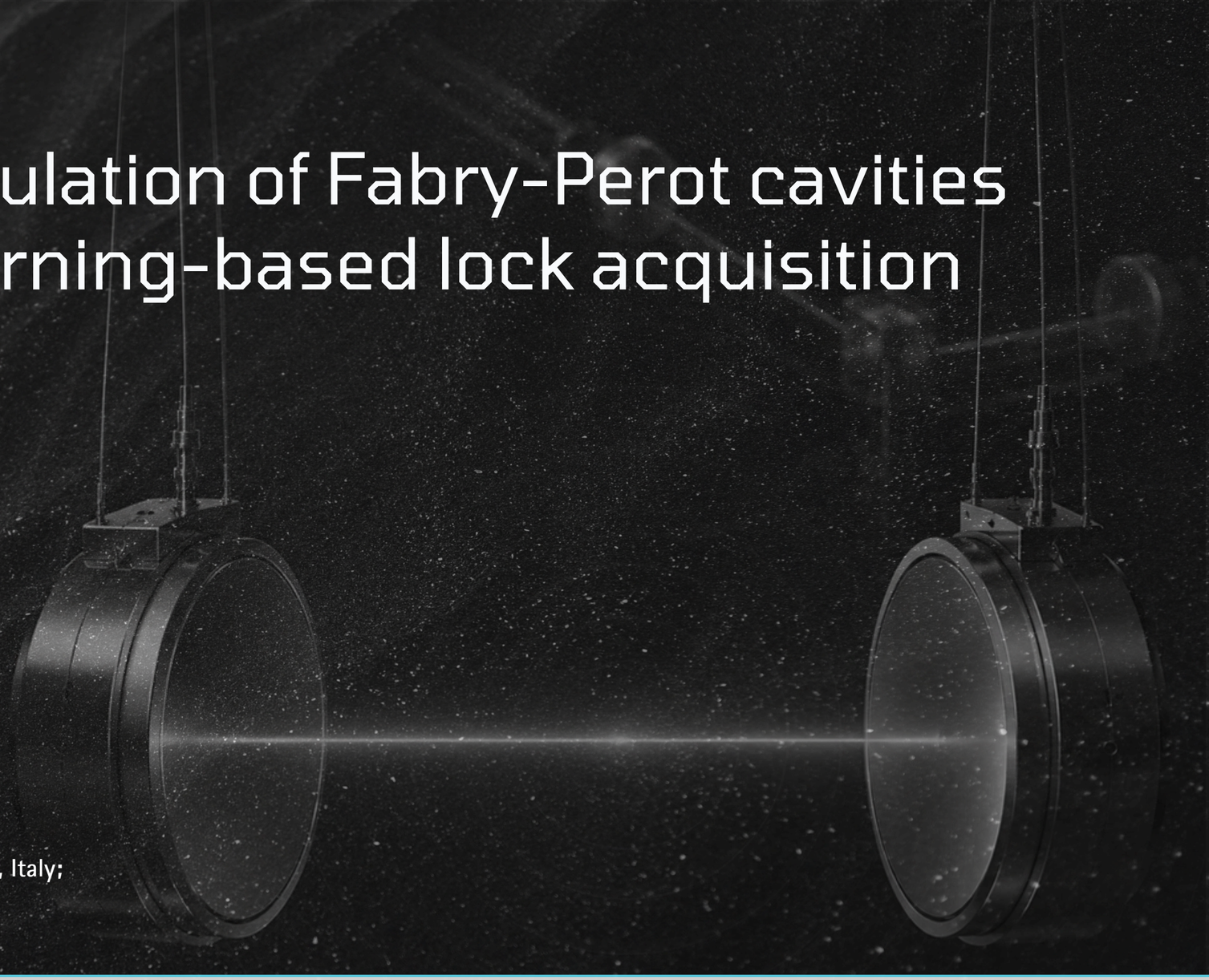
Fast time-domain simulation of Fabry-Perot cavities for Reinforcement Learning-based lock acquisition



The 14th KAGRA International Workshop

16th May 2026

Andrea Svizzeretto[1,2] Mateusz Bawaj[1,2]

[1]Dipartimento di Fisica e Geologia, Università di Perugia, I-06123 Perugia, Italy;
[2]INFN, Sezione di Perugia, I-06123 Perugia, Italy

A.D. 1308 unipg UNIVERSITÀ DEGLI STUDI DI PERUGIA

INFN PERUGIA

VIRGO

EINSTEIN TELESCOPE

CA S

Why time-domain simulation?

When stationary or adiabatic models are not enough

- Fast transients and short optical pulses require explicit propagation in time.
- Moving mirrors couple delayed intra-cavity fields to cavity length variations.
- High-finesse cavities store optical energy: the present depends on the recent past.
- Lock acquisition is a non-stationary control problem due to non linearities of the plant.

Key message

A cavity simulator for control must expose the real transient optical observables.

Target use case

Step-by-step interaction between plant and reinforcement-learning agents.

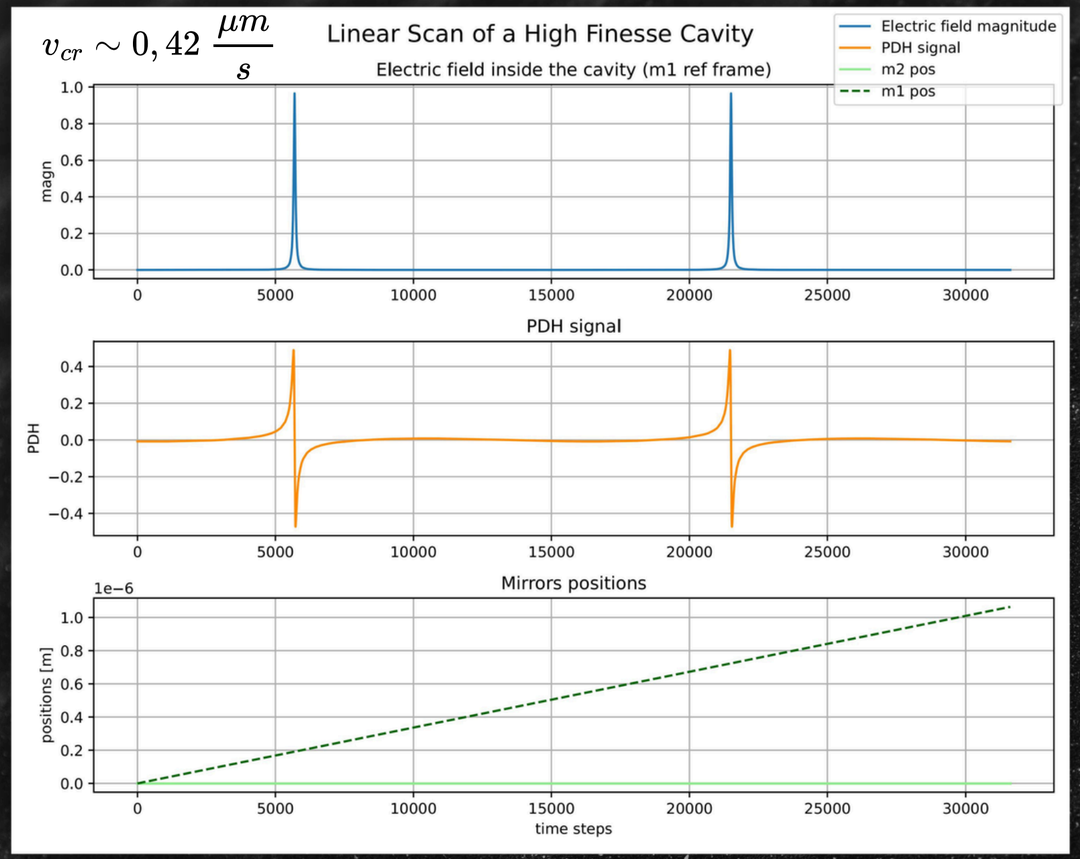
Why

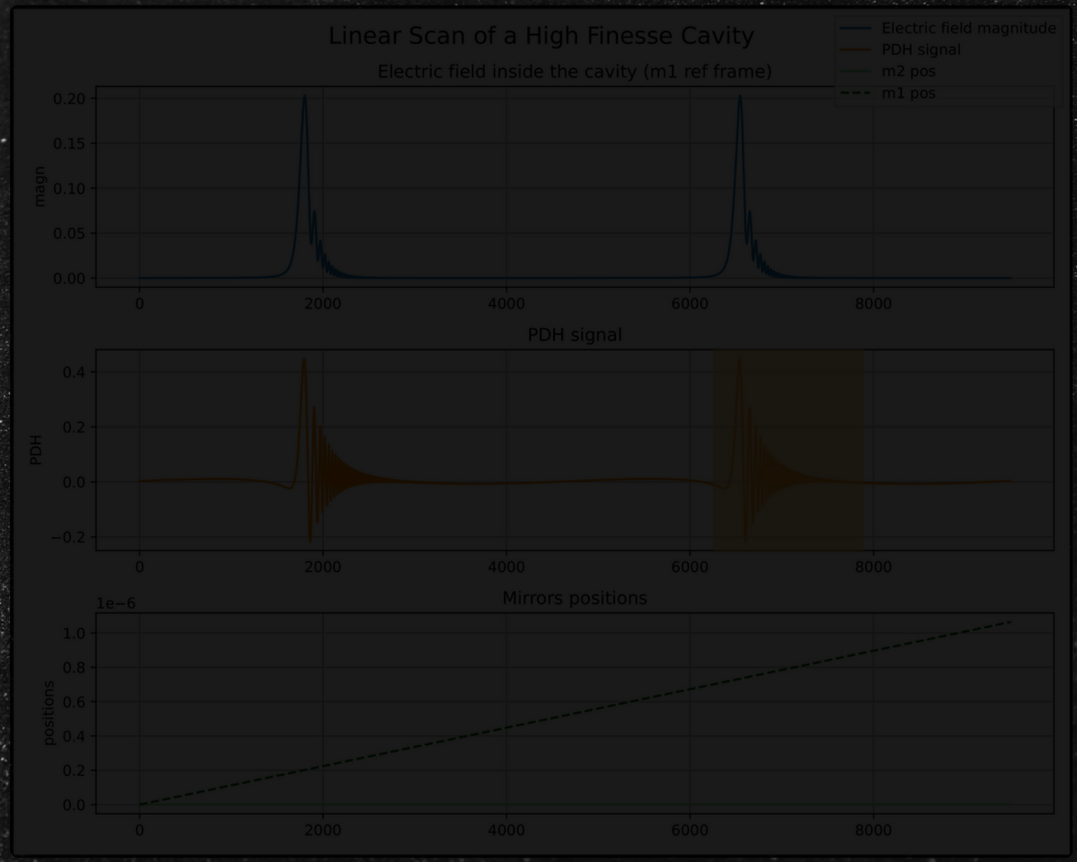
Reinforcement Learning controllers need many safe and repeatable trials before real-hardware deployment.

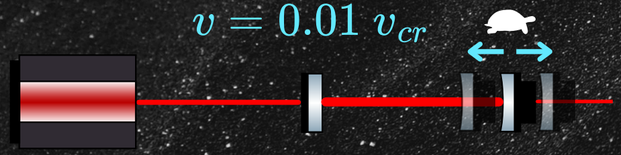
The non-linear regime: resonance crossing and ring-down

Fast mirror motion makes the optical response history-dependent

$v_{cr} \sim 0,42 \ \frac{\mu m}{s}$



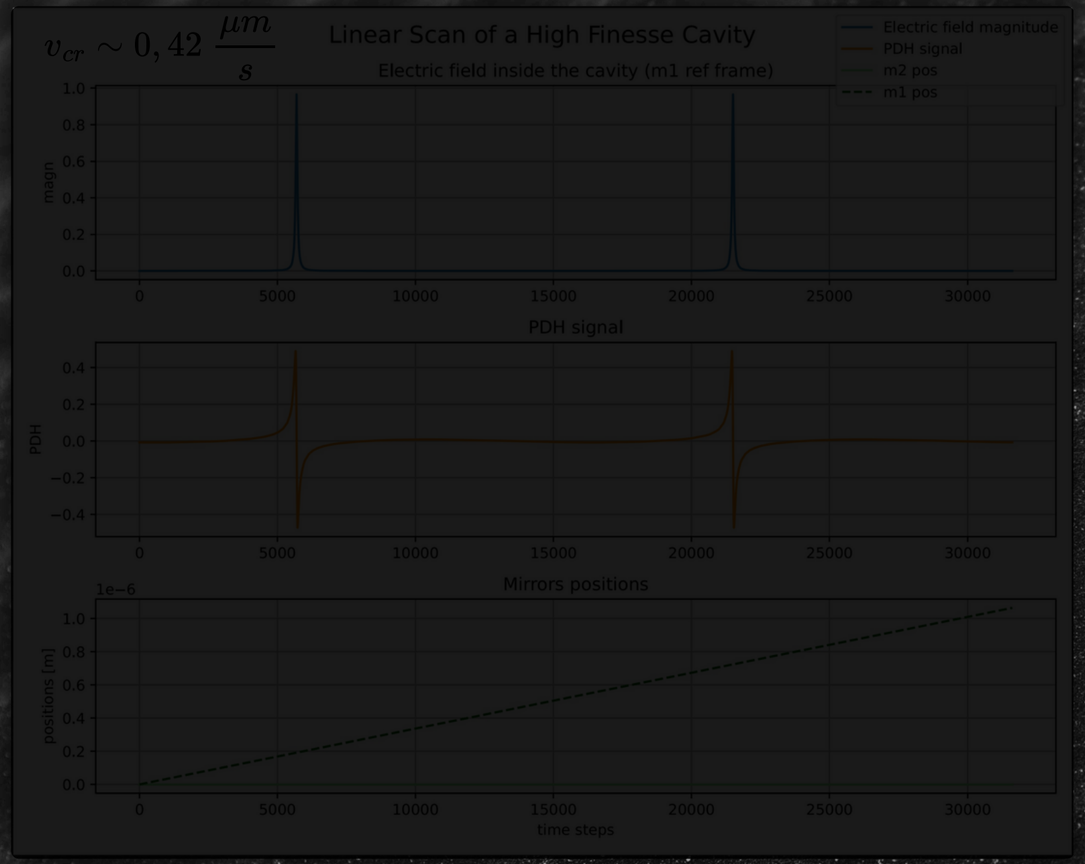


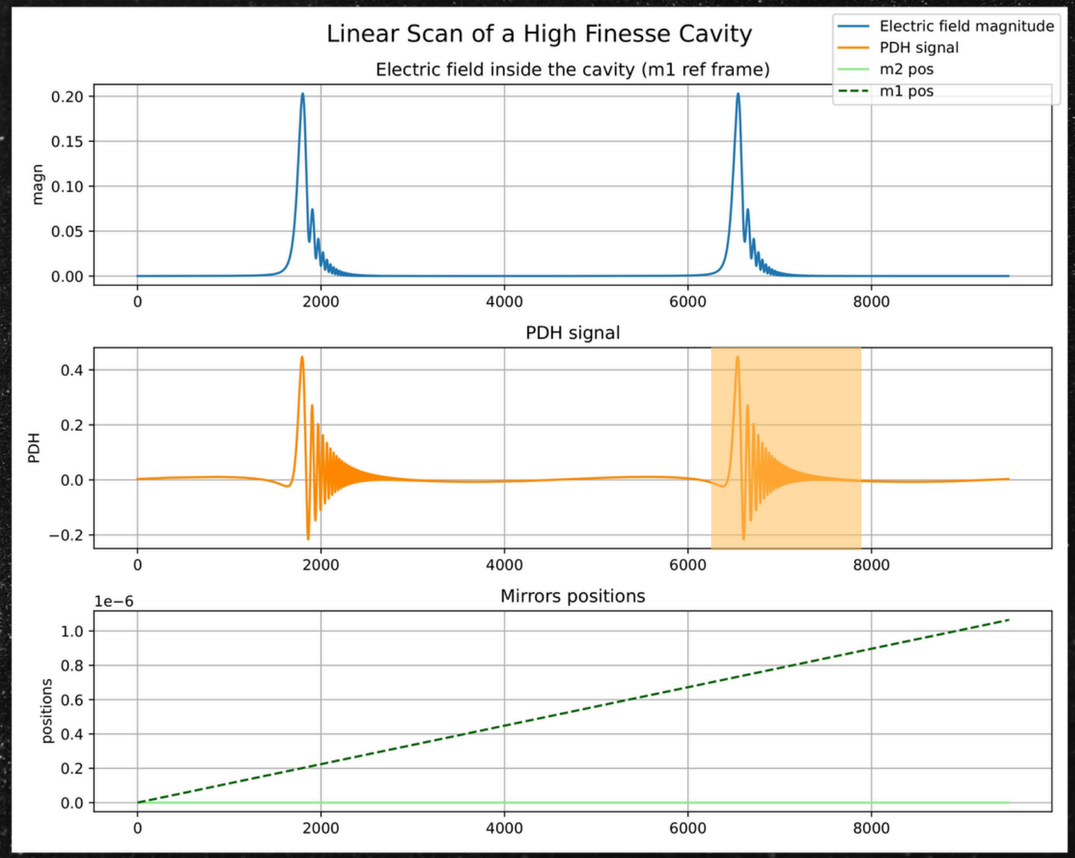


The non-linear regime: resonance crossing and ring-down

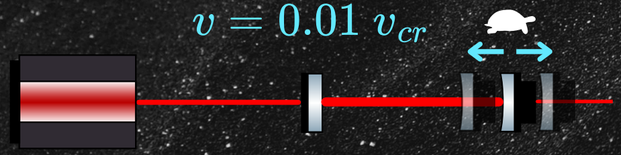
Fast mirror motion makes the optical response history-dependent

$v_{cr} \sim 0,42 \ \frac{\mu m}{s}$





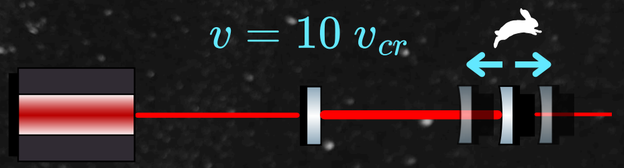
$v = 0.01 \ v_{cr}$



$v \geq v_{\mathrm{cr}} \approx \frac{\lambda \pi c}{4 L \mathcal{F}^2}$

Ring-down Effect

$v = 10 \ v_{cr}$



Design goals

A simulator built around arbitrary dynamics and control interaction

1. Arbitrary boundary conditions

Mirror positions and input electric field can be changed at every simulation step.

Step by step implementation

2. Round-trip consistency

The requested sampling frequency is internally adapted to remain compatible with the cavity round-trip structure.

Adaptive sampling frequency

3. Computational efficiency

Cavity memory is represented with a finite buffer and recursive state updates, avoiding repeated full-history integrations.

Numba backend

Core model: recursive field evolution

The intra-cavity field is a sum over round trips plus a memory tail

$$E(t) = t_{\mathrm{a}} \sum_{n=0}^{N-1} (r_{\mathrm{a}} r_{\mathrm{b}})^n e^{-2ik S_{\mathrm{n}}(t)} E_{\mathrm{in}}(t - 2nT) + (r_{\mathrm{a}} r_{\mathrm{b}})^N e^{-2ikS_N(t)} E(t - 2NT)$$

$$S_{\mathrm{n}}(t) = \sum_{p=0}^{n-1} d(t - 2pT) \qquad d(t) = x_{\mathrm{b}}(t) - x_{\mathrm{a}}(t)$$

Round-trip sum

Each term carries attenuation, phase accumulation, and delayed input field.

Moving mirrors

The accumulated optical path includes the time-dependent input and end mirror positions.

Memory tail

The final recursive term stores the contribution of earlier cavity history.

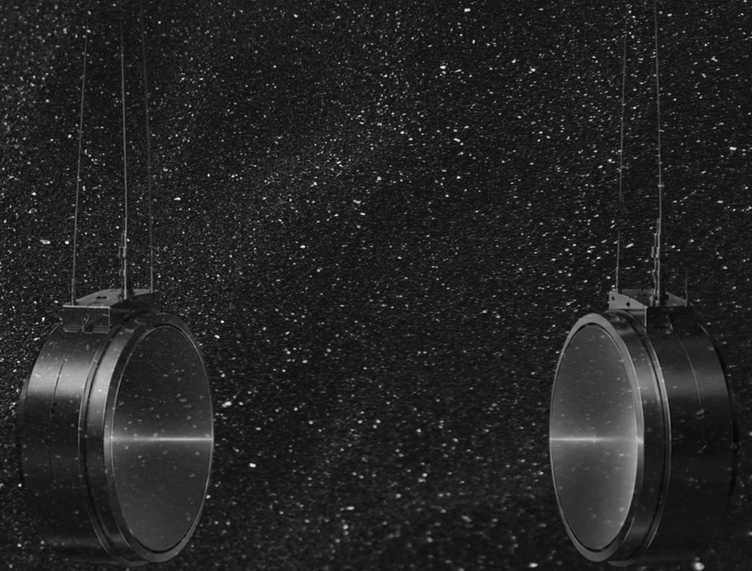
From field to observables

The simulator outputs the signals used by cavity control systems

- Intracavity and transmitted power can be obtained from electric field.
- PDH error signal in the reflected-sideband approximation.

$$V_{\mathrm{PDH}}(t) = -\,\mathrm{Im}\left\{e^{i\gamma} E_{\mathrm{in}}(t)^{*} E(t)\right\}$$

- Both signals can be computed at every control step.
- These observables are exactly what an RL agent can receive as state information.

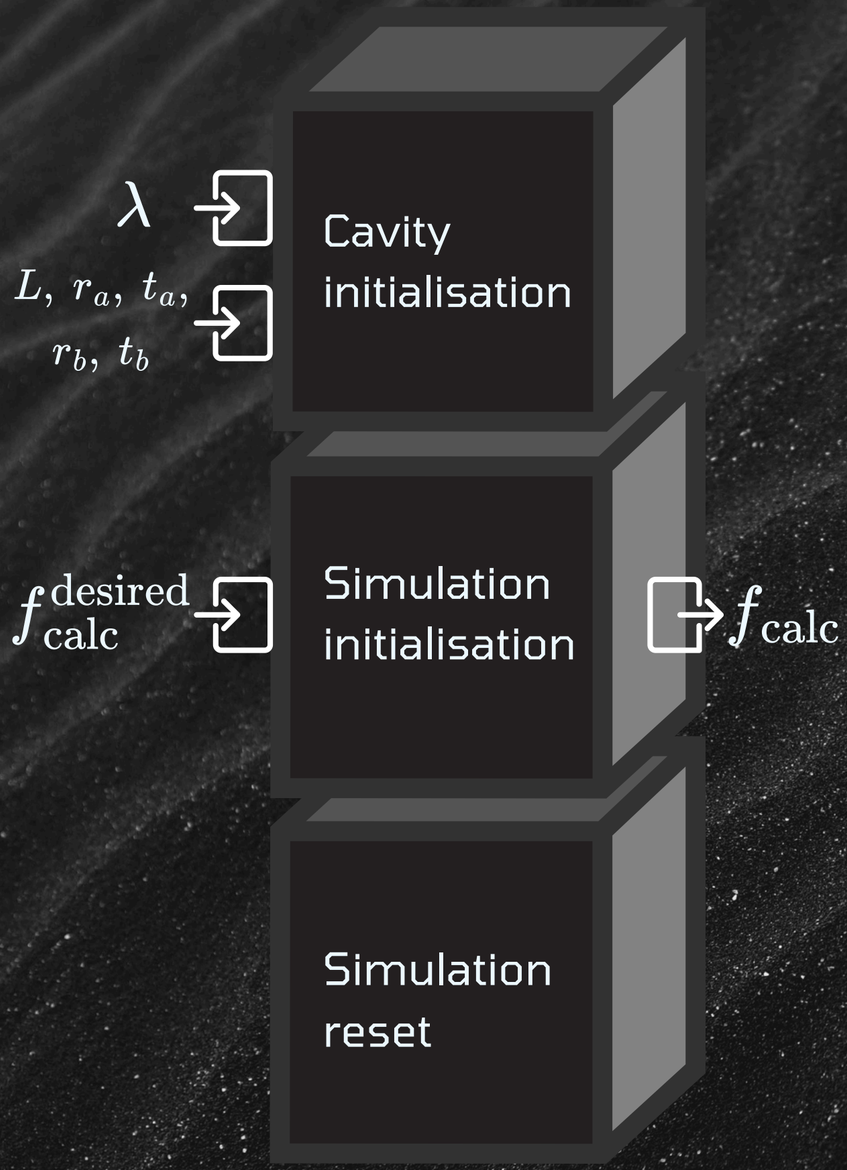


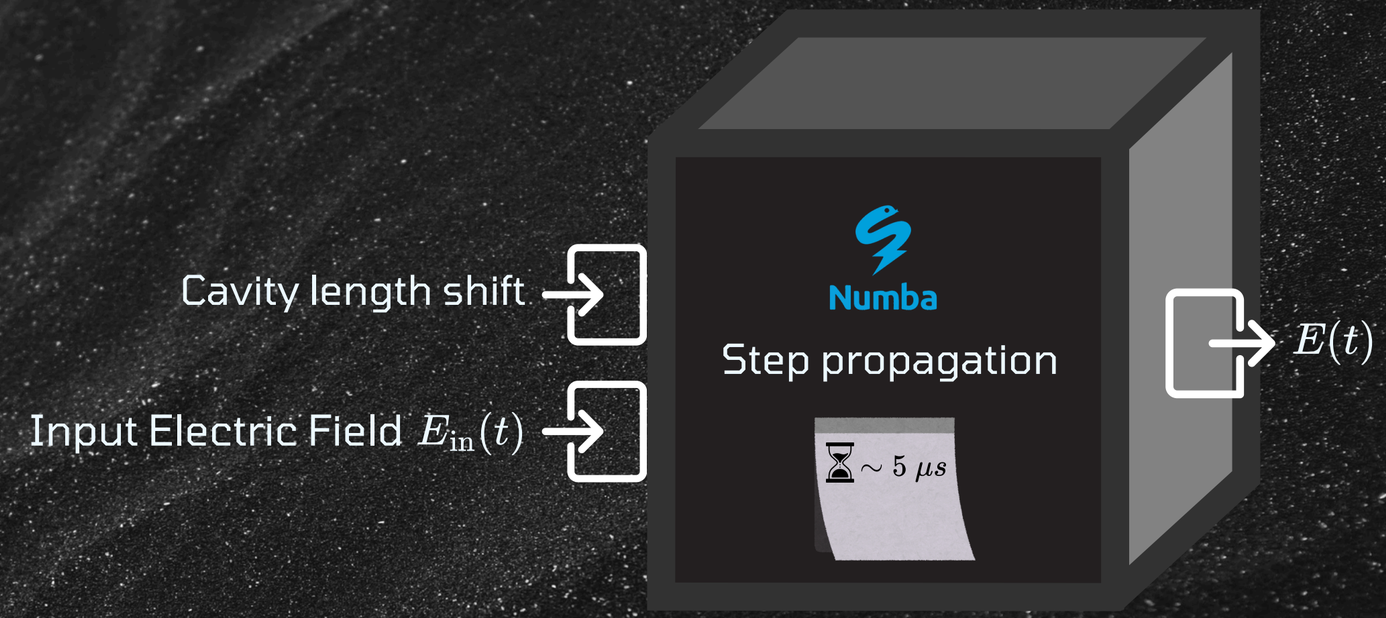
N.B: The cavity is stationary and reactive, once you send an input the length will change immediately and it will remain in that state until you will send another shift.

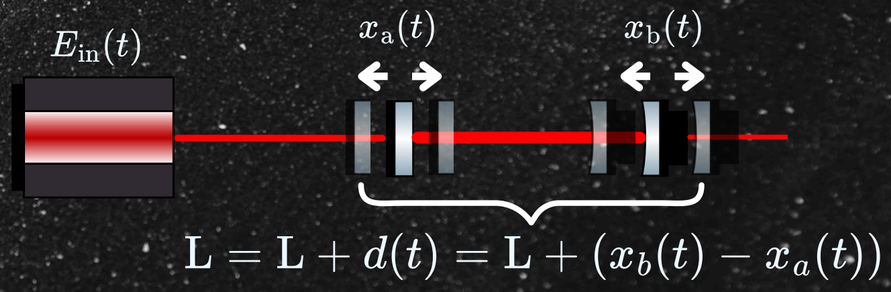
The simulator framework

nspy

A gymnasium-ready interface



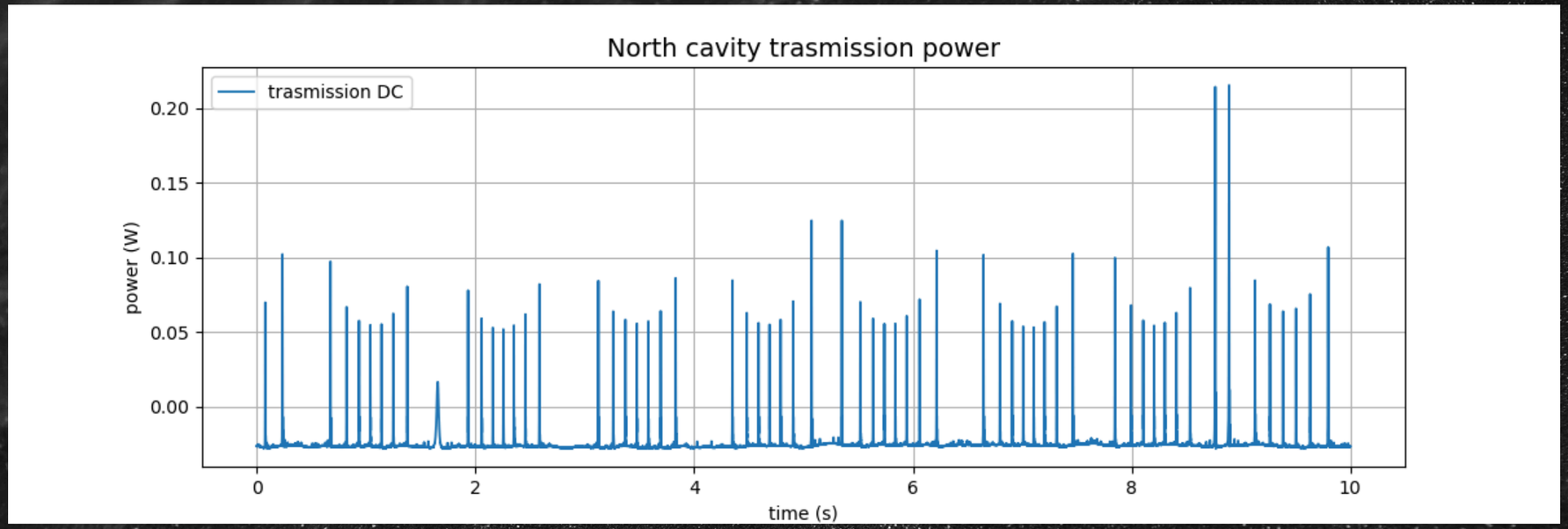




Gymnasium

Simulator validation

Virgo Data comparison



arXiv.2605.13599

Swinging mirror transmission data.

Mirror position and speed reconstruction for each resonance.

Simulator propagation for all the time window.

Why Reinforcement Learning?

Overcoming human performance

Highly suitable in several control tasks and for highly dimensional parameter spaces.

Classic feedback controllers	Deep Reinforcement Learning Models
• Assume linear, time-invariant system responses. • Often require manual tuning and do not adapt to changing dynamics.	• Learn a control policy by interacting with the environment. • Optimizing performance directly based on a reward function, even in the absence of a precise model.

The purpose is not replacing classical controllers but find better control strategies for the non linear regime of the system

- Previous works with several Machine Learning and Reinforcement Learning applications for controlling and alignigning tasks

[2] *"A Deep Learning Technique to Control the Non-linear Dynamics of a Gravitational-wave Interferometer" P. Ma, G. Vajente 2023*

[3] *"First demonstration of neural sensing and control in a kilometer-scale gravitational wave observatory" N. Mukund et al. 2023*

[4] *"Interferobot: aligning an optical interferometer by a reinforcement learning agent" D. Sorokin et al. 2021*

[5] *"Automated alignment of an optical cavity using machine learning" J. Qin et al. 2025*

[6] *"Improving cosmological* reach of a gravitational wave observatory using Deep Loop Shaping*" J. Buchli et al. 2025*

[7] "Continuous control with deep reinforcement learning" T. P. Lillicrap et al. 2019
[8] "Gymnasium: A Standard Interface for Reinforcement Learning Environments" M. Towers et al. 2024

Agent

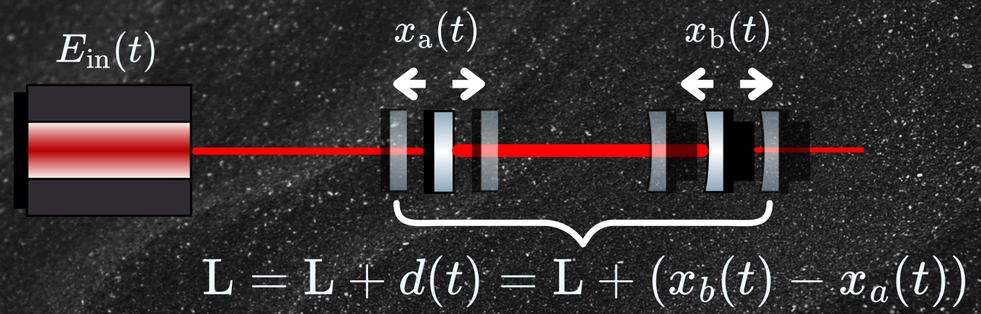
RL Model

Action

$$a_t = \{x_a(t), x_b(t)\}$$

Simulated Environment

Implemented with Gymnasium [10]



Reinforcement Learning Framework

[7] "Continuous control with deep reinforcement learning" T. P. Lillicrap et al. 2019
[8] "Gymnasium: A Standard Interface for Reinforcement Learning Environments" M. Towers et al. 2024

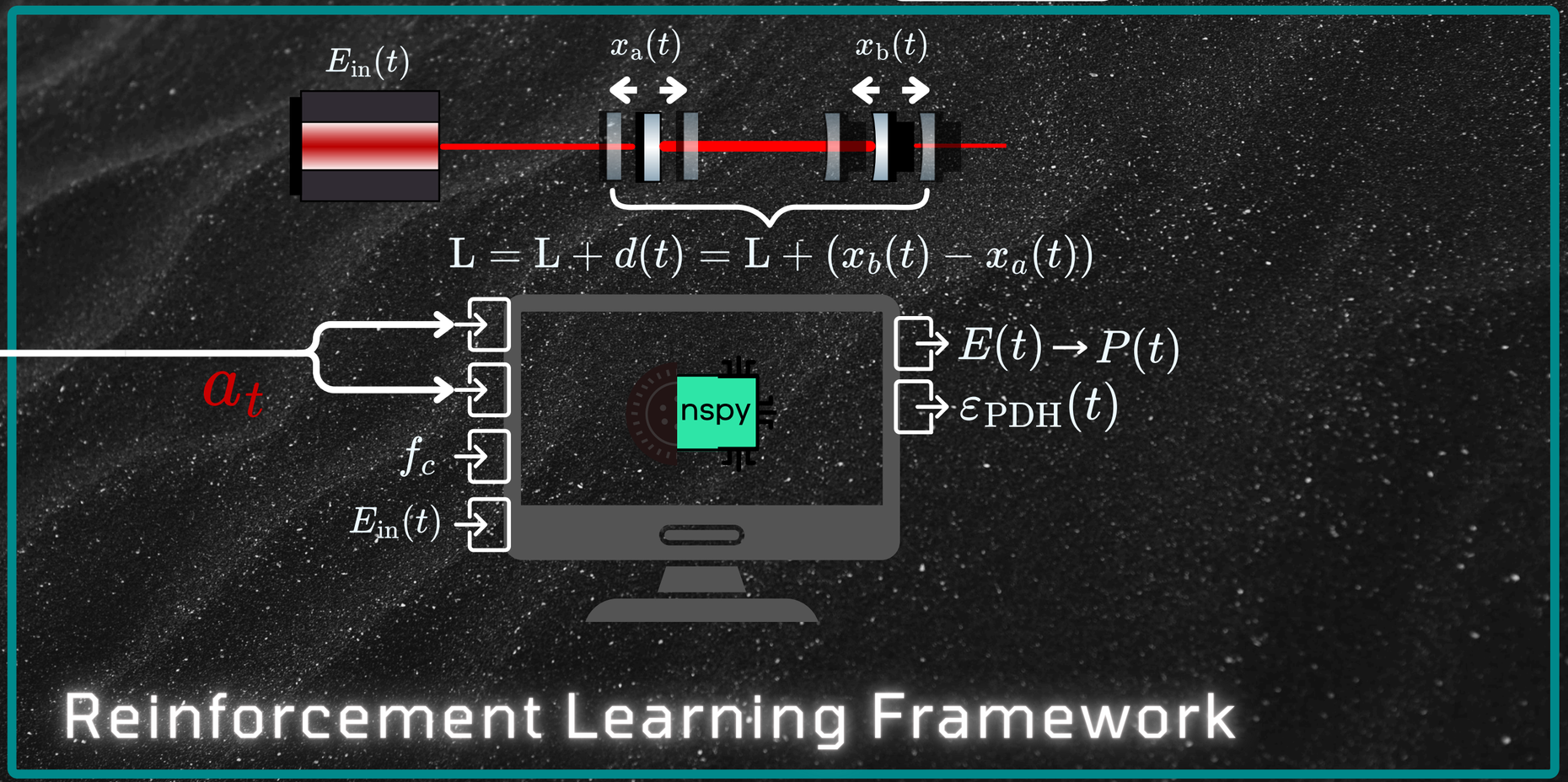
Agent

Action

$a_t = \{x_a(t), x_b(t)\}$

RL Model

Simulated Environment Implemented with Gymnasium [10]



Reinforcement Learning Framework

[7] "Continuous control with deep reinforcement learning" T. P. Lillicrap et al. 2019
[8] "Gymnasium: A Standard Interface for Reinforcement Learning Environments" M. Towers et al. 2024

Agent

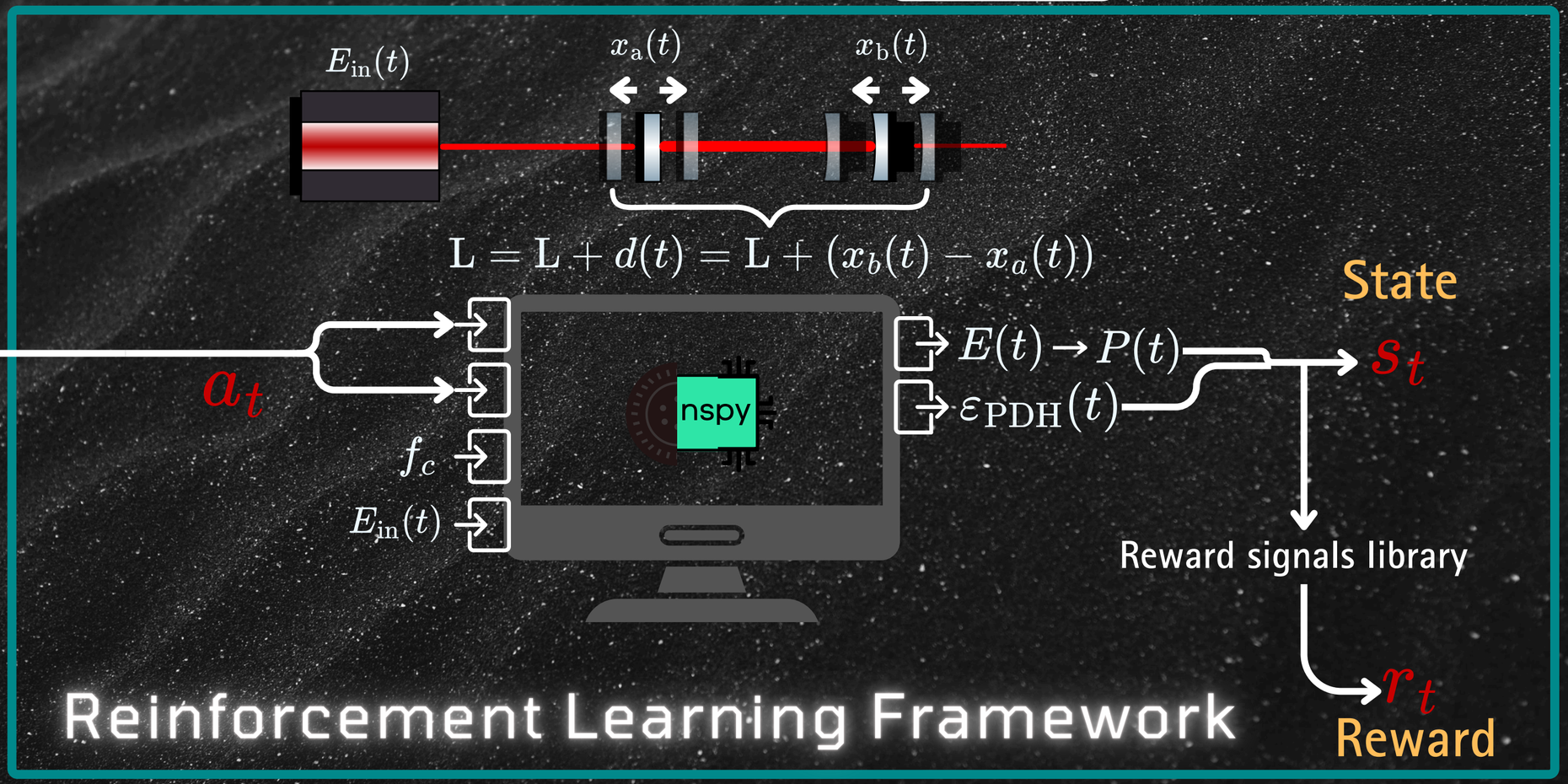
Action

$a_t = \{x_a(t), x_b(t)\}$

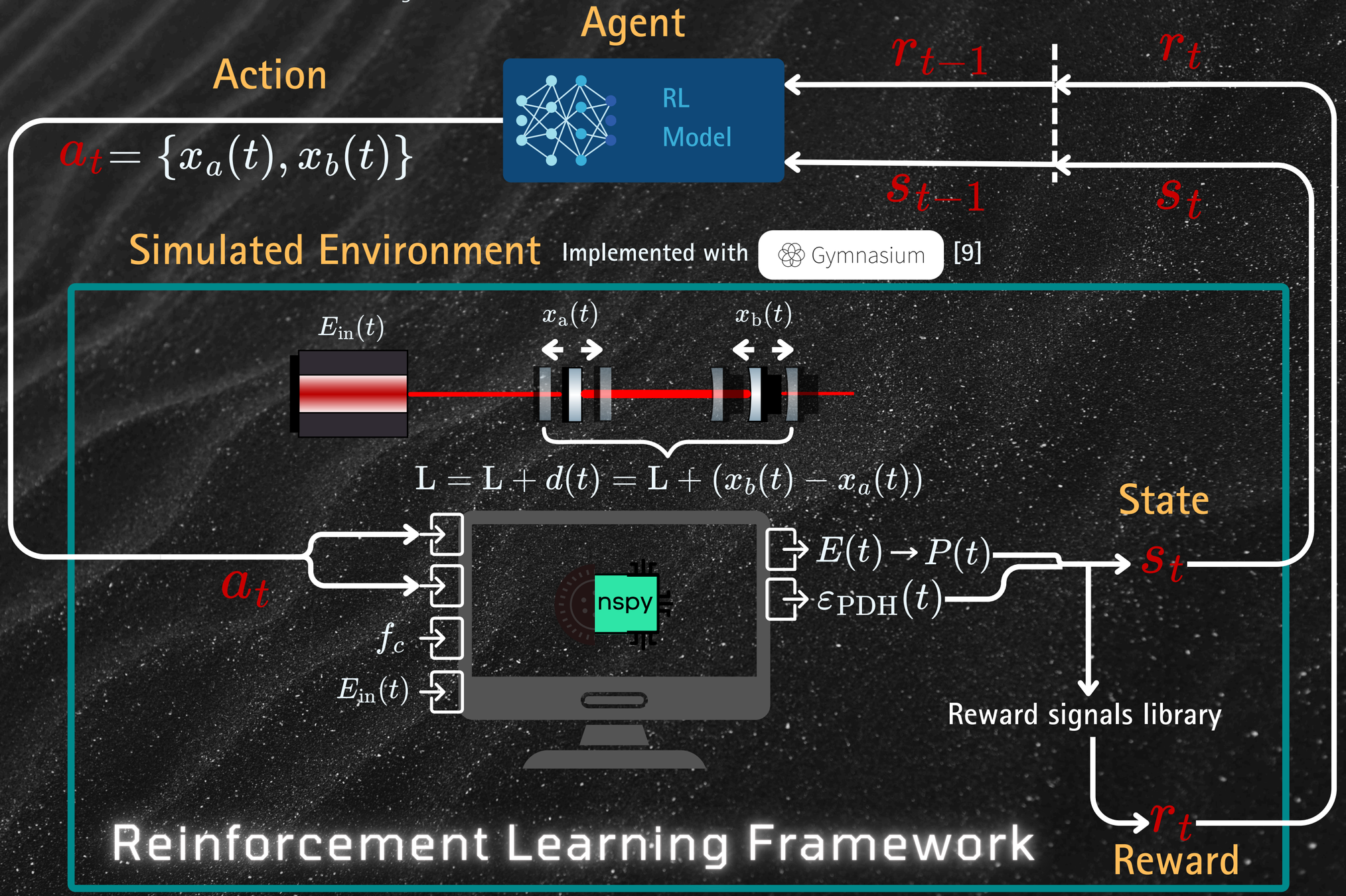
RL Model

Simulated Environment

Implemented with Gymnasium [10]



[7] "Continuous control with deep reinforcement learning" T. P. Lillicrap et al. 2019
[8] "Gymnasium: A Standard Interface for Reinforcement Learning Environments" M. Towers et al. 2024

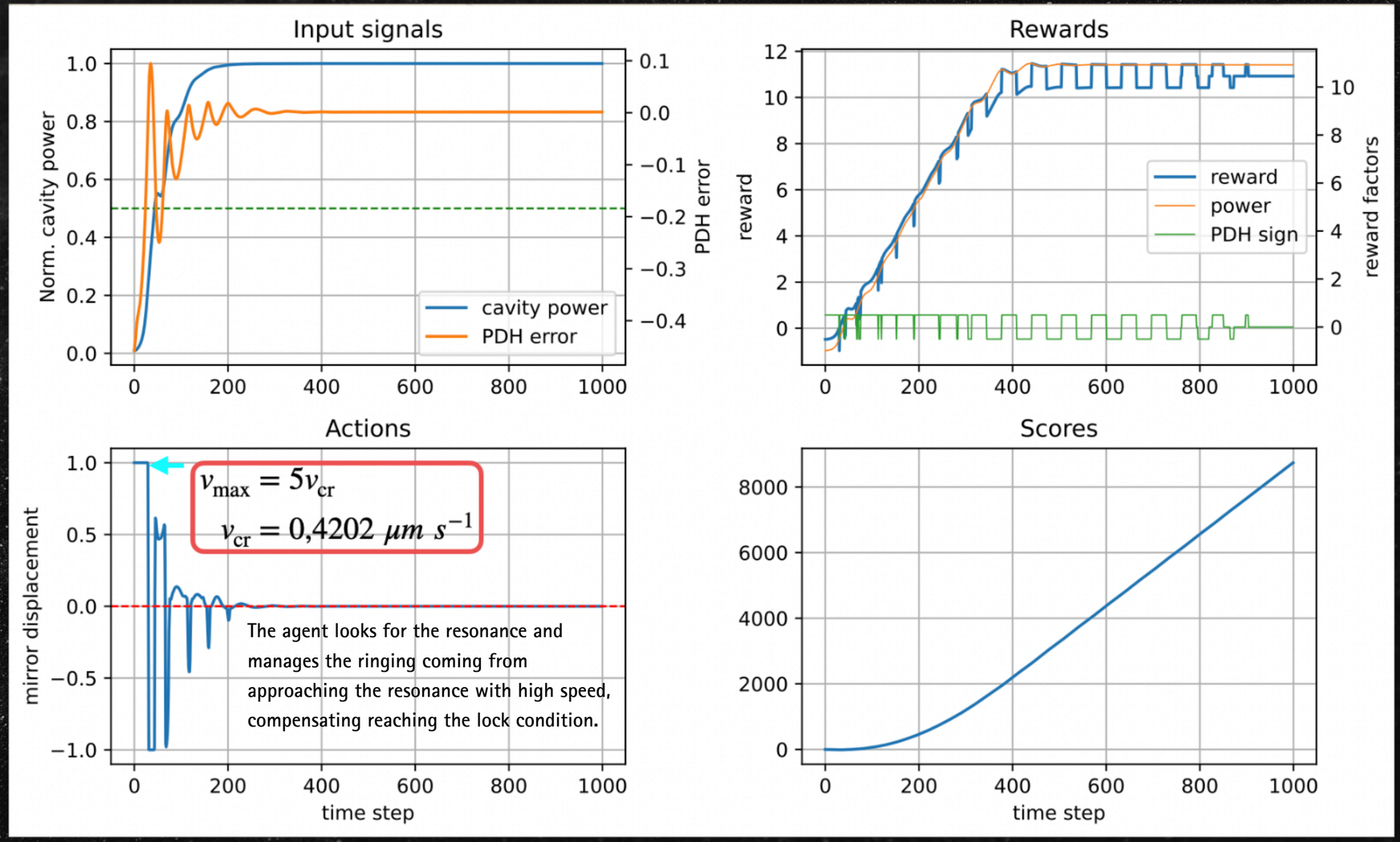


Train and Test

First results

Test of a DDPG model trained for 120000 time-steps.

- Only for the output mirror.
- ~ 200 time-steps to properly lock the cavity.
- The agent approaches the resonance with maximum speed allowed by the environment

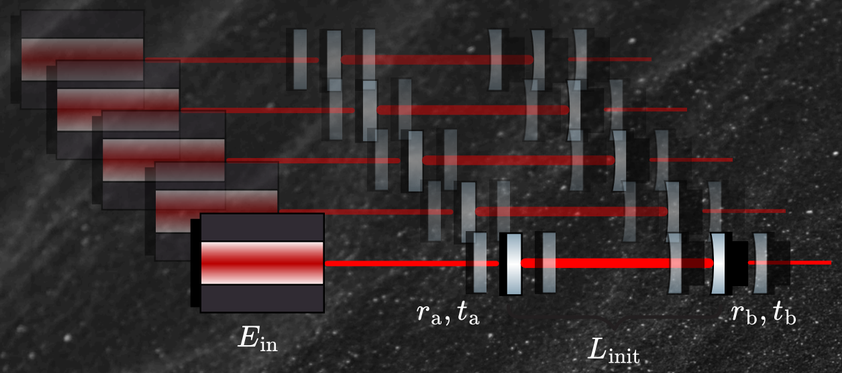


HPC Training campaign

Investigating model-reward configurations

Different models

- DDPG
- TD3
- RecurrentPPO

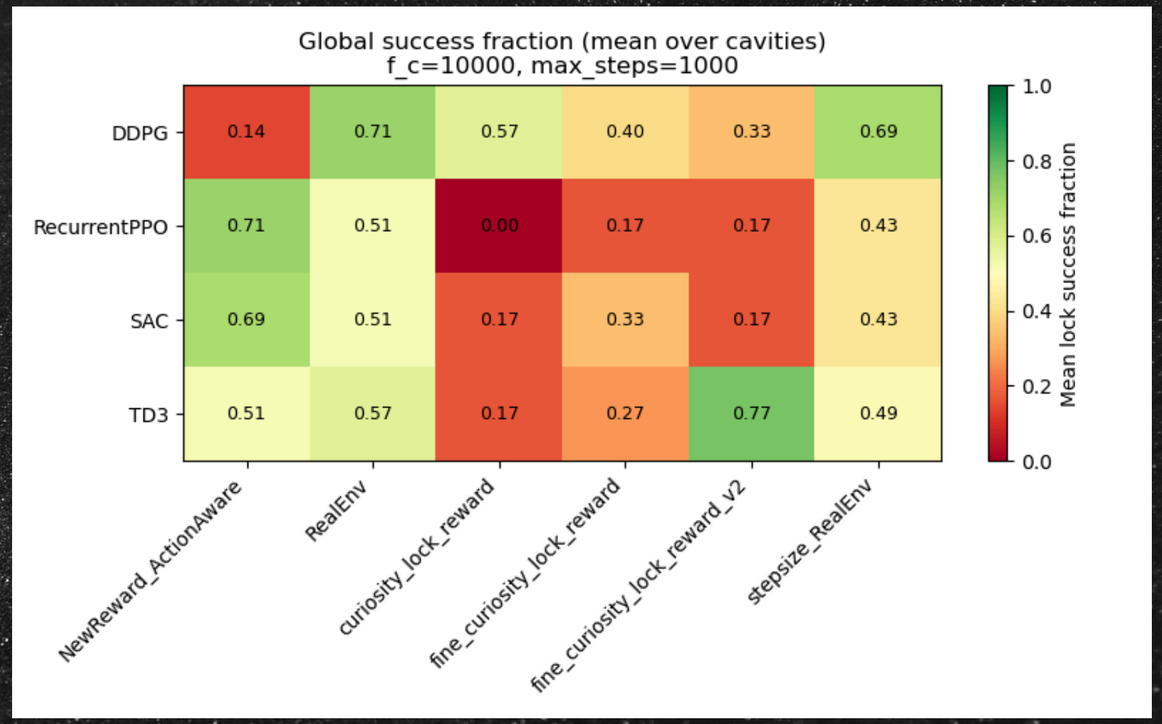


Different cavities

- ARM
- Microcavities
- ...

Domain Randomization

$\langle L_{init}, E_{in}, \text{signal noise} \ldots \rangle$



HPC Infrastructures

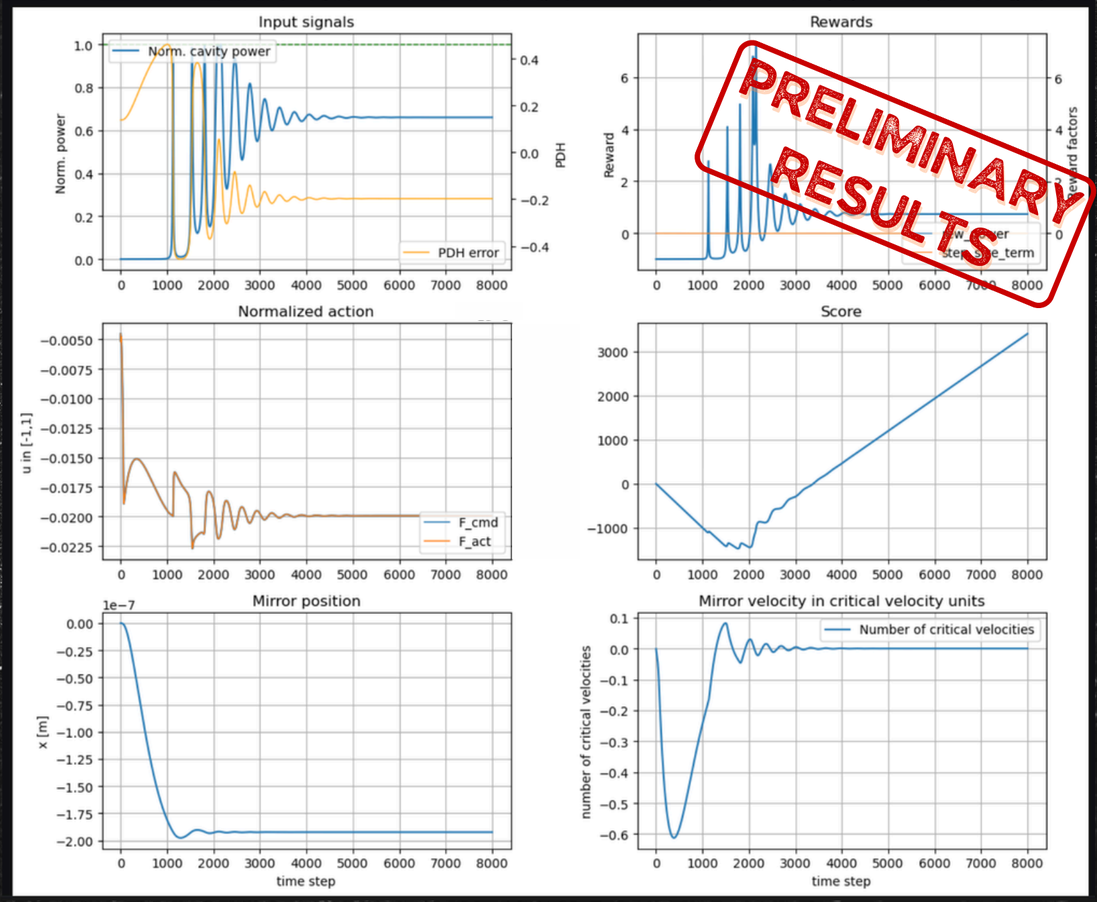
KEY BENEFIT

Run heavy training sessions to find the best model–reward configurations for future SimToReal transfer.

Takeaways and Future Goals

Oreonspy as a bridge between cavity physics and AI control

- Recursive time-domain model preserves cavity memory while remaining computationally efficient.
- Adaptive sampling keeps the simulation consistent with round-trip propagation.
- Validation against Virgo data shows good agreement in a nonlinear ring-down crossing.
- Numba backend supports faster and safe reinforcement learning training.
- NEXT STEP → Implement mirror physics and finite actuator force (WORK IN PROGRESS)



Near future goal

Test a pre-trained model on table top setup → Currently working on the hardware for acquisition and deployment

References

[1] *“An Introduction to Pound-Drever-Hall laser frequency stabilisation” Eric D. Black, 2001*

[2] *"A Deep Learning Technique to Control the Non-linear Dynamics of a Gravitational-wave Interferometer" P. Ma, G. Vajente 2023*

[3] *"First demonstration of neural sensing and control in a kilometer-scale gravitational wave observatory" N. Mukund et al. 2023*

[4] *"Interferobot: aligning an optical interferometer by a reinforcement learning agent" D. Sorokin et al. 2021*

[5] *"Automated alignment of an optical cavity using machine learning" J. Qin et al. 2025*

[6] "Improving cosmological reach of a gravitational wave observatory using Deep Loop Shaping" J. Buchli et al. 2025

[7] *"Dynamics of Laser Interferometric Gravitational Wave Detectors" M. Rakhmanov, Phd Thesis, 2000*

[8] *"Continuous control with deep reinforcement learning" T. P. Lillicrap et al. 2019*

[9] *“Gymnasium: A Standard Interface for Reinforcement Learning Environments” M. Towers et al. 2024*

[10] *"Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: a Survey" Wenshuai Zhao et al. 2021*

“Advanced Virgo Plus: Future Perspectives” Acernese et al. 2023

"New algorithm for the Guided Lock technique for a high-Finesse optical cavity" D. Bersanetti et al. 2019

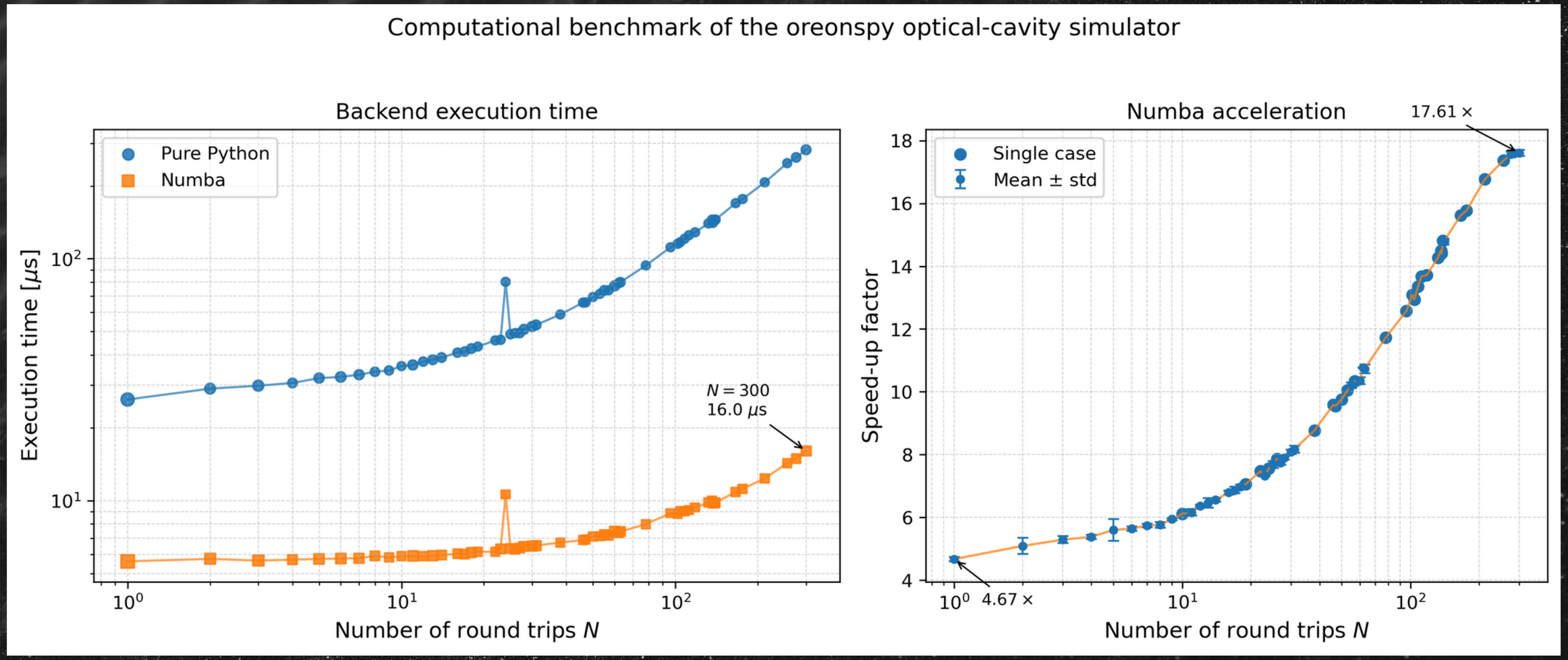
Proceedings

"Partial Observability and Domain Randomization in RL-Based Strategy for Optical Cavity Locking Optimization " A. Svizzeretto et M. Bawaj 2025, Pos

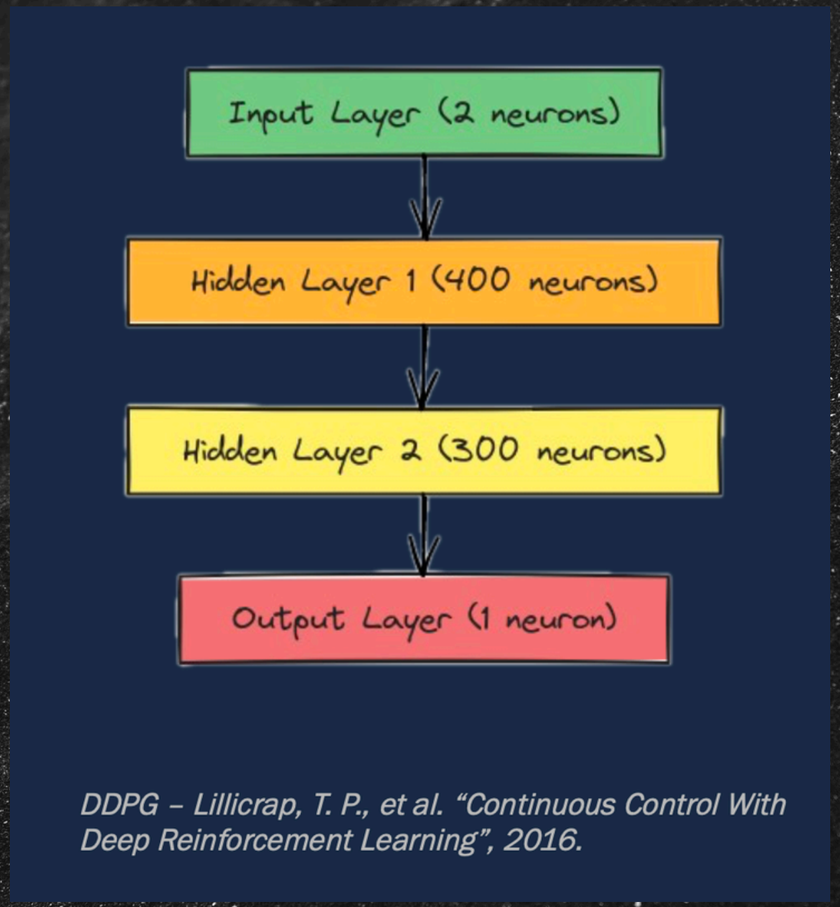
"Applying reinforcement learning to optical cavity locking tasks: considerations on actor-critic architectures and real-time hardware implementation" M. Bawaj et A. Svizzeretto 2025

Thanks for your attention!

BACKUP


Computational benchmark of the oreonspy optical-cavity simulator
Backend execution time
Pure Python
Numba
$N = 300$
16.0 μs
Execution time [μs]
10^2
10^1
10^0
10^1
10^2
Number of round trips N
Numba acceleration
17.61 ×
Single case
Mean ± std
Speed-up factor
18
16
14
12
10
8
6
4
10^0
4.67 ×
10^1
10^2
Number of round trips N

Backup



DDPG – Lillicrap, T. P., et al. "Continuous Control With Deep Reinforcement Learning", 2016.

Backup

DSP for Super Attenuator control 3.125 µs delay,

Real Time DAQ 100 µs

DDPG actor inference in 1.20 ± 0.87 ms.

DAQ a maximum rate of 200 Hz.

Backup

NewReward_ActionAware

$$[r_{\text{power}} = P - \log(1-P) - 1][r_{\text{action}} = 0.8\,\text{sign}(-E\,a)][r = r_{\text{power}} + \frac{1}{2} r_{\text{action}}]$$

RealEnv

$$[r = -|P-1| - \log(|1-P|)]$$

curiosity_lock_reward

$$[g(P) = \frac{1}{1 + e^{-k(P - p_0)}}]$$

$$[r_{\text{power}} = \tanh(3P)][r_{\text{pdh}} = -0.5\,E^2][r_{\text{act}} = -(0.01 + 0.20\,g(P))\,a^2][r_{\text{lock}} = 1 - e^{-0.3\,t_{\text{lock}}}]$$

$$[\Delta P = P_t - P_{t-1}][\Delta E = E_t - E_{t-1}][\text{novelty} = \sqrt{\Delta P^2 + w_E\,\Delta E^2}][r_{\text{int}} = \beta \tanh(\alpha\,\text{novelty})(1 - g(P))]$$

$$[r = r_{\text{power}} + r_{\text{pdh}} + r_{\text{act}} + r_{\text{lock}} + r_{\text{int}}]$$

stepsize_RealEnv

$$[r_{\text{power}} = -|P-1| - \log(|1-P|)] \qquad [\text{err} = (1-P) + E^2][r_{\text{step}} = -0.05 \frac{|a|}{\text{err} + 0.05}]$$

$$[r = r_{\text{power}} + r_{\text{step}}]$$

Backup

fine_curiosity_lock_reward_v2

$$[g_c(P,E) = \exp\left(-\frac{(1-P)^2}{2\sigma_P^2}\right)\exp\left(-\frac{E^2}{2\sigma_E^2}\right)]$$

$$[r_{\text{power}} = \tanh(3P)] \qquad [r_{\text{peak}} = 0.9\exp\left(-\frac{(1-P)^2}{2\sigma_P^2}\right)]$$

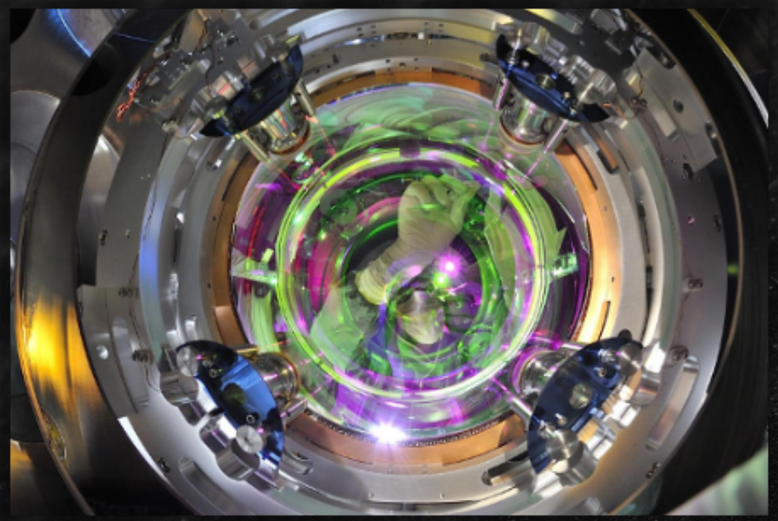
$$[r_{\text{pdh}} = -(0.2 + 5g(P))E^2] \quad [r_{\text{act}} = -(0.01 + 0.25g_c)a^2]$$

$$[r_{\text{align}} = 0.15\ \tanh(3(-Ea))\,g(P)] \quad [r_{\text{int}} = \beta\tanh(\alpha\,\text{novelty})(1-g(P))]$$

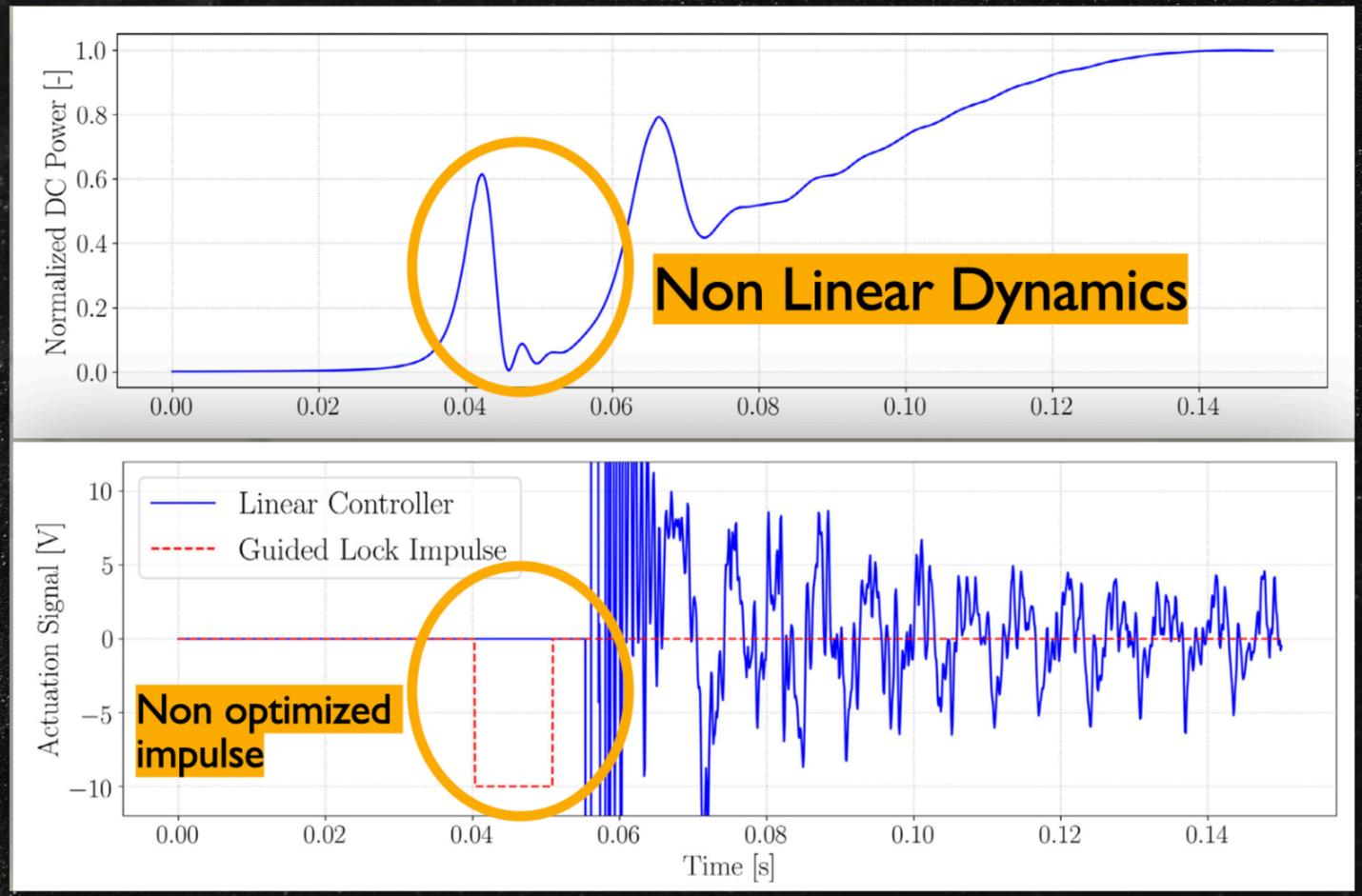
$$[r_{\text{lock}} = 1 - e^{-0.3\,t_{\text{lock}}}]$$

$$[r = r_{\text{power}} + r_{\text{peak}} + r_{\text{pdh}} + r_{\text{act}} + r_{\text{align}} + r_{\text{lock}} + r_{\text{int}}]$$

Modified Guided Lock

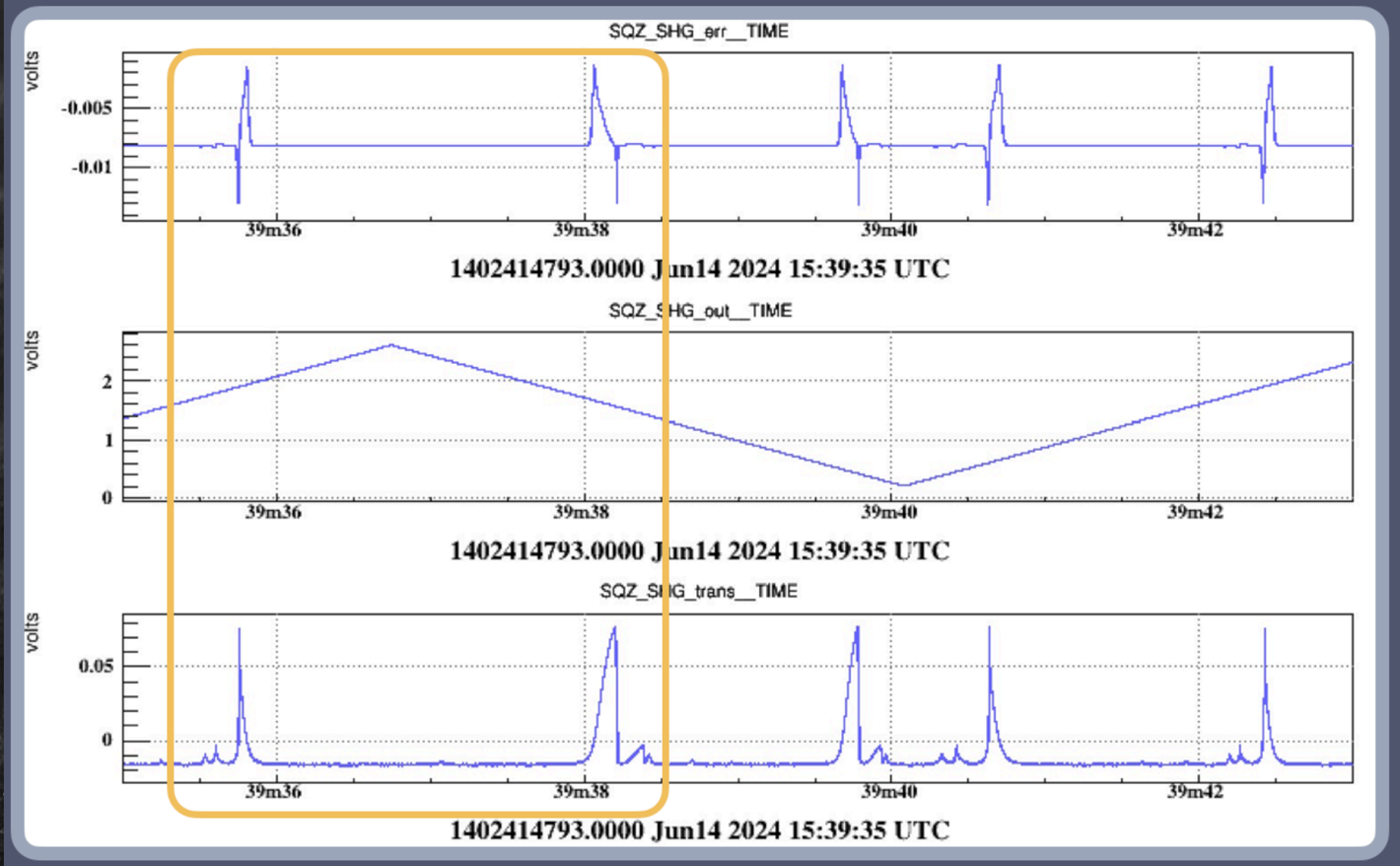


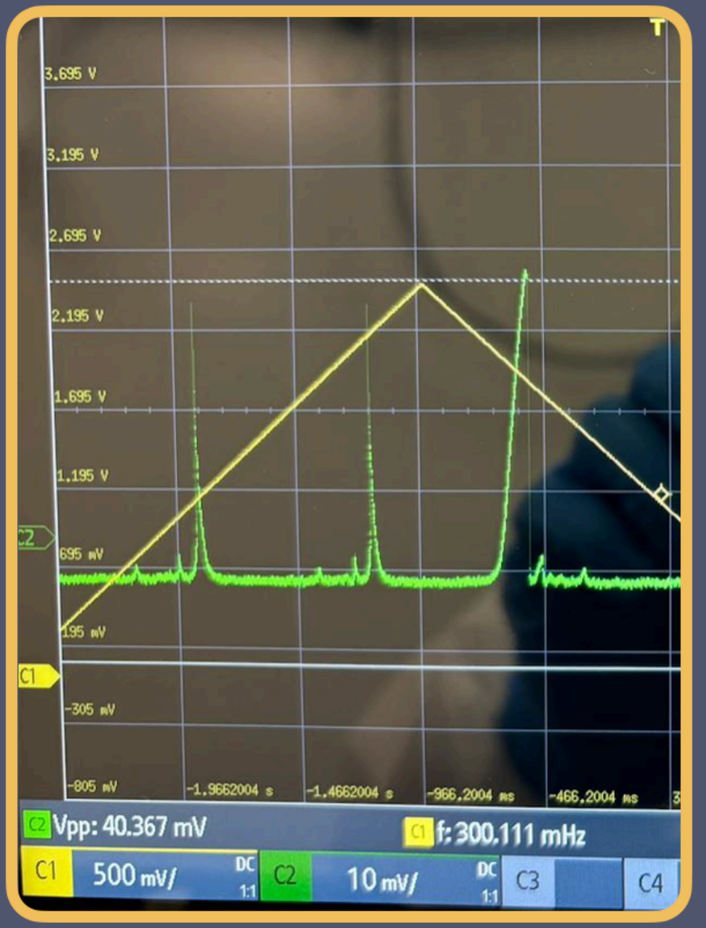
Currently Ring Down effect is managed by actuating with maximum force available.



[7] *"New algorithm for the Guided Lock technique for a high-Finesse optical cavity" D. Bersanetti et al. 2019*

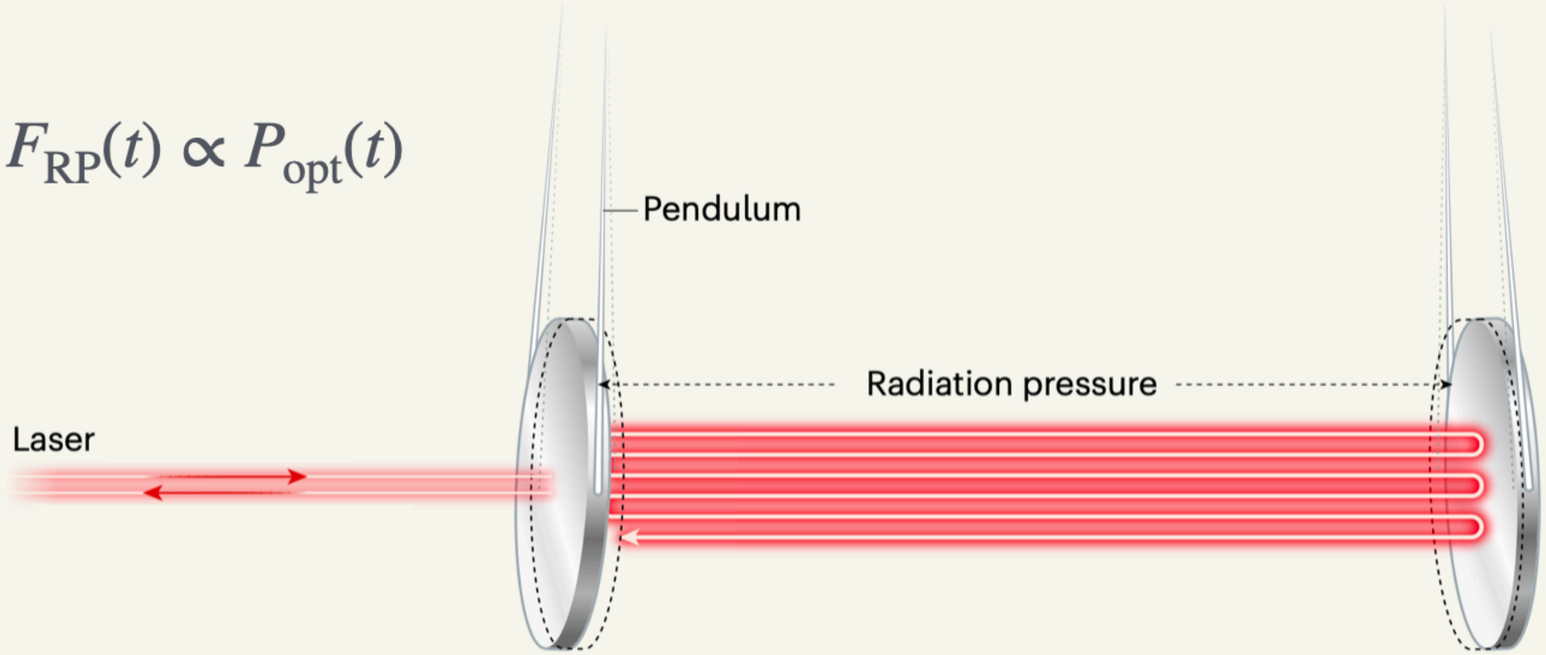
Backup


SQZ_SHG_err__TIME
volts
-0.005
-0.01
39m36
39m38
39m40
39m42
1402414793.0000 Jun14 2024 15:39:35 UTC
SQZ_SHG_out__TIME
volts
2
1
0
39m36
39m38
39m40
39m42
1402414793.0000 Jun14 2024 15:39:35 UTC
SQZ_SHG_trans__TIME
volts
0.05
0
39m36
39m38
39m40
39m42
1402414793.0000 Jun14 2024 15:39:35 UTC


3.695 V
3.195 V
2.695 V
2.195 V
1.695 V
1.195 V
C2
695 mV
195 mV
C1
-305 mV
-805 mV
-1.9662004 s
-1.4662004 s
-966.2004 ms
-466.2004 ms
C2 Vpp: 40.367 mV
C1 f: 300.111 mHz
C1 500 mV/ DC 1:1
C2 10 mV/ DC 1:1
C3
C4

Backup

$$F_{\mathrm{RP}}(t) \propto P_{\mathrm{opt}}(t)$$



In resonance condition, $F_{\mathrm{RP}}(t)$ will be at its maximum.