Emergency Control in Large-Scale Power System using Deep Reinforcement Learning

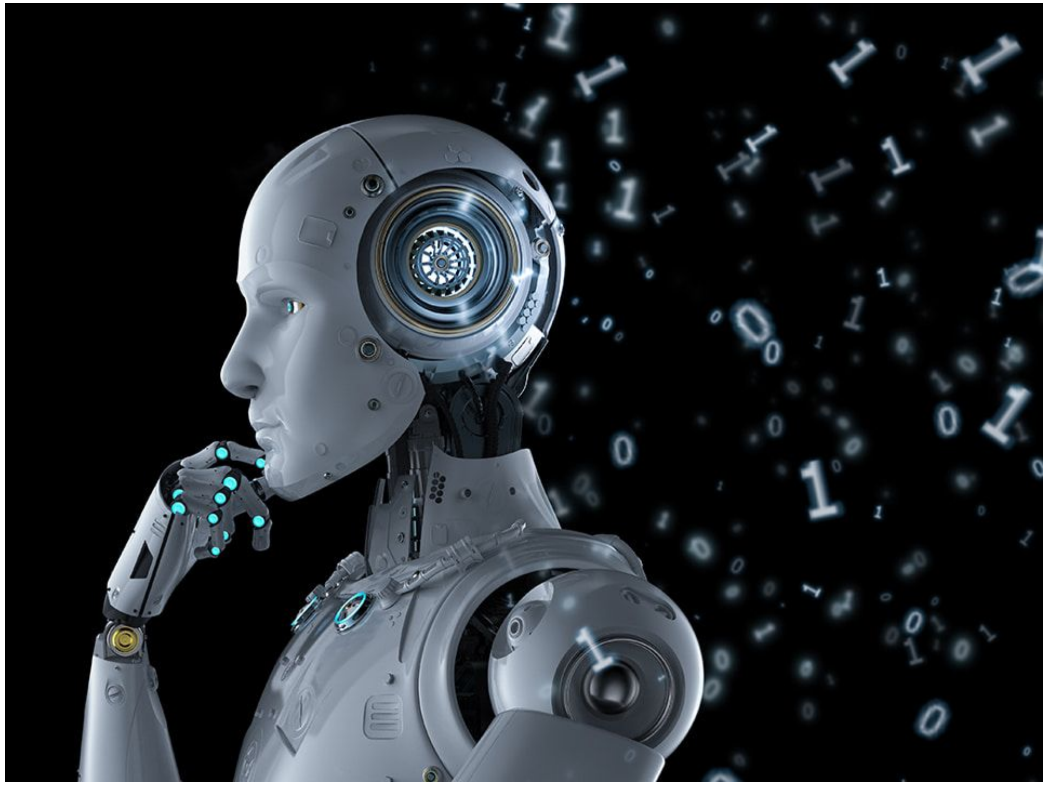
Jie Tan
Robotics @ Google

Deep Reinforcement Learning

[Mastering the game of Go with deep neural networks and tree search, Silver et al. Nature, 2016]

[Outracing champion Gran Turismo drivers with deep reinforcement learning, Wurman et al. Nature, 2022]

Application to Real World Systems



Outline

1. **Background** — Introduction to Deep Reinforcement Learning
2. **Application in power grid** — Power System Emergency Control Using DRL
3. **Extension to large-scale system** — Meta-RL for Large-scale Grid Emergency Voltage Control

Introduction to Deep Reinforcement Learning

Problem Setup

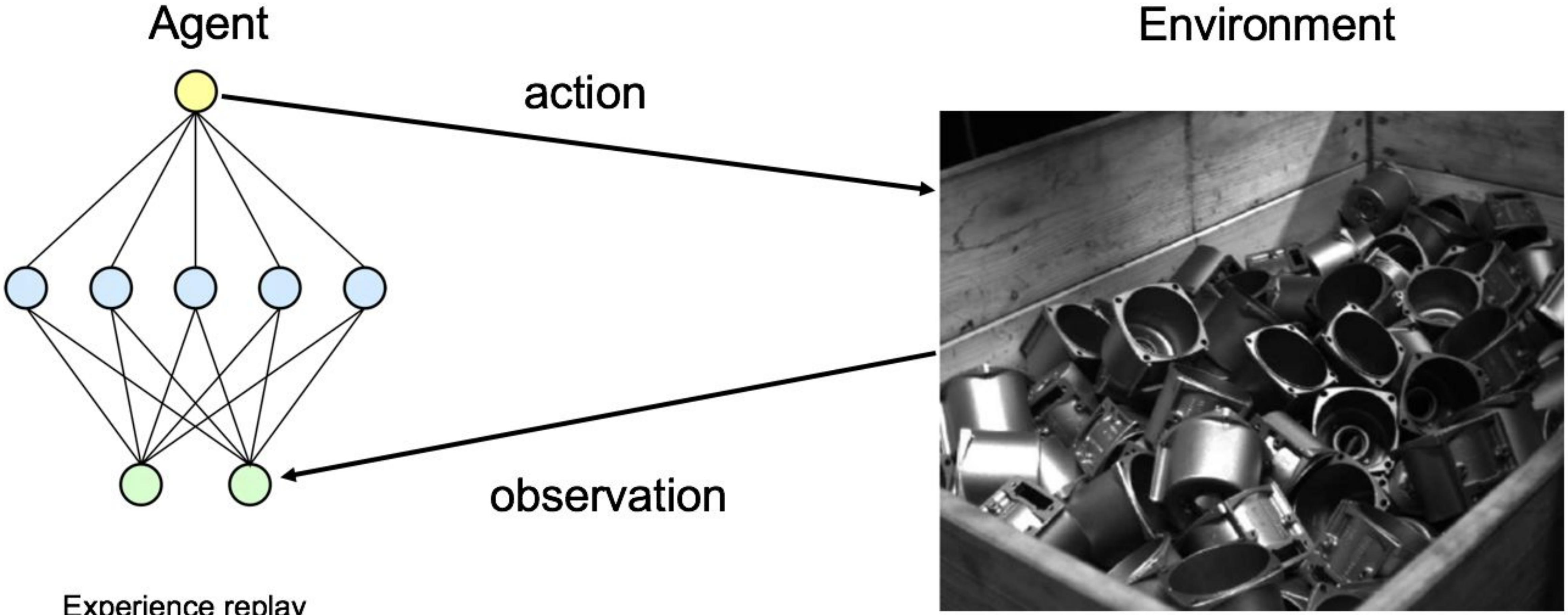
Agent

Environment

action

observation
reward

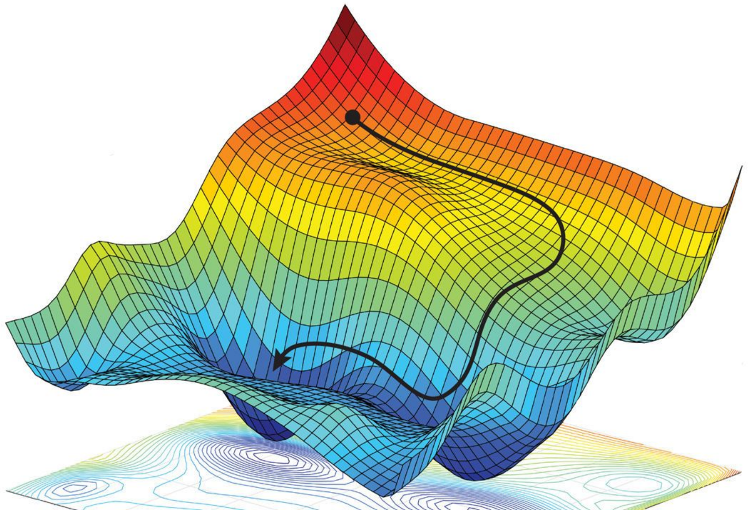
Agent (Policy) Representation



Experience replay
Target networks
Double Q learning
… …

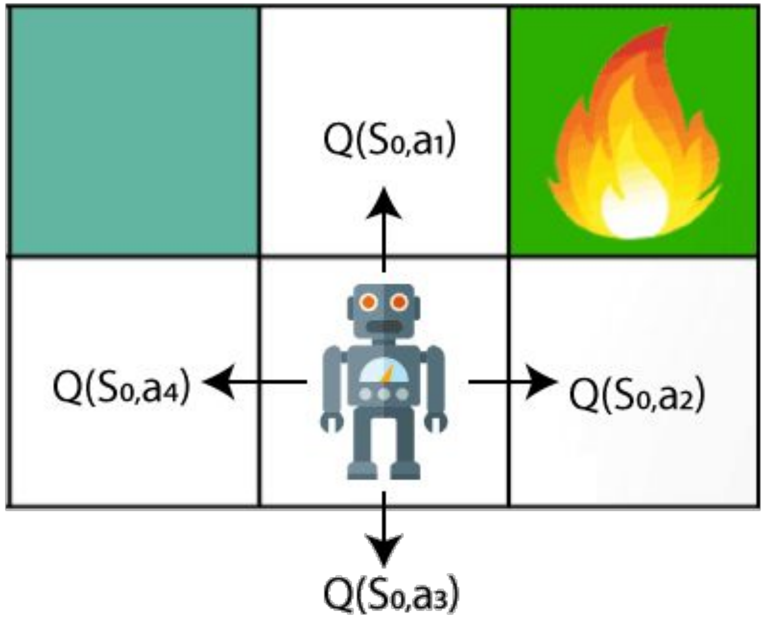
DRL Algorithms

POLICY GRADIENT



$$\nabla \mathbb{E}_\pi[R(\tau))] = \mathbb{E}_\pi[R(\tau)\nabla \log \pi(\tau)]$$

Q-LEARNING BASED



$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \max_a(s_{t+1}, a) - Q(s_t, a_t)]$$

DRL Algorithms in a Nutshell

Current iteration

Step 1 **Exploration**

Execute the policy and add randomness to the actions

Step 2 **Exploitation**

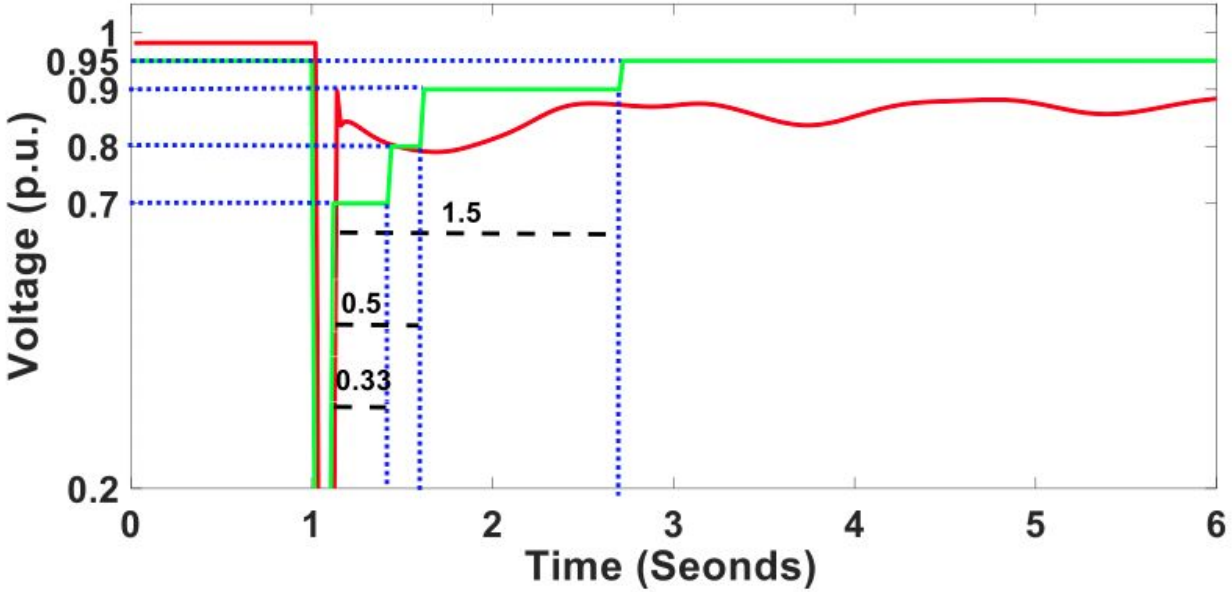
If the result is better than expected, do the same more often in the future

Adaptive Power System Emergency Control Using Deep Reinforcement Learning

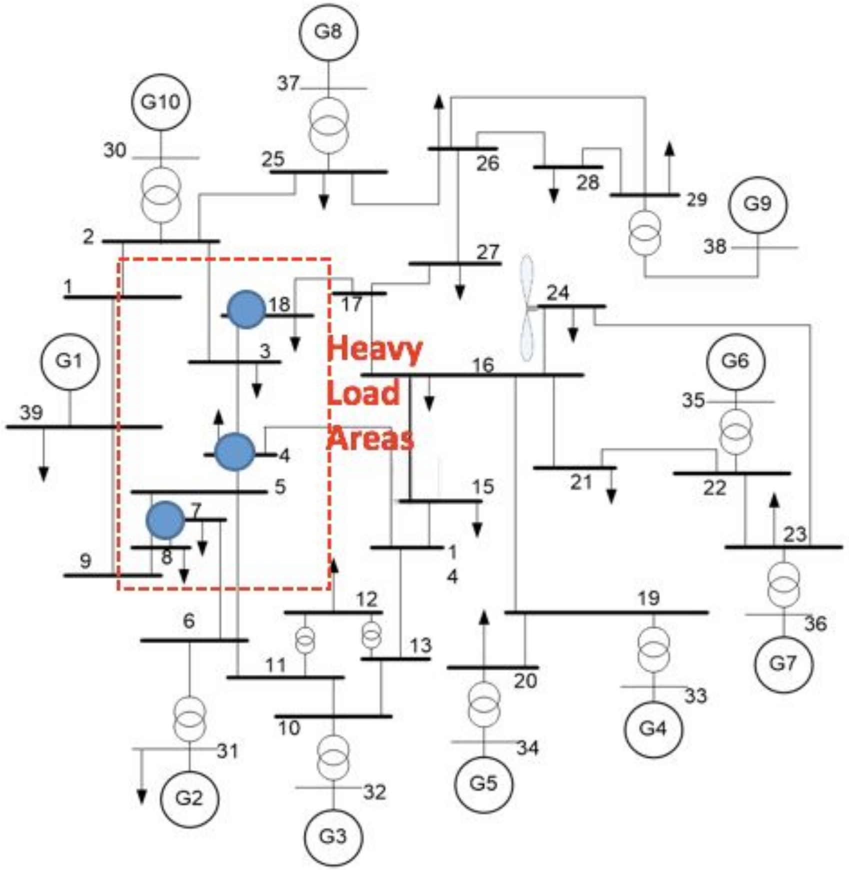
Qiuhua Huang, Renke Huang, Weituo Hao, Jie Tan,
Rui Fan, Zhenyu Huang

IEEE Transactions on Smart Grid, 2019

Fault-Induced Delayed Voltage Recovery (FIDVR)



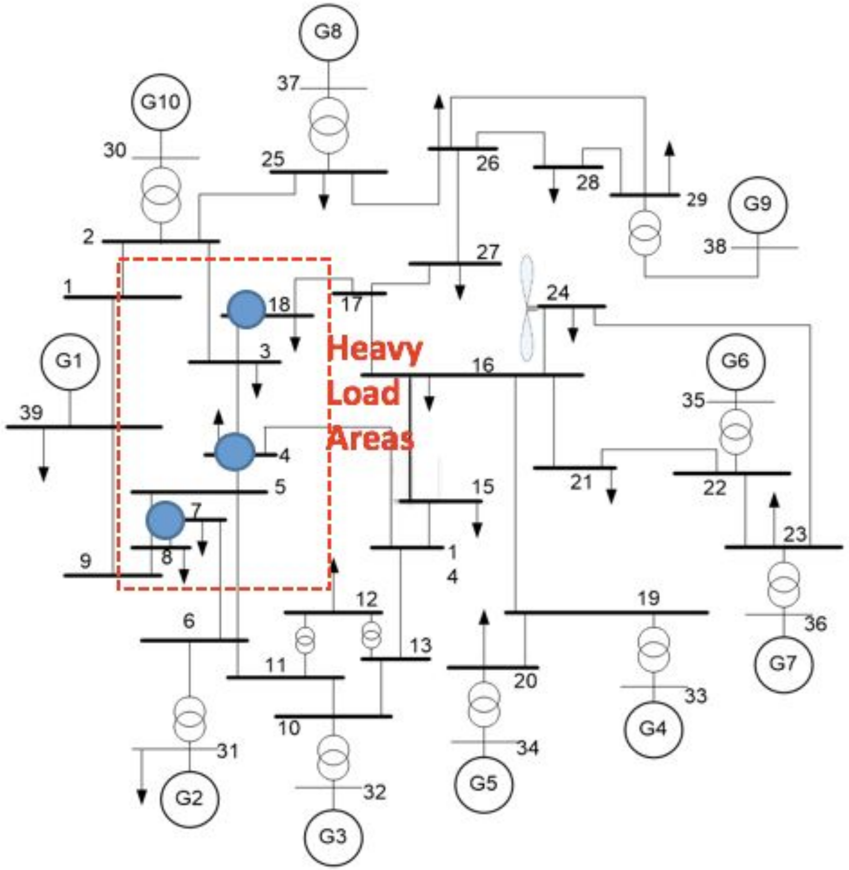
Transient voltage recovery criterion



Substation controlled by RL agent

IEEE 39-bus system model

DRL Formulation

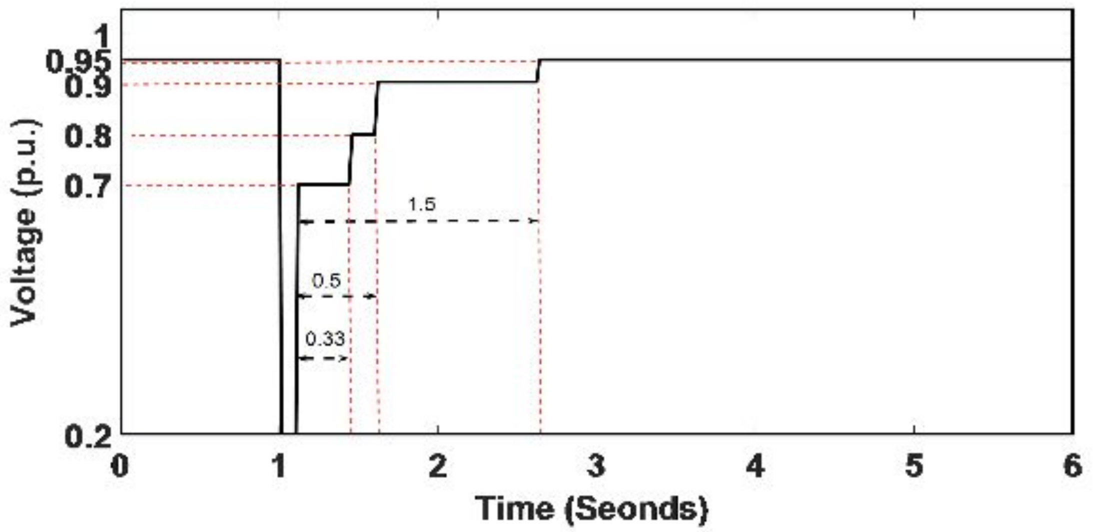


Substation controlled by RL agent

IEEE 39-bus system model

- Observations
  - Voltages and area load levels in the last 10 steps
  - Continuous observation space
- Actions
  - 3 substations could shed load
  - At each bus, at each time step, shed either 0% or 20% of the load
  - 8 dim discrete action space

DRL Formulation: Reward



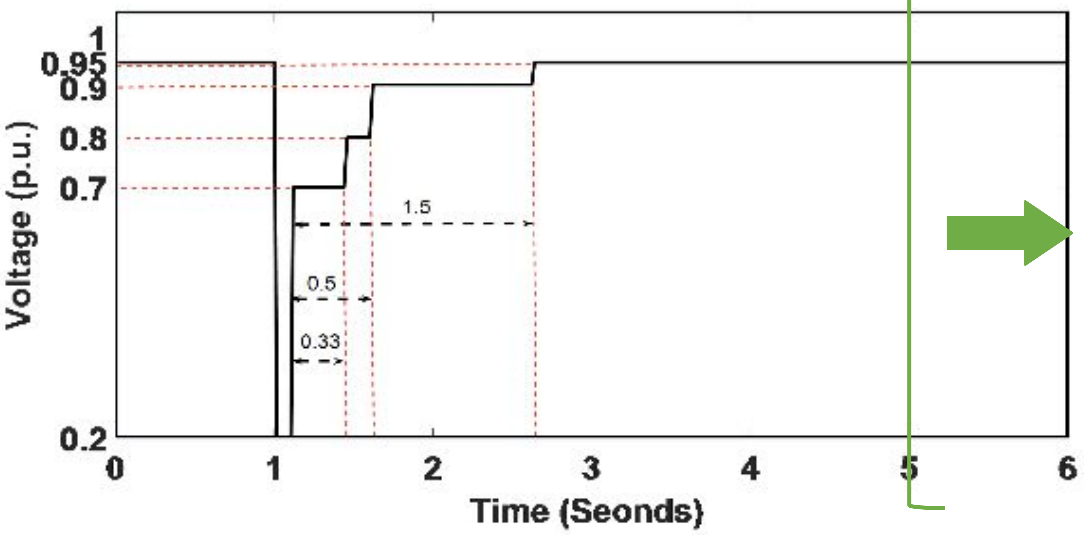
Transient voltage recovery criterion

$$r_t = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Reward function

DRL Formulation: Reward



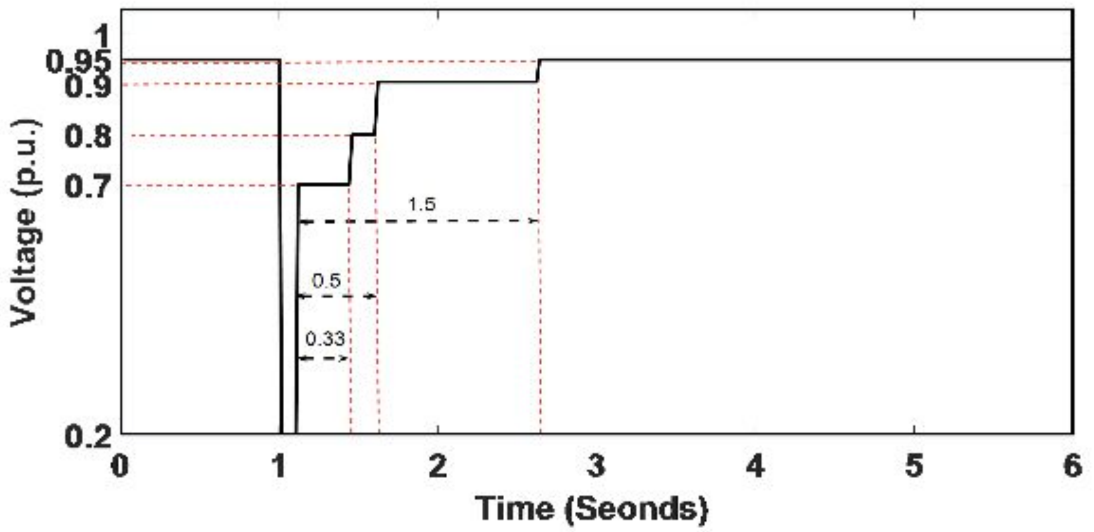
Transient voltage recovery criterion

$$r_t = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Reward function

DRL Formulation: Reward



Transient voltage recovery criterion

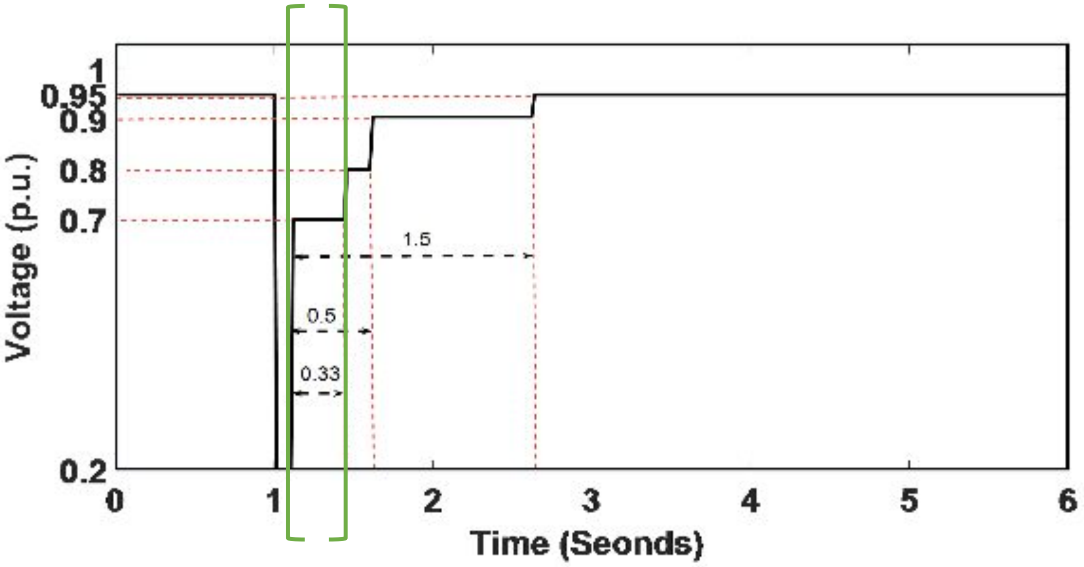
Voltage Criteria

$$r_t = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Reward function

DRL Formulation: Reward



Transient voltage recovery criterion

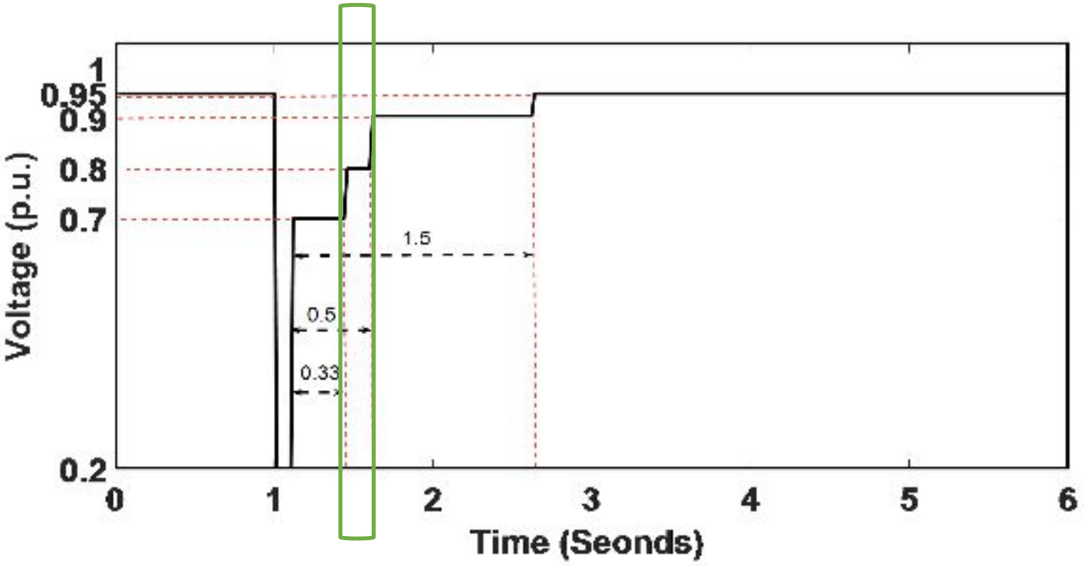
Voltage Criteria

$$r_t = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Reward function

DRL Formulation: Reward



Transient voltage recovery criterion

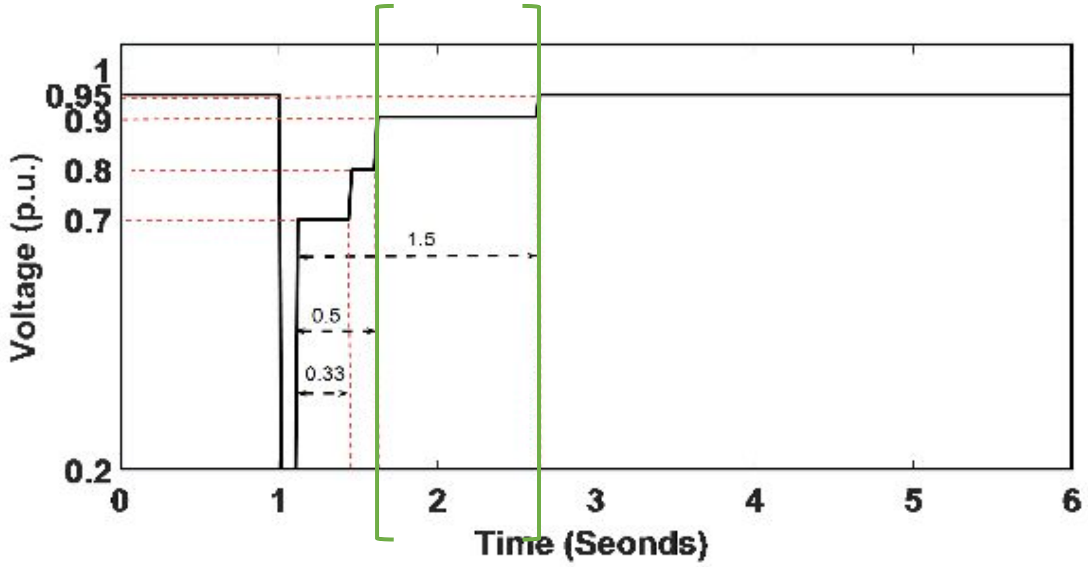
Voltage Criteria

$$r_t = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Reward function

DRL Formulation: Reward



Transient voltage recovery criterion

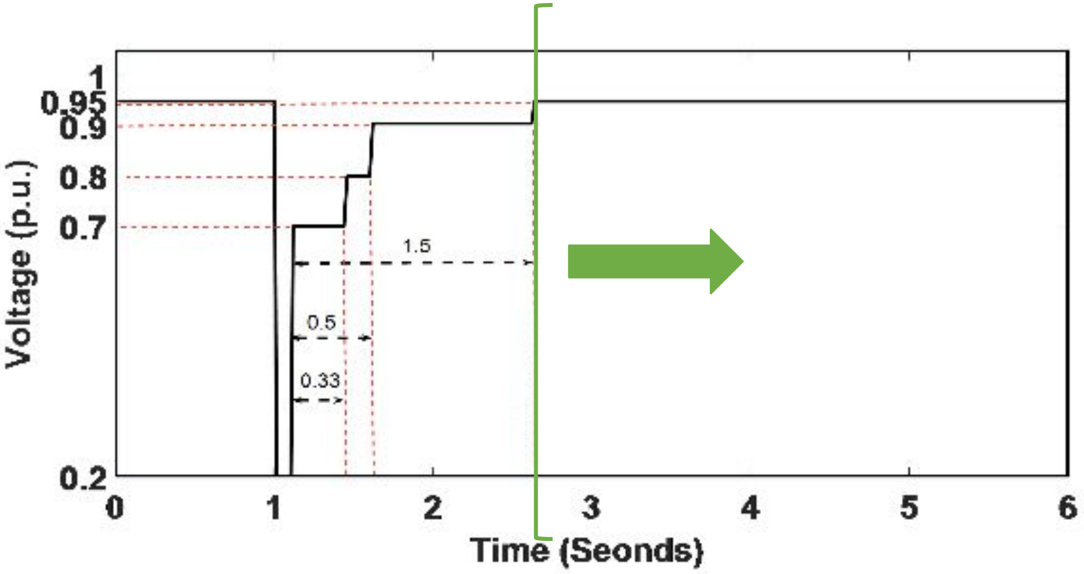
Voltage Criteria

$$r_t = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Reward function

DRL Formulation: Reward



Transient voltage recovery criterion

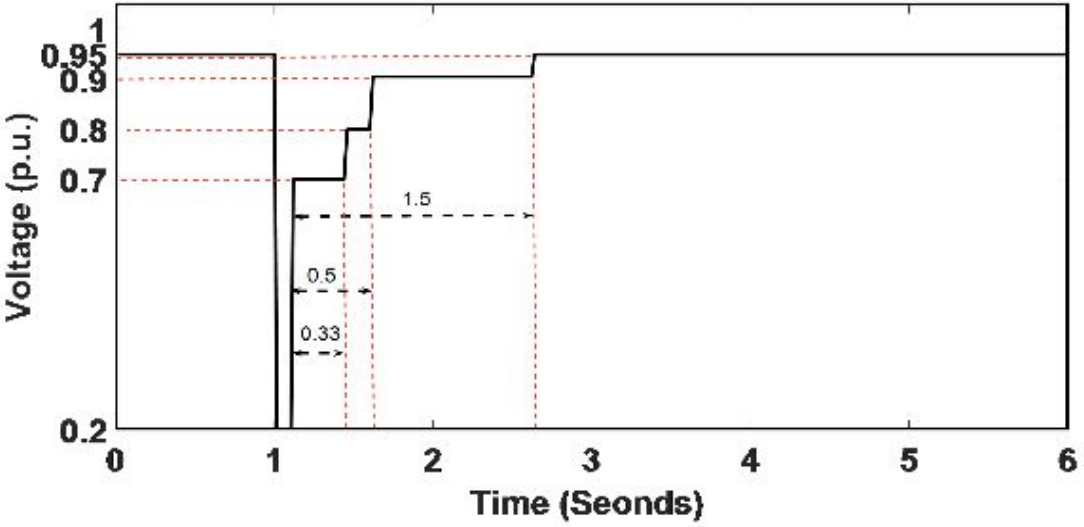
Voltage Criteria

$$r_t = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Reward function

DRL Formulation: Reward



Transient voltage recovery criterion

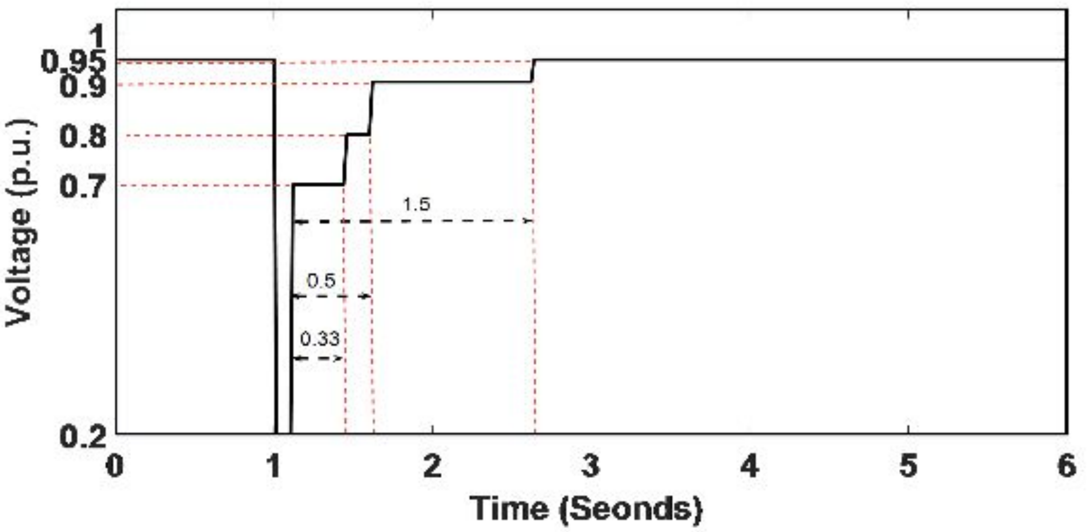
Shedding Amount

$$r_t = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Reward function

DRL Formulation: Reward



Transient voltage recovery criterion

Invalid Action

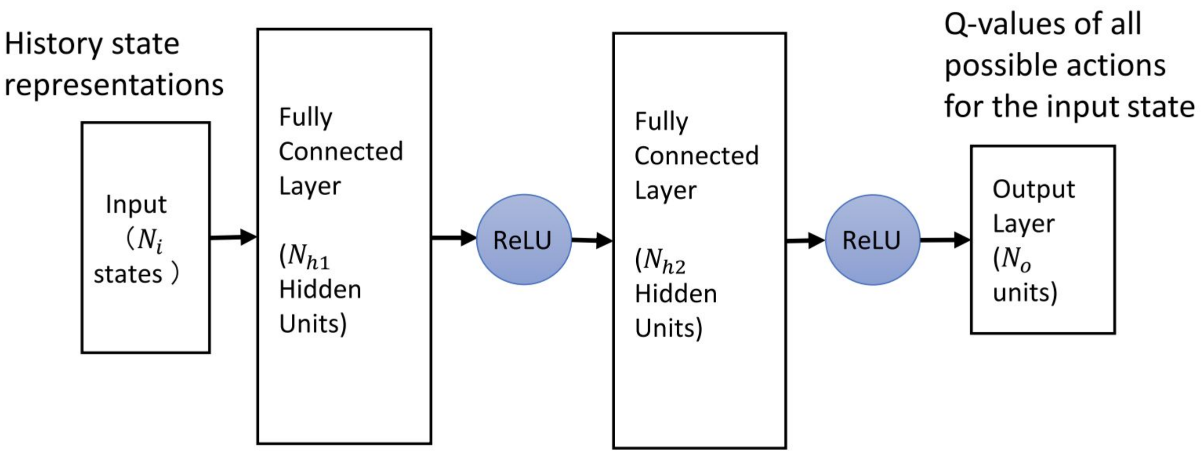
$$r_t = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Reward function

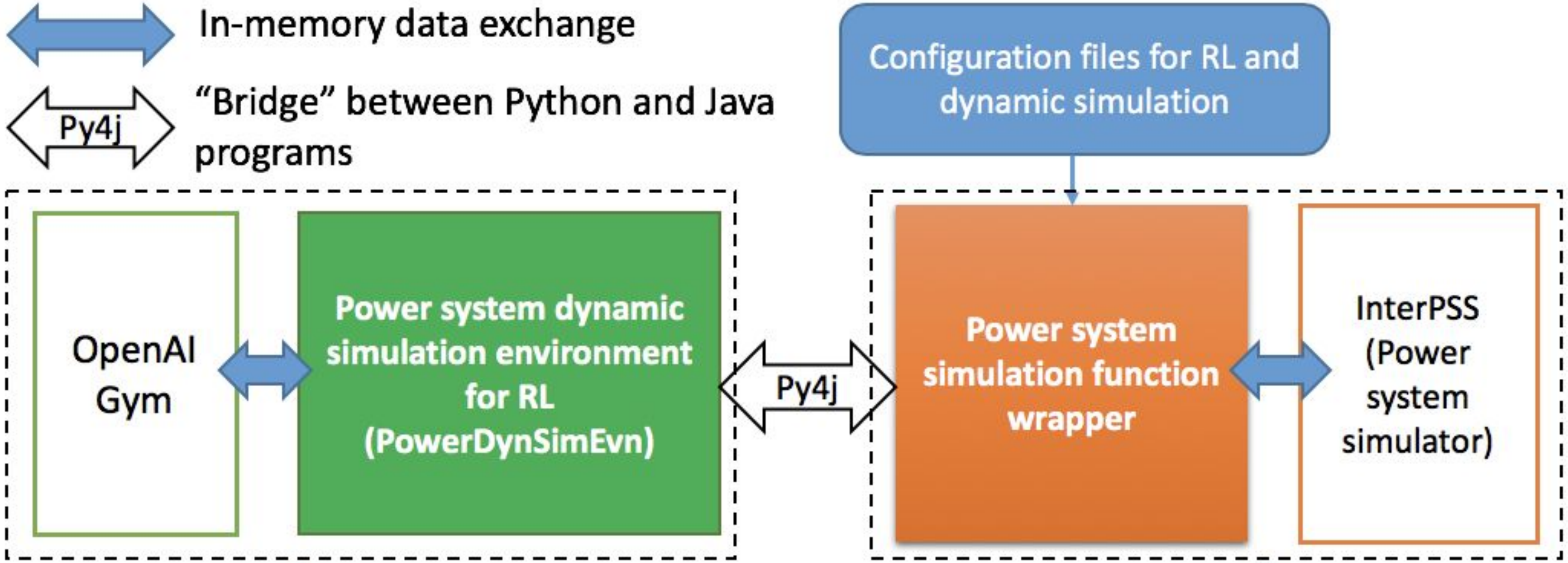
DRL Formulation

- Agent
  - 2-layer fully connected neural network



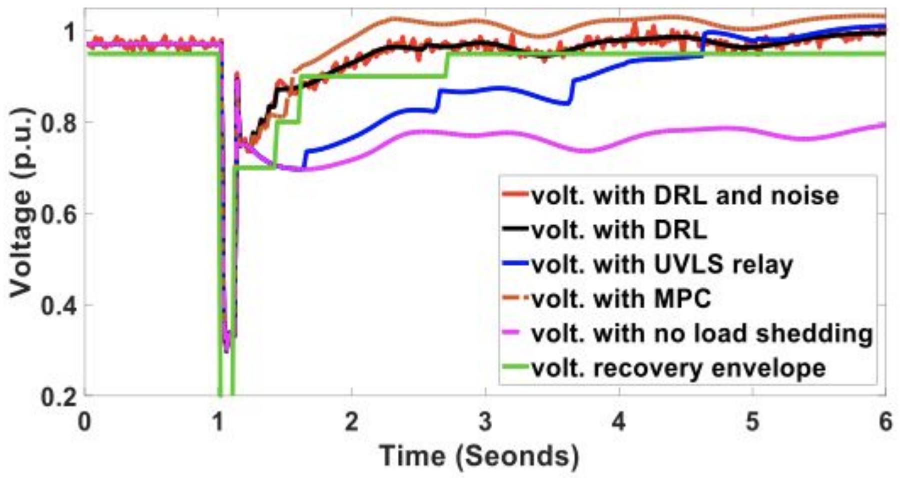
- Training algorithm
  - Deep Q Network (DQN)

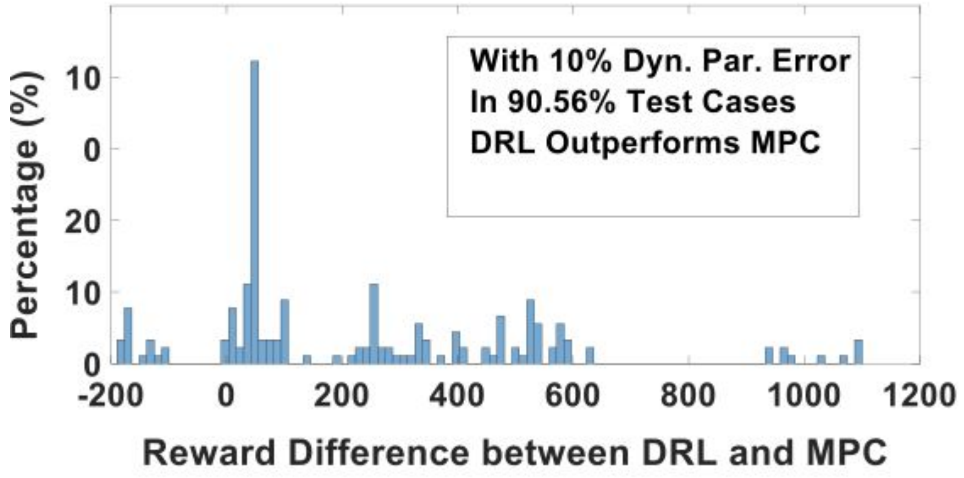
Simulation Environment for Grid Control



RLGC: An open-source platform for developing, testing and benchmarking Reinforcement Learning for Grid Control (https://github.com/RLGC-Project/RLGC)

Experiments and Evaluations





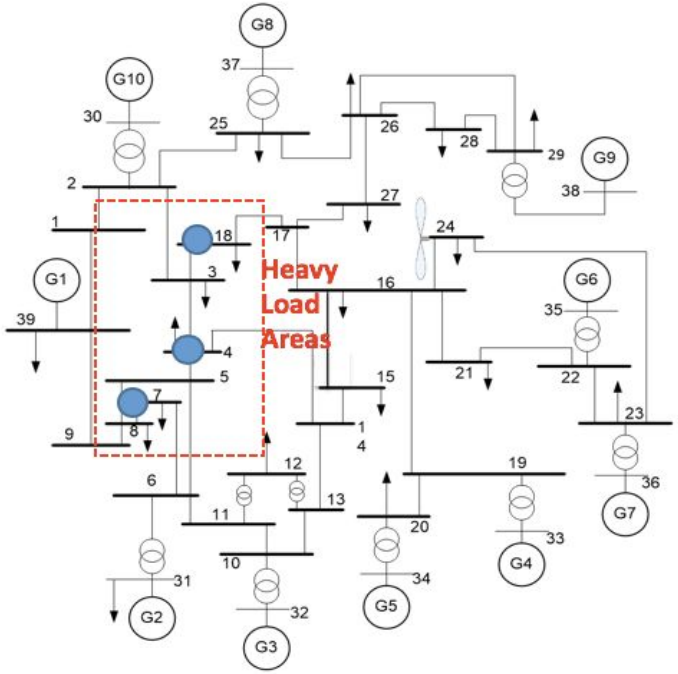
DRL VS Relays

- DRL outperforms Relays for **92.22%** of 462 Test Cases
- For the case as shown, reduce load shedding of **134.64** MW

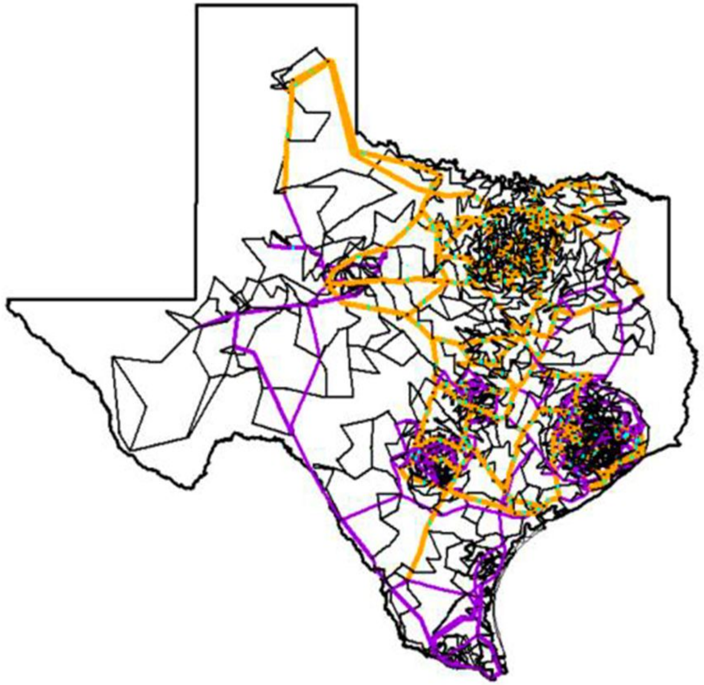
DRL VS MPC

- DRL outperforms MPC for **90.56%** of 462 Test Cases.
- Average Execute Time: **0.13 sec** for DRL, **23.73 sec** for MPC
- For the case as shown, reduce load shedding of **40.64** MW

Large-Scale FIDVR Problem



IEEE 39-bus system model



Texas 2000-bus system

Learning and Fast Adaptation for Grid Emergency Control via Deep Meta Reinforcement Learning

Renke Huang, Yujiao Chen, Tianzhixi Yin, Qiuhua Huang, Jie Tan, Wenhao Yu, Xinya Li, Ang Li, Yan Du

IEEE Transaction on Power Systems, 2022

Challenges of Scale

1 Action space

Number of discrete actions grows exponentially with number of load-shedding buses.

2 Parallelism of RL algorithms

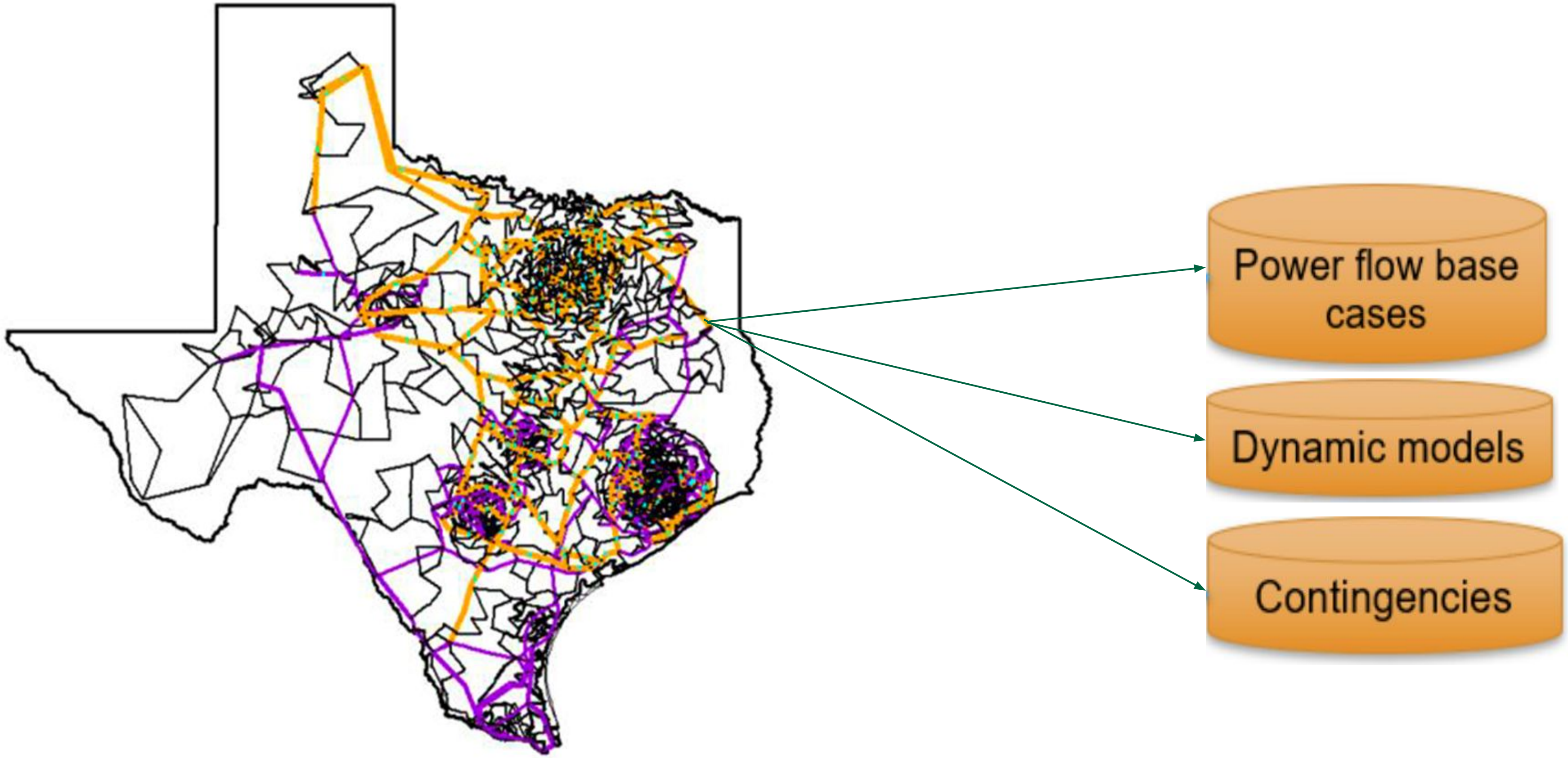
Many reinforcement learning algorithms are inherently sequential.

3 Optimization challenge

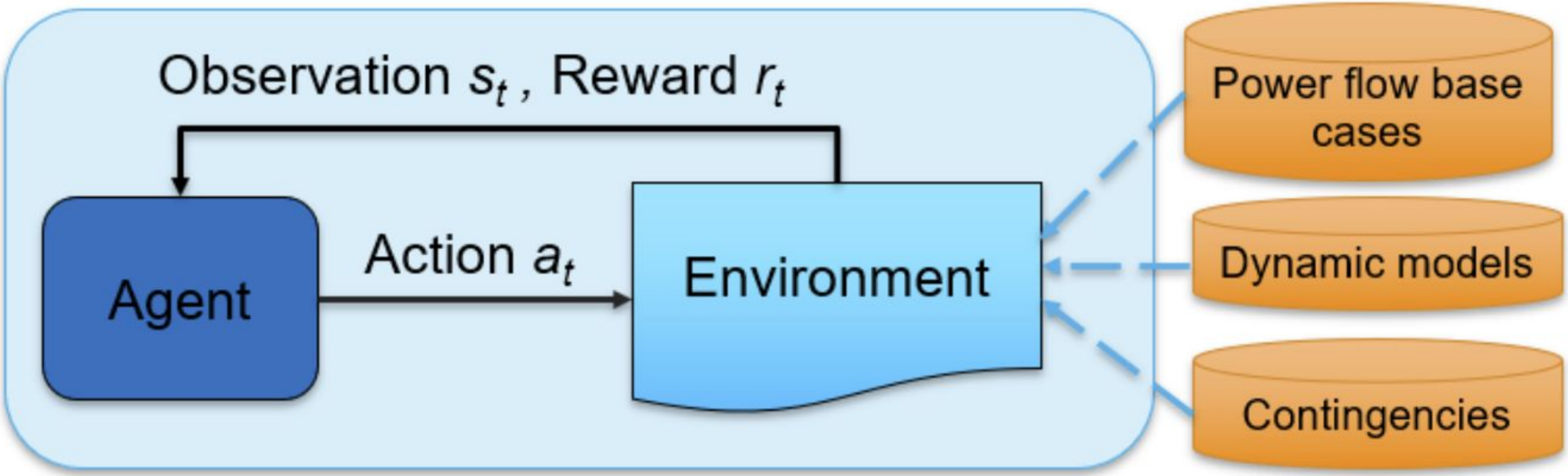
Difficult to find one policy that works optimally in a large number of operation conditions.

Tasks / Operation Conditions

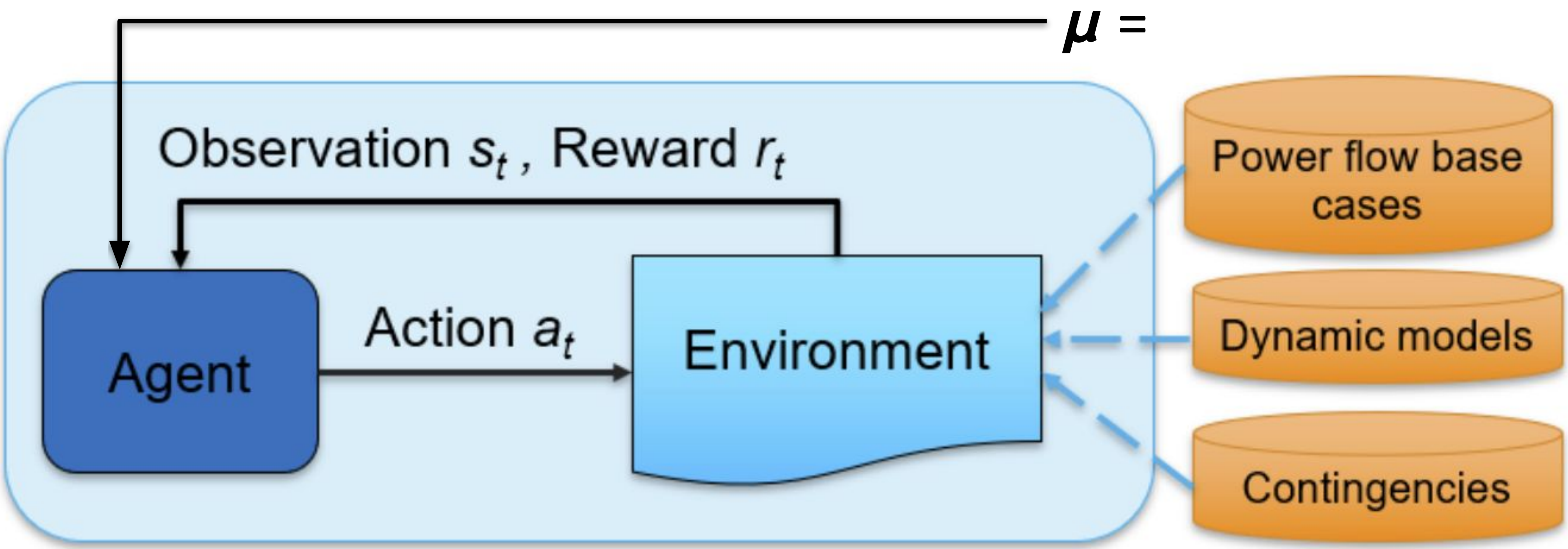
Curse of dimensionality: The number of operation conditions grow exponentially as the grid gets bigger!



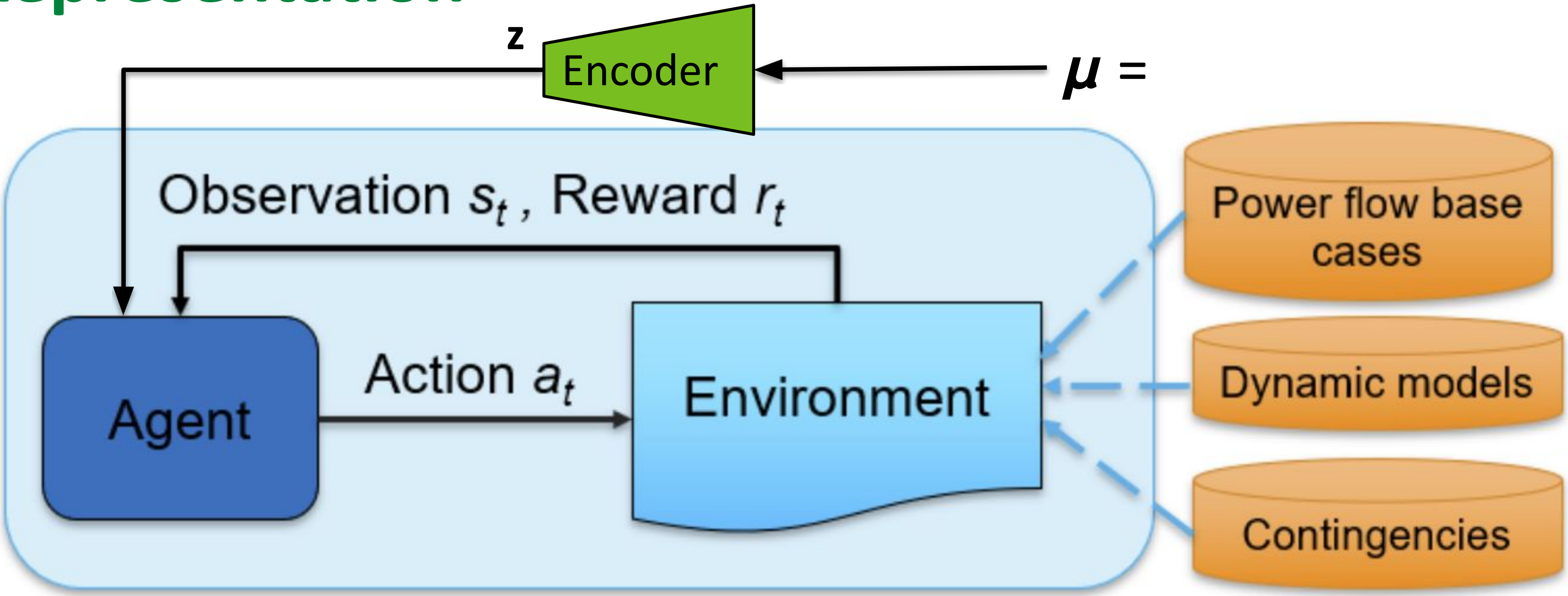
Train One Policy for All Operation Conditions



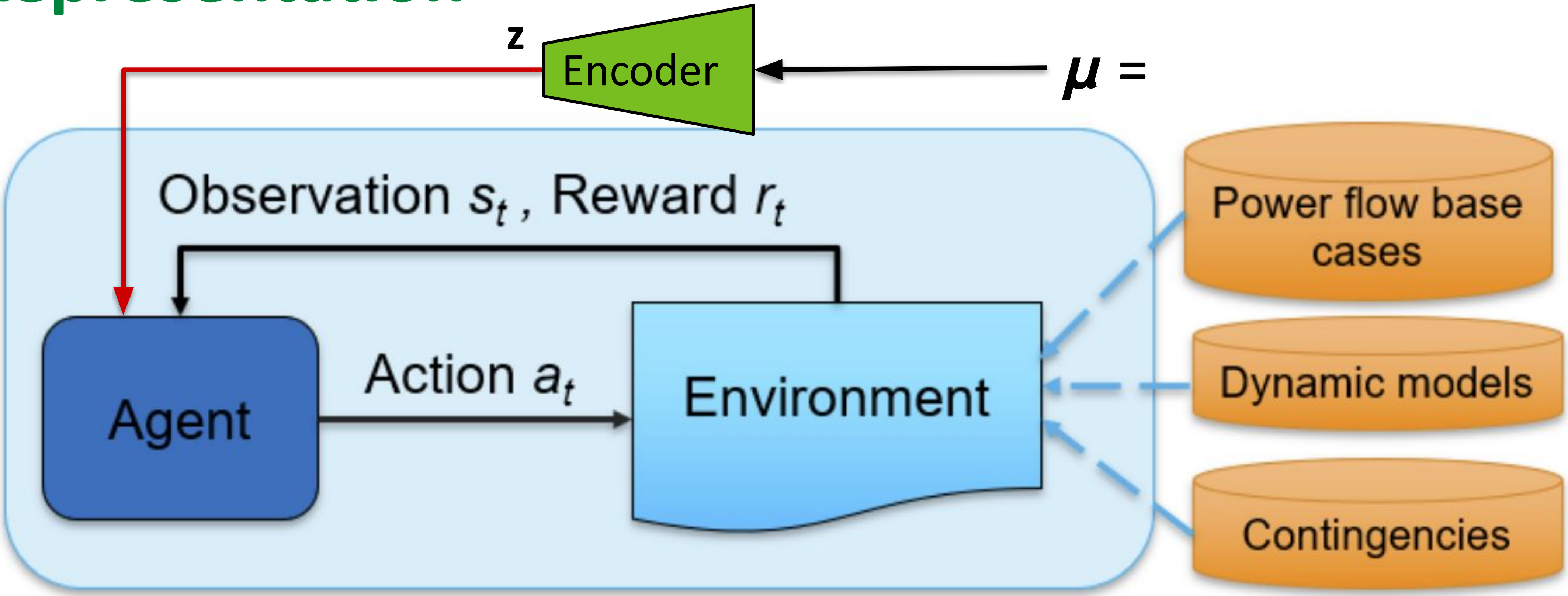
Task-Conditioned (Meta) Policy



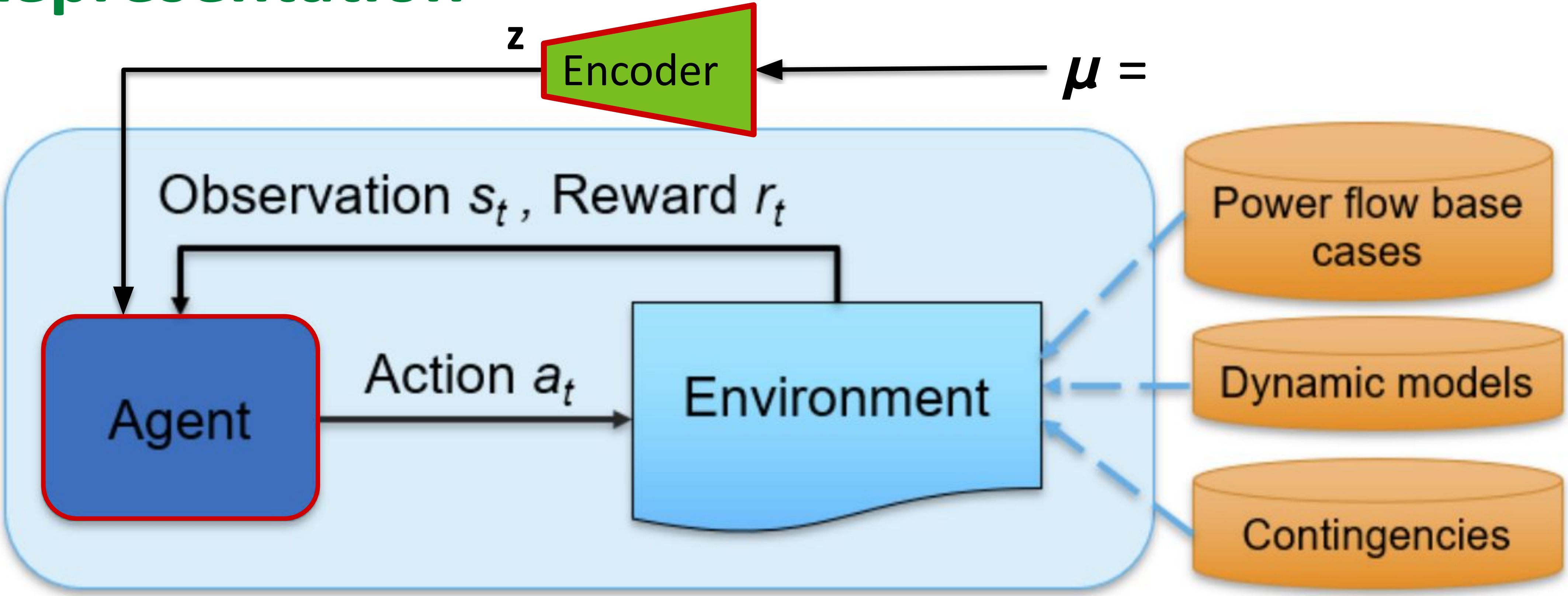
Training Meta Policy with Latent Representation



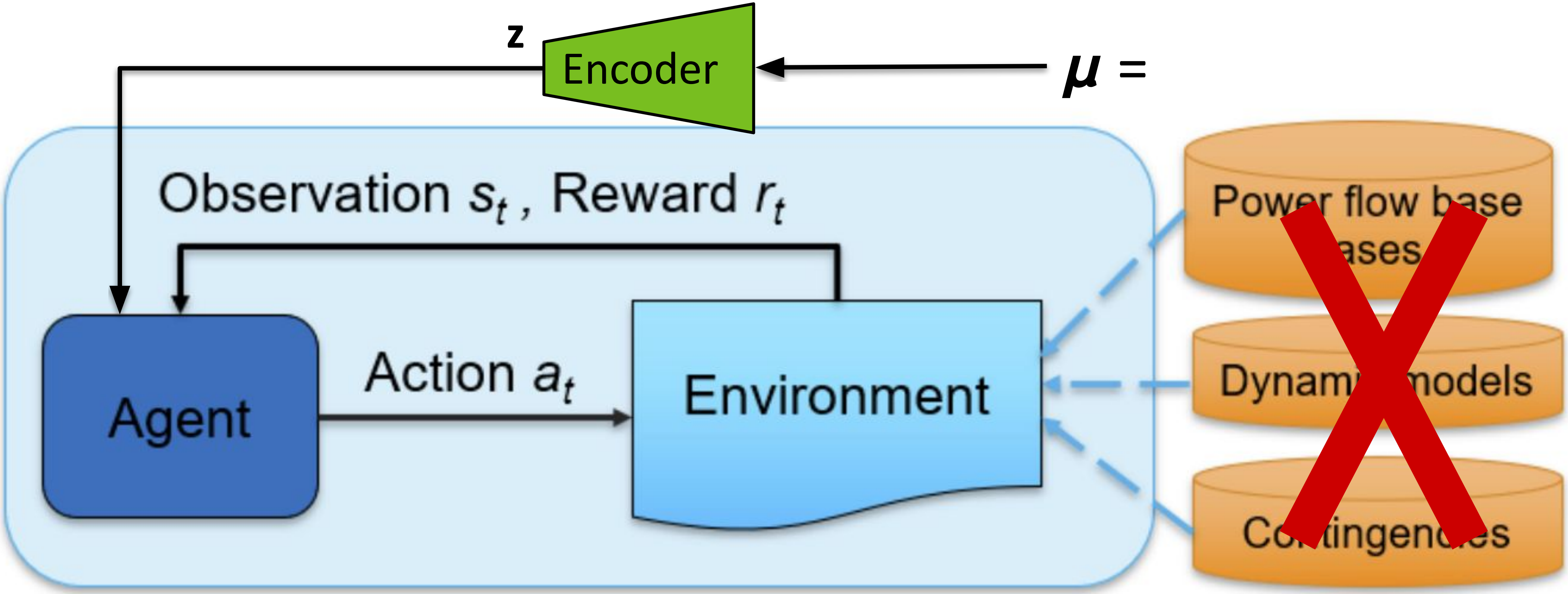
Training Meta Policy with Latent Representation



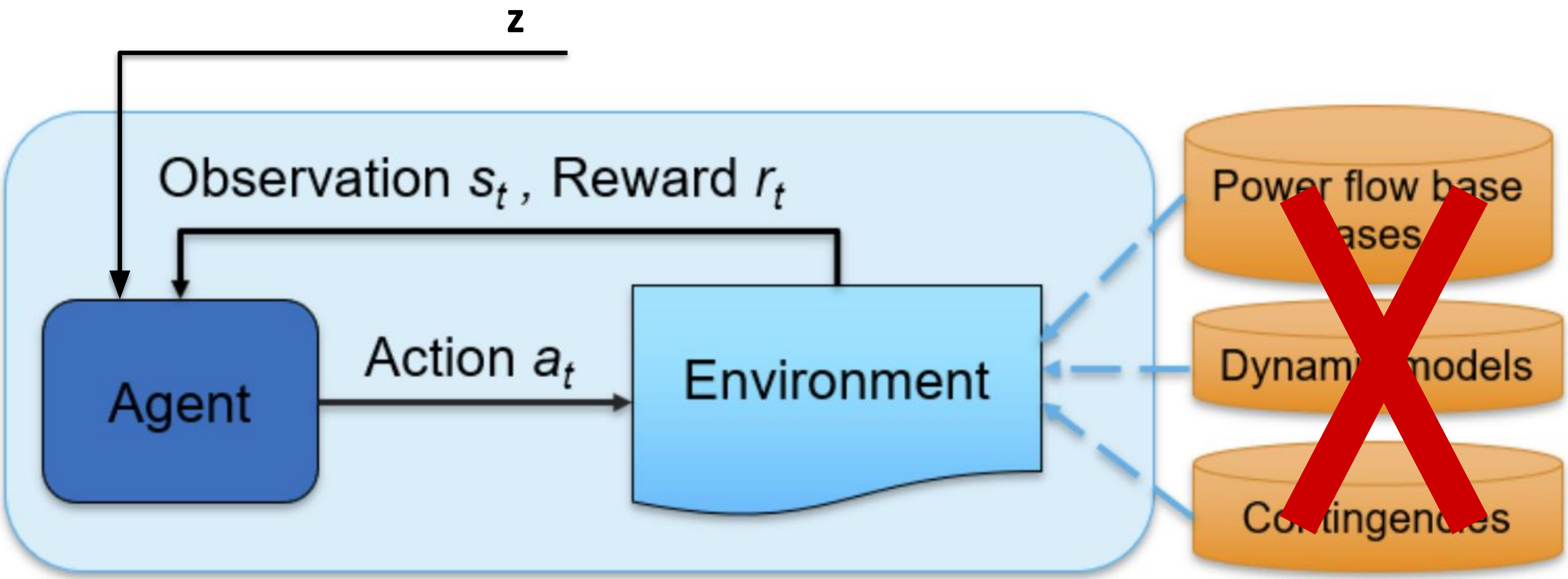
Training Meta Policy with Latent Representation



Adapting Meta Policy in Test Time

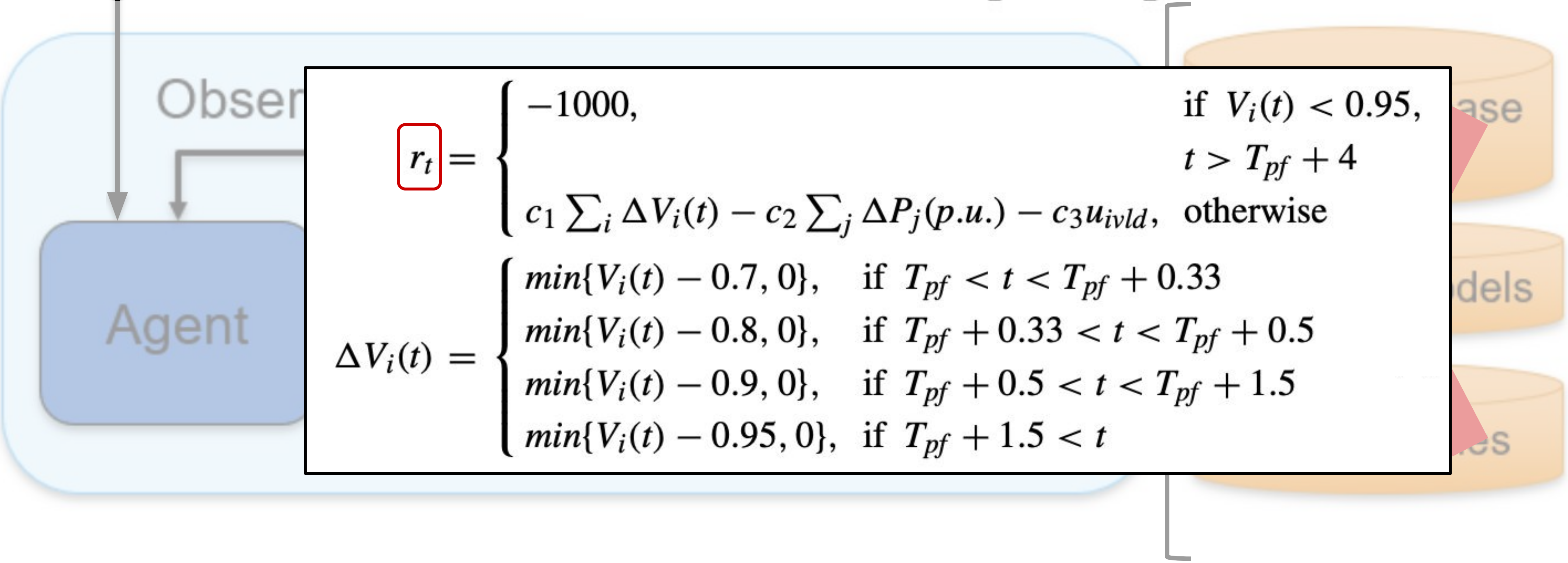


Adapting Meta Policy in Test Time



Adapting Meta Policy in Test Time

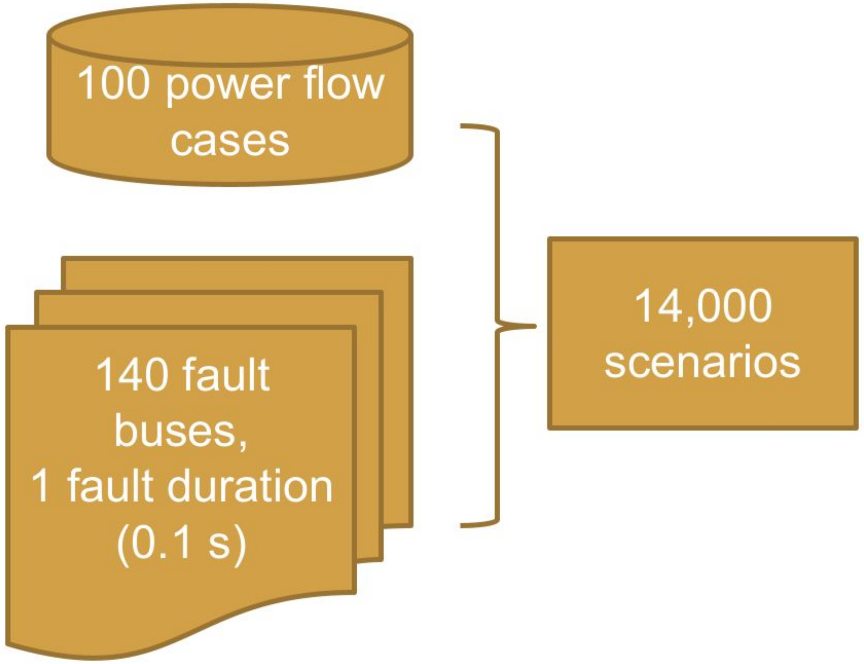
$$\mathbf{z}^* = \arg\max_{\mathbf{z}} \; \mathbb{E}_{\tau \sim p^*(\tau|\pi,\mathbf{z})} \left[\sum_{t=0}^{T-1} \gamma^t \boxed{r_t} \right]$$



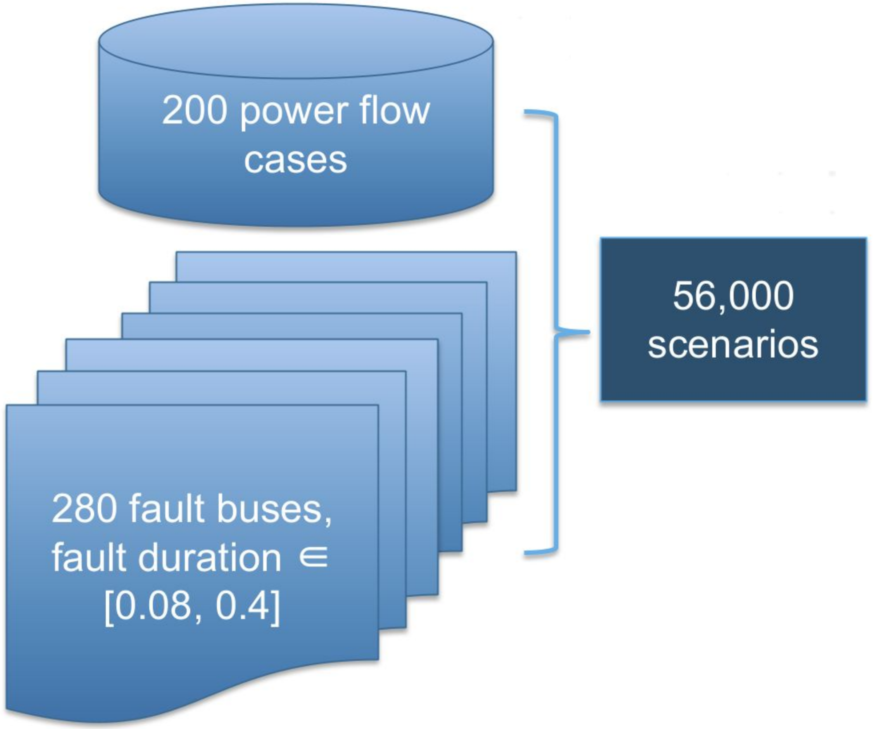
$$\boxed{r_t} = \begin{cases} -1000, & \text{if } V_i(t) < 0.95, \\ & t > T_{pf} + 4 \\ c_1 \sum_i \Delta V_i(t) - c_2 \sum_j \Delta P_j(p.u.) - c_3 u_{ivld}, & \text{otherwise} \end{cases}$$

$$\Delta V_i(t) = \begin{cases} min\{V_i(t) - 0.7, 0\}, & \text{if } T_{pf} < t < T_{pf} + 0.33 \\ min\{V_i(t) - 0.8, 0\}, & \text{if } T_{pf} + 0.33 < t < T_{pf} + 0.5 \\ min\{V_i(t) - 0.9, 0\}, & \text{if } T_{pf} + 0.5 < t < T_{pf} + 1.5 \\ min\{V_i(t) - 0.95, 0\}, & \text{if } T_{pf} + 1.5 < t \end{cases}$$

Training and Testing

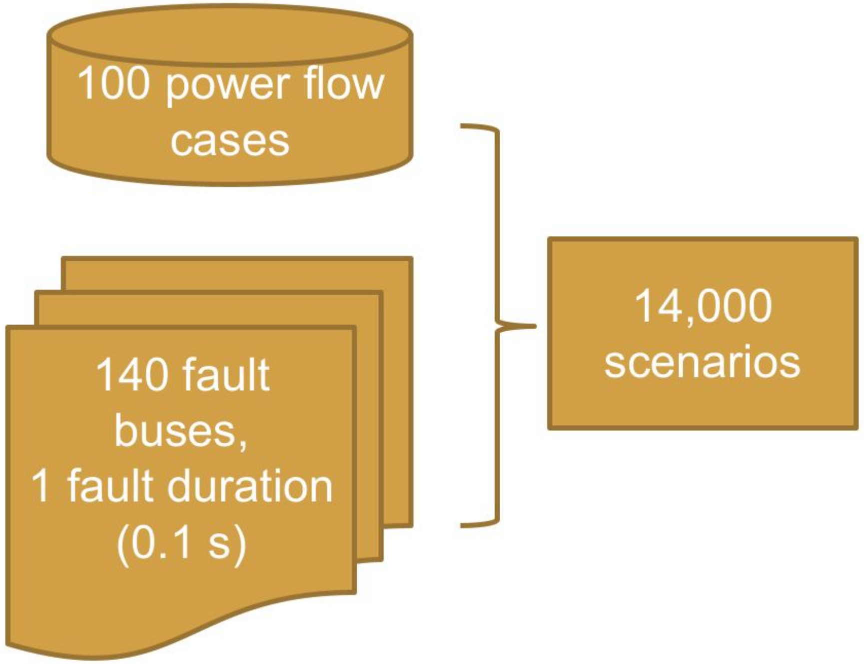


Training tasks

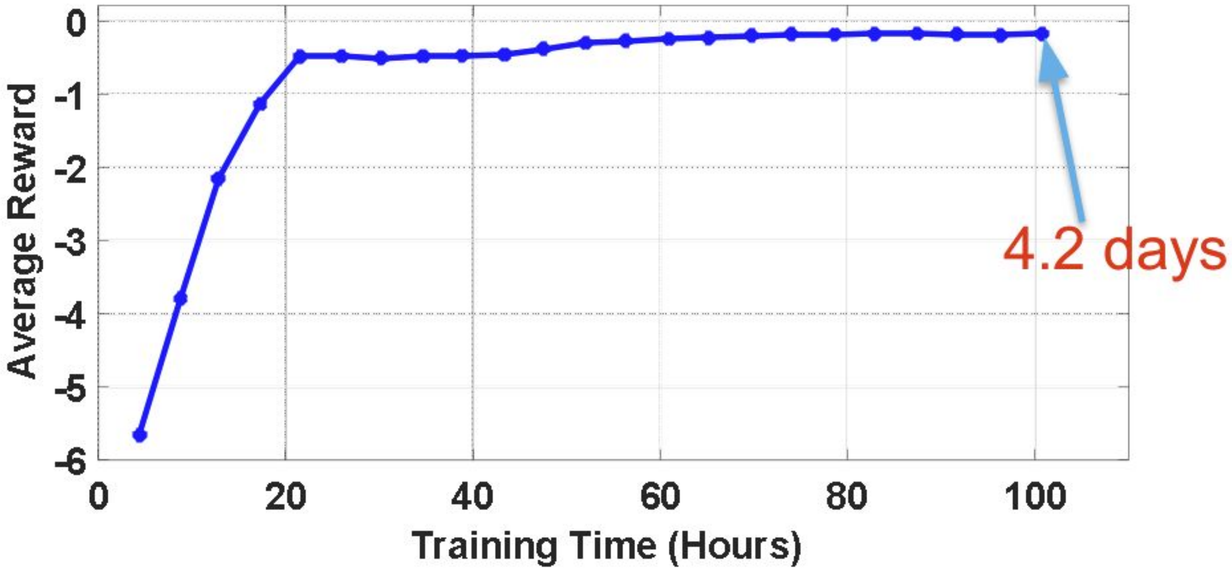


Testing tasks

Training

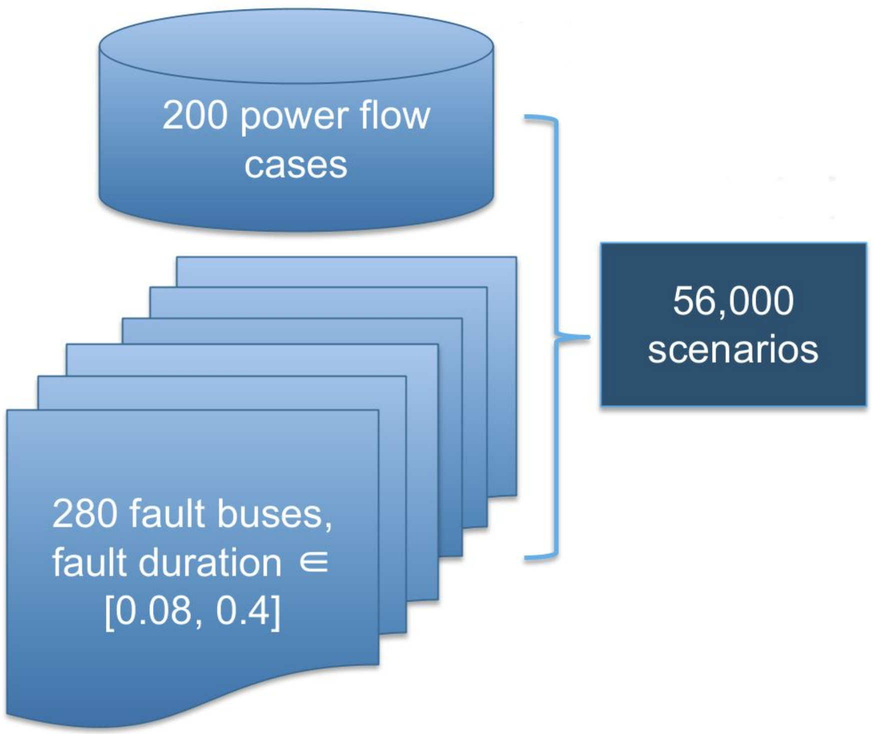


Training tasks

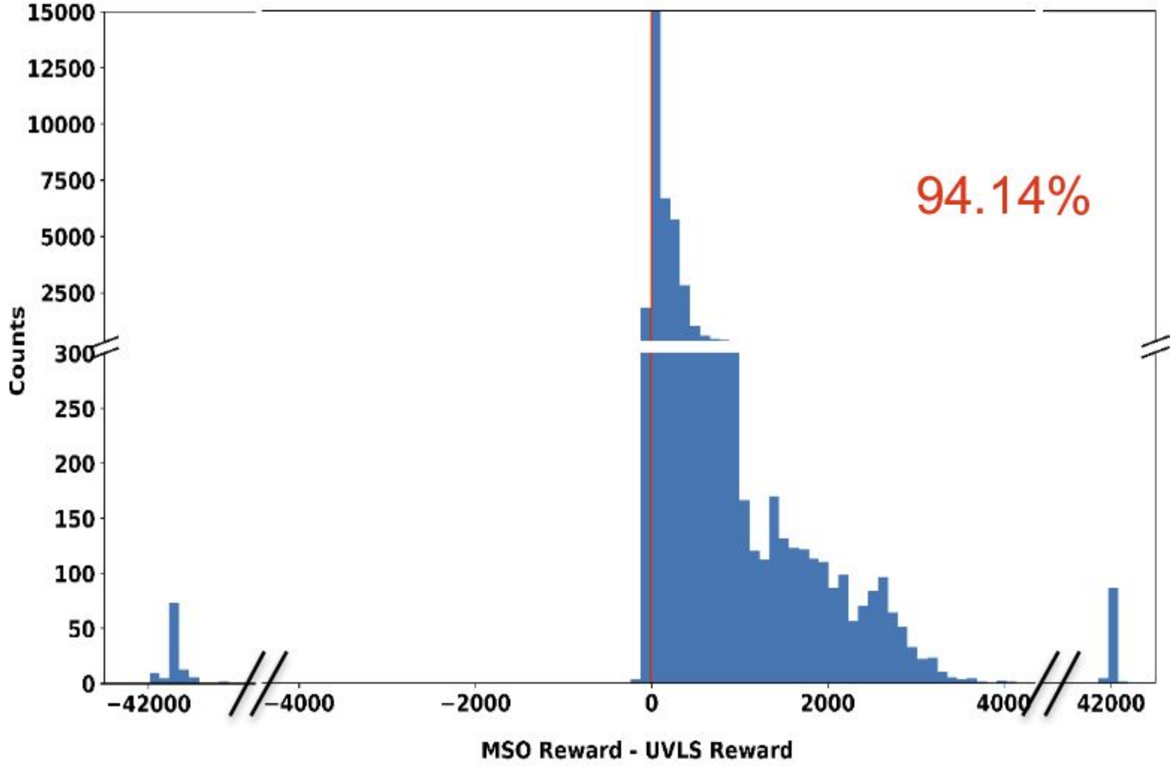


Training curve

Testing



Testing tasks



Performance

Summary

- RL is a powerful tool, which automatically learns state-of-the-art emergency controller for large-scale power grids.

- Many challenges remain:
  - reward design
  - safety
  - sim-to-real gap

- Promising future directions:
  - combine model-based optimal control and model-free learning
  - combine imitation learning with reinforcement learning
  - human-in-the-loop

Thank you