Deep Reinforcement Learning Agents for Energy Saving in Open RAN

Gerry Agluba, Jr.
Prospero Naval, Jr.

University of the Philippines - Diliman

May 29, 2025

Introduction

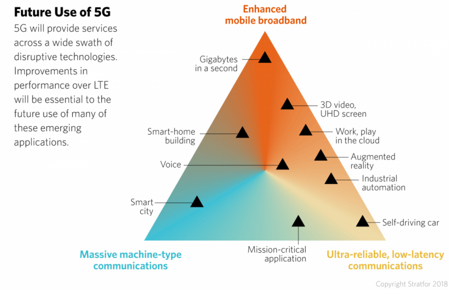

Figure: 5G Use Cases

- The rollout of 5G technology in wireless technology marks a new milestone in the mobile network industry with a new promise on (a) eMBBB, (b) uRLLC and (c) mMTC.

Introduction

- Despite its obvious benefits, 5G technologies comes with new challenges in **Energy Efficiency** [3] primarily because of the increase frequencies, denser network elements and topologies and general increase in transmission and computing resources.
- **Contribution:** Generation of annotated datasets and training agents that can work on multiple RAN scenarios and traffic conditions.

Objective

The main objective of this project is to construct a Deep Reinforcement Learning (DRL) agent that can be deployed as a near-RT RIC controller application (xApps) in an Open RAN environment.

Additionally, this study aims to

- Train different agents based on latest DRL algorithms (Deep Q-Network, PPO)
- Evaluate its performance against a baseline, and well-founded heuristic benchmark

Scope and Limitation

- While the main goal of the study is to ultimately deploy and test the agent in an actual RAN environment. This study was scoped down to using a simulated environment using an open-source RAN simulator (ns3-mmwave-oran). The simulator in its own is limited to simulate a number of base stations (bs) and number of user equipments (ues).
- Moreover, dataset for offline training is scarce in public repository, and thus the agent training (especially in DQN) is only limited to the amount of data observed during explorations.

Methods

Methodology

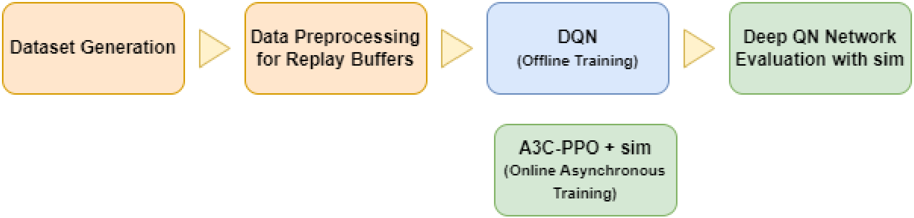

Figure: General RL Pipeline for DQN, A2C-PPO

The general workflow for DRL agents depending on the algorithm used.

- For DQN, Replay Buffers are generated first by series of simulations using random and heuristic policies, before proceeding to offline training and eventually evaluated again wit real simulators.
- On the other hand A2C-PPO is directly trained from the simulators using multiple workers.

Simulation Setup

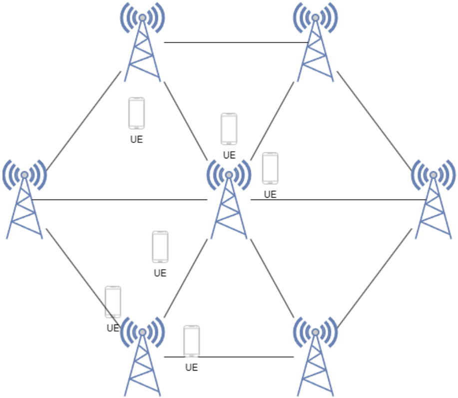

Figure: Simulation Setup

Simulation Setup

The simulated environment is setup as follow

- **7 bs** (6 NR only, 1 NR+LTE)
- **21 ues** uniformly distributed around each bs (3 each) with random walk movement of predefined speed
- Each simulation can be configured over different
  - **traffic scenarios** : *(a) full buffer traffic*, *(b) half nodes in full buffer and half nodes in bursty*, *(c) bursty traffic*, *(d) 0.25 full buffer, 0.25 bursty 3Mbps, 0.25 bursty 0.75Mbps, 0.25 bursty 0.15Mbps*
  - **Data rates**: high/low
  - others ...

State and Action Spaces

Below details the **State Spaces** $\mathcal{S}$ for the agent totalling to **61 states**.

Description	Count
QoS flow volume to transport blocks for downlink	7 (bs) + 1 (agg)
Enegy Saving State Cost	7 (bs)
QoS PDU volume for downlink flows	7 (bs) + 1 (agg)
Radio link failures	7 (bs) + 1 (agg)
Physical Resource Block (PRB) Percent Usage	7x2 (bs)
Ratio of 64QAM transport blocks	7 (bs) + 1 (agg)
EEE KPI	7 (bs)
Zero Count	1

Additionally, the **Action Space** is set $\mathcal{A} = \{0, 1\}^i$ where is i is the number of base station. Each action indicates **0: ES-ON**, **1: ES-OFF** for each base station - totalling to $2^7 = 128$ discrete actions.

Reward Function

The reward function is a function of throughput, energy consumption, coverage and activation cost [2].

$$R = \max_{a} \sum_{t=t_0}^{\infty} \sum_{i=1}^{N} \omega_1 \rho_i(a_i(t)) - \omega_2 \gamma_i(a_i(t)) - \omega_3 \zeta_i(a_i(t)) - \omega_4 \delta_i(a_i(t)) - \omega_2 BsON(a_i(t))$$

subject to $\sum_{j=1}^{k} \omega_i = 1$ and $a_i(t) \in \mathcal{A}$ where $a_i(t)$ is the action for cell i at time t.

Reward Function cont...

1. ρ_i is defined as cell throughput - number of bytes transmitted at the PDCP layer by cell i
2. γ_i is defined as the cell i energy consumption: $\gamma_i(a_i(t)) = EC_i(a_i(t)) \cdot P_{tx,i}$ where EC_i and $P_{tx,i}$ are the total number of PDU transmitted by cell i and its associated transmit power
3. $BsON(a_i(t))$ is defined as the number of active cells at time t
4. $\zeta_i(\cdot)$ and $\delta_i(\cdot)$ are defined respectively as the UE count in Radio Link Failure (RLF) and the cost to activate a single cell

The reward function aims to **maximize the throughput** while **minimizing the energy consumption** and **activation cost**, **decreasing link failures** (increasing coverage).

Dataset Generation for DQN

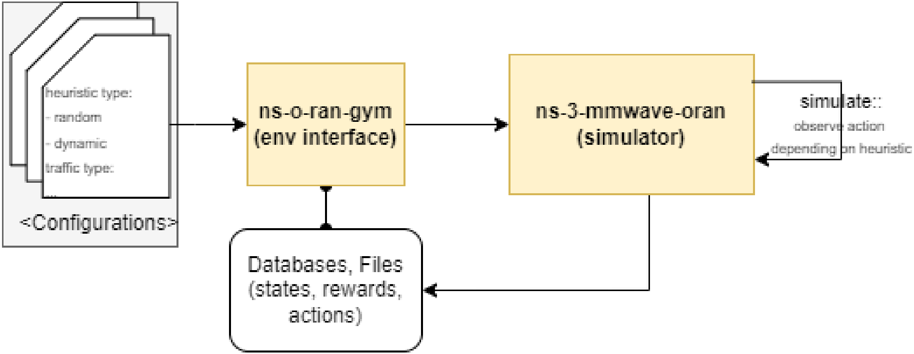

Figure: Generating Data through Simulation

80767 (from 1007 truncated simulated episodes) observations from simulations are collected. Simulations were run from different traffic conditions, data rates, different kinds of heuristic, RNG, etc. using an open-source **ns-3-mmwave-oran** simulator interfaced with gym-capable

Agents

The following agents are sets of agents trained and evaluated in this study.

Agent	Description
DQN	2-layer DQN (with deterministic policy)
PPO	A2C-PPO (8 worker, base)
PPO-2	A2C-PPO (8 worker, w/ reward penalty for similar action)
PPO-3	A2C-PPO (8 worker, w/ action holdout)
PPO-MB-1	A2C-PPO (8 worker, multibinary action)

The agents are compared to the **dynamic** sleeping policy described by Salem[1].

Results

Agent Performance: Throughput & Energy Consumption

Table: Throughput vs Energy Consumption

Agent	**Throughput [Mbps]** (% rel. with benchmark)	**Energy Consumption [W]** (% rel. with benchmark)
benchmark-dynamic	7.59 (+ 0%)	2573.57(+ 0%)
dqn	**11.43(+ 50.59%)** *	2639.18 (+ 2.55%) *
ppo-1	5.03 (- 33.72%)	**2390.49 (- 7.11%)**
ppo-2	4.86 (- 35.96%)	**2348.13 (- 8.76%)**
ppo-3	**5.71 (- 24.76%)**	2418.50 (- 6.03 %)
ppo-mb-1	**5.76 (- 24.11%)**	**2331.05 (- 9.42%)**

* Unreliable result. Deeper investigation suggested that DQN actions are behaves like an always-on policy.

Agent Performance: Throughput & Energy Consumption

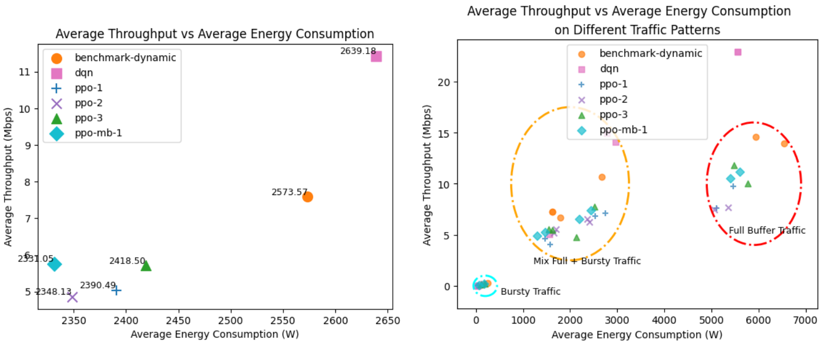

Figure: Scatterplot of Agent Throughput vs Energy Consumption

Agent Performance: Radio Link Failure & ON Cost

Table: Radio Link Failure(RLF) and ON Cost

Agent	**Radio Link Failure**	**ON Cost**
benchmark-dynamic	1.30	0.33
dqn	**0.53***	**0.02***
ppo-1	2.52	2.80
ppo-2	2.49	2.80
ppo-3	**2.11**	1.75
ppo-mb-1	3.00	**1.23**

Agent Performance: Radio Link Failure & ON Cost

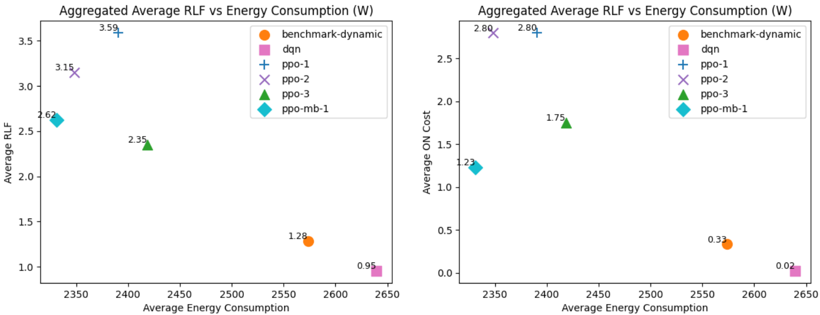

Figure: Scatterplot of Agent RLF, ON-COST vs Energy Consumption

Variance Control: Ideal

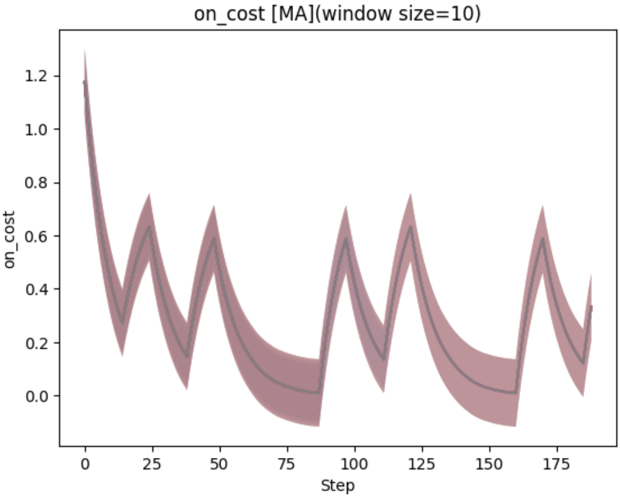

Figure: Ideal Control Scenario

Variance Control: Bang Bang Control!

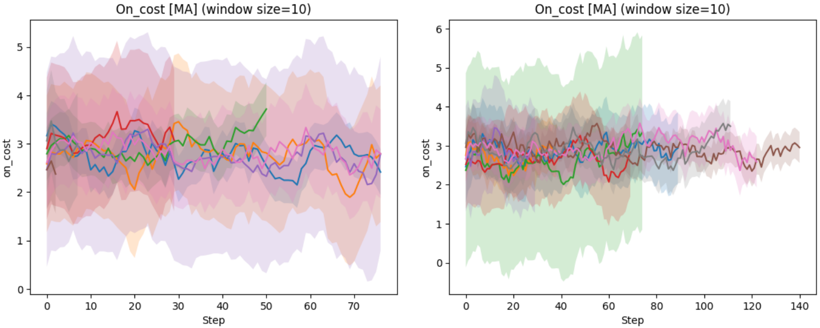

Figure: Bang Bang Control

Variance Control: Mitigations

Mitigate abrupt action changes by imposing action holdout and using multi-binary actions instead.

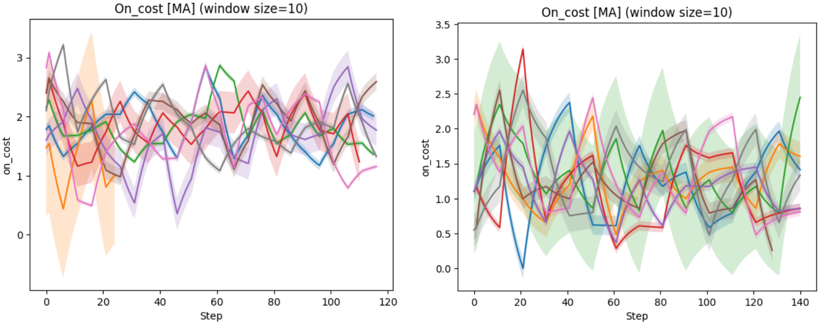

Figure: Bang Bang Control Mitigations

Discussion

1. DQN agent throughput performance was impressive, but a deeper investigation of its policy actions proved in unreliable performance.
2. Disregarding DQN, **ppo-mb-1** comes in with average throughput of 5.76 Mbps, closely followed by ppo-3. **ppo-mb-1** also had the highest energy saving (-9.42%) relative to benchmark-dynamic followed by ppo-2, ppo-1. The throughput vs energy consumption tradeoffs is also reflected to the different traffic conditions 5.
3. **ppo-3** has the lowest RLF followed by ppo-2, ppo-3. On the other hand ppo-mb-1, has had the lowest BS ON cost inching over ppo-3.
4. **ppo-3** and **ppo-mb-1** shows less action variance compared to initial versions of ppo agents: ppo-1, ppo-2.

Conclusion

Conclusion and Future Work

This study was able to construct DRL agents: **(ppo-3, ppo-mb-1)** that were able to **improve energy saving** of RAN **system sacrificing minimal throughput**. We were able to **evaluate agent performance across different traffic scenarios** and **resolve variance control** problems by integrating modification in the agent actor network (discrete to multi-binary) and imposing action holdouts.

It is recommended to evaluat on more realistic traffic scenarios and pseudo-live RAN deployments. The agents are not dynamic with regards to the number of base stations it can support, Future work involves adding a more dynamic state representations robust to any number of base stations.

References

References

- M. Polese, L. Bonati, S. D'Oro, S. Basagni and T. Melodia, "Understanding O-RAN: Architecture, Interfaces, Algorithms, Security, and Research Challenges," in IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 1376-1411, Secondquarter 2023, doi: 10.1109/COMST.2023.3239220.
- https://www.ericsson.com/en/openness-innovation/open-ran-explained
- Shurdi, Olimpjon & Ruci, Luan & Biberaj, Aleksandër & Mesi, Genci. (2021). 5G Energy Efficiency Overview. European Scientific Journal ESJ. 17. 10.19044/esj.2021.v17n3p315.
- Zappone, Alessio. (2016). A Survey of Energy-Efficient Techniques for 5G Networks and Challenges Ahead. IEEE Journal on Selected Areas in Communications. 34. 10.1109/JSAC.2016.2550338.

References

Fatma Ezzahra Salem. Management of advanced sleep modes for energy-efficient 5G networks. Networking and Internet Architecture [cs.NI]. Institut Polytechnique de Paris, 2019.

T. M. Ho, K. -K. Nguyen and M. Cheriet, "Energy Efficiency Learning Closed-Loop Controls in O-RAN 5G Network," GLOBECOM 2023 - 2023 IEEE Global Communications Conference, Kuala Lumpur, Malaysia, 2023, pp. 2748-2753, doi: 10.1109/GLOBECOM54140.2023.10437790

S. Ryoo, J. Jung and R. Ahn, "Energy efficiency enhancement with RRC connection control for 5G new RAT," 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 2018, pp. 1-6, doi: 10.1109/WCNC.2018.837711

References

Zhang, J., Zhang, X., Imran, M., Evans, B., and Wang, W. (2017) Energy Efficiency Analysis of Heterogeneous Cache-enabled 5G Hyper Cellular Networks. In: IEEE Globecom 2016, Washington, DC, USA, 04-08 Dec 2016, ISBN 9781509013289 (doi:10.1109/GLOCOM.2016.7841790)

M. Bordin et al., "Design and Evaluation of Deep Reinforcement Learning for Energy Saving in Open RAN," 2025 IEEE 22nd Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2025, pp. 1-6, doi: 10.1109/CCNC54725.2025.10976108.

A. Lacava, M. Bordin, M. Polese, R. Sivaraj, T. Zugno, F. Cuomo, and T. Melodia. "ns-O-RAN: Simulating O-RAN 5G Systems in ns-3", Proceedings of the 2023 Workshop on ns-3 (2023), DOI:10.1145/3592149.3592161

References

- Lacava, Andrea and Pietrosanti, Tommaso and Polese, Michele and Cuomo, Francesca and Melodia, Tommaso. Enabling Online Reinforcement Learning Training for Open RAN (2024). pp. 577-582 2024 IFIP Networking Conference (IFIP Networking).

- Pritz, Ma and Leung. Jointly-Learned State-Action Embedding for Efficient Reinforcement Learning. CIKM '21: Proceedings of the 30th ACM International Conference on Information & Knowledge Management pp. 1447 - 1456 https://doi.org/10.1145/3459637.3482357

- Fujimoto, et.al. For SALE State-Action Representation Learning for Deep Reinforcement Learning (2023). https://arxiv.org/abs/2306.02451
  stable-baseline-3 Raffin, Hill, Gleave and Kanervisto. Stable-Baselines3: Reliable Reinforcement Learning Implementations (2021). http://jmlr.org/papers/v22/20-1364.html.

Deep Reinforcement Learning Agents for Energy Saving in Open RAN

Gerry Agluba, Jr.
Prospero Naval, Jr.

University of the Philippines - Diliman

May 29, 2025