Grant Agreement No: 101057511

EURO-LABS

EUROpean Laboratories for Accelerator Based Science

HORIZON-INFRA-2021-SERV-01-07 Project EURO-LABS

DELIVERABLE REPORT

Toolkit for optimizing beam delivery (deployed in two facilities)

DELIVERABLE: D5.4

Document identifier:	EURO-LABS-D5.4
Due date of deliverable:	End of Month: AUGUST 2024
Report release date:	01/08/2024
Work package:	WP5 Open, Diverse and Inclusive Science
Lead beneficiary:	GSI
Document status:	Final

Abstract:

*Within EURO-LABS, the "**G**eneric **O**ptimization **F**ramework and **F**rontend" (Geoff[1]) has been deployed in two facilities (CERN and GSI) and has been used for beam optimization as part of Task 5.3. This report focusses on the first results of using Geoff at GSI for automatic optimization for manipulations of various beam species. A short summary of application of Geof at CERN is also given. Geoff has also been so far used successfully for beam steering and beam focusing at the GSI **FR**agment **S**eparator (FRS) used for optimizing radioactive ions beams production. The application of Geoff at GSI has also resulted in the reduction of beam losses of the multi-turn injection into the SIS18 synchrotron from 40% to 15%. Further the injection into SIS18 optimised with Geoff takes ~15 minutes, compared to manual adjustment that can take up to 2 hours. The next*

[1] *Geoff is available under the GNU Public License Version 3 (GPLv3) and hosted on CERN's Gitlab server. The current address is: "https://gitlab.cern.ch/geoff/geoff-app/". It can be reached from EURO-LABS webpage through the link https://web.infn.it/EURO-LABS/results/*

step will be the implantation of Geoff at the laser-driven particle accelerator facility LPA-UHI100 located at CEA-Paris Saclay.

For more information on EURO-LABS, its partners and contributors please see https://web.infn.it/EURO-LABS/

The EUROpean Laboratories for Accelerator Based Science (EURO-LABS) project has received funding from the European Union's Horizon 2020 Research Infrastructure (RI) services advancing frontier knowledge under Grant Agreement no. 101057511. EURO-LABS began in September 2022 and will run for 4 years.

	Name	Partner	Date
Authored by	Sabrina Appel, Nico Madysa, Sandrine Dobosz	GSI, CEA	25/07/2024
Reviewed by	Maria Borge [WP coordinator]	CSIC	29/07/2024
Approved by	Navin Alahari [Scientific coordinator]	GANIL	01/08/2024

TABLE OF CONTENTS

1. INTRODUCTION

EURO-LABS is a network of 33 research and academic institutions from 18 countries (25 beneficiaries and 8 associated partners) from European and non-EU countries, involving 45 Research Infrastructures in the Nuclear physics, Accelerators and Detectors for high energy physics pillars. The project brings together, for the first time, the three research communities of nuclear physics, accelerator and detector technologies for high energy physics, in a pioneering super-community of sub-atomic scientists. It provides effective access to a network of 45 Research Infrastructures (including 3 RIs with Virtual Access) to conduct curiosity-based research, addressing fundamental questions and technological challenges and advancing projects with broad societal impact, fostering knowledge sharing between scientific fields and enhancing Europe's potential for successfully facing future challenges.

One of the main goals of EURO-LABS is to enhance the competitiveness of our research infrastructures and their technical capabilities. This is done with WP2-4 under service improvements. Within WP5, Task 5.3 aims to use Machine Learning (ML) methods to improve beam characteristics, transport efficiency and reproducibility of accelerator tuning, which will reduce the beam preparation time and thus the available time for the experiments. The challenges of applying **M**achine **L**earning (ML) algorithms for accelerators are common and are addressed at several facilities. One part of this automation effort is to provide a framework that allows machine experts, operators and users to solve certain, focused optimization problems and to make these solutions reusable in an operational context. We call this project the "**Generic Optimization Framework and Frontend**", or **Geoff** for short. This document reports on the Deliverable D5.4, which involves the deployment of the new toolkit for beam optimization at least two facilities, CERN and GSI. EURO-LABS finances one scientific staff member at GSI for three years for deployment and implication of Geoff for improved beam delivery at EURO-LABS facilities.

2. THE ML OPTIMIZER TOOLKIT

Geoff (Generic Optimization Framework and Frontend) is an open-source toolkit created at CERN and later introduced at GSI. A collaborative development is continuing between the two institutes. Geoff reduces the complexity of combining many different optimization and ML algorithms with many facility-specific optimization problems. It puts very few restrictions on the formulation of optimization problems and hence is usable for a large variety of cases.

Geoff is based on Python, a programming language that is widely used in scientific research, has a vibrant ecosystem of machine learning algorithms, and is perceived as very beginner-friendly. Both Python and Geoff have proven themselves flexible enough to be quickly adapted to new situations. Unlike in older frameworks, optimization problems in Geoff can be arbitrarily complex; they can use any control systems and even communicate with external simulation tools, as long as they have Python bindings. It is usually embedded into a graphical application, but can also be used in command-line scripting. This implies that the presence of physicists and machine experts in the control room would be required less frequently.

Geoff is available on CERN servers[2] and licensed under the GNU Public License Version 3 (GPLv3). The CERN project page hosts the code, installation instructions, a bug reporting tool (“issue tracker”) and a wiki in which frequently asked questions are answered.

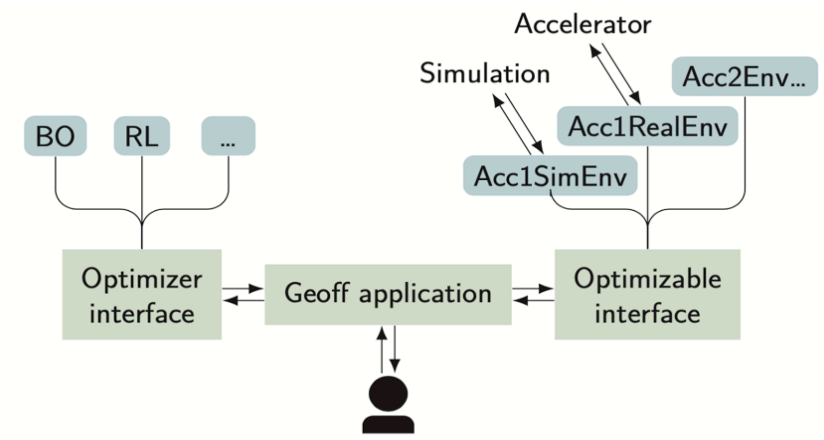

Figure 1: Model of Geoff and its components. The optimizer interface includes BOBYQA, Bayesian optimization (BO), and ***R****einforcement* ***L****earning (RL).*

Geoff conceptually has three components. The first is a program library that provides bindings to a number of well-known optimization and ML libraries, for example: SciPy, Stable Baselines 3, Scikit-Optimize. Thanks to a uniform interface design (see Figure 1, “Optimizer interface”), the user can switch between algorithms without having to adjust any part of their own program.

The second component is a standardized programming interface for optimization problems (see Figure 1, “Optimizable interface”). This encapsulates the problems’ technical details and intricacies so that they can be solved by a generic application. Although Geoff is still in development, experience at CERN shows that the integration of new accelerators and even non-trivial optimization problems is easy and possible within only a few hours. The interface is extensively documented online[3]. A number of written tutorials and examples make it easy to introduce the framework to technical and scientific staff having a moderate programming experience.

The third component is an application with a Graphical User Interface (GUI) that unifies the first two components and exchanges messages between them (see Figure 1, “Geoff application”). This reduces complexity, as optimization problems no longer have to be adapted to each optimization algorithm. This speed up the experimentation and evaluation of algorithms and optimization problems.

The framework is modular and encourages extension and customization to adjust it to each facility's specific needs. For example, the GUI allows optimization problems to generate graphs of their choice that are updated on-line during optimization to monitor the progress. Optimization problems can

[2] https://gitlab.cern.ch/geoff/geoff-app/

[3] https://cernml-coi.docs.cern.ch/

further be made configurable, i.e. declaring settings that the user can change in the GUI before optimization. In such cases, the GUI automatically generates a dialog window in which the user can make these changes.

While solving a new optimization problem, Geoff can also utilize command-line scripting and run in a terminal window to shorten the loop between coding modifications and testing. Figure 2 shows an optimization run that used Geoff for beam steering in the TK[4] transfer channel at GSI. Custom figures can be shown and updated continuously to monitor the algorithm's progress. Shown are the evolution of the measure to be minimized (Figure 2, left, "objective function"), the measured and fitted beam profile at step 25 (Figure 2 right), and the evolution of the optimization parameters (Figure 2, bottom, "actors").

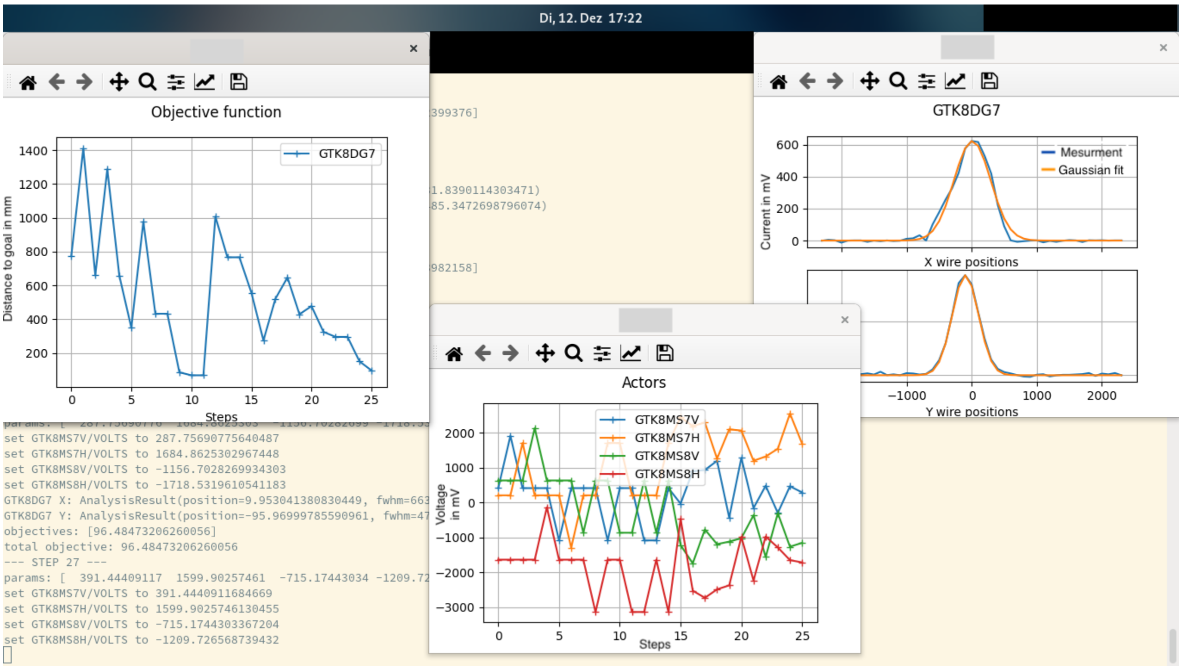

Figure 2: Screenshot of a typical Geoff optimization run in a scripting context: the terminal window with logging messages is in the background, live graphs for monitoring in the foreground.

In the beginning of 2024, development effort on Geoff has been focused on reducing its accrued "technical debt", i.e. to implement updates of its code that have been forgone in 2023 to focus on its introduction at GSI. In particular the software library Gym[5], which Geoff greatly depends on for compatibility with state-of-the-art machine learning algorithms[6], had undergone massive changes, a change of ownership and renamed to Gymnasium[7]. Because of Geoff's intimate relationship with this library, large changes and thorough testing were necessary to ensure a smooth upgrade for all users of Geoff at CERN and GSI. A detailed migration guide has been added to Geoff's documentation to inform its users concisely of all changes that they have to make to their own code to ensure compatibility.

[4] TK is short for German *TransferKanal* ("transfer channel").
[5] https://github.com/openai/gym/
[6] https://stable-baselines3.readthedocs.io/
[7] https://gymnasium.farama.org/

3. APPLICATION OF THE TOOLKIT AT CERN

At CERN, Geoff has been instrumental in optimizing various aspects of accelerator operation. It is now used at almost all CERN accelerators, including Linac 3 and 4, PSB, PS, SPS, LEIR and ISOLDE. Some of the experiments connected to the CERN accelerator complex are part of EURO-LABS project such as ISOLDE and others. Although Geoff is chiefly used as an expert-level tool, operators also use it to solve several daily tasks to reach and maintain a stable operational state of the accelerator. Its integration into the CERN control room has been seamless. Adapting it to new accelerators and optimization tasks often requires only a few hours. This rapid deployment capability has proven beneficial in maintaining high levels of operational efficiency.

At the Super Proton Synchrotron (SPS), Geoff is used for several optimization tasks. One of them is the "ZS alignment", a crucial process for slow extraction. It consists of the precise positioning of seven electromagnetic septa using step motors. Historically, this task required substantial time and effort, 8 hours of manual tuning being rather common. In 2018, the introduction of the Powell algorithm significantly reduced this time to ~45 minutes. In 2021, the BOBYQA algorithm was introduced via Geoff, which brought the time down to just 10 minutes. The reduction in optimization time has had a notable impact: septa alignment is now conducted more frequently, enhancing the overall performance and reliability of the SPS. At the Proton Synchrotron (PS), Geoff is used to run a Reinforcement Learning (RL) algorithm to correct the RF phase and voltage for producing uniform RF splitting in PS for LHC beams. This algorithm has achieved operational state in 2024 (Figure 3). The RL algorithm was trained extensively on simulation data before being successfully transferred to the control room. The core of this system is a **S**oft **A**ctor-**C**ritic (SAC) algorithm, a state-of-the-art RL approach.

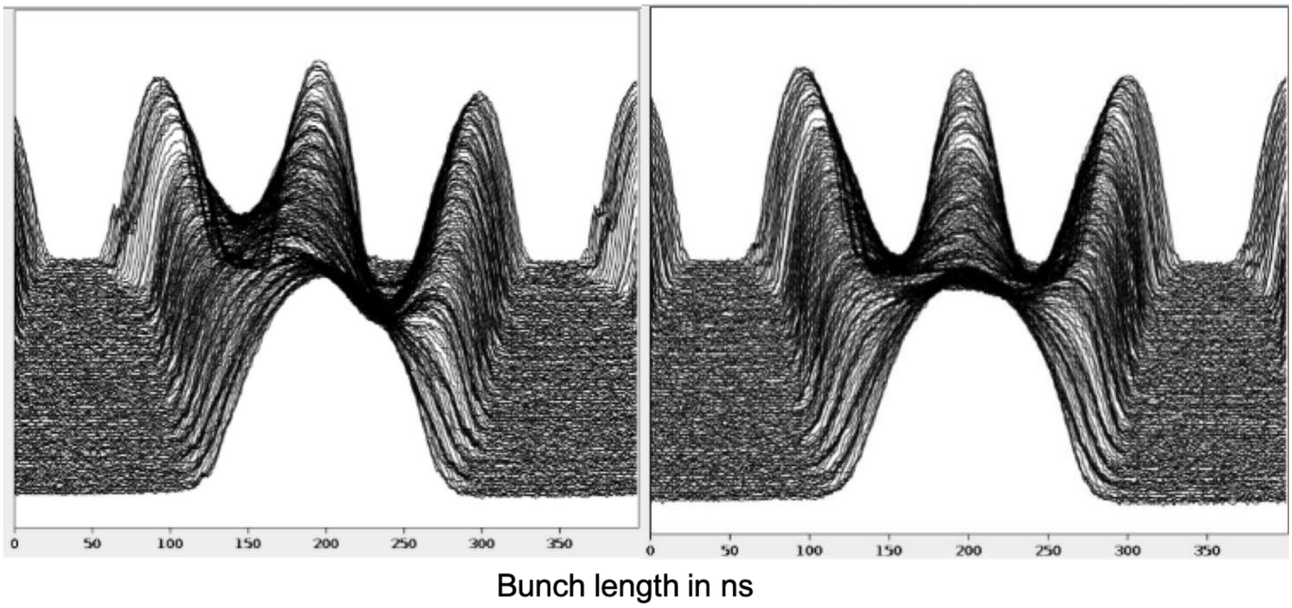

*Figure 3: T*he RL algorithm corrects RF phase and voltage to produce uniform RF splitting in the PS for LHC beams (left before, right after RL correction).

Further examples for successful use cases of Geoff at CERN are

- the automatic drift compensation in the PS multi-turn extraction,

- automatic trajectory steering for beam transfer from PS to SPS and from SPS to LHC,
- fixed target beam steering at PS.

Several more cases where Geoff can be used are currently being investigated both by doctoral students and post-doctoral fellows at CERN.

At present, optimization problems are solved semi-autonomously, i.e. while the algorithm runs from start to completion on its own, it still has to be invoked by an operator, e.g. via a graphical application. Current development efforts at CERN are focused on enabling a fully autonomous operational mode, i.e. controllers that automatically detect when to optimize accelerator settings.

4. APPLICATION OF THE TOOLKIT AT GSI

4.1 MULTI-TURN INJECTION OPTIMIZATION

The Facility for Antiproton and Ion Research (FAIR), will provide antiproton and ion beams of unprecedented intensity and quality to drive the forefront of research using heavy ions and antimatter. The Multi-Turn Injection (MTI) into SIS18 is one of the main bottlenecks to reach the FAIR intensity goals. Because the SIS18 will serve has a booster for the SIS100, hence this will affect the performance of FAIR facility. An important limiting factor for the intensity of intermediate-charge-state ion beams is beam loss-induced vacuum degradation. Because beam losses during injection can trigger the vacuum degradation, injection losses must be minimized. Beam losses during injection may occur both at the septum and at the acceptance of the accelerator. Because the MTI must satisfy Liouville's theorem, four bumper magnets (Figure 4, illustration on the top) create a closed orbit bump with a time variable such that the injection septum deflects the next incoming beam into available horizontal phase space that is close to the previously injected beam.

In November 2023, May 2024 and June 2024, experimental optimization runs were performed at GSI in order to evaluate the use of Geoff. The objective of the optimization was to minimize injection losses in the SIS18 by adjusting five injection parameters and four steerers in the TK. The steerers are used to deflect the beam in the TK to the injection point. The five injection parameters are the initial orbit bump amplitude, the reduction of the orbit amplitude, two additional steerers near the septum and a time offset, which aligns the incoming beam's arrival time. To ensure that optimization always started in a sufficiently bad state and the optimization parameters were randomized before each run. The loss (see Figure 4, right) was estimated by measuring the beam current in the TK (black-dotted), calculating the ideal SIS18 current given a perfect injection (green), and subtracting the measured SIS18 current (black-dashed line). The optimization algorithm of choice was BOBYQA. BOBYQA maintains a quadratic model of the objective function, which must be initialized with $N \geq 2n+1$ evaluations (grey area), where $n=9$ is the number of parameters to be optimized and therefore $N=19$, in Figure 4 on the left. To reduce fluctuations in the objective function, each iteration took the median loss of three measurements.

During the November 2023 run, the beam loss was 45 % after manual tuning by experienced operators; after automatic optimization of all injection parameters, it went down to 15 %. During the May 2024 run, beam was 35 % after manual tuning; 25 % after automatic optimization of only five injection parameters; and 20 % after full optimization (see Figure 4, left).

One optimization takes 15–20 minutes, though the exact duration depends on external factors. In addition, in June 2024, multi-objective optimization has been performed as was planned in the Milestone report MS37.

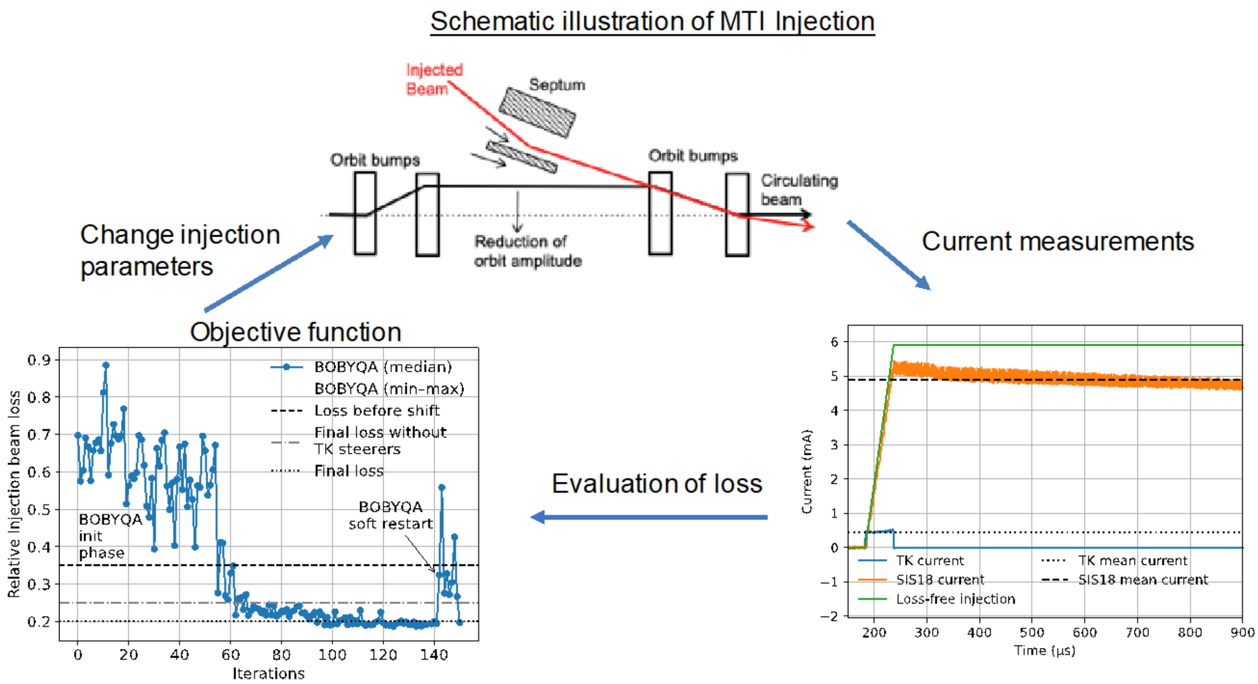

Figure 4: Automatic SIS18 injection optimization with the BOBYQA algorithm carried out in May 2024. The objective function is the beam loss during the injection process and is minimized by varying the injection parameters and TK steerers. On the figure on the left-hand, the gray area marks the initialization phase of the algorithm. Convergences has been reached around the 100^{th} iterations. The figure on the right, shows the relative beam losses that are the difference between the mean SIS18 current (dashed-black) and the ideal current (green), divided by the ideal current. The latter is calculated using the mean TK current (black dotted,) and the expected number of injections.

4.2 FRAGMENT SEPARATOR OPTIMIZATION

The task of a fragment separator is to select and identify the different in-flight nuclides produced in a collision between the primary beam and a fixed target. The nuclides are then used for further studies in nuclear physics or other applications, allowing for instance to identify new rare isotopes and look for new signals that give new insights of our present understanding of these nuclei far from stability.

At GSI, the Fragment Separator (FRS) needs to be set up for experiments at the start of each physics experiment. At present, this process mainly involves manual adjustment by operators and takes 2–3 days. It turns out that certain planned experiments at for e.g. at the Experimental Storage Ring (ESR) are infeasible as the intensity of the fragment beam becomes too low due to optical mismatch between FRS and the downstream ESR.

Within the scope of FAIR, a next-generation fragment separator named Super-FRS is being built. Being at the heart of the NUSTAR collaboration, the Super-FRS is going to be the world leading tool to produce and study rare exotic isotopes and to get insights into the structure of matter and reactions of astrophysical interest. With four times more magnets than the FRS, the complexity is expected to massively and thus increase the required setup time. Both FRS and Super-FRS would benefit from an automated setup, which would increase time available for experiments and make some more experiments feasible.

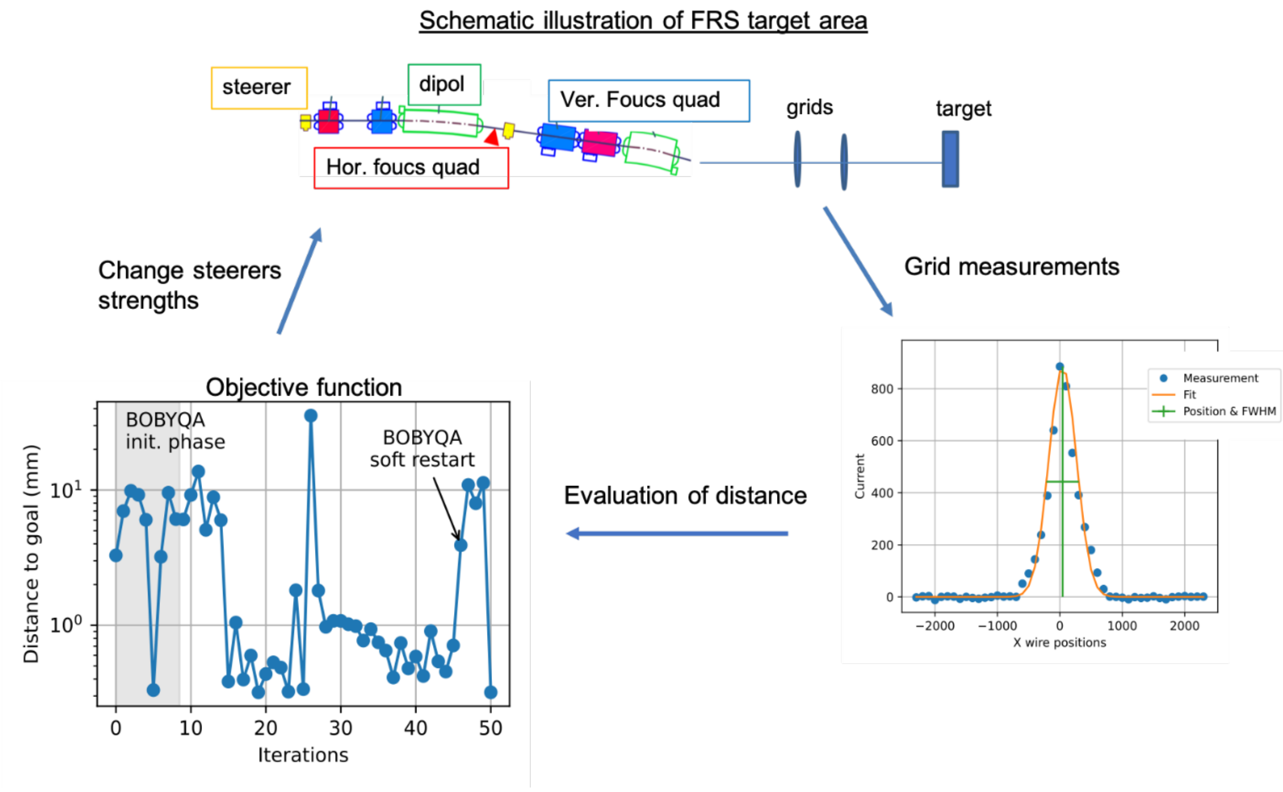

Figure 5: Automatic online steering that uses the BOBYQA algorithm to align the beam at the production target, measured as distance to the goal. The gray area in the left figure marks the initialization phase of the algorithm. Convergence has been reached after ~20 iterations. In the right figure, a Gaussian curve has been fitted to the measurements to estimate beam position and width at the location of the grids.

A first step toward full automation is to make use of numerical optimization algorithms via Geoff. An initial use case that is easily studied is the problem of beam steering with four kicker magnets such that the beam arrives at a specific point and angle on the production target (see Figure 5, top). The beam is measured by two profile grids in front of the target. The profile readings are analyzed by fitting a Gaussian curve to them and estimating its center and width (see Figure 5, right). While the Geoff plugin for the optimization problem was written by us, the analysis code was contributed by the user collaborator. Including it was trivial thanks to modular nature of Geoff. Figure 5 (left) shows the evolution of one optimization run of 50 iterations. Though this one had already converged after 20 iterations, it is not representative as in spite of the randomized initial state, one of the kicker magnet settings happened to be very close to the optimum already. Close towards the end, the algorithm performed a *soft restart*, a mechanism included to prevent possible local minima. All in all, the procedure took about 15 minutes. The algorithm's speed is limited because measurements are bound to the slow-extraction cycle and because the profile grids require multiple cycles to read out both the X and Y coordinates due to the limited bandwidth of the multiplexer that they are attached to.

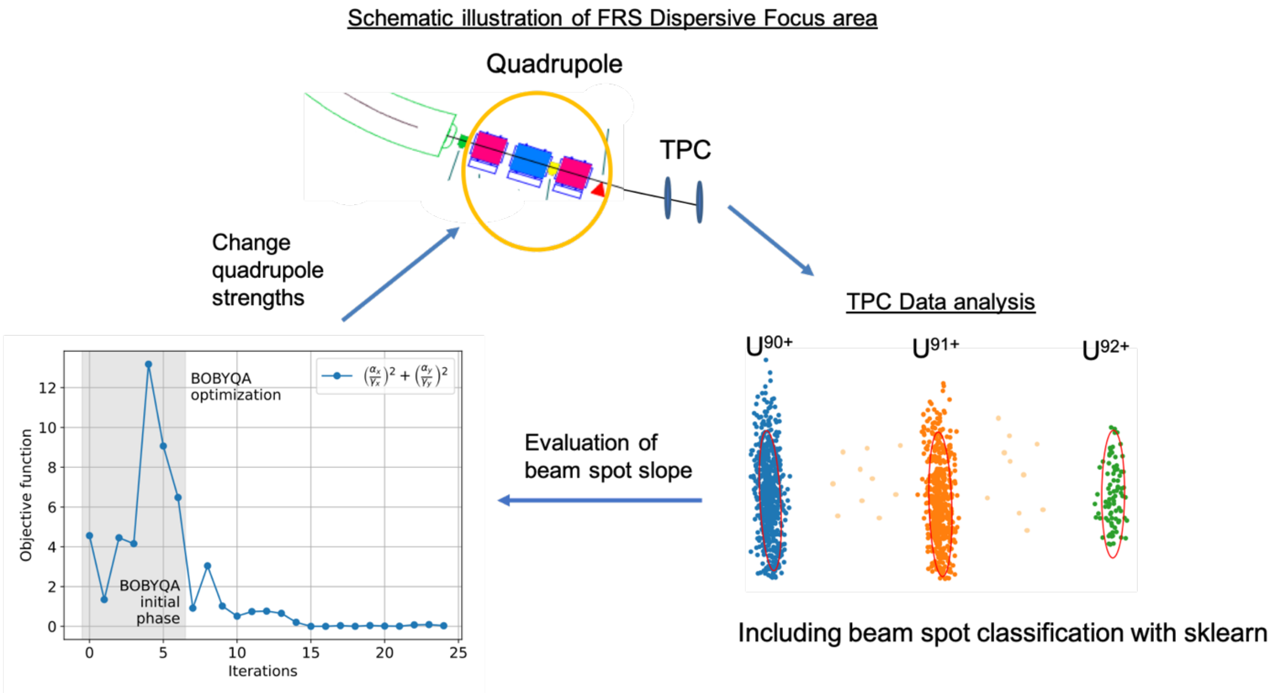

Figure 6: Automatic online focusing that uses the BOBYQA algorithm. The gray area in the left figure marks the initialization phase of the algorithm. Convergences has been reached after about 15 iterations. The objective function (left figure, legend) is a scalar measure of how vertical the U91+ beam spot is oriented (right figure). The BOBYQA algorithm optimizes the strength of the quadrupole magnets (top figure).

The next step up in complexity was to automate the setup of the second focal plane (named also dispersive focus area) of the FRS following the production target, as illustrated in Figure 6. In our experiment, the target was a striping foil that transform the U^{73+} uranium beam into multiple charge states, resulting in three distinct beam spots along with some other charge states. Focusing was performed with the main charge state, U^{91+}.

The goal was to align the U^{91+} beam spot upright by adjusting the quadrupole settings. The beam distribution is measured with a time projection chamber (TPC), which reconstructs the trajectories of all individual particles and is managed by the researcher on a network distinct from the accelerator control systems. Particle trajectories were grouped into beam spots using a classification algorithm from the Python tool *scikit-learn* and has been verified by the user team with simulations. This was easy to do because Geoff plugins can arbitrarily depend on additional packages as mentioned earlier.

In the following analysis step, the Twiss parameters for each beam spot were determined. The Twiss parameters describe the distribution of positions and velocities of the particles in a beam. These parameters were then combined into a single number that represents how vertical the beam spot is. The results are shown in Figure 6 (left). The algorithm was run for 25 iterations, though the optimum was already found after 15.

5. APPLICATION OF THE TOOLKIT AT CEA

5.1 LASER DRIVEN PARTICLE ACCELERATOR LPA-UHI100

Figure 5.1: Photo of the LPA-UHI100 (Laser Plasma Accelerator on UHI100 laser facility), a platform providing an electron beam and an experimental area dedicated to laser-driven electron acceleration studies in plasma media, and applications such as FLASH Radiotherapy, secondary particles generation, diagnostic developments…etc.

The Geoff toolkit will also be used on another type of particle accelerator, the laser-driven particle accelerator facility LPA-UHI100. The facility comprises a 100 TW commercial laser system that generates a pulsed laser beam with an energy of 1.2 J on the target, wavelength of 800 nm, and a pulse width of 25 fs. Focused on a gaseous target, a plasma wave is generated that can accelerate electrons from a few MeV up to 150 MeV at an initial repetition rate of 0.05 Hz. Due to enhanced service achievements from the EURO-LABS project (WP3), the pumping speed in the electron source generation chamber has been increased, allowing now to operate the electron source at 0.2 Hz.

CEA is still waiting for the final authorization to operate the electron source at the nominal performance, as up to now only a temporary authorization for degraded mode has been given (50 MeV maximum energy, 1 shot/min and maximum 15 pC/bunch). It is expected to have a final decision in September 2024. Until then, an automatized acquisition system has been developed as a database for ML Toolkit. The saving process and interfaces with the instruments have been written mostly in Python. This ensures compatibility between the experimental data and Geoff.

For implementation and modification of Geoff in terms of laser and gas target properties, a post-doctoral researcher will be hired. Despite having started the process last year this is not yet complete One candidate has been selected last march 2024 but after administrative processing the clearance to enter the CEA was not given. Hence the selection process has been restarted. Among the candidates interviewed at the end of May, a young scientist has been selected, who has defended his PhD in laser-plasma physics at the end of June. The administrative process for recruitment is in progress. The present plan is to start the adaptation of Geoff to a laser-driven particle accelerator by October 2024. A majority of the work will be carried out by the post-doctoral researcher in collaboration with GSI and CERN.

6. SUMMARY

The objective of Task 5.3 of using ML techniques to optimised beam optimization and control in two facilities has been realised and the planning for its implementation in a third facility of totally different characteristics is planned in the coming months. According with the intermediate results obtained so far, the Geoff ML tool is a great success. In addition to CERN, Geoff has been used successfully at the GSI user facility FRS for beam steering and focusing and also for injection into SI18. The Geoff framework improved the both adjustment time and beam losses substantially compared to a manually tuning by experienced operating staff

7. ANNEX: GLOSSARY

Acronym	**Definition**
API	Application Programming Interface
BOBYQA	Bound Optimization BY Quadratic Approximation
CEA	French Alternative Energies and Atomic Energy Commission, or CEA (Commissariat à l'énergie atomique et aux énergies alternatives)
CERN	European Organization for Nuclear Research
GSI	German facility „GSI Helmholtzzentrum für Schwerionenforschung GmbH"
Gym	Old API standard for reinforcement learning, Gym has been replaced to Gymnasium
Gymnasium	New API standard for reinforcement learning. Gymnasium is a maintained fork of OpenAI's Gym library
LEIR	Low Energy Ion Ring
NUSTAR	Nuclear Structure, Astrophysics and Reactions
PSB	Proton Synchrotron Booster (CERN)
Python binding	Python functions/variables/etc that are calling code in another language
SAC	Soft Actor-Critic
SIS100	German Schwer-Ionen-Synchrotron 100 ("Heavy Ion Synchrotron", 100 indicates max. rigidity, GSI)
SIS18	German Schwer-Ionen-Synchrotron 18 ("Heavy Ion Synchrotron", 18 indicates max. rigidity, GSI)
Super-FRS	Super-Fragment Separator (FAIR)