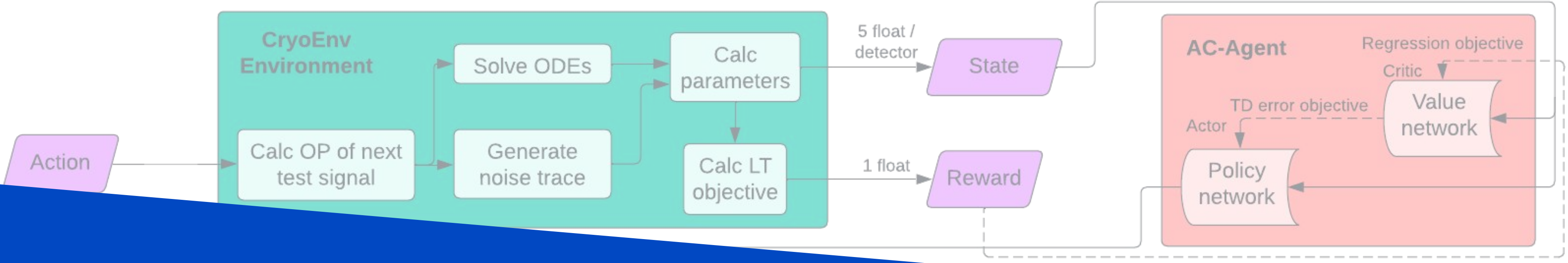


Control of cryogenic dark matter detectors through deep reinforcement learning

Felix Wagner

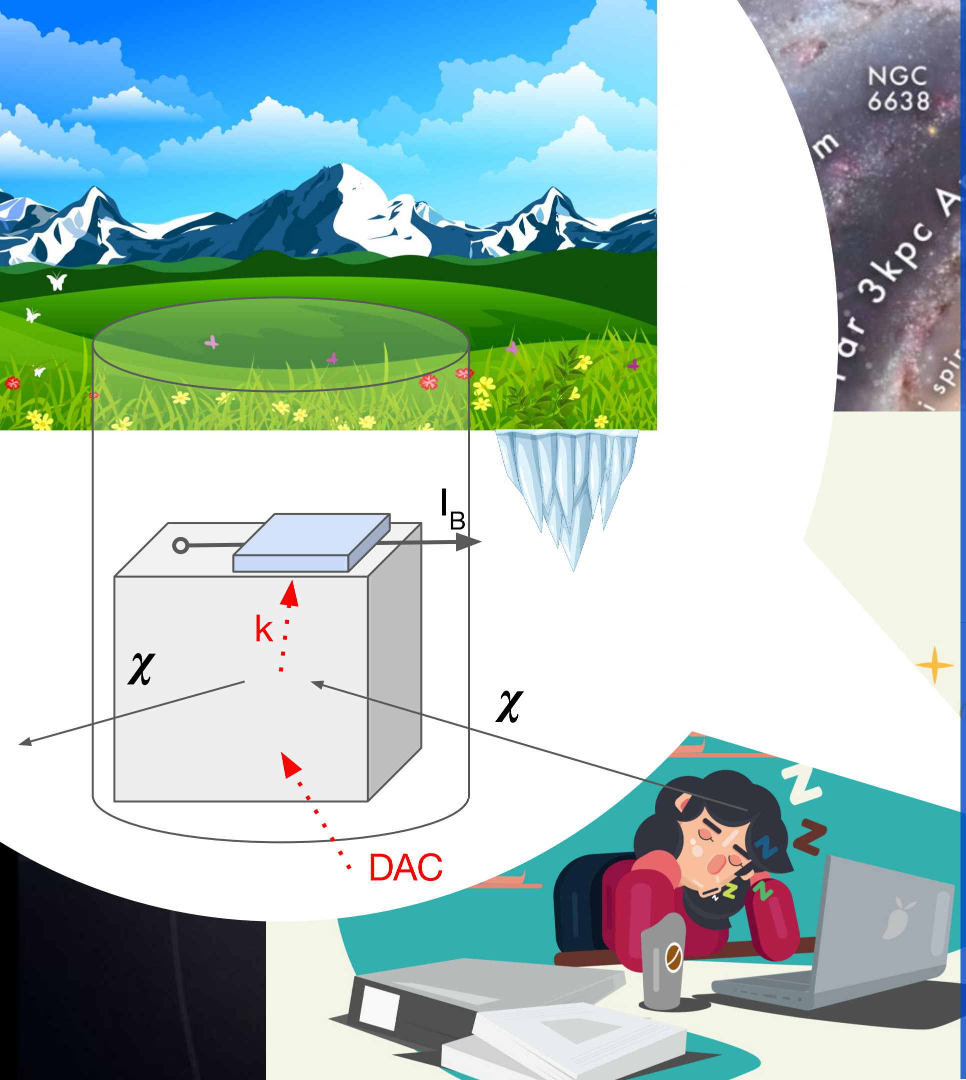
Institute of High Energy Physics of the Austrian Academy of Sciences

21st ACAT, 27th October 2022

ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

HEPHY INSTITUT FÜR HOCHENERGIEPHYSIK

FFG Forschung wirkt.



Direct detection of light DM-nucleus scattering

- earth based detector in underground lab,
- monocristalline target,
- superconducting thermometer,
- operation at O(mK) temperature;
- DM recoils produce phonons, thresholds down to O(10 eV).
- Careful optimization of the **heating (DAC)** and the **bias current (I_B)** of the thermometer are required.
- Standard approach is **time consuming** and requires **manual interventions.**

Direct detection of light DM-nucleus scattering

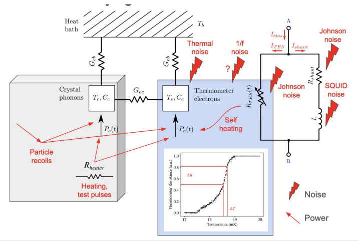
- earth based detector in underground lab,
- monocristalline target,
- superconducting thermometer,
- operation at O(mK) temperature;
- DM recoils produce phonons, thresholds down to O(10 eV).
- Careful optimization of the **heating (DAC)** and the **bias current (I_B)** of the thermometer are required.
- Standard approach is **time consuming** and requires **manual interventions.**
- For future large-scale setups this task needs to be **automated.**

A framework for policy optimization:

This is called a Markov decision process!

function that maps state to probability distribution over actions

Environment



defines the transitions between states according to internal dynamics

Action

Observation *and* ***Reward***

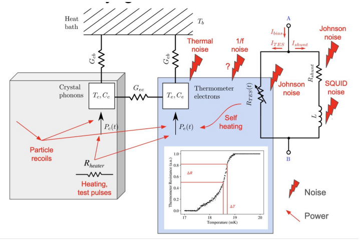
Maximizing rewards over time (returns) is what we want!

receives state, takes action according to policy

Agent

A framework for policy optimization: *hire a Phd student*

Environment



Action

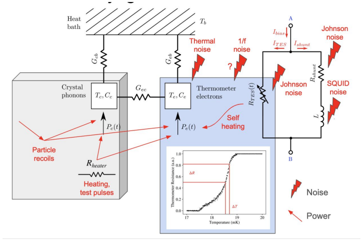
Observation* and *Reward

I just want to get high rewards …

Agent

A framework for policy optimization: *reinforcement learning*

Environment



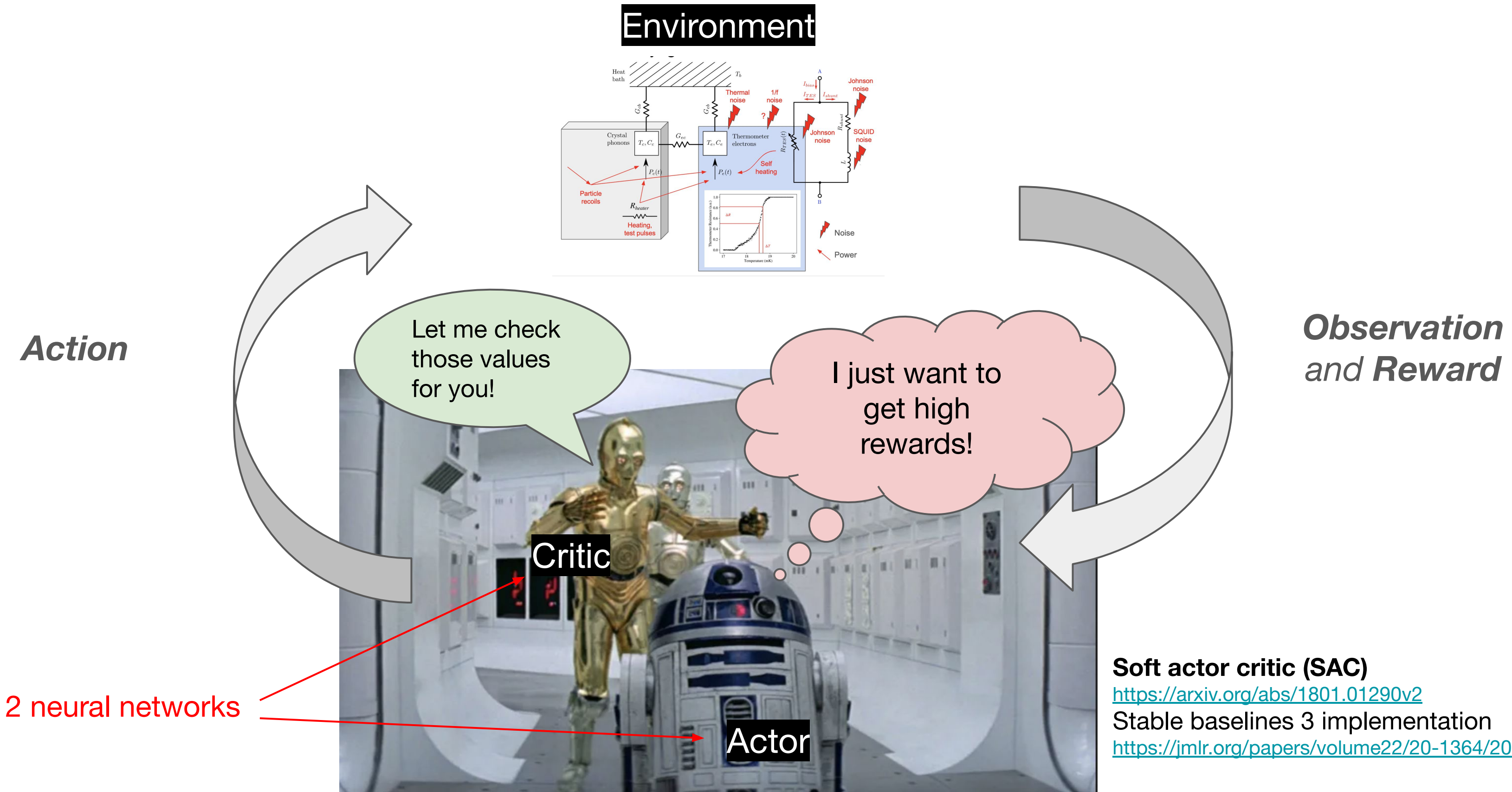
Action

Observation
*and **Reward***

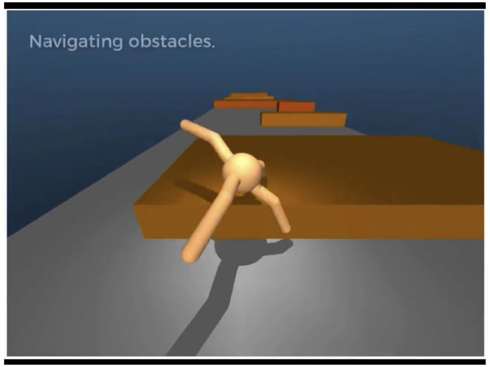
I just want to
get high
rewards!

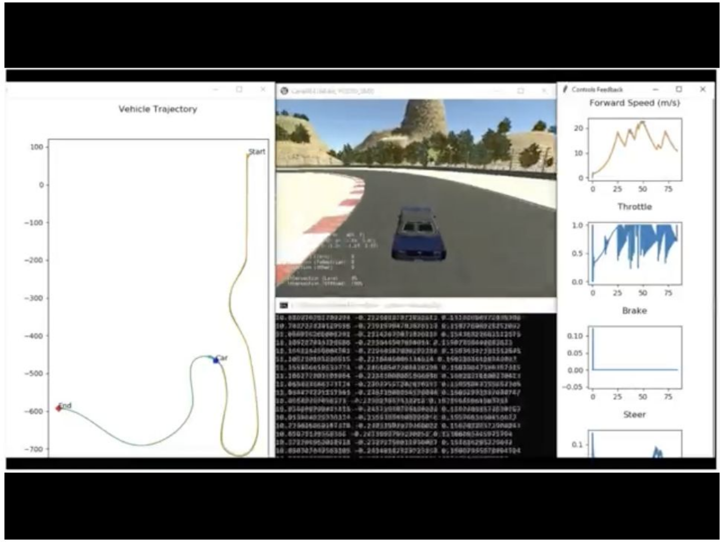
Agent

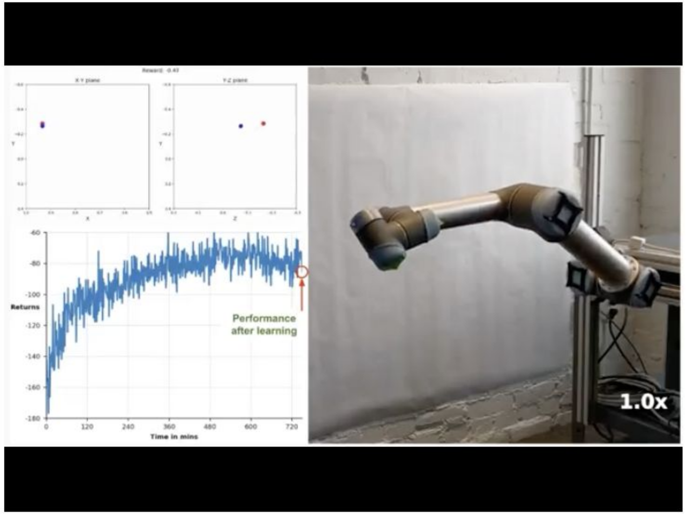
A framework for policy optimization: *reinforcement learning*

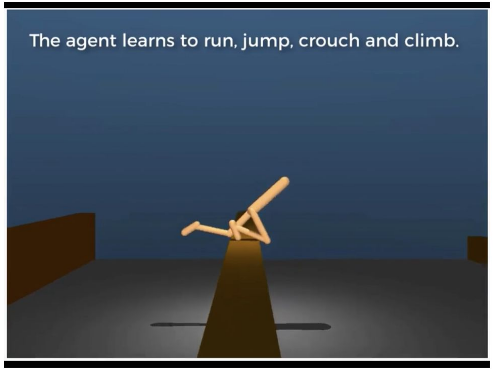


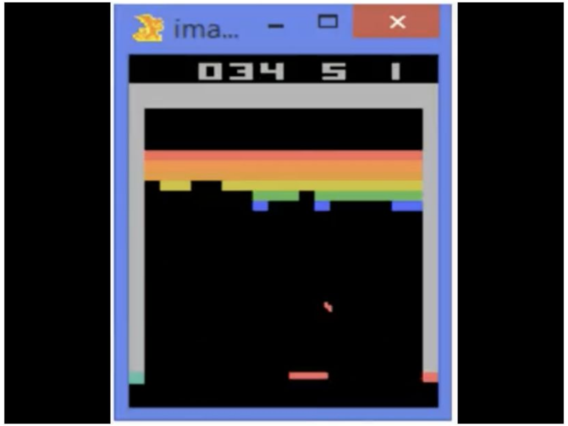
Soft actor critic (SAC)
https://arxiv.org/abs/1801.01290v2
Stable baselines 3 implementation
https://jmlr.org/papers/volume22/20-1364/20-1364.pdf











OpenAI Gym - A framework for reinforcement learning
https://www.gymlibrary.dev/

Examples of reinforcement learning in physics

Article

Magnetic control of tokamak plasmas through deep reinforcement learning

https://doi.org/10.1038/s41586-021-04301-9
Received: 14 July 2021
Accepted: 1 December 2021
Published online: 16 February 2022
Open access
Check for updates

Jonas Degrave[1,3], Federico Felici[2,3]✉, Jonas Buchli[1,3]✉, Michael Neunert[1,3], Brendan Tracey[1,3]✉, Francesco Carpanese[1,2,3], Timo Ewalds[1,3], Roland Hafner[1,3], Abbas Abdolmaleki[1], Diego de las Casas[1], Craig Donner[1], Leslie Fritz[1], Cristian Galperti[2], Andrea Huber[1], James Keeling[1], Maria Tsimpoukelli[1], Jackie Kay[1], Antoine Merle[2], Jean-Marc Moret[2], Seb Noury[1], Federico Pesamosca[2], David Pfau[1], Olivier Sauter[2], Cristian Sommariva[2], Stefano Coda[2], Basil Duval[2], Ambrogio Fasoli[2], Pushmeet Kohli[1], Koray Kavukcuoglu[1], Demis Hassabis[1] & Martin Riedmiller[1,3]

Nuclear fusion using magnetic confinement, in particular in the tokamak configuration, is a promising path towards sustainable energy. A core challenge is to shape and maintain a high-temperature plasma within the tokamak vessel. This requires high-dimensional, high-frequency, closed-loop control using magnetic actuator coils, further complicated by the diverse requirements across a wide range of plasma configurations. In this work, we introduce a previously undescribed architecture for tokamak magnetic controller design that autonomously learns to command the full set of control coils. This architecture meets control objectives specified at a high level, at the same time satisfying physical and operational constraints. This approach has unprecedented flexibility and generality in problem specification and yields a notable reduction in design effort to produce new plasma configurations. We successfully produce and control a diverse set of plasma configurations on the Tokamak à Configuration Variable[1,2], including elongated, conventional shapes, as well as advanced configurations, such as negative triangularity and 'snowflake' configurations. Our approach achieves accurate tracking of the location, current and shape for these configurations. We also demonstrate sustained 'droplets' on TCV, in which two separate plasmas are maintained simultaneously within the vessel. This represents a notable advance for tokamak feedback control, showing the potential of reinforcement learning to accelerate research in the fusion domain, and is one of the most challenging real-world systems to which reinforcement learning has been applied.

Nature **602**, 414–419 (2022).
https://doi.org/10.1038/s41586-021-04301-9

PHYSICAL REVIEW ACCELERATORS AND BEAMS **24**, 104601 (2021)

Real-time artificial intelligence for accelerator control: A study at the Fermilab Booster

Jason St. John,* Christian Herwig, Diana Kafkes, Jovan Mitrevski, William A. Pellico, Gabriel N. Perdue, Andres Quintero-Parra, Brian A. Schupbach, Kiyomi Seiya, and Nhan Tran
Fermi National Accelerator Laboratory, Batavia, Illinois 60510, USA

Malachi Schram
Thomas Jefferson National Accelerator Laboratory, Newport News, Virginia 23606, USA

Javier M. Duarte
University of California San Diego, La Jolla, California 92093, USA

Yunzhi Huang
Pacific Northwest National Laboratory, Richland, Washington 99352, USA

Rachael Keller
Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York 10027, USA

(Received 5 January 2021; accepted 16 August 2021; published 18 October 2021)

Phys. Rev. Accel. Beams **24**, 104601 - 104618 (2021).
https://doi.org/10.1103/PhysRevAccelBeams.24.104601

A control problem with few parameters

Why reinforcement learning?

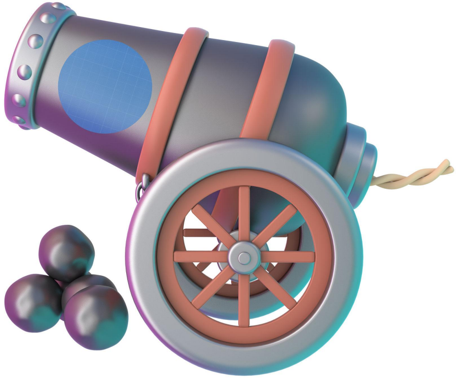
Isn't that like "shooting at sparrows with cannons"?

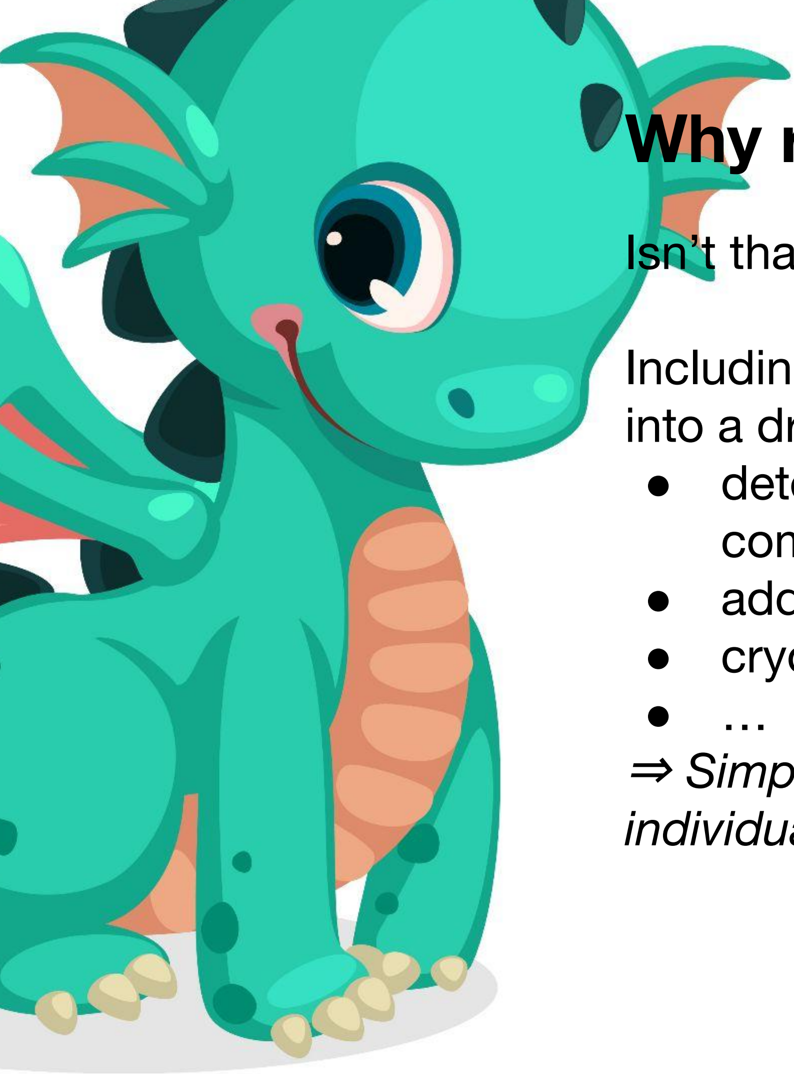
The detector optimization problem is …

- non-linear,
- time-dependent,
- naturally discretized, and
- most parameters are hidden.

⇒ *Simple approaches would need additional constraints!*

Reinforcement learning





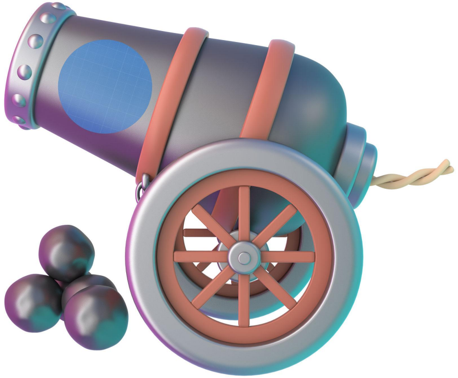
Why reinforcement learning?

Isn't that like "shooting at sparrows with cannons"?

Including more control parameters might "turn the sparrow into a dragon":

- detector concepts with many thermally coupled components,
- additional optimization of magnetic field,
- cryostat parameters,
- …

⇒ *Simple approaches would need individual adaptations!*

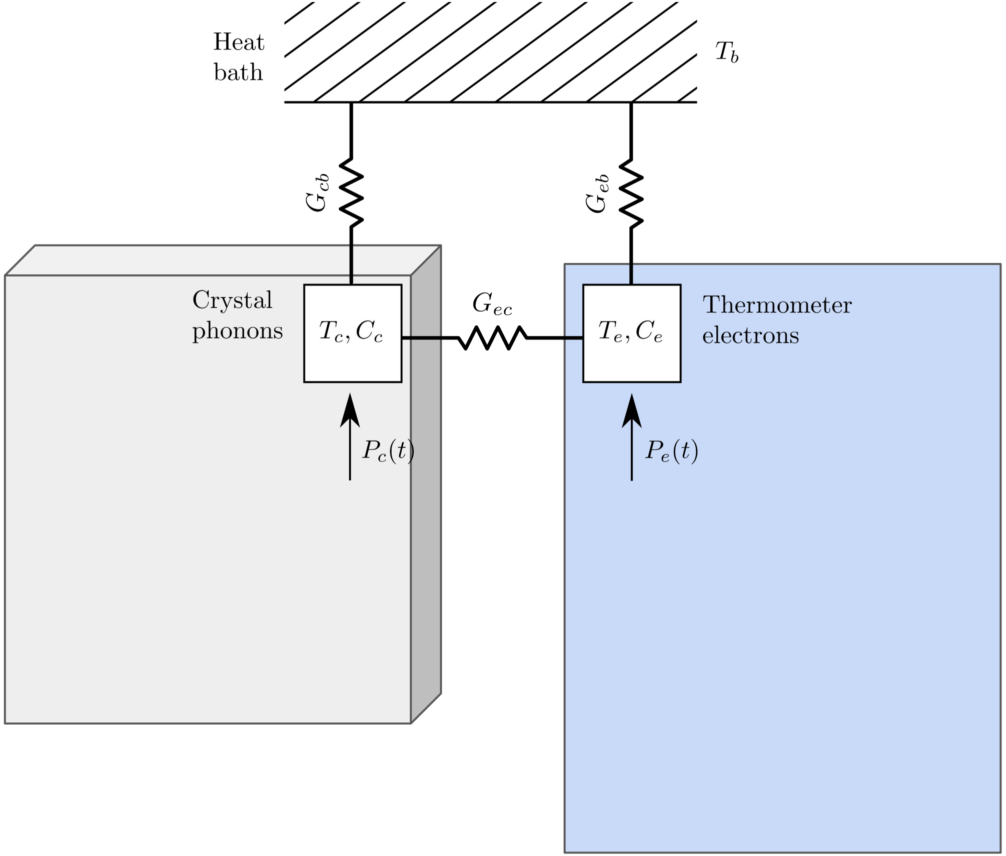


CryoEnv - An OpenAI Gym environment for cryogenic detector optimization

CryoEnv
Code: https://github.com/fewagner/CryoEnv

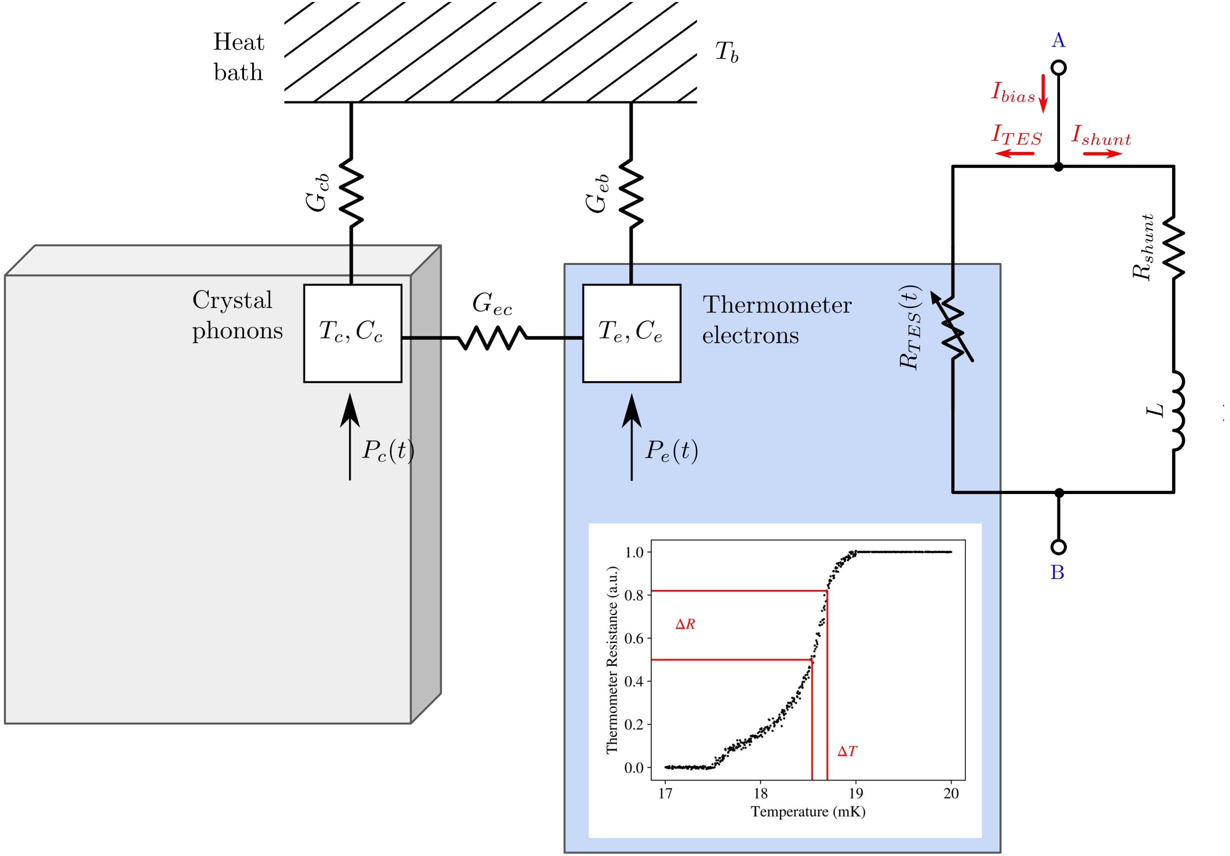
Components of a cryogenic detector

Temperature evolution (T_c, T_e) is dominated by heat capacities (C_c, C_e) of crystal phonons and thermometer electrons and thermal coupling G_{ec} in between and to heat bath (G_{cb}, G_{eb}).



Components of a cryogenic detector

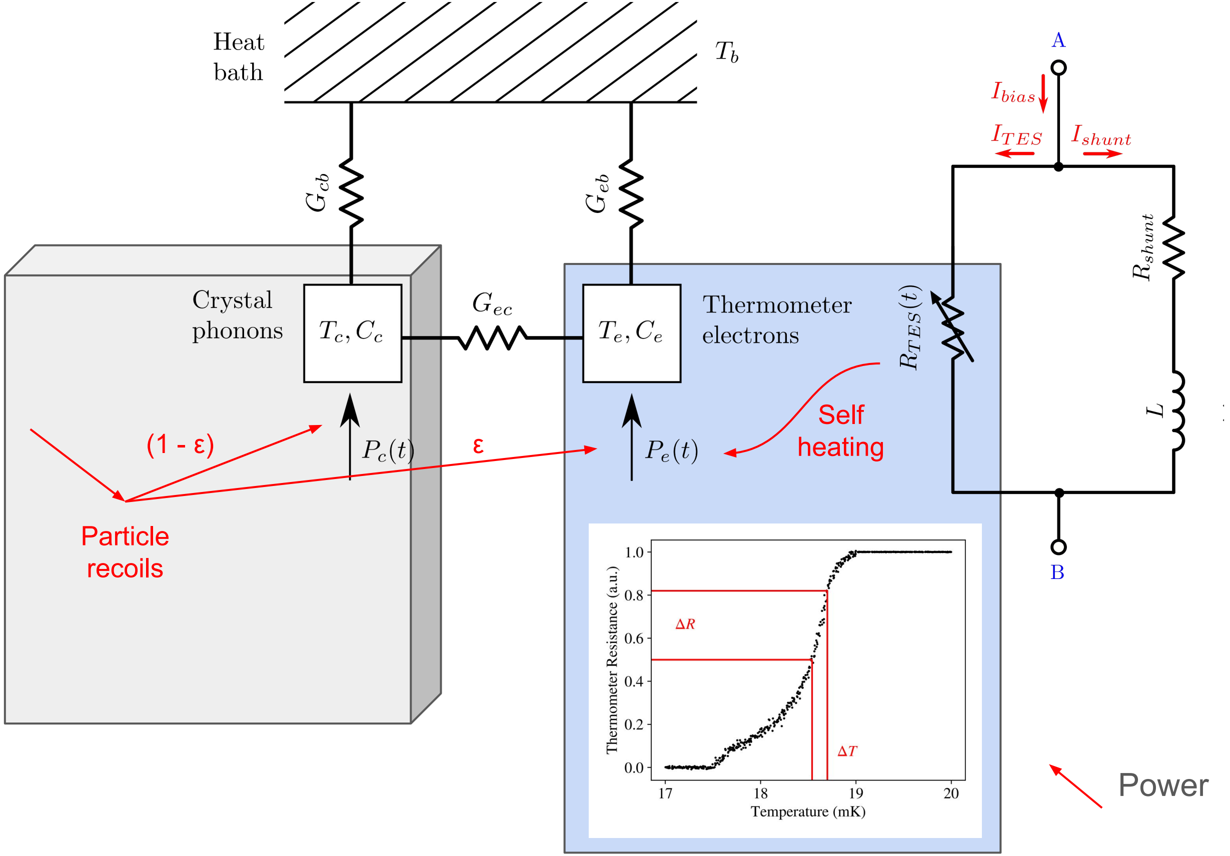
Currents (I$_{TES}$, I$_{shunt}$) in read out circuit are coupled to temperatures through the resistance of the superconducting thermometer R$_{TES}$.



Components of a cryogenic detector

Particle recoils produce power inputs in the crystal (∝ 1 - ε) and thermometer (∝ ε).

Self heating of the thermometer introduces time-dependency.

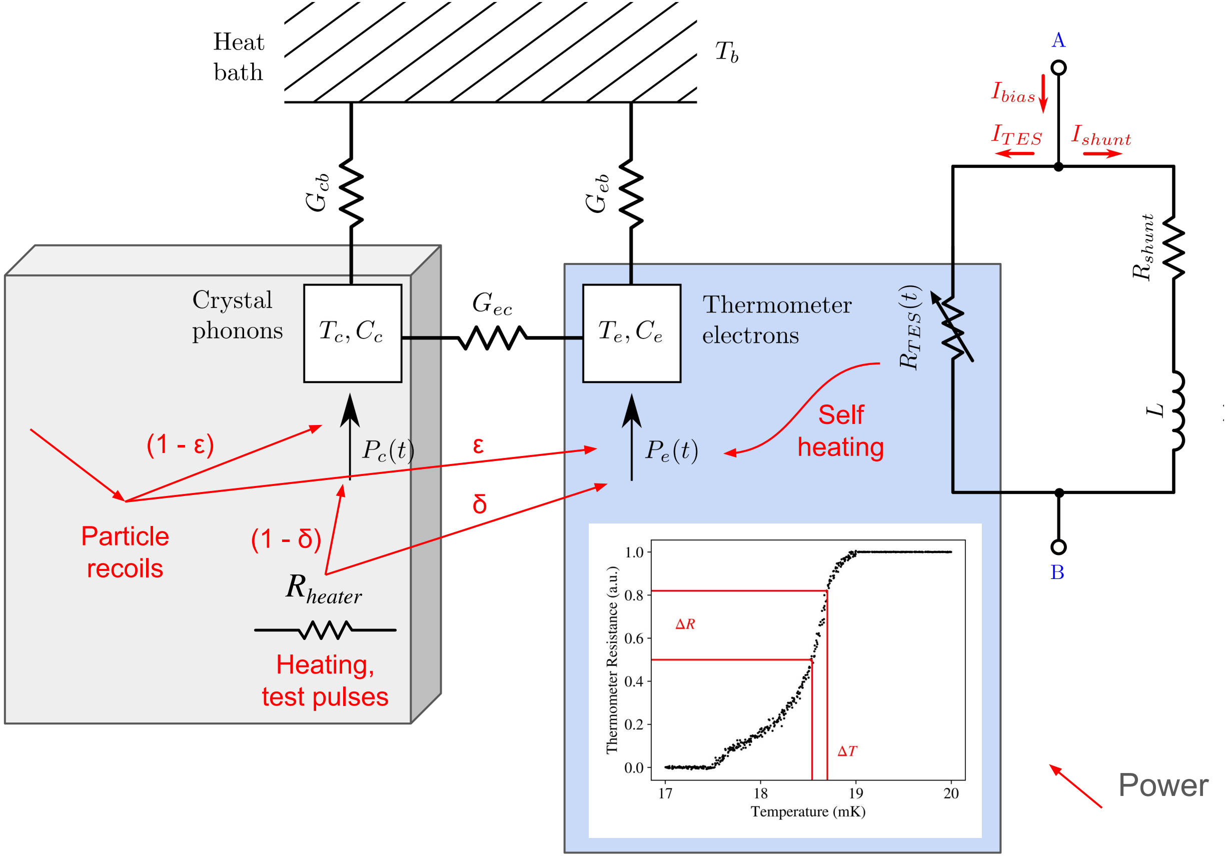


Components of a cryogenic detector

Particle recoils produce power inputs in the crystal (∝ 1 - ε) and thermometer (∝ ε).

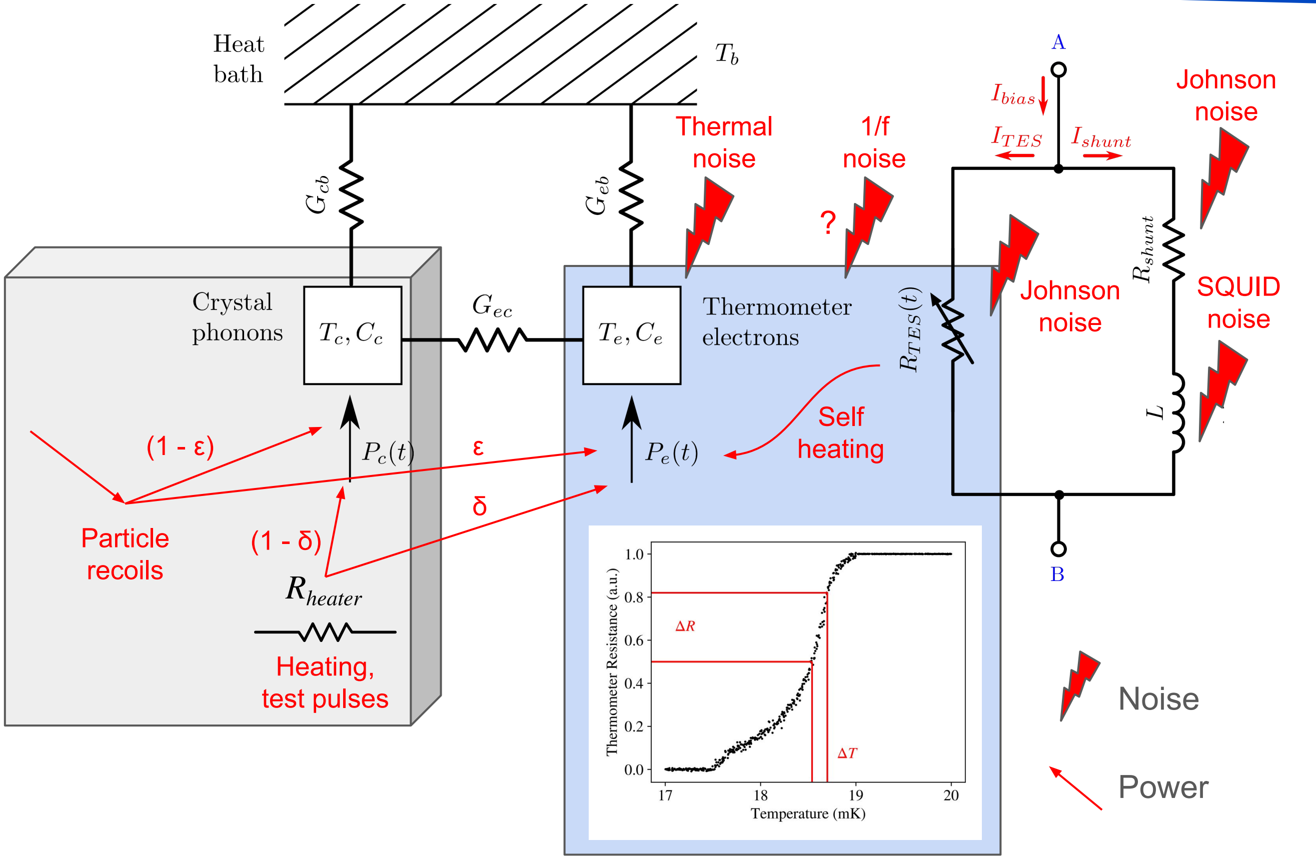
Self heating of the thermometer introduces time-dependency.

A heating resistor in the crystal controls heating and test signals (test pulses).

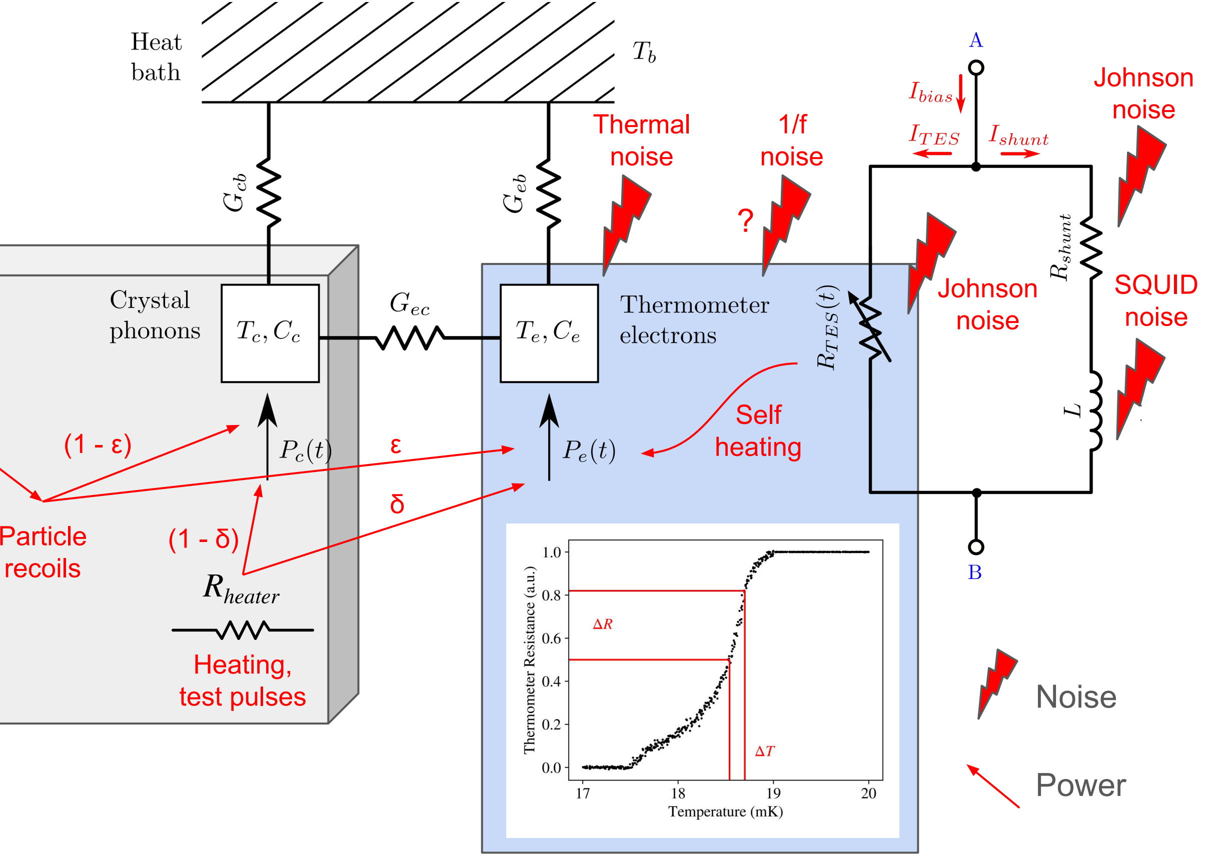


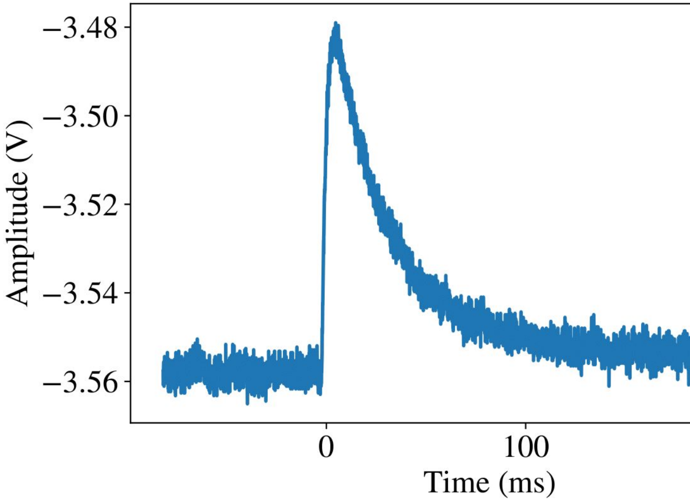
Components of a cryogenic detector

Various noise contributions limit the sensitivity of the detector to low energy recoils.



Components of a cryogenic detector





The digitized signal of a particle recoil is a **pulse-shaped voltage trace.**

Internal parameters are **not observable**!

The cryo ODEs

Cryogenics $$\dot{\underline{T}}(t) = \mathrm{diag}(\underline{C})^{-1}\left(\underline{P}(t, \underline{T}(t), \underline{I_t}(t)) + \mathrm{diag}(\underline{G_b})\left(\underline{T_b} - \underline{T}(t)\right) + \left(\underline{\underline{G}} - \mathrm{diag}(\underline{\underline{G}}\underline{1})\right)\underline{T}(t)\right)$$

Electronics $$\dot{\underline{I}}_t(t) = \mathrm{diag}(\underline{L})^{-1}\left(\mathrm{diag}(\underline{R_s})\underline{I_b} - \mathrm{diag}(\underline{I_t}(t))\left(\underline{R_t}(\underline{T}(t)) + \underline{R_s}\right)\right)$$

System of coupled Ordinary Differential Equations (ODEs) for arbitrary number of thermal components and Transitions Edge Sensors (TESs).

Solve numerically, with interpolation of arbitrary numerical function for TES resistance.

The noise contributions (good summary in E. Pantic, 2008)

advanced

Defs.

$$\Delta I_s(\omega) = \underbrace{\frac{1}{G_{eb}+G_{ETF}} \frac{1}{1+\imath\omega\tau_{eff}} \frac{I_f}{R_f+R_s} \frac{dR_f}{dT_e}}_{S(\omega)} P_e(\omega)$$

$$\tau_{in} \approx C_e/G_{eb}$$

$$\tau_{eff} = \frac{\tau_{in}}{1+\frac{I_f{}^2}{G_{eb}}\frac{dR_f}{dT_e}\frac{R_f-R_s}{R_f+R_s}}$$

$$I_f^2\frac{dR_f}{dT_e}\frac{R_f-R_s}{R_f+R_s} \equiv G_{ETF}$$

$$P_{th}^2 = 4k_B T_e{}^2 G_{eb} \frac{2}{5} \frac{1-\left(\frac{T_{\text{bath}}}{T_e}\right)^5}{1-\left(\frac{T_{bath}}{T_e}\right)^2} \quad \left[\frac{W^2}{Hz}\right] \quad (T_e > T_{\text{bath}})$$

$$\left\langle |\Delta I_{th}(\omega)|^2 \right\rangle = \left\langle |S(\omega)P_{th}(\omega)|^2 \right\rangle$$

Thermal

SQUID

$$\boxed{\left\langle |\Delta I_{sq}(\omega)|^2 \right\rangle = i_{sq}{}^2}$$

Johnson TES

$$\left\langle |\Delta I_{J\_f}(\omega)|^2 \right\rangle = \frac{4k_B T_e R_f}{(R_f+R_s)^2}\left(\frac{\tau_{eff}}{\tau_{in}}\right)^2 \frac{1+\omega^2\tau_{in}{}^2}{1+\omega^2\tau_{eff}^2}$$

$$\left\langle |\Delta I_{J\_s}(\omega)|^2 \right\rangle = \frac{4k_B T_s R_s}{(R_f+R_s)^2}\left(\frac{\tau_{eff}}{\tau_{in}}\right)^2 \frac{\left(1-\frac{I_f^2}{G_{eb}}\frac{dR_f}{dT_e}\right)^2+\omega^2\tau_{in}{}^2}{1+\omega^2\tau_{eff}{}^2}$$

Johnson Shunt

$$\left(\frac{\Delta R_f}{R_f}\right)_{1/f} \propto \frac{1}{\sqrt{\omega}}\left[\frac{1}{\sqrt{Hz}}\right]$$

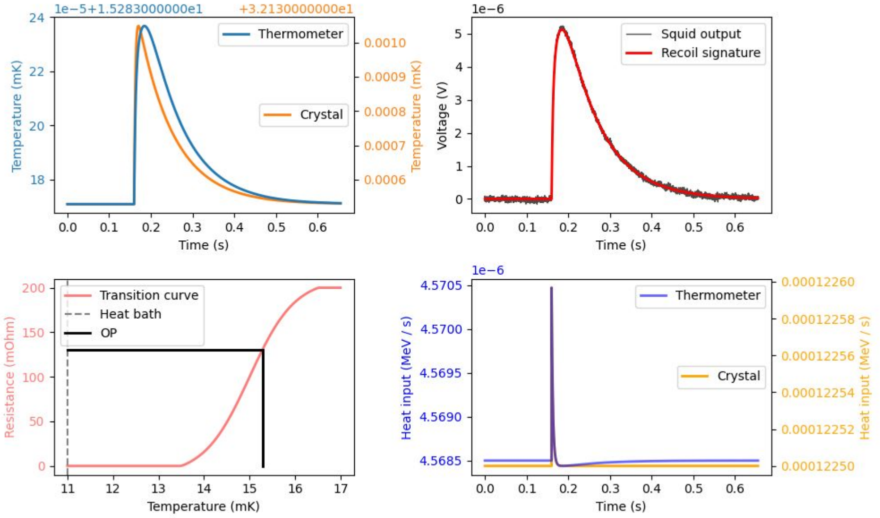
1/f

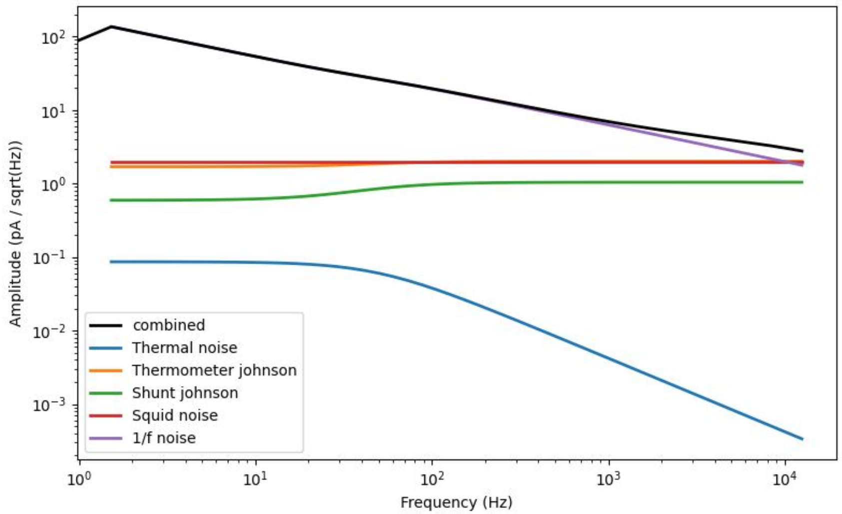
$$\left\langle |\Delta I_{1/f}(\omega)|^2 \right\rangle = \frac{I_f{}^2 R_f{}^2 \left(\frac{\Delta R_f}{R_f}\right)^2_{1/f}}{(R_f+R_s)^2}\left(\frac{\tau_{eff}}{\tau_{in}}\right)^2 \frac{1+\omega^2\tau_{\text{in}}{}^2}{1+\omega^2\tau_{eff}{}^2}$$

Let’s see it in action!

Triggered events can be rendered, ...

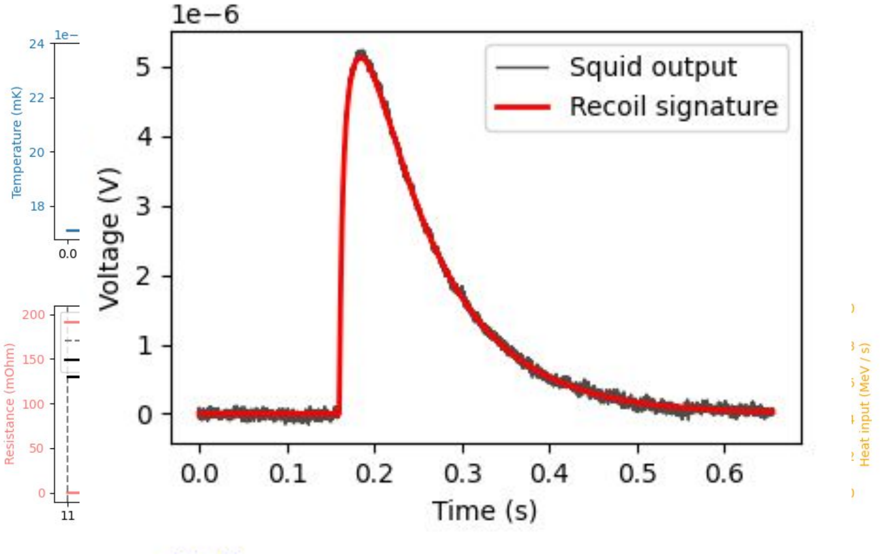
```python
det = cs.DetectorModule()
det.set_control(dac=[-.5], Ib=[-.9], norm=True)
det.wait(5)
det.trigger(er=0., tpa=0.001, verb=True)
det.plot_event()
det.plot_nps()
```

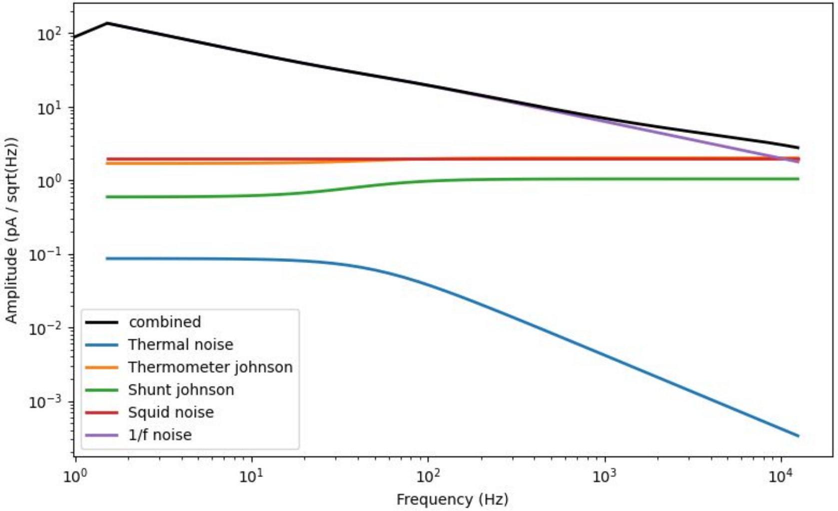




Triggered events can be rendered, …

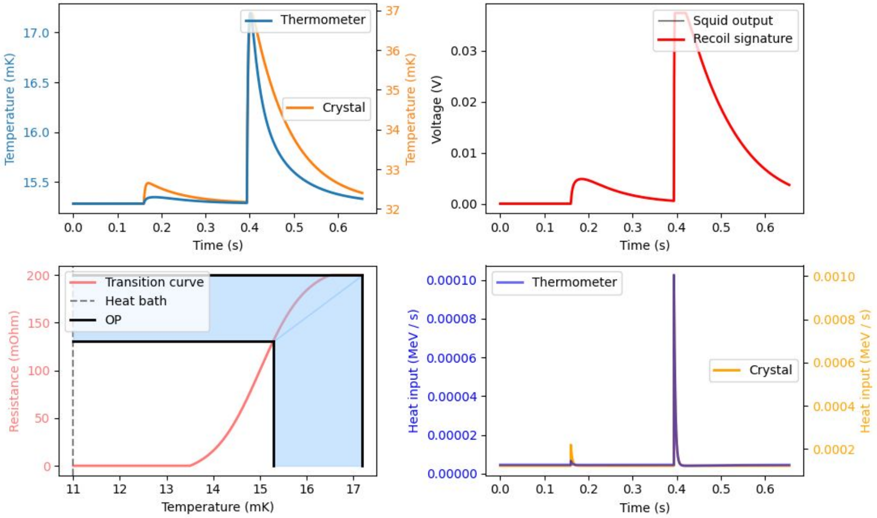
```
det = cs.DetectorModule()
det.set_control(dac=[-.5], Ib=[-.9], norm=True)
det.wait(5)
det.trigger(er=0., tpa=0.001, verb=True)
det.plot_event()
det.plot_nps()
```

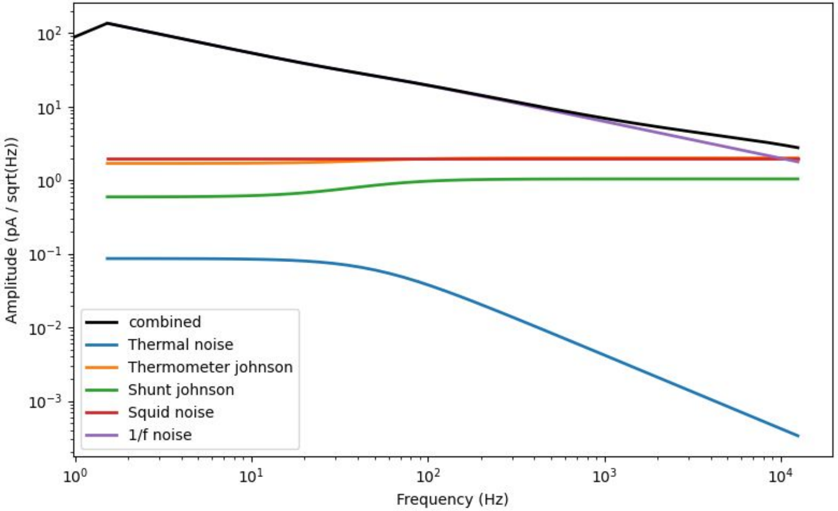




… whoops, pile-up, …

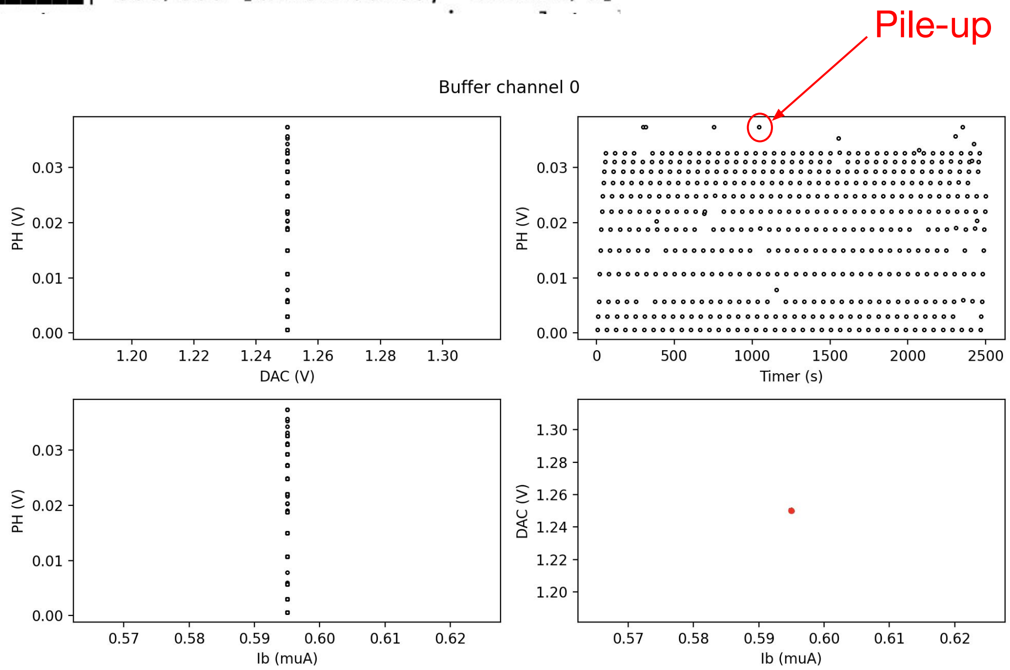
```python
det = cs.DetectorModule(pileup_prob=.99)
det.set_control(dac=[-.5], Ib=[-.9], norm=True)
det.wait(5)
det.trigger(er=0., tpa=1, verb=True)
det.plot_event()
det.plot_nps()
```





… and the pulse height (PH) over time is monitored.

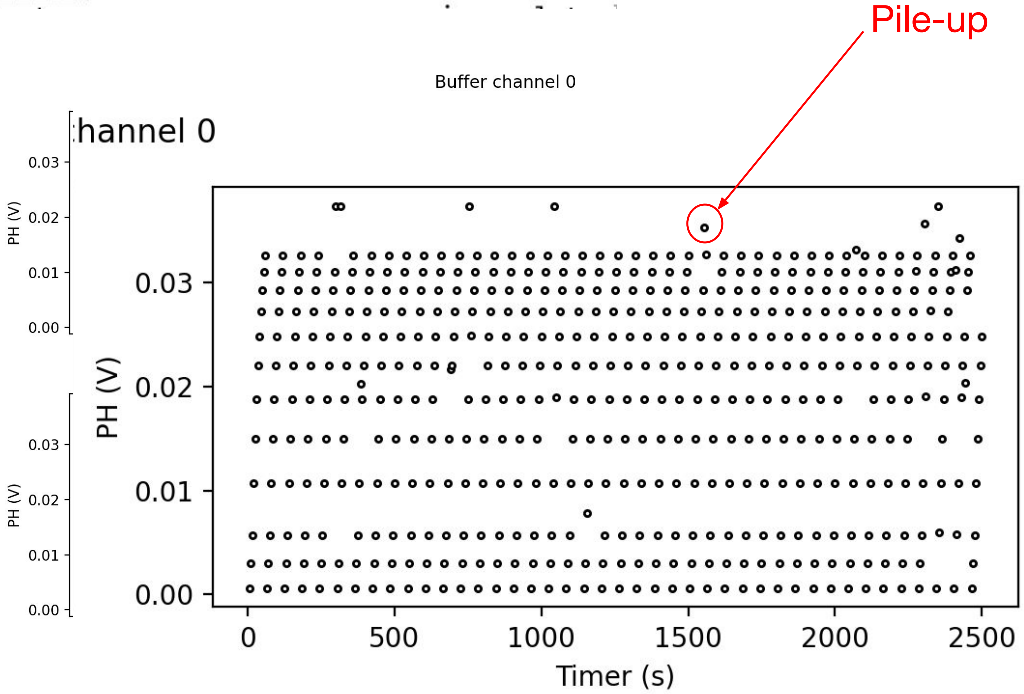
```
felix@Felixs-MacBook-Pro tests % python test_cryosweep.py stable_cryosig --plot
 --pileup_prob 0.05 --Ib -0.9 --dac -0.5 --which stable
100%|████████████████████████████████████████| 500/500 [01:51<00:00,  4.49it/s]
```



42 min continuous measurement time simulated in 1:51 runtime

… and the pulse height (PH) over time is monitored.

```
[felix@Felixs-MacBook-Pro tests % python test_cryosweep.py stable_cryosig --plot
 --pileup_prob 0.05 --Ib -0.9 --dac -0.5 --which stable
 100%|██████████████████████████████████████████| 500/500 [01:51<00:00,  4.49it/s]
```

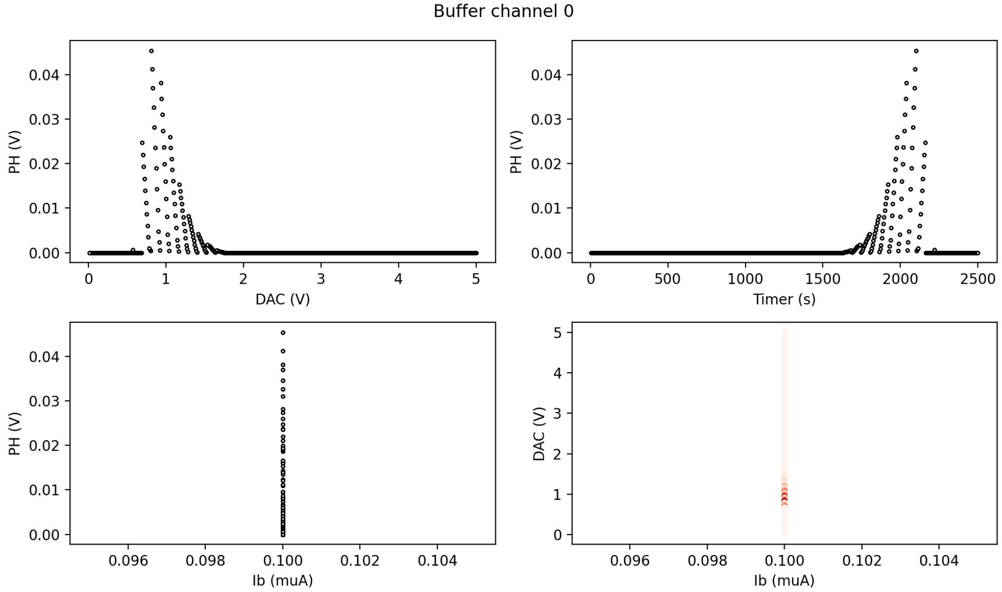


42 min continuous measurement time simulated in 1:51 runtime

Standard approach: find a good OP by sweeping DAC.

```
[felix@Felixs-MacBook-Pro tests % python test_cryosweep.py sweep_cryosig --plot -]
-from 1 --to -1 --dac_ramping_speed 2e-3 --pileup_prob 0. --Ib -1
```

SWEEP



42 min continuous measurement time simulated in 0:37 runtime (empty records are faster integrated than saturated pulses)

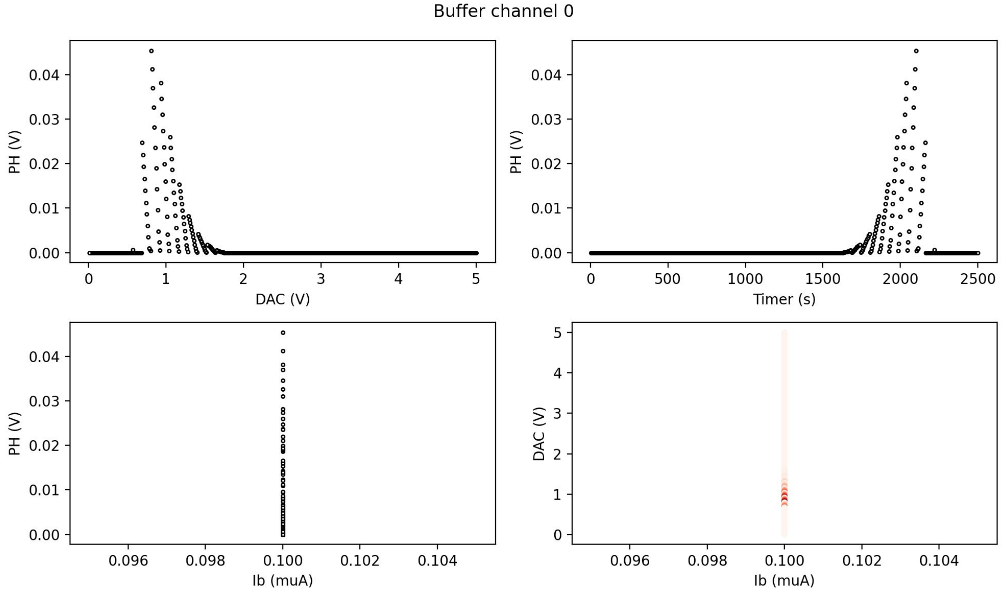
Standard approach: find a good OP by sweeping DAC.

```
[felix@Felixs-MacBook-Pro tests % python test_cryosweep.py sweep_cryosig --plot -]
-from 1 --to -1 --dac_ramping_speed 2e-3 --pileup_prob 0. --Ib -1
```

SWEEP

this is the minimum time required for the standard approach (typically it's 3-4 times this amount!)

42 min continuous measurement time simulated in 0:37 runtime (empty records are faster integrated than saturated pulses)



Can we do better than the standard approach?

Reward function: how do we tell which OP is good?

advanced

First, we need to quantify our objective.
The goal is a Low energy Threshold (LT):

$$\min_{\mathrm{DAC,Ib}} \left(E_{th}(\mathrm{DAC}, \mathrm{Ib}) \right)$$

For low E linear: $E = \mathrm{PCE} \cdot U$

$$\min_{\mathrm{DAC,Ib}} \left(\mathrm{PCE} \cdot U_{th}(\mathrm{DAC}, \mathrm{Ib}) \right)$$

Threshold is a multiple of the resolution

$$\min \left(6.5\, \mathrm{PCE} \cdot RMS \right)$$

PCE = E/U

$$\min \left(6.5\, \frac{E}{U} \cdot RMS \right)$$

Estimate the things with test pulses

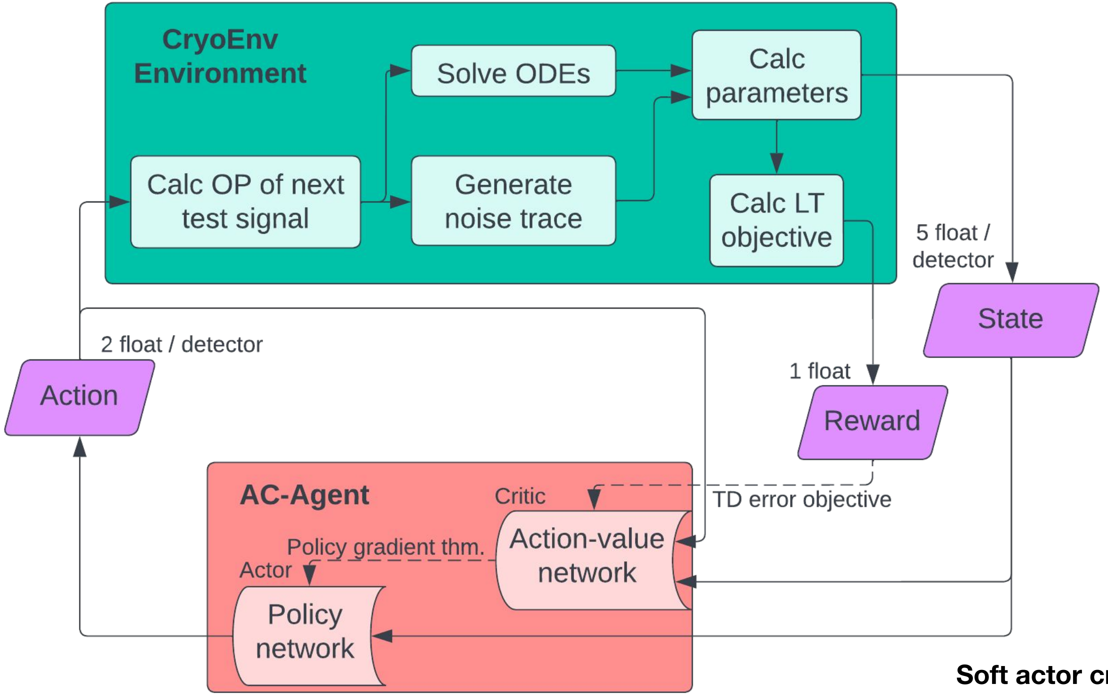
$$\min \left(6.5\, \frac{CPE \cdot TPA}{PH} \cdot RMS \right)$$

Neglect all constants, leads to …

Loss $\min \left(\frac{TPA}{PH} \cdot RMS \right)$ Reward $\max \left(- \frac{TPA}{PH} \cdot RMS \right)$

Training a SAC agent on CryoEnv

Action
(setting DAC and Ib)



Observation
(PH, RMS, TPA, DAC, I_B)
and ***Reward***
(LT objective)

Soft actor critic (SAC)
https://arxiv.org/abs/1801.01290v2
Stable baselines 3 implementation
https://jmlr.org/papers/volume22/20-1364/20-1364.pdf

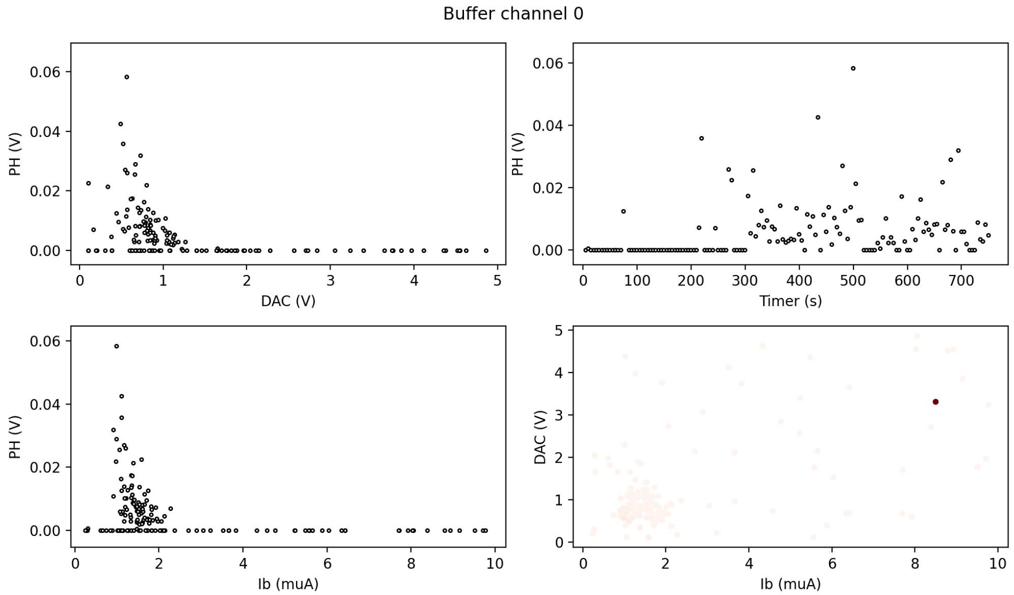
Simple scenario: no pile-up, optimize for one test signal

github.com/fewagner/CryoEnv

```
[felix@Felixs-MacBook-Pro tests % python train_wrapper.py sac_cryosig --n_steps 1]
50 --tpa_queue 1 --pileup_prob 0. --gradient_steps 100 --plot --gamma 0.6
Using cpu device
Wrapping the env with a `Monitor` wrapper
Wrapping the env in a DummyVecEnv.
100%|██████████████████████████████████████████| 150/150 [02:19<00:00,  1.08it/s]
```

TRAINING

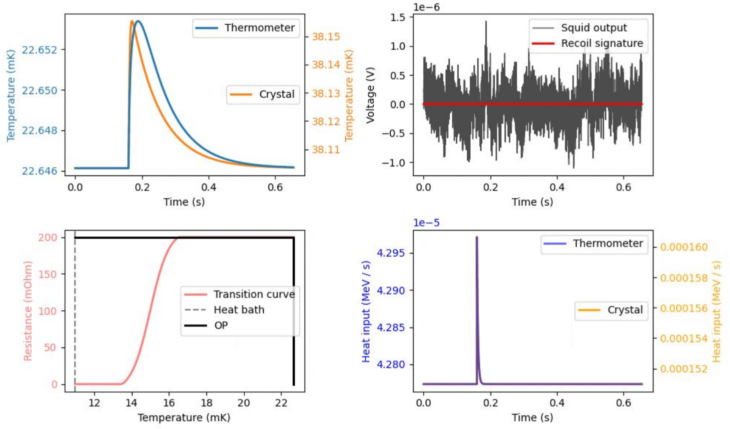
13 min training time simulated in 2:19 runtime
(~2-3 times speed up on GPU expected)



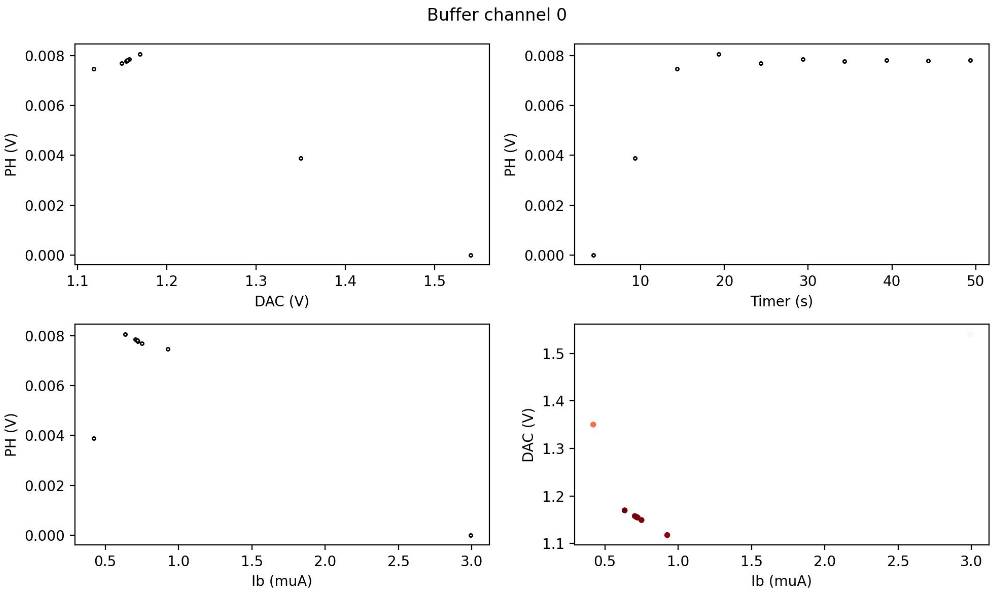
Simple scenario: no pile-up, optimize for one test signal

```
[felix@Felixs-MacBook-Pro tests % python run_wrapper.py sac_cryosig_1 --save_path]
 sac_run_cryosig --n_steps 10 --tpa_queue 1 --pileup_prob 0. --plot
100%|████████████████████████████████████████| 10/10 [00:01<00:00,  5.58it/s]
Average reward: -0.019466353398693147
```

TESTING



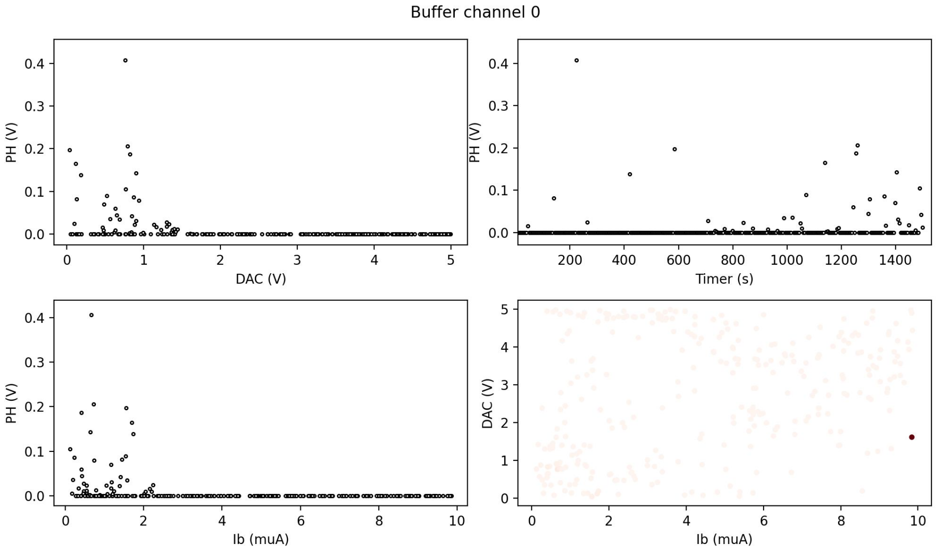
(This animation does not work in the PDF version.)



Challenging scenario: strong pile-up, multiple test signal

```
[felix@Felixs-MacBook-Pro tests % python train_wrapper.py sac_cryosig --n_steps 3]
00 --pileup_prob 0.3 --gradient_steps 50 --plot --gamma 0.6
Using cpu device
Wrapping the env with a `Monitor` wrapper
Wrapping the env in a DummyVecEnv.
100%|████████████████████████████████████████| 300/300 [02:57<00:00,  1.69it/s]
```

TRAINING



26 min training time simulated in 2:57 runtime,
this time with pile up and TPA queue

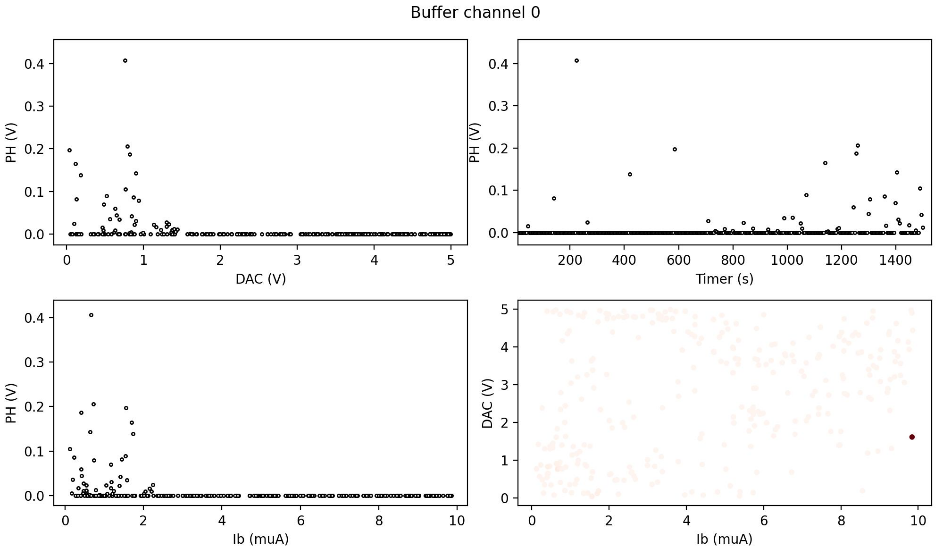
Challenging scenario: strong pile-up, multiple test signal

```
[felix@Felixs-MacBook-Pro tests % python train_wrapper.py sac_cryosig --n_steps 3]
00 --pileup_prob 0.3 --gradient_steps 50 --plot --gamma 0.6
Using cpu device
Wrapping the env with a `Monitor` wrapper
Wrapping the env in a DummyVecEnv.
100%|██████████████████████████████████████████| 300/300 [02:57<00:00,  1.69it/s]
```

TRAINING

this is our estimate for the required training time on a live experiment

26 min training time simulated in 2:57 runtime,
this time with pile up and TPA queue

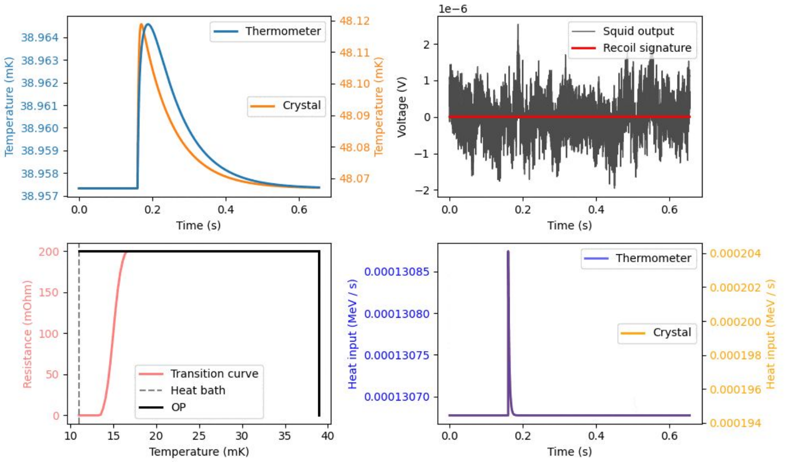


Detector runs robustly! 40% faster than standard approach.

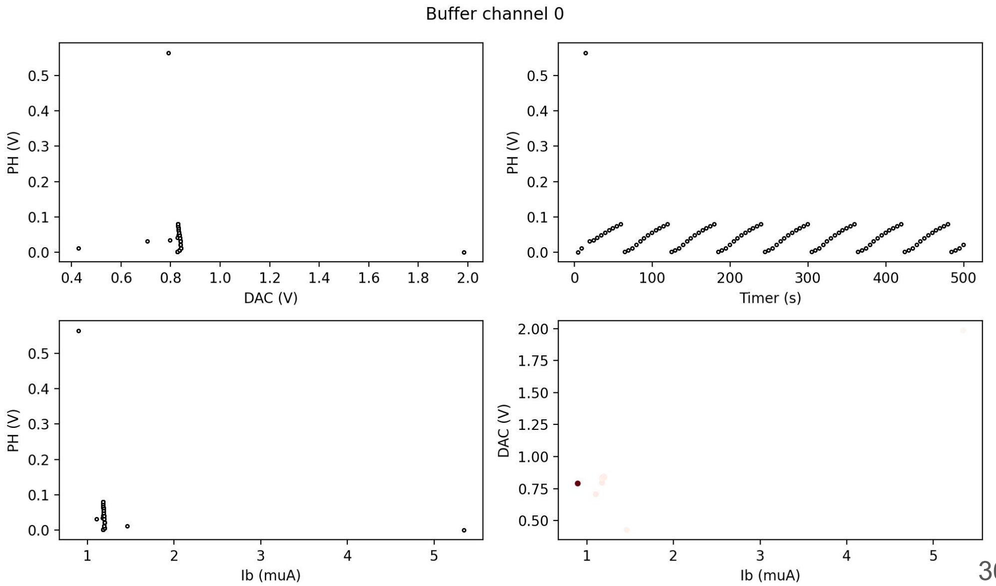
```
[felix@Felixs-MacBook-Pro tests % python run_wrapper.py sac_cryosig_1 --save_path]
 sac_run_cryosig --n_steps 100 --pileup_prob 0.5 --plot
100%|██████████████████████████████████████| 100/100 [00:39<00:00,  2.53it/s]
Average reward: -0.0923125738353164
```

TESTING

The optimal OP is slightly different - is it equivalent, or did we importance sample the rewards by the choice of TPA in training?



(This animation does not work in the PDF version.)



Conclusion, outlook and thanks a lot!

Special thanks for fruitful discussions and support to:
Florian Reindl
Franz Pröbst
Vanessa Zema
Johannes Rothe
Stephan Fichtinger
Wolfgang Waltenberger
Jochen Schieck
Clemens Heitzinger
The CRESST collaboration
The COSINUS collaboration

- Our approach requires **less measurement time** than the standard approach, **no manual interventions**, optimizes directly the **sensitivity**, and is **scalable** to multi-detector setups.
- First **runs in a live measurement environment** are planned for later this year - stay tuned!

Felix Wagner
(PhD)

Special kudos for our fantastic Master's/Bachelor's students:

Daniel Bartolot
(MSc)

Kolos Niedermayer
(Bsc)

No sparrows or dragons were harmed during the making of these slides.