USC University of Southern California

QTAccel: A Generic FPGA based Design for Q-Table based Reinforcement Learning Accelerators

The 27th Reconfigurable Architectures Workshop, 2020

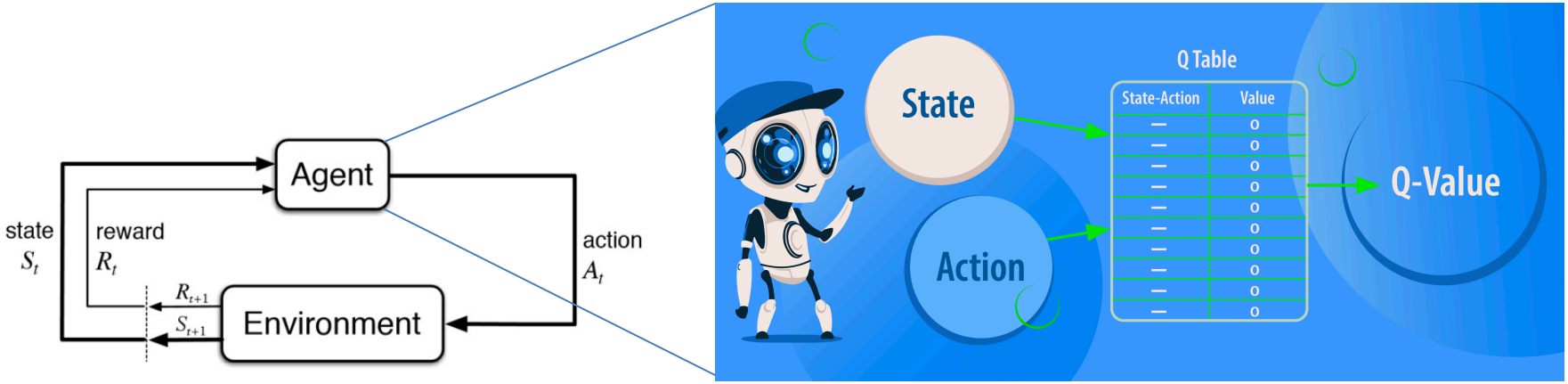
Yuan Meng, Sanmukh Kuppannagari, Rachit Rajat,
Ajitesh Srivastava, Rajgopal Kannany, Viktor Prasanna
Emails: {ymeng643, kuppanna, rrajat, ajiteshs, prasanna}@usc.edu

https://fpga.usc.edu/

University of Southern California

Background: Q-Table based Reinforcement Learning

- Q-values: measuring qualities of state-action pairs
  – stored in a table



- Outperforms Deep Q Network for tractable discrete state space
- Widely used in robotics & games

Introduction: Q learning and SARSA

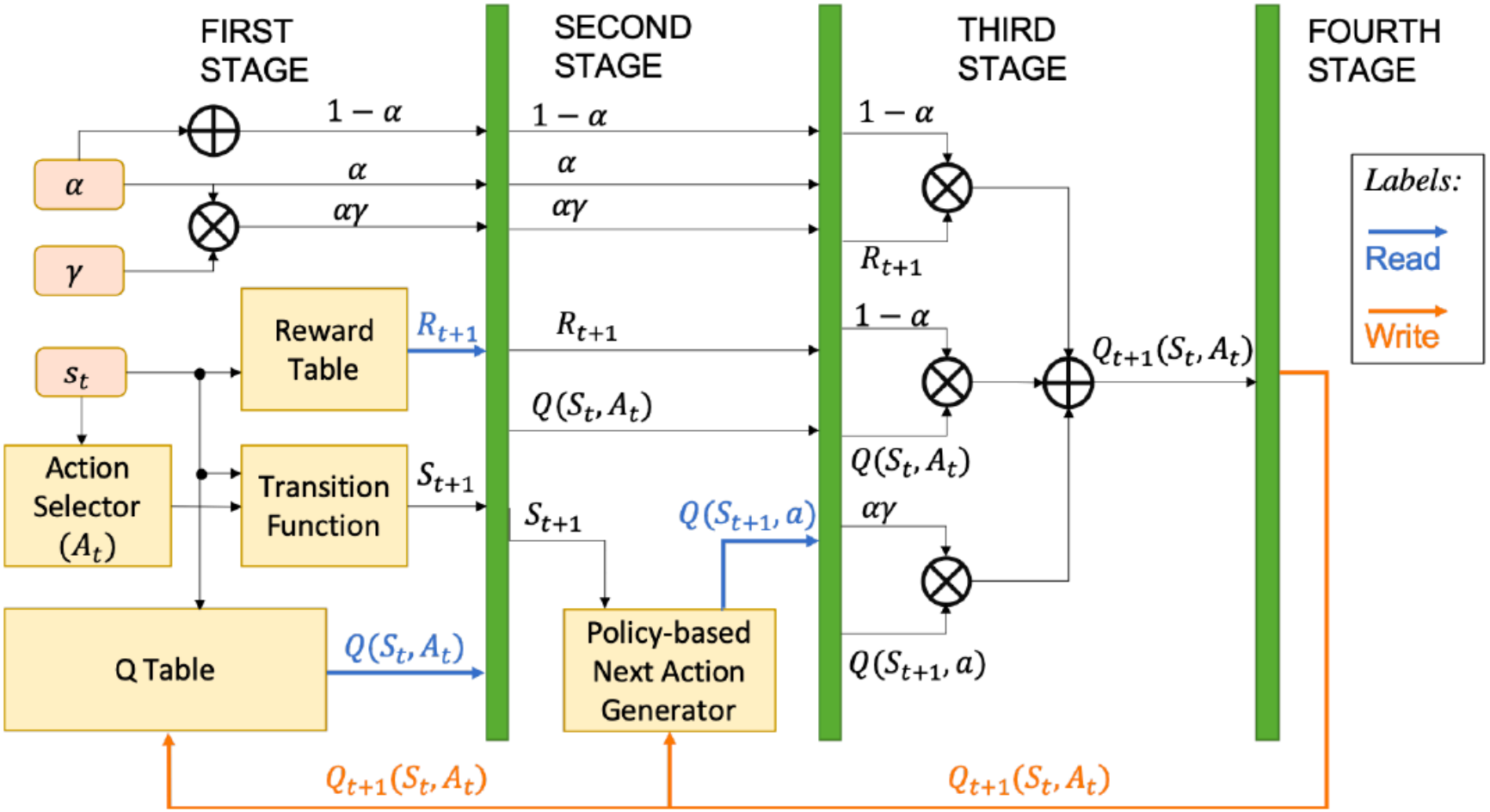
- Q learning: off-policy reinforcement learning algorithm
  - The agent learns optimal policy using absolute greedy policy (maximize future Q value), behaves using another policy
  - ”Q” value is updated through Bell-man equation until convergence

$$Q_{t+1}(S_t, A_t) = Q(S_t, A_t) + \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)]$$

- SARSA: on-policy reinforcement learning algorithm
  - The agent learns optimal policy and behaves using the same policy such as ϵ-greedy policy

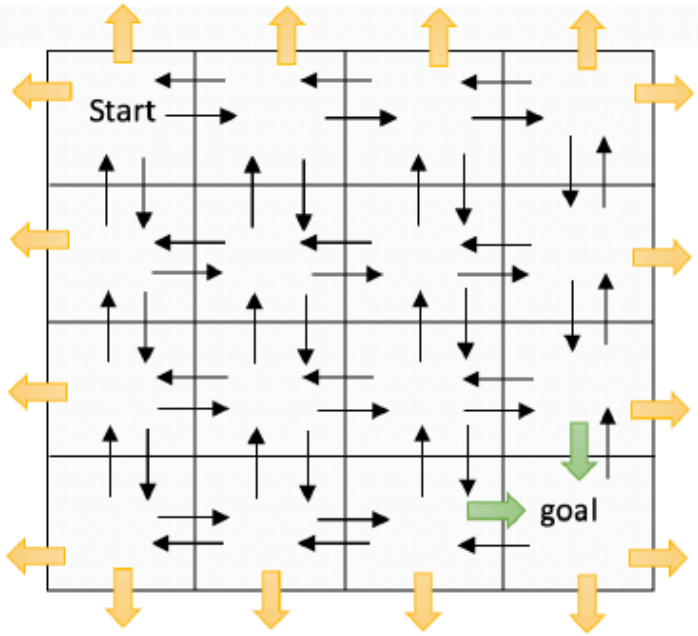
$$Q_{t+1}(S_t, A_t) = Q(S_t, A_t) + \alpha[R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t)]$$

Approach: 4-stage pipeline



Experimental Setup

- Example Grid World Setting

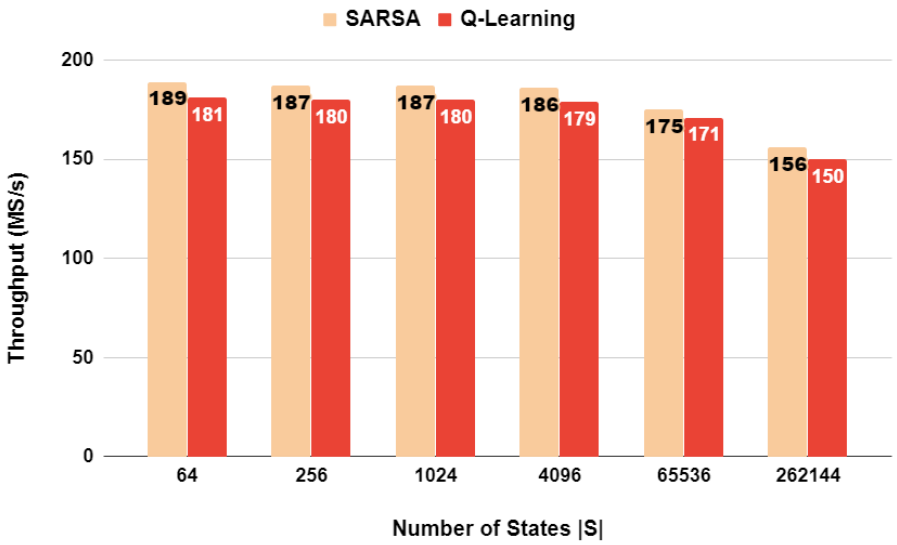


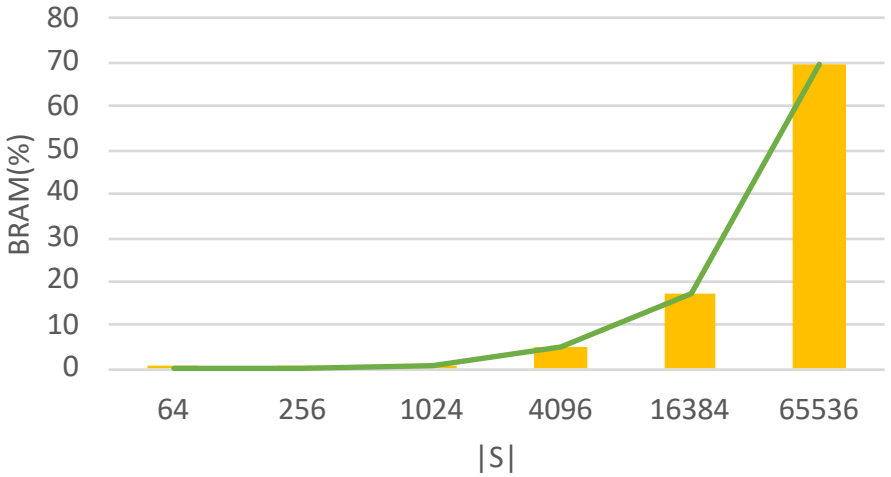
State	Action	Reward
s(x,y)	→	-255
s(x,y)	→	0
s(x,y)	→	+255

Case	1	2	3	4	5	6	7
|S|	64	256	1024	4096	16384	65536	262144
|A|	4,8	4,8	4,8	4,8	4,8	4,8	4,8

Results

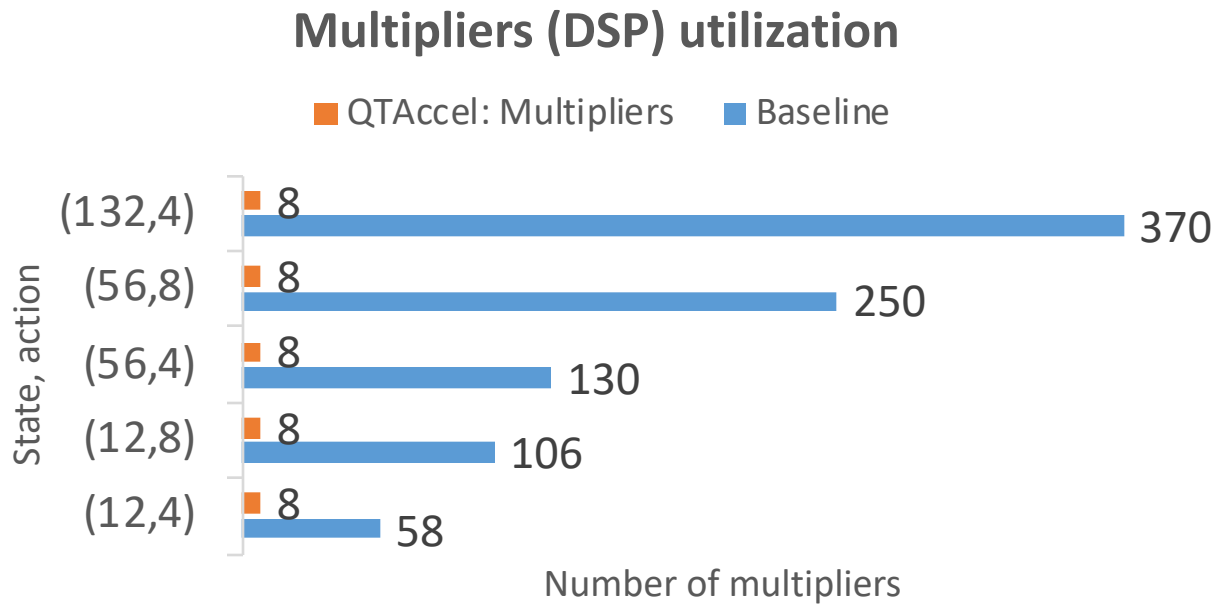
- With 100-200 Mb of on chip memory, our design can support state space of around 1 million
- Application: Well-suited for edge based robotics applications, where state-action space is not huge





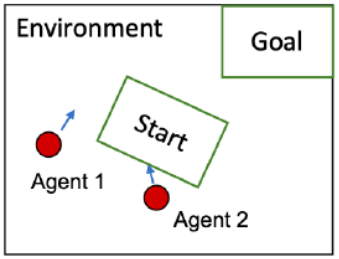
Comparison with State-of-the-art

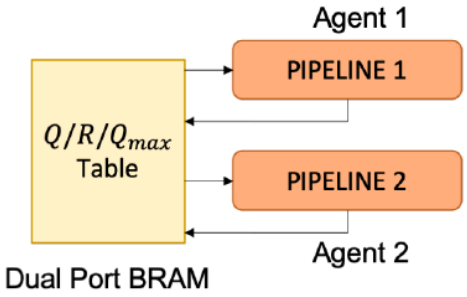
- Low resource utilization compared to state-of-the-art implementation on table-based Q-learning, increasing state-action space does not increase resource utilization



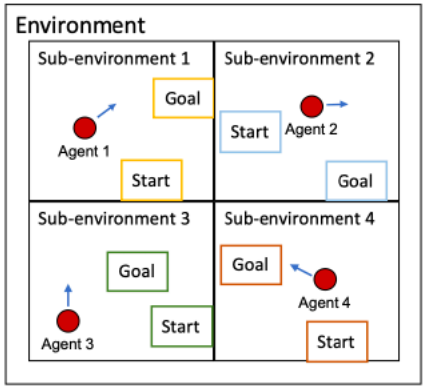
Extension: Multi-agent Settings

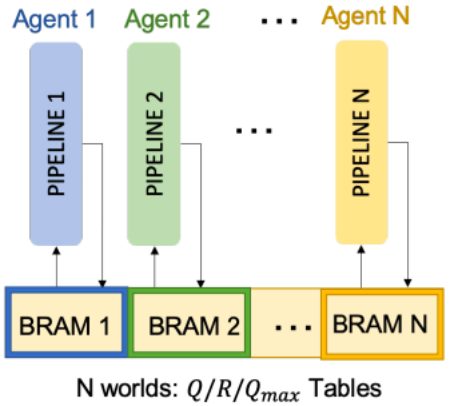
- State Sharing Learners:
  - Two agents cooperates on a task sharing the same environment (i.e. same set of states, actions)





- Independent Learners:
  - Multiple independent agents trained on separate environments





USC University of Southern California

Thank You!

Contact: Yuan Meng (ymeng643@usc.edu)