Tokamak Plasma Operation Control using Multi-Objective Reinforcement Learning in KSTAR

JinSu Kim and Yong-Su Na*

Department of Nuclear Engineering, Seoul National University, Seoul 151-744, Korea
asdwlstn@snu.ac.kr, *ysna@snu.ac.kr

Motivations

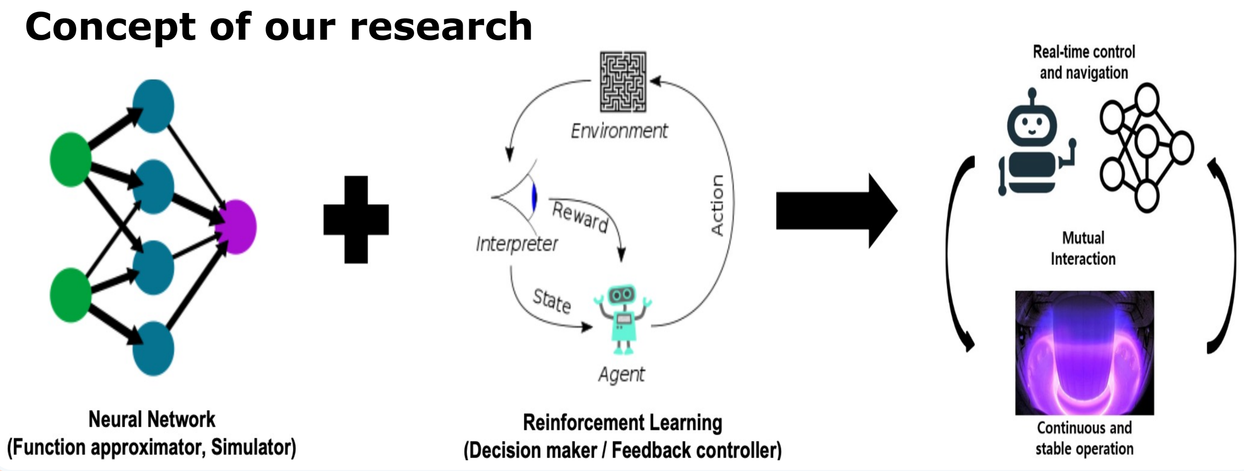
Autonomous Tokamak Plasma Operation Control

- **Achieving a high performance** and **stable** plasma in a steady-state operation is critical challenge for fusion reactors.
- Human trials in experiments alone are not effective in finding optimal conditions, due to **operation limits**[1] and **instabilities**[2,3].

Reinforcement Learning for Multi-targets control

- **Reinforcement Learning** [4,5] offers a promising approach to discover optimal trajectories by data-driven and model-free methods.
- There is a **Pareto-optimal** for target variables in plasma control, indicating that single-objective RL has limits to find optimal policies.

Concept of our research



Objectives

Aims of this research

- **High performance**: to find the optimal way to control $\beta_n \geq 3.0$
- **Shape control:** to control the shape with high performance

Related work

- Seo et al[6] have explored **feedforward beta control** with a **KSTAR simulator based on LSTM**.
- Jonas et al [7] have shown the significant results on **plasma shape control** using MPO algorithm.
- **Multiple 0D parameters control** has been conducted by managing different target variables simultaneously with linear scalarization [8].

Multi-Objective RL for tokamak plasma control

- For multi-target control**, it is necessary to consider the relations** between the controlled variables.
- We used **Generalized Policy Improvement Linear Support (GPI-LS)** [11] to find the set of Pareto-front for **controlling β_n and κ simultaneously** in virtual KSTAR environment.

Dataset setup

1.1. Dataset for training Transformer based simulator

- Input: plasma state $\beta_n, q_{95}, li, \beta_p, \kappa, \delta, R, a$ + controlled variables
  * Controlled variables: P_{EC}+ Z-pos of ECH + P_{NBI}+ I_{PFPC} + I_p
- Output: plasma state $\beta_n, q_{95}, li, \beta_p, \kappa, \delta, R, a$
- Time interval between data points: 50.00ms

1.2. Dataset for training Grad-Shafranov equation solver

- Input: plasma state: $\beta_n, q_{95}, \mathrm{li}, \beta_p, I_p$ + PFPC coil currents configuration
- Output: magnetic flux $\psi \in R^{65\times65}$

1.3. Dataset for training Plasma contour regressor

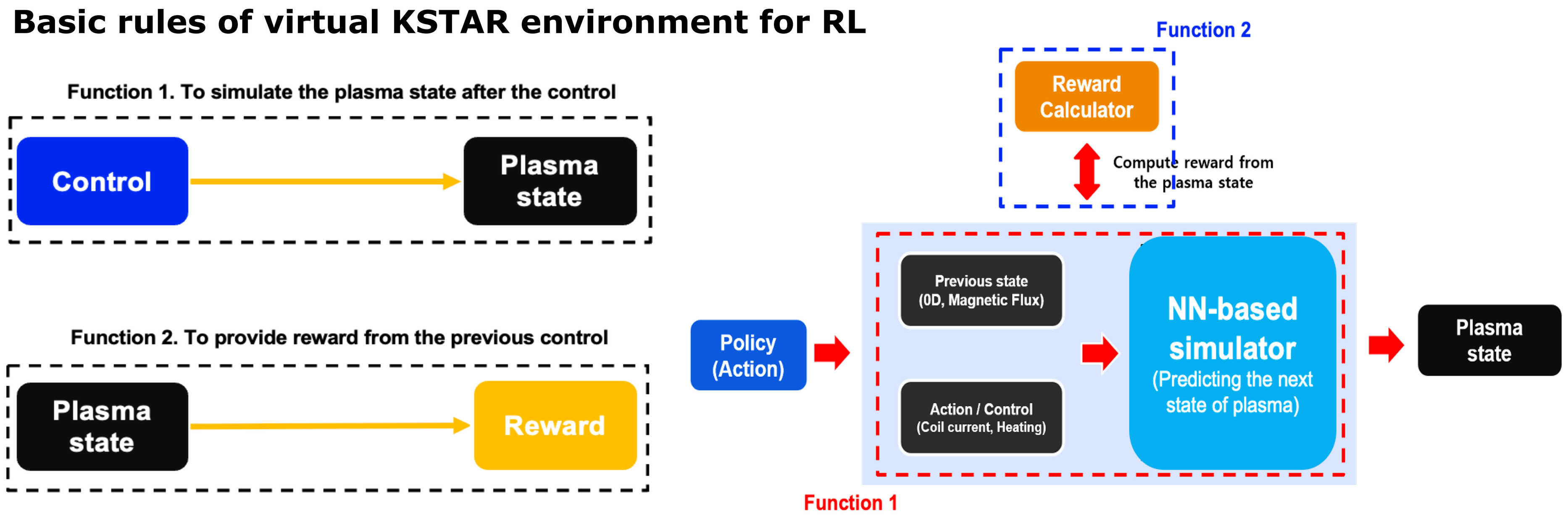
- Input: magnetic flux $\psi \in R^{65\times65}$, plasma state: $\beta_n, q_{95}, \mathrm{li}, \beta_p, I_p$
- Output: magnetic axis (r_{axis}, z_{axis}), 256 contour points $\{(r_1, z_1), \ldots, (r_{256}, z_{256})\}$

1.4. Dataset for training the controller

- Input: plasma state: $\beta_n, q_{95}, \mathrm{li}, \beta_p, I_p$ + controlled variables from past
- Output: P_{EC}+ Z-pos of ECH + P_{NBI}+ I_{PFPC} + I_p for next step

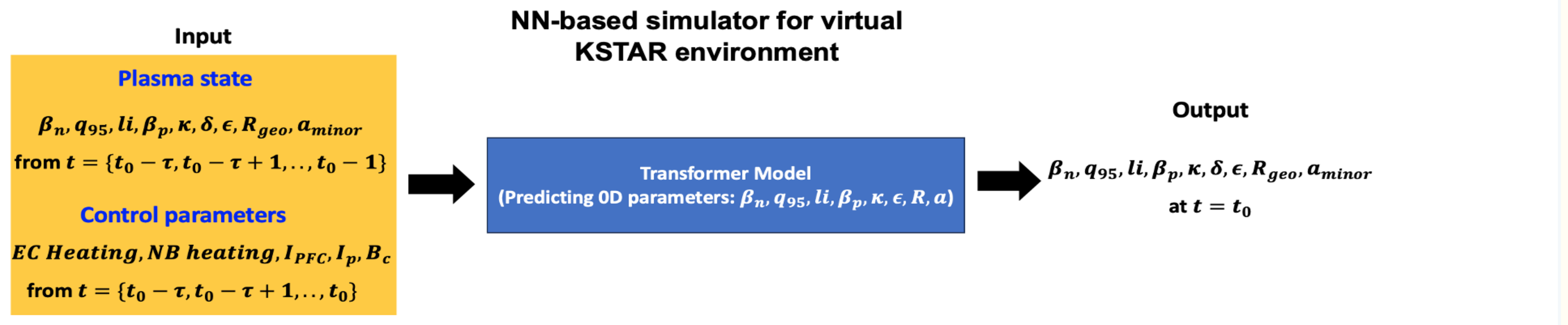
Development of virtual KSTAR environment

Basic rules of virtual KSTAR environment for RL

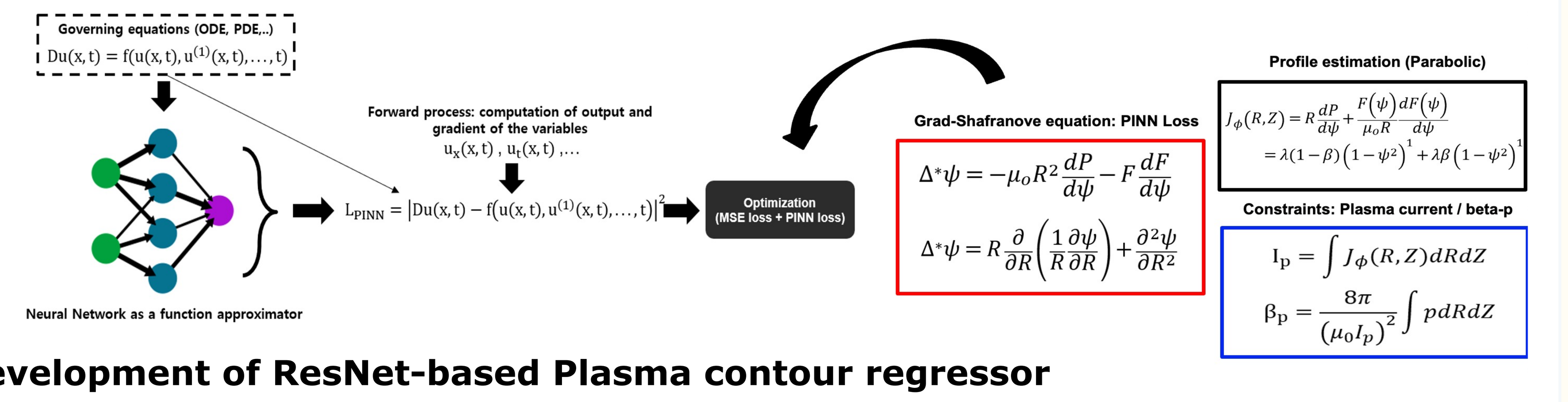


- Integration: Neural Networks for predicting **plasma state parameters + Magnetic flux + LCFS**
- NN-based simulator: Transformer [9] based model for predicting simulator
- Modules for visualizing tokamak plasma: PINN [10]-based Grad-Shafranov solver + ResNet-based contour regressor

Development of Transformer-based plasma state simulator



Development of PINN-based Grad-Shafranov equation solver



Reward Engineering: Reward calculation for target control

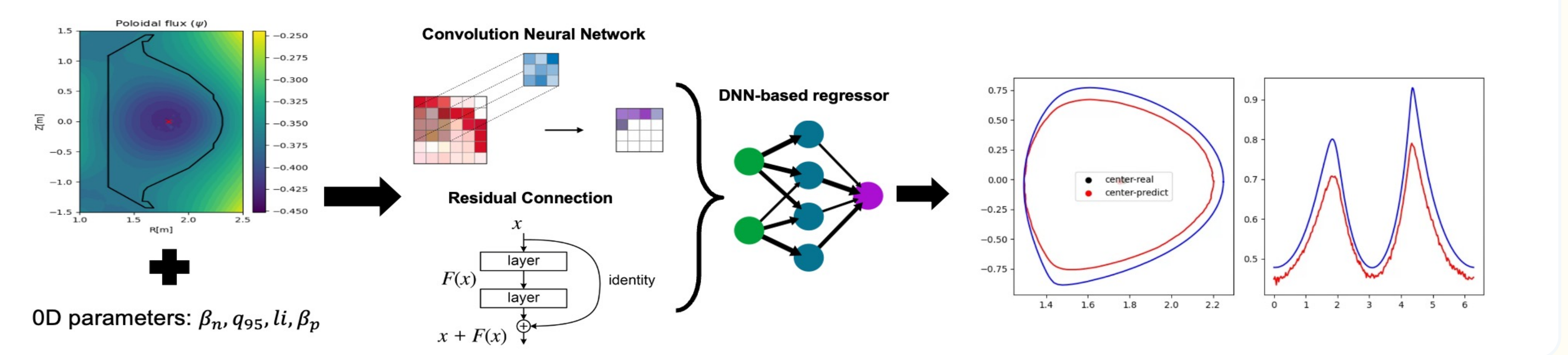
- **Reward**: key component for agent to **provide the feedback** for learning optimal policy
- **Linear scalarization** used for **converting multi-objectives as single-objective task**

Reward function for $target_i$ at state s_t : $R(s_t)_i = \tanh\left(\frac{1}{target\ value \times 0.001 + \sqrt{(target\ value - actual\ value)^2}}\right)$

Linear scalarization for multi-target control: $Total\ Reward = \sum w_i R(s_t)_i$

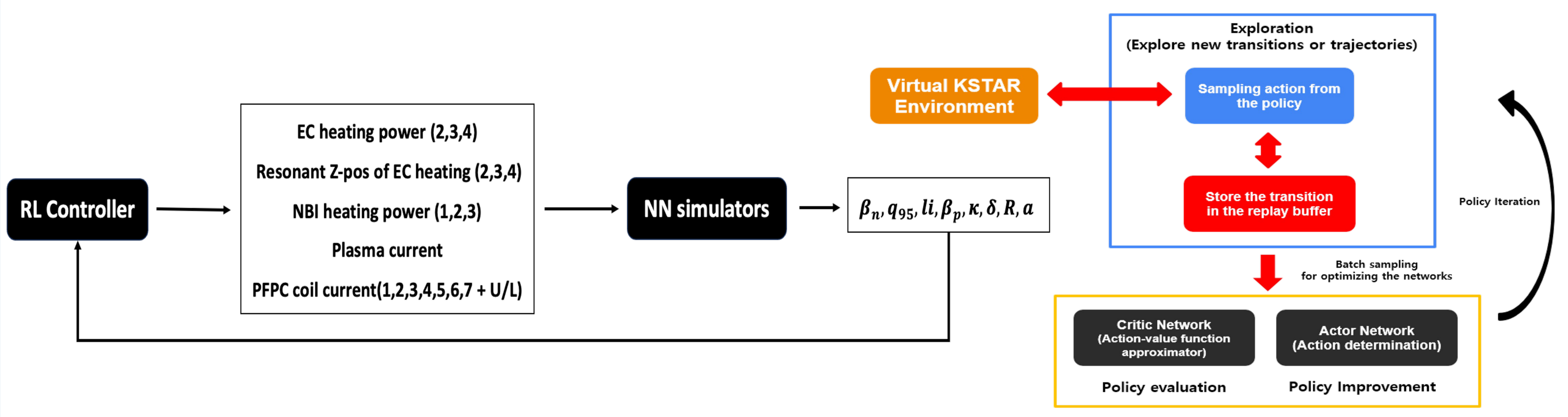
Single-objective reinforcement learning still valid under the linear scalarization

Development of ResNet-based Plasma contour regressor



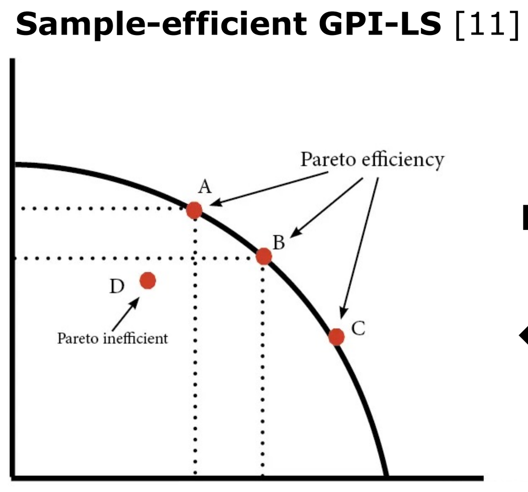
RL framework application

- Mutual interaction between virtual KSTAR environment and RL controller is prerequisite.
- We applied Soft Actor-Critic algorithm[12] for finding the optimal way to approach target values.



Generalized Policy Improvement Linear Support

- If **Pareto-optimal** situations, we can not find the optimal policy satisfying multi-objectives.
- **Sample-efficient GPI-LS** [11] used for searching finite space of corner weights → Pareto frontier



Linear scalarization to find the optimal policy corresponding to the given weight vector

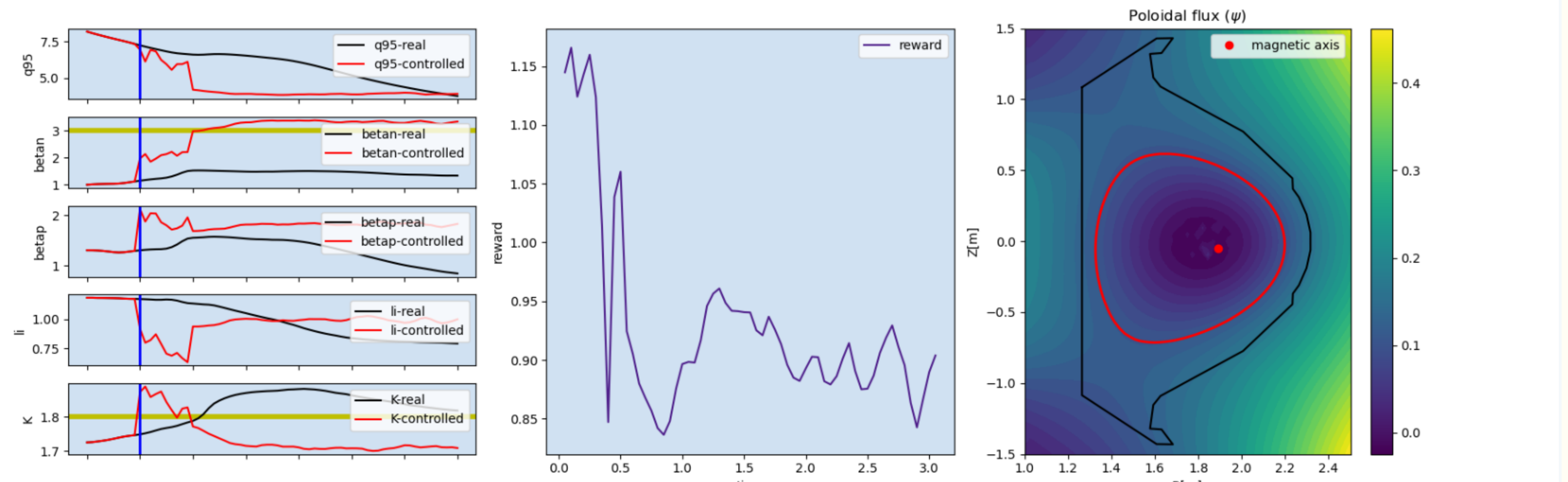
$$\sum_{i=1} w_i r_i(s_t, a_t) = u(\vec{r}(s,a))$$

Given the set of policies, find the new weight vector that guarantee to achieve the maximum possible improvement via GPI-LS algorithm

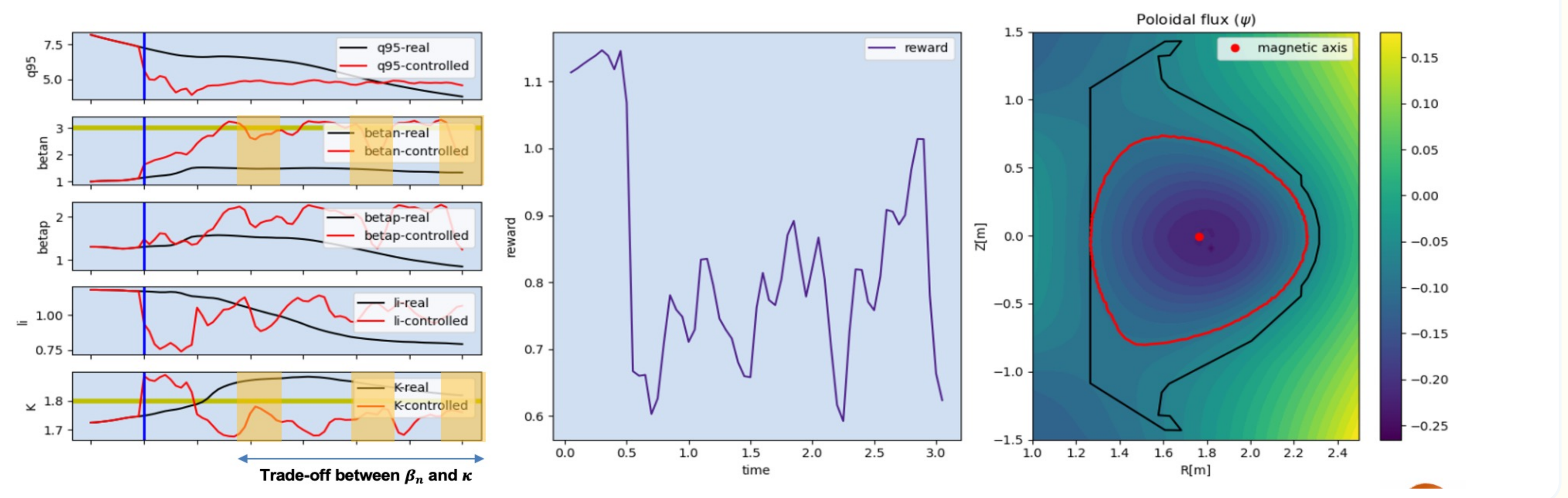
$$\mathrm{argmax}_{\,w \in W_{corner}} (V_w^{GPI} - \max_{\pi \in \Pi} V_w^{\pi})$$

Simulation results

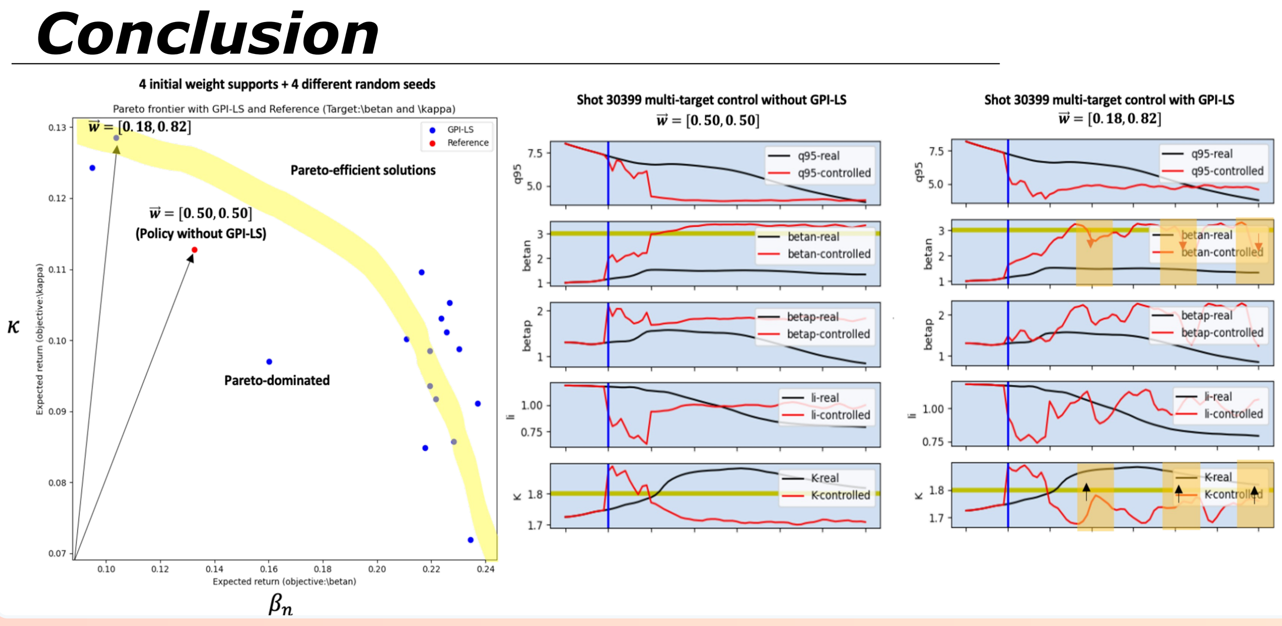
- KSTAR shot 30399 for initial data + SAC control + Target: $\beta_n = 3.0$, $\kappa = 1.8$ (SORL): Elongation X



- KSTAR shot 30399 for initial data + SAC control + Target: $\beta_n = 3.0$, $\kappa = 1.8$ (MORL): Observation of Pareto-optimal



Conclusion



- For multi-targets control, it is necessary to consider the relation between controlled variables including β_n and κ.
- We should find the feasible set of policies and compare them to choose the optimal policy with our preference.
- MORL can find the Pareto-frontier set for multi-target control including plasma performance and shape.
- Virtual KSTAR environment based on NN-simulator and RL application can be used as a navigator for forecasting the feasible and optimistic way to achieve high performance.

References

[1] H.R.Koslowski, "Operational Limits and Limiting instabilities in Tokamak machines", Transactions of fusion science and technology, 61:2T,96-103 (2012)
[2] J.A.Wesson et al, "Disruptions in JET", Nucl.Fusion 29 (1989)
[3] M.F.Turner and J.A.Wesson, "Transprot, Instability and Disruption in Tokamak", Nucl. Fusion 22 (1982)
[4] L.P.Kaelbling, M.L.Littman and A.W.Moore, "Reinforcement Learning : A Survey", Journal of Artificial Intelligence Research, Vol.4 (1996)
[5] Yuxi Li, "Deep Reinforcement Learning : An Overview", arXiv preprint arXiv : 1701.07274 (2017)
[6] J.Seo, Y.-S.Na, B.Kim, C.Y.Lee, M.S.Park, S.J.Park and Y.H.Lee, "Feedforward beta control in the KSTAR tokamak by deep reinforcement learning", Nucl. Fusion 61 (2021)
[7] Degrave,J., Felici, F., Buchli,J. et al, "Magnetic control of tokamak plasmas through deep reinforcement learning", Nature 602, 414-419 (2022)
[8] J.Seo, Y.-S.Na, B.Kim, C.Y.Lee, M.S.Park, S.J.Park and Y.H.Lee, "Development of an operation trajectory design algorithm for control of multiple 0D parameters using deep reinforcement learning in KSTAR", Nucl.Fusion 62 (2022)
[9] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N.Gomez, Lukasz Kaiser and Lllia Polosukhin, "Attention is All You Need", NIPS 2017 (2017)
[10] M. Raissi, P. Perdikaris, G.E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," J. Comput. Phys. 378 (2019)
[11] Lucas N. Alegre, Ana L. C. Bazzan, Diederik M. Roijers, Ann Nowé and Bruno C. da Silva, "Sample-Efficient Multi-Objective Learning via Generalized Policy Improvement Prioritization", AAMAS (2023)
[12] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel and Sergey Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor", ICMR (2018)

2nd International Fusion Plasma Conference (IFPC2023), 21 ~ 25 Aug. 2023, BPEX, Busan, Korea