Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA

Towards DRL agent deployment for lock optimization through HPC training sessions

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

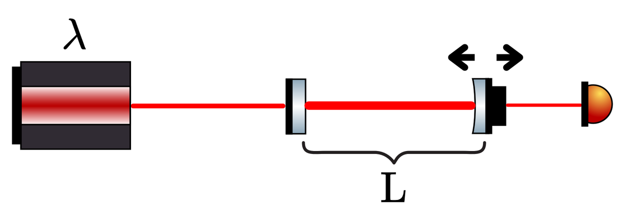
A.D. 1308 unipg UNIVERSITÀ DEGLI STUDI DI PERUGIA · VIRGO · EINSTEIN TELESCOPE · INFN Sez. di Perugia

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

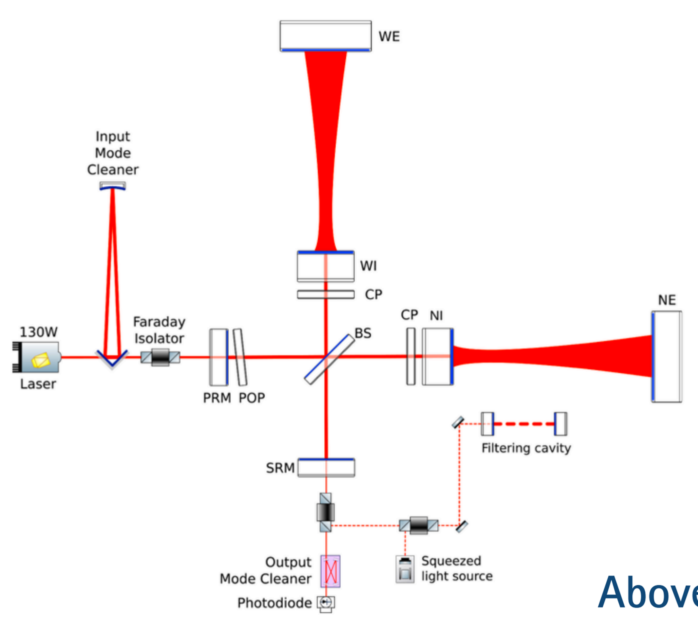
Introduction

Fabry–Perot cavities are critical components in high-sensitivity interferometer as gravitational wave detectors.

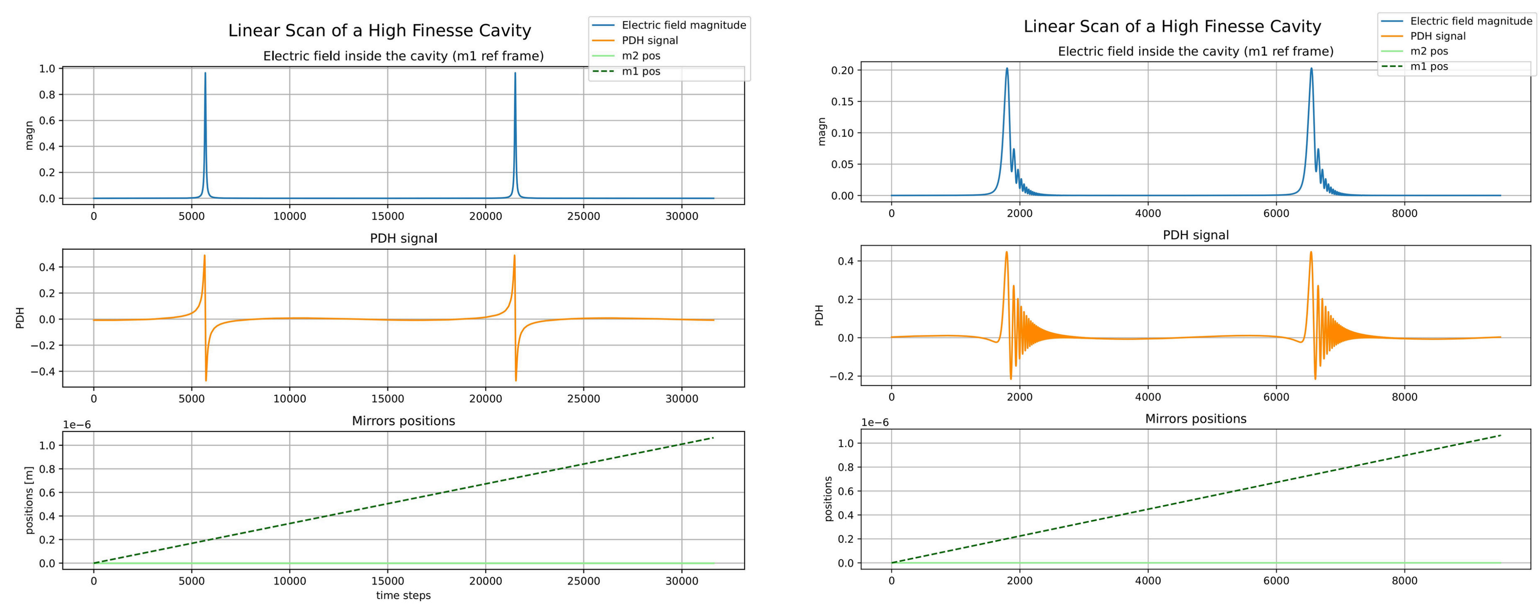
Their operation relies on locking to resonance condition which ensures optical power buildup and guarantees stable interferometric signals.



By changing the length of the cavity with mirror position shifts you can cross resonances where the cavity power is maximised.



$v = 0.01\ v_{cr}$ ← Mirror speed → $v = 10\ v_{cr}$

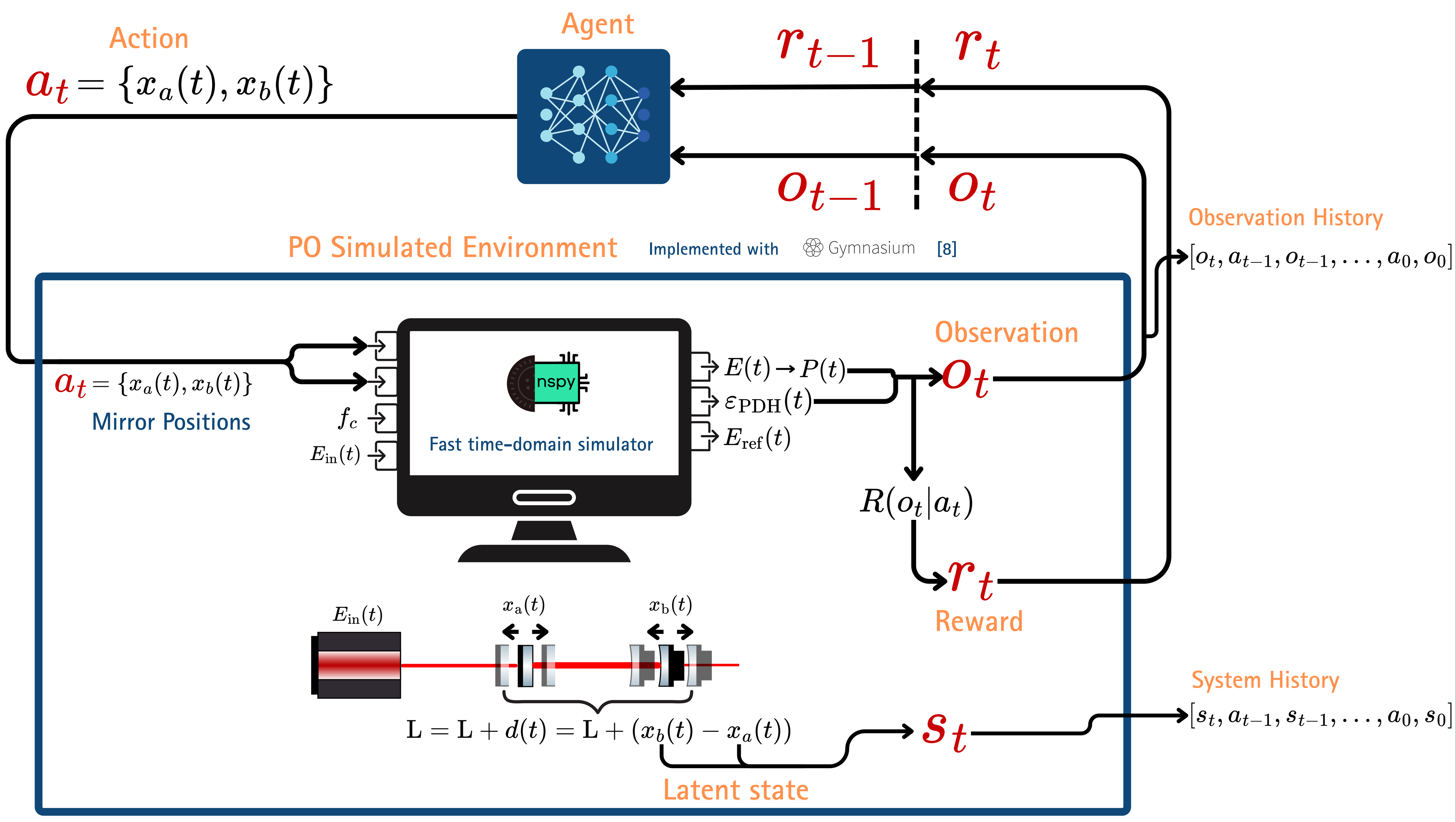


Above a certain threshold non linear effects as ring down arise and make locking procedure harder with classic control. So the idea is using reinforcement learning to manage dynamics.

Reinforcement Learning Framework

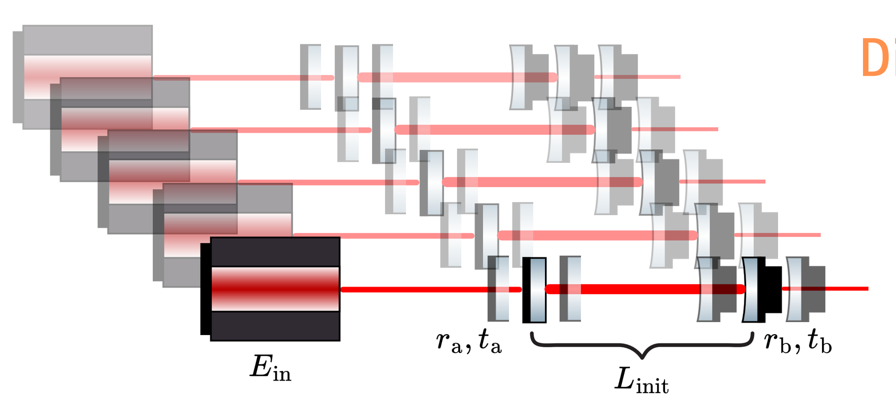
The scheme describes the interaction between the RL agent and a partial observable environment (rigid high finesse cavity). The agent sends actions a_t to the environment, that will change its latent state s_t. The agent will receive the observation o_t as an information together with a scalar value called reward r_t, crucial for the agent to understand the correct behaviour during training in order to reach the task.

Agent is trained firstly on a simulated environment with reality gap as low as possible to facilitate Sim-To-Real transfer, avoiding hardware damage during exploration stage and saving training time.



HPC Training Sessions Plan

Different models
- DDPG
- TD3
- RecurrentPPO



Different cavities
- ARM
- Microcavities
- ...

List of Parameters
$\langle r_a, r_b, t_a, t_b, L_{init}, E_{in}, \dots \rangle$ ← Domain Randomization

LEONARDO CINECA

ADA Cloud

- HPC-driven campaign to train multiple DRL models on a diverse set of simulated optical cavities.
- Each cavity exposes different nonlinear and dynamical regimes, enabling assessment of how model architecture and reward design affect lock acquisition.
- For every model, multiple training runs are performed across cavities and reward functions, with controlled domain randomization to boost robustness and sim-to-real transfer.
- The workflow enables a comparative evaluation and identifies the most resilient model–reward strategies for future deployment in real optical setups.