Cluster counting techniques
The beam test and the preliminary results

Federica Cuna on behalf of FCC and
Cremlin+ collaboration

Rd_FCC collaboration meeting
Dec. 15, 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 871072

Outline

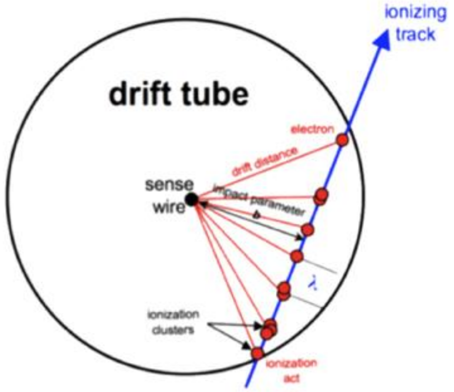
- **Simulation of the cluster counting techniques**
- **Algorithms for cluster reconstruction:**
  - **Louvain algorithm**
  - **Novosibirsk algorithm**
  - **Beijing algorithm**
  - **Lecce algorithm**

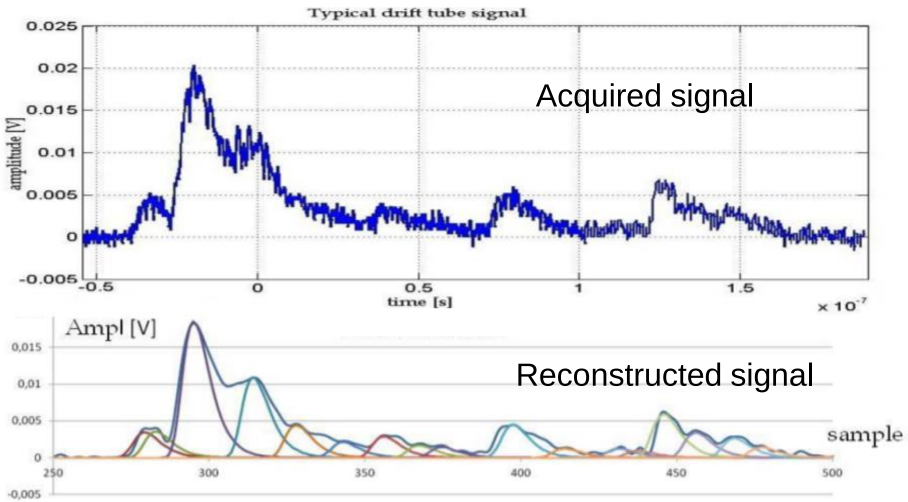
Cluster counting for particle identification

Using the information about energy deposit by a track in a gaseous detector, particle identification can be performed.
The large and intrinsic uncertainties in the total energy deposition represent a limit to the particle separation capabilities.

Cluster counting technique can improve the particle separation capabilities.

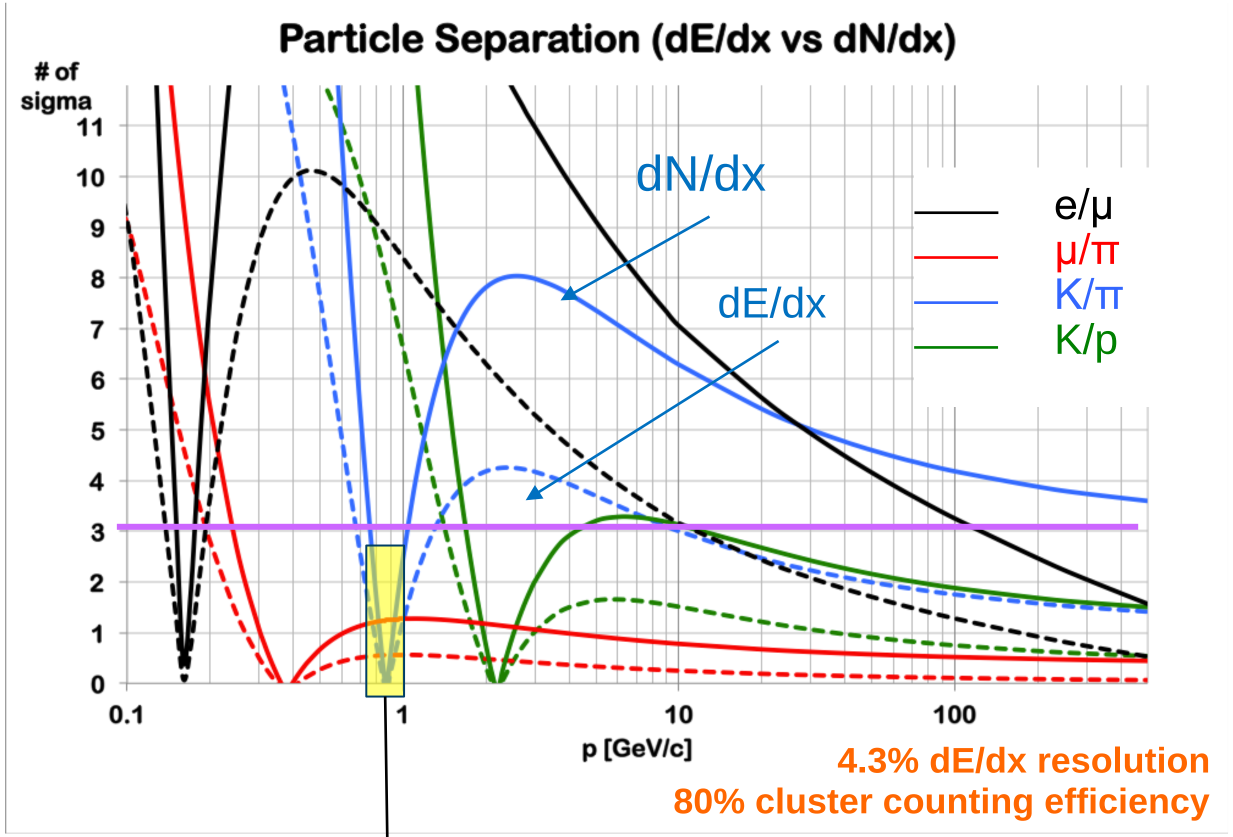
The method consists in singling out, in ever recorded detector signal, the isolated structures related to the arrival on the anode wire of the electrons belonging to a single ionization act (dN/dx).

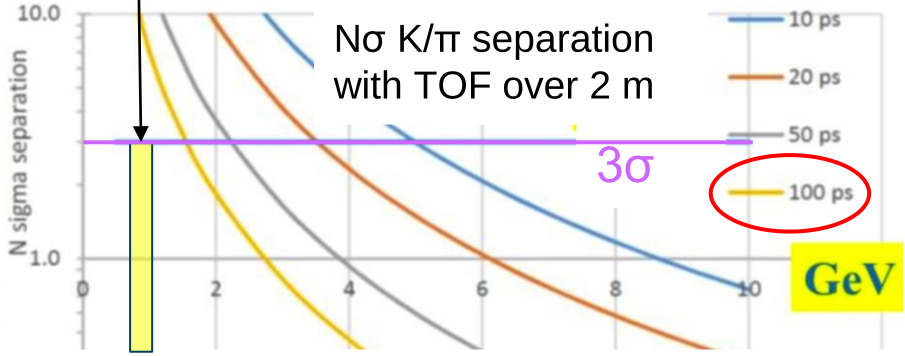




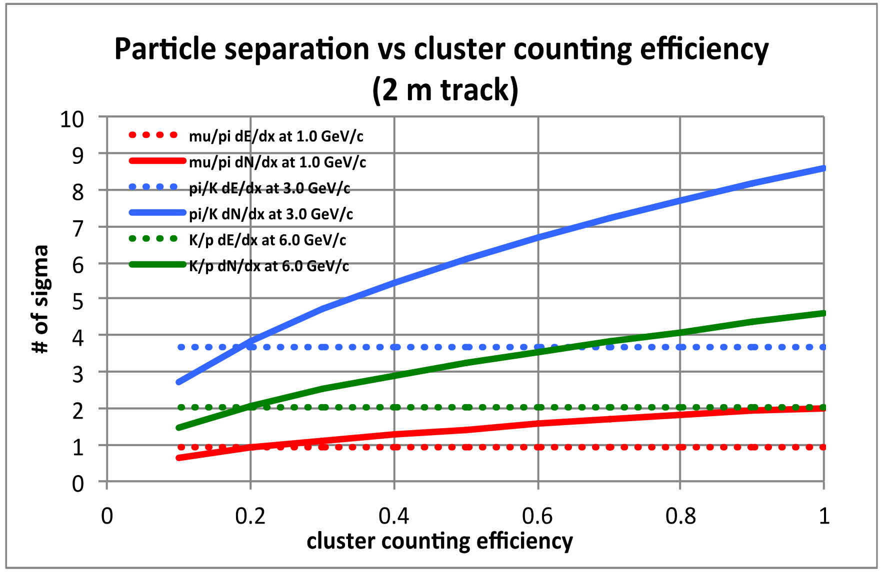
dE/dx	*dN_{cl}/dx*
Truncated mean cut (70-80%) reduces the amount of collected information n ≈ 100 and a 2m track at 1 atm give **σ ≈ 4.3%**	δ_{cl} = 12.5/cm for He/iC_4H_{10}=90/10 and a 2m track give **σ ≈ 2.0%**

Cluster counting for particle identification: expected performance





- 80% cluster counting efficiency.
- Expected excellent K/π separation over the entire range except 0.85<p<1.05 GeV (blue lines)
- Could recover with timing layer



Analytic evaluation, prof F.Grancagnolo
To be checked with test beam and simulations

Cluster counting for particle identification: simulation results

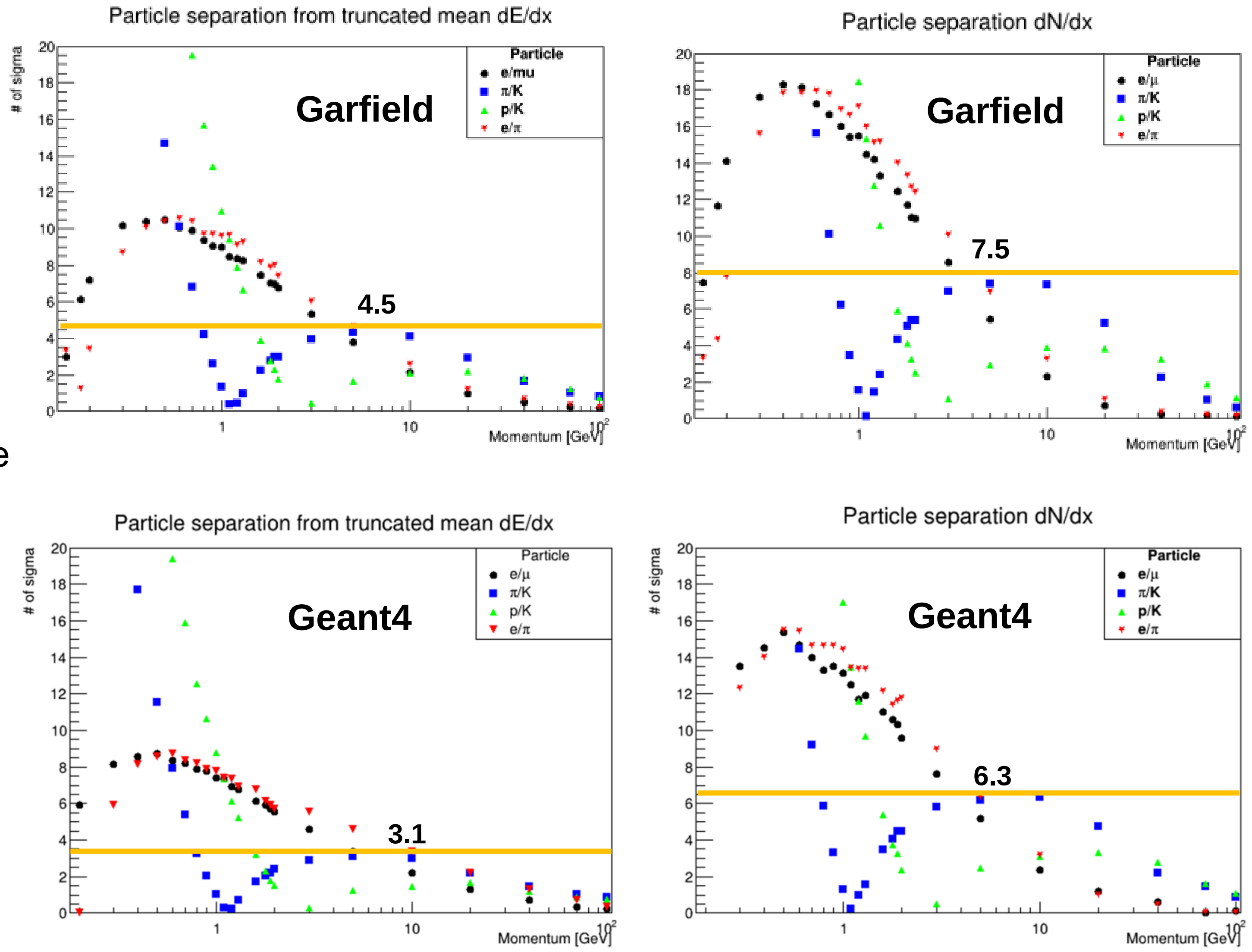
A simulation of the ionization process in 1 cm long side cell of 90% He and 10% iC_4H_{10} has been performed in **Garfield++** and **Geant4**.

Geant4 software can simulate in details a full-scale detector, but the fundamental properties and the performances of the sensible elements have to be parameterized or an “ad hoc” physics model has to be implemented.

Three different algorithms have been implemented to simulate in Geant4*, in a fast and convenient way*, the number of clusters and clusters size distributions, using the energy deposit provided by Geant4.

The simulations confirm the prediction!

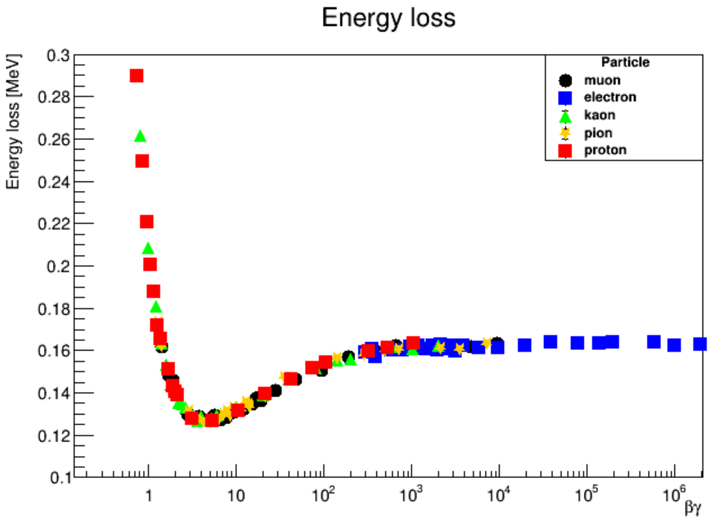
But…

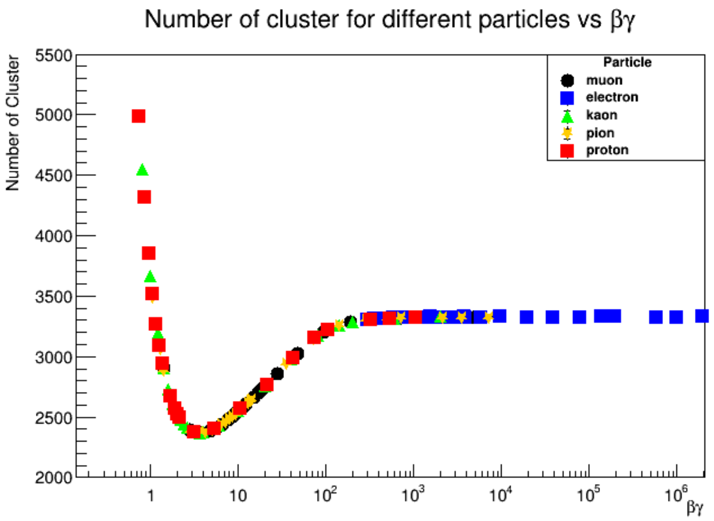


We are assuming a cluster counting efficiency of 100%.

Cluster counting for particle identification: TEST BEAM

- Lack of experimental data on cluster density and cluster population for He based gas, particularly in the relativistic rise region to compare predictions.
- Despite the fact that the Heed model in GEANT4 reproduces reasonably well the Garfield predictions, why particle separation, both with dE/dx and with dNcl/dx, in GEANT4 is considerably worse than in Garfield?
- Despite a higher value of the dNcl/dx Fermi plateau with respect to dE/dx, why this is reached at lower values of βγ with a steeper slope?





These questions are crucial for establishing the particle identification performance at FCCee, CEPC and SCTF

The only way to solve these issues is an experimental measurement!

Cluster counting for particle identification: TEST BEAM

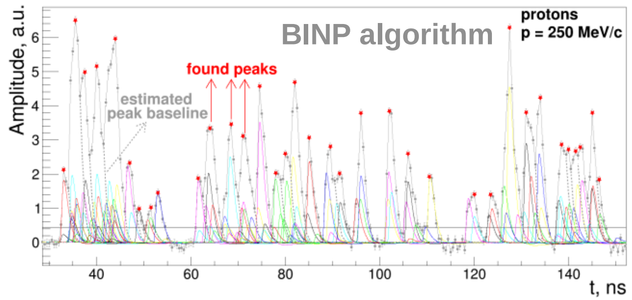
Goals

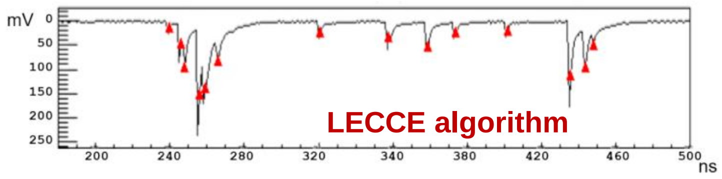
- **Demonstrate the ability to count clusters:**

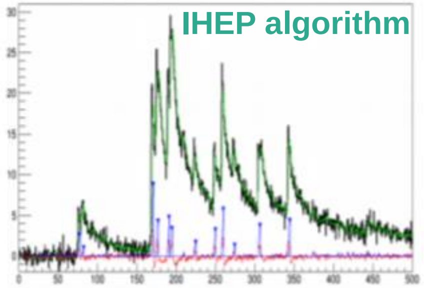
  at a fixed βγ (e.g. muons at a fixed momentum) count the clusters by changing

  - **the cell size (1 – 3 cm)**
  - **the track angle (1- 6 cm)**
  - **the gas mixture (90/10: 12 cl/cm, 80/20: 20 cl/cm)**

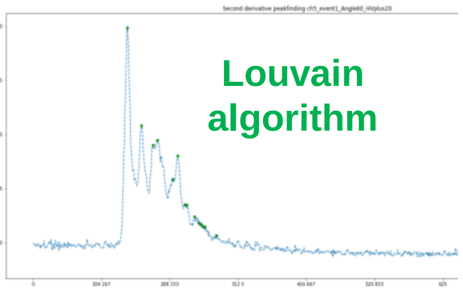
- **Establish the limiting parameters for an efficient cluster counting:**
  - **cluster density as a function of impact parameter**
  - **space charge (by changing gas gain, sense wire diameter, track angle)**
  - **gas gain stability**







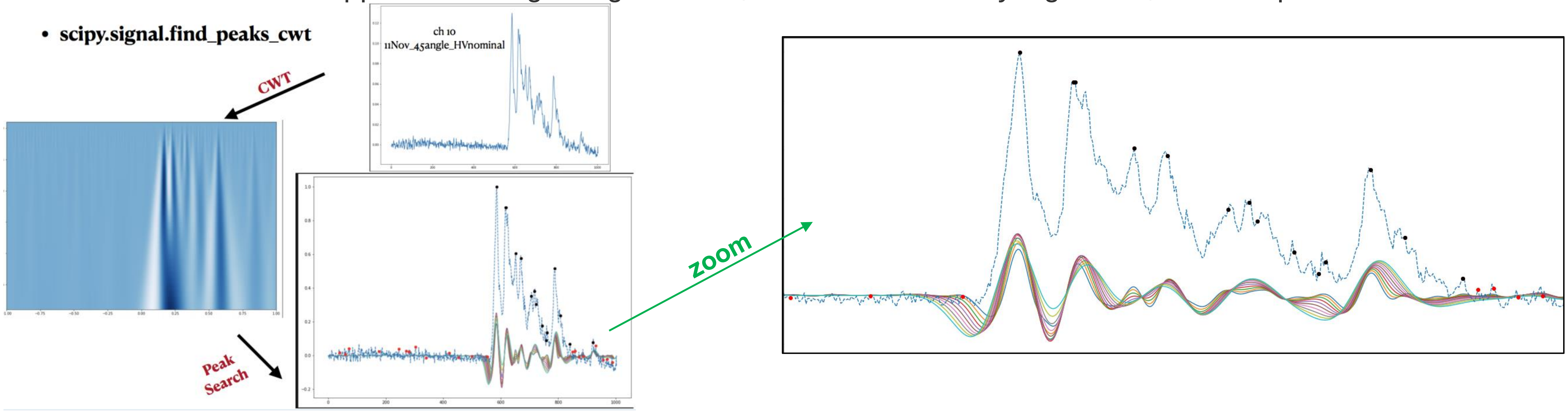
- In optimal configuration, measure the relativistic rise as a function of βγ, both in dE/dx and in dNcl/dx, by scanning the muon momentum from the lowest to the highest value (from a few GeV/c to about 250 GeV/c at CERN/H8).
- Use the experimental results to fine tune the predictions on performance of cluster counting for flavor physics and for jet flavor tagging both in fast and in full simulation.



Louvain algorithm (courtesy of Claudio Caputo)

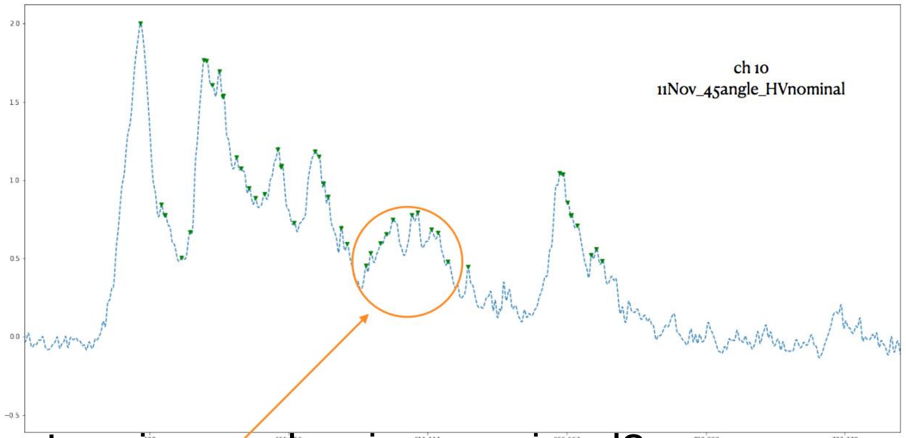
First attempt

The general approach is to smooth *vector* by convolving it with *wavelet*.
Relative maxima which appear at enough length scales, and with sufficiently high SNR, are accepted.



Second attempt

The approach follows the second derivative by implementing the Savitzky-Golay filter



Is noise overlapping our signal?

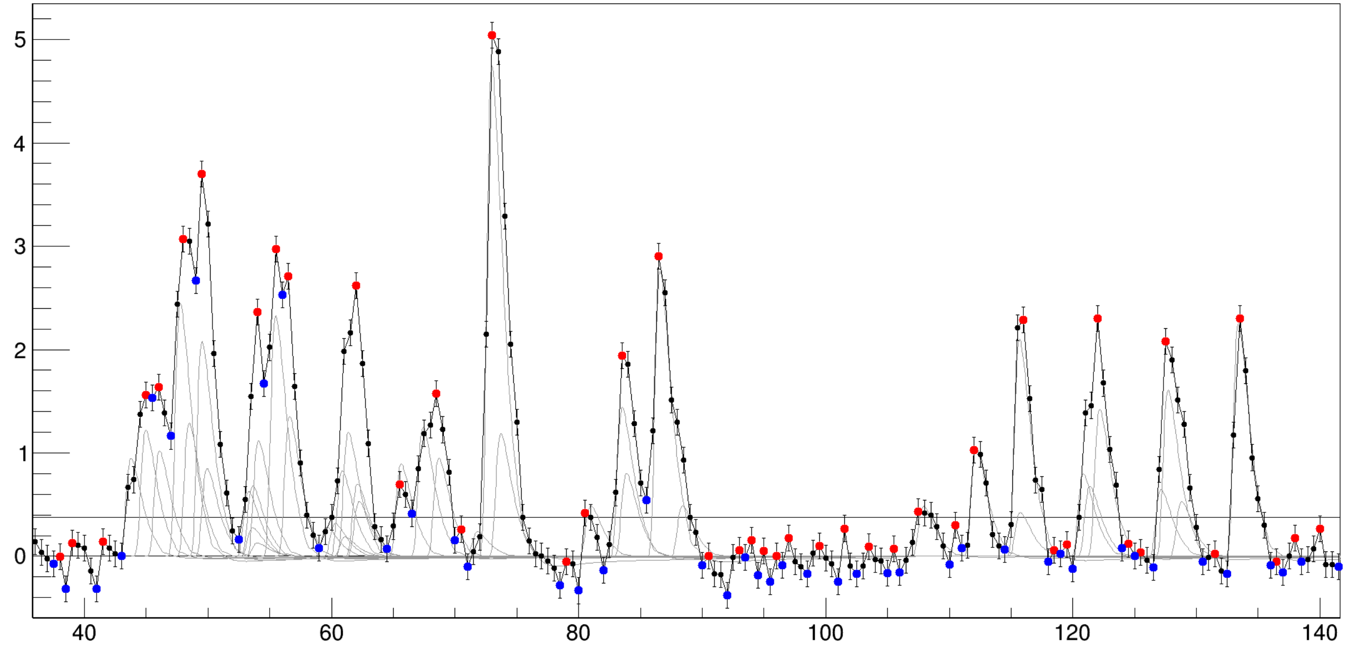
The method to associate peaks to cluster has not been implemented.

Novosibirsk algorithm (courtesy of Slava Ivanov)

The algorithm is based on the dynamic estimation of the **baseline level.**

The waveform contains local minimums and local maximums

- **Each** waveform segment “loc.min.–loc.max.–loc.min” is considered as *peak candidate*
- Peak candidate is identified as *real peak* if it satisfies a *quality criterion.* Currently one peak candidate can give only one real peak
- To calculate the peak quality correctly, one should account for the **baseline** shift, caused by the previous peaks
- Thus, for each peak candidate we should estimate the **baseline** it resides on (“running baseline”)

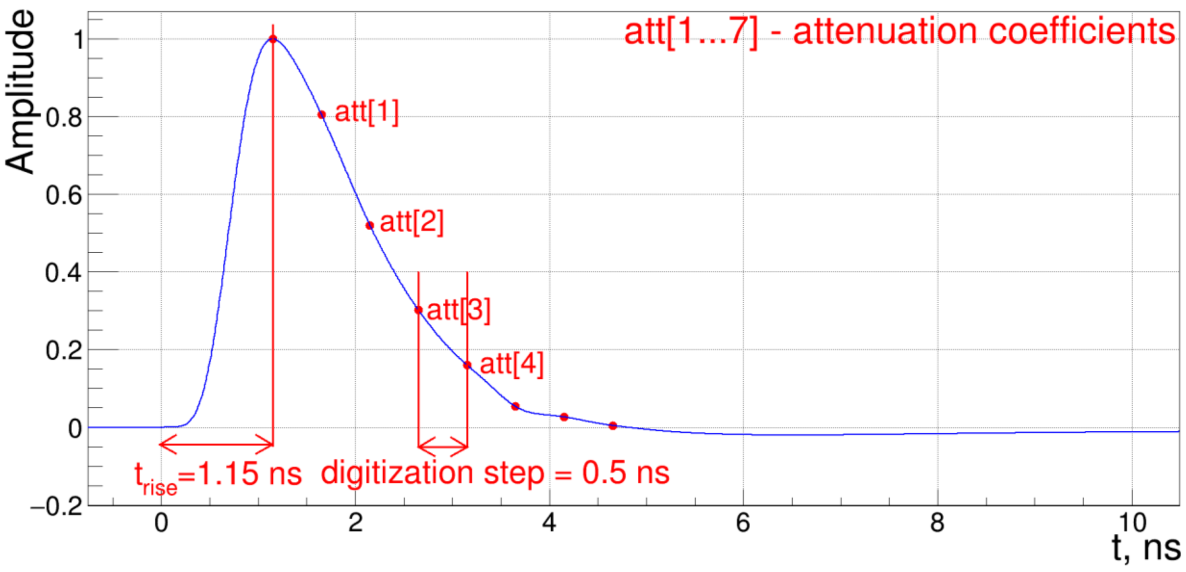


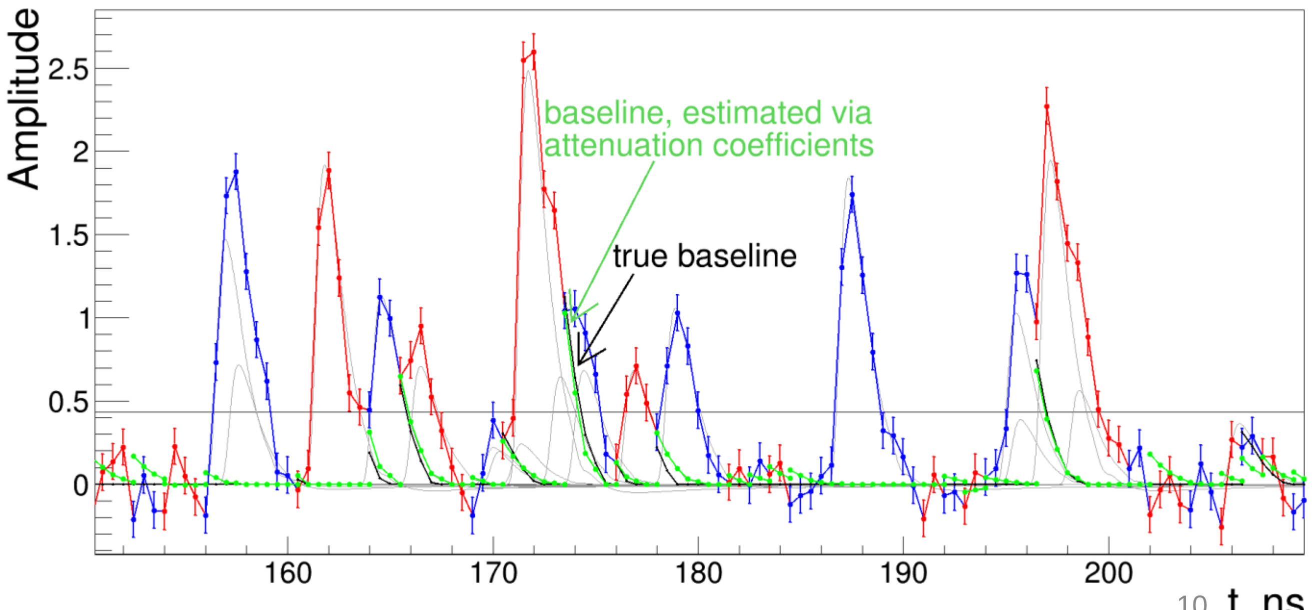
Baseline estimation

To estimate the **baseline** of the current peak candidate the previous 3 peak candidates have to be considered as a real single signal peaks. Their contribution to the baseline is calculated using their amplitudes at their peaks(with subtraction of their baselines) and the attenuation coefficients, calculated for the 7 steps after the signal maximum.

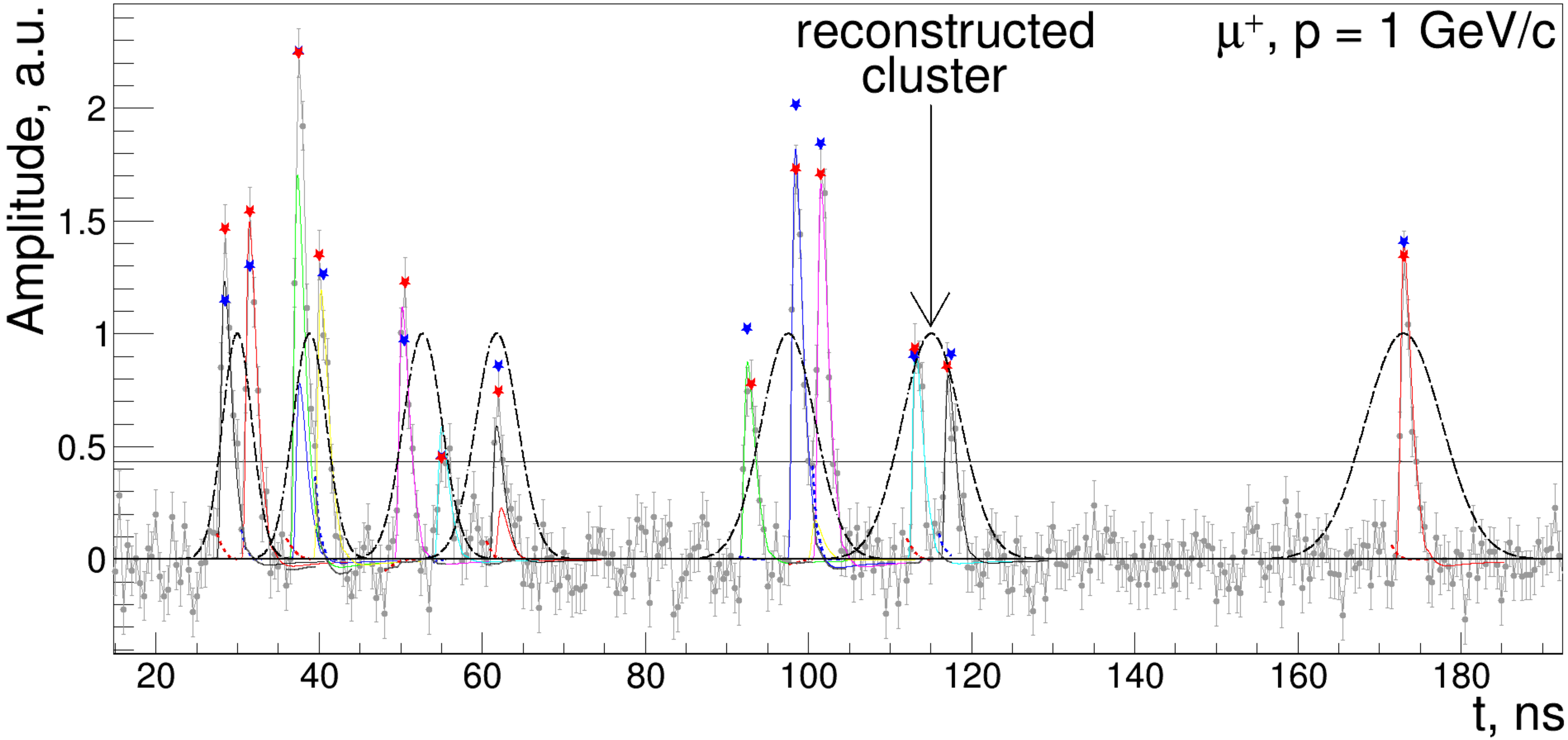
Often due of the overlap of many peaks the signal shape gets deteriorated and this leads to the wrong baseline estimation (as a rule-underestimation).

To overcome this problem the **baseline** has been scaled with the *coefficient to make it equal to the amplitude at the first point of current peak candidate (=first local minimum)*.





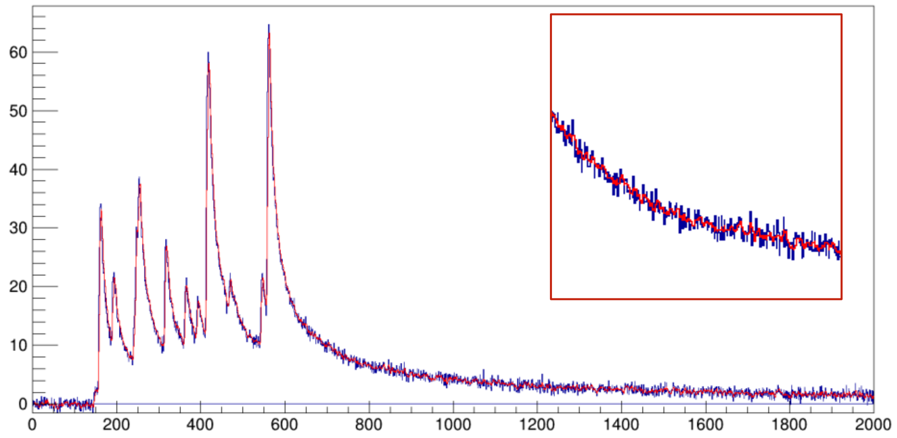
Peaks clusterization algorithm for muons



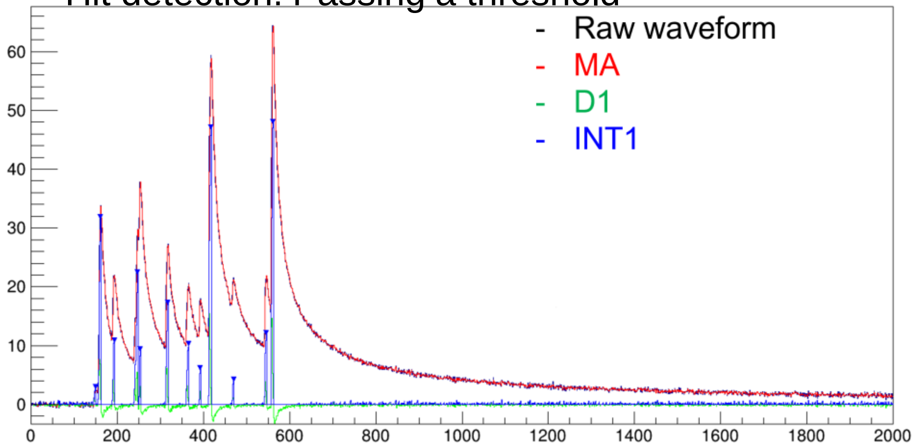
More details on Slava Ivanov talk tomorrow!

Beijing algorithm (courtesy of Guang Zhao)

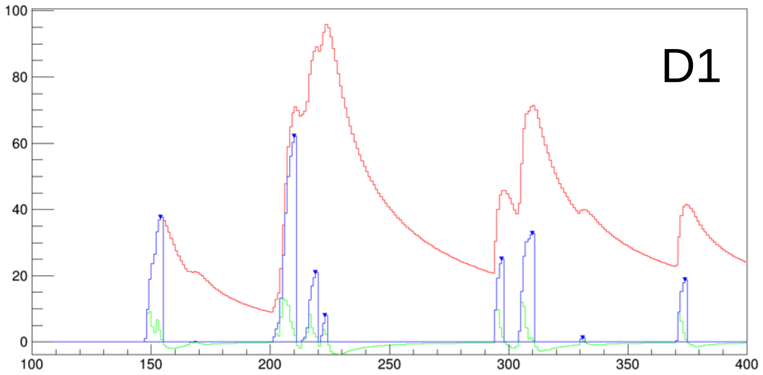
- Noise reduction
  - Filter out high frequency noises in the waveforms in order to improve the S/N ratio
  - Moving average: $\mathbf{MA[i]} = \frac{1}{M} \times \sum_{k=0}^{K<M} S[i-k]$

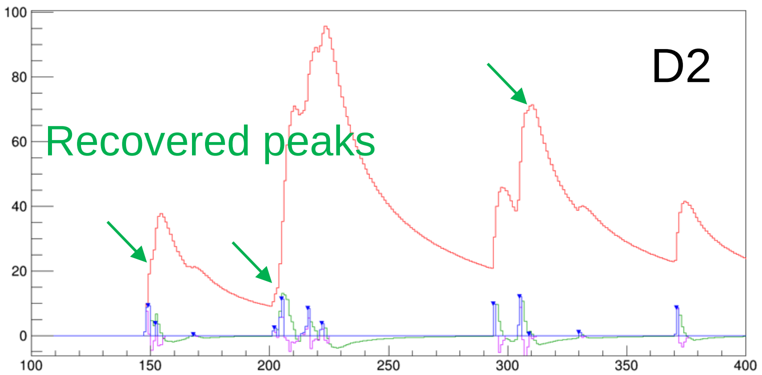


- First derivative and integration
  - First derivative (D1): D1[i] = MA[i] – MA[i – 1]
  - Integration on the positive D1 (**INT1**): recover the rising edge and removing falling edge
  - Hit detection: Passing a threshold



- Second derivative and integration: recover pile-up peaks on the rising edge
  - Second derivative (**D2**): D2[i] = INT1[i] – INT1[i – 1]
  - Integration on the positive D2 (**INT2**)
  - Hit detection: **Passing a threshold**





BEIJING ALGORITHM OPTIMIZATION

- **Detection threshold**

  Minimize the fake rate
- **Moving average size**

  Maximize the counting efficiency

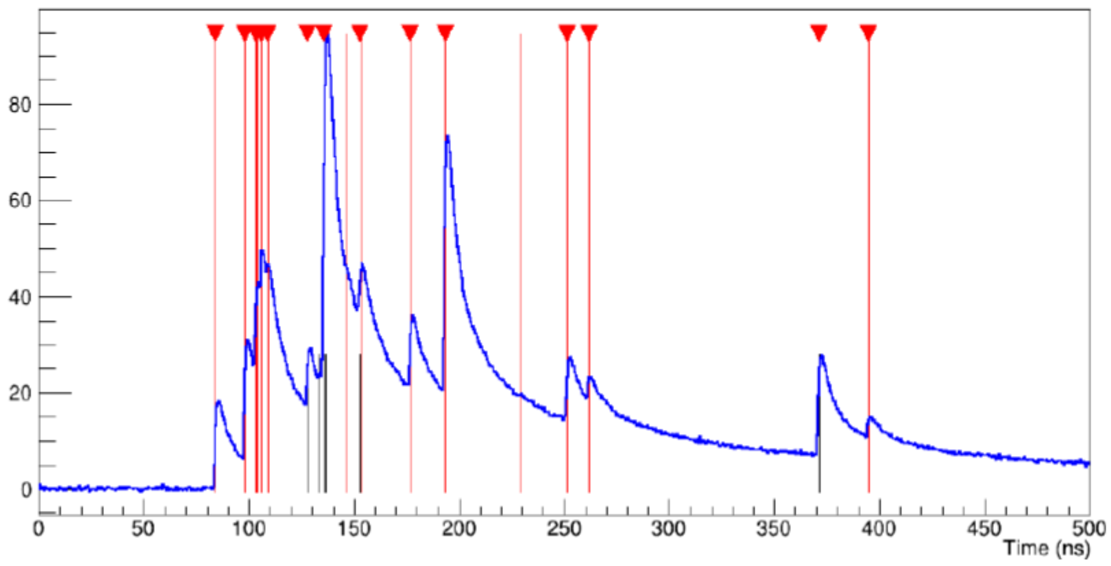
Simulation setup
Gas: 90% He + 10% iC_4H_{10}
Particle: 10 GeV/c pions
Time constant: 1 ns
Noise level: 2%

The MC truth matching algorithm

A detected time is matched to a truth time if

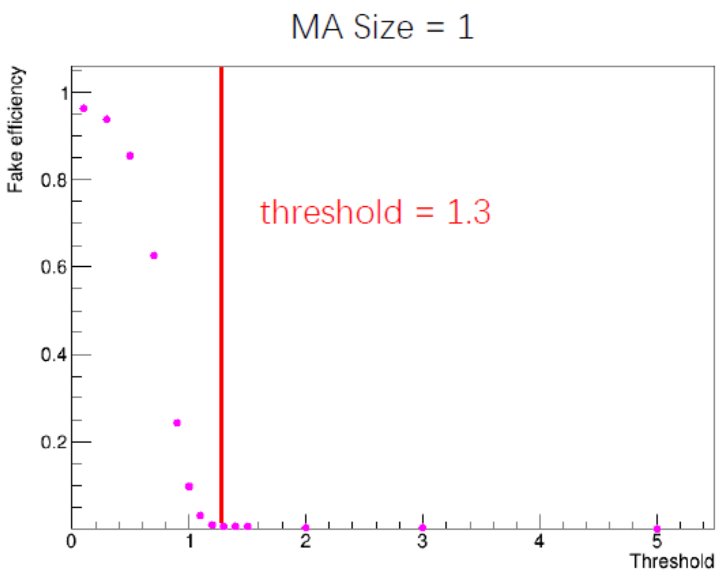
- |Tdet–Ttruth| <2ns
- If the matched truth time is from a primary electron, the detection is defined as a “primary”
- If the matched truth time is from a secondary electron, the detection is defined as a “secondary”
- If a detection is both a primary and secondary, “primary” is set
- Otherwise, “fake” is set

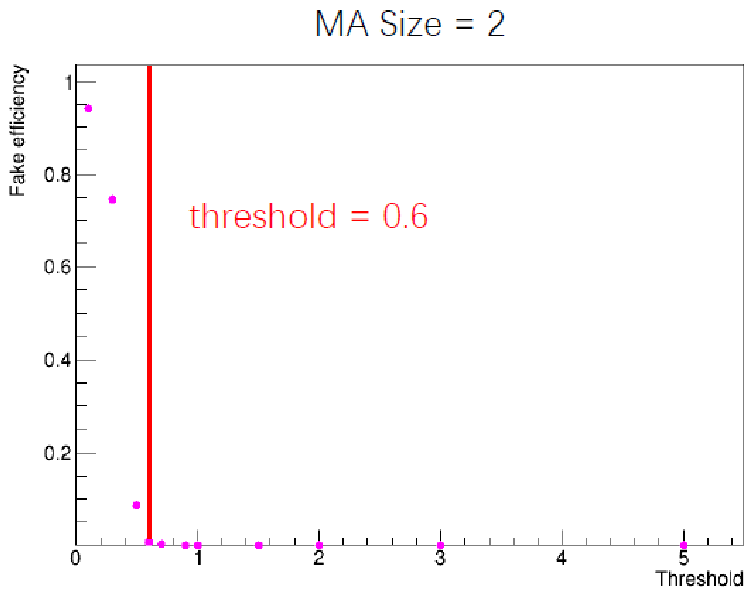
Waveform with “MC truth “ times



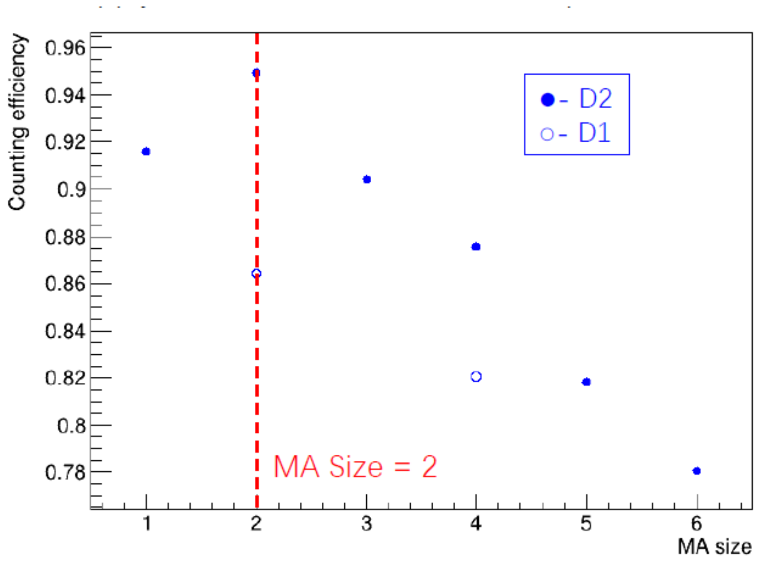
Red line: primaries
-Black line: secondaries
➢ Triangle: detected

Fake rate vs threshold





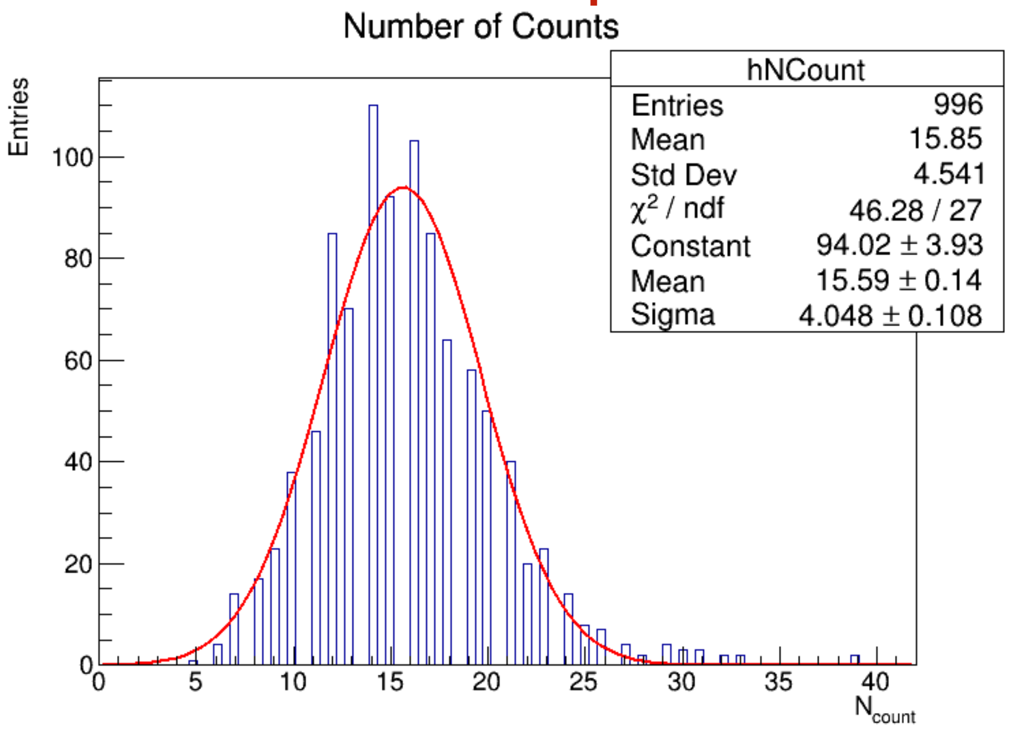
Efficiency vs MA size



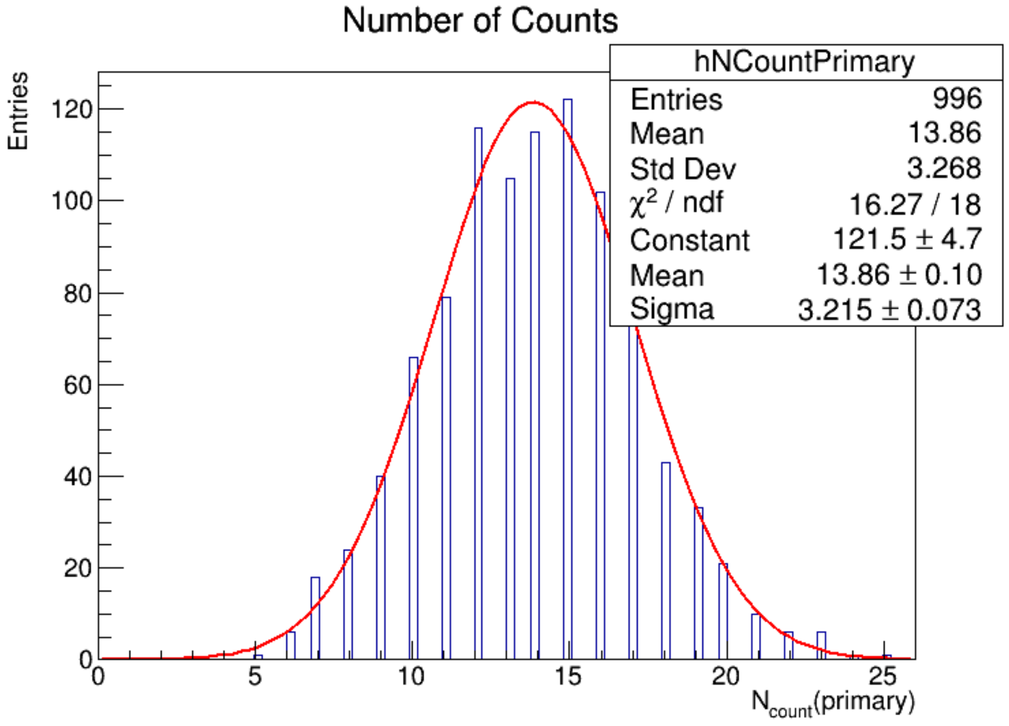
Optimized parameters:
MA = 2
Threshold = 0.6

Number of cluster detected by the algorithm

of detected pulses



of detected pulses (matched to primaries)



- The dN/dx distribution is very like a Gaussian shape
- Only a small portion of counting is related to the secondaries. If this rate is stable, there will be little harm for PID. Need further checks.

The distributions are referred to Monte Carlo studies results

Lecce algorithm

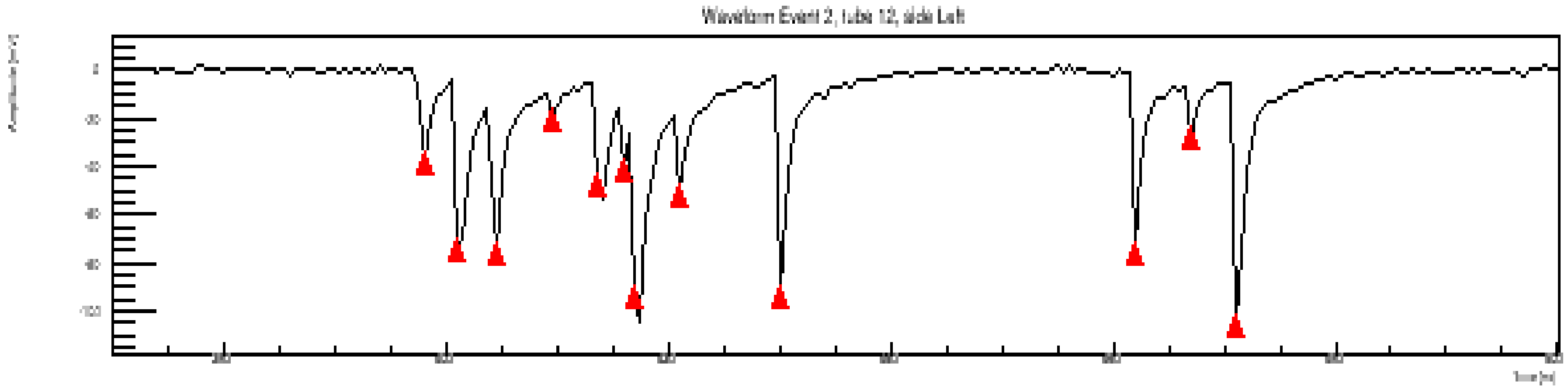
A simple peak finder algorithm has been implemented based on the first and second derivative of the digitized signal function f. For each bin i is defined:

$$f'(i) = \frac{f(i) - \bar{f}(i - \Delta b)}{\Delta b} \qquad\qquad f''(i) = f'(i) - f'(i-1)$$

Δb being the number of bins (signal rise time) over which the average value of f is calculated

A peak (assumed to be an ionization electron) is found when *Δf*, *f'* and *f''* are above a threshold level, defined according to the r.m.s. noise of the signal function *f*, and when the time difference with a contiguous peak is larger than the time bin resolution.

Simulated waveform



Cylindrical cell geometry

He-iC_4H_{10}

Gas gain $4x10^5$

FE gain 10

Peak finder performances

The j-th e^- is *recognizable* when :

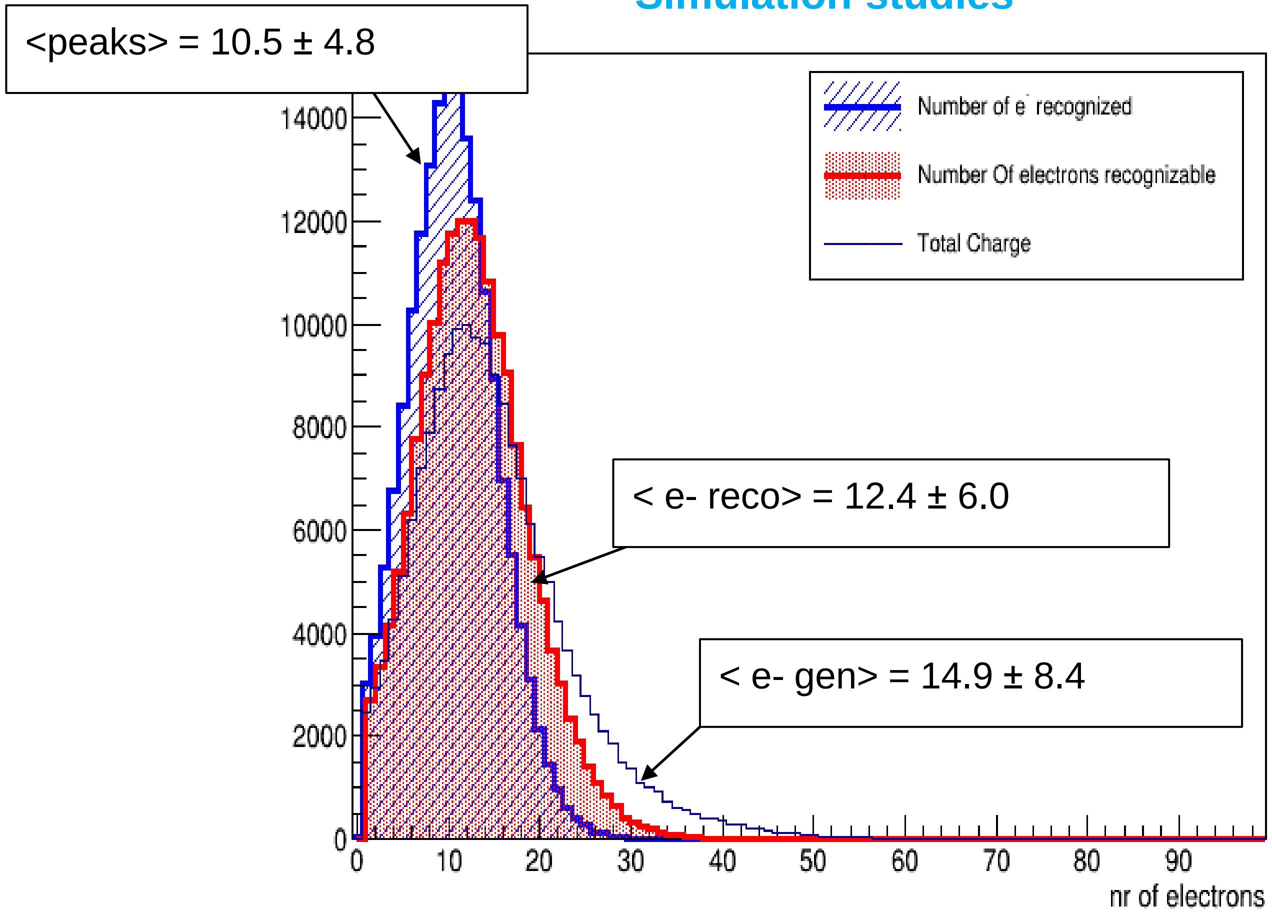
- $t_{j+1} - t_j$ > tcut with tcut 0.6 ps (1.2bin)
- Peak amplitude > threshold

Using these conditions, the nr of skipped electron is 2.5

$$\varepsilon = \frac{e^- \, found}{e^- \, recogn} = 86.7\%$$

$$\varepsilon = \frac{e^- \, found}{e^- \, gen} = 70.5\%$$

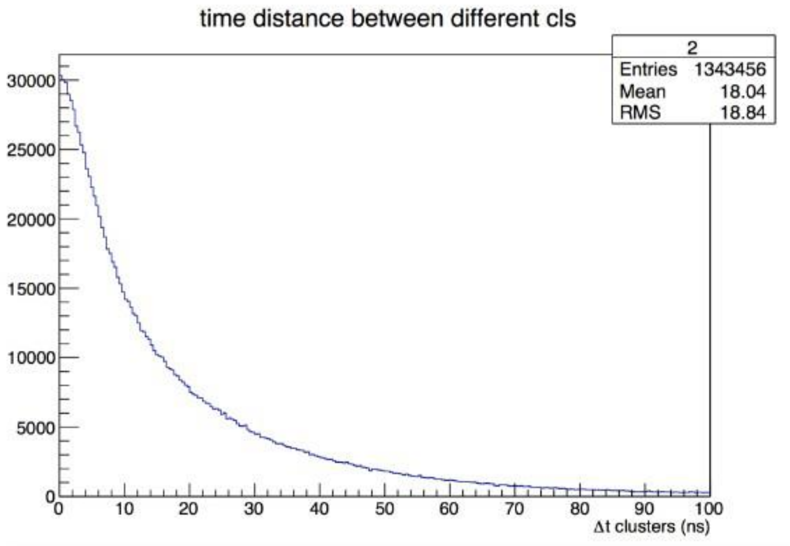
Simulation studies

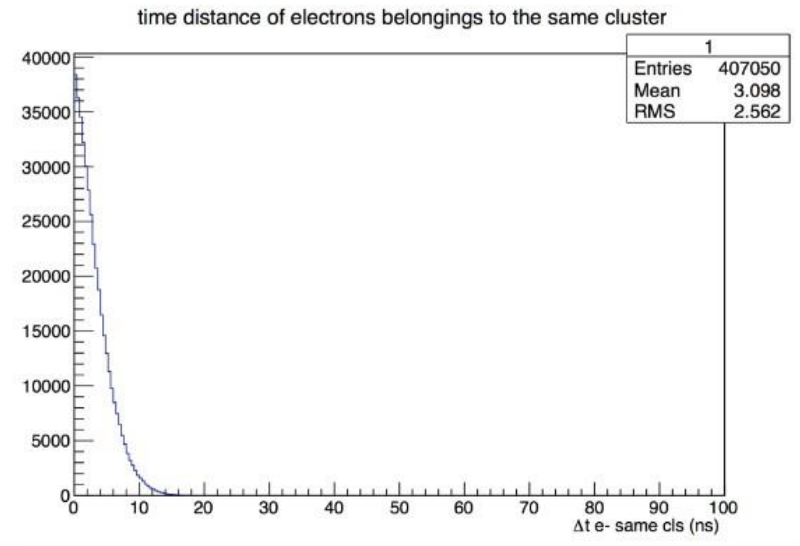


Convert peaks in cluster clusterization

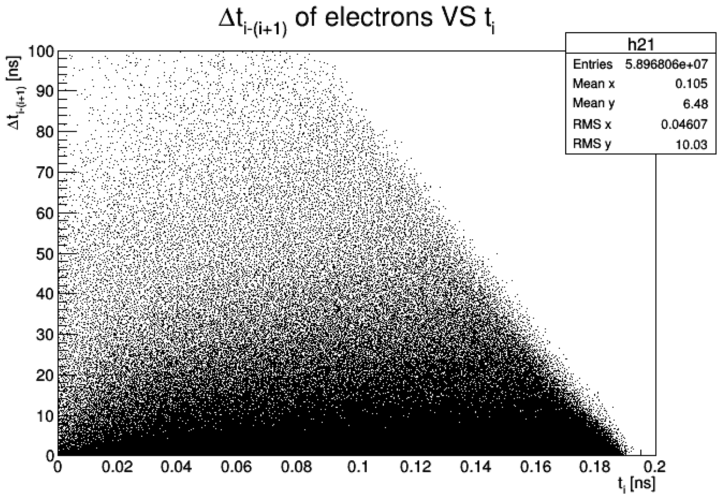
How to convert found peaks in clusters?

- Look to the time difference between electrons belonging to different clusters and those to the same cluster



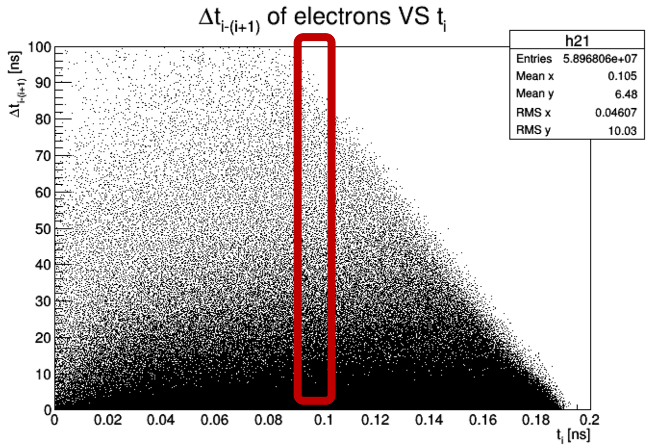


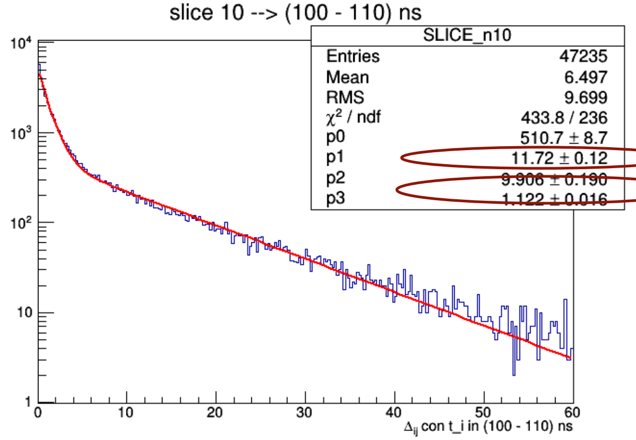
- Event by event plot the difference t_i - t_{i+1} Vs t_i



Convert peaks in cluster clusterization

- Cut into slices along x – axis and fit with a double exponential function



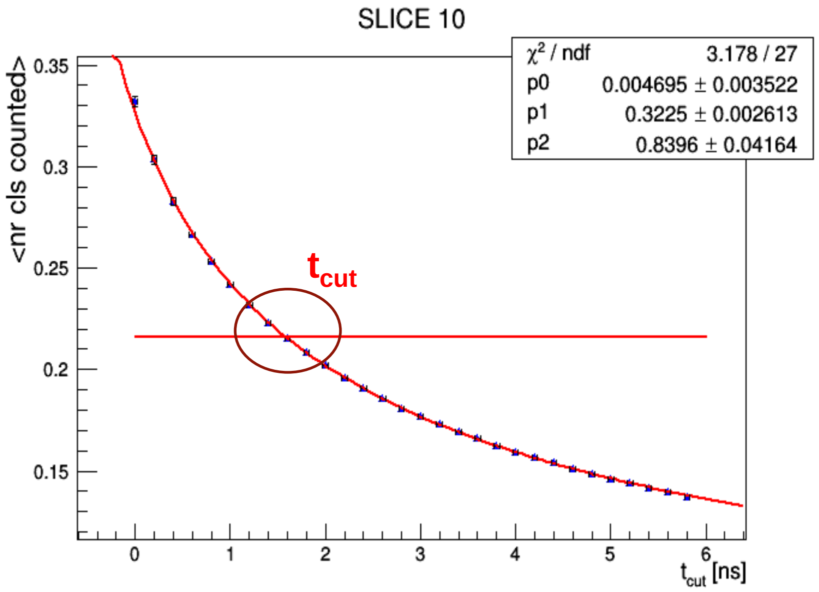


$$f(t) = Ae^{-t/\tau 1} + \mathrm{B}\, e^{-t/\tau 2}$$

τ1 describes time distance for different clusters
τ2 describes time distance for same cluster

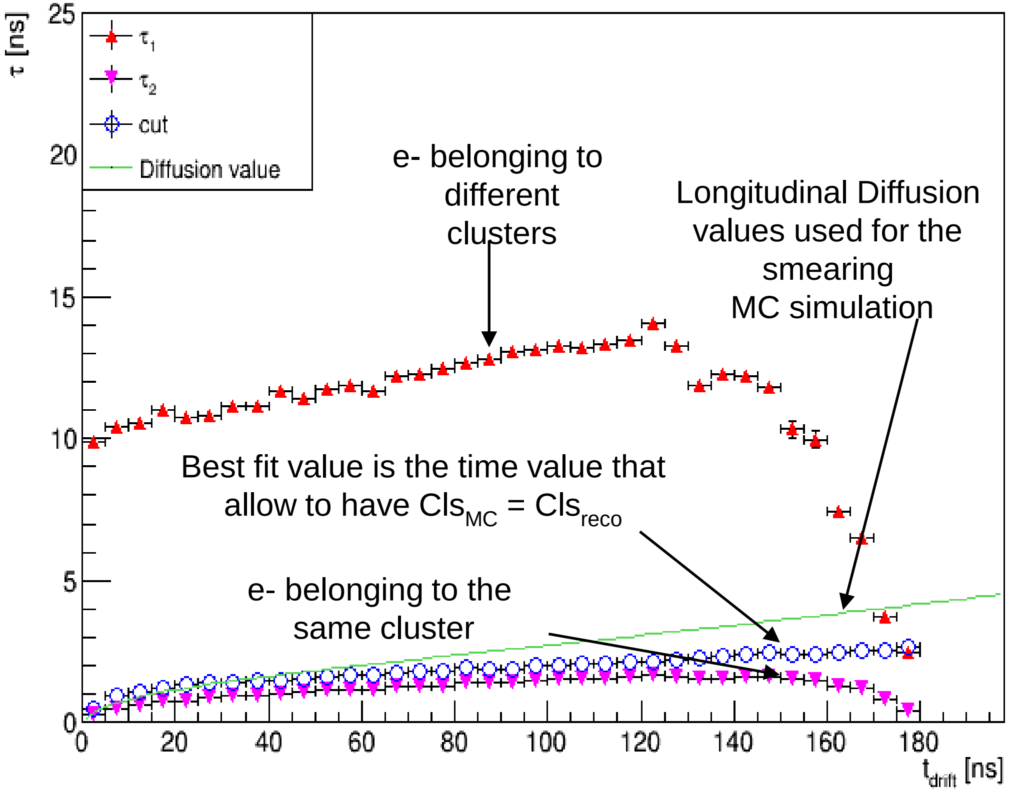
In each slice calculate the right timing cut as the time value that allow to have:

$$\frac{NCls\ counted}{NCls\ MC} = 1$$

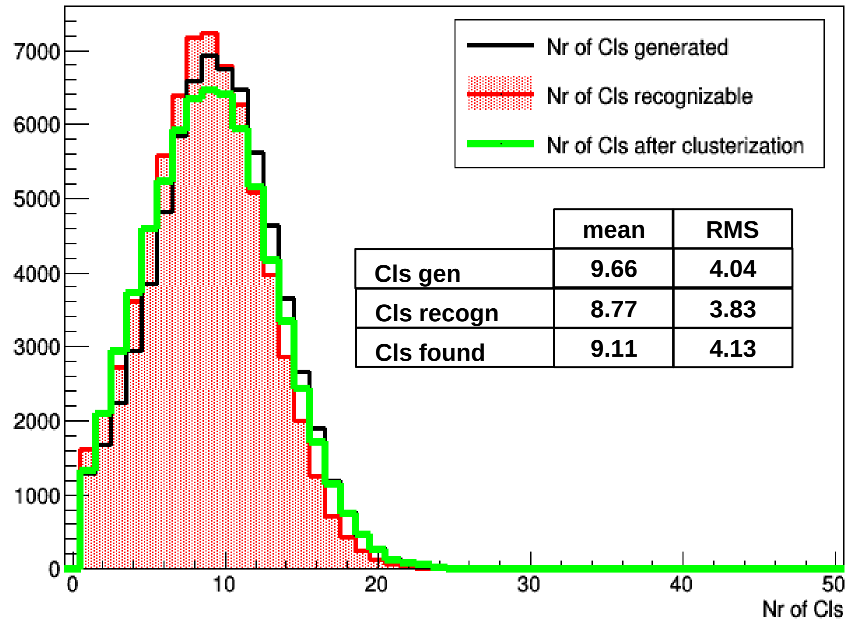


Convert peaks in cluster clusterization

The association of electrons in clusters is based on the time difference between consecutive electrons.
Electrons belonging to same cluster are separated by time differences which are compatible with single electron diffusion.



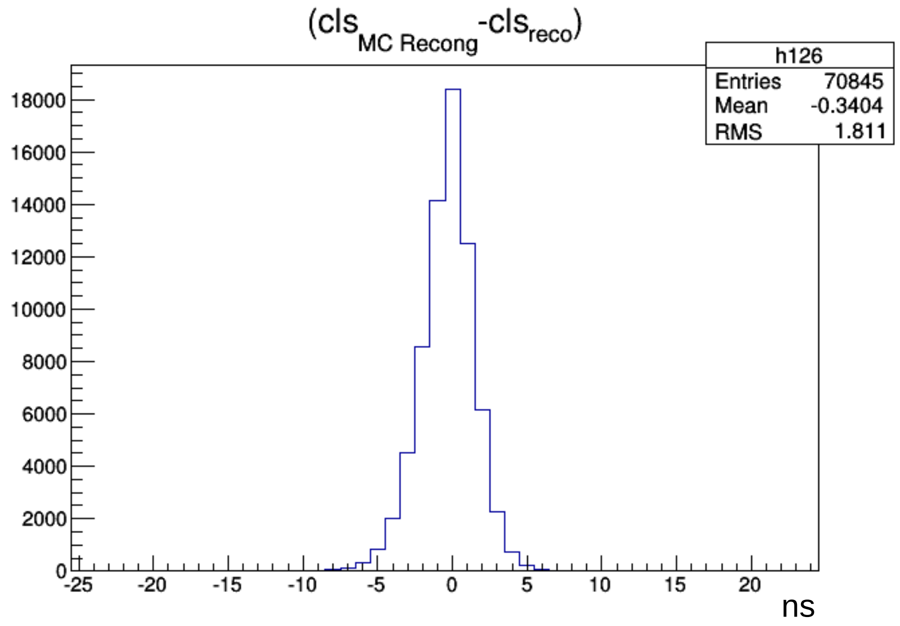
Performances



$$\varepsilon = \frac{Cls\ found}{Cls\ recogn} = 103.9\%$$

$$\varepsilon = \frac{Cls\ found}{Cls\ gen} = 94.3\%$$

Time difference between MC generated cluster and reconstructed cluster

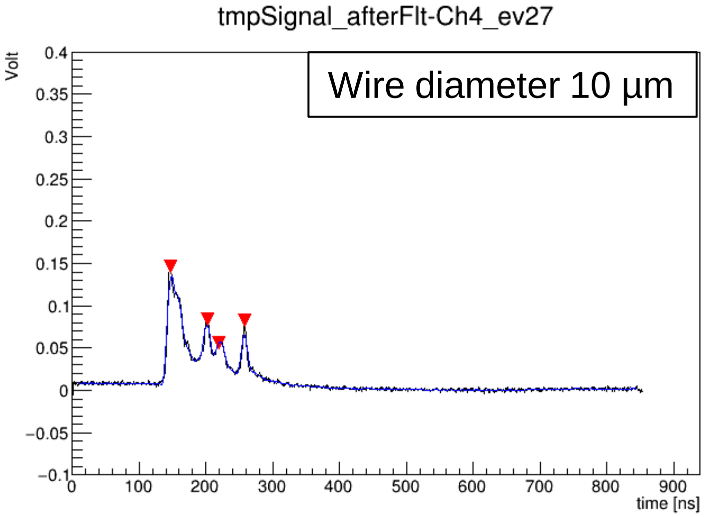


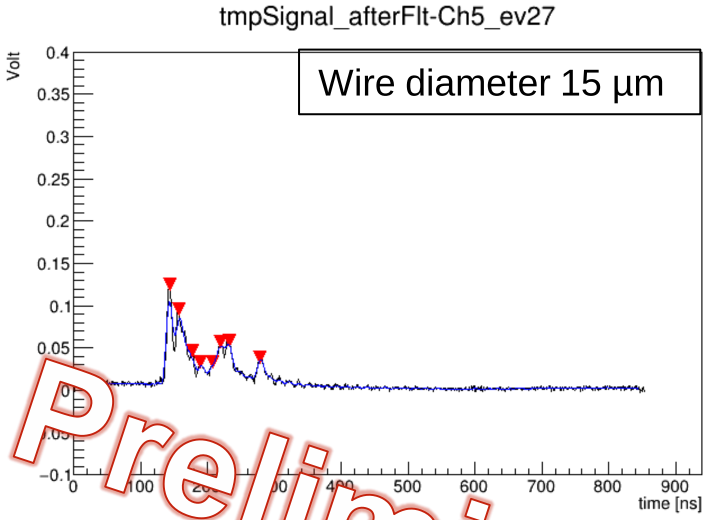
The Lecce algorithm on beam test data

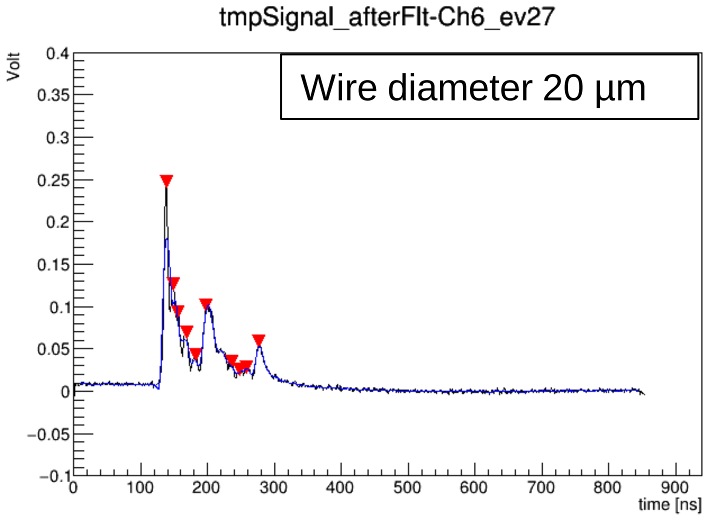
First attempt by using untuned and very tight cuts!

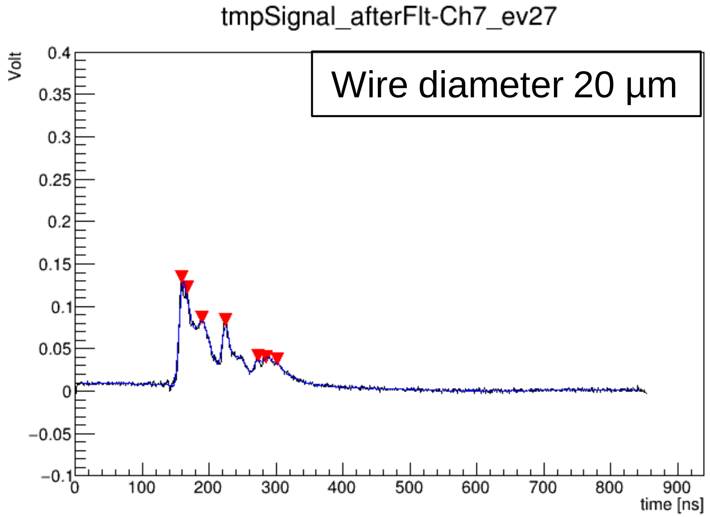
60°
Nominal HV+20V
90/10 He-iC_4H_{10}

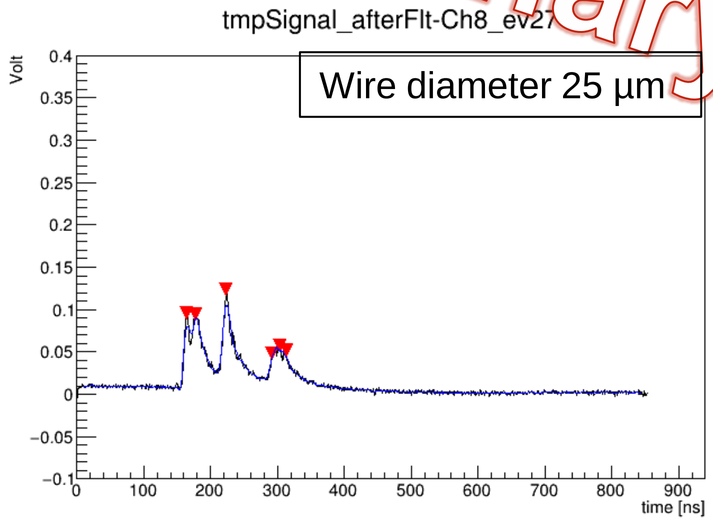
Typical event on the 6 tubes of 1 cm cell size

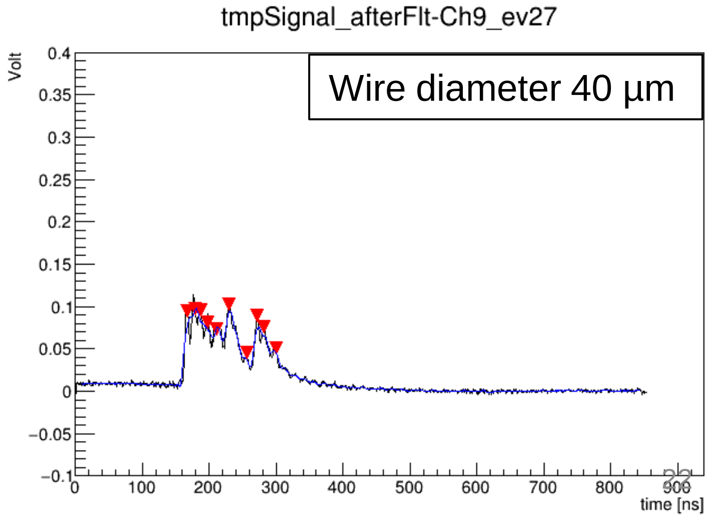






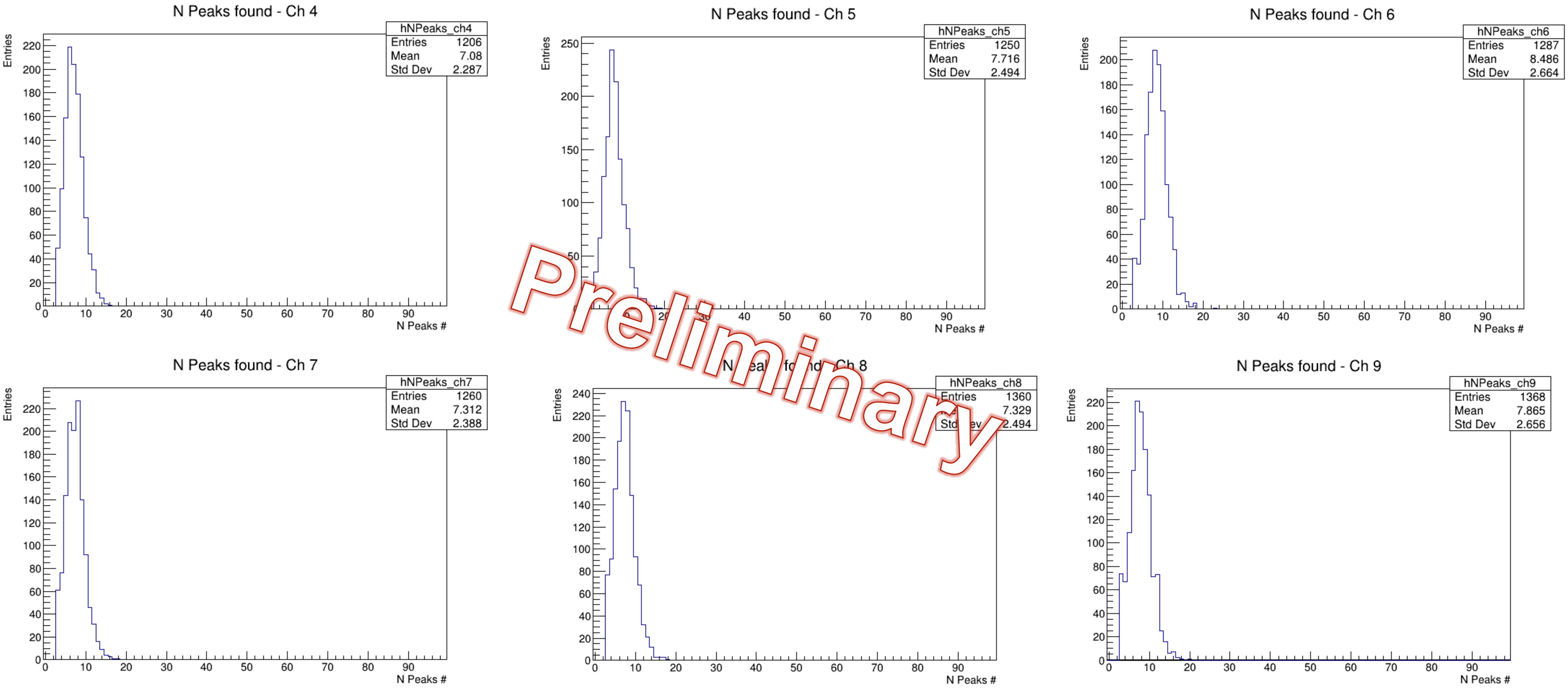






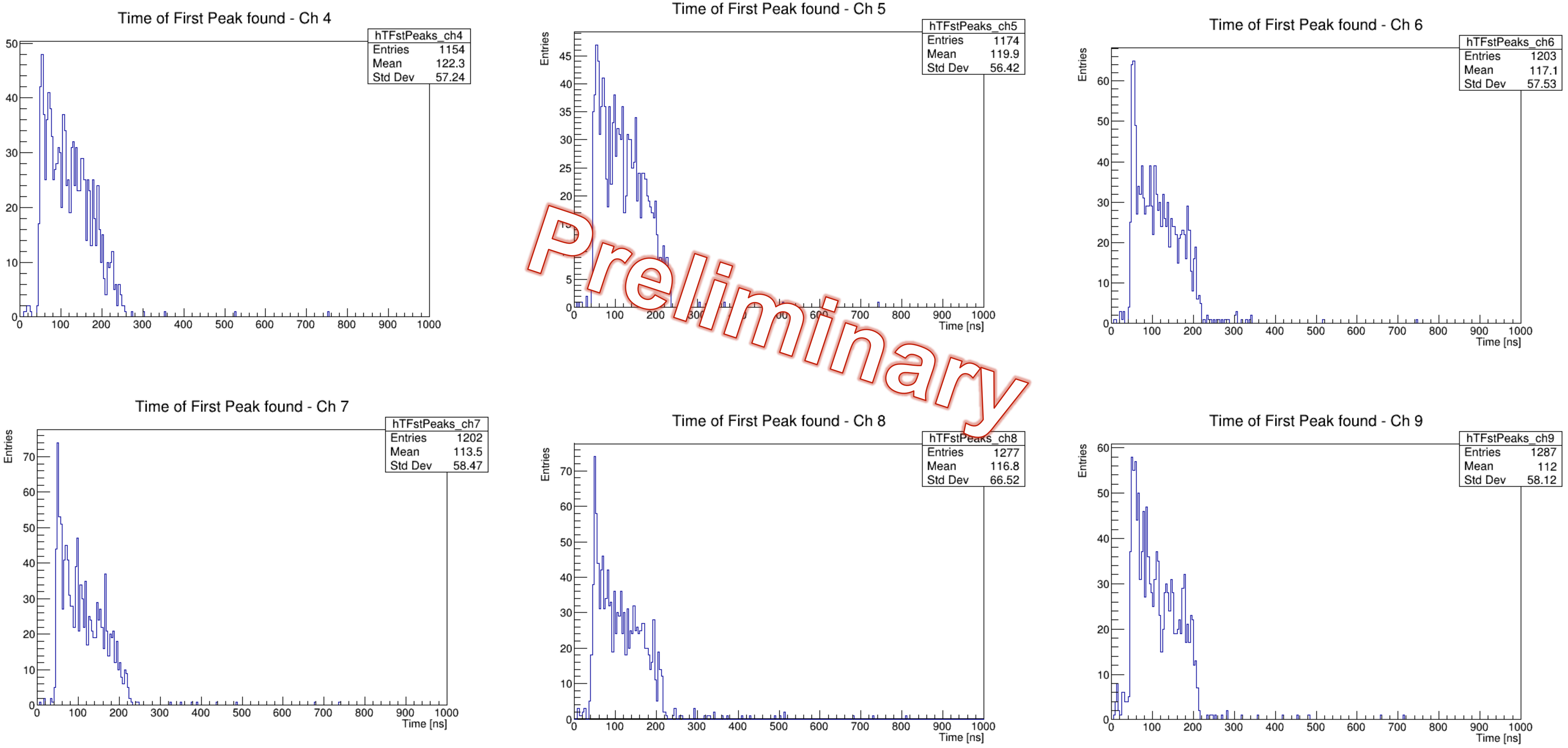
Preliminary

Peaks distribution



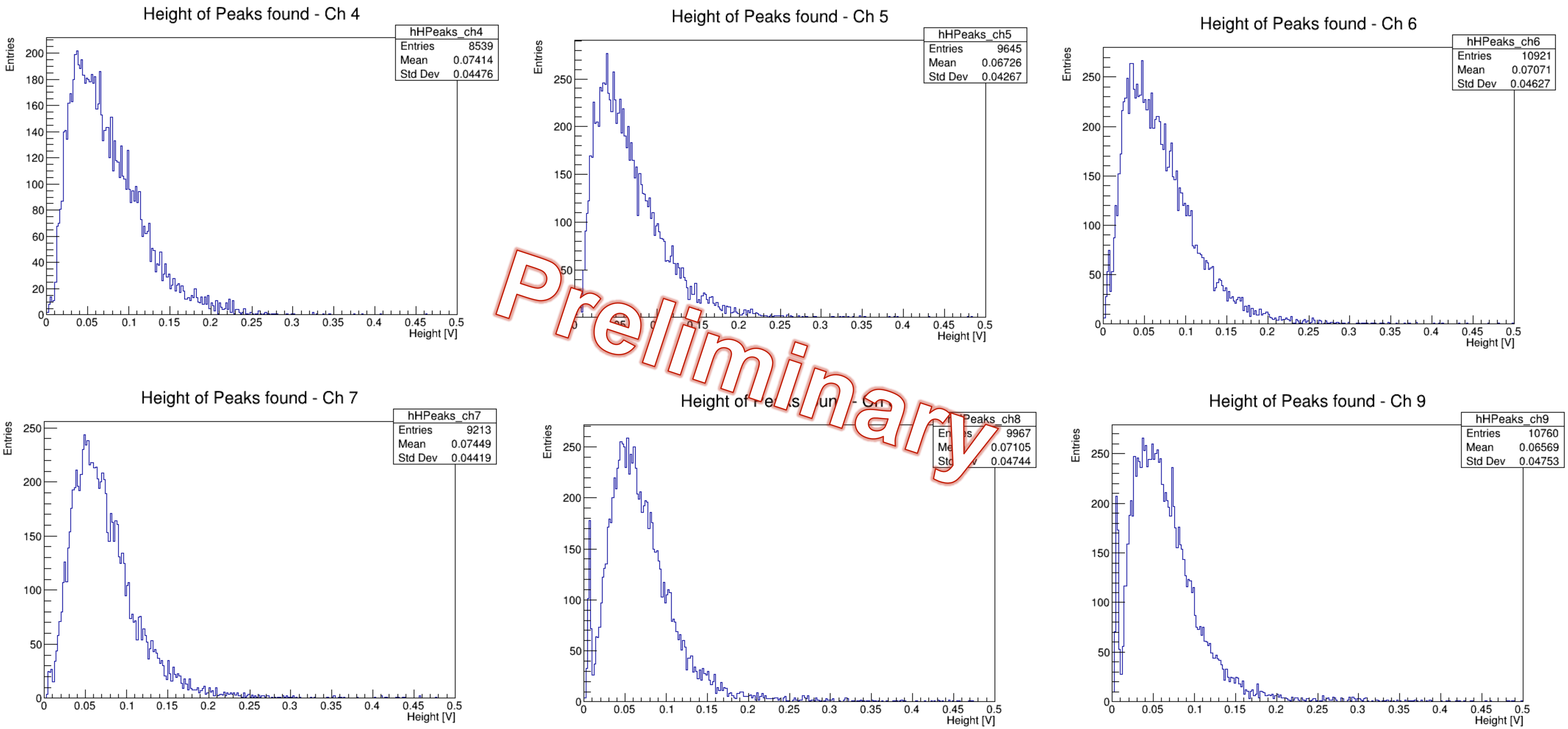
The distributions follow the expected Poissonian shape.
The algorithms applies really tight cuts, so we register an inefficiency
Anyway, this inefficiency need be compared with the recognizable peaks, which are less than the generated ones expected

Drift time distribution



From this distribution, the drift velocity is around 2cm/µs.

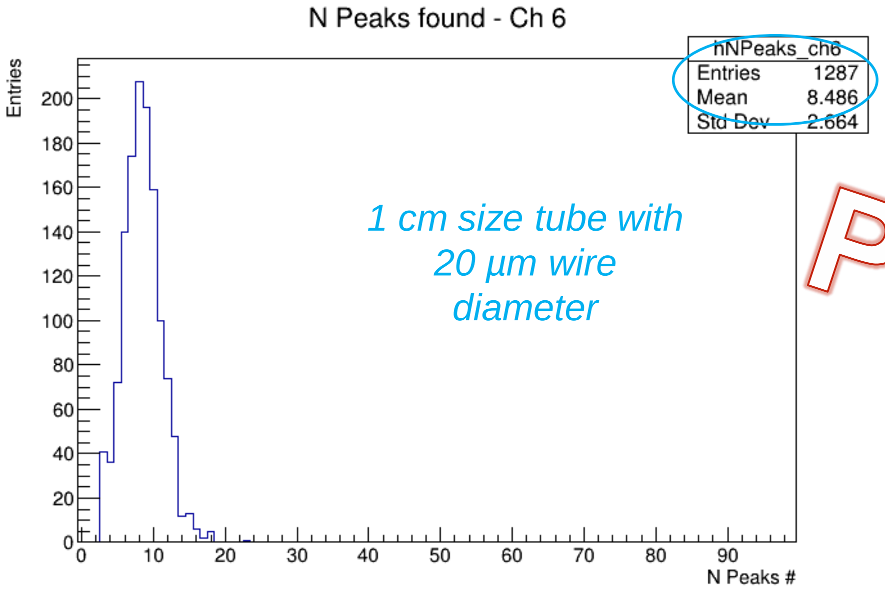
Single electron pulse height



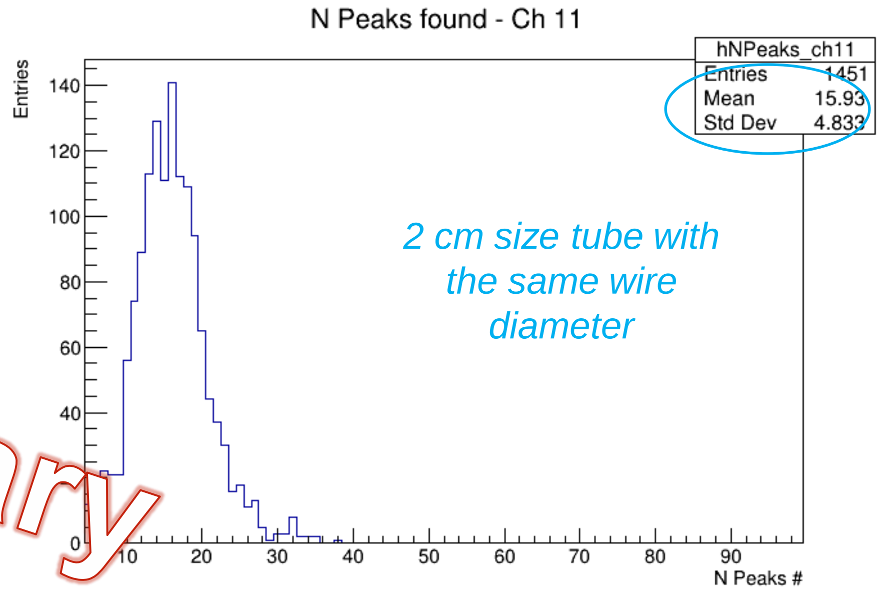
The mean values of the distributions are in agreement with the hardware calculation

Consistency check

The number of clusters generated in 2 cm tube size is expected to be almost twice the number of clusters generated in 1 cm tube size.

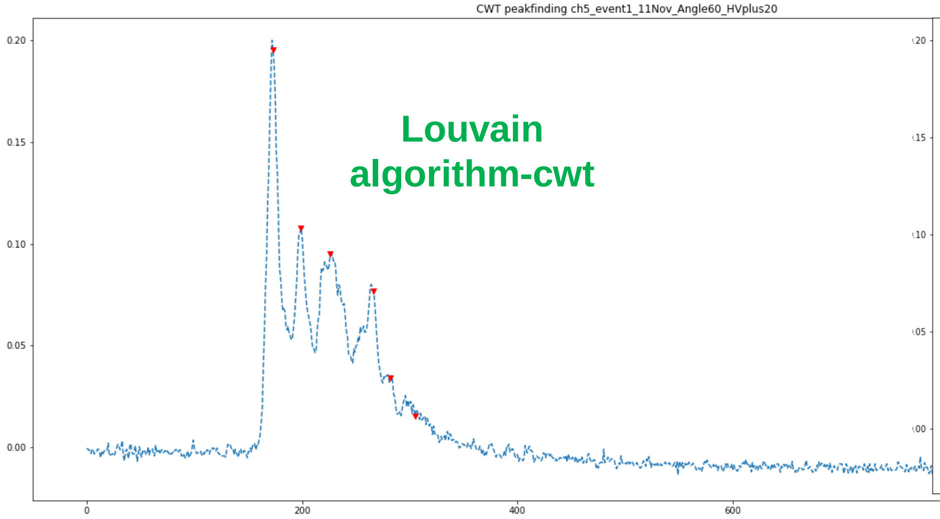


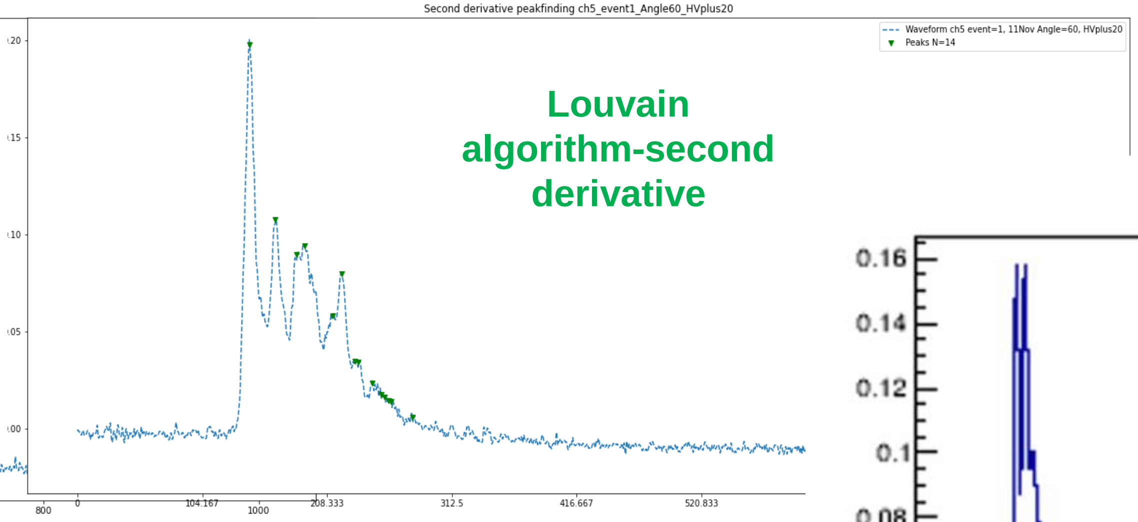
Preliminary

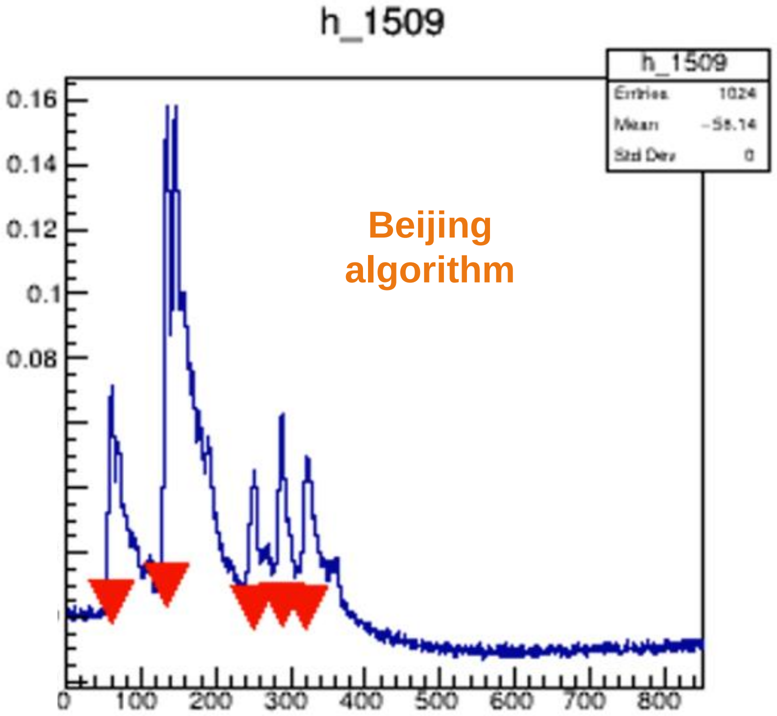


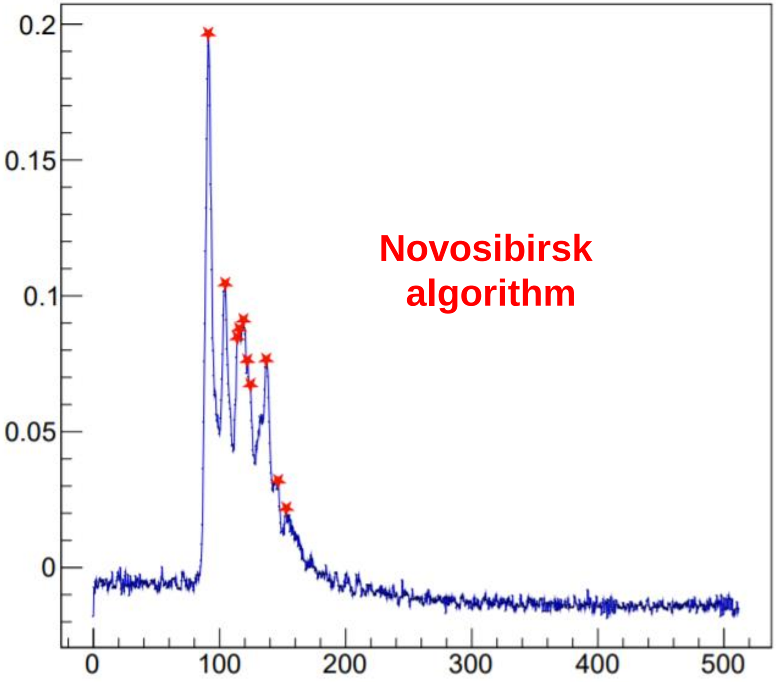
Comparison between the algorithms

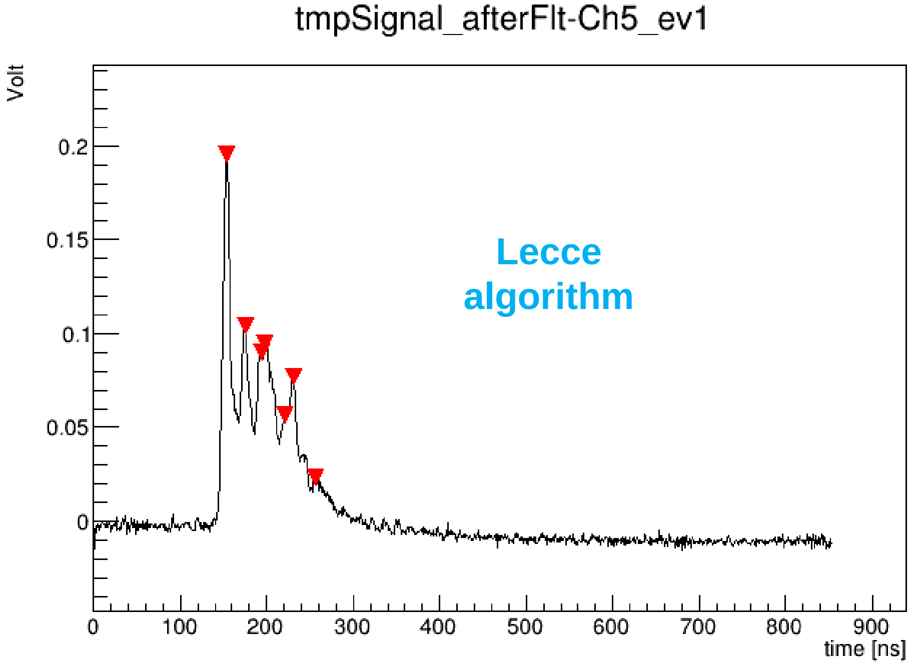
Preliminary











Each algorithms uses cautiously very tight cuts as a starting point of the analysis !!!

Conclusion

The cluster counting technique is a high powerful method to improve the particle identification capabilities: analytic evaluation and simulation confirm its potentials.

- We have already 4 promising algorithms.
- We are working on the optimization of the cuts used by the algorithm.
- We will evaluate their counting efficiency.

The beam test provides us the possibility to study the:

- counting efficiency as a function of gas mixture, gain, geometrical configuration (cell size, sense wires size), arrival time of the first cluster
- cluster density as a function of ionization length and angle
- cluster dimension as a function of gain and cell size
- definition of the optimal condition for the next test whose goal will be the measurement of the relativistic rise of dN/dx and dE/dx

THANK YOU