Pathways to Trustworthy Reinforcement Learning: *Generalization Strategies for Next-Generation Networks*

Ahmad Nagib
School of Computing, Queen's University

Workshop on Advances in Telecommunications Research, Kingston, Canada
June 16, 2025

Outline

1. Why RL for Next-Gen Wireless Networks
2. Practical Challenges of Reinforcement Learning
3. RL Generalization Strategies
4. Open Research Challenges and Future Directions
5. RL Resources
6. Acknowledgments and Q&A

Why RL for Next-Gen Wireless Networks?

RL Fundamentals

- **Formal Definition of RL**[1]:
  - RL is formulated as a Markov Decision Process (MDP) defined by a tuple $(\mathcal{S}, \mathcal{A}, P, R, \gamma)$, where:
    - $\mathcal{S}$: set of states
    - $\mathcal{A}$: set of actions
    - $P(s'|s, a)$: transition probability
    - $R(s, a)$: reward function
    - $\gamma \in [0, 1)$: discount factor
  - The objective is to find a policy $\pi(a|s)$ that maximizes the expected cumulative reward:

$$J(\pi) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right] \tag{1}$$

[1] R. S. Sutton, A. G. Barto, *et al.*, *Reinforcement learning: An introduction*. MIT press Cambridge, 2018, vol. 2

Basic Reinforcement Learning (RL) Interactions

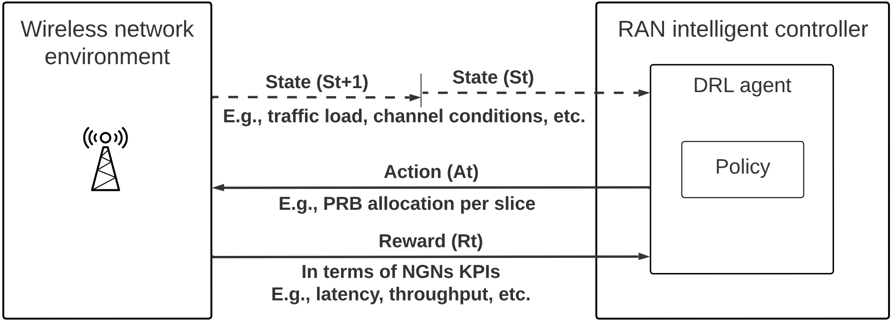

Wireless network environment
RAN intelligent controller
DRL agent
Policy
State (St+1)
State (St)
E.g., traffic load, channel conditions, etc.
Action (At)
E.g., PRB allocation per slice
Reward (Rt)
In terms of NGNs KPIs
E.g., latency, throughput, etc.

Figure: Basic interactions between a DRL agent and the network environment[2].

[2] A. M. Nagib, H. Abou-Zeid, and H. S. Hassanein, "Safe and accelerated deep reinforcement learning-based o-ran slicing: A hybrid transfer learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 310–325, 2024. DOI: 10.1109/JSAC.2023.3336191

Why Reinforcement Learning for Network Optimization?

- **Seamless Integration with Network Control**:
  - RL naturally fits the feedback loop of network operations.
  - Adapts to operator goals and policies.

Why Reinforcement Learning for Network Optimization?

- **Seamless Integration with Network Control**:
  - RL naturally fits the feedback loop of network operations.
  - Adapts to operator goals and policies.
- **Towards Autonomous Networks**:
  - Capable of real-time decision-making in complex environments.
  - Does not require full knowledge of the network system.

Why Reinforcement Learning for Network Optimization?

- **Seamless Integration with Network Control**:
  - RL naturally fits the feedback loop of network operations.
  - Adapts to operator goals and policies.
- **Towards Autonomous Networks**:
  - Capable of real-time decision-making in complex environments.
  - Does not require full knowledge of the network system.
- **Industry Momentum**:
  - Standard bodies and vendors are promoting RL[3,4].
  - Growing recognition of RL's potential in NGWNs.

[3] M. Tsampazi, S. D'Oro, M. Polese, *et al.*, "A comparative analysis of deep reinforcement learning-based xapps in o-ran," in *IEEE Global Communications Conference (GLOBECOM)*, 2023, pp. 1638–1643. DOI: 10.1109/GLOBECOM54140.2023.10437367

[4] T. E. Blog, *Bringing reinforcement learning solutions to action in telecom networks*, https://www.ericsson.com/en/blog/2022/3/reinforcement-learning-solutions, [Accessed 22-01-2024], 2022

Why Reinforcement Learning for Network Optimization?

- **Seamless Integration with Network Control**:
  - RL naturally fits the feedback loop of network operations.
  - Adapts to operator goals and policies.
- **Towards Autonomous Networks**:
  - Capable of real-time decision-making in complex environments.
  - Does not require full knowledge of the network system.
- **Industry Momentum**:
  - Standard bodies and vendors are promoting RL[3,4].
  - Growing recognition of RL's potential in NGWNs.

However, deploying RL in real-world networks comes with significant challenges...

[3] M. Tsampazi, S. D'Oro, M. Polese, *et al.*, "A comparative analysis of deep reinforcement learning-based xapps in o-ran," in *IEEE Global Communications Conference (GLOBECOM)*, 2023, pp. 1638–1643. DOI: 10.1109/GLOBECOM54140.2023.10437367

[4] T. E. Blog, *Bringing reinforcement learning solutions to action in telecom networks*, https://www.ericsson.com/en/blog/2022/3/reinforcement-learning-solutions, [Accessed 22-01-2024], 2022

Practical Challenges of Reinforcement Learning

Practical Challenges of Reinforcement Learning

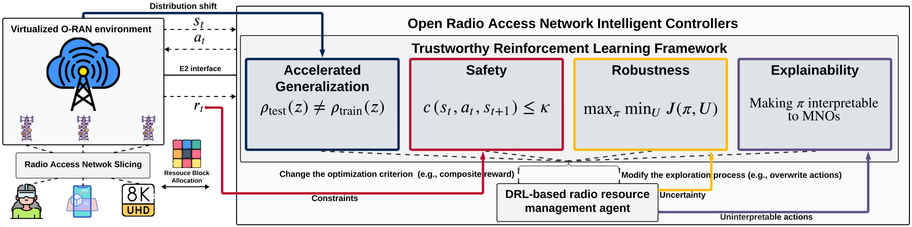

Distribution shift
Virtualized O-RAN environment
s_t
a_t
E2 interface
r_t
Radio Access Netwok Slicing
Resouce Block Allocation
8K UHD
Open Radio Access Network Intelligent Controllers
Trustworthy Reinforcement Learning Framework
Accelerated Generalization
$\rho_{\text{test}}(z) \neq \rho_{\text{train}}(z)$
Safety
$c(s_t, a_t, s_{t+1}) \leq \kappa$
Robustness
$\max_\pi \min_U J(\pi, U)$
Explainability
Making π interpretable to MNOs
Change the optimization criterion (e.g., composite reward)
Modify the exploration process (e.g., overwrite actions)
Constraints
DRL-based radio resource management agent
Uncertainty
Uninterpretable actions

Figure: Dimensions to be addressed for trustworthy DRL for NGWNs[5,6]

[5] A. M. Nagib, "A trustworthy deep reinforcement learning framework for slicing in next-generation open radio access networks," Ph.D. dissertation, School of Computing, Queen's University, 2024

[6] M. Xu, Z. Liu, P. Huang, *et al.*, "Trustworthy reinforcement learning against intrinsic vulnerabilities: Robustness, safety, and generalizability," *arXiv preprint arXiv:2209.08025*, 2022

RL Generalization Strategies

Challenges of Deploying DRL in NGWNs: Ungeneralizable Algorithms

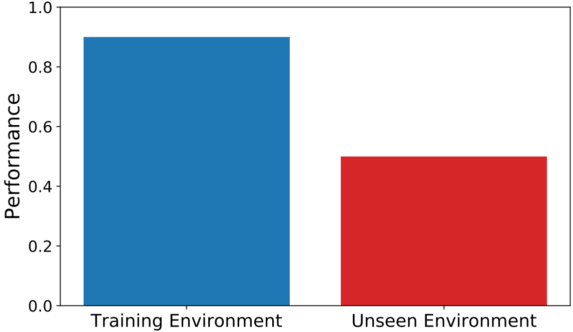


Figure: Challenges in Generalizing from Simulation to Real-World Environments

- Simulation environments often simplify real-world dynamics.
- DRL models may fail to adapt to unforeseen deployment conditions.

Challenges of Deploying DRL in NGWNs: Slow Convergence

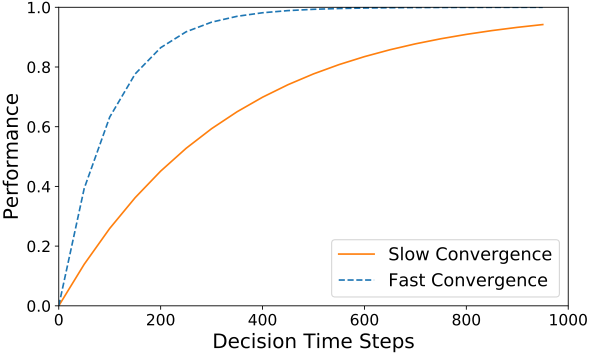
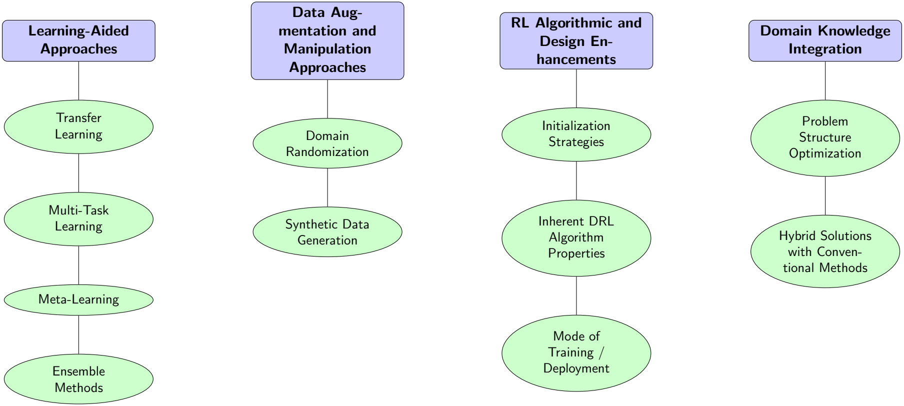


Figure: Challenges in Generalizing from Simulation to Real-World Environments

- DRL models may fail to adapt to unforeseen deployment conditions **quickly**.

Strategies to Enhance DRL Generalization


Learning-Aided Approaches
Transfer Learning
Multi-Task Learning
Meta-Learning
Ensemble Methods
Data Augmentation and Manipulation Approaches
Domain Randomization
Synthetic Data Generation
RL Algorithmic and Design Enhancements
Initialization Strategies
Inherent DRL Algorithm Properties
Mode of Training / Deployment
Domain Knowledge Integration
Problem Structure Optimization
Hybrid Solutions with Conventional Methods

Learning-Based Approaches

Policy Transfer

What Does Transferring a Policy Mean?

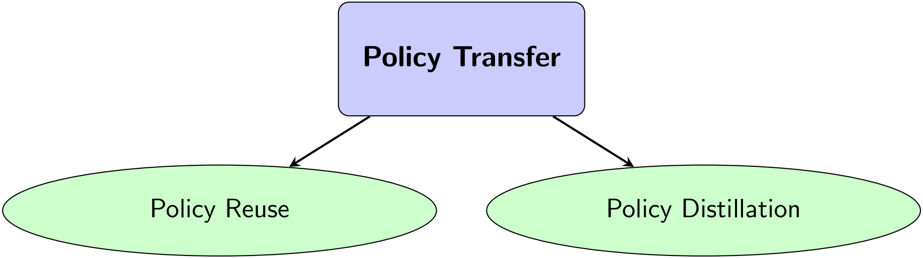
- A policy encodes knowledge about how to act in an environment.
- In Deep RL, policies are typically represented by neural networks:

$$\pi(a \mid s; \theta), \text{ where } \theta \text{ are the network parameters.}$$

- Policy transfer means transferring these learned parameters (or a portion of them) to the target task, potentially with modifications.
- Policy transfer can also be performed by using the output/actions of expert policies to guide the agent in learning a new policy.

[6] Z. Zhu, K. Lin, A. K. Jain, *et al.*, "Transfer learning in deep reinforcement learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 13 344–13 362, 2023. DOI: 10.1109/TPAMI.2023.3292075

Policy Transfer Strategies



Policy Transfer

Policy Reuse

Policy Distillation

$$\pi(\kappa) \;=\; \arg\max_{a} \max_{i} \hat{Q}^{\pi_i}(\kappa, a)$$

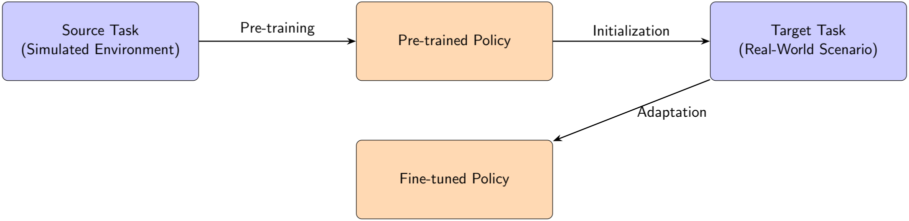
$$\min_L \mathbb{E}_{\tau \sim \pi_E} \left[\sum_{t=1}^{|\tau|} \nabla_L \mathcal{H}^{\times} \left(\pi_E \left(\tau_t \right) \mid \pi_L \left(\tau_t \right) \right) \right]$$

[6] Z. Zhu, K. Lin, A. K. Jain, *et al.*, "Transfer learning in deep reinforcement learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, pp. 13 344–13 362, 2023. DOI: 10.1109/TPAMI.2023.3292075

Policy Reuse: Deployment Examples

1. **Initialization with Expert Policy[7]:**

$$\pi_{\text{learner}}(t=0) = \pi_{\text{expert}}(t=N)$$



Basic Policy Reuse: Initialize policy for a new target task with a pre-trained policy and then fine-tune it with interactions with the target environment.

[7] A. M. Nagib, H. Abou-Zeid, and H. S. Hassanein, "Accelerating reinforcement learning via predictive policy transfer in 6g ran slicing," *IEEE Transactions on Network and Service Management*, vol. 20, no. 2, pp. 1170–1183, 2023. DOI: 10.1109/TNSM.2023.3258692

Policy Reuse Example in Network Slicing

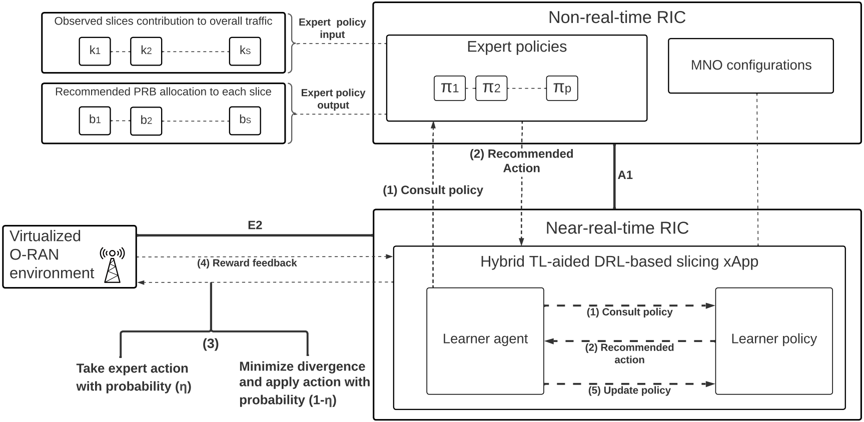
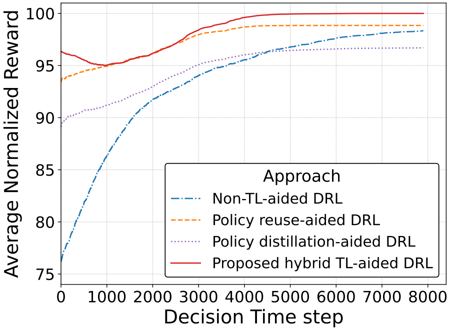


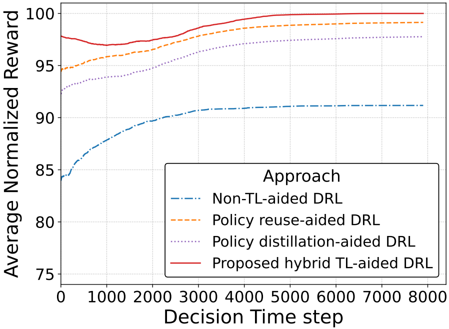
Figure: The policy transfer-aided O-RAN system architecture.

[7] A. M. Nagib, H. Abou-Zeid, and H. S. Hassanein, "Safe and accelerated deep reinforcement learning-based o-ran slicing: A hybrid transfer learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 310–325, 2024. DOI: 10.1109/JSAC.2023.3336191

Similar Traffic in Test: Good performance of policy reuse



(a)

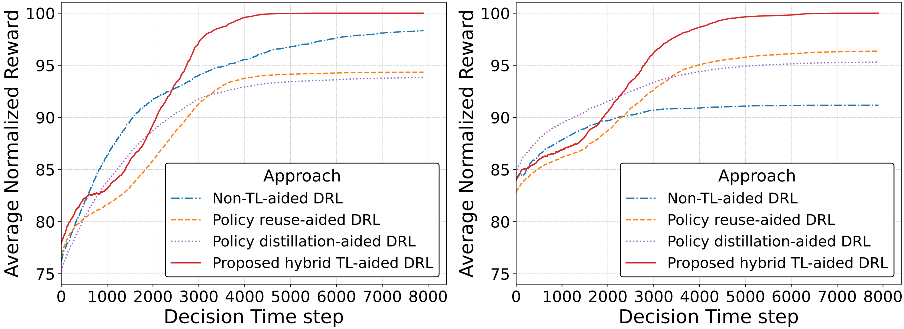


(b)

Figure: Reward convergence performance guided by an expert policy trained using a similar traffic pattern (average of best 64 runs).

[7] A. M. Nagib, H. Abou-Zeid, and H. S. Hassanein, "Safe and accelerated deep reinforcement learning-based o-ran slicing: A hybrid transfer learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 310–325, 2024. DOI: 10.1109/JSAC.2023.3336191

Different Traffic in Test: Poor performance of policy reuse



(a) (b)

Figure: Reward convergence performance guided by an expert policy trained using a different traffic pattern (average of best 64 runs).

[7] A. M. Nagib, H. Abou-Zeid, and H. S. Hassanein, "Safe and accelerated deep reinforcement learning-based o-ran slicing: A hybrid transfer learning approach," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 310–325, 2024. DOI: 10.1109/JSAC.2023.3336191

Learning-Based Approaches

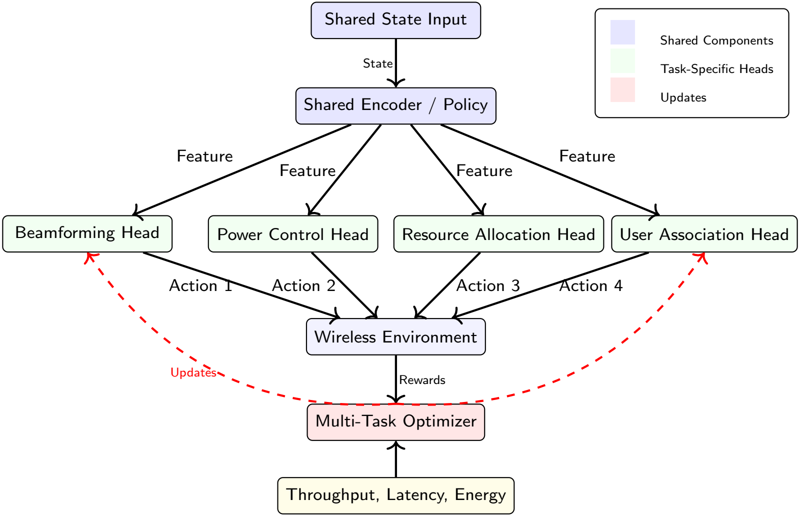
Multi-Task Reinforcement Learning

Multi-Task Reinforcement Learning

- **Concept**:
  - Train an agent across multiple tasks to learn a generalized policy.
- **Method**:
  - Agent learns multiple tasks simultaneously, typically using a shared network architecture for parts of the policy/value function and task-specific components for others.
  - Learn via joint optimization over all tasks.
- **Benefit**:
  - Improves generalization by using domain information from related tasks or across network configurations (e.g. sizes, topologies, or traffic patterns).
  - Reduces costly per-task or per-scenario retraining

[7] N. Vithayathil Varghese and Q. H. Mahmoud, "A survey of multi-task deep reinforcement learning," *Electronics*, vol. 9, no. 9, 2020, ISSN: 2079-9292. DOI: 10.3390/electronics9091363. [Online]. Available: https://www.mdpi.com/2079-9292/9/9/1363

Multi-Task RL



Multi-Task DRL Example for Dynamic MAC Scheduling

- **Task definition:** Network sizes and traffic

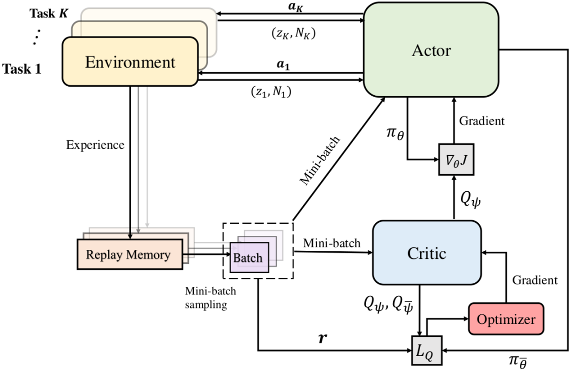


Figure: Illustration of Multi-Task Deep RL in dynamic MAC scheduling[8].

[8] Z. Chen, X. Sun, Y. Jin, *et al.*, "Multi-task reinforcement learning-based multiple access for dynamic wireless networks," *IEEE Transactions on Mobile Computing*, pp. 1–15, 2025. DOI: 10.1109/TMC.2025.3559676

Multi-Task DRL Example for Slicing & Routing

- **Task definition:** Route flows in each network slice

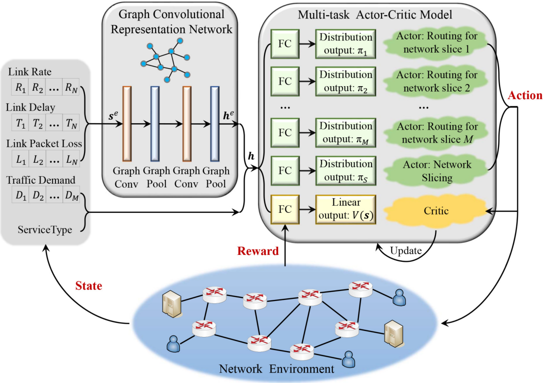


Figure: Architecture of GCN-powered MTDRL model.[9]

[9] T. Dong, Z. Zhuang, Q. Qi, *et al.*, "Intelligent joint network slicing and routing via gcn-powered multi-task deep reinforcement learning," *IEEE Transactions on Cognitive Communications and Networking*, vol. 8, no. 2, pp. 1269–1286, 2022. DOI: 10.1109/TCCN.2021.3136221

Learning-Based Approaches

Meta-Learning

Meta-Learning for DRL Generalization

- Also known as **learning to learn**
- **Definition**: Instead of learning a specific task, learn patterns from various tasks to quickly adapt to new environments.
- There are many approaches to do this each with different objectives on how quickly to adapt, how much generalization is needed, and the meta learning process[10].
- Model-agnostic meta learning (MAML) is one of the popular approaches.
- **Characteristics of Meta-Learning**:
  - Rapid adaptation to new environments/tasks and uses limited data "few shot" in target tasks.
  - More intensive training across multiple tasks.
  - Generally, more complex training process and architectures.

[10] J. Beck, R. Vuorio, E. Z. Liu, *et al.*, "A survey of meta-reinforcement learning," *arXiv preprint arXiv:2301.08028*, 2023

Meta-Learning for DRL Generalization: MAML

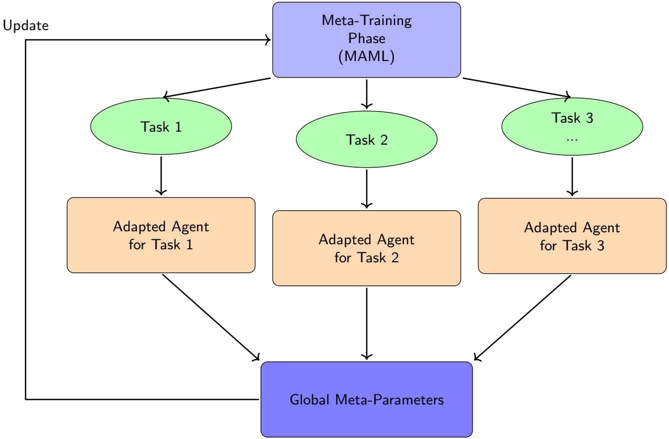


Figure: Meta reinforcement learning conceptual example using Model-Agnostic Meta Learning (MAML)

Meta RL Example for Adaptive Beamforming

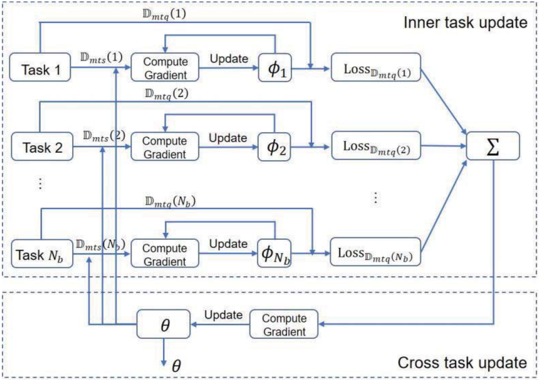


Figure: Workflow of the meta-learning algorithm for beamforming adaptation[11].

[11] Y. Yuan, G. Zheng, K.-K. Wong, *et al.*, "Transfer learning and meta learning-based fast downlink beamforming adaptation," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1742–1755, 2021. DOI: 10.1109/TWC.2020.3035843

Adaptive Beamforming Example: Meta RL vs Transfer RL

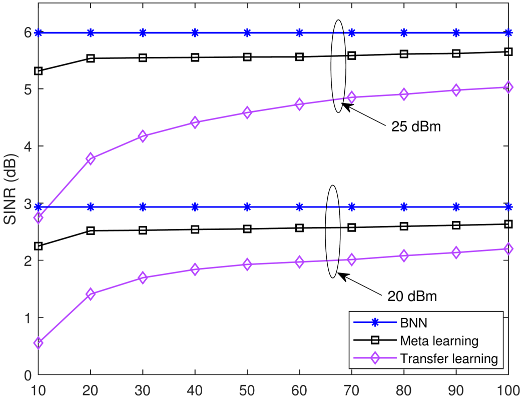


Figure: Comparison of fine-tuning samples for meta and transfer learning.[12]

[12] Y. Yuan, G. Zheng, K.-K. Wong, *et al.*, "Transfer learning and meta learning-based fast downlink beamforming adaptation," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1742–1755, 2021. DOI: 10.1109/TWC.2020.3035843

Comparison: Policy Transfer vs Multi-Task vs Meta-RL[13,14]

Criterion	**Policy Transfer**	**Multi-Task RL**	**Meta-RL**
Deployment Adaptability	Medium (requires similarity)	Medium-High (shared policy)	High (fast adaptation)
Sample Efficiency (Deployment)	High (when similar)	Medium	High
Training Complexity	Low	Medium	High (meta-optimization overhead)
Risk of Negative Transfer	Medium-High	Medium (task interference risk)	Low
Best For	Similar environments with minimal changes	Learning a shared representation for known task distributions	Rapid adaptation to unseen but related tasks

[13] M. Zhao, P. Abbeel, and S. James, "On the effectiveness of fine-tuning versus meta-reinforcement learning," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, *et al.*, Eds., vol. 35, 2022, pp. 26 519–26 531

[14] T. Yu, D. Quillen, Z. He, *et al.*, "Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning," in *Proceedings of the Conference on Robot Learning*, ser. Proceedings of Machine Learning Research, vol. 100, 2020, pp. 1094–1100

Open Research Challenges and Future Directions

Open Challenges in RL-based Wireless Networks

- **Generalization in not a First-Class Goal:** Explicit research focusing purely on few-shot adaptation to unseen wireless environments is less common.
- **Generalization to Out-of-Distribution (OOD) Tasks:** Many wireless papers evaluate on variations of the same task (e.g., different channel conditions, number of users, QoS weights). Achieving broad generalization to diverse, unseen, and out-of-distribution tasks in real deployments is needed.
- **Overfitting to Latest Conditions:** Current approaches are not robust to sequential domain shifts.

Future Perspectives toward Generalizable RL for 6G

- Developing wireless **benchmark challenges** are essential to foster reproducible research that builds on the collective progress of the wireless research community.
  - Foster a culture where **limitations of AI** are encouraged and reported as challenges for others to pursue.
- **Industrial** collaboration to better understand and model the challenges of generalization and RL trustworthiness in general.
- Towards **Foundation** "Generalist" 6G DRL policies:
  - Move beyond parametric variations toward tasks involving qualitatively distinct scenarios.
  - **Continual learning**[15] that is sample efficient.
  - Combining transfer, multi-task and meta-learning for quick adaptation.

[15] M. Caccia, J. Mueller, T. Kim, *et al.*, "Task-agnostic continual reinforcement learning: In praise of a simple baseline," , 2022

RL Resources

RL Resources: Concepts

- **Reinforcement Learning: An Introduction**
  `incompleteideas.net/book/RLbook2020.pdf`
- **RL Theory Seminars:**
  `sites.google.com/view/rltheoryseminars`
- **Safe Reinforcement Learning Online Seminars:**
  `sites.google.com/view/saferl-seminar`
- **Mila Tea Talks:**
  `sites.google.com/lisa.iro.umontreal.ca/tea-talks`
- **Reinforcement Learning Specialization on Coursera**
  `coursera.org/specializations/reinforcement-learning`
- **Reinforcement Learning Mailing List:**
  `groups.google.com/g/rl-list`

RL Resources: Specific to Wireless Networks

- **Single and Multi-Agent Deep Reinforcement Learning for AI-Enabled Wireless Networks: A Tutorial:** `ieeexplore.ieee.org/document/9372298`
- **Ericsson Blog Series on RL:** `ericsson.com/en/blog/2023/11/reinforcement-learning`
- **List of RL Environments for Wireless Networks:** `github.com/ahmadnagib/wireless-rl-envs`

RL Resources: Tools

- **Denny Britz's RL Repository:** github.com/dennybritz/reinforcement-learning
- **MinimalRL-PyTorch:** github.com/seungeunrho/minimalRL
- **Tools for RL in Python** https://neptune.ai/blog/the-best-tools-for-reinforcement-learning-in-python

Generalizable RL Resources

- **Meta-World:** `github.com/Farama-Foundation/Metaworld`
- **Garage:** `github.com/rlworkgroup/garage`
- **RLlib:** `github.com/ray-project/ray/tree/master/rllib`
- **d3rlpy:** `github.com/takuseno/d3rlpy`
- **Quantifying Generalization in RL:** `github.com/openai/coinrun`

Q&A and Acknowledgments

Acknowledgments and Q&A

- This work was done with support from **Dr. Hatem Abou-Zeid** and **Dr. Hossam Hassanein**.
- We encourage community involvement in building Trustworthy RL methods for next-generation wireless networks.
- Reach out to explore opportunities for collaborative research and development.

Thank you for your attention!

Ahmad Nagib