ONLINE REINFORCEMENT LEARNING FOR STRIPPER FOIL AGING COMPENSATION AT THE CERN LOW ENERGY ION RING

B. Rodriguez Mateos*, T. Argyropoulos, V. Kain, A. Lu, A. Menor de Onate,
M. Schenk, M. Slupecki†, European Organization for Nuclear Research, Geneva, Switzerland

Abstract

Stripper foil degradation at the CERN Low Energy Ion Ring (LEIR) poses a significant challenge for beam operations. As the heavy ion beam passes through the stripper foil at the end of the injecting linac, the foil degrades over time, altering the beam energy distribution and reducing the achievable accumulated intensity in the ring. Addressing this operational limitation using traditional control approaches is challenging due to the complex, multi-dimensional nature of the multi-turn injection process. This paper presents a reinforcement learning-based controller to compensate for foil degradation and maintain ring performance. The controller observes longitudinal Schottky spectra encodings and time-of-flight measurements from the linac to adjust the ramping and debunching cavity phases, and electron cooler gun and orbit bump in real time. We demonstrate that pre-training the agent in a data driven surrogate model significantly improves both controller performance and sample efficiency during deployment.

INTRODUCTION

The Low Energy Ion Ring (LEIR) at CERN accumulates lead-ion beams for the Large Hadron Collider (LHC) and fixed target experiments by stacking up to eight consecutive multi-turn injections from Linac3. Each 200 µs-pulse is injected at 200 ms intervals through a six dimensional phase-space painting scheme combining a collapsing horizontal orbit bump with a linearly ramped mean momentum in a high-dispersion region. Between injections, electron cooling compresses and drags the circulating beam longitudinally to free momentum space for the next pulse. After accumulation, the coasting beam is captured into bunches and accelerated to 72 MeV/u for extraction toward the Proton Synchrotron. The quality of this process depends critically on the precise matching between the energy distribution delivered by Linac3 and the LEIR electron-cooler parameters.

A well-known operational limitation is the progressive degradation of the carbon stripper foil, installed at the end of Linac3, used to strip Pb^{29+} ions to Pb^{54+}. As the foil ages under beam bombardment, it causes a gradual shift in the injected momentum distribution, with drift rates in the order of 0.06 ‰ per day [1]. Because electron cooling is energy dependent, even small shifts reduce injection efficiency; in extreme cases, accumulated intensity drops by more than a factor of two, requiring foil replacement approximately every two weeks [1]. Historically, these drifts have been corrected manually by adjusting the last Linac3 accelerating tank and retuning the ramping and debunching RF cavities, a procedure that is time consuming, reactive, and dependent on operator availability. Together, these two latter cavities respectively control the mean energy and the shape of its spread along the 200 µs-pulse.

To move toward automated compensation, Madysa *et al.* [2] constructed a data-driven surrogate model of the LEIR injection dynamics: a β-variational autoencoder (β-VAE) compressed the high-dimensional Schottky spectra into a compact latent representation, while multi-layer perceptrons learned both an intensity model and a dynamics model predicting state transitions as a function of parameter changes. Several RL algorithms were trained offline on this surrogate and shown to recover nominal intensity in fewer than 20 steps when the machine was moderately detuned, establishing the feasibility of RL-based optimization for LEIR. However, limited surrogate generalization prevented online deployment.

Building on this, we presented the first operational results of ML-based beam intensity optimization at LEIR [3], deploying Bayesian Optimization with Upper Confidence Bound (UCB) and Log-Noisy Expected Improvement (Log-NEI) acquisition functions to maximize injection efficiency and minimize RF capture losses, consistently exceeding average operational performance. In parallel, the data-driven surrogate model was improved through an enhanced transition model loss incorporating symmetry and fixed-point constraints to enforce physical consistency and mitigate error accumulation during multi-step rollouts. The offline-trained RL agents reached near-target injection efficiency in approximately three cycles, though performance did not yet meet the operational threshold. In this paper, we present the online deployment of an RL agent that autonomously compensates for stripper-foil aging in Linac3 to maintain optimal injection efficiency into LEIR with close to the full operational range of control parameters, and demonstrate that automated RL-based compensation can sustain beam intensity at nominal levels over extended periods without manual intervention.

PROBLEM FORMULATION

Observation Space

The design of an informative and compact observation space is central to RL agent performance. The observation must capture sufficient information about the machine state to allow the agent to infer corrective actions, while remaining low-dimensional enough for efficient training on limited operational data. The primary diagnostic for the longitudinal beam dynamics during the LEIR injection process is the lon-

* borja.rodriguez.mateos@cern.ch
† also at Department of Communications and Computer Engineering, University of Malta

gitudinal Schottky spectrum, which encodes the momentum distribution of the circulating beam as a function of cycle time throughout the injection plateau. For each machine cycle, the Schottky spectra acquired at successive injection steps are concatenated into a single two-dimensional array spanning frequency and cycle time. This high-dimensional observation is then compressed using a β-VAE [4] into a latent representation of dimension $d_z = 9$. This dimensionality was selected after a systematic study that evaluated reconstruction quality as a function of latent dimension, seeking the lowest dimensionality that preserved the salient spectral features (in particular the mean energy, energy spread, and cooling dynamics) while ensuring that the latent variables remained approximately uncorrelated. The resulting nine-dimensional latent vector provides the agent with a rich summary of the beam state at each cycle.

A key development improving the stability of offline surrogate model training was augmenting the observation space with time-of-flight (ToF) measurements from the Linac3-to-LEIR transfer line. These cycle-bound measurements exploit pairs of capacitive Beam Position Monitors in the transfer line: the reduction in mean energy along the pulse translates into a measurable phase difference $\Delta\varphi$ between the two monitors at the Linac3 bunching frequency, from which the beam velocity can be inferred. For the first injected pulse, a linear fit to the measured $\Delta\varphi(t)$ profile yields two scalars, intercept and slope, providing a compact characterization of the mean beam energy and energy sweep at injection. The $\Delta\varphi$ intercept serves as a direct proxy for stripper foil aging: as the carbon foil degrades under beam bombardment, the mean energy of the stripped Pb^{54+} beam drifts, manifested as a shift in the intercept over days to weeks. Including this observable gives the RL policy explicit access to the dominant source of performance degradation, rather than requiring indirect inference from the Schottky spectra alone, where the foil aging signal is convolved with electron-cooling dynamics and other parameter variations. The full observation is the concatenation of the nine-dimensional β-VAE latent mean vector $\mathbf{z} \in \mathbb{R}^9$ and the two ToF-derived scalars (intercept and slope of $\Delta\varphi$), yielding dimension 11. The ToF features proved essential for stable surrogate model training: since the $\Delta\varphi$ parameters vary slowly and monotonically with foil age, they disambiguate machine states that would appear similar in the Schottky latent space but correspond to different foil lifetimes. This enabled effective reuse of historical data collected across multiple foil installations, as the ToF features provide the context needed to align data from different epochs into a coherent training set.

Action Space

As shown in [1], the energy distribution delivered by Linac3 can be restored after stripper-foil-induced drifts by adjusting the phases of the debunching and ramping RF cavities at the linac exit. The debunching cavity start and end phases control the energy sweep along the 200 µs-pulse, while the ramping cavity phase sets the mean energy offset. Together, these three parameters bring the longitudinal distribution back into the LEIR injection acceptance window. However, correcting the energy distribution alone during a multi-turn injection is insufficient: the electron-cooling process must also be matched to the actual beam energy by adjusting the start and end values of the electron gun voltage program during the injection plateau. In addition, spatial overlap between ion and electron beams at the cooler section is controlled by four corrector magnet settings defining the horizontal and vertical positions and angles of an orbit bump at this location; a mismatch reduces cooling rate and degrades stacking efficiency. These nine parameters (three RF cavity phases, two electron gun voltages, and four cooler bump settings) constitute the action space, with operational bounds listed in Table 1. For the cooler bump parameters, the full safe operational range is used; for the debunching cavity phases, the range extends 20 degrees above and below the zero-crossing point. The agent interacts with the machine on a cycle-by-cycle basis, proposing adjustments $\delta\mathbf{p}$ to these nine parameters based on the current observation. The objective it maximizes is the injection efficiency, $r = \frac{i_{\text{injected}}}{\sum_{j=1}^{8} i_j^{L3}}$, where i_{injected} is the beam intensity in the LEIR ring at the peak of the injection process and i_j^{L3} is the intensity of the j-th pulse measured at the end of the Linac3-to-LEIR transfer line. By normalizing to the total delivered charge, the reward signal isolates the effect of the controlled parameters on the accumulation process, reducing sensitivity to source intensity fluctuations that are beyond the agent's control.

Table 1: Operational Ranges of Action Parameter Values

	Min	**Max**
Electron gun voltage start [V]	2660.0	2695.0
Electron gun voltage end [V]	2620.0	2695.0
Horizontal cooler bump position [mm]	-10.0	16.5
Horizontal cooler bump angle [mrad]	-3.5	3.5
Vertical cooler bump position [mm]	-0.75	5.0
Vertical cooler bump angle [mrad]	0.26	1.65

OPERATIONAL RESULTS

The RL agent was deployed online at LEIR during the 2025 lead-ion run in a cycle-by-cycle closed loop. Two configurations were tested: an agent trained from scratch on the machine (online), and an agent pretrained on the data-driven surrogate model before continued online learning (pretrained). Both approaches used the Soft Actor Critic (SAC) algorithm [5]. The Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm [6] was discarded due to policy collapse, the policy converging to extremal actions and ceasing to explore, consistent with the failure modes identified by Bjorck et al. [7]. The on policy control algorithm, Proximal Policy Optimization (PPO) [8], is more stable during training, but proved insufficiently sample-efficient given the 3.6 s cycle time. SAC was the only algorithm that captured the stochasticity of the LEIR injection process while maintaining adequate sample efficiency. To ensure stable training, several stabilization techniques were adopted

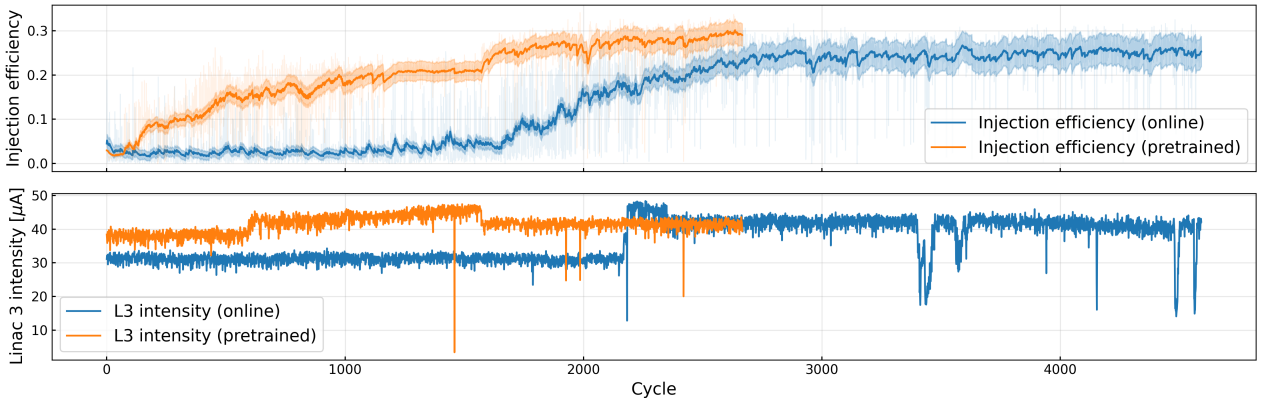

Figure 1: *Top:* Evolution of injection efficiency from Linac3 to LEIR during online training and fine-tuning. Training curves are smoothed with Welford's online algorithm with 0.95 CI [9]. *Bottom:* Intensity measured with a beam current transformer in the Linac3-to-LEIR transfer line.

following [7]. In particular, learning rate warm-up was used during early training to prevent large weight updates that trigger policy collapse when the replay buffer is sparsely populated. Without these measures, the agent's performance was prone to abrupt degradation after initial improvement, mirroring the policy collapse documented in [7].

The results are summarized in Fig. 1. The upper panel shows injection efficiency versus machine cycle index for both agents, averaged over multiple runs. Both start from a deliberately detuned efficiency of approximately 0.03. The pretrained agent improves significantly earlier, reaching above 0.20 within roughly 500 cycles (approximately 2 hours with 1 out of 2 cycles), whereas the online agent requires approximately 1500 cycles. After convergence, both stabilize between 0.20 and 0.30, with the pretrained agent consistently at the upper end. The lower panel shows that injection efficiency was reliably maximized despite an unstable Linac3 source, validating this normalized metric as a reward decoupled from upstream fluctuations. The advantage of pre-training is twofold. First, it improves sample efficiency by roughly a factor of two, reducing the number of cycles needed to reach operational performance from over a thousand to a few hundred. Second, the pretrained agent achieves a higher asymptotic efficiency, suggesting the surrogate model provides a useful inductive bias guiding the policy toward a better region of parameter space. These results demonstrate that RL holds promise for an autonomous controller for compensating stripper foil aging at LEIR, maintaining injection performance without manual intervention. Further long term evaluations are needed in order to confirm its effectiveness. To further characterize the pretrained agent, Fig. 2 shows repeated evaluation episodes in which machine parameters are randomized to a detuned configuration and the agent recovers operational efficiency. The agent consistently recovers injection efficiency from near-zero values to within the operational target (0.25–0.30) in fewer than ten cycles in the majority of episodes, corresponding to less than 36 seconds of beam time. This rapid recovery is particularly relevant for operational scenarios

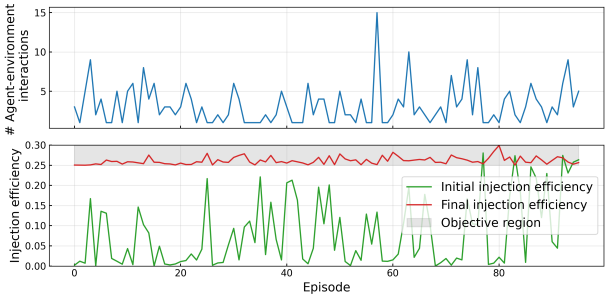

Figure 2: Evaluation of the pretrained agent over repeated episodes starting from randomized machine configurations. *Top:* Episode length in number of iterations. *Bottom:* Initial (green) and final (red) injection efficiency for each episode. The shaded band indicates the operational target region.

such as beam commissioning, foil exchanges or machine restarts.

LIMITATIONS AND OUTLOOK

Despite the successful online deployment, two main limitations remain. First, the sample efficiency, while adequate for proof-of-concept operation, is not yet sufficient for routine use during tightly scheduled physics runs; even with surrogate-model pre-training, substantial fine-tuning is needed before reaching operational performance. Second, sim-to-real transfer is hampered by the high noise floor of the injection process ($\sigma \approx 0.0133$ relative to a maximal efficiency of ~0.3), driven primarily by shot-by-shot variations in Linac3 delivery: ion source fluctuations, transport jitter, and pulse-to-pulse energy spread introduce cycle-to-cycle variability not attributable to the agent's actions. This limits the surrogate model's ability to learn accurate dynamics, degrading the pretrained policy quality and requiring extensive online fine-tuning. Future work will focus on improving both surrogate model fidelity and online sample efficiency as well as long time evaluations of the control algorithm.

REFERENCES

[1] S. Hirlaender *et al.*, "Energy Dependence of the Reproducibility and Injection Efficiency of the LINAC3-LEIR Complex", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 3188–3191. doi:10.18429/JACoW-IPAC2019-WEPTS040

[2] N. Madysa, R. Alemany-Fernandez, N. Biancacci, B. Goddard, V. Kain, and F. M. Velotti, "Automated Intensity Optimisation Using Reinforcement Learning at LEIR", in *Proc. IPAC'22*, Bangkok, Thailand, Jun. 2022, pp. 941–944. doi:10.18429/JACoW-IPAC2022-TUPOST040

[3] B. Rodriguez Mateos, T. Argyropoulos, F. Carlier, V. Kain, M. Schenk, and M. Slupecki, "Operational results of data-driven automated intensity optimization at CERN's LEIR", in *Proc. IPAC'25*, Taipei, Taiwan, Jun. 2025, pp. 2913–2916. doi:10.18429/JACoW-IPAC2025-THPM109

[4] C. P. Burgess *et al.*, "Understanding disentangling in β-VAE", arXiv preprint, 2018. doi:doi:10.48550/arXiv.1804.03599

[5] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor", in *Proc. 35th Int. Conf. on Machine Learning (ICML'18)*, Stockholm, Sweden, Jul. 2018, vol. 80, pp. 1861–1870. doi:10.48550/arXiv.1801.01290

[6] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods", in *Proc. 35th Int. Conf. on Machine Learning (ICML'18)*, Stockholm, Sweden, Jul. 2018, vol. 80, pp. 1587–1596. doi:10.48550/arXiv.1802.09477

[7] J. Bjorck, C. P. Gomes, and K. Q. Weinberger, "Is High Variance Unavoidable in RL? A Case Study in Continuous Control", in *Proc. 10th Int. Conf. on Learning Representations (ICLR'22)*, Held online, Apr. 2022. doi:10.48550/arXiv.2110.11222

[8] J. Schulman *et al.*, "Proximal Policy Optimization Algorithms", arXiv preprint, 2017. doi:doi:10.48550/arXiv.1707.06347

[9] A. A. Efanov, S. A. Ivliev, and A. G. Shagraev, "Welford's algorithm for weighted statistics", in *Proc. 3rd International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE)*, Moscow, Russia, Mar. 2021, pp. 1-5. doi:10.1109/REEPE51337.2021.9387973