CONFERENCE PRE-PRINT

ADVANCED MAGNETIC PLASMA CONTROL ENABLED BY REINFORCEMENT LEARNING

G.F. SUBBOTIN, D. SOROKIN, A. GRANOVSKIY, I. KHARITONOV, E. ADISHCHEV, E.N. KHAIRUTDINOV, M.R. NURGALIEV
Next Step Fusion
Bertrange, Luxembourg
Email: gs@nextfusion.org

Abstract

The development of stable and efficient magnetic confinement for high-temperature plasmas is a critical challenge in the pursuit of fusion energy. Traditional control systems in tokamaks rely on complex, model-based algorithms that are computationally intensive and depend on accurate plasma state reconstructions. This paper presents a alternative approach to plasma shape and position control utilizing deep reinforcement learning (RL). We have successfully developed and experimentally validated an RL-based controller on the DIII-D tokamak that directly maps magnetic sensor data to actuator commands, bypassing the need for explicit state reconstruction. Our controller demonstrates robust and precise control over plasma shape and position, maintaining stability even in the presence of significant plasma perturbations. This reconstruction-free method represents a significant advancement, offering a computationally efficient and highly adaptable solution for real-time plasma control in current and future fusion devices.

1. INTRODUCTION

Nuclear fusion is widely regarded as a promising solution for sustainable energy, offering clean, safe, and nearly limitless power without carbon emissions. Among the various fusion concepts, the tokamak has emerged as the most successful device for achieving magnetic confinement of high-temperature plasmas. In tokamaks, toroidal (TF) and poloidal (PF) field coils work together to provide confinement, stability, and active plasma control. While TF coils, along with the plasma current, generate the rotational transform essential for confinement, PF coils are responsible for shaping, positioning, and stabilizing the plasma, as well as controlling divertor and exhaust configurations. Because instabilities in tokamak plasmas can develop on millisecond timescales, real-time plasma control systems (PCS) are required to ensure stable, high-performance operation.

Traditional PCS approaches rely on equilibrium reconstruction codes such as rtEFIT [1], LIUQE [2], and Equinox [3], which solve the Grad–Shafranov equation using magnetic diagnostics to infer plasma shape and current profiles. Although highly effective, these methods require continuous reconstruction and significant computational resources. Recent advances in machine learning have introduced an alternative: reinforcement learning (RL)-based controllers that operate directly on raw diagnostic signals, bypassing reconstruction and reframing control from a conventional “sensor–actuator” model to a state-oriented framework.

In this work, we explore RL-based magnetic control on the DIII-D tokamak, where the control task is formulated as a partially observable Markov decision process and solved using the Soft Actor-Critic algorithm. Training is carried out with NSFsim [4], a high-fidelity simulator that reproduces plasma dynamics by coupling magnetic and kinetic equilibrium evolution. The resulting controllers demonstrate strong robustness across scenarios, successfully stabilizing high-performance H-mode plasmas, handling H–L transitions, and maintaining control under varying auxiliary heating conditions.

2. REINFORCEMENT LEARNING FOR PLASMA CONTROL

The principal training scheme is illustrated in Figure 1a). Plasma evolution is modeled using NSFsim, which solves the free-boundary Grad–Shafranov equation coupled with 1D core transport and provides plasma responses to PF-coil current variations. The RL agent is trained in this environment with the Soft Actor-Critic (SAC) [5] algorithm, employing an asymmetric Actor–Critic architecture. The Actor receives noisy measurements from magnetic probes, flux loops, and coil currents, while the Critic is granted privileged access to exact diagnostic values, last closed flux surface (LCFS) geometry, magnetic axis, X-points, and their temporal derivatives. The Actor generates low-level control commands that nonlinearly establish currents in shaping PF coils.

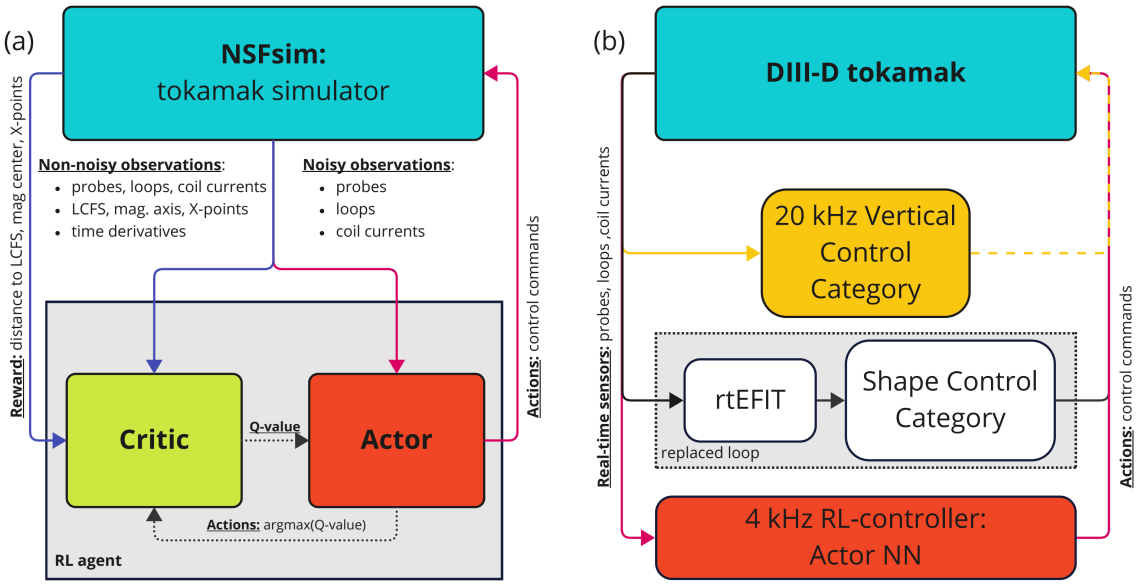

Fig. 1. The training (a) and test (b) environments scheme. Elements with the same color correspond between figures (a) and (b).

The learning objective is to maximize a cumulative discounted reward quantifying the similarity between the evolving and target plasma shapes. The reward equals unity for a perfect match and decreases toward zero as deviations increase. The Critic network, is implemented as a 552–256–1 multilayer perceptron (MLP) with ReLU activations, while the Actor is a compact 132–256–18 MLP with ReLU activations that maps real-time sensor inputs directly to actuator commands. Once trained, the Actor can be deployed to the DIII-D plasma control system (PCS) without the need for equilibrium reconstruction.

Each training episode lasts 1 second and is initialized with a plasma state defined by current, toroidal field, PF-coil currents, pressure and flux gradients ($dp/d\psi$, $FdF/d\psi$), electron and ion temperatures, and effective ionic charge Z_{eff}, extracted from EFIT reconstructions and experimental data. A reference plasma shape, derived from selected experimental shots, serves as the target for control. Small discrepancies between NSFsim and EFIT equilibria, arise from diagnostic uncertainties and different consideration of eddy currents in passive structures.

Robustness against real-machine variability is ensured by exposing the agent to a wide distribution of plasma and hardware conditions. Two complementary training strategies were adopted. In the first, each episode begins with perturbed plasma parameters: T_e, T_i, and Z_{eff} values at the magnetic axis and separatrix are randomized within the intervals wider than usual operational space of DIII-D, while initial PF-coil currents are offset by up to ±100 A. Subsequent plasma kinetic evolution is modeled with NSFsim using Bohm–gyroBohm transport coefficients [6]. Only Ohmic heating is considered; randomized initial temperatures emulate both hot and cold plasmas, thereby reducing computational overhead. At each time step, Gaussian noise consistent with experimental measurements is added to diagnostic signals. In this setting, plasma temperature decreases throughout each episode, exposing the agent to a range of β_p values. Controllers trained with this strategy are denoted RL24.

The second strategy employs fitted profiles of electron temperature, electron density, and ion temperature, which are scaled by random multipliers in the range (0, 4] at the start of each episode. This procedure maintains constant kinetic profiles, and therefore constant β_p, over the course of training. Such treatment improves the agent's ability to manage the Shafranov shift, as discussed later. Controllers trained with this approach are denoted RL25.

3. EXPERIMENTAL RESULTS

Four different RL controllers were tested during the 2024–2025 experimental campaign. These controllers differ in the reference shots used for training, which determine the set of magnetic probes and loops employed, the active PF coils used for control, and the plasma conditions. Two RL24 controllers were trained on H-mode (RL24-1, #200563) and L-mode (RL24-2, #186093) reference discharges. Two RL25 controllers (RL25-1 and RL25-2) were trained on an H-mode reference (#193593). RL25-1 included more magnetic sensors in its observations – 48

probes and 30 loops, compared with 46 probes and 23 loops for RL25-2. Both RL25 controllers were tested during a single experimental day.

As an initial proof of concept, the controller is tested in discharge 201381 with assistance from the fast vertical stability algorithm to sustain plasma position. In the subsequent discharge (201382), the control interval is 3 seconds to evaluate controller performance beyond the training duration. In this case, the 20 kHz vertical stabilization from isoflux is disabled after 1 seconds from handover. The corresponding plasma boundary evolution is shown in Figure 2.

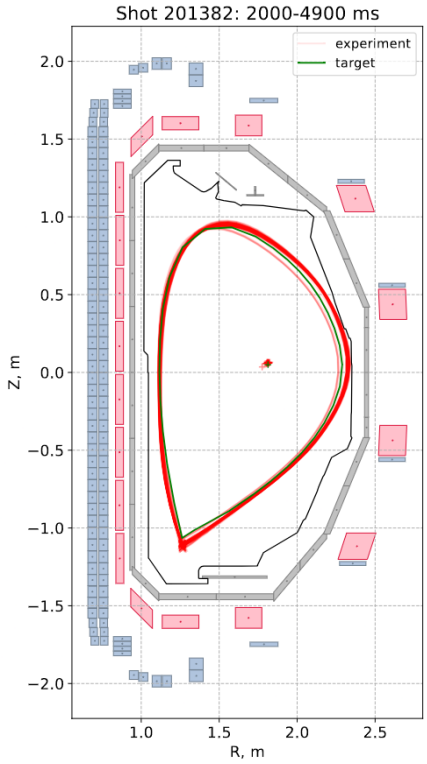

Fig. 2. Plasma boundary shape evolution in shot 201382. Red contours show the evolution of the plasma boundary with a 20 ms time step. While each contour is drawn with transparency, their overlap shows a small dispersion around quasi-steady-state equilibrium. The target is given by a green line.

The RL-based controller maintained the plasma magnetic axis at the target location despite the increased heating power relative to the reference case. However, the plasma outer boundary shifted closer to the first wall. This expansion resulted from the combination of elevated plasma pressure and approximately 15% lower total PF-coil currents commanded by the controller, which collectively led to an increased plasma volume. Time traces of plasma shape parameters, the magnetic axis, and X-point coordinates are shown in Figure 3a). Importantly, disabling the 20 kHz vertical stabilization did not degrade control quality, which remained stable until the end of the discharge. The most significant discrepancy was observed in the vertical position of the X-point.

Additional tests demonstrated cross-regime applicability. While controller RL24-1 was trained on an H-mode reference, RL24-2 was trained on an L-mode reference with only short neutral beam blips of 2 MW. Nevertheless, RL24-2 successfully controlled plasma shape in H-mode operation with continuous neutral beam injection of 7.5 MW. The corresponding performance is shown in Figure 3b). Operation in a substantially different transport regime introduced larger deviations from the target shape, which were corrected once neutral beam injection was switched off.

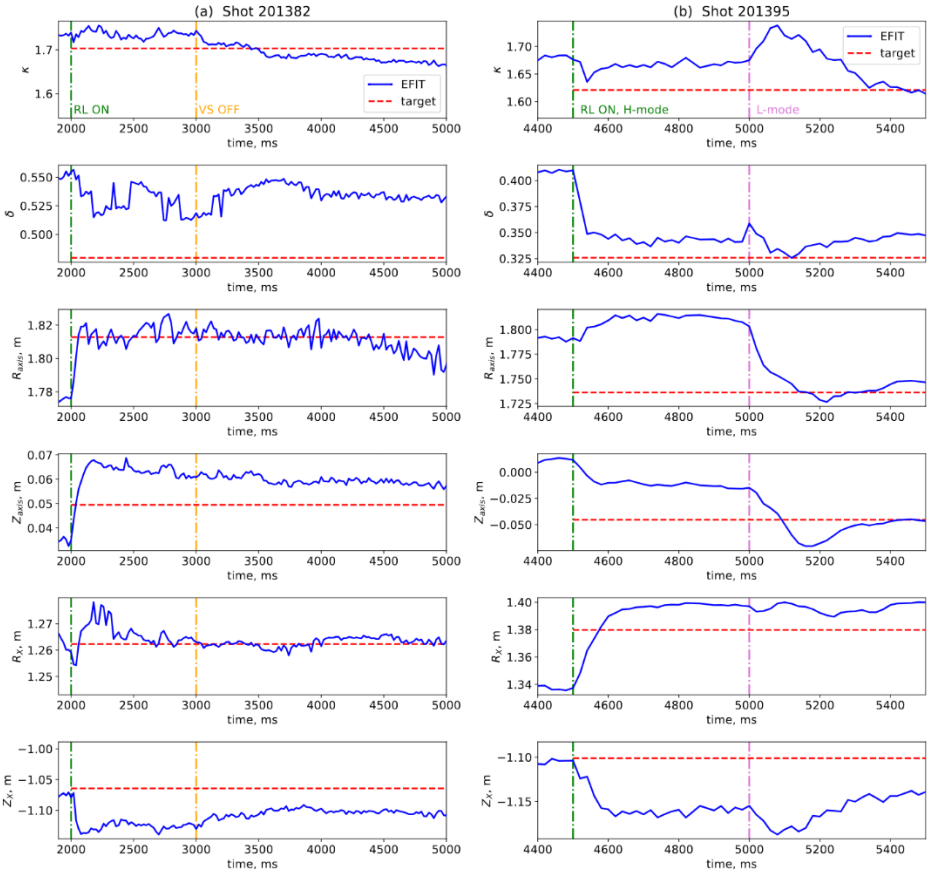

Fig. 3. Comparison between the target values that were used for training the RL-controller and the experimental result. Illustrated values are elongation (κ), up-down average triangularity (δ), coordinates for magnetic axis (R_{axis}, Z_{axis}), and X-point (RX, ZX). (a) Shot 201382 with H-mode regime and 3 s control interval by controller RL24 #1. At time t = 3 s, fast vertical stability is turned off. (b) Shot 201395 with H-L transition at t = 5 s due to switching NBI off.

Improved control performance is demonstrated by RL25 controllers during the second testing campaign. As shown in Figure 4, the overall control error of the plasma boundary and magnetic axis position was maintained within 1.5 cm and 1 cm, respectively. Owing to their different treatment of plasma pressure, RL25 agents exhibited more robust behavior under variations in the kinetic profiles. In contrast to RL24 controllers, no visible displacement of the plasma center is observed during changes in neutral beam injection (NBI) power.

Quantitative assessment of control quality is carried out using metrics such as errors in the horizontal and vertical positions of the magnetic axis and X-point, as well as the mean deviation of the separatrix from the target shape. These statistics are evaluated over the last 100 ms of RL25 operation across 11 discharges with identical target configurations. While the separatrix deviation metric provides only a global measure of shape control accuracy, it nonetheless offers a useful indication of controller performance. For RL25-2, the mean shape deviation is less than 1.2 cm in most cases (see Figure 5). Higher errors were concentrated in the upper region of the plasma column and near the X-point, consistent with the analysis of X-point positioning accuracy.

Figure 6 presents the error statistics for plasma center and X-point positions. The magnetic axis is maintained within 0.7 cm of the reference in most cases, with horizontal control performing better than vertical (0.25 cm versus 0.5 cm error, respectively). A similar trend is observed for the X-point, although the difference is more pronounced: horizontal errors are typically 0.4 cm, whereas vertical errors reached up to 4 cm. This limitation is attributed primarily to discrepancies between the plasma models used during training and the actual experimental conditions. Larger errors in the vertical X-point position are expected, given its sensitivity to the balance between plasma current density and external coil currents. Since edge current density distributions is not explicitly constrained during training, X-point positioning is more susceptible to modeling inaccuracies.

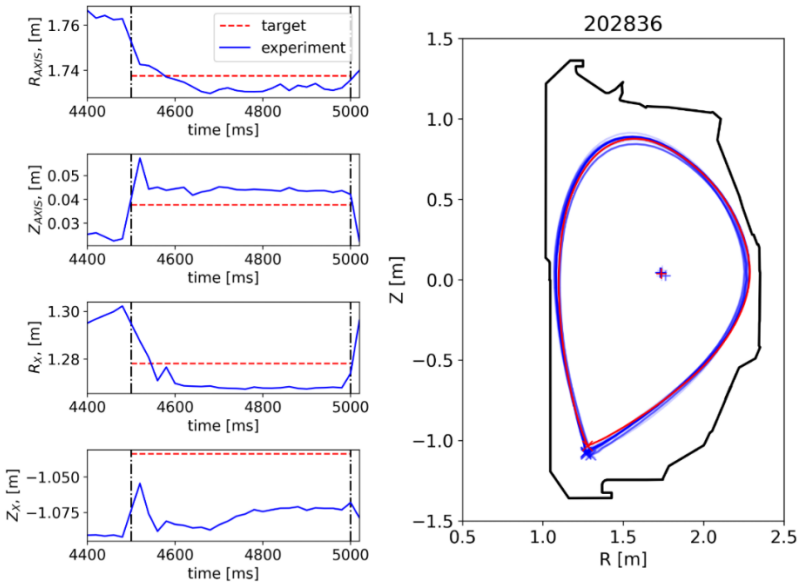

Fig. 4. Plasma axis and X-point position timetraces (left) compared with target values (red line) and shape time evolution (right) in blue line compared with target shape (red line)

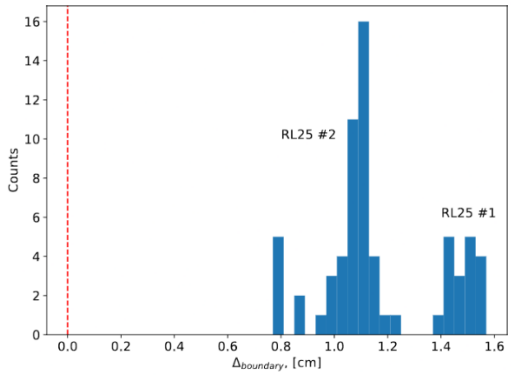

Fig. 5. Control errors distribution - separatrix shape control, statistics over 11 shots with I_p=1 MA, last 100 ms to avoid transient effects influence

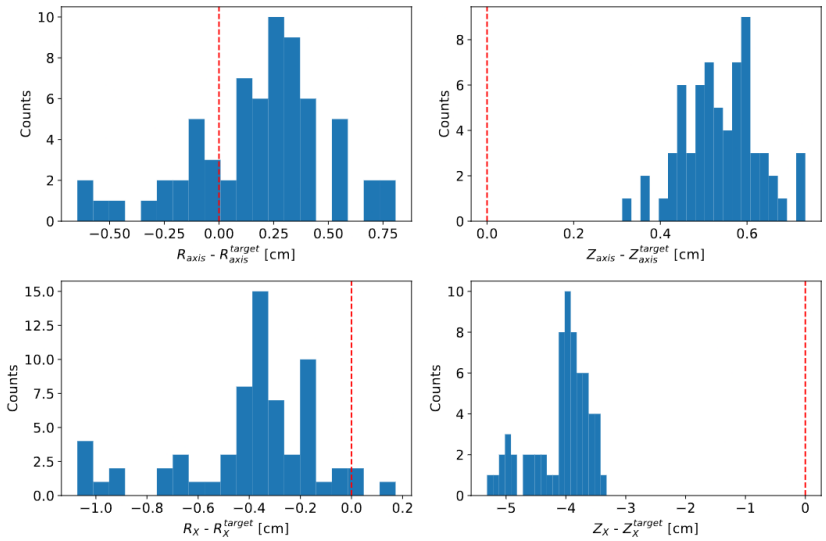

Fig. 6. Control errors distribution - magnetic axis position and X-point position, statistics over 11 shots with I_p=1 MA, last 100 ms to avoid transient effects influence

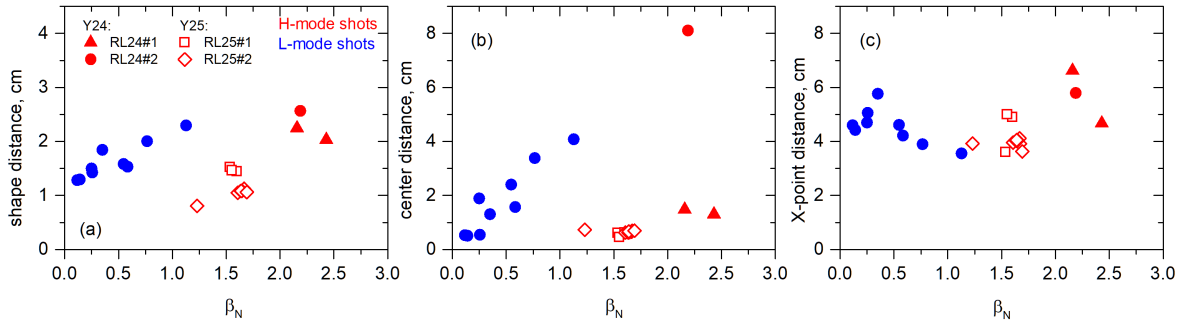

Fig. 7. Change in reward components during β_N scan with P_{NBI} for both RL-controllers. While errors in shape and magnetic axis position increase with β_N, X-point position mismatch is insensitive to it.

The dependence of controller performance on β_N is summarized in Figure 7. Reward components, including shape deviation and offsets in magnetic axis and X-point positions, are plotted against β_N values obtained under varying NBI power. For RL24 controllers, shape and axis mismatches remained below 2 cm at low normalized beta (0.5–0.7) and were smallest in Ohmic conditions (no NBI). Errors increased systematically with higher normalized beta. This behavior reflects the bias of the RL24 training procedure toward Ohmically heated regimes. However, the RL24-1 controller, which was trained with a high-power H-mode reference, displayed reduced axis errors under high-beta conditions. By contrast, RL25 controllers performed consistently across all tested β_N values, without degradation in either shape or position control.

4. CONCLUSION AND DISCUSSION

This work demonstrates the successful development, integration, and experimental validation of reinforcement learning (RL)-based controllers for plasma shape and position control on the DIII-D tokamak. By directly mapping raw diagnostic signals to actuator commands, these controllers eliminate the need for real-time equilibrium reconstruction and provide fast and efficient control at the computational cost of simple matrix operations. Training uses NSFsim, which couples magnetic equilibrium and kinetic transport, and allows the agent to learn robust strategies under varied heating and plasma conditions.

Two training strategies are explored. Controllers from the RL24 series, trained with randomized Ohmic-like conditions, show strong performance in H-mode and during transient events such as H–L transitions. However, these controllers remain sensitive to variations in heating power, with increased shape deviations and X-point errors at higher β_N. By contrast, RL25 controllers, trained with scaled kinetic profiles and constant β_N, achieve improved robustness against Shafranov shift variations. Experimental validation confirms that RL25 maintains plasma boundary and magnetic axis within 1.5 cm and 1 cm, respectively, even under drastic changes in NBI power and pellet fueling. Error statistics show that the plasma center remains typically within 0.7 cm of the reference, with horizontal positioning outperforming vertical control. The largest discrepancies appear in the vertical X-point location, reflecting limitations of the training models and the sensitivity of X-point positioning to edge current density distributions not explicitly constrained during training.

These results highlight the significant potential of RL methods for advanced plasma control. Compared to traditional approaches such as model predictive control, RL does not require linearized or simplified models during training, and therefore handles nonlinear and multi-physics dependencies. Randomization during training also defines the operational domain, ensuring robust performance within known limits while flagging operation outside this range as failure modes for machine protection.

Looking forward, further progress requires the incorporation of heating and current-drive models, more accurate transport treatments, and faster plasma simulation tools with synthetic diagnostics. Ultimately, RL-based methods, in combination with model predictive control and other approaches, provide a comprehensive and flexible framework for robust plasma control in future fusion power plants, addressing challenges such as limited diagnostics, constrained actuator availability, and long-pulse operation.

ACKNOWLEDGEMENTS

This material is based upon work supported partly by Next Step Fusion S.a.r.l and by the U.S. Department of Energy, Office of Science, Office of Fusion Energy Sciences, using the DIII-D National Fusion Facility, a DOE Office of Science user facility, under Award(s) DE-FC02–04ER54698.

REFERENCES

[1] J.R Ferron et.al. Nuclear Fusion, 38(7):1055–1066, July 1998.
[2] J.-M. Moret et.al. Fusion Engineering and Design 91:1–15, February 2015.
[3] J Blum et.al. Journal of Physics: Conference Series, 135:012019, November 2008.
[4] R. Clark et.al. Fusion Eng. Des. 211, 114765, 2025.
[5] T. Haarnoja et.al. arXiv:1801.01290, 2018.
[6] M. Erba et.al. Nucl. Fusion, 38(7):1013, 1 July 1998