signal processing in cryogenic experiments

Jules COLAS
CNRS, IP2I

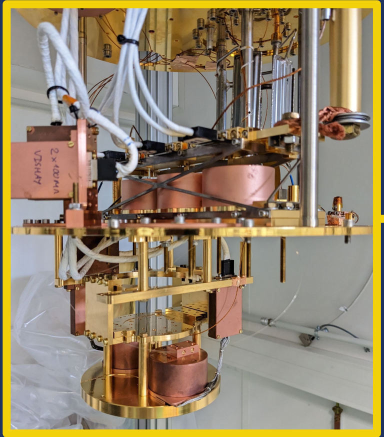
What is “a signal” ?

“everything can be a signal”

Signal and noise distinction is arbitrary, the signal is the information we care about, the noise is the sum of all other perturbations.

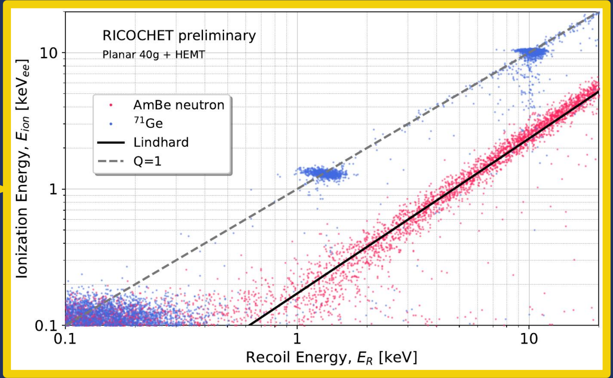
“a signal is what is not a noise”

A little bit of context



detector in a fridge

electronics / DAQ

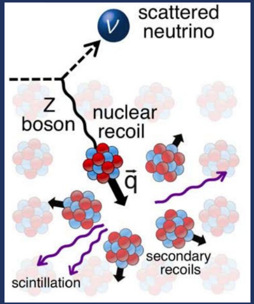


discoveries (at best)
fun stuff (at least)

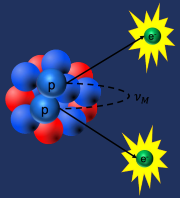
Signal processing **extracting information**

science/experiment interplay

the science determines the detection process
the experiment adds noises



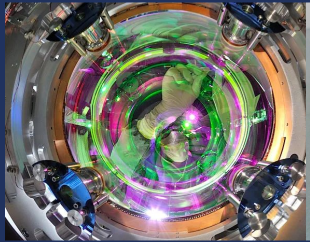
CEνNS



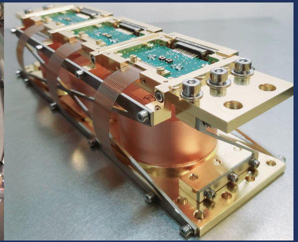
$0\nu\beta\beta$

general methods

same algorithms
different experiments



LIGO, GW



RICOCHET, CEνNS

Outlook of this lecture

Basics (15 min)

basic knowledge
useful definitions
tips

Case study (25 min)

~~bolometer~~ calorimeter experiment processing pipeline
(RICOCHET like)

Basics of spectral analysis

Fourier transform

Windowing

Convolution and correlation

Power Spectral Density estimation

Fourier transform definition

frequency index

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-i2\pi \frac{k}{N} n}$$

time index

frequency signal

time signal

size of signal

Note: we will restrict the discussion to the **discrete** Fourier transform and x_n can be complex or real valued.

Fourier transform cheat sheet

Parseval's theorem

$$\sum_{n=0}^{N-1} x_n y_n^* = \frac{1}{N} \sum_{k=0}^{N-1} X_k Y_k^*$$

Plancherel's theorem

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X_k|^2$$

signal energy is **conserved** through FT

Convolution theorem

$$\{u * v\}(x) = \mathcal{F}^{-1}\{U \cdot V\}$$

Dirac δ is the **neutral element** of convolution

Correlation

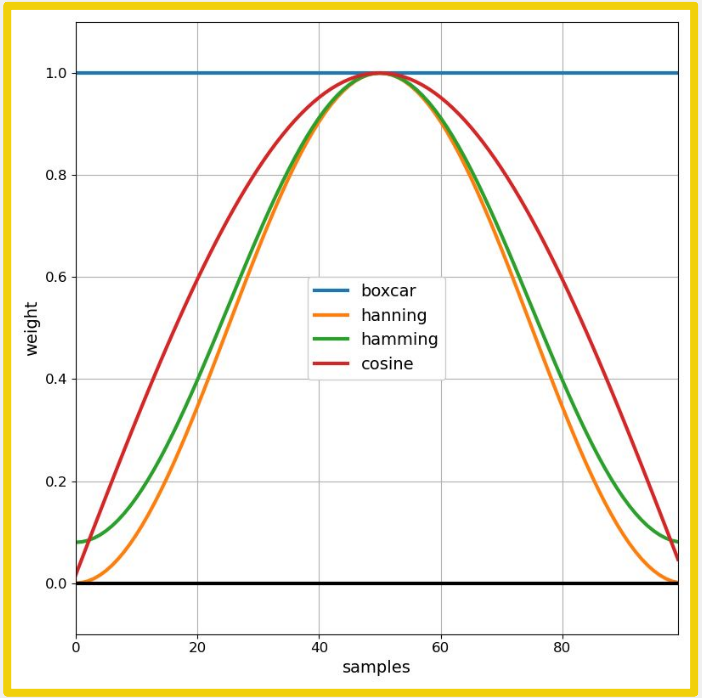
$$\mathcal{F}\{f \star g\} = \overline{\mathcal{F}\{f\}} \cdot \mathcal{F}\{g\}$$

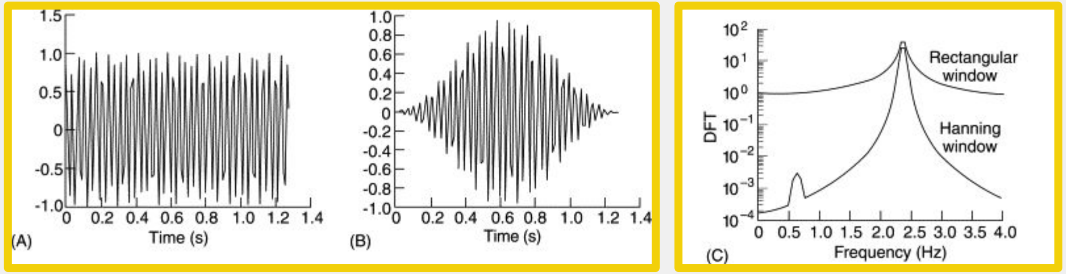
measure of similarity between f and g

Fourier transform windowing

windowing forces the signal to be null at the borders of **observation window → reduce frequency leakage**

there is a lot of windowing functions
but there is **no universal rule** to choose one





[Structural Dynamics Fundamentals and Advanced Applications, 2020](https://www.sciencedirect.com/topics/engineering/hanning-window)
https://www.sciencedirect.com/topics/engineering/hanning-window

Spectral estimation definition

Wiener-Khinchin

$$S(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} r_{xx}(k) e^{-i\omega k}$$

Autocorrelation

$$r_{xx}(\tau) = \mathbb{E}\left[x(t)^* \cdot x(t-\tau)\right]$$

direct link between statistical description and system evolution

one can define the **cross power spectral density** as the FT of the **cross-correlation** of two signal x, y

$$S_{xy}(f) = \sum_{n=-\infty}^{\infty} R_{xy}(\tau_n) e^{-i2\pi f \tau_n} \, \Delta\tau$$

in signal processing we often use the **cross-covariance**

$$\mathrm{R}_{XY}(\tau) = \mathrm{E}\left[X_{t-\tau} \overline{Y_t}\right]$$

→

$$\mathrm{K}_{XY}(\tau) = \mathrm{E}\left[(X_{t-\tau} - \mu_X) \overline{(Y_t - \mu_Y)}\right]$$

Spectral estimation coherency

Coherency definition

Power Spectral Density between x and y

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)}$$

Power Spectral Densities of x and y

Coherency estimate is **valid only for linear systems**
results can be inaccurate in case of a non-linear dynamic

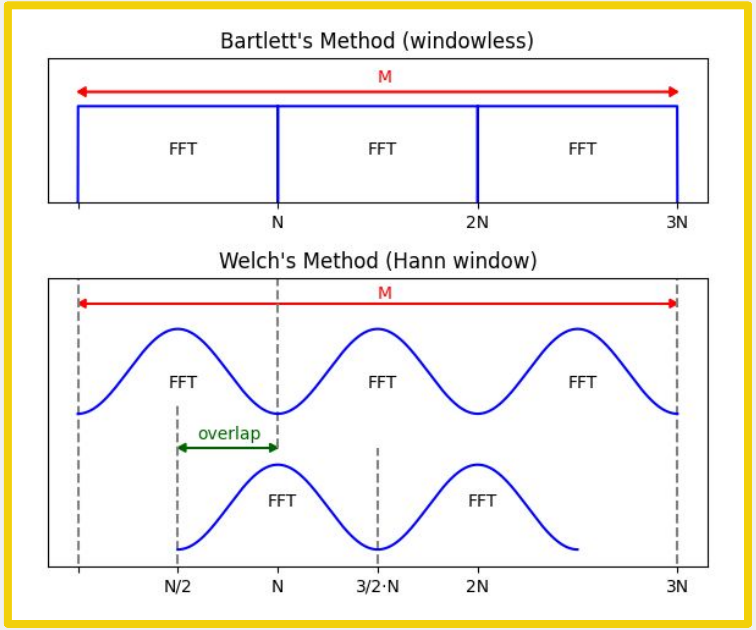
Spectral estimation a lot of methods...

- ***Non-parametric** methods* for which the signal samples can be **unevenly spaced in time**
- Least-squares spectral analysis, Lomb–Scargle periodogram, Non-uniform discrete Fourier transform
- ***Non-parametric** methods* for which the signal samples must be **evenly spaced in time**
  - Periodogram, Bartlett's method, Welch's method a windowed version of Bartlett's method that uses overlapping segments
  - Multitaper is a periodogram-based method that uses multiple tapers
  - Singular spectrum analysis
  - Short-time Fourier transform
  - Critical filter is a nonparametric method based on information field theory that can deal with noise, incomplete data, and instrumental response functions
- ***Parametric techniques***
  - Autoregressive model (AR) estimation, Moving-average model (MA) estimation, Autoregressive moving-average (ARMA)
  - MUltiple SIgnal Classification (MUSIC) is a popular superresolution method.
  - Maximum entropy spectral estimation is an *all-poles* method useful for SDE when singular spectral features are expected.
- ***Semi-parametric techniques***
  - SParse Iterative Covariance-based Estimation (SPICE) estimation, and the more generalized (r , q) -SPICE.
  - Iterative Adaptive Approach (IAA) estimation
  - Lasso, similar to least-squares spectral analysis but with a sparsity enforcing penalty.

Source: wikipedia

Spectral estimation tips

The Bartlett/Welch methods



IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3058744.

Median / mean

median is **less sensitive to outliers** than mean

more robust than standard Welch method

need to correct the **median bias**

$$\alpha = \sum_{\ell=1}^{n} \frac{(-1)^{\ell+1}}{\ell}$$

median can be used to estimate **σ** (RMS)

σ = k . MAD
MAD = median absolute deviation
k ~ 1.4826

In Python:
scipy.signal.welch(..., average="median"),
scipy.stats.median_abs_deviation(..., scale="normal")

Case study → multi-channel calorimeter

RICOCHET
A Coherent Neutrino Scattering Program

Goal 1: electronic noise estimation (including correlations)

Goal 2: automatic pulse detection and amplitude estimation

Constraints lowest energy possible + “noisy” environment

Colas, J., Billard, J., Ferriol, S. *et al.*
Development of Data Processing and Analysis Pipeline for the Ricochet Experiment.
J Low Temp Phys 211, 310–319 (2023).
https://doi.org/10.1007/s10909-022-02907-5

Degradation model

describes how the **detector measurement line** impacts the **signal of interest**
we are looking for “u”, but we measure “x”

today, we’ll go with this one

additive noise $X = U + B$ ←

cross-talk $X = U + H.Y$

linear transformation $X = A . (U + B) + C$ (example)

Note: the degradation model will define which method is best suited for your problem (need additional hypothesis)

Processing pipeline

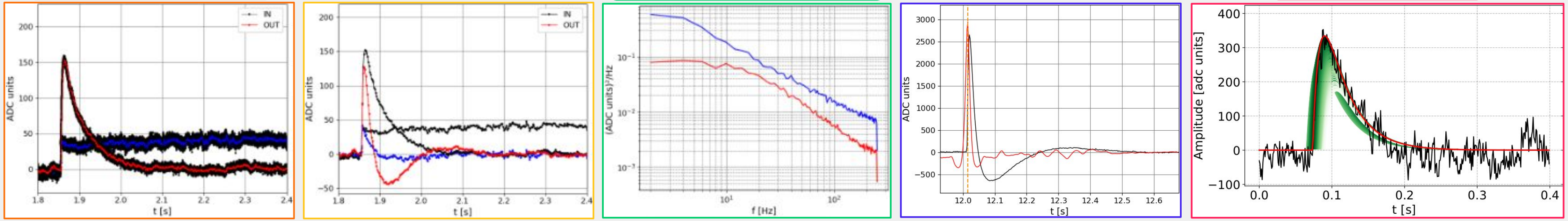
raw data file → **processing** → processed file

extract reduced quantities

combiner

downsampler → pre-filter → noise estimator → trigger → minimizer

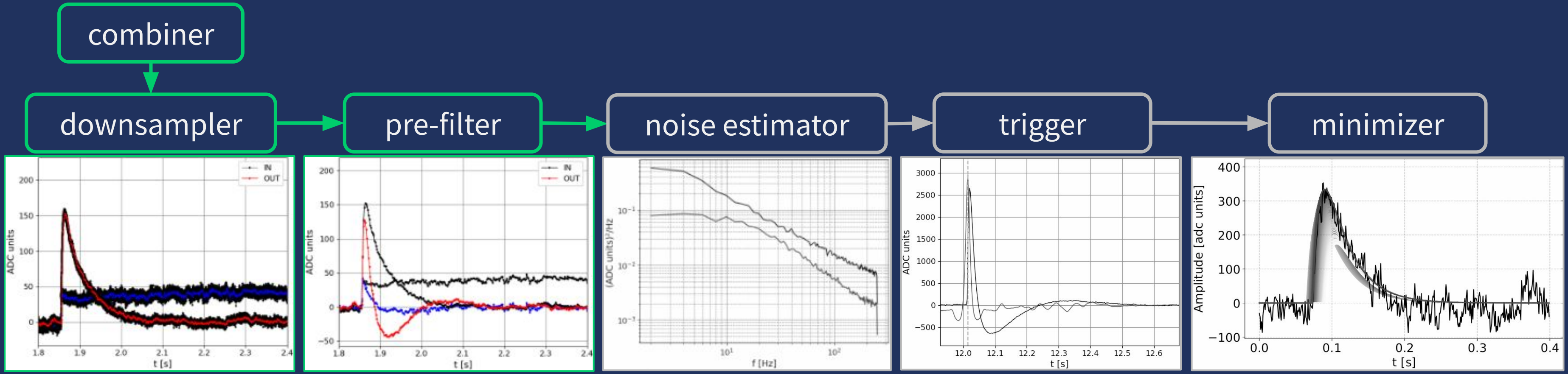
RQ's



continuous data stream → reduced quantities

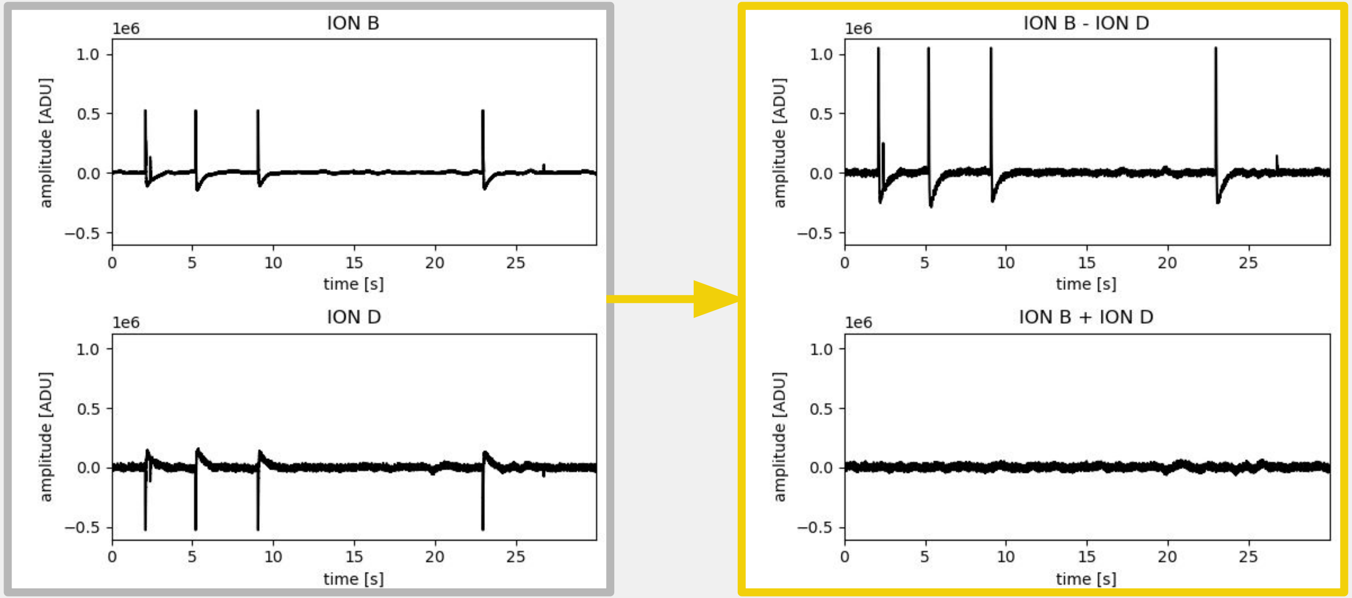
Pre-processing

Goal : prepare data for more reliable or faster analysis



continuous data stream → “clean” data stream

Pre-processing combine channels

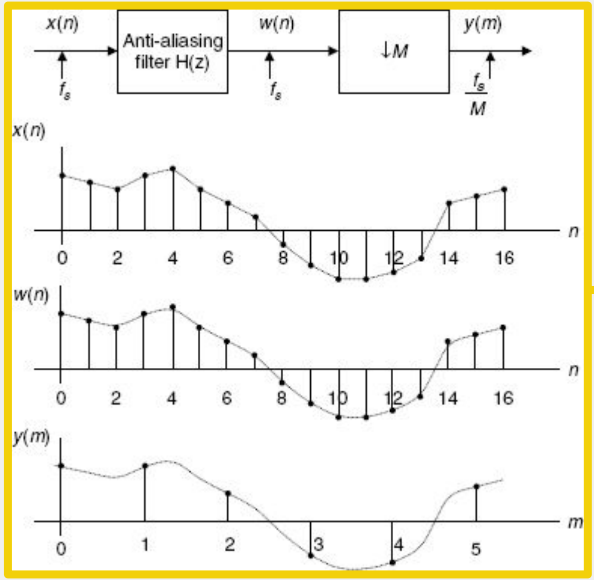


Sometime, signal processing can be trivial

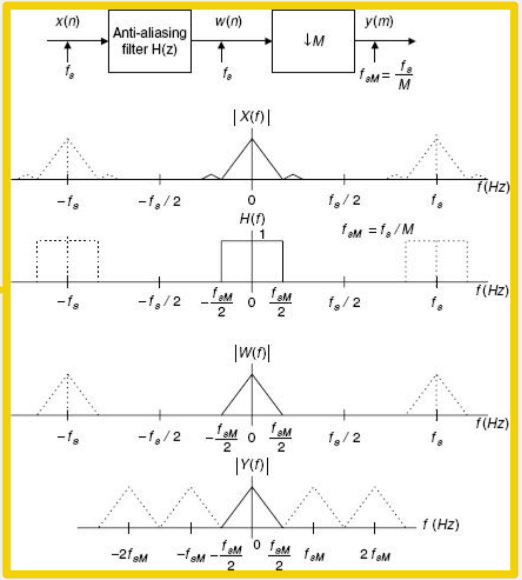
e^--h^+ collection → differential signal → summing them will improve SNR by sqrt(2)

Pre-processing downsampling

Time



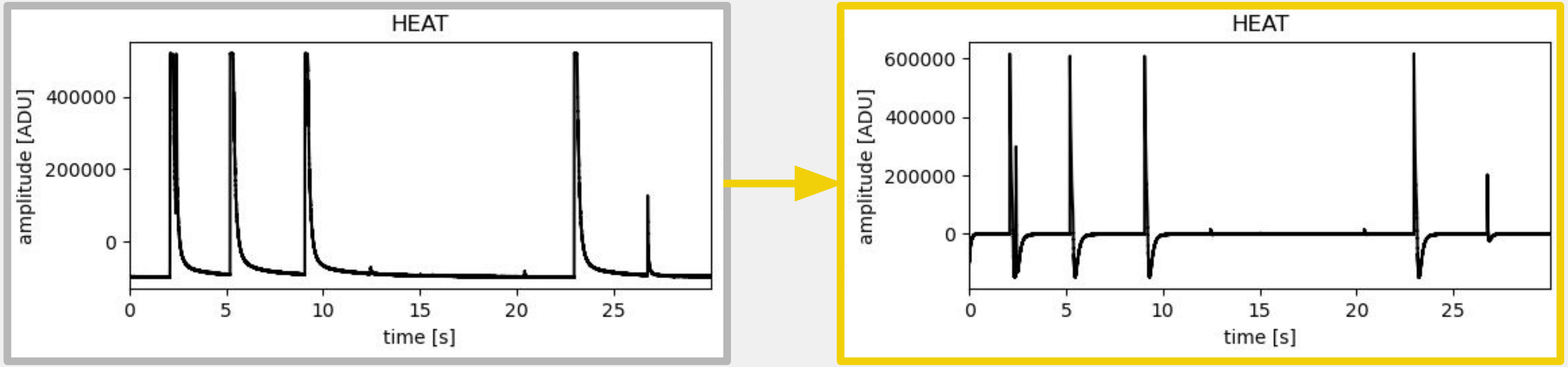
Frequency



Downsampling reduces the data size, thus improves the processing speed and (**may**) degrade performances reducing the number of sample will induces **aliasing → you need to filter these aliases before decimation**

Source: https://www.eetimes.com/multirate-dsp-part-1-upsampling-and-downsampling/

Pre-processing high-pass filtering



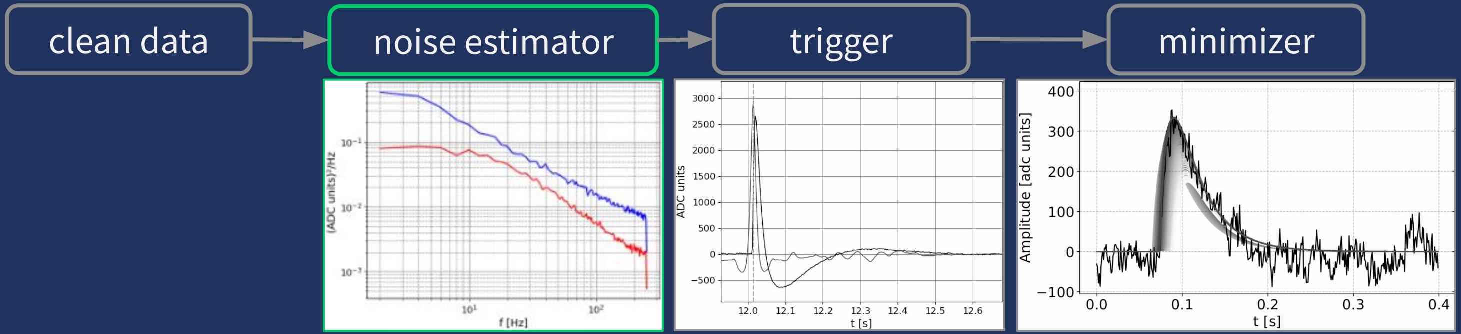
It’s necessary to define an observation window of size **tw**, so it’s not possible to (**accurately**) observe frequencies below **2/tw → remove them !**

Ex: If **tw** = 1 s, **fc = 2/tw = 2 Hz**

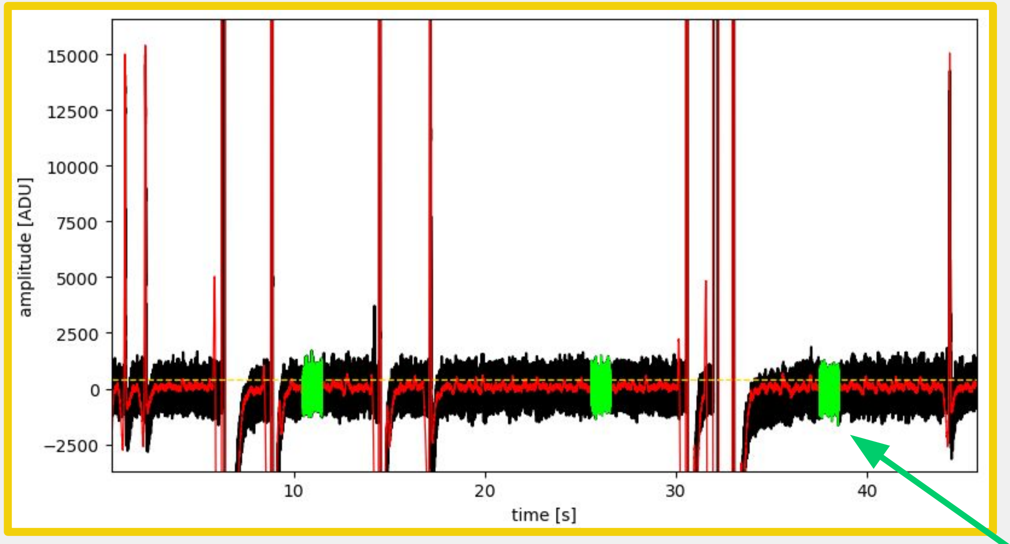
Noise estimation

Goal : estimate electronic noise floor PSD/CSD

Constraints : lot of pulses (wrt. det. dynamic), stationary noises (ergodicity), …



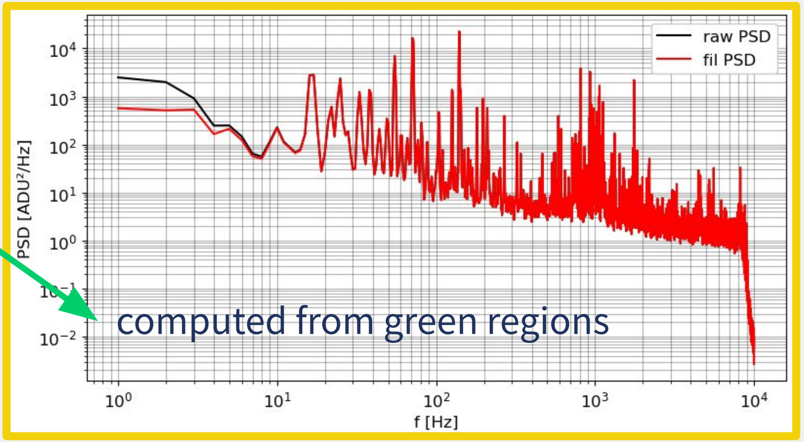
Noise estimation



Principle

we know the signal we are looking for: **pulses**

we can use a **similarity** measurement to identify the sections of data which are **"as different as possible of a signal"**



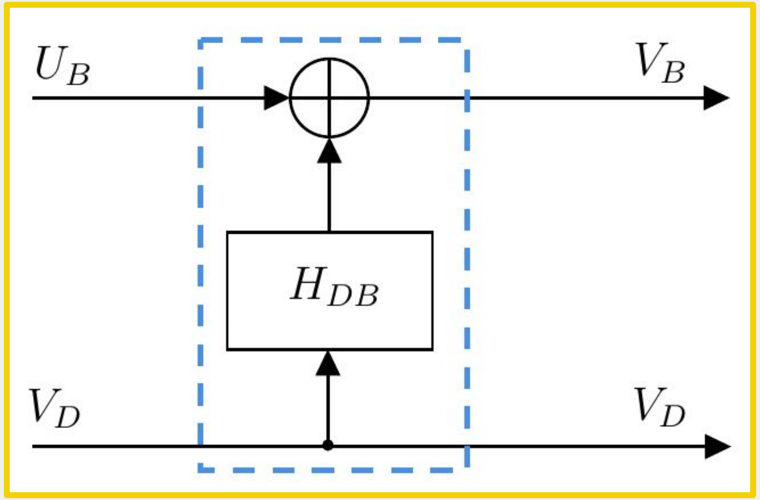
Note: for a more precise description of this algorithm see my Phd thesis

Exercise 1: noise decorrelation

We can use many approaches to clean our data, let's talk about noise decorrelation

signal of interest

"a noise"



signal measured

"a noise"

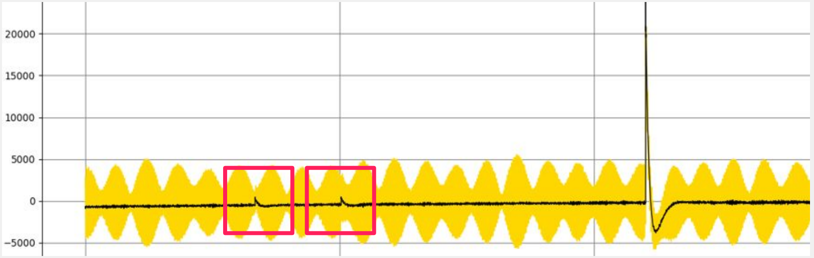
Question 1: find the coupling transfer function that minimize the energy of decorrelated signal

Question 2: determine the relation between the PSD of U_B and V_B

Exercise 1: noise decorrelation (results)

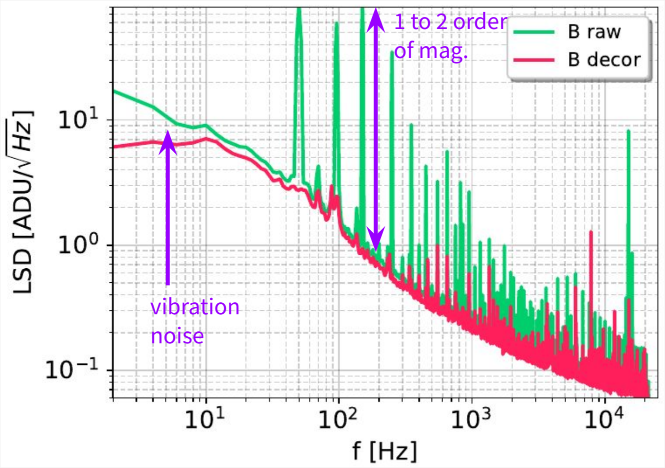
Question 1

$$H_{DB}(f) = \frac{\Re(V_B^*(f)V_D(f))}{|V_D(f)|^2}$$



Question 2

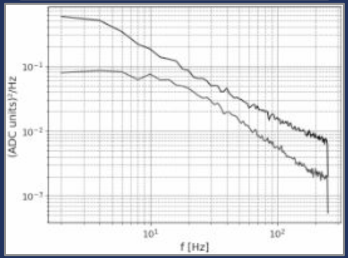
PSD(U_B) = **(1-C_{BD})**.PSD(V_B)

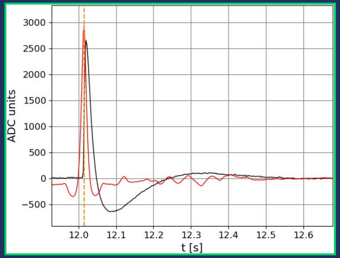


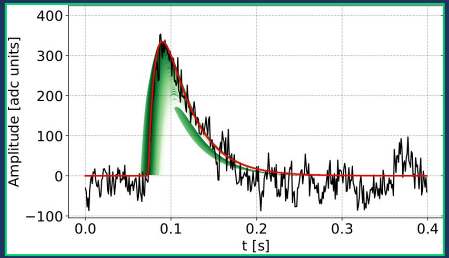
Automatic pulse detection

Goal 1 : detect pulses at the lowest energy possible
Goal 2 : estimate the amplitude of each individual pulse

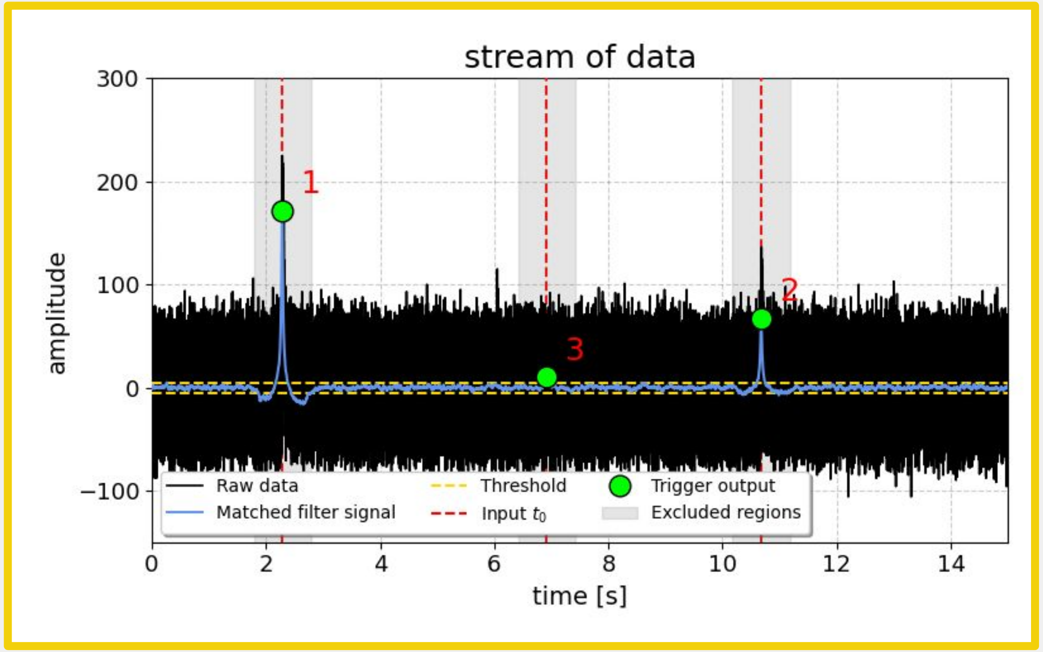
clean data → noise estimator → trigger → minimizer







Automatic pulse detection



matched filter $H = c\frac{P^*}{J}$

pulse template

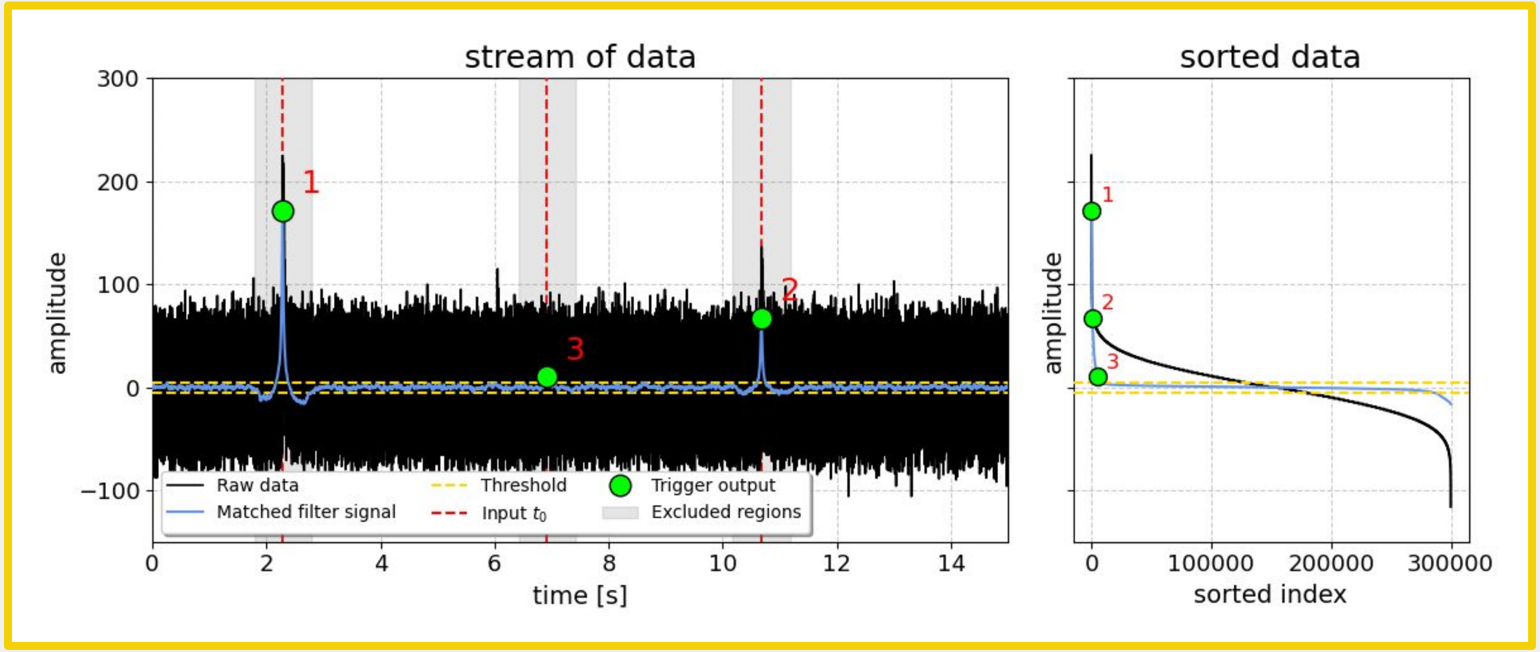
electronic noise psd

normalization factor

the matched filter **optimizes the SNR**

it implies a **lower detection threshold**
→ useful for low energy physics

Automatic pulse detection



1. sort by amp.
2. take first sample
3. exclude in time
4. take next
5. go back to 3

Exercise 2: amplitude estimation

We identified the approximate position of pulses → **what are their amplitudes ?**
we want an **objective function to minimize → chi²**

Question 1: how to estimate the amplitude "a" from chi² expression ?

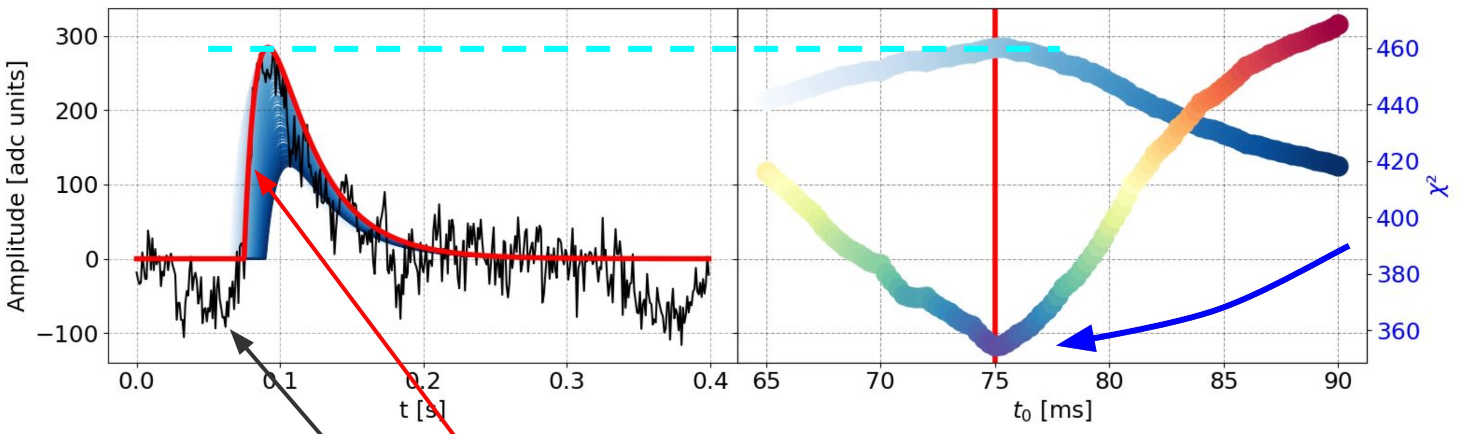
$$\chi^2 = \sum_f \frac{|D - a.M|^2}{J}$$

This expression suppose a gaussian variance likelihood

in short, minimizing chi² is equivalent to maximizing the likelihood

Exercise 2: amplitude estimation

Frequential pulse fitting illustration



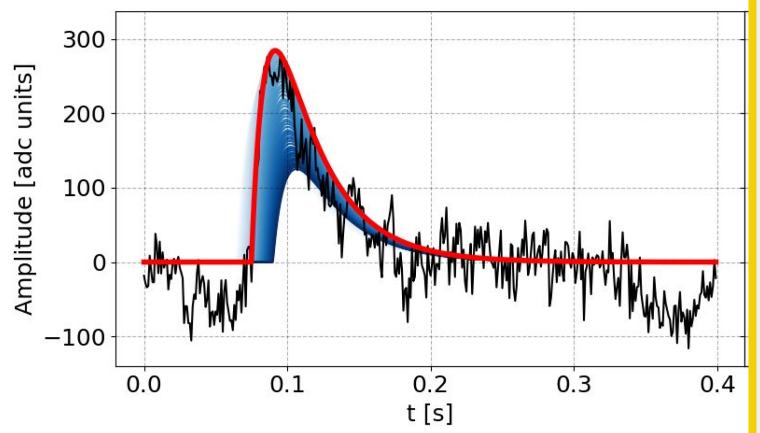
$$\chi^2 = \sum_f \frac{|D - a.\, M|^2}{J}$$

χ^2 can be seen as a measure of **goodness of fit**

min χ^2 ←→ **max â**

Energy resolution

What is the energy resolution ?

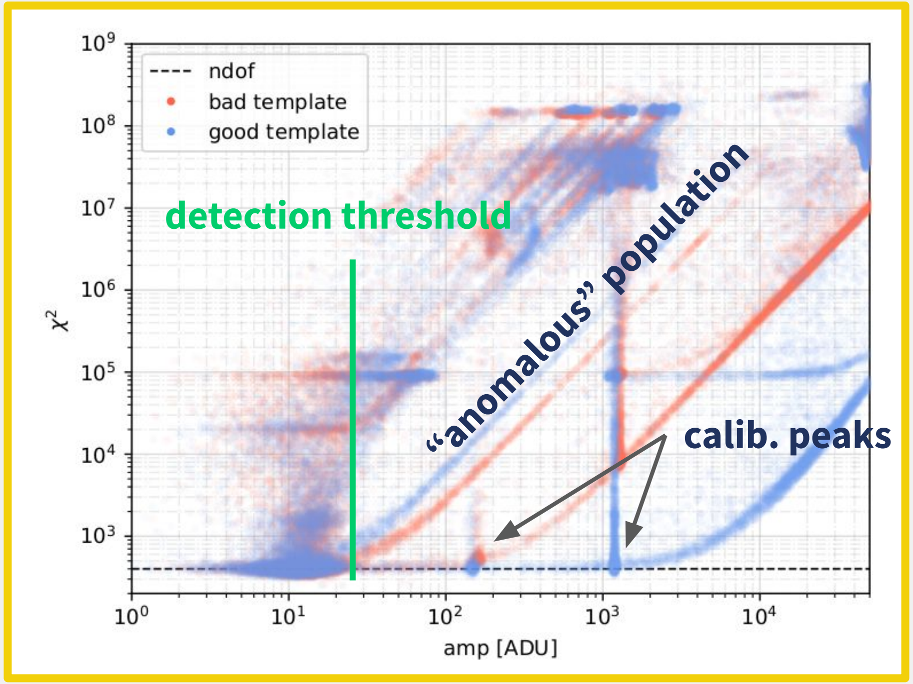


Formula

$$\sigma^2 = \left(\frac{1}{2} \frac{\partial^2 \chi^2}{\partial \hat{a}^2} \right)^{-1}$$

The timing resolution can also be derived from χ^2

What we get at the end ...



goodness of fit

amplitude

Conclusion

- Signal processing (SP) **can be simple** (ex: addition of two signals)

- There is a bunch of **unused and interesting methods**
  - interact with SP experts ! (it's a very wide field)

- A cryogenic detector is a **complex system** which combine hardware and software
  - SP is a crucial step in the detection pipeline

- SP method to use depends on the **sensor conditioning**
  - detector design
  - amplifiers type

" Information is the resolution of uncertainty. " Claude Shannon

Literature

Non-linear methods:

“Improving the Performance of Cryogenic Calorimeters with Nonlinear Multivariate Noise Cancellation Algorithms”

arXiv:2311.01131v2 [physics.ins-det] 6 Feb 2024

Decorrelation methods:

“Automatic cross-talk removal from multi-channel data”

arXiv:gr-qc/9909083v1 27 Sep 1999

“Noise correlation and decorrelation in arrays of bolometric detectors”

arXiv:1203.1782v1 [physics.data-an] 8 Mar 2012

Threshold reduction:

“Lowering the energy threshold of large-mass bolometric detectors”

arXiv:1012.1263v1 [astro-ph.IM] 6 Dec 2010

“Processing the signals from solid-state detectors in elementary-particle physics.”

Riv. Nuovo Cim. **9**, 1–146 (1986) https://doi.org/10.1007/BF02822156

Parameter/resolution estimation:

“Optimum filter-based analysis for the characterization of a high-resolution magnetic microcalorimeter towards the DELight experiment”

arXiv:2310.08512v1 [hep-ex] 12 Oct 2023

“When “Optimal Filtering” Isn’t”

arXiv:1611.07856v1 [physics.data-an] 23 Nov 2016

\+ variational minimization, adaptative filtering (active denoising), wavelets, non-gaussian noise hypothesis, machine learning, …

Likelihood vs. chi^2

N pulses fitted simultaneously with beta_n the “local” parameters (t0, a, …) and theta the “global” shape parameter

$$\mathcal{L}(\vec{\theta}|D_0,\ldots,D_N) = \prod_f \prod_{n=1}^{N} e^{-\frac{1}{2}\frac{|D_n(f)-\overline{D}_n(f)|^2}{J(f)}}$$

$$-\log\mathcal{L}(\vec{\theta}) = \frac{1}{2}\sum_{n=1}^{N}\chi^2(\vec{\theta};\vec{\beta}_n)$$

$$\left.\frac{\partial \log\Big(\mathcal{L}(\vec{\theta}|D_0,\ldots,D_N)\Big)}{\partial\vec{\theta}}\right|_{\vec{\theta}=\hat{\vec{\theta}}} = 0$$

$$\equiv \sum_{n=1}^{N}\left.\frac{\partial\chi^2(\vec{\theta};\vec{\beta}_n)}{\partial\vec{\theta}}\right|_{\vec{\theta}=\hat{\vec{\theta}}} = 0$$

minimizing the chi^2 = **maximizing** the likelihood

Generalization

mono-channel

$$\chi^2 = \sum_f \frac{|D - a.\, M|^2}{J}$$

noise PSD

multi-channels

$$\chi^2(\vec{a}; t_0) = \sum_f \left[\vec{Y} - \vec{a}\vec{P}(\vec{\theta}; t_0)\right]^\dagger J^{-1} \left[\vec{Y} - \vec{a}\vec{P}(\vec{\theta}; t_0)\right]$$

cross-covariance matrix