ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPM032

TOWARD AUTONOMOUS CONTROL: REINFORCEMENT LEARNING FOR IMPROVING ACCELERATOR PERFORMANCE

A. Gilardi[1,†] A. Aksoy[1], L. Bonnard[1], R. Corsini[1], W. Farabolini[1], L.A. Foldesi[1,2], O. Franek[1,3], D. Gamba[1], E. Granados[1], V. Kain[1], P. Korysko[1], A. Malyzhenkov[1], A. Mostacci[4], B. Rodriguez Mateos[1], A. Petersson[1,5], A. Pollastro[6], V. Rieker[1,7], M. Schenk[1], K. N. Sjøbæk[7], G. Tangari[1,4], L.M. Wroe[1]
[1]European Organization for Nuclear Research CERN, Geneva, Switzerland
[2]University of Zurich, Zurich, Switzerland
[3]Czech Technical University in Prague, Prague, Czech Republic
[4]Sapienza University of Rome, Rome, Italy
[5]Lund University, Lund, Sweden
[6]University of Naples Federico II, Naples, Italy
[7]University of Oslo, Oslo, Norway

Abstract

Particle accelerators like CLEAR (CERN Linear Electron Accelerator for Research) are essential tools for advancing various scientific fields. Automating their operation to ensure stability and reproducibility is crucial for future large-scale projects. This paper explores the first steps toward autonomous control of the CLEAR beamline, focusing initially on beam steering and advancing to complex tasks like quadrupole alignment, vital for operational stability. Reinforcement Learning (RL) agents that adapt in real-time via beam screens measurements were trained and tested. The approach was optimized for sampling efficiency, and to avoid the high cost and invasiveness of traditional data collection schemes in accelerator environments. The method enables single-shot optimization for real operations, reducing the need for manual intervention. Results show that after a few hours of training, the method is effective for single-step corrections all the way to the end of the CLEAR beamline. The already proven advantages of this technique is driving further development by the CLEAR research team.

INTRODUCTION

Particle accelerators are complex machines supporting applications from fundamental physics to industry and medicine [1]. Thousands are currently operational worldwide, with about half employed for industrial applications, such as material processing and sterilization, and about a third used in hospitals [2]. These facilities are critical to modern science and technology, driving advances in both research and healthcare.

Maintaining optimal accelerator performance presents significant challenges. Beam steering and alignment must be achieved with high precision to ensure stable operation, but manual tuning by experts becomes increasingly time-consuming and less reproducible as systems grow in scale and complexity.

† email: antonio.gilardi@cern.ch

Future facilities, employing high-energy LINACs in high-energy colliders, will demand unprecedented beam quality and stability, making automation essential. Even small drifts or imperfect corrections can severely degrade beam performance. To address these challenges, intelligent, automated control systems are being actively developed. Machine Learning (ML) techniques, particularly Reinforcement Learning (RL), are being explored to improve efficiency, precision, and reproducibility in beam tuning [3, 4]. This work presents an RL-based autonomous control system for beam steering and orbit alignment at the CERN Linear Electron Accelerator for Research (CLEAR) [5].

STATE OF THE ART

Beam steering is a control task that can often be well approximated using a linear lattice model [6]. This allows for the application of analytical algorithms and optimization techniques. While the first-order behavior of an ideal beamline can be simulated with relative ease, accurately modeling lattice imperfections remains a significant challenge. Such imperfections are difficult to measure accurately and even harder to replicate in simulation, which limits the effectiveness of purely model-based correction schemes. Consequently, analytical orbit correction methods may fail to provide satisfactory single-shot corrections due to discrepancies between the model and the real measurements.

A widely used traditional method for orbit correction is one-to-one steering, in which the beam is sequentially recentered at each beam position monitor using the preceding upstream corrector magnet [7–11]. One-to-one steering is valued for its simplicity and does not require a detailed beam dynamics model, making it suitable for local orbit correction. However, this approach can produce non-optimal or "zig-zag" beam orbits that deviate from the ideal trajectory. Furthermore, local correction methods, typically, use higher corrector strength and results in higher order distortions which effectively spoil beam quality.

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPM032

Modern requirements, such as achieving single-shot optimization across the whole accelerator, call for more holistic approaches. Rather than correcting locally, an optimal decision must consider the global state of the beam across the entire machine. This need for global, coordinated correction motivates the exploration of advanced, data-driven control strategies.

In this context, RL has emerged as a promising candidate for autonomous beam control—especially in large-scale accelerators, where the high number of variables and their nonlinear effects lead to an explosion of the parameter space. RL is a branch of machine learning in which an agent learns optimal control policies through interaction with its environment and feedback in the form of rewards, rather than explicit programming. Recent advancements in deep RL have demonstrated remarkable performance in mastering high-dimensional control problems across various fields, including in the particle accelerator community [12–15]. Preliminary studies suggest that RL-based methods can manage the high-dimensional and dynamic nature of accelerator optimization problems, enabling more autonomous and robust operation. Given RL's demonstrated ability to address complexity and uncertainty, it is a natural fit for the beam steering problem, where traditional approaches may fall short.

CERN LINEAR ELECTRON ACCELERATOR FOR RESEARCH

CLEAR is a 200 MeV electron LINAC designed as a flexible test facility for accelerator R&D [16]. Its modular layout supports a wide range of beam energies, intensities, and configurations, enabling a diverse experimental programme [17]. Due to this versatility, identical beam conditions are rarely repeated, rendering fixed reference orbits and precomputed steering solutions impractical. Instead, beam alignment is typically performed manually for each setup, with operators adjusting magnet settings based on feedback from beam position monitors until acceptable orbit quality is achieved. This manual tuning process is time-consuming, requires expert knowledge, and may not always yield optimal alignment—particularly under evolving experimental conditions. In addition, unlike large accelerator complexes, CLEAR lacks a fully instrumented continuous orbit feedback system.

CLEAR is a suitable place to test an RL-based solution to autonomously steer the beam in real time, adapting dynamically to the varying machine state. On the other hand, a significant limitation of the facility is the sparse availability of non-invasive, real-time Beam Position Monitor (BPM) coverage along the full beamline [17]. As a result, destructive screen monitors, such as Yttrium Aluminum Garnet (YAG) screens or optical transition radiation (OTR) foils, had to be used to capture transverse beam profiles and measure the beam position [18].

METHODOLOGY

Beam trajectory control in accelerators is achieved using dipole steering magnets to compensate for misalignments in focusing elements and beamline components. A good control of the beam trajectory is essential to preserve emittance and maintain beam quality [6, 19]. An RL-based beam steering system was developed using a custom Gymnasium environment [20], interfaced with the Stable-Baselines3 (SB3) library [21]. The agent operates in a continuous action space, with each action representing an adjustment to the normalized current of a steering magnet. Observations consist of measured beam positions obtained from the screens. The agent interacts with the environment in a closed loop: at each iteration, it computes magnet adjustments, observes the outcome on beam position, and updates its policy based on the reward.

The Twin Delayed Deep Deterministic Policy Gradient (TD3) algorithm [22] was selected for its robustness and improved training stability, benefiting from clipped double Q-learning, target policy smoothing, and delayed policy updates [23].

Hardware control is handled via CERN's JAPC and PyJAPC interfaces [24], enabling real-time control of magnet currents and diagnostic devices. All current commands are clipped within predefined safety bounds to protect the equipment.

Due to their destructive effect on the beam, screens are used in a sequential measurement strategy. All required screens are inserted at the start of an episode. At each time step, the beam strikes the most upstream screen, which is then retracted before the subsequent pulse. This process continues until all screens are retracted, allowing one beam profile per pulse without insertion delays. This sequential imaging strategy is devised for minimizing the impact from the destructive nature of screens.

At each step, the screen image is processed to calculate the position of the beam centroid. The centroid of the beam intensity distribution is computed to infer beam position, and to convert from pixel coordinates to real-space positions, accounting for camera geometry, a homography-based transformation is applied [25]. The result is that for each screen insertion, one obtains an estimate of the beam's transverse offset at that location. Once all the screens are extracted, a scalar reward is computed based on the deviation from the target orbit on each screen, encouraging centering and minimizing positional error.

Each interaction with the environment is stored by the agent as a transition tuple (s, a, r, s') in a replay buffer, where s is the current state (i.e. the observed beam positions), a is the action taken (i.e. the applied magnet current adjustments), r is the received reward (quantifying beam alignment accuracy), and s' is the resulting next state (the updated beam position observation). During training, mini-batches of these transitions are sampled from the buffer, and the policy is updated via gradient descent to minimize the temporal-difference error between the predicted and target Q-values. The agent's goal is to improve its policy so that future actions will yield higher rewards.

Episodes terminate when all screens have been used or an early stopping condition is met. Upon reset, screens

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPM032

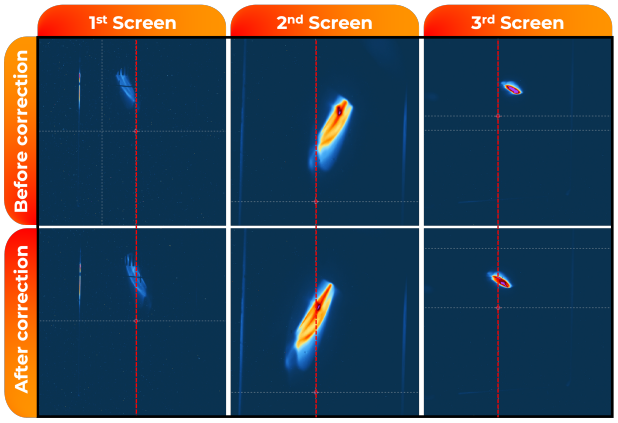

Figure 1: Visual comparison of the beam orbit before (top) and after (bottom) correction for the horizontal plane. Significant misalignment relative to the reference axis (red dashed lines) is observed prior to correction, while improved centering achieved in single-shot correction.

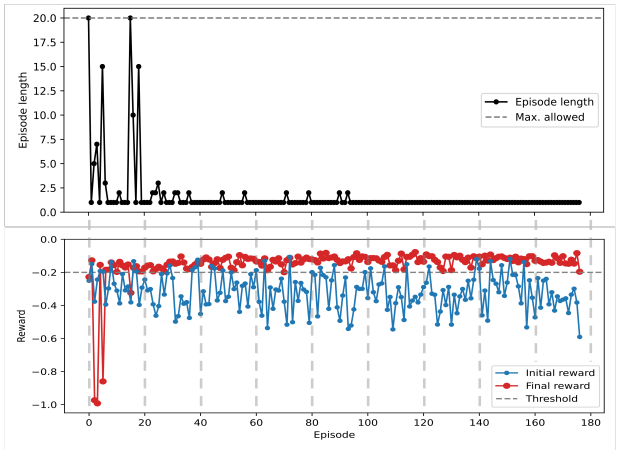

Figure 2: Training progress of the RL agent. Episode length (top) decreases as the agent learns to complete the steering task more efficiently. Reward (bottom) with final values (red) consistently exceeding the defined threshold (dashed line).

are reinserted and magnet settings randomized within safe limits. The agent continues training over multiple episodes, refining its steering policy iteratively. This approach was implemented in a single axis of the transverse plane but can be extended to both axes with minimal modifications.

RESULTS AND OUTLOOK

The results in Fig. 1 demonstrate the effectiveness of the proposed RL-based correction method for the horizontal plane. Visual comparison shows a marked improvement in beam alignment relative to the reference axis, confirming the impact of the developed approach.

The training performance is summarized in Fig. 2. A rapid decline in episode lengths (top) and a steady improvement in final reward values (bottom) illustrate fast policy convergence and stable learning behavior. The reported rewards correspond to those obtained from the initial beam orbit (following random corrector initialization) and from the corrected orbit after the agent's action.

After only a few tens of episodes, the agent learned to reliably correct random initial orbits in a single step, surpassing the reward threshold. This demonstrates not only the efficiency of training but also the agent's ability to generalize to varying machine conditions during deployment.

Training convergence was typically achieved within approximately two hours of run time, validating the robustness and practicality of the method for routine accelerator operation. An autonomous correction of beam trajectories was achieved within a few pulses, while respecting machine safety constraints and instrumentation limits. This resulted in a reproducible, software-in-the-loop system that progressively refined beam alignment over successive iterations, outperforming, for instance in speed, manual tuning and static correction algorithms. The deployment of an RL controller led to improved consistency and speed in orbit recovery, even under varying beam parameter conditions, thus ensuring reproducible operating conditions with minimal downtime.

Beyond orbit steering, similar RL-based strategies could be adapted for targeting other beam properties, such as minimizing beam size, controlling beam shape, or maximizing transmission efficiency. Promising preliminary results have also been observed when applying similar techniques to quadrupole alignment tasks.

Despite the encouraging results, some limitations were identified. The primary diagnostic used—destructive screen-based imaging—required mechanical insertion and removal, introducing latency and disrupting normal beam operation. Consequently, real-time, continuous feedback was not possible during routine runs, as the agent could only train and act in dedicated measurement modes. Future work will focus on integrating non-invasive diagnostics, such as BPMs, into the RL environment. Access to fast, non-destructive feedback would enable in situ training during standard operation, significantly accelerating learning and allowing real-time corrections without beam interruption. Additionally, future efforts will evaluate the center position of the screen using an in-vacuum reference Structured Laser Beam, providing a more stable and precise alignment baseline [26].

ACKNOWLEDGEMENTS

Thanks to the CERN School of Computing for offering a valuable learning experience. The authors also acknowledge financial support from the PNRR MUR project PE0000013-FAIR and from the European Union's Horizon Europe Research and Innovation programme under Grant Agreement No. 101057511 (EURO-LABS).

ISBN: 978-3-95450-248-6 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2025-THPM032

REFERENCES

[1] K. Peach, P. Wilson, and B. Jones, "Accelerator science in medical physics", *Br. J. Radiol.*, vol. 84, no. special_issue_1, pp. S4–S10, 2011.

[2] The Nobel Prize, "Accelerators and Nobel Laureates", https://www.nobelprize.org/prizes/themes/accelerators-and-nobel-laureates

[3] V. Kain *et al.*, "Sample-efficient reinforcement learning for CERN accelerator control", *Phys. Rev. Accel. Beams*, vol. 23, no. 12, p. 124801, 2020. doi:10.1103/PhysRevAccelBeams.23.124801

[4] I. Kante, "Machine Learning for beamline steering," arXiv preprint arXiv:2311.07519, Nov. 2023. [Online]. Available: https://arxiv.org/abs/2311.07519

[5] R. Corsini *et al.*, "The future of the CLEAR facility: consolidation, ongoing upgrades and its evolution towards future e-facilities at CERN", presented at the IPAC'25, Taipei, Taiwan, Jun. 2025, paper TUPM027, this conference.

[6] H. Wiedemann, *Particle Accelerator Physics*, Cham, Switzerland: Springer Cham, 2015.

[7] M. G. Minty and F. Zimmermann, *Measurement and Control of Charged Particle Beams*, Heidelberg , Germany: Springer Berlin, 2003.

[8] Y. Zhao and A. Latina, "Optimization of the compact linear collider rings-to-main-linac at 380 GeV", *Phys. Rev. Accel. Beams*, vol. 28, no. 2, p. 021003, Feb. 2025. doi:10.1103/PhysRevAccelBeams.28.021003

[9] N. Blaskovic Kraljevic and D. Schulte, "Beam-Based Beamline Element Alignment for the Main Linac of the 380 GeV Stage of CLIC", in *Proc. IPAC'19*, Melbourne, Australia, May 2019, pp. 465–467. doi:10.18429/JACoW-IPAC2019-MOPMP018

[10] M. Aiba and M. B?Âge, "Beam-based Alignment of an X-FEL Undulator Section Utilizing Corrector Pattern", in *Proc. FEL'12*, Nara, Japan, Aug. 2012, paper TUPD27, pp. 293–296.

[11] P. Tenenbaum, L. Hendrickson, and T. O. Raubenheimer, "Developments in Beam-Based Alignment and Steering of the Next Linear Collider Main Linac", in *Proc. PAC'01*, Chicago, IL, USA, Jun. 2001, paper FPAH069, pp. 3837–3839.

[12] A. Ibrahim, D. Derkach, A. Petrenko, F. Ratnikov, and M. Kaledin, "Optimisation of the accelerator control by reinforcement learning: a simulation-based approach", *arXiv:2503.09665*, Mar. 2025. doi:10.48550/arXiv.2503.09665

[13] X. Chen, Y. Jia, X. Qi, Z. Wang, and Y. He, "Orbit correction based on improved reinforcement learning algorithm", *Phys. Rev. Accel. Beams*, vol. 26, no. 4, p. 044601, Apr. 2023. doi:10.1103/PhysRevAccelBeams.26.044601

[14] X. Pang, S. Thulasidasan, and L. Rybarcyk, "Autonomous control of a particle accelerator using deep reinforcement learning", *arXiv:2010.08141*, Dec. 2020. doi:10.48550/arXiv.2010.08141

[15] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "A brief survey of deep reinforcement learning", *arXiv:1708.05866*, Sep. 2017. doi:10.1109/MSP.2017.2743240

[16] D. Gamba *et al.*, "The CLEAR user facility at CERN", *Nucl. Instrum. Methods Phys. Res., Sect. A*, vol. 909, pp. 480–483, Nov. 2018. doi:10.1016/j.nima.2017.11.080

[17] R. Corsini, W. Farabolini, P. Korysko, A. Malyzhenkov, V. Rieker, and K. N. Sjobak, "Status of the CLEAR User Facility at CERN and its Experiments", in *Proc. LINAC'22*, Liverpool, UK, Aug.-Sep. 2022, pp. 753–757. doi:10.18429/JACoW-LINAC2022-THPOPA05

[18] V. Schlott, "Free-Electron Laser Beam Instrumentation and Diagnostics", in Synchrotron light sources and free-electron lasers: accelerator physics, instrumentation and science applications, Cham, Switzerland: Springer Cham, 2020, pp. 779–826.

[19] P. Tenenbaum and T. O. Raubenheimer, "Resolution and systematic limitations in beam-based alignment", *Phys. Rev. ST Accel. Beams*, vol. 3, no. 5, p. 052801, May 2000. doi:10.1103/PhysRevSTAB.3.052801

[20] L. Kennedy, "Deep reinforcement learning-based longitudinal optimization and control of the proton synchrotron booster at CERN", MS Thesis, Department of Physics, University of Strathclyde, Glasgow, Scotland, 2023. https://cds.cern.ch/record/2930268

[21] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: reliable reinforcement learning implementations", *J. Mach. Learn. Res.*, vol. 22, no. 268, pp. 1–8, Now. 2021.

[22] OpenAI TD3 (Twin Delayed Deep Deterministic policy gradient), https://spinningup.openai.com/en/latest/algorithms/td3.html

[23] Stable-Baselines3, "TD3 (Twin Delayed Deep Deterministic policy gradient)", https://stable-baselines3.readthedocs.io/en/master/modules/td3.html

[24] Pyjapc: Python to FESA/LSA/INCA Interface via JAPC, https://gitlab.cern.ch/scripting-tools/pyjapc

[25] R. Hartley, *Multiple View Geometry in Computer Vision*, Cambridge, UK: Cambridge University Press, Jan. 2011.

[26] K. Polak, J.-C. Gayde, and M. Sulc, "Structured laser beam in non-homogeneous environment", CERN, Geneva, Tech. Rep. CERN-BE-2023-014, 2022. https://cds.cern.ch/record/2849070