Autonomous Fabry-Perot cavity locking via deep reinforcement learning

Mateusz Bawaj[1,2]
mateusz.bawaj@unipg.it

Andrea Svizzeretto[1,2]
andrea.svizzeretto@dottorandi.unipg.it

June 17th, 2025

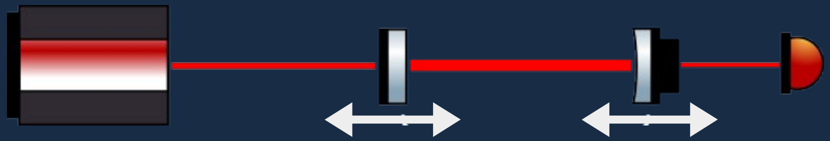
EUROPEAN AI FOR FUNDAMENTAL PHYSICS CONFERENCE
EuCAIFCon 2025

Cagliari, Sardinia, Italy

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

Problem statement

What is the optimal way to actuate cavity mirrors to reach resonance under real-world conditions?



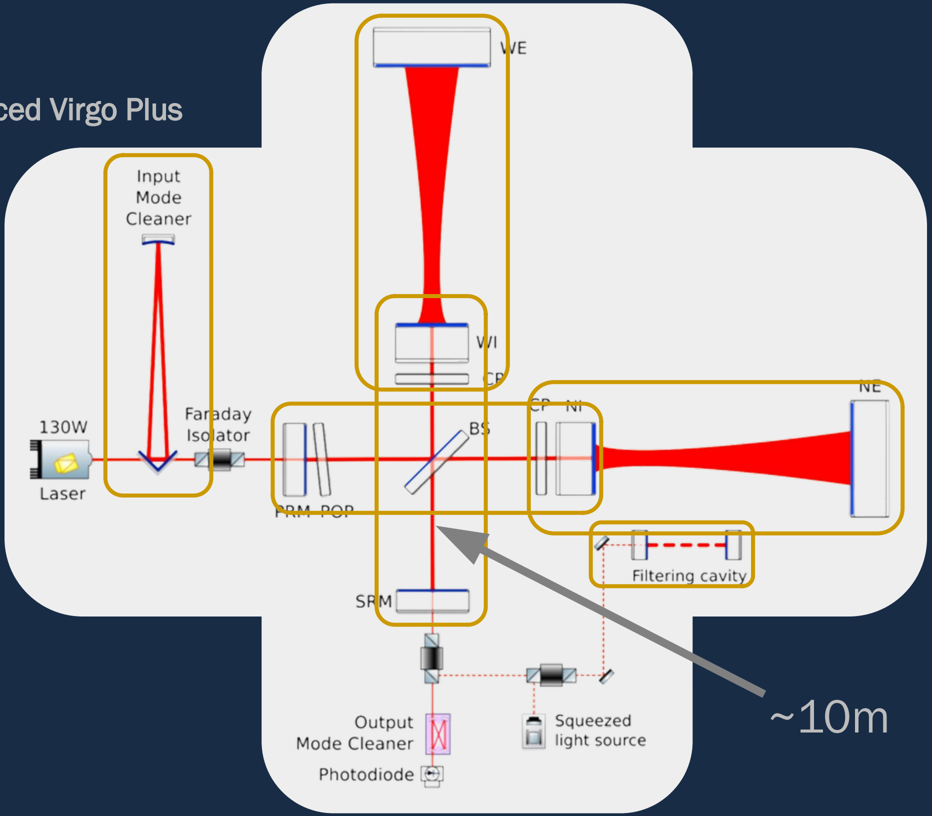
Let's try with a **feedback loop**:

- plant
- continuous plant state measurement
- ML model inference process run after each state measurement
- actuator action

= implementation of an AI controller

Plant – first approximation

Advanced Virgo Plus



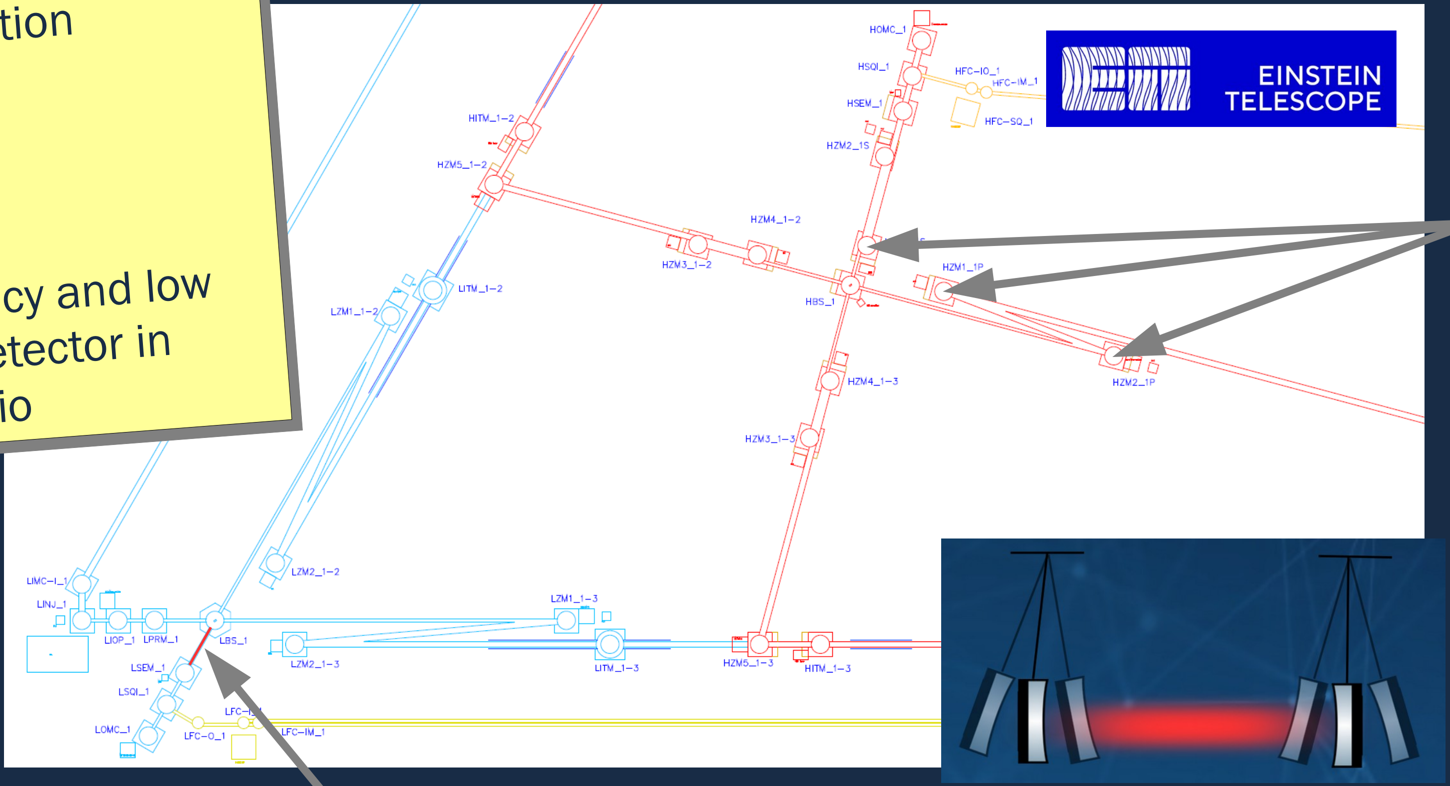
- seismic isolation
- laser injection
- suspended interferometer
- suspended optical cavities
- quantum state injection system
- thermal compensation

Acernese et al. "Advanced Virgo Plus: Future Perspectives", 2023 J. Phys.: Conf. Ser. 2429 012040

Plant – Einstein Telescope

Two configuration scenarios:
- triangle
- 2L

High-frequency and low frequency detector in each scenario

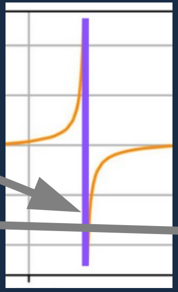


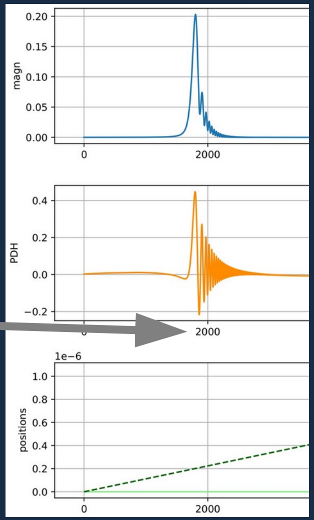
~10m

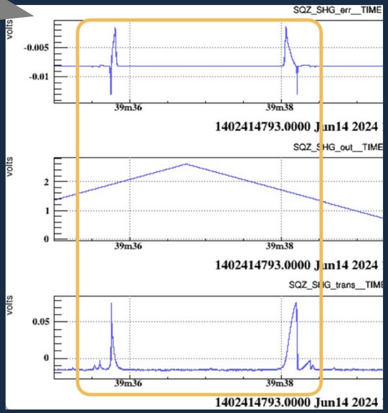
J. Degallaix et al. "Optical layout update", 2024 ET-0123A-24

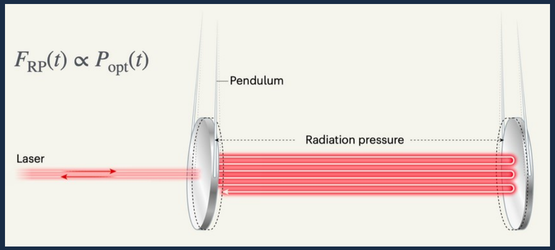
Non-linearities we expect

- Narrow linear range of the error signal
- Dynamics
- Hysteresis effect (during scan)
- Response of the actuator
- Time-dependency
- Cross-coupling between DoFs



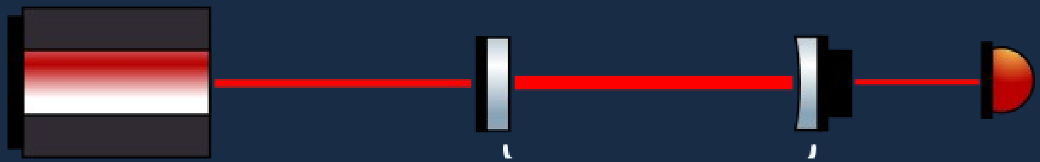


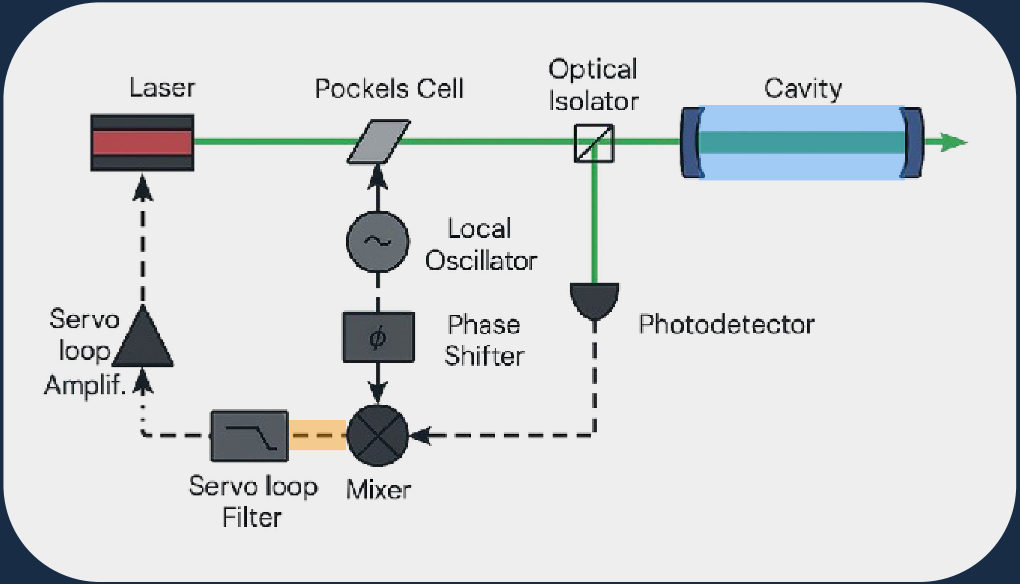


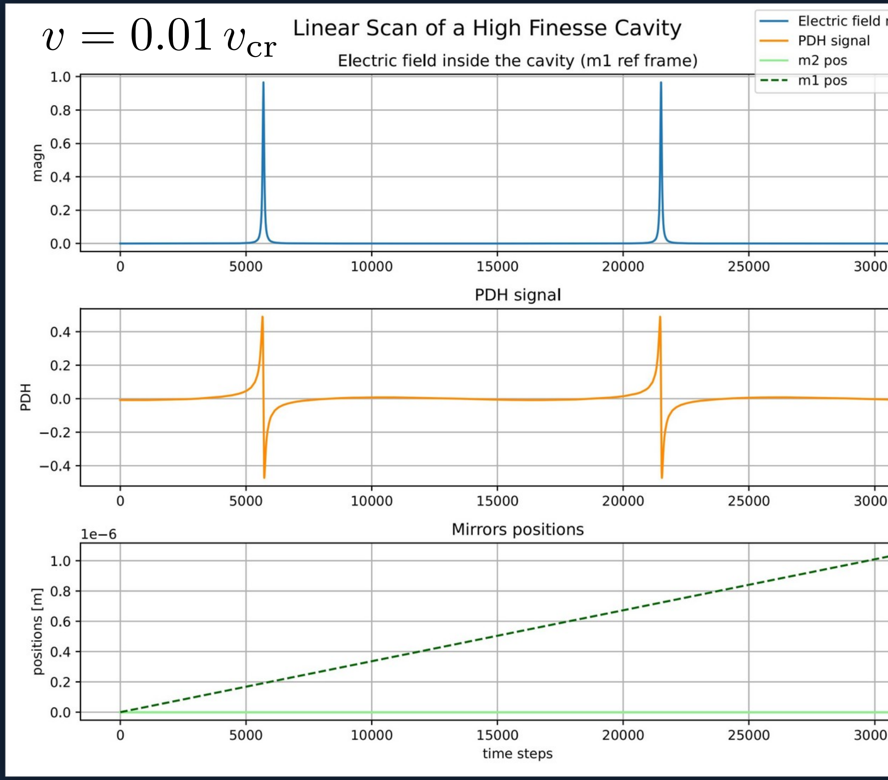


Sequino V., et al. 2020. "Quantum Fluctuations Have Been Shown to Affect Macroscopic Objects." NATURE 583:31–32.

Plant – longitudinal DoF







Black, Eric D. 2001. "An Introduction to Pound–Drever–Hall Laser Frequency Stabilization." American Journal of Physics 69 (1): 79–87.

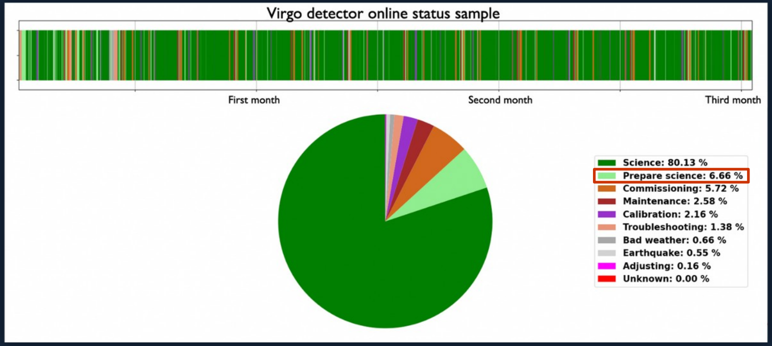
Classical strategies

- Narrow linear range of the error signal ◂ Conditional execution
- Dynamics ◂ Conditional execution/Adiabatic regime
- Hysteresis effect (during scan) ◂ Adiabatic regime
- Response of the actuator ◂ Linearisation
- Time-dependency ◂ Recurrent retuning
- Cross-coupling between DoFs ◂ Filtering & suboptimal gain

Why to use ML, RL in our case?

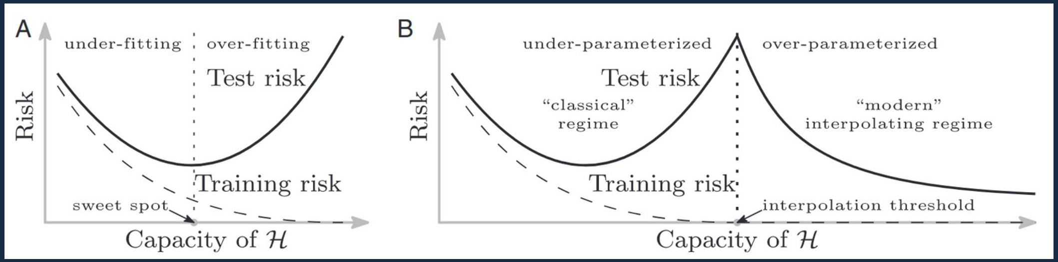
To speed-up lock acquisition!

- Ability to learn non-linear functions
- Modern over-parameterized models effectively handle over-fitting
- Direct performance optimization using a reward function, even without an explicit system model
- Reinforcement Learning enables learning control policies through interaction with the environment
- Inherently suitable for MIMO (Multi-Input Multi-Output) systems



Classical risk curve

Deep risk curve



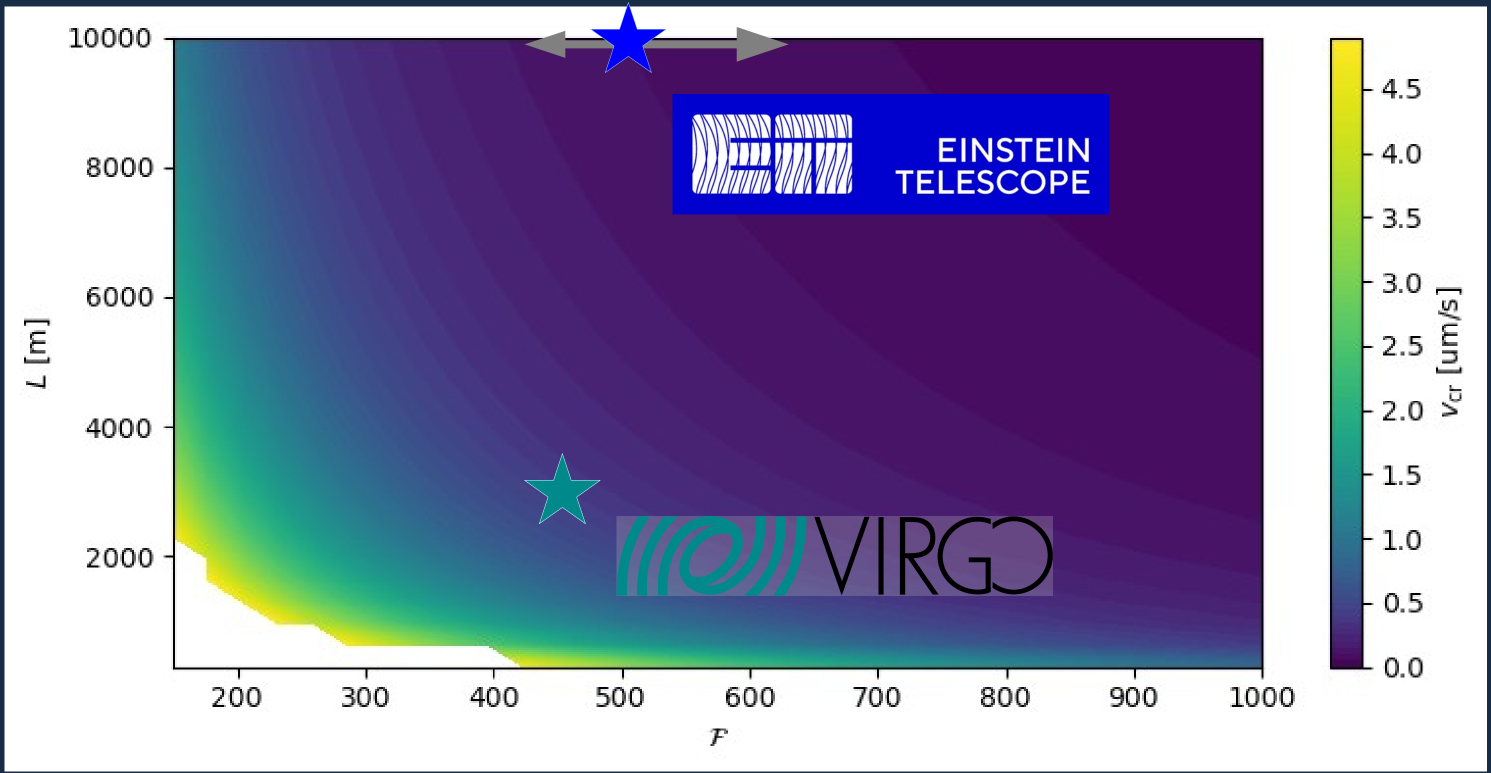
Belkin, M., D. Hsu, S. Ma, and S. Mandal. 2019. "Reconciling Modern Machine-Learning Practice and the Classical Bias–Variance Trade-Off." PNAS 116 (32): 15849–54.

Longitudinal lock acquisition

critical velocity:

$$v_{\mathrm{cr}} = \frac{\lambda \pi c}{4L\mathcal{F}}$$

where:
L – cavity length
$\mathcal{F}$ – cavity Finesse



Current control system & DSP

Each UDSPT unit contains:

- 6 analog-to-digital and 6 digital-to-analog converters (24-bit, 1MSPS);
- FPGA and DSP.

Units in the same create communicate via back panel.

Units of different crate communicate via network.

Each unit can become a TOLM.

There are 20 crates in Advanced Virgo

It means ~200 UDSPT boards in total.

UDSPT unit boards in a μTCA chassis. On the right there is the NAT-MCH switch. Entire crate has ~800 GFLOPS of computing power available.

Nguyen, C., et al. "Automated Source of Squeezed Vacuum States Driven by Finite State Machine Based Software." Review of Scientific Instruments 92, no. 5 (May 1, 2021): 054504.

Implementation attempt
simulated environment

Input variables: mirrors position $x_a(t)$, $x_b(t)$, input electric field $E_{in}(t)$
Simulation parameters: mirrors reflectivity r_a, r_b, computation frequency

$$E(t) = t_a \sum_{n=0}^{N-1} (r_a r_b)^n \, e^{-2\imath k S_n(t)} \, e^{-2\imath k x_a(t)} \, E_{\mathrm{in}}(t - 2nT) + \\ + (r_a r_b)^N \, e^{-2\imath k S_N(t)} \, E(t - 2NT)$$

where:

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

Numba

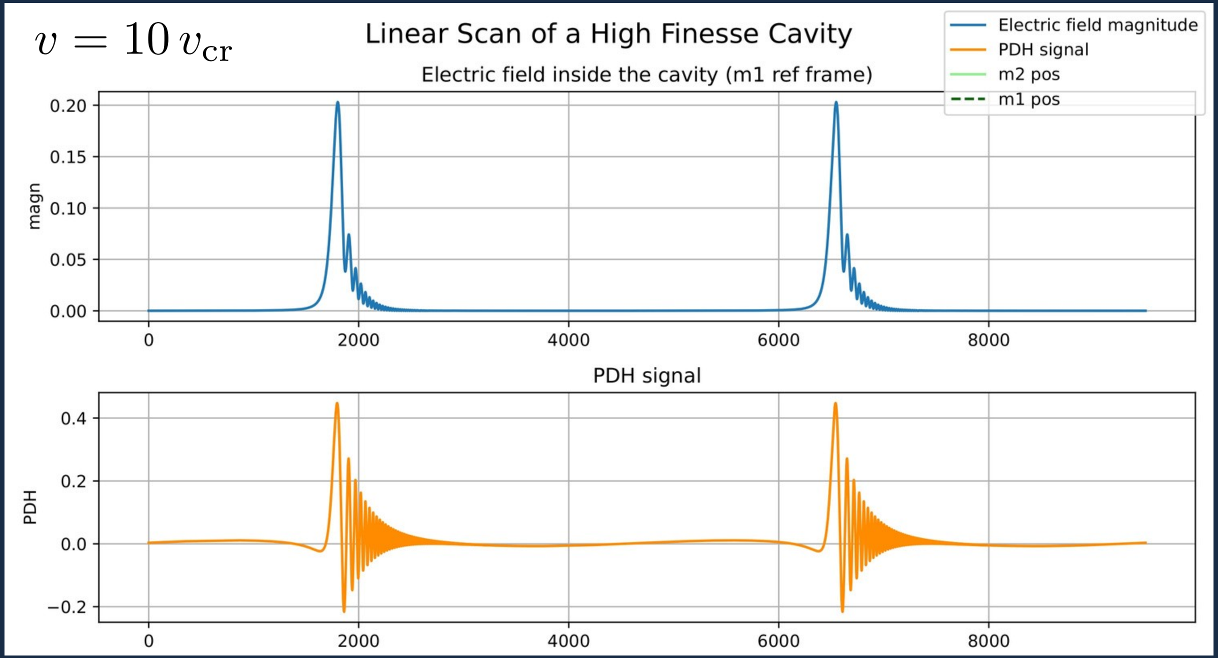
We used Numba to accelerate the loop and we reached 50us of execution time.

M. Rakhmanov, "Dynamics of Laser Interferometric Gravitational Wave Detectors" ,Phd Thesis, 2000

Implementation attempt

simulated environment

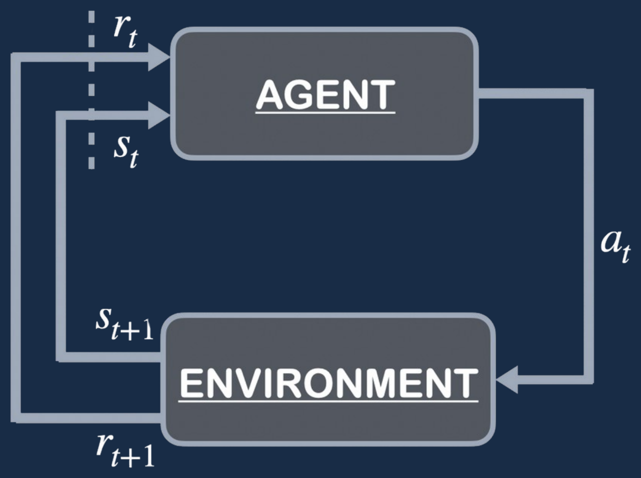
We are able to simulate response of a wide variety of cavities in time-domain in an efficient way.

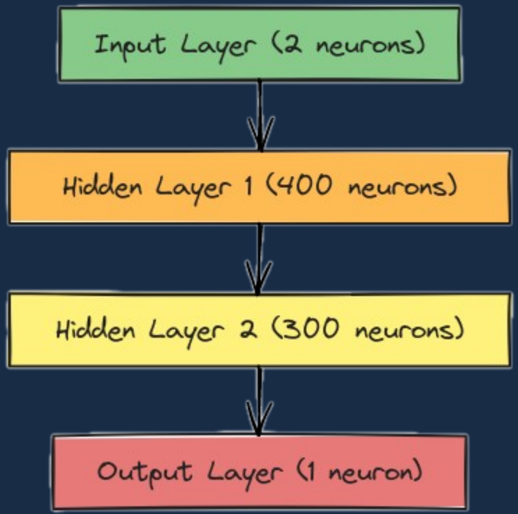


We used Numba to accelerate the loop and we reached 50us of execution time.

Implementation attempt
ML agent

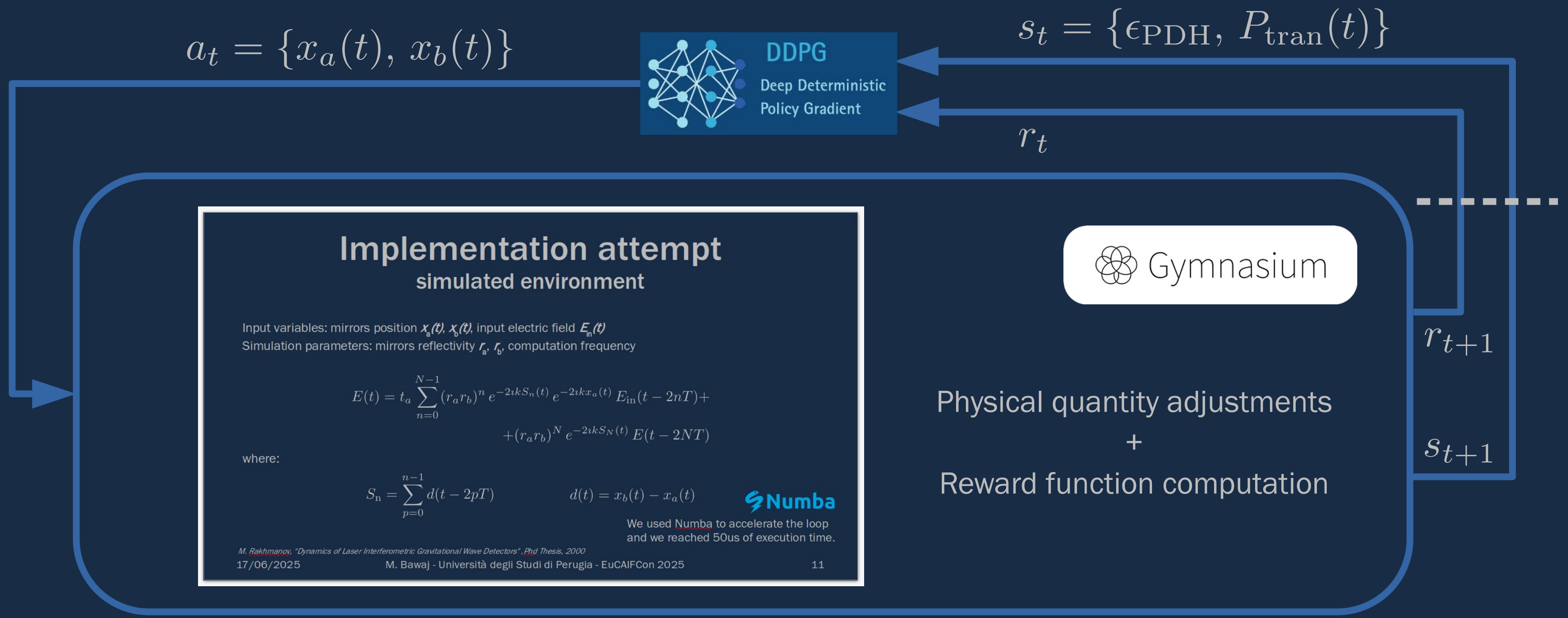
s_t – current state of the environment
a_t – action chosen by the agent
r_t – reward generated by the reward funciton





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

Implementation attempt
simulated environment



Output variables: Pound-Drever-Hall error signal $\varepsilon_{\mathrm{PDH}}(t)$, transmitted electric field power $P_{\mathrm{tran}}(t)$

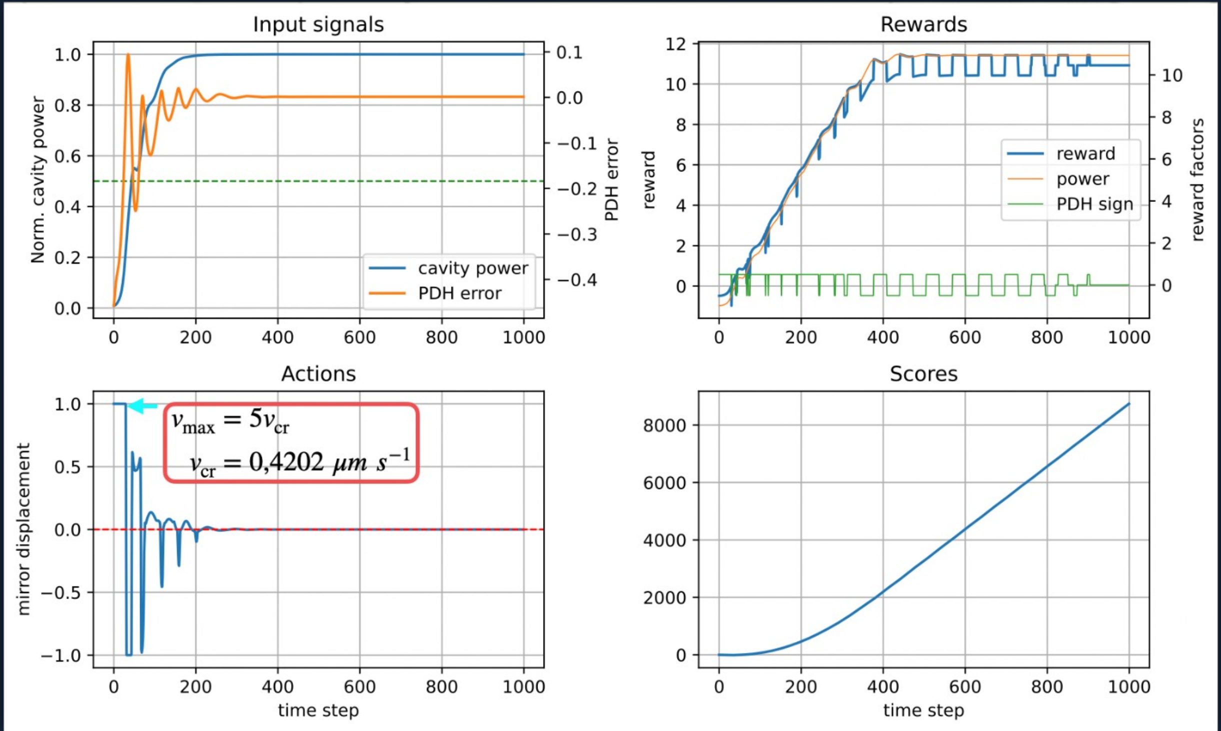
Towers, Mark, et al. 2024. "Gymnasium: A Standard Interface for Reinforcement Learning Environments." arXiv.

Implementation attempt
preliminary results

Test of a DDPG model trained for 120 000 time-steps. Only for the output mirror action.

~200 time-steps are needed to properly lock the cavity.

The maximum speed allowed for the agent to approach the resonance is limited by the environment.

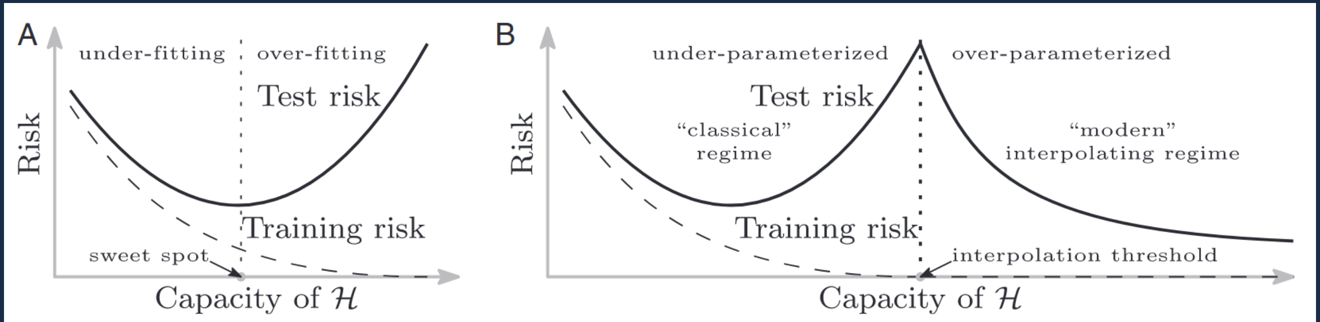


Application

Slow inference in large models may return us to the classical trade-off between complexity and generalization.

Classical risk curve

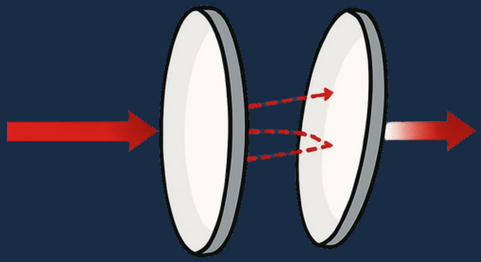
Deep risk curve



Network size	<1k	>1M
Inference time	~0,5–5 ms	~10–30 ms
Computing perf.	1MFLOPS	1GFLOPS
Hardware	CPU	GPU
Precision	float is available	float is expensive

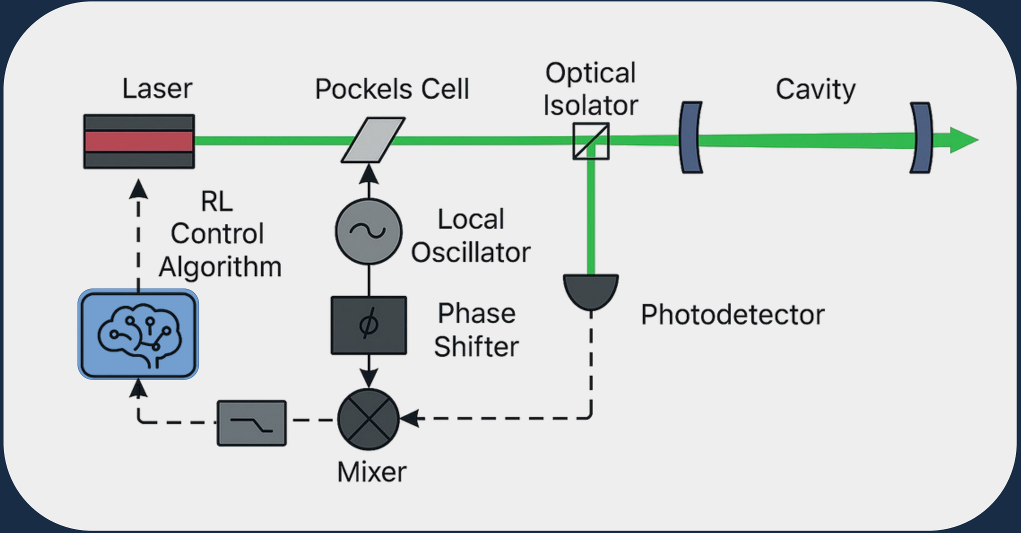
Belkin, M., D. Hsu, S. Ma, and S. Mandal. 2019. "Reconciling Modern Machine-Learning Practice and the Classical Bias–Variance Trade-Off." PNAS 116 (32): 15849–54.

Sim to real transfer



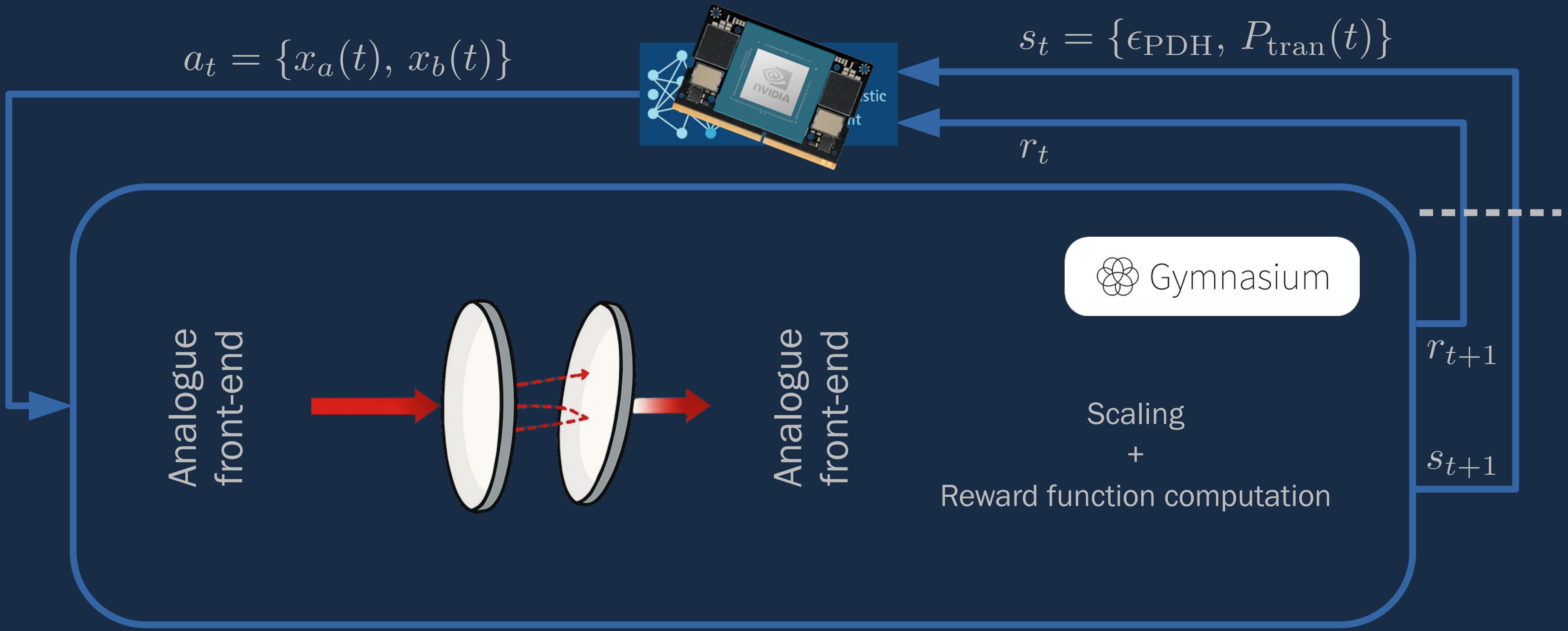
Challenges:

- Sub-optimal sampling due to long inference
- Non-Gaussian noise in the system
- Suspended mirrors
- Angular misalignment
- Long delay in interaction



Zhao, W., et al. "Sim-to-Real Transfer in Deep Reinforcement Learning for Robotics: A Survey." In 2020 IEEE Symposium Series on Computational Intelligence (SSCI), 737–44. Canberra, ACT, Australia: IEEE.

Sim to real transfer



Output variables: Pound-Drever-Hall error signal $\varepsilon_{PDH}(t)$, transmitted electric field power $P_{tran}(t)$

Towers, Mark, et al. 2024. "Gymnasium: A Standard Interface for Reinforcement Learning Environments." arXiv.

Conclusions

- ML/RL has already proved to be successful in several control tasks
- We hope it will also in our!

To widen the field of application:

- we need complete time-domain simulators
- we need low-latency RT hardware with ML acceleration capabilities

Similar attempts

Other works about AI/ML application to GW detector control:

- Sorokin, D., et al. 2021. “Interferobot: Aligning an Optical Interferometer by a Reinforcement Learning Agent.” arXiv.
- Ma, P. X., and G. Vajente. 2023. “A Deep Learning Technique to Control the Non-Linear Dynamics of a Gravitational-Wave Interferometer.” arXiv.
- Mukund, N., et al. 2023. “First Demonstration of Neural Sensing and Control in a Kilometer-Scale Gravitational Wave Observatory.” arXiv.
- Qin, J., et al. 2025. “Automated Alignment of an Optical Cavity Using Machine Learning.” Classical and Quantum Gravity 42 (4): 045003.

- More coming soon…

Thank you for your attention