ILLINOIS TECH
College of Computing

Reinforcement Learning Algorithms for a Reconfigurable Intelligent Surfaces Optimization Problem

Authors: Joshua Lin, Ziru Chen

Mentor: Lin Cai *(Associate Professor, IIT/ECE)*

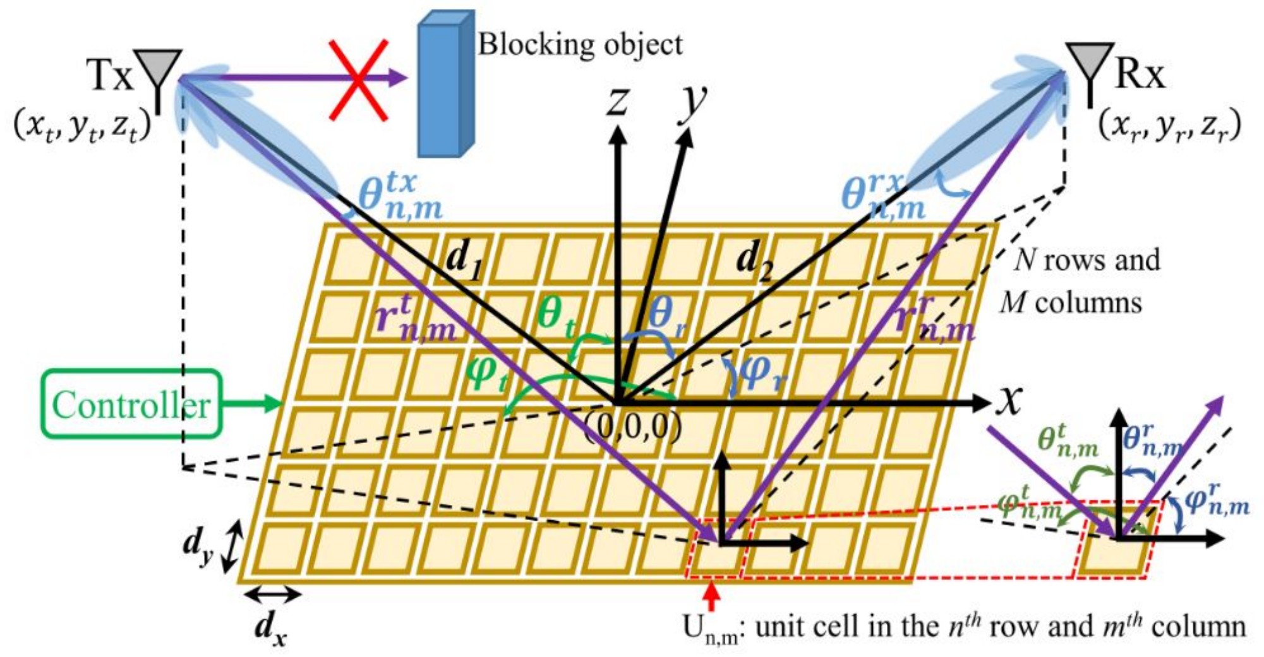
WILLIAM FREMD
HIGH SCHOOL

Abstract

Reconfigurable Intelligent Surfaces (RIS) are used as low cost solutions to enhance wireless communication signals [1]-[2]. They consist of a 2D array of cells capable of adjusting the phase shifts of incident signals in real time to amplify signal strength to a receiver. However, determining the optimal phase shift for each cell is challenging due to the numerous possibilities of large arrays [3]. This poster explores the use of online Reinforcement Learning (RL) algorithms to optimize these RIS coefficients for maximizing the output of the wireless communication system.

Background

- For a signal sent to a RIS surface reflects off its various cells, traveling a different distance to each one.
- The signal then experiences a phase shift and "bounces" off in the direction of the receiver.
- Signals that reach the receiver with the same phase shift are amplified.



[4]

Challenges

- Calculating reward for different distances to allow the environment to be expandable for different distances – Normalize the reward function based on a theoretical maximum strength
- To avoid catastrophic forgetting – reduce rate of learning (0.0001)
- Difficult variables to calculate – Made simplifications such as lambda, G values.

Problem Statement

We attempt to maximize the power received as given by this equation:

Theorem 1: The received signal power in RIS-assisted wireless communications is as follows [4]

$$P_r = P_t \frac{G_t G_r G d_x d_y \lambda^2}{64\pi^3} \times \left| \sum_{m=1-\frac{M}{2}}^{\frac{M}{2}} \sum_{n=1-\frac{N}{2}}^{\frac{N}{2}} \frac{\sqrt{F^{combine}_{n,m}}\,\Gamma_{n,m}}{r^t_{n,m} r^r_{n,m}} e^{\frac{-j2\pi(r^t_{n,m}+r^r_{n,m})}{\lambda}} \right|^2$$

P_r = Power Received
P_t = Power Transmitted
G_t, G_r, G = Antenna Gain (Simplified)
N, M = Maximum rows, columns of the RIS grid
dx, dy = Size of cell: Distance between cells in the x/y directions
$r^t_{n,m}, r^r_{n,m}$ = Distance from cell (n, m) to receiver/transmitter
λ = Fixed Wavelength
$F^{combine}_{n,m}$ = Accounts for the effect of normalized power radiation

$$\Gamma_{n,m} = A * e^{j\phi}$$

Where A is constant and ϕ can be adjusted

Method of Approach

By modeling RIS as arrays of adjustable phase shift coefficients for each cell, we create an optimization problem for the maximum received signal strength. To solve for the optimal coefficients, we employ the RL algorithms Proximal Policy Optimization (PPO) and Advantage Actor-Critic (A2C).

In our environment,

$$G_t, G_r, G, \lambda, F^{combine}_{n,m}, A$$ are held constant.

Our reward function is thus defined as:

$$reward = 10 \cdot \frac{P_r}{P_{r_{max}}} - 5$$

where we have:

$P_{r_{max}}$ = Theoretical Maximum Power Received

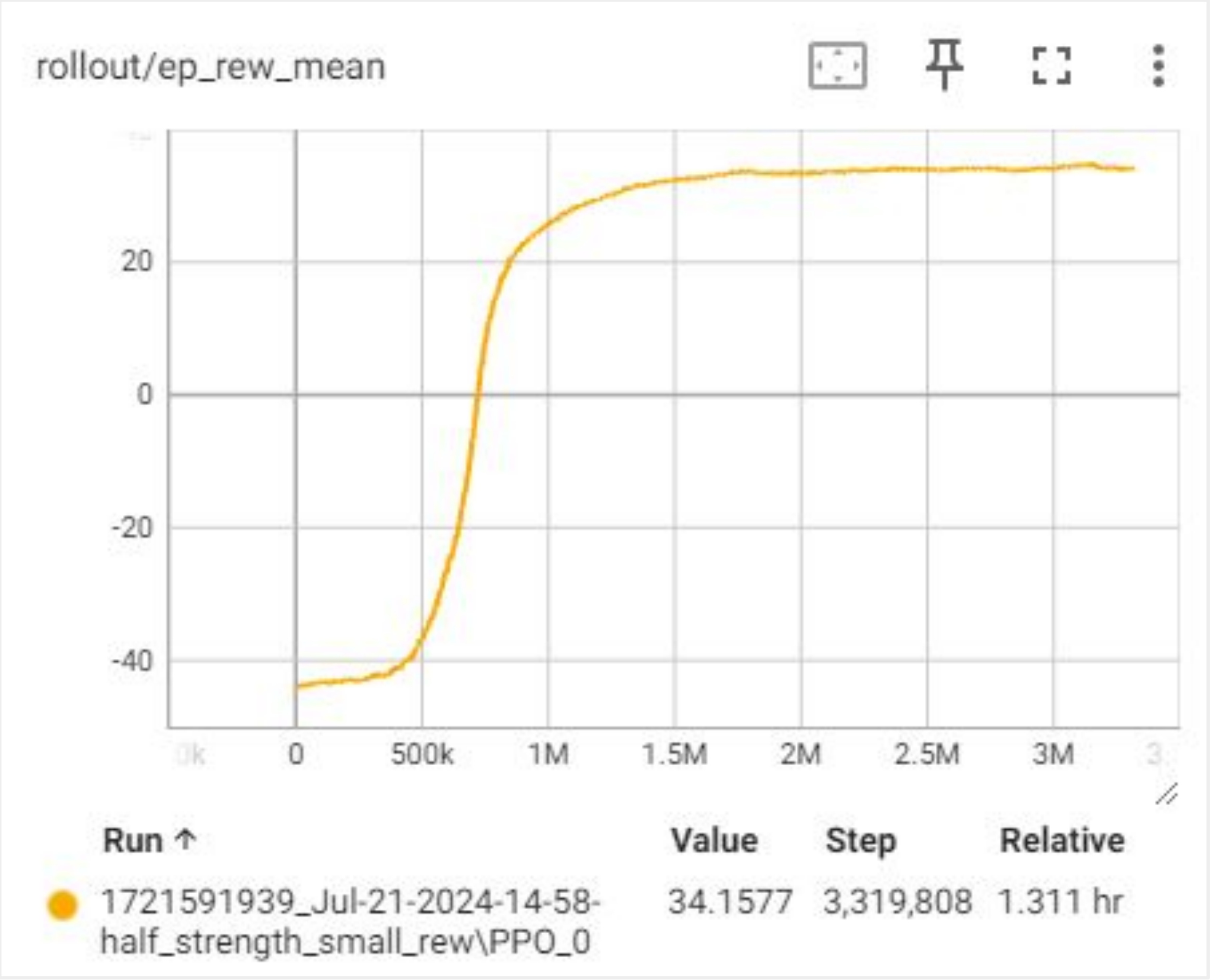
$$P_{r_{max}} = P_t \frac{G_t G_r G d_x d_y \lambda^2}{64\pi^3} \times \left(\sum_{m=1-\frac{M}{2}}^{\frac{M}{2}} \sum_{n=1-\frac{N}{2}}^{\frac{M}{2}} \frac{\sqrt{F^{combine}_{n,m}}\,A}{r^t_{n,m} r^r_{n,m}} \right)^2$$

Therefore, our environment's reward varies from -5 to 5. We run PPO and A2C algorithms with this equation, for 10 step episodes before each reset. 10 step episodes results in an inflation of the returned reward–this implies maximum and minimum function of -50 to 50.

We use a 4x4 grid to model the cells in our surface. Our algorithm will then set a continuous action space for each of the cells.

Results

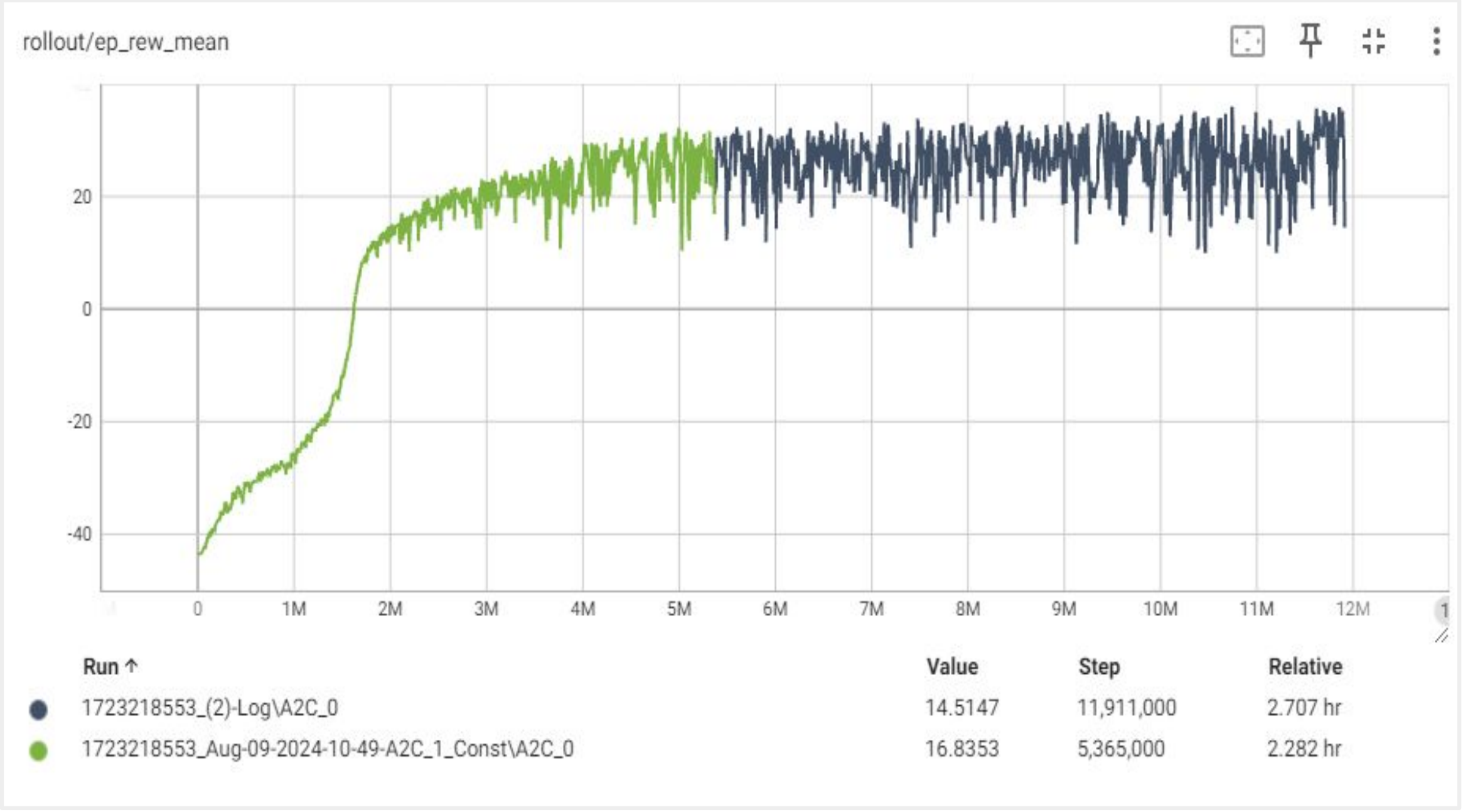
PPO:


rollout/ep_rew_mean

Run ↑	Value	Step	Relative
● 1721591939_Jul-21-2024-14-58-half_strength_small_rew\PPO_0	34.1577	3,319,808	1.311 hr

Both algorithms only run a few million steps before leveling off at a reward of around 30.

A2C:


rollout/ep_rew_mean

Run ↑	Value	Step	Relative
● 1723218553_(2)-Log\A2C_0	14.5147	11,911,000	2.707 hr
● 1723218553_Aug-09-2024-10-49-A2C_1_Const\A2C_0	16.8353	5,365,000	2.282 hr

Conclusion

Running trials of PPO and A2C algorithms shows that both PPO and A2C converges at a maximum of 30, which is only 80% of the maximum reward value which can be returned.

Their failure to maximize the function implies current limitations of these deep learning algorithms to find alternative paths for optimization.

References

[1] Taha, A., Zhang, Y., Mismar, F. B., & Alkhateeb, A. (2020). Deep Reinforcement Learning for Intelligent Reflecting Surfaces: Towards Standalone Operation. IEEE Transactions on Wireless Communications. https://doi.org/10.1109/spawc48557.2020.9154301

[2] Feng, K., Wang, Q., Li, X., & Wen, C. (2020). Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems. IEEE Wireless Communications Letters, 9(5), 745–749. https://doi.org/10.1109/lwc.2020.2969167

[3] Huang, C., Mo, R., & Yuen, C. (2020). Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning. IEEE Journal on Selected Areas in Communications, 38(8), 1839–1850. https://doi.org/10.1109/jsac.2020.3000835

[4] Tang, W., Chen, M. Z., Chen, X., Dai, J. Y., Han, Y., Di Renzo, M., Zeng, Y., Jin, S., Cheng, Q., & Cui, T. J. (2020). Wireless communications with reconfigurable intelligent surface: path loss modeling and experimental measurement. IEEE Transactions on Wireless Communications, 20(1), 421–439. https://doi.org/10.1109/twc.2020.3024887

Real Time Communications
Conference and Expo at Illinois Tech
IEEE International Conference
IEEE RTC 2024 - October 8-9, 2024 - IIT, Chicago, IL, USA

IEEE