Machine Learning Activities at KIT for Accelerators

Andrea Santamaría García, Tobias Boltz, Erik Bründermann, Bastian Härer, Akira Mochihashi, Michael Nasse, Carl Sax, Chenran Xu, Anke-Susanne Müller

DESY IPC Seminar (26-01-2021)
Machine Learning Activities for Accelerators

KARA
- Timing Modes for Advanced Light Sources: Control of the Micro-Bunching Instability with Reinforcement Learning
  Tobias Boltz, Bastian Härer, Andrea Santamaría García
- Bayesian Optimization of the Injection Efficiency
  Chenran Xu, Akira Mochihashi

FLUTE
- Machine Learning Towards Autonomous Accelerators: Control of the longitudinal bunch profile with Reinforcement Learning
  Andrea Santamaría García, Chenran Xu, Erik Bründermann

Test facility for accelerator physics
Synchrotron light source
Control of the Micro-Bunching Instability with Reinforcement Learning

Laboratory for Applications of Synchrotron Radiation (LAS)

Project: **Timing Modes for Advanced Light Sources (TiMo)**
BMBF-Verbundforschung

Project partners:
- Helmholtz-Zentrum Berlin
- TU Dortmund

**GOAL AT KIT**

Development of a longitudinal feedback system to control the micro-bunching instability in short bunch operation mode with Reinforcement Learning (RL) to **tailor Coherent Synchrotron Radiation (CSR)** emission.
Coherent Synchrotron Radiation (CSR)

There is a strongly increased radiation at wavelengths longer than the bunch length ($\sigma_z < \lambda$) where the synchrotron radiation is emitted coherently (CSR).

CSR = high radiation power $\Rightarrow$ interesting for users! $\Rightarrow$ operation in short bunch mode (low $\alpha$ optics)

- microstructures appear making the CSR power fluctuate

We want a high average, low variance CSR

Image courtesy of A.-S. Müller

Simulation code: Parallelized VFP solver Inovesa

T. Boltz et al, MOPGW017, IPAC’19
Influencing the Microbunching Instability

The **CSR self-interaction** contributes to the effective potential that the beam is subjected to, and is continuously changing during micro-bunching dynamics.

\[ V_{\text{eff}}(q) = V_{\text{RF}}(q) + V_{\text{CSR}}(q) \]

**IDEA**: compensate the effect of the CSR perturbation and go back to the original restoring force provided by the accelerating voltage by **modulating the RF voltage (amplitude)**.

\[ V_{\text{RF}} = \hat{V}(t) \sin(2\pi f_{\text{RF}} t) \]

\[ \hat{V}(t) = \hat{V}_0 + A_{\text{mod}} \sin(2\pi f_{\text{mod}} + \phi_{\text{mod}}) \]

Trying the idea in simulation with a constant modulation:

- Initial damping, but quickly out of sync...we need **dynamic control**!

Images courtesy of T. Boltz
Testing the idea with manual control

\[ V_{RF} = \hat{V}(t) \sin(2\pi f_{RF} t) \]

\[ \hat{V}(t) = \hat{V}_0 + A_{mod} \sin(2\pi f_{mod} + \varphi_{mod}) \]

High average, low variance CSR. So far so good…now to Reinforcement Learning!
The Reinforcement Learning Ecosystem

Reinforcement learning does not train on a static set of data, it works on a dynamic system where time matters, where the system acted upon is evolving in time.

**AGENT**

Interacts with the environment through actions that are solely influenced by a reward signal.

May include one or more of these components:

- **Policy**: agent's behaviour function (how the agent picks its actions)
- **Value function**: how good is each state and/or action.
- **Model**: agent's representation of the environment.

executes action → receives observation and scalar reward → updates policy / value function

**REWARD**

Scalar feedback signal (many reward signals are valid, but need to be chosen wisely)

**GOAL**

Maximization of expected cumulative reward

**MARKOV PROCESSES**

Markov Decision Processes (MDPs) formally describe RL environments

Markov process = sequence of Markov states

→ The current state contains enough information to characterize what happens next (the future is independent of the past, given the present)
Illustrating Reinforcement Learning concepts

The agent needs to get from state 0 to state 15 to get out of the maze

<table>
<thead>
<tr>
<th>Reward</th>
<th>Actions</th>
<th>Policy</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1 for each step, +1 at state 15</td>
<td>Up, down, right, left</td>
<td>Random</td>
<td>Steps</td>
</tr>
</tbody>
</table>

A random policy seems not to be the best choice! Only 6 steps are needed to get out

Value function obtained by analytically solving Bellman's equation ($O(n^3)$ for n states)

How to improve in the next episode?

Many types of agents:

- **Value based**: contains a value function, policy is implicit (no policy)
- **Policy based**: does not store the value function, only the policy
- **Actor Critic**: stores both the policy and value function
- **Model free**: the agent simply relies on some trial-and-error experience for action selection
- **Model based**: the agent exploits a previously learned model to accomplish the task
The plethora of RL algorithms

**Value Optimization**
- DQN
- MMC
- PAL
- NAF
- NEC
- QR-DQN
- Bootstrapped DQN
- Categorical DQN
- DDQN
- N-Step Q-Learning
- Dueling DDQN with PER
- UCB via Q-ensembles
- Rainbow

**Policy Optimization**
- Actor Critic
- Policy gradient
  - DDPG
  - PPO
  - SAC
  - ACER
  - DDPG with HER
  - DDPG HAC
  - TD3

**Explored algorithms**
- DDPG: Deep Deterministic Policy Gradient
- PPO: Proximal Policy Optimization
- SAC: Soft Actor Critic
- TD3: Twin Delayed DDPG
Applying Reinforcement Learning to our case

**Action**
Modulation of the RF amplitude
\[ \hat{V}(t) = \hat{V}_0 + A_{mod} \sin(2\pi f_{mod} + \varphi_{mod}) \]

**Reward**
\[ R = \mu_{CSR} - w \sigma_{CSR} \]
where \( w \) is a weight
Could we improve the reward definition?

**Experimental observable (state definition):**
CSR signal
We can consistently and reliably measure the THz emission thanks to KAPTURE II and provides information about the micro-bunching dynamics. Input: (8x1) feature vector

**Simulation observable (state definition):**
Charge distribution
Input: (256x256) matrix + (5x1) feature vector

Images courtesy of T. Boltz
Observation vector based on the CSR signal

- $\mu_{CSR}$ is the normalized mean of the CSR power signal in the last time period.
- $\sigma_{CSR}$ is the normalized standard deviation of the CSR power signal in the last time period.
- $m_{trend}$ is a slow trend of the CSR power signal.
- $a_{f_{main}}$ is the amplitude of the main frequency in the Fourier transformed CSR signal.
- $f_{main}$ is the main frequency in the Fourier transformed CSR signal.
- $\varphi_{f_{main}}$ is the phase of the main frequency in the Fourier transformed CSR signal.
- $\Delta \theta_{RF}$ is the relative phase between the CSR signal and the applied RF signal (amplitude modulation).
- $c_{term}$ models the termination condition (difference between the last reward and the one 10 steps prior).
Training the agents (CSR signal)

**Algorithm: PPO**

- The NNs are updated at every step
- Performance (reward) drops to zero after a certain number of episodes with the PPO agent
- The results on the next slide show peak performance / the best deployable agent

Images courtesy of T. Boltz
Can we improve performance by adjusting the exploration noise?

Episode is terminated because it does not improve
→ Exploration noise = Ornstein-Uhlenbeck

Quickly finds a stable regime and the CSR emission is considerably improved
→ The exploration noise is built into the algorithm and it’s naturally reduced at low rewards
Evolution of the actions with time (PPO)

Images courtesy of T. Boltz

It cannot be appreciated in the RF voltage modulation by the naked eye but…

Small changes in the modulations frequency stabilize the CSR emission

1 step = 0.25 synchrotron periods (chosen small enough for the agent to be able to react to the changing micro-structure dynamics)
In practice: we need hardware!

Fast feedback for real-time optimization

- Institute of Data Processing and Electronics (IPE)
- Institute for Anthropomatics and Robotics (IAR)

Cooperation between:

Achieved:
- RL framework developed on HighFlex 2 → easy deployment of ML algorithms on an FPGA
- The DDPG algorithm was tested with a very low latency of 17 μs in the scope of a PhD thesis

New PhD position open! (2021-2024).
Contact: Michele Caselle

First steps at real-time control of physical processes with ML

Edmund Blomley, Tobias Boltz, Miriam Brosi, Michele Caselle, Timo Dritschler, Melvin Klein, Christoph Pohl, Weijia Wang
T. Boltz, W. Wang et al, TUCPL06, ICALEPCS 2019
Control of the Micro-Bunching Instability with Reinforcement Learning

ACHIEVEMENTS

- Manual control of the microbunching instability in simulation with the Vlasov-Fokker-Planck solver Inovesa through RF amplitude modulation
- Research of different RL algorithms (DDPG, TD3, PPO and SAC) and their application to control in simulation
- First experimental studies: capability of influencing the microbunching instability with RF amplitude modulations of the kicker cavity in a closed loop demonstrated at KARA
- Development of an RL framework for easier deployment of ML algorithms on FPGAs
- DDPG algorithm tested on the FPGA with a very low latency (17 μs)
Control of the Micro-Bunching Instability with Reinforcement Learning

OPEN QUESTIONS

**Benchmark Problem (Learning Process)**
- Reproducibility (fully deterministic computation)
- Choosing suitable exploration noise
- Instability of the agent’s learning process
- Reward function and termination condition
- Choosing the RL algorithm (e.g., on-/off-policy learning)

**Extension of Control**
- Feasibility of control based on solely the CSR signal
- Signal generalization to different bunch currents
- Generalization to different machine settings
Bayesian Optimization of the Injection Efficiency

- We want to improve the injection rate from the booster to the storage ring.

- Manual trial-and-error tuning is time consuming, depends on the operator’s experience, and can easily get stuck in local optima.

- Bayesian optimization has been successfully implemented in other facilities and can converge to the global optimum.

C. Xu, Master thesis, KIT, to be published.
Preparing and testing the model

Determination of the Gaussian Process (GP) hyperparameters (signal variance, characteristic length-scales, and noise variance) through measurements. A GP model is trained to recreate the observed structure.

Testing the model in simulation:

- Left: grid scan of parameter space (~1h)
- Right: Bayesian Optimization (3 min)

Injection efficiency averaged over 10 runs for two different acquisition functions

Plots courtesy of C. Xu
Correction for beam lifetime dependency with stored current

- The beam lifetime depends on the storage ring current (Touschek scattering)
- If left unaccounted for, the injection efficiency values will decrease independently of the optimization algorithm

Plots courtesy of C. Xu
Bayesian Optimization of the Injection Efficiency

ACHIEVEMENTS

- Development of a Bayesian Optimization algorithm with Gaussian Processes with three different acquisition functions and up to 9 input parameters

Optimized injection is roughly two times faster!

Plots courtesy of C. Xu
Machine Learning Toward Autonomous Accelerators

The "Autonomous Accelerator" is a two-year project funded by Helmholtz AI, one of the five platforms initiated by the Helmholtz Information and Data Science Incubator.

Start: September 2020

- DESY PI (coordinating): Annika Eichler
- KIT PI: Erik Bründermann
- KIT Research associate: Andrea Santamaría García
- DESY Research associate: Oliver Stein
- KIT Doctoral researcher: Chenran Xu
- ARES contact: Florian Burkart
Goals & Challenges

Goals

- Control of the longitudinal bunch profile (to transversal bunch profile control, to automatic start-up)
- Transfer Learning: apply algorithms to two similar facilities (ARES and FLUTE)

Challenges

- High-dimensional, continuous state and action spaces
- Low repetition rate
- Nonlinear (collective) effects
- (Sub)-femtosecond requirements on bunch duration

FLUTE

Spatial Light Modulators

- Laser pulse length
- Laser pulse shape
- Laser spot size
- Laser spot position
- Magnet current

Photoinjector

RF gun

- RF amplitude
- RF phase
- Beam arrival
- Time

Traveling wave structure

- Temperature profile
- Amplitude
- Phase

Magnetic chicane

- R56 travel difference of average beam energy

Institute for Synchrotron Radiation
TU Darmstadt, 13.12.2010

A.-S. Müller – Coherent Synchrotron Radiation in Storage Rings

FLUTE

Spatial Light Modulators

- Laser pulse length
- Laser pulse shape
- Laser spot size
- Laser spot position
- Magnet current

Photoinjector

RF gun

- RF amplitude
- RF phase
- Beam arrival
- Time

Traveling wave structure

- Temperature profile
- Amplitude
- Phase

Magnetic chicane

- R56 travel difference of average beam energy

Institute for Synchrotron Radiation
TU Darmstadt, 13.12.2010

A.-S. Müller – Coherent Synchrotron Radiation in Storage Rings
Spatial Light Modulators

- The 3D shape of the electron beam is directly correlated to the shape of the laser light that will hit the cathode of the photoinjector:
  - e.g., laser pulse length → electron bunch length

- Spatial Light Modulators (SLMs) provide a flexible way of shaping the laser pulse in space and time

- Nonlinear effects inherent to laser manipulation (transportation, compression, third harmonic generation) can distort the original laser pulse

- Goal: use the virtual cathode signal as feedback to iteratively optimize the SLM setting by training a NN to learn the nonlinear mapping between the initial and final laser shapes
First steps with test setup

Generation of phase image for SLM input with Gerchberg–Saxton algorithm

Capturing the image with a camera and loading it in Python for loss calculations and on-the-loop image corrections with ML

Interesting topic tackled at the Stanford Computational Imaging Lab (Neural Holography)

http://www.computationalimaging.org/publications/neuralholography/

Images courtesy of C. Sax and C. Xu
End of 2020 Outlook

FINISHED

Bayesian optimization of the injection efficiency
Master thesis by Chenran Xu finished

ONGOING

Control of the micro-bunching instability with Reinforcement Learning
PhD by Weijia Wang finished
PhD by Tobias Boltz finishing this year

STARTED

Machine Learning toward autonomous accelerators
PhD by Chenran Xu started

STARTING

Helmholtz Innovation Pool (AI): ACCLAIM

Let’s keep in touch in the international machine learning for accelerators community!
Let me know if you would like to show your work: andrea.santamaria@kit.edu

The One World charged particLe accElerator (OWLE) Colloquium & Seminar Series

Given the impossibility of travel during the COVID-19 crisis the (OWLE) seminar series was established as an inter-institutional global online colloquium and seminar(s).

The OWLE-Colloquium is aimed at giving researchers a platform to share research and development results of very broad interest.

The OWLE-ML seminar series has a topical focus on machine learning and experimental demonstration of AI-ML.

Colloquium talks are held via Zoom once a month on the first Tuesday at 1:30 PM EDT (19:30 CEST, 10:30 AM PST).
Seminars are held every second and last Tuesdays at 2:30 PM EDT (20:30 CEST, 11:30 AM PST).

https://sites.google.com/view/owle/