2nd collaboration workshop on Reinforcement Learning for Autonomous Accelerators (RL4AA'24)

Monday, February 5, 2024 - Wednesday, February 7, 2024

Universität Salzburg (Paris-Lodron-Universität)

Book of Abstracts

Contents

Applying Reinforcement Learning to IFMIF-DONES HVAC optimisation 1

Quantum annealing for sample-efficient reinforcement learning 1

Reinforcement Learning for CERN's accelerators 2

Practical Microsecond Real-Time Reinforcement Learning 2

Optimising non-linear Injection using Reinforcement Learning 3

Enhancing Autonomy of Unmanned Surface Vehicles through Integrated Perception and Control 4

Explainability in Reinforcement Learning: An Application for Powertrain Control 4

Utilizing Machine Learning-optimized Piecewise Polynomial Models in Mechatronics . . 5

Exploring the Dynamics of Reinforcement Learning in Aerospace Control 6

Optimising Quantum Gate Fidelity with Deep Reinforcement Learning 6

Data-Driven Model Predictive Control for Automated Optimization of Injection into the SIS18 Synchrotron 7

Samlpe Alignment in Neutron Beamlines Using Reinforcement Learning 8

Reinforcement Learning for FLASH Dose Delivery Optimization 8

The RL4AA'23 kick-off workshop at KIT 9

Real-time control with reinforcement learning at KARA 9

Applying Reinforcement Learning to IFMIF-DONES HVAC optimisation 10

Optimization at the GSI/FAIR accelerator facility 11

Designing and Running Real-World RL Experiments 11

Safe and fast Model-based Reinforcement Learning based on Gaussian Processes demonstrated on CERN AWAKE 12

Rapid Adaptation in Accelerator Control: Leveraging Meta-Reinforcement Learning for Handling Dynamic System Variations 12

Beam Trajectory Control with Lattice-Agnostic Reinforcement Learning 13

RL Crash Course and Question Round 14

The Geometry of Reinforcement Learning: Insights from the Dual Linear Program 14

Optimization and Control at DESY - RL and Beyond 14

ML applications at BESSY 15

Considerations on Reinforcement Learning feasibility for the automatic setup of controlled longitudinal emittance blow-up in the CERN SPS 15

Comparing Q-Learning of Single-Agent and Multi-Agent Reinforcement Learning 16

RL Activities Overview at SLAC (preliminary title) 17

Adaptive X-ray Fluorescence Imaging 17

Robustly Safe Bayesian Optimization 17

Zero-shot Optimiser Learning for Particle Accelerators under Partial Observability . . . 18

Learning to Do or Learning While Doing: Reinforcement Learning and Bayesian Optimisation for Online Continuous Tuning 18

Cheetah: Bridging the Gap Between Machine Learning and Particle Accelerator Physics with High-Speed, Differentiable Simulations 19

Towards real-world RL 19

Posters / 37

Applying Reinforcement Learning to IFMIF-DONES HVAC optimisation

Authors: Antonio Manjavacas[1]; Manuel A. Vázquez-Barroso[1]; Juan Gómez-Romero[1]; Francisco Martín-Fuertes[2]

[1] *University of Granada*

[2] *CIEMAT*

Corresponding Authors: jgomez@decsai.ugr.es, manvazbar@ugr.es, francisco.martin-fuertes@ciemat.es, manjavacas@ugr.es

As a critical radiological facility, the International Fusion Materials Irradiation Facility - DEMO Oriented Neutron Source (IFMIF-DONES) will implement effective measures to ensure the safety of its personnel and the environment. To enable the proper implementation of these measures, the ISO 17873 standard has been adopted throughout the design process of the facility. The proposed dynamic confinement measures outlined in this standard require a thorough design of the nuclear Heating, Ventilation and Air Conditioning (HVAC) system to ensure effective containment barriers, stable pressure levels and proper treatment of effluents. However, the design and control of such a critical system presents several challenges, as numerous factors influence pressure stability within the facility.

Despite these challenges, recent advances in Deep Reinforcement Learning (DRL) algorithms have demonstrated their effectiveness in solving complex continuous control problems in a variety of domains. In this work, we evaluate the performance of DRL algorithms in controlling the nuclear HVAC system of IFMIF-DONES. For this purpose, we use a MELCOR simulation model of the particle accelerator facility as a training environment and adapt the functionalities of this simulator to enable the continuous control of the air inlet flow rates.

Possible contributed talk:

Yes

Are you a student?:

Yes

Contributed Talks / 38

Quantum annealing for sample-efficient reinforcement learning

Author: Michael Schenk[1]

Co-authors: Elías Combarro [2]; Michele Grossi [1]; Verena Kain [1]; Kevin Li [1]; Sofia Vallecorsa [1]

[1] *CERN*

[2] *University of Oviedo*

Corresponding Authors: kevin.shing.bruce.li@cern.ch, combarro@gmail.com, michael.schenk@cern.ch, verena.kain@cern.ch, michele.grossi@cern.ch, sofia.vallecorsa@cern.ch

Free energy-based reinforcement learning (FERL) using clamped quantum Boltzmann machines (QBM) has demonstrated remarkable improvements in learning efficiency, surpassing classical Q-learning algorithms by orders of magnitude. This work extends the FERL approach to multi-dimensional optimisation problems and eliminates the restriction to discrete action-space environments, opening doors for a broader range of real-world applications. We will discuss the results obtained with

quantum annealing, employing both a simulator and D-Wave quantum annealing hardware, as well as a comparison to classical RL methods. We will cover how the algorithms are evaluated for control problems at CERN, such as the AWAKE electron beam line, and for classical RL benchmarks of varying degree of complexity.

Possible contributed talk:

Yes

Are you a student?:

No

Invited Talks / 39

Reinforcement Learning for CERN's accelerators

Author: Verena Kain[1]

Co-authors: Michael Schenk [1]; Francesco Velotti [1]

[1] *CERN*

Corresponding Authors: michael.schenk@cern.ch, verena.kain@cern.ch, francesco.maria.velotti@cern.ch

CERN has a long tradition of model-based feedforward control with a high-level of abstraction. With the recently approved project "Efficient Particle Accelerators", the CERN management commits to go one step further and invest heavily into automation on all fronts. The initiative will therefore also further push data-driven surrogate models, sample-efficient optimisation and continous control algorithms into the current control system. Reinforcement Learning has been part of the CERN algorithm suite before many numerical optimisation algorithms. The many decades old CERN machines do however not easily provide for RL to be used - black-box optimisation algorithms are more easily integrated. This contribution will summarise RL controllers in the making for CERN and will mainly focus on CERN's RL vision - offline RL, the importance of being able to deal with partially observable systems, and the necessity for continuously learning controllers.

Possible contributed talk:

No

Are you a student?:

No

Contributed Talks / 40

Practical Microsecond Real-Time Reinforcement Learning

Author: Luca Scomparin[None]

Co-authors: Andrea Santamaria Garcia [1]; Michele Caselle [1]; Chenran Xu [2]

[1] *KIT*

[2] *IBPT*

Corresponding Authors: andrea.santamaria@kit.edu, michele.caselle@kit.edu, luca.scomparin@kit.edu, chenran.xu@kit.edu

Reinforcement Learning (RL) has been successfully applied to a wide range of problems. When the environment to control does not exhibit stringent real-time constraints, currently available techniques and computational infrastructures are sufficient. At particle accelerators, however, it is often possible to encounter stringent requirements on the time necessary for an action to be chosen, that in some extreme cases can fall in the microsecond scale.

These challenging conditions also present some benefits. For instance, the data throughput of the real-world environment can be orders of magnitude greater compared to a simulation. This opens the possibility of online training without the issues linked to transferring a simulation-trained agent to the real world.

In this contribution, real-time constraints and how they affect RL algorithms will be introduced, followed by a description of FPGAs and heterogeneous hardware platforms. This is then used to motivate the architecture of the state-of-the-art KINGFISHER RL system. Finally, an in-depth discussion of the use-cases where this approach can be beneficial will be provided, together with basic guidelines for structuring RL problems in a more hardware-friendly way.

Possible contributed talk:

Yes

Are you a student?:

Yes

Student Session / 41

Optimising non-linear Injection using Reinforcement Learning

Author: Alexander Schütt[1]

Co-authors: Pierre Schnizer [2]; Christoph Knochenhauer [3]

[1] *Helmholtz-Zentrum Berlin*

[2] *pierre.schnizer@helmholtz-berlin.de*

[3] *TU Munich*

Corresponding Authors: alexander.schuett@helmholtz-berlin.de, pierre.schnizer@helmholtz-berlin.de

Synchrotron light source storage rings aim to maintain a continuous beam current without observable beam motion during injection. One element that paves the way to this target is the non-linear kicker (NLK). The field distribution it generates poses challenges for optimising the topping-up operation.

Within this study, a reinforcement learning agent was developed and trained to optimise the NLK operation parameters. We present the models employed, the optimisation process, and the achieved results.

Possible contributed talk:

Yes

Are you a student?:

Yes

Student Session / 42

Enhancing Autonomy of Unmanned Surface Vehicles through Integrated Perception and Control

Authors: Juan Montoya Bayardo[None]; Simon Hirlaender[1]; Christian Borgelt[2]

[1] *PLUS University Salzburg*

[2] *Uni Salzburg*

Corresponding Author: juan.montoya-bayardo@stud.sbg.ac.at

The Sonobot Unmanned Surface Vehicle (USV), developed by EvoLogics, is a system platform tailored for hydrographic surveying in inland waters. Despite its integrated GPS and autopilot system for autonomous mission execution, the Sonobot lacks a collision avoidance system, necessitating constant operator monitoring and significantly limiting its autonomy.

Recognizing the untapped potential of USVs for integrating advancements in autonomous vehicles, machine learning, and control theory, we propose a two-layered system: a perception layer for object detection and an algorithmic layer for collision-free path selection. The novelty of our perception layer lies in the integration of a Stereo Camera, LiDAR, and Front Looking sonar for robust obstacle detection.

For the algorithmic layer, we engineered a simple yet powerful cost function. Our preliminary results demonstrate the ability to calculate a collision-free trajectory for the Sonobot using this cost function in conjunction with a Model Predictive Controller (MPC).

We invite discussion on the potential of testing the MPC against Reinforcement Learning and the possibility of combining MPC and RL to further enhance the autonomy and efficiency of USVs.

Possible contributed talk:

Yes

Are you a student?:

Yes

Contributed Talks / 43

Explainability in Reinforcement Learning: An Application for Powertrain Control

Author: Catherine Laflamme[1]

Co-authors: Jörg Doppler [2]; Sven Domnika [2]; Leonhard Czarnetzki [1]; Bence Palvolgyi [3]; Zsolt Viharos [3]

[1] *Fraunhofer Austria Research GmbH*

[2] *Bosch Engineering GmbH*

[3] *SZTAKI: Institute for Computer Science and Control*

Reinforcement learning (RL), a subgroup of machine learning, has gained recognition for its astonishing success in complex games, however it has yet to show similar success in more real-world scenarios. In principle, the ability for RL to generalise past experience, act in real time, and its resilience to new states makes it particularly attractive as a robust decision-making support for real-world scenarios. However, such scenarios bring unique challenges that aren't present in the game-like domains, such as complex and contradictory reward functions and a necessity for explainability. In this presentation we will discuss some of these challenges in the context of using RL for automotive

powertrain control. We will discuss the problem setup, including reward definition, as well as one approach to explainability. This approach is to first learn a neural network based policy (which can learn effectively and efficiently) and then extrace a rule-based policy (which is easier to interpret and can be directly implemented in current control software). The results are benchmarked with an optimised MATLAB policy, using a simulink simulation.

Possible contributed talk:

Yes

Are you a student?:

No

Student Session / 44

Utilizing Machine Learning-optimized Piecewise Polynomial Models in Mechatronics

Authors: Hannes Waclawek[1]; Stefan Huber[2]

[1] *osef Ressel Centre for Intelligent and Secure Industrial Automation*

[2] *Josef Ressel Centre for Intelligent and Secure Industrial Automation*

Corresponding Authors: hannes.waclawek@fh-salzburg.ac.at, stefan.huber@fh-salzburg.ac.at

The success and fast pace of Machine Learning (ML) in the past decade was also enabled by modern gradient descent optimizers embedded into ML frameworks such as TensorFlow. In the context of a doctoral research project, we investigate how these optimizers can be utilized directly, outside of the scope of neural networks. This approach holds the potential of optimizing explainable models with only few model parameters allowing to derive properties for direct, physical explanation and interpretation, like velocity, acceleration or jerk. This is highly beneficial for use in the field of mechatronics. However, while modern gradient gradient descent optimizers shipped with ML frameworks perform well in neural nets, results show that most optimizers have limited capabilities when applied directly to PP models. Domain-specific model requirements like C^k-continuity, acceleration or jerk limitation as well as spectral or energy optimization pose the need for developing appropriate loss functions, novel algorithms as well as regularization techniques in order to improve optimizer performance.

In this context, we investigate piecewise polynomial models as they occur (and are required) in 1D trajectory planning tasks in mechatronics. Utilizing TensorFlow optimizers, we optimize our PP model towards multi-targeted loss functions suitable for fitting of C^k-continuos PP functions which can be deployed in an electronic cam approximation setting. We enhance capabilities of our PP base model by utilizing an orthogonal Chebyshev basis along with a novel regularization method improving convergence of the approximation and continuity optimization targets. We see a possible application of this approach in Deep Reinforcement Learning applied to Control Theory. By exchanging the black box that is a neural network with an explainable PP model, we foster utility of Reinforcement Learning in designing cyber-physical control systems.

Possible contributed talk:

Yes

Are you a student?:

Yes

Student Session / 45

Exploring the Dynamics of Reinforcement Learning in Aerospace Control

Authors: Georg Schäfer[1]; Jakob Rehrl[1]; Stefan Huber[1]; Simon Hirlaender[2]

[1] *Josef Ressel Centre for Intelligent and Secure Industrial Automation*

[2] *PLUS University Salzburg*

Corresponding Authors: georg.schaefer@fh-salzburg.ac.at, jakob.rehrl@fh-salzburg.ac.at, stefan.huber@fh-salzburg.ac.at

The Quanser Aero2 system is an advanced laboratory experiment designed for exploring aerospace control systems, featuring two motor-driven fans on a pivot beam for precise control. Its capability to lock axes individually offers both single degree of freedom (DOF) and two DOF operation. The system's non-linear characteristics and adaptability to multivariable configurations make it especially interesting for control theory research.

In this study, we use Reinforcement Learning (RL) to control the Aero2 system. To keep complexity low in a first step, this work focuses on the single DOF setup. A RL agent is trained on a simulation to develop a policy for orienting the beam to a specific tilt, using the target's tilt deviation and velocity as the state space. To further reduce complexity, the second motor uses reversed polarity voltage from the first motor, resulting in a single action, enabling an in-depth analysis of the learning behaviour of the employed agents.

Even if we reduce the action space to one dimension by exploiting the symmetry of the two rotors, the given balancing task could not be solved with the default configuration of the used Proximal Policy Optimization (PPO) agent. We identified that a reduction of the number of units in each hidden fully connected layer of the agent networks is necessary to solve the task. However, detailed visualisations of the development of the policy over time revealed a transition from stable to volatile action choices in the long term, which is unexpected according to the current state of the literature. Future research will focus on the underlying causes of the observed volatility, giving insights into the dynamic nature of RL.

Possible contributed talk:

No

Are you a student?:

Yes

Posters / 46

Optimising Quantum Gate Fidelity with Deep Reinforcement Learning

Authors: Leander Grech[1]; Gianluca Valentino[1]; Tony John George Apollaro[1]; Mirko Consiglio[1]

[1] *University of Malta*

Corresponding Author: leander.grech@um.edu.mt

Noisy intermediate-scale quantum (NISQ) computers work by applying a set of quantum gates to an initial ground state, to transform it into a final state that represents the solution to complex computational problems, such as molecular energy evaluation or optimising for the shortest routes in the travelling salesman problem. The effective realisation of NISQ computers requires the implementation of multi-qubit gates with high fidelity, and low leakage of information to surrounding qubits and higher energy levels. The adherence to these constraints is essential for achieving universal fault-tolerant quantum computing.
Quantum optimal control theory focuses on designing electromagnetic pulse shapes, which are transmitted to the qubits in order to precisely guide their evolution from an initial quantum state to a final state. While several quantum gate synthesis algorithms such as GRadient Ascent Pulse Engineering (GRAPE), Chopped RAndom Basis (CRAB) and Gradient Optimization of Analytic conTrols (GOAT) have been developed, they face scalability challenges as the number of qubits increases.
Recently, the quantum computing community has begun to explore Reinforcement Learning (RL) techniques as an alternative to quantum optimal control algorithms, however, none of these methods tune the pulse parameters directly. RL-based approaches involve an agent interacting with a simulated quantum environment by taking actions that modify the pulse parameters in either the time or frequency domain to achieve the desired quantum state transformation. The effectiveness of these transformations is mainly measured by the fidelity between the achieved quantum state and the desired state. This fidelity metric is integral to the feedback loop, serving as part of the reward mechanism for the RL agent. Other metrics, such as leakage to higher energy states of a qubits must also be taken into consideration for the development of efficient and fast gates.

Possible contributed talk:

Yes

Are you a student?:

No

Posters / **47**

Data-Driven Model Predictive Control for Automated Optimization of Injection into the SIS18 Synchrotron

Author: Sabrina Appel[1]

Co-authors: Simon Hirlaender [2]; Nico Madysa [1]

[1] *GSI*

[2] *PLUS University Salzburg*

Corresponding Authors: simon.hirlaender@plus.ac.at, n.madysa@gsi.de, s.appel@gsi.de

In accelerator labs like GSI/FAIR, automating complex systems is key for maximising physics experiment time. This study explores the application of a data-driven model predictive control (MPC) to refine the multi-turn injection (MTI) process into the SIS18 synchrotron, departing from conventional numerical optimisation methods. MPC is distinguished by its reduced number of optimisation steps and superior ability to control performance criteria, effectively addressing issues like delayed outcomes and safety concerns, including septum protection.

The study focuses on a highly sample-efficient MPC approach based on Gaussian processes, which lies at the intersection of model-based reinforcement learning and control theory. This approach merges the strengths of both fields, offering a unified and optimised solution and yielding a safe and fast state-based optimisation approach beyond classical reinforcement learning and Bayesian optimisation.

Our study lays the groundwork for enabling safe online training for the SS18 MTI issue, showing great potential for the application of data-driven control in similar scenarios.

Possible contributed talk:

Yes

Are you a student?:

No

Posters / 48

Samlpe Alignment in Neutron Beamlines Using Reinforcement Learning

Author: Morgan Henderson[1]

Co-authors: Stuart Calder [2]; Matthew Kilpatrick [1]; Jonathan Edelen [1]

[1] *RadiaSoft LLC*

[2] *ORNL*

Corresponding Authors: caldersa@ornl.gov, jedelen@radiasoft.net, kilpatrick@radiasoft.net, mhenderson@radiasoft.net

RadiaSoft has been developing machine learning (ML) methods for automating processes within the accelerator landscape for the past five years. One critical area of this work has been the full automation of sample alignment at neutron and x-ray beamlines to ensure both high quality experimental data and efficient use of operator hours. Historically, sample alignment has been a manual or a semi-automated process requiring significant levels of human intervention (particularly for time-intensive processes such as temperature scans). Due to the need for both visual and detector-based alignment of samples and the execution of corresponding beamline controls, ML methods, and reinforcement learning (RL) in particular, are well-suited for this application. Here we provide an overview of both the visual and detector-based aspects of the sample alignment problem and describe our plans and early results for applying RL to the controls portion of sample alignment for neutron beams. We will also discuss how our current work will be extended to x-ray beamlines.

Possible contributed talk:

Yes

Are you a student?:

Contributed Talks / 49

Reinforcement Learning for FLASH Dose Delivery Optimization

Authors: Jonathan Edelen[1]; Morgan Henderson[1]

Co-authors: Auralee Edelen [2]; Matthew Kilpatrick [1]; Jorge Diaz Cruz [2]

[1] *RadiaSoft LLC*

[2] *SLAC*

Corresponding Authors: dejorge@slac.stanford.edu, kilpatrick@radiasoft.net, jedelen@radiasoft.net, mhenderson@radiasoft.net, edelen@slac.stanford.edu

RadiaSoft is developing machine learning methods to improve the operation and control of industrial accelerators. Because industrial systems typically suffer from a lack of instrumentation and a noisier environment, advancements in control methods are critical for optimizing their performance. In

particular, our recent work has focused on the development of pulse-to-pulse feedback algorithms for use in dose optimization for FLASH radiotherapy. The PHASER (pluridirectional high-energy agile scanning electronic radiotherapy) system is of particular interest due to the need to synchronize 16 different accelerators all with their own noise characteristics. This presentation will provide an overview of the challenges associated with dose optimization for a PHASER-like system, a description of the toy model used to evaluate different control schema, and our initial results using RL for dose delivery optimization.

Possible contributed talk:

Yes

Are you a student?:

Posters / **50**

The RL4AA'23 kick-off workshop at KIT

Authors: Andrea Santamaria Garcia[1]; Simon Hirlaender[2]; Chenran Xu[3]; Jan Kaiser[4]

[1] *KIT*

[2] *PLUS University Salzburg*

[3] *IBPT*

[4] *DESY*

Corresponding Authors: andrea.santamaria@kit.edu, chenran.xu@kit.edu, simon.hirlaender@plus.ac.at, jan.kaiser@desy.de

Reinforcement Learning (RL) is a unique learning paradigm that is particularly well-suited to tackle complex control tasks, can deal with delayed consequences, and learns from experience without an explicit model of the dynamics of the problem. These properties make RL methods extremely promising for applications in particle accelerators, where the dynamically evolving conditions of both the particle beam and the accelerator systems must be constantly considered.
While the time to work on RL is now particularly favourable thanks to the availability of high-level programming libraries and resources, its implementation in particle accelerators is not trivial and requires further consideration.
In this context, the Reinforcement Learning for Autonomous Accelerators (RL4AA) international collaboration was established to consolidate existing knowledge, share experiences and ideas, and collaborate on accelerator-specific solutions that leverage recent advances in RL.
The collaboration was launched in February 2023 during the RL4AA'23 workshop at the Karlsruhe Institute of Technology. This workshop included introductory lectures to RL, a hands-on tutorial on how to apply RL to an accelerator problem, an overview talk on RL in the field of accelerators, and advanced discussions in dedicated working groups on modern RL challenges.

Possible contributed talk:

No

Are you a student?:

No

Invited Talks / **51**

Real-time control with reinforcement learning at KARA

Authors: Andrea Santamaria Garcia[1]; Luca Scomparin[None]; Michele Caselle[1]; Erik Bründermann[2]; Andreas Kopmann[1]; Johannes Steinmann[3]; Chenran Xu[4]; Anke-Susanne Müller[1]

[1] *KIT*

[2] *Karlsruhe Institute of Technology (KIT) - IBPT*

[3] *KIT / IBPT*

[4] *IBPT*

Corresponding Authors: anke-susanne.mueller@kit.edu, erik.bruendermann@kit.edu, michele.caselle@kit.edu, chenran.xu@kit.edu, luca.scomparin@kit.edu, andreas.kopmann@kit.edu, andrea.santamaria@kit.edu, johannes.steinmann@kit.edu

Reinforcement Learning (RL) has demonstrated its effectiveness in solving control problems in particle accelerators. A challenging application is the control of the microbunching instability (MBI) in synchrotron light sources. Here the interaction of an electron bunch with its emitted coherent synchrotron radiation leads to complex non-linear dynamics and pronounced fluctuations.

Addressing the control of intricate dynamics necessitates meeting stringent microsecond-level real-time constraints. To achieve this, RL algorithms must be deployed on a high-performance electronics platform. The KINGFISHER system, utilizing the AMD-Xilinx Versal family of heterogeneous computing devices, has been specifically designed at KIT to tackle these demanding conditions. The system implements an experience accumulator architecture to perform online learning purely through interaction with the accelerator while still satisfying strong real-time constraints.

The preliminary results of this innovative control paradigm at the Karlsruhe Research Accelerator (KARA) will be presented. Notably, this represents the first experimental attempt to control the MBI with RL using online training only and running on hardware.

Possible contributed talk:

Yes

Are you a student?:

No

Student Session / 52

Applying Reinforcement Learning to IFMIF-DONES HVAC optimisation

Authors: Antonio Manjavacas Lucas[1]; Manuel A. Vázquez-Barroso[1]; Juan Gómez-Romero[1]; Francisco Martín-Fuertes[2]

[1] *University of Granada*

[2] *CIEMAT*

Corresponding Authors: manjavacas@ugr.es, manvazbar@ugr.es, francisco.martin-fuertes@ciemat.es, jgomez@decsai.ugr.es

As a critical radiological facility, the International Fusion Materials Irradiation Facility - DEMO Oriented Neutron Source (IFMIF-DONES) will implement effective measures to ensure the safety of its personnel and the environment. To enable the proper implementation of these measures, the ISO 17873 standard has been adopted throughout the design process of the facility. The proposed dynamic confinement measures outlined in this standard require a thorough design of the nuclear Heating, Ventilation and Air Conditioning (HVAC) system to ensure effective containment barriers, stable pressure levels and proper treatment of effluents. However, the design and control of such a critical system presents several challenges, as numerous factors influence pressure stability within the facility.

Despite these challenges, recent advances in Deep Reinforcement Learning (DRL) algorithms have demonstrated their effectiveness in solving complex continuous control problems in a variety of domains. In this work, we evaluate the performance of DRL algorithms in controlling the nuclear HVAC system of IFMIF-DONES. For this purpose, we use a MELCOR simulation model of the particle

accelerator facility as a training environment and adapt the functionalities of this simulator to enable the continuous control of the air inlet flow rates.

Possible contributed talk:

Yes

Are you a student?:

Yes

Invited Talks / 53

Optimization at the GSI/FAIR accelerator facility

Authors: Sabrina Appel[1]; Nico Madysa[1]

[1] *GSI*

Corresponding Authors: s.appel@gsi.de, n.madysa@gsi.de

The complexity of the GSI/FAIR accelerator facility demands a high level of automation in order to maximize time for physics experiments. This talk will give an overview of different optimization problems at GSI, from transfer lines to synchrotrons to the fragment separator. Starting with a summary of previous successful automation, the talk will focus on the latest developments in recent months, such as the optimization of multi-turn injection in the SIS18 synchrotron. The introduction of a Python bridge to the settings management system LSA and the integration of GeOFF (Generic Optimization Framework & Frontend) enabled and facilitated beam-based optimization with numerical algorithms and machine learning. Geoff is an open-source framework that harmonizes access to a number automation techniques and simplifies the transition towards and between them.

Possible contributed talk:

Yes

Are you a student?:

No

Keynote / 54

Designing and Running Real-World RL Experiments

Author: Antonin Raffin[None]

Corresponding Author: antonin.raffin@dlr.de

This talk covers the challenges and best practices for designing and running real-world reinforcement learning (RL) experiments.
The idea is to walk through the different steps of RL experimentation (task design, choosing the right algorithm, implementing safety layers) and also provide practical advice on how to run experiments and troubleshoot common problems.

Slides are also online: https://araffin.github.io/slides/design-real-rl-experiments/

Possible contributed talk:

Yes

Are you a student?:

Posters / 55

Safe and fast Model-based Reinforcement Learning based on Gaussian Processes demonstrated on CERN AWAKE

Authors: Simon Hirlaender[1]; Verena Kain[2]; Sabrina Pochaba[None]; Lukas Lamminger[3]; Zevi Della Porta[2]

[1] *PLUS University Salzburg*

[2] *CERN*

[3] *Plus University Salzburg*

Corresponding Authors: verena.kain@cern.ch, sabrina.pochaba@salzburgresearch.at, simon.hirlaender@plus.ac.at

Reinforcement Learning (RL) is emerging as a valuable method for controlling and optimizing particle accelerators, learning through direct experience without a pre-existing model. However, its low sample efficiency limits its application in real-world scenarios. This paper introduces a model-based RL approach using Gaussian processes to address this efficiency challenge. The proposed RL agent successfully controlled the trajectory in CERN's AWAKE facility with limited interactions, outperforming traditional numerical optimizers. Unlike these optimizers, which require exploration for each use, the RL agent quickly learns and can be applied for single or few-shot control, including online stabilization of accelerators. The method is also capable of respecting state constraints and non-stationary environements, which is demonstrated in simulations.
This method represents a significant step forward in applying RL to accelerator control for practical use.

Possible contributed talk:

No

Are you a student?:

No

Posters / 56

Rapid Adaptation in Accelerator Control: Leveraging Meta-Reinforcement Learning for Handling Dynamic System Variations

Authors: Simon Hirlaender[1]; Lukas Lamminger[2]; Verena Kain[3]; Sabrina Pochaba[None]; Andrea Santamaria Garcia[4]; Jan Kaiser[5]; Chenran Xu[6]; Annika Eichler[5]

[1] *PLUS University Salzburg*

[2] *Plus University Salzburg*

[3] *CERN*

[4] *KIT*

[5] *DESY*

[6] *IBPT*

Corresponding Authors: andrea.santamaria@kit.edu, sabrina.pochaba@salzburgresearch.at, verena.kain@cern.ch, chenran.xu@kit.edu, simon.hirlaender@plus.ac.at, jan.kaiser@desy.de

In typical reinforcement learning applications for accelerators, system dynamics often vary, leading to
decreased performance in trained agents. In certain scenarios, this performance degradation is severe,
necessitating retraining. However, employing meta-reinforcement learning in conjunction with an appropriate simulation can enable an agent to rapidly adapt to environmental changes. This concept is illustrated by meta-training an agent within a simulated environment replicating the electron line of
CERN's AWAKE experiment. The task involves guiding the electron towards a specific trajectory.
During the simulation, the quadrupoles of the segment are varied randomly, and action masking is employed to mimic magnetic control faults. Our findings reveal that the agent can quickly adjust to specific system configurations with minimal steps. This methodology holds potential for application in any Partially Observable Markov Decision Process (POMDP) characterised by slowly evolving hidden parameters.

Possible contributed talk:

No

Are you a student?:

No

Posters / **57**

Beam Trajectory Control with Lattice-Agnostic Reinforcement Learning

Author: Chenran Xu[1]

Co-authors: Erik Bründermann [2]; Anke-Susanne Müller [3]; Andrea Santamaria Garcia [3]; Jan Kaiser [4]; Annika Eichler [4]

[1] *IBPT*

[2] *Karlsruhe Institute of Technology (KIT) - IBPT*

[3] *KIT*

[4] *DESY*

Corresponding Authors: jan.kaiser@desy.de, andrea.santamaria@kit.edu, anke-susanne.mueller@kit.edu, chenran.xu@kit.edu, erik.bruendermann@kit.edu

In recent work, it has been shown that reinforcement learning (RL) is capable of outperforming existing methods on accelerator tuning tasks. However, RL algorithms are difficult and time-consuming to train, and currently need to be retrained for every single task. This makes fast deployment in operation difficult and hinders collaborative efforts in this research area. At the same time, modern accelerators often reuse certain structures, such as transport lines consisting of several magnets, within or across facilities, leading to similar tuning tasks.
In this contribution, we use different methods, such as domain randomization, to allow an agent trained in simulation to easily be deployed to a group of similar tasks. Preliminary results show that this training method is transferable and allows the RL agent to control the beam trajectory at similar lattice sections of two different real linear accelerators. We expect that future work in this direction will enable faster deployment of learning-based tuning routines, and lead towards the ultimate goal of autonomous operation of accelerator systems and transfer of RL methods to most accelerators.

Possible contributed talk:

No

Are you a student?:

RL Crash Course / 58

RL Crash Course and Question Round

Authors: Andrea Santamaria Garcia[1]; Sabrina Pochaba[None]

[1] *KIT*

Corresponding Authors: andrea.santamaria@kit.edu, sabrina.pochaba@salzburgresearch.at

Student Session / 59

The Geometry of Reinforcement Learning: Insights from the Dual Linear Program

Authors: Nikola Milosevic[1]; Johannes Müller[2]

Co-authors: Nico Scherf [3]; Semih Cayçı [2]

[1] *Max Planck Institute for Human Cognitive and Brain Sciences*

[2] *RWTH Aachen*

[3] *MPI CBS*

Corresponding Author: nmilosevic@cbs.mpg.de

Reinforcement Learning (RL) has become a cornerstone of machine learning, showcasing remarkable success in addressing real-world control problems and providing insights into cognitive processes in the brain. However, navigating the intricacies of modern RL proves challenging due to its numerous moving parts, escalating agent complexity, and the application of deep learning in a non-i.i.d. setting. The inherent challenge of intuitively reasoning about RL stems, in part, from its time-dependent and recursive nature. During this presentation, we explore the *dual linear program* and the intuitions it can offer. What traditionally serves as a theoretical construct for proving theorems emerges as a valuable tool for developing intuitions and facilitating the exploration of higher-level questions. We will focus on two practical demonstrations that underscore the significance of this perspective: 1) designing policy optimization algorithms and 2) pretraining RL agents. During the first half of this presentation, I will review the dual linear program and its geometry, aiming to uncover novel policy optimization strategies. In the second part, I will provide a preview of how the linear program can be generalized to *convex* MDPs, resulting in pretraining objectives similar to representation learning with the Variational Autoencoder.

Possible contributed talk:

Yes

Are you a student?:

Yes

Invited Talks / 60

Optimization and Control at DESY - RL and Beyond

Author: Annika Eichler[1]

Co-authors: Jan Kaiser [1]; Jannis Lübsen [2]

[1] *DESY*
[2] *TUHH*

Corresponding Authors: jan.kaiser@desy.de, jannis.luebsen@tuhh.de, annika.eichler@desy.de

DESY has many years on experience on optimization and control of particle accelerators. Reinforcement learning has been explored within the last three years. In this talk the results of this investigation are summarized and an outlook is given. Further control and optimization challenges for operation are presented and discussed.

Possible contributed talk:

No

Are you a student?:

No

Invited Talks / 61

ML applications at BESSY

Author: Gregor Hartmann[1]

[1] *HZB*

Corresponding Author: gregor.hartmann@helmholtz-berlin.de

In order to improve BESSY's experimental environment, several ML-based applications are used at HZB. These efforts cover challenges arising at the accelerator, beamlines and detectors at the experiment. This talk provides on overview of these activities focussing on RL and providing insights in the optimization of a beamline, tuning of an e-gun as well as electron beam positioning in BESSY's storage ring. The limitations of RL and the reason to also use user ML-techniques are also discussed and presented by various examples.

Possible contributed talk:

No

Are you a student?:

No

Contributed Talks / 62

Considerations on Reinforcement Learning feasibility for the automatic setup of controlled longitudinal emittance blow-up in the CERN SPS

Author: Niky Bruchon[1]

[1] *CERN*

Corresponding Author: niky.bruchon@cern.ch

Despite the spreading of Reinforcement Learning (RL) applications for optimizing the performance of particle accelerators, this approach is not always the best choice. Indeed, not all problems are suitable to be solved via RL. Before diving into such techniques, a good knowledge of the problem, the available resources, and the existing solutions is recommended. An example of the complexities related to RL solutions is the automatic setup of controlled longitudinal emittance blow-up in the CERN SPS. Several criticalities, such as the data availability and the increasing problem dimensions, limited the development of an operational tool based on RL. Therefore, the released software relies on generic optimizers only, even if promising results with Bayesian optimization were achieved.

Possible contributed talk:

Yes

Are you a student?:

No

Student Session / 63

Comparing Q-Learning of Single-Agent and Multi-Agent Reinforcement Learning

Authors: Sabrina Pochaba[None]; Simon Hirlaender[1]; Roland Kwitt[None]; Peter Dorfinger[None]

[1] *PLUS University Salzburg*

Reinforcement Learning (RL) is a rising subject of Machine Learning (ML). Especially Multi-Agent RL
(MARL), where more than one agent interacts with an environment by learning to solve a task, can model
many real-world problems. Unfortunately, the Multi-Agent case yields more difficulties in the already challenging field of Reinforcement Learning, like scalability issues, non-stationarity or non-unique learning goals.
To better understand these problems, we compare Single-Agent RL with MARL in the simple board game Tic-Tac-Toe. This game is a two-player zero-sum game, meaning that two adversarial players compete
against each other during the game by setting their marks (x or o) on a 3x3 board. If one player has three
of his marks in one line (vertical, horizontal or diagonal), this player wins the game and ends it. If neither
of the players gets three marks in one line until all fields of the 3x3 boards are filled, the game ends with a
draw.
We study the learning of a Single- and a Multi-Agent system playing Tic-Tac-Toe, using a Q-Learning algorithm that describes the learning of the agent in one formula. As typical in RL, the agent interacts with
an environment during learning, which is, in this case, the 3x3 board of the Tic-Tac-Toe game. The two
playing agents set their marks one after another. At the end of each game, every agent gets a reward based on
the game's outcome. During learning, the agent tries to maximize his reward, which leads to a well-playing
strategy, namely the policy.
We show that a Single-Agent RL agent only performs as well as the opponent against whom he is trained,

while the agents in the MARL scenario learn an optimal strategy against every possible opponent. Addition-
ally, the agents in the MARL learn more quickly than the ones in the Single-Agent case.
We will use these results to set up a MARL setting in network communications. In this scenario, all communicating electronic devices are different agents, that should communicate in a reliable way, using as
many resources as possible for a quick communication without disturbing the communication of the other
devices

Possible contributed talk:

Yes

Are you a student?:

Yes

Invited Talks / 66

RL Activities Overview at SLAC (preliminary title)

Corresponding Author: edelen@slac.stanford.edu

Contributed Talks / 68

Adaptive X-ray Fluorescence Imaging

Author: Daniel Ratner[1]

[1] *SLAC*

Corresponding Author: dratner@slac.stanford.edu

Possible contributed talk:

Are you a student?:

Student Session / 69

Robustly Safe Bayesian Optimization

Author: Jannis Lübsen[1]

Co-author: Annika Eichler [2]

[1] *TUHH/DESY*

[2] *DESY*

Corresponding Author: jannis.luebsen@tuhh.de

Safety guarantees for Gaussian processes require the assumption that the true hyperparameters are known. However, this assumption usually does not hold in practice. In this talk, a method is introduced to overcome this issue which estimates confidence intervals of hyperparameters from their posterior distributions. Finally, it can be shown that via appropriate scaling safeness can be robustly guaranteed with high probability.

Possible contributed talk:

No

Are you a student?:

Yes

Posters / 70

Zero-shot Optimiser Learning for Particle Accelerators under Partial Observability

Authors: Jan Kaiser[1]; Oliver Stein[1]; Annika Eichler[1]

[1] *DESY*

Corresponding Authors: jan.kaiser@desy.de, annika.eichler@desy.de

In the quest to harness the full potential of particle accelerators for scientific research, the need for precision and efficiency in their operation is paramount. Traditional tuning methods, while effective, fall short in optimising performance swiftly and accurately, leading to underutilisation of valuable beam time. This study applies deep reinforcement learning to autonomously tune transverse beam parameters of the ARES linear accelerator at DESY, pioneering in particular the use of domain randomisation to achieve the zero-shot transfer of a policy trained on a computationally cheap simulation to the real particle accelerator facility. We demonstrate that our approach significantly enhances tuning speed and operational efficiency, thereby promising to accelerate scientific discoveries by making experimental setups more reproducible and efficient.

Possible contributed talk:

No

Are you a student?:

Posters / 71

Learning to Do or Learning While Doing: Reinforcement Learning and Bayesian Optimisation for Online Continuous Tuning

Authors: Jan Kaiser[1]; Chenran Xu[2]; Annika Eichler[1]; Andrea Santamaria Garcia[3]; Oliver Stein[1]; Erik Bründermann[4]; Willi Kuropka[1]; Hannes Dinter[1]; Frank Mayet[1]; Thomas Vinatier[1]; Florian Burkart[1]; Holger Schlarb[1]

[1] *DESY*

[2] *IBPT*

[3] *KIT*

[4] *Karlsruhe Institute of Technology (KIT) - IBPT*

Corresponding Authors: andrea.santamaria@kit.edu, jan.kaiser@desy.de, annika.eichler@desy.de, erik.bruendermann@kit.edu, chenran.xu@kit.edu

In the pursuit of optimising particle accelerators, the choice of method for autonomous tuning is critical for enhancing performance and operational efficiency. This study delves into comparing deep reinforcement learning-trained optimisers (RLO) and Bayesian optimisation (BO) for this purpose, motivated by the need to address the complex, dynamic nature of accelerators. Through simulation and real-world applications at the ARES accelerator at DESY, the research assesses RLO's adaptability and speed against BO's deployment simplicity. The findings illuminate the trade-offs involved in selecting an optimisation method, offering guidance on balancing performance improvements with practical deployment considerations in particle accelerator tuning.

Possible contributed talk:

No

Are you a student?:

Posters / **72**

Cheetah: Bridging the Gap Between Machine Learning and Particle Accelerator Physics with High-Speed, Differentiable Simulations

Authors: Jan Kaiser[1]; Chenran Xu[2]; Annika Eichler[1]; Andrea Santamaria Garcia[3]

[1] *DESY*

[2] *IBPT*

[3] *KIT*

Corresponding Authors: chenran.xu@kit.edu, jan.kaiser@desy.de, annika.eichler@desy.de, andrea.santamaria@kit.edu

The optimisation and control of particle accelerators present significant challenges due to the limited availability of beam time, high computational costs, and the complexity of the underlying physics. Machine learning has emerged as a powerful tool to address these challenges, but its application is hindered by the scarcity of high-quality data and the computational intensity of traditional simulation methods. Here, we introduce *Cheetah*, a Python-based simulation package that leverages PyTorch for differentiable beam dynamics simulations, offering a solution that is up to seven orders of magnitude faster than existing codes. Cheetah's automatic differentiation capability enables the integration of machine learning techniques directly with particle accelerator physics, facilitating advanced applications such as reinforcement learning-trained optimisers, gradient-based actuator tuning and system identification, physics-based priors for Bayesian optimisation, and modular neural network beam dynamics surrogate modelling. Our results demonstrate the practical utility of Cheetah in real-world settings, including the successful application on five example applications, highlighting its potential to significantly advance the field of accelerator physics through the integration of machine learning methodologies.

Possible contributed talk:

No

Are you a student?:

Keynote / **73**

Towards real-world RL