FEEDBACK DESIGN FOR CONTROL OF THE MICRO-BUNCHING INSTABILITY BASED ON REINFORCEMENT LEARNING

T. Boltz*, M. Brosi, E. Bründermann, B. Haerer, P. Kaiser, C. Pohl, P. Schreiber, M. Yan, T. Asfour, A.-S. Müller
Karlsruhe Institute of Technology, Karlsruhe, Germany

This contribution is largely based on [1].

Abstract

The operation of ring-based synchrotron light sources with short electron bunches increases the emission of coherent synchrotron radiation in the THz frequency range. However, the micro-bunching instability resulting from self-interaction of the bunch with its own radiation field limits stable operation with constant intensity of CSR emission to a particular threshold current. Above this threshold, the longitudinal charge distribution and thus the emitted radiation vary rapidly and continuously. Therefore, a fast and adaptive feedback system is the appropriate approach to stabilize the dynamics and to overcome the limitations given by the instability. In this contribution, we discuss first efforts towards a longitudinal feedback design that acts on the RF system of the KIT storage ring KARA (Karlsruhe Research Accelerator) and aims for stabilization of the emitted THz radiation. Our approach is based on methods of adaptive control that were developed in the field of reinforcement learning and have seen great success in other fields of research over the past decade. We motivate this particular approach and comment on different aspects of its implementation.

MICRO-BUNCHING INSTABILITY

Modern ring-based synchrotron light sources commonly offer a dedicated short-bunch operation mode in which the bunch length is compressed in order to support dedicated experiments. At the KIT storage ring KARA (Karlsruhe Research Accelerator), this enables the reduction of the bunch length down to several picoseconds. While the high degree of longitudinal compression leads to an increased emission of coherent synchrotron radiation (CSR) in the THz frequency range, it also causes a strong self-interaction of the electron bunches with their own emitted CSR. Above a given threshold current, that depends on several machine parameters [2], this CSR self-interaction causes the formation of dynamically changing micro-structures in the longitudinal charge distribution and hence fluctuating CSR emission. The phenomenon is thus referred to as micro-bunching or microwave instability. The effect of the CSR self-interaction on the longitudinal beam dynamics is conveniently described by the CSR wake potential

$$V_{\mathrm{CSR}}(q,t) = \int_{-\infty}^{\infty} \widetilde{\rho}(\omega,t) Z_{\mathrm{CSR}}(\omega) e^{i\omega q} \mathrm{d}\omega \,, \qquad (1)$$

where $q = (z - z_s)/\sigma_{z,0}$ denotes the generalized longitudinal position, $\widetilde{\rho}(\omega)$ the Fourier-transformed longitudinal bunch profile and $Z_{\mathrm{CSR}}(\omega)$ the CSR-induced impedance of the storage ring. As an additional contribution to the effective potential the bunch is exposed to, besides the accelerating RF potential, this acts as a dynamic perturbation to the temporal evolution of the longitudinal charge distribution. The entire process can be simulated using the KIT-developed Vlasov-Fokker-Planck (VFP) solver Inovesa [3] , which has shown great qualitative agreement with measurements at KARA [4]. Figure 1 illustrates the micro-bunching dynamics in the longitudinal phase space (left) and the corresponding fluctuations of the emitted CSR power (right) simulated with Inovesa.

Depending on the application at hand, the occurrence of micro-structures can also be quite desirable as it increases the radiated power at frequencies corresponding to the size of the present structures. Thus, in order to tailor the CSR emission to each application individually, this contribution is concerned with the development of a longitudinal feedback that establishes extensive control over the micro-bunching dynamics and thereby enables, both, excitation and mitigation of the occurring micro-structures.

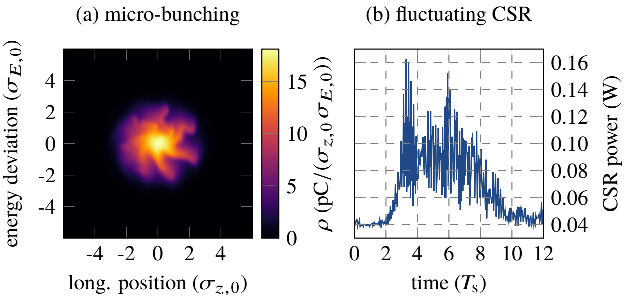

Figure 1: (a) The CSR self-interaction of the bunch causes the formation of micro-structures in the longitudinal charge distribution. (b) Their dynamic evolution leads to fluctuations in the emitted CSR power (T_s denoting the synchrotron period).

APPROACH TO CONTROL

As briefly discussed in [5], we find that the instability is largely driven by the CSR wake potential's perturbation of the restoring force provided by the RF system. Particularly, the slope of the effective potential at the synchronous position is modified considerably during the micro-bunching dynamics. To exert control, we thus aim to recover the strength of the restoring force in order to compensate a major part

* tobias.boltz@kit.edu

of the perturbation caused by the CSR wake potential. As the perturbation is, according to Eq. (1), dependent on the bunch profile and therefore on the evolution of the charge distribution, the compensation mechanism has to dynamically adjust to this as well. As an empirically effective and feasible approach we therefore aim for an RF amplitude modulation scheme

$$V_{\mathrm{RF}}(t) = \hat{V}(t)\sin(2\pi f_{\mathrm{RF}}\, t)\,, \tag{2}$$

$$\hat{V}(t) = \hat{V}_0 + A_{\mathrm{mod}}(t)\sin(2\pi f_{\mathrm{mod}}(t)\, t + \varphi_{\mathrm{mod}})\,, \tag{3}$$

in which the modulation amplitude $A_{\mathrm{mod}}(t)$ and frequency $f_{\mathrm{mod}}(t)$ are rapidly adjusted. This yields a sequential decision problem in which we would like to determine the ideal choice of $A_{\mathrm{mod}}(t_i)$ and $f_{\mathrm{mod}}(t_i)$ at every time step t_i. Given that the micro-bunching dynamics occur at time scales comparable to the synchrotron period, the step width Δt of the sequence should be chosen in the same order of magnitude. As a promising approach to solve this task, the following section briefly introduces the basic concept of reinforcement learning. A more detailed introduction can be found in [6].

REINFORCEMENT LEARNING

Reinforcement learning (RL) is an active sub-field of machine learning which led to spectacular results in recent applications, see e.g. [7, 8]. It differs from other forms of machine learning in that its learning paradigm does not require a pre-existing data set. Instead learning takes place in an iterative process based on the general concept of trial-and-error search. The learner or decision maker, usually called the *agent*, continuously interacts with an *environment* while seeking to improve its behavior. At each iteration the agent perceives the current *state* S_t of the environment and is faced with the task to choose an *action* A_t. Based on the chosen action, the environment yields a scalar *reward* R_t and transitions to the next state S_{t+1}. Thereby, the agent's objective is defined as maximizing the cumulative reward received over time.

Formally, the reinforcement learning problem is described as a Markov decision process (MDP). In its most rigorous form, the MDP demands a perfect fulfillment of the Markov property, which puts a specific restriction on the sequence of states: The probability of transitioning to state S_{t+1} may only depend on the previous state S_t and not on any other state visited in the past $(S_1, \ldots, S_{t-1})$. If this condition is satisfied, it guarantees that the agent is provided with the necessary information to choose the optimal action in every encountered state. While this rigorous formalism is very useful for modeling a wide range of problems and allows precise theoretical statements, the Markov property can sometimes be difficult to fulfill in practical applications.

FEEDBACK DESIGN

For the sequential decision problem denoted in Eq. (3) the definition of a Markovian process is straightforward. In order to simulate the longitudinal beam dynamics VFP solvers require an initial charge distribution in the longitudinal phase space and a set of constant parameters. Subsequently, the temporal evolution of this distribution is calculated in an iterative manner. At each step, the calculation is entirely based on the preceding distribution. Hence, choosing the charge distributions $\psi_t(z, E)$ as the state signal

$$S_t \doteq \psi_t(z, E) \tag{4}$$

yields a Markov process, fully satisfying the Markov property introduced in the previous section.

As mentioned above, we are primarily interested in tailoring the emission of CSR to individual applications. We thus define the reward function based on the observed CSR signal

$$R_t \doteq R_t(P_{t,\mathrm{CSR}})\,. \tag{5}$$

In case of trying to mitigate the instability, the damping of the micro-structures in phase space corresponds to a stabilization of the emitted CSR power as it removes the fluctuation caused by the micro-bunching dynamics. One way to express this objective in a scalar reward function is

$$R_t \doteq w_1 \mu_{t':t} - w_2 \sigma_{t':t}\,, \tag{6}$$

where $\mu_{t':t}$ and $\sigma_{t':t}$ denote the mean and standard deviation of the time series $P_{t,\mathrm{CSR}}$ in the interval $[t', t]$, and $w_{1,2} > 0$ are simple weighting factors. As a complementary approach, trying to deliberately excite the micro-structures to increase the emission of CSR in the desired frequency range $[f_1, f_2]$, the reward function may simply be defined as the emitted power in this bandwidth

$$R'_t \doteq \mu_{t':t}(f_1, f_2)\,. \tag{7}$$

Finally, the formal definition of the action space corresponding to Eq. (3) is chosen as

$$A_t \in \{A_{\mathrm{mod}}(t) \times f_{\mathrm{mod}}(t)\}\,. \tag{8}$$

Combining the above stated definitions of S_t, R_t and A_t yields a fully functional MDP to which we can apply established RL solution methods. The VFP solver Inovesa was already adjusted to support these efforts and first tests of training an agent on simulation data are currently ongoing.

Feasibility of the State Signal

While the definition of the state signal in Eq. (4) provides the theoretical comfort of perfectly fulfilling the Markov property, measuring the longitudinal charge distribution in phase space at real storage rings is a major challenge. To make this approach more applicable in practice we would thus like to use information provided by diagnostic systems that are more commonly available. For now, the simplest and most robust way to acquire information about the state of the micro-bunching at KARA is by measuring the emitted CSR power $P_{t,\mathrm{CSR}}$ in the THz frequency range, e.g. [4, 9]. As $P_{t,\mathrm{CSR}}$ is strongly correlated with the micro-bunching

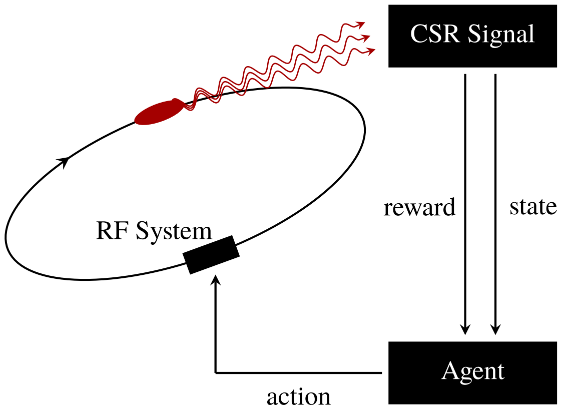

Figure 2: General feedback scheme using the CSR power signal to construct both, the state and reward signals of the Markov decision process (MDP).

dynamics, we consider the following alternative definition of the state signal

$$S_t \doteq S_t(P_{t,\mathrm{CSR}}) \ . \tag{9}$$

The resulting general feedback scheme is illustrated in Figure 2. Whether or not the CSR signal can provide enough information for the decisions the agent is confronted with has to be verified empirically. Ideally, the condensed information yields a fast learning process and convergence to a satisfying extent of control over the micro-bunching dynamics. If the CSR signal turns out to be insufficient, the state signal should be augmented with complementary information about the longitudinal phase space restoring the Markov property as closely as possible.

SUMMARY

In order to establish extensive control over the micro-bunching dynamics in short electron bunches, a fast and adaptive longitudinal feedback is required which is capable of adjusting to the dynamic perturbation caused by the CSR self-interaction. Given that the CSR wake potential explicitly depends on the current state of the charge distribution, the action or countermeasure should, in general, be expected to be state-dependent as well.

In this contribution, we outline a general feedback scheme that is designed to make use of reinforcement learning methods in order to accomplish this challenging task.

ACKNOWLEDGEMENT

T. Boltz and P. Schreiber acknowledge the support by the DFG-funded Doctoral School “Karlsruhe School of Elementary Particle and Astroparticle Physics: Science and Technology (KSETA)”. This research is in part supported by the Innovationspool AMALEA (Accelerating Machine Learning for Physics) in the Helmholtz Association’s Programme “Matter and Technologies”.

REFERENCES

[1] T. Boltz *et al.*, “Feedback Design for Control of the Micro-Bunching Instability based on Reinforcement Learning”, in *Proc. 10th Int. Particle Accelerator Conf. (IPAC’19)*, Melbourne, Australia, May 2019, pp. 104–107. `doi:10.18429/JACoW-IPAC2019-MOPGW017`

[2] K. L. F. Bane, Y. Cai, and G. Stupakov, “Threshold studies of the microwave instability in electron storage rings”, *Phys. Rev. ST Accel. Beams*, vol. 13, p. 104402, 2010.

[3] P. Schönfeldt *et al.*, “Parallelized Vlasov-Fokker-Planck solver for desktop personal computers”, *Phys. Rev. Accel. Beams*, vol. 20, p. 030704, 2017. `https://github.com/Inovesa/Inovesa`

[4] J. L. Steinmann *et al.*, “Continuous bunch-by-bunch spectroscopic investigation of the microbunching instability”, *Phys. Rev. Accel. Beams*, vol. 21, p. 110705, 2018.

[5] T. Boltz *et al.*, “Perturbation of Synchrotron Motion in the Micro-Bunching Instability”, in *Proc. 10th Int. Particle Accelerator Conf. (IPAC’19)*, Melbourne, Australia, May 2019, pp. 108–111. `doi:10.18429/JACoW-IPAC2019-MOPGW018`

[6] R. S. Sutton and A. G. Barto, *Reinforcement Learning*. Cambridge, MA, USA: MIT Press, 2018.

[7] V. Mnih *et al.*, “Human-level control through deep reinforcement learning”, *Nature*, vol. 518, pp. 529–533, 2015.

[8] D. Silver *et al.*, “Mastering the game of Go with deep neural networks and tree search”, *Nature*, vol. 529, pp. 484–489, 2016.

[9] M. Brosi *et al.*, “Fast mapping of terahertz bursting thresholds and characteristics at synchrotron light sources”, *Phys. Rev. Accel. Beams*, vol. 19, p. 110701, 2016.