Reinforcement Learning for Science

Mar 20

Tailin Wu, Westlake University

Website: ai4s.lab.westlake.edu.cn/course

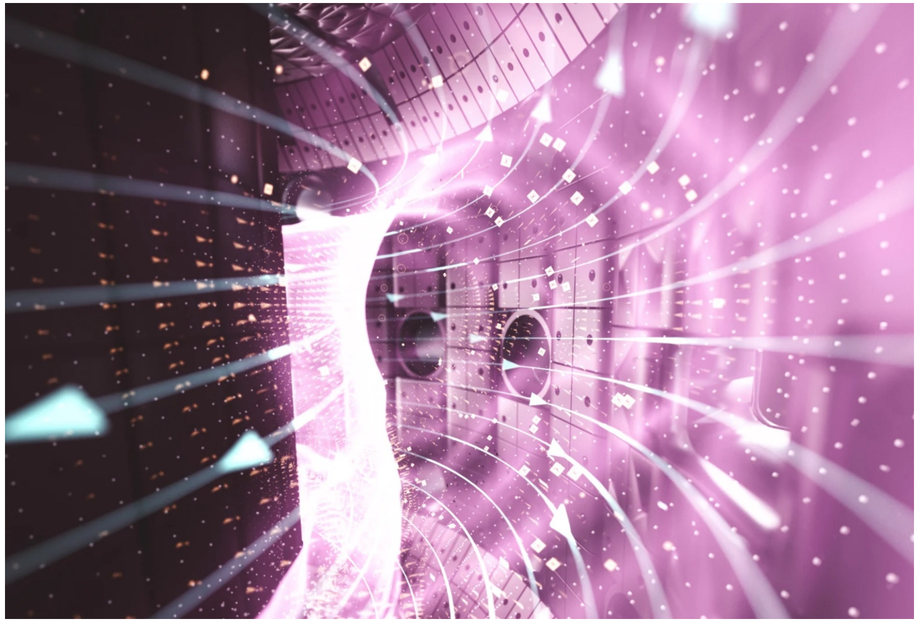
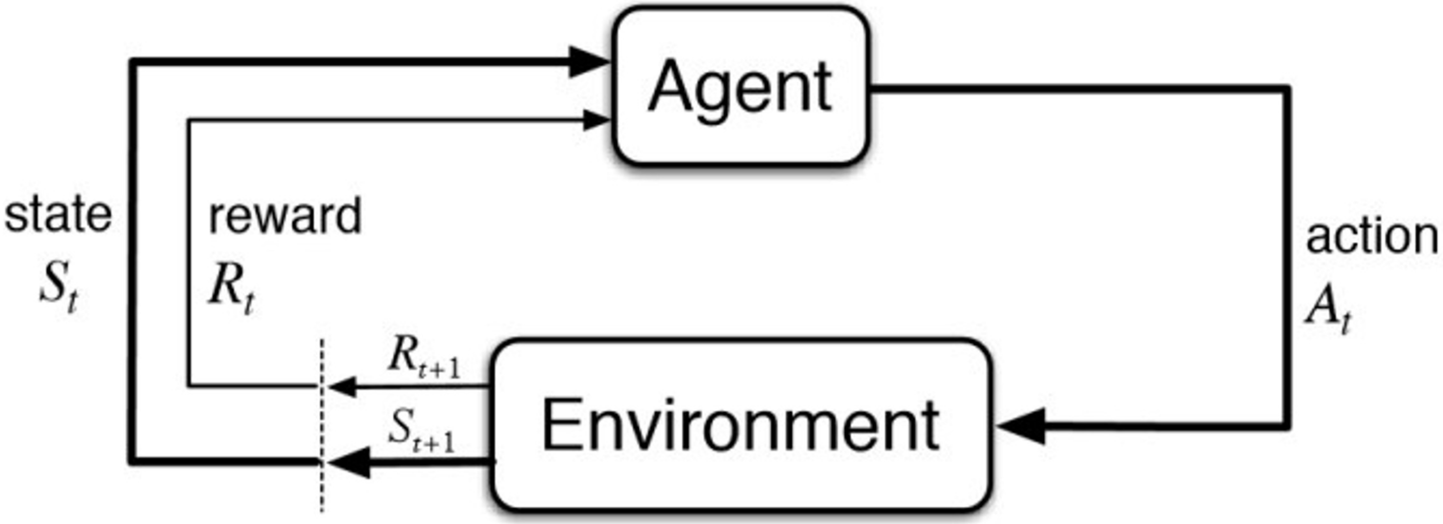


Image from: DeepMind

Markov Decision Process (MDP): Setup



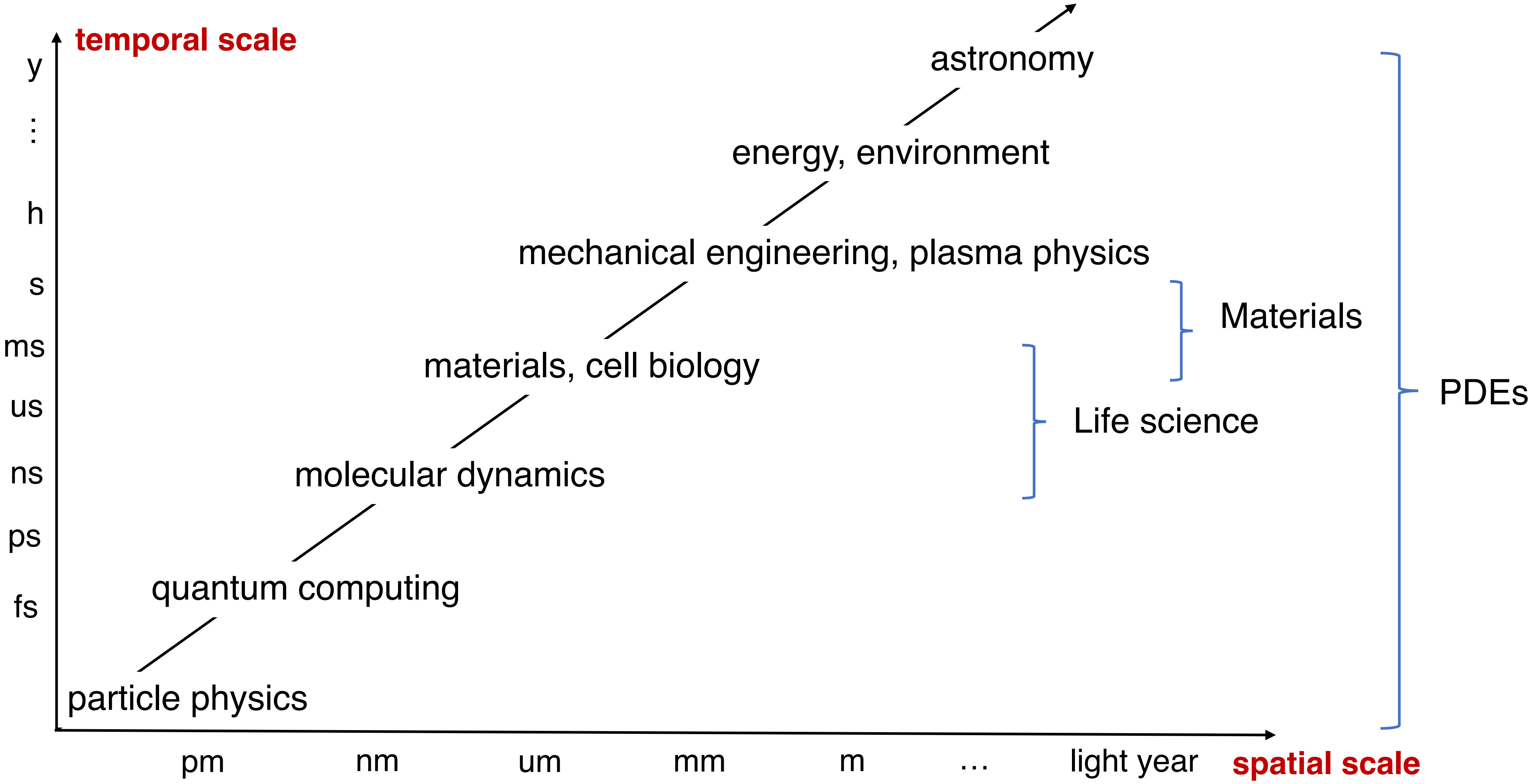
Goal: Maximize the long-term expected reward w.r.t. to the policy $\pi(A_t|S_t)$

$$\max_{\pi(A_t|S_t)} \mathbb{E}_t[R_t]$$

Recent Deep RL papers in *Nature/Science*

Paper	Publisher	Application
Avoiding fusion plasma tearing instability with deep reinforcement learning	*Nature* 2024	Tokamak control
Champion-level drone racing using deep reinforcement learning	*Nature* 2023	Drone racing
Top-down design of protein architectures with reinforcement learning	*Science* 2023	Protein design
Dense reinforcement learning for safety validation of autonomous vehicles	*Nature* 2023	Autonomous driving
Magnetic control of tokamak plasmas through deep reinforcement learning	*Nature* 2022	Tokamak control
Discovering faster matrix multiplication algorithms with reinforcement learning	*Nature* 2022	Matrix multiplication
A graph placement methodology for fast chip design	*Nature* 2021	Chip design
A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play	*Science* 2018	Board game

Application in AI for Science: from microscopic to macroscopic



temporal scale

y
⋮
h
s
ms
us
ns
ps
fs

astronomy

energy, environment

mechanical engineering, plasma physics

materials, cell biology

molecular dynamics

quantum computing

particle physics

Materials

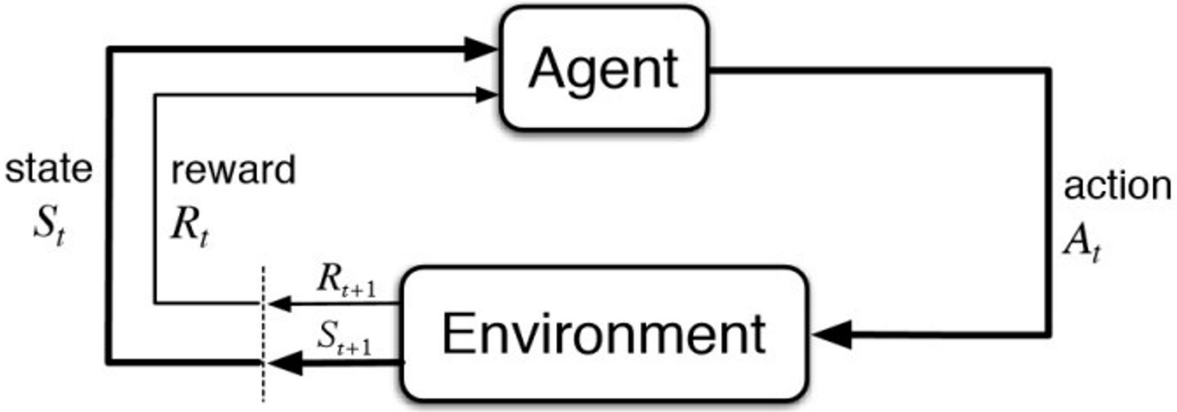
Life science

PDEs

pm
nm
um
mm
m
…
light year

spatial scale

How to Apply RL in AI for Science



1. Define the task
Specify:

- State S
- Action A
- Reward R

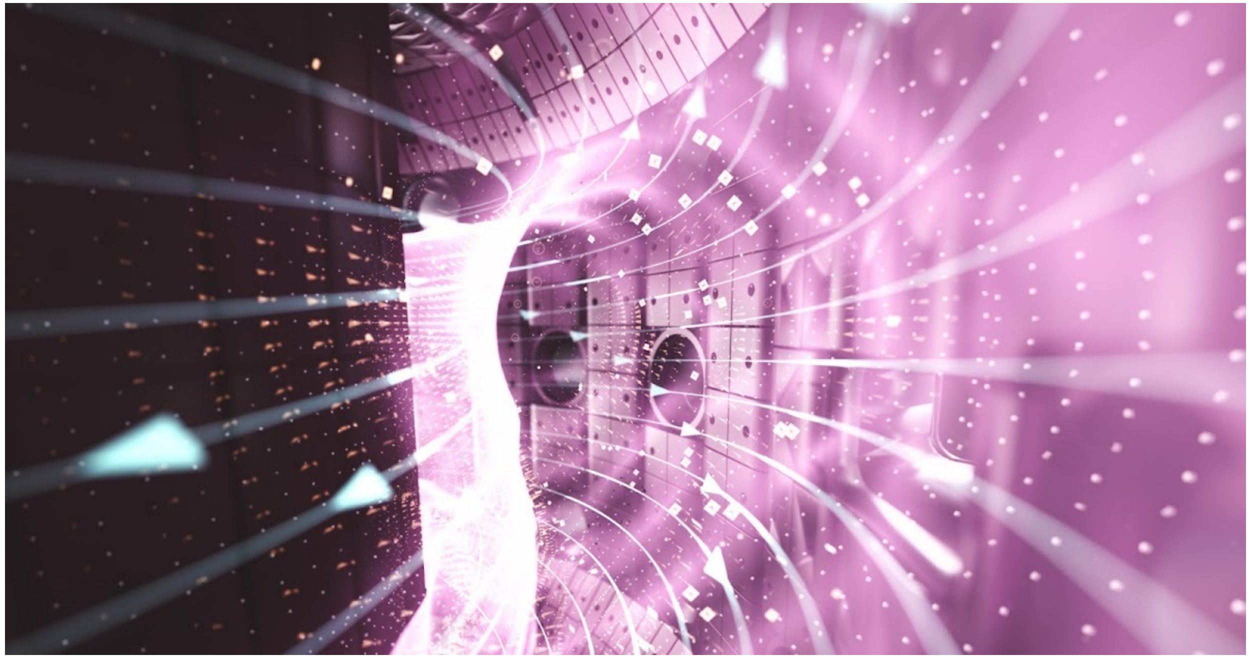
Learn:

- Policy $\pi_\theta(A|S)$

2. Choose an appropriate RL algorithm

RL for Science: Case study

- Deep RL for controlled nuclear fusion

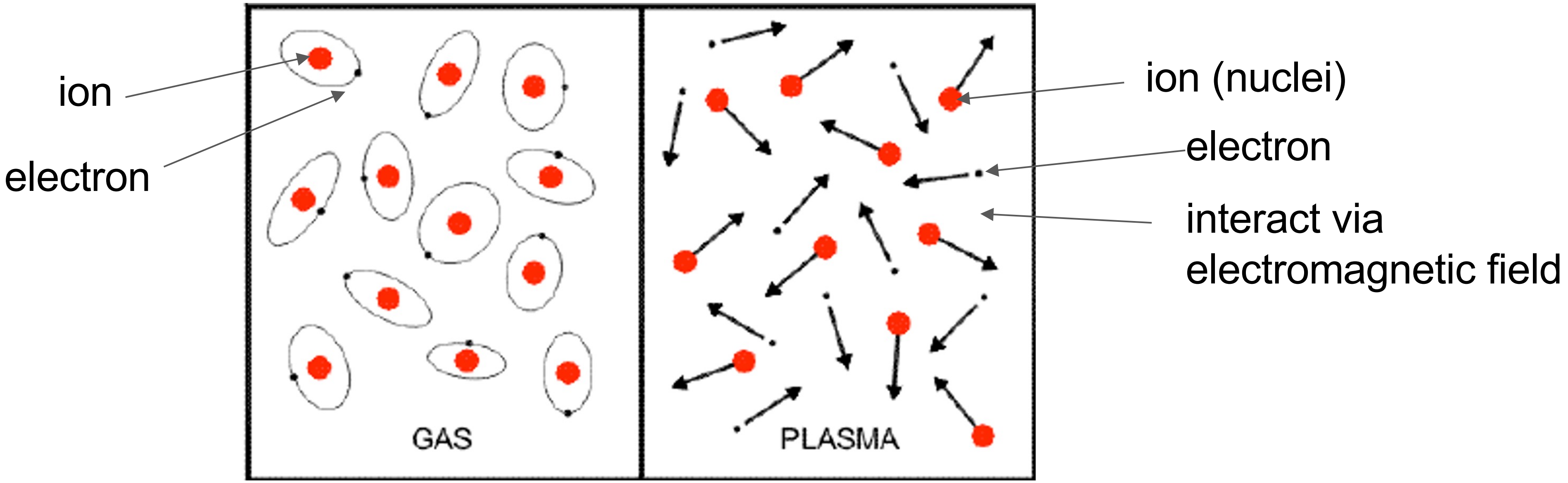


Preliminaries: plasma（等离子体）

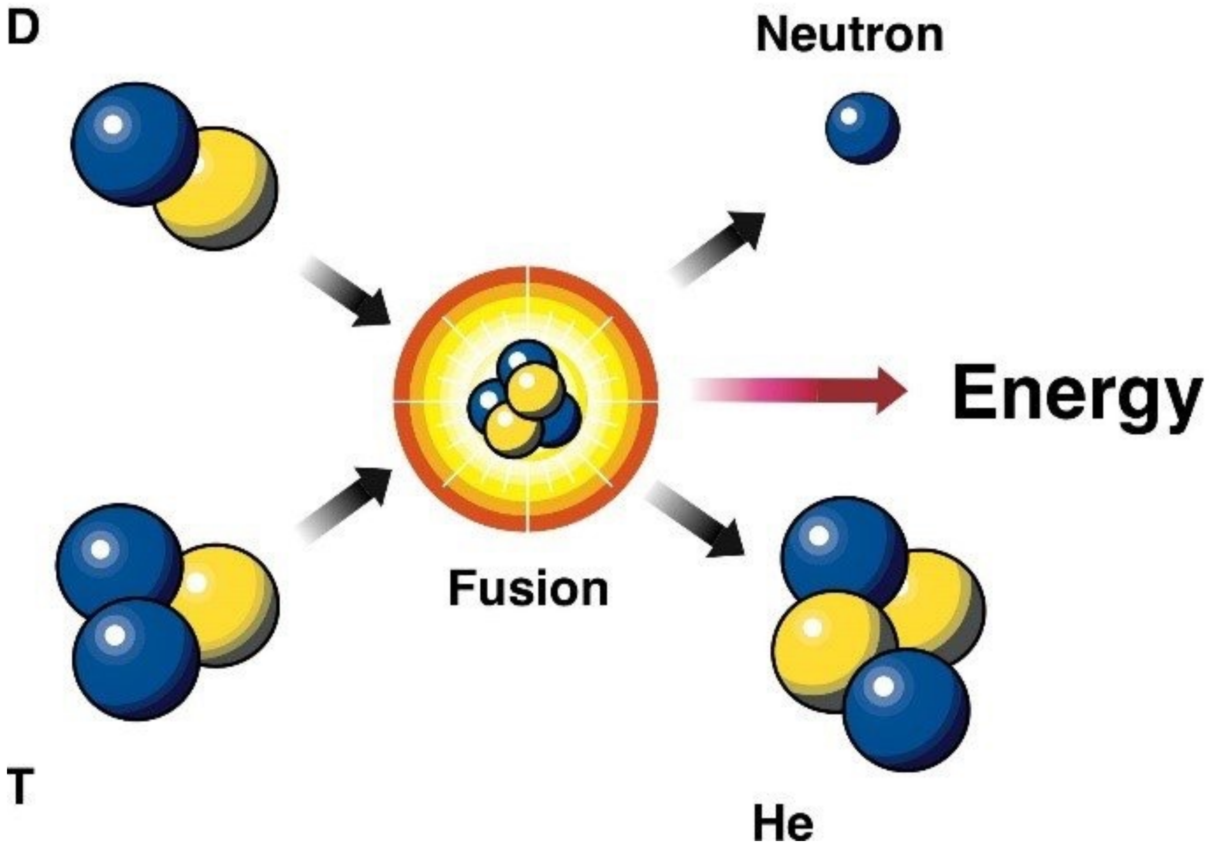
Plasma: Consisting of energetic ions and *free* electrons, interacted via electromagnetic (EM) field.

Examples: fire, lightning, sun, nuclear fusion

It is one of the four fundamental states of matter. It is the *dominant* form of ordinary matter in the universe.



Preliminaries: Nuclear fusion



Preliminaries: Why nuclear fusion?

1. Percentage of mass transferred to energy: $E = mc^2$

- Chemical: 0.0000001%
- Nuclear fission: 0.1%
- **Nuclear fusion: 0.4%**
- Black hole: 40%
- Matter + anti-matter: 100%

2. Inexhaustible supply of fusion fuels:

Deuterium can be distilled from all forms of water, while tritium will be produced during the fusion reaction as fusion neutrons interact with lithium. The reserve on Earth is able to fulfil the needs for **millions of years**.

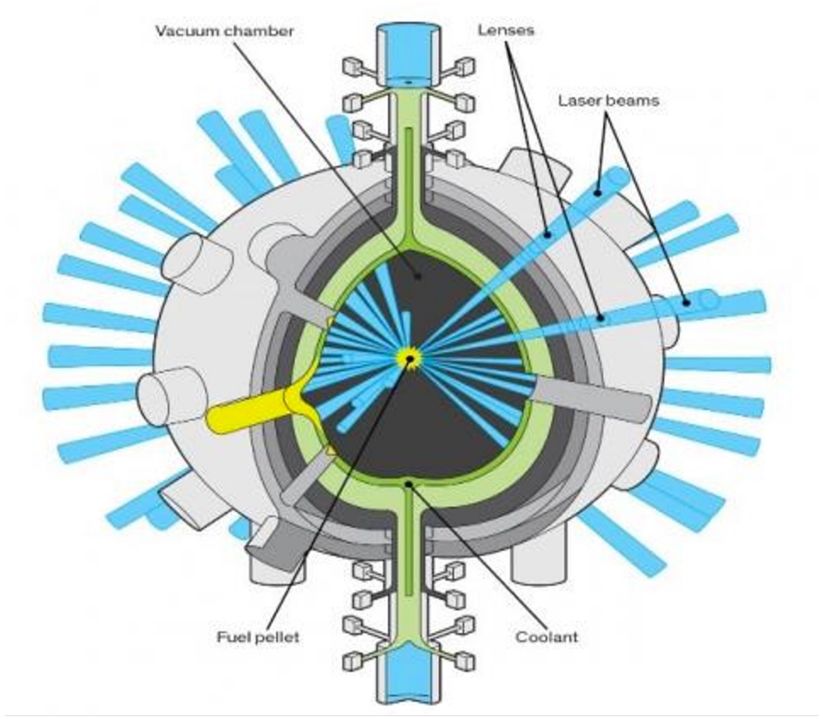
3. Environment friendly:

- No CO_2
- No long-lived radioactive waste
- No risk of meltdown

Preliminaries: Two major ways of controlled nuclear fusion

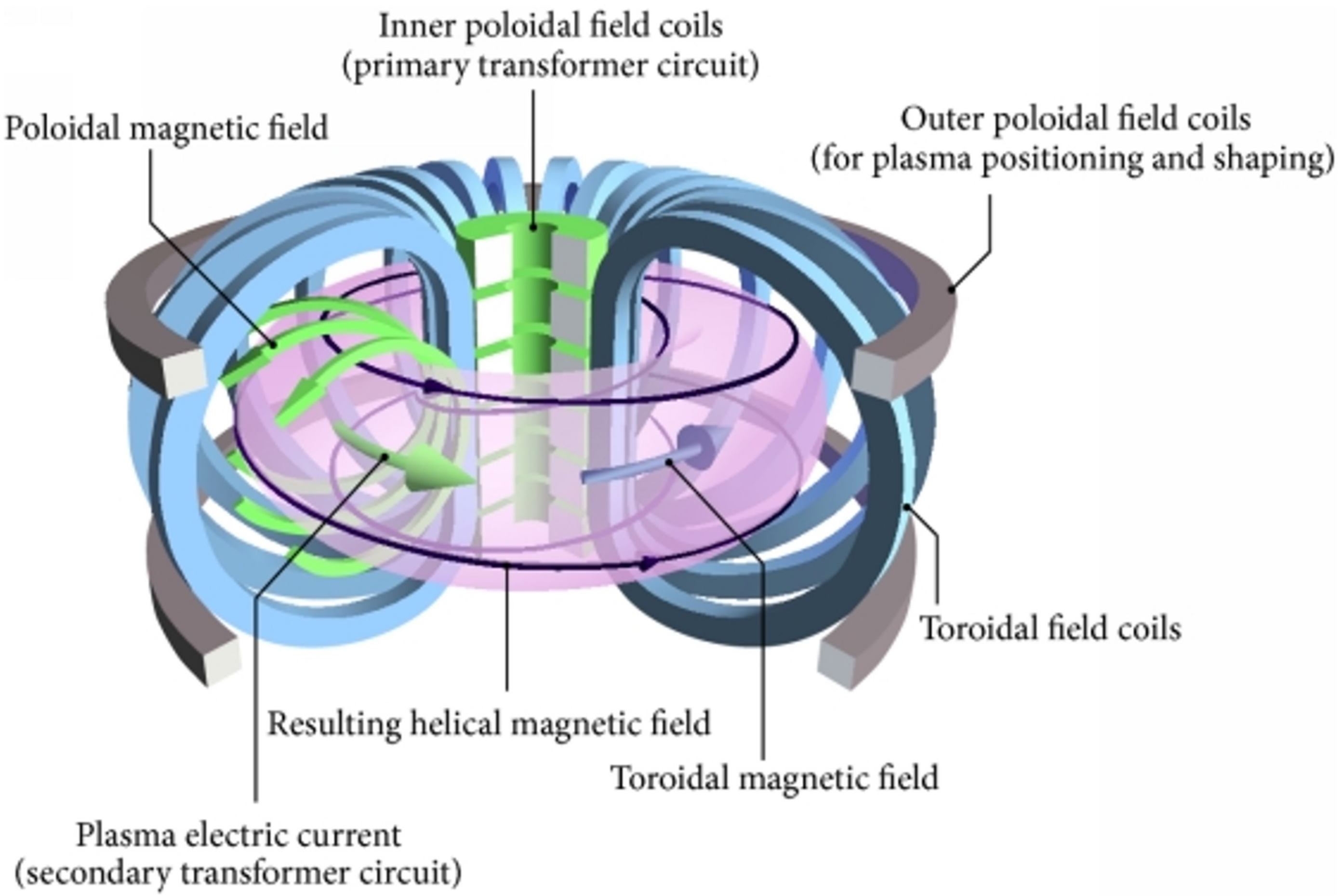
The temperature required for confining the fusion plasma are so hot (>10 million °C), and cannot be confined via any material. Two main ways of confinement:

1. Inertial confinement（惯性约束）：

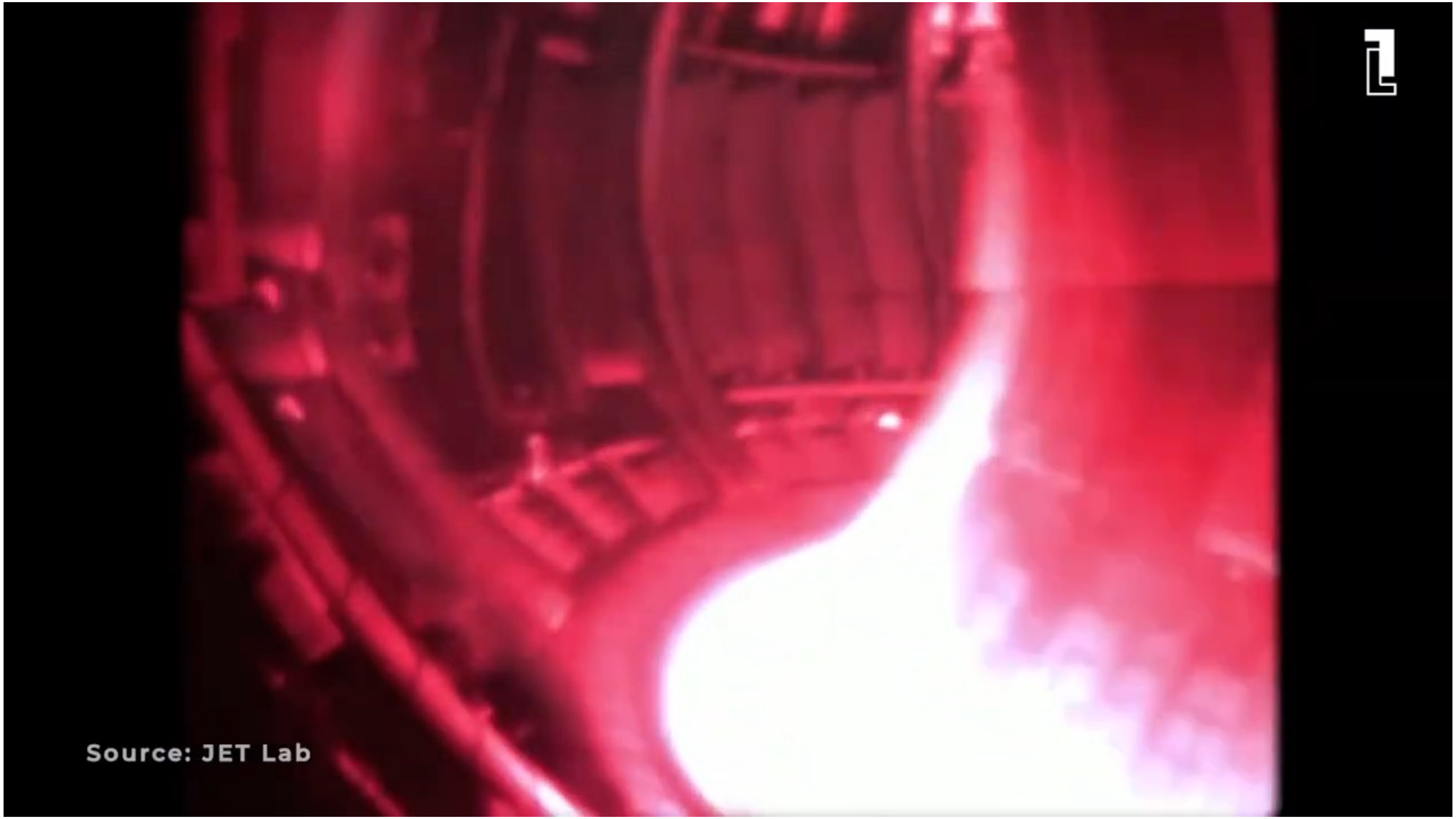


Preliminaries: Two major ways of controlled nuclear fusion

2. Magnetic confinement（磁约束）, using Tokamak (current work)


Inner poloidal field coils
(primary transformer circuit)
Poloidal magnetic field
Outer poloidal field coils
(for plasma positioning and shaping)
Toroidal field coils
Resulting helical magnetic field
Toroidal magnetic field
Plasma electric current
(secondary transformer circuit)

Controlled nuclear fusion using Tokamak



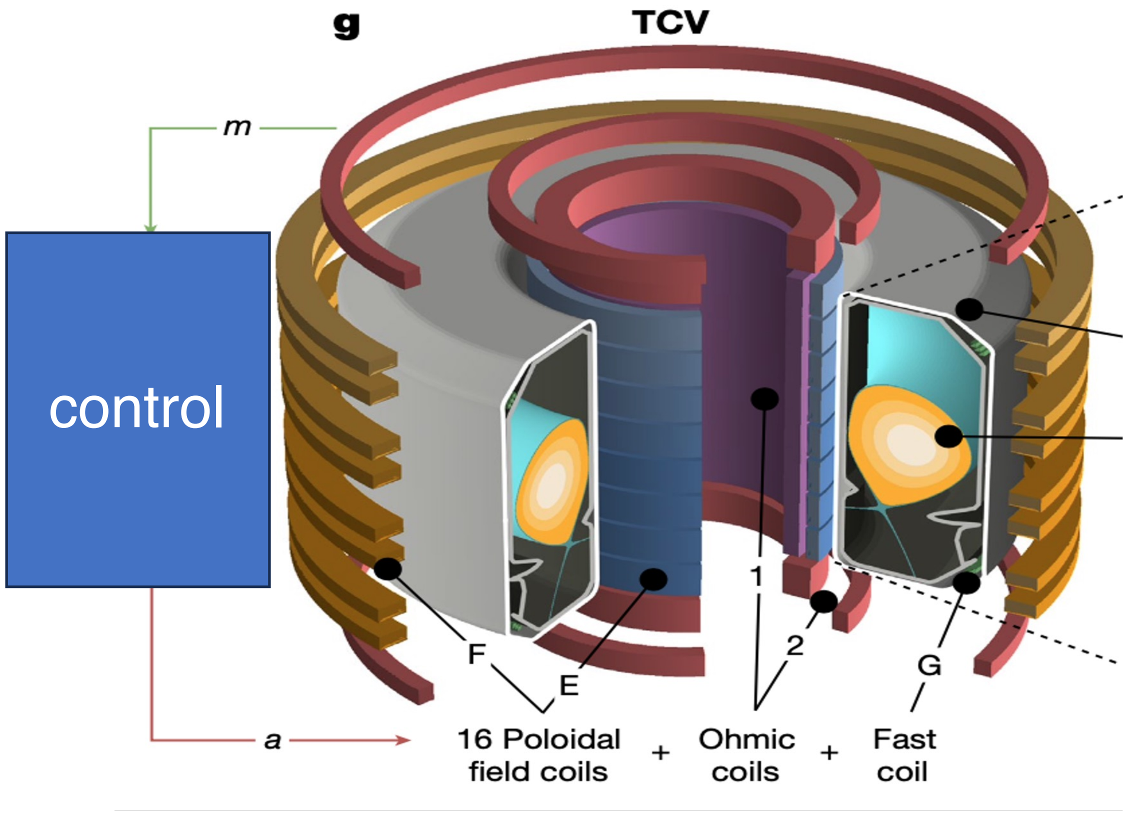
Source: JET Lab

1. What is the problem?

Task: To shape and maintain a high-temperature plasma within the tokamak vessel.

Each time step t have observation and needs to output a control signal:

观测 m: input observation, R^{92}
控制 a: output control, R^{19}



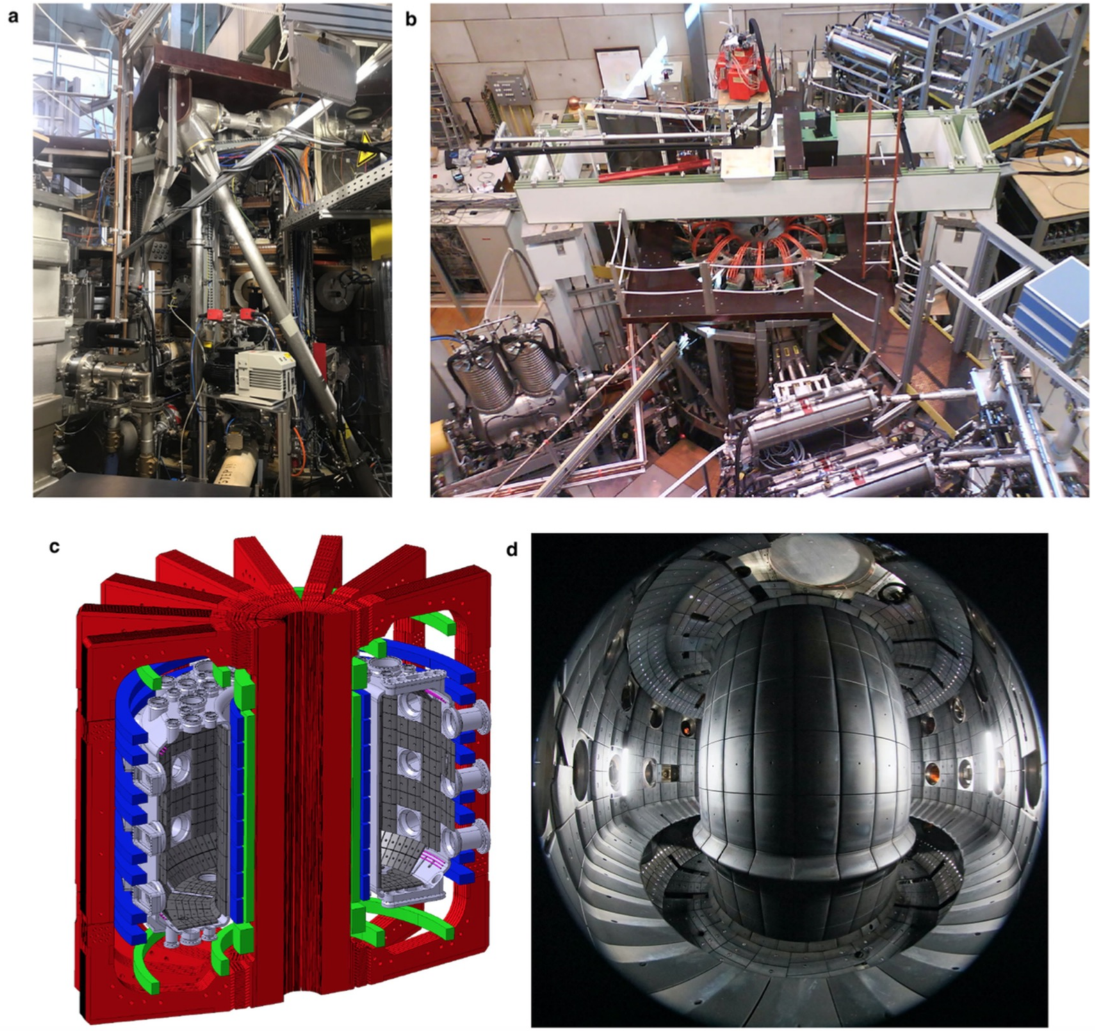
1. What is the problem? Device overview

Tokamak à configuration variable (TCV) (present work)

- Plasma height: 1.40m
- Major radius: 0.88m
- Plasma life span: 2s maximum
- Toroidal magnetic field: 1.43T
- Additional heating power: 4.5MW

ITER (cost $22 billion):

- Major radius: 6.2 m
- Magnetic field: 11.8 T
- Heating power: 320 MW
- Fusion power: 500 MW
- Discharge duration: up to 1000 s



Extended Data Fig. 1 | Pictures and illustration of the TCV. a, b Photographs showing the part of the TCV inside the bioshield. **c** CAD drawing of the vessel and coils of the TCV. **d** View inside the TCV (Alain Herzog/EPFL), showing the limiter tiling, baffles and central column.

2. Why is it important?

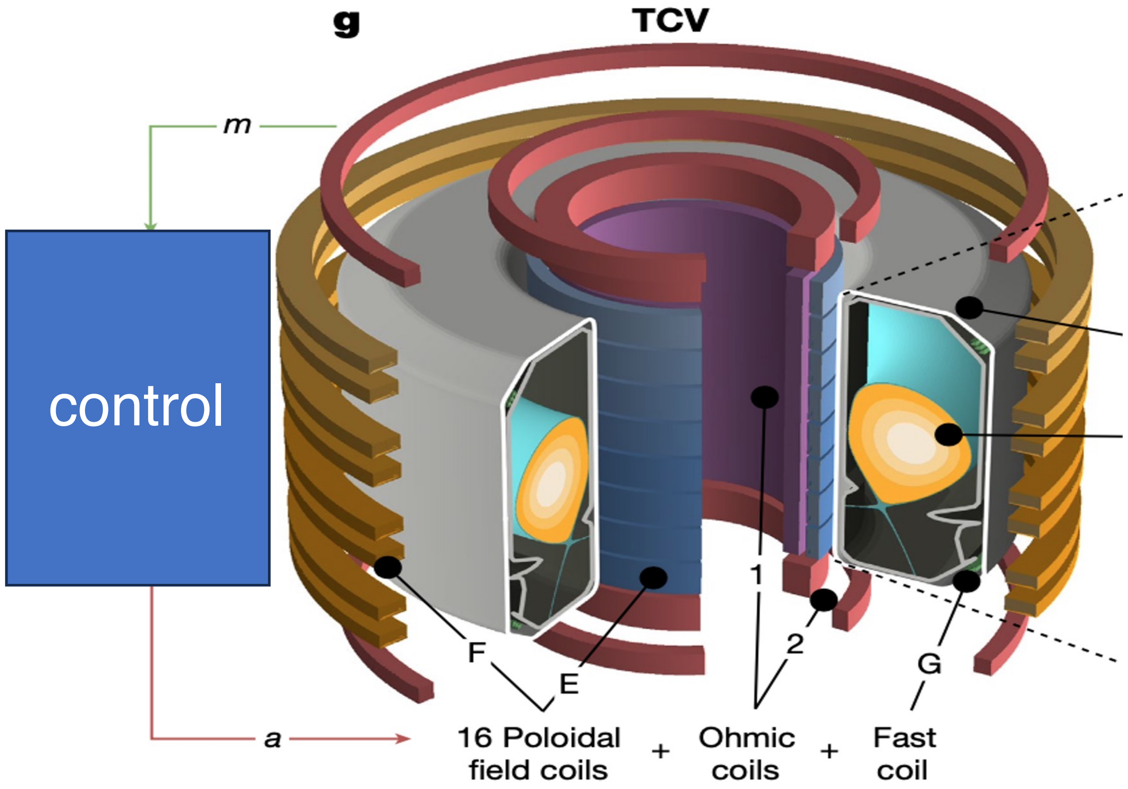
The effective control of plasma within a tokamak will **pave the way** for commercial nuclear fusion, which allows to produce energy energy that is

(1) Virtually unlimited;

(2) Environmentally friendly.

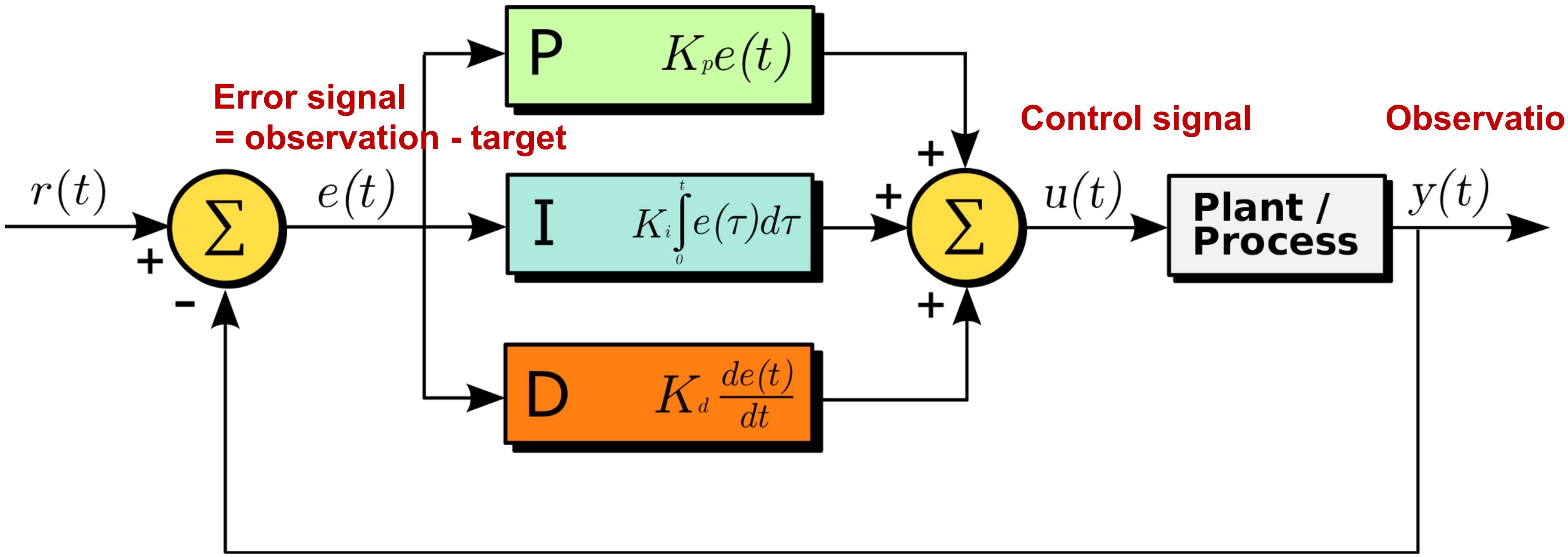
3. Why is it hard?

This requires high-dimensional, high-frequency, closed-loop control using magnetic actuator coils, further complicated by the diverse requirements across a wide range of plasma configurations.



4. Limitation of prior methods: PID control

Proportional–integral–derivative (PID) control:



4. Limitation of prior methods: PID control

Pros: Effective

Cons:

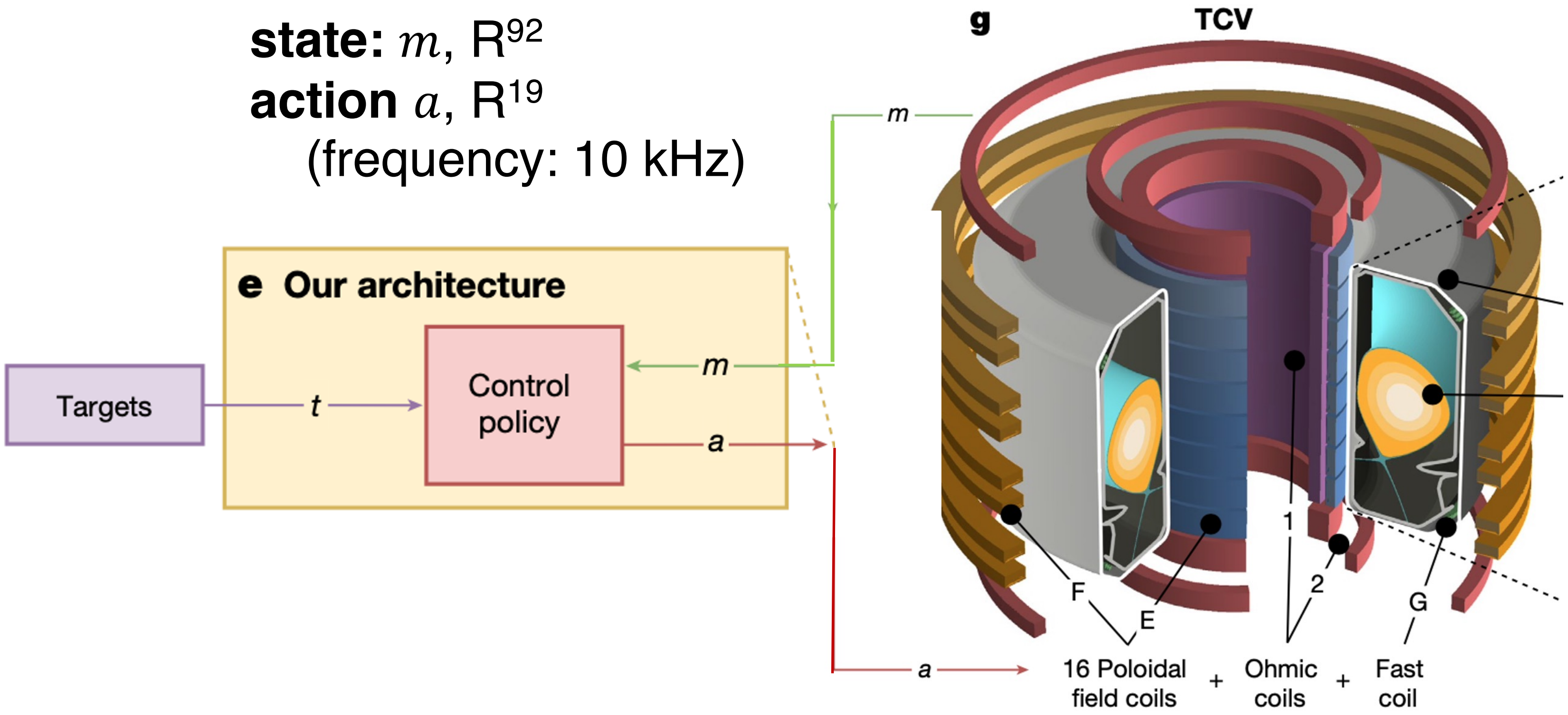
(1) The controllers are designed on the basis of linearized model dynamics

(2) Requires substantial engineering effort, design effort and expertise whenever the target plasma configuration is changed, together with complex, real-time calculations for equilibrium estimation

5. Main components of the proposed method [1]

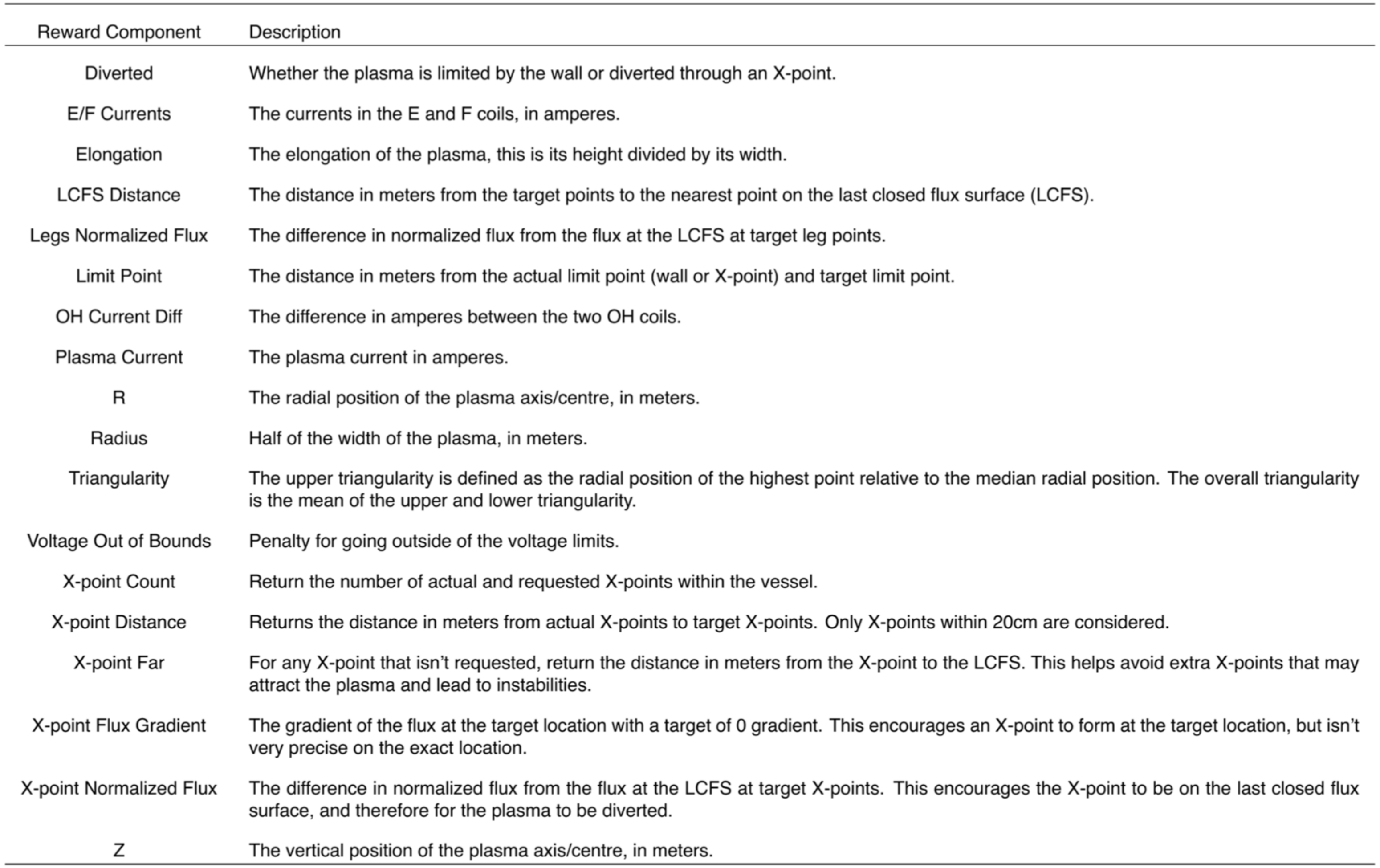
[1] Degrave, Jonas, et al. "Magnetic control of tokamak plasmas through deep reinforcement learning." *Nature* 602.7897 (2022): 414-419.

state: m, R^{92}
action a, R^{19}
(frequency: 10 kHz)



5. Main components of the proposed method

Reward: The target values g of the objectives are often time-varying (e.g., the plasma current and boundary target points), and are sent to the policy as part of the observations: $\pi(a|s, g)$.



Reward Component	Description
Diverted	Whether the plasma is limited by the wall or diverted through an X-point.
E/F Currents	The currents in the E and F coils, in amperes.
Elongation	The elongation of the plasma, this is its height divided by its width.
LCFS Distance	The distance in meters from the target points to the nearest point on the last closed flux surface (LCFS).
Legs Normalized Flux	The difference in normalized flux from the flux at the LCFS at target leg points.
Limit Point	The distance in meters from the actual limit point (wall or X-point) and target limit point.
OH Current Diff	The difference in amperes between the two OH coils.
Plasma Current	The plasma current in amperes.
R	The radial position of the plasma axis/centre, in meters.
Radius	Half of the width of the plasma, in meters.
Triangularity	The upper triangularity is defined as the radial position of the highest point relative to the median radial position. The overall triangularity is the mean of the upper and lower triangularity.
Voltage Out of Bounds	Penalty for going outside of the voltage limits.
X-point Count	Return the number of actual and requested X-points within the vessel.
X-point Distance	Returns the distance in meters from actual X-points to target X-points. Only X-points within 20cm are considered.
X-point Far	For any X-point that isn't requested, return the distance in meters from the X-point to the LCFS. This helps avoid extra X-points that may attract the plasma and lead to instabilities.
X-point Flux Gradient	The gradient of the flux at the target location with a target of 0 gradient. This encourages an X-point to form at the target location, but isn't very precise on the exact location.
X-point Normalized Flux	The difference in normalized flux from the flux at the LCFS at target X-points. This encourages the X-point to be on the last closed flux surface, and therefore for the plasma to be diverted.
Z	The vertical position of the plasma axis/centre, in meters.

5. Main component of the proposed method

Training:

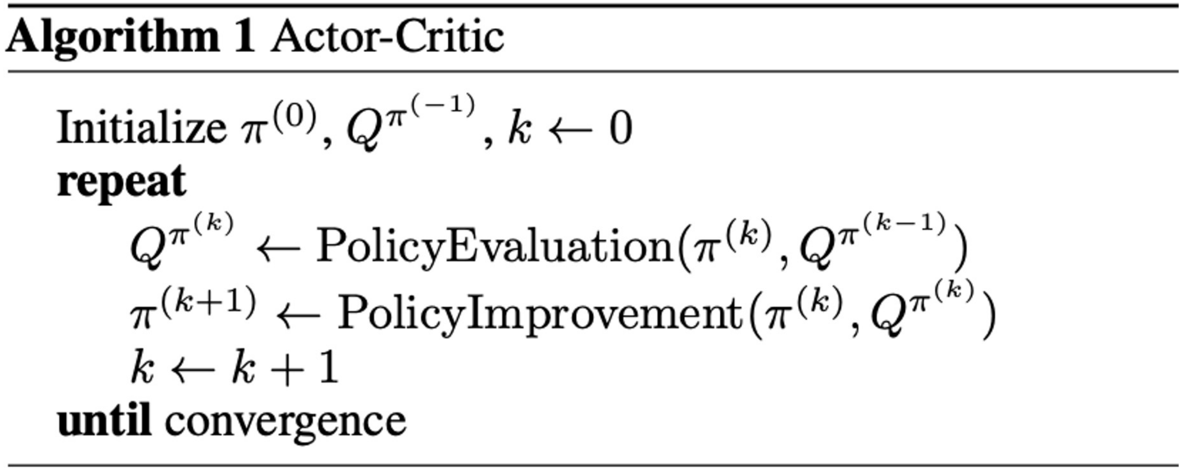
Perform training within a simulated environment using a solver.

Inference:

Directly deploy it in the device.

RL method:

Maximum a posteriori policy optimization (MPO) [1].



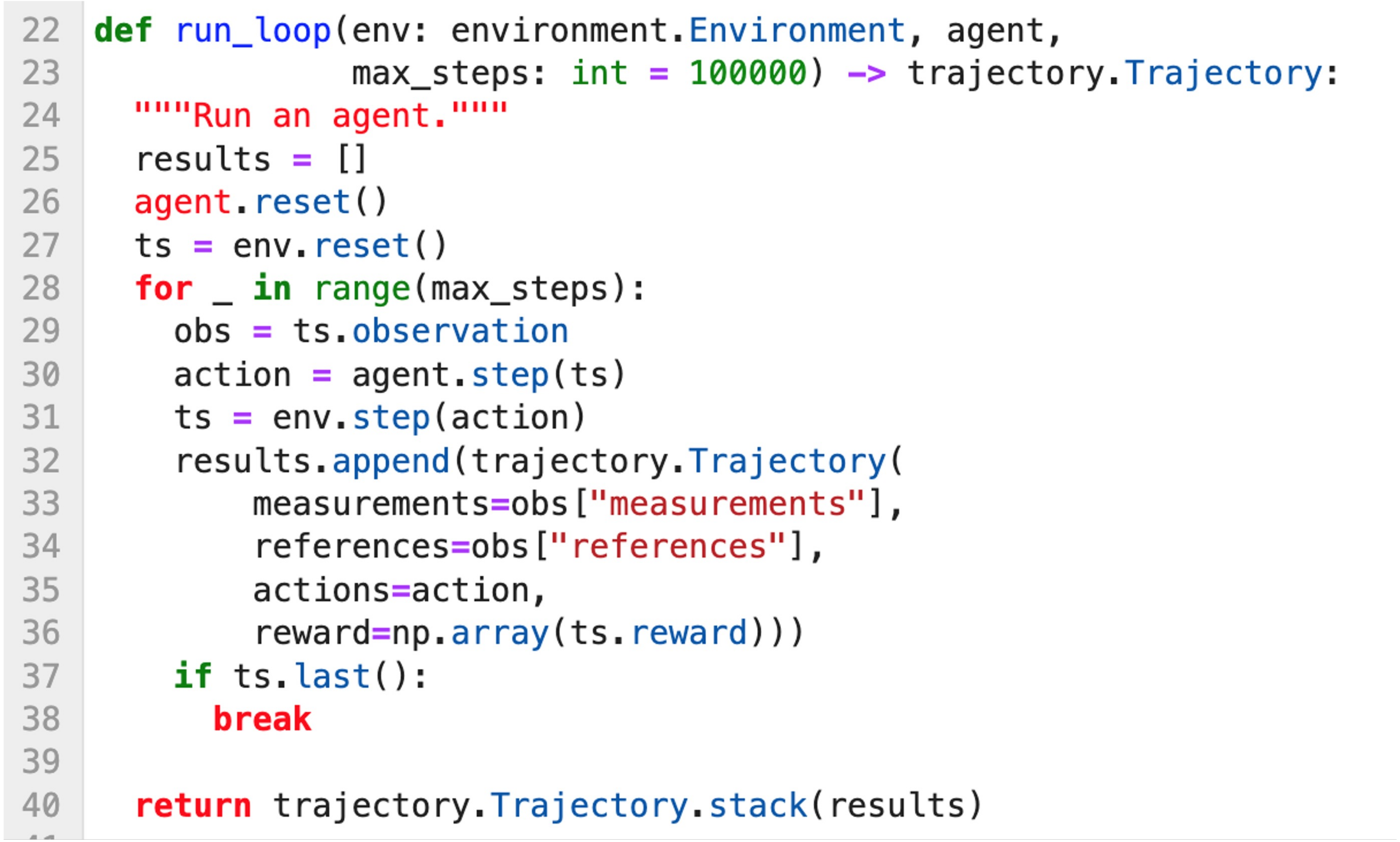
Algorithm 1 Actor-Critic

```
Initialize $\pi^{(0)}$, $Q^{\pi^{(-1)}}$, $k \leftarrow 0$
repeat
    $Q^{\pi^{(k)}} \leftarrow \text{PolicyEvaluation}(\pi^{(k)}, Q^{\pi^{(k-1)}})$
    $\pi^{(k+1)} \leftarrow \text{PolicyImprovement}(\pi^{(k)}, Q^{\pi^{(k)}})$
    $k \leftarrow k + 1$
until convergence
```

Actor π: small MLP, must be fast.
Critic Q^{π}: LSTM, can be large, only used in training

[1] Abdolmaleki, Abbas, et al. "Maximum a posteriori policy optimisation." ICLR 2018

Inference code

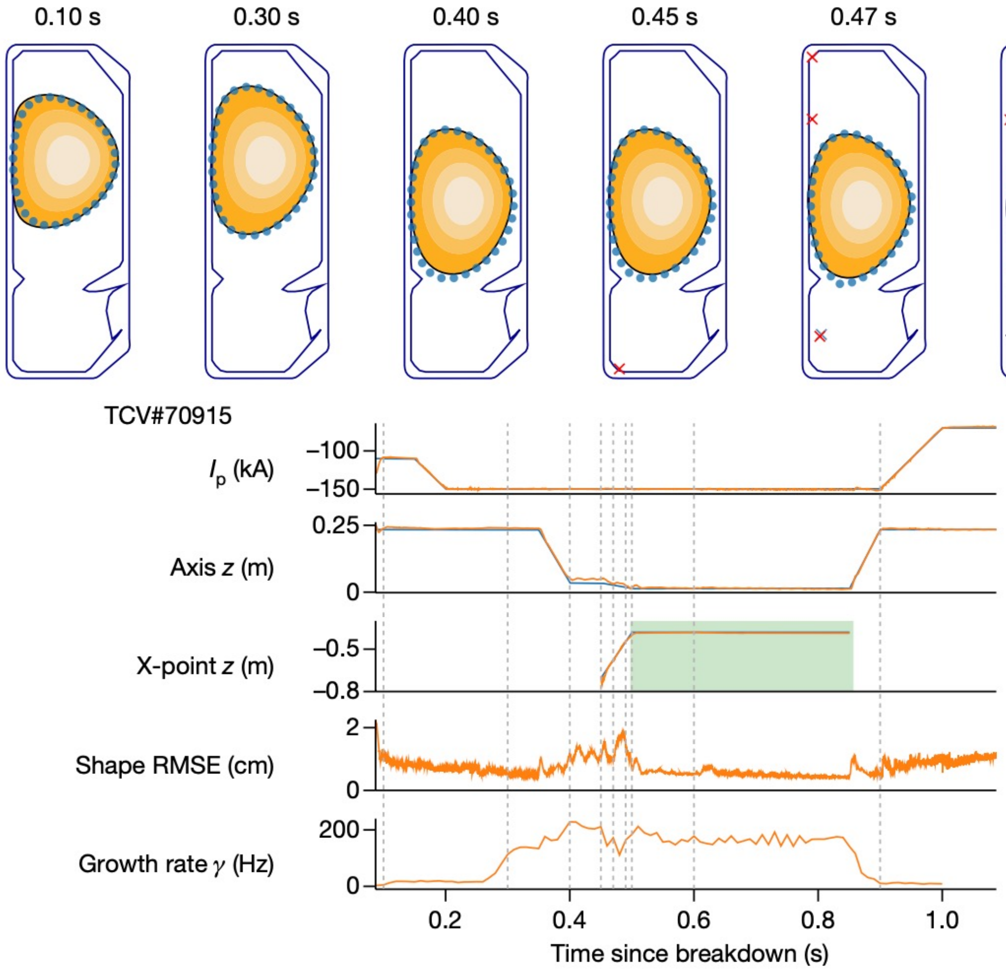


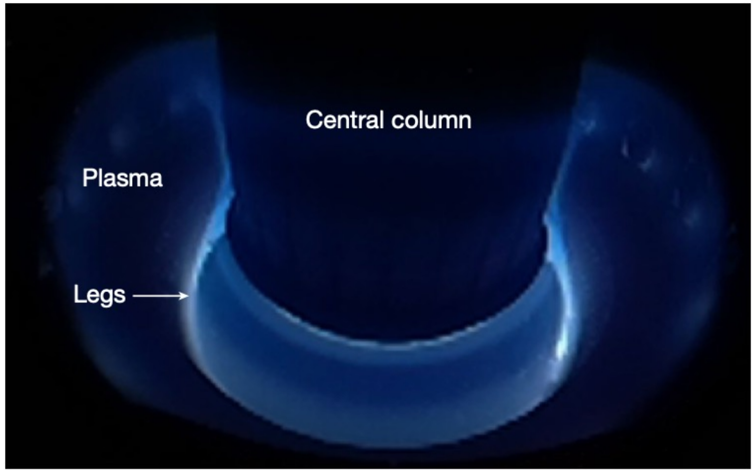
```python
def run_loop(env: environment.Environment, agent,
             max_steps: int = 100000) -> trajectory.Trajectory:
  """Run an agent."""
  results = []
  agent.reset()
  ts = env.reset()
  for _ in range(max_steps):
    obs = ts.observation
    action = agent.step(ts)
    ts = env.step(action)
    results.append(trajectory.Trajectory(
        measurements=obs["measurements"],
        references=obs["references"],
        actions=action,
        reward=np.array(ts.reward)))
    if ts.last():
      break

  return trajectory.Trajectory.stack(results)
```

6. Main results

The position and shape (orange line) matches well with the target (blue)

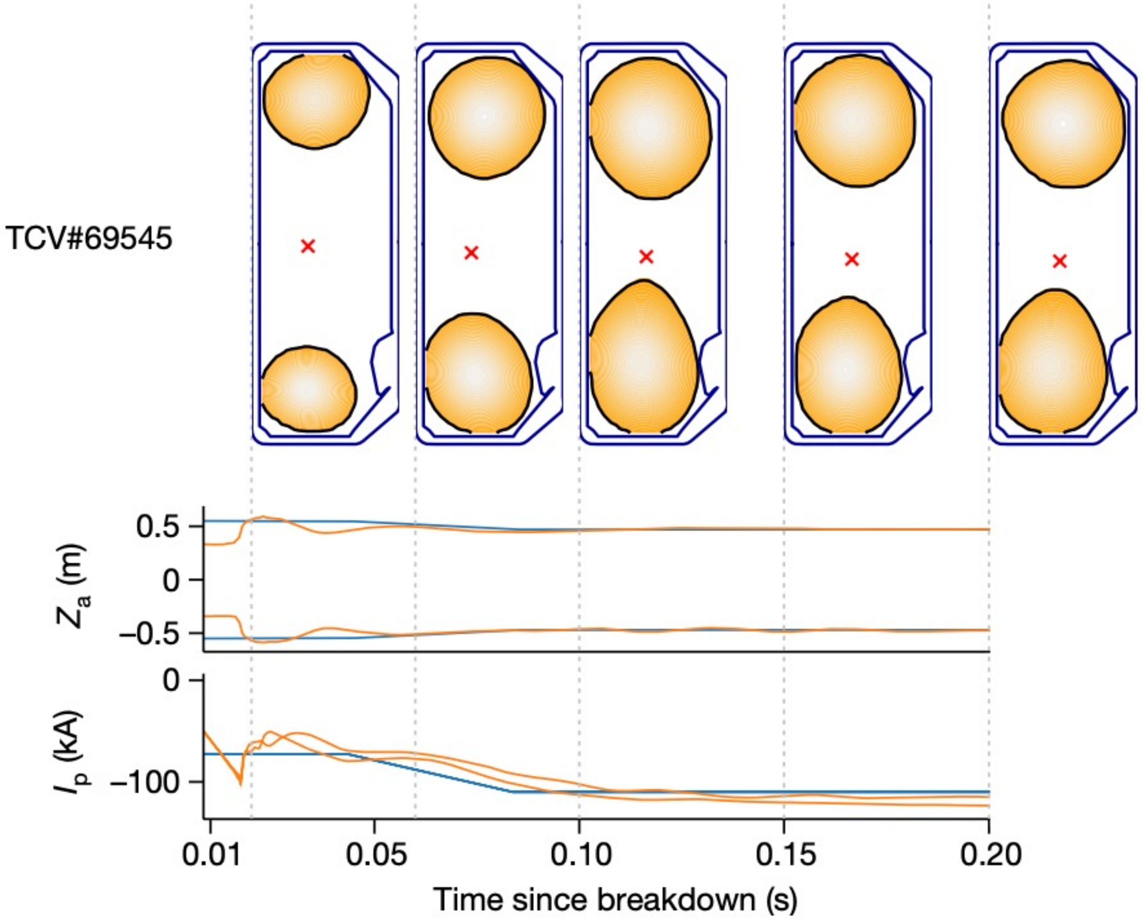


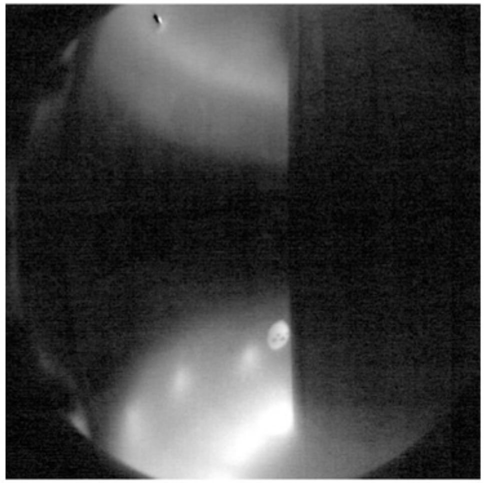


Inside view at 0.6 s

6. Main results

First demonstration of double droplet shape:





Other Deep RL work in AI4Science: Life science (1)

Protein:

1. Wang, Yi, et al. "Self-play reinforcement learning guides protein engineering." *Nature Machine Intelligence* 5.8 (2023): 845-860.
2. Lutz, Isaac D., et al. "Top-down design of protein architectures with reinforcement learning." *Science* 380.6642 (2023): 266-273.
3. Lee, Minji, et al. "Protein sequence design in a latent space via model-based reinforcement learning." (2022).
4. Xu, Xiaopeng, et al. "AB-Gen: antibody library design with generative pre-trained transformer and deep reinforcement learning." *Genomics, Proteomics & Bioinformatics* (2023).

Molecules:

1. Jeon, Woosung, and Dongsup Kim. "Autonomous molecule generation using reinforcement learning and docking to develop potential novel inhibitors." *Scientific reports* 10.1 (2020): 22104.
2. Korshunova, Maria, et al. "Generative and reinforcement learning approaches for the automated de novo design of bioactive compounds." *Communications Chemistry* 5.1 (2022): 129.
3. Mazuz, Eyal, et al. "Molecule generation using transformers and policy gradient reinforcement learning." *Scientific Reports* 13.1 (2023): 8799.
4. Polykovskiy, Daniil, et al. "Molecular sets (MOSES): a benchmarking platform for molecular generation models." *Frontiers in pharmacology* 11 (2020): 565644.

Other Deep RL work in AI4Science: Life science (2)

Molecules (continued):

5. Hu, Xiuyuan, et al. "De novo Drug Design using Reinforcement Learning with Multiple GPT Agents." *Advances in Neural Information Processing Systems* 36 (2024).
6. Popova, Mariya, Olexandr Isayev, and Alexander Tropsha. "Deep reinforcement learning for de novo drug design." *Science advances* 4.7 (2018): eaap7885.

RNA:

1. Whatley, Alexander, Zhekun Luo, and Xiangru Tang. "Improving RNA secondary structure design using deep reinforcement learning." *arXiv preprint arXiv:2111.04504* (2021).
2. Eastman, Peter, et al. "Solving the RNA design problem with reinforcement learning." *PLoS computational biology* 14.6 (2018): e1006176.

Genomics:

1. Nicholls, Hannah L., et al. "Reaching the end-game for GWAS: machine learning approaches for the prioritization of complex disease loci." *Frontiers in genetics* 11 (2020): 521712.
2. Karami, Mohsen, et al. "Revolutionizing genomics with reinforcement learning techniques." *arXiv preprint arXiv:2302.13268* (2023).

Other Deep RL work in AI4Science: Fluid control (1)

Cylinder:

1. Chen, Wenjie, et al. "Deep reinforcement learning-based active flow control of vortex-induced vibration of a square cylinder." *Physics of Fluids* 35.5 (2023). (SAC)
2. Wang, Qiulei, et al. "DRLinFluids: An open-source Python platform of coupling deep reinforcement learning and OpenFOAM." *Physics of Fluids* 34.8 (2022). (SAC)
3. Wang, Qiulei, et al. "Dynamic feature-based deep reinforcement learning for flow control of circular cylinder with sparse surface pressure sensing." *arXiv preprint arXiv:2307.01995* (2023). (SAC & PPO)
4. Tang, Hongwei, et al. "Robust active flow control over a range of Reynolds numbers using an artificial neural network trained through deep reinforcement learning." *Physics of Fluids* 32.5 (2020). (PPO)
5. Xu, Hui, et al. "Active flow control with rotating cylinders by an artificial neural network trained by deep reinforcement learning." *Journal of Hydrodynamics* 32.2 (2020): 254-258. (PPO)
6. Rabault, Jean, et al. "Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control." *Journal of fluid mechanics* 865 (2019): 281-302. (PPO)
7. Wang, Zhicheng, et al. "Deep reinforcement transfer learning of active control for bluff body flows at high Reynolds number." *Journal of Fluid Mechanics* 973 (2023): A32. (TD3)
8. Zheng, Changdong, et al. "Data-efficient deep reinforcement learning with expert demonstration for active flow control." *Physics of Fluids* 34.11 (2022). (SAC)

Other Deep RL work in AI4Science: Fluid control (2)

Point:

1. Mei, Jiazhong, J. Nathan Kutz, and Steven L. Brunton. "Observability-Based Energy Efficient Path Planning with Background Flow via Deep Reinforcement Learning." *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023. (PPO)
2. Gunnarson, Peter, et al. "Learning efficient navigation in vortical flow fields." *Nature communications* 12.1 (2021): 7143. (V-RACER)

Foil:

1. Novati, Guido, and Petros Koumoutsakos. "Remember and forget for experience replay." *International Conference on Machine Learning*. PMLR, 2019. (V-RACER)
2. Wang ZP, Lin RJ, Zhao ZY, et al. Learn to flap: foil non-parametric path planning via deep reinforcement learning. *Journal of Fluid Mechanics*. 2024;984:A9. (PPO)

Fish:

1. Verma, Siddhartha, Guido Novati, and Petros Koumoutsakos. "Efficient collective swimming by harnessing vortices through deep reinforcement learning." *Proceedings of the National Academy of Sciences* 115.23 (2018): 5849-5854. (DRQN)
2. Mandralis, Ioannis, et al. "Learning swimming escape patterns for larval fish under energy constraints." *Physical Review Fluids* 6.9 (2021): 093101. (V-RACER)

Other Deep RL work in AI4Science: Materials science (1)

Materials:

1. Rajak, Pankaj, et al. "Autonomous reinforcement learning agent for chemical vapor deposition synthesis of quantum materials." *npj Computational Materials* 7.1 (2021): 108.
2. Zamaraeva, Elena, et al. "Reinforcement learning in crystal structure prediction." *Digital Discovery* 2.6 (2023): 1831-1840.
3. Zheng, Bowen, Zeyu Zheng, and Grace X. Gu. "Designing mechanically tough graphene oxide materials using deep reinforcement learning." *npj Computational Materials* 8.1 (2022): 225.
4. Govindarajan, Prashant, et al. "Learning Conditional Policies for Crystal Design Using Offline Reinforcement Learning." *Digital Discovery* (2024).
5. Pandey, Ashish, et al. "Reinforcement learning based carbon nanotube growth automation." *2021 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. IEEE, 2021.
6. Pan, Elton, Christopher Karpovich, and Elsa Olivetti. "Deep reinforcement learning for inverse inorganic materials design." *arXiv preprint arXiv:2210.11931* (2022).

Other Deep RL work in AI4Science: Materials science (2)

Meta-materials/composite/polymer:

1. Sui, Fanping, et al. "Deep reinforcement learning for digital materials design." *ACS Materials Letters* 3.10 (2021): 1433-1439.
2. Gongora, Aldair E., et al. "Designing composites with target effective young’s modulus using reinforcement learning." *Proceedings of the 6th Annual ACM Symposium on Computational Fabrication*. 2021.
3. Ma, Ruimin, Hanfeng Zhang, and Tengfei Luo. "Exploring high thermal conductivity amorphous polymers using reinforcement learning." *ACS Applied Materials & Interfaces* 14.13 (2022): 15587-15598.
4. Rosafalco, Luca, et al. "Reinforcement learning optimisation for graded metamaterial design using a physical-based constraint on the state representation and action space." *Scientific Reports* 13.1 (2023): 21836.