Data analysis tools for the HOLMES experiment

NuMass - Milano 2022

Luca Origo on the behalf of the HOLMES collaboration

Istituto Nazionale di Fisica Nucleare

Università degli Studi di Milano Bicocca

Slides outline

- **The HOLMES experiment**
  - **measurement, sensitivity**
- Data handling
- Pulse analysis
- Parameter estimation
- Background rejection

The HOLMES experiment

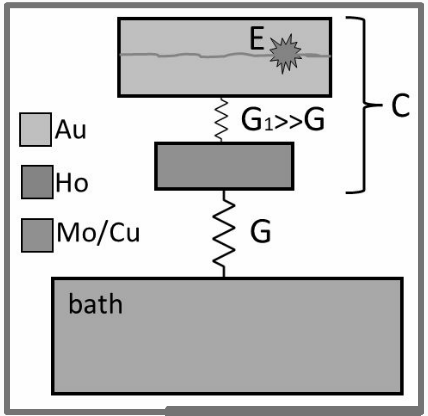
- **direct** and calorimetric m_ν measurement
- ^{163}Ho electron capture
- sensitivity extrapolated from spectral fit

no model-dependence

Assessing a measurement only relying on the energy-momentum conservation principle.

source ⊂ detector

The decay energy is entirely absorbed except for the neutrino contribution.



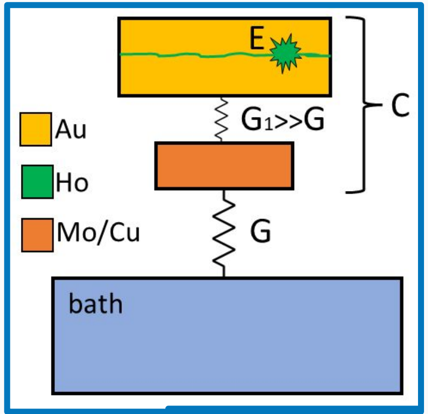
TES scheme

The HOLMES experiment

- direct and **calorimetric** m_ν measurement
- ^{163}Ho electron capture
- sensitivity extrapolated from spectral fit

source ⊂ detector

The decay energy is entirely absorbed except for the neutrino contribution.



TES scheme

no model-dependence

Assessing a measurement only relying on the energy-momentum conservation principle.

KATRIN
KARLSRUHE TRITIUM NEUTRINO EXPERIMENT

ECHo

PROJECT 8

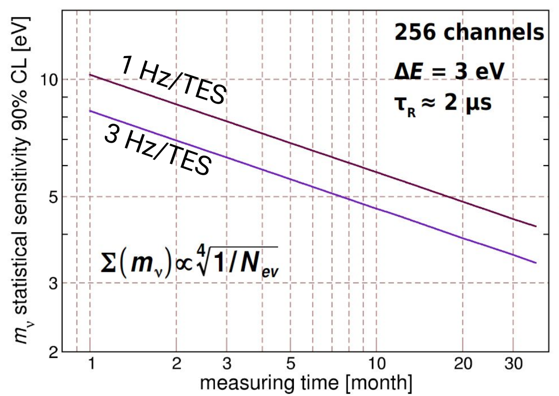
The HOLMES experiment

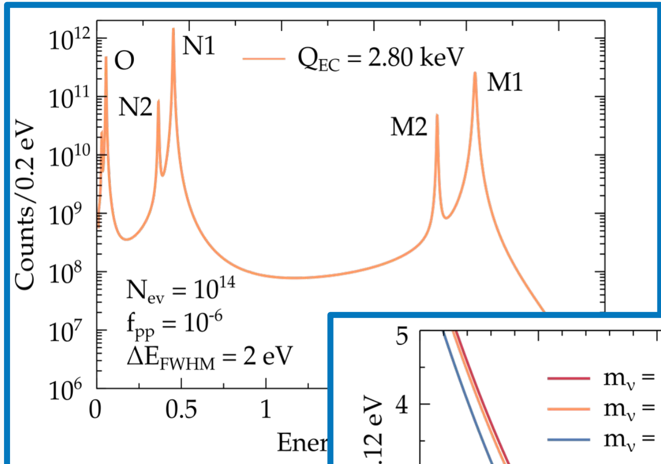
- direct and calorimetric m_ν measurement
- ^{163}Ho electron capture
- **sensitivity** extrapolated from spectral fit

ROI lineshape(E_c) ∝ $\sqrt{(Q-E_c)^2 - m_\nu^2}$

$$^{163}\text{Ho} + e^- \rightarrow {}^{163}\text{Dy}^{(*)} + \nu_e$$

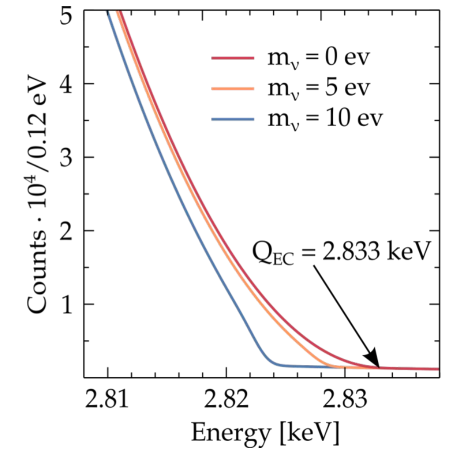
E_c





simulated spectrum

ROI



The HOLMES experiment

Data analysis for the m_ν measurement

Energy resolution as improved as possible

...

...

Dead time as low as possible

...

...

Time resolution as improved as possible

...

...

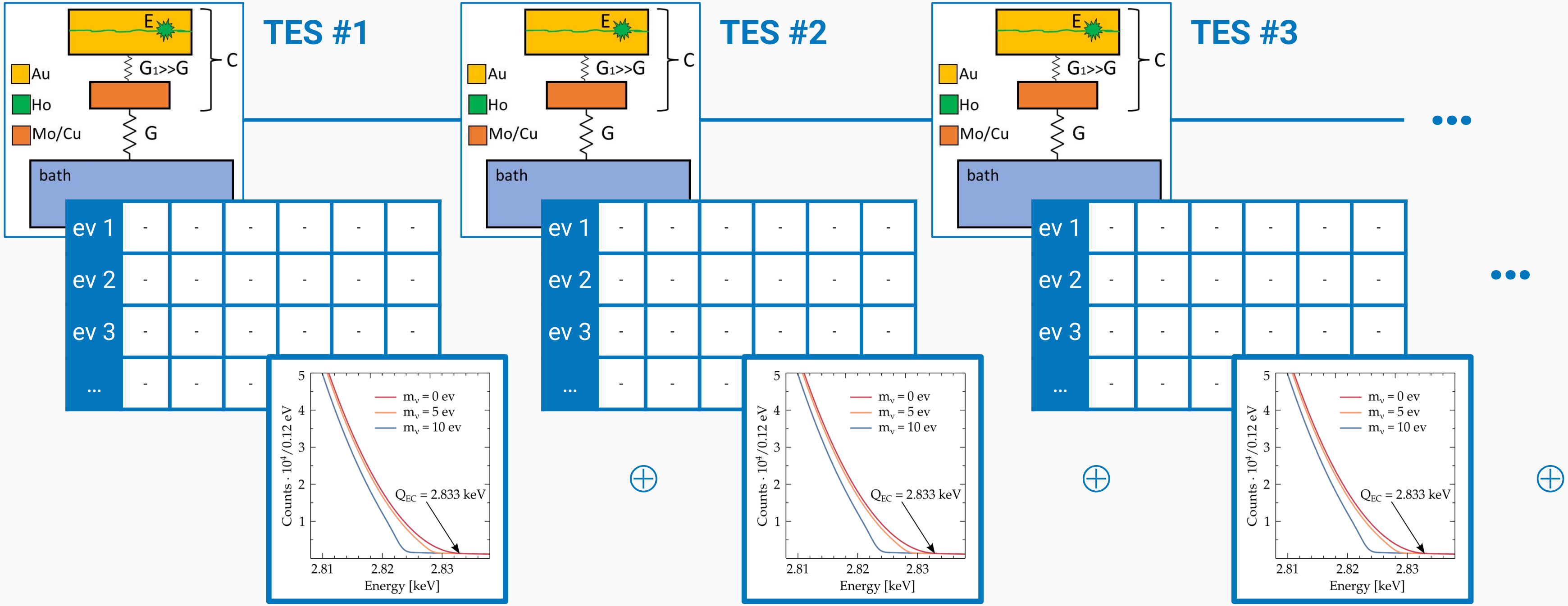
Slides outline

- The HOLMES experiment
- **Data handling**
  - **data taking, compression and tagging**
- Pulse analysis
- Parameter estimation
- Background rejection

Data handling

- **Signals collected in arrays**
- **t_{samp}=2µs, n_{pts}=1024, τ_R~15µs, τ_D~350µs**

- Python matricial operations with HDF5 file
- Data compression through parametrization



Data handling

- Signals collected in arrays
- t_{samp}=2μs, n_{pts}=1024, τ_R~15μs, τ_D~350μs

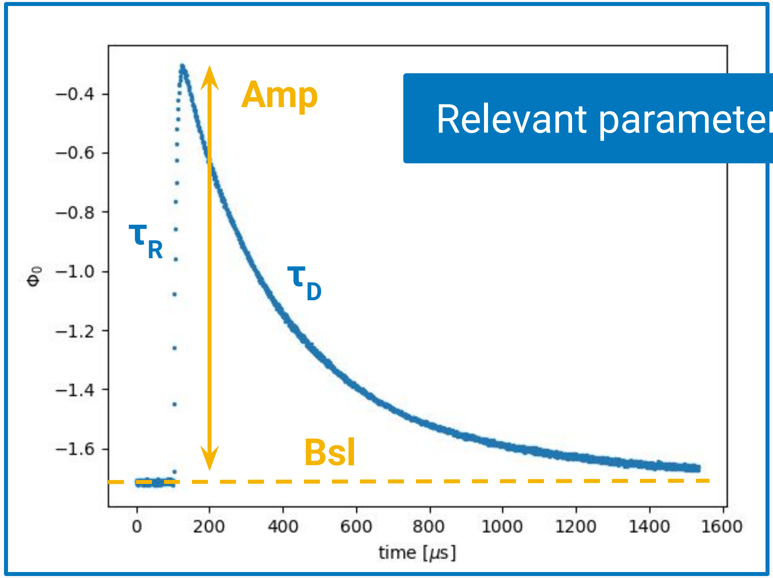
- **Python matricial operations with HDF5 file**
- **Data compression through parametrization**

N_{ev} x n_{pts}

	pt 1	pt 2	…	…	pt n-1	pt n	t_0
ev 1							
ev 2							
…							
ev N							

N_{ev} x n_{par}

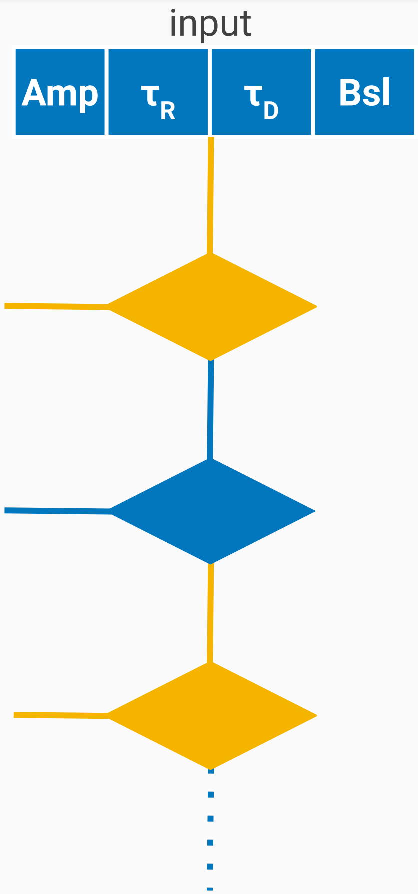
	Amp	τ_R	τ_D	Bsl	t_0
ev 1					
ev 2					
ev 3					
…					



Relevant parameters only

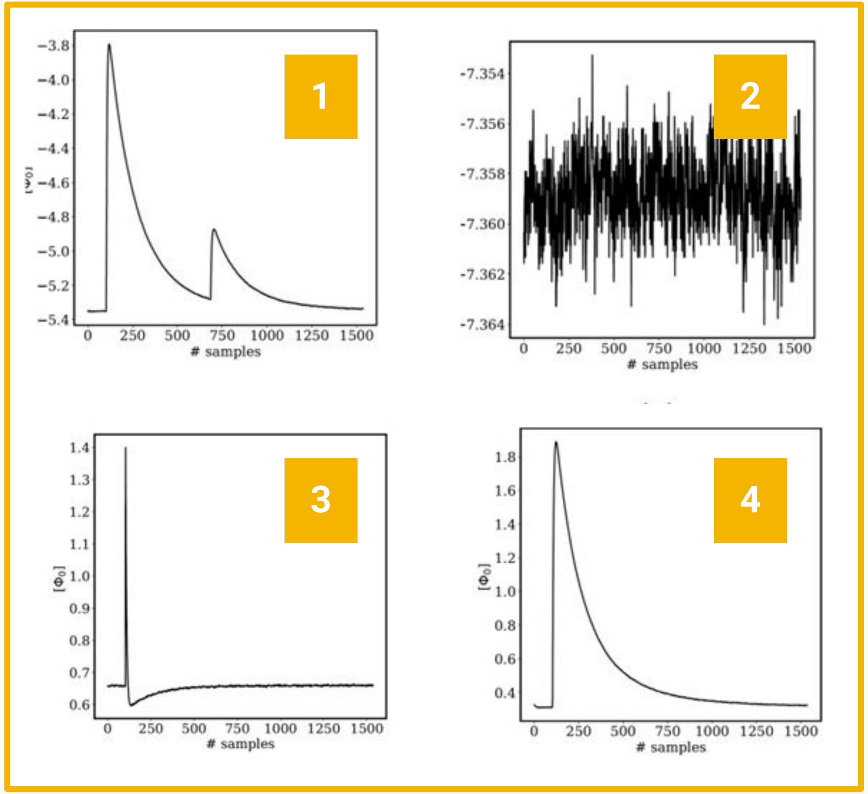
Data handling

- Parametrization ⟶ offline events tagging (**clustering algorithms** under development to automate the operation)
  - 5 parameters **thresholds**:
    - **empty** (useful ⟶ noise spectrum analysis)
    - **strange**
    - **multiple** (descent pile-up)
    - **bad-baseline**
    - **coincidence** (cosmic muons)
  - Untagged events are labeled as **'good'**



Data handling

- Parametrization ⟶ offline events tagging (**clustering algorithms** under development to automate the operation)
  - 5 parameters **thresholds**:
    - **empty** (2) (useful ⟶ noise spectrum analysis)
    - **strange** (3)
    - **multiple** (1) (descent pile-up)
    - **bad-baseline** (4)
    - **coincidence** (cosmic muons)
  - Untagged events are labeled as **'good'**



Slides outline

- The HOLMES experiment
- Data handling
- **Pulse analysis**
  - **optimum filter**
  - **correction algorithms**
- Parameter estimation
- Background rejection

Pulse analysis

- Amplitude estimation by means of **optimum filter application**
  - Signal-to-Noise ratio is maximized
  - An average signal is required

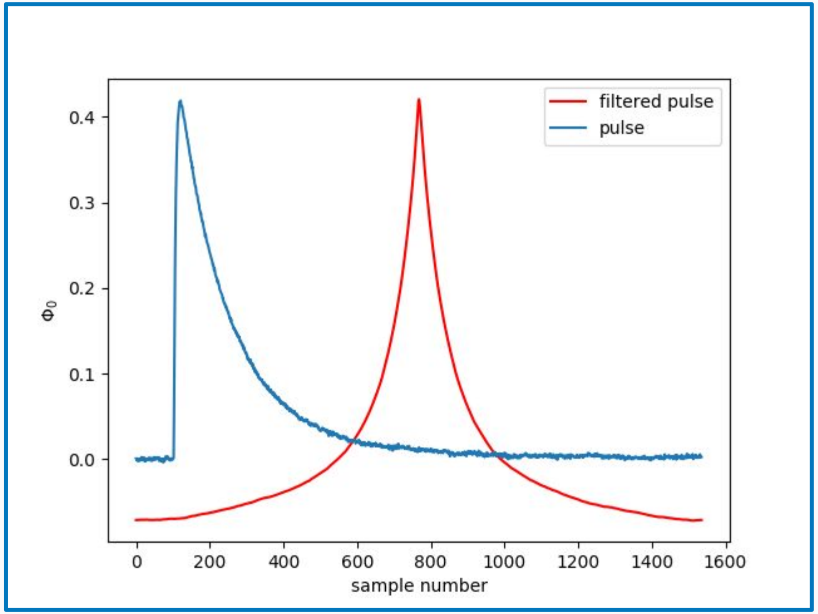
- **Assumptions:**
  - noise is ergodic
  - signal is well-sampled
  - signal modeled as:

$$\mathbf{s}[i] = K(E) \cdot \mathbf{m}[i] + \mathbf{n}[i]$$

signal

average

noise



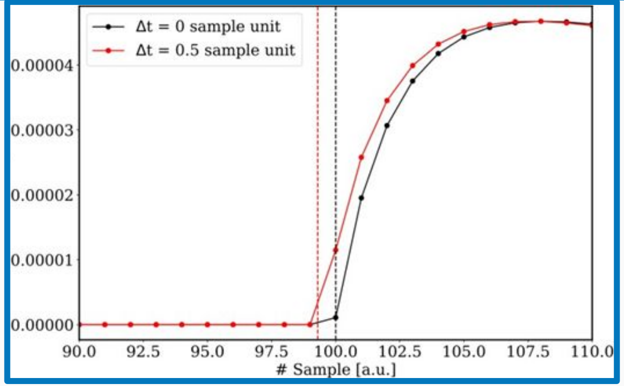
Pulse analysis

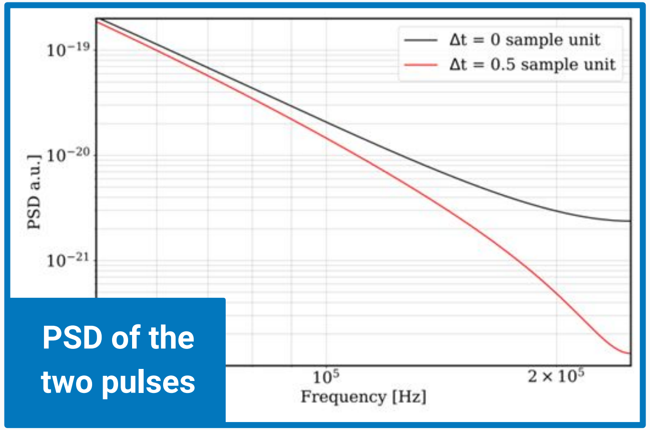
- What's the real arrival time of each signal?

$$\Delta t = |t_{true} - t_o|$$

- Discrete sampling produces an amplitude smearing
  - ⬆️ sampling frequency ⇔ effect ⬇️
  - ⬇️ points on the pulse's rise ⇔ effect ⬆️
- Solution: moving average to smooth the signal's rise
  - finding the best one that optimizes our spectral resolution
- The **arrival time correction** avoids a distortion of the energy spectrum

Rising edge samples of two simulated pulses





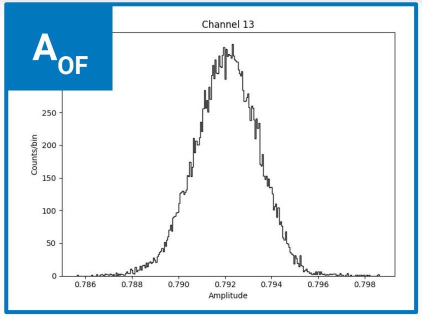
PSD of the two pulses

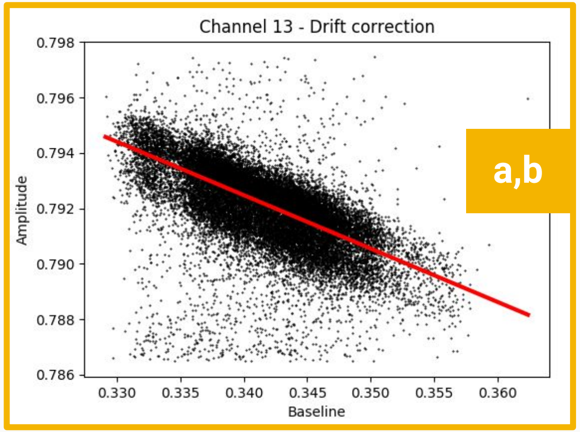
Pulse analysis

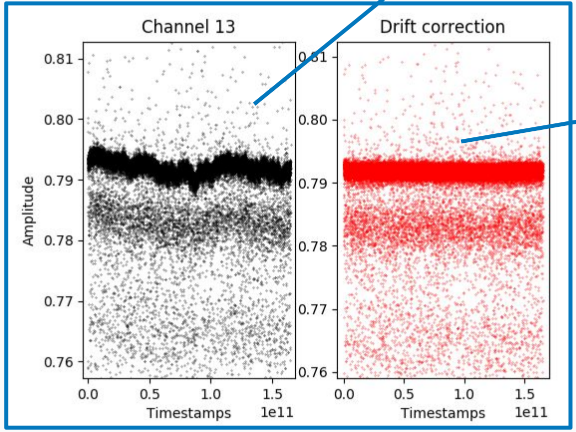
- Detector's gain depends on the baseline **B**
  - parameters (a,b) estimated from linear regression

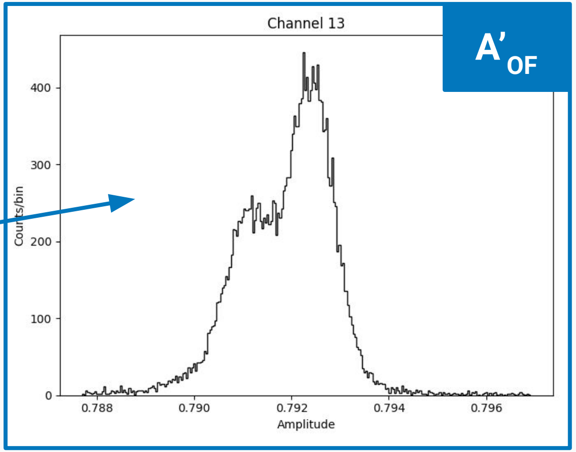
$$A'_{OF} = A_{OF} - f(B) = A_{OF} - a \cdot B - b$$

- Also the **drift correction** avoids a distortion of the energy spectrum





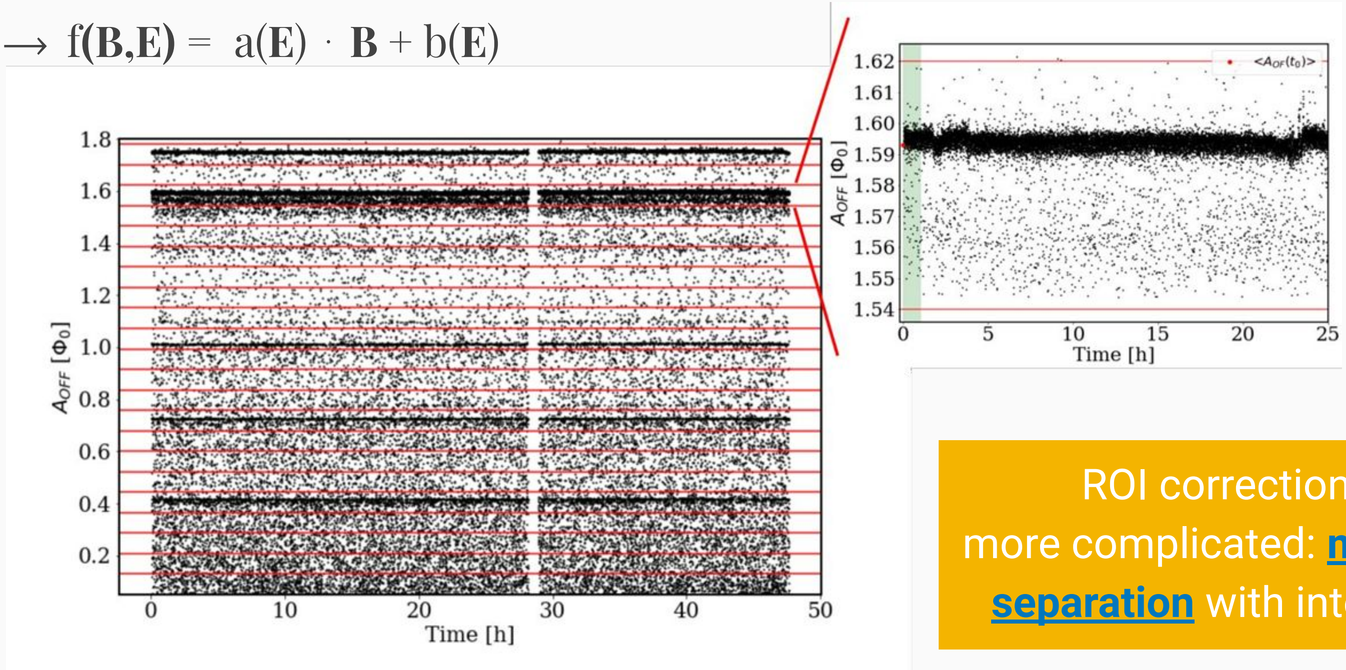




Pulse analysis

- Correction parameters depend on the mean amplitude (⇔ energy) of the considered dataset

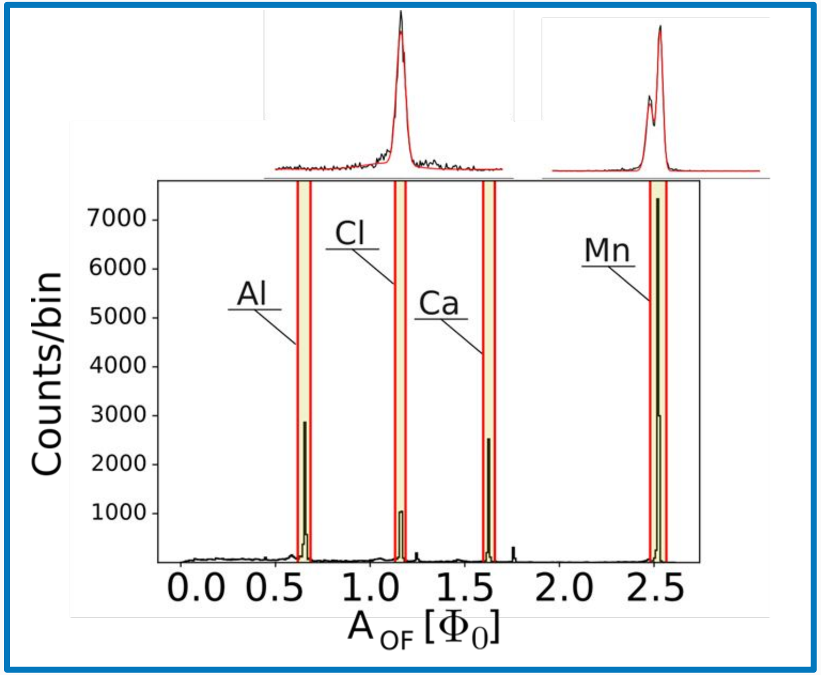
$$f(\mathbf{B}) \longrightarrow f(\mathbf{B},\mathbf{E}) = a(\mathbf{E}) \cdot \mathbf{B} + b(\mathbf{E})$$



ROI correction gets more complicated: **multi-region separation** with interpolation

Pulse analysis

- **Energy calibration** delivered by ^{163}Ho EC characteristic peaks
  - extra-sources near/inside the ROI (Ca, Cl)
- Each TES have its own $\mathbf{E = f(A_{OF})}$
  - aiming at a parallel spectrum analysis
- During test measurements 4 calibration sources in (1,6) keV region



The only calibration-dependent quantity is the Q-value which will be a free-parameter in the spectral fit.

An extremely precise calibration is not required

Slides outline

- The HOLMES experiment
- Data handling
- Pulse analysis
- **Parameter estimation**
  - **Bayesian approach**
- Background rejection

Parameter estimation

- Very useful application
  - extracting quantities from a fit
- Frequentist vs Bayesian
  - same performances in simple problems
  - the latter provides a more natural involvement of systematic errors
  - priors have an ambiguous role
  - ✖ fixed parameter's value
  - ✔ updating parameter's distribution (to sample from)

Bayesian approach:

posterior = $\dfrac{\text{prior} \times \text{Likelihood}}{\text{model evidence}}$

prior: precedent informations

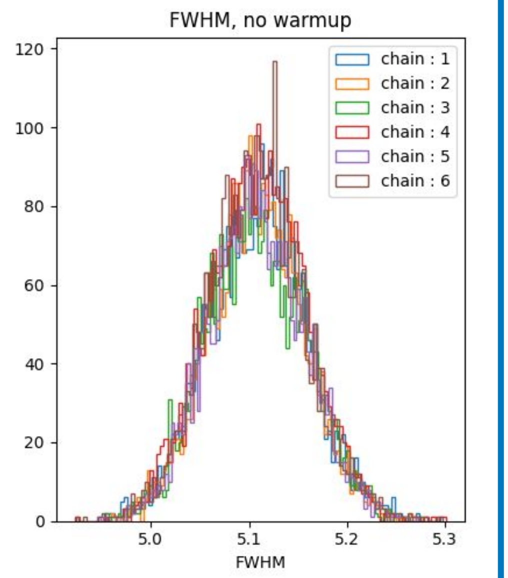
Likelihood: fitting the data to the model

posterior → model evidence: marginalized Likelihood (normaliz. factor difficult to compute)

Parameter estimation

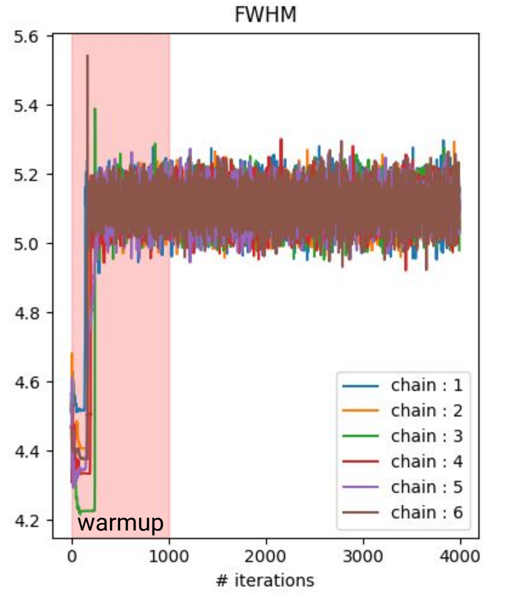
- [Markov Chain](#) generation
  - Sequence of states (=parameter's values) tending to a stationary distribution

- [Markov Chain Monte Carlo](#)
  - Computing posterior starting from a data-set, its likelihood model and all the parameters priors



- [Stan](#) is a software for bayesian inference that exploits MCMC

- It probes the parameters space looking for high-probability density regions



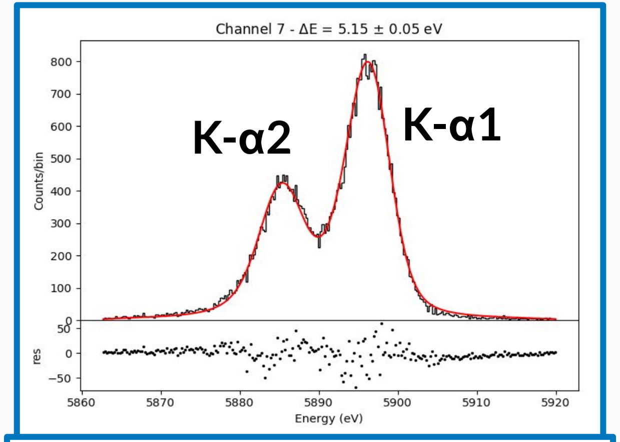
Parameter estimation

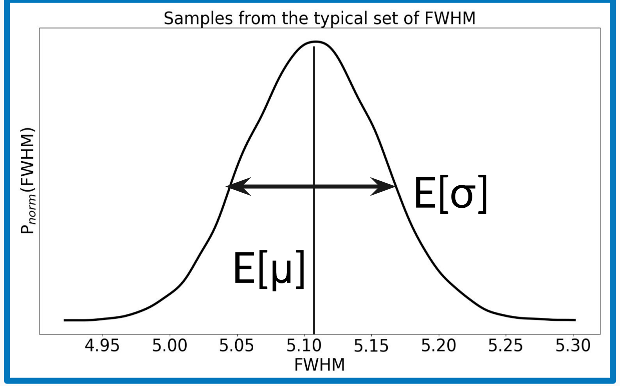
- Frequentist vs Bayesian application in a **FWHM estimation** from a calibration peak (^{55}Mn @ 5.895 keV)

- Comparable outcomes:

Frequentist	5.15 ± 0.05 eV
Bayesian	E[μ] = 5.11 eV E[σ] = 0.05 eV

- Future application on ^{163}Ho more complex spectrum





Slides outline

- The HOLMES experiment
- Data handling
- Pulse analysis
- Parameter estimation
- **Background rejection**
    - **pile-up discrimination algorithms**

Background rejection

HOLMES background sources	
unresolved pile-up (typically on the rise)	main contribution considering ~100 Bq/detector
^{166m}Ho β-decay	inside the detectors, low Q-value
natural radioactivity	Monte Carlo studies
cosmic rays	studied with plastic scintillators

Methods for the undesired events discrimination:

- Wiener filter
- DSVP algorithm

A reduction of the pile-up contribution implies an improvement of our time resolution and, in particular, of our m_ν sensitivity.

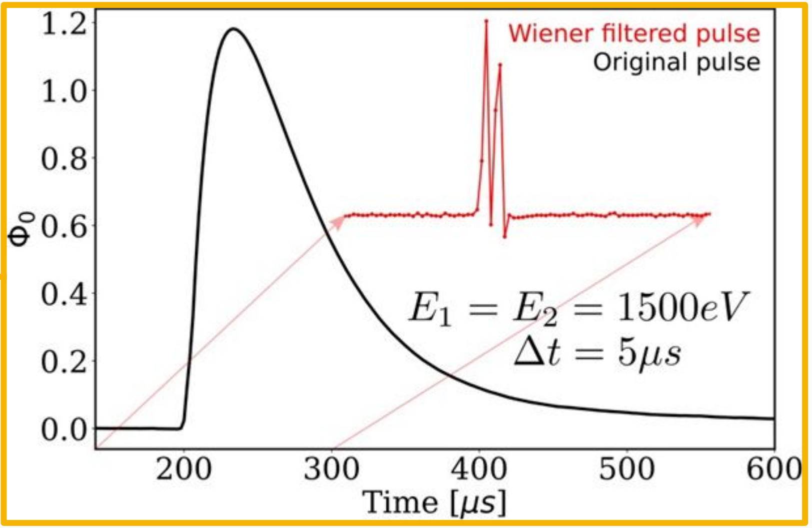
Background rejection

Wiener filter

- Each event is filtered in order to **recover the time profile of the energy deposition**
  $\longrightarrow$ the detector response is deconvolved

Single event $\longrightarrow$ single delta pulse

Pile-up event $\longrightarrow$ multi-delta/broadened delta pulse

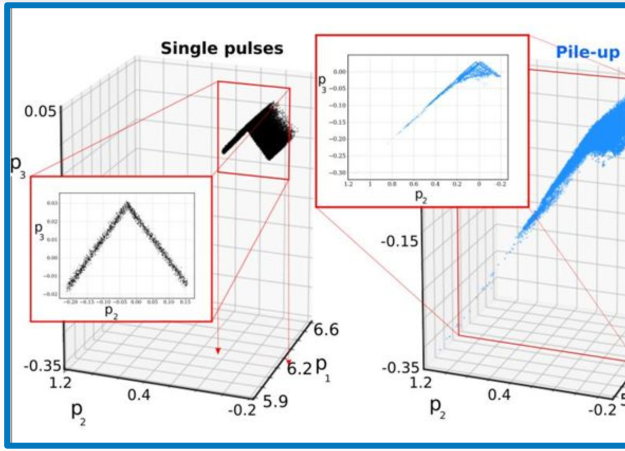


- discrimination through **Wiener shape parameters**: delta width at a given height, delta points above the latter and delta maximum

Background rejection

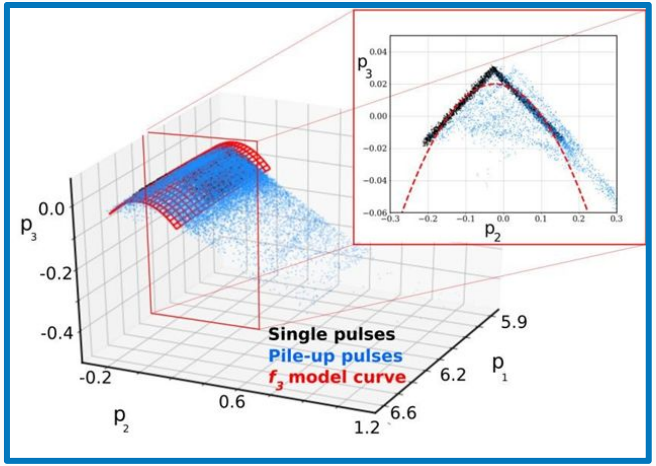
DSVP (Discrimination through Singular Vectors Projections)

- Unsupervised learning technique that discriminate pulses **looking at the average data 'morphology'**
  - a data-set with $N_{good} >> N_{bad}$ is required to create a reduced parameter space



- Iterative procedure that
  - finds discrimination (hyper-surfaces) thresholds
  - removes events different from the average

- More on this technique is presented [Here].



Background rejection

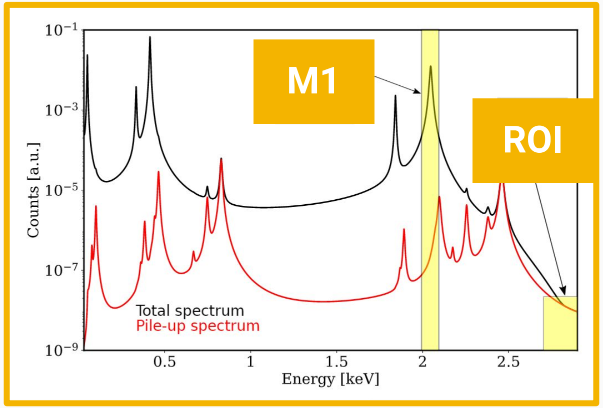
A promising strategy:

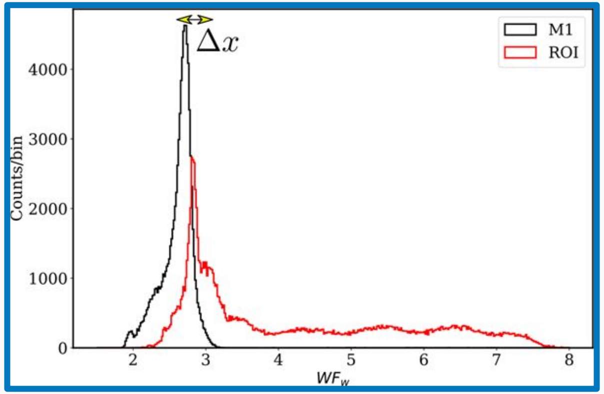
- Using the Wiener transfer function (computed @ M1 EC-peak) for an initial cleaning
  - to reach the $N_{good} > N_{bad}$ condition @ ROI
  - cuts on Wiener **width shape parameter** ($\mathbf{WF_W}$)

$$\mathbf{WF_W}^{\text{max.min}}(\text{ROI}) = \mathbf{WF_W}^{\text{max.min}}(\text{M1}) + \mathbf{\Delta x}$$

- Applying DSVP on the filtered ROI data-set

Simulating datasets…



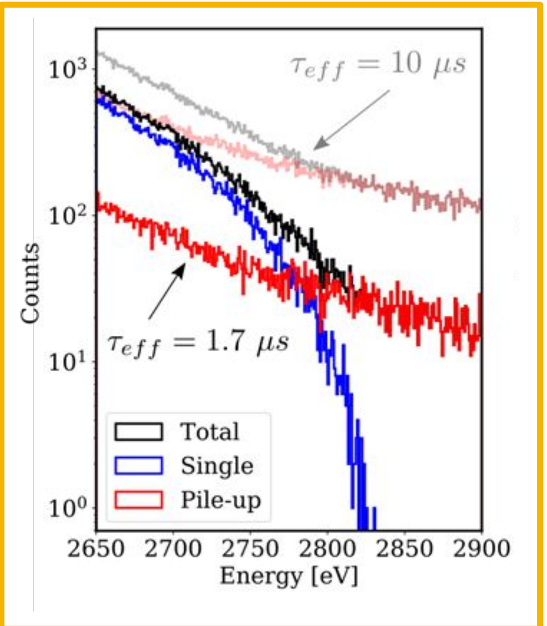


Background rejection

To evaluate the behavior of our algorithms, we define an **effective time resolution** $\boldsymbol{\tau}_{\mathbf{eff}}$:

$$\boldsymbol{\tau}_{\mathbf{eff}} = (f_{pp}|_{after} \,/\, f_{pp}|_{before}) \cdot \delta\tau$$

Simulations assume the first level data reduction reach a time resolution ($\delta\tau$) of 10μs, corresponding to f_{pp}~2 (@ **300Hz/TES**)



- **Inside the ROI**:
  - $\boldsymbol{\tau}_{\mathbf{eff}}$ after Wiener ~ 3μs (f_{pp}~0.6)
  - $\boldsymbol{\tau}_{\mathbf{eff}}$ after Wiener + DSVP ~ 1.5μs (f_{pp}~0.3)
- The pile-up fraction **over the entire EC spectrum** decreases from 10^{-3} to 10^{-4}

HOLMES analysis focuses on the

^{163}Ho EC spectrum reconstruction

that can lead to a m_ν upper

boundary assessment.

Conclusions

Conclusions

HOLMES analysis focuses on the **^{163}Ho EC spectrum reconstruction** that can lead to a m_ν upper boundary assessment.

Clustering algorithms for the **first level data reduction** are under development.

Clustering algorithms for the **first level data reduction** are under development.

HOLMES analysis focuses on the **^{163}Ho EC spectrum reconstruction** that can lead to a m_ν upper boundary assessment.

The energy is well-estimated, **the intrinsic resolution of our TESs is recovered**.

Conclusions

Conclusions

HOLMES analysis focuses on the **^{163}Ho EC spectrum reconstruction** that can lead to a m_ν upper boundary assessment.

The energy is well-estimated, **the intrinsic resolution of our TESs is recovered**.

Clustering algorithms for the **first level data reduction** are under development.

A bayesian tool for the **parameter estimation** is under testing.

Conclusions

HOLMES analysis focuses on the ^{163}**Ho EC spectrum reconstruction** that can lead to a m_ν upper boundary assessment.

Clustering algorithms for the **first level data reduction** are under development.

The energy is well-estimated, **the intrinsic resolution of our TESs is recovered**.

A bayesian tool for the **parameter estimation** is under testing.

Applying an iterative DSVP routine on a Wiener filtered dataset can be a good strategy to **reduce the pile-up fraction**.

Conclusions

HOLMES analysis focuses on the **^{163}Ho EC spectrum reconstruction** that can lead to a m_ν upper boundary assessment.

Clustering algorithms for the **first level data reduction** are under development.

The energy is well-estimated, **the intrinsic resolution of our TESs is recovered**.

A bayesian tool for the **parameter estimation** is under testing.

A novel technique for reducing the dead time of the experiment exploiting the **matrix optimum filter** is under study [**Here**].

Applying an iterative DSVP routine on a Wiener filtered dataset can be a good strategy to **reduce the pile-up fraction**.

Thanks for the attention!

ν

$\bar{\nu}$