TempoRL: Learning When to Act

André Biedenkapp[1], Raghu Rajan[1], Frank Hutter[1], and Marius Lindauer[2]

[1]University of Freiburg, [2]Leibniz University Hannover

**universität freiburg**

Leibniz Universität Hannover

In a Nutshell

1. We propose a **proactive way of doing RL**
2. We introduce **skip-connections into MDPs**
   - Use of action repetition
   - Faster propagation of rewards
3. We **propose a novel algorithm** using skip-connections
   - Learn *what* action to take & *when* to make a new decision
   - Condition *when* on *what*
4. We **evaluate our approach in a variety of settings**
   - Tabular Q-learning on Grid-World (see below)
   - DQN on featurized environments (LunarLandar, MountainCar)
   - DDPG on featurized environments (Pendulum)
   - DQN with image states on Atari
     (Pong, Qbert, Freeway, MsPacman, BeamRider)

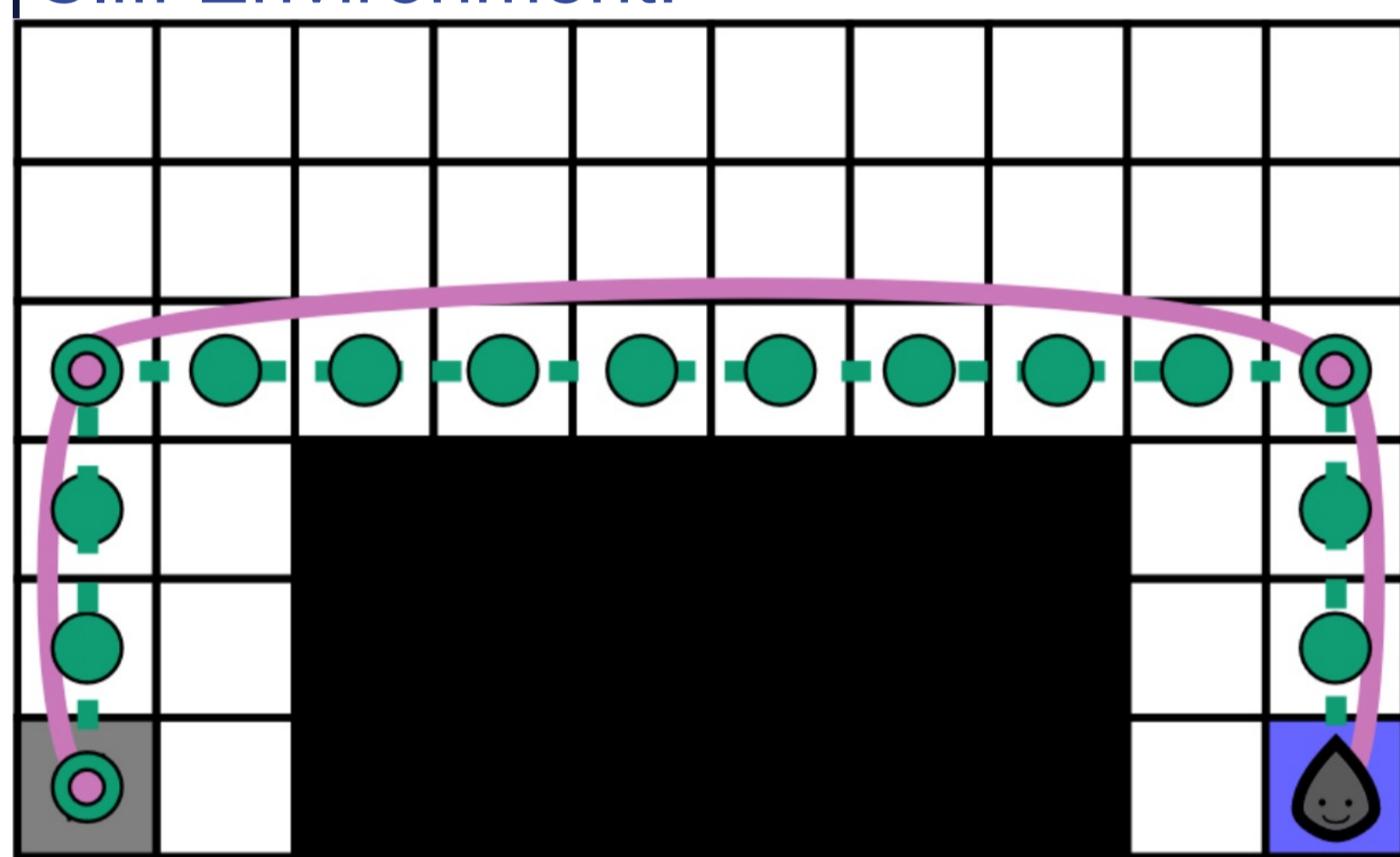## Information Propagation w/ Skip Connections

What action?: $Q^\pi(s, *) \to a$ When to switch?: $Q^\pi(s, *|a) \to skip$
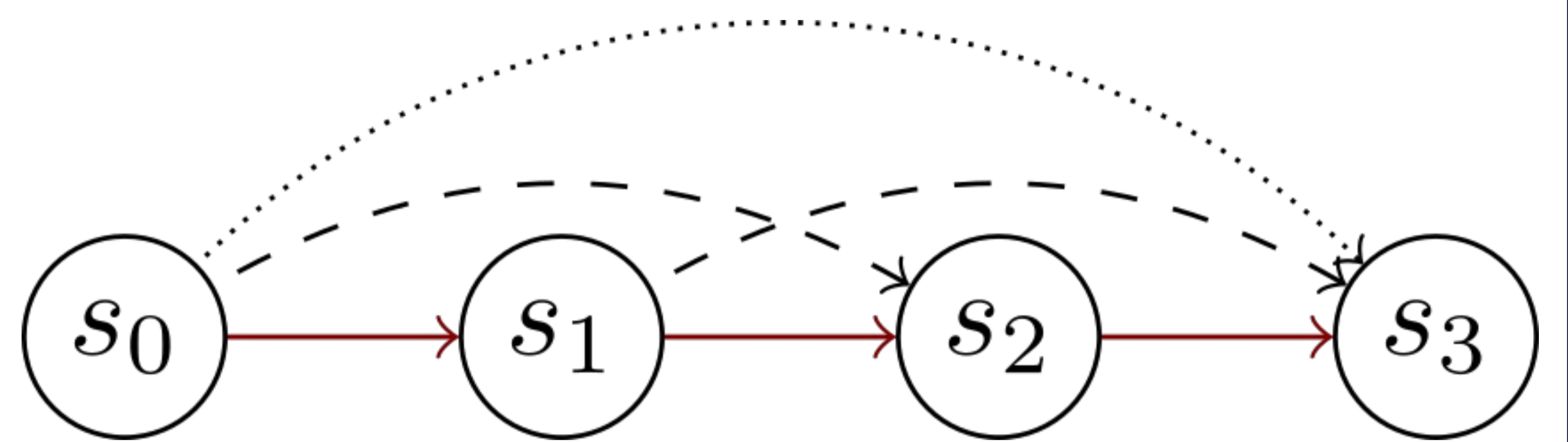
- Action repetition introduces skips
- Information can be propagated faster along skips
- With large skips, multiple smaller skips can be observed

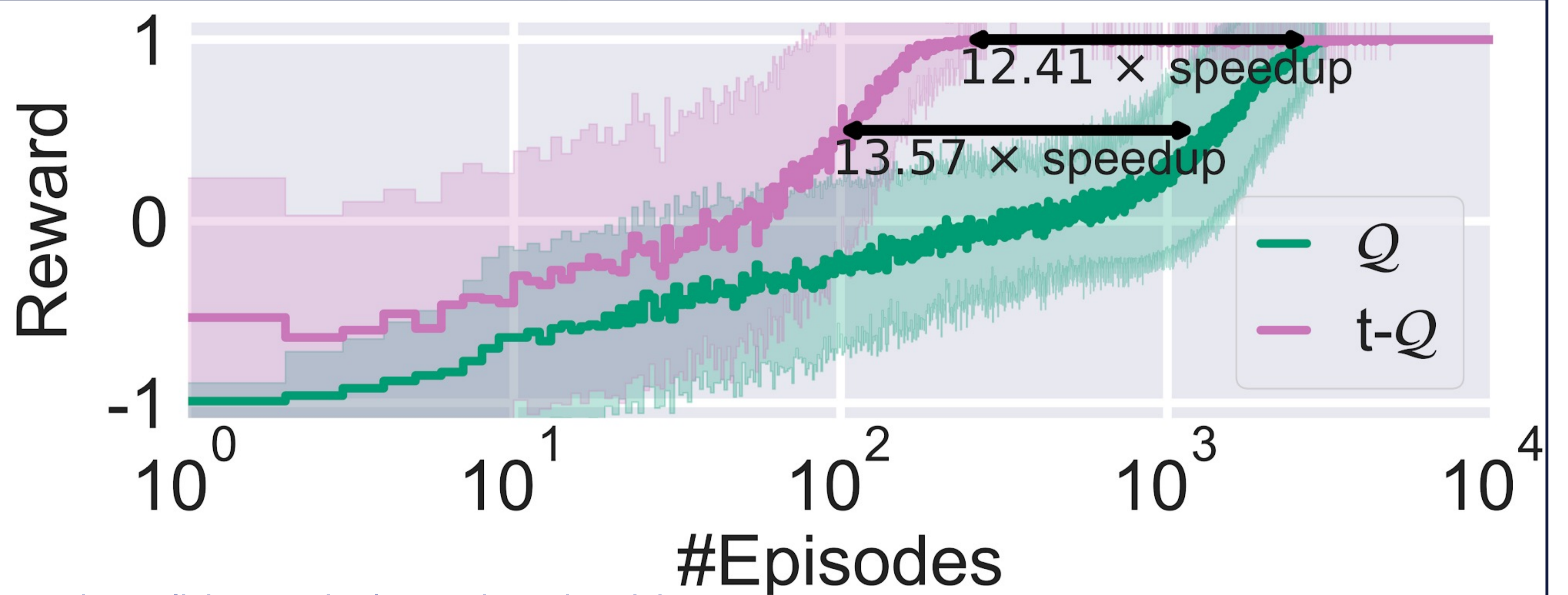## Evaluation Performance of Tabular Q-learning Agents
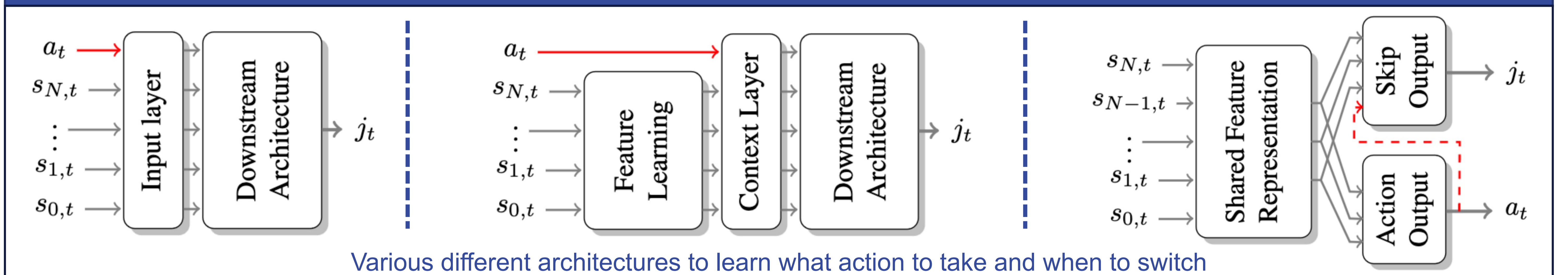
Cliff Environment:

Sparse reward:
  +1 for reaching goal
  -1 for falling down cliff

Vanilla Q-learning: 16 decisions
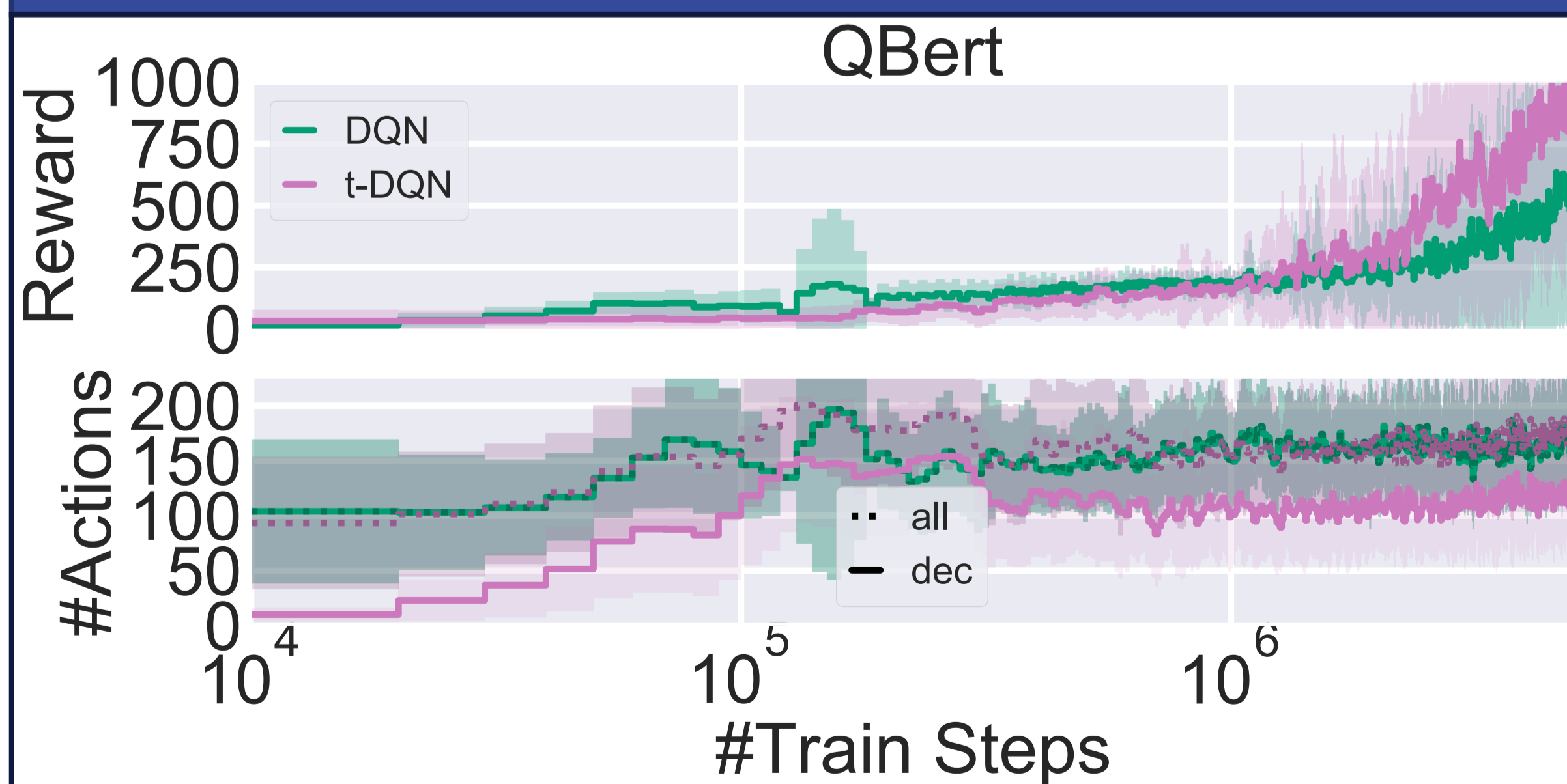TempoRL Q-learning: 3 decisions → proactive policies can be learned much quicker

12.41 × speedup
13.57 × speedup

Moving to Deep RL

Various different architectures to learn what action to take and when to switch

Deep RL - Atari Results

QBert

- Overhead in learning *when* to act
- Our TempoRL DQN lags behind before outperforming vanilla DQN while requiring drastically fewer decision (depicted as *dec* in the Figure)

- In environments where action repetition is highly beneficial TempoRL DQN learns faster and better policies than vanilla DQN

Freeway
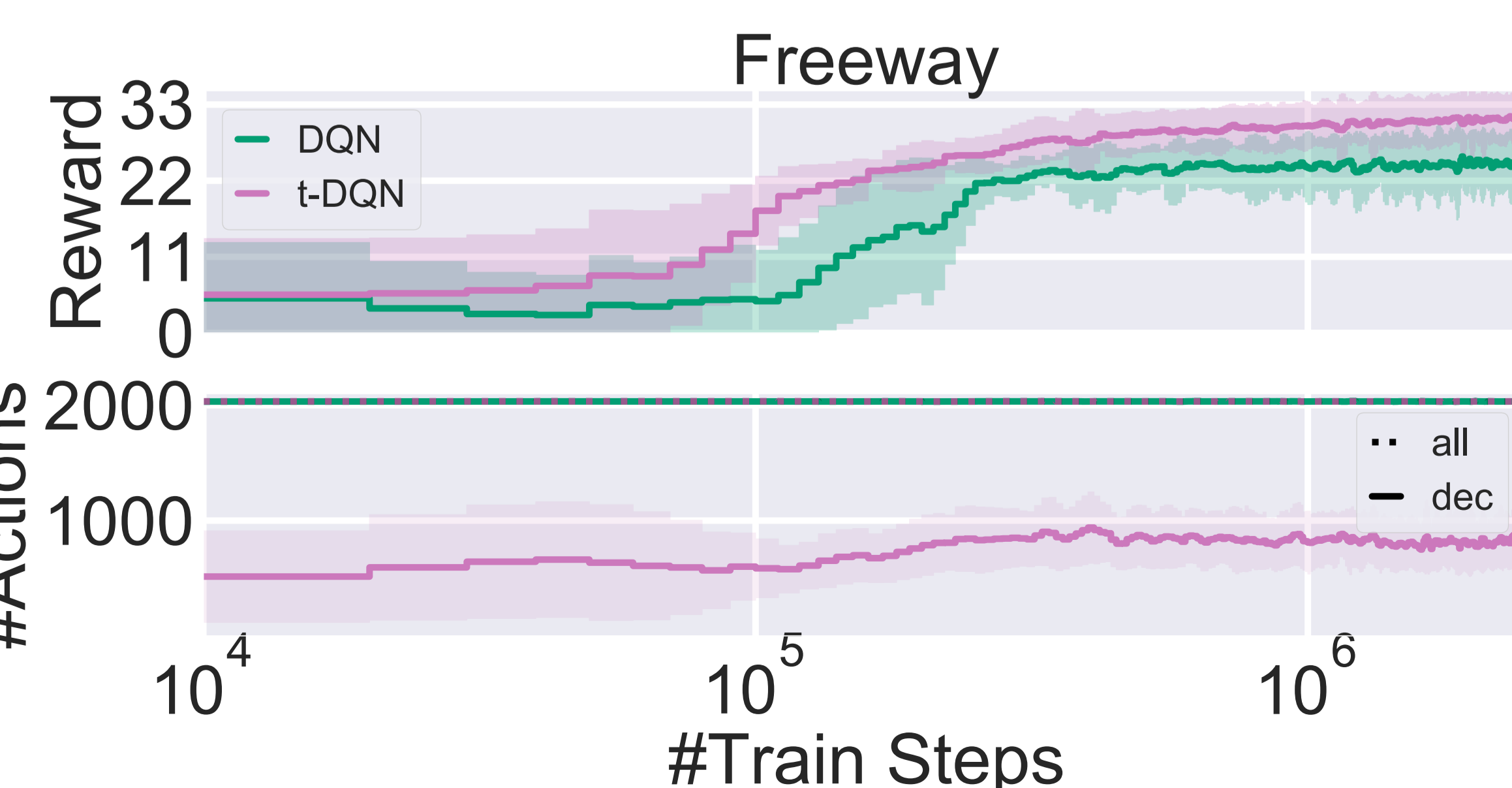
Future Work

- Distributional TempoRL
- Changing TempoRL exploration

AutoML.org

https://github.com/automl/TempoRL