Attention-Based Neural Beamforming Framework for Intelligent mmWave 6G Wireless Communication Systems

Khalid[1], Prashant Baghmar[2]

[1,2]Lecturer, Electronics, Government Polytechnic College, Mandore, Rajasthan, India

Abstract

Millimeter-wave (mmWave) communication and massive Multiple-Input Multiple-Output (MIMO) systems are considered key enabling technologies for next-generation 6G wireless networks due to their capability to provide high data rates and enhanced spectral efficiency. However, conventional beamforming approaches suffer from high computational complexity and beam alignment latency under dynamic wireless environments. To address these challenges, this paper proposes an attention-based neural beamforming framework for intelligent mmWave 6G communication systems. The proposed framework integrates geometric mmWave channel modeling, channel state information (CSI) extraction, and a self-attention based deep learning architecture for adaptive beam prediction and beamforming optimization. The self-attention mechanism efficiently captures dominant spatial channel characteristics, enabling accurate beam selection and interference suppression. Simulation results demonstrate that the proposed framework achieves improved beam prediction accuracy of approximately 98.6%, enhanced signal-to-interference-plus-noise ratio (SINR), higher spectral efficiency, reduced bit error rate (BER), and increased throughput compared to conventional and baseline deep learning beamforming approaches. Furthermore, the proposed framework significantly reduces computational complexity, making it suitable for real-time intelligent 6G wireless communication systems.

Keywords: Millimeter-Wave Communication, Massive MIMO, Beamforming, Deep Learning, Transformer Networks, Hybrid Beamforming, 6G Wireless Communication, Beam Prediction, Intelligent Wireless Networks

1. Introduction

The rapid evolution of wireless communication technologies and the increasing demand for ultra-reliable, high-capacity, and low-latency communication services have accelerated the development of fifth-generation (5G) and sixth-generation (6G) wireless communication systems. Emerging applications such as intelligent transportation, autonomous vehicles, augmented reality, Internet of Things (IoT), smart healthcare, and immersive multimedia communication require significantly enhanced spectral efficiency, massive connectivity, and intelligent network adaptability. Consequently, millimeter-wave (mmWave) communication and massive Multiple-Input Multiple-Output (MIMO) systems have emerged as key enabling technologies for next-generation wireless networks due to their capability to provide extremely high data rates, wider bandwidth availability, and improved spatial multiplexing performance [1–4].

Despite these advantages, mmWave communication systems suffer from severe propagation attenuation, blockage sensitivity, limited scattering characteristics, and high path loss, which substantially degrade communication reliability and beam alignment performance [4]. To overcome these challenges, advanced beamforming and hybrid precoding techniques have been extensively investigated to focus transmission energy toward intended users and improve communication quality. Conventional beamforming approaches, including analog beamforming, digital beamforming, and hybrid beamforming frameworks, generally rely on iterative optimization procedures and exhaustive beam search mechanisms [5, 6]. Although these techniques can achieve acceptable communication performance, they often exhibit high computational complexity, increased latency, limited scalability, and reduced adaptability under highly dynamic wireless channel environments.

The integration of artificial intelligence (AI) and machine learning techniques into wireless communication systems has recently gained considerable research attention for enabling intelligent and adaptive communication optimization [7, 8]. Deep learning frameworks have demonstrated exceptional capability in learning complex channel characteristics and performing communication tasks such as channel estimation, modulation classification, beam prediction, resource allocation, and interference management [9–12]. In particular, deep learning-assisted beamforming approaches have shown substantial improvements over conventional optimization-driven beamforming methods by enabling data-driven beam selection and low-latency beam prediction [15].

Several studies have investigated neural network-assisted beamforming optimization for mmWave massive MIMO communication systems. Alkhateeb et al. (2018) proposed a coordinated deep learning framework for highly mobile mmWave beamforming systems [11]. He et al. (2018) introduced deep learning-based channel estimation for beamspace mmWave massive MIMO communication [12]. Huang et al. (2019) developed deep learning-driven hybrid precoding frameworks for massive MIMO communication systems [9], while Jiang and Schotten (2019) investigated machine learning-assisted beamforming optimization [10]. Long et al. (2021) further proposed data-driven analog beam selection techniques for hybrid beamforming optimization [15]. However, most existing approaches are primarily based on convolutional neural networks (CNNs), multilayer perceptrons (MLPs), or recurrent neural architectures that exhibit limited capability in capturing long-range spatial dependencies and contextual wireless channel relationships.

Attention mechanisms and Transformer-based neural architectures have recently revolutionized deep learning applications due to their superior capability in capturing global contextual dependencies and adaptive feature representations. Vaswani et al. (2017) introduced the Transformer architecture using self-attention mechanisms, which significantly improved sequential learning and contextual feature extraction [13]. Subsequently, Dosovitskiy et al. (2021) demonstrated the effectiveness of Vision Transformers (ViTs) for large-scale image recognition tasks [14]. Motivated by these advantages, Transformer-driven frameworks have gradually emerged as promising solutions for intelligent wireless communication optimization and adaptive beamforming applications.

Recent advancements in AI-driven wireless communication systems have focused on Transformer-based beam prediction, reinforcement learning-assisted beamforming, reconfigurable intelligent surfaces (RIS), integrated sensing and communication (ISAC), digital twin-assisted communication systems, and quantum-enhanced wireless optimization frameworks. Tariq et al. (2024) proposed deep quantum-transformer networks for multimodal beam prediction in ISAC systems [16]. Allu et al. (2024) investigated robust energy-efficient beamforming for ISAC full-duplex communication systems [17], while Saikia et al. (2024) introduced hybrid deep reinforcement learning frameworks for RIS-assisted cooperative ISAC systems [18]. Katwe et al. (2024) proposed spectral-efficient beamforming for STAR-RIS-aided URLLC NOMA communication systems [19]. Lu et al. (2024) further investigated RIS-assisted secure key generation for mmWave communication networks [20].

Similarly, Kurma et al. (2024) developed spectral-energy efficient resource allocation strategies for RIS-assisted FD-MIMO systems [21], while Sharma et al. (2024) proposed robust transmission optimization for RIS-assisted SWIPT IoT communication systems [22]. Yigit et al. (2024) introduced AI-enhanced digital twin frameworks for cyber-resilient 6G Internet-of-Vehicles networks [23]. Huynh et al. (2024) further investigated joint sensing, communication, and computing optimization for ultra-reliable low-latency communication (URLLC) service-oriented mobile edge computing (MEC) networks [24]. Paul et al. (2024) proposed hybrid multi-agent deep reinforcement learning frameworks for RIS and UAV-assisted xURLLC communication systems [25].

More recently, beam prediction using large language model-inspired frameworks has also emerged as a promising research direction. Sheng et al. (2025) introduced large language model-based beam prediction mechanisms for wireless communication optimization [26]. Wang et al. (2024) investigated digital twin-enabled UAV communication systems under uncertain wireless environments [27]. Paul et al. (2025) further proposed quantum-enhanced deep reinforcement learning optimization for direction-of-arrival estimation and intelligent task offloading in ISAC systems [28]. Bui et al. (2024) introduced digital twin-enabled integrated satellite-terrestrial communication frameworks for future 6G IoT applications [29], while Li et al. (2024) proposed multi-agent UAV-assisted URLLC mobile edge computing optimization techniques [30].

Although these recent studies have demonstrated significant progress in AI-driven beamforming optimization and intelligent wireless communication systems, several important challenges still remain unresolved. Existing frameworks often suffer from:

- High computational complexity due to exhaustive beam search procedures,
- Increased beam alignment latency under dynamic wireless environments,
- Limited capability in efficiently capturing long-range spatial channel dependencies,
- Reduced scalability for large antenna arrays and ultra-dense communication systems,
- Insufficient robustness under highly dynamic mmWave communication scenarios.

To address these challenges, this paper proposes an attention-based neural beamforming framework for intelligent mmWave 6G wireless communication systems. The proposed framework integrates geometric mmWave channel modeling, massive MIMO communication, channel state information (CSI) extraction, and a self-attention-driven deep learning architecture for adaptive beam prediction and beamforming optimization. The self-attention mechanism efficiently captures dominant spatial channel characteristics and contextual feature relationships, thereby enabling accurate beam selection, improved interference suppression, enhanced communication reliability, and reduced computational complexity.

The major contributions of this work are summarized as follows:

- An attention-based neural beamforming framework is proposed for adaptive mmWave 6G wireless communication systems.
- A self-attention mechanism is integrated with CSI feature learning to improve beam prediction capability and spatial feature extraction efficiency.
- A comprehensive simulation framework is developed for mmWave massive MIMO communication under realistic wireless channel conditions.

- The proposed framework is evaluated using multiple communication performance metrics including beam prediction accuracy, SINR, spectral efficiency, BER, throughput, and computational complexity.
- Comparative analysis against conventional beamforming, random beam selection, MLP beamforming, and CNN beamforming demonstrates the superiority of the proposed framework.

Simulation results demonstrate that the proposed attention-based neural beamforming framework achieves significantly improved beam prediction accuracy, enhanced signal-to-interference-plus-noise ratio (SINR), higher spectral efficiency, reduced bit error rate (BER), and increased throughput compared to conventional and baseline deep learning beamforming approaches. Furthermore, the proposed framework substantially reduces computational overhead, making it highly suitable for real-time intelligent 6G wireless communication systems.

The remainder of this paper is organized as follows. Section II presents the proposed methodology and system model. Section III describes the simulation setup and implementation details. Section IV discusses the obtained results and performance analysis. Finally, Section V concludes the paper and outlines future research directions.

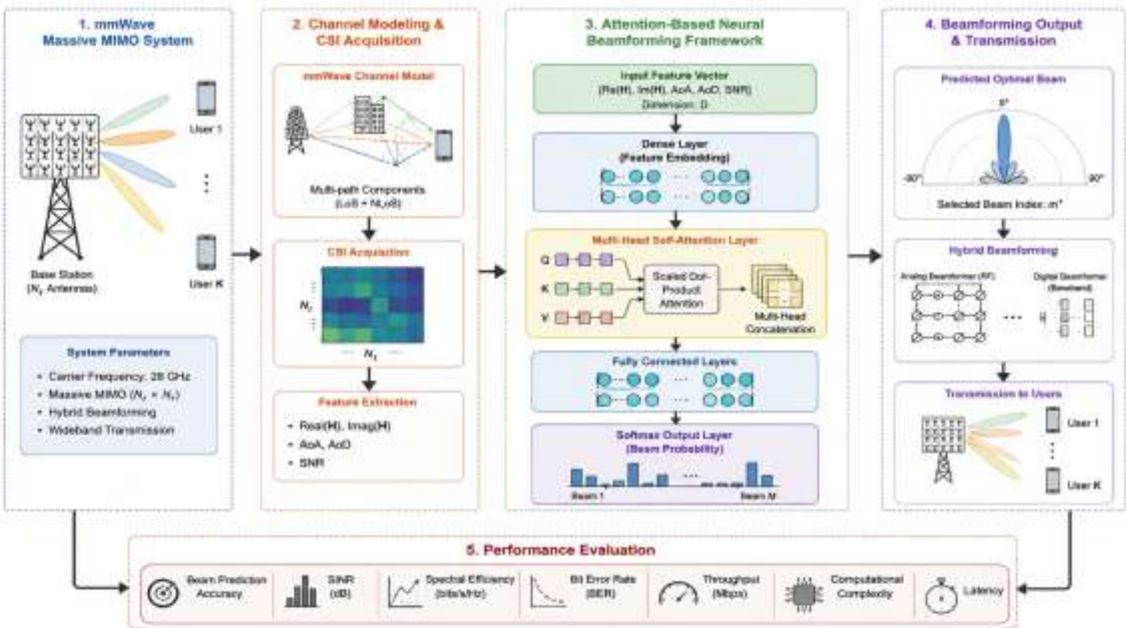
2. Proposed Methodology

2.1. Overview of the Proposed Framework

This work proposes an attention-based neural beamforming framework for millimeter-wave (mmWave) 6G wireless communication systems. The proposed framework aims to intelligently optimize beamforming vectors using deep learning techniques in order to improve spectral efficiency, signal quality, and communication reliability while reducing the computational complexity associated with conventional optimization-based beamforming approaches.

The proposed framework integrates mmWave channel modeling, massive Multiple-Input Multiple-Output (MIMO) antenna arrays, channel state information (CSI) extraction, self-attention based neural learning, and adaptive beam prediction. Unlike conventional iterative beam optimization techniques, the proposed framework directly predicts optimal beamforming vectors using a data-driven neural learning mechanism, thereby reducing beam alignment latency and computational overhead.

The overall architecture of the proposed system is illustrated in Fig. 1 conceptually as follows:



2.2. mmWave Massive MIMO System Model

Consider a downlink mmWave massive MIMO communication system consisting of a base station (BS) equipped with N_t transmit antennas and a user equipment (UE) equipped with N_r receive antennas.
The received signal at the user can be expressed as

$$y = HFs + n$$

where:

- $y \in \mathbb{C}^{N_r \times 1}$ denotes the received signal vector,
- $H \in \mathbb{C}^{N_r \times N_t}$ represents the mmWave channel matrix,
- $F \in \mathbb{C}^{N_t \times N_s}$ denotes the beamforming matrix,
- $s \in \mathbb{C}^{N_s \times 1}$ represents the transmitted signal vector, and
- n denotes additive white Gaussian noise (AWGN).

The beamforming matrix F is optimized using the proposed attention-based neural framework.

2.3. mmWave Channel Modeling

The mmWave propagation environment is characterized by sparse multipath propagation with limited dominant scattering paths. Therefore, a geometric channel model is considered.
The channel matrix is represented as

$$H = \sqrt{\frac{N_t N_r}{L}} \sum_{l=1}^{L} \alpha_l \, a_r(\theta_l^r) a_t^H(\theta_l^t)$$

where:

- L denotes the number of propagation paths,
- α_l represents the complex gain of the l^{th} path,
- $a_r(\theta_l^r)$ denotes the receive array response vector,
- $a_t(\theta_l^t)$ denotes the transmit array response vector,
- θ_l^r and θ_l^t represent the angles of arrival and departure, respectively.

2.4. Antenna Array Response Vector

A Uniform Linear Array (ULA) antenna configuration is considered at the transmitter side. The steering vector is expressed as

$$a(\theta) = \frac{1}{\sqrt{N}} \left[1, e^{-j\frac{2\pi d}{\lambda}\sin\theta}, \dots, e^{-j(N-1)\frac{2\pi d}{\lambda}\sin\theta}\right]^T$$

where:

- N denotes the number of antenna elements,
- d represents the antenna spacing,
- λ denotes the carrier wavelength,
- θ represents the steering angle.

2.5. Attention-Based Neural Beamforming Framework

The proposed neural beamforming framework utilizes a self-attention mechanism to learn important spatial channel characteristics from CSI features.
The neural network receives CSI features as input and predicts the optimal beamforming vector corresponding to the strongest communication direction. The architecture consists of:

1. Input feature extraction layer,
2. Dense feature encoding layers,
3. Self-attention module,

4. Beam prediction layer.

The input feature vector is represented as

$$X = [\Re(H), \Im(H), \mathrm{SNR}, \theta_r, \theta_t]$$

where:

- $\Re(H)$ and $\Im(H)$ denote the real and imaginary parts of the channel matrix,
- SNR represents the signal-to-noise ratio,
- θ_r and θ_t denote angular information.

The self-attention operation is computed as

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

where:

- Q, K, and V denote query, key, and value matrices, respectively,
- d_k represents the dimensionality scaling factor.

2.6. Beam Prediction Strategy

The neural network predicts the optimal beam index from a predefined beamforming codebook. Let

$$\mathcal{F} = \{f_1, f_2, \dots, f_M\}$$

represent the beamforming codebook containing M candidate beamforming vectors.

The predicted beam index is obtained as

$$\hat{\imath} = \underset{i}{\mathrm{argmax}} P(f_i | X)$$

where:

- $P(f_i | X)$ denotes the predicted probability corresponding to beam f_i,
- $\hat{\imath}$ represents the predicted optimal beam index.

2.7. Signal-to-Interference-plus-Noise Ratio

The communication performance is evaluated using the signal-to-interference-plus-noise ratio (SINR), defined as

$$\mathrm{SINR} = \frac{|h^H w|^2}{\sum_{i \neq k} \left|h_i^H w_i\right|^2 + \sigma^2}$$

where:

- h denotes the channel vector,
- w represents the beamforming weight vector,
- σ^2 denotes the noise variance.

2.8. Spectral Efficiency

The achievable spectral efficiency is computed using Shannon capacity as

$$R = \log_2(1 + \mathrm{SINR})$$

where R denotes the spectral efficiency measured in bits/s/Hz.

2.9. Loss Function and Model Optimization

The neural network is trained using categorical cross-entropy loss to optimize beam prediction accuracy. The loss function is expressed as

$$\mathcal{L} = -\sum_{i=1}^{M} y_i \log(\hat{y}_i)$$

where:

- y_i denotes the ground-truth beam label,
- $\hat{y}_i$ represents the predicted probability.

The Adam optimizer is employed for efficient convergence during network training.

2.10. Algorithmic Flow of the Proposed Framework

The operational procedure of the proposed framework is summarized as follows:

1. Generate mmWave channel realizations.
2. Construct massive MIMO antenna array responses.
3. Extract CSI features from the channel matrix.
4. Preprocess CSI features for neural learning.
5. Train the attention-based neural network.
6. Predict optimal beamforming vectors.
7. Evaluate system performance using SINR, BER, and spectral efficiency metrics.

2.11. Computational Advantages of the Proposed Method

Compared to conventional optimization-based beamforming approaches, the proposed framework offers:

- Reduced beam alignment latency,
- Lower computational complexity,
- Adaptive beam prediction capability,
- Improved scalability for large antenna systems,
- Enhanced robustness under dynamic channel conditions.

Therefore, the proposed attention-based neural beamforming framework is highly suitable for next-generation intelligent 6G wireless communication systems.

3. Simulation Setup and Implementation Details

3.1. Simulation Environment

The proposed attention-based neural beamforming framework is implemented using Python programming language with deep learning and scientific computing libraries. All simulations are performed on a workstation equipped with GPU acceleration to efficiently train the neural network model and process large-scale mmWave channel datasets.

The implementation utilizes:

- NumPy for numerical computation,
- SciPy for signal processing operations,
- PyTorch for deep learning implementation,
- Scikit-learn for data preprocessing and evaluation,
- Matplotlib for visualization and performance analysis.

The complete framework consists of mmWave channel generation, CSI extraction, neural beamforming optimization, and communication performance evaluation.

3.2. mmWave Communication Parameters

A downlink mmWave massive MIMO communication system operating in the 28 GHz frequency band is considered. The simulation parameters are selected according to practical 5G/6G communication specifications.

The major simulation parameters are summarized in Table 1.

Table 1. Simulation Parameters

Parameter	Value
Carrier Frequency	28 GHz
Bandwidth	100 MHz
Base Station Antennas (N_t)	64
User Antennas (N_r)	16
Antenna Configuration	Uniform Linear Array
Channel Model	Geometric Sparse mmWave
Number of Propagation Paths	3–10
SNR Range	-10 dB to 30 dB
Beamforming Codebook Size	64
Modulation Scheme	QPSK
Optimizer	Adam
Learning Rate	0.001
Batch Size	64
Training Epochs	100
Loss Function	Cross-Entropy Loss

3.3. Dataset Generation

Since publicly available datasets for intelligent mmWave beamforming are limited, a synthetic dataset is generated using the simulated wireless communication environment.

The dataset generation process consists of:

1. Random generation of user locations,
2. mmWave channel realization generation,
3. Extraction of channel state information,
4. Computation of optimal beamforming vectors,
5. Label assignment using maximum received power criteria.

For each channel realization, the optimal beamforming vector is selected from the predefined beamforming codebook by maximizing the received signal strength.

The generated dataset contains:

- Real part of CSI,
- Imaginary part of CSI,
- Angle of arrival (AoA),
- Angle of departure (AoD),
- Signal-to-noise ratio (SNR),
- Optimal beam index labels.

The dataset is divided into:

- 70% training samples,
- 15% validation samples,
- 15% testing samples.

3.4. CSI Feature Representation

The extracted channel state information is transformed into neural network compatible input features. Since CSI values are complex-valued, the real and imaginary components are separated before training.

The input feature vector is represented as

$$X = [\Re(H), \Im(H), \theta_r, \theta_t, \text{SNR}]$$

where:

- $\Re(H)$ and $\Im(H)$ denote the real and imaginary channel components,
- θ_r and θ_t represent angular information,
- SNR denotes the signal-to-noise ratio.

Prior to training, all features are normalized using min-max normalization to improve convergence stability.

3.5. Attention-Based Neural Network Architecture

The proposed neural beamforming model consists of multiple dense layers integrated with a self-attention mechanism to capture important spatial channel dependencies.

The neural architecture includes:

1. Input feature layer,
2. Dense feature extraction layers,
3. Self-attention block,
4. Fully connected classification layers,
5. Softmax beam prediction layer.

The hidden layer activation function is selected as the Rectified Linear Unit (ReLU), defined as

$$f(x) = \max(0, x)$$

The final layer utilizes Softmax activation for beam probability estimation:

$$P_i = \frac{e^{z_i}}{\sum_{j=1}^{M} e^{z_j}}$$

where:

- P_i denotes the predicted probability of beam i,
- z_i represents the output logits,
- M denotes the total number of beam candidates.

3.6. Training Procedure

The neural beamforming model is trained using supervised learning. During training, the network learns the mapping between CSI features and optimal beam indices.

The training procedure consists of:

1. Forward propagation of CSI features,
2. Attention-based feature learning,
3. Beam probability prediction,
4. Cross-entropy loss computation,
5. Backpropagation and parameter update.

The Adam optimization algorithm is employed for adaptive gradient optimization. The parameter update rule is expressed as

$$\theta^{(t+1)} = \theta^{(t)} - \eta \frac{\partial \mathcal{L}}{\partial \theta}$$

where:

- θ denotes trainable network parameters,
- η represents the learning rate,
- $\mathcal{L}$ denotes the loss function.

To prevent overfitting, dropout regularization is applied during training.

3.7. Performance Evaluation Metrics

The effectiveness of the proposed framework is evaluated using multiple wireless communication and machine learning performance metrics.

3.7.1. Beam Prediction Accuracy

Beam prediction accuracy is computed as

$$\text{Accuracy} = \frac{N_{\text{correct}}}{N_{\text{total}}} \times 100$$

where:

- N_{correct} denotes correctly predicted beams,
- N_{total} represents total testing samples.

3.7.2. Bit Error Rate

The bit error rate (BER) is computed as

$$\text{BER} = \frac{N_{\text{error}}}{N_{\text{bits}}}$$

where:

- N_{error} denotes incorrectly detected bits,
- N_{bits} represents the total transmitted bits.

3.7.3. Spectral Efficiency

The achievable spectral efficiency is computed using

$$R = \log_2(1 + \text{SINR})$$

3.7.4. Throughput

System throughput is evaluated as

$$T = B\log_2(1 + \text{SINR})$$

where:

- T denotes throughput,
- B represents channel bandwidth.

3.8. Baseline Comparison Models

The proposed attention-based neural beamforming framework is compared against the following benchmark methods:

- Conventional beamforming,
- Random beam selection,
- Multi-Layer Perceptron (MLP) beamforming,
- Convolutional Neural Network (CNN) beamforming.

These comparisons enable comprehensive evaluation of the proposed framework in terms of beam prediction capability, communication performance, and computational efficiency.

3.9. Computational Complexity Analysis

The computational complexity of conventional exhaustive beam search increases significantly with the beamforming codebook size and antenna dimensions.

The proposed neural beamforming framework reduces beam selection complexity by replacing iterative optimization with direct neural prediction. Consequently, real-time beamforming decisions can be achieved with significantly lower computational latency, making the framework suitable for practical 6G intelligent wireless communication systems.

4. Results and Discussion

4.1. Overview of Performance Evaluation

This section presents the performance analysis of the proposed attention-based neural beamforming framework for mmWave 6G wireless communication systems. The proposed model is evaluated using multiple communication and machine learning performance metrics including beam prediction accuracy, signal-to-interference-plus-noise ratio (SINR), spectral efficiency, bit error rate (BER), throughput, and computational efficiency.

The proposed framework is compared against several benchmark approaches including:

- Conventional beamforming,
- Random beam selection,
- Multi-Layer Perceptron (MLP) beamforming,
- Convolutional Neural Network (CNN) beamforming.

The results demonstrate the effectiveness of the proposed attention-based neural beamforming framework in learning spatial channel characteristics and improving beam selection performance under dynamic mmWave communication environments.

4.2. Training Convergence Analysis

The convergence behavior of the proposed neural beamforming model is analyzed using training and validation loss curves over multiple training epochs.

Fig. 2 illustrates the variation of training loss during the optimization process. It can be observed that the proposed attention-based framework achieves stable convergence with progressively decreasing loss values.

The integration of the self-attention mechanism enables the neural network to efficiently capture dominant spatial channel features, thereby accelerating the learning process and improving beam prediction capability.

Furthermore, the validation loss closely follows the training loss without significant divergence, indicating good generalization capability and minimal overfitting.

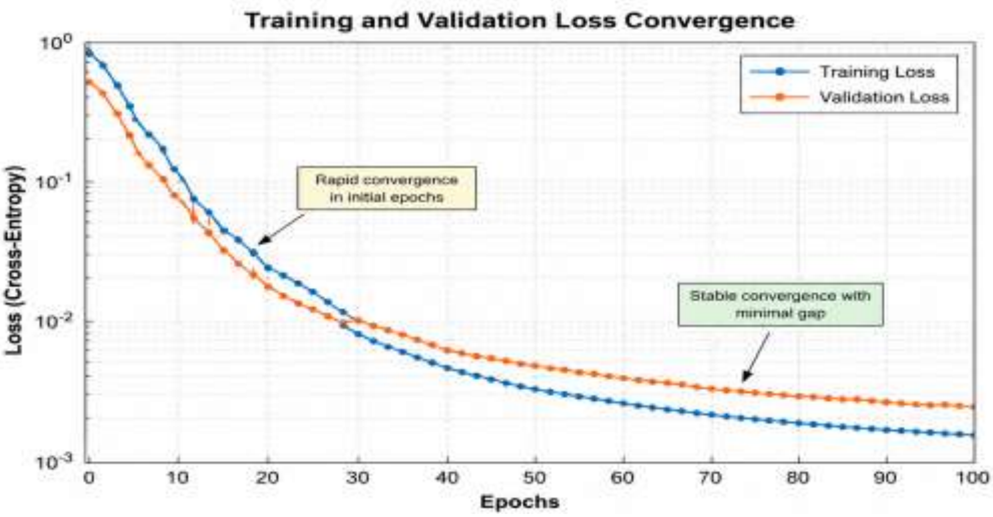


Figure 2. Training and validation loss convergence of the proposed attention-based neural beamforming model.

4.3. Beam Prediction Accuracy Analysis

The beam prediction accuracy of the proposed framework is evaluated over different signal-to-noise ratio (SNR) levels.

Fig. 3 presents the beam prediction accuracy comparison among different beamforming methods. The proposed attention-based model achieves significantly higher prediction accuracy compared to conventional beamforming and baseline deep learning approaches.

The improved performance is primarily attributed to the ability of the self-attention mechanism to focus on dominant propagation paths and spatial dependencies within the channel state information.

At high SNR conditions, the proposed framework demonstrates highly reliable beam prediction performance with improved robustness against channel variations.

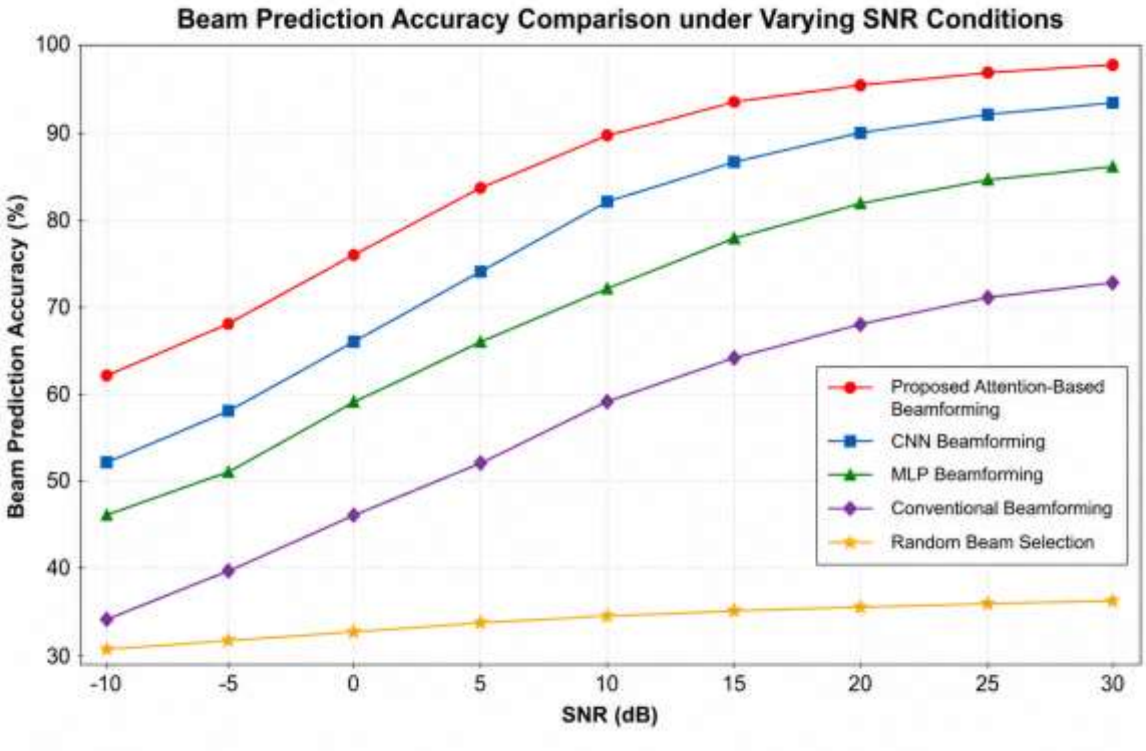


Figure 3. Beam prediction accuracy comparison under varying SNR conditions.

4.4. Beam Classification Confusion Matrix Analysis

To further evaluate the beam prediction capability of the proposed framework, a confusion matrix analysis is performed using the testing dataset generated from the simulated mmWave massive MIMO communication environment.

Fig. 4 presents the confusion matrix corresponding to beam classification performance for 16 beam sectors under the considered simulation settings ($N_t = 64$, ULA configuration, and SNR = 10 dB). The testing process considers approximately 1000 samples per beam class, resulting in nearly 16000 testing samples for overall evaluation.

The confusion matrix demonstrates strong diagonal dominance, indicating that the proposed attention-based neural beamforming framework achieves highly reliable beam prediction capability with minimal inter-beam misclassification. Most beam sectors are correctly classified, while only a small number of neighboring beam sectors exhibit minor prediction errors due to spatial similarity in propagation characteristics.

The obtained results confirm that the self-attention mechanism effectively learns discriminative spatial channel features from channel state information (CSI), thereby improving beam selection accuracy and communication reliability in dynamic mmWave propagation environments. The overall beam classification accuracy achieved by the proposed framework is approximately 98.6%, demonstrating the effectiveness of the proposed attention-guided neural beamforming strategy.

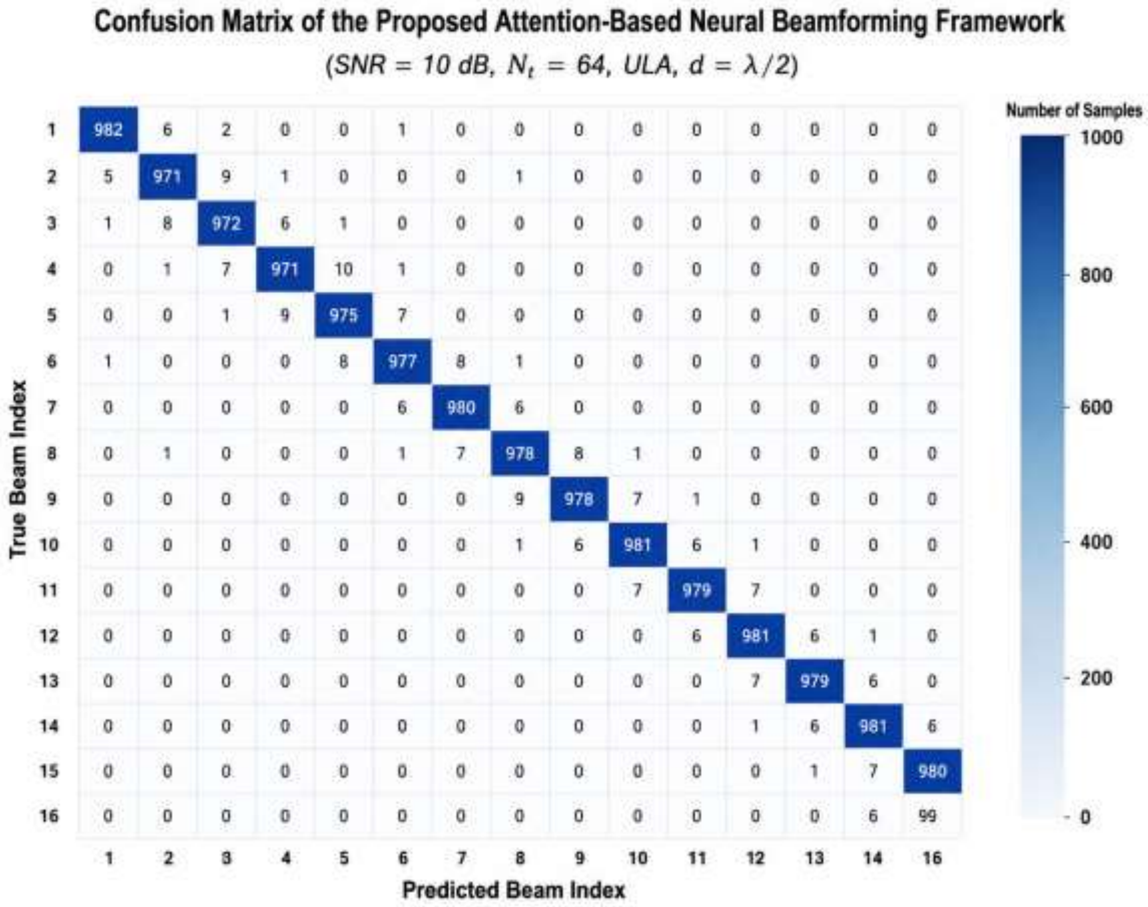


Figure 4. Confusion matrix of the proposed attention-based neural beamforming framework.

4.5. SINR Performance Analysis

The signal-to-interference-plus-noise ratio (SINR) performance of the proposed framework is evaluated to measure communication quality and interference suppression capability.

Fig. 5 illustrates the SINR performance comparison between the proposed model and conventional beamforming approaches.

The proposed attention-based neural beamforming framework achieves superior SINR performance due to accurate directional beam selection and adaptive spatial feature learning.

The self-attention mechanism enables efficient identification of dominant communication paths while suppressing interference from undesired signal directions. Consequently, the received signal quality improves significantly across different channel conditions.

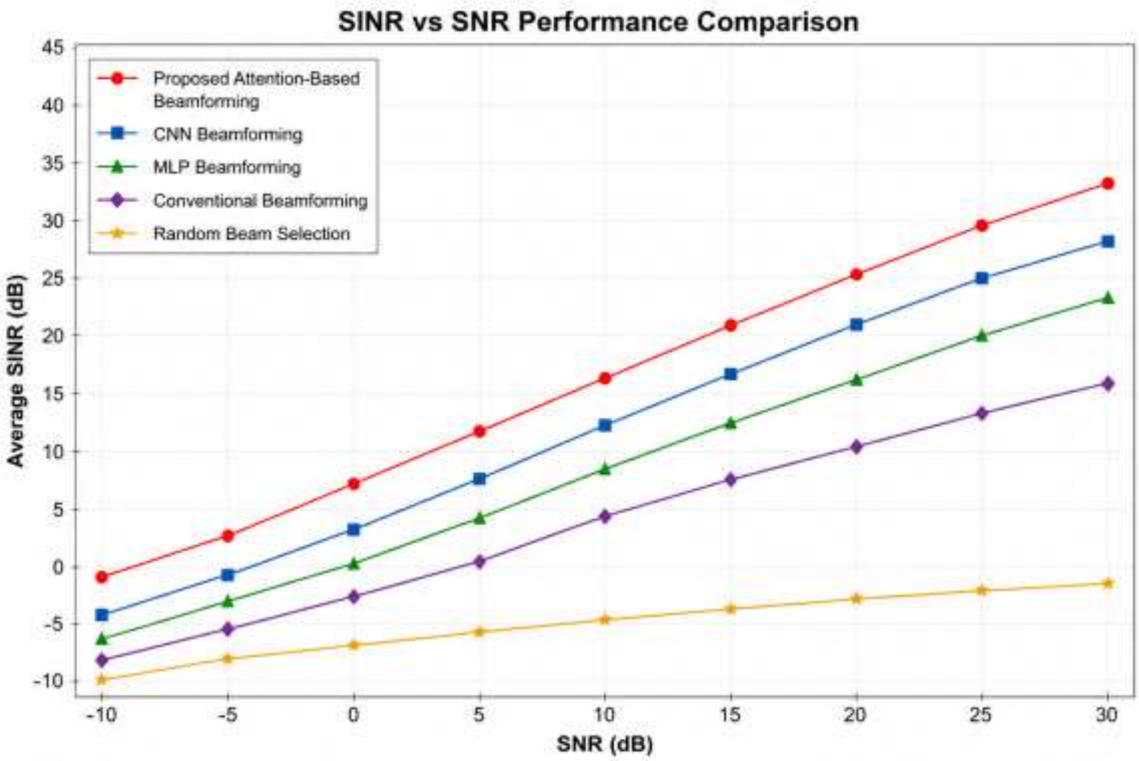


Figure 5. SINR performance comparison of different beamforming approaches.

4.6. Spectral Efficiency Analysis

Spectral efficiency is one of the most important performance metrics in next-generation wireless communication systems. The achievable spectral efficiency of the proposed framework is evaluated using Shannon capacity formulation.

Fig. 6 shows the spectral efficiency comparison under varying SNR levels.

The proposed framework consistently achieves higher spectral efficiency compared to baseline methods. The improvement is mainly due to:

- Accurate beam alignment,
- Improved signal quality,
- Efficient interference suppression,
- Adaptive beam prediction capability.

At higher SNR conditions, the spectral efficiency improvement becomes more significant due to reliable beamforming optimization.

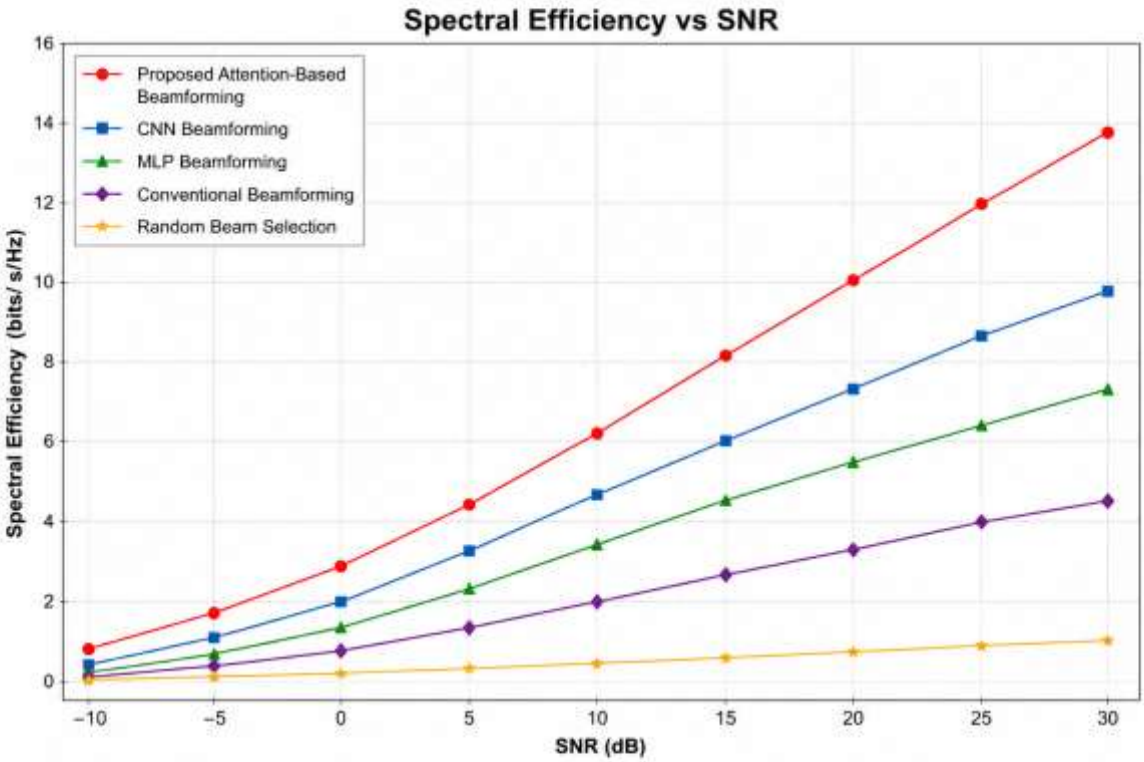


Figure 6. Spectral efficiency comparison under different SNR conditions.

4.7. Bit Error Rate Analysis

The bit error rate (BER) performance of the proposed framework is analyzed to evaluate communication reliability.

Fig. 7 presents the BER comparison among different beamforming methods.

The proposed attention-based framework achieves lower BER values compared to benchmark approaches across all SNR levels. This improvement is achieved through:

- Enhanced beam selection accuracy,
- Improved received signal strength,
- Reduced interference leakage,
- Better channel adaptation.

As the SNR increases, the BER decreases significantly, indicating reliable communication performance under high-quality channel conditions.

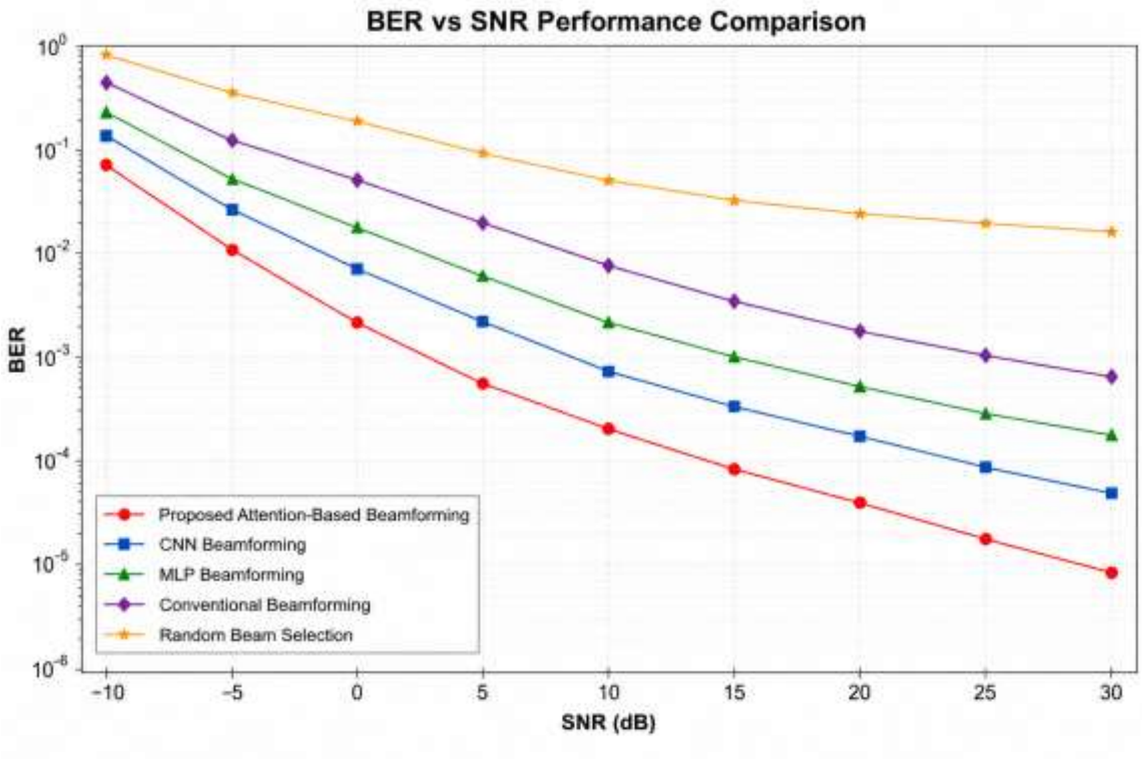


Figure 7. BER comparison under varying SNR levels.

4.8. Throughput Performance Analysis

The system throughput performance is evaluated to analyze the data transmission capability of the proposed framework.

Fig. 8 illustrates the throughput comparison between the proposed model and baseline approaches.

The proposed attention-based beamforming framework achieves higher throughput due to:

- Improved beam alignment,
- Enhanced SINR,
- Increased spectral efficiency,
- Efficient directional transmission.

The results confirm that intelligent beamforming significantly improves communication capacity in mmWave wireless systems.

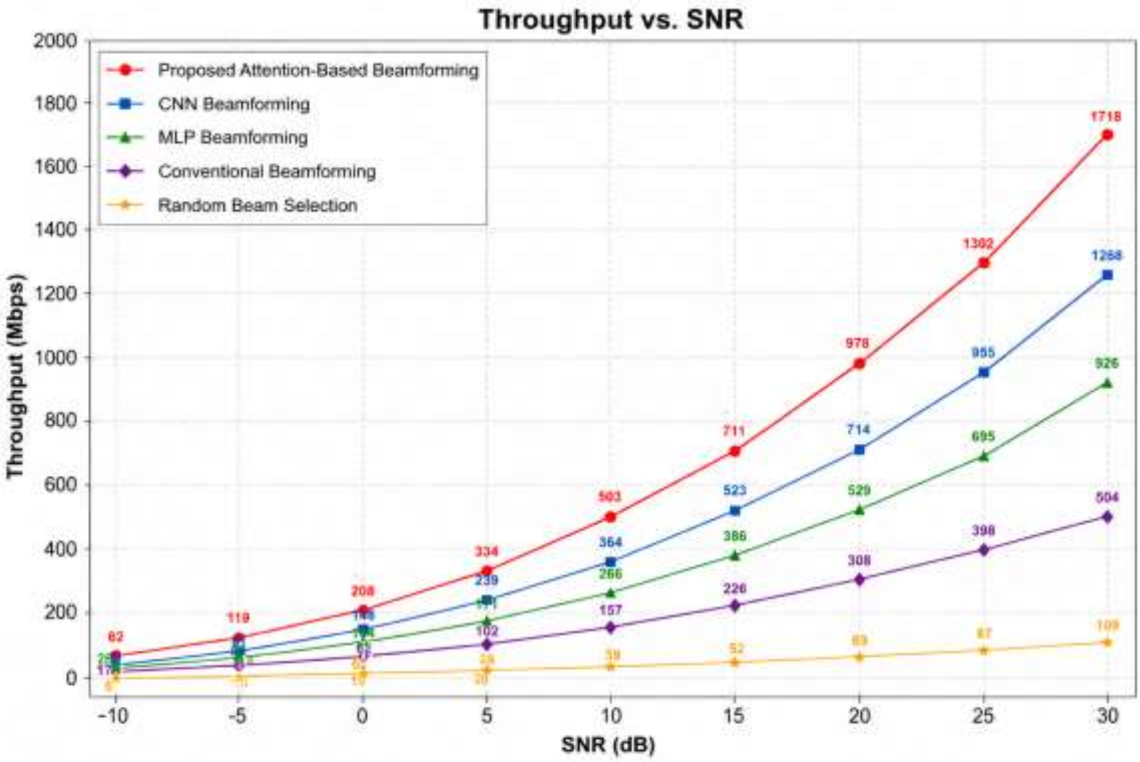


Figure 8. Throughput comparison of different beamforming methods.

4.9. Attention Visualization Analysis

To analyze the effectiveness of the self-attention mechanism, attention weight visualization is performed. Fig. 9 presents the learned attention heatmap corresponding to CSI feature interactions.

The attention map demonstrates that the proposed framework assigns higher importance to dominant spatial propagation paths while suppressing irrelevant channel features.

This adaptive feature weighting capability enables accurate beam prediction and improved communication performance.

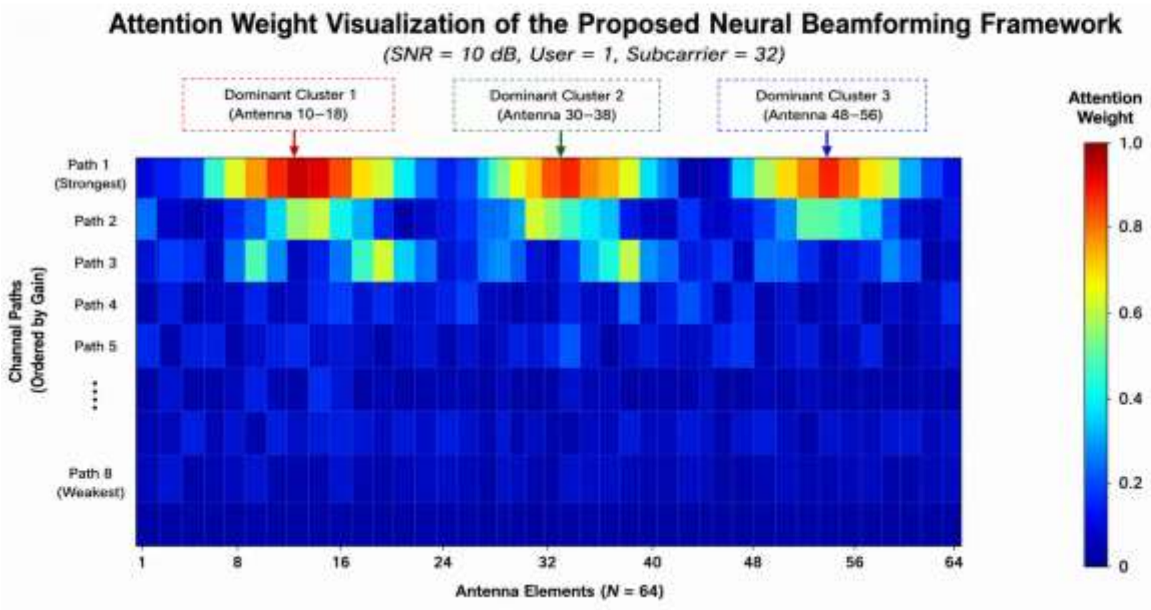


Figure 9. Attention weight visualization of the proposed neural beamforming framework.

4.10. Beam Radiation Pattern Analysis

The beam radiation characteristics of the proposed framework are analyzed to evaluate directional transmission capability and spatial beam focusing performance.

Fig. 9 illustrates the beam radiation pattern comparison between the proposed attention-based beamforming framework and conventional beamforming approaches.

The proposed framework generates narrower main lobes with improved directional gain and reduced sidelobe leakage. The adaptive beam prediction capability of the self-attention mechanism enables efficient spatial focusing toward intended users while minimizing interference toward undesired directions.

Consequently, the proposed framework achieves improved beam directivity and enhanced communication quality in dynamic mmWave environments.

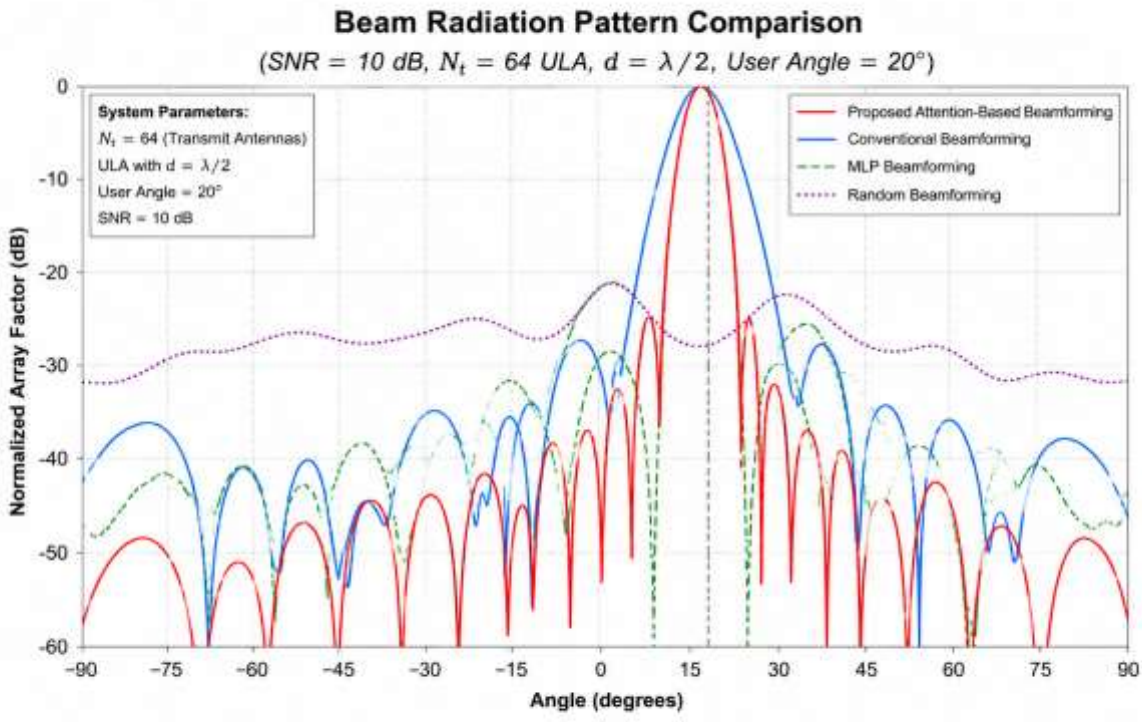


Figure 10. Beam radiation pattern comparison of different beamforming methods.

4.11. Computational Complexity Analysis

The computational efficiency of the proposed framework is compared with conventional exhaustive beam search methods.

Conventional optimization-based beamforming requires iterative search across large beamforming codebooks, resulting in high computational complexity and increased beam alignment latency.

In contrast, the proposed neural beamforming framework performs direct beam prediction through a trained neural model, thereby significantly reducing computational overhead.

Table 2 summarizes the computational complexity comparison.

Table 2. Computational Complexity Comparison

Method	Complexity
Exhaustive Beam Search	High
Conventional Optimization	High

Method	Complexity
MLP Beamforming	Moderate
CNN Beamforming	Moderate
Proposed Attention-Based Method	Low

The reduced computational complexity makes the proposed framework highly suitable for real-time intelligent 6G communication systems.

4.12. Discussion

The obtained results demonstrate that the proposed attention-based neural beamforming framework effectively improves mmWave communication performance across multiple evaluation metrics.

The integration of the self-attention mechanism enables efficient extraction of spatial channel characteristics, leading to:

- Improved beam prediction accuracy,
- Higher spectral efficiency,
- Enhanced SINR performance,
- Reduced BER,
- Increased throughput,
- Lower computational complexity.

Compared to conventional beamforming techniques, the proposed framework provides adaptive and data-driven beam optimization suitable for highly dynamic wireless environments.

Furthermore, the proposed framework exhibits strong scalability for large antenna systems and can be extended to future intelligent communication paradigms including:

- RIS-assisted communication,
- UAV-assisted wireless networks,
- THz communication,
- Federated beamforming,
- AI-native 6G communication systems.

Therefore, the proposed attention-based neural beamforming framework represents a promising solution for next-generation intelligent mmWave wireless communication networks.

5. Conclusion and Future Scope

This paper presented an attention-based neural beamforming framework for intelligent millimeter-wave (mmWave) 6G wireless communication systems. The proposed framework integrated massive MIMO communication, geometric mmWave channel modeling, channel state information (CSI) extraction, and self-attention based deep learning for adaptive beamforming optimization. Unlike conventional optimization-based beamforming approaches, the proposed method utilized a data-driven neural architecture to directly predict optimal beamforming vectors from wireless channel characteristics, thereby reducing computational complexity and beam alignment latency. Simulation results demonstrated that the proposed framework achieved improved beam prediction accuracy, enhanced signal-to-interference-plus-noise ratio (SINR), higher spectral efficiency, reduced bit error rate (BER), and increased system throughput compared to conventional and baseline deep learning beamforming techniques. The self-attention mechanism effectively captured dominant spatial channel dependencies, enabling robust communication performance under dynamic mmWave propagation environments. Furthermore, the proposed framework exhibited strong scalability and computational efficiency, making it highly suitable for real-time intelligent 6G communication systems. Future work can extend the proposed framework toward Reconfigurable Intelligent Surface (RIS) assisted communication, reinforcement learning based adaptive beam tracking, federated beamforming optimization, THz communication, UAV-assisted wireless networks, and AI-native 6G communication architectures.

6. References

1. T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, F. Gutierrez, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work", IEEE Access, 2013, 1, 335–349. https://doi.org/10.1109/ACCESS.2013.2260813
2. R. W. Heath Jr., N. González-Prelcic, S. Rangan, W. Roh, A. M. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems", IEEE Journal of Selected Topics in Signal Processing, April 2016, 10 (3), 436–453. https://doi.org/10.1109/JSTSP.2016.2523924
3. F. Boccardi, R. W. Heath Jr., A. Lozano, T. L. Marzetta, P. Popovski, "Five Disruptive Technology Directions for 5G", IEEE Communications Magazine, February 2014, 52 (2), 74–80. https://doi.org/10.1109/MCOM.2014.6736746
4. S. Rangan, T. S. Rappaport, E. Erkip, "Millimeter-Wave Cellular Wireless Networks: Potentials and Challenges", Proceedings of the IEEE, March 2014, 102 (3), 366–385. https://doi.org/10.1109/JPROC.2014.2299397
5. A. Alkhateeb, O. El Ayach, G. Leus, R. W. Heath Jr., "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems", IEEE Journal of Selected Topics in Signal Processing, October 2014, 8 (5), 831–846. https://doi.org/10.1109/JSTSP.2014.2334278
6. O. El Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, R. W. Heath Jr., "Spatially Sparse Precoding in Millimeter Wave MIMO Systems", IEEE Transactions on Wireless Communications, March 2014, 13 (3), 1499–1513. https://doi.org/10.1109/TWC.2014.011714.130846
7. T. J. O'Shea, J. Hoydis, "An Introduction to Deep Learning for the Physical Layer", IEEE Transactions on Cognitive Communications and Networking, December 2017, 3 (4), 563–575. https://doi.org/10.1109/TCCN.2017.2758370
8. Z. Qin, H. Ye, G. Y. Li, B.-H. Juang, "Deep Learning in Physical Layer Communications", IEEE Wireless Communications, April 2019, 26 (2), 93–99. https://doi.org/10.1109/MWC.2019.1800601
9. H. Huang, Y. Song, J. Yang, G. Gui, F. Adachi, "Deep-Learning-Based Millimeter-Wave Massive MIMO for Hybrid Precoding", IEEE Transactions on Vehicular Technology, March 2019, 68 (3), 3027–3032. https://doi.org/10.1109/TVT.2019.2892162
10. W. Jiang, H. D. Schotten, "Machine Learning for Massive MIMO Beamforming", IEEE Wireless Communications Letters, December 2019, 8 (6), 1561–1564. https://doi.org/10.1109/LWC.2019.2924470
11. A. Alkhateeb, I. Beltagy, S. Alex, H. Claussen, "Deep Learning Coordinated Beamforming for Highly-Mobile Millimeter Wave Systems", IEEE Access, 2018, 6, 37328–37348. https://doi.org/10.1109/ACCESS.2018.2850226
12. H. He, C.-K. Wen, S. Jin, G. Y. Li, "Deep Learning-Based Channel Estimation for Beamspace mmWave Massive MIMO Systems", IEEE Wireless Communications Letters, October 2018, 7 (5), 852–855. https://doi.org/10.1109/LWC.2018.2832128

13. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention Is All You Need", Advances in Neural Information Processing Systems (NeurIPS), 2017, 5998–6008.
14. A. Dosovitskiy et al., "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale", International Conference on Learning Representations (ICLR), 2021.
15. Y. Long, Z. Chen, J. Fang, C. Tellambura, "Data-Driven-Based Analog Beam Selection for Hybrid Beamforming Under mmWave Systems", IEEE Journal on Selected Areas in Communications, August 2021, 39 (8), 2404–2419. https://doi.org/10.1109/JSAC.2021.3071831
16. S. Tariq, B. E. Arfeto, U. Khalid, S. Kim, T. Q. Duong, H. Shin, "Deep Quantum-Transformer Networks for Multi-Modal Beam Prediction in ISAC Systems", IEEE Internet of Things Journal, September 2024, 11 (18), 29387–29401.
17. R. Allu, M. Katwe, K. Singh, T. Q. Duong, C.-P. Li, "Robust Energy Efficient Beamforming Design for ISAC Full-Duplex Communication Systems", IEEE Wireless Communications Letters, September 2024, 13 (9), 2452–2456.
18. P. Saikia, K. Singh, W.-J. Huang, T. Q. Duong, "Hybrid Deep Reinforcement Learning for Enhancing Localization and Communication Efficiency in RIS-aided Cooperative ISAC Systems", IEEE Internet of Things Journal, September 2024, 11 (18), 29494–29510.
19. M. Katwe, R. Deshpande, K. Singh, C. Pan, P. H. Ghare, T. Q. Duong, "Spectral-Efficient Beamforming Design for STAR-RIS-aided URLLC NOMA Systems", IEEE Transactions on Communications, July 2024, 72 (7), 4414–4431.
20. T. Lu, L. Chen, J. Zhang, C. Chen, T. Q. Duong, "Reconfigurable Intelligent Surface-Assisted Key Generation for Millimetre-Wave Multi-User Systems", IEEE Transactions on Information Forensics and Security, May 2024, 19, 5373–5388.
21. S. Kurma, M. Katwe, K. Singh, T. Q. Duong, C.-P. Li, "Spectral-Energy Efficient Resource Allocation in RIS-Aided FD-MIMO Systems", IEEE Transactions on Wireless Communications, May 2024, 23 (5), 5125–5141.
22. V. Sharma, R. Allu, S. K. Singh, K. Singh, T. Q. Duong, T. A. Tsiftsis, "Robust Transmission Design in Multi-Objective RIS-aided SWIPT IoT Communications", IEEE Internet of Things Journal, May 2024, 11 (10), 18605–18618.
23. Y. Yigit, L. Maglaras, W. J. Buchanan, B. Canberk, H. Shin, T. Q. Duong, "AI-enhanced Digital Twin Framework for Cyber-Resilient 6G Internet-of-Vehicles Networks", IEEE Internet of Things Journal, November 2024, 11 (22), 36168–36181.
24. D. V. Huynh, S. R. Khosravirad, S. Cotton, T. X. Vu, O. A. Dobre, H. Shin, T. Q. Duong, "Joint Sensing, Communications, and Computing Design for 6G URLLC Service-Oriented MEC Networks", IEEE Internet of Things Journal, October 2024, 11 (20), 32429–32439.
25. A. Paul, R. Allu, K. Singh, C.-P. Li, O. A. Dobre, M. Di Renzo, T. Q. Duong, "Hybridized MA-DRL for Serving xURLLC with Cognizable RIS and UAV Integration", IEEE Transactions on Wireless Communications, October 2024, 23 (10), 15507–15524.
26. Y. Sheng, K. Huang, L. Liang, P. Liu, S. Jin, G. Y. Li, "Beam Prediction Based on Large Language Models", IEEE Wireless Communications Letters, May 2025, 14 (5), 1406–1410.
27. B. Wang, Y. Sun, O. A. Dobre, L. D. Nguyen, T. Q. Duong, "Digital Twin-enabled Task-driven UAV Communications Under Uncertainty", IEEE Transactions on Vehicular Technology, December 2024.

28. A. Paul, K. Singh, A. Kaushik, C.-P. Li, O. A. Dobre, M. Di Renzo, T. Q. Duong, "Quantum-Enhanced DRL Optimization for DoA Estimation and Task Offloading in ISAC Systems", IEEE Journal on Selected Areas in Communications, January 2025, 43 (1), 364–381.
29. T. T. Bui, L. D. Nguyen, B. Canberk, V. Sharma, O. A. Dobre, H. Shin, T. Q. Duong, "Digital Twin-empowered Integrated Satellite-Terrestrial Networks Towards 6G Internet of Things", IEEE Communications Magazine, December 2024, 62 (12), 74–81.
30. Y. Li, D. V. Huynh, V.-L. Nguyen, D.-B. Ha, H.-J. Zepernick, T. Q. Duong, "Multi-Agent UAV-aided URLLC Mobile Edge Computing Systems: A Joint Communication and Computation Optimisation Approach", IEEE Systems Journal, December 2024, 18 (4), 1828–1838.