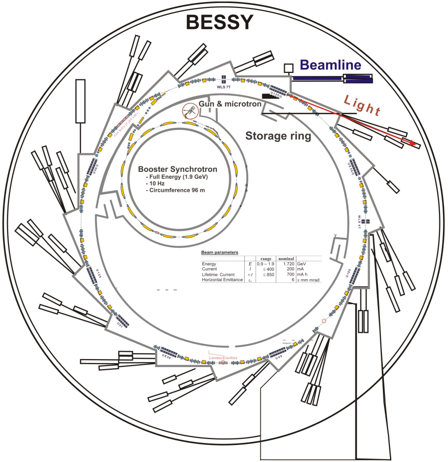
AMALEA:
HZB Final Report

Luis Vera Ramírez, Tom Mertens, Pierre Schnizer, Gregor Hartmann

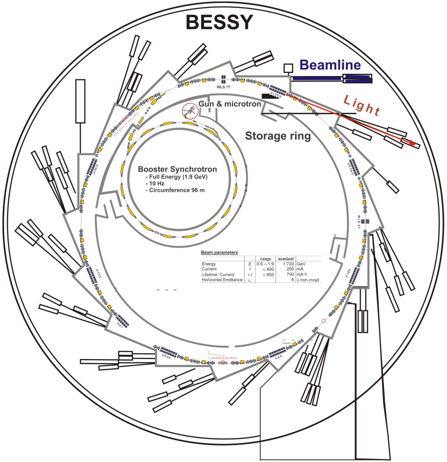
Helmholtz-Zentrum Berlin

luis.vera_ramirez@helmholtz-berlin.de

Table of Contents



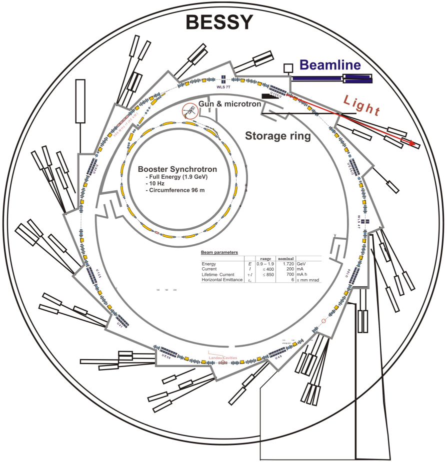
HZB Helmholtz Zentrum Berlin



ML @ BESSY II: Overview

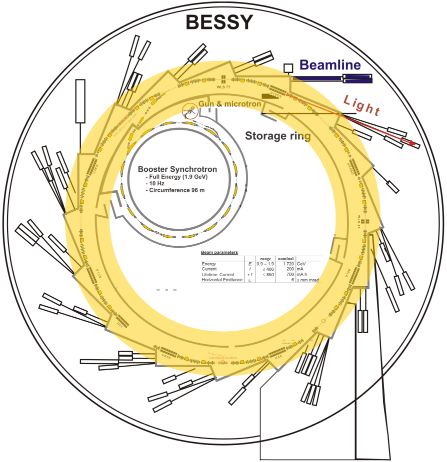
HZB Helmholtz Zentrum Berlin

Device	Injector Ring accelerator	Undulator	Beamline	Experiment
Data	• Archive, Diagnostic • Simulations • Online optimization	• Diagnostic • Raytracing • Scans/online data	• Demands • Simulations • Beamtimes	
Methods	• SVR-RFF • DNN • Deep-RL-Control • RNN, LSTM	• Autoencoder • CNN, MLP, GBoost • Dataloader • Tensor product • kNN, auto-diff.	• Reasonable random generator	
Agent	Operator	Beamline scientist	(Random-) User	



		range	nominal	
Energy	E	0.9 – 1.9	1.720	GeV
Current	I	≤ 400	200	mA
Lifetime- Current	$\tau \cdot I$	≤ 850	700	mA h
Horizontal Emittance	ε_x		6	π mm mrad

Table of Contents



- Beam lifetime: defined via the **current decay rate**

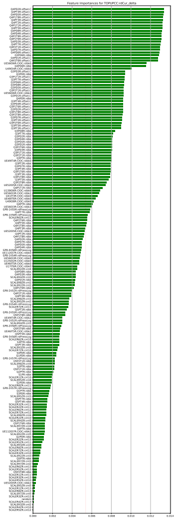
$$\frac{1}{\tau} = -\frac{\mathring{I}}{I}$$

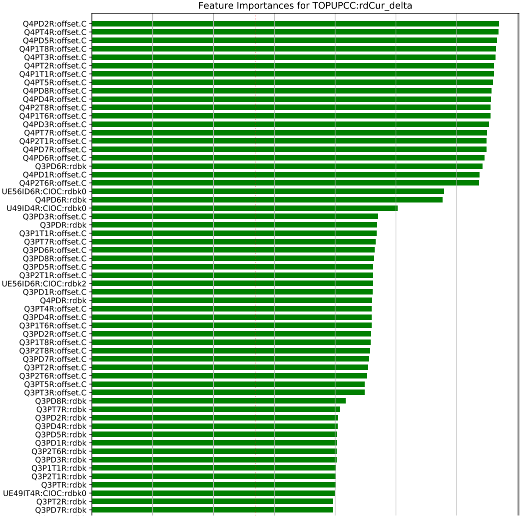
- EPICS Variable can't be used - very delayed
- First approach: exact calculation from measurements - unstable due to measurement errors

$$\frac{1}{\tau} = -\frac{\ln(I_t) - \ln(I_{t_0})}{t - t_0}$$

- Final approach: **piecewise linear regression** with k previous measurements (experiments with $k = 20$)

$$\frac{1}{\tau} \approx -\frac{1}{I_t} \frac{\sum_{i=0}^{k} (I_{t-i} - I_{t_0})(t - i - t_0)}{\sum_{i=0}^{k} (t - i - t_0)^2}$$

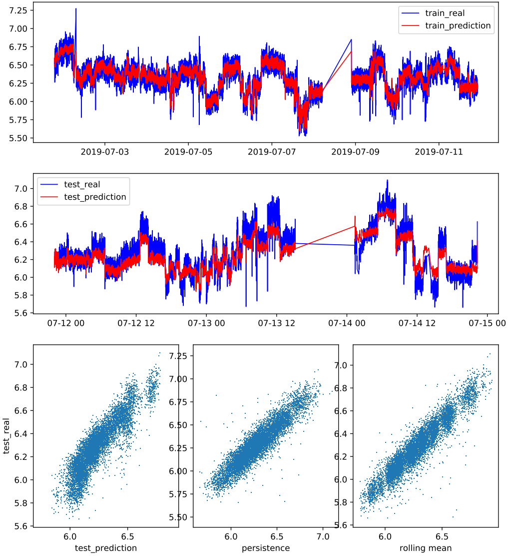




- **185 input variables**: undulator gaps and shifts (21), quadrupole (58) and sextupole (7) currents, quadrupole offsets (38), local beam loss measurements (49)...
- Evenly distributed feature importances - **quadrupoles** (offsets) and **undulator** gaps stand out.

Beam Lifetime Prediction: SVR-RFF and DNN with chronological split

HZB Helmholtz Zentrum Berlin



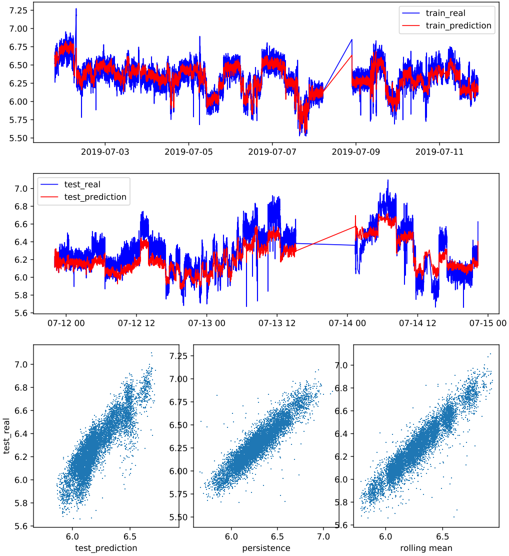
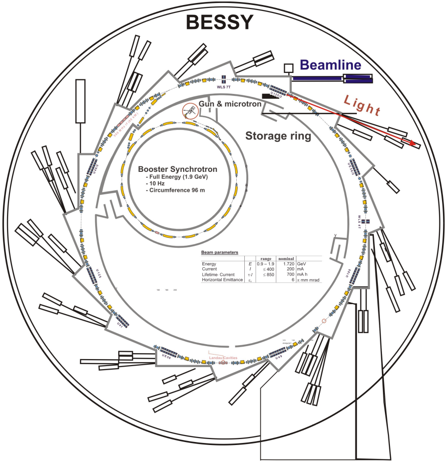
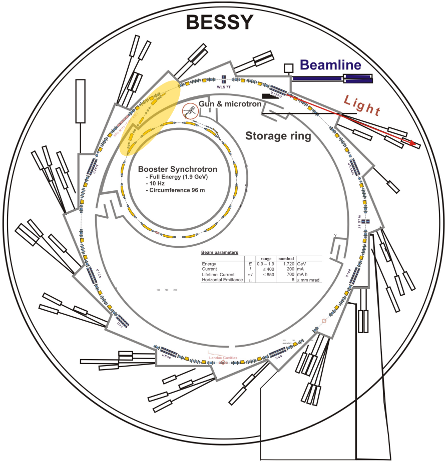


Table of Contents



		range	nominal	
Energy	E	0.9 – 1.9	1.720	GeV
Current	I	≤ 400	200	mA
Lifetime- Current	$\tau \cdot I$	≤ 850	700	mA h
Horizontal Emittance	ε_x		6	π mm mrad

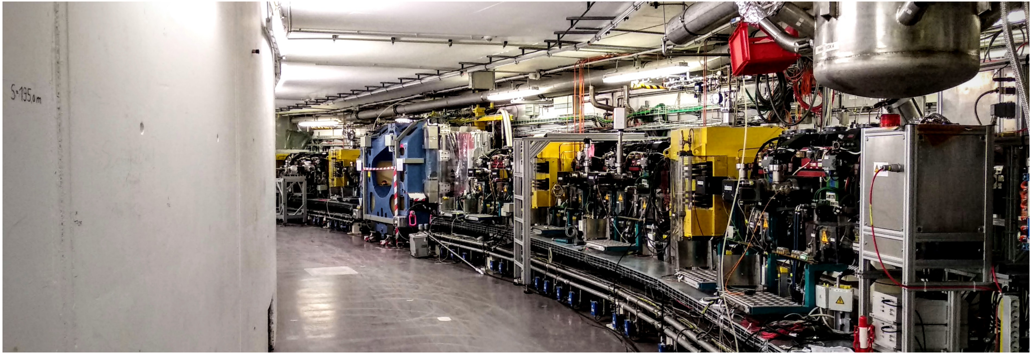
Table of Contents



Injection Efficiency: Case description

We can define the injection efficiency as the **fraction of current increase** generated in the storage ring by the charge accelerated in the booster.

Injection efficiency is known to be affected by temperature and nowadays needs **manual tuning** $\rightarrow$ suitable for **RL-based optimization**



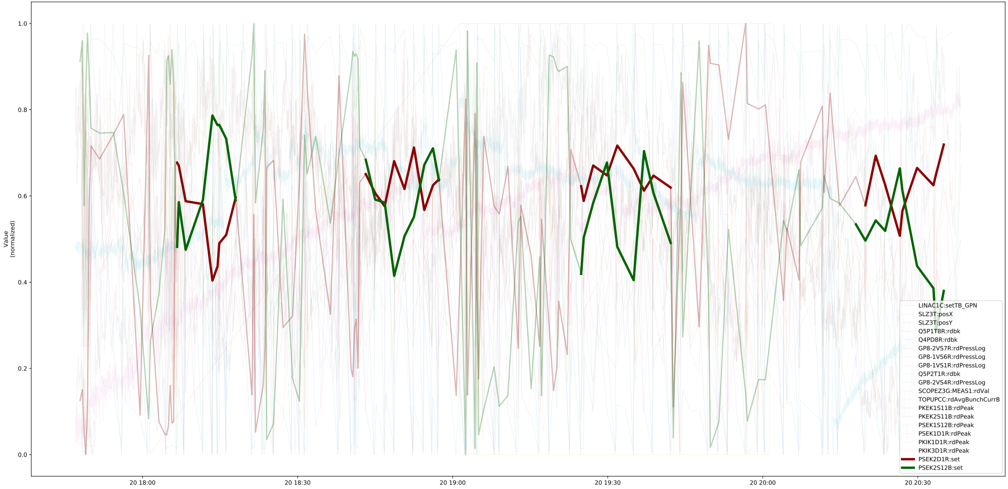
Injection Efficiency: Case description

- Action variables:
  - **Deflection angle into the storage ring**, generated by the 2^{nd} septum.
  - **Deflection angle into the booster**, generated by the 2^{nd} septum.
- State variables (19):
  - Number of **bunches** generated by the LINAC (1, 3 or 5).
  - Injection angle **mismatch**, measured by the beam position in the transfer line (x,y).
  - **Current** measured during the booster acceleration phase.
  - Measured **loss rate after extraction** from the booster.
  - Power supply currents into **quadrupoles** (3 variables).
  - Collisions with rest gas particles and vacuum pressures (4 variables).
  - Peak intensity of the remaining ring and booster septa and kickers (6 variables).
- Reward: **injection efficiency** of the last shot.

Observed states:

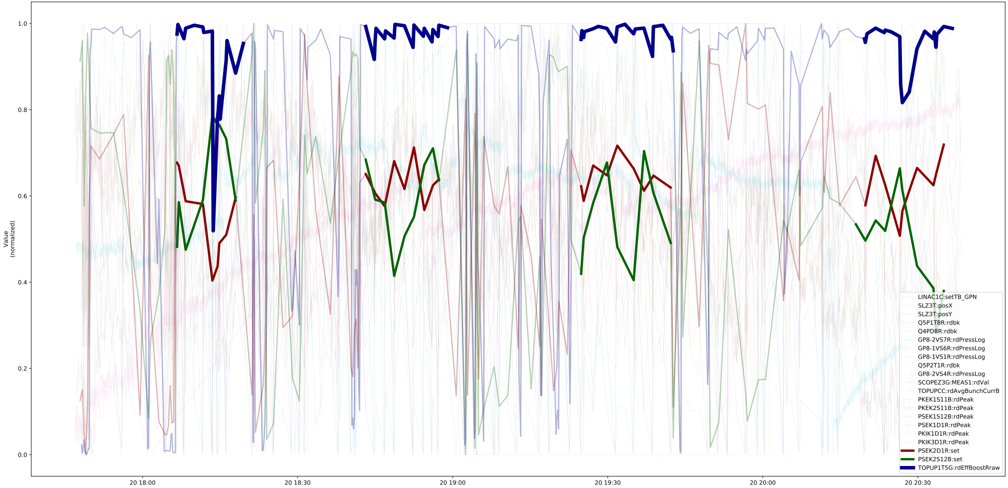
Injection Efficiency: Test @ Machine Commisioning (20/12/19)

Observed states + actions:



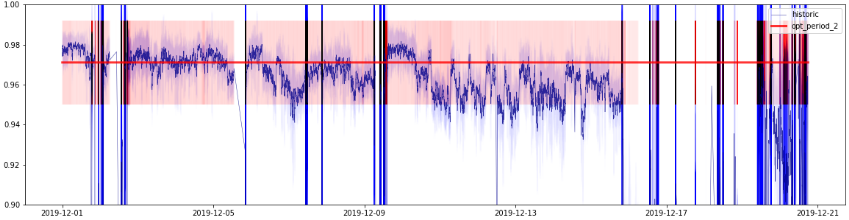
Injection Efficiency: Test @ Machine Commisioning (20/12/19)

Observed states + actions + rewards:

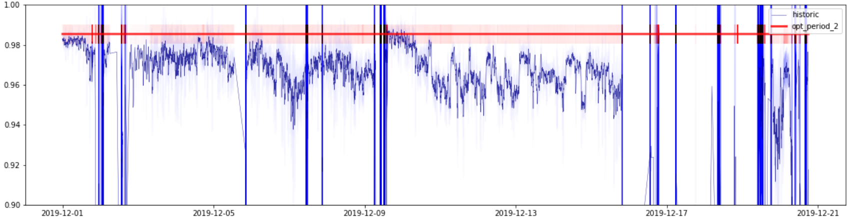


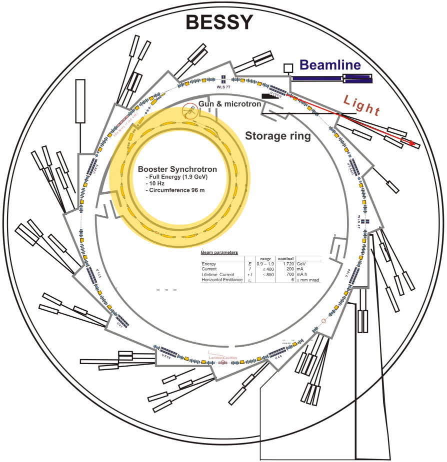
Injection Efficiency: Results comparison

The 2nd exploitation period (18:42 - 19:01) achieved a mean efficiency of $97.11 \pm 2.09\%$



Restricted to multibunch injections the mean efficiency was $98.56 \pm 0.46\%$





BESSY
Beamline
Light
Gun & microtron
Storage ring
Booster Synchrotron
- Full Energy (1.9 GeV)
- 10 Hz
- Circumference 96 m
Beam parameters
range
nominal
Energy E 0.9 – 1.9 1.720 GeV
Current I ≤ 400 200 mA
Lifetime- Current τ·I ≤ 850 700 mA h
Horizontal Emittance ε 6 π mm mrad

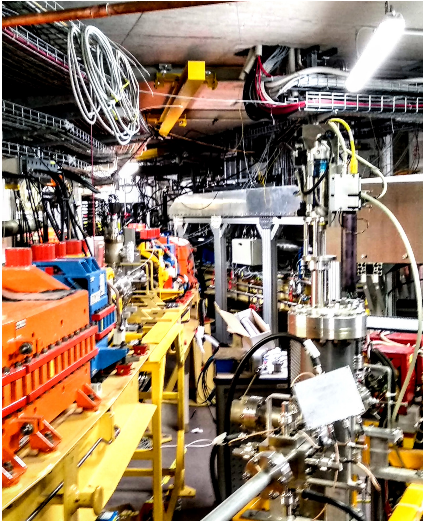
Booster Current Optimization: Case description

HZB Helmholtz Zentrum Berlin

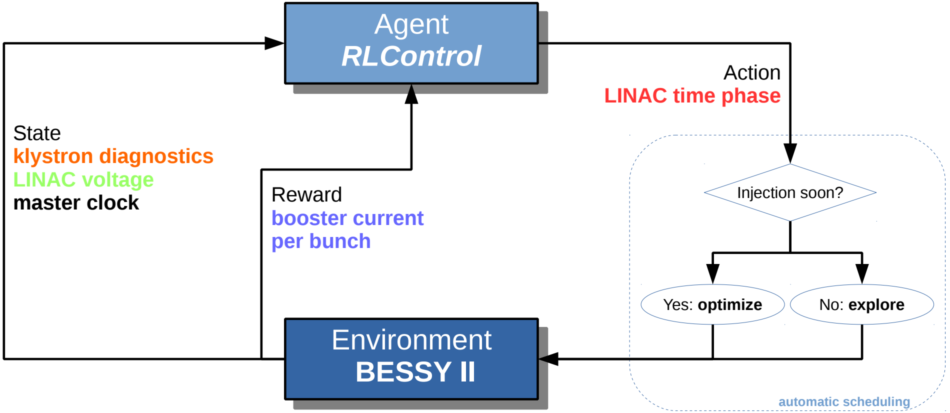
After long interruptions of the machine operation, the booster current tends to be low - as for today, **manual parameter tuning** is required.

We seek an automatized, RL-based solution.

- State variables:
  - High (radio) frequency - master clock.
  - Voltage in LINAC.
  - Klystron current diagnostic measurements.
- Action variable: **time phase in LINAC.** Observations show that this parameter does not affect the injection efficiency.
- Reward: (normalized) **booster current per bunch.**

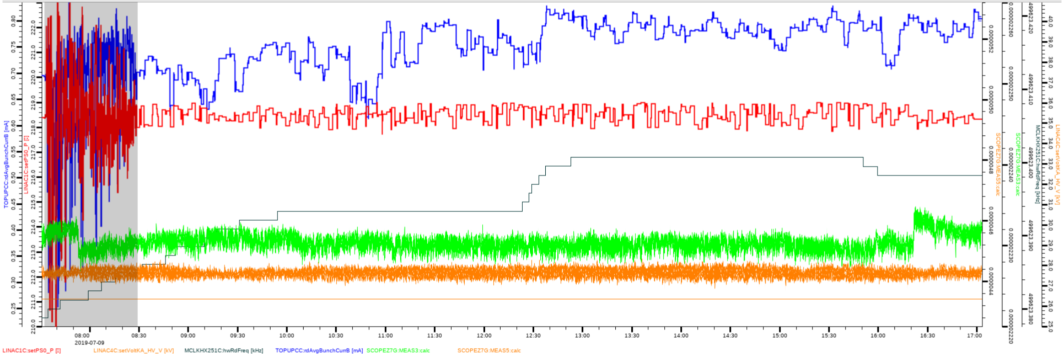


Booster Current Optimization: Automatic scheduling



Exploration is scheduled in the meantime between injections to avoid disturbing user activity - optimization activated shortly before each injection.

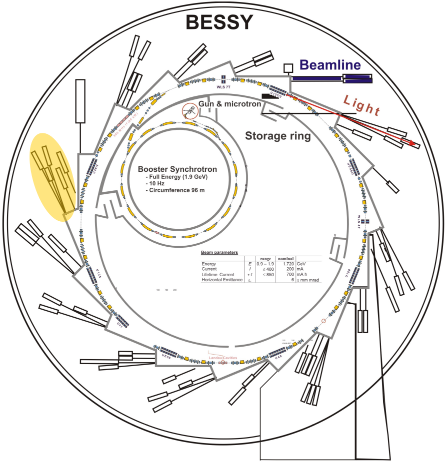
HZB Helmholtz Zentrum Berlin



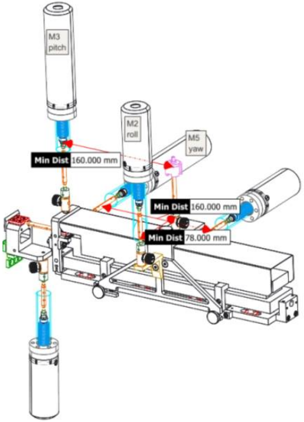
- Reward in blue, action in red - remaining lines correspond to state variables.
- Pretraining with 30 days of historical data. Exploration with automatic schedule shaded - **first hour**.
- The agent optimizes (and learns) successfully during the next **8.5 hours of user operation**.

Table of Contents

HZB Helmholtz Zentrum Berlin



- **UE112 PGM-1 is a low-energy high-flux beamline** with permanent meV RIXS experiment and open port for flexible end stations.
- The RMU (refocussing mirror unit) consists of **two mirrors** (M4 and M5), each parameterized by **3 translations and 3 rotations.**
- Only the 6 rotation axes are **motorized** - translation has to be set manually.



Given ***any* initial translation and rotation configuration**, is it possible to reach a **footprint of size** $\sim 1 \times 20\,\mu$m (disregarding the position at the screen) just **modifying the rotation**?

Beamline Raytracing: Simulation with RAY-UI

HZB Helmholtz Zentrum Berlin

- P. Baumgärtel, M. Witt, J. Baensch, M. Fabarius, A. Erko, F. Schäfers and H. Schirmacher, RAY-UI: A Powerful and Extensible User Interface for RAY, AIP Conf. Proc. 1741, 040016 (2016)
- P. Baumgärtel, P. Grundmann, T. Zeschke, A. Erko, J. Viefhaus, F. Schäfers, and H. Schirmacher, RAY-UI: New Features and Extensions, AIP Con. Proc. 2054, 060034 (2019)

Contact

Dr. Peter Baumgärtel

(030) 8062 - 15154

Email

Prof. Dr. Hartmut Schirmacher

Email

Business card

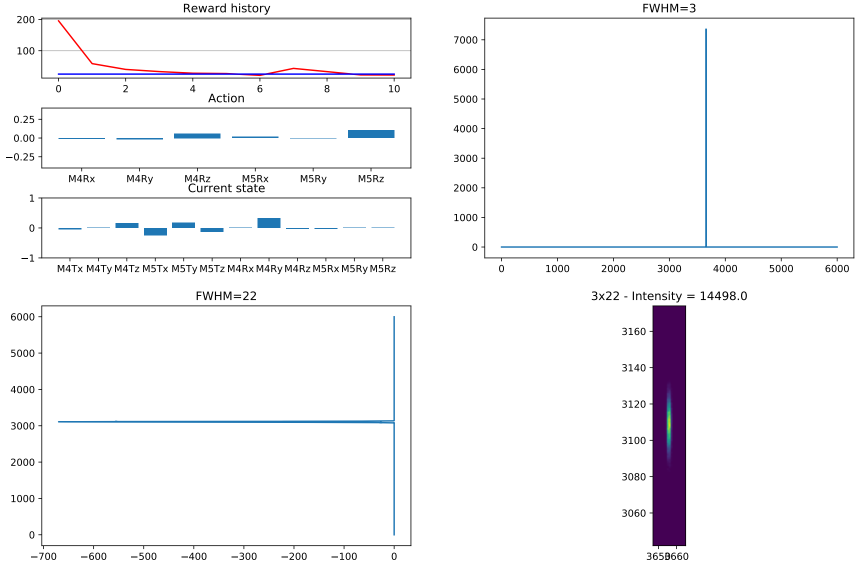
W-i-p:
- CUDA
- Coherence

http://hz-b.de/ray

15

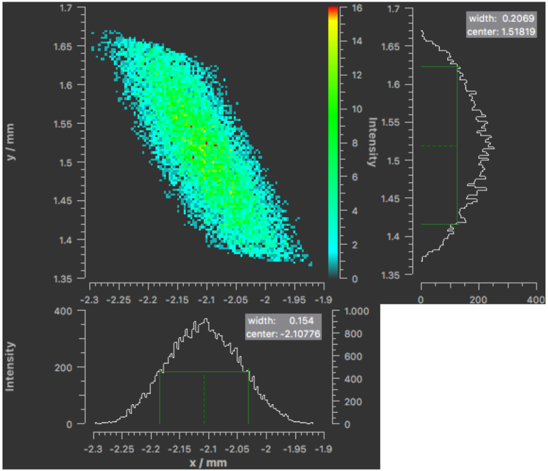
Tests at the real beamline had to be postponed...

Beamline Raytracing: A (nice) test run



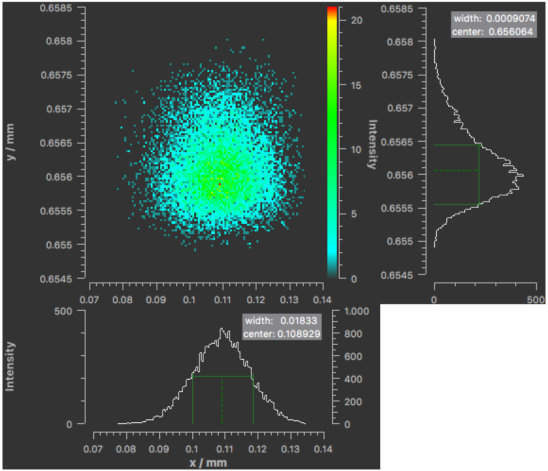
Beamline Raytracing: A (nice) test run - Before and after with RAYUI

UE112-PGM1_rayUI_v12-04_190807_100eV_harm1_Refl_mac; 2020-09-09 17:29



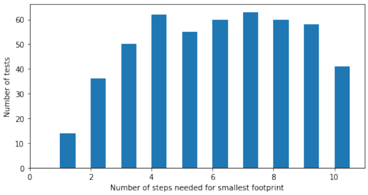
RAY-UI (Changeset ID: c473858ae6f8)

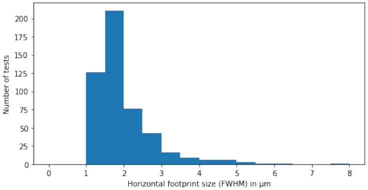
UE112-PGM1_rayUI_v12-04_190807_100eV_harm1_Refl_mac; 2020-09-09 17:31

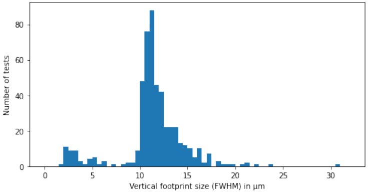


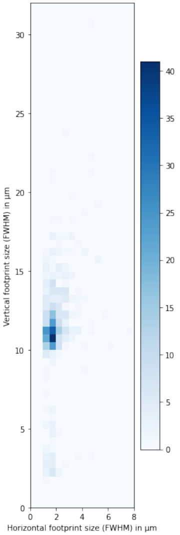
RAY-UI (Changeset ID: c473858ae6f8)

We analyze **499 test runs** of the agent retracing and pick the **best footprint within 10 steps**. The initial position of the 12 parameters (translations and rotations) are sampled from a gaussian with $\sigma = 0.2$ (w.r.t. normalized parameter ranges)

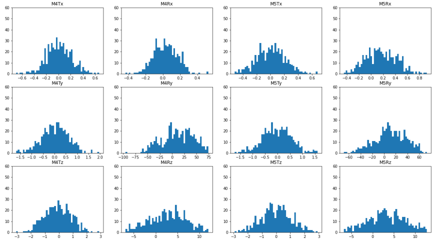


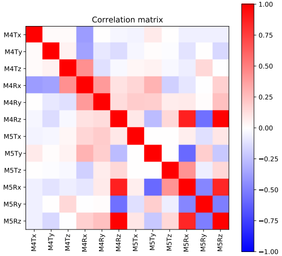




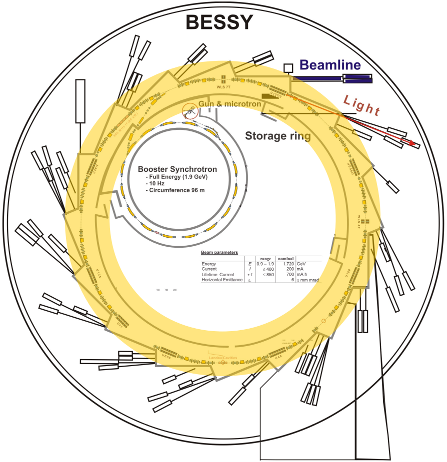


Beamline Raytracing: Correlation analysis of optimal positions



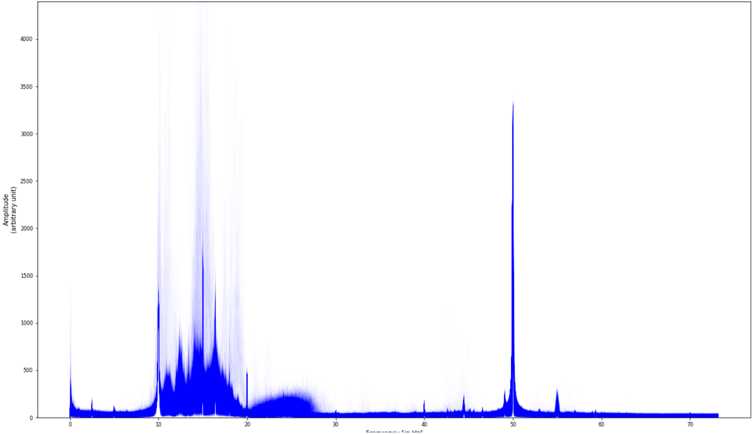


HZB Helmholtz Zentrum Berlin

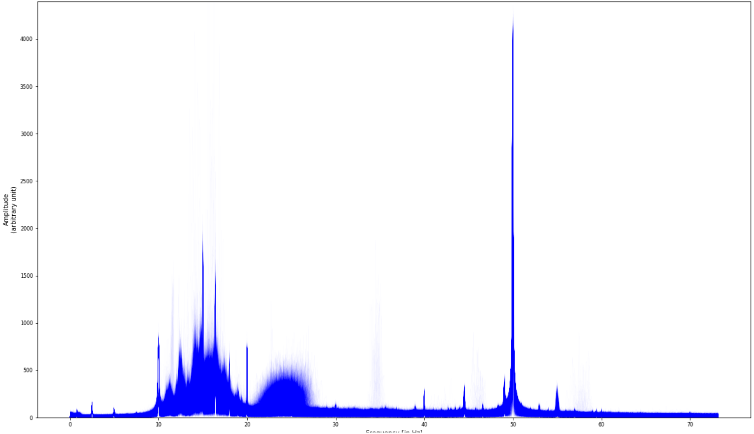


HZB Helmholtz Zentrum Berlin

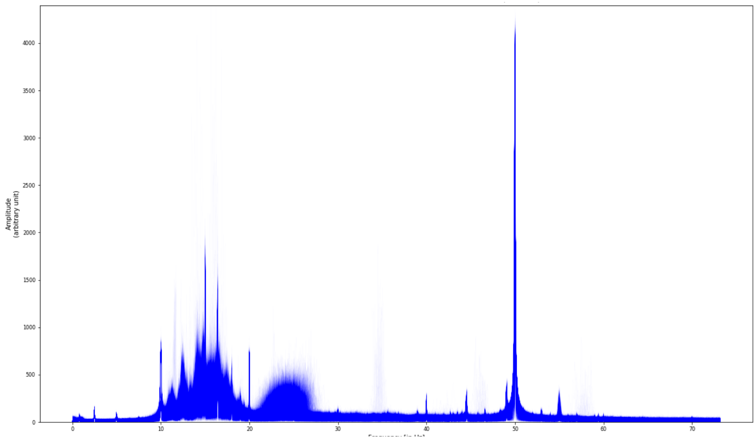
Horizontal beam motion spectrum:



Horizontal beam motion spectrum **with fast orbit correction**:

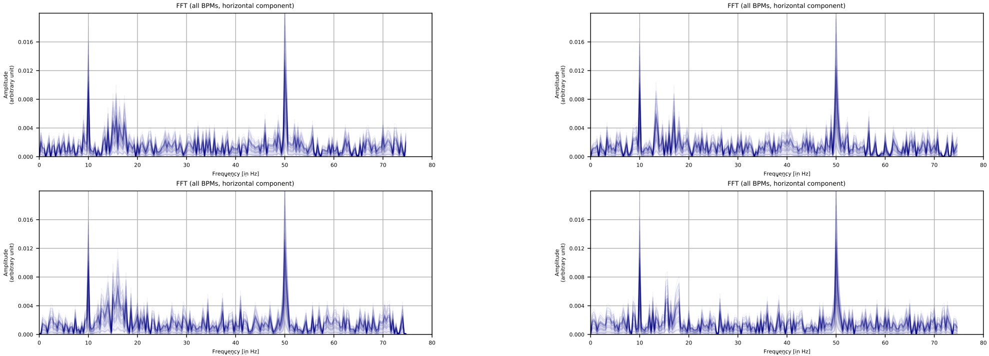


Horizontal beam motion spectrum **with fast orbit correction**:

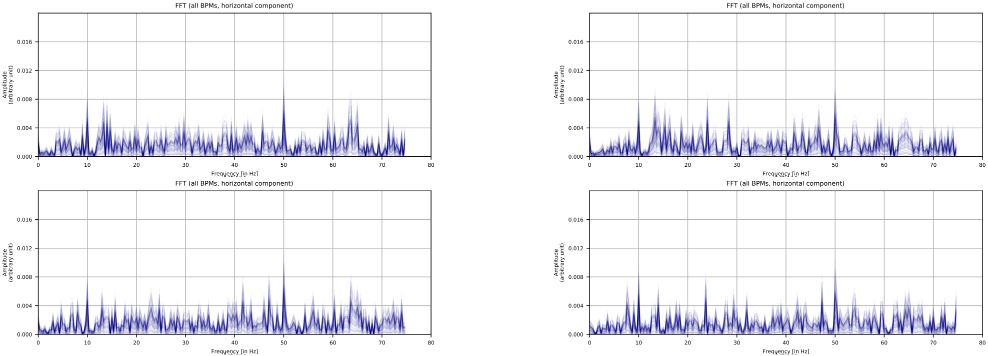


Is it possible to mitigate the perturbations in the range 10-30Hz with a RL Agent?

Mitigation of Harmonic Orbit Perturbations: OCELOT simulations



Simulation of perturbed beam motion spectrum (horizontal quadrupole offset perturbed)



Simulation of perturbed beam motion spectrum **corrected with RL-Agent**

Mitigation of Harmonic Orbit Perturbations: First Tests @ BESSY II

HZB Helmholtz Zentrum Berlin

- In July 2020 we managed to set up the infrastructure for RL-based correction of harmonic perturbations during machine commissioning.

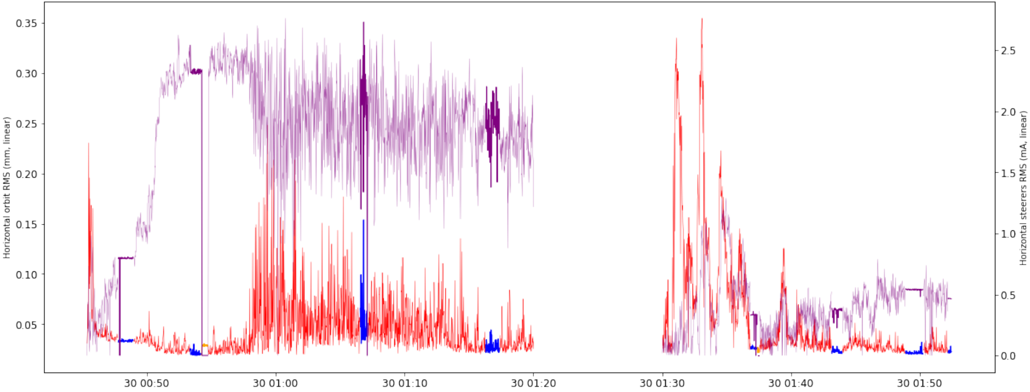
$\rightarrow$ First *plausibility tests* of the Naus-based framework **up to 20Hz** were carried out succesfully.

- In September 2020 we carried out new tests, focussing on the code performance in order to accelerate the interaction loop and so get first meaningful learning results.

$\rightarrow$ A direct zmq-communication with the mBox (fast orbit correction infrastructure) was established, allowing an acceleration of the RL-interaction loop **up to 100Hz**.

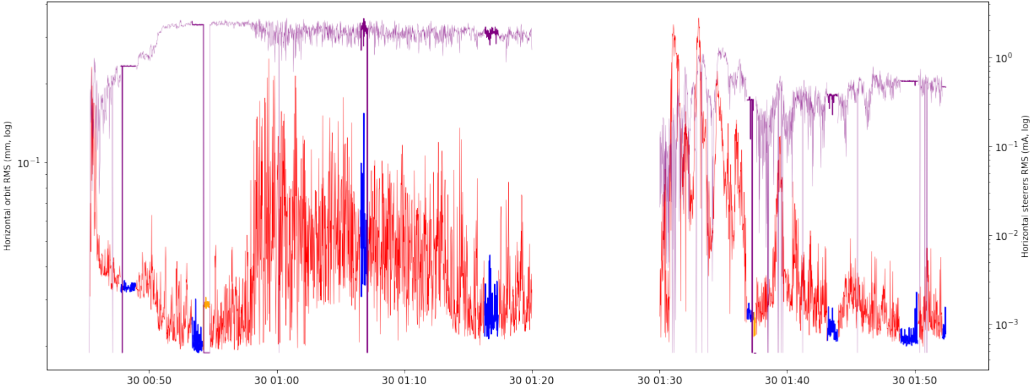
Mitigation of Harmonic Orbit Perturbations: Some test runs (30/09/20)

Comparison of BPM and steerers motion (RMS, Archiver data):



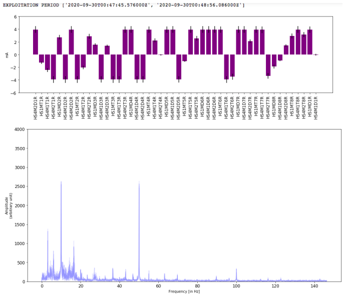
BPM RMS norm: red = exploration, blue = exploitation, orange = agent off (comparison)
Learning rate: learn at every step at $\sim$ 26.6 Hz (left); learn after 50 steps at $\sim$ 100 Hz (right)

Comparison of BPM and steerers motion (RMS, Archiver data):

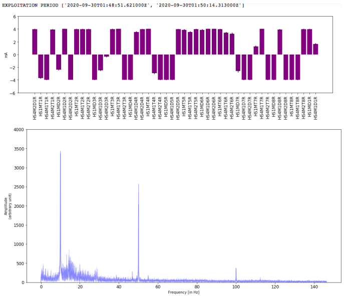


BPM RMS norm: red = exploration, blue = exploitation, orange = agent off (comparison)
Learning rate: learn at every step at ∼ 26.6 Hz (left); learn after 50 steps at ∼ 100 Hz (right)

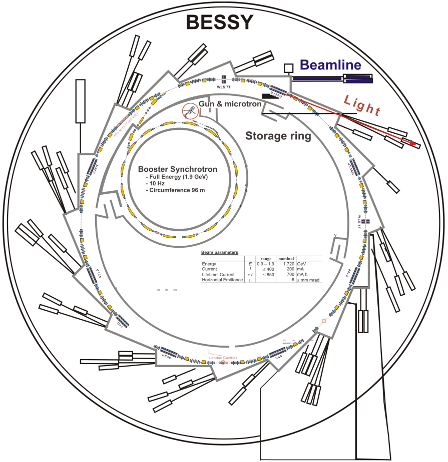
Comparison of perturbed and mitigated spectra (Archiver data):



Beam Motion Spectrum: blue = exploitation, orange = agent off (comparison)
Learning rate: learn at every step at ∼ 26.6 Hz



Beam Motion Spectrum: blue = exploitation, orange = agent off (comparison)
Learning rate: learn after 50 steps at ∼ 100 Hz



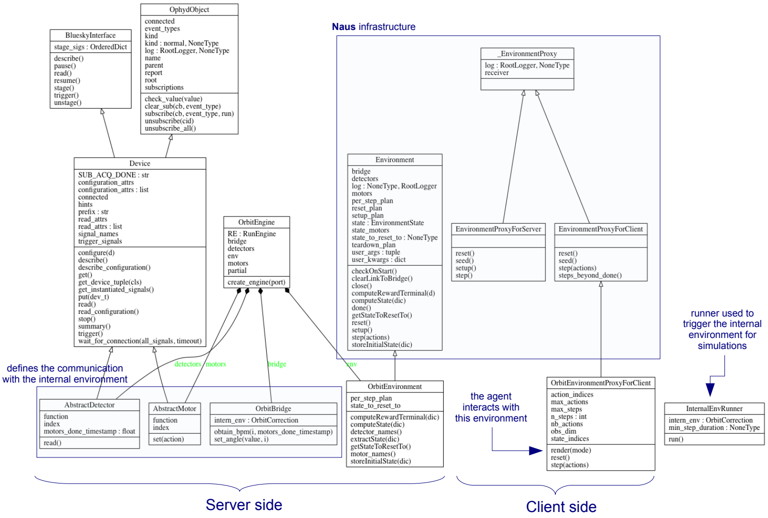
BESSY
Beamline
Light
Gun & microtron
Storage ring
Booster Synchrotron
- Full Energy (1.9 GeV)
- 10 Hz
- Circumference 96 m
Beam parameters
range nominal
Energy E 0.9 – 1.9 1.720 GeV
Current I ≤ 400 200 mA
Lifetime- Current τ·I ≤ 850 700 mA h
Horizontal Emittance ε 6 π mm mrad

Digital Twin

- A computer model of the real machine allowing, among others:
  - simulations with actual machine setting
  - forecast of machine performance
  - linear and non-linear modelling

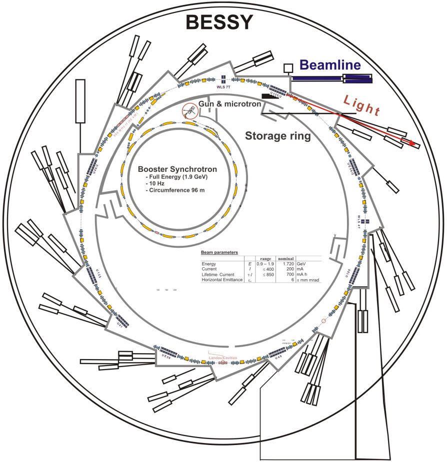
From the ML-perspective it will represent a major improvement $\rightarrow$ faster development since less commissioning time needed, common interfaces...

Digital Twin: First Use Cases - Orbit Correction



We aimed to create an interaction framework whose interfaces remained completely unchanged when training a RL-model with simulations or at the real machine. For this we used **Naus**, based on **bluesky** and **ophyd**

Table of Contents



Outlook → ACCLAIM

- Harmonic Orbit Perturbations - further tests and development
- Roll-out of Anomaly Detection systems
- Further development of the digital twin - ACCLAIM Postdoc position already advertised: `https://recruitingapp-5181.de.umantis.com/Vacancies/1352/Description/2?lang=eng`
- Further projects: e.g. ML for bERLinPro electron gun

Table of Contents

[BKH16] Jimmy Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. *ArXiv*, abs/1607.06450, 2016.

[LHP+16] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

[PHD+17] Matthias Plappert, Rein Houthooft, Prafulla Dhariwal, Szymon Sidor, Richard Y. Chen, Xi Chen, Tamim Asfour, Pieter Abbeel, and Marcin Andrychowicz. Parameter space noise for exploration. *CoRR*, abs/1706.01905, 2017.

[SB18] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018.

[SLH+14] David Silver, Guy Lever, Nicolas Heess, Thomas Degris, Daan Wierstra, and Martin Riedmiller. Deterministic policy gradient algorithms. In Eric P. Xing and Tony Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 387–395, Bejing, China, 22–24 Jun 2014. PMLR.

[ZM18] Xiaoqin Zhang and Huimin Ma. Pretraining deep actor-critic reinforcement learning algorithms with expert demonstrations. *CoRR*, abs/1801.10459, 2018.

Backup: Beam lifetime prediction results

HZB Helmholtz Zentrum Berlin

- Data from 2019-07-01 19:00:00 until 2019-07-16 19:00:00, restricted to **top-up** and **multibunch**.
- 80% (31631 samples) is used for training and 20% for test (7908 samples).
- Tests both with **random** and **chronological** split.
- Baselines:
  - Test set average.
  - Persistence: previous target measurement.
  - *Moving persistence*: moving average of the last 5 target measurements.

Test set	Algorithm	RMSE				R^2		
		Avg.	Pers.	Mov. pers.	Model	Pers.	Mov. pers.	Model
Random 20%	`ExtraTrees`	0.201319	0.099248	0.091464	0.068175 ± 0.000038	0.756961	0.79359	0.885322 ± 0.000128
	`SVR-RFF`				0.077432 ± 0.000216			0.852064 ± 0.000825
	`DNN`				0.069457 ± 0.000342			0.880964 ± 0.001177
Last 20%	`ExtraTrees`	0.231393	0.095732	0.078776	0.194755 ± 0.000952	0.828836	0.884099	0.291586 ± 0.006932
	`SVR-RFF`				0.121407 ± 0.003349			0.724506 ± 0.015291
	`DNN`				0.125046 ± 0.005757			0.707345 ± 0.027032

Backup: DDPG

- **Deep Deterministic Policy Gradient** [LHP+16]: Actor-Critic Reinforcement Learning algorithm for continuous environments.
- The **Q-function** (estimation of the future reward for a given state-action pair) and the **policy** (map between states and actions) are *approximated* with Neural Networks.

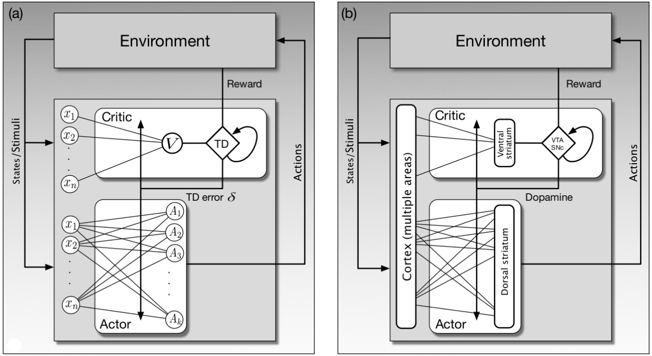


Figure: From [SB18]

HZB Helmholtz Zentrum Berlin

- DDPG needs several implementation tricks: delayed target networks ($Q_{\tilde{\omega}}$, $\mu_{\tilde{\theta}}$), replay buffer...
- **Critic update** with stochastic behavior policy β and loss[1]:

$$L(\omega) = \mathbb{E}_{s_t \sim \rho^\beta,\, a_t \sim \beta}\left[\Big(Q_\omega(s_t, a_t) - (r(s_t, a_t) + \gamma Q_{\tilde{\omega}}(s_{t+1}, \mu_{\tilde{\theta}}(s_t))\Big)^2\right]$$

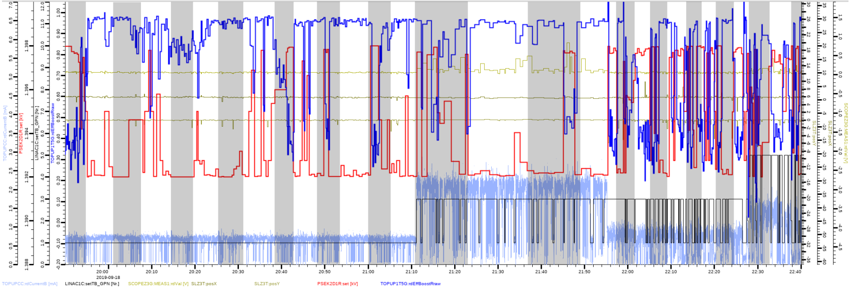
- **Actor update** - (off-policy) Deterministic Policy Gradient Theorem ([SLH$^+$14]): for the performance objective $J_\beta(\mu_\theta) = \mathbb{E}_{s\sim\rho^\beta}\left[Q^{\mu_\theta}(s, \mu_\theta(s))\right]$,

$$\nabla_\theta J_\beta(\mu_\theta) \approx \mathbb{E}_{s\sim\rho^\beta}\left[\nabla_\theta \mu_\theta(s) \nabla_a Q^{\mu_\theta}(s,a)\big|_{a=\mu_\theta(s)}\right]$$

[1] In [SLH$^+$14], $\rho^\beta(s') := \int_S \sum_{t=1}^\infty \gamma^{t-1} p_0(s) p(s \to s' | t, \beta) ds$

Backup: DDPG Test Issues

- Long training time and normalization problems $\rightarrow$ improved through **demonstration with historical data** (inspired by [ZM18]).
- Slow *reaction time* to reward modifications $\rightarrow$ improved with low γ.
- Non-optimal exploration $\rightarrow$ improved with **Parameter Space Noise** ([PHD^{+}17]).



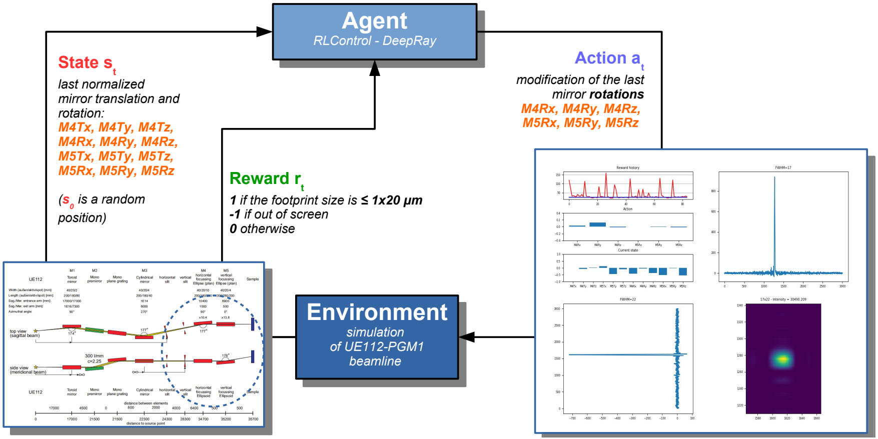
Backup: DDPG Settings for Injection Efficiency Tests

- Neural networks: `relu` used as inner activation function and `adam` as optimizer (lr = 0.001).
  - Critic network: five hidden layers (50+100+50+20+10 neurons) and concatenates actions at the first hidden layer. Linear activation at the output layer.
  - Actor network: three hidden layers (50+20+10 neurons). `tanh` used as activation for the output layer.
- Data preprocessing: $[-1, 1]$ linear normalization, one year historical data reduced to injections.
- Parameter Space Noise: $\delta = 0.01$.
- Training parameters: $\gamma = 0.2$, pretraining with 10000 steps (2000 before actor training), warm-up with 32 steps, target model update rate = 0.1.

Backup: DDPG Settings for Booster Current Tests

- Neural networks: `relu` used as inner activation function and `adam` as optimizer (lr = 0.001).
  - Critic network: five hidden layers (25+50+25+10+5 neurons) and concatenates actions at the first hidden layer. Linear activation at the output layer.
  - Actor network: three hidden layers (25+10+5 neurons), all of them with layer normalization ([BKH16]). `tanh` used as activation for the output layer.
- Data preprocessing: $[-1, 1]$ linear normalization, historical data downsampled to 60 seconds.
- Parameter Space Noise: $\delta = 0.01$.
- Training parameters: $\gamma = 0.2$, pretraining with 10000 steps (2000 before actor training), warm-up with 32 steps, target model update rate = 0.1.
- *Brute-force* synchronization: update every **2 seconds** through EPICS.

Backup: Beamline Raytracing RL Loop



- **Step**: rotation update (with fixed maximum step size) and raytracing
- **Episode**: end after a fixed number of steps or after getting reward ± 1

Backup: DDPG Settings for Beamline Raytracing

- Neural networks: `tanh` used as inner activation function and `adam` as optimizer (lr = 0.001).
  - Critic network: five hidden layers (250+500+250+100+50 neurons) and actions concatenated at the first hidden layer. Linear activation at the output layer.
  - Actor network: three hidden layers (250+100+50 neurons).
- Data preprocessing: $[-1, 1]$ linear normalization.
- Parameter Space Noise: $\delta = 0.05$.
- Training parameters: $\gamma = 0.99$, warm-up with 32 steps, target model update rate = 0.01

Backup: Configuration - Orbit Simulation Tests

Environment:

- Simulated perturbation applied to **Q4M2D1R.dx** (horizontal offset of the quadrupole) with resolution = 150Hz
- x component of the BPM (beam position monitor) **BPMZ6D1R** read with **windows size = 30** ($\rightarrow$ **state**)

$\rightarrow$ the RL Agent *stays in the time domain!*

- Horizontal steerer **HS4M2D1R** modified ($\rightarrow$ **action**)
- The **reward** is defined through $2e^{-c\sqrt{\frac{\sum_i x_i^2}{N}}} - 1$, where x_i denotes the horizontal component at each BPM, N is the number of BPMs and c is a normalization constant (in these simulations $c = 10000$)

HZB Helmholtz Zentrum Berlin

DDPG agent:

- Actor network with 500-200-100 neurons, `relu` as inner activation function (output with `tanh`)
- Critic network 500-(500+action)-200-100 neurons, `relu` as inner activation function
- $\gamma = 0.9$
- Exploration with
  - Ornstein-Uhlenbeck process
  - Parameter Space Noise ([PHD$^+$17])

Backup: DDPG Settings for Orbit Tests at BESSY II

DDPG agent:

- Actor network with 1250-500-250 neurons, `relu` as activation function (output with `tanh`)
- Critic network 1250-(1250+action)-500-250 neurons, `relu` as activation function
- $\gamma = 0.1$
- Exploration with Ornstein-Uhlenbeck process

Environment - BESSY II:

- **State**: all active BPMs (102) with window size 10.
- **Action**: all horizontal steerers (48) modified up to $\pm$ 4 mA.
- **Reward**: exponential transformation of the BPM norm: $2e^{-10\sqrt{\frac{\sum_{i=1}^{102} x_i^2}{102}}} - 1$