Mechanical and Aerospace Engineering

Princeton Plasma Control
control.princeton.edu

SEOUL NATIONAL UNIVERSITY

Tokamak operation design and control with deep reinforcement learning

Presented by Jaemin Seo

js7146@princeton.edu

Collaborators

Princeton University

A. Jalalvand, R. Conlin, A. Rothstein, J. Abbate, and E. Kolemen

Seoul National University

B. Kim, C.Y. Lee, M.S. Park, S.J. Park, Y.H. Lee, and Y.-S. Na

KSTAR

KFE
KOREA INSTITUTE OF FUSION ENERGY

DIII-D NATIONAL FUSION FACILITY

GENERAL ATOMICS

2022. 10. 14

26th Workshop on MHD Stability Control

Contents

Motivation: RL in fusion research

- **MHD control with AI (Google DeepMind & EPFL & TCV team, *Nature*, 2022)**

Article

Magnetic control of tokamak plasmas through deep reinforcement learning

https://doi.org/10.1038/s41586-021-04301-9
Received: 14 July 2021
Accepted: 1 December 2021
Published online: 16 February 2022
Open access
Check for updates

Jonas Degrave[1,2], Federico Felici[2,3✉], Jonas Buchli[1,3✉], Michael Neunert[1,2], Brendan Tracey[1,3✉], Francesco Carpanese[1,2,3], Timo Ewalds[1,2], Roland Hafner[1,2], Abbas Abdolmaleki[1], Diego de las Casas[1], Craig Donner[1], Leslie Fritz[1], Cristian Galperti[2], Andrea Huber[1], James Keeling[1], Maria Tsimpoukelli[1], Jackie Kay[1], Antoine Merle[2], Jean-Marc Moret[2], Seb Noury[1], Federico Pesamosca[2], David Pfau[1], Olivier Sauter[2], Cristian Sommariva[2], Stefano Coda[2], Basil Duval[2], Ambrogio Fasoli[2], Pushmeet Kohli[1], Koray Kavukcuoglu[1], Demis Hassabis[1] & Martin Riedmiller[1,2]

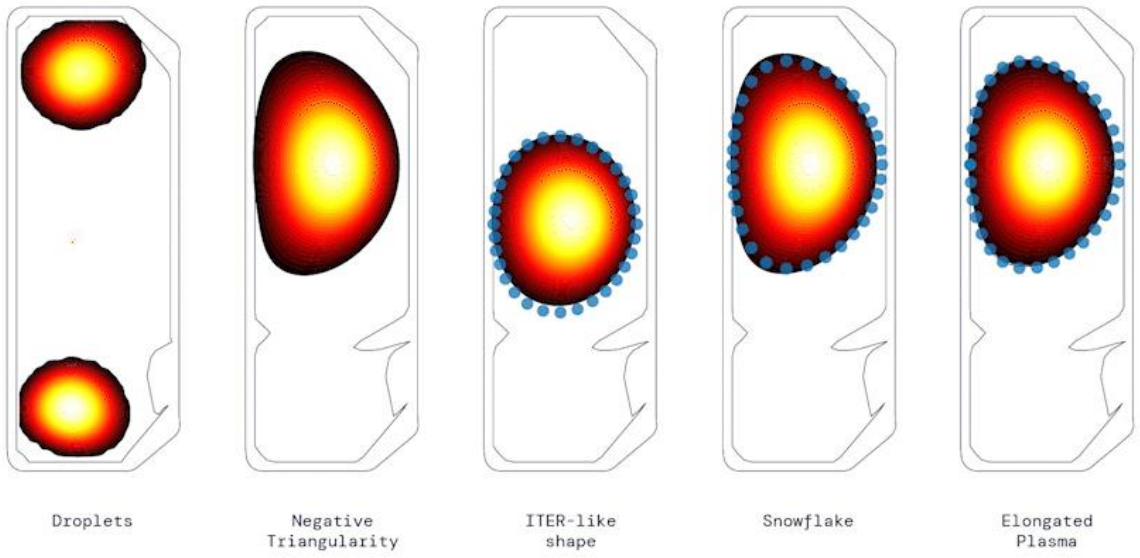
Nuclear fusion using magnetic confinement, in particular in the tokamak configuration, is a promising path towards sustainable energy. A core challenge is to shape and maintain a high-temperature plasma within the tokamak vessel. This requires high-dimensional, high-frequency, closed-loop control using magnetic actuator coils, further complicated by the diverse requirements across a wide range of plasma configurations. In this work, we introduce a previously undescribed architecture for tokamak magnetic controller design that autonomously learns to command the full set of control coils. This architecture meets control objectives specified at a high level, at the same time satisfying physical and operational constraints. This approach has unprecedented flexibility and generality in problem specification and yields a notable reduction in design effort to produce new plasma configurations. We successfully produce and control a diverse set of plasma configurations on the Tokamak à Configuration Variable[1,2], including elongated, conventional shapes, as well as advanced configurations, such as negative triangularity and 'snowflake' configurations. Our approach achieves accurate tracking of the location, current and shape for these configurations. We also demonstrate sustained 'droplets' on TCV, in which two separate plasmas are maintained simultaneously within the vessel. This represents a notable advance for tokamak feedback control, showing the potential of reinforcement learning to accelerate research in the fusion domain, and is one of the most challenging real-world systems to which reinforcement learning has been applied.

Tokamaks are torus-shaped devices for nuclear fusion research and are a leading candidate for the generation of sustainable electric power. A main direction of research is to study the effects of shaping the distribution of the plasma into different configurations[3–5] to optimize the stability, confinement and energy exhaust, and, in particular, to inform the first burning-plasma experiment, ITER. Confining each configuration within the tokamak requires designing a feedback controller that can manipulate the magnetic field[6] through precise control of several coils that are magnetically coupled to the plasma to achieve the desired plasma current, position and shape, a problem known as the tokamak magnetic control problem.

The conventional approach to this time-varying, non-linear, multivariate control problem is to first solve an inverse problem to precompute a set of feedforward coil currents and voltages[7,8]. Then, a set of independent, single-input single-output PID controllers is designed to stabilize the plasma vertical position and control the radial position and plasma current, all of which must be designed to not mutually interfere[6]. Most control architectures are further augmented by an outer control loop for the plasma shape, which involves implementing a real-time estimate of the plasma equilibrium[9,10] to modulate the feedforward coil currents[8]. The controllers are designed on the basis of linearized model dynamics, and gain scheduling is required to track time-varying control targets. Although these controllers are usually effective, they require substantial engineering effort, design effort and expertise whenever the target plasma configuration is changed, together with complex, real-time calculations for equilibrium estimation.

A radically new approach to controller design is made possible by using reinforcement learning (RL) to generate non-linear feedback controllers. The RL approach, already used successfully in several challenging applications in other domains[11–13], enables intuitive setting of performance objectives, shifting the focus towards what should be achieved, rather than how. Furthermore, RL greatly simplifies

[1]DeepMind, London, UK. [2]Swiss Plasma Center - EPFL, Lausanne, Switzerland. [3]These authors contributed equally: Jonas Degrave, Federico Felici, Jonas Buchli, Michael Neunert, Brendan Tracey, Francesco Carpanese, Timo Ewalds, Roland Hafner, Martin Riedmiller. ✉e-mail: federico.felici@epfl.ch; buchli@deepmind.com; btracey@deepmind.com



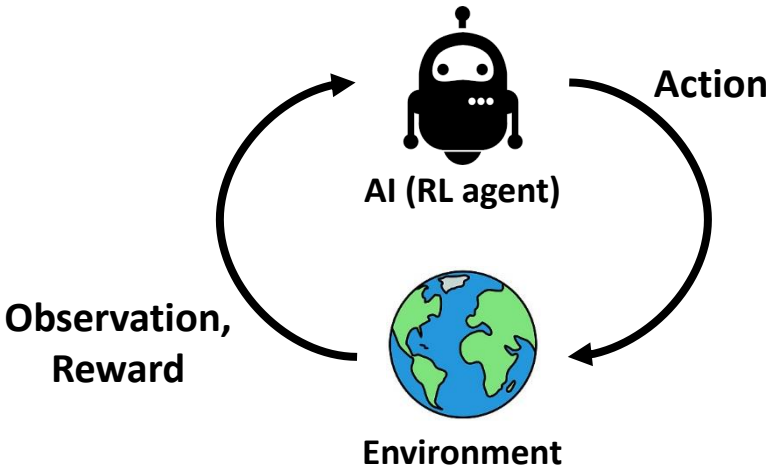
- With RL technique, the plasma current, shape and position were successfully controlled in TCV.
- It demonstrated the huge potential of **RL** × **Fusion** research.

[1] J Degrave et al, Nature 602 (2022) 414

What is RL?

[1] V Mnih et al, Nature 518 (2015) 519

- **Deep reinforcement learning (RL) [1]** ***"Finding the best decision-making policy"***



Time

Action	Bad	Good	Bad	Good	Good
Reward	-1	+1	-1	+1	+1

- The AI takes the action that yields a higher reward.
- It gradually learns the **optimal decision-making policy**.

- **Difficulties of RL application in fusion research**

- RL requires a reliable training environment (simulation).

  But we don't have a <u>perfect</u> all-in-one tokamak simulator including gyro-kinetic, MHD stability, H&CD, …

- RL typically requires $> 10^5$ simulation iterations to train.

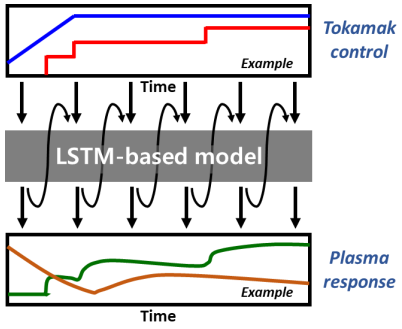
  But reliable theory-based simulation (TGLF, EPED, NUBEAM, …) <u>takes minutes to hours</u> for a single step.

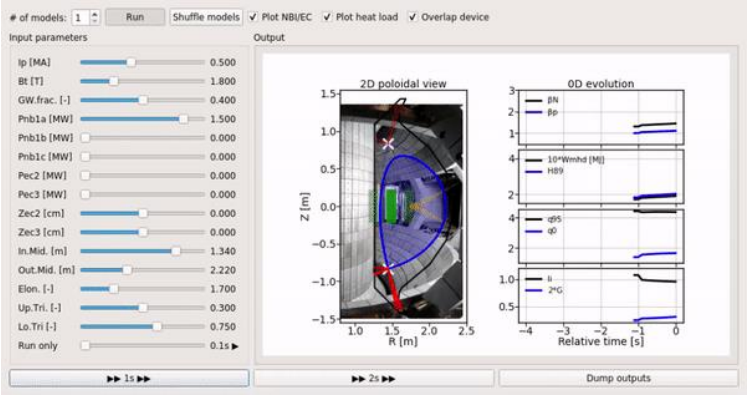
Introduction: What I did

- **Therefore, I've done [1, 2]**

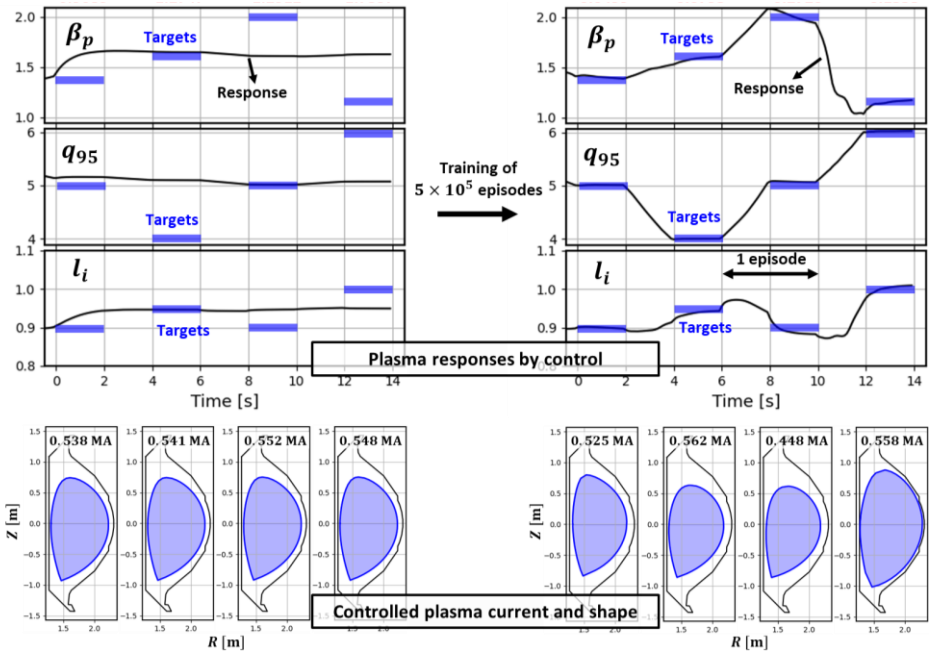
[1] J Seo et al, Nucl. Fusion 61 (2021) 106010
[2] J Seo et al, Nucl. Fusion 62 (2022) 086049

1. **Make a fast surrogate environment**
   - We have 10-year exp data in KSTAR.
   - So, let's build a <u>data-driven</u> simulator!





2. **Then, use RL for tokamak operation design**
   - On the simulator above, we trained an AI that controls the tokamak actuators to achieve given targets.



3. **Validate in KSTAR experiments**

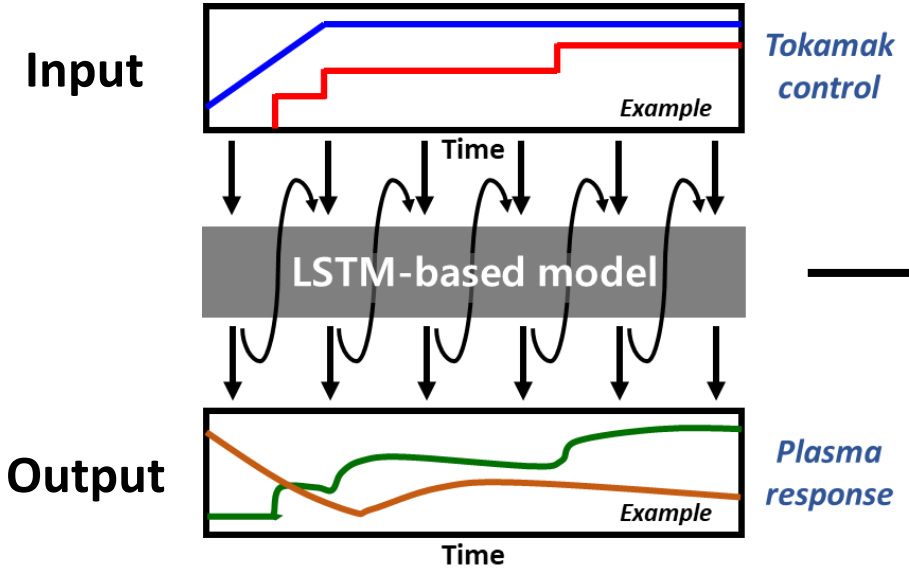
Contents

Surrogate model for the KSTAR tokamak

- **Idea**

  Input *Output*

  - We have experimental data of both tokamak actuators and plasma responses.
  - Then, we can train the **LSTM-based neural network** that mimics the plasma response.



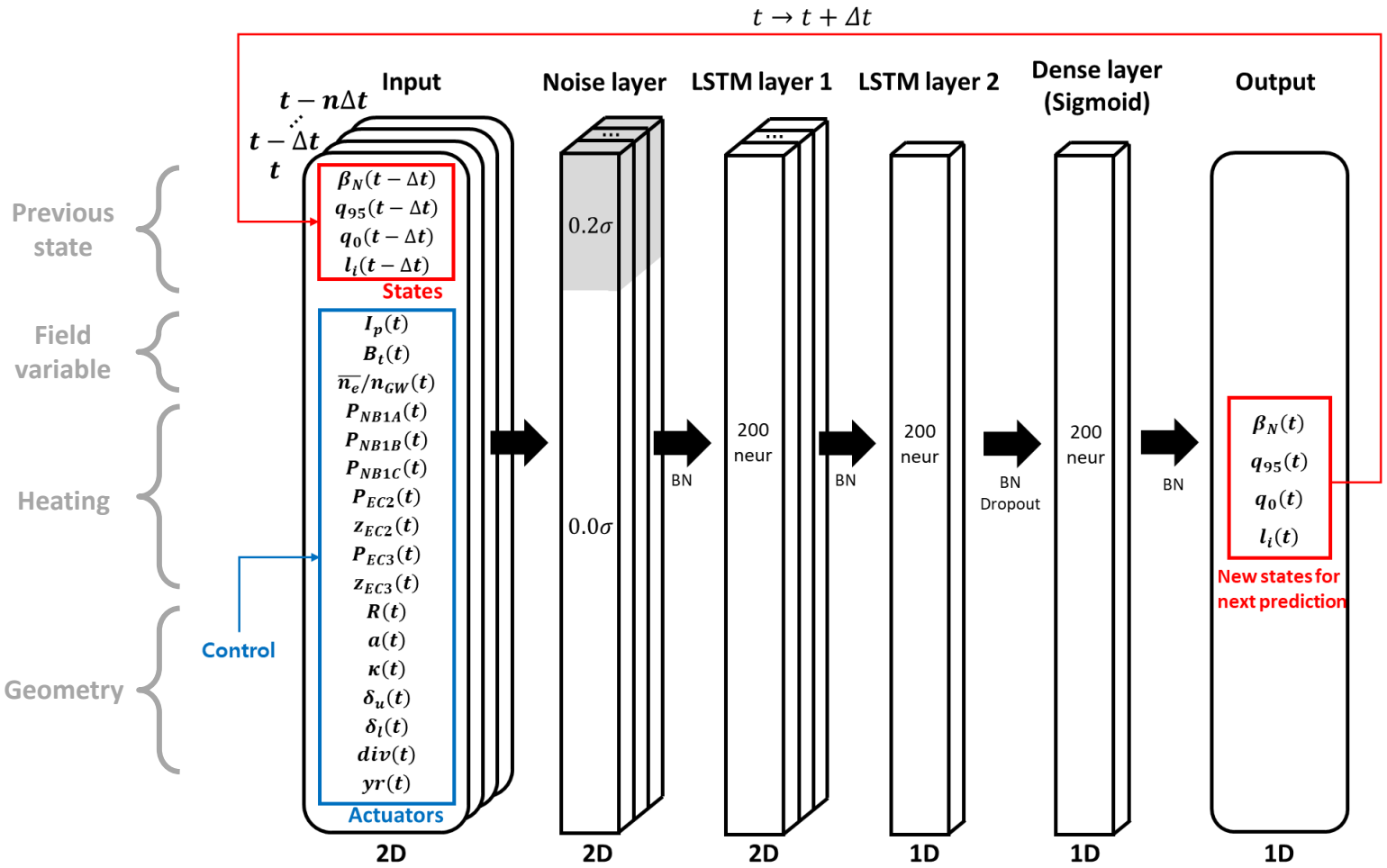
Tokamak control → Plasma current, heating power, Boundary shape, …

→ **What we want to make: Predicting plasma states from actuators**

Plasma response → Plasma beta (β_N, β_p), safety factor (q_{95}, q_0), internal inductance (l_i)…

Surrogate model for the KSTAR tokamak

- **Neural network diagram**



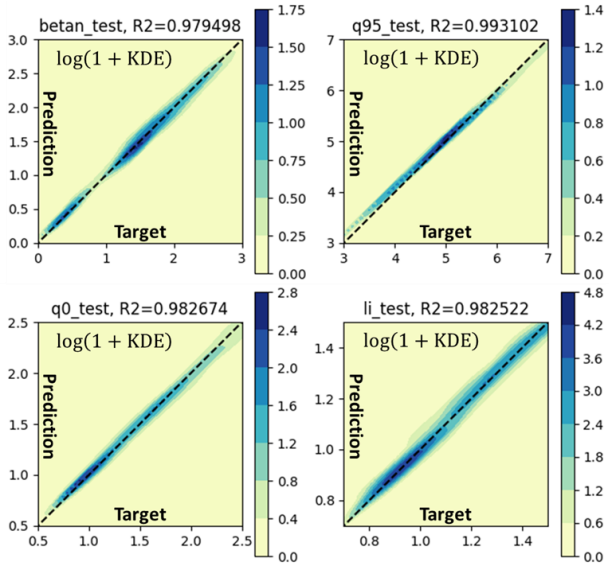
- **Inputs** are the time-series of the previous plasma state, field variables, heating, and geometric information.

- **Outputs** are $\beta_N, q_{95}, q_0, l_i$. Others ($W_{MHD}, \beta_p$) can be also estimated from these.

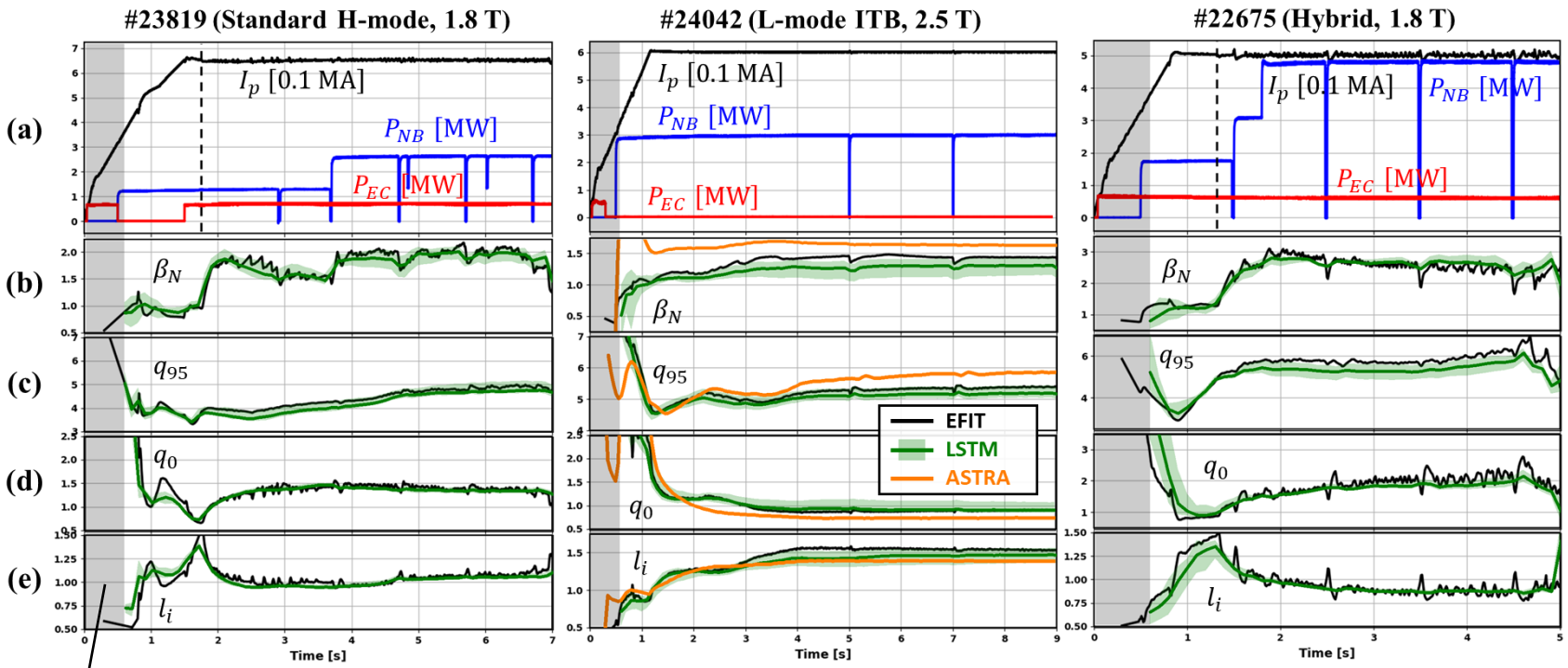
- To minimize overfitting and undesired prediction, noise layer, dropout, and ensemble averaging were applied.

Surrogate model for the KSTAR tokamak

- **Predictive simulation in KSTAR with the trained model**



Regression plot for testset

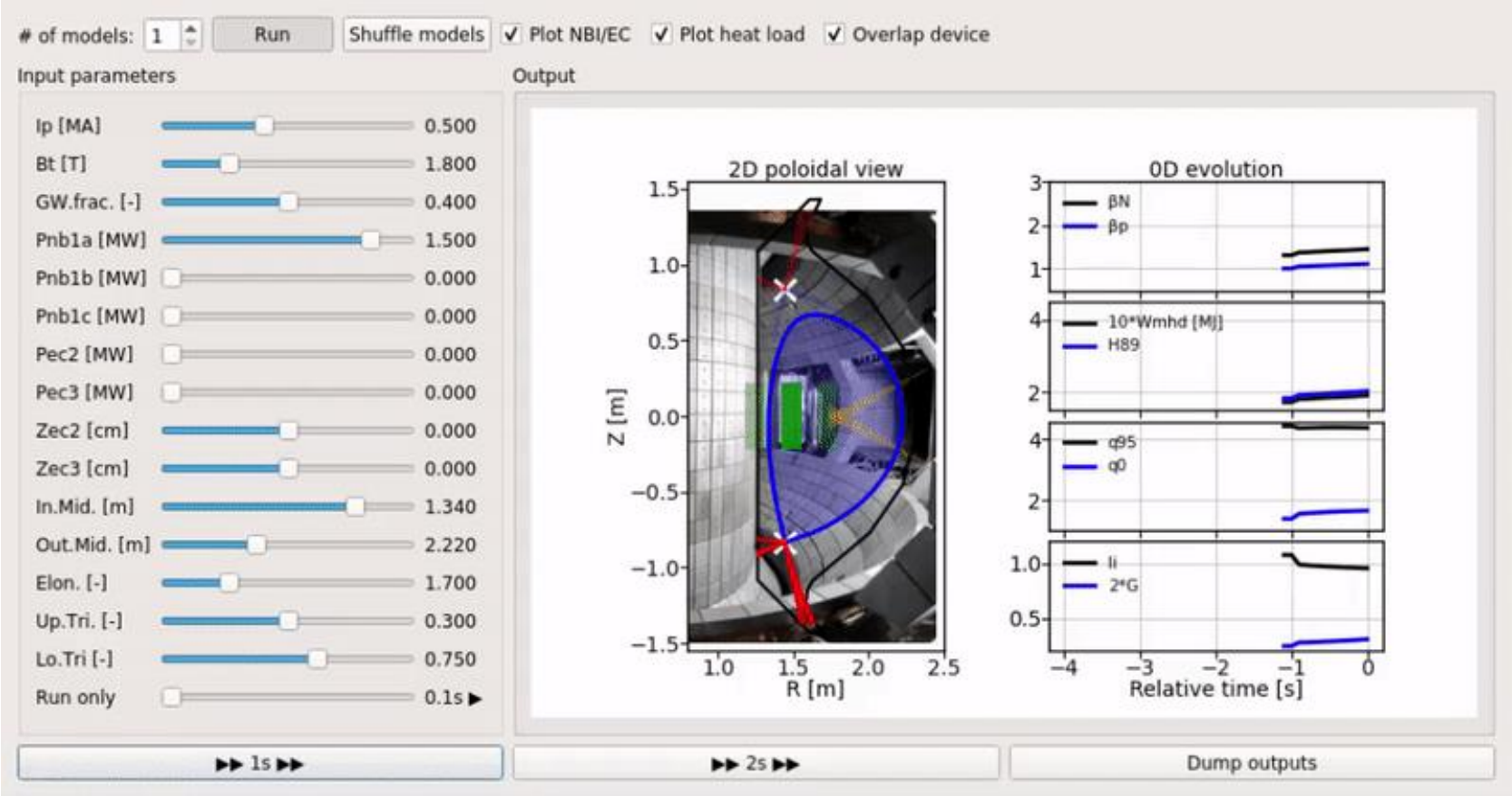


- If only the initial condition (grey box) is given, the entire discharge can be predicted from the actuator scenario.

- The model successfully predicts the plasma evolution in different scenarios.

Surrogate model for the KSTAR tokamak

- **Interactive GUI for predictive modeling**



- We can perform a virtual interactive experiment in real-time.

- This tool can be used for the **environment of the RL training.**

https://github.com/jaem-seo/KSTAR_tokamak_simulator

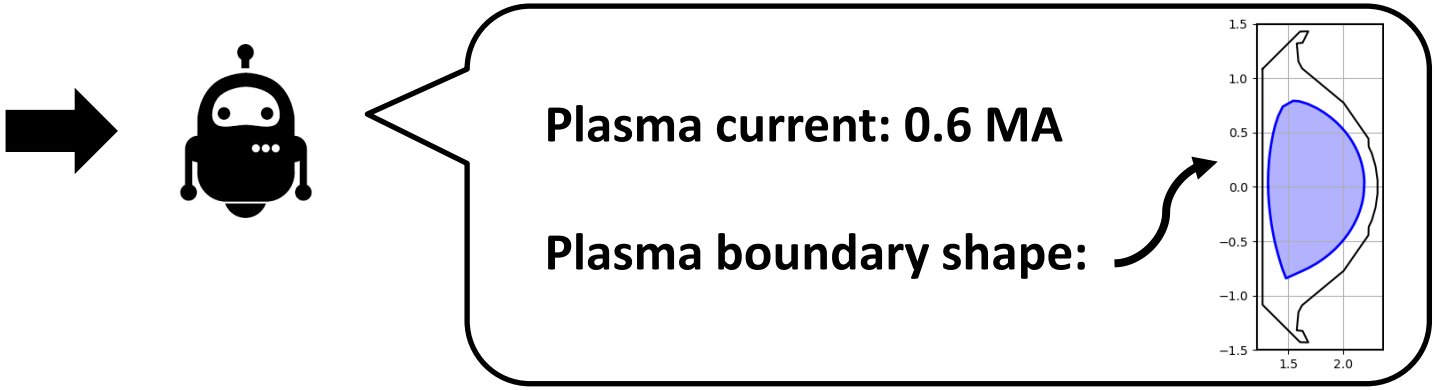
Tokamak operation design with RL

- **Idea**
  - Now, we have a training environment for RL.
  - Then, we want an AI that designs and suggests a possible operation trajectory to achieve a given target plasma state.

Goal

Ex) How do we operate a tokamak to achieve $\beta_p = 2, q_{95} = 5, l_i = 1$ at a given available heating condition ($P_{NB} = 5$ MW)?

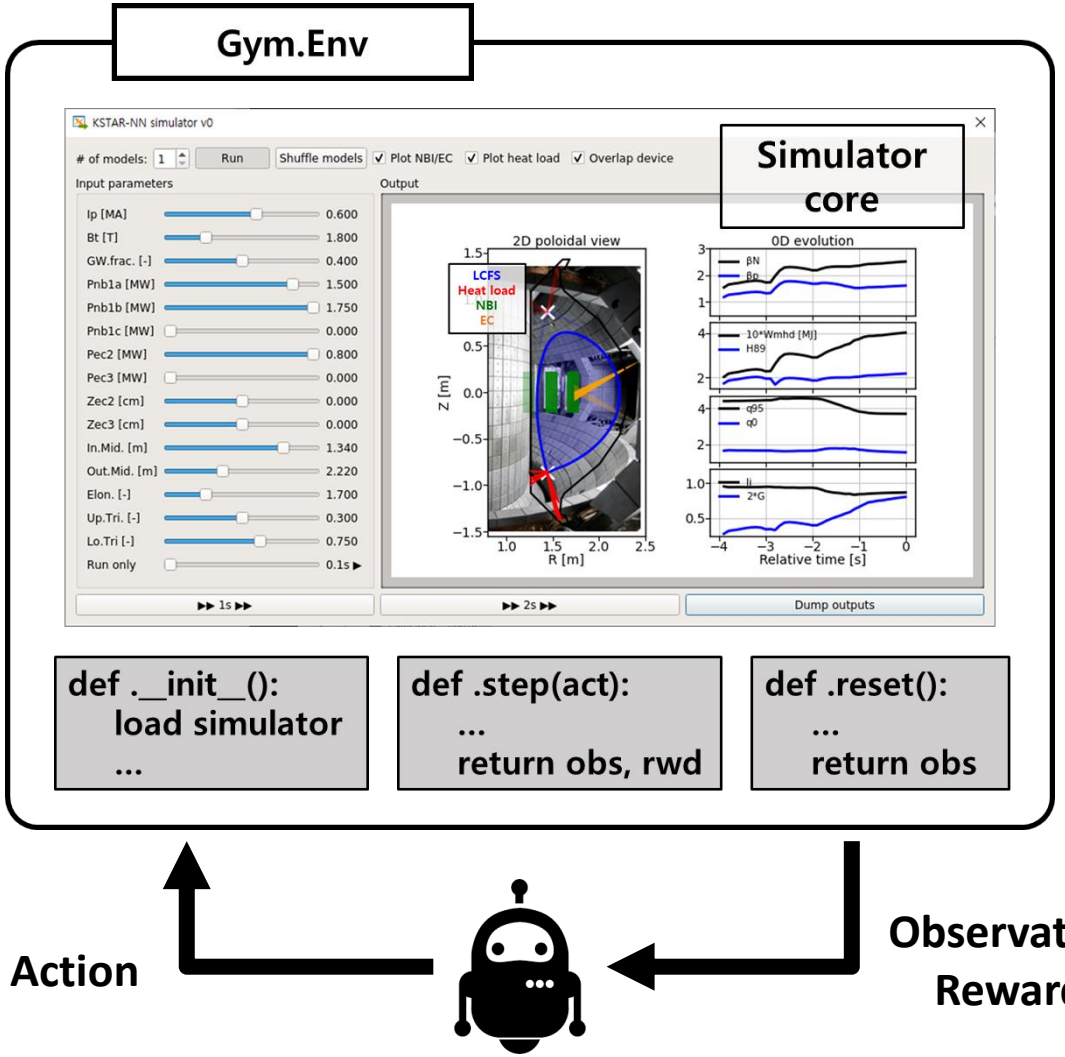
Constraints



Tokamak operation design with RL

[1] G Brockman et al, ArXiv preprint, ArXiv1606.01540 (2016)
[2] S Fujimoto et al, 35th ICML (2018)
[3] A Hill et al, https://github.com/hill-a/stable-baselines (2018)

- **RL training for operation design**



- For the AI to interact with the simulator, we need to wrap it into a standard format, *OpenAI Gym* environment [1].

- The AI was trained by *TD3* [2] implementation from *Stable Baselines* [3].

Action: $\{I_p, \kappa, \delta_u, \delta_l, R_{in}, R_{out}\}$

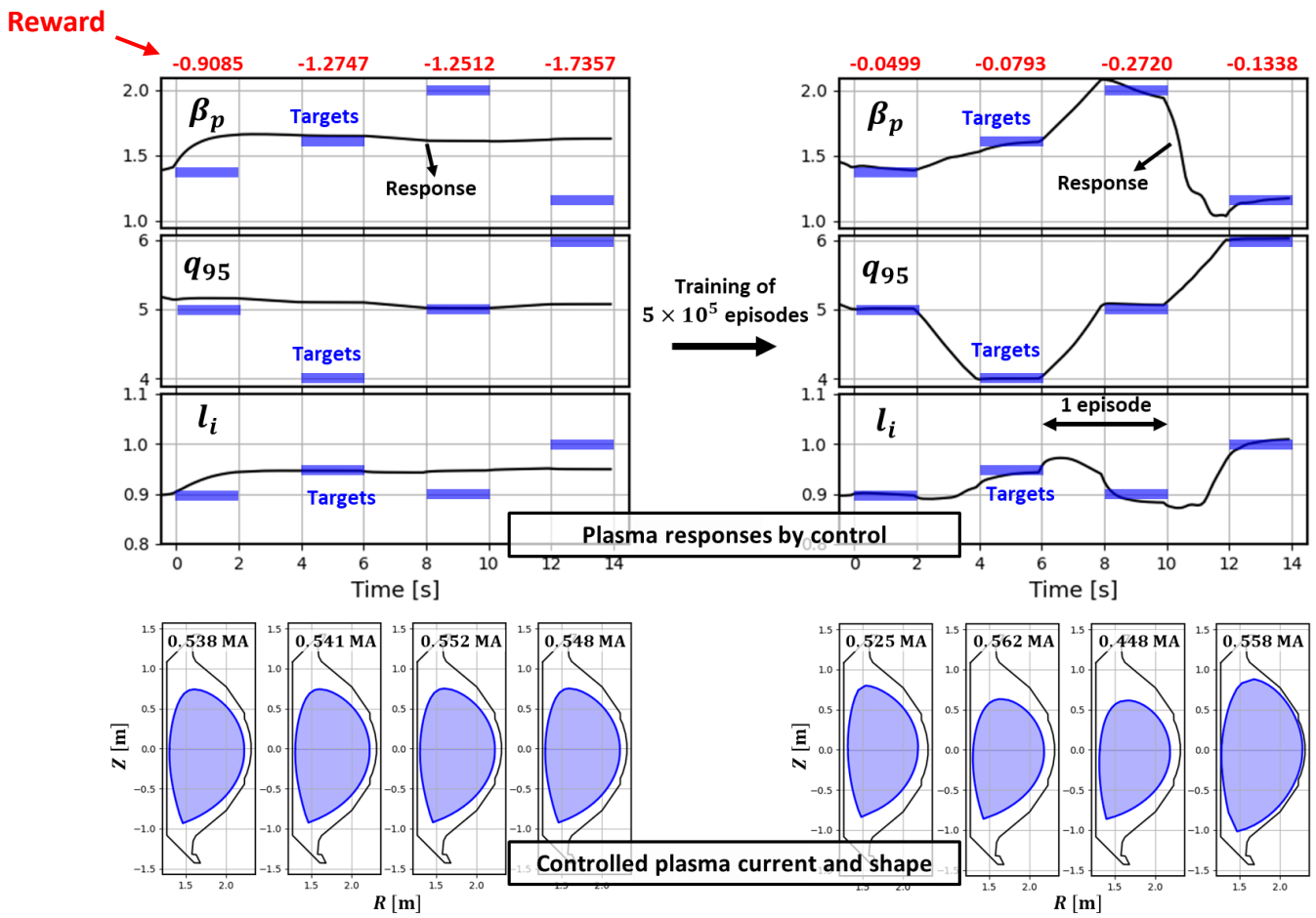
Observation: Previous action, $\{\beta_p, q_{95}, l_i\}_{old}$, $\{\beta_p, q_{95}, l_i\}_{target}$, $\{P_{NB}'s\}$

Reward: $-\mathrm{RMS}\left(\frac{y-y_{target}}{\epsilon_y}\right)_{y=\beta_p,q_{95},l_i}$

The closer to the target, the higher

Tokamak operation design with RL

- **Validation in the simulation**



- After enough training, the AI determines reasonable solution of I_p and the boundary shape to reach the target of multiple parameters.

https://github.com/jaem-seo/AI_tokamak_control

Tokamak operation design with RL

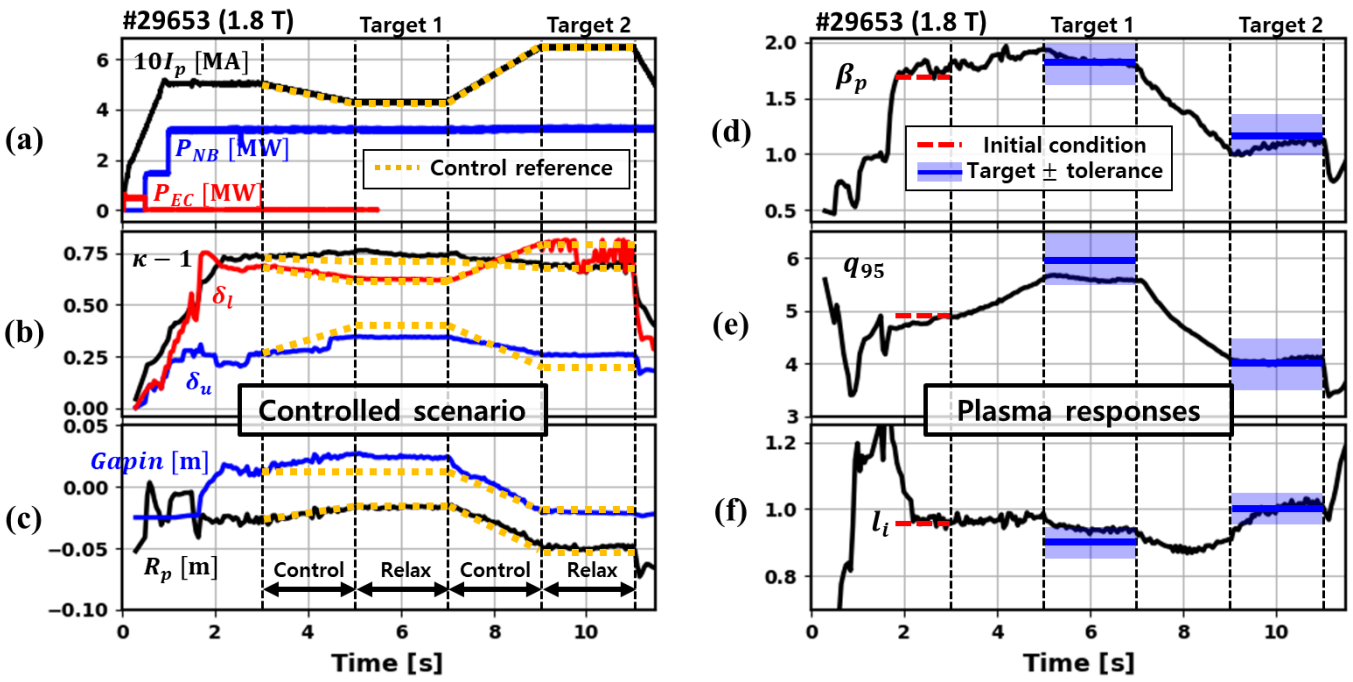
- **Validation in the KSTAR experiment**

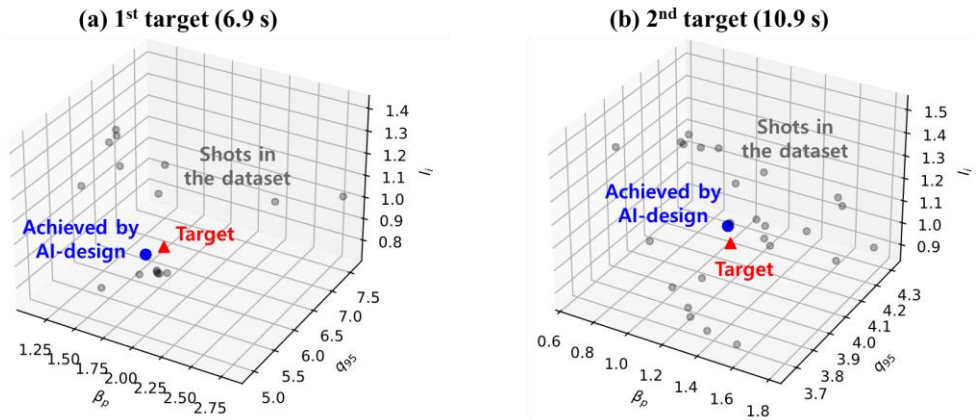
Target setting

(β_p, q_{95}, l_i)

$= (1.8, 6.0, 0.9)$

$\& \ (1.2, 4.0, 1.0)$





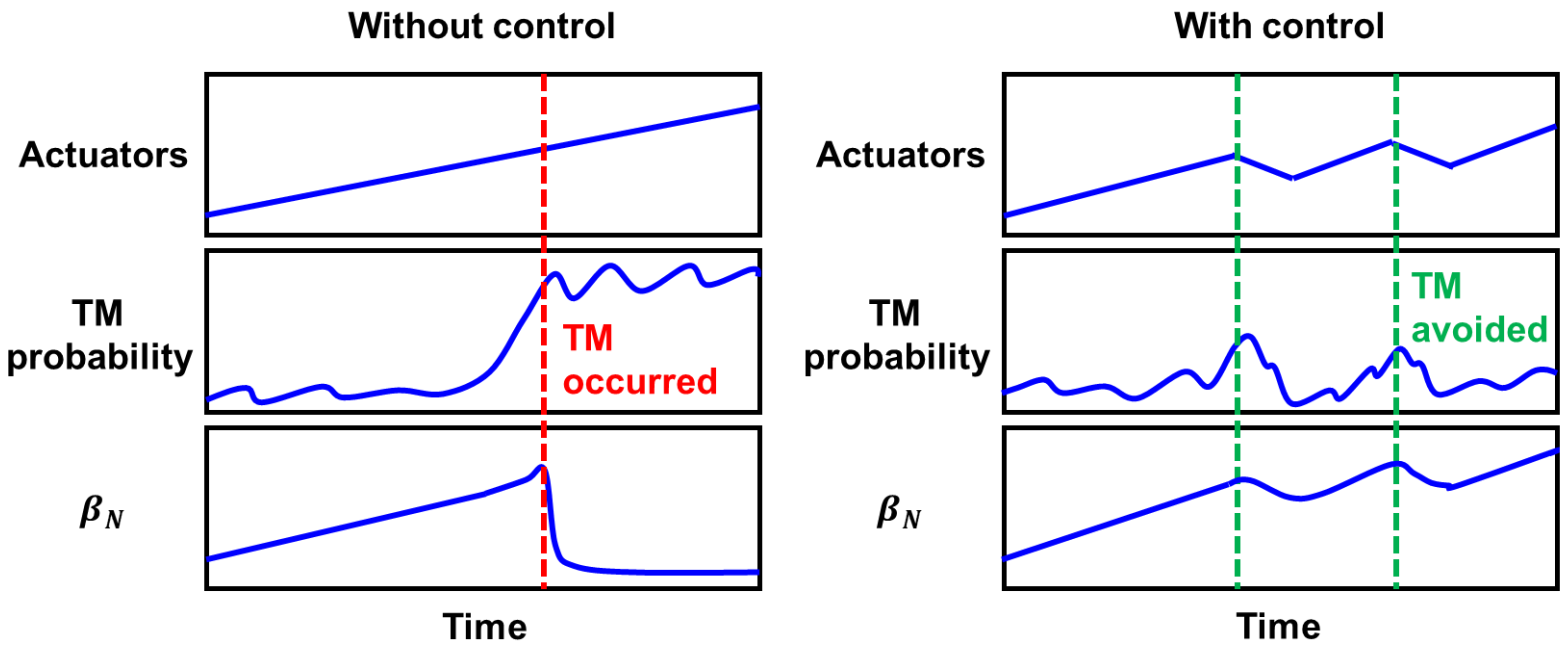
- The plasma response followed the preset targets.
- However, actual <u>shape controls were not perfect</u>.

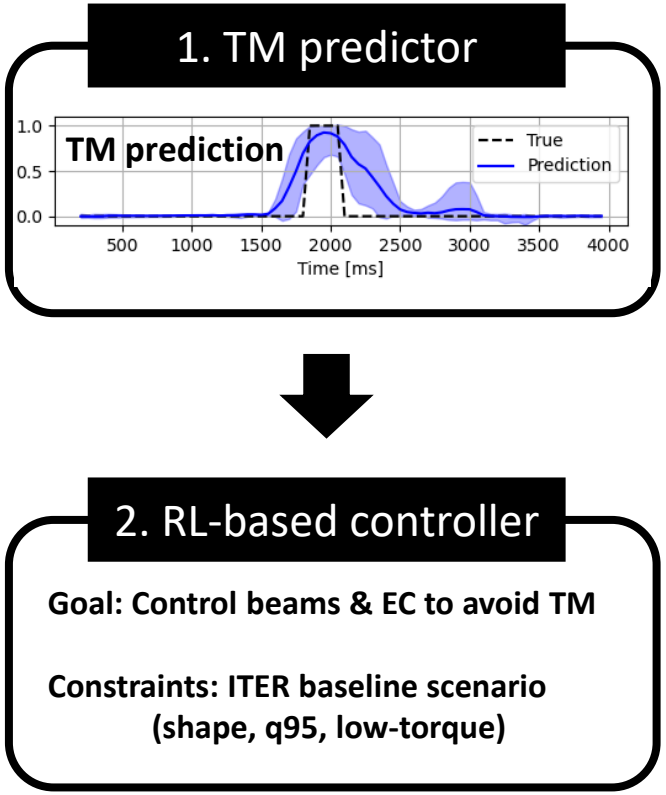
→ **Control/diagnostics uncertainty should be reflected later.**

Contents

Tearing mode avoidance in DIII-D

- **Idea**





- For high-performance plasma, we should control plasma near the marginally stable regime.

- By avoiding the tearing mode, we can pursue higher performance.

Real-time version of CAKE (RT-CAKE)

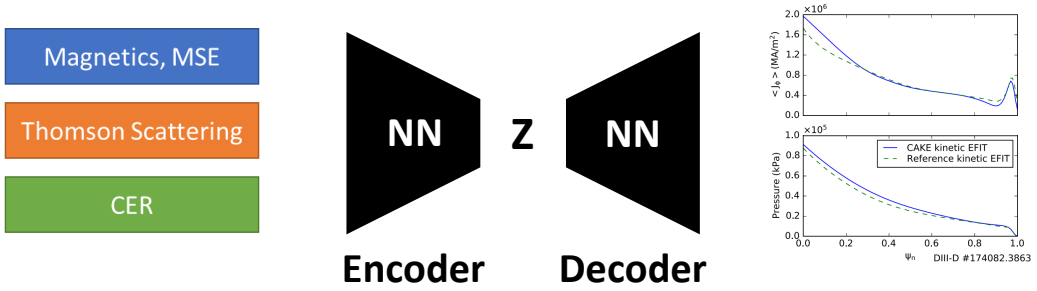
[1] Z A Xing et al, Fus. Eng. Des. 163 (2021) 112163

- **CAKE (Consistent Automatic Kinetic Equilibrium reconstruction) [1]**

Automated, but <u>not real-time-feasible</u> (∵Many iterations, NBI and bootstrap calculation)

- Lots of works (stability, prediction, control) demand real-time kinetic equilibrium.

- **RT-CAKE** can be theoretically possible with <u>Encoder-Decoder</u> type NN.

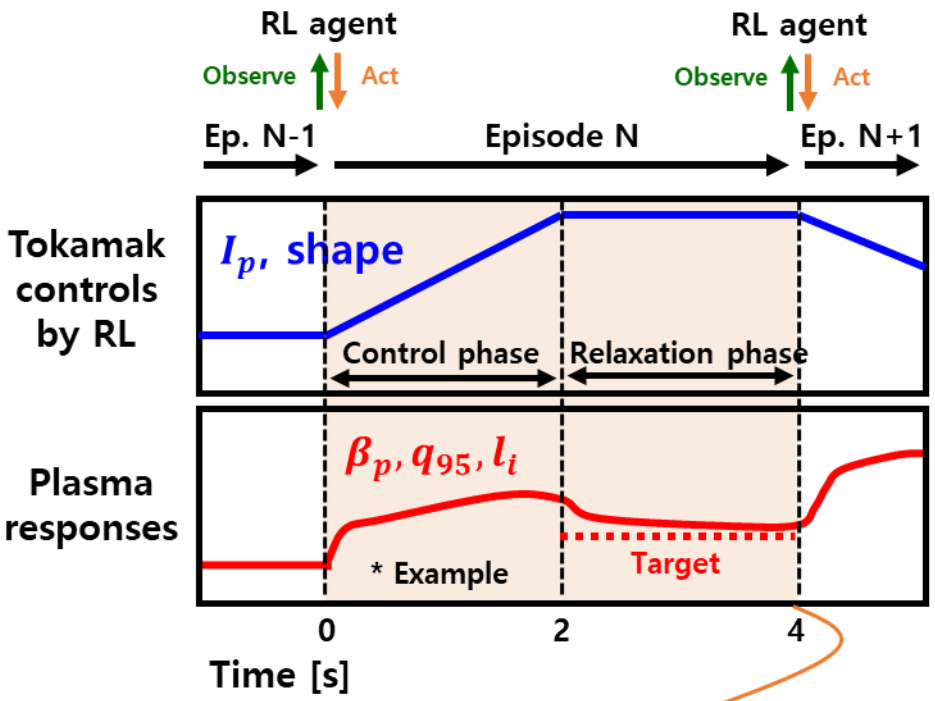


Thank you

Back-ups

Tokamak operation design with RL

- **RL training for operation design (1 episode)**



➡ **Reward,** $r = -\mathrm{rms}\left(\frac{y(t=4s)-y_{target}}{\epsilon_y}\right)_{y=\beta_p,q_{95},l_i}$

$\times\left(5\times10^5\right)$ **episodes**

- When the next target of $\{\beta_p, q_{95}, l_i\}$ is randomly given, the AI determines next $\{I_p$, boundary shape$\}$.
- A single episode consists of **control phase** and **relaxation phase**.
- In the control phase, the control variables are varying along the AI solution, and in the relaxation phase, they maintain for enough saturation of the plasma state.
- At the end of the episode, the reward is estimated.
- The target values and the available NBI powers are also randomly reset for the next episode.

Tokamak operation design with RL

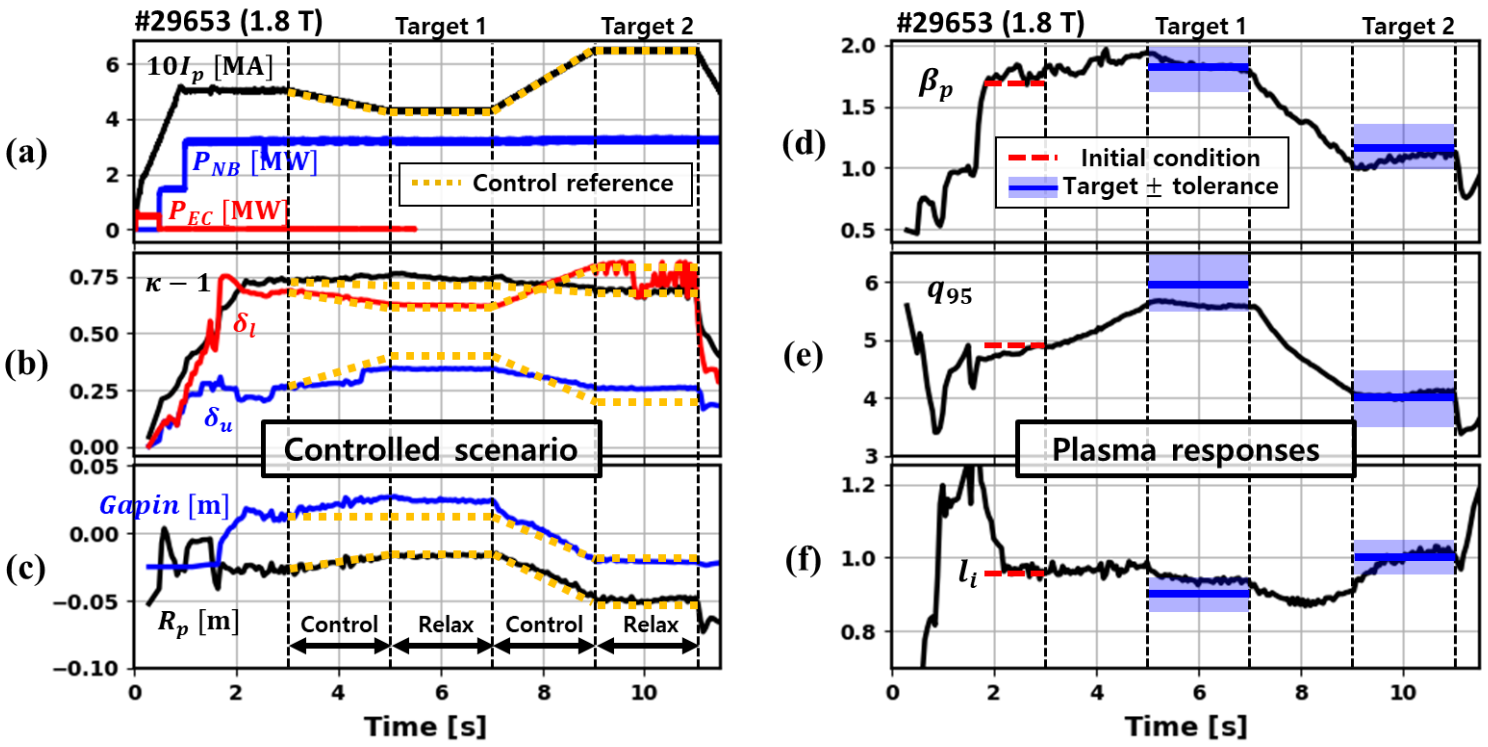
- **Validation in the KSTAR experiment**

Target setting

(β_p, q_{95}, l_i)

$= (1.8, 6.0, 0.9)$

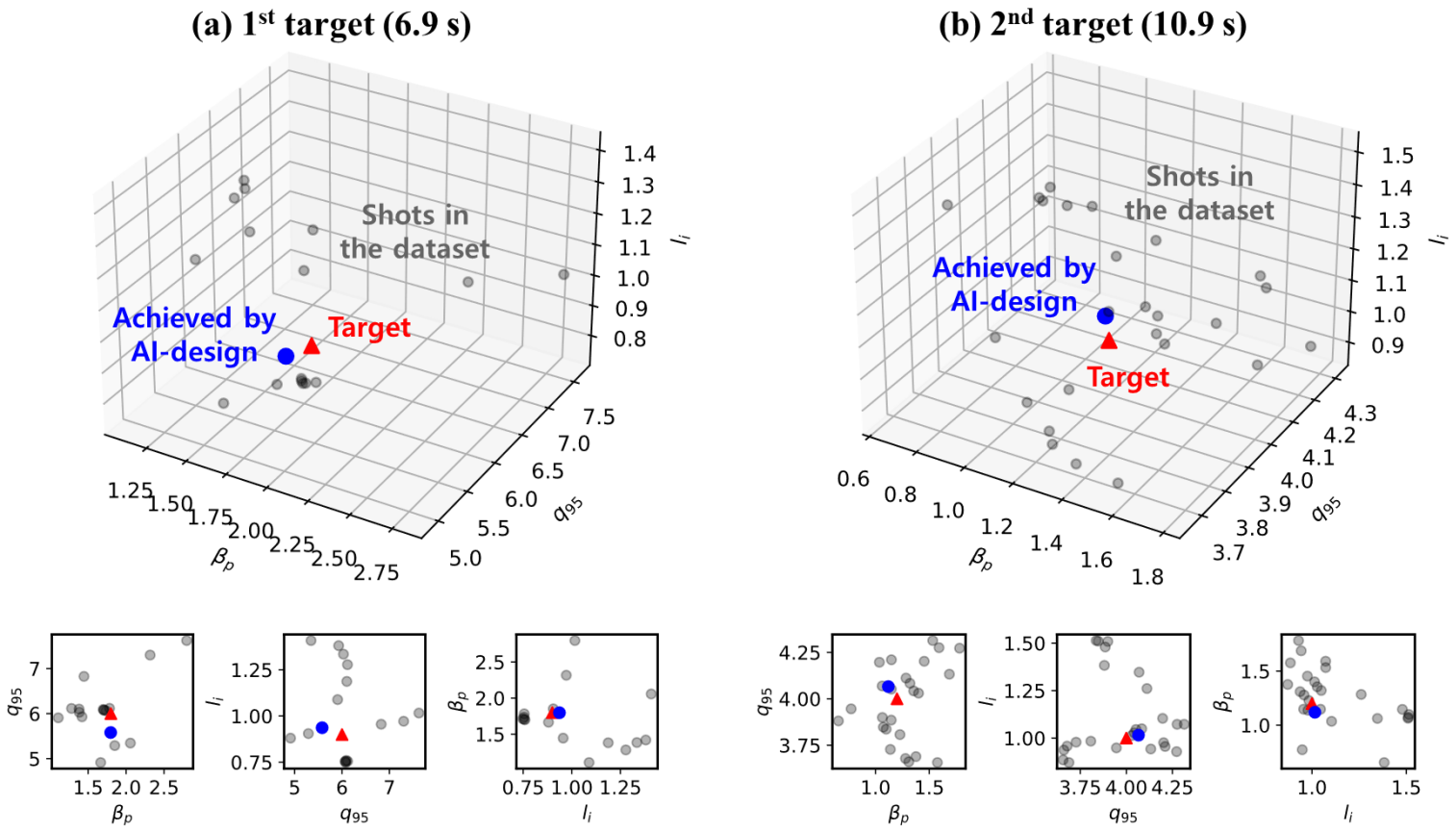
$\& \ (1.2, 4.0, 1.0)$



- After we set the target, the AI determines the operation trajectory. Then we did experiment with that trajectory.
- The plasma response followed the preset targets.
- However, actual <u>shape controls were not perfect</u>. → **Control/diagnostics uncertainty should be reflected later.**

Tokamak operation design with RL

- **Validation in the KSTAR experiment**



Goal & Achieved in parametric space, with other shots under similar condition

- Compared to other shots under similar operating condition (gray), the AI-designed operation yields an achievement closer to the target.

- Even under similar condition, the plasma response can differ according to the previous history.

Similar conditions in dataset:
$B_t \pm 0.05$ T, $I_p \pm 0.05$ MA,
$P_{NB} \pm 0.1$ MW, $\kappa \pm 0.05$, $\delta_{u,l} \pm 0.1$

Surrogate model for the KSTAR tokamak

- **Prediction of plasma states**
  - We want the model to predict the plasma responses by learning the pattern of the experimental data.
  - Plasma response is time-dependent. → RNN or **LSTM**