OPTIMIZING ENERGY EFFICIENCY PERFORMANCE IN RIS-ASSISTED NEAR-FIELD MIMO SYSTEM USING DEEP RL

Amjad Iqbal[a], Ala'a Al-Habashna[a b], Gabriel Wainer[a] and Gary Boudreau[c]

[a] Department of System and Computer Engineering at Carleton University, Ottawa, Canada
[b] School of Computing and Informatics, Al Hussein Technical University, Amman, Jordan
[c] Ericsson Canada, Kanata, Canada

ABSTRACT

Multiple-input-multiple-output (MIMO) technology improves the capacity and reliability of wireless communication networks. However, high hardware complexity and energy usage often impact their scalability. Reconfigurable intelligent surfaces (RISs) act as a promising solution to tackle these challenges owing to their energy-efficient design, but the isolation between RIS elements limits the ability to control wave propagation fully. Integrating RIS with MIMO offers new ideas and solutions to these limitations. In the scenario where the base station (BS) and users are closely placed, the antenna operates in near-field regimes, causing wavefronts to exhibit spherical rather than planar ones, making traditional far-field optimization techniques ineffective. To overcome such challenges, we propose an advanced deep reinforcement learning approach to jointly optimize BS beamforming and RIS phase shifts to maximize the energy efficiency (EE) performance in a near-field network. The proposed approach achieves up to 10-22% improvement in EE compared to conventional methods.

Keywords: 5G/B5G/6G, MIMO, RIS, energy efficiency, deep reinforcement learning.

1 INTRODUCTION

The need for higher throughput, density, and diverse applications drives the development of sixth-generation (6G) networks. Massive multiple-input multiple-output (mMIMO) technology has gained considerable attention to meet these requirements. mMIMO increases the number of antennas, developing extremely large antenna arrays (ELAAs) [1], which change electromagnetic wave nature. Electromagnetic waves change from plane to spherical wavefronts, especially in the near-field region, which can be used for enhanced spatial resolution, precise control for beamforming, and better localization. However, studying near-field effects requires accurate modeling and innovative solutions to manage increasing complexities and power requirements for future wireless communication systems.

Reconfigurable intelligent surfaces (RIS) are an emerging solution to address these issues and offer a scalable and power-efficient feature to near-field systems [2]. RIS offers precise manipulation of electromagnetic waves by deploying a large number of controllable elements for optimal signal transmission and beamforming [3]. Each element in RIS could be individually adjusted for the phase and amplitude of incidence waves, intelligently shaping the transmitted wave's direction and strengthening the transmission efficiency. This makes RIS a promising technology for addressing the traditional limitations of wireless communication networks. RIS is primarily used to optimize phase shifts for efficient signal reflection; however, the effectiveness of RIS significantly depends on the accurate channel state information (CSI), helping to determine the optimal configuration for beamforming and signal enhancement. Obtaining accurate CSI in RIS-aided networks is always challenging due to its passive

Proc. of the 2025 Annual Simulation Conference (ANNSIM'25),
May 26-29, 2025, Universidad Complutense de Madrid, Madrid, Spain
J. L. Risco-Martín, G. Rabadi, D. Cetinkaya, R. Cárdenas, S. Ferrero-Losada, and A. Bany Abdelnabi, eds.

reflective nature. Therefore, advanced techniques are needed to unlock RIS's full potential in improving the performance of future wireless communication systems [4].

As the RIS elements are passive, minimum mean square error (MMSE)/least squares algorithms can be applied to estimate the cascaded channel link (i.e., BS-RIS-user) [5], [6]. RIS's location, elements, and reflective effects can be adjusted based on particular communication conditions, making it easy to scale and accommodate different requirements. RIS-assisted networks have a multiplicative fading impact; thus, the equivalent path loss is the product rather than the sum of two specific path losses (i.e., the path loss of the BS-RIS link and the RIS-user link [7]). Therefore, large-scale RIS deployments could significantly alter the propagation and manipulation of electromagnetic fields (EMF) [8].

EMF radiation patterns can be separated into far-field region and near-field region and can be determined by the Rayleigh distance (i.e., $2\,D^2/\lambda$), where D is the largest array aperture dimension, and λ is the transmission wavelength [9]. In the case of an RIS-assisted framework, if the array aperture of the RIS is not large and the Rayleigh distance is small, there is wave scattering in the far field of the RIS. However, this far-field assumption may not appear in some scenarios when the distance between the RIS and the user is short. Specifically, when the number of RIS elements increases, the array aperture becomes large, and the Rayleigh distance increases. Thus, it is essential to consider a near-field channel model to describe the transmission of the signal [10]. For instance, a near-field channel model was used to maximize the weighted sum rate for the RIS-assisted MIMO network [11]. Similarly, in [12], the channel estimation method was investigated for near-field networks to maximize the sum rate. In [13], energy efficiency (EE) performance was enhanced by optimizing the transmit power and RIS elements. All the methods above use a static model with fixed CSI, and an objective function is recalculated at the start of each iteration. This results in high computational complexity, high pilot overhead, and noise vulnerability, making such approaches less effective for dynamic scenarios.

In these cases, we can use machine learning (ML), and especially deep learning (DL) [14], which is well-suited for complex and data-driven tasks and has the capabilities of learning and feature extraction. In near-field systems, DL models can be trained offline while the deployment can be done online, being effective for real-time applications [15]. For instance, in [16], the received signal strength was used to identify optimal near-field codewords to maximize the sum rate. In [17], the optimal angle and distance were jointly predicted for the near-field codebooks using a deep neural network (DNN). A transmission architecture using DL for near-field MIMO [18] was used to achieve a maximum sum rate while considering the effect of imperfect CSI. A convolutional neural network (CNN) was used to capture the features from the complex CSI and maximize the achievable rate [19]. However, these approaches do not scale well in dynamic scenarios, as a large amount of pilot overhead is needed for codebook optimization and accurate CSI estimation. To address this challenge, deep reinforcement learning (DRL) can learn policies directly from interacting with the environment, reducing reliance on extensive pilot signaling.

Here, we propose an advanced near-field training approach for RIS-assisted MIMO networks. We introduce a low-complexity on-policy (i.e., proximal policy optimization (PPO)) algorithm, which adapts to real-time channel conditions without relying on prior historical data or environment-specific assumptions. We built the model into a simulation environment design that captures the near-field effects. The simulation framework builds the foundation, validates the mathematical model, and explores how near-field properties influence key system performance metrics (e.g., EE). This simulation framework complements the theoretical model, offering insights. The results to be discussed later show that based on simulation analysis, we can achieve up to 10-22% better EE performance than the traditional approach under similar conditions.

2 SYSTEM MODEL AND PROBLEM FORMULATION

We consider a near-field RIS-assisted communication network shown in Figure 1. We assume that a single multi-antenna BS serves multiple users with multiple antennas. Without loss of generality and avoiding confusion, we suppose a realistic environment scenario where several propagation obstacles obstruct the direct communication between the BS and the users, which is common in modern wireless networks and

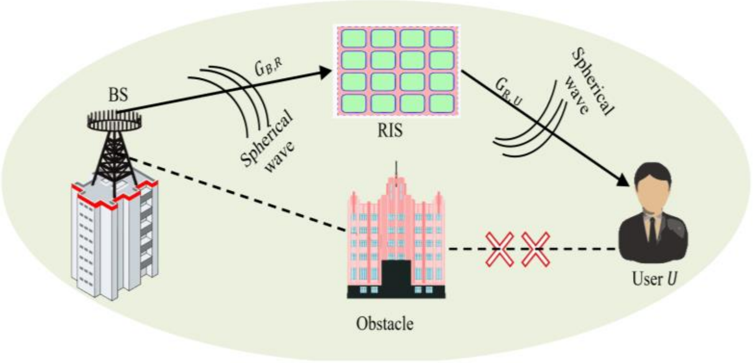

Figure 1. RIS-Assisted Near-field MIMO Network

results are derived solely from simulations; thus, RIS is required to facilitate such communication. The BS and user are equipped with N_b-elements and N_u-elements arranged in a uniform linear array, respectively. The RIS has passive reflecting elements and is composed of a uniform planar array configured with $R = R_xR_y$, where R_x and R_y are the number of elements along the horizontal and vertical axes. We use a near-field channel model, where the system operates at a carrier frequency f_c leading to the transmission wavelength $\lambda_c = c/f_c$, such that c indicates the speed of light. The array apertures for each element, i.e., BS, RIS, and user, is represented as D_B, D_R and D_U and can be expressed mathematically $D_B = (N_b - 1)d$, $D_R = \sqrt{[(R_x - 1)d]^2 + \left[\left(R_y - 1\right)d\right]^2}$, and $D_U = (N_u - 1)d$, respectively.

2.1.1 Channel Model

The channel characteristics for the proposed system model are categorized into two regions: the channel between BS and RIS and the channel between RIS and user, followed by the Rayleigh distance $\mathfrak{R}$

$$\mathfrak{R} = \frac{2D^2}{\lambda} \tag{1}$$

We consider a three-dimensional topology for the proposed communications system model where the BS antennas are located on the x-axis, which implies that the coordinates of the BS array's midpoint are (0, 0, 0). This means that the coordinate of the $\boldsymbol{n_{\mathscr{b}}}$-th antenna at the BS can be represented as

$$q_B(n_{\mathscr{b}}) = (\tilde{n}_{\mathscr{b}}d, 0, 0) \tag{2}$$

such that $\tilde{n}_{\mathscr{b}} = n_{\mathscr{b}} - \frac{N_B-1}{2}$. Furthermore, we assume that the user antennas are parallel to the x-axis. Thus, the coordinates of the user midpoint array are (x_U, y_U, z_U), where the coordinate of the n_u-th is defined as

$$q_U(n_u) = (x_U + \tilde{n}_u d, y_U, z_U), \tag{3}$$

where $\tilde{n}_u = n_u - \frac{N_U-1}{2}$. Similarly, we assume all RIS elements are parallel to the XY plane. The RIS midpoint coordinates are denoted as (x_R,y_R,z_R). Hence, the coordinates of RIS for the (r_x,r_y)-th elements can be represented as

$$q_R\left(r_x, r_y\right) = \left(x_R + \tilde{r}_x d, y_R + \tilde{r}_y d, z_R\right), \tag{4}$$

where $\tilde{r}_x = r_x - \frac{R_x-1}{2}$ and $\tilde{r}_y = r_y - \frac{R_y-1}{2}$. As a result of the above geometry, the channel path between BS and RIS can be modeled as follows:

$$G_{B,R} = \left[g_{B,R}^{1}, \ldots, g_{B,R}^{r_{xy}}, \cdots, g_{B,R}^{R}\right]^T, \tag{5}$$

where $g_{B,R}^{r_{xy}} = \left[\mathscr{k}_{r_{xy}},1e^{-j\frac{2\pi c}{f_c}d_{r_{xy,1}}^B},\cdots,\mathscr{k}_{r_{xy}},N_B e^{-j\frac{2\pi c}{f_c}d_{r_{xy,N_B}}^B}\right]^T$, and $\mathscr{k}_{r_{xy}},n_b$ indicates the free space path loss between the BS n_b-th antenna and the RIS (r_x,r_y)-th elements. The distance between the BS n_b-th antenna and the RIS (r_x,r_y)-th elements can be calculated as

$$d^B_{r_{xy}},n_b = \left\| q_B(n_b) - q_R(r_x,r_y) \right\|_2, \tag{6}$$

We further expand $d^B_{r_{xy}},n_b$ as

$$\begin{aligned} d^B_{r_{xy}},n_b &= \sqrt{(x_R + \tilde{r}_x d - \tilde{n}_{\mathcal{b}} d)^2 + \left(y_R + \tilde{r}_y d\right)^2 + z_R^2} = \sqrt{\left(c^B_{r_{xy}}\right)^2 + (\tilde{n}_{\mathcal{b}} d)^2 - 2\tilde{n}_{\mathcal{b}} d c^B_{r_{xy}} \sin\alpha^B_{r_{xy}} \sin\beta^B_{r_{xy}}} \\ &\overset{(a)}{\approx} c^B_{r_{xy}} - \tilde{n}_{\mathcal{b}} d \sin\alpha^B_{r_{xy}} \sin\beta^B_{r_{xy}} + \frac{(\tilde{n}_{\mathcal{b}} d)^2 (1 - \sin^2\alpha^B_{r_{xy}} \sin^2\beta^B_{r_{xy}})}{2c^B_{r_{xy}}} \end{aligned} \tag{7}$$

where $c^B_{r_{xy}}$, $\alpha^B_{r_{xy}}$ and $\beta^B_{r_{xy}}$ represents the midpoint distance (0,0,0) between BS and the RIS $\left(r_x, r_y\right)$-elements, the azimuth and elevation angle, respectively. The approximation (a) is obtained by the Taylor series, i.e., $\sqrt{1+a} \approx 1 + \frac{a}{2} - \frac{a^2}{8}$.

Similarly, the channel path between the RIS and the user can be represented as

$$G_{R,U} = \left[g^1_{R,U}, \dots, g^{n_u}_{R,U}, \cdots, g^{N_U}_{R,U}\right]^T, \tag{8}$$

where $g^{n_u}_{R,U} = \left[\mathcal{k}_{n_u},1 e^{-j\frac{2\pi c}{f_c} d^U_{n_{u,1}}}, \cdots, \mathcal{k}_{n_u},R e^{-j\frac{2\pi c}{f_c} d^U_{n_{u,R}}}\right]^T$, and $\mathcal{k}_{n_u},r_{xy}$ represent the free space path loss between the RIS (r_x,r_y)-th elements and the user n_u-th antenna. The distance between the RIS (r_x,r_y)-th elements and the user n_u-th antenna is expressed by

$$d^U_{n_u},r_{xy} = \left\| q_U(n_u) - q_R(r_x,r_y) \right\|_2, \tag{9}$$

We further expand $d^U_{n_u},r_{xy}$ as

$$d^U_{n_u},r_{xy} \approx c^U_{r_{xy}} - \tilde{n}_u d \sin\alpha^U_{r_{xy}} \sin\beta^U_{r_{xy}} + \frac{(\tilde{n}_u d)^2 (1 - \sin^2\alpha^U_{r_{xy}} \sin^2\beta^U_{r_{xy}})}{2c^U_{r_{xy}}} \tag{10}$$

where $c^U_{r_{xy}}$, $\alpha^U_{r_{xy}}$ and $\beta^U_{r_{xy}}$ indicates the distance between the user array and the RIS $\left(r_x, r_y\right)$-elements, the azimuth and elevation angle, respectively. In the case of near-field, the line-of-sight (LoS) channel for the BS-RIS and RIS-user links has degrees of freedom (DoF) given by $l = \min\{N_B, N_U, R\}$. In large-scale arrays, l will be significantly greater than 1, even without non-LoS paths. One of the main advantages of near-field communications is the higher DoF, allowing support for multiple data streams, $q > 1$, without relying on abundant environmental scattering. This results in enhanced DoF in RIS-aided near-field communications. Considering the above details, the channel for the proposed system model is given as

$$G_i = G_{B,R} \Theta G_{R,U}. \tag{11}$$

where Θ represents the RIS phase shift matrix and can be expressed as $\Theta = \mathrm{diag}(\phi_1, \dots \phi_r, \dots \phi_R)$, such that $\phi_r = e^{-j\theta_r}$, and $\theta_r \in [0, 2\pi]$ denoting the RIS phase-shift coefficient for the r-th elements. The received signal of the user can be expressed as

$$y_u = U^H G_i W x + \eta, \tag{12}$$

such that $U \in \mathbb{C}^{N_U \times q}$ is the combing matrix, $W \in \mathbb{C}^{N_B \times q}$ is the beamforming matrix, $x \in \mathbb{C}^{q \times 1}$ is the symbol vector which satisfies $\mathbb{E}[xx^H] = I_q$, and $\eta \sim \mathcal{CN}\left(0, \sigma_\eta^2\right)$ indicates the additive white Gaussian noise (AWGN). The achievable data rate (δ) for the user can be expressed mathematically as

$$\delta_i = \log_2\left(1 + \frac{|U^H G_i W|^2}{\sum_{i \neq k} |U^H G_k W|^2 + \sigma_\eta^2}\right). \tag{13}$$

We maximize EE's performance by jointly optimizing beamforming at the BS and phase shift matrices at RIS. We define the EE as the ratio of the total sum rate (δ_i) to the total power consumption (W).

$$EE = \frac{\sum_{u=1}^{U} \delta_i}{\sum P_T} \tag{14}$$

Where $P_T = P_t + P_{circuit} + P_{RIS}$, such that P_t represent total transmit power, $P_{circuit}$ indicates static circuit power and is equivalent to $P_{circuit} = P_{BS} + P_U$, whereas P_{RIS} denotes the RIS power consumption.

2.2 Problem Formulation

Codebook-based training schemes are typically employed in wireless communication networks' for near-field beamforming and phase shift optimization. However, these methods usually have a large codebook size and incur large training overhead in near-field practical scenarios. In the near-field scenario, spherical wavefront propagation is considered, where the channel gain depends on both the distance and angular variation between the transmitter and receiver. To address this issue, our primary focus is to jointly optimize the BS beamforming (W) and RIS phase shift (Θ) directed toward the legitimate UE's location. We aim to achieve optimal beamforming (W) at BS and phase shift (Θ) at RIS without relying on a predefined codebook to maximize the EE performance. The optimization problem using the beamforming matrix and phase shift is formulated as

$$\max_{W,\Theta} EE \tag{15a}$$

$$s.t. \|W\|^2 \le P_{max}, \tag{15b}$$

$$\Theta \in (0,2\pi), \quad \forall_r \tag{15c}$$

$$\delta_i \ge \delta_{min}, \quad \forall_u \tag{15d}$$

such that constraint (15b) represents the maximum power value at the BS, while constraint (15c) ensures that the phase shift values at the RIS are bounded between 0 and 2π, and constraint (15d) indicates the user QoS satisfaction. Unlike far-field scenarios, near-field phase shift designs must consider both focusing and steering effects, as the proximity between BSs, RISs, and UEs introduces spatial variations that impact signal propagation.

The CSI of the proposed systems is updated at each time step t, leading to increased network complexity, non-convexity, and significant non-linearity to solve the optimization problem. Furthermore, the channel model introduces additional DoF due to the channel's sensitivity to both distance and angle, increasing the challenge of the optimization problem. Traditional approaches, such as AO and exhaustive search methods, can be used to solve such optimization problems. However, such approaches require high computation capabilities and are not feasible when dealing with large and dynamic-scale network problems. Therefore, we introduce an advanced ML algorithm to tackle this issue. Specifically, we design a model-free DRL that can flexibly adapt to learn the behavior of the environment and find optimal beamforming vectors and phase shifts under complex channel conditions.

3 PROPOSED SOLUTION USING DRL

DRL is a sub-branch of DL, which relies on the learning process where agents interact with an environment and learn subsequently. This allows the agent to use DRL models for online learning by generating the sample independently. DRL algorithms can be categorized into value-based and policy-based and can be used to solve the discrete and continuous action space. In the case of the value-based, Q-learning, Deep Q-learning, and Double deep Q-learning are standard algorithms used to solve the discrete action variable space problem. However, such algorithms are suitable to support the smaller action space problem. The policy-based DRL is the fine-tuning RL technique that can solve problems with continuous variables action space. In the case of policy-based, off-policy and on-policy are the two approaches for training the DRL algorithm. In off-policy approaches (e.g., deep deterministic policy gradient (DDPG)), the agent can learn from past experiences stored in a replay buffer, potentially accelerating learning by not strictly following the current policy during updates. However, such an approach is unstable due to discrepancies between the behavior and target policies. Moreover, the actor and critic network must be updated at each time step t, which can require significant computational resources. Furthermore, accurate hyperparameter tunning is also challenging in the case of dynamic and complex environments, leading to suboptimal performance. On the other hand, the on-policy approaches (e.g., proximal policy optimization (PPO)) tackle these challenges by leveraging a clipped surrogate objective function, which helps update the present policy incrementally without deviating too far from the old policy, maintaining stability and reducing the risk of divergence. Furthermore, PPO does not rely on replay buffers, which simplifies implementation and reduces computational complexity. Therefore, we used the PPO approach to effectively solve the optimization problem defined in (15). The following section explains the DRL formulation and proposed PPO approach.

3.1 DRL Formulation

DRL relies on the iterative learning process where an agent constantly interacts with the environment to optimize the performance. The agent observes the current state $s_t \in \mathcal{S}$ and execute an action $a_t \in \mathcal{A}$ at each time step t, where $\mathcal{S}$ and $\mathcal{A}$ indicate the set of possible states and actions, respectively. After executing the action, the agent receives feedback from the environment in the form reward r and moves to the next state s_{t+1} by following the policy π. The state, action, and reward are modeled as a Markov decision process (MDP) and defined as $\{\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma\}$, where $\mathcal{P}$ is the transition state when the agent moves from one state to another, and γ is the discount factor. The agent aims to find the optimal policy to maximize the cumulative reward function over time. In this work, the agent is modeled as BS and RIS, and the state, action, and reward for our problem formulation are defined as

- *State space*: At each time step t, the state space $\mathcal{S}$ observes the environment and collects the CSI, including the angle and distance information between BS-RIS and RIS-UEs. We express the state space as

$$\mathcal{S} = \left[G_{B,R} + G_{R,U} + \Theta W\right] \tag{16}$$

- *Action space:* The action space $\mathcal{A}$ contains the information on the angle and distance between BS-RIS and RIS-UEs, and the aim is to find the optimal beamforming matrix W at the BS and phase shift Θ at the RIS at each time step t, respectively.
- *Reward:* The reward function r is to maximize the EE performance and can be formulated as

$$r = EE = \frac{\sum_{u=1}^{U} \delta_i}{\sum P_T} \tag{17}$$

3.2 Proximal Policy Optimization

PPO is the advanced version of the DRL algorithm derived from trust region policy optimization (TRPO) and is less complex than TRPO regarding computational value [20]. The neural network used in the PPO algorithm is composed of the actor and critic network. The actor network is responsible for the action selection based on the current policy and generates the probability distribution over the possible action. The critic network evaluates and predicts the Q-value for taking action a_t in a particular state s_t and capture the impact of angle and distance factor within the Q-value function (i.e., $Q(s_t, a_t)$) to enhance the performance of the near-field MIMO system. Following the actor and critic network, the objective function is defined as

$$g = \mathbb{E}\left[\sum_{t=1}^{T} \nabla_\psi \log \pi_\psi \,(s_t|a_t) A_{\pi_\psi}\right] \tag{18}$$

where $\mathbb{E}[\cdots]$ is the expectation used in the optimization process and approximated through sampling to empirically estimate the objective function's gradient. ∇_ψ indicated the gradient of the policy parameters in DNN. A_{π_ψ} represents the advantage function that helps to prevent the algorithm from overfitting and minimize the variance at each time step t and can be defined as

$$A_{\pi_\psi} = Q_{\pi_\psi}(s_t, a_t) - V_{\pi_\psi}(s_t), \tag{19}$$

To mitigate the complexity of the KL-divergence constraint defined in TRPO, the PPO algorithm utilizes a clipped surrogate objective (CSO) function, which uses a clip function to limit policy updates to a specific range over multiple training steps, reducing the algorithm's complexity. The CSO is designed to prevent the more extensive weight update of the PPO algorithm and can be formulated as

Algorithm 1: Proposed PPO Algorithm

```
1.  **Initialize** actor and critic network parameters hyperparameters
2.  **for** episode $e \leftarrow 1$ to $\mathcal{E}$ **do**
3.    **Initialize** environment and reset the initial state $s_o$ with CSI ($G_{B,R}$,$G_{R,U}$)
4.    **for** timestep $t \leftarrow 1$ to $T$ **do**
5.      Observe the current state $s_t$
6.      Sample action $a_t$ from the policy $\pi_\psi$, such that $a_t \sim \pi_\psi(a_t|s_t)$.
7.      Execute action $a_t$, observe the reward $r$ and move to the next state $s_{t+1}$
8.      Store all these experience $(s_t, a_t, r_t, s_{t+1})$ in the replay memory $\mathcal{D}$
9.      Update policy and value function at every timestep from $\mathcal{D}$
10.     Compute the CSO by calculating the probability ratio and applying the clip factor to limit
        policy updates for equation (20)
11.     Calculate the final objective function for equation (22) following the squared error loss $H_t(\psi)$.
12.     Calculate the advantage function using equation (19).
13.     Update policy parameter $\psi$ with gradient ascent as $\psi \leftarrow \psi + \alpha\nabla_\psi O^t_{Final}(\psi)$,
14.     **Repeat** steps 5-13 until the algorithm converges.
15.   **End for**
16. **End for**
```

$$O^t_{CLIP}(\psi) = \mathbb{E}_t\left[\min\left(\eta_t(\psi), clip(\eta_t(\psi), 1-\sigma, 1+\sigma)\right)A_{\pi_\psi}\right], \tag{20}$$

where $\eta_t(\psi)$ represent the probability ratio and indicate the choices of selection of policies and defined as

$$\eta_t(\psi) = \frac{\pi_\psi(a_t|s_t)}{\pi_{\psi_{\text{old}}}(a_t|s_t)} \tag{21}$$

$\pi_{\psi_{\text{old}}}$ denotes the policy at $t-1$ instant. Clipping the probability ratio ensures the minimum required similarity between two consecutive policies. σ is the clipped factor hyperparameter. Thus, the final objective function of the proposed algorithm is represented as

$$O^t_{Final}(\psi) = \mathbb{E}_t\left[O^t_{CLIP}(\psi) - c_1 H_t(\psi) + c_2\mathbb{E}_{\pi_\psi}(s_t)\right], \tag{22}$$

where c_1, c_2 and $H_t(\psi)$ indicates the coefficients and value squared error loss and $H_t(\psi) = \left(V_{\pi_\psi}(s_t) - V_{target}\right)^2$. The detailed pseudocode for the proposed framework is given in Algorithm 1.

4 SIMULATION RESULTS ANALAYSIS

In this section, we validate the simulation results for the proposed PPO algorithm and compare its performance with the benchmark schemes. The channel matrices for both channels are randomly generated following the near-field Rayleigh distribution (i.e., $\Re = 2D^2/\lambda$). The simulation parameters and hyperparameters for the proposed algorithm are defined in Table 1.

4.1 Benchmark schemes

We considered three benchmark algorithms and compared their performance with the proposed PPO approach. A brief overview of these benchmark algorithms is as follows:

• Near-field beam training using DL: A CNN-based beam training method is proposed to solve the challenges in the ELAA system by optimizing beamforming in the near-field, where users experience a spherical wavefront [19]. Although the CNN model avoids the complexity of codebook searches, it may

Table 1. Simulation Parameters

Parameter/Hyperparameter	Value
Noise power σ_η^2	-105 dBm
Carrier frequency f_c	45 GHz
Transmit power P_t	40 dBm
Clip factor σ	0.2
Total number of episodes $\mathcal{E}$	1000
Number of time steps T	10000
Antenna spacing d	$\lambda_c/2$
Experience replay buffer $\mathcal{B}$	100000
Discount factor γ	0.9
Learning rate α	10^{-5}
Batch size $\mathcal{D}$	64
Number of hidden layers	3
Neuron per hidden layer	256, 256, 64

lack adaptability in dynamic environments where channel conditions are updated constantly. This means the model entirely relies on the predefined dataset, and its generalizability to unseen scenarios can be limited. Moreover, this approach does not handle joint optimization of beamforming and RIS phase shifts.

- Hierarchical beam training using alternating optimization (AO): An OA algorithm solves the joint optimization problem in RIS-assisted near-field MIMO networks [21]. In this case, the objective function is recalculated at the beginning of each iteration, leading to high computational complexity and convergence issues, especially in complex and dynamic scenarios. Moreover, a fixed codebook is designed for the specific channel models, leading to suboptimal performance when real-time adaptability is required or channel conditions change with the user mobilities.
- Far-field optimization: In the case of far-field scenarios, beamforming methods permit the assumption that the electromagnetic waves exhibit planar wavefronts as the distance between the transmitter and the receiver far exceeds the Rayleigh distance. Most of the optimization techniques for the far field usually employ fixed, directionally focused codebooks or analytical beamforming schemes to steer beams toward the intended users' directions, helping to improve the signal quality [22] and [23].

The simulation results for the behavior of the loss function across episodes under different power levels and SNR values are presented in Figure 2. The loss function decreases monotonically with increasing episodes. It can be observed from Figure 2 that when the episodes are close to 650-800, the loss function starts to converge, and optimization process becomes static, and no further improvement occurs. It can be revealed from the given result that a higher power (i.e., P_t=40dBm) obtains comparatively lower loss values than the lower power (e.g., P_t=20dBm) since a higher power value increases the signal strength and reduces the complexity of the optimization process for the proposed framework. Likewise, the impact of SNR is also shown. It indicates that the higher SNR value (i.e., 40dB) provides a much lower loss than the lower SNR value (i.e., 10dB), showing the importance of the channel conditions in reducing interference and noise. Figure 2 highlights that the higher values of power and SNR achieve the lowest loss and provide faster convergence compared to the lower values of power and SNR. Balancing power and channel quality significantly impacts the system's performance.

In Figure 3, we show a simulation result where all approaches converge. The proposed PPO approach achieves the highest EE with every increasing episode and starts to converge around 650 episodes with no further improvement. This highlights the ability of the proposed PPO approach to effectively and adaptively optimize the performance of EE in high-dimensional scenarios. The near-field DL approach achieves comparatively high EE performance and reaches 5.2 bits/J; however, it takes more episodes to converge (i.e.,750 episodes), indicating its limitation in solving highly complex problems. The other approaches (i.e., AO and Far-field) converge when episodes reach 600 and 500, respectively. This indicates that these two approaches require fewer episodes to converge; however, the achieved EE performance is much less than

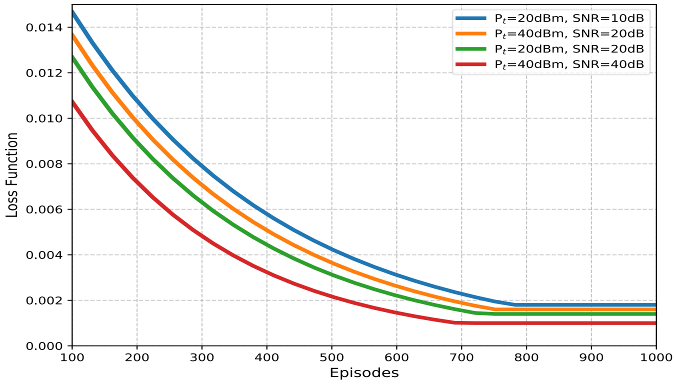

Figure 2. Loss Function with respect to episodes

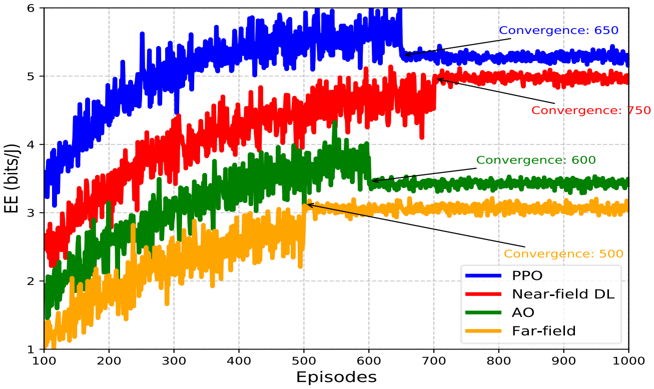

Figure 3. Convergence Analysis

the other ML approaches due to limited applicability in highly dynamic systems. This shows the effectiveness of the ML approaches for highly dynamic scenarios. To conclude, this analysis demonstrates the superiority of PPO in achieving higher EE, although it takes more episodes to converge, making it a promising solution for RIS-assisted near-field MIMO.

In Figure 4, simulation results show the performance of the objective function (i.e., EE) for our proposed method with respect to episodes and compare their performance with the benchmark approaches. It can be seen from the given Figure that the proposed PPO approach provides outstanding performance and surpasses the EE performance significantly over the benchmark approaches. This demonstrates the effectiveness of the proposed PPO's approach suitability in the complex and dynamic environment. Moreover, the action function is evaluated more efficiently when employing the advantage function in PPO, facilitating the agent's efficient learning. This helps the PPO agent to identify and optimize the optimal beamforming and phase shift configurations, thereby improving the objective function (i.e., EE). On the other hand, moderate EE performance is achieved using the DL beam training for near-field. Specifically, CNN is configured based on various layers (e.g., padding and kernel size) to extract the CSI features. This approach primarily relies on pre-collected and offline training, limiting its effectiveness in a dynamic environment. Additionally, this approach is lacking in exploring and adaptively choosing appropriate beamforming and phase shift strategies for complex and dynamic scenarios. Meanwhile, the EE achieved by the traditional approach (i.e., AO) is relatively lower than all other approaches. The reason is that such approaches are effective when solving small and static network problems. However, such approaches always face difficulties when handling high-dimensional and non-linear optimization problems. Finally, the EE achieved by the far field is comparatively lower than all other approaches. This approach does not consider spherical waves and the spatial non-stationary characteristic of near-field beamforming, leading to suboptimal performance. To conclude, the proposed PPO approach achieves higher EE performance on every episode than the other approaches, highlighting the effectiveness of the proposed PPO for the RIS-assisted

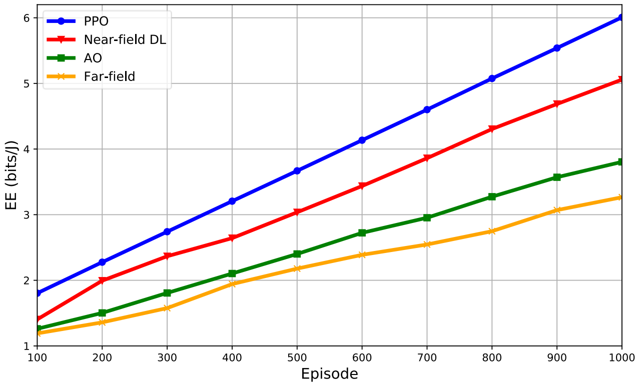

Figure 4. EE performance vs. episode

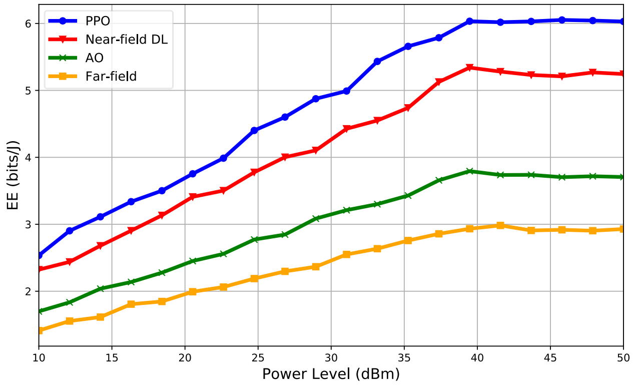

Figure 5. EE vs. power

near-field MIMO networks.

The performance of EE for all the approaches over the different power values is shown in Figure 5. It can be noticed from the results that the EE is increasing smoothly with increasing power values for all approaches. However, the proposed approach outperforms 10-22% then the other approaches on each power value. Moreover, once the power value reaches 40dBm, the performance of all the approaches becomes static with no further increment in the EE performance. This is due to the constraints' contradiction defined in (15b).

Finally, the simulation results for EE versus RIS elements are presented in Figure 6. We observe that the EE constantly improves with increasing the RIS elements for all approaches. It can be noticed that the proposed approach achieves better EE performance by increasing the RIS elements and can effectively

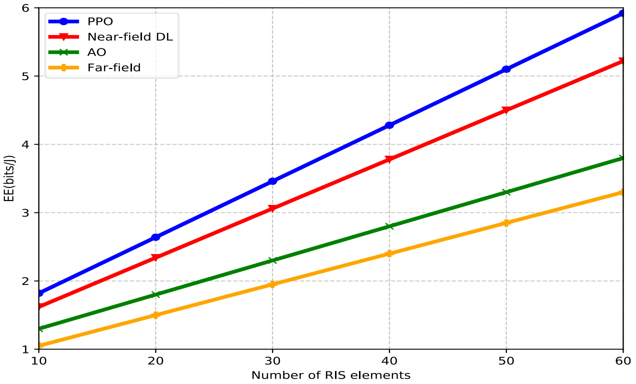

Figure 6. EE versus the number of RIS elements

amplify the larger RIS size. This is because the PPO method finds better police to predict accurately channel characteristics and optimize the beamforming and phase shift to achieve near-optimal performance. On the other hand, the near-field DL approach achieves relatively higher performance than the other approaches. However, such an approach is based on the predefined channel model and limits its ability to adapt to dynamically changing environments or unknown channel characteristics. This limitation may result in suboptimal performance in scenarios where the channel deviates from the predefined model.

5 CONCLUSIONS

We investigated the performance of EE in RIS-assisted MIMO networks considering the near-field scenario. We solved the EE as a joint optimization problem (beamforming and phase shift) by leveraging the angular or distance information embedded in the received signal from BS-RIS and RIS-user links. The optimization problem was non-convex due to the outdated CSI at each time step t and was challenging to solve using the traditional approaches. Thus, we leveraged an on-policy DRL approach (i.e., PPO) that efficiently approximates the optimal solution for beamforming at BS and phase shift at RIS in the formulated problem. The proposed algorithm's effectiveness and the mathematical framework's validity were demonstrated through comprehensive simulation results, and performance was compared with near-field DL and traditional approaches (i.e., near-field AO and far-field approaches). The PPO approach took the benefits of clipping surrogates' function during the training and utilized the clip parameter to restrict the policy update, preventing excessive policy shifts and enhancing training. This enabled PPO to enhance the performance of EE compared to other approaches. Based on the simulation results, the system performance was significantly improved, confirming the advantages of the PPO framework for near-field channels.

ACKNOWLEDGMENTS

This work is funded by Ericsson Canada and the Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] F. A. Pereira De Figueiredo, “An Overview of Massive MIMO for 5G and 6G,” *IEEE Lat. Am. Trans.*, vol. 20, no. 6, pp. 931–940, 2022, doi: 10.1109/TLA.2022.9757375.

[2] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M. S. Alouini, and R. Zhang, “Wireless communications through reconfigurable intelligent surfaces,” *IEEE Access*, vol. 7, no. June 2018, pp. 116753–116773, 2019, doi: 10.1109/ACCESS.2019.2935192.

[3] J. Du, X. Luo, X. Li, M. Zhu, K. M. Rabie, and F. Kara, “Semi-Blind Joint Channel Estimation and Symbol Detection for RIS-Empowered Multiuser mmWave Systems,” *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 362–366, 2023, doi: 10.1109/LCOMM.2022.3212083.

[4] J. An, C. Xu, L. Gan, and L. Hanzo, “Low-Complexity Channel Estimation and Passive Beamforming for RIS-Assisted MIMO Systems Relying on Discrete Phase Shifts,” *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 1245–1260, 2022, doi: 10.1109/TCOMM.2021.3127924.

[5] H. Liu, X. Yuan, and Y. J. A. Zhang, “Matrix-Calibration-Based Cascaded Channel Estimation for Reconfigurable Intelligent Surface Assisted Multiuser MIMO,” *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2621–2636, 2020, doi: 10.1109/JSAC.2020.3007057.

[6] W. K. Ghamry and S. Shukry, “Channel estimation for RIS-aided MIMO systems in MmWave wireless communications with a few active elements,” *Cluster Comput.*, vol. 27, no. 10, pp. 14247–14267, 2024, doi: 10.1007/s10586-024-04627-9.

[7] M. Najafi, V. Jamali, R. Schober, and H. V. Poor, “Physics-Based Modeling and Scalable Optimization of Large Intelligent Reflecting Surfaces,” *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2673–2691, 2021, doi: 10.1109/TCOMM.2020.3047098.

[8] M. Cui and L. Dai, "Channel Estimation for Extremely Large-Scale MIMO: Far-Field or Near-Field?," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2663–2677, 2022, doi: 10.1109/TCOMM.2022.3146400.

[9] S. Ye *et al.*, "Extremely Large Aperture Array (ELAA) Communications: Foundations, Research Advances and Challenges," *IEEE Open J. Commun. Soc.*, vol. 5, no. November, pp. 7075–7120, 2024, doi: 10.1109/OJCOMS.2024.3486172.

[10] S. Chen *et al.*, "Hybrid - Field Full-Dimensional Channel Estimation for Reconfigurable Intelligent Surfaces with Extremely-Large Aperture," *IEEE Wirel. Commun. Netw. Conf. WCNC*, pp. 1–5, 2024, doi: 10.1109/WCNC57260.2024.10571039.

[11] A. Papazafeiropoulos, P. Kourtessis, S. Chatzinotas, D. I. Kaklamani, and I. S. Venieris, "Near-Field Beamforming for Stacked Intelligent Metasurfaces-Assisted MIMO Networks," *IEEE Wirel. Commun. Lett.*, vol. 13, no. 11, pp. 3035–3039, 2024, doi: 10.1109/LWC.2024.3438840.

[12] J. Tian, Y. Han, S. Jin, X. Li, J. Zhang, and M. Matthaiou, "Near-Field Channel Reconstruction in Sensing RIS-Assisted Wireless Communication Systems," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 2, pp. 12223–12238, 2024, doi: 10.1109/TWC.2024.3389026.

[13] M. Rihan, A. Zappone, S. Buzzi, D. Wübben, and A. Dekorsy, "Energy efficiency maximization for active RIS-aided integrated sensing and communication," *Eurasip J. Wirel. Commun. Netw.*, vol. 2024, no. 1, 2024, doi: 10.1186/s13638-024-02346-8.

[14] A. Archi, H. A. Saadi, and S. Mekaoui, "Applications of Deep Reinforcement Learning in Wireless Networks-A Recent Review," *Proc. - 2023 2nd Int. Conf. Electron. Energy Meas. IC2EM 2023*, vol. 1, no. Ic2em, pp. 1–8, 2023, doi: 10.1109/IC2EM59347.2023.10419512.

[15] M. R. Khan, C. L. Zekios, S. Bhardwaj, and S. V. Georgakopoulos, "A Deep Learning Convolutional Neural Network for Antenna Near-Field Prediction and Surrogate Modeling," *IEEE Access*, vol. 12, no. March, pp. 39737–39747, 2024, doi: 10.1109/ACCESS.2024.3377219.

[16] W. Liu, H. Ren, C. Pan, and J. Wang, "Deep Learning Based Beam Training for Extremely Large-Scale Massive MIMO in Near-Field Domain," *IEEE Commun. Lett.*, vol. 27, no. 1, pp. 170–174, 2023, doi: 10.1109/LCOMM.2022.3210042.

[17] G. Jiang and C. Qi, "Near-Field Beam Training Based on Deep Learning for Extremely Large-Scale MIMO," *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 2063–2067, 2023, doi: 10.1109/LCOMM.2023.3289513.

[18] Y. Wang, Z. Gao, S. Chen, C. Hu, and D. Zheng, "Deep Learning–Based Channel Extrapolation and Multiuser Beamforming for RIS-aided Terahertz Massive MIMO Systems over Hybrid-Field Channels," *Intell. Comput.*, vol. 3, pp. 1–20, 2024, doi: 10.34133/icomputing.0065.

[19] J. Nie, Y. Cui, Z. Yang, W. Yuan, and X. Jing, "Near-field Beam Training for Extremely Large-scale MIMO Based on Deep Learning," *IEEE Trans. Mob. Comput.*, vol. PP, no. X, pp. 1–11, 2024, doi: 10.1109/TMC.2024.3462960.

[20] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *arXiv:1707.06347v2*, pp. 1–12, 2017, [Online]. Available: arxiv:1707.06347

[21] S. Lv, Y. Liu, X. Xu, A. Nallanathan, and A. L. Swindlehurst, "RIS-aided Near-Field MIMO Communications: Codebook and Beam Training Design," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 2, pp. 12531–12546, 2024, doi: 10.1109/TWC.2024.3393412.

[22] A. Iqbal, A. Al-Habashna, G. Wainer, G. Boudreau, and F. Bouali, "PPO-BASED ENERGY EFFICIENCY MAXIMIZATION FOR RIS-ASSISTED MULTI-USER MISO SYSTEMS," *2024 IEEE 100th Veh. Technol. Conf.*, pp. 1–6, 2024, doi: 10.1109/VTC2024-Fall63153.2024.10757460.

[23] H. Zhang, S. Ma, Z. Shi, X. Zhao, and G. Yang, "Sum-Rate Maximization of RIS-Aided Multi-User MIMO Systems with Statistical CSI," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 7, pp. 4788–4801, 2023, doi: 10.1109/TWC.2022.3228910.

AMJAD IQBAL is a post-doctoral fellow at the Department of Systems and Computer Engineering at Carleton University, Canada. He received his PhD in Telecommunication from Universiti Tunku Abdul Rahman (UTAR), Malaysia. His research interests include 5G/B5G/6G, RIS, MIMO, and Machine Learning for Wireless Communication. His email address is amjadiqbal3@cunet.carleton.ca.

ALA'A AL-HABASHNA received his Master of Engineering degree from Memorial University of Newfoundland in 2010, and his PhD degree from Carleton University in 2018, both in Electrical and Computer Engineering. Currently, Dr. Al-Habashna is an Adjunct Research Professor at Carleton University and a senior researcher at Statistics Canada, Ottawa, Canada. His current research interests include 5G wireless networks, IoT applications, multimedia communication over wireless networks, discrete-event modeling and simulation, signal detection and classification, cognitive radio systems, and applied machine learning and computer vision. His email is alaaalhabashna@sce.carleton.ca.

GABRIEL WAINER is a Professor in the Department of Systems and Computer Engineering, Carleton University (Ottawa, ON, Canada). His current research interests are related with modeling methodologies and tools, parallel/distributed simulation, and real-time systems. He is a Fellow of SCS. His email is gwainer@sce.carleton.ca. His website is www.sce.carleton.ca/faculty/wainer.

Gary Boudreau is a 5G Systems Architect at Ericsson Canada. His research interests include digital and wireless communications as well as digital signal processing. His email address is gary.boudreau@ericsson.com.