Advanced Magnetic Plasma Control Enabled by Reinforcement Learning

G.F. Subbotin[1], D. Sorokin[1], A. Granovskiy[1], I. Kharitonov[1], E. Adishchev[1], E.N. Khairutdinov[1], M.R. Nurgaliev[1], R. Clark[2], H. Shen[3], W. Choi[3], J. Barr[3], and D.M. Orlov[2]

[1] *Next Step Fusion, Luxembourg, Mondorf-les-Bains, 56A Avenue François Clément, L-5612*
[2] *Center for Energy Research, University of California, San Diego. San Diego, CA 92093*
[3] *General Atomics, San Diego, CA, 92186*
e-mail: gs@nextfusion.org

A novel reinforcement learning (RL) based controller for shape and position control is tested on the DIII-D tokamak in various plasma regimes. The developed approach for controller design is flexible for any target specification as well as any diagnostic setup. Fusion Power Plant sets significant challenges for control due to the limited diagnostics set and reduced actuators efficiency. RL approach is a promising technology that enables high-precision robust control by bringing model-based features and fast analysis of complex sensor measurements in real time into the control loop.

Developed methods to create such controllers employ reinforcement learning. With the developed machine-agnostic pipeline, training a single controller takes less than 24 hours. RL has gained significant attention in the fusion industry in recent years. It is a promising approach that offers the possibility to train optimal non-linear multiple-input and multiple-output (MIMO) controllers for real-time plasma control. RL has been applied for magnetic plasma control on TCV [1] and HL-3 [2] as well as for stability control on DIII-D [3], and for kinetic control on KSTAR [4]. The studies referenced in [2-4] rely on surrogate environments trained using historical data.

The initial experimental results demonstrate that RL-controllers trained using this method are effective in both H- and L-mode plasma regimes and can handle discrete events like H-L transitions. Despite training episodes being limited to 1 second, the RL-controller has proven capable of maintaining plasma control for extended durations. An example of the plasma shot controlled by RL-controller is shown in Figure 1.

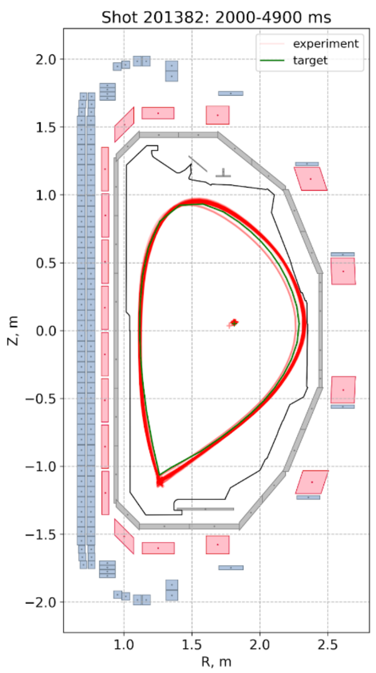

Fig. 1. Plasma boundary shape evolution. Red contours show the evolution of the plasma boundary with a 20 ms time step. While each contour is drawn with transparency, their overlap shows a small dispersion around quasi-steady-state equilibrium. The target is given by a green line.

Out-of-domain tests are carried out by varying injected neutral beam (NBI) power. The RL-controller trained on L-mode reference with low β_N=0.37 and an average 0.9 MW NBI power managed to control shape in the H-mode plasma with β_N=2.4 and P_{NBI} = 7.5 MW. Increasing β_N from 0.1 to 2.4 (by P_{NBI} = 0.0 to 7.5 MW) allowed for an investigation of the influence of plasma pressure on control errors. In shots with β_N<0.6-0.7 shape and plasma position are controlled within a 2 cm range. However, in shots with high β_N>2 errors increase up to 3 cm in shape and 8 cm in magnetic axis position, nearing the termination threshold implemented during training. Improving control quality can be achieved by incorporating more precise models for auxiliary heating and current drive.

The training utilized the Soft Actor-Critic (SAC) algorithm [5], which employs two separate neural networks: the Critic and the Actor, as illustrated in Figure 2. The RL agent was trained using NSFsim [6], a 2D Grad-Shafranov equilibrium and 1D transport code, as the simulation environment. The Actor neural network, after training, functions as a non-linear MIMO RL controller designed for deployment within the Plasma Control System (PCS). It relies exclusively on signals available in real-time, such as magnetic probes, flux loops, and coil currents. To ensure robustness, these observations are artificially noised, with the noise dispersion calibrated based on an analysis of experimental signals. The Critic neural network has access to privileged information about the plasma state and its dynamics. It observes non-noisy signals from magnetic probes, flux loops, and coil currents, as well as the (*R*, *Z*) coordinates of the last closed flux surface, magnetic axis, and X-point. This additional information allows the Critic to better estimate future rewards, thereby improving the training performance. The reward function is constructed using three components that quantify deviations from the given target state: (1) shape distance, (2) magnetic center distance, and (3) X-point distance.

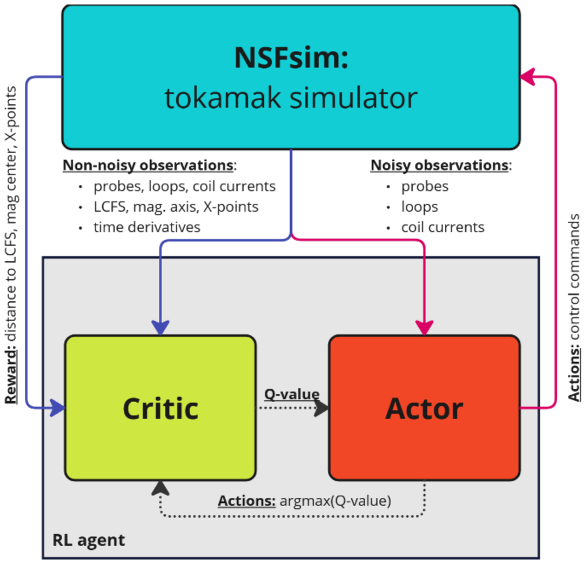

Fig. 2. Training RL-agent with NSFsim environment. Critic and Actor neural networks observe state differently: Actor sees only noisy signals that are available from real-time diagnostics and Critic has non-noisy signals as well as additional information about plasma shape and position

By using raw sensor signals instead of an external equilibrium reconstruction (1–1.5 ms on a 65×65 grid), the evaluation time is cut to <40 μs per cycle. This approach supports fine spatial grids, is scalable to larger devices, and is limited primarily by the simulation's computational speed. Partial observability from kinetic terms is addressed by randomizing initial states, coil currents, gradient profiles (dp/dψ, FdF/dψ), temperatures, and Z_{eff}, ensuring the RL controller is robust to disturbances in real experiments.

The environment is equipped with an actuator layer to provide choppers with low-level control commands. The poloidal field (PF) coil system at DIII-D uses a common feeding power bus, which brings significant challenges to simulating plasma response during agent training. A model of choppers, together with power supplies, is used to accurately represent PF coil control.

This work was supported in part by the US Department of Energy under DE-FC02–04ER54698 and by Next Step Fusion

References

[1]. Degrave, J., Felici, F., Buchli, J. et al. Nature 602, 414–419 (2022)
[2]. Niannian Wu, Zongyu Yang, Rongpeng Li et.al. arXiv:2409.09238
[3]. Seo, J., Kim, S., Jalalvand, A. et al. Nature 626, 746–751 (2024)
[4]. Jaemin Seo et al 2021 Nucl. Fusion 61 106010
[5]. T.Haarnoja et.al. arXiv:1801.01290
[6]. R. Clark et al. Fusion Engineering and Design 211 (2025) 114765