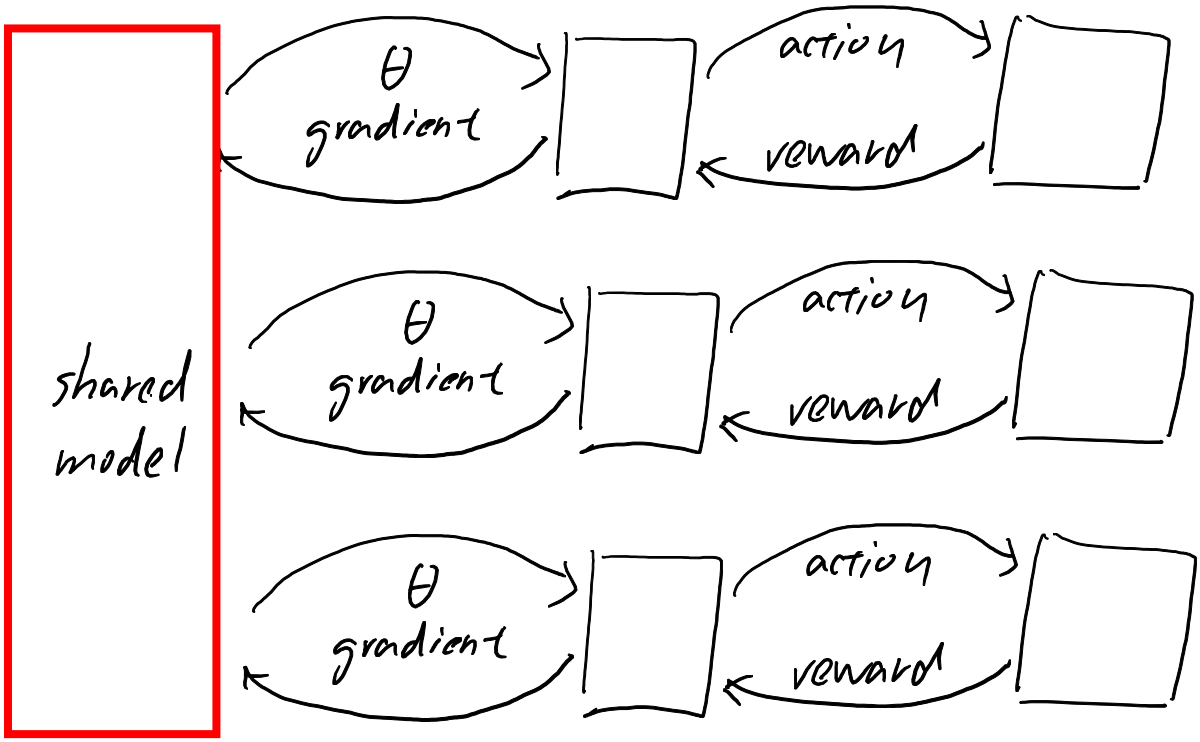
- Distributed GPU training
  - Massive distributed DQN (Gorila)
    - faster
    - enable on-policy
    - enable continuous & recurrent training

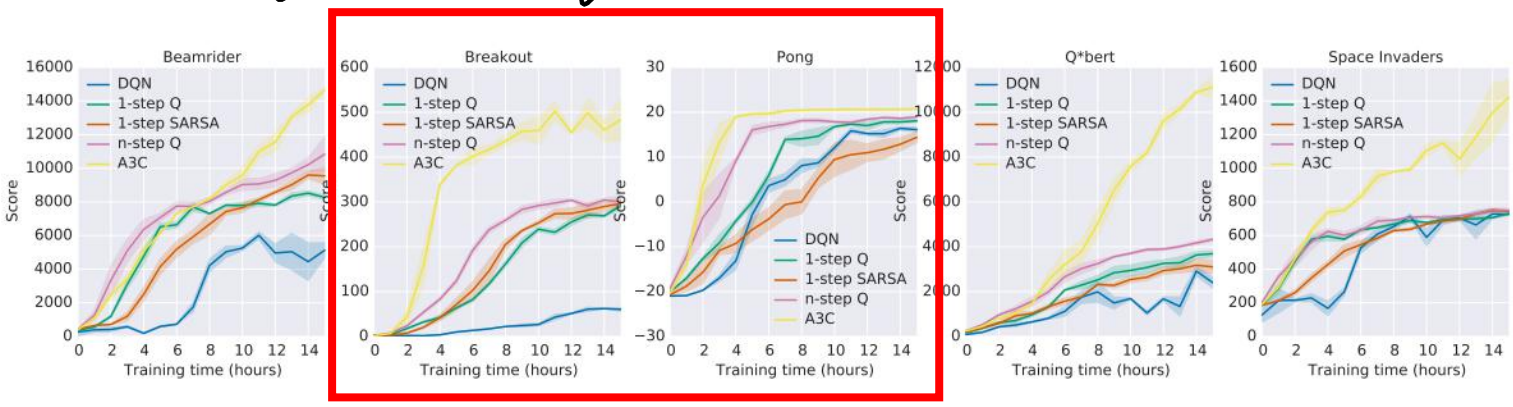
- AsyncRL:



  - works for both QL & PG, on & off policy
  - RMSProp stablize training a lot
  - Applied to 1-step QL, n-step QL, AC

- Test result on Pong & break-out:
  - Set-up: 16 core CPU

- Result: Accuracy:



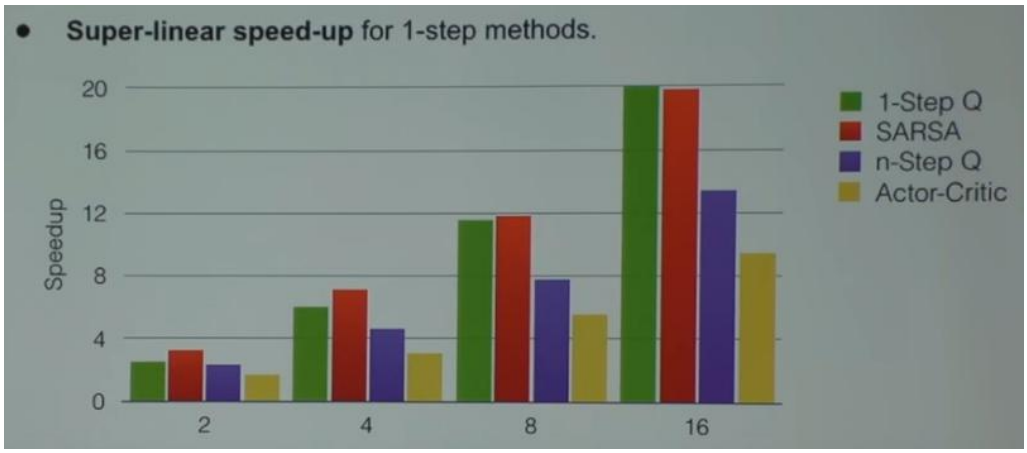
AC is much better, even with CPU training.

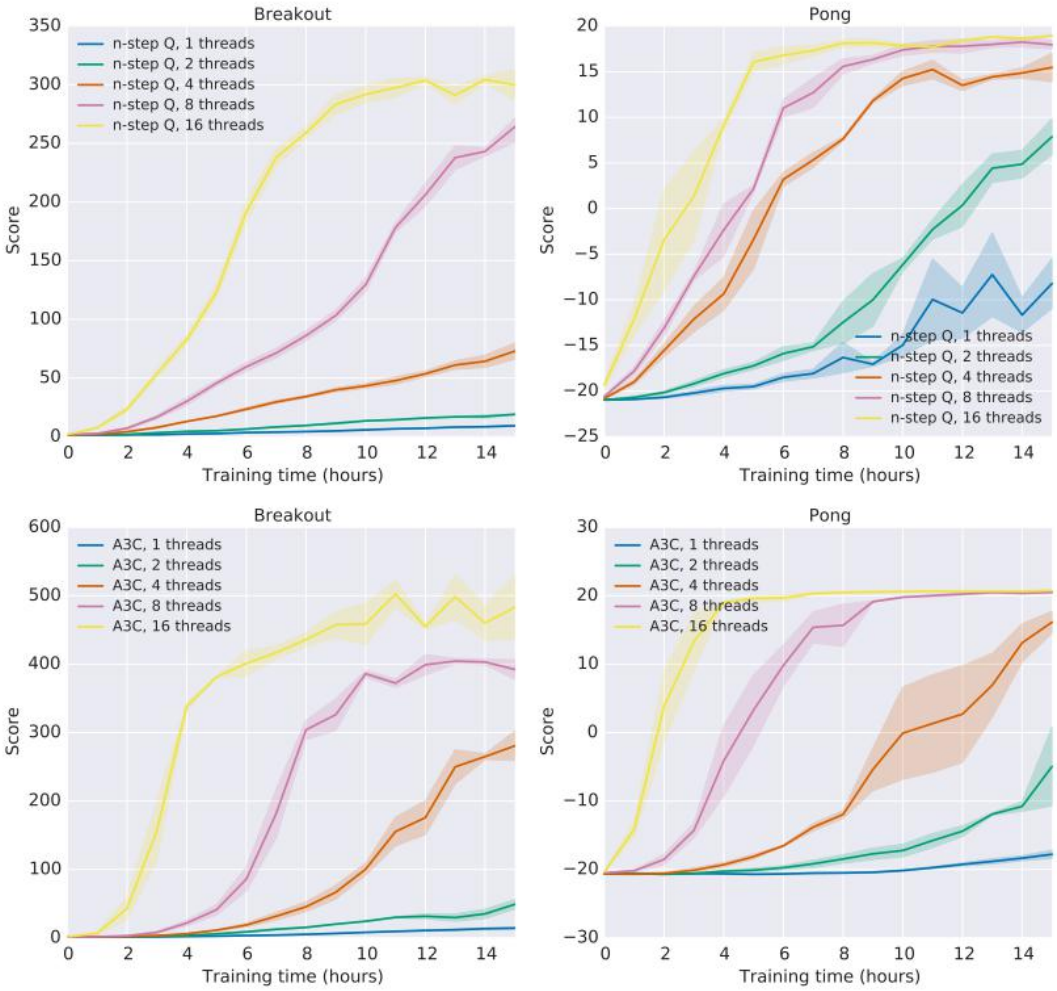
Time:

Method	Training Time	Mean	Median
DQN	8 days on GPU	121.9%	47.5%
Gorila	4 days, 100 machines	215.2%	71.3%
D-DQN	8 days on GPU	332.9%	110.9%
Dueling D-DQN	8 days on GPU	343.8%	117.1%
Prioritized DQN	8 days on GPU	463.6%	127.6%
A3C, FF	1 day on CPU	344.1%	68.2%
A3C, FF	4 days on CPU	496.8%	116.6%
A3C, LSTM	4 days on CPU	623.0%	112.6%

compared to human

Time VS # of threads:





RNN
LSTM is needed to play games that requires memory, ie Maze