EPFL

Magnetic Control of Tokamak Plasmas through Deep Reinforcement Learning

Brendan Tracey - Google DeepMind, London

on behalf of the DeepMind-EPFL team

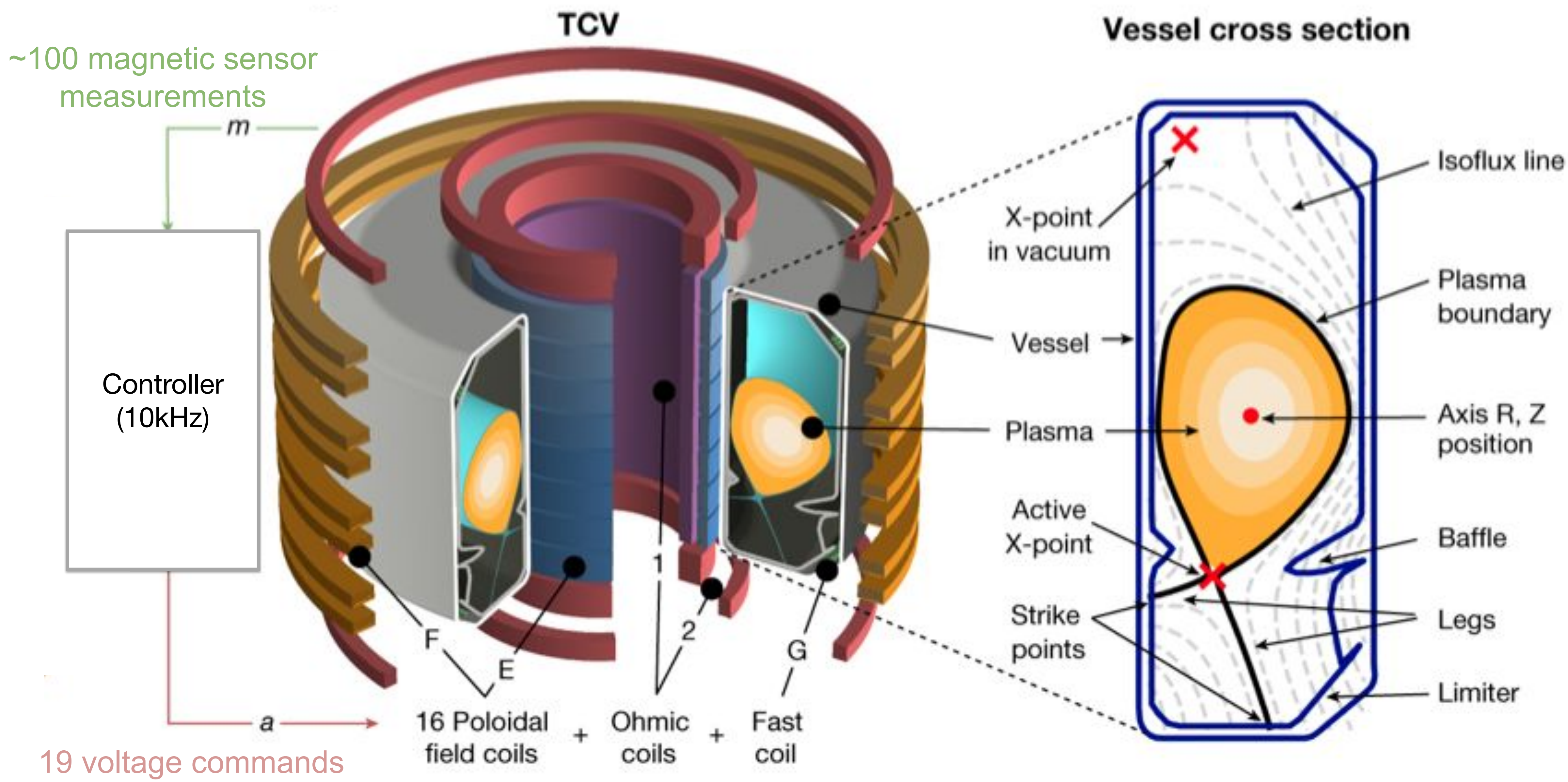
Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Abbas Abdolmaleki, Diego de las Casas, Craig Donner, Leslie Fritz, Cristian Galperti, Andrea Huber, James Keeling, Maria Tsimpoukelli, Jackie Kay, Antoine Merle, Jean-Marc Moret, Seb Noury, Federico Pesamosca, David Pfau, Olivier Sauter, Cristian Sommariva, Stefano Coda, Basil Duval, Ambrogio Fasoli, Ian Davies, Andrea Michi, Yuri Chervonyi, Pushmeet Kohli, Koray Kavukcuoglu, Demis Hassabis & Martin Riedmiller

Workshop on AI for Accelerating Fusion and Plasma Science
November 2023

SWISS PLASMA CENTER

DeepMind

Axisymmetric Equilibrium Control



Axisymmetric Tokamak Plasma Control

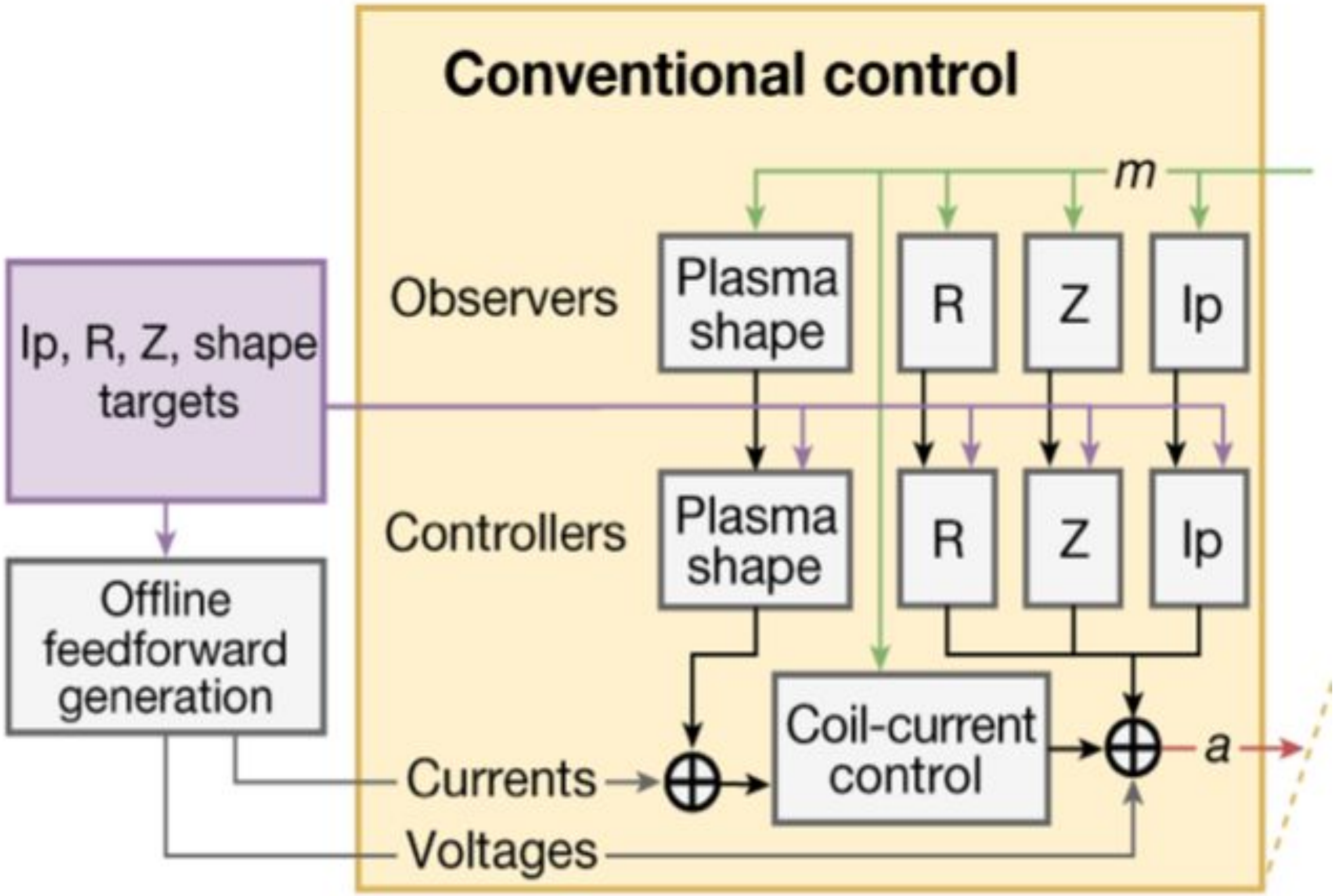
- **Need to control:**
  - **Total plasma current I_p**
  - **Radial position R (by vertical magnetic fields)**
  - **Vertical position Z (by radial magnetic fields - *unstable for elongated plasmas*)**
  - **Plasma shape: Last Closed Flux Surface (LCFS)**

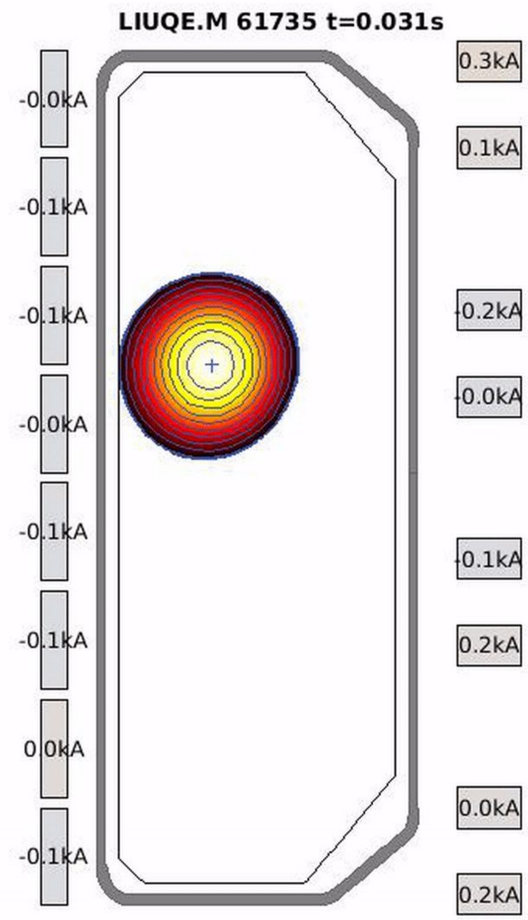
EPFL

DeepMind

Traditional Solutions

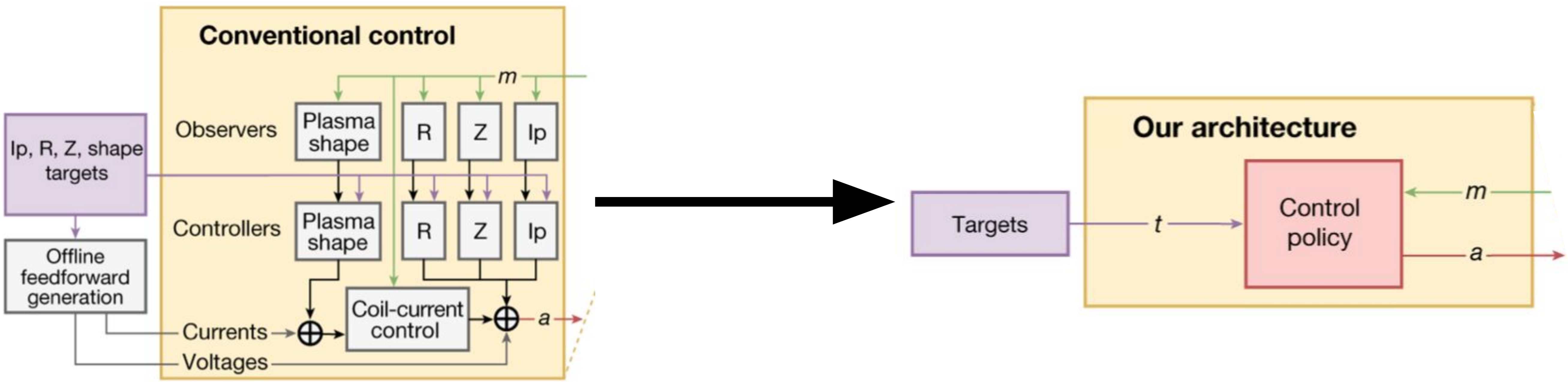
1. Choose combinations of coils to use to control each quantity
2. Pre-compute feedforward coil currents and voltages
3. Design feedback controllers
4. Tune individual control parameters for each (hopefully orthogonal) control loop





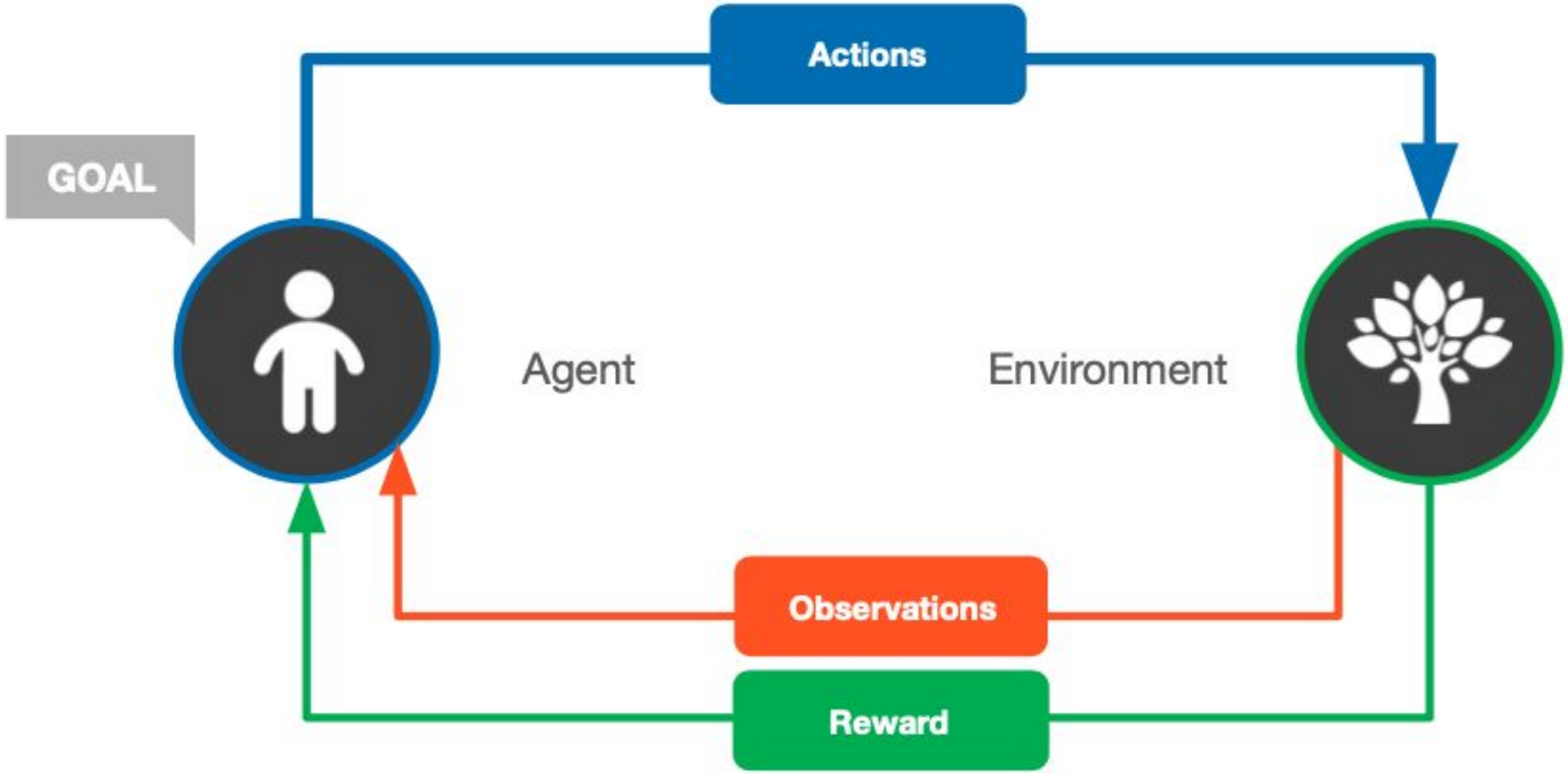
EPFL DeepMind

Reinforcement Learning Solution



- Single Integrated Controller
- No feedforward generation
- No separate error estimation

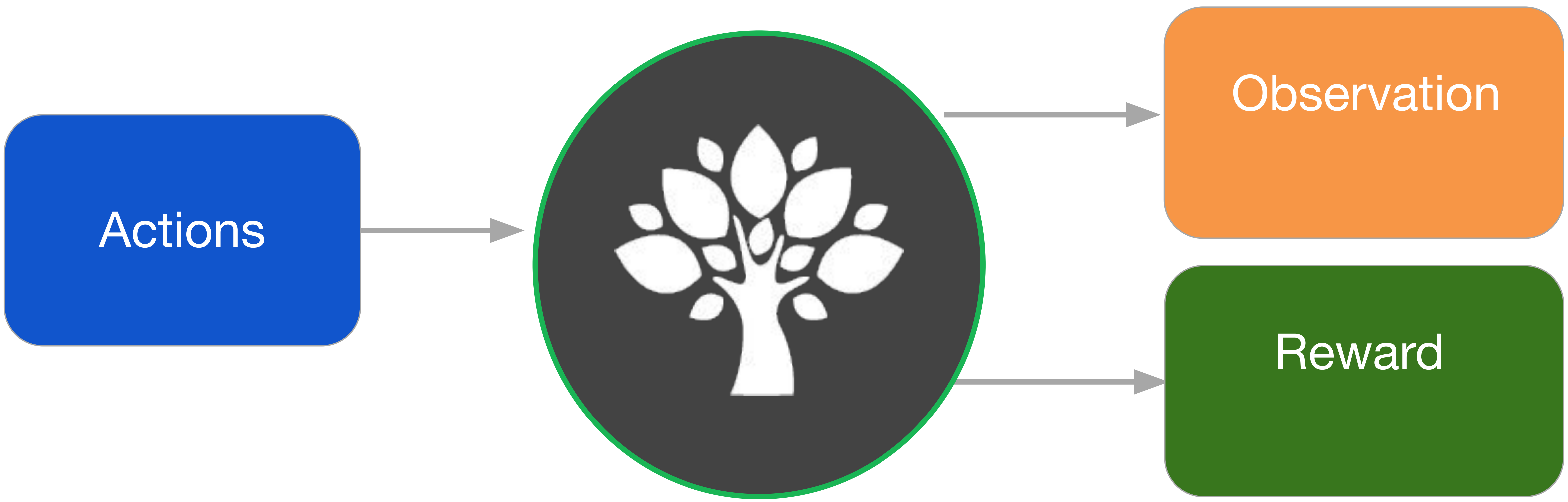
Reinforcement Learning



[Figure and RL slide material from hereon: courtesy A. Abdolmaleki]

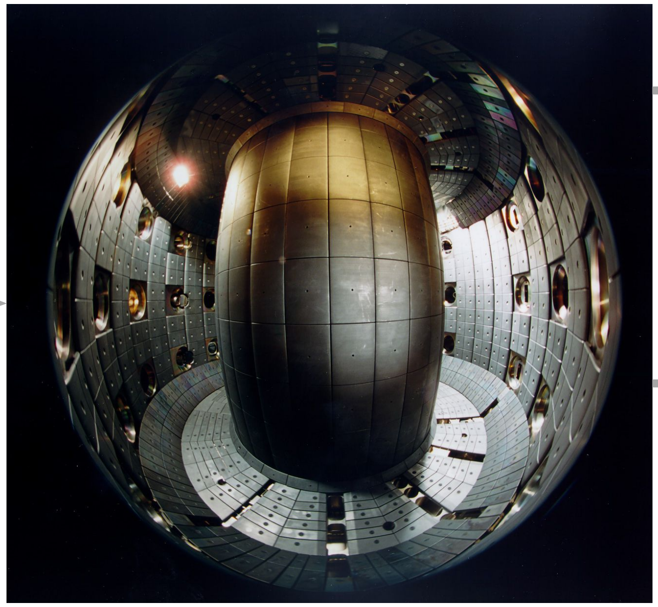
Learn an action selection function (π) through <u>trial-and-error</u> to achieve high reward

What is an environment?



Environment

19 Actions:
Control Coil Voltages



TCV Tokamak

Reward

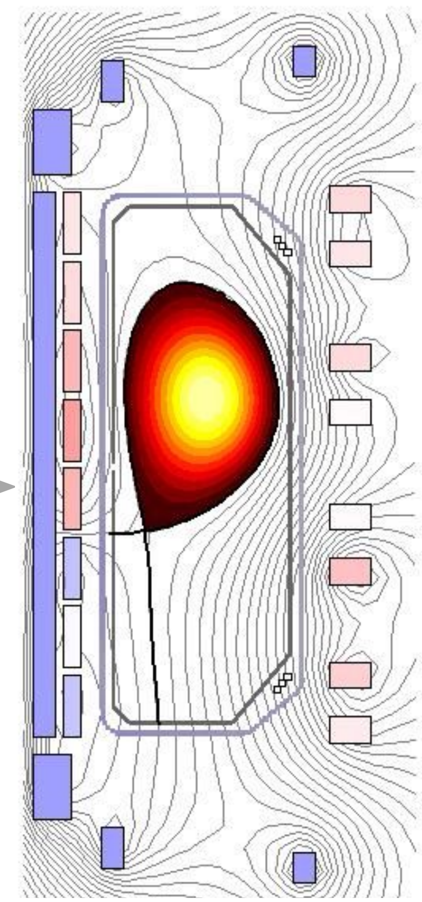
92 Observations:
- 34 Flux Loops
- 38 Magnetic Probes
- 20 Coil Currents

Training on Hardware is Impractical

EPFL DeepMind

Training Environment

19 Control Coil Voltages



Reward

92 Observations:

- 34 Flux Loops
- 38 Magnetic Probes
- 20 Coil Currents

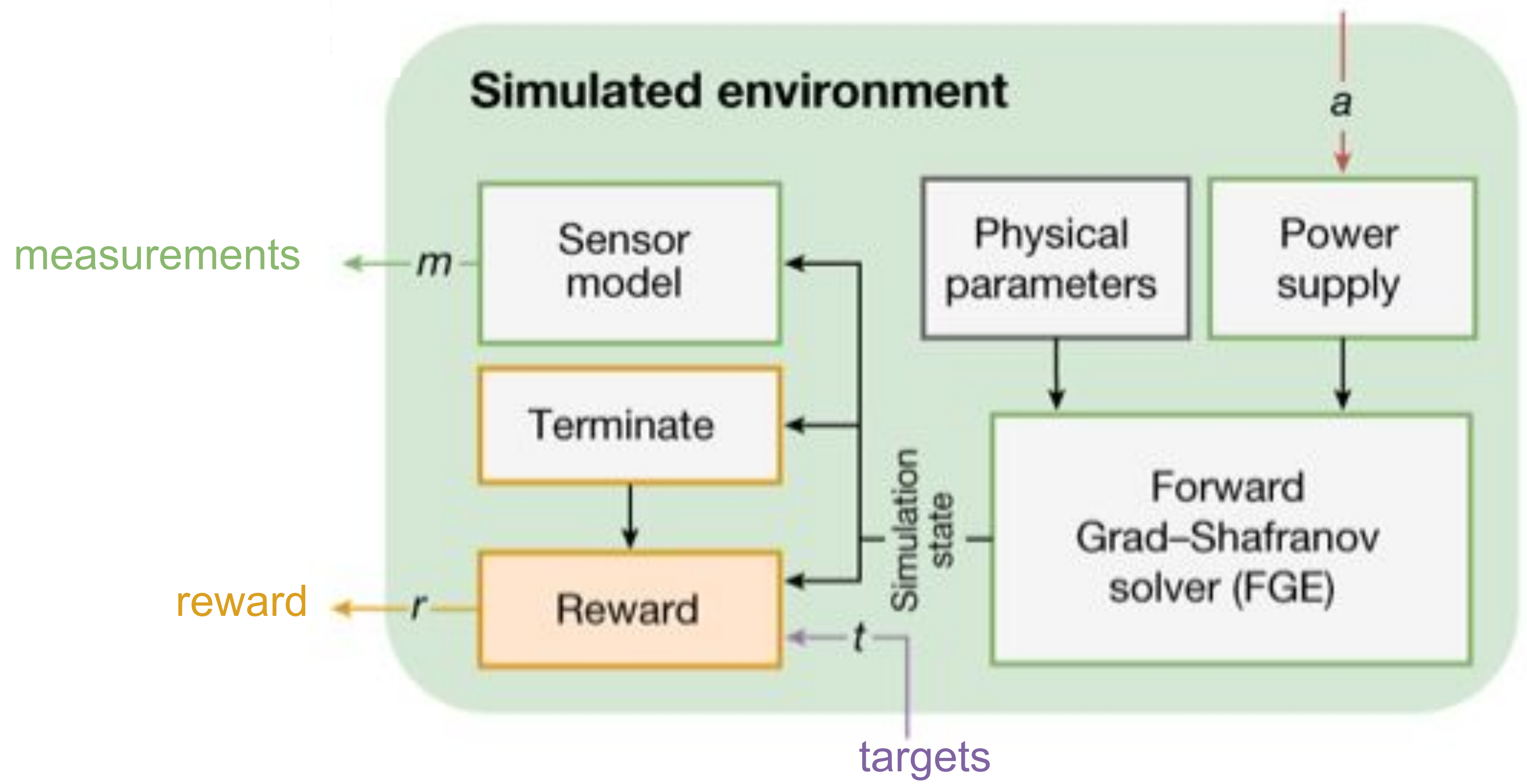
FGE Simulator [Carpanese EPFL PhD 2020]

- Free boundary Grad-Shafranov solver
- Circuit equations for conductors

EPFL

DeepMind

Full Simulation Model

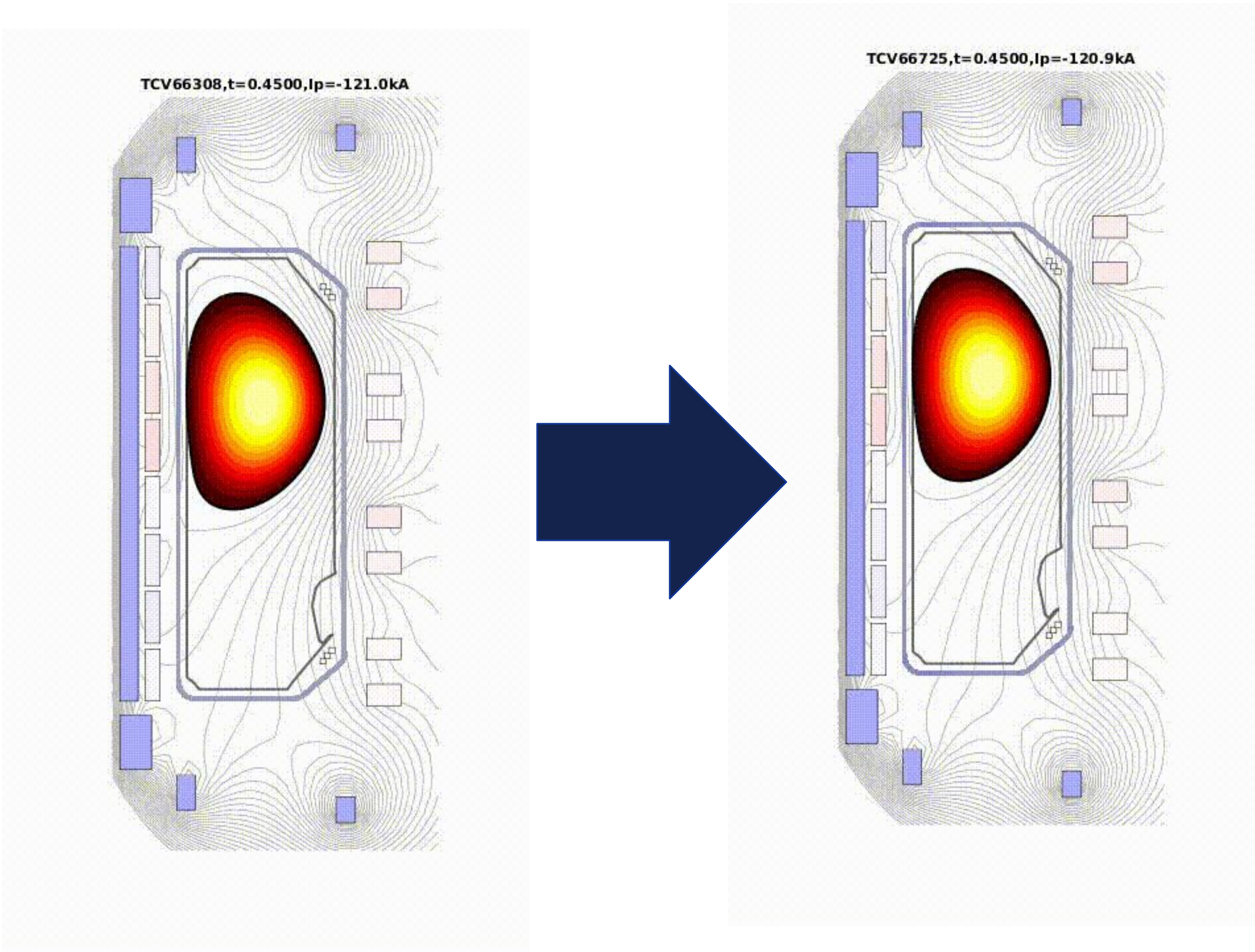


- A lot of physics/engineering know-how goes into the simulator
- Need for domain experts: plasma physicists / tokamak engineers & modelers

Reward Design

Goals

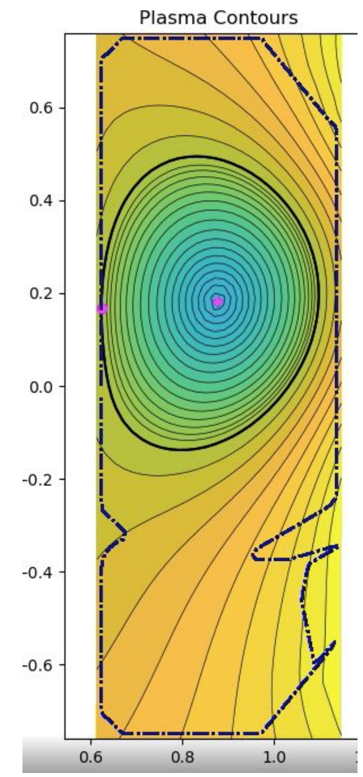
1. **Keep the plasma alive**
2. Stabilize the plasma location
3. Shape Control



Shot 66308:
(0.024s real time)

Shot 66725:
(0.55s real time)

Reward Design

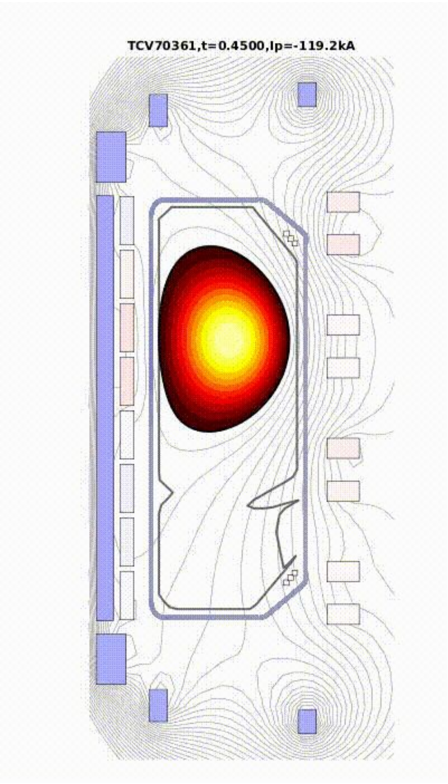


Stabilize:

- R Centroid
- Z Centroid
- Plasma Current (Ip)

Goals

1. Keep the plasma alive
2. **Stabilize the plasma location**
3. Shape Control



Stabilizing the position through control
Shot 70361: 0.55s real time

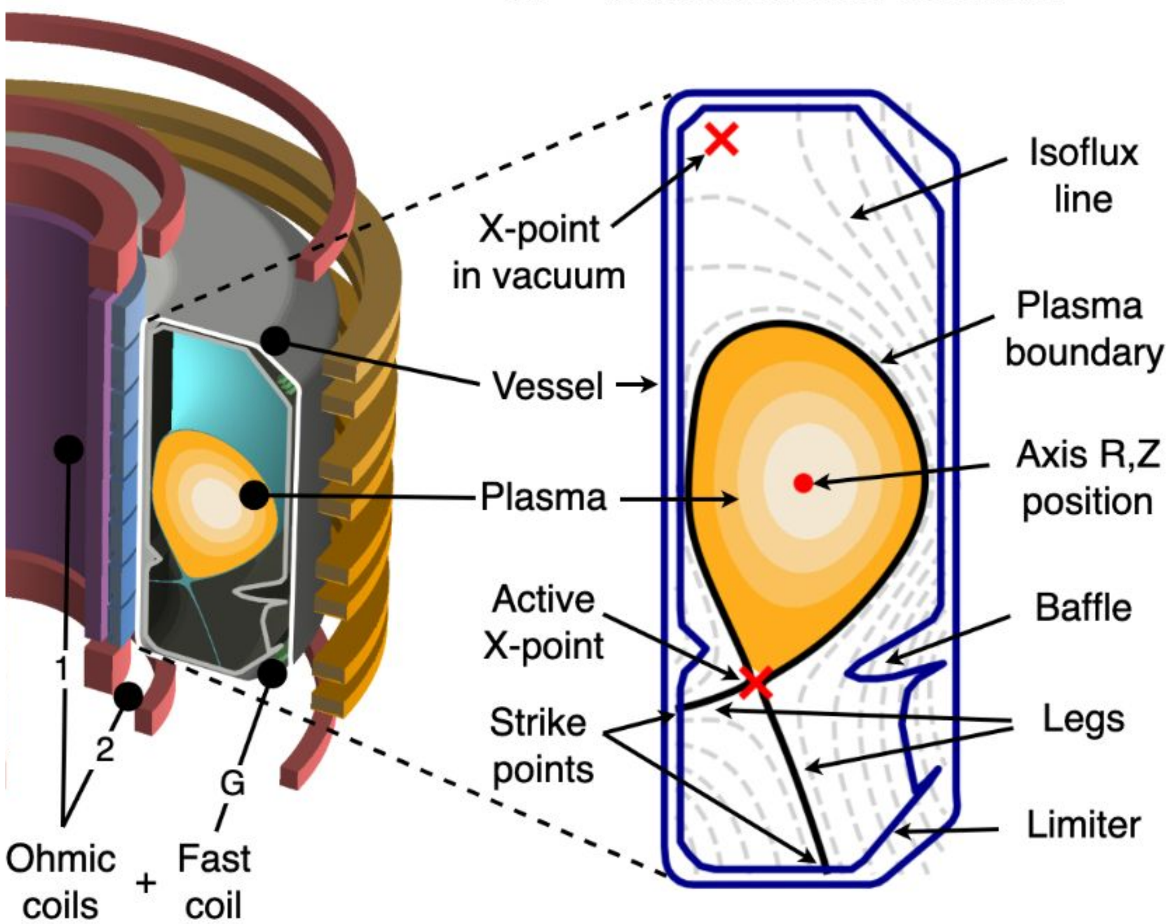
Reward Design

Goals

1. Keep the plasma alive
2. Stabilize the plasma location
3. **Shape Control**

$$\texttt{SoftPlus}(x) = \left[2 \cdot \sigma\left(f_{\text{scale}}(x, \text{good}, \text{bad}, 0, \zeta)\right)\right]_0^1,$$

$$\texttt{SmoothMax}(x_{1\ldots n}, w_{1\ldots n}, \alpha) = \frac{\sum_{i=1}^{n} w_i x_i e^{\alpha x_i}}{\sum_{i=1}^{n} w_i e^{\alpha x_i}}.$$



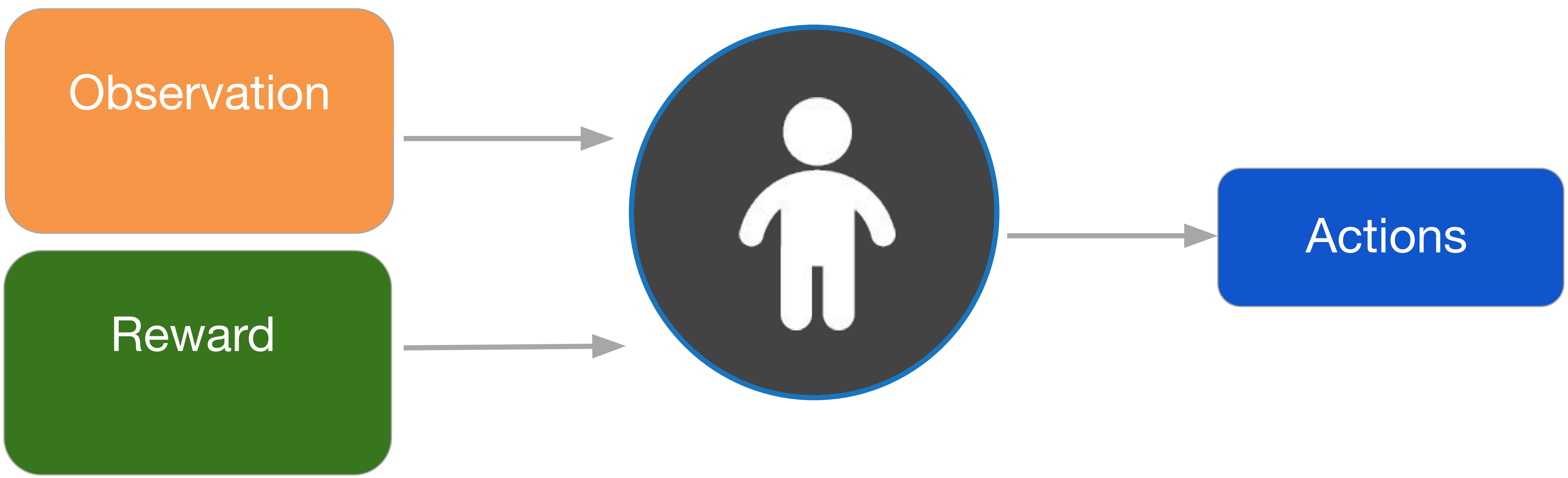
Shape Control – Must match:

1) Target shape outline
2) Active X-point location
3) Passive X-point locations
4) Leg locations
5) Plasma current

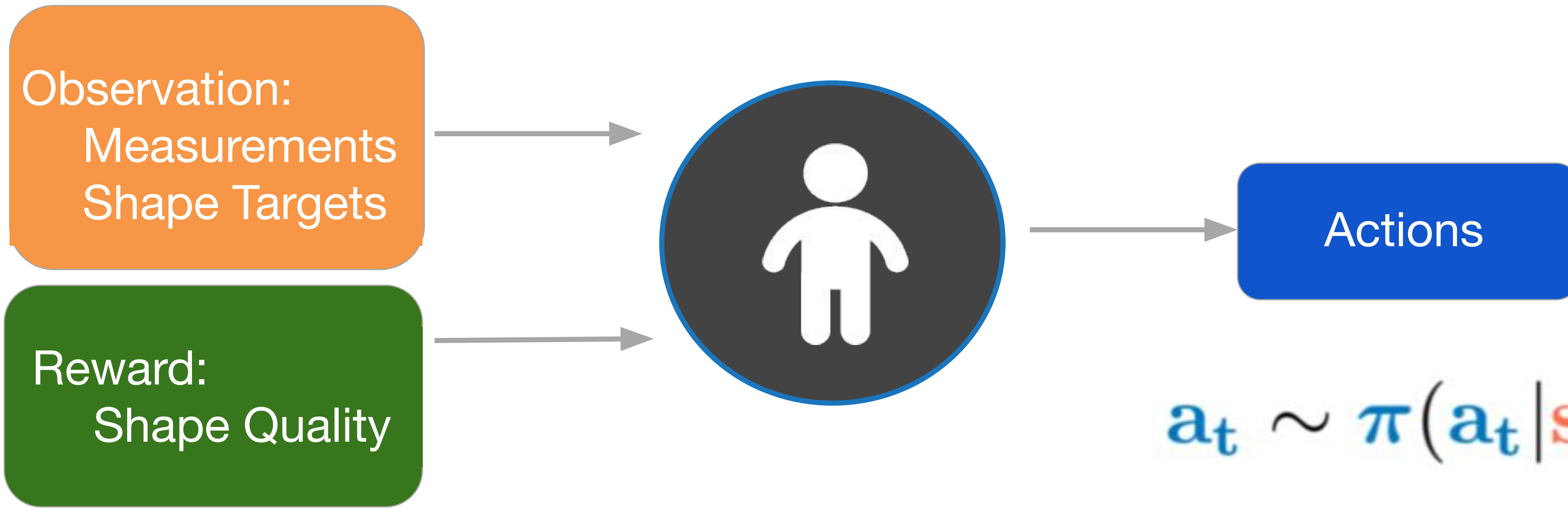
Whilst maintaining OH coil current stability

EPFL DeepMind

What is an Agent?



Tokamak Agent



$$a_t \sim \pi(a_t | s_t)$$

Aim

Find **optimal policy** – maximize discounted sum of future rewards

Q Function: State-Action Values

State-action value function maps an **action** in a given **state** to **expected rewards**.

$$Q^{\pi}(s_t, a_t) = \mathbb{E}_{\pi}\left[r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots | s_t, a_t\right]$$

It is equal to **expected total discounted reward** for an agent starting from state **s** and performing action **a** and following its **policy**.

Actor-Critic Methods

- **Critic estimates the *Q function* from data generated by interacting with environment**

- **Actor learns a policy π by ascending (via gradient-based methods) the *Q function* learned by the critic**

Actor-Critic Methods

- Critic estimates the *Q function* from data generated by interacting with environment

- Actor learns a policy π by ascending (via gradient-based methods) the *Q function* learned by the critic

- Actor critic approaches allow asymmetry between actor and critic
  - The critic can benefit from privileged information unknown to the actor

Actor-Critic Methods

- **Critic estimates the *Q function* from data generated by interacting with environment**

- **Actor learns a policy π by ascending (via gradient-based methods) the *Q function* learned by the critic**

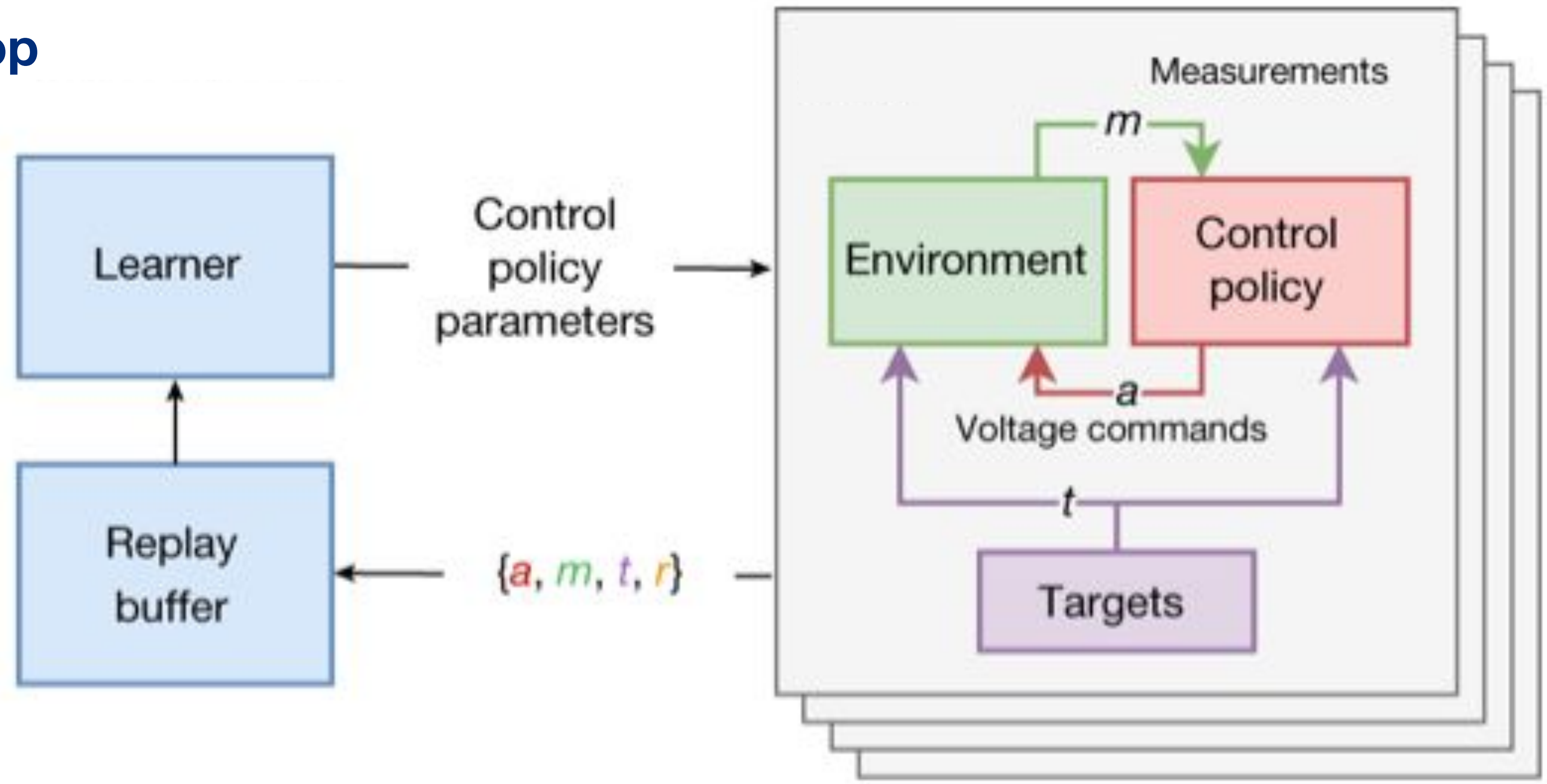
- **Actor critic approaches allow asymmetry between actor and critic**
  - **The critic can benefit from privileged information unknown to the actor**

- **We use Maximum A Posteriori Policy Optimisation**
  (MPO, Abdolmaleki et al. 2018)

EPFL DeepMind

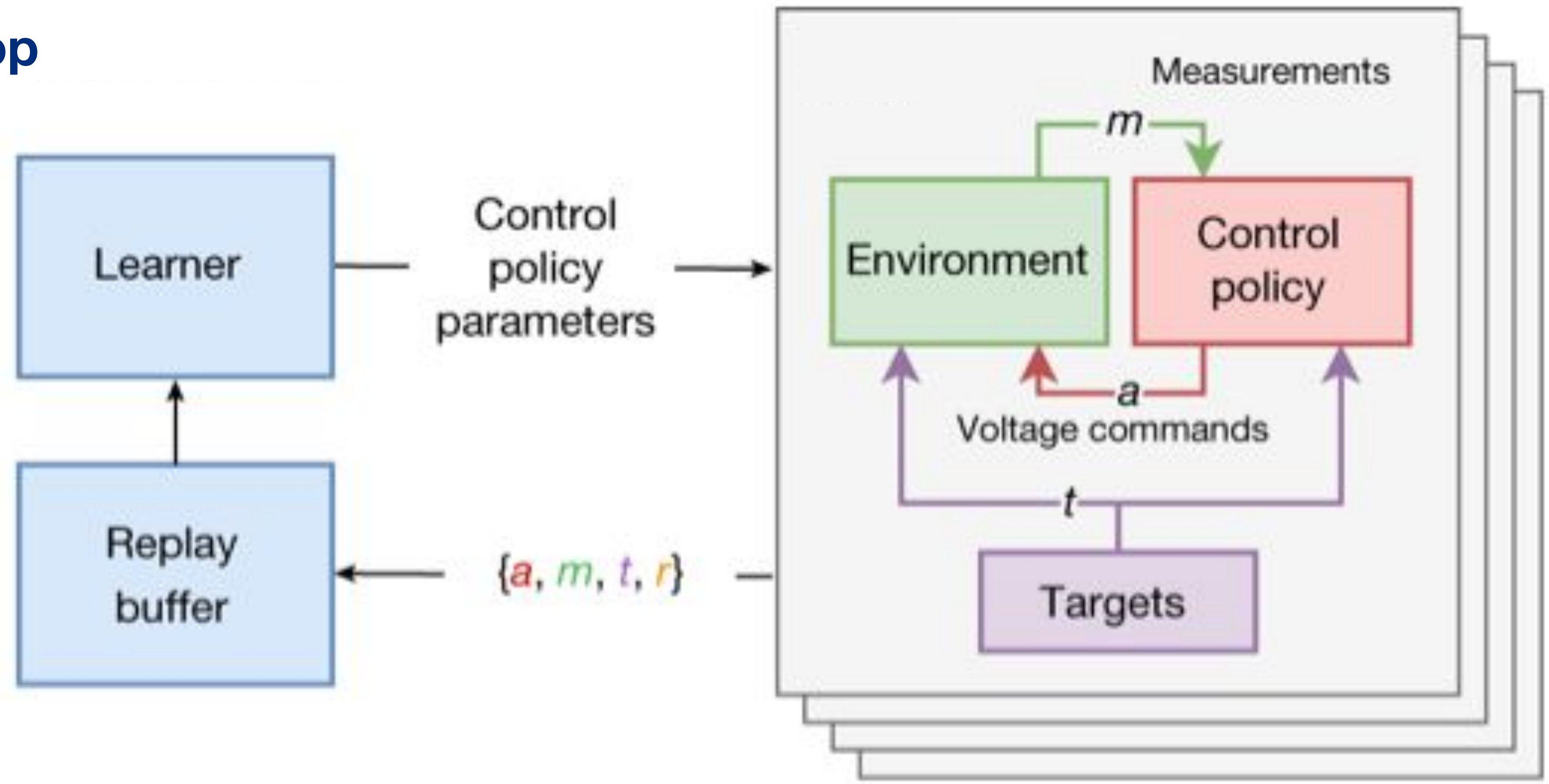
Creating an Agent

Learning Loop

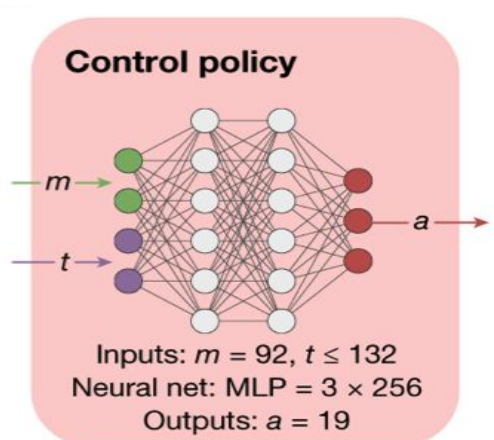


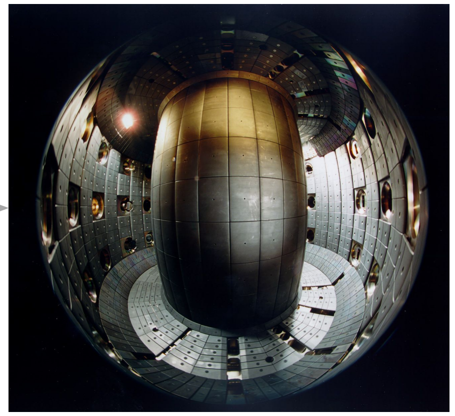
Creating an Agent

Learning Loop



Deployment





Transferring simulation-trained agents to the TCV tokamak

Deployment mostly just worked!

Iterations on reward design needed to achieve stability

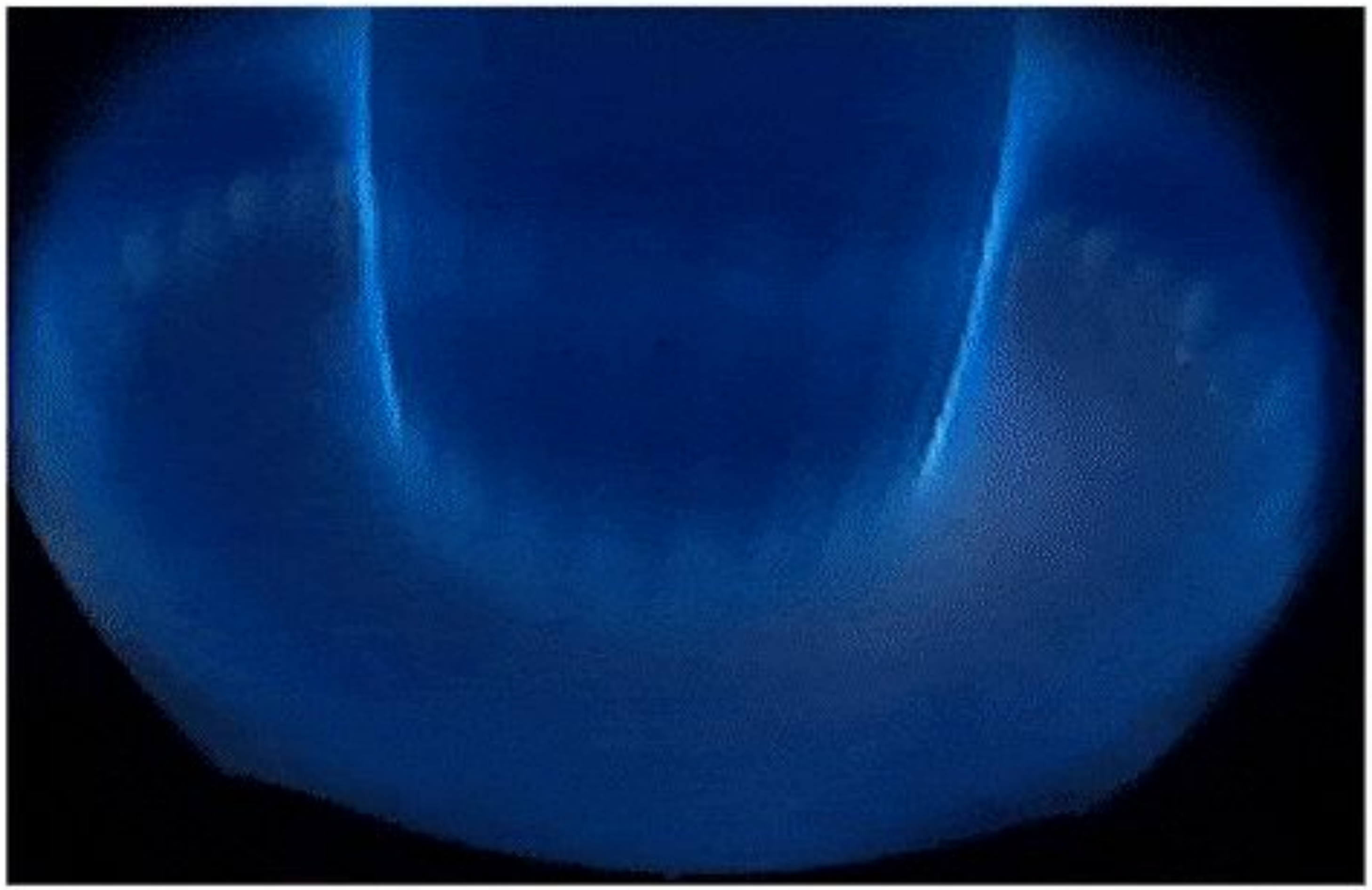
Simulator upgrades required to attain a successful agent

Targeted environment variation:

- **Measurement noise**
- **Plasma parameters (resistivity, plasma pressure ratio …)**
- **Power supply**

EPFL DeepMind

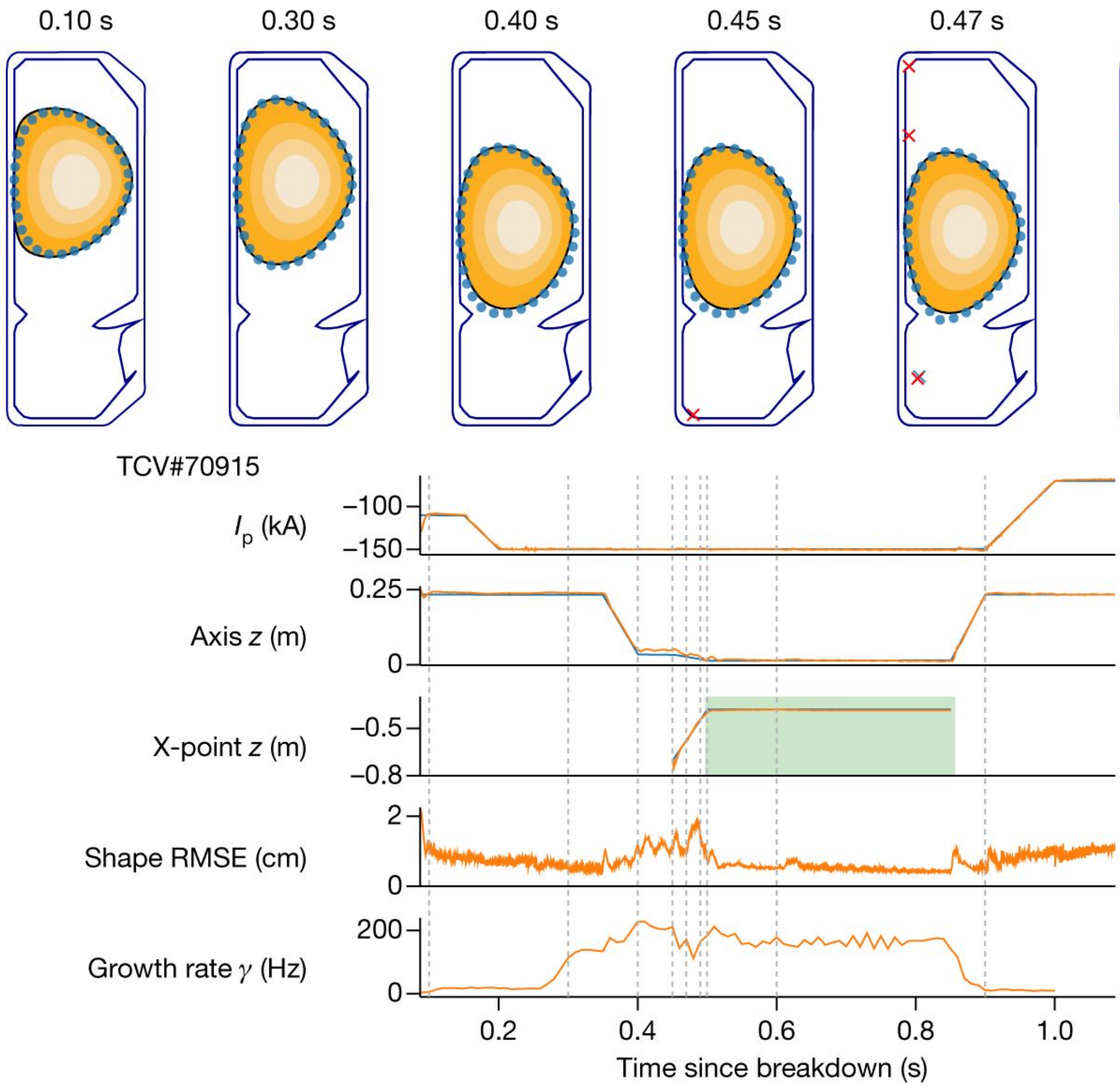
Result - demonstration shot

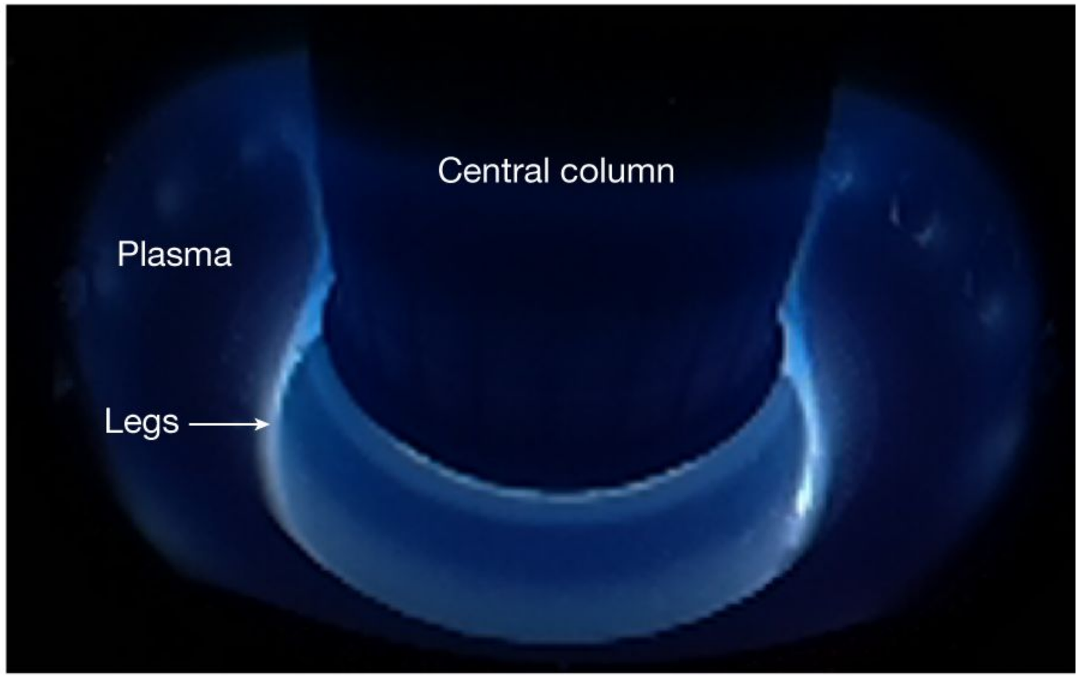


View from inside the tokamak

Plasma state reconstruction

Demonstration shot

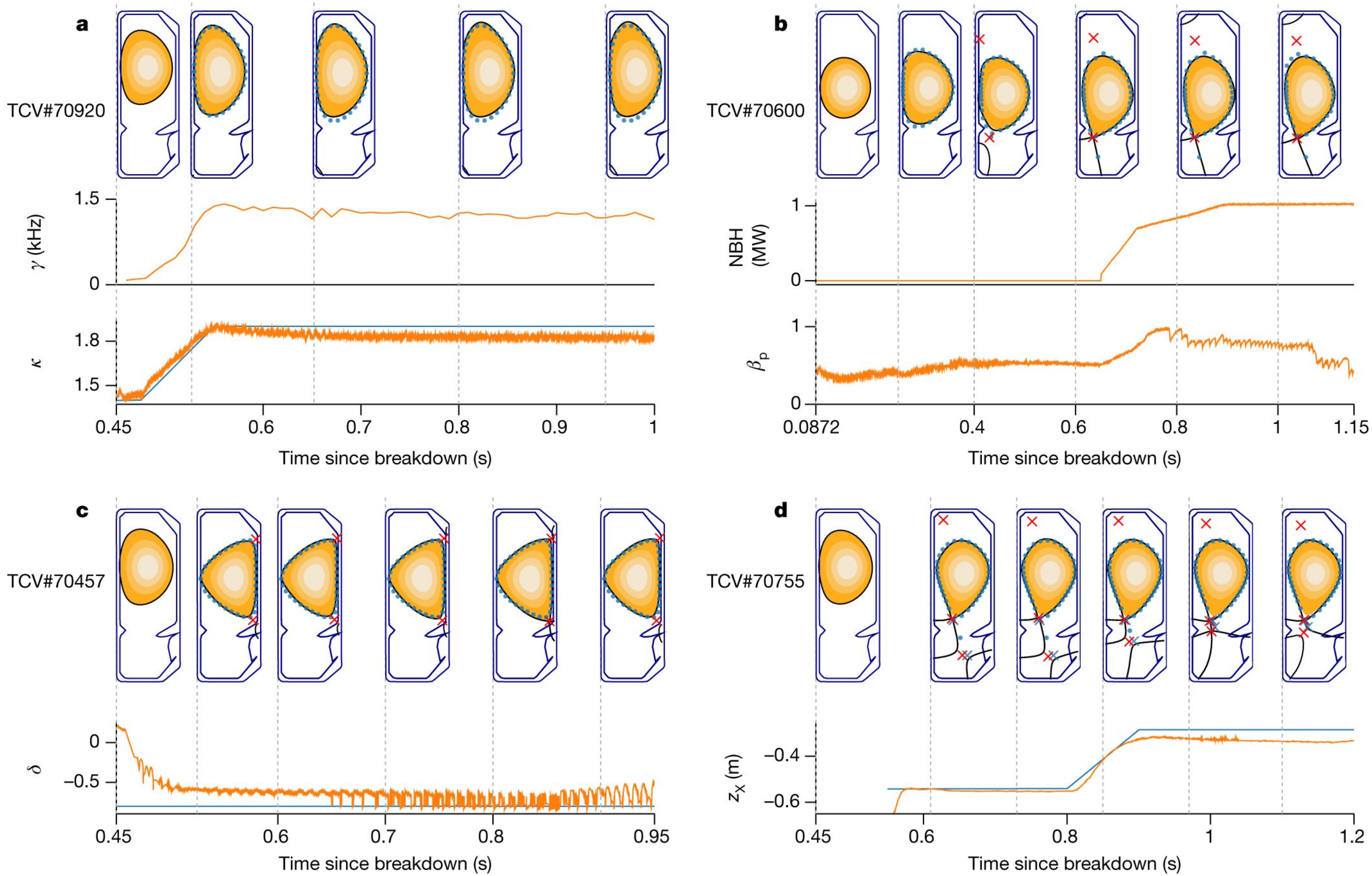




Inside view at 0.6 s

EPFL DeepMind

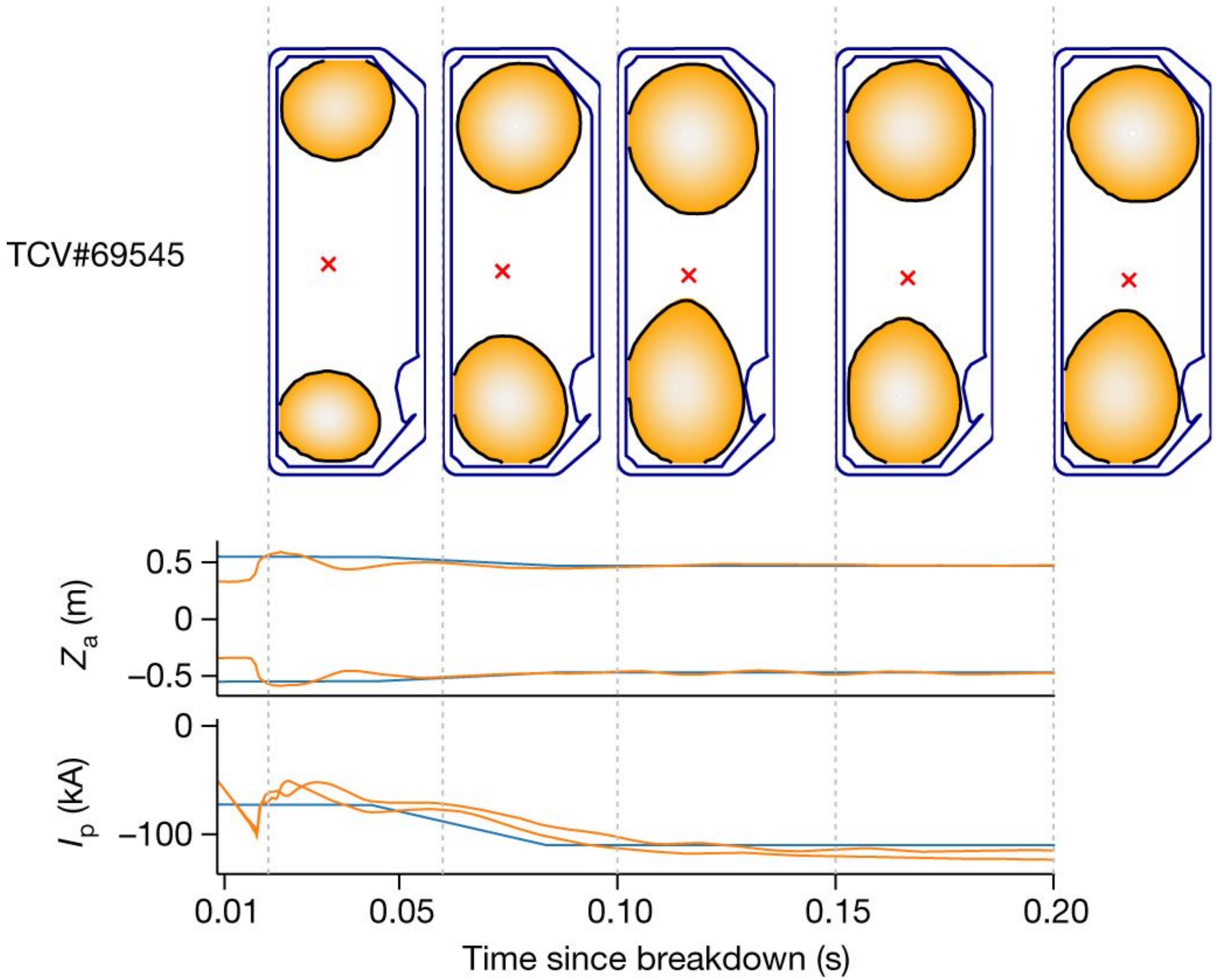
Various plasma shapes controlled in in TCV with RL

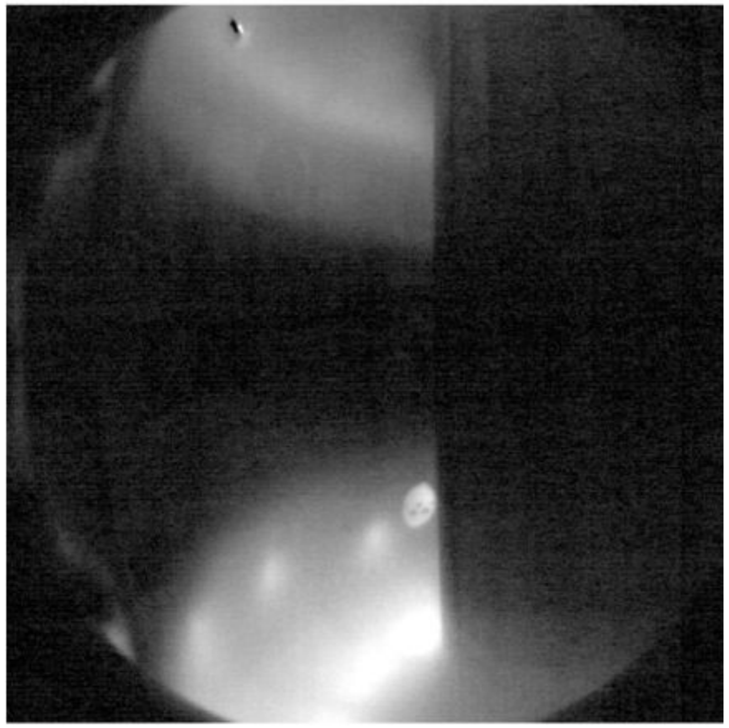


EPFL

DeepMind

Opening new frontiers for TCV: Droplet plasmas

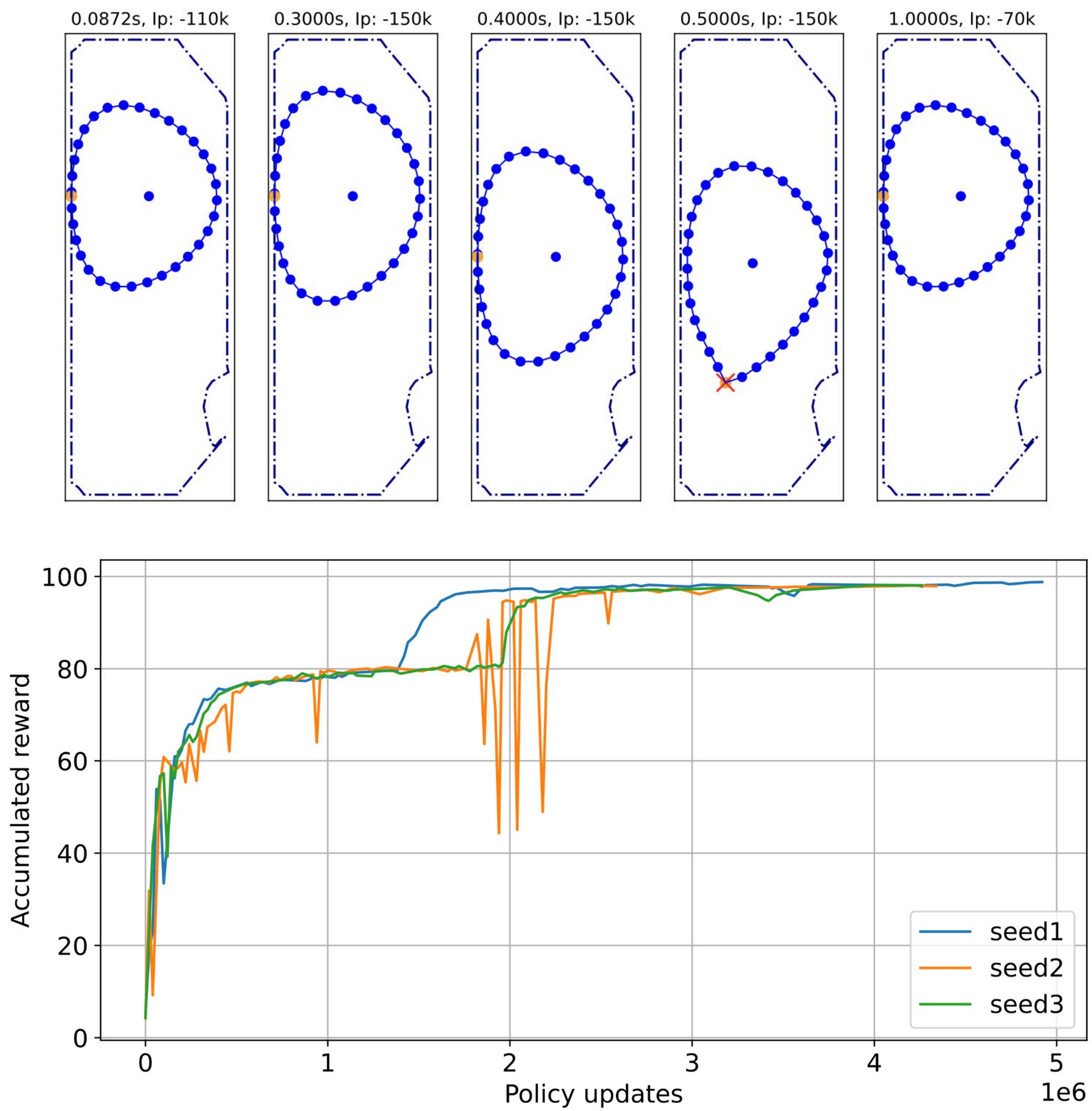




EPFL

DeepMind

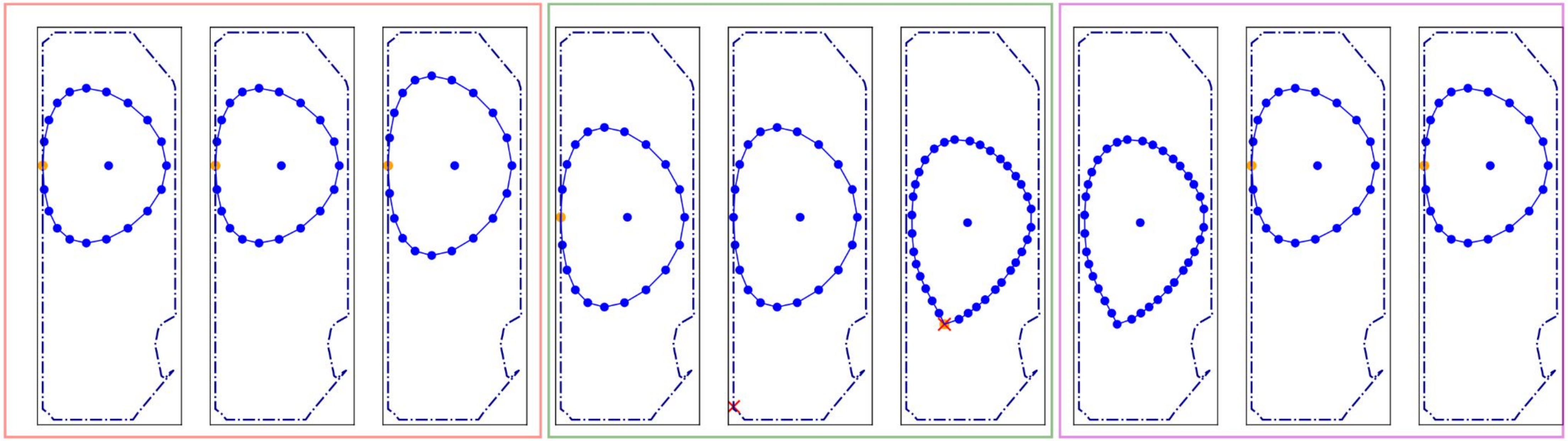
Exploration Challenges



Episode Chunking

Chunk (3k)	Chunk (3k)	Chunk (3k)

Episode (10k steps)



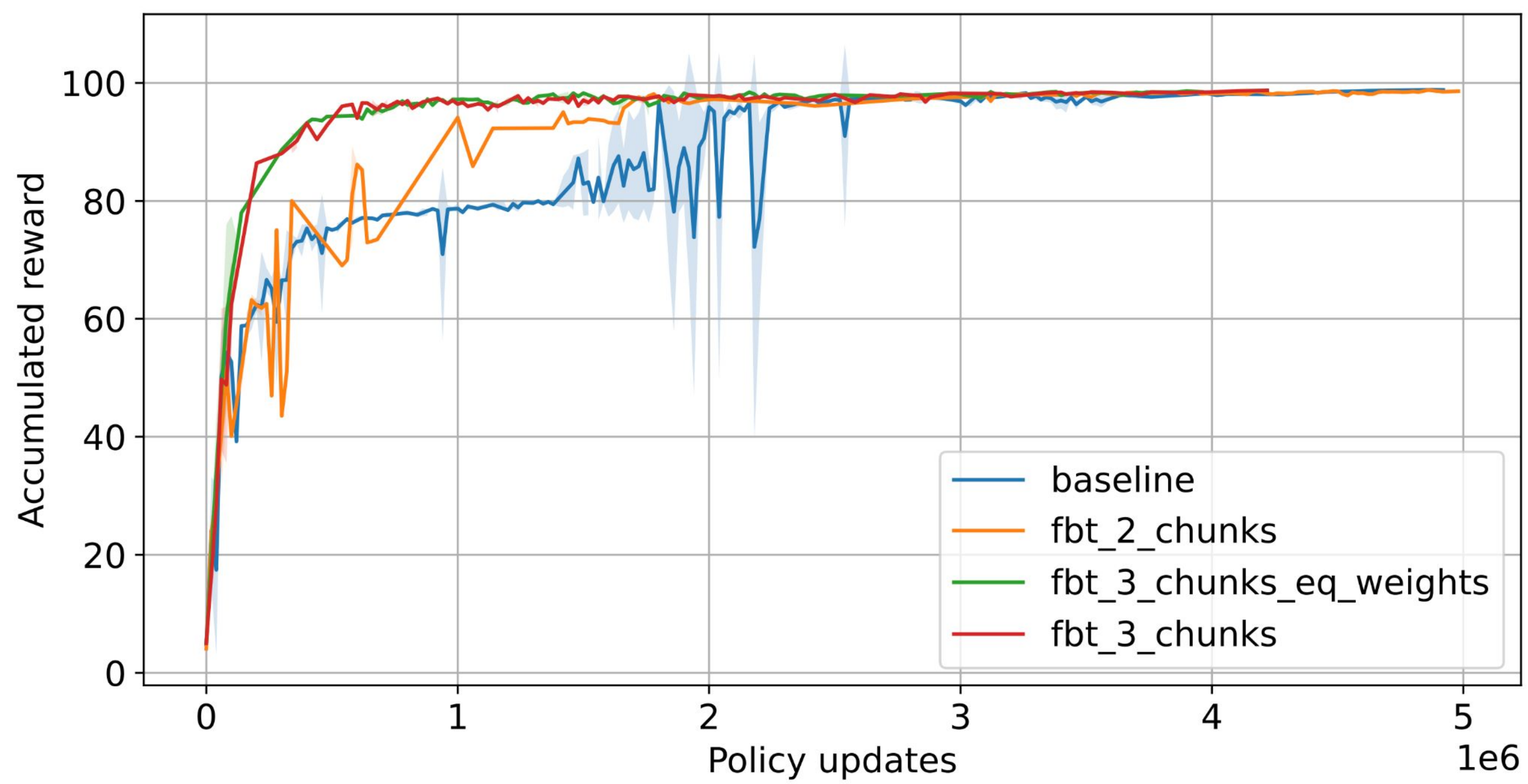
EPFL

DeepMind

Episode Chunking

Chunk (3k) Chunk (3k) Chunk (3k)

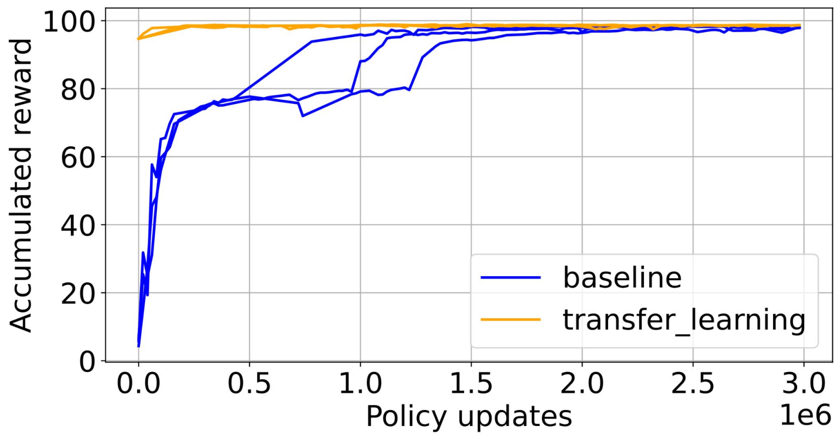
Episode (10k steps)



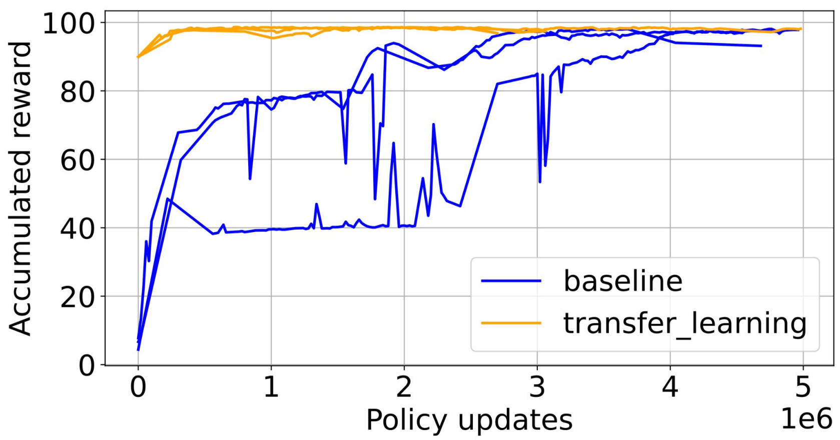
Transfer Learning

Shift Ip

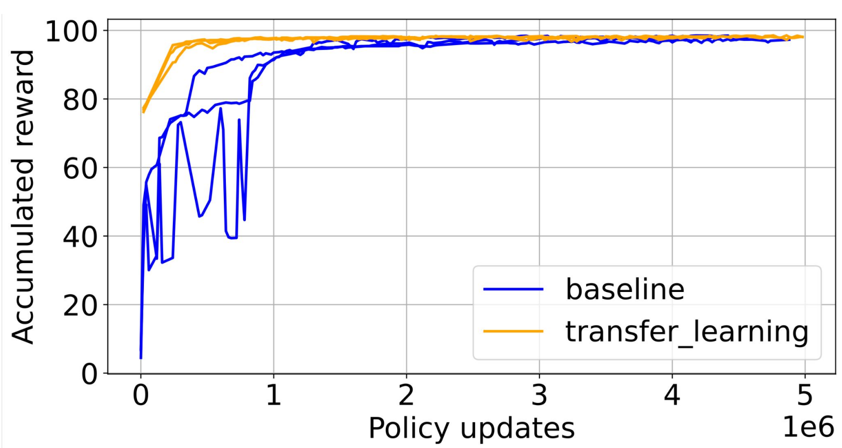
10kA



20kA

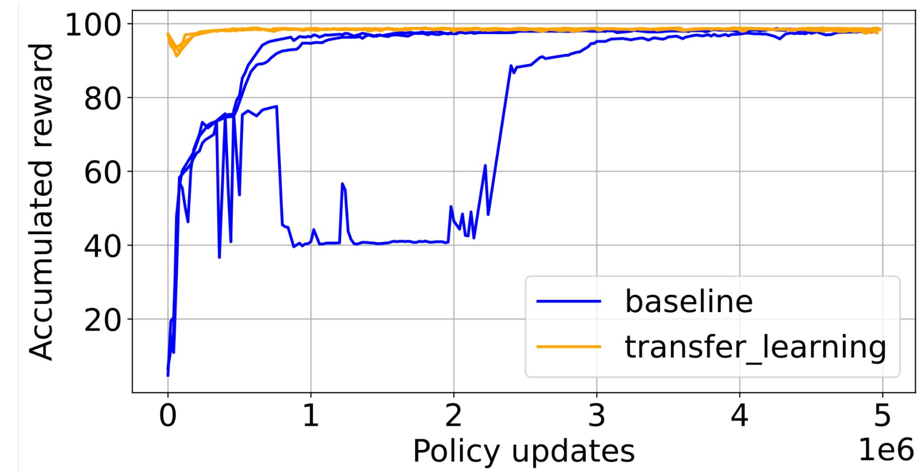


50kA

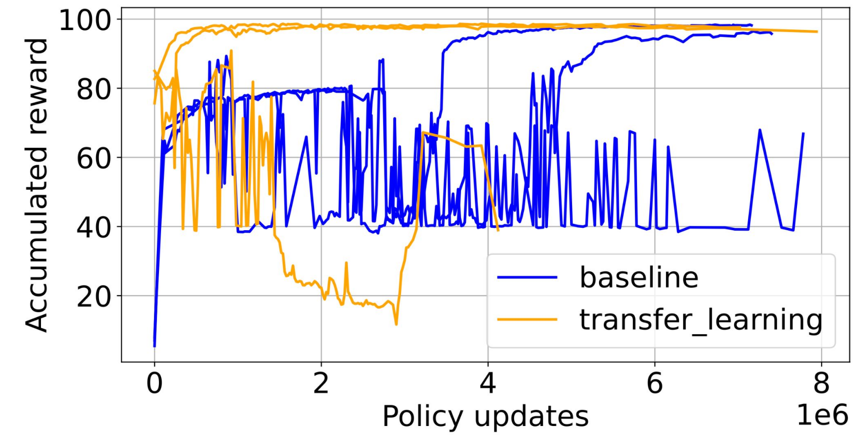


Shift Shape

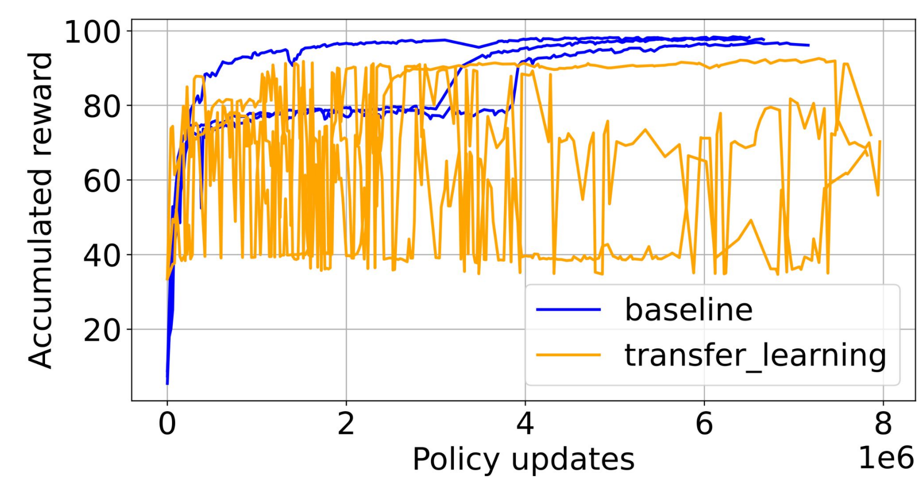
2cm



10cm



20cm



Shape Accuracy

LCFS Error

Experiment	I_p Error (%)	LCFS Mean RMSE (cm)
Baseline	0.353 ± 0.221	0.567 ± 0.221
Narrow Reward	**0.238** ± 0.076	**0.201** ± 0.057
Reward Schedule	0.450 ± 0.321	0.490 ± 0.196

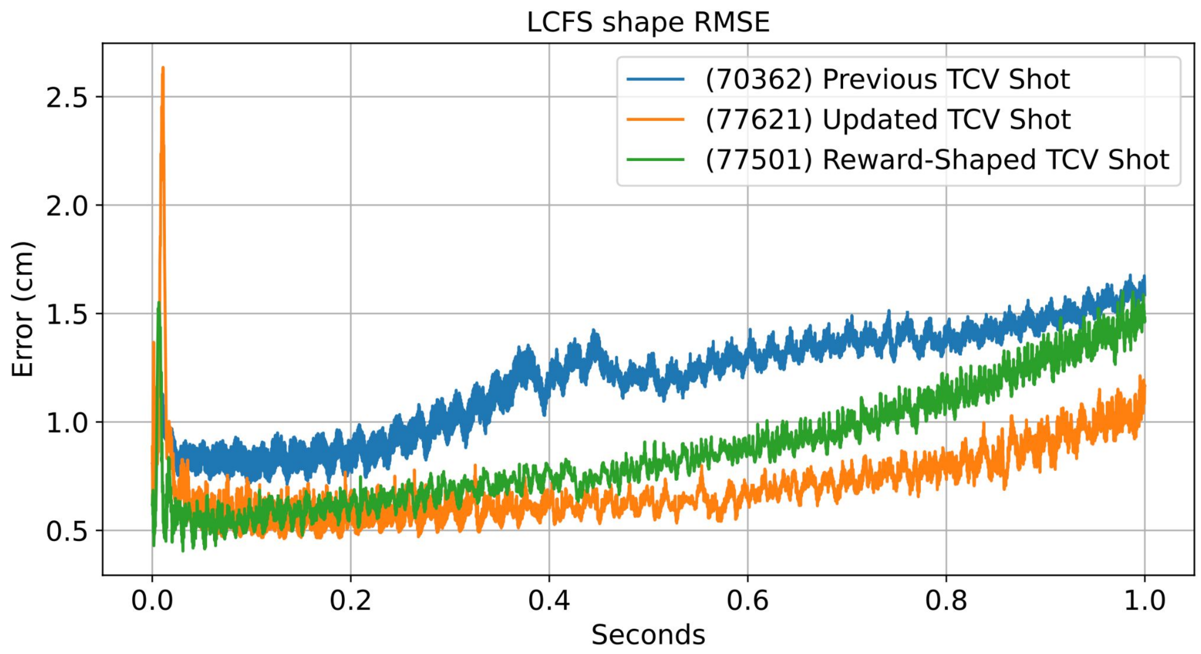
X-point accuracy

Experiment	I_p Error (%)	LCFS Mean RMSE (cm)	X-point Location Error (cm)
Baseline	0.848 ± 1.710	1.122 ± 1.460	0.669 ± 0.491
X-Point Fine Tuned	0.717 ± 0.624	0.845 ± 0.097	**0.289** ± 0.027
Narrow X-Point Reward	6.143 ± 4.602	4.536 ± 3.268	1.199 ± 1.102
Additional Training	**0.502** ± 0.423	**0.723** ± 0.159	0.541 ± 0.112

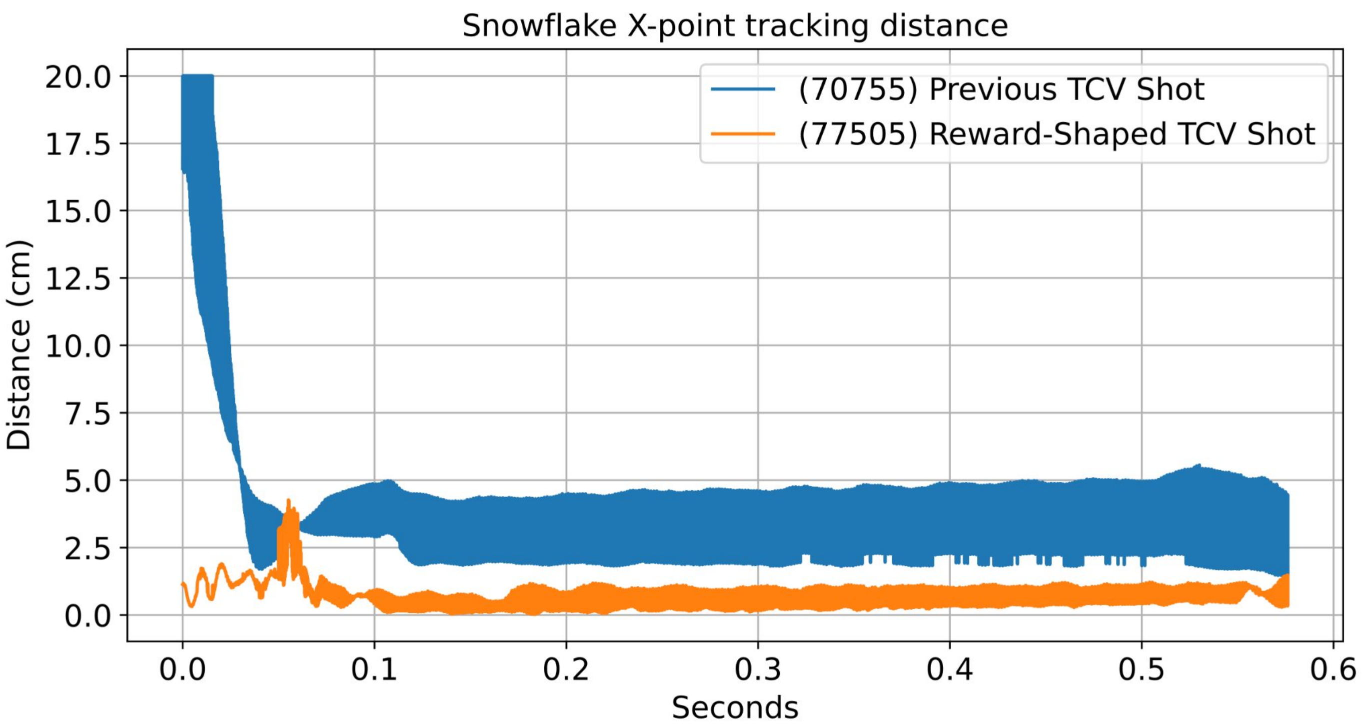
EPFL DeepMind

Hardware Shape Accuracy

LCFS Error



X-point accuracy



Summary

- Large Opportunities available in speeding up design
  - If can avoid the exploration problem – Do!
  - Can jumpstart training from pre-existing data

- Need to be careful with treating results in simulation as truth
  - Can significantly increase accuracy in sim
  - Need to push again on improving sim2real transfer

Contrasting Classic Control and RL

Traditional controllers (MIMO PID)	Reinforcement Learning Implementation
Separate error for each control loop	Single reward function
Error computed online	No explicit error signals or estimation
Separate complex state estimators and tuning of multiple control loops	Joint (and potentially generalising) solution to entire stabilization/control problem
Domain knowledge required for problem definition and separate controller design	Domain knowledge is in simulator. Just define reward functions
Tuning of several control parameters	Reward function engineering
(Usually) Clear relation between parameters and aspects of control performance	Black-box agent
Integral control nominally gives zero steady-state error on desired quantities	No certainty of zero steady-state errors in case of external disturbances

EPFL

DeepMind

Outlook

- Generalist agents - No need to retrain for new references

- Expand simulator capabilities to broaden the horizon of possibilities

- Co-design: simultaneously optimize tokamak design (plasma shape, sensors, coil, vessel placement) together with controller

Conclusions

- **Demonstrated RL for closed-loop magnetic control of tokamak plasmas, trained in simulation and <u>tested on a real device</u>**
  - **Implementing 10kHz controller with 100+ measurements, 20 actions is a milestone for RL on real-world systems in terms of complexity**
  - **<u>Models</u> are sufficiently accurate to perform the required simulations**
- **Bright future for more applications of RL**
  - **For accelerating fusion science: improving plasma performance & design new devices**
  - **For application to more complex real-world systems, in particular where good models exist**

SWISS
PLASMA
CENTER

Successful multidisciplinary collaboration!

Tight integration between teams to understand and control this challenging system

EPFL DeepMind