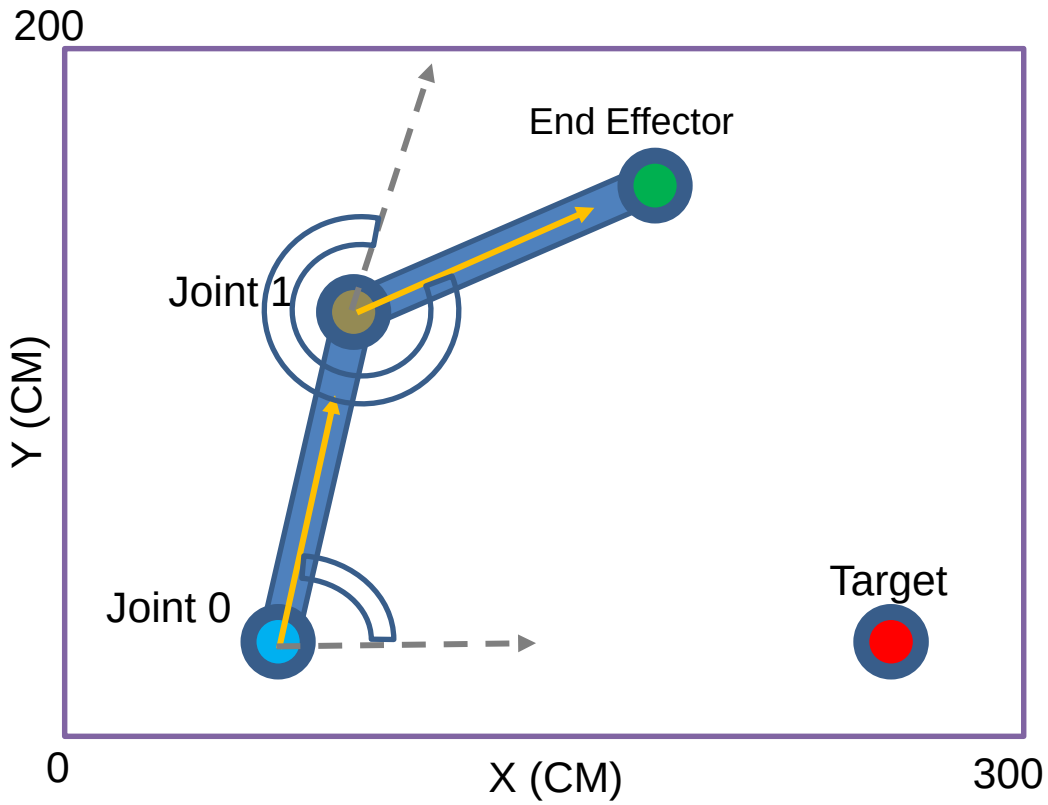
Robot control with Deep Reinforcement Learning

Hadi Beik-Mohammadi

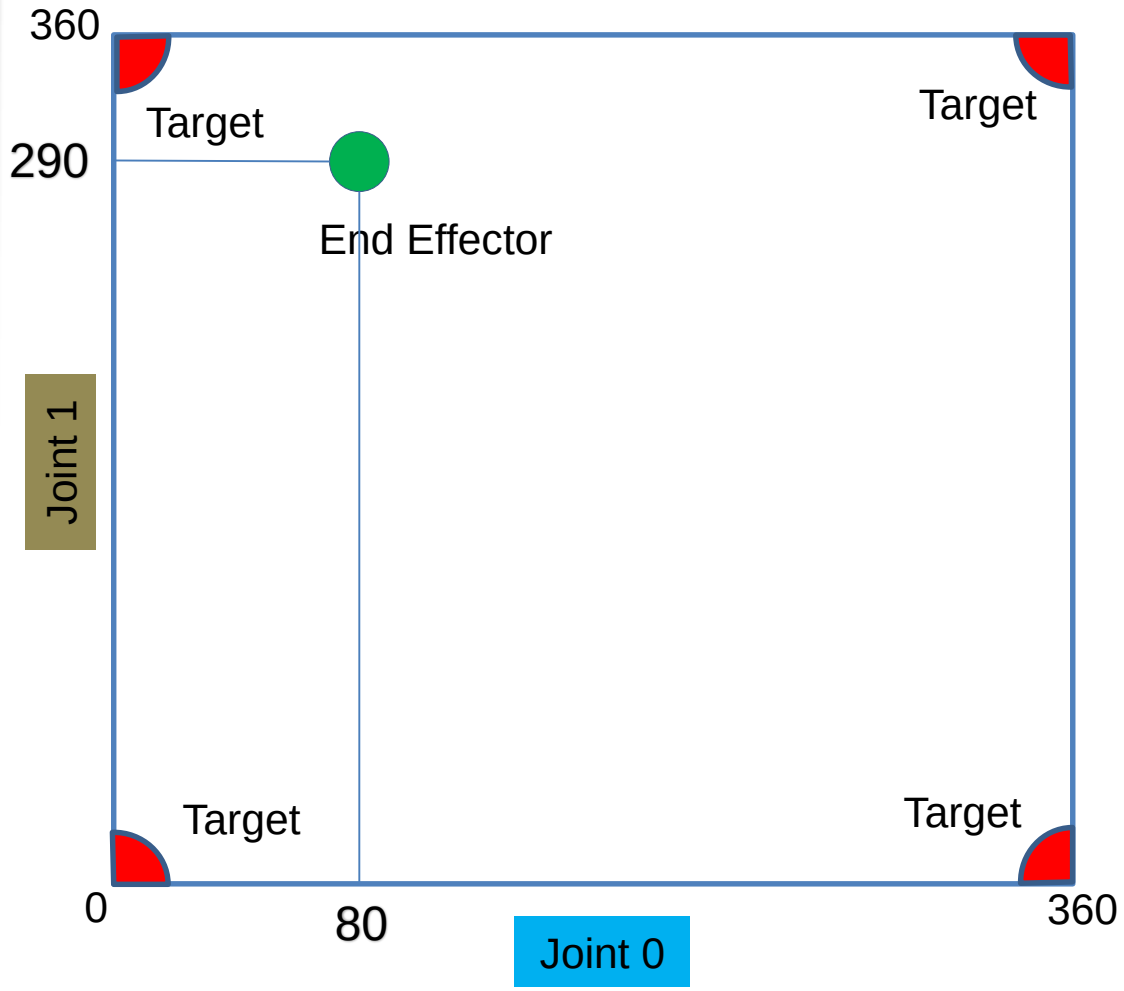
INTELLIGENT ROBOTICS SEMINAR TALK, DECEMBER 2017

- Inverse and Forward Kinematic
- How to Learn a behavior
- Methods
  - Inverse Recurrent Model
  - Deep Deterministic Policy Gradient
- Conclusion

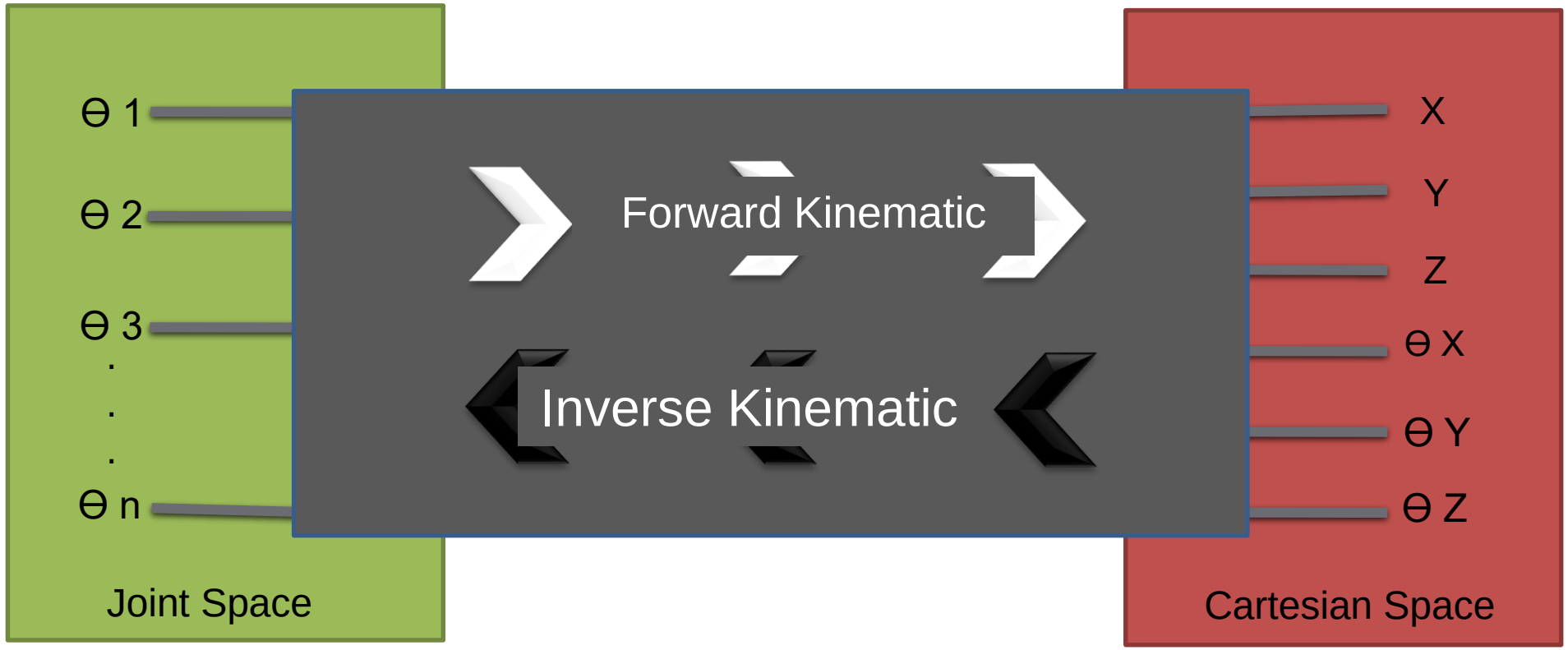
Cartesian Space



Joint Space



Forward and Inverse Kinematic



The brain evolved, not to think or feel, but to control movement

Daniel Wolpert, nice TED talk

How to build agents that learn behaviors in a dynamic world?

Learning a behavior:
Learning to map sequences of observations to actions, for a particular goal

What supervision does an agent need to learn purposeful behaviors in dynamic environments?

- Rewards: sparse feedback from the environment whether the desired behavior is achieved
- Demonstrations
- Specifications/Attributes of good behavior

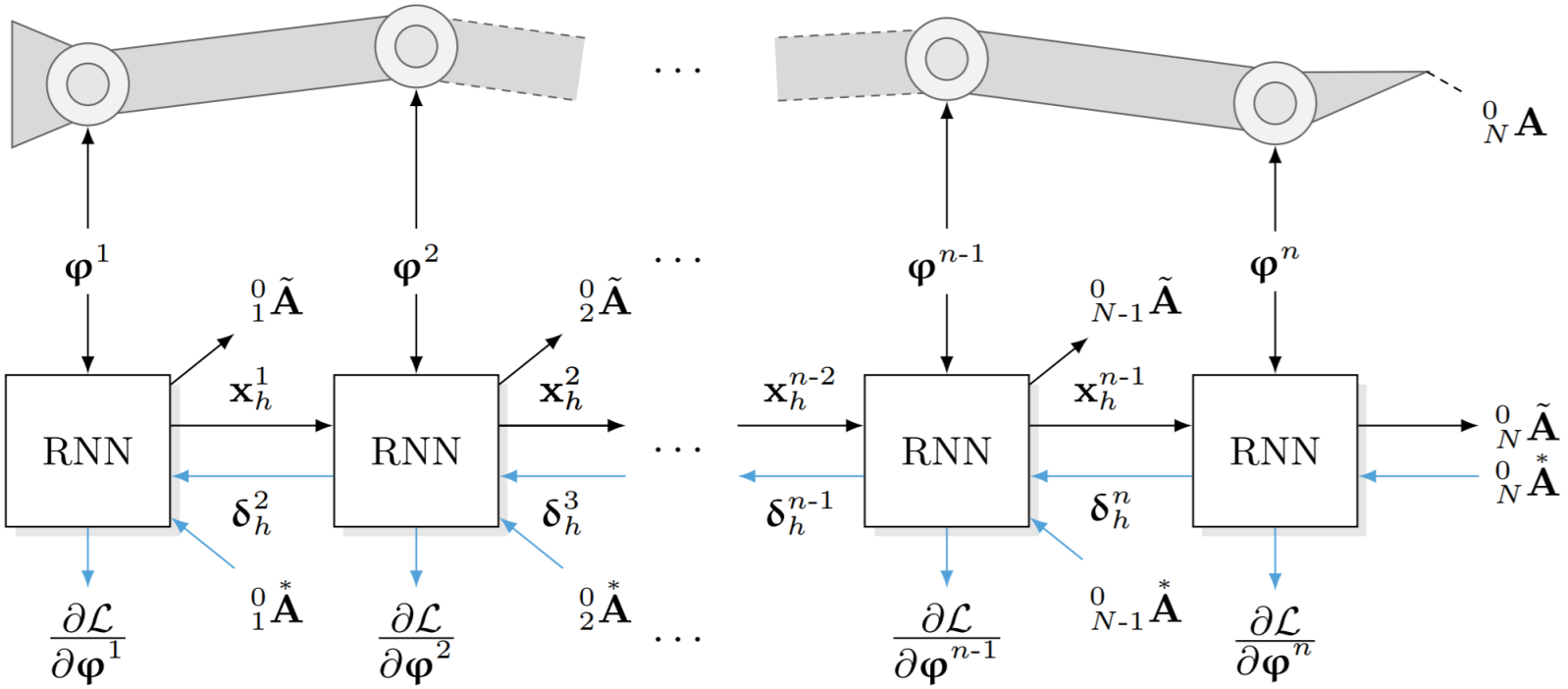
Inverse Recurrent Model (IRM)[1]

- Control Snake-Like Many Joint Robot Arms
- BPTT on recurrent forward models
- Recurrent Neural Networks LSTM
- Offline

Deep Deterministic Policy Gradient (DDPG)[2]

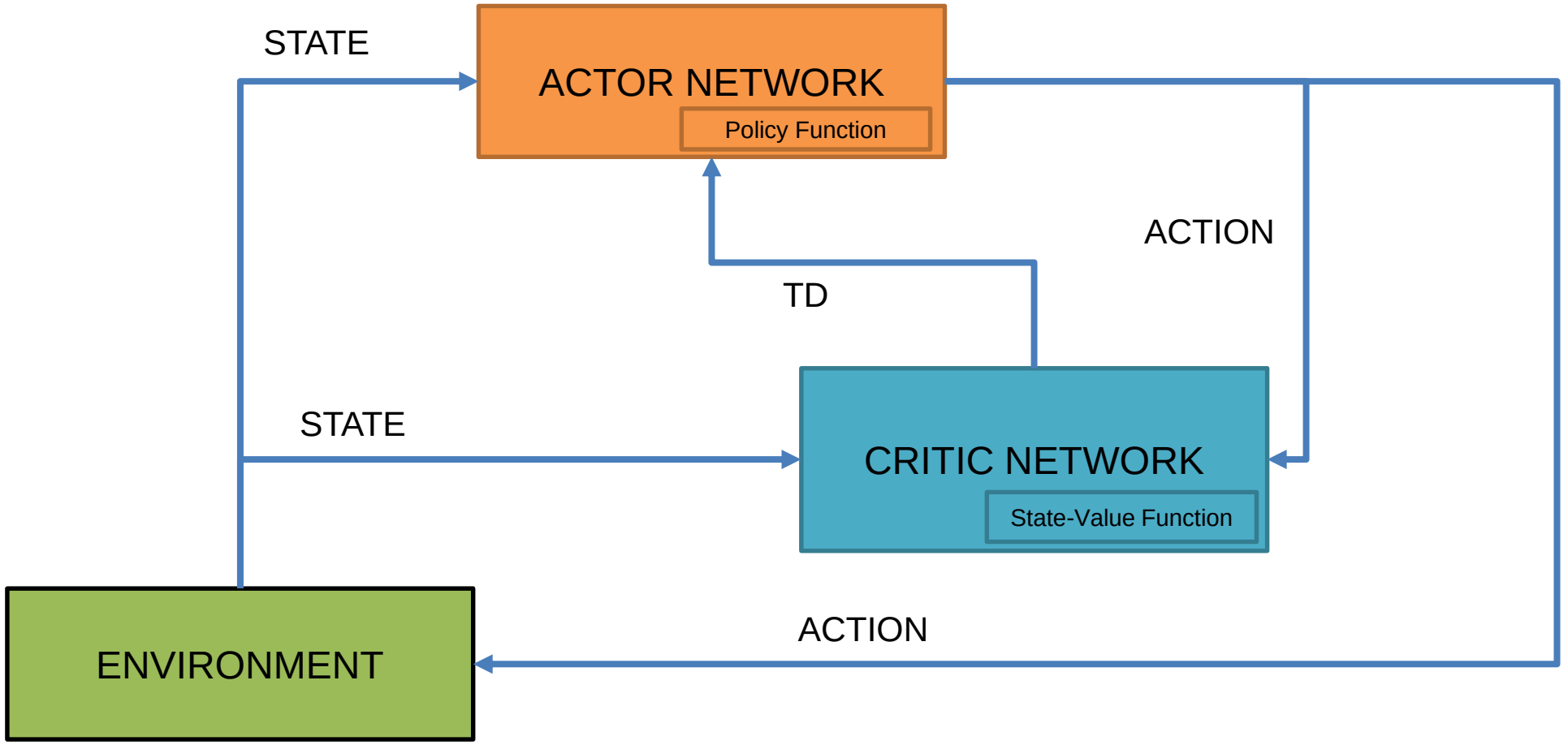
- Deep Reinforcement Learning Method
- Actor Critic Network
- Continuous Action Domain
- Model Free
- Online

Inverse Recurrent Model (IRM)



[1]

Deep Deterministic Policy Gradient (DDPG)

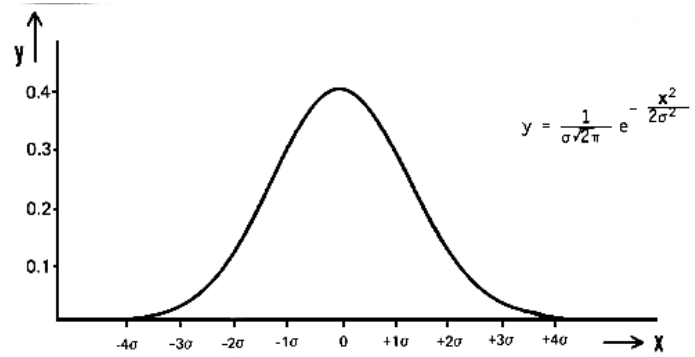


Deep Deterministic Policy Gradient (DDPG)

https://youtu.be/tJBIqkC1wWM

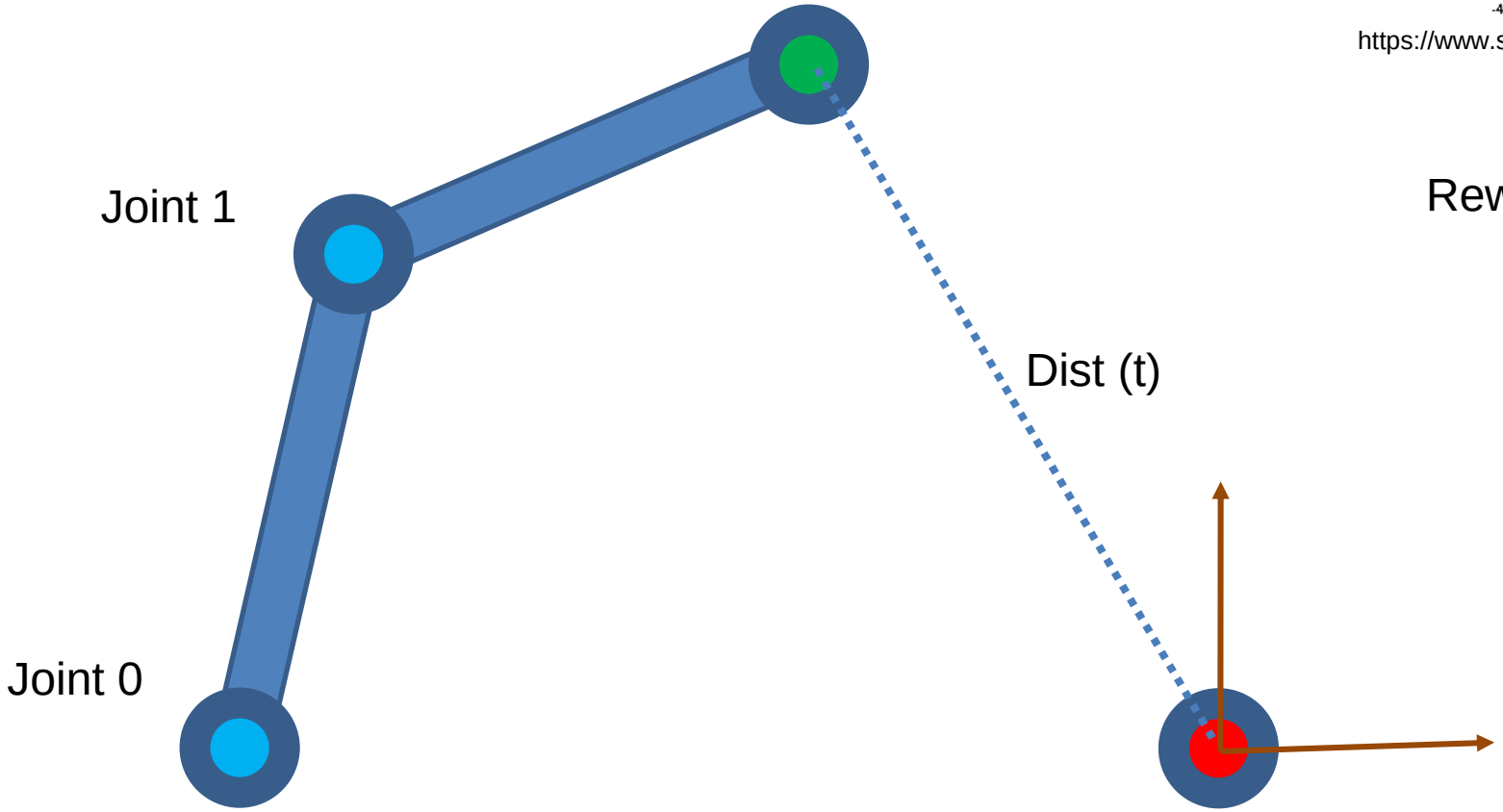
Rewarding

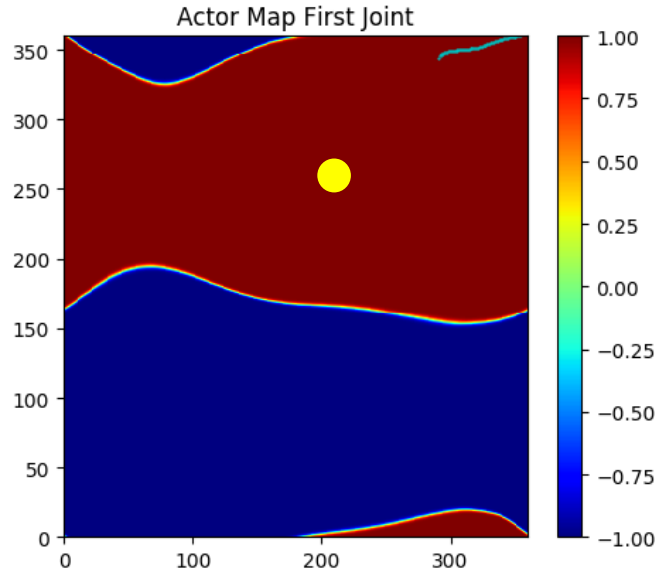
Reward 1 = Gaussian(Dist(t))

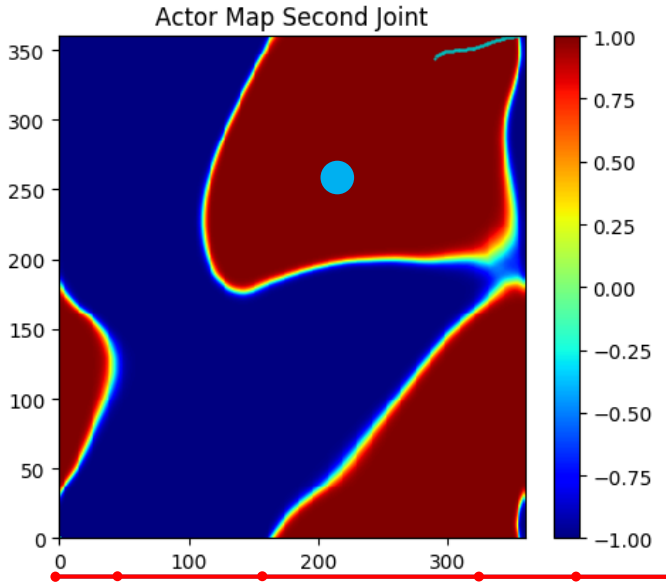


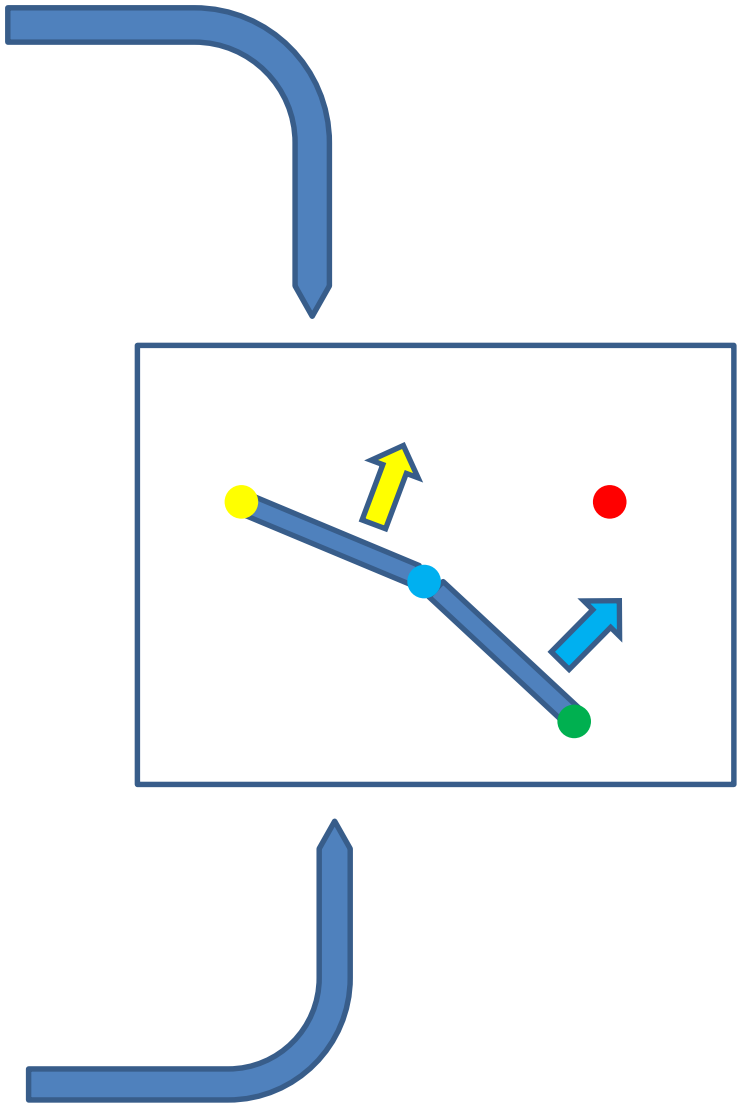
https://www.sfu.ca/sonic-studio/handbook/Graphics/Gaussian.gif

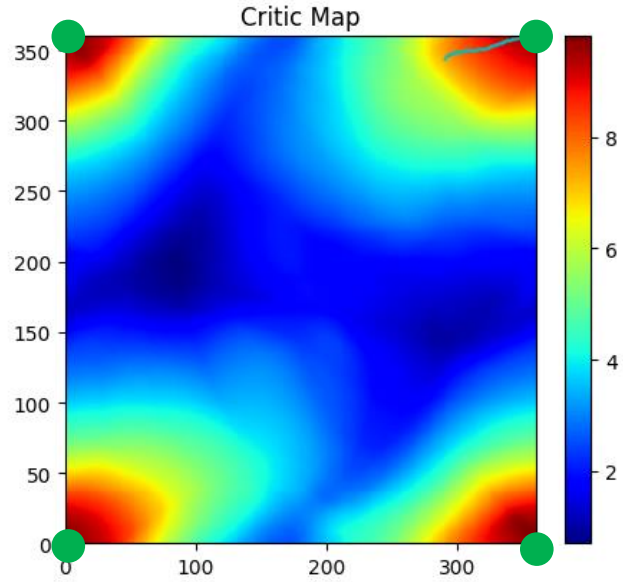
Reward 2 = Dist(t-1) – Dist(t)



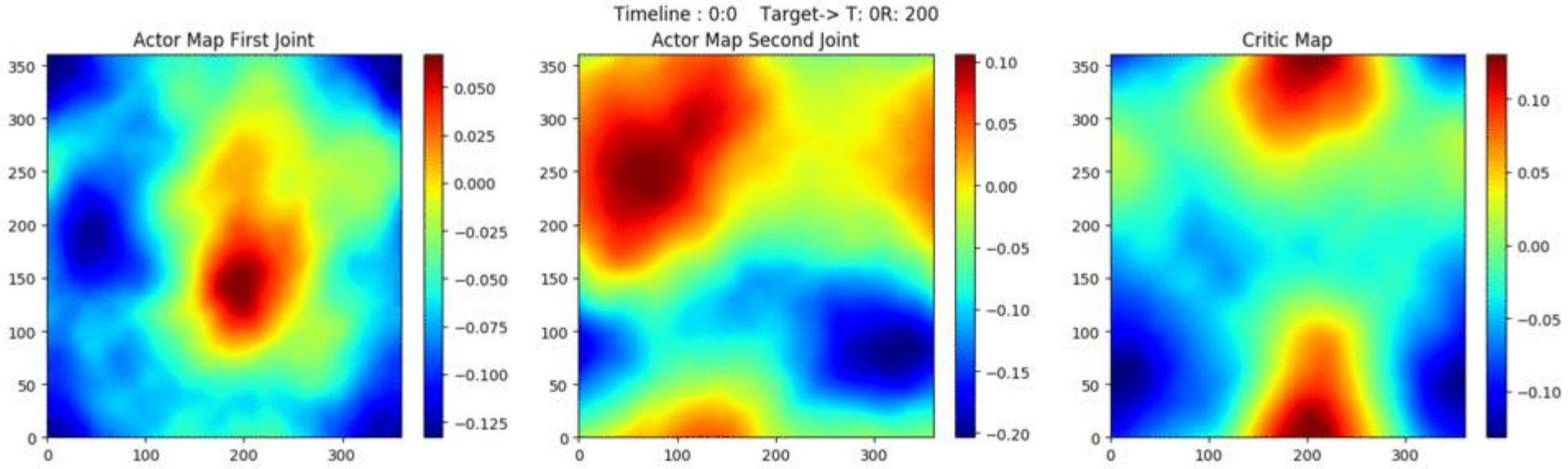

Actor Map First Joint
350
300
250
200
150
100
50
0
0
100
200
300
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Actor Map Second Joint
350
300
250
200
150
100
50
0
0
100
200
300
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00




Critic Map
350
300
250
200
150
100
50
0
0
100
200
300
8
6
4
2

2 DOF Manipulator Actor Critic maps during Learning



Pros:

- Operate over continuous action spaces
- Algorithm can learn policies end-to-end
- Model-Free

Cons:

- No Proof for learning
- No Guarantee for results
- Requires a large number of training episodes to find solutions

References

- [1] Sebastian Otte , Adrian Zwiener , and Martin V. Butz, Inherently Constraint-Aware Control of Many-Joint Robot Arms with Inverse Recurrent Models
- [2] Continuous control with deep reinforcement learning, Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, Daan Wierstra
- [3] Deep Reinforcement Learning and Control, Spring 2017, CMU 10703

INTELLIGENT ROBOTICS SEMINAR TALK, DECEMBER 2017