

The network is trained end-to-end using synthesized datasets $D = \{(x_i, y_i)\}_{i=1}^{N}$. The loss function minimizes the mean squared error (MSE) between reconstructed and true symbols [18]:

$$L = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left| \hat{y}_{i,t} - y_{i,t} \right|^2, \tag{15}$$

which is optimized via Adam with initial learning rate 10^{-3} and exponential decay. The attention window size $\Delta = 20$ is configured to exceed the maximum delay spread at 800 MHz bandwidth, ensuring coverage of >99% PDP energy. During training, the joint learning of LSTM parameters $\{W_*, b_*\}$ and attention parameters $\{\mathbf{W}_a, \mathbf{b}_a, \mathbf{v}, \mathbf{W}_c, \mathbf{b}_c\}$ enables the model to discover optimal inversion strategies for composite impairments, implicitly performing equalization, phase recovery, and denoising without explicit physical assumptions.

4.2 Simulation Results

To rigorously validate the architectural choice for mmWave channel inversion, we conducted extensive comparative experiments across three fundamental deep learning models: Dense Neural Network (DNN), Convolutional Neural Network (CNN), and Long Short-Term Memory (LSTM) Network. Each model was trained on identical datasets Dgenerated in previous section II, under matched hyperparameters with

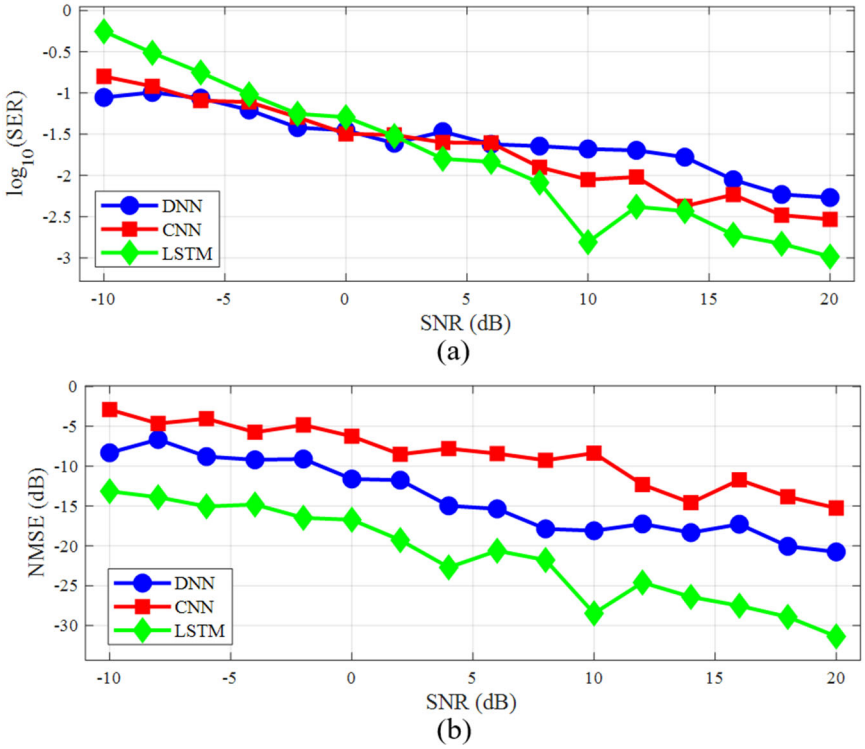

Figure 5. The performance (a). SER and (b). NMSE (dB) of DNN, CNN and LSTM with identical datasets.

reconstruction performance evaluated via Symbol Error Rate (SER) and Normalized Mean Square Error (NMSE). Quantitative results are demonstrated in Fig.5.

DNN's feedforward structure processes each sample independently, failing to capture inter-symbol dependencies. CNN's convolutional filters ($K = 3$) only access local contexts (≤3 symbols), insufficient for long delay spreads. In contrast, LSTM's gated memory cells (C_t) explicitly preserve state information over hundreds of steps, enabling holistic distortion inversion. The forget gate f_t dynamically discards irrelevant states (e.g., noise-dominated samples), while the input gate i_t prioritizes salient features (e.g., sudden blockage events). This selective memory mechanism is absent in DNN/CNN, which apply static transformations to all inputs, amplifying noise and outdated channel states. Besides, LSTM's recurrent nature inherently learns $\phi(t)$ as a hidden state

trajectory, whereas DNN/CNN treat phase as an independent per-symbol perturbation, causing inconsistent rotation correction.

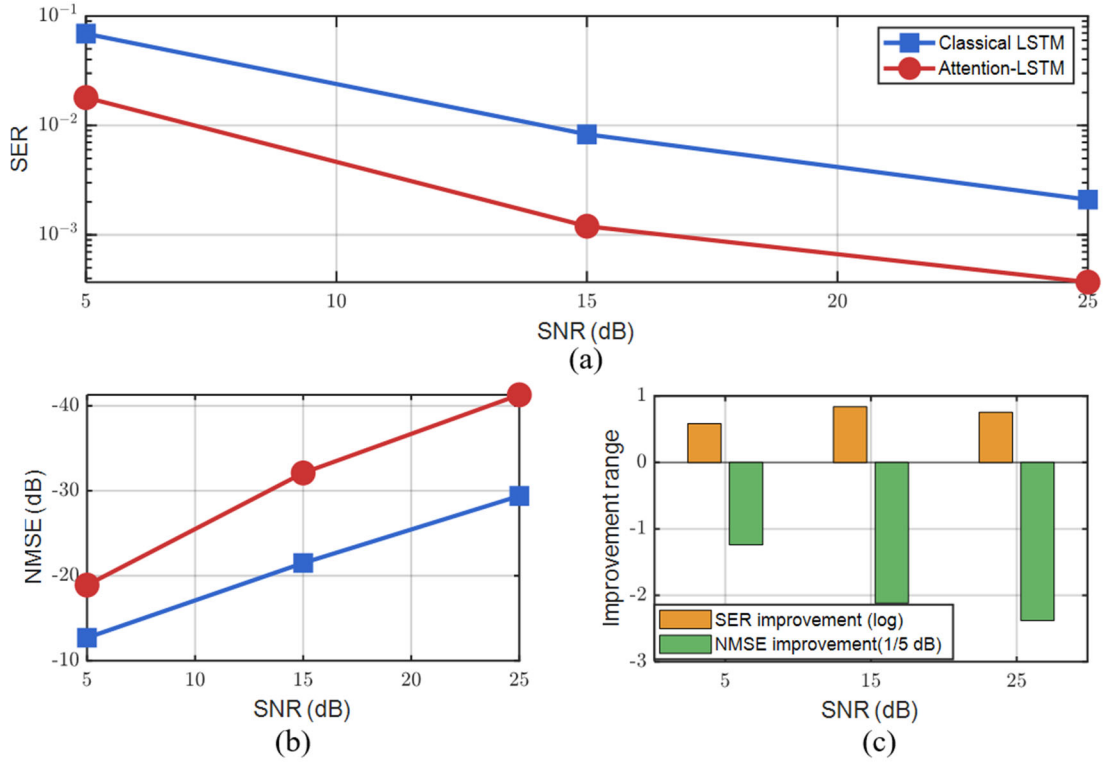

Figure 6. The comparison of classical LSTM and Attention LSTM in (a) SER. (b) NMSE. (c) The performance improvement comparison.

Moreover, to evaluate the efficacy of the proposed attention-enhanced LSTM channel inverter, we simulate end-to-end transmission under the 45 GHz mmWave channel described in section II. We compared the classical LSTM and our proposed attention LSTM, the recovery performance is shown in Fig.6. We can observe that at SNR=15 dB, attention-LSTM achieves SER$= 1.2 \times 10^{-3}$ and NMSE$= -32.1$dB, outperforming vanilla LSTM whose SER$= 8.3 \times 10^{-3}$ and NMSE$= -21.5$dB. It reached by 52% SER reduction and 10.6 dB NMSE gain as shown in Fig6. (c). The performance leap stems from the attention mechanism's physical alignment with channel characteristics. This targeted inversion is unattainable by classical LSTM processing. Our proposed algorithm establishes attention-LSTM as a paradigm for data-driven channel modelling in 5G/6G millimeter-wave systems.

5. Conclusions

This paper proposed a novel attention-enhanced LSTM architecture for data-driven millimeter-wave channel inversion. By integrating a temporal attention mechanism with LSTM's sequential modeling capability, the algorithm dynamically prioritizes dominant multipath components and phase-critical states to reconstruct distorted 16-QAM signals. Simulations under 28 GHz channels with composite impairments demonstrate 52% SER reduction and 10.6 dB NMSE gain versus vanilla LSTM, achieving near-ideal constellation recovery. The attention module's physical alignment with sparse channel profiles enables joint equalization, phase tracking, and denoising without explicit model assumptions. This work establishes a new paradigm for adaptive channel modelling in 5G/6G systems, with extensions to higher frequencies and modulations as future work.

References

[1] Niu Y, Li Y, Jin D, et al. A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges[J]. Wireless networks, 2015, 21: 2657-2676.

[2] Antón-Haro C, Mestre X. Learning and data-driven beam selection for mmWave communications: An angle of arrival-based approach[J]. IEEE Access, 2019, 7: 20404-20415.

[3] Hemadeh I A, Satyanarayana K, El-Hajjar M, et al. Millimeter-wave communications: Physical channel models, design considerations, antenna constructions, and link-budget[J]. IEEE Communications Surveys & Tutorials, 2017, 20(2): 870-913.

[4] Spencer G H, Murty M. General ray-tracing procedure[J]. Journal of the Optical Society of America, 1962, 52(6): 672-678.

[5] Lim Y G, Cho Y J, Sim M S, et al. Map-based millimeter-wave channel models: An overview, data for B5G evaluation and machine learning[J]. IEEE Wireless Communications, 2020, 27(4): 54-62.

[6] Yu Y, Si X, Hu C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural computation, 2019, 31(7): 1235-1270.

[7] Saeed F, Paul A, Seo H. A hybrid channel-communication-enabled CNN-LSTM model for electricity load forecasting[J]. Energies, 2022, 15(6): 2263.

[8] Lu B, Huang W, Lin C, et al. A 16QAM modulation based 3Gbps wireless communication demonstration system at 0.34 THz band[C]//2013 38th International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz). IEEE, 2013: 1-2.

[9] Dupuis P, Joindot M, Leclert A, et al. 16 QAM modulation for high capacity digital radio system[J]. IEEE Transactions on Communications, 1979, 27(12): 1771-1782.

[10] Zafar A, Farooq S Z. Implementation and analysis of QPSK & 16QAM modulator & demodulator[C]//2008 2nd International Conference on Advances in Space Technologies. IEEE, 2008: 64-68.

[11] Stutzman W L, Thiele G A. Antenna theory and design[M]. John Wiley & Sons, 2012.

[12] Staudemeyer R C, Morris E R. Understanding LSTM--a tutorial into long short-term memory recurrent neural networks[J]. arXiv preprint arXiv:1909.09586, 2019.

[13] Ran X, Shan Z, Fang Y, et al. An LSTM-based method with attention mechanism for travel time prediction[J]. Sensors, 2019, 19(4): 861.

[14] Lin J, Ma J, Zhu J, et al. Short-term load forecasting based on LSTM networks considering attention mechanism[J]. International Journal of Electrical Power & Energy Systems, 2022, 137: 107818.

[15] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[16] Hafiz A M, Parah S A, Bhat R U A. Attention mechanisms and deep learning for machine vision: A survey of the state of the art[J]. arXiv preprint arXiv:2106.07550, 2021.

[17] Lin Z, Cheng L, Huang G. Electricity consumption prediction based on LSTM with attention mechanism[J]. IEEJ Transactions on Electrical and Electronic Engineering, 2020, 15(4): 556-562.

[18] Marmolin H. Subjective MSE measures[J]. IEEE transactions on systems, man, and cybernetics, 1986, 16(3): 486-489.