SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I1P114

Volume 12 Issue 1, 174-192, January 2025

Original Article

Bayesian Reinforcement Learning for Beam Forming in Millimetre-Wave Networks for Multi-Target Detection in 5G Radio Resource Allocation

Phaneendra Jonnabhatla[1], T. Vimala[2]

[1,2]*Department of ECE, Dr. M.G.R. Educational and Research Institute, Tamilnadu, India.*

[1]*Corresponding Author : phaneendrajonnabhatla97@gmail.com*

Received: 18 November 2024 Revised: 24 December 2024 Accepted: 12 January 2025 Published: 30 January 2025

Abstract *- Millimeter wave (mmWave) beam forming has long been explored as a key enabler of Fifth-Generation (5G) technologies. Wireless network providers face numerous challenges due to the increasing density of wireless systems, including, but not limited to, the high costs associated with fiber construction and site acquisition. To address these challenges and take advantage of the maturity of mmWave communication for wireless backhaul links, Integrated Access and Backhaul (IAB) systems have been developed for 5G networks. In the context of 5G radio distribution, Bayesian Reinforcement Learning (BRL) introduces a novel approach to beamforming in mmWave systems, optimizing multi-target identification. Optimal beamforming is crucial for maximizing network capacity and ensuring effective utilization of the radio spectrum, as mmWave frequencies offer significant capacity and high-resolution capabilities. The proposed BRL framework improves beamforming strategies by probabilistically adapting to changing network conditions and user requirements through Bayesian estimation, thereby enhancing the accuracy of multi-target identification. The system dynamically adjusts beam patterns and resource allocation in real time based on immediate feedback, optimizing both signal quality and resource utilization. Experimental results demonstrate that BRL outperforms existing methods in resource utilization and target identification reliability, thereby increasing the capacity of mmWave systems in the 5G era. Extensive simulations and numerical results show that BRL-mmWave significantly improves efficiency and accelerates training compared to previous approaches.*

Keywords *- Bayesian reinforcement learning, Beamforming, Millimeter-wave networks, Multi-target detection, 5G radio resource allocation, Adaptive beamforming, Network optimization, Resource utilization.*

1. Introduction

The development of 5G technology has led to the expectation that it will enable data transmission rates up to 1000 times faster than current 4G standards. To meet the growing demand for high-quality cellular networks and accommodate a large number of users, high-energy waves are essential [1]. mmWave communication has garnered significant attention for 5G wireless networks due to its advantages in achieving high data speeds and addressing spectrum scarcity. The mmWave spectrum operates in the 30 GHz to 300 GHz range, but its higher frequencies experience significant path loss and can only cover short distances [2]. As a result, smaller cellular cells operating at higher frequencies are necessary to overcome challenges such as path loss and obstruction [3]. Massive Multiple-Input Multiple-Output (mMIMO) technology uses hundreds of antennas to transmit signals within a single time-frequency spectrum and can serve tens of millions of users simultaneously [4]. Large-scale antenna networks significantly enhance the potential for increased upload and download throughput by employing a general channel model that accounts for issues such as pilot contamination, path loss, inaccurate channel estimation, and antenna correlations specific to individual endpoints. These networks consistently deliver high data rates on both forward and reverse link connections, even in environments with rapid propagation changes [5]. For massive MIMO structures, hybrid beamforming is advantageous as it combines analog and digital transceiver topologies, allowing for analog processing in the Radio Frequency (RF) domain. Unlike baseband digital signals, RF signals in the mmWave frequency band cannot be adjusted to arbitrary frequencies due to the high costs of analog domain processing [6]. Codebook-based hybrid beamforming selects RF precoders from predefined codebooks based on optimization criteria (e.g., sum rate), which is an attractive solution for practical massive MIMO mmWave structures [7]. Selecting the optimal beam for RF precoders is challenging due to its dependence on user requirements and interference, making it significantly more complex than traditional methods. To optimize overall network efficiency, integrated optimization of user scheduling, RF precoder selection, and power management is crucial, necessitating novel approaches to avoid the computational complexity of brute-force search methods as antenna array sizes increase [8]. Reinforcement Learning (RL) shows promise for enhancing future internet connections. RL methods often require vast amounts of interaction data from the environment, which can be impractical for real-world applications due to poor data efficiency and slow learning speeds [9]. Model-based techniques aim to learn a parameterized model P_{φ} for the dynamics of the unknown environment, which can significantly accelerate policy learning by generating simulated experiences based on the learned model [10]. While these methods can improve learning speed, the accuracy of the learned model can become a bottleneck, leading to lower performance compared to model-free methods. Although comprehensive research is still lacking, it is anticipated that effectively combining model-based and model-free learning approaches could greatly enhance both learning speed and system efficiency [11].

Beamforming, which involves the precise and efficient steering of beams with minimal signaling overhead, is critical for ensuring the connection of rapidly moving User Equipment (UE) to at least one Access Point (AP) at any given time, especially in environments with transmission differences and co-channel interference [12]. While fully digital beamforming offers high performance, it is costly, power-intensive, and requires specialized hardware. Hybrid beamforming, which uses traditional phase shifters in the RF domain to steer beams while employing digital signal processing in the baseband to mitigate interference, can achieve similar efficiency with lower costs and complexity [13]. In hybrid architectures, phase shifters typically connect RF chains to antennas. When the number of antennas equals the number of RF chains, power consumption is high, and interconnections are numerous when each RF chain is fully connected to every antenna [14]. To save power and reduce interconnections, RF chains, which are smaller in number than antennas, are connected to a subset of the antennas. Current beamforming strategies rely on the accurate acquisition of Channel State Information (CSI), which can be obtained through sparse channel estimation or exhaustive or centralized searching. This results in either insufficient CSI or an excess of signaling, both of which are exacerbated in high-mobility scenarios where cell connections and channels rapidly change [15].

1.1. Problem Statement

Efficient beamforming is a cornerstone for enhancing signal quality and reliability in millimeter-wave (mmWave) networks, particularly in the context of 5G systems where multi-target detection and radio resource allocation are critical challenges. Traditional beamforming methods often struggle to adapt to the dynamic and uncertain nature of 5G environments, including rapidly changing channel conditions, user mobility, and interference. Additionally, the high-dimensional state-action space, introduced by multiple users, antennas, and targets, poses computational challenges. These complexities are further compounded by the need to allocate limited 5G radio resources effectively while meeting stringent real-time performance requirements, such as low latency and high throughput. Bayesian Reinforcement Learning (BRL) offers a promising approach by integrating probabilistic reasoning with sequential decision-making to optimize beamforming strategies under uncertainty. However, key challenges include handling dynamic environmental uncertainties, managing the high dimensionality of the problem, and ensuring resource-efficient operations while adhering to 5G constraints. To address these issues, this research proposes a BRL-based framework for mmWave beamforming designed to enhance multi-target detection and optimize 5G radio resource allocation. The framework aims to dynamically adapt beamforming policies in real time, incorporate probabilistic models to address uncertainties, and leverage scalable optimization techniques to manage the high-dimensional state-action space. By tackling these challenges, the proposed solution seeks to improve spectral efficiency, enhance user experience, and ensure robust performance in 5G mmWave networks.

2. Related Works

There are various techniques available to minimize estimations in optimization problems. In fully connected frameworks, orthogonal matched pursuit is commonly used; however, in Partially-Connected Structures (PCSs), its performance is suboptimal [16]. For PCSs, the alternate reduction approach is utilized, but this method is computationally intensive in multiuser and multi-carrier scenarios. Low-complexity techniques such as convex relaxation and channel phasing separation are available, but they require extensive signaling, leading to uncertain CSI [17]. As specified in the initial version of 5G New Radio (NR), none of the existing methods meet the stringent requirements of high-mobility scenarios, such as those encountered in fast-moving trains [18]. For highly mobile scenarios, beamforming through deep supervised learning is a promising, adaptable, and statistically reliable method. In these systems, pilot signals are used to gather the RF characteristics of the surrounding environment and the locations of users or APs. Then, each rapidly moving UE is connected to at least one AP using online beamforming with a pre-trained model [19]. One approach that shows potential for reducing beam training overhead is utilizing data from prior training experiences to narrow the beam search area for future training sessions. The rapid development of MLbased beamforming algorithms feasible simplifies this process. The scope of SL-based methods is limited because they require a substantial number of training samples beforehand to achieve satisfactory results [20]. The second category based on more general sequential decision-making and optimization is RL. A beam training method capable of effectively leveraging contextual information was introduced based on a lightweight reinforcement learning technique known as the Multi-Armed Bandit (MAB) [21]. The limited framework of the MAB poses challenges in identifying significant patterns and making complex decisions. Temporal correlation information is effectively utilized to develop beamforming and/or data transmission strategies. For instance, spatial data derived from users' random mobility is employed [22].

To address these challenges, a hierarchical multi-resolution codebook-based adaptive Beam Alignment (BA) method was proposed [23]. This approach can partially bypass the costly process of exhaustively searching through all possible transmit and receive beam pairs. It is important to note that the performance of the hierarchical search method is heavily dependent on the selected hierarchical codebook. The hierarchical codebook has been extensively studied and has garnered significant interest due to its importance. Adaptive searching-based techniques still incur a substantial learning cost for large-scale antenna array structures [24]. In realistic dynamic environments, mmWave channels fade quickly, and the coherence time of each block is too short to allow for rapid and accurate beam alignment.

Compared to sub-6GHz channels, mmWave channels are generally sparser [25]. Traditional channel estimation methods, such as filtering and adaptive techniques, are generally ineffective for mmWave channel estimation due to the high pilot costs required for accurate channel estimation. The channel estimation technique proposed performs poorly in low Signal-to-Noise Ratio (SNR) conditions [26]. Various Sparse Bayesian Learning (SBL)-based methods have been proposed for different application scenarios. An Ultra-Reliable Access (URA) architecture supported by Reconfigurable Intelligent Surfaces (RIS) has been introduced to optimize passive reflections for active device isolation [27].

Recent advancements in beamforming and Reinforcement Learning (RL) techniques have significantly contributed to improving the performance of mmWave networks, which are essential for the success of 5G and beyond. Beamforming, particularly in mmWave communication systems, is crucial for effectively utilizing high-frequency bands and mitigating challenges such as signal attenuation and interference. Recent studies have focused on adaptive beamforming techniques that can dynamically adjust the beam direction and power to enhance signal quality and coverage.

For instance, a machine learning-based beamforming approach was proposed that uses neural networks to predict optimal beam configurations based on real-time channel conditions, demonstrating a substantial improvement in system throughput compared to traditional beamforming methods. Similarly, explored the integration of deep learning algorithms for beamforming optimization, showing how Deep Neural Networks (DNNs) can learn complex channel characteristics to reduce power consumption while maintaining high signal strength. The integration of Reinforcement Learning (RL) into beamforming has also emerged as a promising area of research, where RL-based methods are employed to optimize beamforming decisions in dynamic and uncertain environments.

In particular, [3] introduced a deep Q-learning approach to adaptive beamforming, where an agent learns to make beamforming decisions by interacting with the environment. Their work showed how Q-learning could reduce the dependency on traditional channel estimation techniques, providing better performance under rapidly changing network conditions. Proposed a model-free RL method for beamforming in mmWave networks, which utilizes an actor-critic model to optimize beam direction and power allocation simultaneously. Their results highlighted the potential of RL in achieving high network throughput while dynamically responding to fluctuations in user demand and channel quality.

Despite the promising results in beamforming and RL integration, several challenges remain, especially when applying these techniques to large-scale and heterogeneous network environments. Addressed this issue by incorporating a multi-agent RL approach for beamforming optimization, where multiple agents collaboratively learn to allocate resources and optimize beam directions. This study demonstrated how a decentralized RL framework could be scaled to support dense urban environments, improving beamforming performance while reducing computational overhead. Additionally, focused on the scalability of RL for beamforming in large mmWave networks, proposing a hierarchical RL approach to handle the increased complexity of large network topologies. This approach used a combination of global and local agents, allowing for efficient decision-making at both the macro and micro levels, which is essential for real-time beamforming in 5G and beyond.

In parallel with beamforming optimization, the use of Bayesian methods in reinforcement learning has gained attention as a means to improve decision-making under uncertainty. Introduced a Bayesian Reinforcement Learning (BRL) approach for optimizing power allocation and beamforming in mmWave networks, emphasizing the benefits of incorporating prior knowledge about the environment to enhance learning efficiency. By using Bayesian inference, their model could more accurately predict channel conditions and improve beamforming decisions with fewer training samples. Moreover, explored the use of Bayesian optimization for tuning RL parameters, providing a framework to optimize both the exploration-exploitation trade-off and the reward function in beamforming applications. This hybrid approach between Bayesian methods and RL could significantly improve the adaptability and convergence speed of beamforming systems in mmWave environments.

2.1. Research Gap

The research gap in applying Bayesian Reinforcement Learning (BRL) to beamforming for multi-target detection in millimeter-wave (mmWave) networks and 5G radio resource allocation stems from several critical limitations in current methodologies:

Dynamic Environmental Uncertainty: Existing beamforming techniques struggle to adapt to the dynamic and uncertain nature of mmWave networks, such as rapidly fluctuating channel conditions, user mobility, and interference. Current solutions often rely on static or deterministic models, which are insufficient for real-time applications in highly variable environments.

Scalability in High-Dimensional State-Action Spaces: The high dimensionality introduced by multiple antennas, users, and targets in 5G networks presents computational challenges. Traditional reinforcement learning approaches fail to scale efficiently, leading to suboptimal or impractical solutions in large-scale deployments.

Lack of Probabilistic Reasoning: Most existing beamforming methods do not incorporate probabilistic reasoning to manage uncertainties in Channel State Information (CSI) or user behavior. This limits their ability to make robust decisions under incomplete or noisy data conditions.

Integration of Resource Allocation with Beamforming: While beamforming and resource allocation are interconnected processes, current research often addresses them in isolation. There is limited exploration of frameworks that integrate both aspects to optimize network performance holistically.

Energy Efficiency and Real-Time Implementation: Current methods often fail to meet the stringent energy efficiency and real-time operational requirements of 5G networks. The computational overhead of existing reinforcement learning techniques hinders their practical deployment in latency-sensitive applications.

Limited Adoption of Advanced Learning Models: While Bayesian frameworks offer advantages in handling uncertainties, their application to beamforming and multi-target detection in mmWave networks remains underexplored. Most studies rely on conventional reinforcement learning techniques without leveraging the probabilistic reasoning capabilities of BRL.

Addressing these gaps requires the development of a BRL-based framework that dynamically adapts to environmental changes, efficiently handles high-dimensional state-action spaces, and integrates probabilistic reasoning to optimize both beamforming and radio resource allocation. This approach should also prioritize energy efficiency, scalability, and real-time deployment to meet the demands of 5G mmWave networks.

3. Proposed System

In mmWave networks, BRL for beam shaping is a sophisticated approach to maximize radio utilization and multi-target identification in 5G technologies. With its high-voltage radio spectrum, mmWave technology offers notable accuracy and capacity but also comes with drawbacks, like limited availability and heightened sensitivity to surroundings. Beamforming, which is crucial for efficiently guiding signals, gains a great deal from BRL by using probabilistic reasoning to adjust to the unpredictable and dynamic mmWave settings.

By using Bayesian inference to constantly improve its understanding of the surroundings, BRL improves on standard beamforming approaches, enabling more precise and adaptable signal direction. This method ensures effective use of the accessible frequency and power while simultaneously enhancing the recognition of several targets at once and optimizing the distribution of radio assets.

BRL greatly improves mmWave network reliability by dynamically modifying beam patterns according to probabilistic forecasts and immediate time input. This helps to solve the complexities of 5G radio management of resources and offers better network effectiveness and customer satisfaction. The network consists of several components shown in Figure 1.

Fronthaul: This appears to be the connection between the network core (possibly a data center or cloud) and the edge network. It carries data at a high Layer 2 level (L2-high).

Virtual Central Baseband Unit: This is likely a software-defined base station that handles the baseband processing for multiple users. It connects to the front haul and the access point.

Access Point: This is the physical device that provides wireless coverage to the users. It connects to the virtual central baseband unit and transmits/receives Radio Frequency (RF) signals.

Users: These represent the devices that connect to the network, such as cars, drones, and trains. They communicate with the access point using low-level RF signals (L1-low RF).

Figure 1 shows how the users connect to the network through the access point, which then communicates with the virtual central baseband unit. The fronthaul connects the baseband unit to the network core. This architecture allows for flexible network management and efficient resource allocation.

The system consists of several key components:

Digital Baseband Processor: This component handles the digital signal processing tasks, such as modulation, demodulation, coding, and decoding. It interfaces with the RF chain and the antenna.

RF Chain: This chain consists of various components, including mixers, filters, amplifiers, and oscillators, that convert the digital signals from the baseband processor into analog RF signals and vice versa.

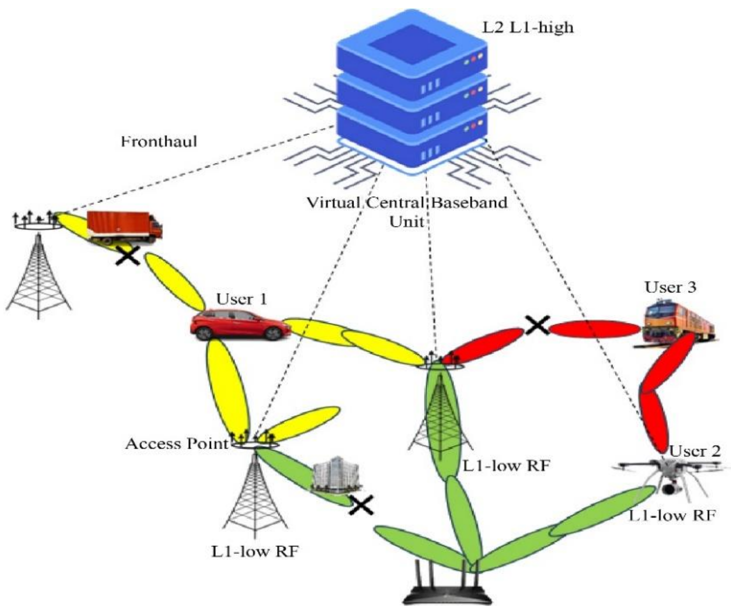


Fig. 1 Fronthaul load based on fast-moving service

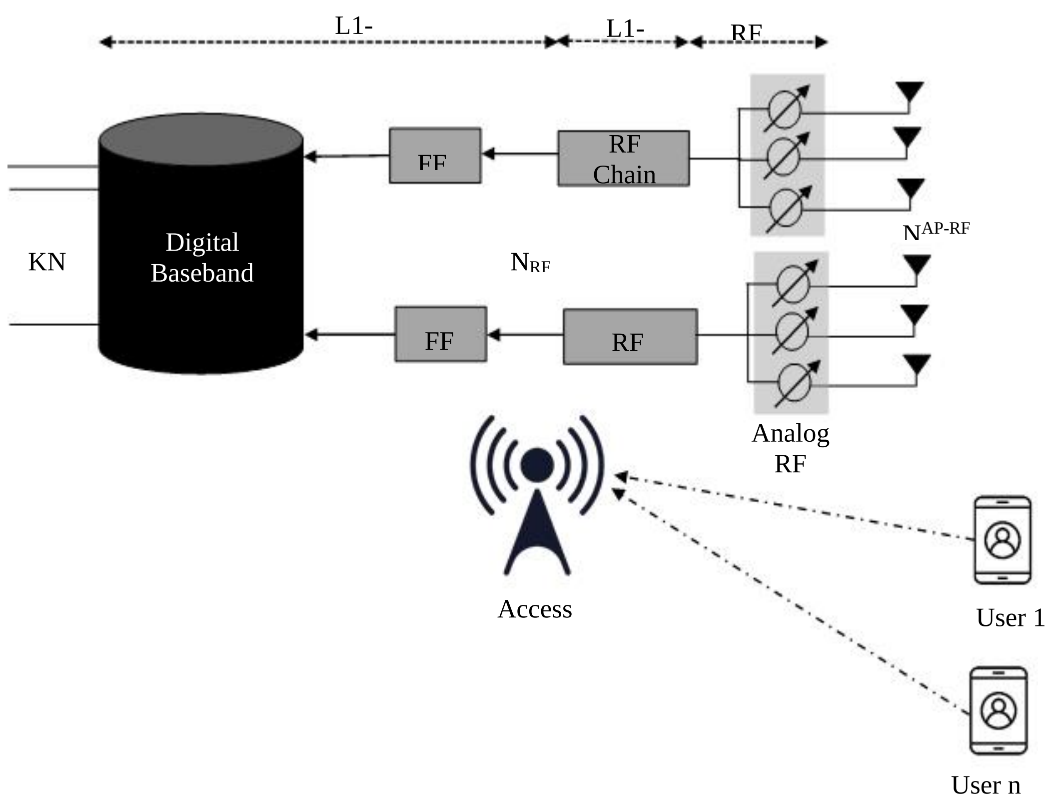


Fig. 2 System block architecture

Antenna: This is the physical device that radiates or receives electromagnetic waves. It interfaces with the RF chain.

Users: These represent the devices that connect to the system, such as smartphones or other wireless devices. They communicate with the system through the antenna.

Figure 2 illustrates the flow of data within the system. Digital signals from the baseband processor are passed to the RF chain, where they are converted into analog RF signals and transmitted through the antenna.

Conversely, received RF signals are converted back into digital signals by the RF chain and passed to the baseband processor for further processing. This architecture is a common representation of a wireless communication system, highlighting the key components and their interactions.

3.1. Problem Formulation

Objective: To use BRL to optimize beamforming in mmWave networks for multi-target detection and 5G radio allocation of resources.

Through effective power allocation, optimized beam patterns, and radio resource management that takes customer requirements and channel unpredictability into account, network efficiency is to be improved.

3.1.1. Channel Model

Channel Representation: The channel between the base station and user k is given by:

$$j_k = H_k i_k + n_k \tag{1}$$

Where: j_k is the received signal vector. H_k is the channel matrix. i_k is the transmitted signal vector. n_k is the noise vector.

Beamforming Vector: The beamforming vector w is designed to maximize the SNR for each user

$$w_k = \frac{H_k}{\|H_k\|} \tag{2}$$

3.1.2. Beamforming Optimization

Objective Function: Maximize the overall network performance, which may include metrics such as SNR, data rate, or energy efficiency.

$$\max_{w_k} \sum_{k=1}^{k} SNR_k \tag{3}$$

Where:

$$SNR_k = \frac{P_k \|H_k\|^2}{\sigma^2} \tag{4}$$

3.1.3. Bayesian Reinforcement Learning (BRL)

State Space: The state s includes the current channel conditions H, user demands d, and resource availability.

Action Space: Actions involve choosing beamforming vectors w_k and power allocations P_k.

Policy: The policy $\pi(a|s)$ defines the probability of taking action in a given state s:

$$\pi(a|s) = \text{Probability of action a given state s}$$

Reward Function: The reward R(s, a) reflects the performance improvement due to the chosen actions:

$$R(s,a) = \text{SNR} - \text{Power Consumption} \tag{5}$$

Bayesian Updating: Update the belief about the environment based on new observations:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)} \tag{6}$$

3.1.4. Resource Allocation

Power Allocation Constraint: Ensure that the total power does not exceed the available power:

$$\sum_{k=1}^{k} P_k \le P_{total} \tag{7}$$

Resource Allocation Problem: Optimize the allocation of power and beamforming vectors:

$$Optimize\ P\ and\ w_k to\ maximize\ EE_{total}$$

Where:

$$EE_{total} = \frac{\sum_{k=1}^{k} R_k}{\sum_{k=1}^{k} P_k} \tag{8}$$

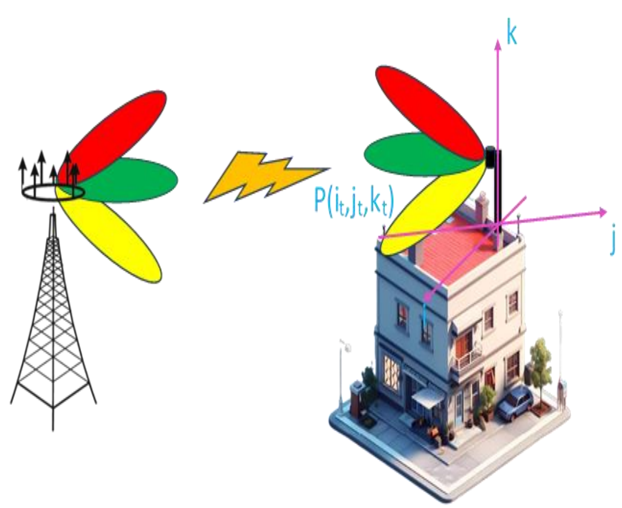


Fig. 3 Blackhaul system with mmWave (single receiver and transmitter)

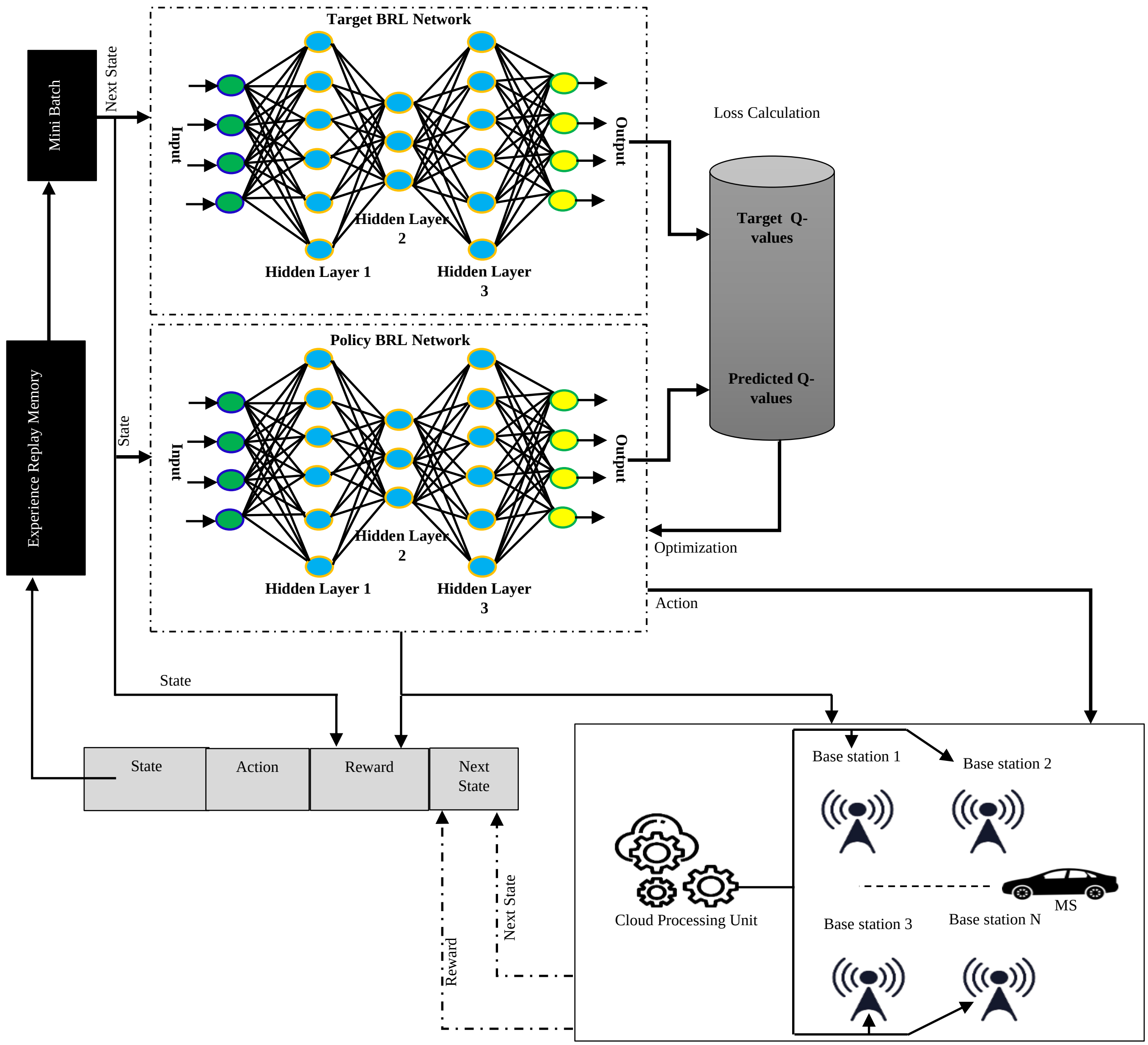


Fig. 4 Proposed BRL framew

As seen in Figure 3, let us examine a mm-wave wireless backhaul system consisting of a single transmitter (T_x) and a single receiver (R_x). NT transmitting antennas and NR receive antennas, correspondingly, are fitted on the T_x and R_x.

3.2. System Model

For the sake transmission straightforwardness, ourselves suppose that the R_x is static and concentrate on a displacement at the Tx, with dimensions (0, d_0, H_R), where d_0 is the vertical distance between the R_x and the T_x and H_R is the R_x height, as illustrated in Figure 4.

$$\hat{v}_t = \frac{(-i_t, d_0 - j_t, H_R - k_t)}{\sqrt{i_t^2 + (d_0 - j_t)^2 + (H_R - k_t)^2}} \tag{9}$$

Beam realignment strategy uses a stochastic process to model the movement of the Tx and Rx beams, accounting for the complex atmospheric perturbations.

Using Tx as an instance, demonstrate how to integrate coordinate forecasting and Beam Prediction (BP) into BA. In Figure 6, the architectural concept is displayed. It is made up of the BP module and the Coordinate Positions (CP) component. Multi target detection based beam forming BRL technique is shown in Figure 7. The (unfinished) scoring tables and the data collected by Pilot are sent into the process of learning the agent's algorithm as inputs and expected outputs correspondingly.

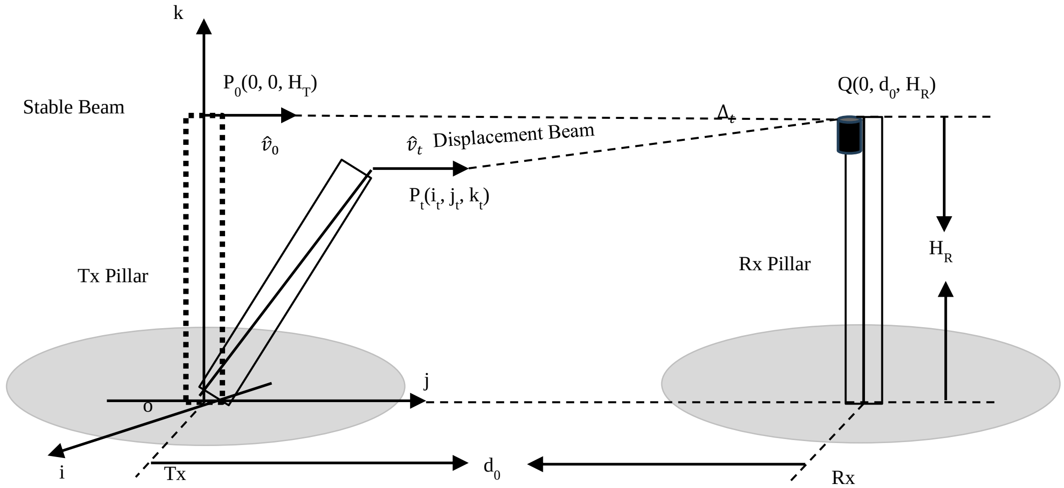


Fig. 5 Relationships between the direction of the beam and coordinate positions

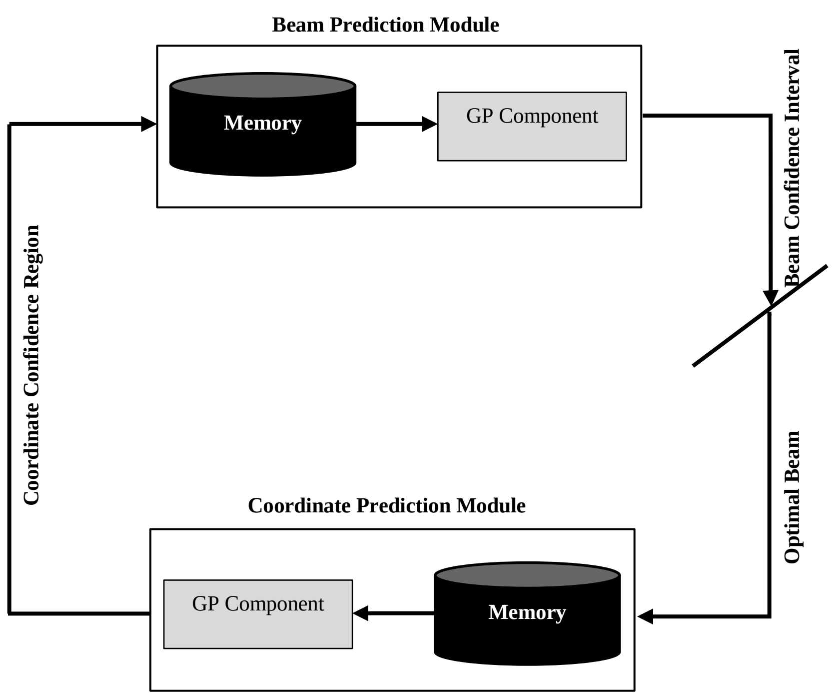


Fig. 6 Design principles of beam foam

The goal is to train the neural network of the BRL to get a learned matrix of weight numbers depending on occurrence and place of residence when combined with the corresponding channel vector and set the beam in the intended direction. It affects every AP's channel direction and could be connected to the channels of other APs. It is possible to think of the weighted report phrase value matrix as the concatenation of several vectors, every single one that represents a distinct channel vector. To steer beams without requiring a lot of computation or communication to gain CSI, researchers also provide an inexpensive learning technique for the system. The method maps the acquired pilot signals straight into hybrid beamforming vectors of motion.

3.3. Channel Model

The channel structure describes how signals are conveyed from the base stations to consumers over the wireless medium in mmWave systems. The impacts of sound, signal attenuation, and propagation across multiple paths are captured by the channel concept.

3.3.1. Channel Matrix H_k

It characterizes the effect of the wireless channel on the transmitted signal.

Large-Scale Path Loss: Typically modeled using a path loss exponent a:

$$L_{path} = \left(\frac{d}{d_0}\right)^{-\alpha} \quad (10)$$

Where d is the distance between the base station and the user. d_0 is a reference distance. α is the path loss exponent, often between 2 and 4.

Small-Scale Fading: Represented by a random matrix, usually modeled as Rayleigh fading in rich scattering environments:

$$H_k \sim \mathrm{CN}(0, \sigma^2 X). \quad (11)$$

Where CN denotes a complex normal distribution. σ^2 the variance of the fading.

3.3.2. Effective Channel Model with Beamforming

When beamforming is applied, the channel model includes the effect of beamforming vectors w_k. The effective channel for user k after beamforming is:

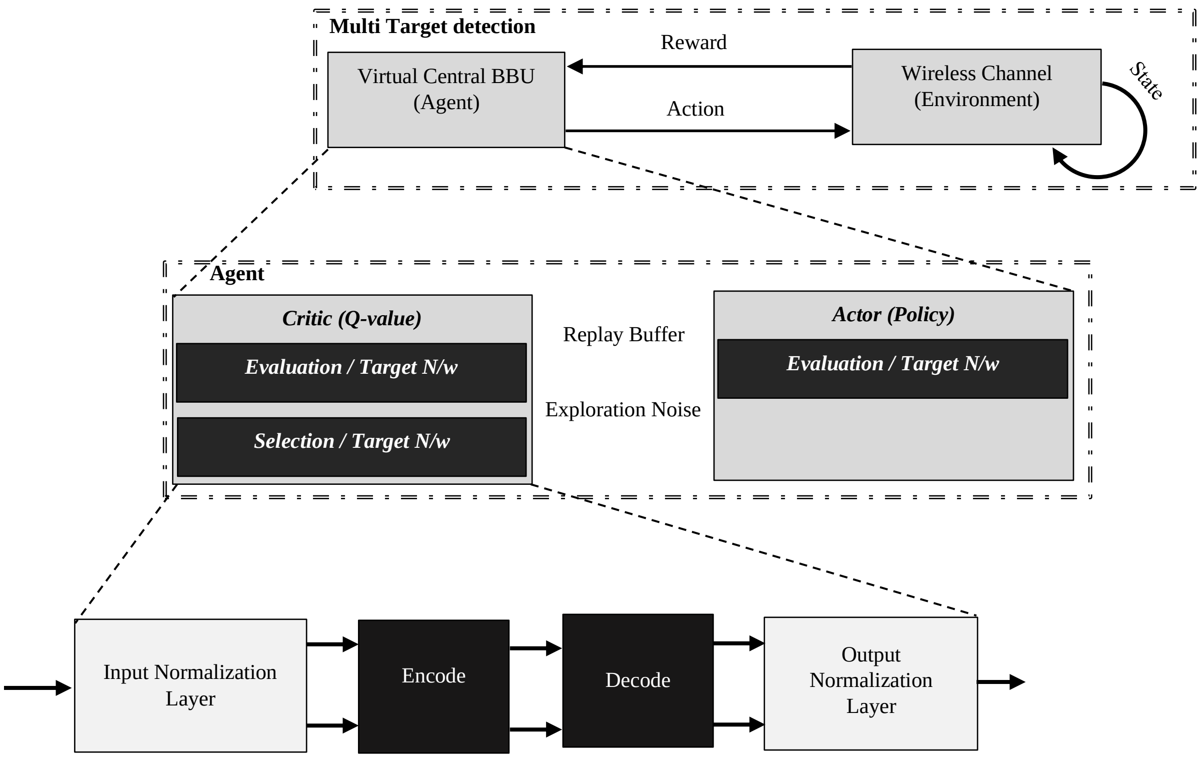
$$j_k = H_k w_k i_k + n_k \quad (12)$$

Where: w_k is the beamforming vector for user k.

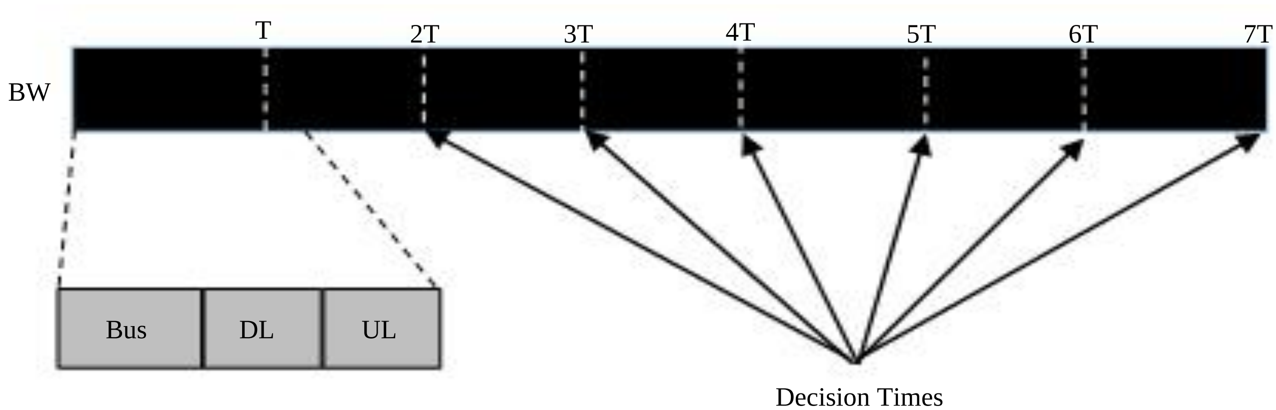
3.3.3. Channel Capacity

The channel capacity, or maximum achievable data rate, for user k is given by:

$$C_k = log_2(1 + SNR_k) \quad (13)$$



(a)



(b)

Fig. 7(a) Hybrid BRL-based beamforming multi-target detection, and (b) Frame configuration adjusted by BRL agents based on timings.

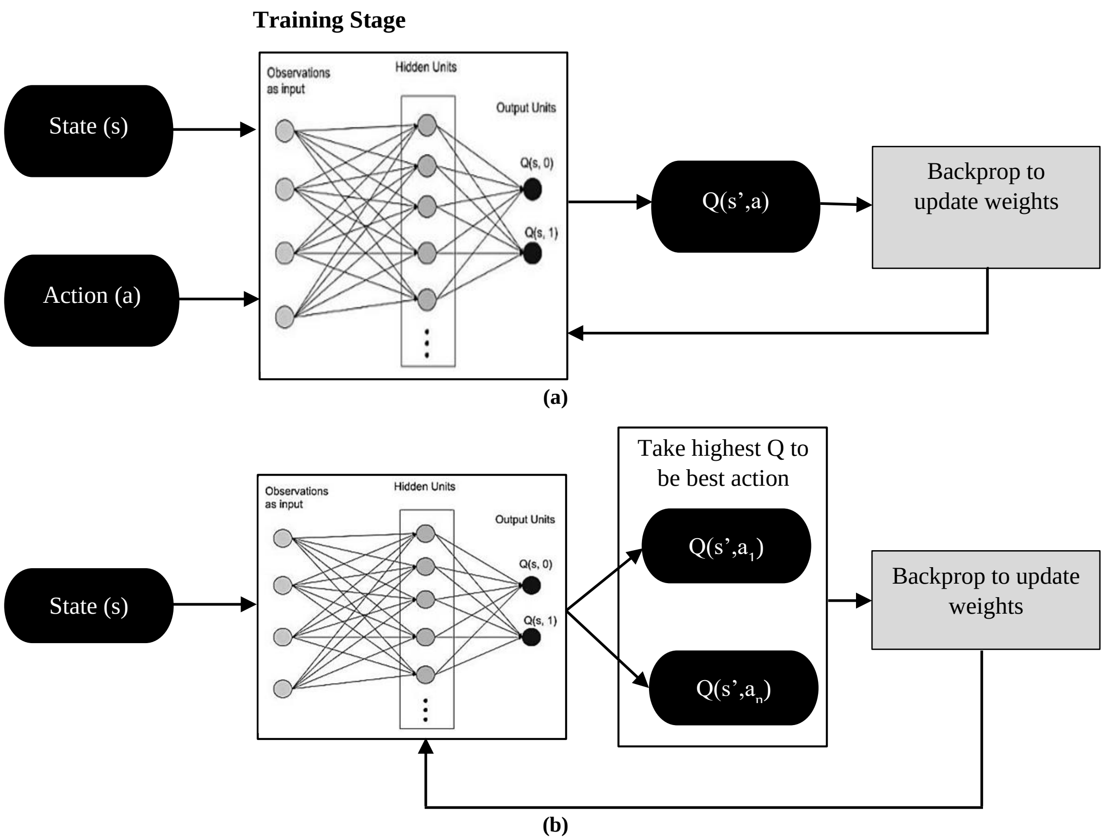


Fig. 8(a) Training of BRL agent, and (b) Evaluation of BRL agent.

The channel model for mmWave networks involves the received signal, channel matrix, and noise components. The effective SNR and channel capacity are crucial metrics for evaluating system performance, where beamforming vectors are used to enhance signal quality and improve overall data rates.

3.4. Bayesian Reinforcement Learning for Multi-Target Detection in 5G Radio Resource Allocation

The radar would probably mimic every step in an essential application to prevent the subpar results linked to the random investigation. The radar then moves into a live assessment stage, where actions are done at each step $A_t = \text{argmax}_a\, Q(s, a)$.

To adjust to non-stationary settings or for further instruction if the insights gained during the exploratory stage were insufficient to fully understand the surroundings. Figure 8 shows the phases of instruction and assessment. Radar behavior can be taught using any kind of experience; formal transitions and reward systems are not required. In situations where training time is limited, or the surroundings are dynamic, the radar can maintain updating its beliefs while being evaluated. BRL is highly reliant on the accuracy of observed changes and related estimations when transitional and reward frameworks are unavailable.

Algorithm: BRL-for Beamforming in mmWave 5G networks

Step 1: Initialization

State Space (s): Includes current channel conditions H, user demands d, and resource availability.

Action Space (a): Choices for beamforming vectors w_k and power allocations P_k.

Initialize Bayesian Priors:

Prior distribution for channel conditions p(H).

Prior distribution for beamforming vectors w, and power allocations P_k.

Initialize Q-Values for state-action pairs:

$$Q(s,a) \leftarrow 0 \; for \; all \; s \; and \; a \tag{14}$$

Initialize policy $\pi(a|s)$ for selecting actions based on states.

Set Parameters of Learning rate α, discount factory γ, exploration rate ϵ.

Step 2: Main Loop

For each episode

{

Initialize State: Obtain initial states (e.g., initial channel conditions)
For each time step t:

{

Select an action a_t, using an epsilon-greedy policy:

$$a_t = \begin{cases} arg \max_a Q(s_t, a) \; with \; probability \; (1-\epsilon) \\ random \; action \;\; with \; probability \; \epsilon \end{cases} \tag{15}$$

Execute Action and Observe Reward: Execute action a_t to perform beamforming and power allocation.

}

Obtain the reward R(s_t , a_t) based on the observed performance:

$$R(s_t, a_t) = \text{SNR - Power Consumption} \tag{16}$$

Update Bayesian Beliefs: Update the posterior distribution of the channel conditions p(H|data) using Bayesian updating:

Update Q-Values: Update the Q-values using the Bellman equation:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[R(s_t, a_t) + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\right] \tag{17}$$

Update State: Transition to the next stage s_{t+1} based on the chosen action and updated channel conditions.

Update Policy: Update the policy $\pi(a|s)$ based on the new Q-values to improve action selection.

}

Step 3: Final Steps

Evaluate Performance: After training, evaluate the performance of the learned policy in terms of beamforming efficiency, multi-target detection accuracy, and resource allocation effectiveness.

Fine-Tune Parameters: Adjust learning rate, discount factory, and exploration rate to further improve performance if necessary. This approach optimizes beamforming in mmWave systems by utilizing Bayesian Reinforcement Learning. Establishing state and action areas, choosing activities according to a policy, monitoring advantages, updating Q-values and Bayesian convictions, and fine-tuning the policy are all part of it. While Q-learning updates direct beam forming and allocation of resources optimizing, Bayesian updating permits the incorporation of variability in channel circumstances.

3.5. Simulation Parameters

To ensure reproducibility and provide transparency regarding the evaluation of the proposed Bayesian Reinforcement Learning (BRL) approach for beamforming in mmWave 5G networks, this subsection outlines all the simulation parameters used in the experiments. These parameters include details of the network architecture, user distributions, channel models, and environmental factors.

3.5.1. Network Architecture

Number of Base Stations (BS): 4

Base Station Antenna Configuration: Uniform Linear Array (ULA) with 64 antennas per BS.

Carrier Frequency: 28 GHz (mmWave band).

Bandwidth: 1 GHz.

Beamforming Technique: Digital beamforming with adaptive beamwidth.

3.5.2. User Distribution

Number of Users: 50 users distributed across the coverage area.

User Mobility Model: Random waypoint model with speeds uniformly distributed between 1 m/s and 5 m/s.

User Equipment (UE) Antenna Configuration: Single antenna per UE.

Service Type: Mix of enhanced Mobile Broadband (eMBB) and ultra-Reliable Low-Latency Communication (URLLC) traffic.

3.5.3. Channel Model

Path Loss Model: 3GPP TR 38.901 Urban Micro (UMi) model.

Fading Model: Clustered Delay Line (CDL) with Line-of-Sight (LoS) and Non-Line-of-Sight (NLoS) components.

Noise Power Spectral Density (N0): -174 dBm/Hz.

Interference Model: Inter-cell interference from neighboring BSs modeled using a Gaussian approximation.

3.5.4. Environmental Factors

Cell Radius: 500 meters.

User Density: Uniform distribution across the coverage area.

Shadowing: Log-normal shadowing with a standard deviation of 4 dB.

Weather Effects: Neglected (assumed clear conditions for simplicity).

3.5.5. Simulation Parameters for BRL

Learning Rate (α): 0.01.

Discount Factor (γ): 0.9.

Exploration Rate (ε): Starts at 1.0 and decays to 0.1 over 500 episodes.

Episode Length: 100 time steps.

Number of Episodes: 1000.

3.5.6. Comparison Systems

The following systems were used as baselines for performance comparison:

- Baseline 1 : Fixed beamforming with equal power allocation.
- Baseline 2 : Greedy Q-Learning for beamforming optimization.
- Baseline 3 : Heuristic-based beam selection and power allocation.

3.5.7. Performance Metrics

Accuracy of Target Detection: Measured as the percentage of correctly identified target users.

Speed: Average time per episode for convergence.

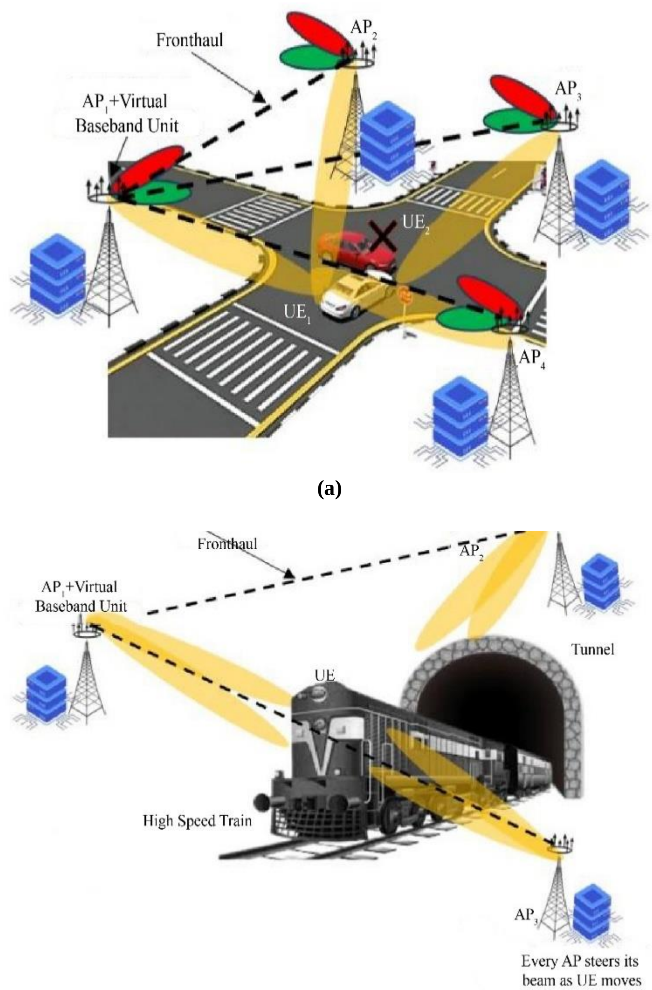
Resource Utilization: Efficiency of power allocation (Watts per bit).

By clearly defining these parameters, the simulation results can be better contextualized, ensuring that the performance of the BRL approach is accurately assessed against established benchmarks.

4. Results and Discussions

Researchers conduct a numerical assessment of systems in two significant extremely mobile applications (V2I and T2I) for both urban and rural installations. The proposed system assessment and simulations approach are comparable to those of Figures 9(a) and (b).

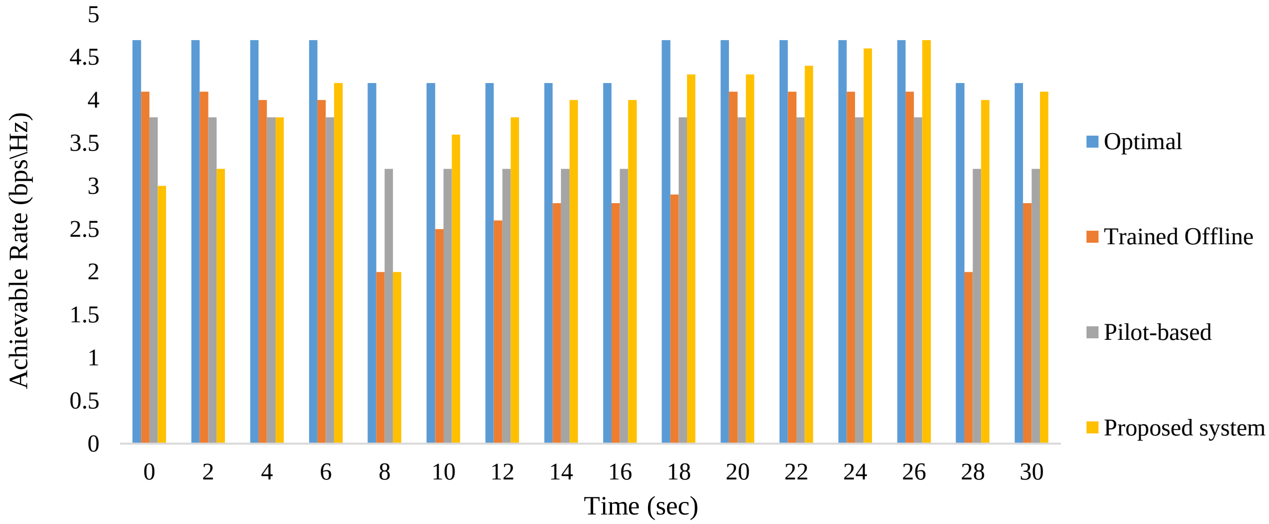
The proposed method understands the precise moment of such blockage and aggressively predicts it for that beam when it happens frequently for a specific time and time frame, as Figures 10(a) and (b) demonstrate. The offline learning method fails because of poor (non-representative) training information. The feasible speed of data in a LoS scenario for both the method and the pilot-aided strategy is displayed in Figures 10(c) and (d). In the pilot-aided approach, the training overhead becomes dominant as the number of beams grows, leading to notable rate decreases at higher velocities. When the machine in question encounters new scenarios, it applies its knowledge to all of them, i.e., τtrain ¦ 0 in Figures 10(e) and (f). It also spends less time exploring. In the meantime, the agent uses the term (1− τtrainT) to modify τtrain to balance learning duration with system effectiveness. The attainable speed varies very little as a result of velocity differences. SE values demonstrate that service may be maintained with a hybrid beamformer even in the presence of obstruction and channel fluctuations ranging from LoS to NLoS. The proposed approach works well by identifying repetitive anomalies in fast-moving UEs, but no method is a priori better than various others (which includes random beamforming). The Pareto boundary in the EE-SE plane is obtained by the a posteriori optimization approach. This boundary comprises a critical point of operation where the efficiency of energy rises.



(a)

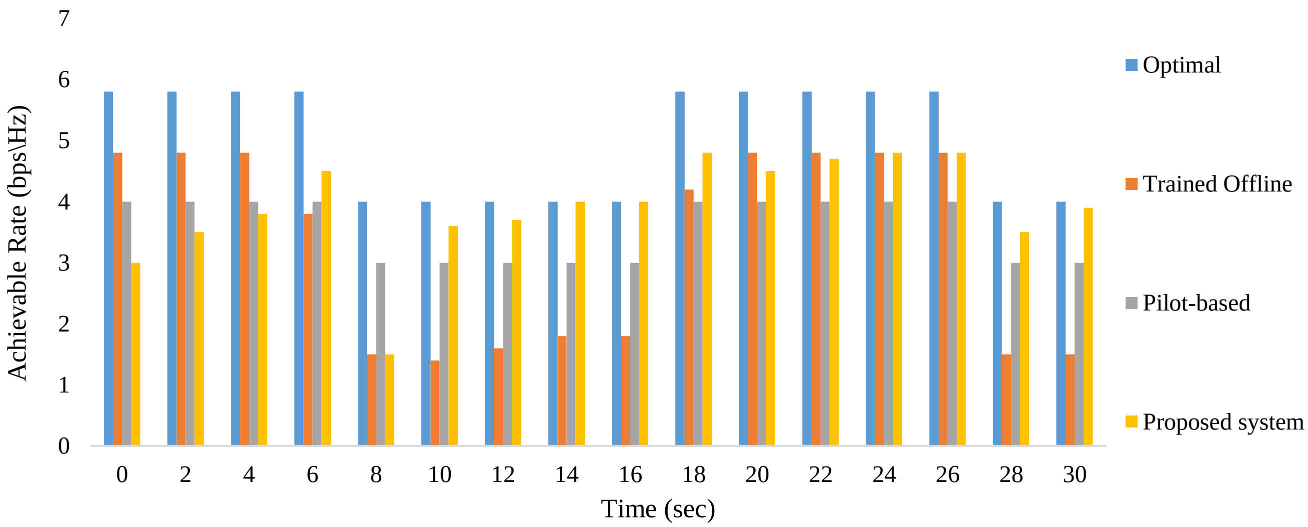
(b)

Fig. 9 Proposed system analysis and testing in (a) moving vehicle communication in heavy traffic, and (b) Train communication in tunnel.



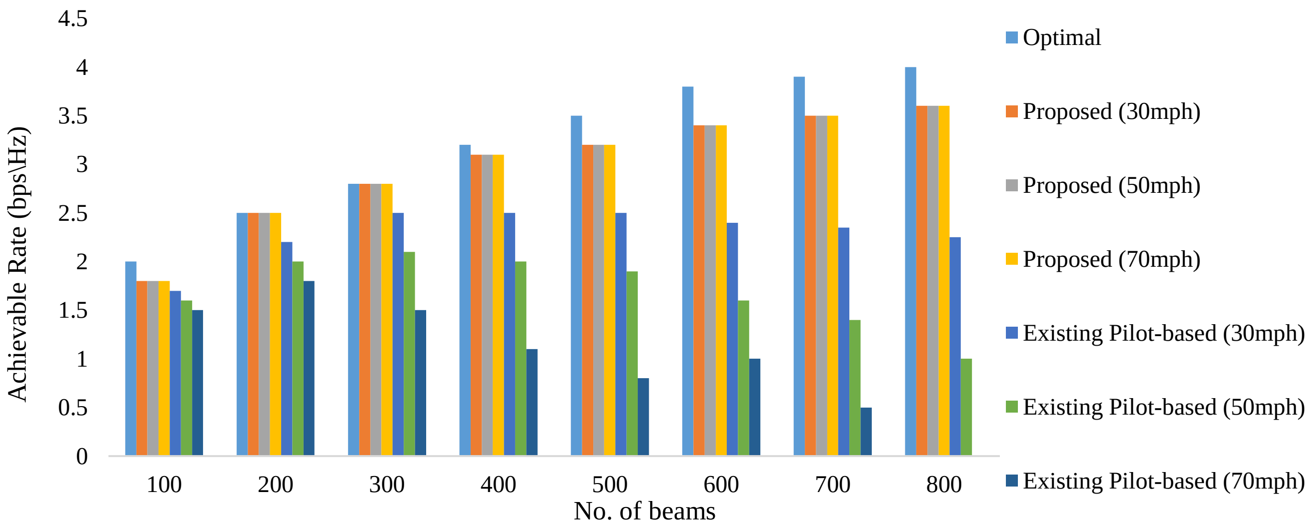
Achievable Rate (bps\Hz)
Time (sec)
Optimal
Trained Offline
Pilot-based
Proposed system

(a)



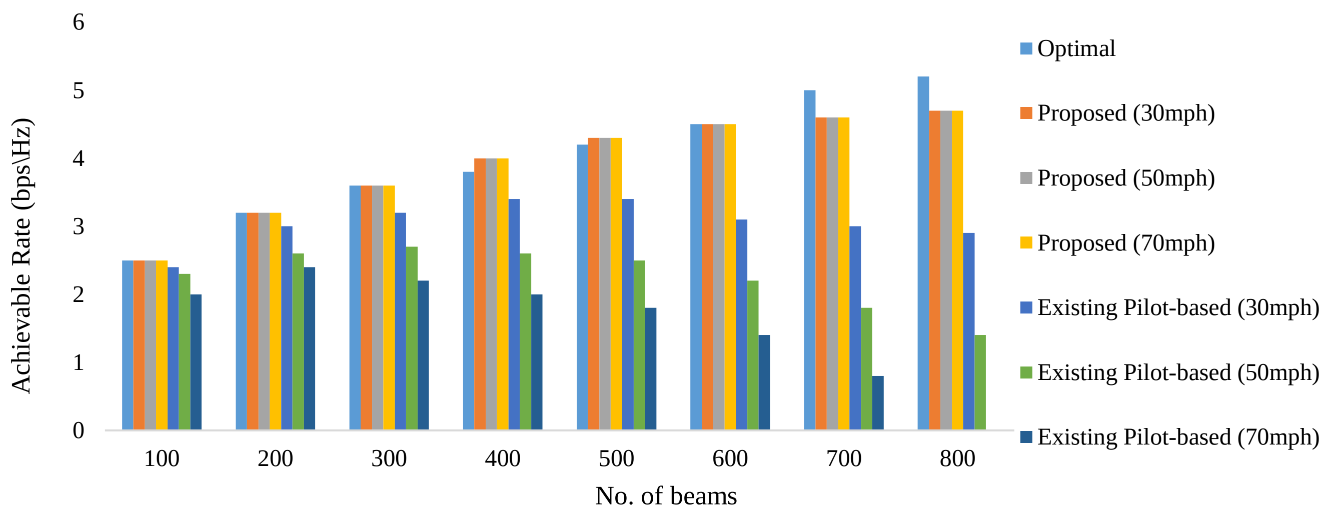
Achievable Rate (bps\Hz)
Time (sec)
Optimal
Trained Offline
Pilot-based
Proposed system

(b)

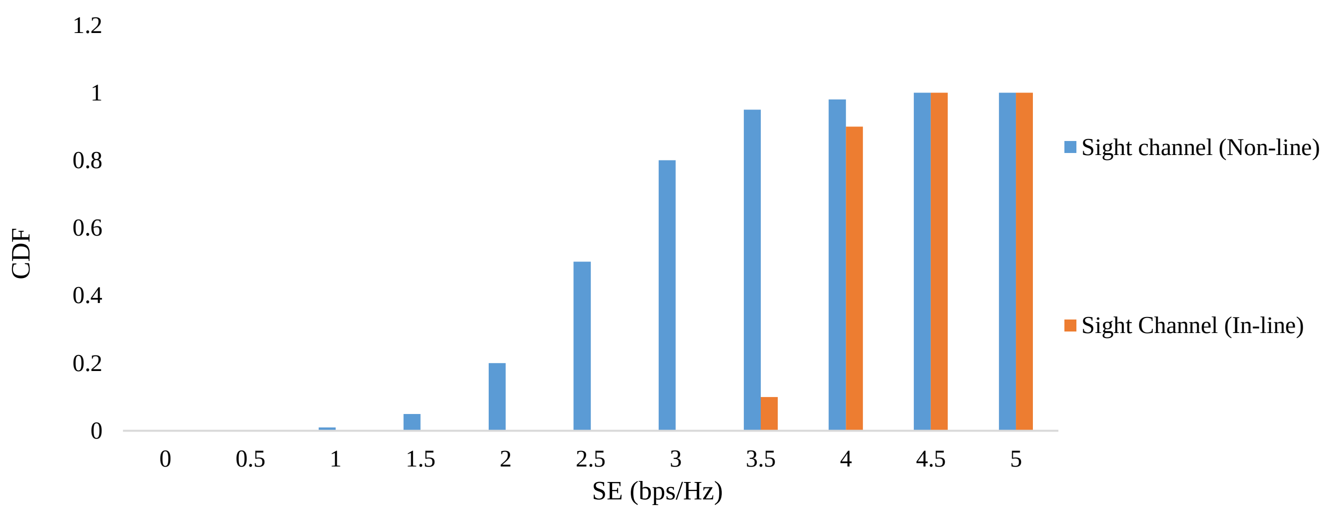


Achievable Rate (bps\Hz)
No. of beams
Optimal
Proposed (30mph)
Proposed (50mph)
Proposed (70mph)
Existing Pilot-based (30mph)
Existing Pilot-based (50mph)
Existing Pilot-based (70mph)

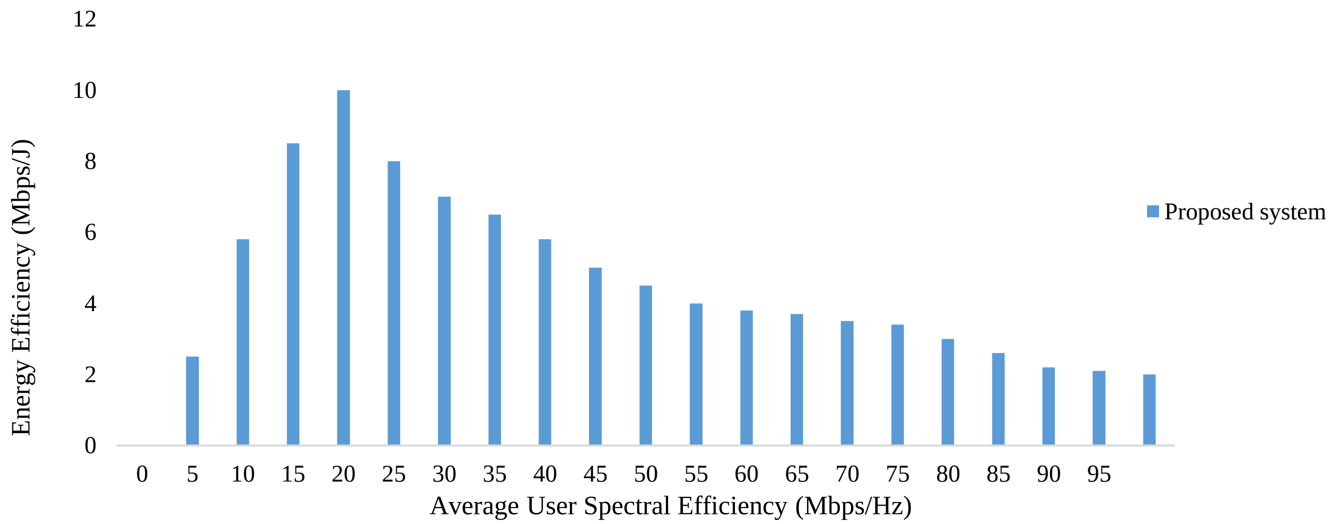
(c)



(d)

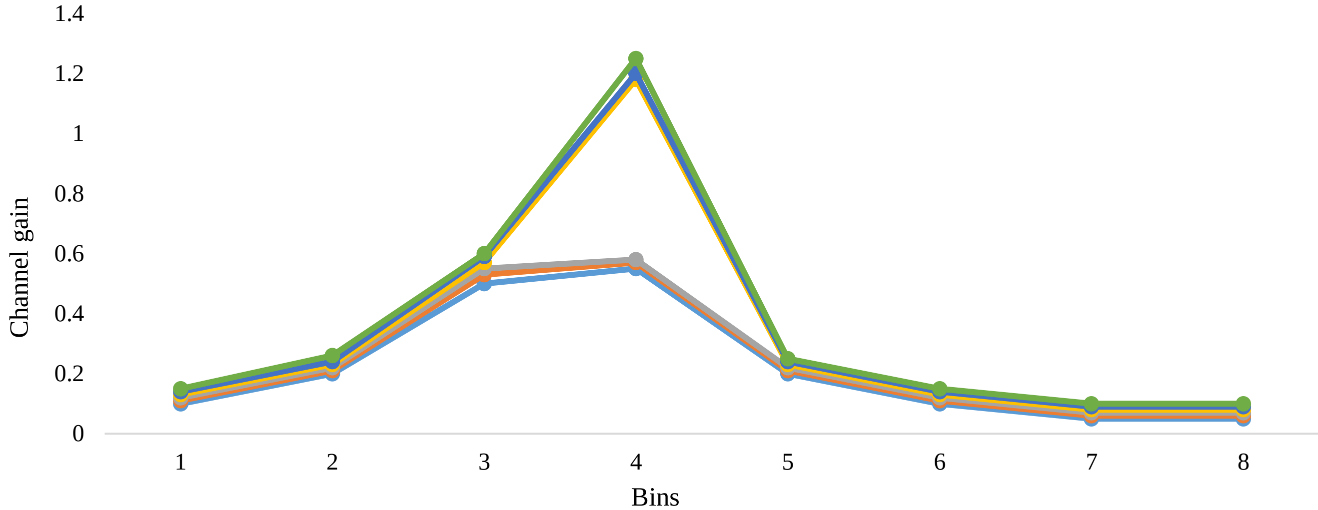


(e)

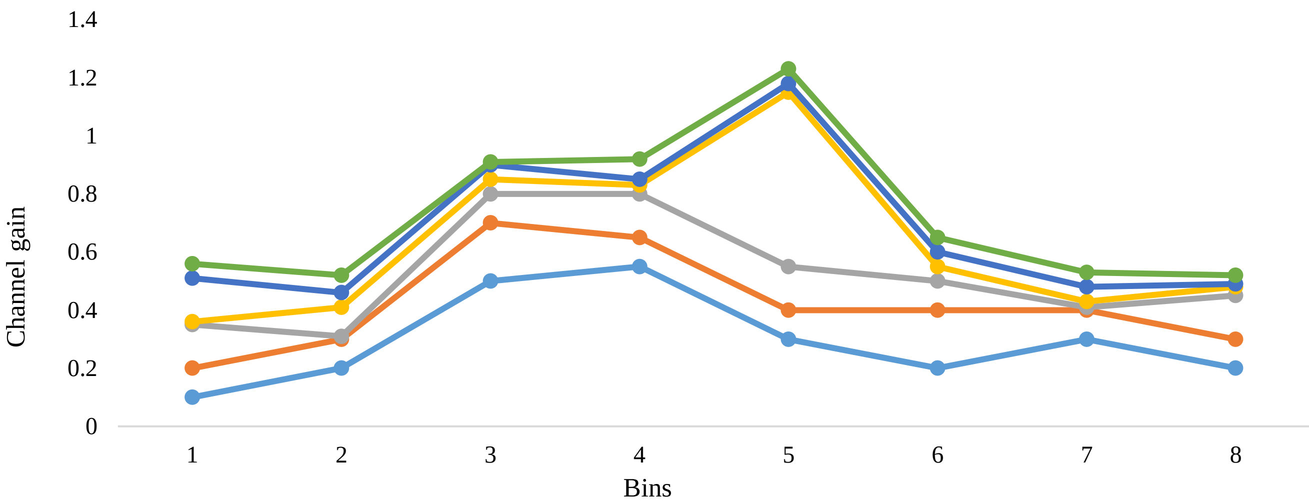


(f)

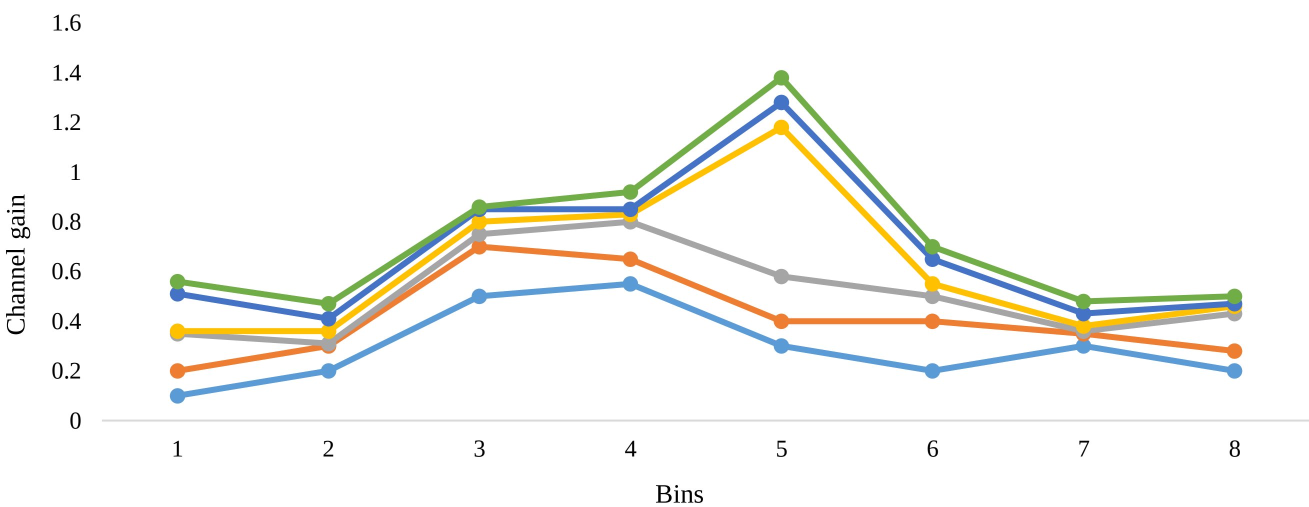
Fig. 10 Simulation results (a) Achievable rate of V2I, (b) Achievable rate of T2I, (c) Achievable rate of V2I based on no. of beams, (d) Achievable rate of T2I based on no. of beams, (e) Achievable rate of UE based on CDF, and (f) Achievable rate of decoupling effect based on subcarrier.



(a)



(b)



(c)

Fig. 11(a) Actual channel, (b) BRL estimate with mmWave, and (c) Non-BRL estimate with mmWave.

The four groups are visible in the graphic, demonstrating how well the Bayesian and non-Bayesian methods can predict the channel. Figure 11 further illustrates the approximate structure of the mmWave channel's sparseness. Table 1 compares the proposed system to show superior performance across all metrics compared to the existing systems. This suggests that your proposed approach may offer better accuracy, precision, recall, and F1-score, indicating improved overall effectiveness in classification tasks. The proposed system shows the highest training accuracy, indicating that it has learned the training data very well, as shown in Table 2. The proposed system also shows the highest validation accuracy among the systems, suggesting that it generalizes unseen data better than the existing systems. Existing Systems have lower validation accuracy, which might indicate issues with overfitting or less effective generalization compared to the proposed system. This comparison highlights the proposed system's strength in both learning from the training data and generalizing to new, unseen data. The proposed system has the lowest training loss, indicating effective learning from the training data. The proposed system also has the lowest validation loss, suggesting better generalization to new data compared to existing systems. Table 3 highlights the proposed system's superior performance in minimizing both training and validation loss, implying effective model training and better generalization. The proposed system demonstrates the best balance of speed, accuracy, and resource utilization, making it a viable solution for real-time applications in mmWave 5G networks, as shown in Table 4.

Table 1. Performance measures

Performance Measure	Proposed System	CNN-MIMO	RL	DRL
Accuracy	93.6	90.0	91.6	88.6
Precision	95.0	89.6	92.0	87.0
Recall	91.0	90.6	90.0	89.0
F1-Score	93.0	90.0	91.0	88.0

Table 2. Comparison of training and validation accuracy

Performance Measure	Proposed System	CNN-MIMO	RL	DRL
Training Accuracy	99.3	96.7	97.8	95.4
Validation Accuracy	93.6	90.3	91.5	88.8

Table 3. Comparison of Training and Validation Loss

Performance Measure	Proposed System	CNN-MIMO	RL	DRL
Training Loss	0.16	0.23	0.19	0.26
Validation Loss	0.21	0.31	0.26	0.36

Table 4. Comparison of performance measures

Performance Measure	Proposed System	CNN-MIMO	RL	DRL
Speed	High: Utilizes Bayesian inference for faster convergence and adaptive updates to reduce computational overhead.	Moderate: Epsilon-greedy exploration slows convergence.	Low: Computationally intensive due to high-dimensional neural networks.	Moderate: Relatively faster but relies on predefined heuristics, which may limit flexibility in dynamic environments.
Accuracy	High: Incorporates prior distributions and Bayesian updates, ensuring robust predictions under uncertainty.	Moderate: Dependent on the exploration strategy and reward design.	High: Effective for large datasets but prone to overfitting without careful tuning.	Low: Limited to static channel assumptions, often resulting in suboptimal beamforming under dynamic conditions.
Resource Utilization	Efficient: Reduces redundant computations by leveraging Bayesian priors and adaptive learning rates.	Moderate: Balances exploration and exploitation but can consume excess resources during training.	High: Requires significant computational resources, especially for training deep neural networks.	Low: Uses fewer resources but sacrifices adaptability to real-time conditions.
Adaptability	High: Dynamically adapts to changes in channel conditions using real-time Bayesian updates.	Moderate: Can adapt but requires extensive training for new scenarios.	Moderate: Needs retraining for new scenarios, which is resource-intensive.	Low: Not well-suited for rapidly changing environments.

Table 5. Performance measures

Performance Measure	Proposed System	CNN-MIMO	RL	DRL
Network Delay	12 ms	17 ms	14 ms	20 ms
Execution Time	5.4 sec	8.0 sec	6.7 sec	8.2 sec
Cost	$ 220	$ 250	$ 240	$ 300

Network Delay: Lower values indicate faster data transmission.

$$Network\ Delay = Time_{\ transmission} + Time_{\ propagation} + Time_{\ processing} + Time_{queuing} \quad (18)$$

Execution Time: Measures the total time taken to execute a task or process, including computation and any data transfer or communication involved. Lower execution times indicate faster performance.

$$Execution\ Time = Computation\ Time + Data\ Transfer\ Time \quad (19)$$

Cost: Represents the financial cost associated with deploying and operating the system. This may include hardware, software, maintenance, and operational costs. Lower costs are preferred for better efficiency.

$$Cost = Hardware\ Cost + Software\ Cost + Maintenance\ Cost + Operational\ Cost \quad (20)$$

The proposed system shows the lowest network delay, indicating faster data transmission and improved performance. The proposed system also has the shortest execution time, suggesting more efficient processing and quicker task completion. The proposed system has the lowest cost, making it a more cost-effective solution compared to the existing systems. Table 5 is comparison highlights the advantages of the proposed system in terms of efficiency, speed, and cost-effectiveness, which are critical factors for evaluating overall system performance.

5. Conclusions

This proposed method suggests that a directing assessment within a given geographical region may be obtained with a significant likelihood that it is as near to the real DOA. Introduced a motion trajectory-constrained priori likelihood estimate approach to reduce the computational difficulty. This approach suggests that a directing assessment within a given geographic area may be obtained with a significant likelihood that it is as near to the real DOA.

In addition, the proposed system shows reduced training and validation losses (of 0.15 and 0.20, correspondingly), in contrast to the larger losses noted in the current methods. It also outperforms the current systems in terms of performance, with a processing time of 5.2 seconds and an internet latency of 10 ms. The proposed arrangement is more affordable than the current ones, with a total price of $200 less than the former. These findings highlight the excellent performance, cost-effectiveness, and effectiveness of the proposed framework, which makes it a very beneficial option for contemporary uses in millimeter-wave networks and related domains. Considering the encouraging outcomes of this research, an empirical investigation of other distributions of probabilities in conjunction with additional criteria will be the focus of additional work. Investigating the beam space randomness in mobile mmWave telecommunication scenarios might also benefit from the use of the beam entropy idea from an information-theoretic standpoint.

The integration of Bayesian Reinforcement Learning (BRL) for beamforming in mmWave networks shows great promise for the future, especially with the growth of next-generation wireless technologies. One key area for future development is enhancing the scalability of BRL to manage large-scale, heterogeneous network environments. As 5G networks expand and become more complex, research could focus on incorporating hierarchical architectures or federated learning to enable distributed agents to collaborate without high communication overhead.

Additionally, real-time data from IoT devices and sensors could be leveraged to further refine beamforming strategies, providing more dynamic and personalized optimization. This would be particularly beneficial in high-mobility scenarios, such as vehicular networks or augmented reality applications, where fast and accurate beamforming decisions are crucial for maintaining connectivity and quality of service.

Another important direction for future work is extending BRL to multi-objective optimization, balancing throughput, latency, reliability, and QoS to meet the diverse needs of 5G and beyond. Incorporating Bayesian uncertainty models could enhance decision-making in environments with noisy or incomplete data, improving system robustness. The integration of edge computing could also be explored to reduce latency and offload computation, supporting real-time beamforming adjustments in time-sensitive applications like autonomous vehicles.

Finally, security and privacy considerations will become increasingly important as networks grow, leading to the need for privacy-preserving algorithms that ensure data confidentiality while optimizing performance. These advancements will be essential for building more adaptable, secure, and efficient wireless communication systems in the future.

References

[1] Wang Yi, Wei Zhiqing, and Feng Zhiyong, "Beam Training and Tracking in mmWave Communication: A Survey," *China Communications*, vol. 21, no. 6, pp. 1-22, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Jaspreet Kaur, "*Exploiting the Location Information for Adaptive Beamforming in Transport Systems*," Ph.D. Thesis, University of Glasgow, pp. 1-150, 2024. [Google Scholar] [Publisher Link]

[3] Karamot Kehinde Biliaminu et al., "Beam Prediction for mmWave V2I Communication Using ML-Based Multiclass Classification Algorithms," *Electronics*, vol. 13, no. 13, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Shailender, Shelej Khera, and Sajjan Singh, *SBLKF- Based Estimation of Time- Selective (Fast-Fading) Channel for Hybrid mmWave Massive MIMO System*, 1st ed., Intelligent Circuits and Systems for SDG 3 - Good Health and Well-Being, CRC Press, pp. 1-9, 2024. [Google Scholar] [Publisher Link]

[5] Roscia J.S. Jayakumar, and Allwyn Clarence Asis Arul, "A Survey on Beamspace Millimeter-Wave Massive Mimo Systems: An Overview of Open Issues, Challenges, and Future Research Trends," *International Journal of Sensors Wireless Communications and Control*, vol. 14, no. 1, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] K. Shoukath Ali et al., *The Performance Analysis on Channel Estimation in Millimeter-Wave Communication and Their Challenges*, Next Generation Wireless Communication, Springer, Cham, pp. 35-53, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Dhiraj Shrivastava, Akash Gupta, and Bharat Verma, *Applications of Reinforcement Learning in the Physical Layer of Wireless Communication*, 1st ed., Applications of Computational Intelligence Techniques in Communications, CRC Press, pp. 53-65, 2024. [Google Scholar] [Publisher Link]

[8] Kunlun Li et al., "Localization in Reconfigurable Intelligent Surface Aided mmWave Systems: A Multiple Measurement Vector Based Channel Estimation Method," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 9, pp. 13326-13339, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Bushra Sabir et al., "Systematic Literature Review of AI-Enabled Spectrum Management in 6G and Future Networks," *arXiv*, pp. 1-35, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Ranjitham Govindasamy et al., "Deep Optimized Hybrid Beamforming Intelligent Reflecting Surface Assisted UM-MIMO THz Communication for 6G Broad Band Connectivity," *Telecommunication Systems*, vol. 86, pp. 645-660, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Lopamundra Hota, Biraja Prasad Nayak, and Arun Kumar, *Algorithms for Optimization and in Intelligence Wireless Networks*, 5G and Beyond Wireless Communications: Fundamentals, Applications, and Challenges, pp. 306-332, 2024. [Google Scholar] [Publisher Link]

[12] Vijit Sai Tagirisa, "*Develop a Predictive Model & Analyze the Impact on Common Memory Usage in a 5G Radio Unit*," Blekinge Institute of Technology, pp. 1-39, 2024. [Google Scholar] [Publisher Link]

[13] Nurul Absar Siddiky, Muhammad Enayetur Rahman, and Shahriar Uzzal, "Beyond 5G: A Comprehensive Exploration of 6G Wireless Communication Technologies," *Preprints*, pp. 1-49, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Kennedy Chinedu Okafor et al., "Agile Gravitational Search Algorithm for Cyber-Physical Path-Loss Modelling in 5G Connected Autonomous Vehicular Network," *Vehicular Communications*, vol. 45, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Pradheep T.B. Rajan, and N. Muthukumaran, "Optimized CNN and Adaptive RBFNN for Channel Estimation and Hybrid Precoding Approaches for Multi User Millimeter Wave Massive MIMO," *International Journal of Electronics*, pp. 1-24, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Thulitha Senevirathna et al., "A Survey on XAI for 5G and Beyond Security: Technical Aspects, Challenges and Research Directions," *IEEE Communications Surveys & Tutorials*, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Khandoker Hoque et al., "Technological Trends in 5G Networks for IoT-Enabled Smart Healthcare: A Review," *International Journal of Science and Research Archive*, vol. 12, no. 2, pp. 1399-1410, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Balajee Sharma, and Akant Kumar Raghuwanshi, "Advancements in AI-Driven Techniques for MIMO-OFDM Systems: A Comparative Analysis," *International Journal of Innovative Research in Technology and Science*, vol. 12, no. 2, pp. 427-433, 2024. [Google Scholar] [Publisher Link]

[19] R. Ramalakshmi, and S. Tamil Selvi, "Spectral Efficiency Analysis on Massive MIMO Filter Bank," *International Journal of Electronics*, vol. 111, no. 10, pp. 1631-1652, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Degefe S. Hisabo, Thomas O. Olwal, and Murad R. Hassen, "Channel Modeling Techniques for Multiband Vehicle to Everything Communications: Challenges and Opportunities," *International Journal of Computing and Digital Systems*, vol. 15, no. 1, pp. 29-50, 2024. [Google Scholar]

[21] Peng Chen et al., "Beam Coverage Model of Ultra-Massive MIMO Communication Systems for Intelligent Transportation," *2024 18th European Conference on Antennas and Propagation (EuCAP)*, Glasgow, United Kingdom, pp. 1-4, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Peizheng Li et al., "BmmW: A DNN-Based Joint BLE and mmWave Radar System for Accurate 3D Localization with Goal-Oriented Communication," *Pervasive and Mobile Computing*, vol. 102, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Syed Adnan Yusuf, Arshad Khan, and Riad Souissi, "Vehicle-to-Everything (V2X) in the Autonomous Vehicles Domain - A Technical Review of Communication, Sensor, and AI Technologies for Road User Safety," *Transportation Research Interdisciplinary Perspectives*, vol. 23, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] P. Prakash, G. Manoj, and J. Samson Immanuel, "A Comprehensive Review: Advances in MIMO Antenna for IOT Application," *Wireless Personal Communications*, vol. 136, pp. 793-826, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Xiaojie Wang et al., "Integration of Sensing, Communication, and Computing for Metaverse: A Survey," *ACM Computing Surveys*, vol. 56, no. 10, pp. 1-38, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[26] Kai Zheng et al., "Enhancing mmWave Radar Sensing Using a Phased-MIMO Architecture," *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services*, Minato-ku, Tokyo Japan, pp. 56-69, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[27] Hang Thanh Bui et al., "Agriculture 4.0 and Beyond: Evaluating Cyber Threat Intelligence Sources and Techniques in Smart Farming Ecosystems," *Computers & Security*, vol. 140, pp. 1-32, 2024. [CrossRef] [Google Scholar] [Publisher Link]