Interferobot: aligning an optical interferometer by a reinforcement learning agent

M1 Hiroki Fujimoto

Oct. 2 @Ando Lab Seminar

Paper

- https://arxiv.org/abs/2006.02252

Interferobot: aligning an optical interferometer by a reinforcement learning agent

Dmitry Sorokin[1,4]
dmitrii.sorokin@phystech.edu

Alexander Ulanov[1]
a.e.ulanov@gmail.com

Ekaterina Sazhina[1,4]
sazhinaekaterina@gmail.com

Alexander Lvovsky[1,2,3]
alex.lvovsky@physics.ox.ac.uk

[1]*Russian Quantum Center, Moscow, Russia* [2]*University of Oxford, United Kingdom*
[3]*P. N. Lebedev Physics Institute, Moscow, Russia* [4]*Moscow Institute of Physics and Technology*

Abstract

Limitations in acquiring training data restrict potential applications of deep reinforcement learning (RL) methods to the training of real-world robots. Here we train an RL agent to align a Mach-Zehnder interferometer, which is an essential part of many optical experiments, based on images of interference fringes acquired by a monocular camera. The agent is trained in a simulated environment, without any hand-coded features or a priori information about the physics, and subsequently transferred to a physical interferometer. Thanks to a set of domain randomizations

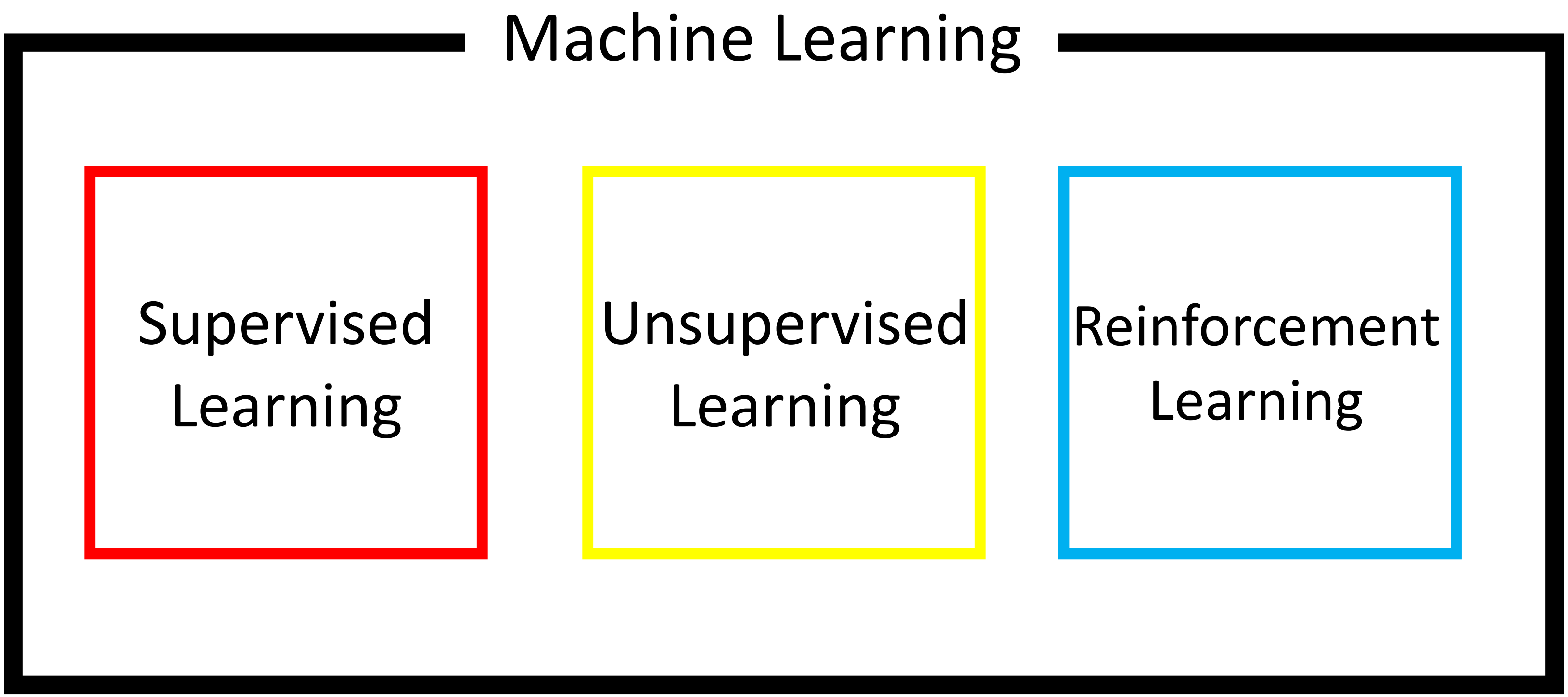
Contents

- Introduction of Reinforcement Learning
  - ◆What is Reinforcement Learning?
  - ◆Defining the Reinforcement Learning problem
- Interferobot:
  - ◆What is Interferobot?
  - ◆Training the Interferobot in simulated environment
  - ◆Experiment with physical environment

Reinforcement Learning

What is Reinforcement Learning?

There are three basic fields in Machine learning



What is Reinforcement Learning?

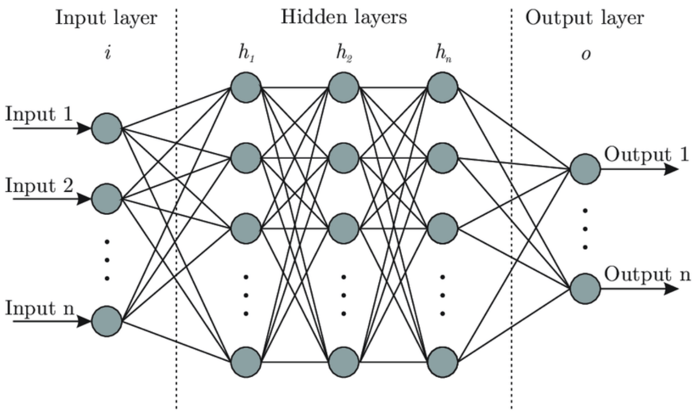
- Supervised learning

Supervised learning is used for analysis or prediction.

Predict output for the input.

Input: image

Neural Network



(https://www.researchgate.net/publication/321259051_Prediction_of_wind_pressure_coefficients_on_building_surfaces_using_Artificial_Neural_Networks)

Output: prediction

Cat : 95%
Lion : 5%

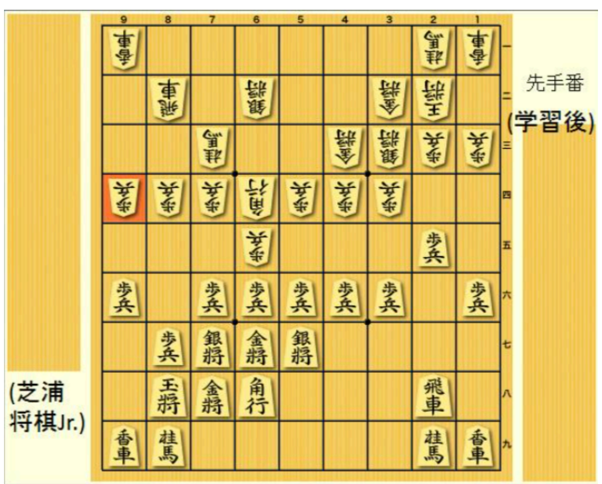
What is Reinforcement Learning?

- Reinforcement learning (RL)

Reinforcement learning is used to control the environment.

Agent learns optimal action through interaction with the environment.

environment



(ipsj.ixsq.nii.ac.jp)

Agent

recognizes the state of the environment

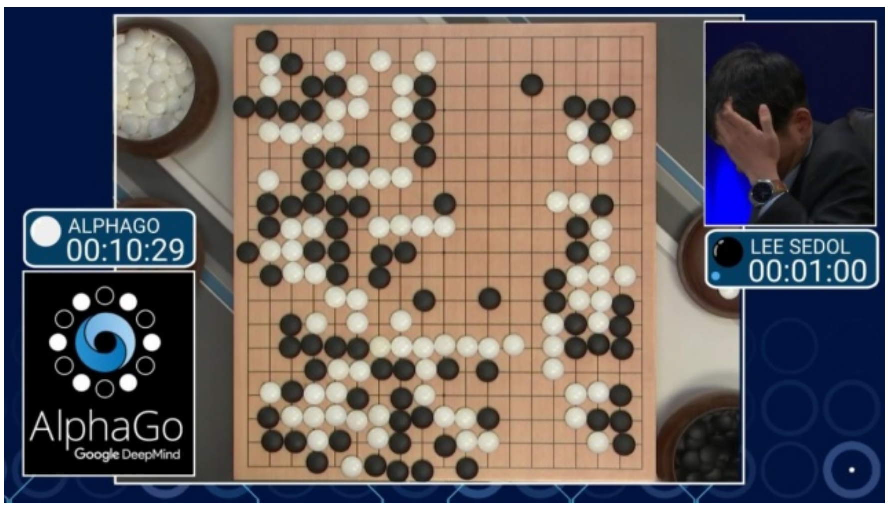
selects the optimal action

Take the optimal action

What is Reinforcement Learning?

- Examples in games

Alpha Go



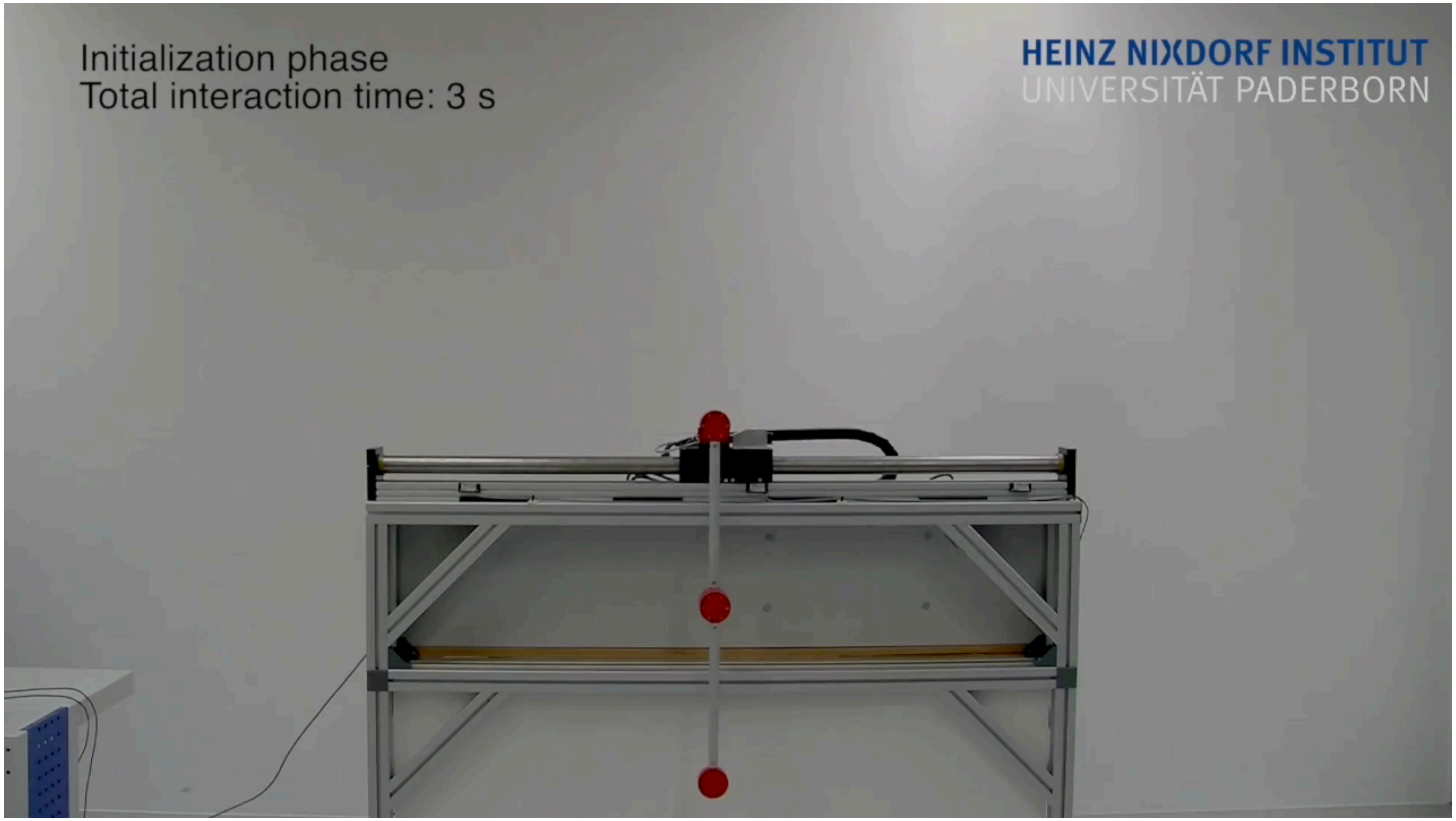
(https://www.engadget.com/2016/03/12/watch-alphago-vs-lee-sedol-round-3-live-right-now/?_ga=2.241475077.649112291.1601463824-230945953.1601463824)

Agent57

(https://arxiv.org/abs/2003.13350)

What is Reinforcement Learning?

- Example in physical robotics



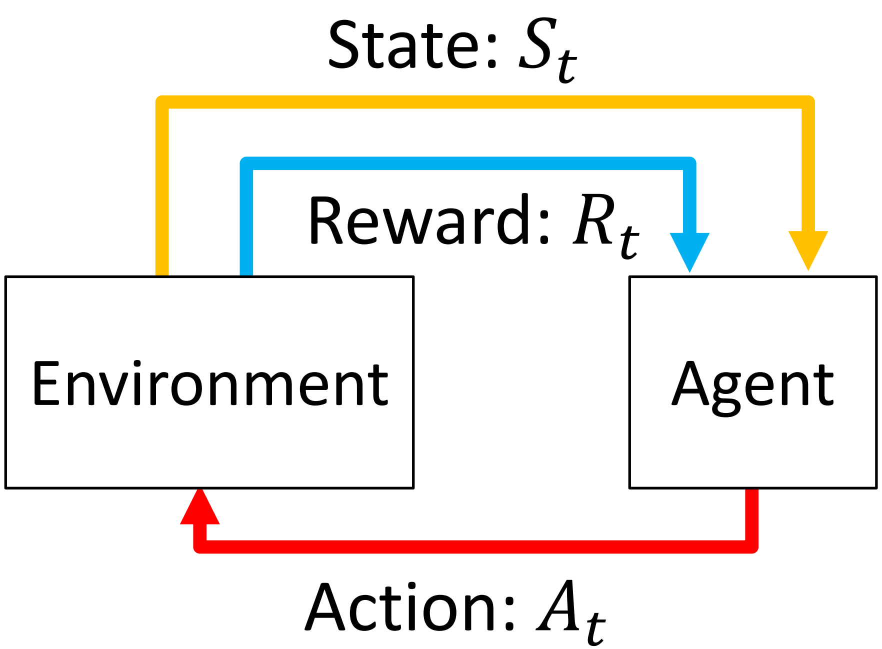
[A Reinforcement Learning Strategy for the Swing-Up of the Double Pendulum on a Cart, Procedia Manufacturing, Volume 24, 2018, Pages 15-20]

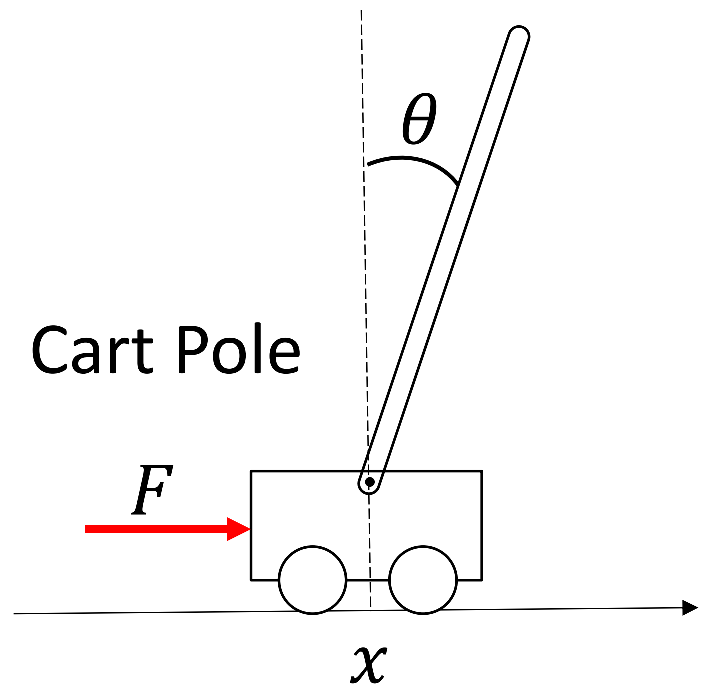
Defining the RL problem

State of cart pole: $S_t = (\theta, \dot{\theta}, x, \dot{x})$

Reward to Agent: $R_t = \cos\theta$

Actions that agent takes: $A_t \in \{\pm F_0\}$



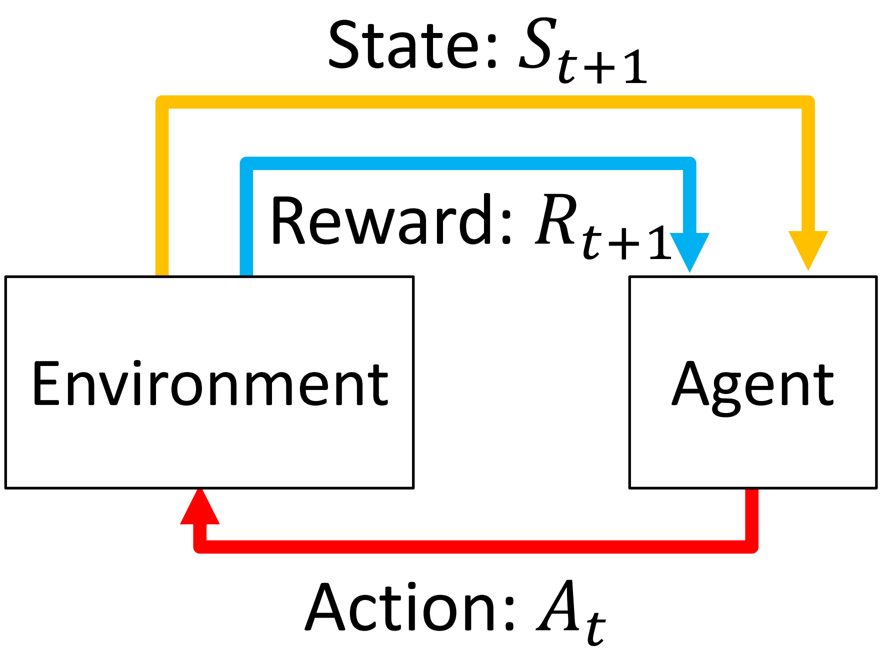


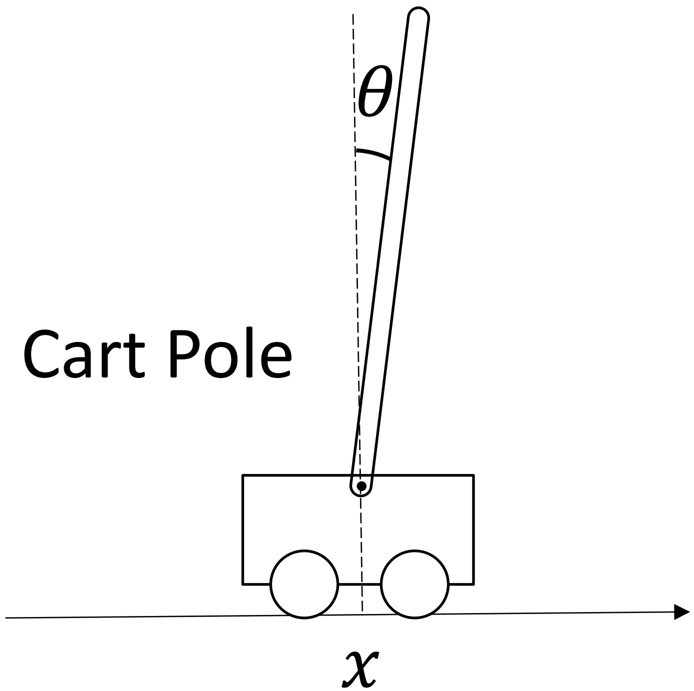
Defining the RL problem

State of cart pole: $S_t = (\theta, \dot{\theta}, x, \dot{x})$

Reward to Agent: $R_t = \cos\theta$

Actions that agent takes: $A_t \in \{\pm F_0\}$



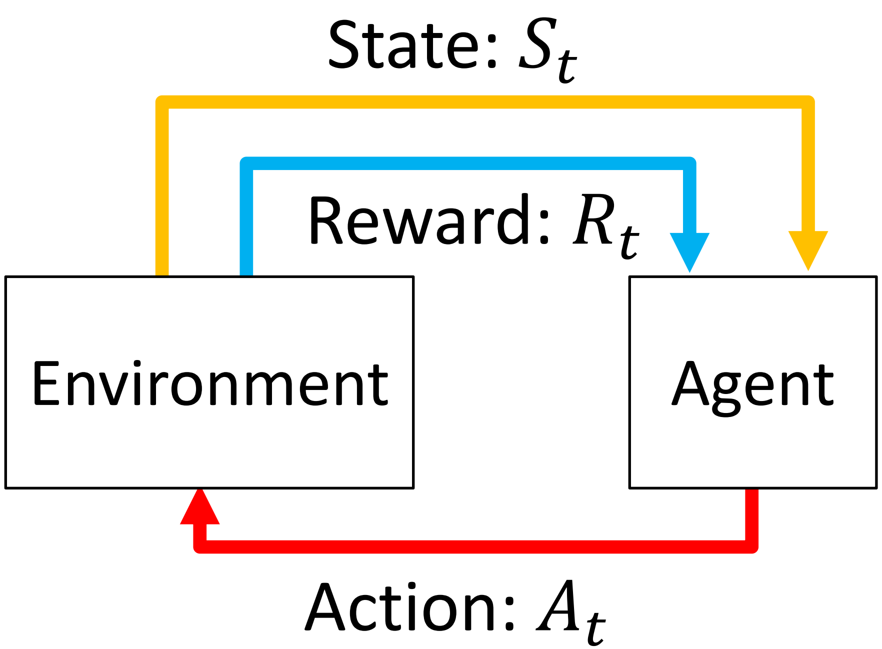


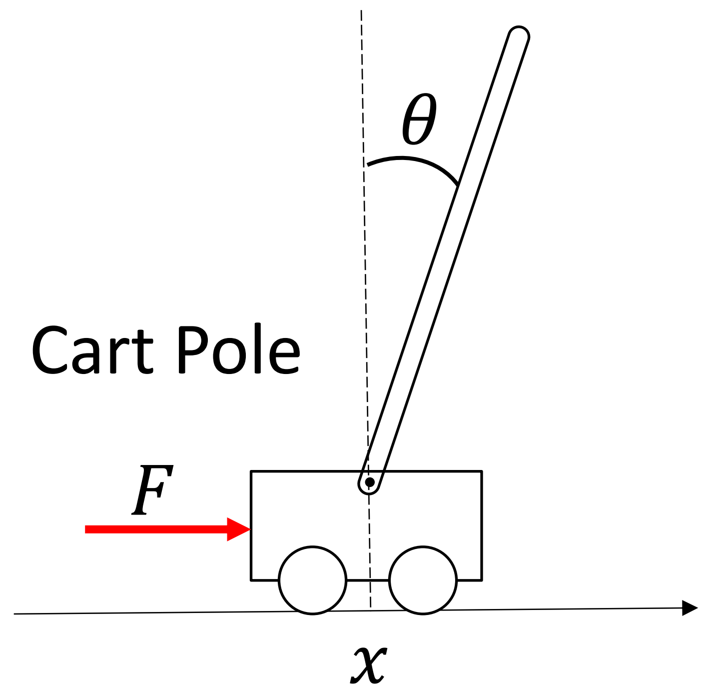
Defining the RL problem

Total rewards (return): $G = \sum_{t=0}^{T} R_t$

We want the agent that obtains the largest return.

Many algorithms: Q-Learning, SARSA, Policy gradient method, Actor-Critic method, etc.





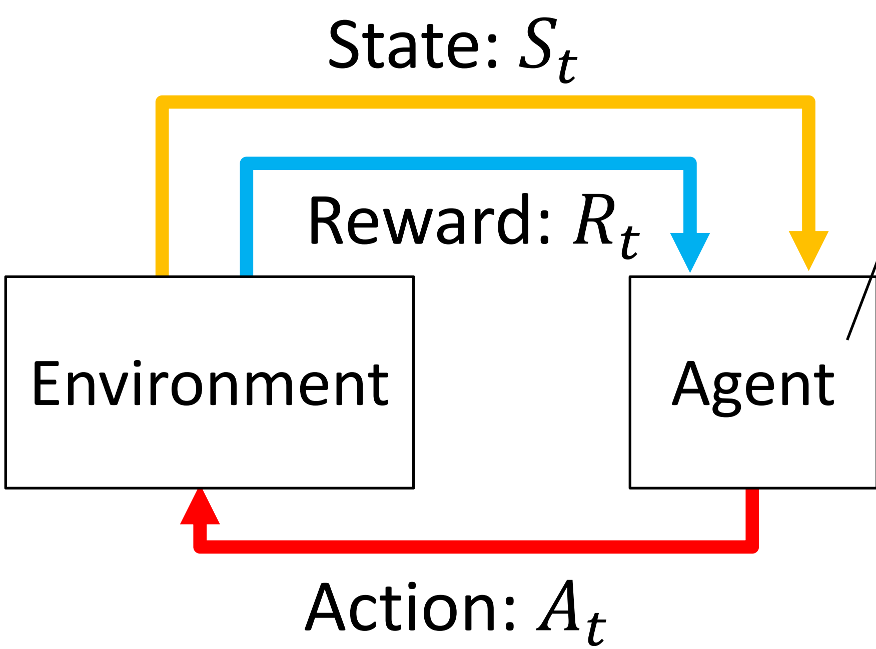
Q-Learning

Total rewards (return): $G = \Sigma_{t=1}^{T} R_t$

State-action value function (Q function):

$$Q(s, a) = \mathbb{E}[G | S_0 = s, A_0 = a]$$

≈Value of choosing action: a when the state is s



Based on observed R_t and Bellman eq., <u>update</u> the estimated $Q(s, a)$

⬇

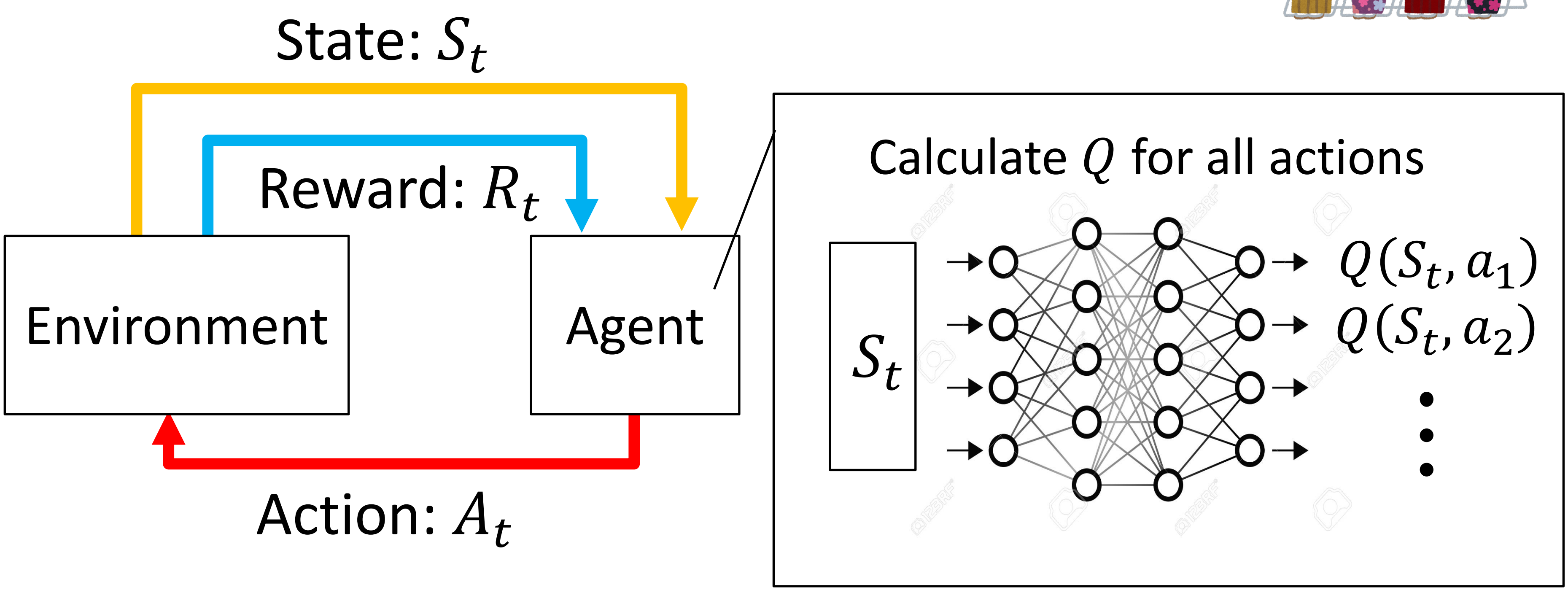
Choose $arg \max_{a} Q(s, a)$

Deep Q Network

- Deep Q Network (DQN)

  Combination of Q-Learning and Deep Learning

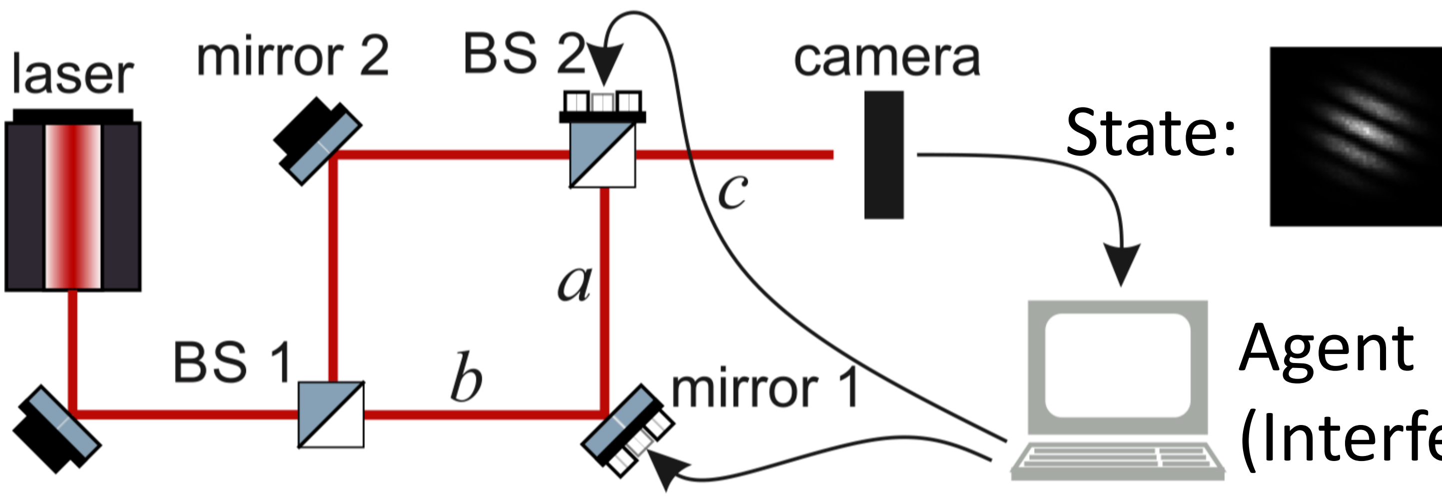
  Estimate $Q(s, a)$ with Neural Network



Training the Interferobot

What is Interferobot?

Interferobot is an agent trained to align Mach-Zehnder interferometer.



Agent (Interferobot)

Action: tilting mirror 1 and BS 2

Training in simulated environment

Agent is trained through interaction (trial and error) with the environment.

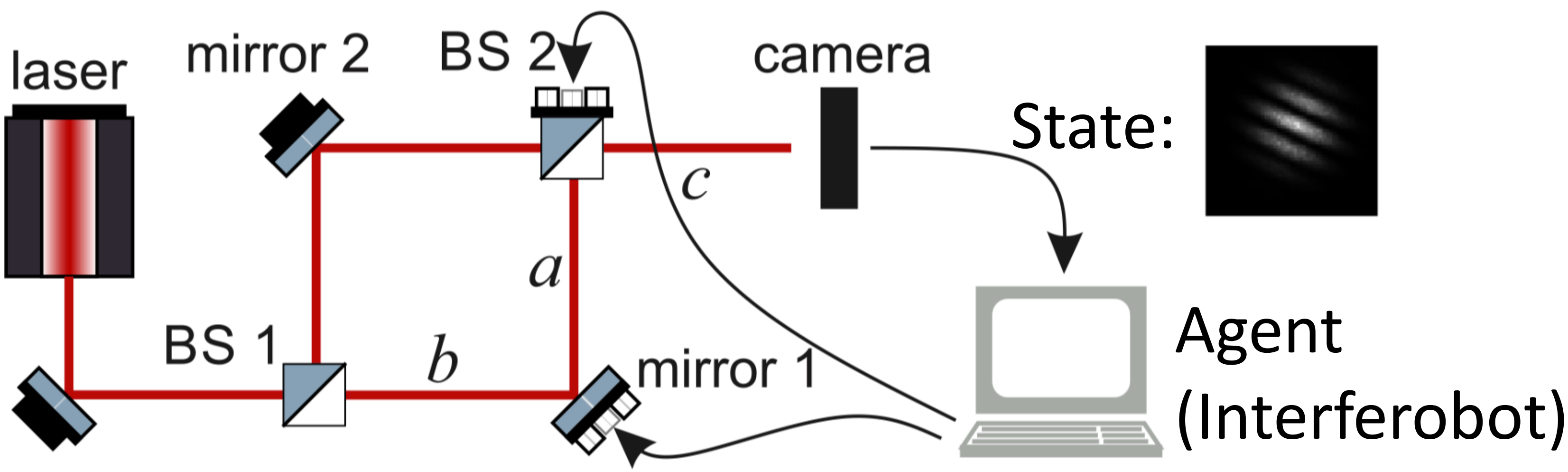
Training in physical environment takes long time.

⬇

Train the agent in simulated environment

Simulator of Mach-Zehnder interferometer

- Configuration
  - Gaussian beam with a plane wavefront with constant radius
  - PZT actuator for alignment is on mirror 1 and BS 2

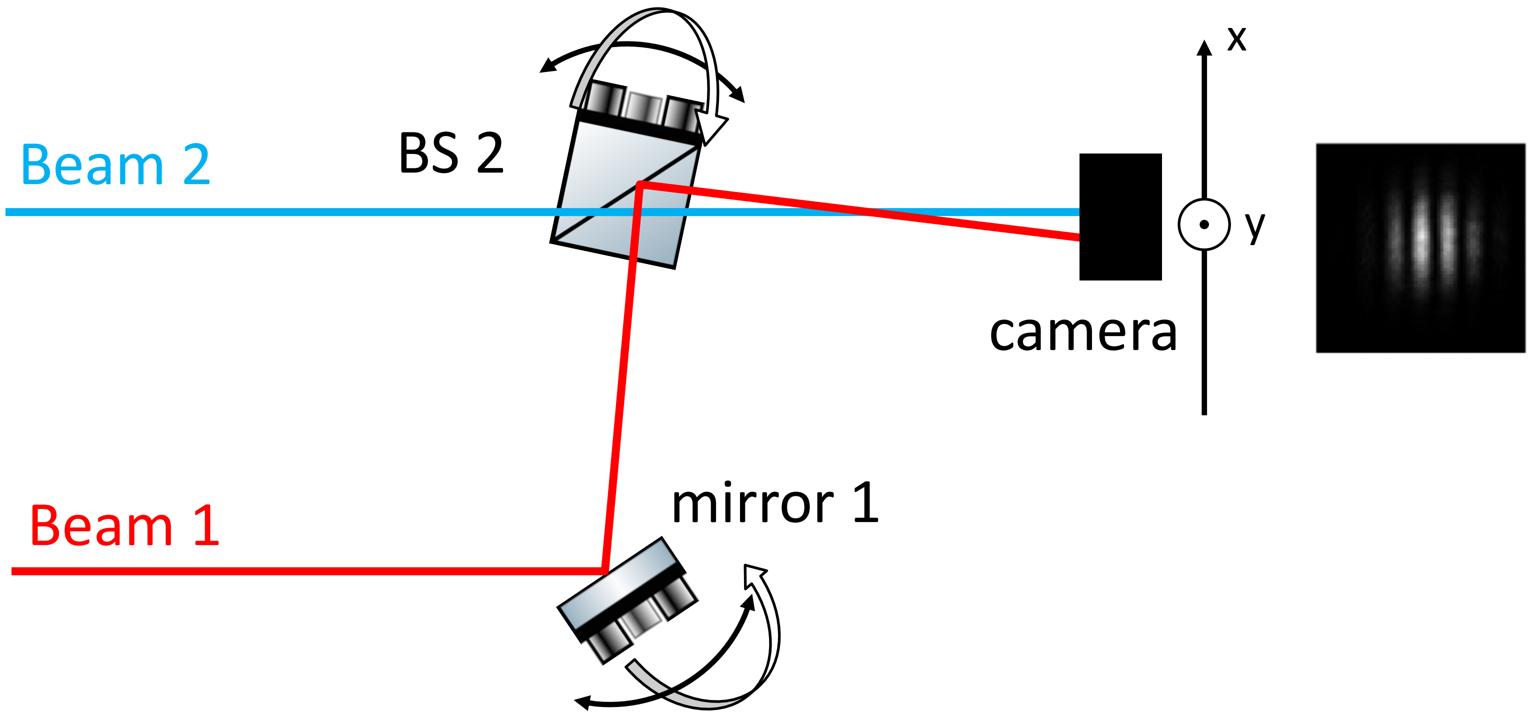


parameter	a	b	c	Beam radius	Wavelength
value	20 cm	30 cm	10 cm	950 μm	635 nm

Simulator of Mach-Zehnder interferometer

- Initial condition

At the beginning of each training, mirror 1 and BS 2 are tilted randomly within $\pm\alpha$ from aligned state

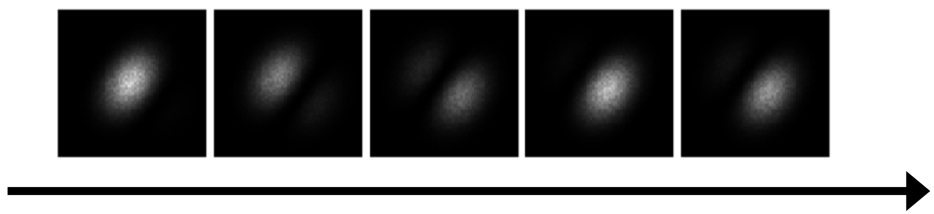


parameter	$\alpha_{M1,x}$	$\alpha_{M1,y}$	$\alpha_{BS2,x}$	$\alpha_{BS2,y}$
value [rad]	$5.2 \cdot 10^{-3}$	$3.7 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$

Simulator of Mach-Zehnder interferometer

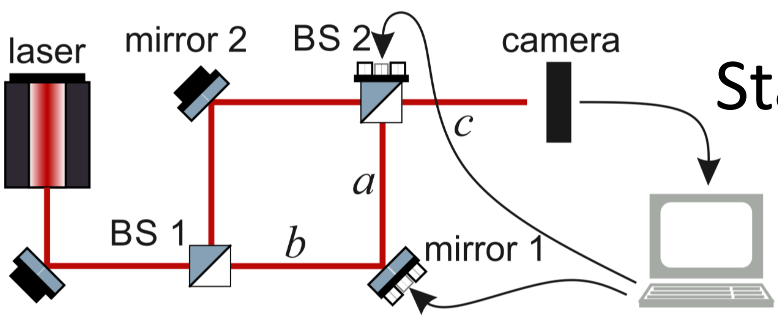
- State of Mach-Zehnder interferometer

Interferometric pattern changes with the optical path difference ΔL: $0 \rightarrow \lambda$

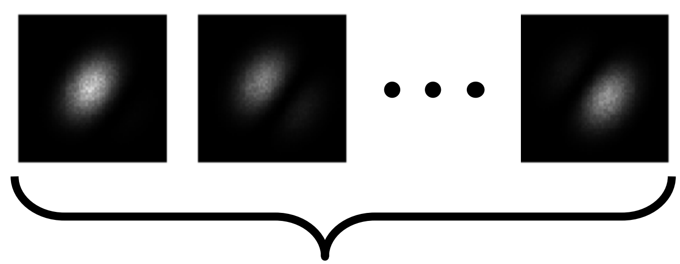


ΔL (changed by PZT on mirror2)

State = 16 images of 64×64 pixels acquired by camera



State:



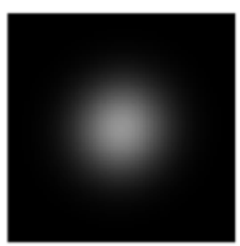
16 images

Simulator of Mach-Zehnder interferometer

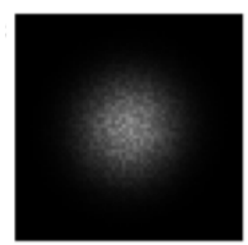
- Domain randomization

Adding noise to the simulated environment can reduce the gap between the simulated and real environments.

- Vary beam radius by ± 20 % randomly
- Rescale the brightness of observed images by ± 30 % randomly
- Add white noise to each pixel, etc.



Simulated image



Simulated image with noise

Simulator of Mach-Zehnder interferometer

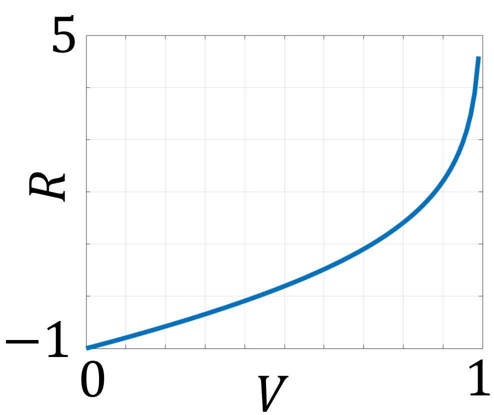
- Reward

Visibility: $V = \dfrac{\max_t(I_{\text{tot}}) - \min_t(I_{\text{tot}})}{\max_t(I_{\text{tot}}) + \min_t(I_{\text{tot}})}$, $\quad 0 \leq V \leq 1$

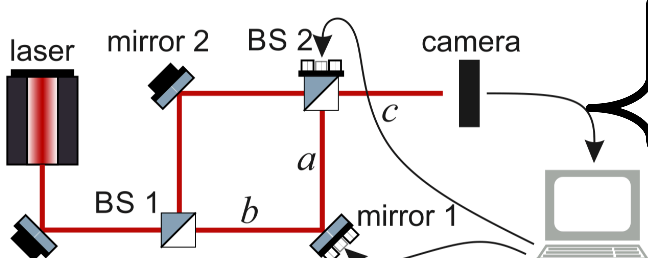
If V is used as reward

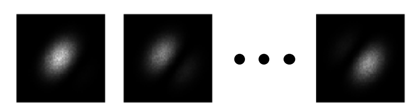
- Agent never gets penalty ($\because V \geq 0$)
- Agent is not rewarded for fine-tuning (eg. $V = 0.95$ vs. $V = 0.98$)

Visibility itself is not suitable for reward.



$\boxed{\text{Reward: } R = V - \log(1 - V) - 1}$



State: $S_t =$ 

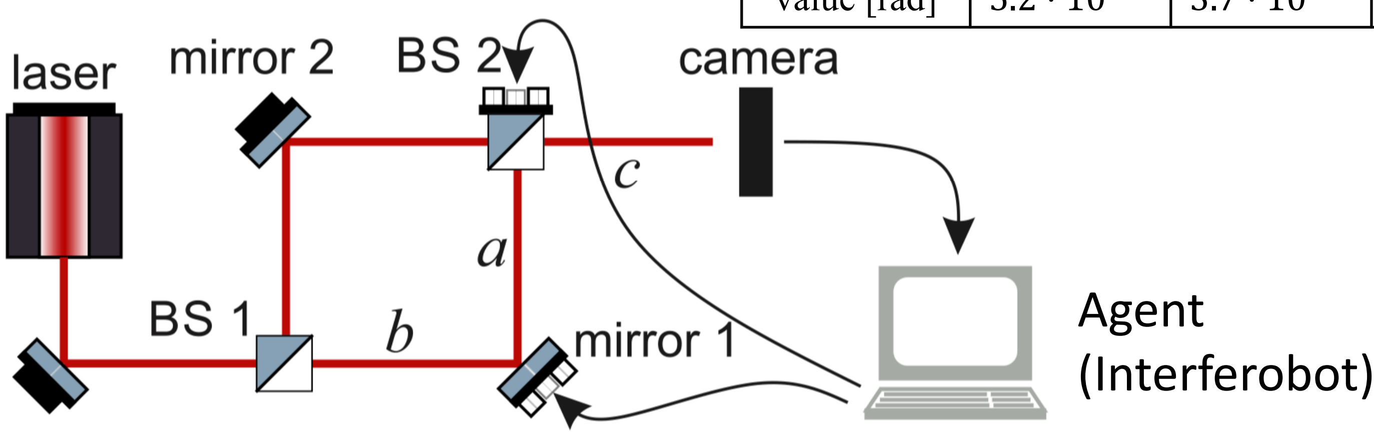
Reward: $R_t = V_t - \log(1 - V_t) - 1$

Simulator of Mach-Zehnder interferometer

- Actions of Interferobot

$$\begin{cases}\text{Do nothing} \\ \text{Tilt mirror 1 horizontally by } [\pm 0.01, \pm 0.05, \pm 0.1] \times \alpha_{M1,x} \\ \text{Tilt mirror 1 vertically by } [\pm 0.01, \pm 0.05, \pm 0.1] \times \alpha_{M1,y} \\ \text{Tilt BS 2 horizontally by } [\pm 0.01, \pm 0.05, \pm 0.1] \times \alpha_{BS2,x} \\ \text{Tilt BS 2 vertically by } [\pm 0.01, \pm 0.05, \pm 0.1] \times \alpha_{BS2,y}\end{cases}$$

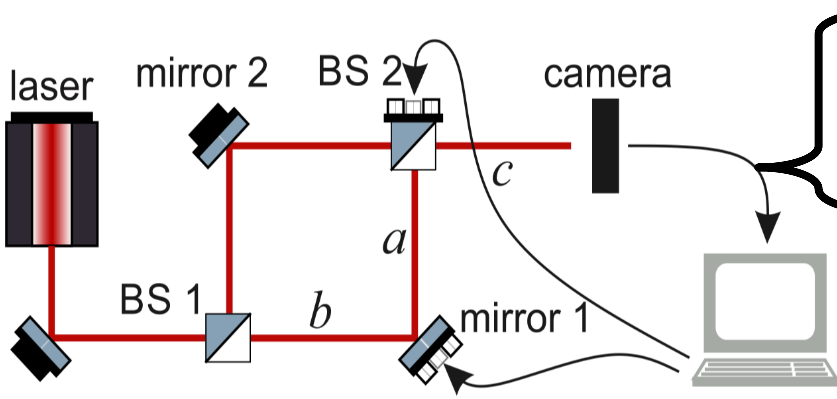
parameter	$\alpha_{M1,x}$	$\alpha_{M1,y}$	$\alpha_{BS2,x}$	$\alpha_{BS2,y}$
value [rad]	$5.2 \cdot 10^{-3}$	$3.7 \cdot 10^{-3}$	$2.6 \cdot 10^{-3}$	$1.8 \cdot 10^{-3}$

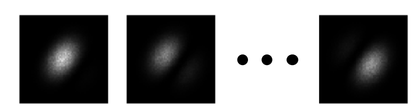


Simulator of Mach-Zehnder interferometer

- Conditions for training

1. Repeat interaction for 100 steps (1 episode)
   Interferobot is trained with double dueling DQN
2. Reset the simulated environment and begin new episode
3. Repeat 1 and 2 for 5×10^4 times (10 hours)

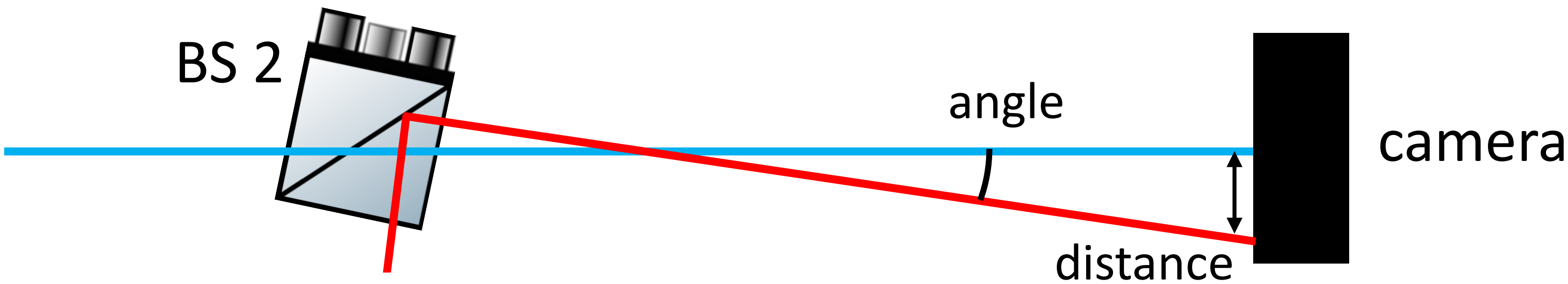


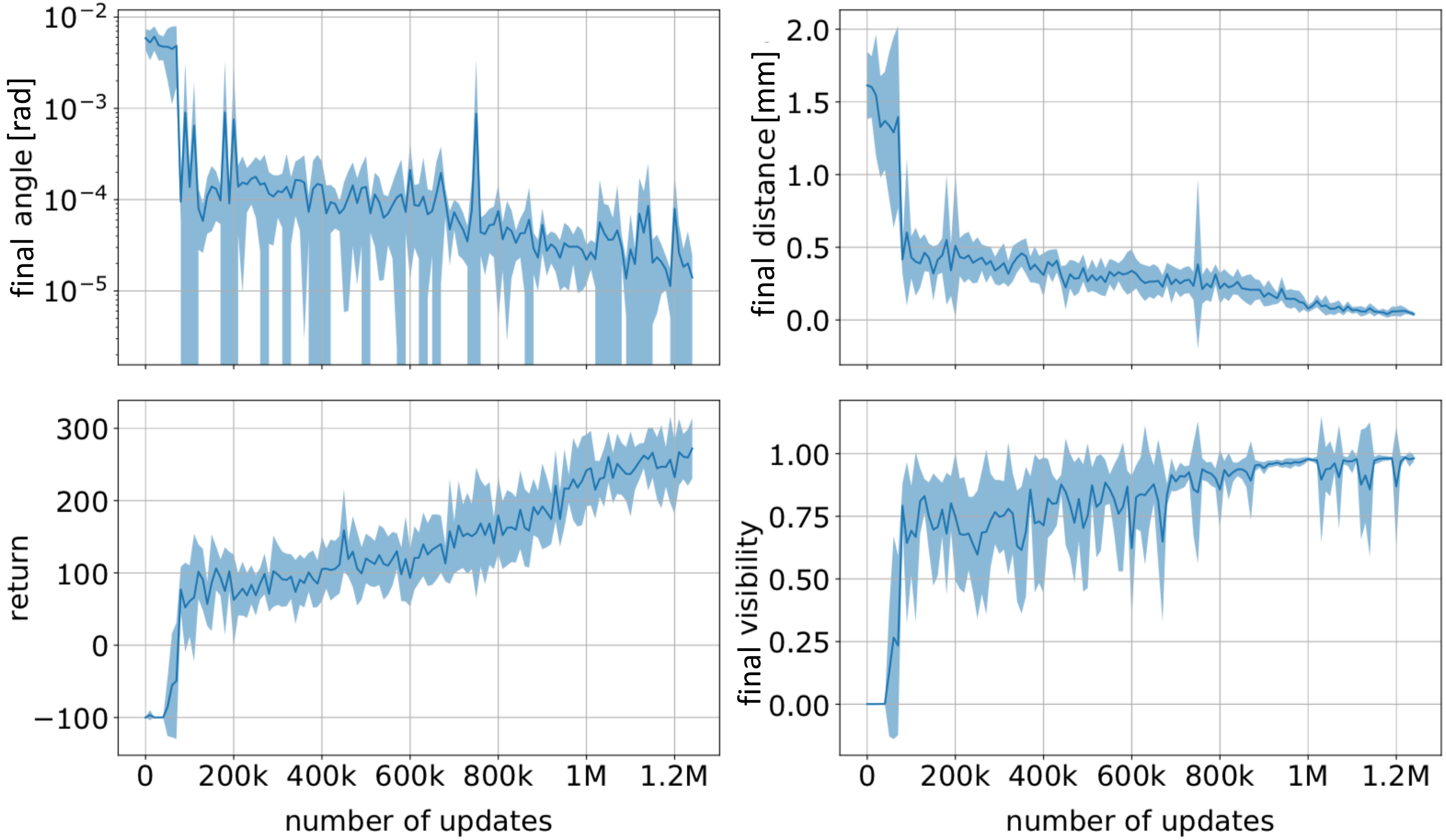
State: $S_t =$ 

Reward: $R_t = V_t - \log(1 - V_t) - 1$

Action: $A_t \in \left\{ \begin{array}{l} \text{Do nothing} \\ \text{Tilt mirror 1 horizontally} \\ \text{Tilt mirror 1 vertically} \\ \text{Tilt BS 2 horizontally} \\ \text{Tilt BS 2 vertically} \end{array} \right\}$

Result in the simulated environment





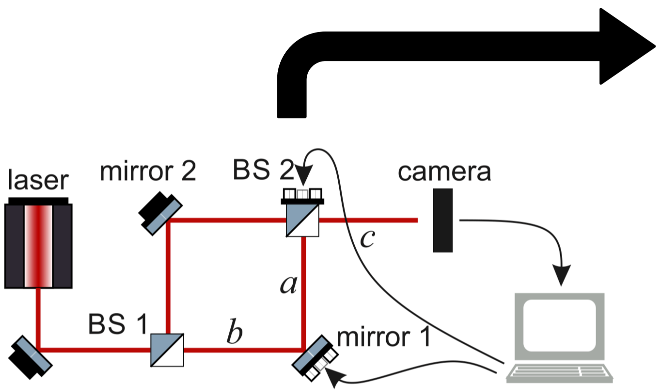
Experiment with the physical environment

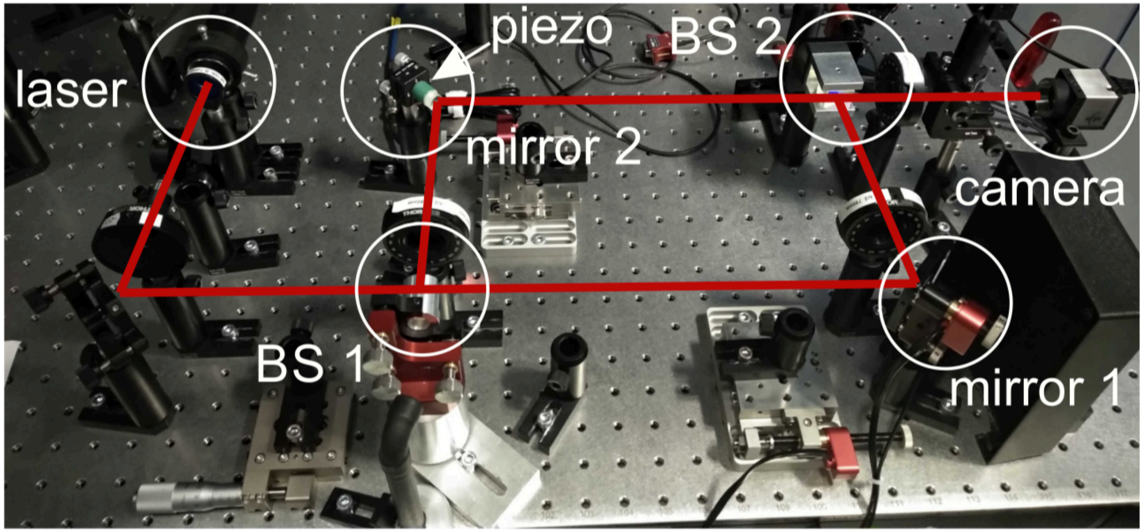
Experiment with the physical environment

Prepare the same Mach-Zehnder interferometer as the simulated one.

Let the trained Interferobot align physical Mach-Zehnder interferometer

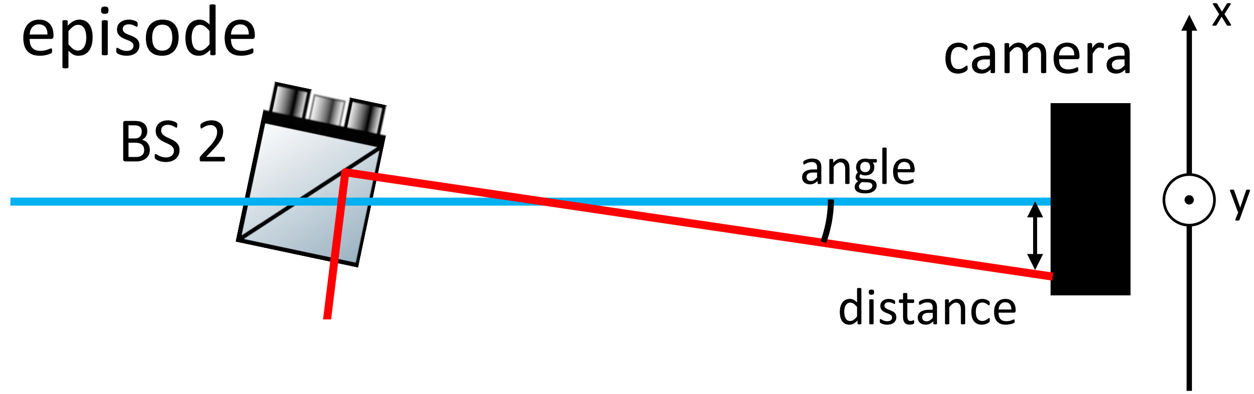
Experiments were conducted for 100 times (100 episodes)

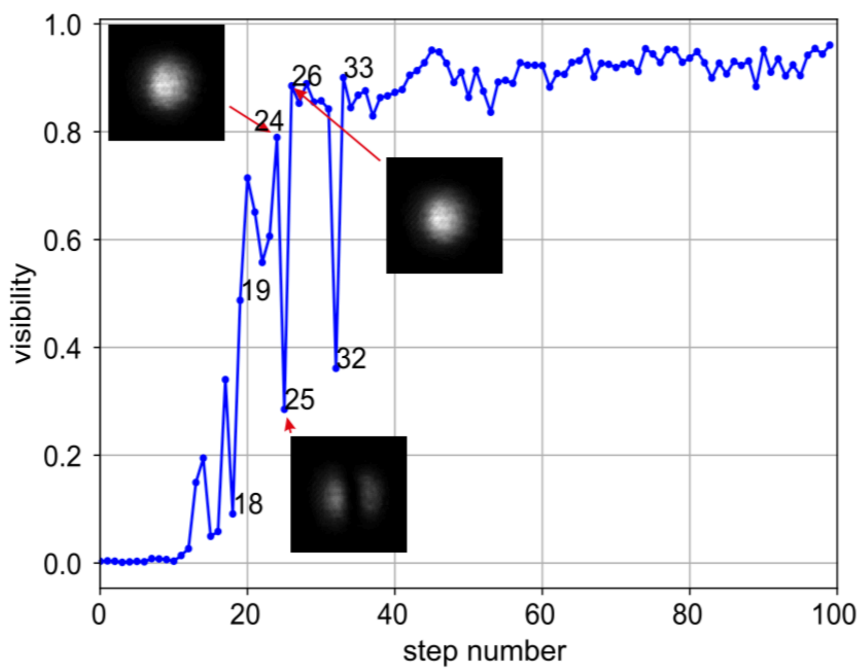


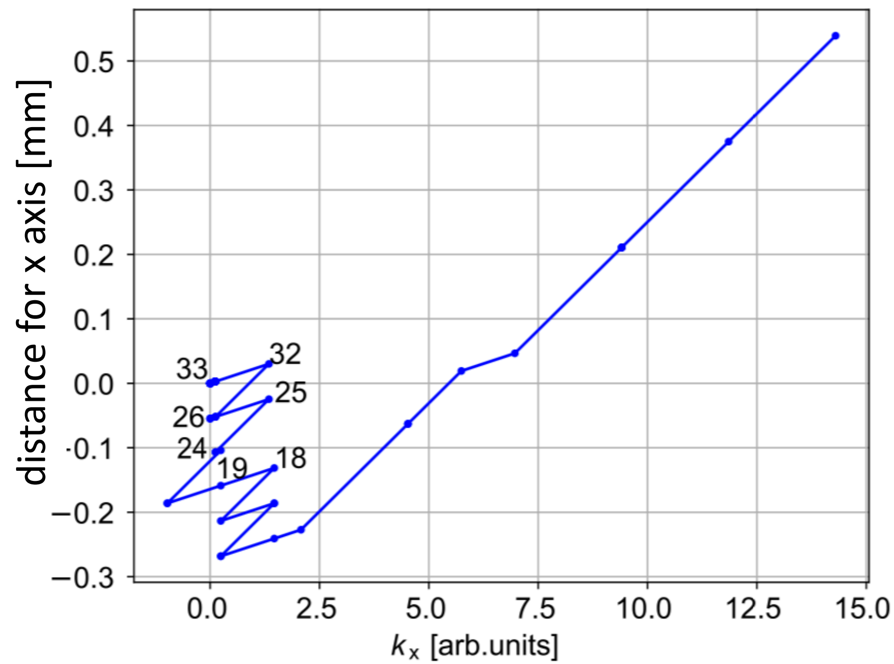


Experiment with the physical environment

- Result in 1 episode



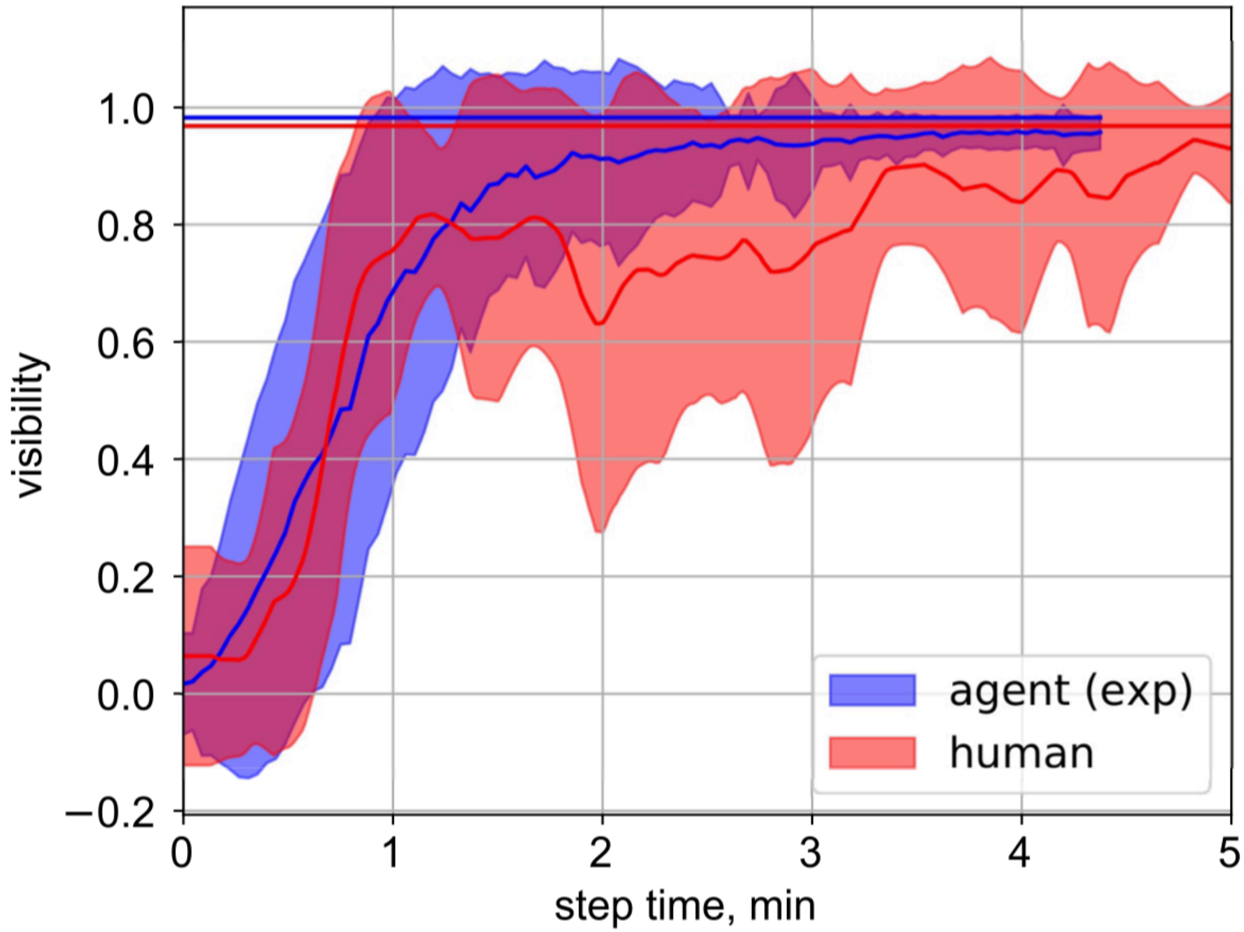




~ angle

Experiment with the physical environment

- Comparison with human's performance



Summary

- Training in simulated environment can reduce the training time.

- Interferobot performed well in both simulated and physical environments.

- Interferobot outperformed human in aligning Mach-Zehnder interferometer.

Thank you for listening