Gossip-based Actor-Learner Architectures for Deep Reinforcement Learning

NeurIPS'19

Mido Assran, Josh Romoff, Nicolas Ballas, Joelle Pineau, Mike Rabbat

facebook Artificial Intelligence Research

McGill Mila

Motivation

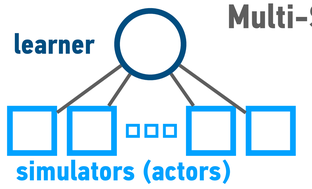
Parallelizing Deep Reinforcement Learning

*Typically tradeoff **computational efficiency** and **sample efficiency***

Want parallelization methods that are both computationally efficient and sample efficient

Current Approaches

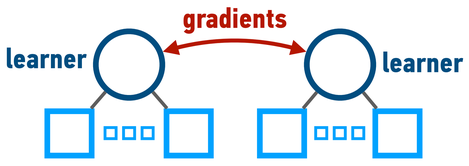
Multi-Simulator Architecture



- *Decorrelate observations*
- *Increase sampling rate*

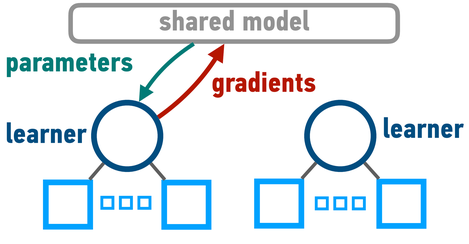
Parallelizing Multi-Simulator Architectures

Option-1 (Canonical Approach: Synchronize Gradients)



- *Average grad. after each update (synchronous)*

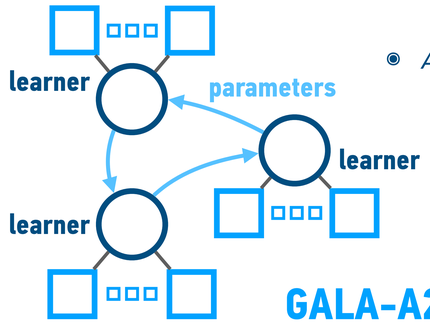
Option-2 (Asynchronous Parameter Server: A3C)



- *Write grad. to shared model (asynchronous)*

Asynchronous methods for deep reinforcement learning, Mnih et al., ICML 2016
OpenAI Baselines, Dhariwal et al., 2017
Recurrent experience replay in distributed reinforcement learning, Kapturowski et al., ICLR 2018
Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures, Espeholt et al., ICML 2018
Accelerated methods for deep reinforcement learning, Stooke and Abbeel, 2019

Gossip-based Actor-Learner Architectures (GALA)

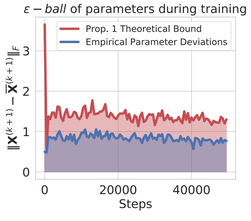


- *Asynchronously average models with neighbours ("gossip")*

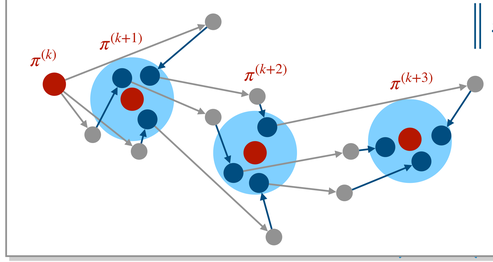
*** Inspired from gossip-algorithms in control-systems lit.*

GALA-A2C

1. *Learner sends policy, π_i, to actors*
2. *Actors step through simulators in parallel and send back observations*
3. *Learner computes policy & critic gradients; updates models*
4. *Asynchronously average parameters with neighbours*



Theoretical ϵ-ball guarantees (GALA)



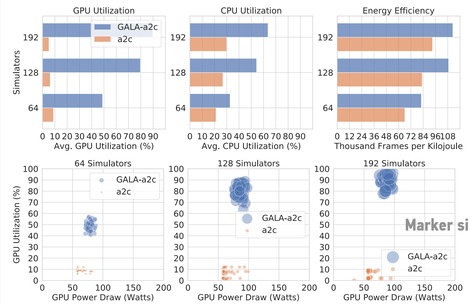
$$\left\| \pi_i^{(k)} - \bar{\pi}^{(k)} \right\| \leq \epsilon = \alpha \frac{\tilde{\beta} L}{1 - \beta}$$

ϵ-ball characterized by: learning-rate, communication graph, upper bound on message delays

Empirical Evaluation

Atari-2600 games; NVIDIA V100 GPU, 48 CPUs

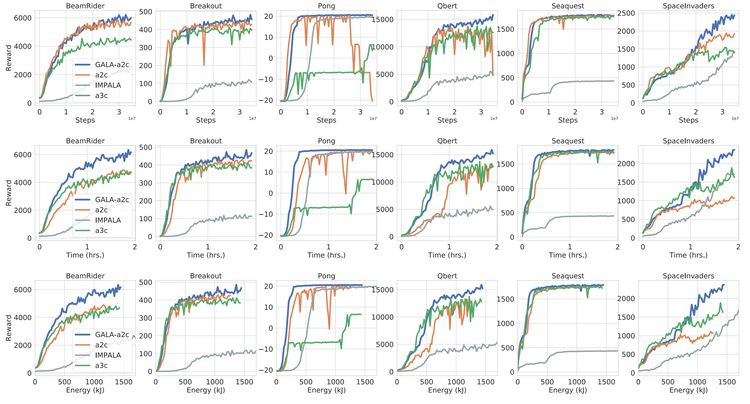
Hardware Utilization & Frame Rates



GALA-A2C improves computational and energy efficiency of **A2C**, and achieves higher frame-rates

Game Scores

GALA-A2C maintains sample efficiency of **A2C** and is competitive with state-of-the-art



Comparing Training Curves for A3C, IMPALA, A2C, and GALA-A2C