RL4AA'26

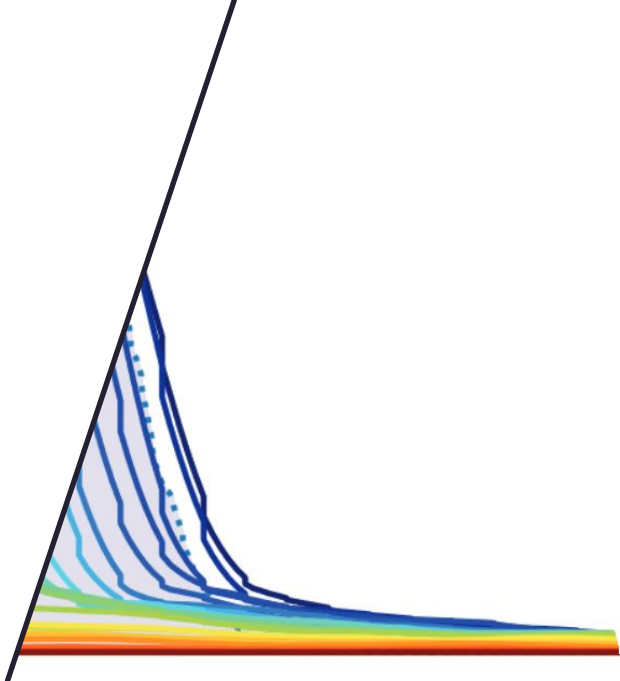
Dreaming of Schottky Spectra: Building World Models for LEIR robust automation

Borja Rodriguez Mateos
borja.rodriguez.mateos@cern.ch

CERN

Acknowledgements: Theodoros Argyropoulos, Michael Schenk, Verena Kain, Joel Wulff, Adrián Menor de Oñate, Francesco Maria Velotti

March 30, 2026



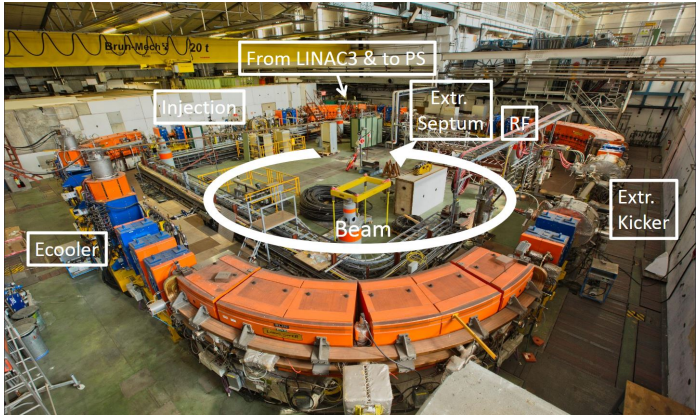
Agenda

1. Introduction
2. World Modelling
3. Building the World Model
4. Offline training
5. Online results
6. Discussion

1. Introduction

1.1 The Low Energy Ion Ring (LEIR)

A practical machine for data-driven optimization



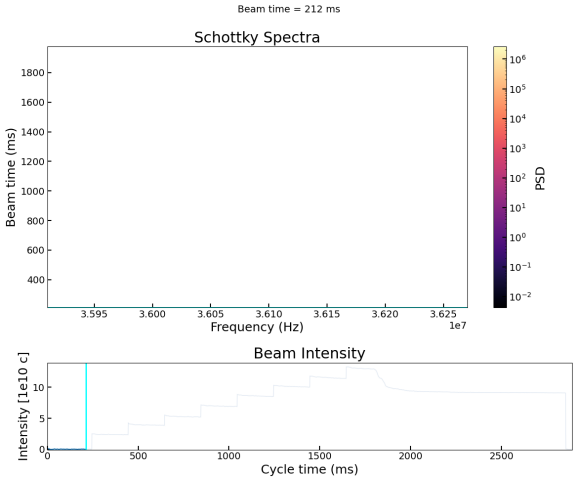
Supervision and operation of the machine:

- Complex system by design
- Many hours of manual maintenance/recovery of performance
- High repetition rate
- Low energy beam

→ Some concepts will be skipped here, you can check them out in last year talk here

1.2 High intensity Pb^{54+} beams

Beam accumulation in LEIR



High intensity Lead beams are obtained through:

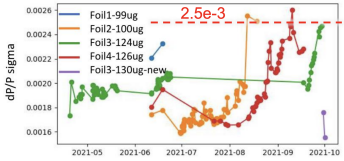
- Accumulation of eight 200ms spaced pulses from the linac 3 in the ring
- 6 dimensional phase space painting through a multi-turn injection
- Circulating beam is cooled and dragged in phase space for accomodating for the next incoming pulse

1.3 Aging of the Stripper Foil

How to compensate for Stripper foil aging

Stripping foil has a limited lifespan and degrades over time

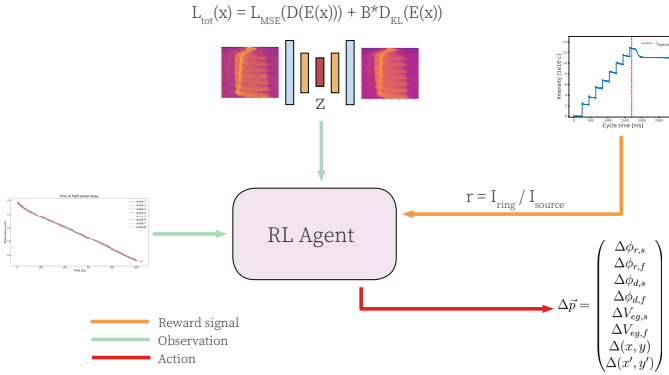
- Impacts the mean ion energy and changes the energy distribution width
- Decreases the injection efficiency over time and needs to be compensated manually
- Foil changed when $\frac{\Delta p}{p}$ standard deviation exceeds 2.3 x 10^{-3}
- Performance drop can be attenuated by tuning ramping and debunching cavities and adjusting the cooling



Pb^{54+} produced : $1.6x10^{15}$ $1.3x10^{15}$ $1.0x10^{15}$ $1.1x10^{15}$

1.4 Reinforcement Learning at LEIR

Automating stripper foil aging compensation



- RL agent observes encoded Schottky spectra and ToF measurement
- Tries to maximize injection efficiency from start of transfer line in Linac 3 to the ring before capture, i.e. $r = \frac{i_{\text{ring}}}{\sum_{j=1}^{8} i_{\text{pulse},j}}$
- Acts on debunching and ramping cavities, cooler bump and electron gun voltage (9 parameters)
- Can use pre-trained agents or train online

2. World Modelling

2.1 Definition

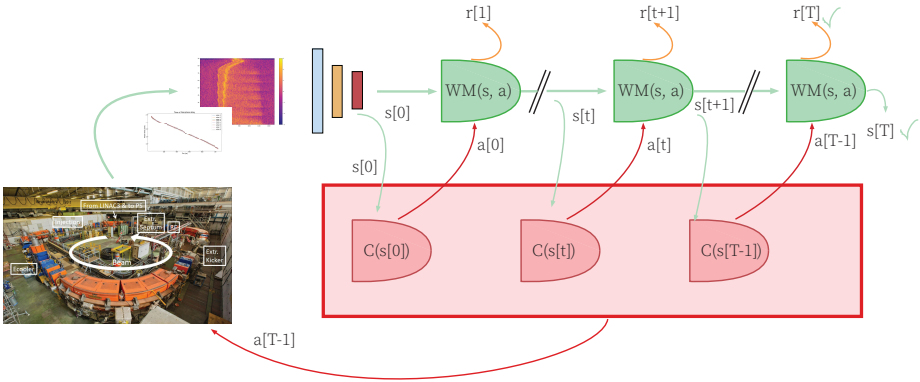
What are World Models?

- World models is a broad term that refers to any learned model of the environment's dynamics
- It approximates the transition and reward function of the environment
- Predicts next state and reward given current state and action
- Can be promising for robust and sample efficient autonomous control of particle accelerators. With a good World Model:
  - One can add guardrails and check for safety constraints in the learned world
  - Plan decisions in advance without consuming machine time

"An internal model of external reality that allows an agent to try out alternatives and react to future situations before they arise" — "The Nature of Explanation", Craik, 1943

2.2 World Modelling at LEIR

Using experimental data for world model building



- World Model $\mathrm{WM}(s_t, a_t) = s_{t+1}, r_{t+1}$ predicts the next state and reward
- Controller $C(s_t)$ outputs best action for observed state that maximizes return
- After achieving satisfying reward after T timesteps in the World Model the controller applies it to the machine

2.3 World Model flavours

What World Models are we testing?

MDN-RNN:

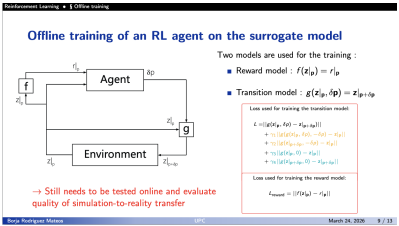
- VAE to compress high-dimensional Schottky spectra into a low-dimensional latent code z
- MDN-RNN to predict future latent states as a mixture of Gaussians
- Original implementation of World Models generative NN
- Suffers from hallucinations

Diffusion based:

- Conditional diffusion model to generate future state-reward trajectories, conditioned on the current state, action and return-to-go
- Actions are not jointly generated by the diffusion model but by a separately trained inverse dynamics model (IDM) reconstructs missing actions
- Controller is a standard one-step TD algorithms namely TD3BC

Decoupled MLP based Reward and Transition Model:

- Two separate MLP maps (encoded observation, action) $\rightarrow$ next encoded observation and encoded observation $\rightarrow$ reward



3. Building the World Model

3.1 Data collection for surrogate model

Exploration of the 9-dimensional parameter space

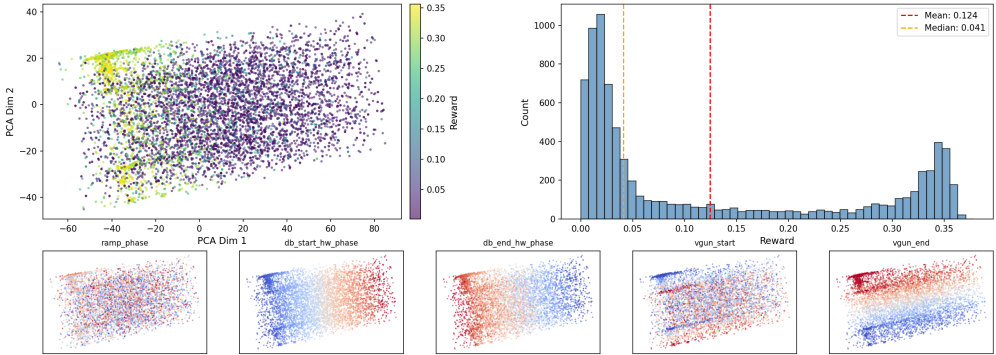
	Sobol	Active Learning	High β-UCB	Total
Number of samples	$\approx$ 4500	$\approx$ 3500	$\approx$ 1500	9500

Number of samples drawn from each sampling strategy to build the data-driven surrogate model

- Active Learning is based on trying to minimize the integrated posterior variance of a model of the objective computed with MC integration
- Sobol sequences generate low discrepancy quasi-random sequences of points
- High β-UCB acquisition function to have a "finer mesh" around points with better injection efficiency
- Need to collect data with different stripper foil ages for it to be coherent
- Dataset best injection efficiency: $r = 0.365$
- Even for the most similar (z, a) pairs in the training set (nearest neighbours with distance ¡ 0.003), the reward varies with a spread of 0.017 IE

3.2 2-Dimensional data representation

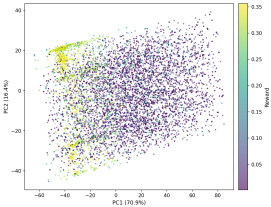
Principal component analysis

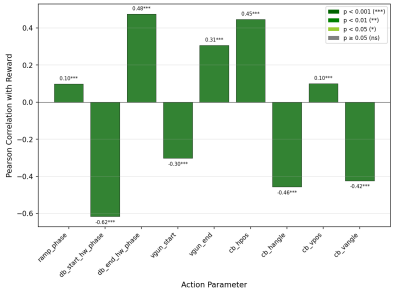


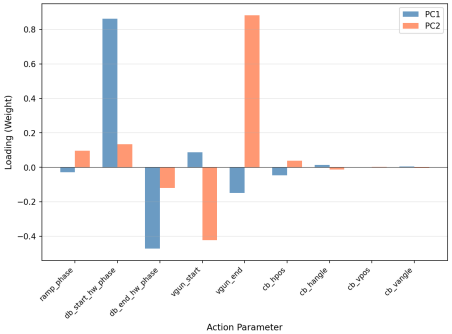
The debunching cavity and the end of the electron gun voltage profile seems to be the most impactful on the reward value

3.3 Statistical significance of parameters

Principal component analysis





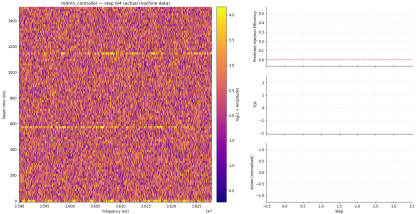


All parameters are statistically significant ($p < 0.05$)

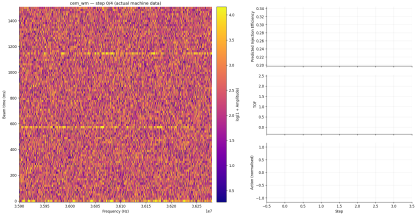
4. Offline training

4.1 Dreaming of Schottky Spectra

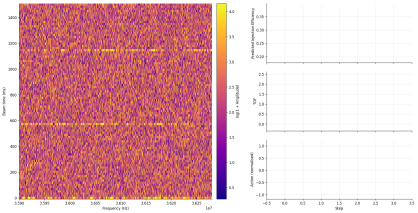
Training in the World Model



MDN-RNN + CMA-ES



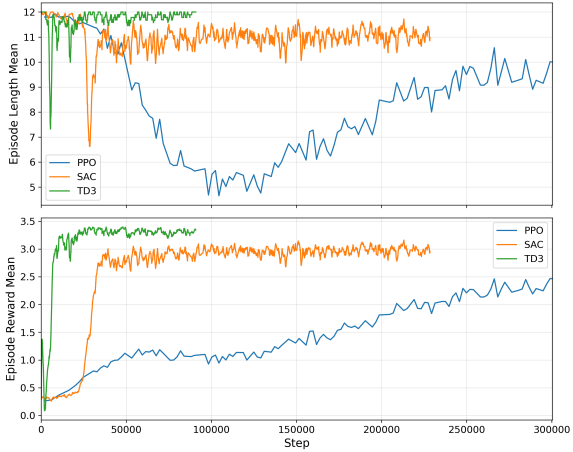
DAWM + CEM

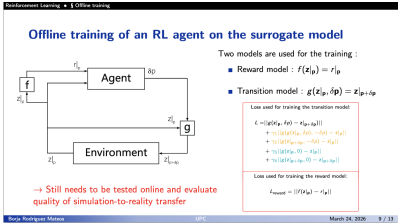


DAWM + TD3BC

4.2 Offline training on decoupled

Performances on the data-driven surrogate model

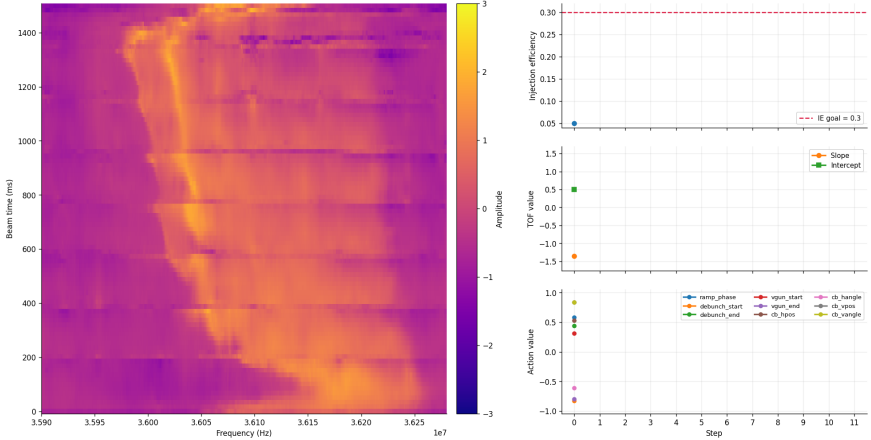




- Lower discount factor due to the setting of the problem ($\gamma = 0.73$)
- Maximum number of steps of 12 per episode

4.3 Offline training Visualization

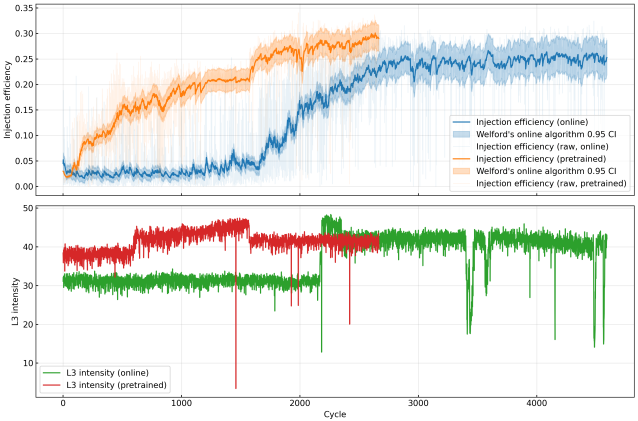
Evaluation of the SAC agent



5. Online results

5.1 Online training the RL agent

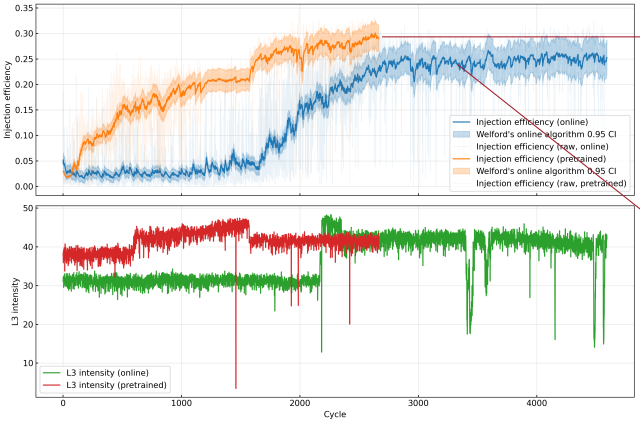
Comparison of online training vs pre-trained on a surrogate

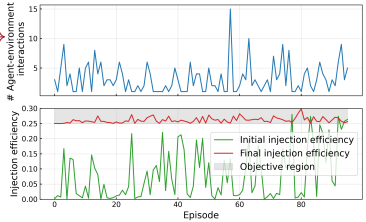


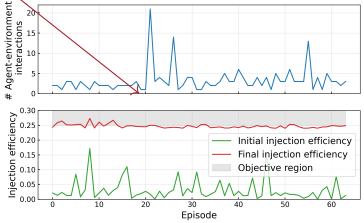
- Tested training online and pre-trained with the decoupled reward and transition NN World Model
- Soft Actor Critic (SAC) agent is used in both cases with same hyperparameters with similar ageing of the stripper foil
- Improved performance and sample efficiency by pre-training it on the world model

5.2 Evaluation of the trained agent

Comparison of online training vs pre-trained on a surrogate







In how many iterations can the agent re-establish beam performance from a random parameter configuration?

6. Discussion

6. Discussion

Things left to do in 2026 for World Modelling

- Need to test online the static diffusion based world model during 2026 run starting in April
- Add feedback from accelerator to improve world model based on current accelerator observations
- Increase training data and capacity of World Model to other past configurations (Pb data from 2023, 2024,...)
- Augment World Model to several types of other heavy ions tested at LEIR (Mg, O, ...)

Borja Rodriguez Mateos

Geneva, March 30, 2026

borja.rodriguez.mateos@cern.ch