ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-FRAG002

DEVELOPMENT AND EXPERIMENTAL VALIDATION OF A MACHINE LEARNING-BASED METHODOLOGY FOR CYCLOTRON BEAM CONTROL: RESULTS FROM THE PSI HIPA FACILITY*

M. Haj Tahar†, M. Bocchio, M. Busch, W. Joho, S. Marquie, E. Solodko,
Transmutex SA, Vernier, Switzerland
A. Barchetti, C. Baumgarten, J. Grillenberger, M. Sapinski, M. Schneider, J. Snuverink,
Paul Scherrer Institute, Villigen, Switzerland

Abstract

Transmutex SA is developing an accelerator-driven system (ADS) designed to generate clean energy while reducing the lifetime of radioactive waste. Such a subcritical reactor concept requires high reliability and a high degree of accelerator automation to ensure operational effectiveness.

To address these demands, a machine learning (ML) methodology was developed and experimentally validated for automatic beam control in cyclotrons. This work reports the first practical demonstration of machine-learning-based beam control in a high power cyclotron, representing a significant step for this class of accelerators.

The validation experiments were performed on the injector ring of the High Intensity Proton Accelerator (HIPA) at the Paul Scherrer Institute (PSI), whose design closely matches the injector concept developed by Transmutex. Key challenges were addressed, including the identification of suitable observables and actuators, adapting the ML model to the accelerator response dynamics, and integrating ML-based control with existing feedback loops. The approach reliably aligned the beam with the reference trajectory, improving extraction efficiency while minimizing losses.

Over an extensive 12-day operational test campaign, remarkably long in the context of real-time ML experiments, the model demonstrated robust performance across a range of operational scenarios, including varying beam currents and different turn numbers.

These results show that machine learning can enhance operational efficiency, reduce operator workload, and increase automation in cyclotron-driven systems.

INTRODUCTION AND MOTIVATION

The operation of Accelerator-Driven Systems (ADS) requires accelerators to be tuned with high stability, reproducibility, and efficiency. Traditional manual approaches, although effective in research environments, cannot meet these industrial demands.

Machine learning (ML) offers a promising path forward. By enabling automated, data-driven tuning, ML has the potential to simplify operational control, reduce reliance on expert operators, increase reliability, and improve efficiency of accelerator operation.

At Transmutex, these ideas are being developed within the framework of the START (Subcritical Transmutation by Accelerated Reliable Technology) system and particularly its cyclotron design. Given the architectural similarity between the START cyclotron design and the HIPA accelerator complex at the Paul Scherrer Institute (PSI), the Injector 2 at PSI serves as an ideal experimental platform. PSI granted twelve days of dedicated beam time, offering a unique opportunity to validate ML-based beam control under realistic operating conditions.

EXPERIMENTAL SETUP

HIPA Injector 2 Overview

Injector 2 is a four-sector cyclotron designed to accelerate protons from an injection energy of 870 keV to a final energy of 72 MeV. Its main components include four large sector magnets (SM1–SM4), two double-gap RF resonators (CI1 and CI3), and two recently installed single-gap resonators (CI2 and CI4). Extraction is performed by the septum magnet (AXA), which deflects the last orbit toward the extraction channel. The beam is then bent by the next dipole (AXB) into the transfer line where final beam current is measured by current transformer (MXC1). Finally the beam is stopped by a dedicated beam dump (BX2). A schematic layout is shown in Fig. 1.

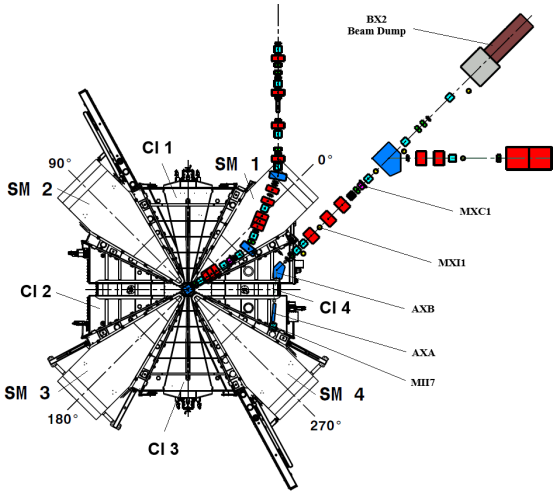

Figure 1: Schematic layout of Injector 2 and its main components.

* Work supported by Paul Scherrer Institute.
† malek@transmutex.com

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-FRAG002

Sector Magnets and Trim Coils

Each sector magnet is equipped with:

- A pair of main coils (AIHS) providing the primary magnetic field,
- Nine pairs of trim coils (TI3-TI11), symmetrically placed above and below the median plane, used for local orbit correction and optimization of extraction conditions.

In addition SM1 and SM3 magnets have another pairs of coils (TIA/TIB and TI2) at their pole tips for the first orbits fine-tuning.

A representative 3D view of SM1 is shown in Fig. 2.

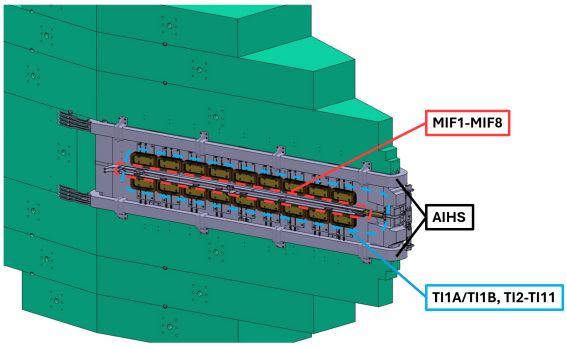

Figure 2: Sector magnet SM1 with phase measurement devices, main and trim coils.

RF Acceleration System

Beam acceleration is provided by four RF resonators. CI1, the first double-gap resonator, is particularly critical since it determines the initial turn trajectory and remains fixed during tuning. CI3, another double-gap resonator, and the newly added single-gap resonators CI2 and CI4 increase energy gain and operational flexibility. Together, they determine energy per turn, beam phase stability, and radial trajectory.

Beam Phase Monitoring

Eight phase measurement probes (MIF1–MIF8) are installed along the radius of SM1 (see Fig. 2). These devices measure the beam phase turn by turn, providing an indirect but highly sensitive diagnostic of the beam trajectory.

Loss Monitoring

Extraction losses are monitored using:

- Ionization chambers MII7 and MXI1,
- The KXAI collimator set attached to AXA septum.

These devices provide location- and magnitude-resolved loss monitoring, ensuring both safe and efficient operation.

INTERACTION WITH INJECTOR 2

The experimental interface to Injector 2 relies on the EPICS (Experimental Physics and Industrial Control System) framework, the standard control protocol at the PSI HIPA facility. The Python-based PyEPICS library was employed to provide direct access to EPICS process variables (PVs) for reading diagnostics and setting machine actuators.

All control commands were executed from a dedicated Linux server within a protected PSI subnet, ensuring secure and real-time communication with the accelerator. Two types of PVs were relevant: "IST" channels, reporting measured values, and "SOL" channels, representing the requested set values. While IST readings may exhibit noise or calibration offsets, SOL values more reliably reflect the commanded machine state. For this reason, SOL channels were applied for controllable parameters such as coil currents and resonator voltages, whereas IST channels were reserved for diagnostics such as temperatures or phase probes.

This configuration provided a consistent and stable representation of the accelerator state, enabling reliable data acquisition and safe actuation during ML-driven tuning experiments.

TUNING STRATEGY AND CAMPAIGN OVERVIEW

To ensure safe deployment of machine learning on Injector 2, tuning was carried out at a beam current of 20 µA. This level was chosen because it balances measurement quality with operational safety: stable phase readings, low beam power, small beam size allowing wider action space, and meaningful loss monitor signals.

Two actuator classes were identified for ML control.

- High-impact actuators:
  1. AIHS (main coils current) – sets the central magnetic field and determines the radial trajectory, directly affecting the extraction position.
  2. CI3V (resonator 3 voltage) – governs the energy gain per turn and thereby influences beam phase and extraction location.
- Fine-tuning actuators:

  The trim coils on the sector magnets (TI1A/B, TI2–TI11). While their effect is more localized, they are indispensable for precision shaping of the magnetic field and achieving optimal trajectory control.

Operational boundaries for AIHS and CI3V were empirically scanned to map safe versus interlock regions. Central reference values were then chosen within these ranges to serve as targets for ML-based fine-tuning.

Experimental Campaign Overview

The experimental campaign was designed to evaluate ML-based tuning across a representative set of operating scenarios. Five configurations were tested, each associated with

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-FRAG002

a specific turn number, which was adjusted by controlling the peak voltages of resonators 2 and 4. The corresponding settings are summarised in Table 1. These configurations represent the following operational regimes:

- Nominal operation with all four resonators active (Stage 4),
- Reduced operation with three resonators (Stages 0-2),
- Degraded mode with only two resonators (Stage 3).

Table 1: Resonators Reference Configuration per Experimental Stage

		Resonators setup, [kVp]			
Stage	**Turn N**	**Res1**	**Res2**	**Res3**	**Res4**
0	72	430	429	451	0
1	73	430	401	449	0
2	74	430	371	448	0
3	89	430	0	449	0
4	60	430	428	448	428

Each day of the campaign followed a repeatable cycle:

1. Setup Phase (morning): Operators configured Injector 2 for the target turn number, optimized the beam at reference current (~2 mA), measured beam profiles, and defined the ML action space.
2. Tuning Phase (daytime): ML agents trained at low current (~20 μA).
3. Run Phase (evening/night): Agents were left in autonomous control overnight to evaluate robustness and stability.
4. High-Power Test (final stage only): A full day of ML control at elevated current to test performance under demanding conditions.

The full campaign schedule is presented in Figure 3.

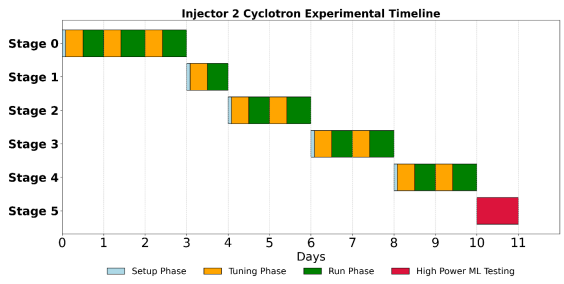

Figure 3: Phase breakdown of the Injector 2 cyclotron experiment over the 12-day campaign.

HISTORICAL DATA ANALYSIS AND FEATURE ENGINEERING

Historical data was used to formulate the ML task and pre-train models before deploying them on the real machine. It provided a safe and rich basis for identifying relevant features, defining actionable spaces, and ensuring the ML agent was exposed to realistic system dynamics without risk to the accelerator.

Data was retrieved from the PSI archiving system and resampled to a fixed 200 ms interval by forward-filling the latest values, ensuring synchronised time series across all channels.

Only machine states corresponding to the resonator configurations used during the planned experiments were retained. Notably, full operation with four resonators could not be included, as the fourth resonator was only installed at the end of 2024.

Preprocessing Pipeline

Preprocessing ensured that the dataset was clean, stable, and physically meaningful by applying filters for beam stability, physical validity, outliers, balancing across periods, and scaling to a uniform range.

Feature Selection and Importance

Beyond the predefined actionable features and target variables, feature engineering focused on aggregating redundant signals such as temperatures and identifying a concise set of additional inputs relevant for ML training.

Correlation analysis confirmed that resonator voltages and main coil current strongly influence MIF phases, while trim coils provide finer local corrections. Based on expert input and exploratory models, two parameters-CI3V and AIHS-were assigned higher importance, as they dominantly control global beam behaviour. Among outputs, downstream MIFs closer to extraction were given greater weight, reflecting their operational criticality.

After filtering and balancing, the final dataset contained approximately 27 million uniformly sampled points, equivalent to over 60 days of Injector 2 operation. This structured dataset formed the foundation for simulator pre-training, surrogate model construction, and ultimately the safe integration of ML into Injector 2 tuning.

ADAPTING ACCELERATOR PHYSICS FOR ML INTEGRATION

The main machine learning method applied was a model-free reinforcement learning (RL) algorithm, Twin Delayed Deep Deterministic Policy Gradient (TD3) [1]. It was selected for its robustness in continuous control tasks and deployed through a pipeline that combined simulation-based pretraining, physics-informed reward shaping, and real-time interlock handling.

To ensure robustness in case RL convergence was unstable or too slow, Bayesian Optimization (BO) was implemented as a backup method.

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-FRAG002

Reward Shaping

A domain-informed reward function was created to capture accelerator physics priorities and operational safety. It combined spatially resolved phase deviations, beam loss behaviour, actuator usage, and interlock risk, aligning the agent's training objectives with operator practices. Candidate reward designs were tested in a BMAD-based [2] simulation of Injector 2 beam dynamics before deployment.

A central element was the weighted root-mean-square (RMS) phase error at the eight radial probes (MIF1–MIF8), with progressively larger weights assigned downstream near extraction. The error is defined as:

$$\varepsilon_{\text{phase}} = \left[\sum_{i=1}^{8} w_i \left(\varphi_i - \varphi_i^{\text{ref}} \right)^2 \right]^{1/2}, \tag{1}$$

where φ_i is the measured phase at probe i, φ_i^{ref} the reference phase, and w_i the weight emphasising importance near extraction.

Phase and beam-loss penalties were scaled using hyperbolic tangent functions:

$$R_{\text{beam}} = -\tfrac{1}{2} \left[\tanh\left(\frac{\varepsilon_{\text{phase}}}{10} \right) + \tanh\left(\frac{\Delta I_{\text{loss}}}{30} \right) \right]. \tag{2}$$

Trim-coil usage was penalised to discourage excessive local corrections:

$$P_{\text{trim}} = -\lambda \frac{1}{N} \sum_{j=1}^{N} |I_j|, \tag{3}$$

where I_j is the unnormalised current applied to trim coil j, $N = 12$ is the number of coils, and λ a scaling factor.

The final reward is:

$$R = R_{\text{beam}} + P_{\text{trim}}, \tag{4}$$

with an additional large negative penalty in case of interlock events (beam current drop > 20% below nominal).

Interlock Handling

A two-layer interlock mechanism was integrated in parallel with the HIPA protections to enable safe autonomous exploration. Interlocks were triggered when:

- phase signals (MIF1–MIF8) became invalid or out of range,
- beam-loss diagnostics (KXAI, MII7, MXI1) exceeded thresholds or returned invalid values, or
- the beam current readings MXC1 deviated by more than ±20% from nominal.

The response protocol distinguished recoverable from persistent events: a first interlock caused the system to roll back to the last best-known action with a small perturbation; a second consecutive interlock within 15 seconds forced the environment into safe mode, halting ML actions until MXC1 stabilised within ±20% for at least 40 seconds.

All interlock events were logged with timestamp, type (single/double), and cause (MIF, losses, MXC1), enabling later analysis of fault conditions and refinement of the safe action space.

ML DEPLOYMENT

The RL deployment on Injector 2 was formulated with an action space of 14 normalized variables (12 trim coils, the AIHS main coil, and the CI3V resonator voltage) and an observation space consisting of beam phases (MIF1–MIF8), beam losses (MII7 and MXI1), beam current (MXC1), turn number, and selected environmental metrics such as magnet and air temperatures. The agent's objective was to match the measured beam phases to a predefined reference profile while minimizing beam losses and penalizing excessive trim coil usage. All quantities were normalized to the range $[-1, 1]$. Figure 4 [3] illustrates the RL paradigm applied to Injector 2, where correctors define the action space and diagnostics form the observation space.

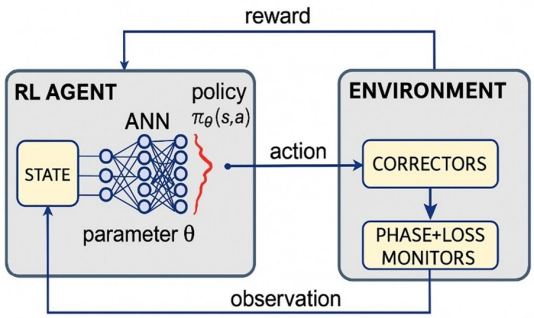

Figure 4: The RL paradigm applied to cyclotron tuning.

To enable safe and efficient training, two complementary simulation environments were developed. A physics-based environment, implemented with BMAD tracking simulations, was used to validate convergence properties of the RL algorithm, tune the scaling of parameters, and explore reward formulations. A data-driven environment, based on a surrogate model trained on historical Injector 2 data, was applied for pretraining and to reduce the need for online training. Together, these environments enabled the safe design of reward functions, shortened experimental learning cycles, and ensured robustness of the deployed agent.

RESULTS

Each selected turn, corresponding to a distinct resonator configuration, was treated as an independent tuning task. Performance was assessed using three indicators: the initial and final episode reward, the number of steps required to reach a satisfactory state, and the frequency of interlocks as a measure of operational safety.

The main metrics were defined as follows:

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-FRAG002

Table 2: Summary of RL Agent Training Performance Across Turns (Convergence is Defined as Sustaining <10 Steps per Episode)

Turn	Resonators Active	Pretraining	Convergence (timesteps / time)	Avg. Final Reward	Interlocks (total / after convergence)
72	3 (R1,R2,R3)	No	535 / ~4 h	−0.06	21 / 3
73	3 (R1,R2,R3)	Yes	291 / ~2 h 20 m	−0.06	12 / 0
74	3 (R1,R2,R3)	Yes	1117 / ~5 h 50 m	−0.06	51 / 3
89	2 (R1,R3)	Yes	114 / ~52 m	−0.055	0 / 0
60	4 (R1,R2,R3,R4)	No	217 / ~2 h 3 m	−0.06	4 / 1

- *Initial reward:* machine state before the agent's correction.
- *Final reward:* state after the agent's correction.
- *Convergence reward threshold:* $R > -0.08$.
- *Convergence:* any episode solved in <10 steps, sustained thereafter.

Table 2 summarises the performance across all training stages. Pretraining improved convergence in certain cases, but didn't perform well when the surrogate state diverged from the actual machine conditions.

Turn 72 - First Deployment From Scratch

The agent exhibited a rapid reduction of episode length, converging in fewer than 60 episodes. The final reward improved from approximately −0.7 to better than −0.05. After convergence, phases MIF7 and MIF8 were aligned within ±1° of the reference, losses were suppressed well below alarm thresholds, and trim-coil usage was reduced. Interlocks appeared only in early episodes and disappeared after convergence. The learning dynamics and MIF8 phase alignment are shown in Fig. 5 and Fig. 6, respectively.

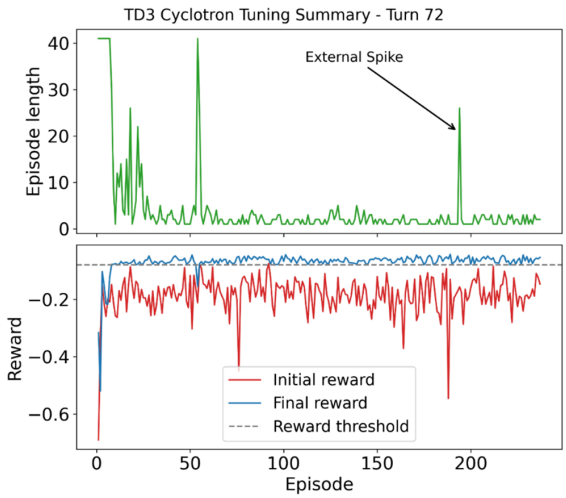

Figure 5: Learning curve and reward for turn 72.

Beam losses and coil efficiency improved markedly during training. As shown in Fig. 7, beam losses decreased by

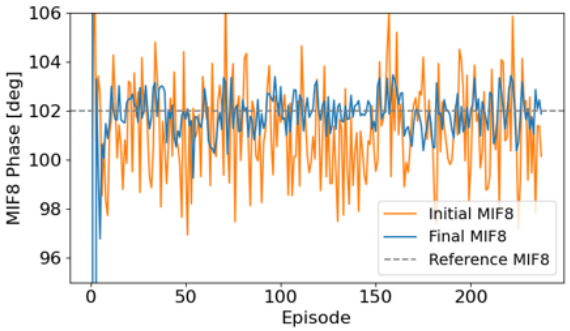

Figure 6: Phase alignment (MIF8) for turn 72.

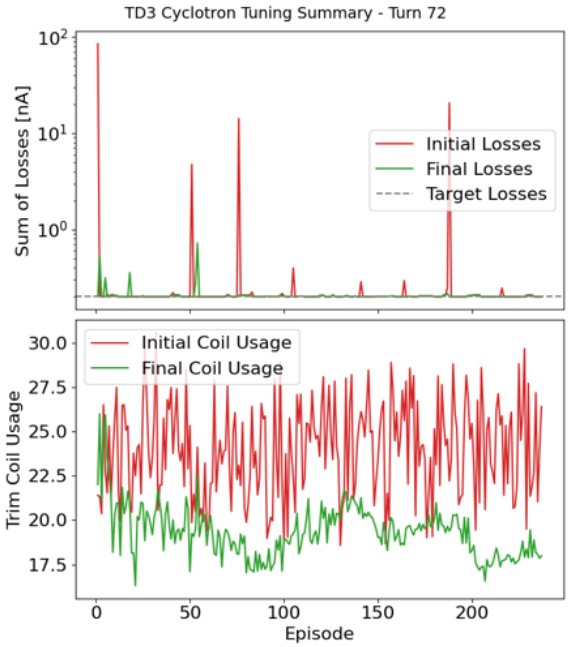

Figure 7: Beam loss suppression and trim coil efficiency for turn 72.

nearly two orders of magnitude, reaching levels far below the 30 nA warning threshold. At the same time, the mean absolute trim-coil currents were reduced, demonstrating that the agent converged to efficient control strategies requiring only minimal corrective action.

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-FRAG002

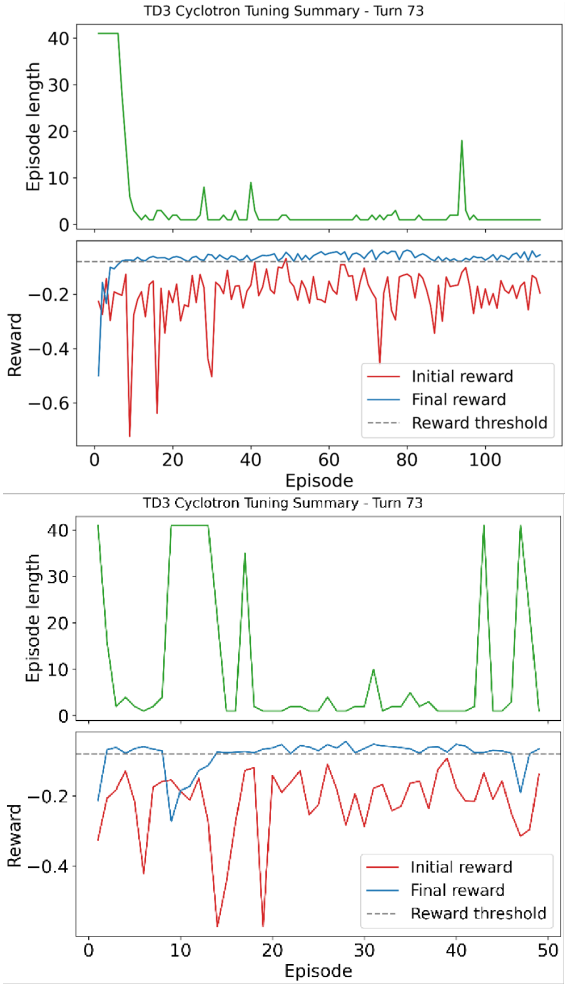

Figure 8: Learning curve and reward for turn 73 comparison: pretrained (top) vs. from scratch (bottom).

Turn 73 - Deployment With Surrogate Pretraining

Relative to Turn 72, convergence was faster, with final rewards stabilising near −0.05. Phase alignment accuracy remained within ±1°, losses stayed below thresholds, and interlocks were eliminated after early episodes. An ablation comparison indicated that the pretrained policy reached stable high reward within a few episodes, whereas training from scratch showed larger variance and slower convergence. The learning curve comparison of both approaches is presented in Fig. 8.

During training at Turn 73, the agent consistently suppressed beam losses, which remained well below the 10–30 nA warning threshold after convergence. The average absolute trim-coil usage also decreased, indicating that stable beam conditions were maintained with less corrective effort compared to the initial episodes.

Turn 74 - Transfer From Turn 73

When transferring the pretrained actor from Turn 73, adaptation required about 50 episodes owing to smaller turn separation. This run had the longest convergence time among all turns. Final phases remained within ±1.5° and interlocks disappeared after initial episodes. Initial beam losses occasionally exceeded 10 nA, reflecting the higher sensitivity of this configuration. However, the agent adapted rapidly, driving losses below 1 nA once convergence was reached. Simultaneously, the mean trim-coil usage steadily decreased, confirming that efficient and stable beam transport was achieved under more demanding conditions.

Turn 89 - Most Degraded Configuration

Despite the most degraded configuration, convergence was the fastest: only 114 timesteps from scratch. Final phases were aligned within ±1.5°, and beam losses decreased by nearly two orders of magnitude while trim-coil usage stabilised. No interlocks were observed. This may reflect reduced action space of the model for this particular turn number. Beam losses were initially close to 10 nA. As training progressed, losses fell by nearly two orders of magnitude. The mean trim-coil currents also stabilised at moderate values, indicating that the agent identified a compact set of efficient corrections sufficient to maintain clean extraction without excessive actuator usage.

Contrary to expectations, surrogate-based pretraining degraded performance. The pretrained agent failed to reach the −0.08 reward threshold consistently and showed large fluctuations. This outcome indicates that, in this regime, pretraining introduced a detrimental bias, underlining the need for close alignment between surrogate data and real machine dynamics.

Turn 60 - Nominal Configuration (All Resonators On)

In the nominal configuration, convergence was achieved in fewer than 20 episodes. Final phases were within ±1.5° of the reference, with stable low losses and minimal interlocks. The beam losses decreased steadily throughout training and remained well below operational thresholds after convergence. Trim-coil usage was consistently low, highlighting that the agent benefited from the larger turn separation to maintain beam quality with minimal corrective intervention.

Key Observations

RL model achieved reproducible tuning with interlocks confined to early exploration, learned policies operated in safe regions. The value of pretraining was turn-dependent-beneficial when the surrogate distribution matched machine conditions (e.g., Turn 73), and detrimental under mismatch (e.g., Turn 89). Policies generally required adaptation across turns, indicating the need for turn-specific or multi-turn training.

CONCLUSION AND FUTURE WORK

This work presents the first successful deployment of reinforcement learning for real-time tuning of a high-intensity cyclotron, using the Injector 2 machine at PSI as an experimental platform. Over the course of two weeks, a TD3-based

ISBN: 978-3-95450-255-4 ISSN: 2226-0358 doi: 10.18429/JACoW-ICALEPCS2025-FRAG002

RL agent was trained and tested to optimize trim coil currents and cavity voltages based on beam phase measurements and loss signals. The results demonstrate that the agent can reliably learn to minimize phase error and beam losses while respecting machine safety constraints, showing performance comparable to manual expert tuning in several operating conditions.

Key contributions include the development of a real-time Gym-compatible control environment interfaced with EPICS, the use of machine-learning-compatible reward functions informed by physics and operational safety, and the successful management of beam interlocks and dynamic feedback signals during learning.

These outcomes provide a solid foundation for extending ML-based control to more complex and higher-power accelerator systems. In particular, future work will focus on the staged deployment of this approach across the full HIPA accelerator chain. This effort aims to not only improve tuning speed and reliability at HIPA, but also to play a key role in the accelerated commissioning of new cyclotron-based systems, such as Transmutex, as well as future high-intensity beamline configurations, including HIMB.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the management of the Paul Scherrer Institut (PSI) for their support and for providing the unique opportunity of nearly two weeks of dedicated beam time for this study. Special thanks are extended to the PSI operator team for their technical assistance, expert advice, and patience throughout the experimental campaign.

The TRANSMUTEX team is acknowledged for their continued backing and for their vision in integrating machine lerning into accelerator control.

Finally, the authors would like to thank the CERN colleagues Verena Kain and Michael Schenk for the fruitful discussions and valuable advices during CERN School of Computing that helped shape this work.

REFERENCES

[1] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods", in *Proc. 35th Int. Conf. Mach. Learn. (ICML)*, Stockholm, Sweden, July 2018, vol. 80, pp. 1587–1596. http://proceedings.mlr.press/v80/fujimoto18a.html

[2] D. Sagan, "Bmad: A relativistic charged particle simulation library", *Nucl. Instrum. Meth. Phys. Res. Sect. A*, vol. 558, pp. 356–359, 2006. doi:10.1016/j.nima.2005.11.001

[3] V. Kain *et al.*, "Sample-efficient reinforcement learning for CERN accelerator control", *Phys. Rev. Accel. Beams*, vol. 23, p. 124801, 2020. doi:10.1103/PhysRevAccelBeams.23.124801