Control of cryogenic dark matter detectors through deep reinforcement learning

Felix Wagner
Institute of High Energy Physics
of the Austrian Academy of Sciences

HEPHY INSTITUT FÜR HOCHENERGIEPHYSIK

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

1 Introduction

Cryogenic phonon detectors are one of the leading technologies to reach **sensitivity to light dark matter interactions** in direct detection experiments.

They consist of a target crystal equipped with a superconducting thermometer, and a SQUID-based readout circuit. The system requires the **careful optimization** of the heating of the thermometer (DAC) and the bias current in the readout circuit (I_B). The standard approach for this is **time consuming** and requires **manual interventions**.

For future large-scale setups this task needs to be **automated**. We show in a simulation, that this is possible with reinforcement learning.

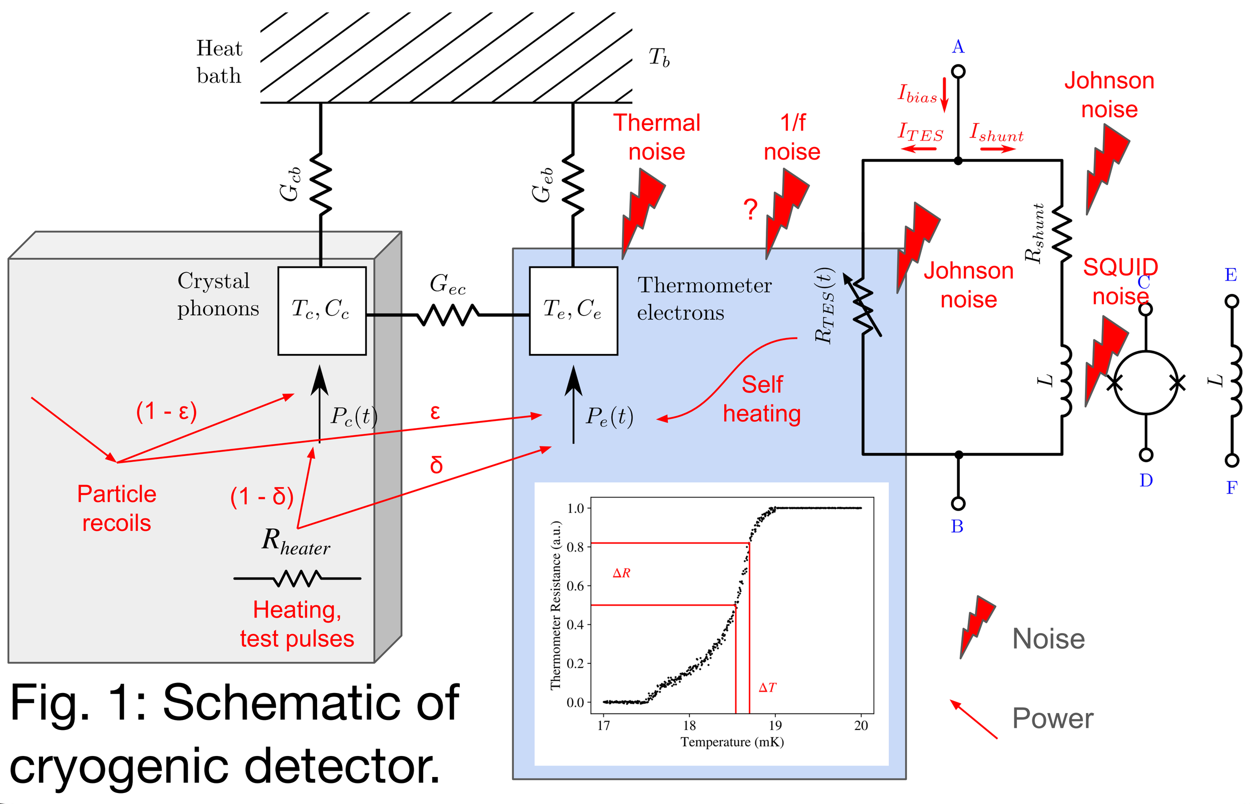

Fig. 1: Schematic of cryogenic detector.

2 Cryogenic detector response simulation

We built a simulation of cryogenic detectors. The system is governed by its **thermal and electronic dynamics** (two ODEs) [1], and noise contributions [2, 3]:

$$\dot{\underline{T}}(t) = \mathrm{diag}(\underline{C})^{-1}\left(\underline{P}\left(t, \underline{T}(t), \underline{I_t}(t)\right) + \mathrm{diag}\left(\underline{G_b}\right)\left(\underline{T_b} - \underline{T}(t)\right) + (\underline{\underline{G}} - \mathrm{diag}(\underline{\underline{G}}1))\underline{T}(t)\right)$$
$$\dot{I}_t(t) = \mathrm{diag}(\underline{L})^{-1}\left(\mathrm{diag}\left(\underline{R_s}\right)\underline{I_b} - \mathrm{diag}\left(\underline{I_t}(t)\right)\left(\underline{R_t}(\underline{T}(t)) + \underline{R_s}\right)\right)$$

Heat capacities C, thermal couplings G, temperatures T, thermometer current I_t and resistances R_t and R_s are not directly observable.

A self heating of the thermometer introduces history-dependency. We can describe the system as a **Partially Observable Markov Decision Process (POMPD)** with only one observable: the **SQUID output.**

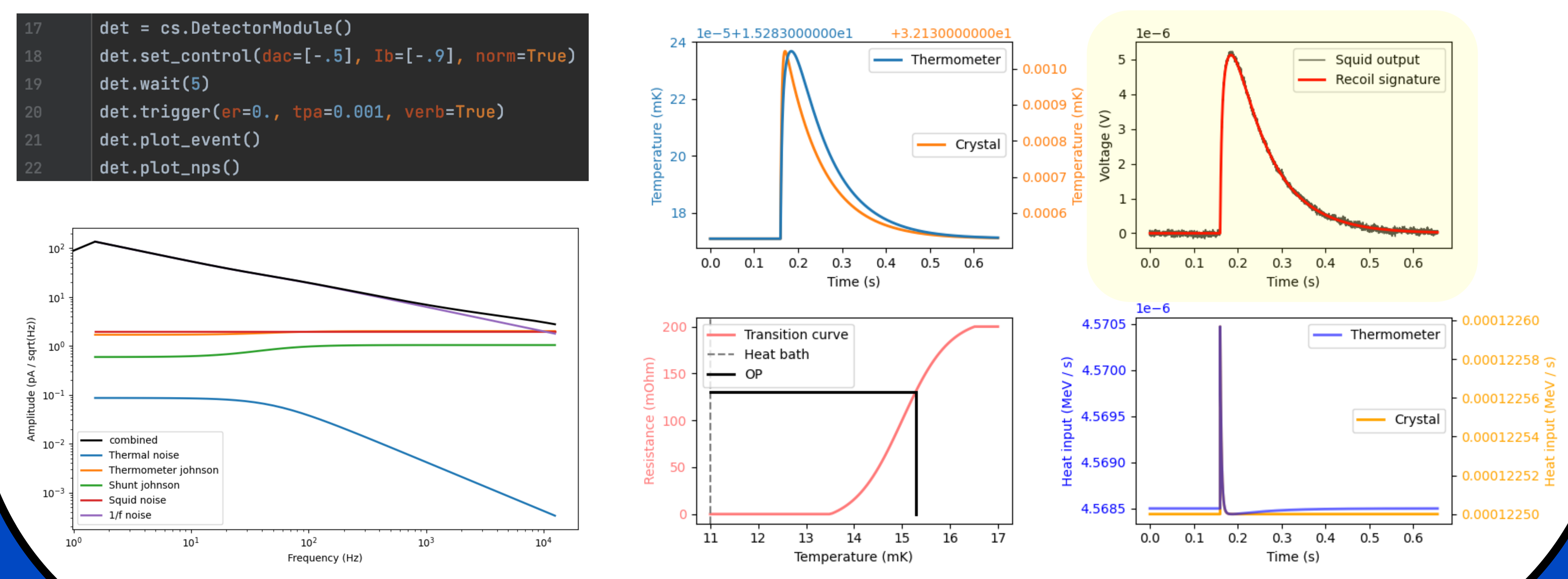

Fig. 2: Rendered output of simulator.

3 Reinforcement learning

The control problem is modelled as the interplay of an **agent** with an **environment**: in each time step the agent performs an **action** and receives a **reward** and new **state**. The agent learns to maximize the rewards over time.

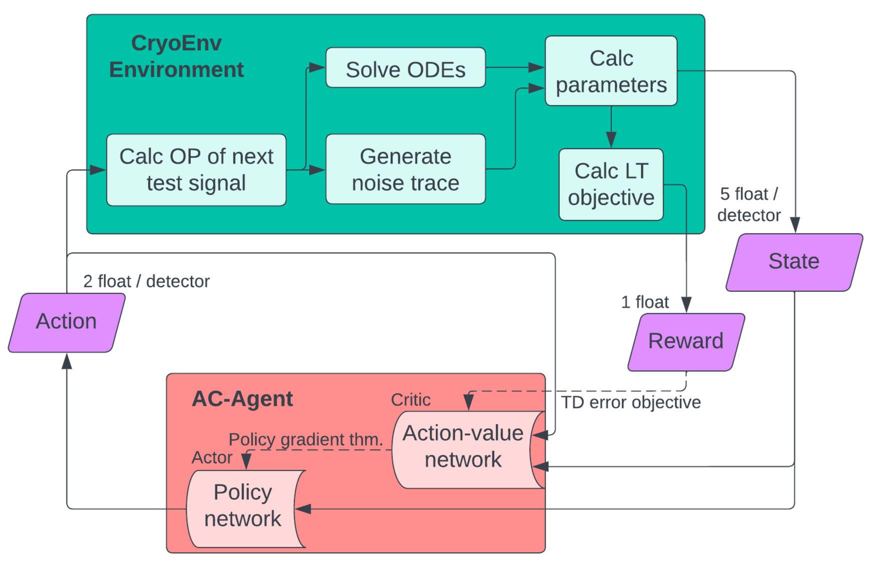

Fig. 3: Actor critic flow diagram.

State: DAC, I_B, TPA, PH, RMS.
Actions: Target DAC and I_B.
Reward: $\max\left(-\frac{\mathrm{TPA}}{\mathrm{PH}} \cdot \mathrm{RMS}\right)$

We use an OpenAI Gym environment and a Stable Baselines 3 Soft Actor Critic (SAC) agent [4, 5, 6]. **Online learning** can be realized with independent control/training threads, and communication through an MQTT feed (WIP).

TPA ... energy of injected signal
PH ... strength of response
RMS ... noise of response

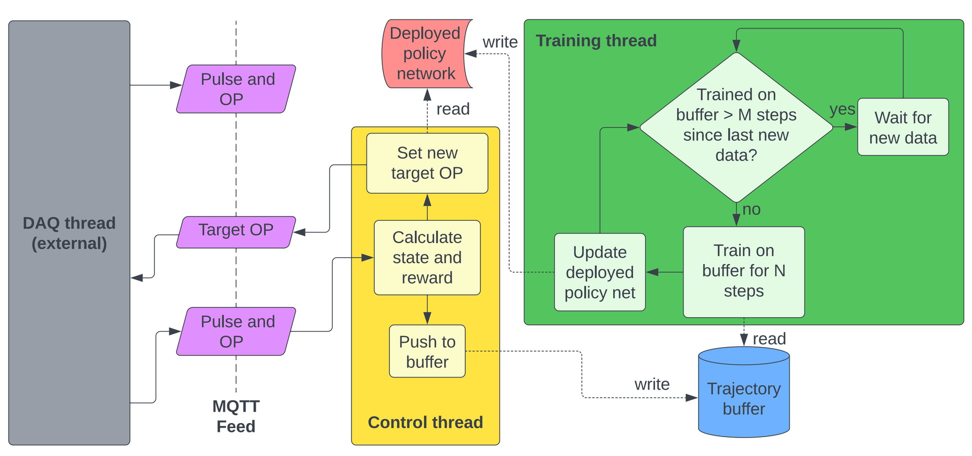

Fig. 4: Online learning flow diagram.

4 Fast and automatic optimization in simulated environment

Standard approach:

Sweep the DAC for several settings of I_B, choose setting with strongest response. Each sweep takes in our simulation **42 min measurement time equivalent**.

Reinforcement learning (ours):

In simulation, training takes **26 min measurement time equivalent** with realistic conditions (pile-ups, etc).

Conclusion

Our approach requires **less measurement time, no manual interventions,** optimizes directy the **sensitivity** and is **scalable** to multi-detector setups.

First runs in a live measurement are planned for the near future - stay tuned!

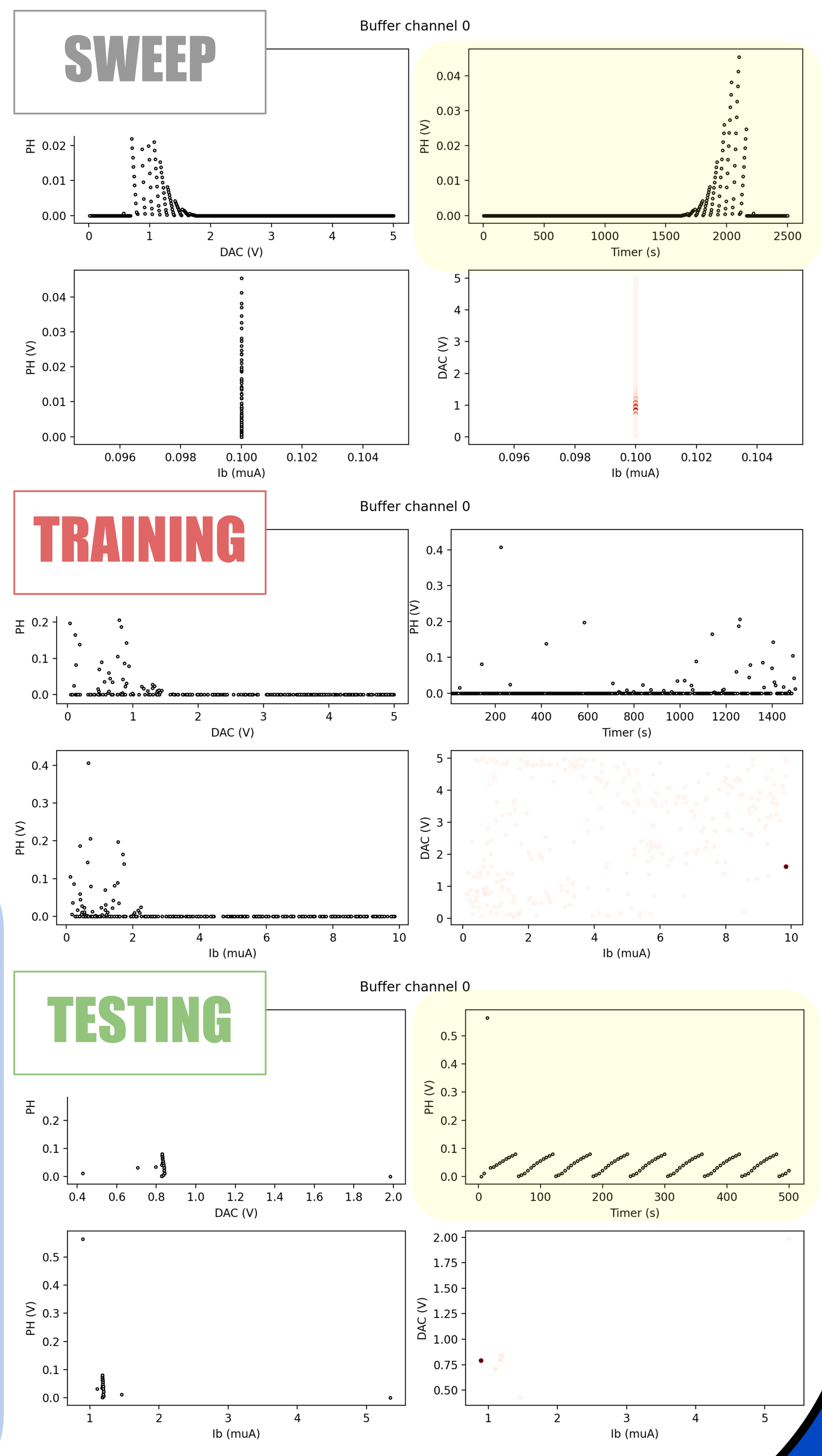

Fig. 5: Trajectory monitors.

References

[1] J Low Temp Phys 100, 69–104 (1995), https://doi.org/10.1007/BF00753837
[2] E. Pantic, PhD Thesis, TU Munich (2008)
[3] Comp Phys Comm 181, 1982-1985 (2010), https://doi.org/10.1016/j.cpc.2010.09.003
[4] arXiv:1606.01540, https://doi.org/10.48550/arXiv.1606.01540
[5] JMLR 22, 1-8 (2021), http://jmlr.org/papers/v22/20-1364.htmlv
[6] PMLR 80, 1861-1870 (2018), https://proceedings.mlr.press/v80/haarnoja18b.html

Acknowledgements

Contributions to detector optimization efforts were made by Daniel Bartolot and Kolos Niedermayer within their master- and bachelor thesis at TU Vienna. We thank many members of the CRESST and COSINUS collaborations for inspiring discussions, especially: Florian Reindl, Franz Pröbst, Vanessa Zema, Johannes Rothe, Stephan Fichtinger, Jochen Schieck; and Wolfgang Waltenberger from the CMS collaboration. We are grateful for the discussions within the Rare Event Search group at HEPHY Vienna and Clemens Heitzinger's group at TU Vienna. This work was supported by the Austrian research promotion agency (FFG), project ML4CPD. The computational results presented were obtained using the Vienna CLIP cluster.

FFG Forschung wirkt.