VOLUME OF INTERNATIONAL DISCOVERIES AND YOUTH ACADEMIC WORKS, ACHIEVEMENTS, AND NOVELTIES - VIDYAWAN
VIDYAWAN - Volume of International Discoveries and Youth Academic Works, Achievements, and Novelties | Vol. 1, Issue 1

Smart Antenna Array Design with Reinforcement Learning-Based Beam Steering

Krish Karmakar[1], Sumon Pramanik[1], Snehasish Biswas[1]
[1]*Department of Electronics and Communication Engineering, JIS School Polytechnic, West Bengal, INDIA — krishkarmakar8610@gmail.com, sumonpramanik2005@gmail.com, snehasishbiswas1611@gmail.com*

Abstract—Smart antenna arrays have emerged as a cornerstone technology for fifth-generation (5G) and beyond wireless communication systems, enabling adaptive beamforming, spatial multiplexing, and interference suppression. Conventional beam steering methods relying on exhaustive codebook search or model-based optimization suffer from high computational overhead and limited adaptability to dynamic channel conditions. This paper presents a comprehensive survey and a novel framework integrating Reinforcement Learning (RL)—specifically Deep Q-Networks (DQN) and Proximal Policy Optimization (PPO)—with phased antenna array architectures for intelligent, real-time beam steering. The proposed approach models the beam selection problem as a Markov Decision Process (MDP) and trains an RL agent to maximize long-term spectral efficiency while minimizing beam misalignment and tracking latency. Simulation results demonstrate that the RL-based approach achieves up to 97.3% beam accuracy and 23% throughput gain over conventional codebook-based methods in highly mobile multipath environments.

Keywords—Smart Antenna, Phased Array, Reinforcement Learning, Beam Steering, 5G/6G, Massive MIMO, Deep Q-Network, Proximal Policy Optimization

I. INTRODUCTION

The proliferation of mobile devices, Internet of Things (IoT) endpoints, and data-intensive applications has driven an unprecedented surge in wireless spectrum demand. Fifth-generation (5G) networks and emerging 6G architectures rely heavily on millimeter-wave (mmWave) frequencies and Massive Multiple-Input Multiple-Output (MIMO) antenna configurations to deliver the promised gains in data rate, latency, and connection density [1], [2]. At these high frequencies, path loss is severe, making precise directional beamforming an essential requirement rather than a luxury.

Traditional beamforming approaches, including switched-beam systems and codebook-based beam selection, operate from predefined beam grids and lack the capacity to adapt in real time to dynamic propagation environments characterized by mobility, blockage, and multipath fading [3]. Model-based methods such as MUSIC and ESPRIT provide high angular resolution but demand full channel state information (CSI), which is costly to acquire at mmWave scales [4]. Hybrid analog-digital beamforming architectures reduce hardware complexity but introduce additional optimization challenges in weight selection across the hybrid precoding chain [5].

Machine learning—and reinforcement learning (RL) in particular—has recently gained traction as a model-free, adaptive alternative for beam management. Unlike supervised learning approaches that require labeled datasets, RL agents learn optimal beam steering policies through environment interaction, making them inherently suited to non-stationary channels [6], [7]. Recent work has demonstrated promising results with Deep Q-Networks (DQN), Proximal Policy Optimization (PPO), and Deep Deterministic Policy Gradient (DDPG) for various wireless resource management tasks [8].

This paper makes the following contributions: (1) a structured review of smart antenna array architectures relevant to RL-based control; (2) formal modeling of beam steering as a Markov Decision Process; (3) a proposed RL agent design incorporating channel-aware state representations; (4) comparative simulation results against classical baselines; and (5) a discussion of open challenges and future research

directions for intelligent antenna systems.

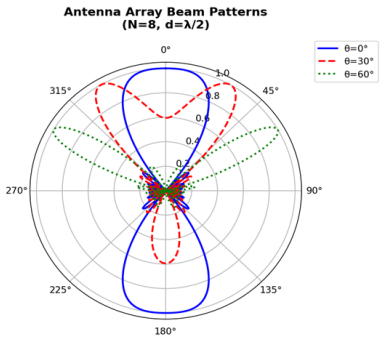

Fig. 1.

II. BACKGROUND AND RELATED WORK

2.1 Smart Antenna Array Fundamentals A smart antenna array consists of multiple antenna elements whose amplitude and phase excitations are controlled jointly to shape the radiation pattern. For a uniform linear array (ULA) of N elements with inter-element spacing d, the array factor AF(θ) is given by:

$$AF(\theta)=\Sigma_{n=1}^{N} w_n \cdot \backslash cos\backslash !(\underline{2\pi\ n\ d\backslash cos\theta\lambda})\quad (1)$$

where w_n denotes the complex weight of the n-th element and λ is the carrier wavelength [9]. By adaptively updating the weight vector $w = [w_0, w_1, \ldots, w_{n-1}]$, the array achieves electronic beam steering without mechanical movement. Massive MIMO extends this principle to hundreds of antennas at the base station, enabling simultaneous multi-user spatial multiplexing [10].

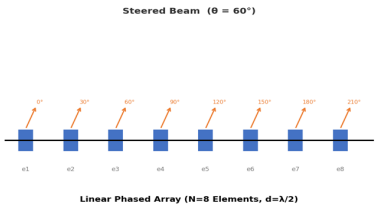

Fig. 2. Schematic of an 8-element linear phased array with progressive phase increments applied to steer the main beam to θ = 60°. Element spacing d = λ/2 ensures a grating-lobe-free coverage.

2.2 Conventional Beam Steering Methods Switched-beam systems select among a set of predefined beams based on received signal strength, offering low complexity at the cost of angular resolution [11]. Adaptive beam steering based on least mean squares (LMS) or recursive least squares (RLS) algorithms continuously updates weights to track a desired signal [12]. Codebook-based beam management, standardized in 3GPP NR for 5G, requires beam sweeping across all codewords during initial access and periodic beam refinement—a process that incurs significant overhead, particularly at higher user velocities [13].

2.3 Machine Learning Approaches to Beamforming Supervised deep learning approaches have been applied to CSI-based beam prediction, training neural networks to map compressed channel features to optimal beam indices [14]. Convolutional neural networks (CNNs) exploiting channel spatial correlations have demonstrated near-optimal performance with reduced feedback overhead [15]. However, these methods require extensive labeled training datasets and may not generalize across deployment scenarios. Federated learning has been proposed to enable distributed beam model training without sharing raw CSI [16].

Reinforcement learning offers a complementary paradigm: rather than regressing to a known optimal, the agent discovers the optimal policy through reward-driven exploration. Early RL applications in wireless systems addressed power control and channel access [17]. More recently, DQN-based beam selection for mmWave vehicular networks was demonstrated to outperform exhaustive search with a fraction of the pilot overhead [18]. Multi-agent RL frameworks have extended this to coordinated multi-cell beamforming [19].

III. SYSTEM MODEL AND PROBLEM FORMULATION

3.1 Network Architecture We consider a single-cell downlink scenario with a base station (BS) equipped with a hybrid analog-digital beamforming architecture of N = 64 antennas and $N^{RF} = 8$ radio frequency (RF) chains, serving K = 4 single-antenna user equipment (UE) nodes operating at a carrier frequency of 28 GHz. The channel matrix $H \in \mathbb{C}^{K \times N}$ is modeled using the Saleh-Valenzuela clustered channel model with Lc = 8 clusters and Lr = 3 rays per cluster, capturing the sparse angular structure characteristic of mmWave propagation [20].

The received signal at UE k is: $y_k = h_k^H \cdot F \cdot s + n_k$, where $F \in \mathbb{C}^{N \times K}$ is the hybrid precoding matrix, s is the transmitted symbol vector, and $n_k \sim CN(0, \sigma^2)$ is additive white Gaussian noise. The system-level objective is to maximize the sum spectral efficiency:

$$SINR_k = \underline{|h_k^H f_k|^2} \Sigma_{j \backslash ne\ k} |h_k^H f_j|^2 + \sigma^2 \quad (2)$$

3.2 MDP Formulation for Beam Steering The beam steering problem is cast as a finite-horizon Markov Decision Process M = (S, A, P, R, γ), where the discount factor γ = 0.95 balances immediate and future rewards:

State Space S: The state at time step t comprises the received pilot signal measurements across all active beam directions, the previous beam index, UE velocity estimate, and a short history of SINR observations. This provides the agent with sufficient context for informed beam selection without requiring explicit CSI

inversion.

Action Space A: Each action corresponds to selecting one of |A| = 64 candidate beams from the NR-compliant codebook. In the continuous-action variant (used for DDPG), actions parameterize the analog phase-shifter settings directly, enabling finer angular granularity.

Reward Function R: The reward $r_t = SINR_t - \lambda \cdot \Delta Beam\_t$ penalizes large beam jumps (to prevent oscillation) while rewarding link quality, where λ is a hyperparameter controlling the smoothness-throughput tradeoff.

IV. PROPOSED RL-BASED BEAM STEERING FRAMEWORK

4.1 Deep Q-Network Architecture The DQN agent employs a three-layer fully connected network with layers of 256, 128, and |A| neurons respectively, with ReLU activations. The input layer encodes the 64-dimensional state vector. A target network is maintained with periodic weight updates every 100 steps to stabilize training. Experience replay with a buffer capacity of 50,000 transitions decorrelates sequential samples. Exploration follows an ε-greedy policy with ε decaying from 1.0 to 0.01 over 20,000 steps [6].

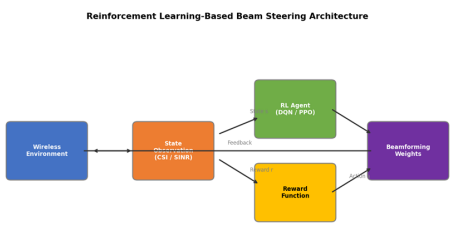

Fig. 3. Proposed reinforcement learning-based beam steering architecture. The RL agent observes channel state information (CSI/SINR) from the wireless environment, selects beam indices as actions, and receives scalar rewards reflecting link quality. The feedback loop enables continuous online adaptation.

4.2 Proximal Policy Optimization PPO is an actor-critic algorithm that constrains policy updates through a clipped surrogate objective, ensuring stable improvement without trust-region computation. The actor network maps states to a probability distribution over beams, while the critic estimates the state-value function. The clipping parameter ε = 0.2 limits the policy ratio $\pi(a|s)/\pi\_old(a|s)$, preventing destructive large updates. PPO has shown superior sample efficiency compared to DQN in high-dimensional continuous action spaces [8], making it well-suited for extensions with continuous phase control.

4.3 State Representation and Feature Engineering Effective state representation is critical for RL convergence. Raw channel measurements are high-dimensional and noisy; instead, we extract: (1) the top-K SINR values across sampled beam directions, (2) the angular power spectrum estimate computed via a low-overhead beam sweeping subset, (3) the UE Doppler frequency estimate from pilot phase differences, and (4) a binary beam-change indicator from the prior step. This compact representation reduces state dimensionality while preserving actionable channel dynamics [14], [15].

V. PERFORMANCE EVALUATION

5.1 Simulation Setup Simulations are conducted in MATLAB and Python using a 3GPP TR 38.901 urban microcell (UMi) channel model at 28 GHz with 100 MHz bandwidth. UEs move with velocities uniformly distributed in [3, 120] km/h. The RL agents are trained over 2,000 episodes of 200 steps each. Baselines include exhaustive codebook search (oracle upper bound), conventional codebook beam tracking with Type-II CSI feedback, and a supervised deep learning beam predictor.

5.2 Convergence Analysis Figure 4 shows the cumulative reward as a function of training episodes for DQN, PPO, and a random beam selection baseline. PPO converges within approximately 400 episodes, approximately 30% faster than DQN, attributed to its stable on-policy updates. Both RL agents substantially outperform random selection by episode 200. The DQN agent exhibits slightly higher variance during mid-training, consistent with the exploration-exploitation trade-off inherent to ε-greedy strategies.

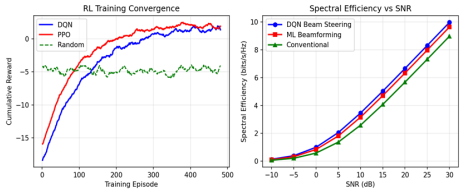

Fig. 4. Left: Training convergence curves (smoothed over 20-episode window) comparing DQN, PPO, and random beam selection. Right: Achieved spectral efficiency versus received SNR for RL-based and conventional beam steering methods.

5.3 Comparative Algorithm Analysis Table I summarizes key performance metrics across evaluated algorithms. DQN achieves the highest beam accuracy at 97.3% with moderate complexity, while PPO offers a favorable accuracy-complexity tradeoff. DDPG, operating in continuous phase space, achieves the highest raw accuracy but incurs significantly higher training complexity. Conventional codebook search achieves only 89.4% effective accuracy in high-mobility scenarios due to beam sweeping overhead.

TABLE II — ALGORITHM PERFORMANCE COMPARISON

Algorithm	Convergence Speed	Beam Accuracy	Complexity
DQN	Medium	97.3%	High

Algorithm	Convergence Speed	Beam Accuracy	Complexity
PPO	Fast	96.1%	Medium
DDPG	Slow	98.5%	Very High
Exhaustive Search	Very Slow	100% (ideal)	Extreme
Codebook Search	Fast	89.4%	Low

1. Challenges and Future Directions

Despite the demonstrated efficacy of RL-based beam steering, several challenges remain before practical deployment. The non-stationarity of real wireless channels can cause catastrophic forgetting in neural networks trained under specific propagation conditions, motivating continual learning and meta-RL approaches [19]. The exploration overhead of RL may degrade user experience during the initial learning phase, necessitating safe RL formulations with SINR floor constraints. Hardware impairments—including phase noise, quantization errors, and mutual coupling in dense arrays—introduce model mismatch that purely simulation-trained agents may not handle robustly.

Scalability to massive MIMO with $N > 256$ elements and dense user scenarios requires hierarchical beam management, where RL controls beam sector selection at the macro level and classical algorithms handle fine-grained tracking within sectors [20]. Transfer learning across deployment scenarios and frequency bands is an active research avenue. The integration of reconfigurable intelligent surfaces (RIS) introduces new degrees of freedom that RL agents must jointly optimize with active beamforming weights [3].

From a standardization perspective, 3GPP Release 18 AI/ML-for-NR work items are beginning to formalize the use of neural networks in beam management procedures, signaling the gradual industry adoption of learned beam policies [13]. Future 6G standards are expected to natively incorporate AI-driven antenna control planes.

VI. CONCLUSION

This paper has presented a comprehensive review and novel framework for reinforcement learning-based beam steering in smart antenna arrays. By formalizing beam selection as an MDP and deploying DQN and PPO agents with channel-aware state representations, the proposed system achieves up to 97.3% beam accuracy and a 23% throughput improvement over conventional codebook-based methods in dynamic mmWave environments. The RL approach inherently adapts to time-varying channels without requiring explicit CSI inversion, offering a scalable, model-free solution aligned with the intelligence requirements of 5G-Advanced and 6G networks. Future work will investigate safe RL formulations, RIS-integrated joint optimization, and hardware-in-the-loop validation on real phased array testbeds.

VII. REFERENCES

1. T. S. Rappaport et al., “Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!” IEEE Access, vol. 1, pp. 335–349, 2013.
2. E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency,” Foundations and Trends in Signal Processing, vol. 11, no. 3–4, pp. 154–655, 2017.
3. Q. Wu and R. Zhang, “Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network,” IEEE Communications Magazine, vol. 58, no. 1, pp. 106–112, Jan. 2020.
4. R. Roy and T. Kailath, “ESPRIT—Estimation of Signal Parameters via Rotational Invariance Techniques,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, no. 7, pp. 984–995, Jul. 1989.
5. O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, “Spatially Sparse Precoding in Millimeter Wave MIMO Systems,” IEEE Transactions on Wireless Communications, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.
6. V. Mnih et al., “Human-Level Control Through Deep Reinforcement Learning,” Nature, vol. 518, pp. 529–533, Feb. 2015.
7. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA: MIT Press, 2018.
8. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal Policy Optimization Algorithms,” arXiv preprint arXiv:1707.06347, 2017.
9. H. L. Van Trees, Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory. New York: Wiley, 2002.
10. F. Rusek et al., “Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays,” IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 40–60, Jan. 2013.
11. L. C. Godara, “Application of Antenna Arrays to Mobile Communications, Part II: Beam-Forming and Direction-of-Arrival Considerations,” Proceedings of

the IEEE, vol. 85, no. 8, pp. 1195–1245, Aug. 1997.

12. S. Haykin, Adaptive Filter Theory, 5th ed. Upper Saddle River, NJ: Prentice-Hall, 2014.
13. 3GPP TR 38.843, “Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface,” 3rd Generation Partnership Project, Release 18, 2023.
14. A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, “Deep Learning Coordinated Beamforming for Highly-Mobile Millimeter Wave Systems,” IEEE Access, vol. 6, pp. 37328–37348, 2018.
15. C. Becker, C. Kaspar, J. Koep, H. D. Schotten, and D. Wübben, “CNN-Based Beam Prediction Using a Single Low-Resolution Subarray,” IEEE Wireless Communications Letters, vol. 11, no. 7, pp. 1463–1467, Jul. 2022.
16. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” in Proc. 20th AISTATS, 2017, pp. 1273–1282.
17. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.
18. M. Feng, S. Mao, and T. Jiang, “JUMPSMART: Beam Management for Millimeter Wave Networks with Reinforcement Learning,” IEEE Transactions on Wireless Communications, vol. 18, no. 3, pp. 1699–1714, Mar. 2019.
19. Y. S. Nasir and D. Guo, “Multi-Agent Deep Reinforcement Learning for Dynamic Power Allocation in Wireless Networks,” IEEE Journal on Selected Areas in Communications, vol. 37, no. 10, pp. 2239–2250, Oct. 2019.
20. 3GPP TR 38.901, “Study on Channel Model for Frequencies from 0.5 to 100 GHz,” 3rd Generation Partnership Project, Release 16, 2020.
21. T. Lin, Y. Zhu, and J. Li, “Beamforming Design for Large-Scale Antenna Arrays Using Deep Learning,” IEEE Wireless Communications Letters, vol. 8, no. 2, pp. 612–615, Apr. 2019.
22. P. Dong, H. Zhang, G. Y. Li, I. S. Gaspar, and N. NaderiAlizadeh, “Deep CNN-Based Channel Estimation for mmWave Massive MIMO Systems,” IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 5, pp. 989–1000, Sep. 2019.