ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-TUPS61

PRELIMINARY RESULTS ON THE REINFORCEMENT LEARNING-BASED CONTROL OF THE MICROBUNCHING INSTABILITY

L. Scomparin*, J. Becker, E. Blomley, E. Bründermann, M. Caselle, T. Dritschler, A. Kopmann, A. Mochihashi, A.-S. Müller, A. Santamaria Garcia, M. Schuh, J. L. Steinmann, M. Weber, C. Xu
Karlsruhe Institute of Technology, Karlsruhe, Germany

Abstract

Reinforcement learning (RL) is applied to control the microbunching instability (MBI) in synchrotron light sources. Here the interaction of an electron bunch with its emitted coherent synchrotron radiation leads to complex non-linear dynamics and pronounced fluctuations. Addressing the control of intricate dynamics necessitates meeting stringent microsecond-level real-time constraints. To achieve this, RL algorithms must be deployed on a high-performance electronics platform. The KINGFISHER system, utilizing the AMD-Xilinx Versal family of heterogeneous computing devices, has been specifically designed at Karlsruhe Institute of Technology (KIT) to tackle these demanding conditions. The system implements an experience accumulator architecture to perform online learning purely through interaction with the accelerator while still satisfying strong real-time constraints. The preliminary results of this innovative control paradigm at the Karlsruhe research accelerator (KARA) will be presented. Notably, this represents the first experimental attempt to control the MBI with RL using online training only.

INTRODUCTION

Electron storage rings are a possible source for the production of bright, high repetition rate, terahertz-range radiation [1]. Strong coherent emission is possible when microstructures smaller than the wavelength of the emitted radiation appear in the bunch charge distribution. This can be achieved through the microbunching instability (MBI), where finger-like microstructures in the longitudinal phase-space interact with their own emitted coherent synchrotron radiation (CSR). The microstructure rotation in the phase-space due to synchrotron motion, produces CSR fluctuations at a multiple of the synchrotron frequency, called the bursting or "finger" frequency[2]. When the self-interaction is strong enough the phase-space distribution is periodically blown-up, dispersing the microstructures and thus stopping the CSR. The instability is triggered again when the synchrotron radiation damping increases the charge density [2–4].

At the Karlsruhe research accelerator (KARA), this phenomena is observed in the short bunch operation mode with low momentum compaction factor α_c at an energy of 1.3 GeV, where the bunch length is reduced compared to the optics at 2.5 GeV for regular synchrotron operation and photon science experiments. The bursting and slow-bursting frequencies have a timescale of ≈ 30 kHz and of ≈ 200 Hz respectively, albeit being strongly dependent on the specific settings of accelerating voltage and optics [2]. These instabilities hinder the usage of the CSR for material science or medical applications, where the radiation output typically needs to be constant for the detection systems.

A candidate technique to control these instability is reinforcement learning (RL), a class of machine learning (ML) algorithms [5]. The application of RL methods to real-world systems is usually hindered by the large quantities of training data required, which at accelerators with low repetition rates becomes prohibitive to collect. This issue is overcome by pre-training, or fully training, on a simulated version of the environment.

For control problems like the MBI, simulations are computationally expensive, necessitating significantly more time compared to conducting tests directly on the accelerator. The rate at which training data is produced at KARA is sufficient to allow training without simulation, but leads to stringent real-time constraints, not allowing the use of the conventional implementations of RL algorithms. In this work, we use the *experience accumulator* technique described in [6] to perform real-time online training.

So far, the two main attempts at controlling the MBI have been provided in [7] and [8]. These two approaches are very different. The work of [7] experimentally verifies the use of classical control techniques to target the slow-bursting. In [8] a simulation study uses RL techniques to control the bursting behavior, in conditions where the slow-bursting is not present.

In this work, we perform experimental tests of RL control of the slow-bursting. Nonetheless the system described can also be applied to the problem of controlling "finger" bursting.

METHODS

The system described in [5, 6] to control the horizontal betatron oscillations (HBO) at KARA was adapted to control the MBI. A schematic representation is shown in Fig. 1. The CSR from a single bunch is detected with a Schottky diode, whose signal is then digitized and sent over a digital Aurora fiber link by a custom KAPTURE-2 [9] and HighFlex 2 system. The RL agent inference is performed on the KINGFISHER platform using an AMD-Xilinx Versal VCK190 evaluation board.

Thanks to the systematic study carried out in [10], it was shown that accelerating voltage modulations through the

* luca.scomparin@kit.edu

ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-TUPS61

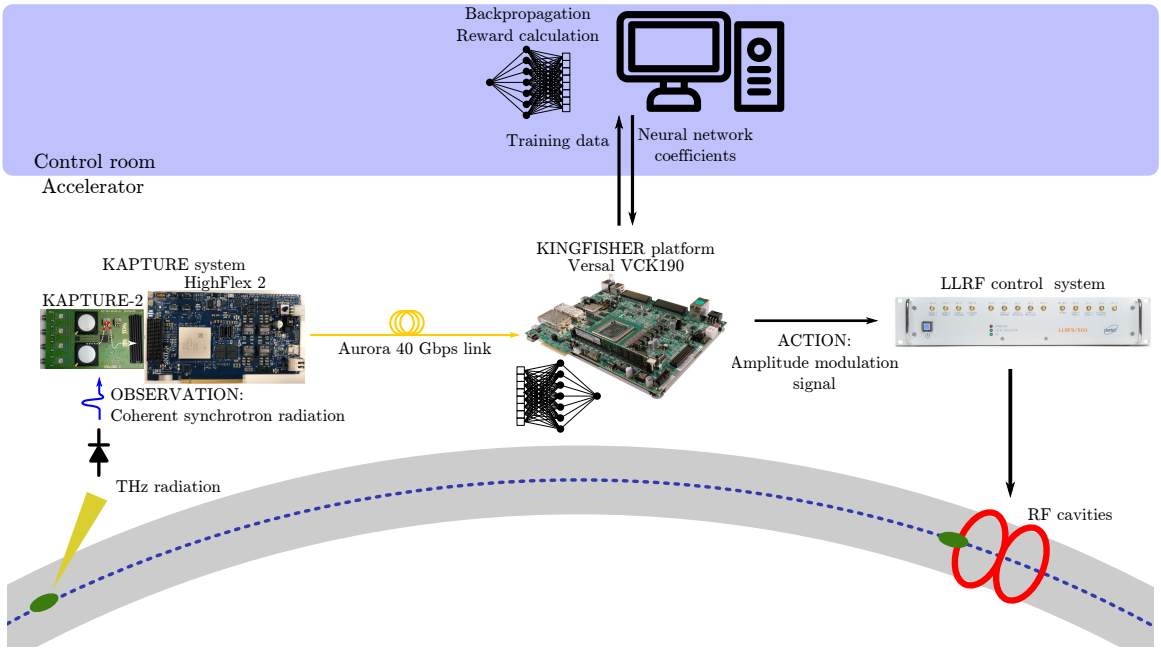

Figure 1: Schematic of the experimental setup employed at KARA. On the accelerator side, following a counter-clockwise signal path, the light is digitized and fed into the AMD-Xilinx Versal VCK190 evaluation card, were the RL agent resides. The training data is sent to a computer in the control room, while the action is applied to the LLRF system.

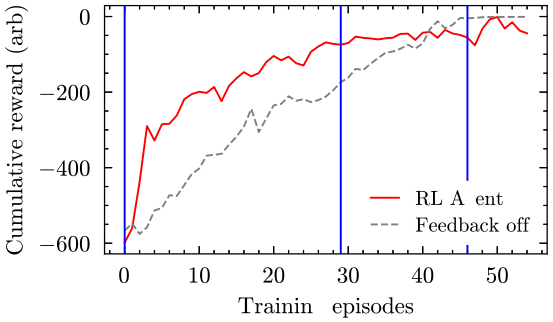

Figure 2: Reward as a function of training step with and without feedback from the RL agent. The blue lines mark the episodes shown in Fig. 3.

radiofrequency (RF) system can affect the instability. Thus, the action signal from the agent was applied as a voltage modulation to the Dimtel LLRF9 [11] low-level radiofrequency (LLRF) system controlling one of the two RF stations of KARA, consisting of two cavities each. In order to do this, a custom serial interface was added to the unit that allows an external device to add an 11 bit amplitude and 13 bit phase offset to the LLRF set-point, corresponding respectively to full scale values of ± 100 kV and $\pm 5.6°$. A new sample is requested by the LLRF every 6 revolutions, corresponding to a sample rate of ≈ 450 kHz.

The signal processing chain, with the agent inference, is analogous to the one in [6]. The RL agent policy is encoded into a 64 neuron, one hidden layer, fully connected neural network (NN). A window of the latest 64 CSR signal samples is used as the input to the NN. A design like this would make the agent sensitive only to those number of samples, but, as discussed in the introduction, the MBI has different timescales that might be of interest for control experiments. Specifically, slow-bursting is characterized by oscillations that would not be perceivable with this setup. Because of this, two decimation stages are applied, each with its own decimation factor and 128-coefficients finite input response (FIR) filters. In order to adapt the decimated rate back to the one requested by the LLRF, an interpolation stage is added to the action. Similarly, a 128-coefficient FIR filter is used. Gaussian noise is added to the output action in order to drive the exploration of the agent.

The same experience accumulator architecture described in reference [6] is employed. A low-latency real-time agent is deployed to the AI Engines of the Versal board and interacts with KARA at a rate of ≈ 28 kHz with a decimation factor of 96, with the interaction being monitored and stored in memory. The training is asynchronously performed on a computer in the control room (as shown in Fig. 1) using the Stable-baselines3 [12] implementation of the proximal policy optimization (PPO) algorithm [13]. The new NN coefficients are then uploaded to Versal to obtain new training data. For the experiments described in this work, 2048 decimated data samples were taken, with a decimation factor of 96, corresponding to 72 ms or $\approx$ 670 synchrotron periods. The *reward engineering* is performed live on KARA, as the reward function can be modified at training time. The accelerator working conditions are reported in Table 1.

DISCUSSION AND OUTLOOK

Due to the low beam lifetime in low-α_c operation mode, the beam current decays rapidly. Given the strong depen-

ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-TUPS61

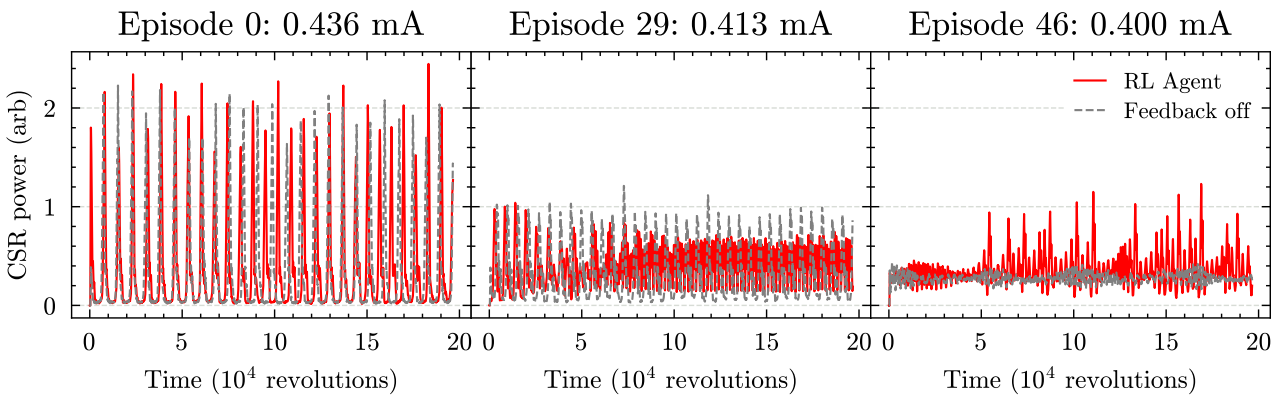

Figure 3: CSR signal in three training episodes showing the strong dependence on beam current. An RL agent data acquisition (red) is compared with one without feedback (grey), in order to high-light the effect of the controller.

Table 1: KARA Machine Parameters Used During the Experiment

Parameter	Value
Energy	1.3 GeV
RF frequency	499.750 MHz
RF voltage	767 kV
Synchrotron frequency	9.3 kHz
α_c	6.7×10^{-4}
Bunch current	0.1 mA to 0.7 mA

dence of the MBI dynamics with bunch current [2, 8], a reference acquisition without agent action, i.e. with the feedback switched off, was taken after each training data acquisition. Several different reward functions and decimation/interpolation settings were tested. An example of the episode reward as a function of the number of training steps is shown in Fig. 2. The reward function was defined as

$$R(x_i) = -(x_i - \bar{x})^2, \quad (1)$$

where x_i is the i-th decimated CSR signal sample, and $\bar{x}$ is the average of the signal over the entire episode. This function was chosen to try and minimize the variance of the signal. The decay of the bunch current leads to a decrease in CSR output, in turn increasing the reward. Thus, it is important to compare the agent reward with a baseline acquisitions with the feedback switched off, in order to disentangle the contribution to the reward increase due to the agent and the one due to the current decay. Three of these training episodes are shown in Fig. 3, with their corresponding reward shown in Fig. 2. In the left panel, the untrained agent is not affecting the slow-bursting behavior, while the bursting is filtered out by the decimation filter, so it is present but not visible. In the no-action signal of the central panel, the bursting is still present, albeit with a lower amplitude due to the decay of the current. The RL agent managed to maintain the fluctuations at a lowered and stable level, after an initial transient. In the right panel, the current falls below the slow-bursting threshold. The RL agent at this point did not train fast enough to adapt to the quickly changing dynamics, and is thus exciting the instability. Despite the varied results, it is important to notice that the RL agent consistently performs better than the no-action baseline, as shown in the accumulated reward over time (Fig. 2), except at the very end, precisely due to the transition below threshold.

The fact that the controller does not perform well at high currents might indicate a fundamental characteristic of the controllability of the MBI, for which studies are so-far missing in literature.

A potential effect impacting the effectiveness of the agent is the chosen Gaussian noise. Specifically, the noise has high-frequency components that are likely filtered out by the response of the RF cavities. A smoother exploration noise could potentially mitigate this phenomenon. Additionally, the problem is partially observable, meaning that the observations given to the agent are not sufficient to fully know the state of the system, in this case represented by the phase space distribution. Adding the latest actions to the observation vector usually gives more information about the current state.

CONCLUSION

A promising approach to control the MBI has been introduced, alongside the first experimental implementation of RL employing solely online training for this problem. The current hardware design allows adaptation to the wide variety of timescales the instability presents. Moreover, clear improvements are proposed that will increase the performance of the trained RL agent allowing a systematic study of this technique, currently underway at KARA.

This work represents a first step towards an autonomous system capable of tailoring terahertz radiation to the user's needs, thanks to the self-learning capabilities of RL.

ACKNOWLEDGEMENTS

This research has been supported by the Helmholtz Association within the Innovation Pool project ACCLAIM - "Accelerating Science with Artificial Intelligence and Machine Learning". The authors acknowledge the support of Dmitry Teytelman in modifying the LLRF hardware to accept an external digital modulation signal.

ISBN: 978-3-95450-247-9 ISSN: 2673-5490 doi: 10.18429/JACoW-IPAC2024-TUPS61

REFERENCES

[1] A.-S. Müller and M. Schwarz, "Accelerator-based thz radiation sources," in *Synchrotron Light Sources and Free-Electron Lasers*. 2020, pp. 83–117. doi:10.1007/978-3-030-23201-6_6

[2] M. Brosi, "In-depth analysis of the micro-bunching characteristics in single and multi-bunch operation at KARA," Ph.D. dissertation, Karlsruher Institut für Technologie (KIT), 2020. doi:10.5445/IR/1000120018

[3] J. L. Steinmann *et al.*, "Continuous bunch-by-bunch spectroscopic investigation of the microbunching instability," *Phys. Rev. Accel. Beams*, vol. 21, p. 110 705, 11 2018. doi:10.1103/PhysRevAccelBeams.21.110705

[4] S. Funkner *et al.*, "Revealing the dynamics of ultrarelativistic non-equilibrium many-electron systems with phase space tomography," *Scientific Reports*, vol. 13, no. 1, 2023. doi:10.1038/s41598-023-31196-5

[5] L. Scomparin *et al.*, "KINGFISHER: A Framework for Fast Machine Learning Inference for Autonomous Accelerator Systems," in *Proc. IBIC'22*, Kraków, Poland, 2022, pp. 151–155. doi:10.18429/JACoW-IBIC2022-MOP42

[6] L. Scomparin *et al.*, *Reinforcement learning on-the-edge through experience accumulators at a particle accelerator*, 2024.

[7] C. Evain *et al.*, "Stable coherent terahertz synchrotron radiation from controlled relativistic electron bunches," *Nature Physics*, vol. 15, no. 7, pp. 635–639, 2019. doi:10.1038/s41567-019-0488-6

[8] T. Boltz, "Micro-bunching control at electron storage rings with reinforcement learning," Ph.D. dissertation, Karlsruher Institut für Technologie (KIT), 2021. doi:10.5445/IR/1000140271

[9] M. Caselle *et al.*, "KAPTURE-2. a picosecond sampling system for individual thz pulses with high repetition rate," *Journal of Instrumentation*, vol. 12, no. 01, p. C01040, 2017. doi:10.1088/1748-0221/12/01/C01040

[10] A. Santamaria Garcia *et al.*, "Systematic study of longitudinal excitations to influence the microbunching instability at KARA," English, in *Proc. IPAC'23*, Venice, Italy, 2023, pp. 2681–2684. doi:10.18429/JACoW-IPAC2023-WEPA018

[11] "LLRF9 product page." (Last accessed in 2024), https://www.dimtel.com/products/llrf9

[12] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: Reliable reinforcement learning implementations," *Journal of Machine Learning Research*, vol. 22, no. 268, pp. 1–8, 2021. http://jmlr.org/papers/v22/20-1364.html

[13] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, *Proximal policy optimization algorithms*, 2017.