Software & Analysis Status

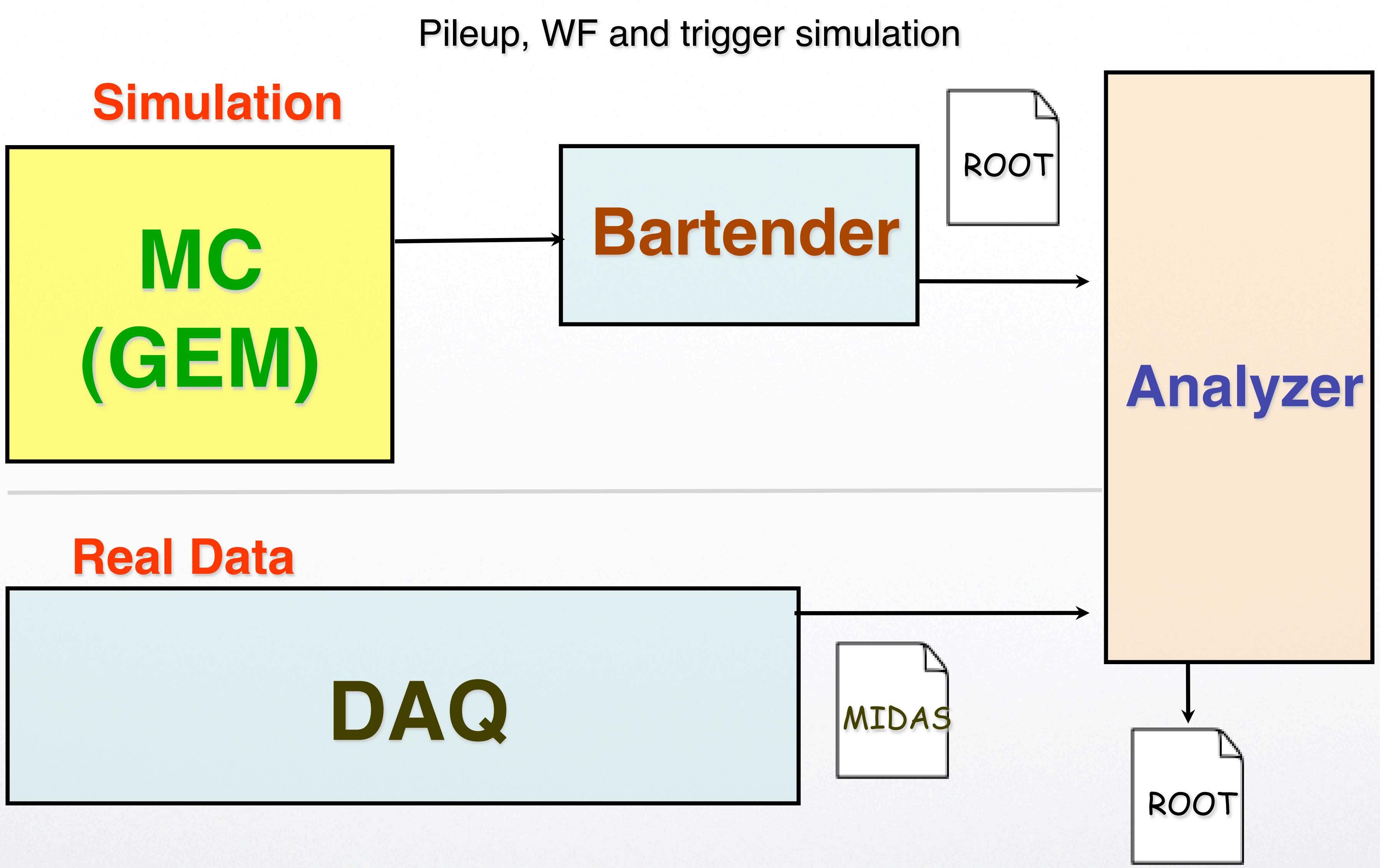
R. Sawada
MEG Review Meeting
PSI, 18 July 2007

Outline

- Software framework (Just for reminding)
- Online
- Simulation
- Analysis
  - Reconstruction
  - Physics analysis
- Computing
- Man power
- Summary

Abbreviations
XEC : Liquid xenon calorimeter
TC : Timing counter
TICP : TC Phi measuring counter (Bars)
TICZ : TC Z measuring counter (Fibers)
DCH : Drift chambers
RD : Radiative decay
WF : Waveform
B.R. : Branching ratio
AIF : Annihilation in flight
MC : Monte-Carlo simulation

MEG software



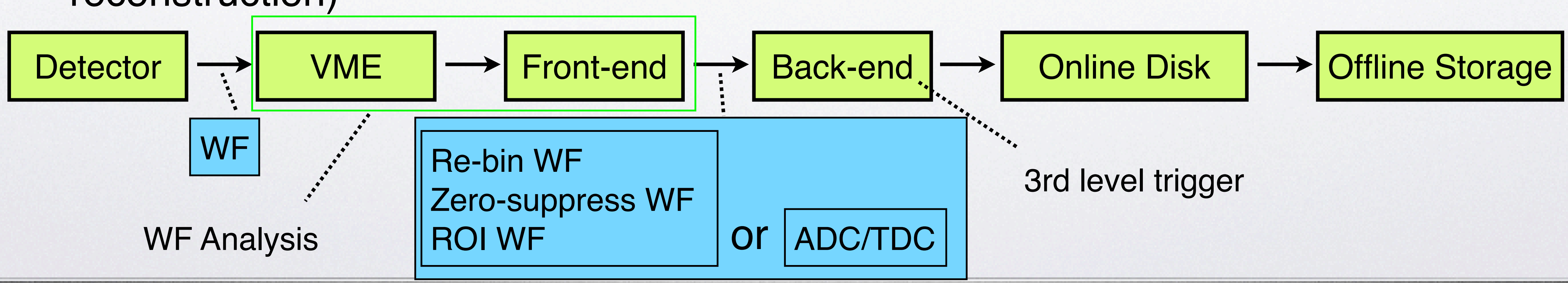
Status of software group

- **Sub groups have regular internet meetings**
  - DCH, TC, XEC, data reduction and physics working group
- **New module for “global” analysis was added**
  - Event pre-selection, sub-detecter inter-calibration, parameters for physics analysis
- **Developers of analyzer is increasing, especially students**
  - Thanks to ROME(automatic code generation system developed in MEG), new comers start development smoothly
- **Quality control of software is getting important**
  - Standard test procedure of software was introduced.
  - Everybody makes the check of software with a standard data set, with identical configuration files at every time after commits

Online
(data reduction)

Data reduction (software aspect)

- Maximum DAQ rate is limited by data size (waveform).
- Data reduction working group is developing data reduction methods with taking into account impact to analysis result.
- Methods
  - Waveform data reduction
    - Zero-suppression. Skip recording waveform outside of some criteria. (e.g. smaller pulse than a threshold)
    - Decrease data points to recored by averaging several points (Re-binning).
    - Decrease time window for recording (ROI)
  - Record ADC/TDC instead of WF for calibration events. (ADC/TDC simulated by software)
  - 3rd level trigger in DAQ back-end (software event selection by using online fast reconstruction)



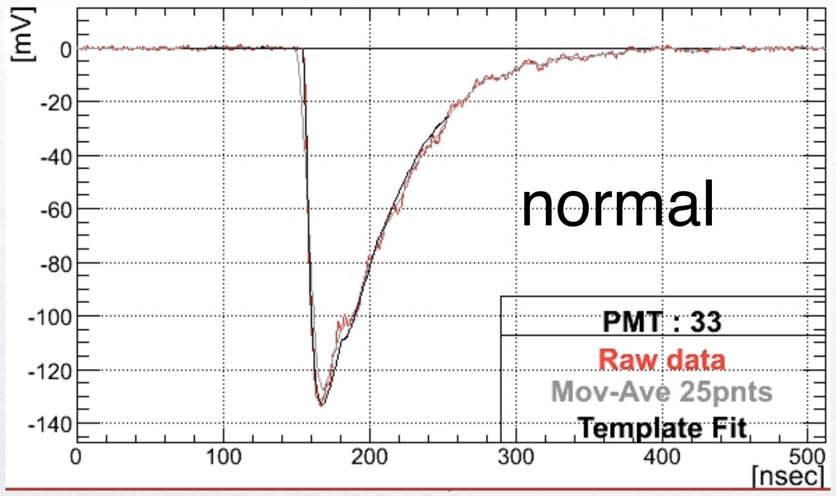
Calorimeter

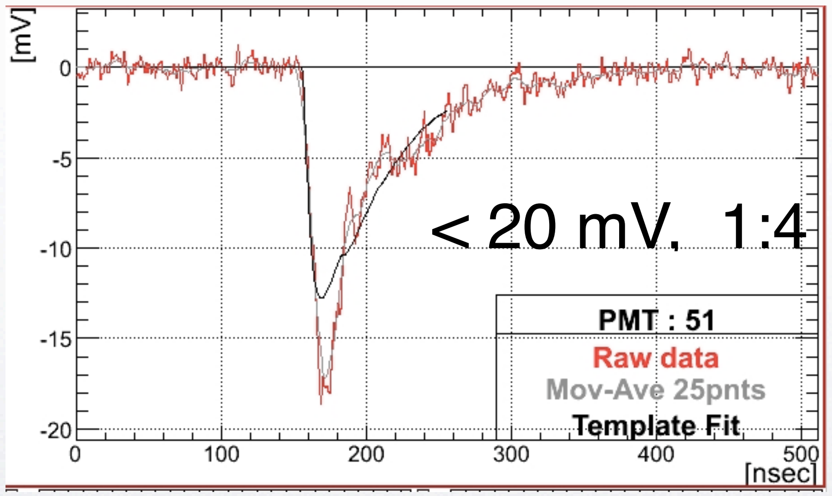
Energy reconstruction : 70% of PMTs are needed

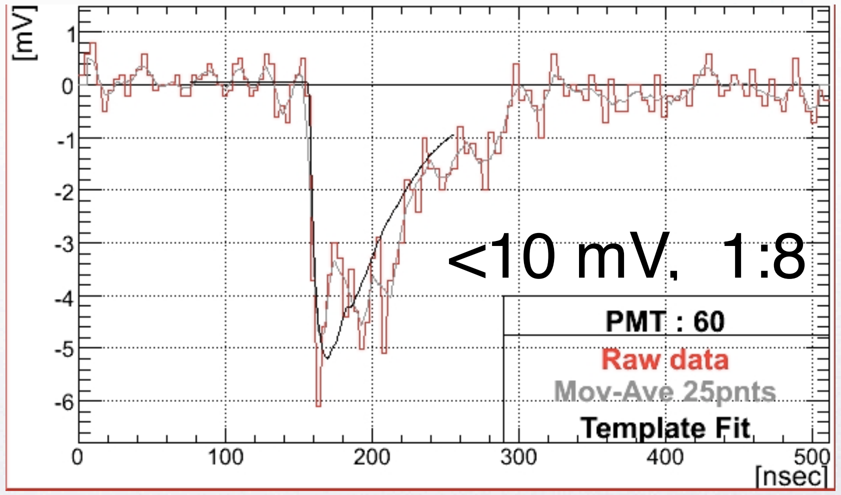
Time reconstruction : Small pulses do not give important information for reconstruction

⇨ **Re-binning is suitable**

Example







Drift chamber

2.7 wires per plane out of 9 are expected to have a hit.

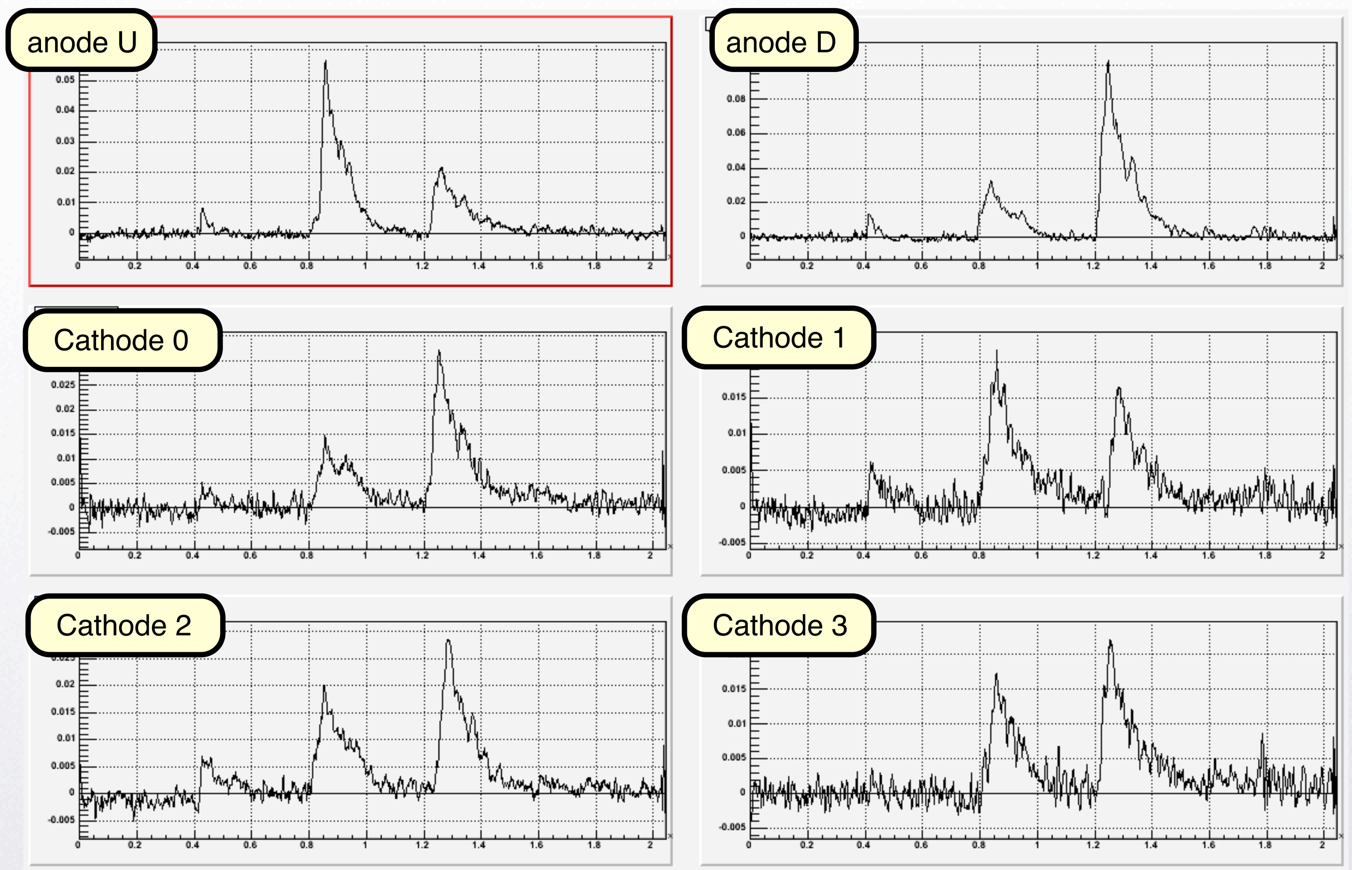
⟹ **Factor 3.2 reduction by zero suppression**

For Michel or RD events 700 nsec time window is enough instead of 2000 nsec

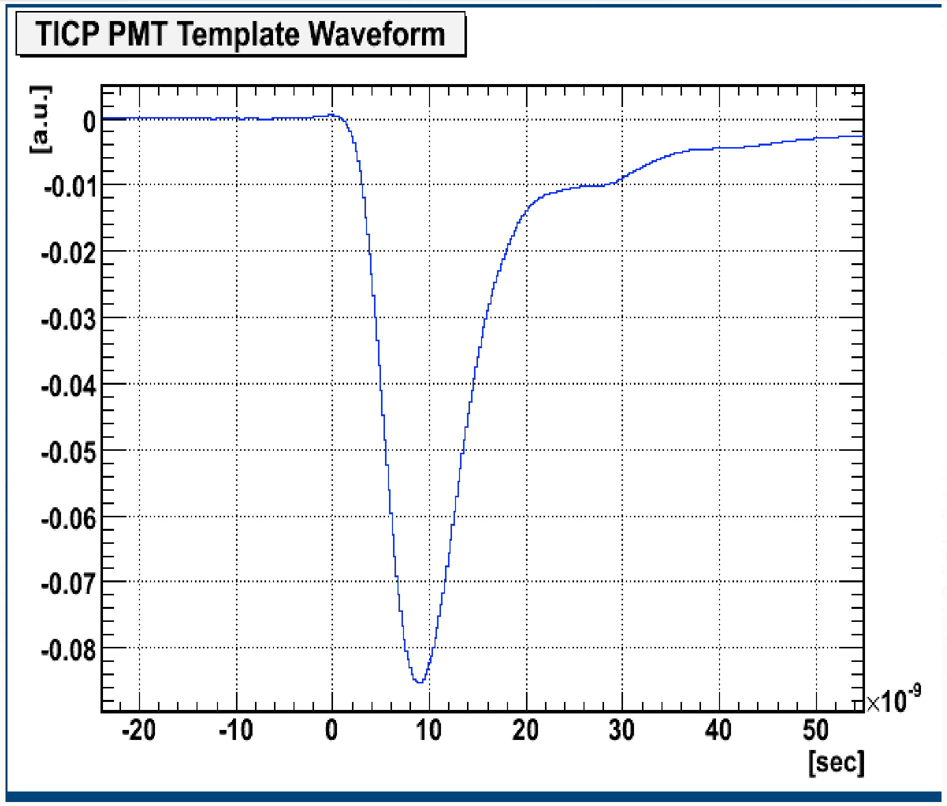
⟹ **Factor 1.4 reduction by reduced time window for recording**

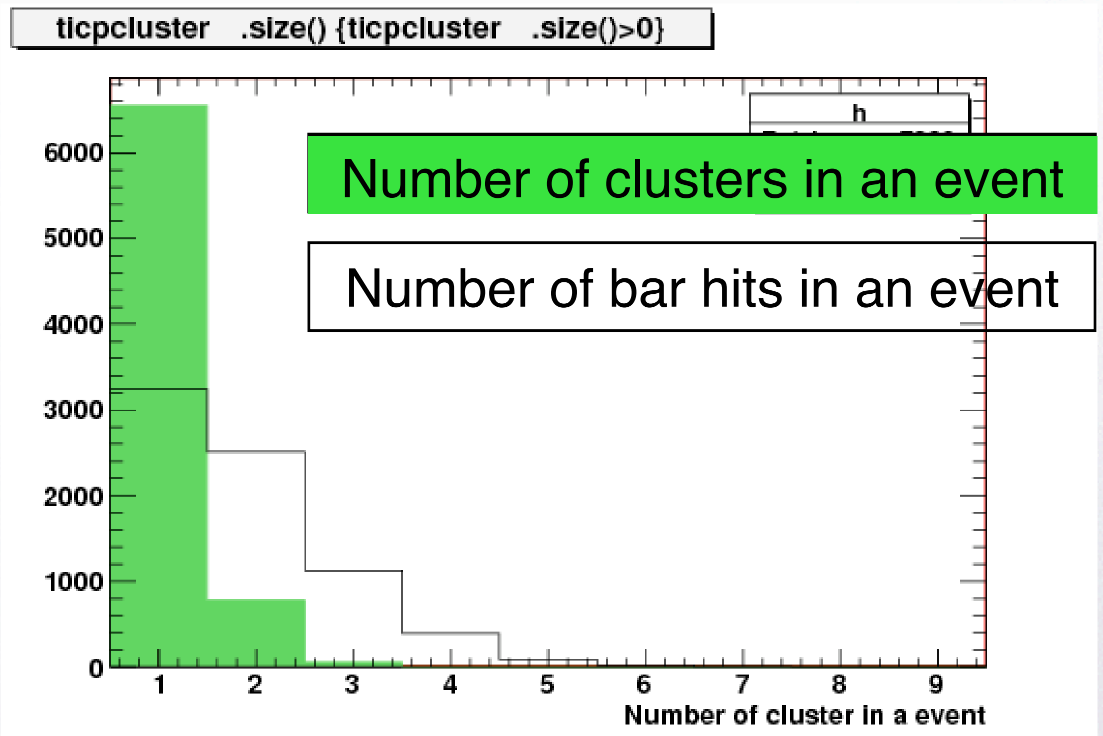
Topological relation between channels has to be taken into account.
(i.e. When a wire has a hit, cathode and anode in the cell and adjacent cells are recorded)

3 cells × (2 anodes + 4 cathodes) = 18 channels



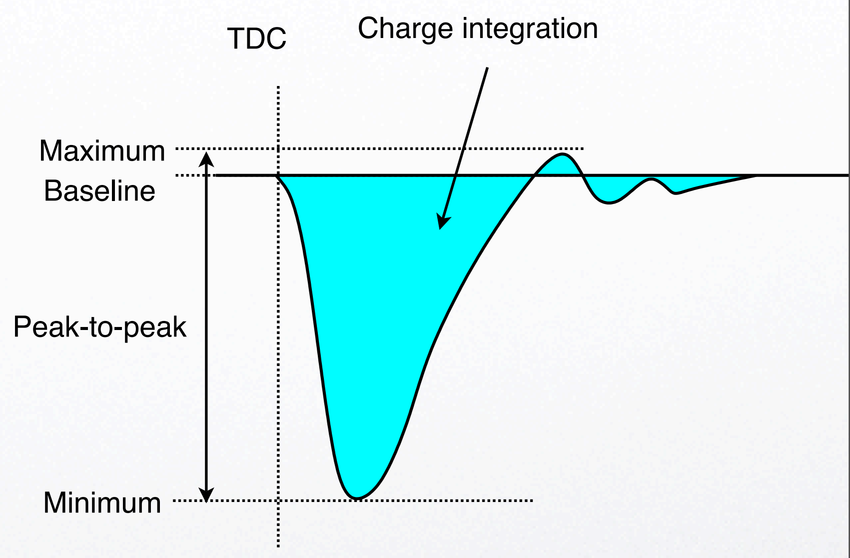
- Data size is smaller than DCH or XEC.
  - TC:120 ch, XEC:846 ch, DCH:1728 ch
- Pulse width (~50 nsec) is narrow compared to DRS full time window (512 nsec).
  - **Decresing time window to record will decrease data size**
- Number of bar hits is small (~2 bars out of 30)
  - **Zero-suppression will decrease data size**





Status

- **C functions library of waveform analysis** in online was prepared.
  - Baseline, Peak-to-peak voltage, Maximum, Minimum, Charge integration, Leading edge TDC, Double threshold TDC, Constant fraction TDC.
- **Simple zero suppression** with a single voltage threshold is implemented in DAQ front-end
- **Data reduction for DCH** is implemented in online front-end



- **To do**
  - Waveform analysis in DAQ front-end or back-end.
  - ADC/TDC DAQ mode instead of waveform
  - Study of sub-detector specific data reduction. Estimation of data reduction factor and impact to analysis
  - 3rd level trigger

Simulation

MC(GEM) updates

- Event generation
  - AIF probability enhanced mode. (Preliminary)
  - Radiative decay
    - More flexible configuration keys to specify generation condition. (energy cuts and opening angle)
    - Calculation of phase volume (branching ratio) of specified kinematic range
- Geometry updates
  - Geometry for this year’s setup
  - Details (target, honeycomb panel of calorimeter, alpha source....)
- Better simulation of drift electron in chambers. (diffusion included)

MC(Bartender) updates

- Reorganization of MC hit structures of each modules (DCH, TIC[P/Z], XEC) to be easier to compare with reconstruction.
- Waveform simulation updates.
  - DCH : Response function from X-ray source data
  - XEC : Precise TTS simulation, Attenuators in DRS, Clipping before splitter.
- Several performance improvements for collecting data in signal region.
  - 1.2 sec/event/CPU for prompt background
  - 1.3 sec/event/CPU for accidental background

Analysis

Figures and numbers will be shown.
These are all preliminary, just for status report.

All studies are using MC data.
(Pure signal, or assuming 3E7 muon stopping rate)

Drift chamber

DCH task organization

HitRec

HitRec2

Waveform → Hit

ClusterFinder

ClusterFinder2

ClusterFinder3

ClusterFinder4

Clustering Hits

TrackFinder

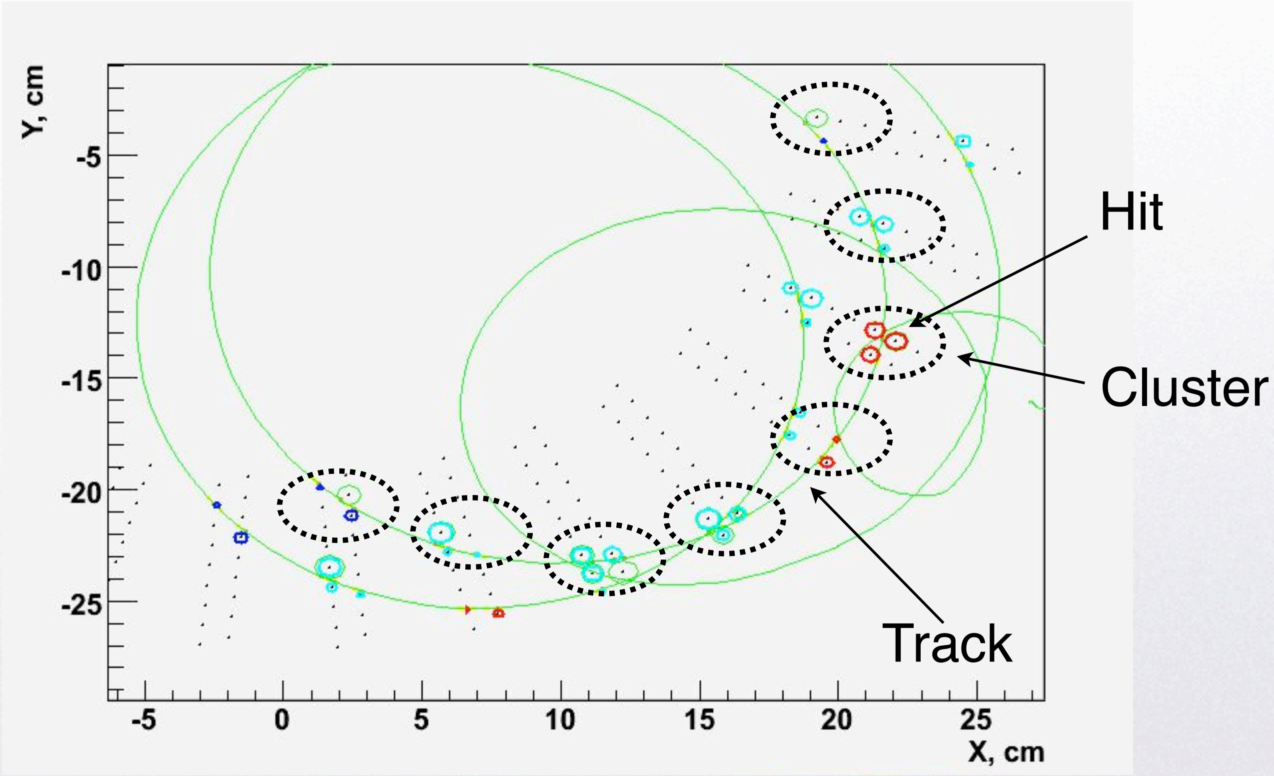
TrackFinder2

Finding track candidates
Grouping hits associating with a track

MinuitFit

TrackFit

Track fitting



Several algorithms at each stages are being developed in parallel

Hit Finding

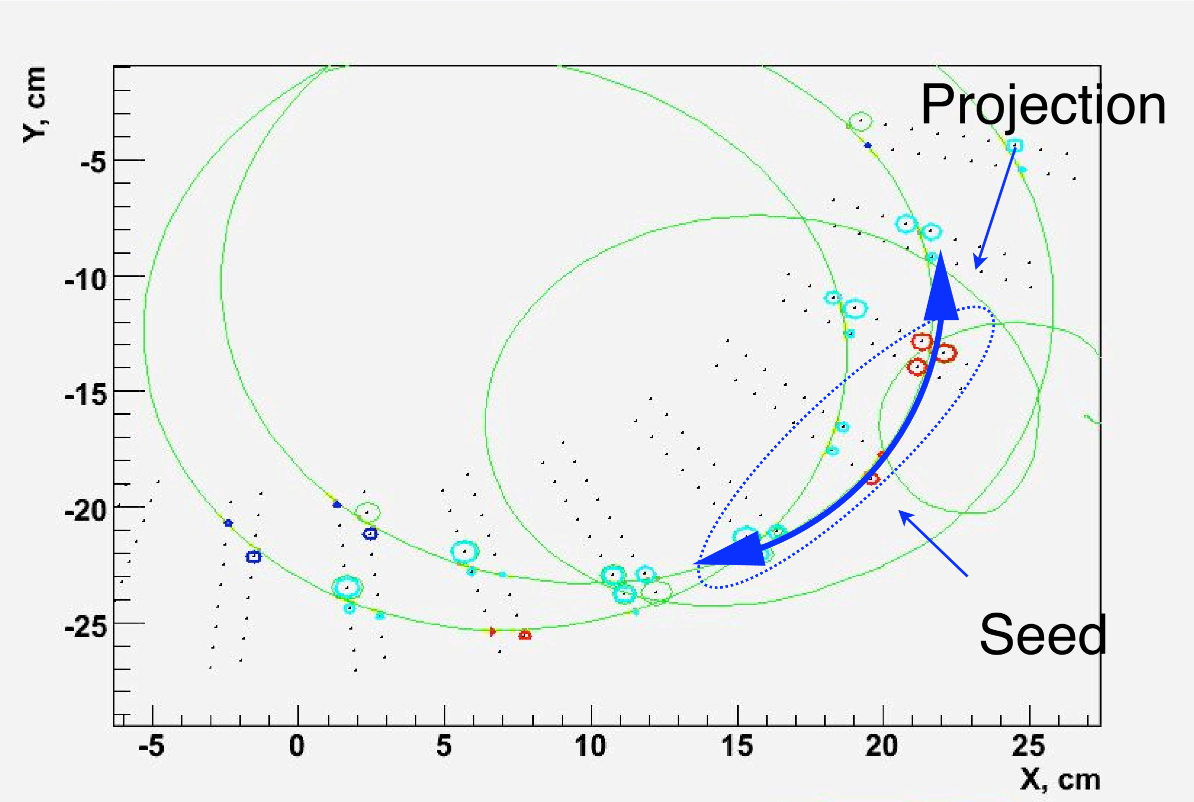
- Goals
  - Correct for waveform noise (high frequency, low frequency, coherent, incoherent)
  - Remove far out of time hits
  - Find best leading edge time
  - Find integrated charge
  - Find Z coordinate from both charge division on anodes and charge distribution on pads
  - Compare shape of hits at two ends, on pads
- Status
  - Waveform analysis implemented with noise reduction.
  - Charge integration in time region of recognized pulse is implemented
  - Z coordinate from both wires and pads are implemented

Cluster Finding

- Goals
  - Identify cells associated with single particle passage
  - Identify cells hit by more than one track passing
  - Reduce noise into next stage
  - Get better axial (z) and radial (r) coordinate
  - Provide a precise single coordinate for the fitting
- Status
  - Two algorithms are implemented
    - Algorithm 1 – Combine two hits into a cluster according to Z position and cell#.
    - Algorithm 2 – Find clusters using adjacent hits, average R coordinate, splitting of large clusters using z information. coordinates are improved at track finding stage.

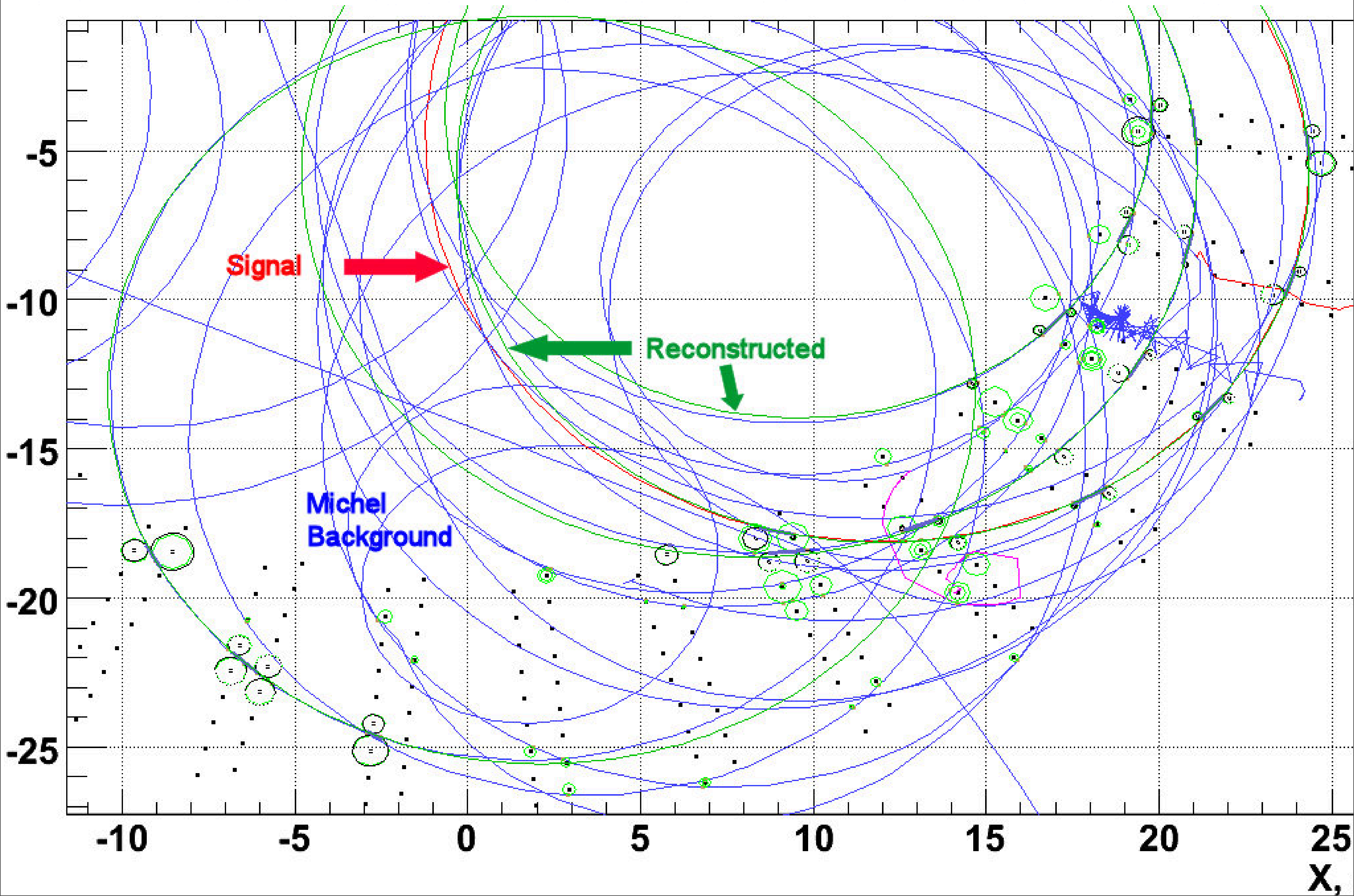
Track Finding

- Goals
  - Find hits associated with particle
  - Resolve left-right ambiguity for fitter
    (maybe good enough for trigger)
  - Provide first estimate of kinematical quantities
  - Reduce extra hits in input to fitter



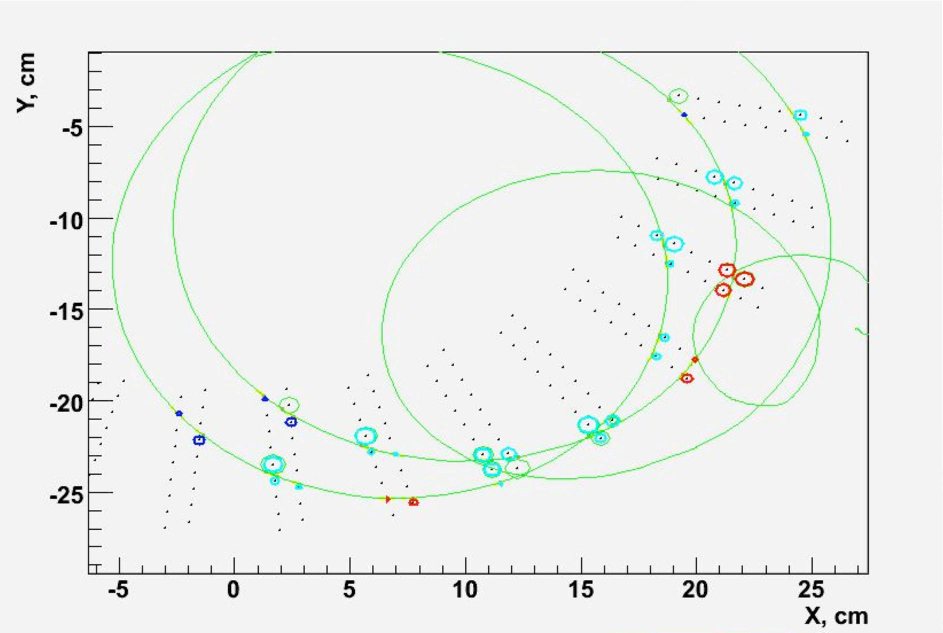
- Status
  - Two algorithms are implemented. “make seed” → “project track” → “connect clusters”
    - Algorithm-1:Progressive track finder based on 3-cluster track seeds at large radius, successively projecting to next chamber in each direction
    - Algorithm-2:Progressive track finder based on 3-cluster track seeds from younger chamber#, successively projecting to next chamber to direction of increasing chamber#.

Algorithm-1: Event display



Algorithm-2(Seeding)

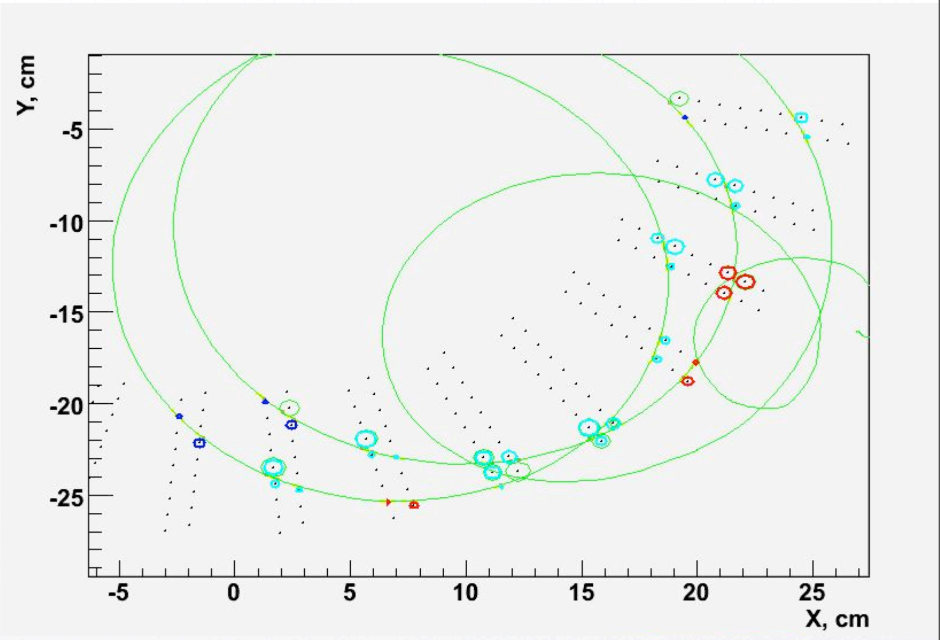
- Look for any 3 clusters on consecutive chambers with certain criteria as track seed
  - middle cluster should exceed a threshold in R, to eliminate contaminations
  - R, Z windows applied to select seed clusters on each side
  - missing a chamber allowed



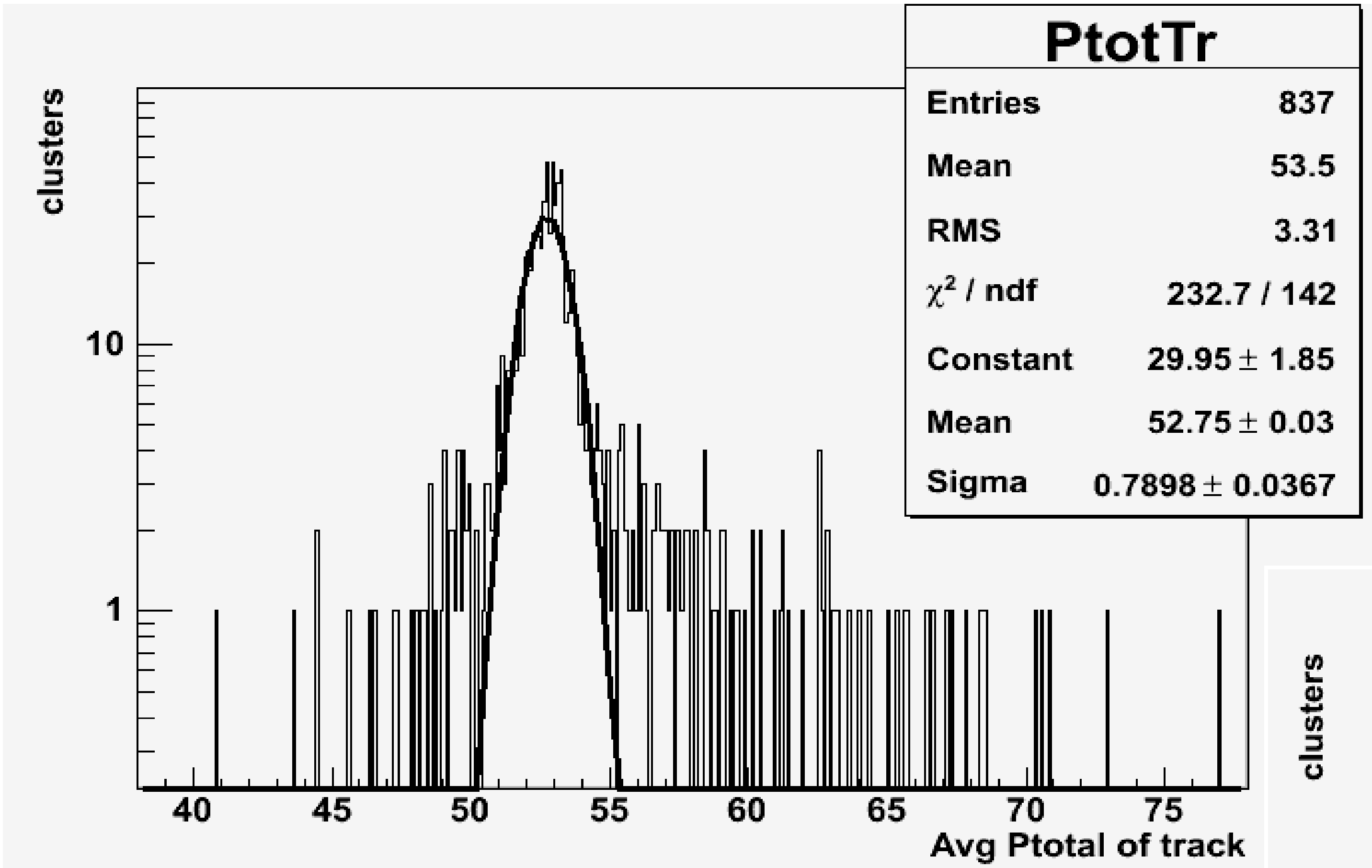
- Optimize T0 from the seed
  - Find smallest time in all hits in the cluster (EstT0)
  - Calculate a track circle from the seed
  - get cluster angles
  - Calculate drift circles by TXY functions, with input hit time - EstT0 and cluster angles
  - Calculate the deviation of drift XY point to the track circle
  - Decrease EstT0 by 5 ns, iterate above steps until we minimize the deviation, take this EstT0 as T0

Algorithm-2(Tracking)

- Projection using invariant of motion for both R and Z instead of linear Z, and circular R projection
- Extend tracks in both directions, increasing chamber # and decreasing
- Improve cluster positions while extending the tracks by the cluster angles and T0 gotten from track
- Resolve left-right ambiguities in the process of refining the cluster positions
- Split tracks if there is more than one cluster falls into the projection window
- Missing one chamber is allowed but not yet looking for signal hits on them
- Trash all tracks that have less than minimum number of clusters (set to 5 now)

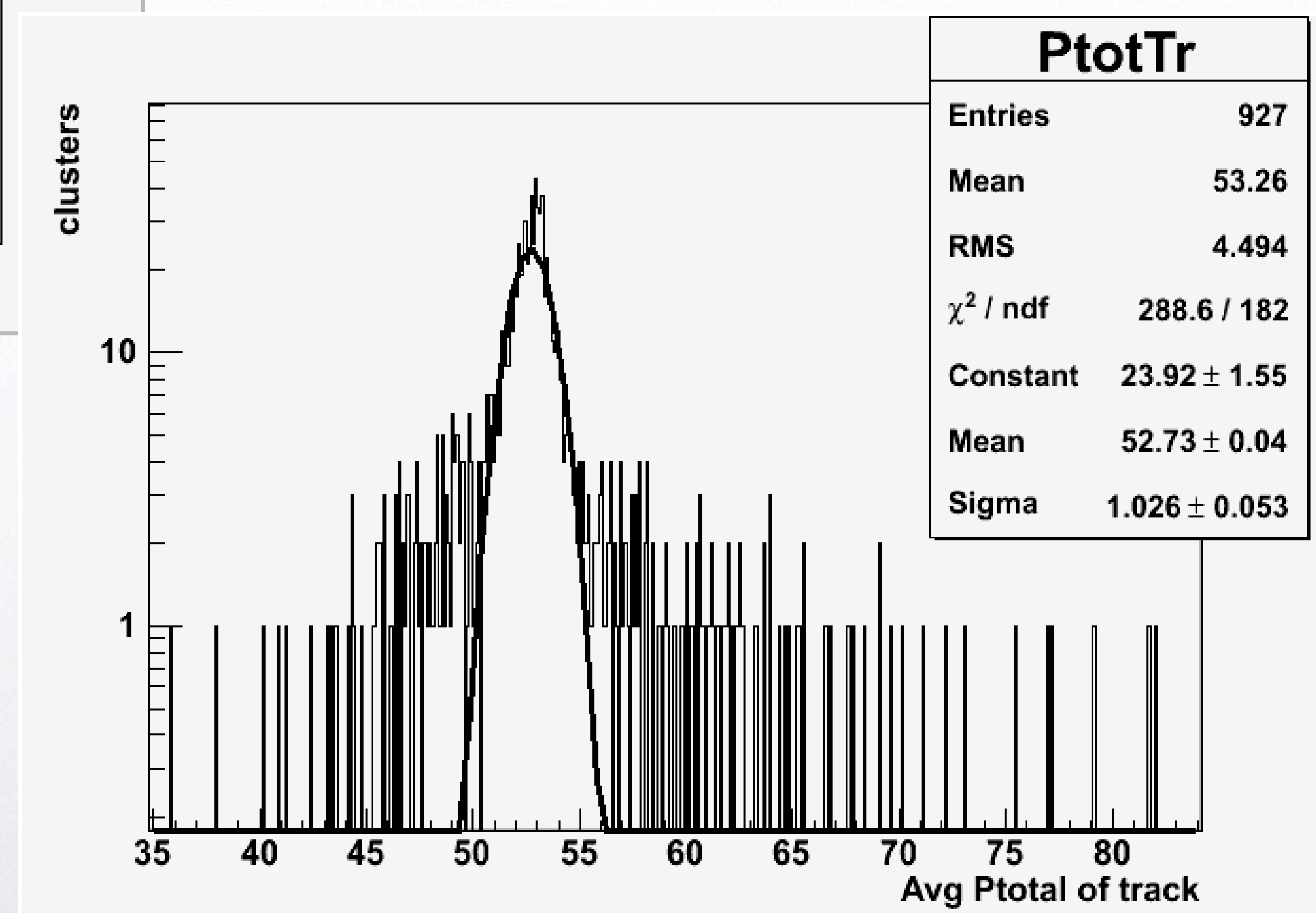


Algorithm-2 result : Momentum Resolution



1.5% in σ for signal events

1.9% for signal with 3e+7 background

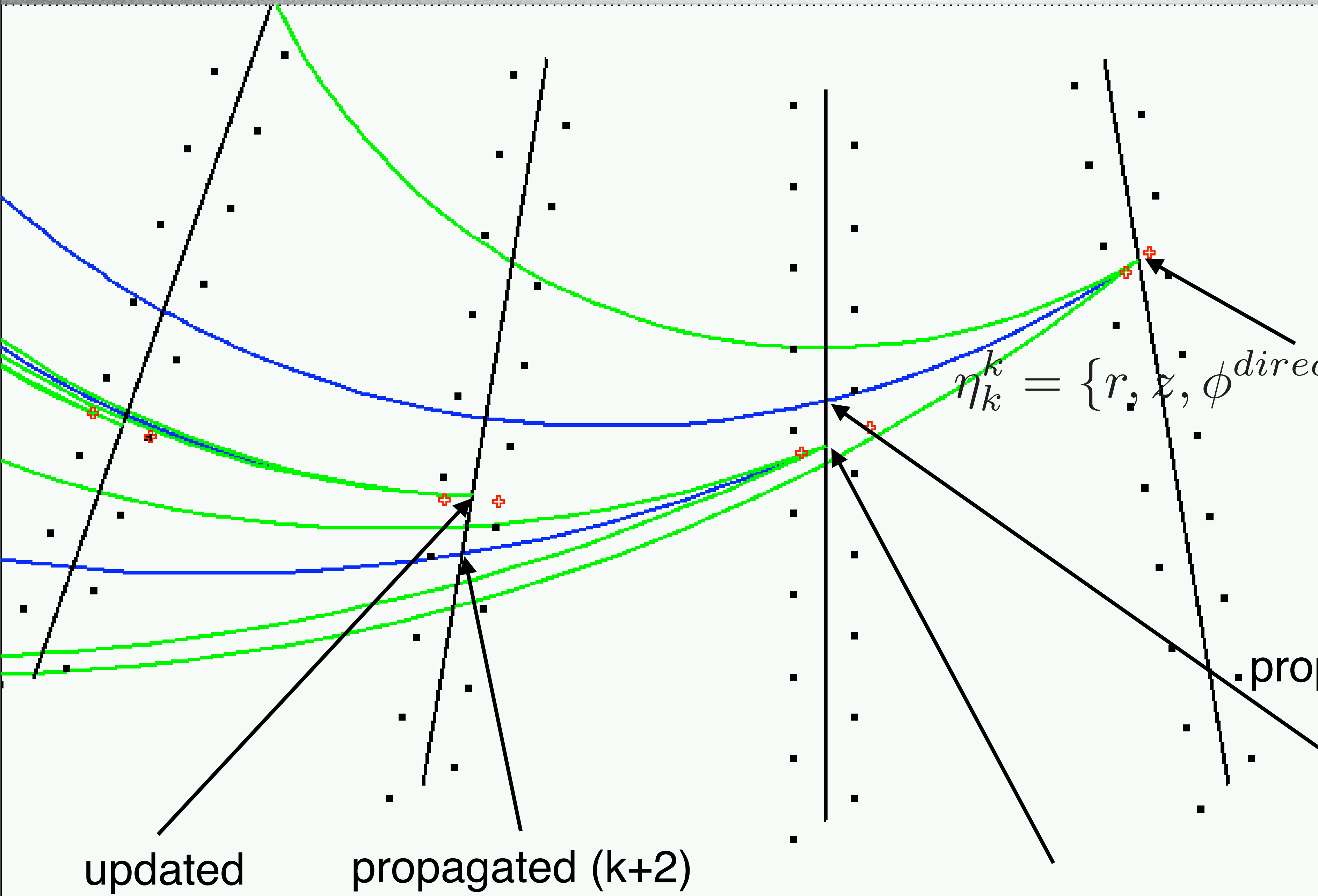


This is resolution before fitting stage.
This could be used for pre-selection, but this is not the final result.

Track Fitting

- Goals
  - Find best kinematical quantities
  - Determine quality of fit, identify poorly fit tracks
  - Determine precision of measurements
- Kalman filter exists
  - Finds kinematical quantities and state vector of tracks at different points in trajectory
- Alternative fitter using MINUIT
  - Make use of integration of equation of motion in vacuum.
  - This is written in Fortran code. Need to be implemented in analyzer.

Principle of Kalman Filter



parameter vector
with error matrix
on the current plane

$$\eta_k^k = \{r, z, \phi^{direction}, \tan(\theta^{direction}), 1/p\}$$

$$C_k^k = <\delta\eta\delta\eta>$$

propagated to next plane

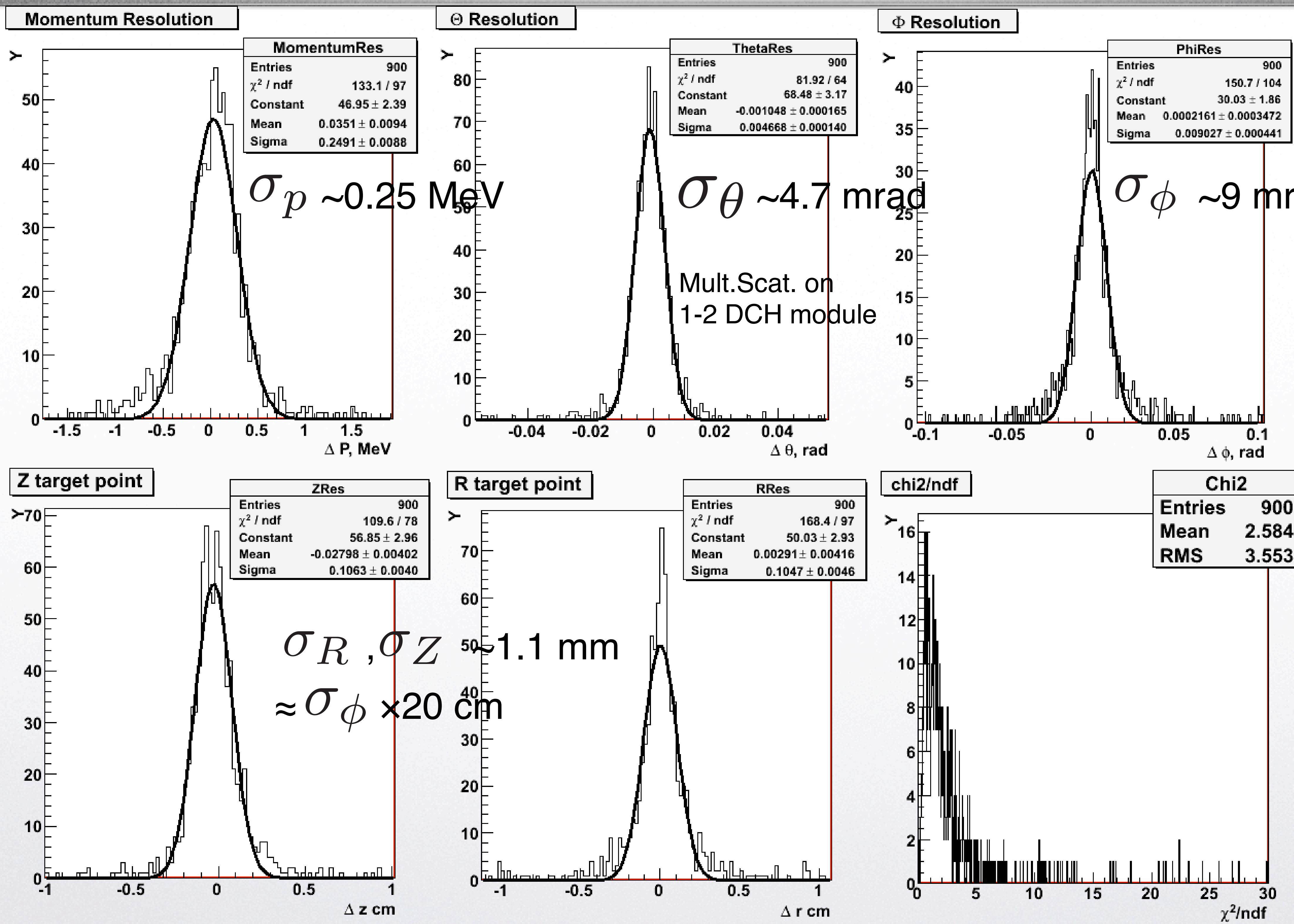
$$\eta_k^{k+1}, C_k^{k+1}$$

$$\eta_{k+1}^{k+1}, C_{k+1}^{k+1}$$

Updated by weighted mean of
parameter vector and measurement

Recursive least-squares estimation
Equivalent to global least-squares method including all correlations between measurements due to multiple scattering.

Resolution of Signal Events in full DCH reconstruction chain



σ_p ~0.25 MeV

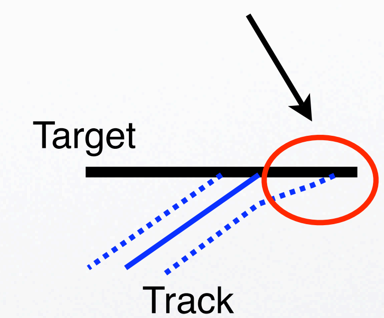
σ_θ ~4.7 mrad

Mult.Scat. on
1-2 DCH module

σ_ϕ ~9 mrad

σ_R ,σ_Z ~1.1 mm
$\approx \sigma_\phi$ ×20 cm

additional rotation in magnetic field, for 1 mm additional path if track not normal to target

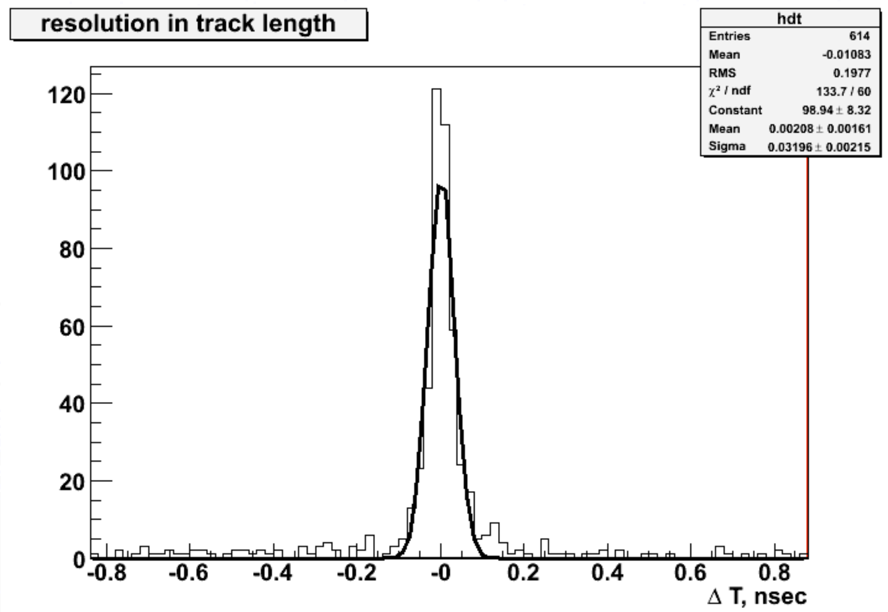


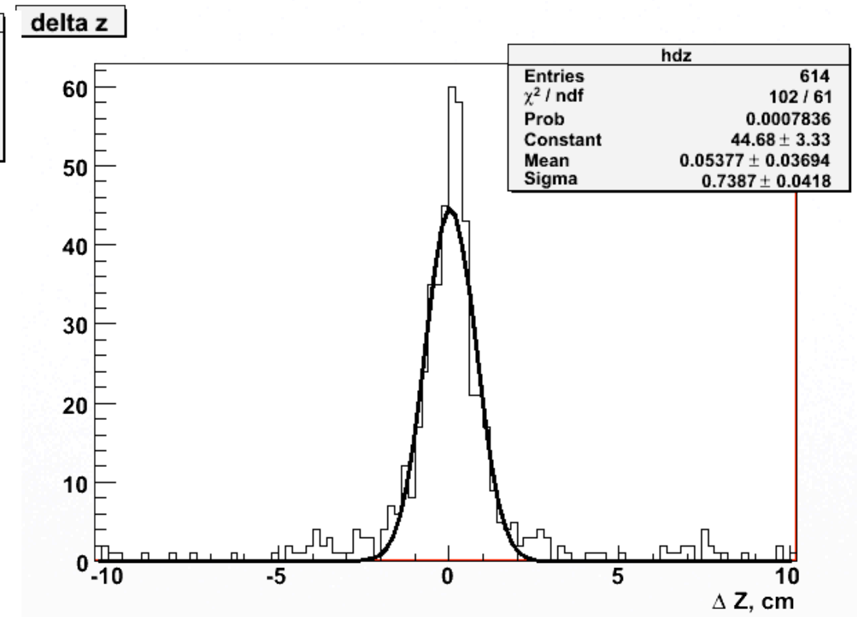
Uncertainty from T0, DCHTXY map,fake hits, ... With michel background

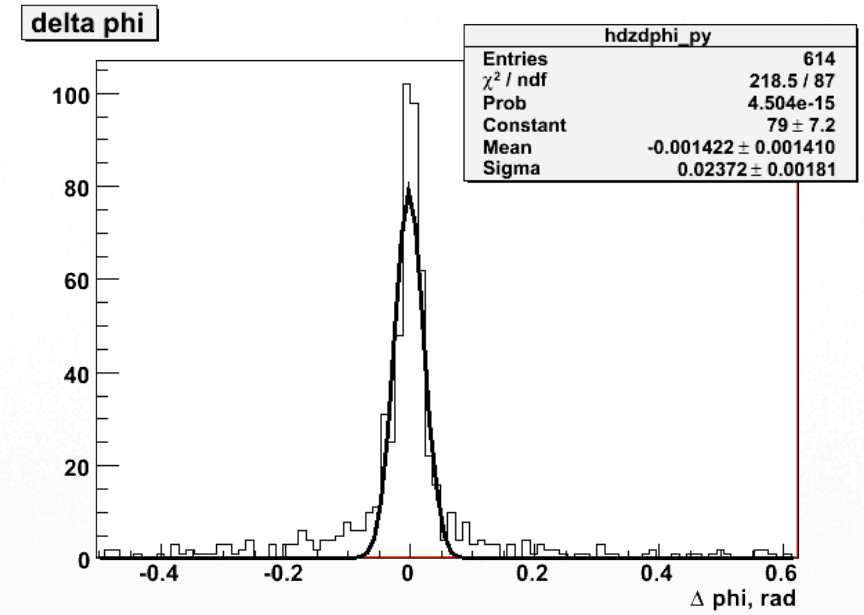
Resolution of track on extrapolation to TC (comparison with MC TICZ hit)

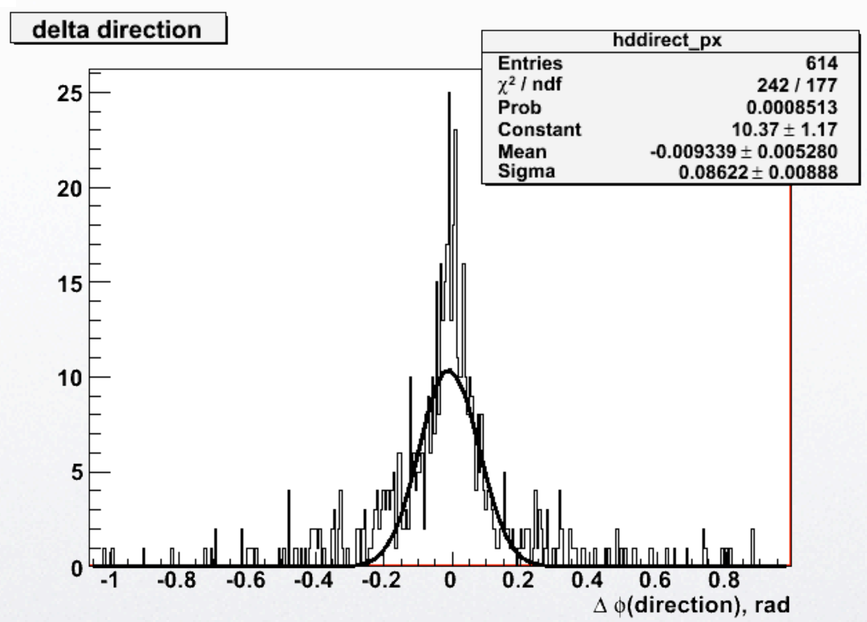
$\sigma_{track\text{-}length} \sim 0.03$ nsec

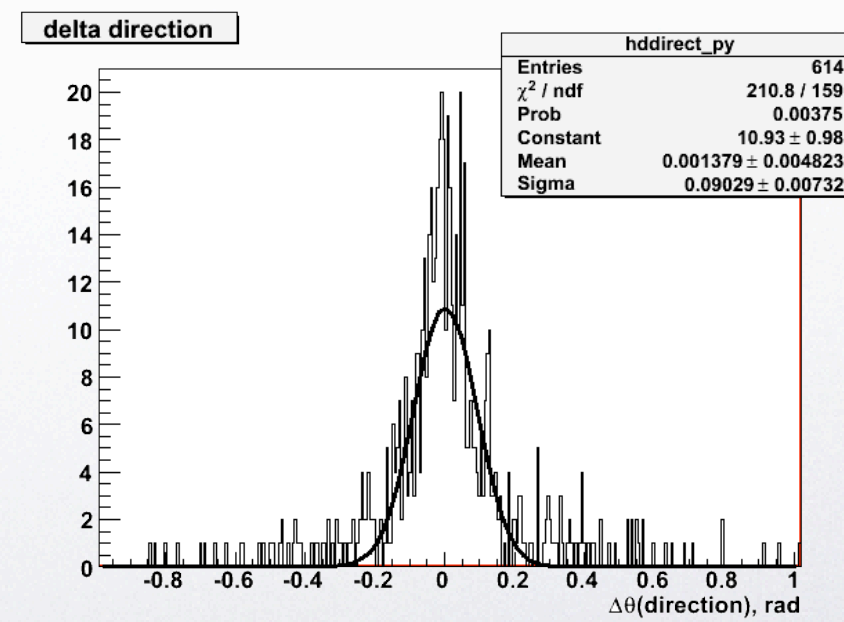
$\sigma_z, \sigma_R \sim 0.7$ cm











$\sigma_{direction} \sim 0.1$ rad

Timing counter

TC Analysis

Waveform→ Q,T of PMT

Hit in bar

TICP

PMT analysis

Hit Reconstruction

TICZ

TICZ Hit clustering

Clustering

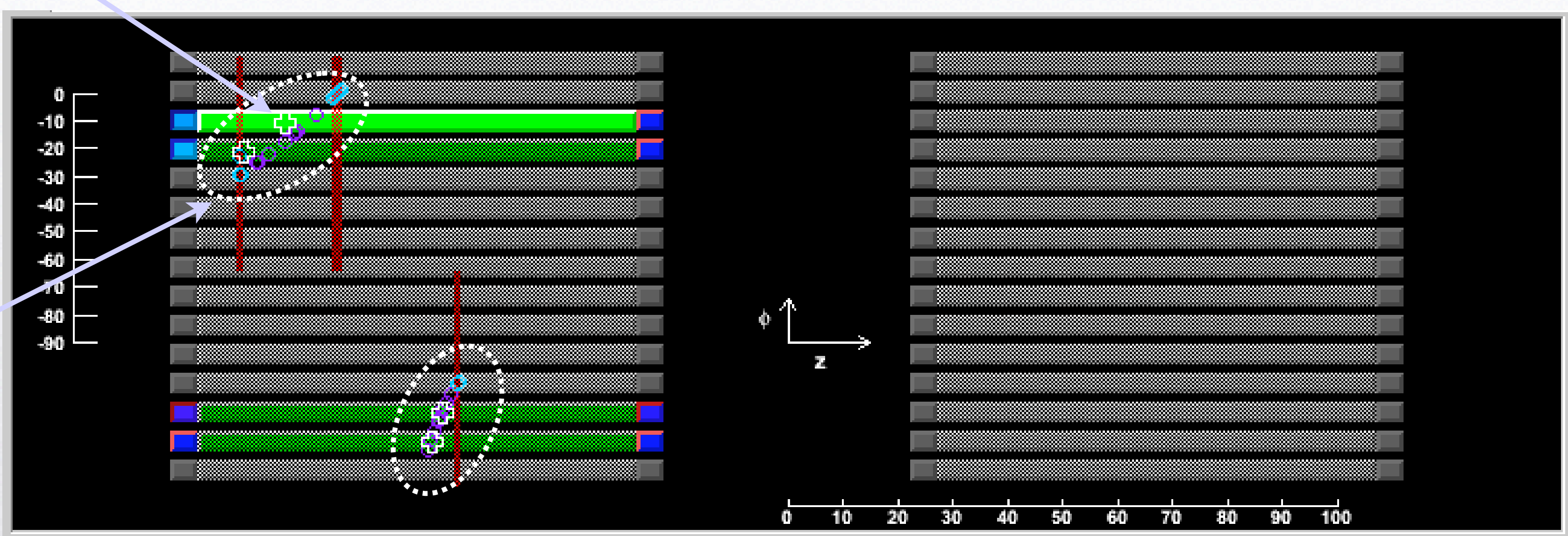
Track Fitter

Combined clustering of TICP and TICZ

Hit clustering of fiber counter

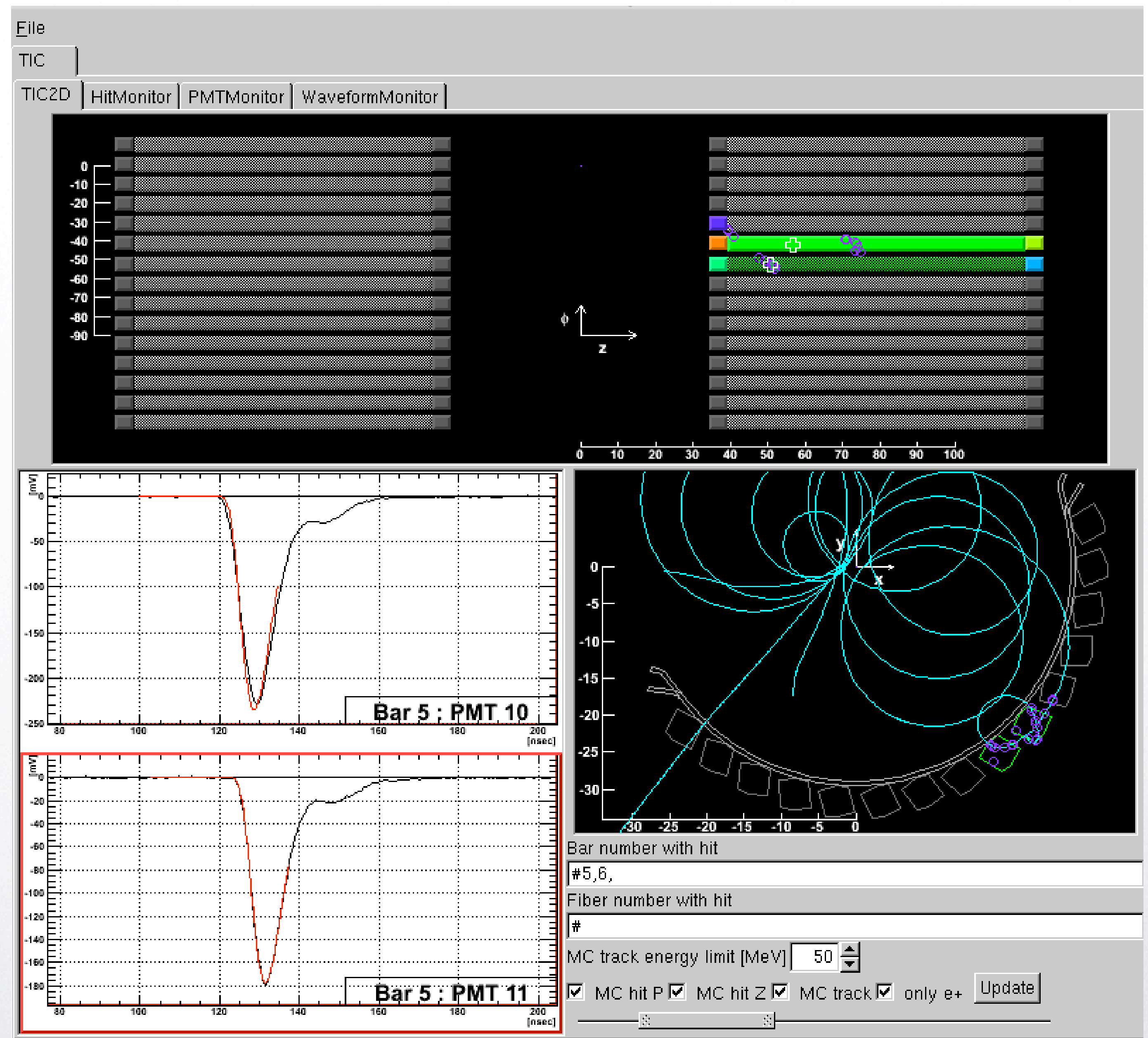
Hit

Cluster

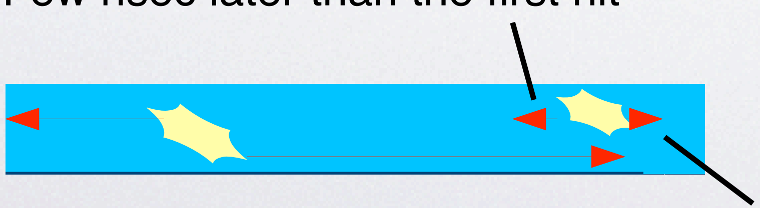


- Pulse finding
- Charge integration
- Time estimation
  - Template fitting
- Double hit identification
  - This is source of tail events of distribution
  - χ^2 of WF template fitting

Example of double hit



Few nsec later than the first hit



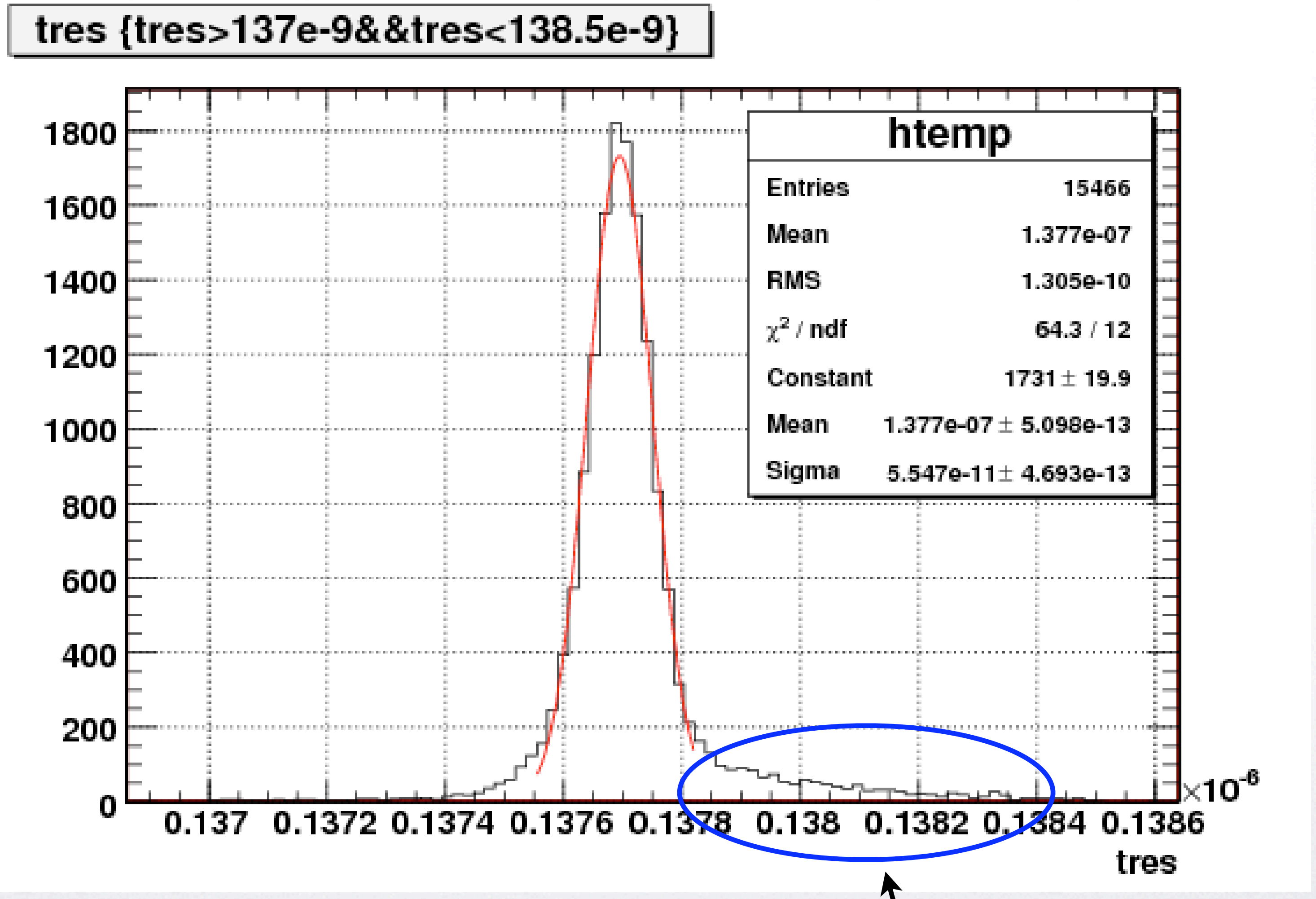
Few nsec earlier than the first hit

TICP: Hit reconstruction : Time

- Impact time with average time
  - (t1 + t2) / 2   average of both side PMT
  - Independent of z position

Pure signal

σT = 55 psec



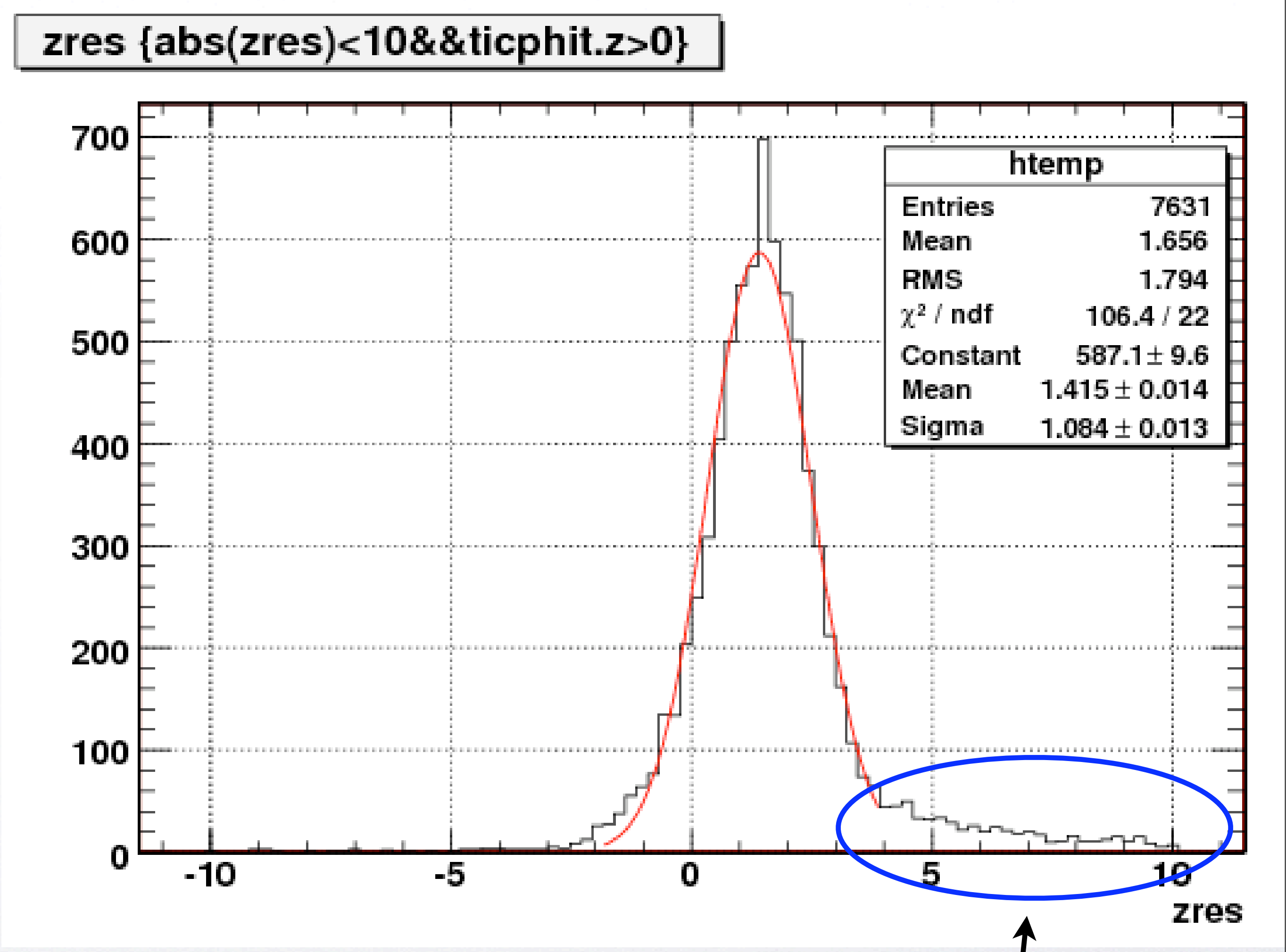
double hits events

- Z estimation from time difference
  - z = (t1 - t2) * Veff / 2 + zc + α
  - Need effective light velocity (each bar)

Pure signal

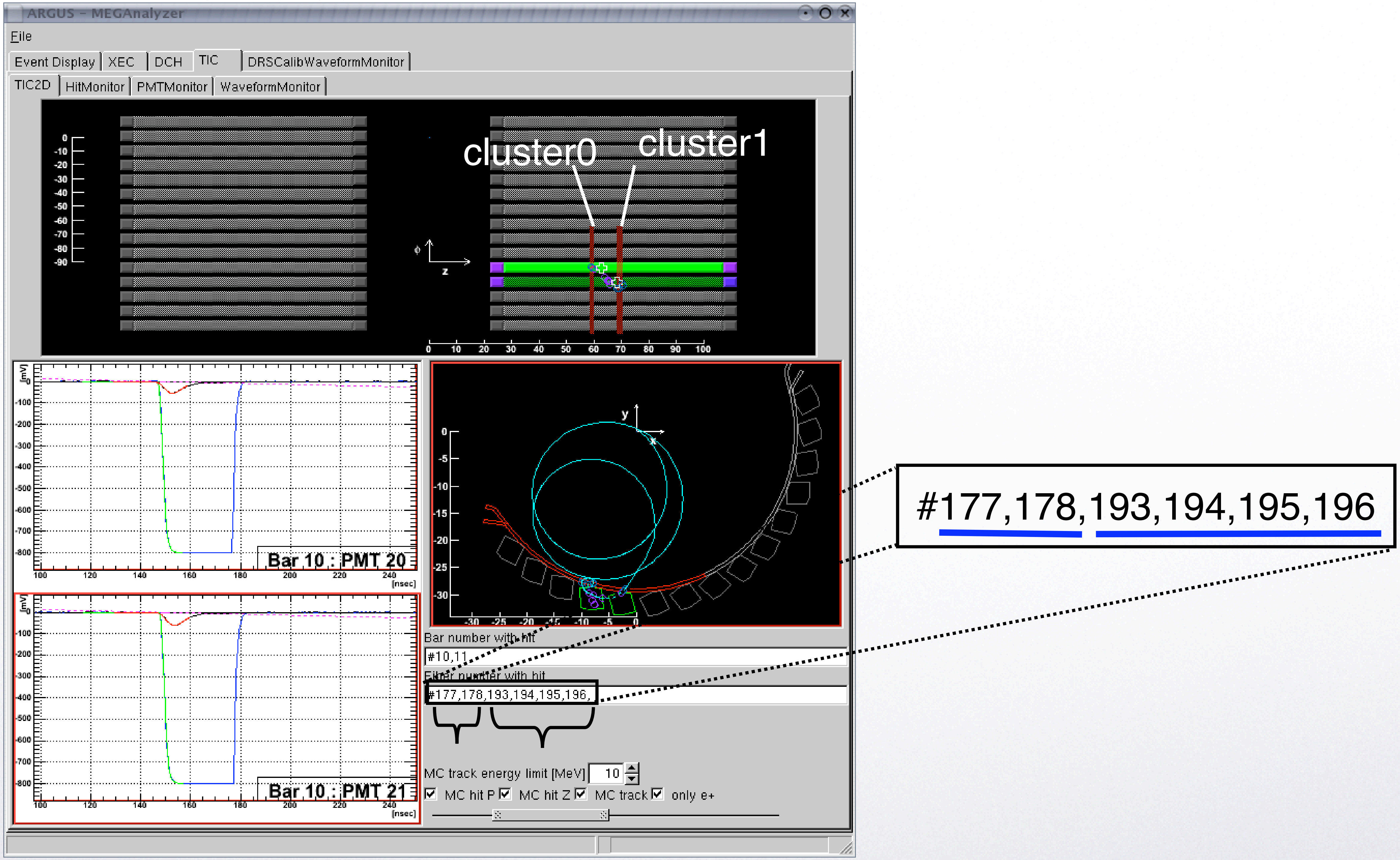
σz = 1.1 cm

(with only TICP counters)



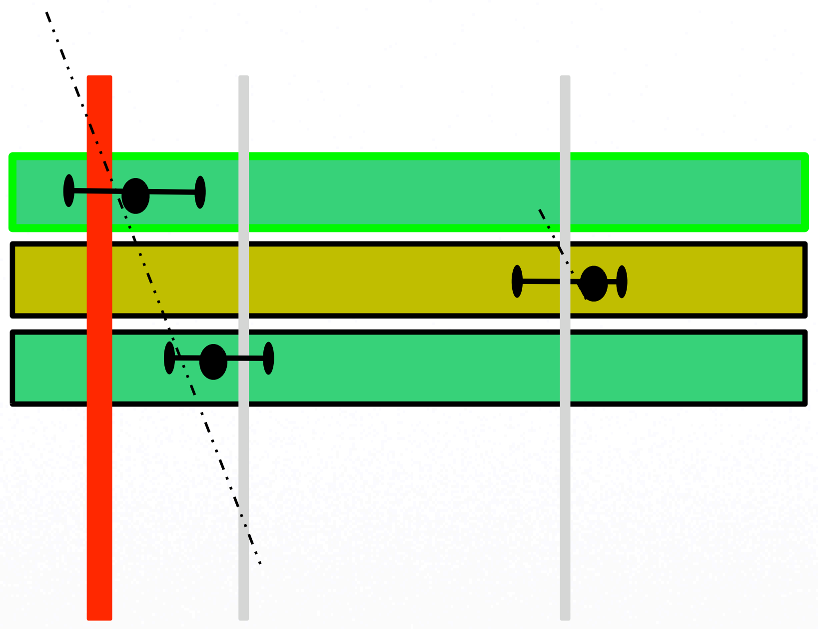
double hits events

TICZ Clustering



Simply clustering consecutive hits

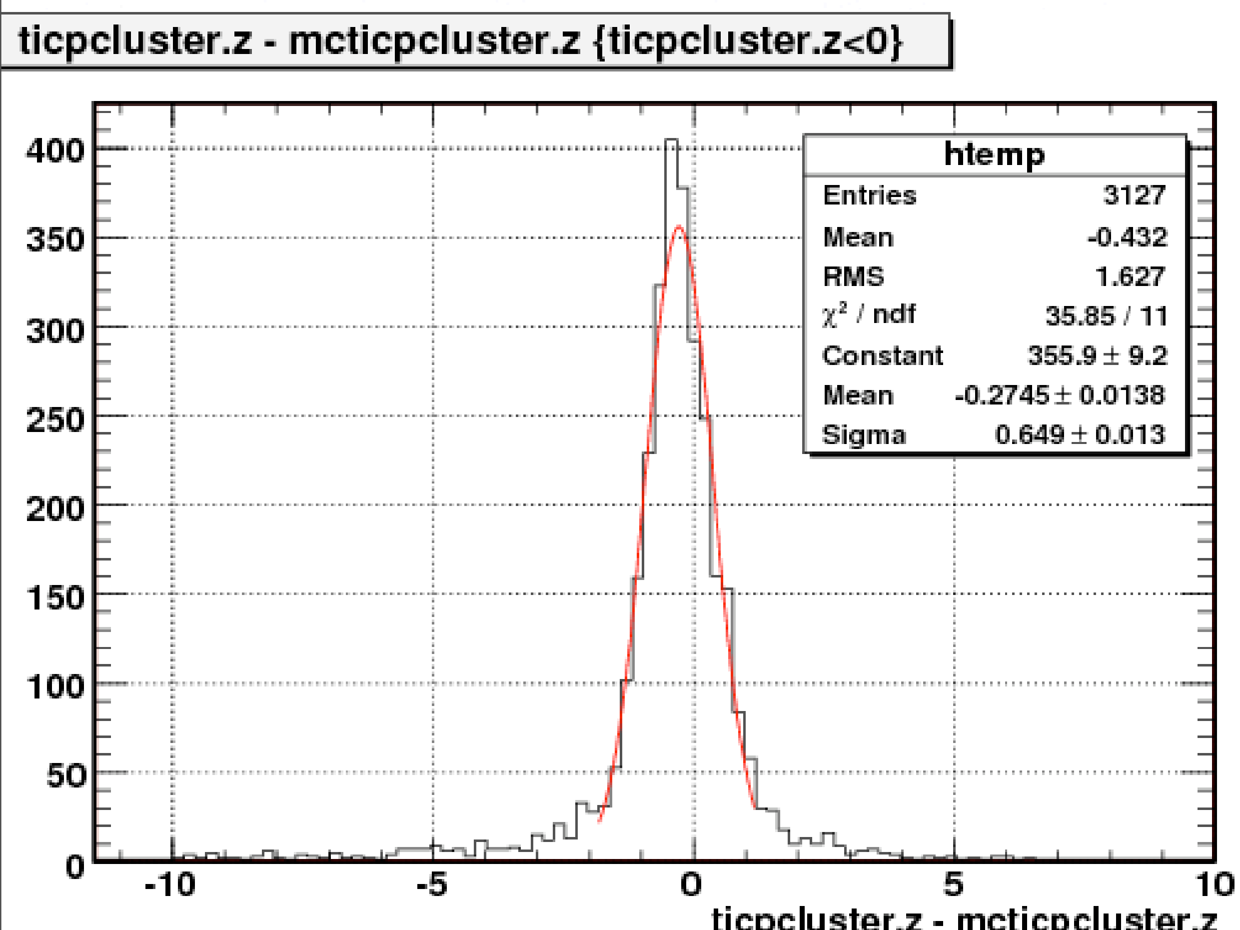
Preliminary



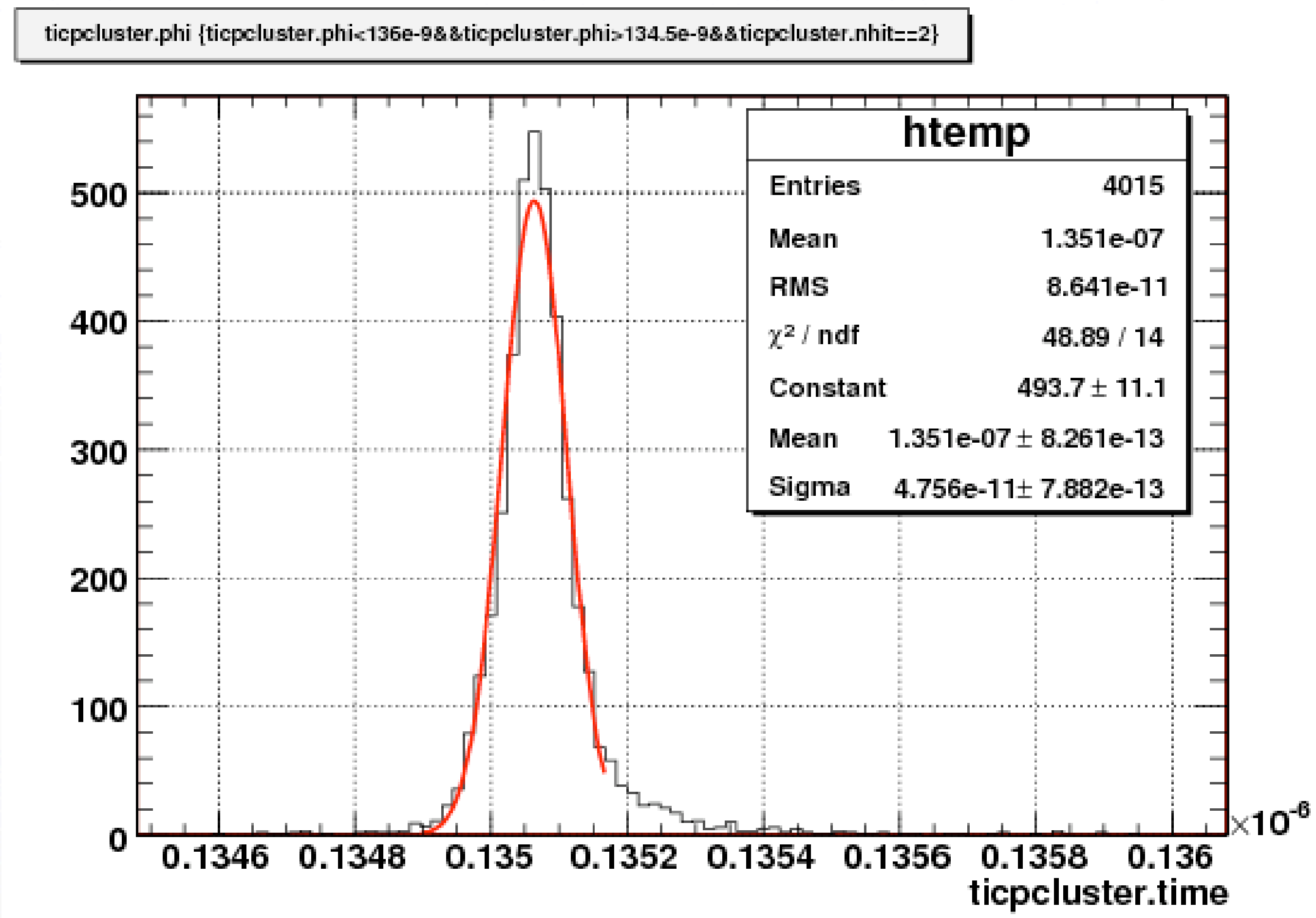
- Clustering
  - Clustering TICP hits according to time
  - Not require consecutiveness
  - Use position only for 'good hits' for average
  - Search for 'front TICZ cluster'
- Computation of cluster properties (z, time)
  - If there is 'front TICZ cluster', combine z position
  - If second 'good hit' is adjacent with the earliest hit in the cluster, take weighted average of hits.

Cluster reconstruction result

Z



Time



For the events with two adjacent hits,

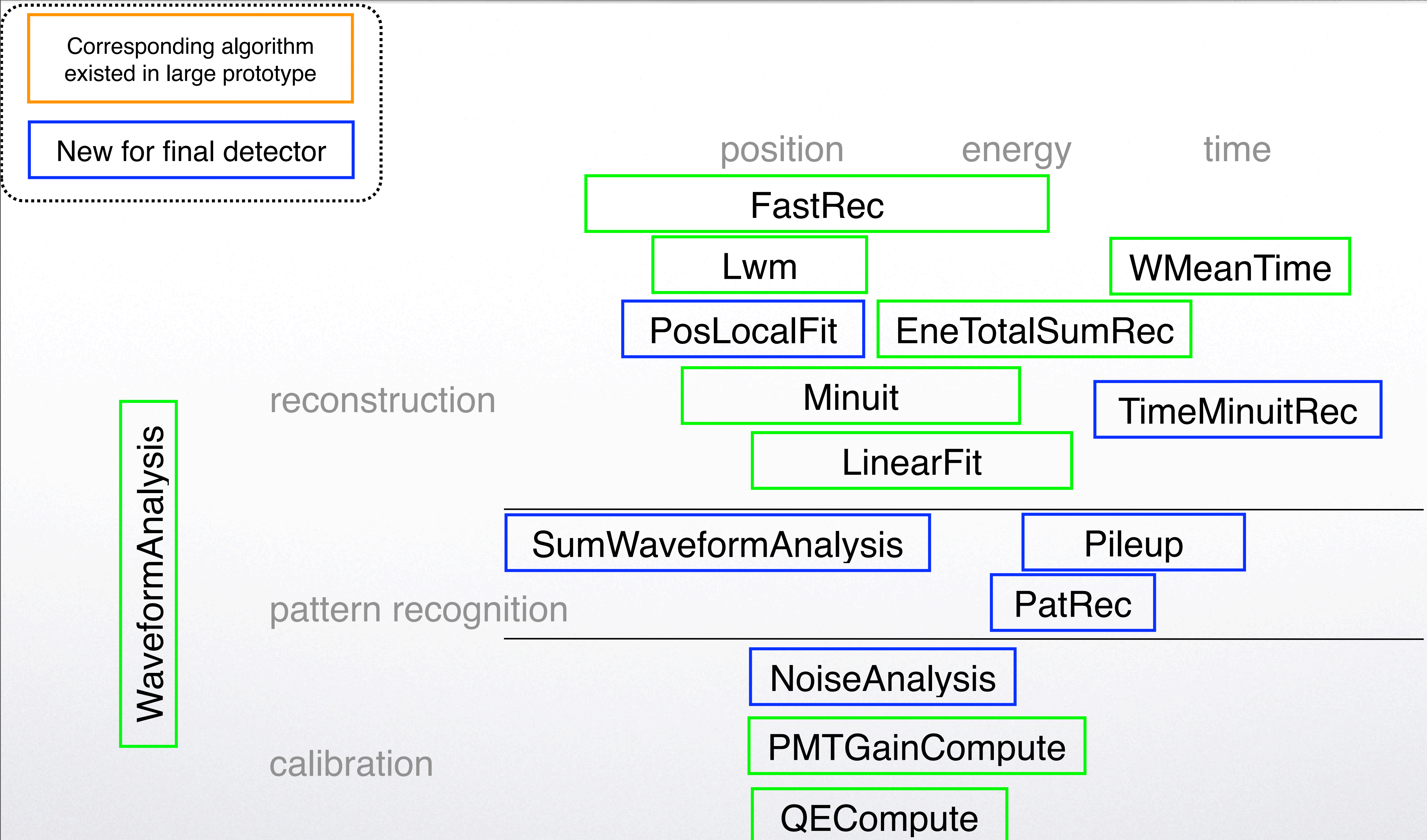
σz = 0.65 cm

σT = 48 psec

With cut of energy loss > 5MeV, it improves to 43 psec

Calorimeter

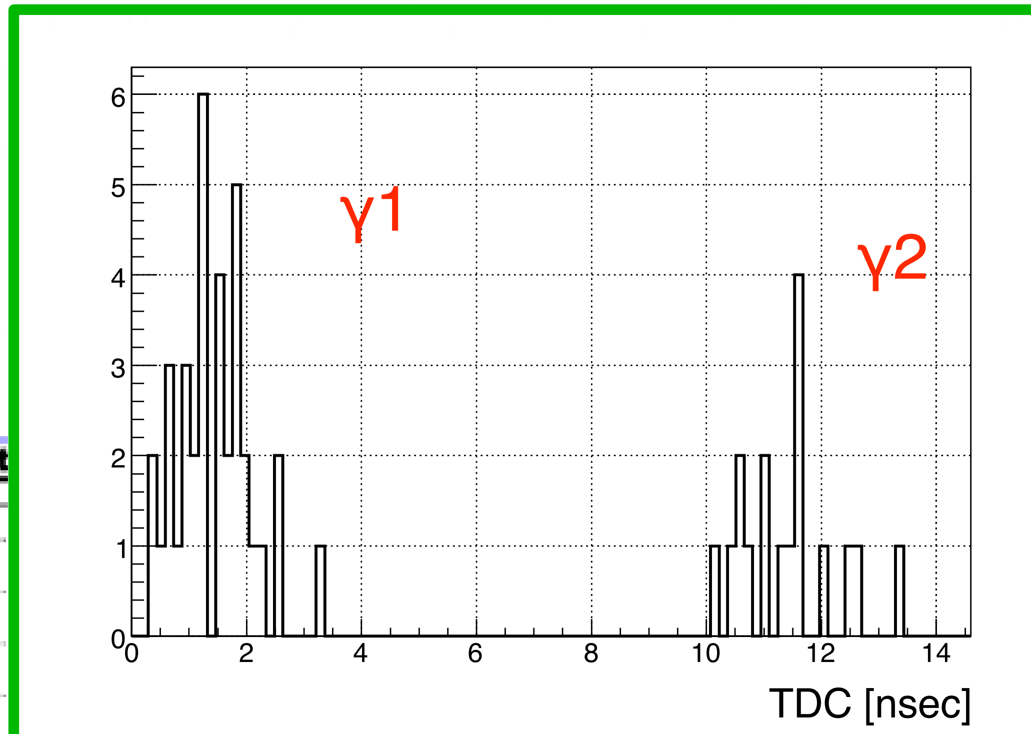
XEC Tasks



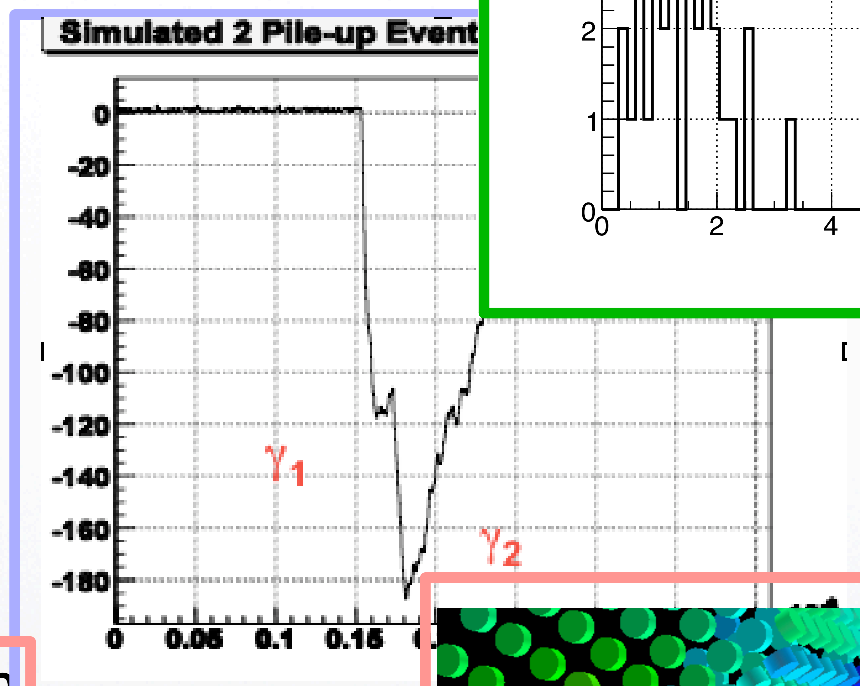
Corresponding algorithm existed in large prototype
New for final detector
position
energy
time
FastRec
Lwm
WMeanTime
PosLocalFit
EneTotalSumRec
reconstruction
Minuit
TimeMinuitRec
LinearFit
WaveformAnalysis
SumWaveformAnalysis
Pileup
PatRec
pattern recognition
NoiseAnalysis
PMTGainCompute
calibration
QECompute

Pileup Identification

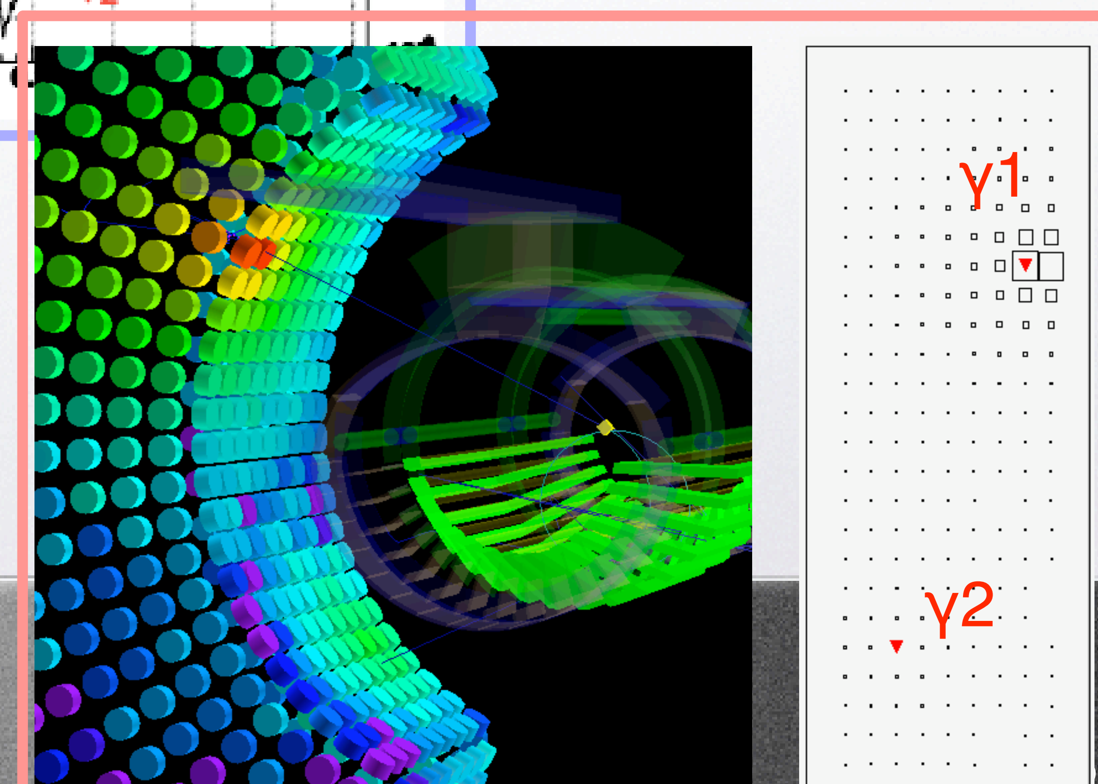
- Time separation
  - Max time difference between earliest and latest PMT
  - Two different parameter set
    - time strict & low threshold methods



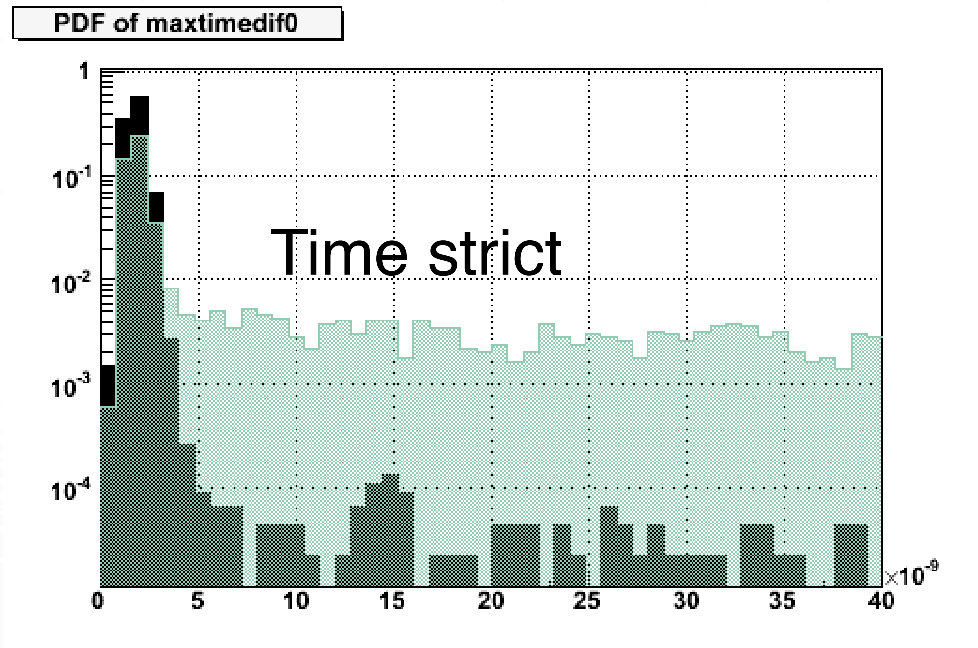
- Waveform analysis
  - Peak search of sum waveform
  - Differential method

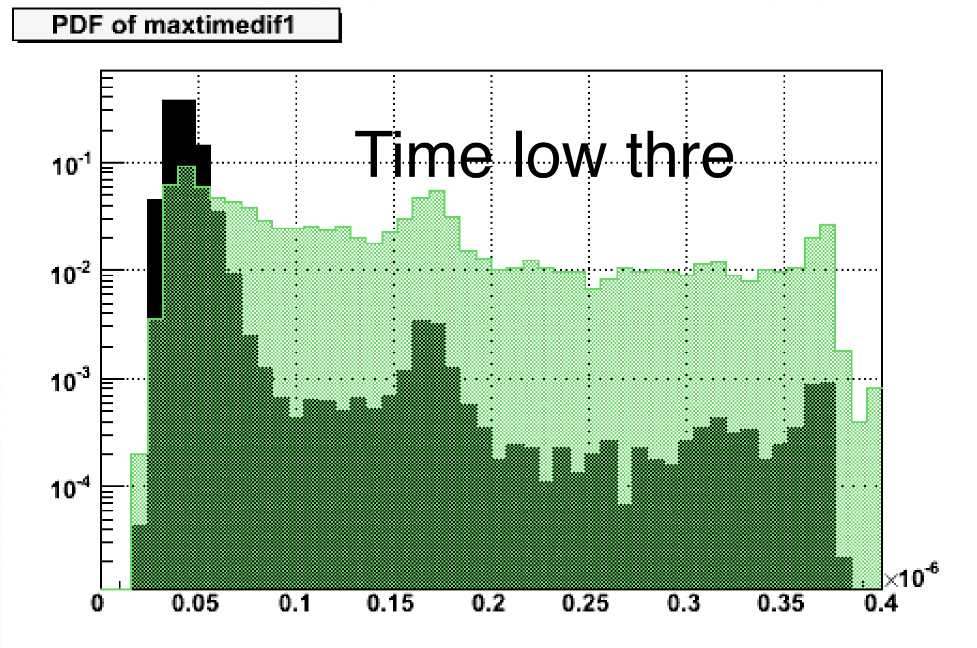


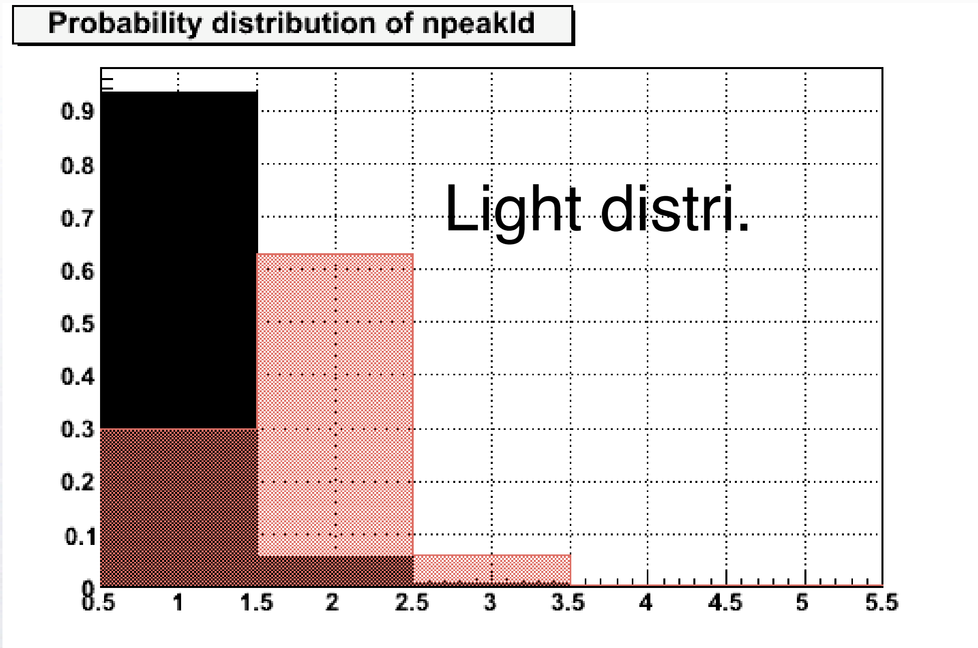
- Pattern recognition of light distribution
  - Three different algorithms
    - Peak search of X,Y projection
    - 2D peak search
    - Dipole

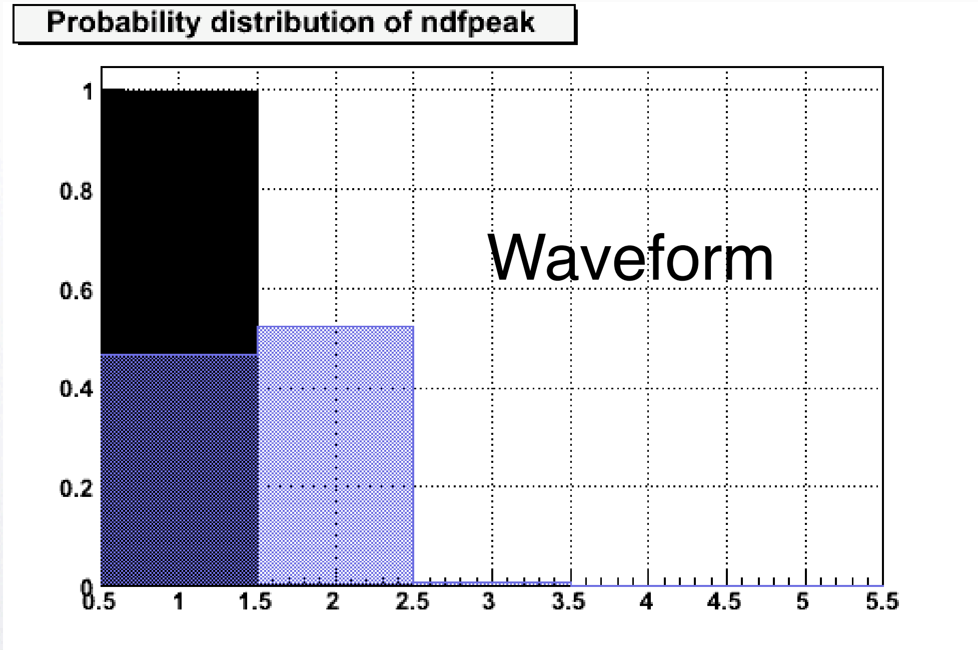


PDF from MC of several algorithms







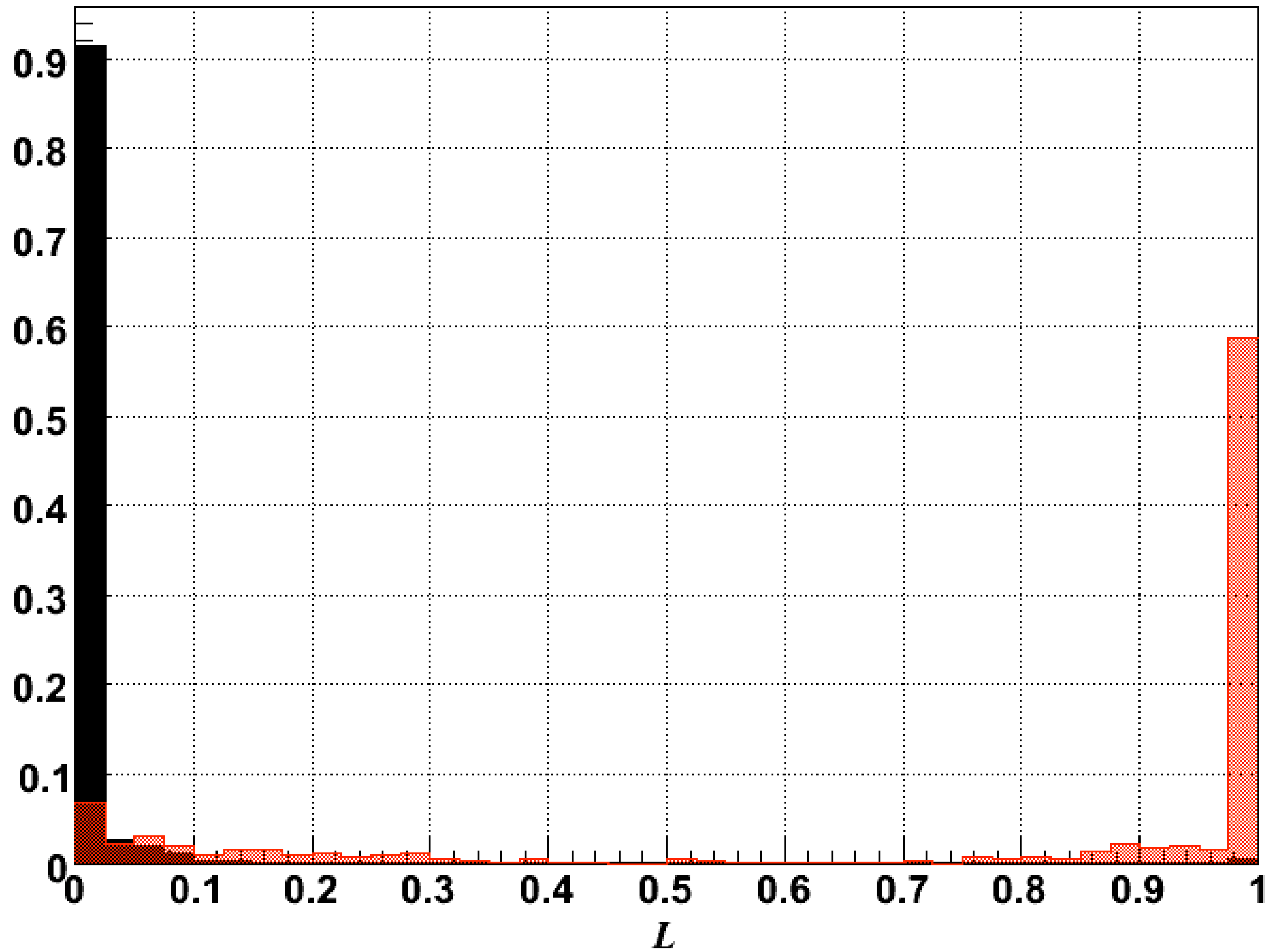


Black histograms are distribution for single gamma events
Color histograms are distribution for pile-up events

Combination of several methods (Likelihood)

- Likelihood
  - L = A * P(pileup) * p(pileup | ***X***)

    = A * P(pileup) * p(pileup | T) * P(pileup | LD) * P(pileup | WF)



Threshold	0.5	0.1	0.05
ID Power	76 %	86 %	91 %
Miss-ID	1.3 %	2.8 %	6.0 %

Vertical scale is normalized

Physics

Outline

- Maximum likelihood

1. How to build likelihood function?
2. Sensitivity study using PDF from known detector performance. (using different data set from 1.)

- Blind $\mu \rightarrow e\gamma$ analysis
- Possible analysis procedure

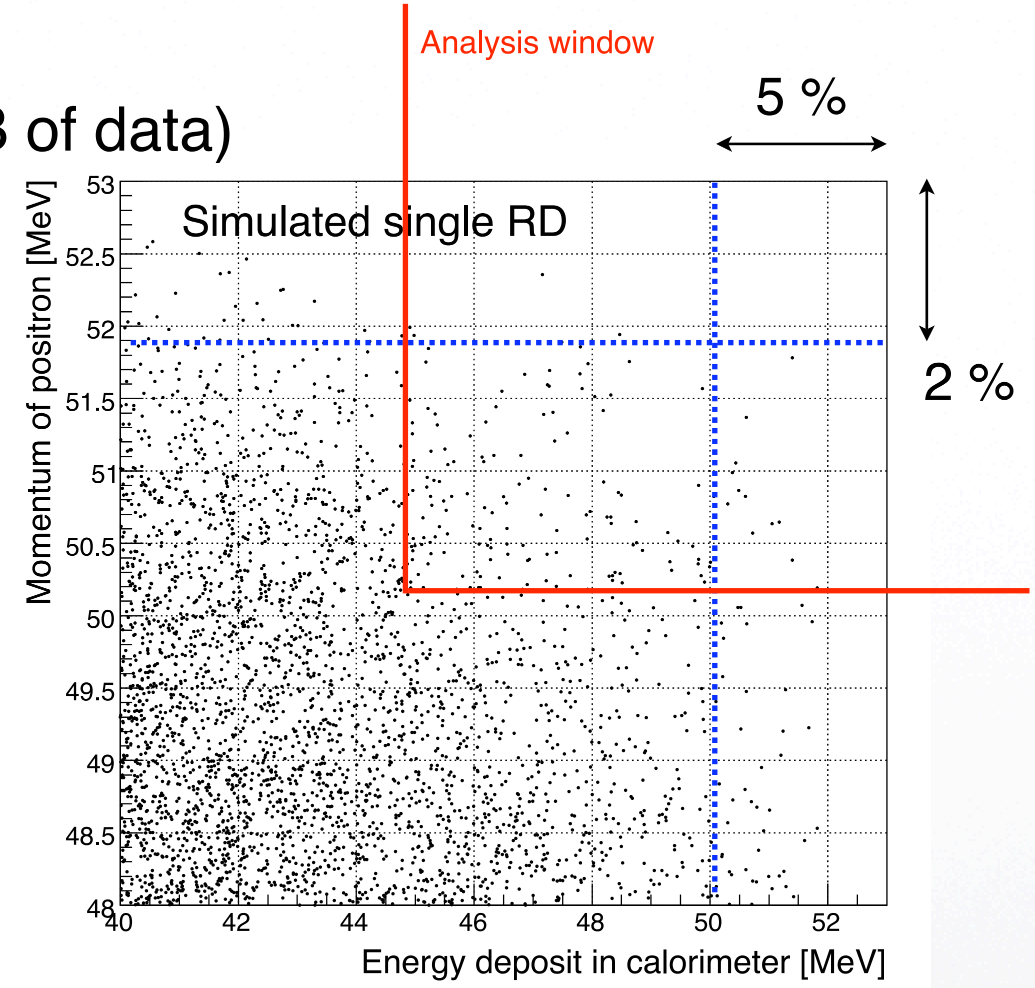
Maximum likelihood

- Building blocks
  - Measured parameters for i-th event:
    - $x_i = (E_e, E_\gamma, T_{e\gamma}, \theta_{e\gamma})$
  - Number of events:
    - s (signal), s' (radiative decay), b (accidental background), N = s+s'+b (total)
  - Probability density function (PDF):
    - S(xi) (signal), S'(xi) (radiative decay), B(xi) (accidental)
- Partial probability to measure xi
  - $P(x_i) = (sS(x_i) + s'S'(x_i) + bB(x_i)) / N$
- Likelihood function which is maximized for "best" estimators
  - $L(s) = \Pi P(x_i) = \Pi(sS(x_i) + s'S'(x_i) + (N-s-s')B(x_i)) / N$

How to build PDF ?

MC Production & Analysis procedure

- MC Production (>1 week with 19 CPUs and 650 GB of data)
  - Enhanced statistics around signal region
    - δx=0.1, δy=0.25, δt=0.75nsec δθeγ=115mrad
    - (47.5 < Ee, 44.9 < Eγ)
  - Signal (3×10^4), accidental BG (10^5), prompt BG (10^5)
  - Pileups assuming muon rate of 3×10^7/sec
  - Trigger simulation
  - Only RD γ, no AIF γ



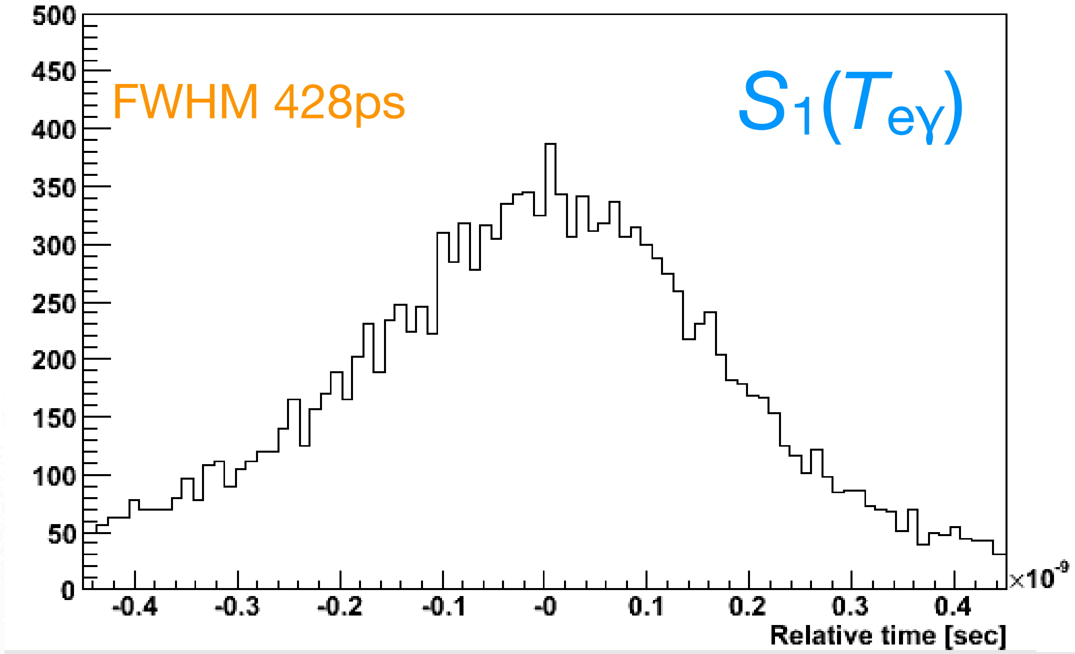
- Analysis
  - Window : δx=0.05, δy=0.15, δt=0.45nsec δθeγ=70mrad
  - Reconstruction algorithm (Already implemented reconstructions in analyzer)
  - γ
    - pileup rejection, weighted charge sum, depth cut (>3.5cm), weighted mean time, Minuit fitting position
  - Positron
    - Kalman fitting with MC hits
  - **Not with best resolutions**

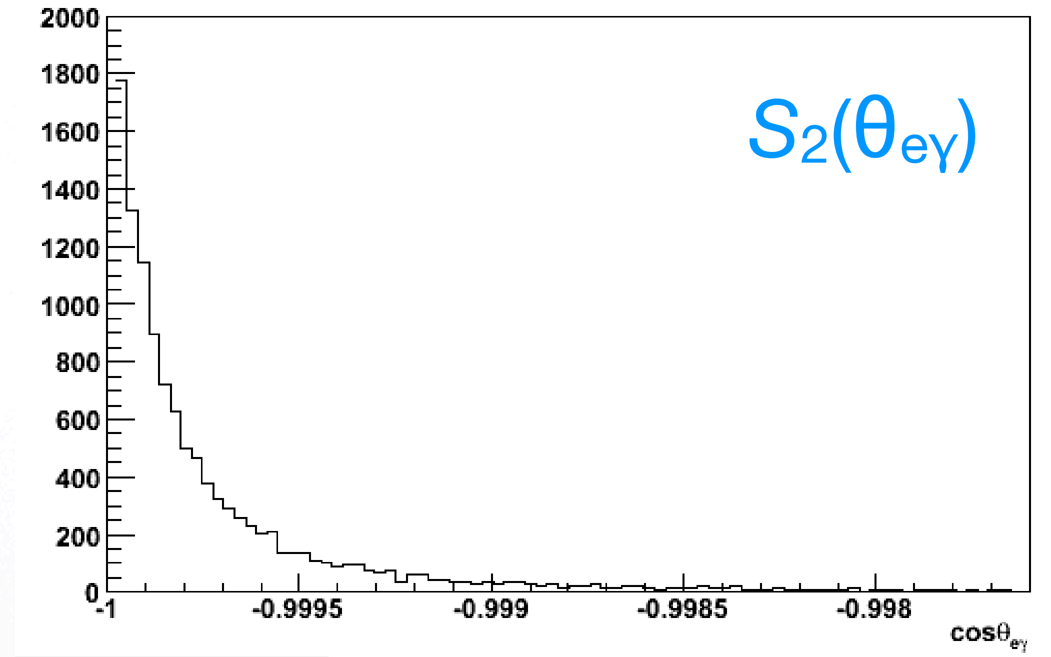
Calculation of Signal PDF

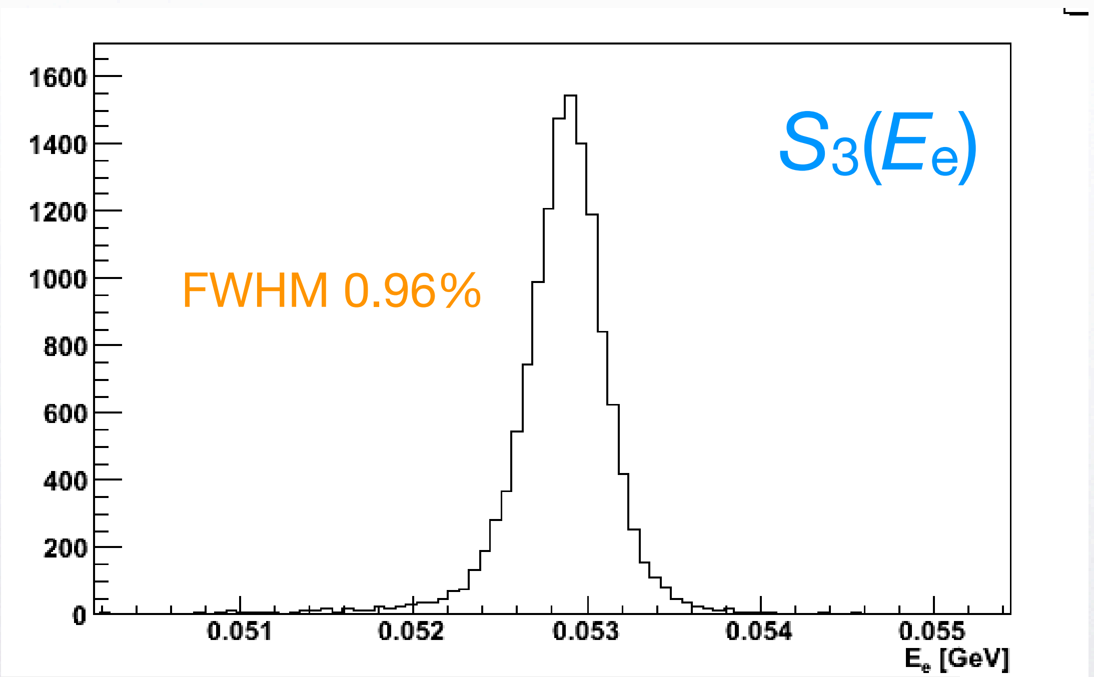
- $S(x_i) = S_1(T_{e\gamma})S_2(\theta_{e\gamma})S_3(E_e)S_4(E_\gamma)$
- All parameters are statistically independent.
- Response functions

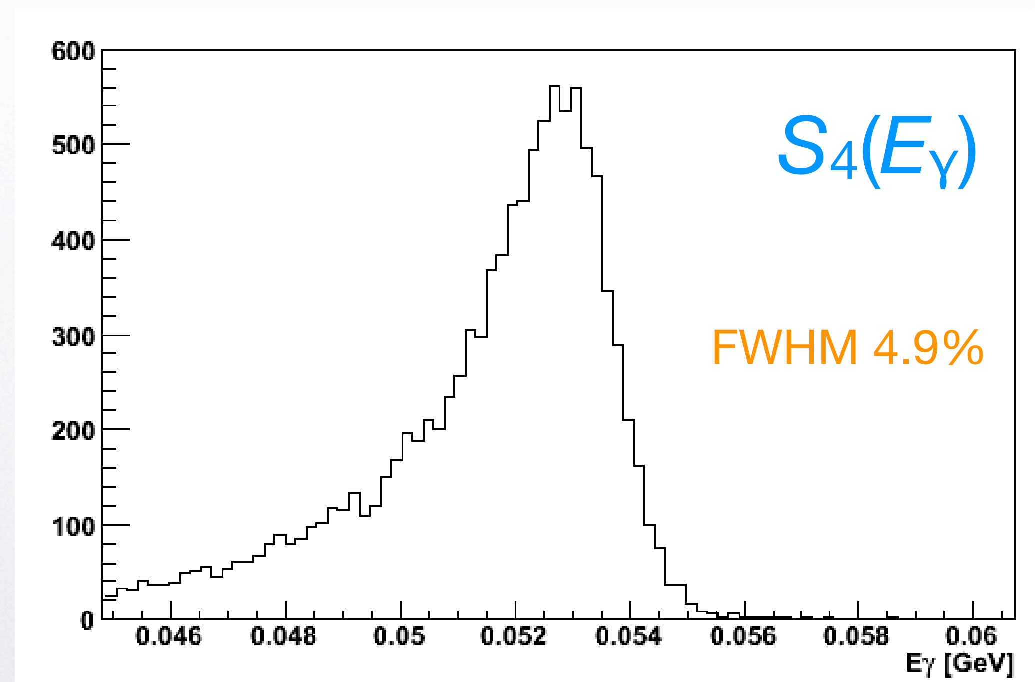
	How to obtain?	Remarks
$S_1(T_{e\gamma})$	RD data	Energy dependence
$S_2(\theta_{e\gamma})$	MC	How to check?
$S_3(E_e)$	Measured Michel spectrum	
$S_4(E_\gamma)$	π^0 data (55MeV-γ)	How to scale to 52.8MeV?

Calculation of Signal PDF







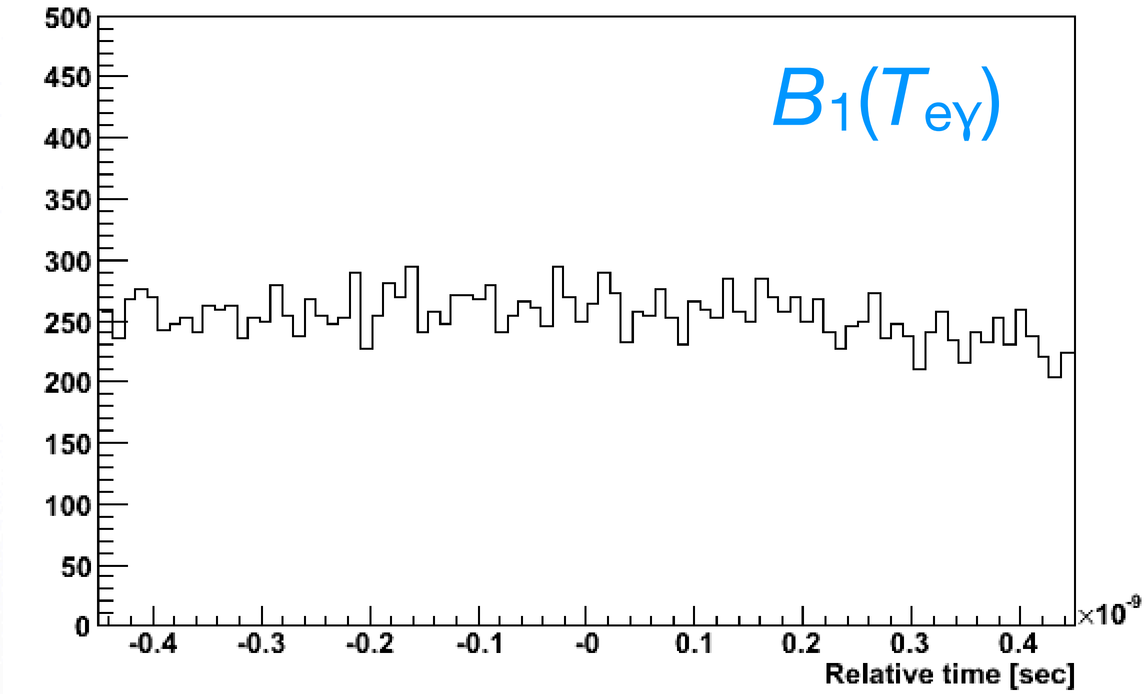


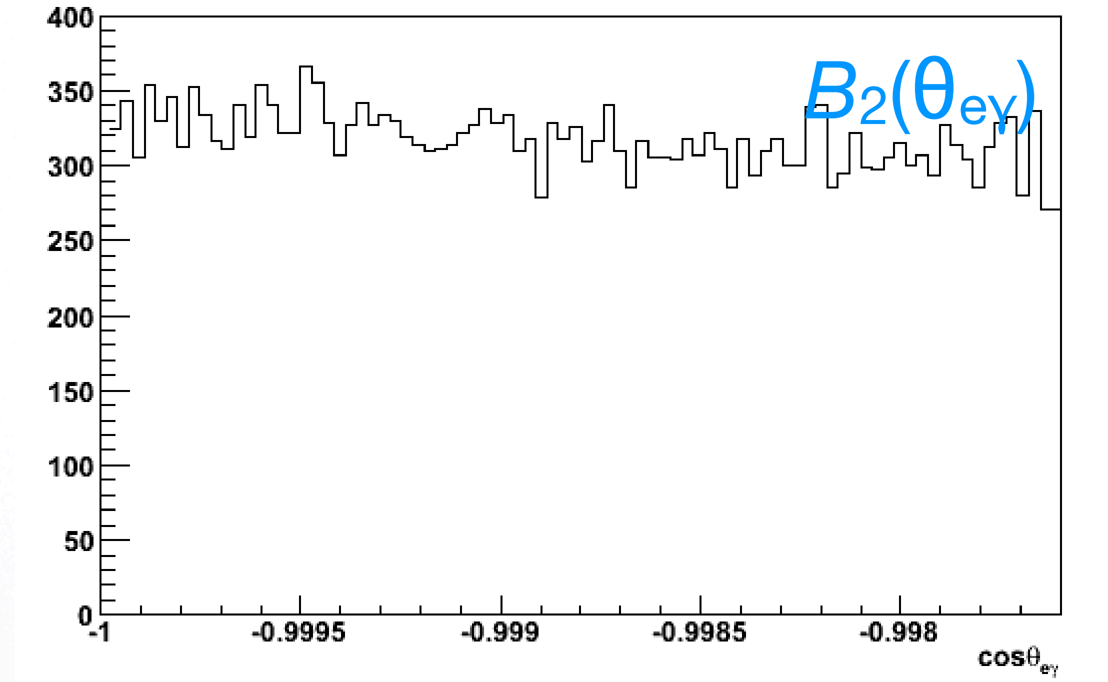
Calculation of Accidental BG PDF

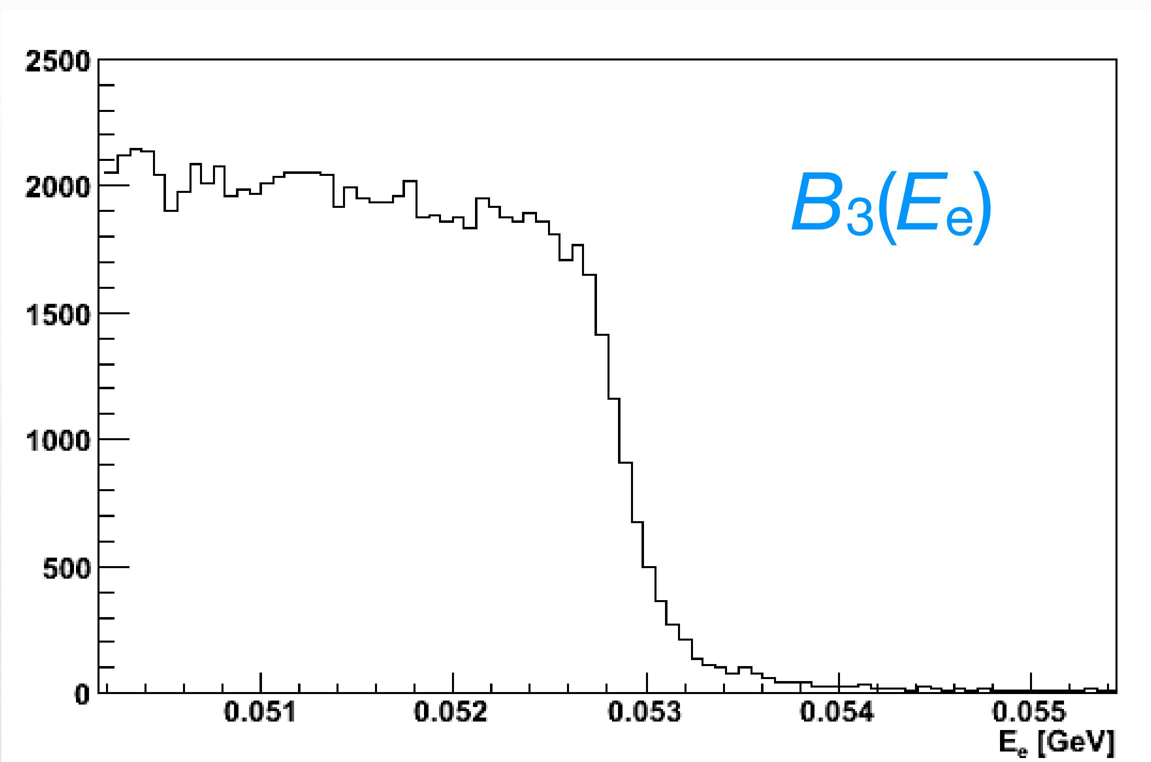
- $B(xi) = B1(Te\gamma)B2(\theta e\gamma)B3(Ee)B4(E\gamma)$
- All parameters are statistically independent.

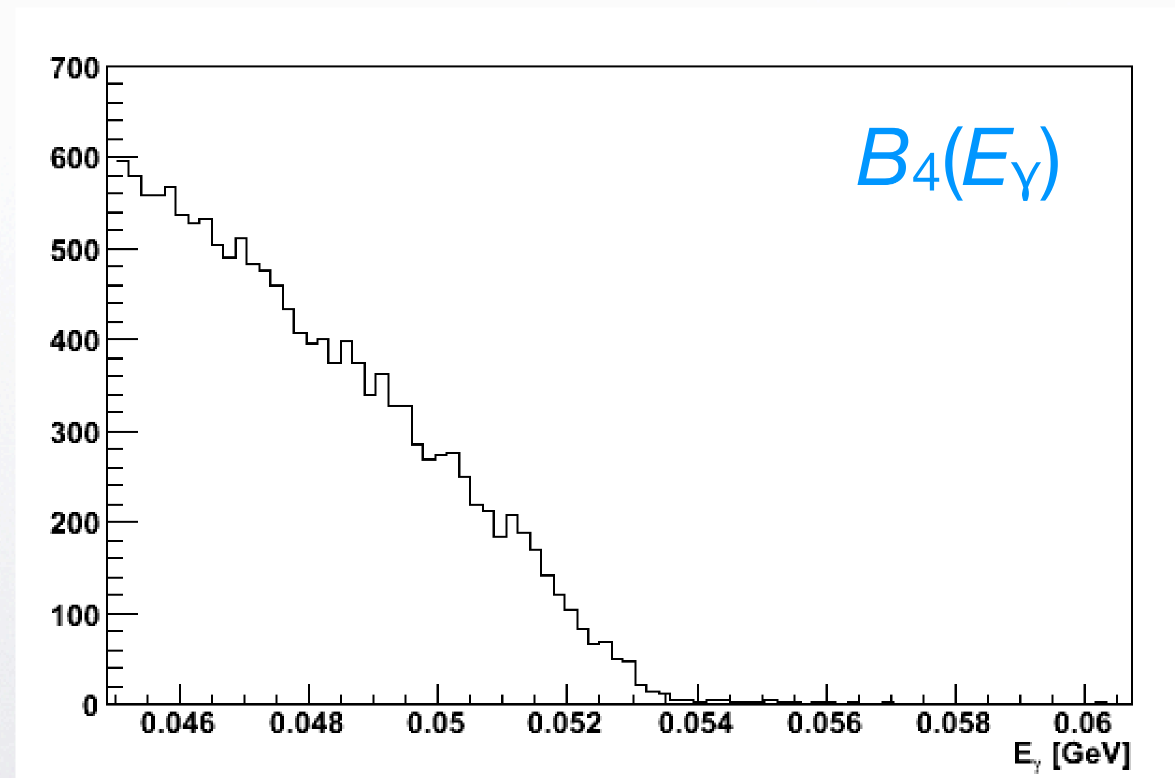
	How to obtain?	Remarks
$B1(Te\gamma)$	Flat distribution	
$B2(\theta e\gamma)$	BG data	Flat distribution?
$B3(Ee)$	Measured Michel spectrum	Need special run?
$B4(E\gamma)$	Measured Xe single spectrum	Need special run?

Calculation of Accidental BG PDF







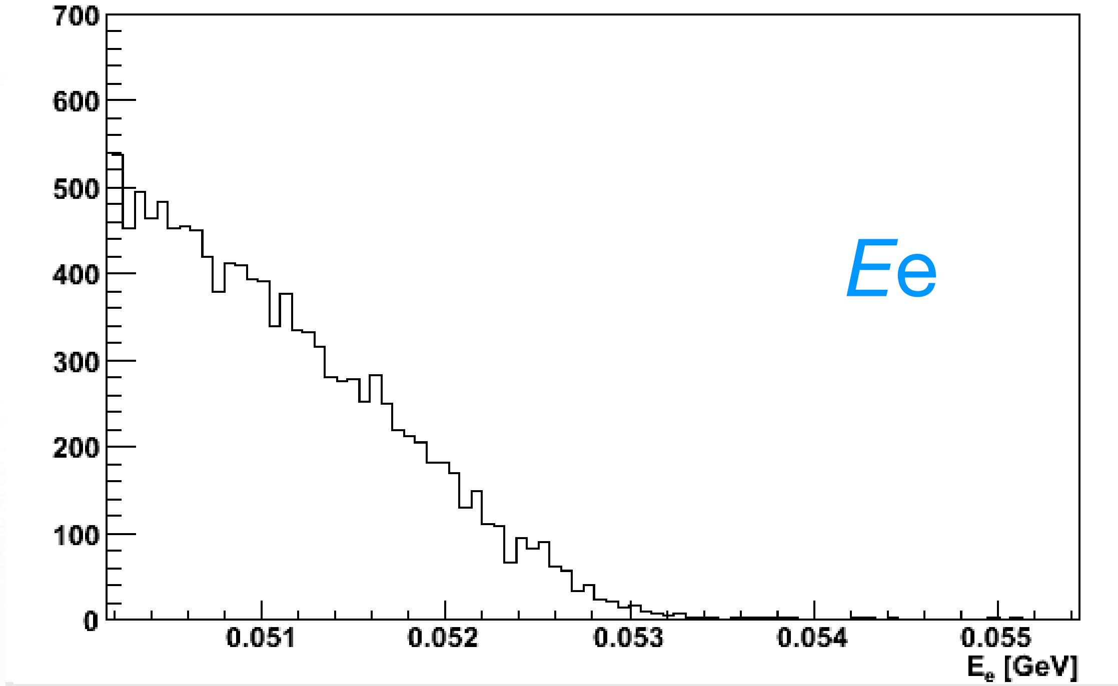


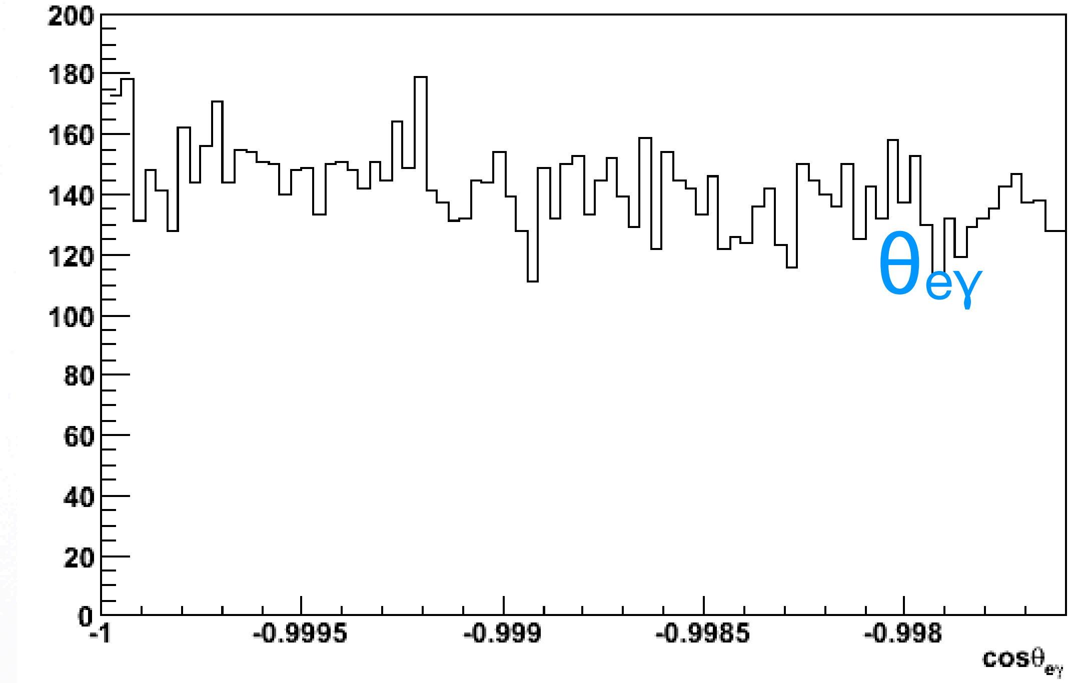
Calculation of Radiative Decay PDF

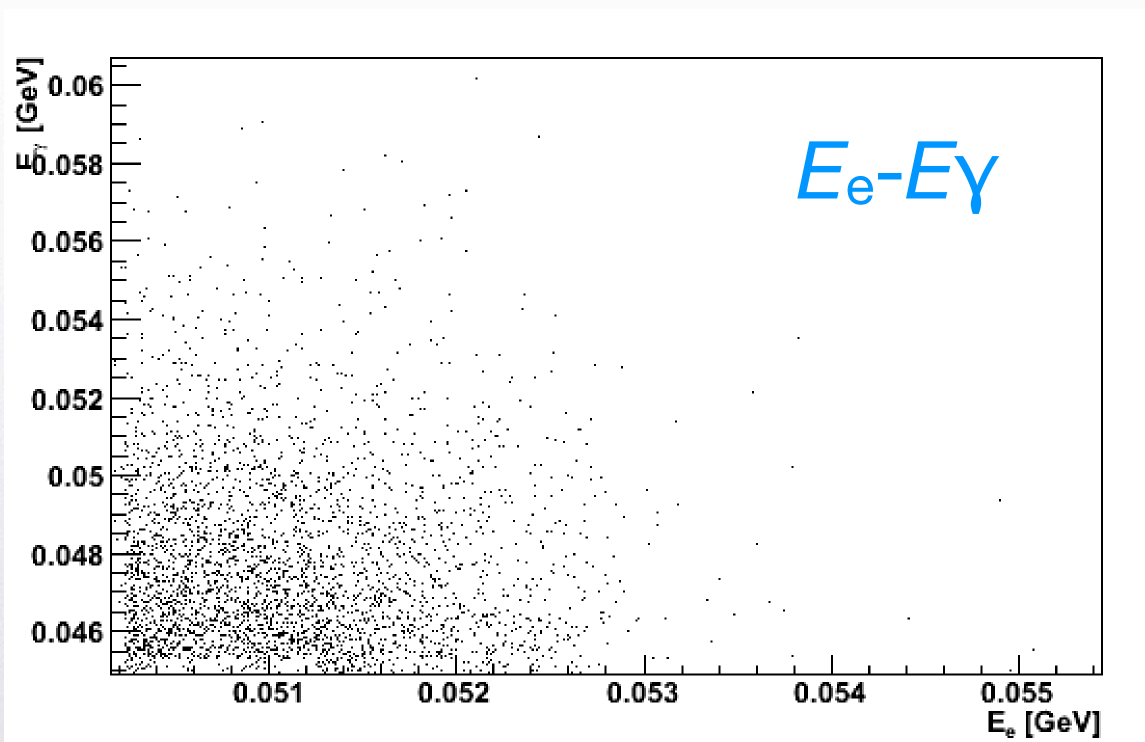
- $S'(x_i) = S'_1(T_{e\gamma})S'_2(\theta_{e\gamma}, E_e, E_\gamma)$
  - $S'(x_i) = S'(T_{e\gamma}, \theta_{e\gamma}, E_e, E_\gamma)$ if there is an energy dependence in timing resolution.

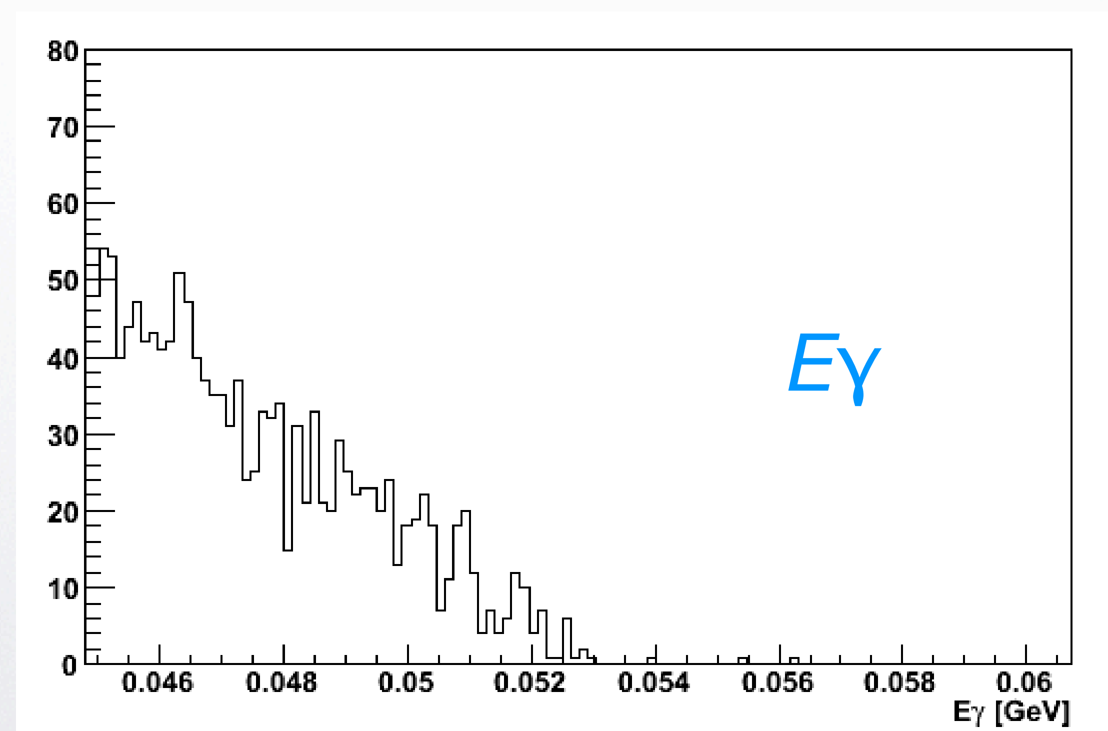
	How to obtain?	Remarks
$S'(T_{e\gamma}, \theta_{e\gamma}, E_e, E_\gamma)$	MC	from RD data?

Calculation of Radiative Decay PDF

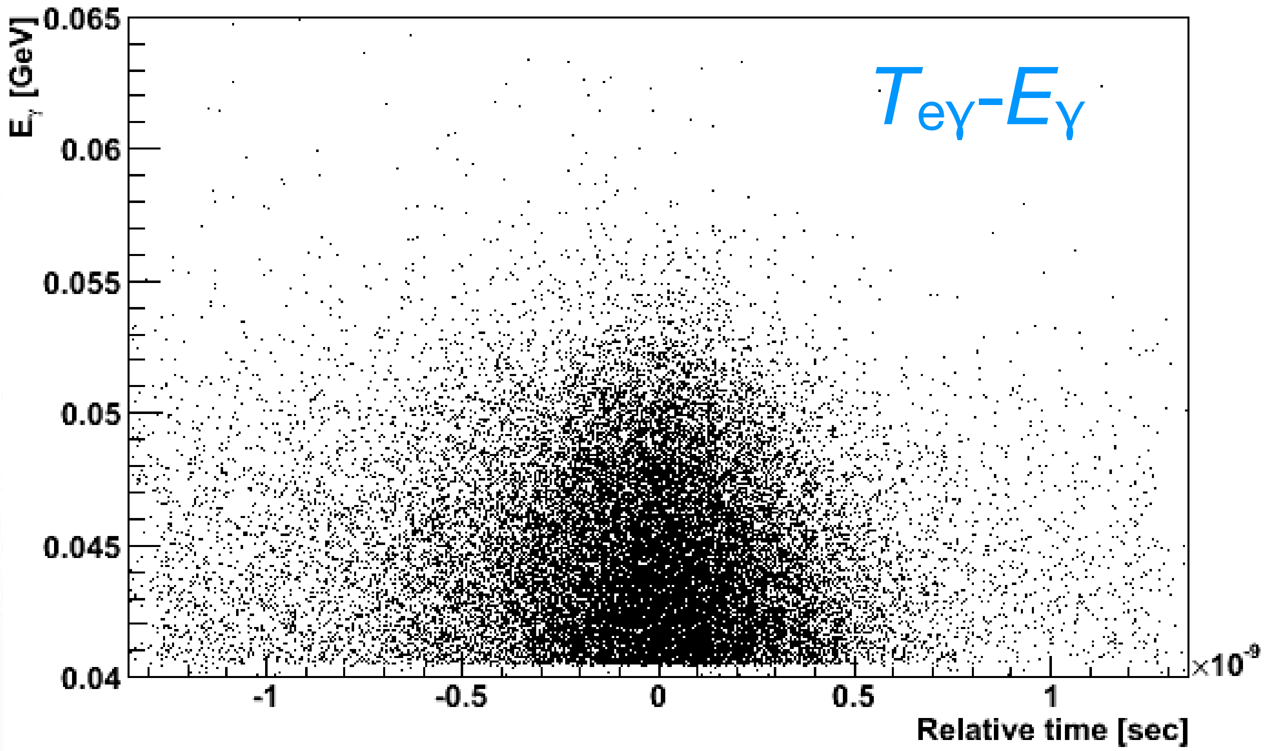


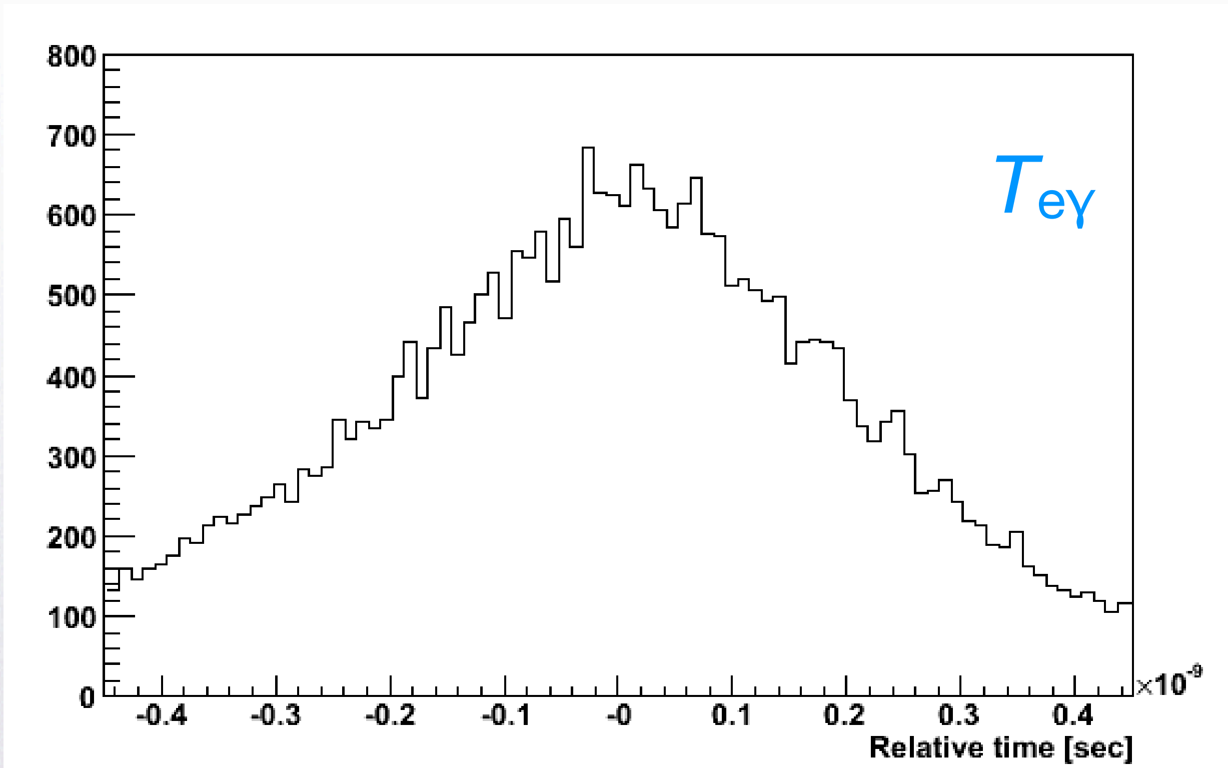




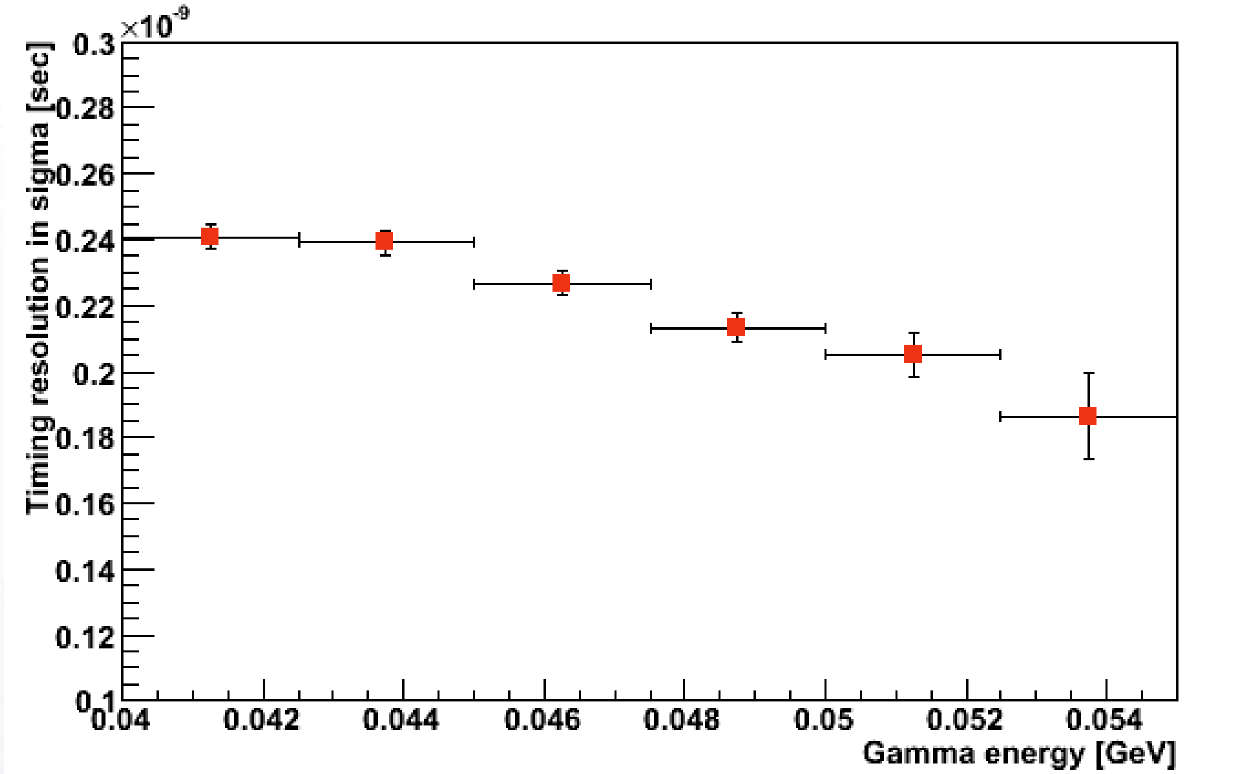


Calculation of Radiative Decay PDF





E_γ dependence of timing resolution



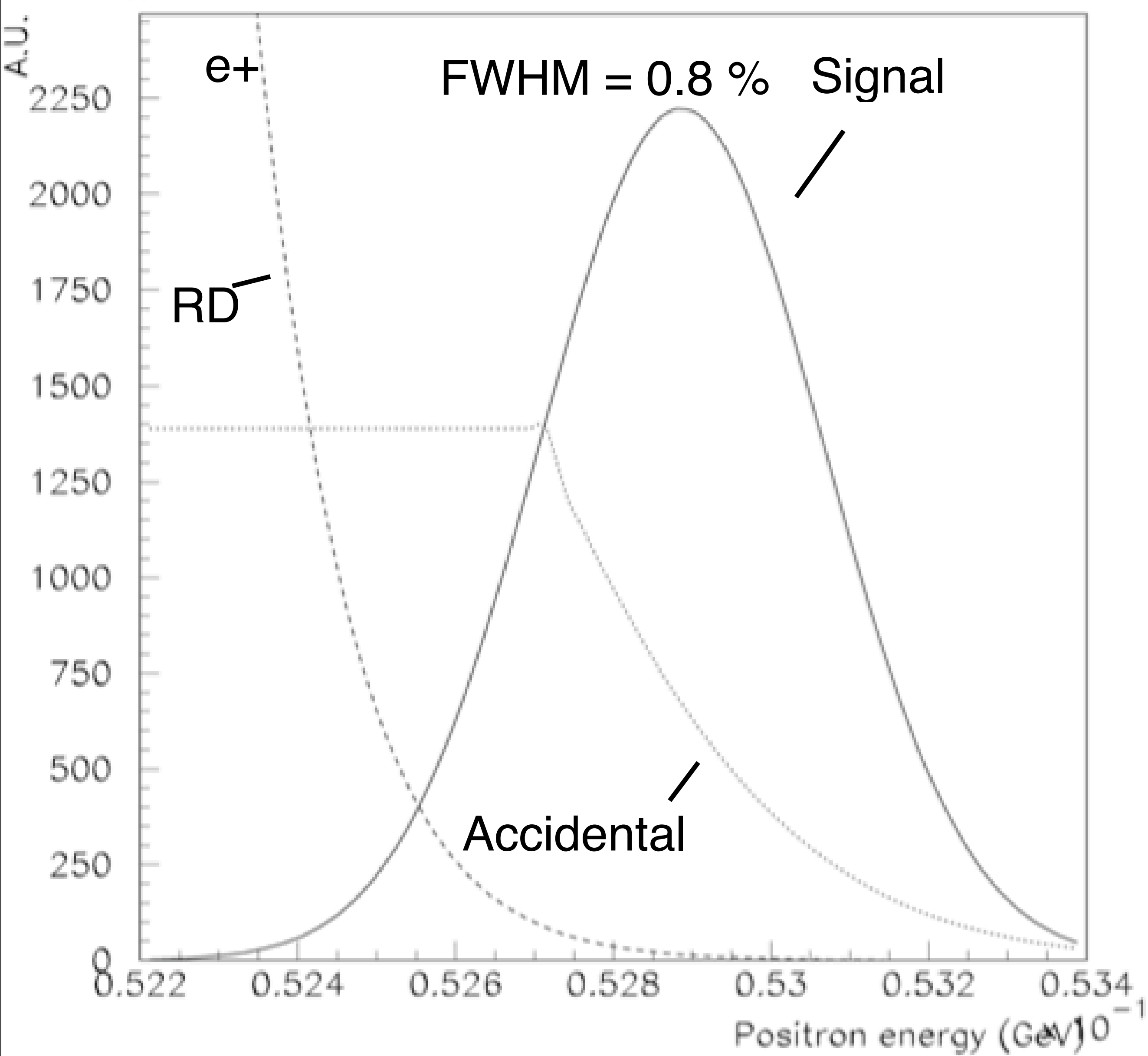
Sensitivity study

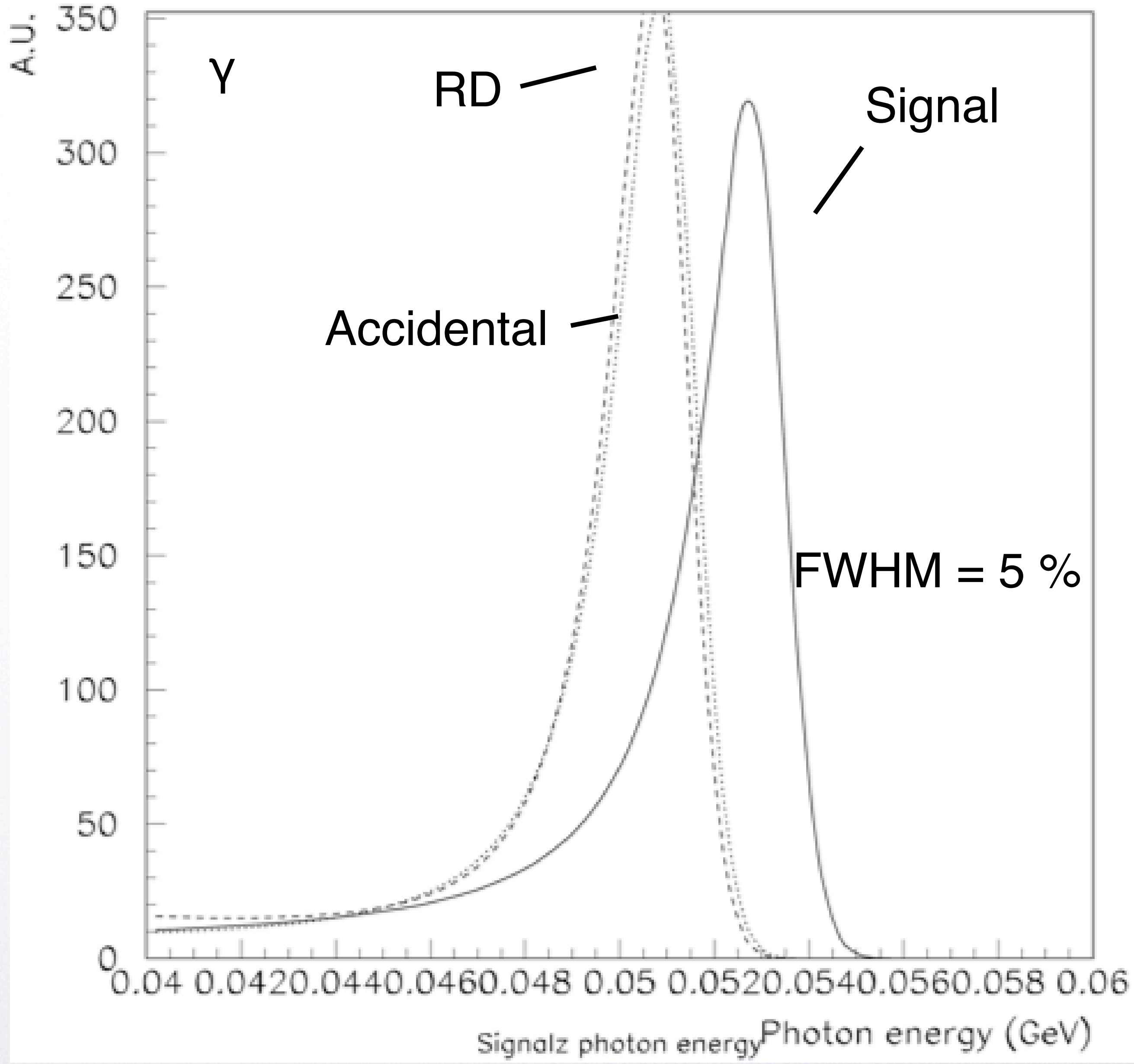
- Sensitivity determination study using PDF assuming known detector performance (shown in following slides)
- Based on Feldman-Cousins prescription.
- Efficiency 0.65(positron) × 0.4 (gamma)
- Muon stopping rate : 3×10^{7}
- Running time : 3 years

Data set is different from previous slides, since two studies were done in parallel.

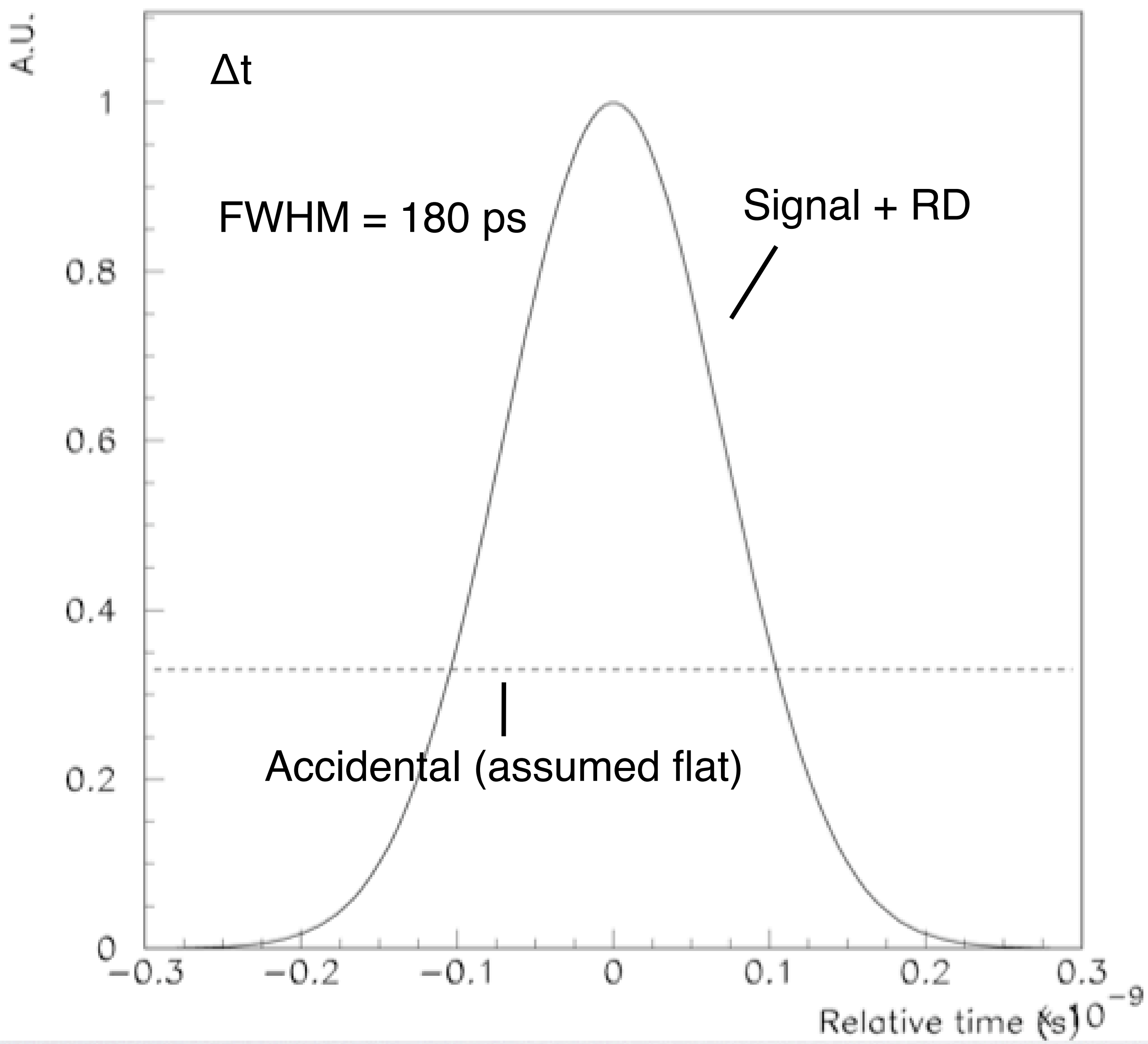
G. Feldman & R. Cousins, PRD 57 (1998) 3873

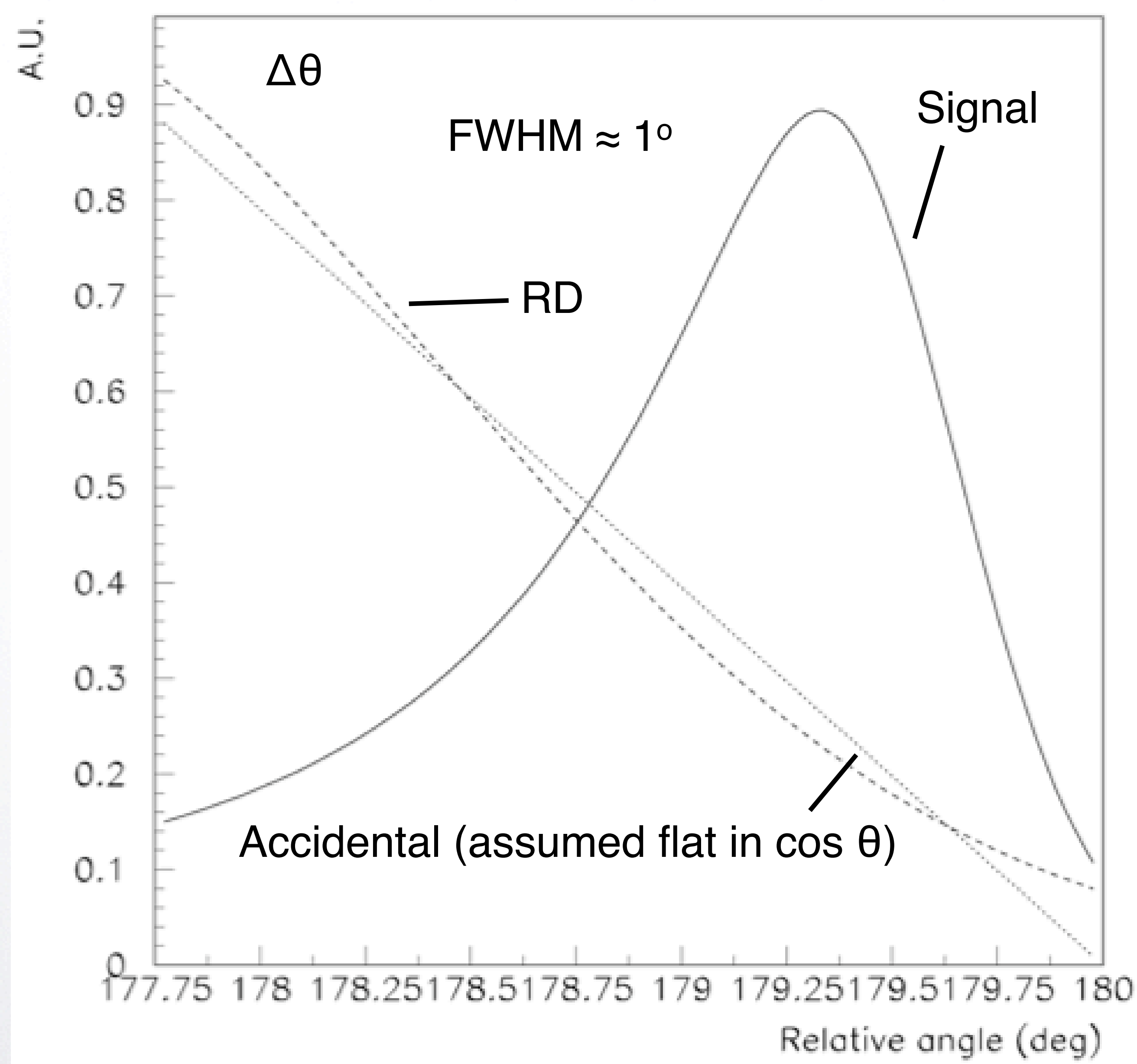
Sensitivity study : PDF 1



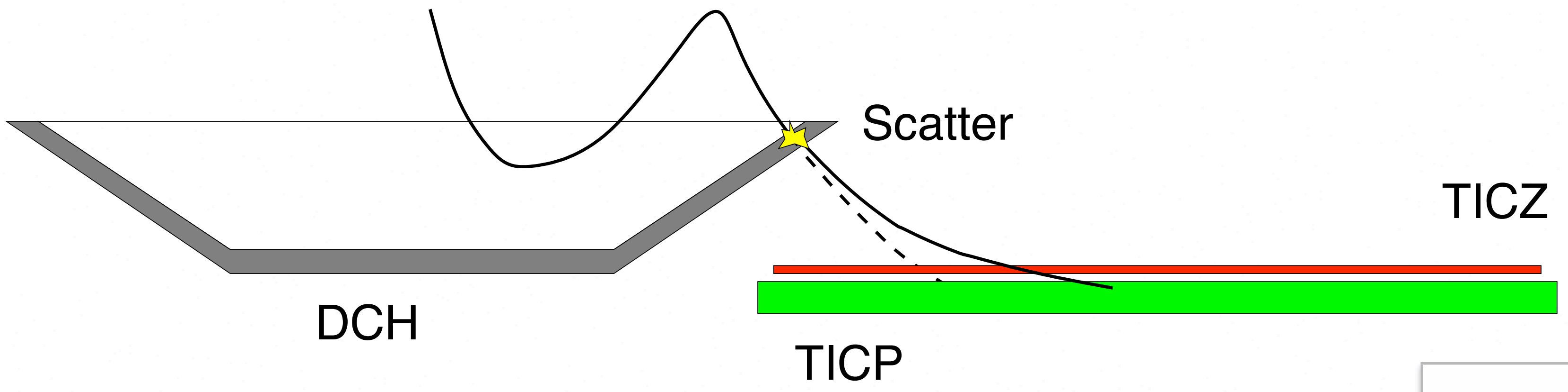


Sensitivity study : PDF 2





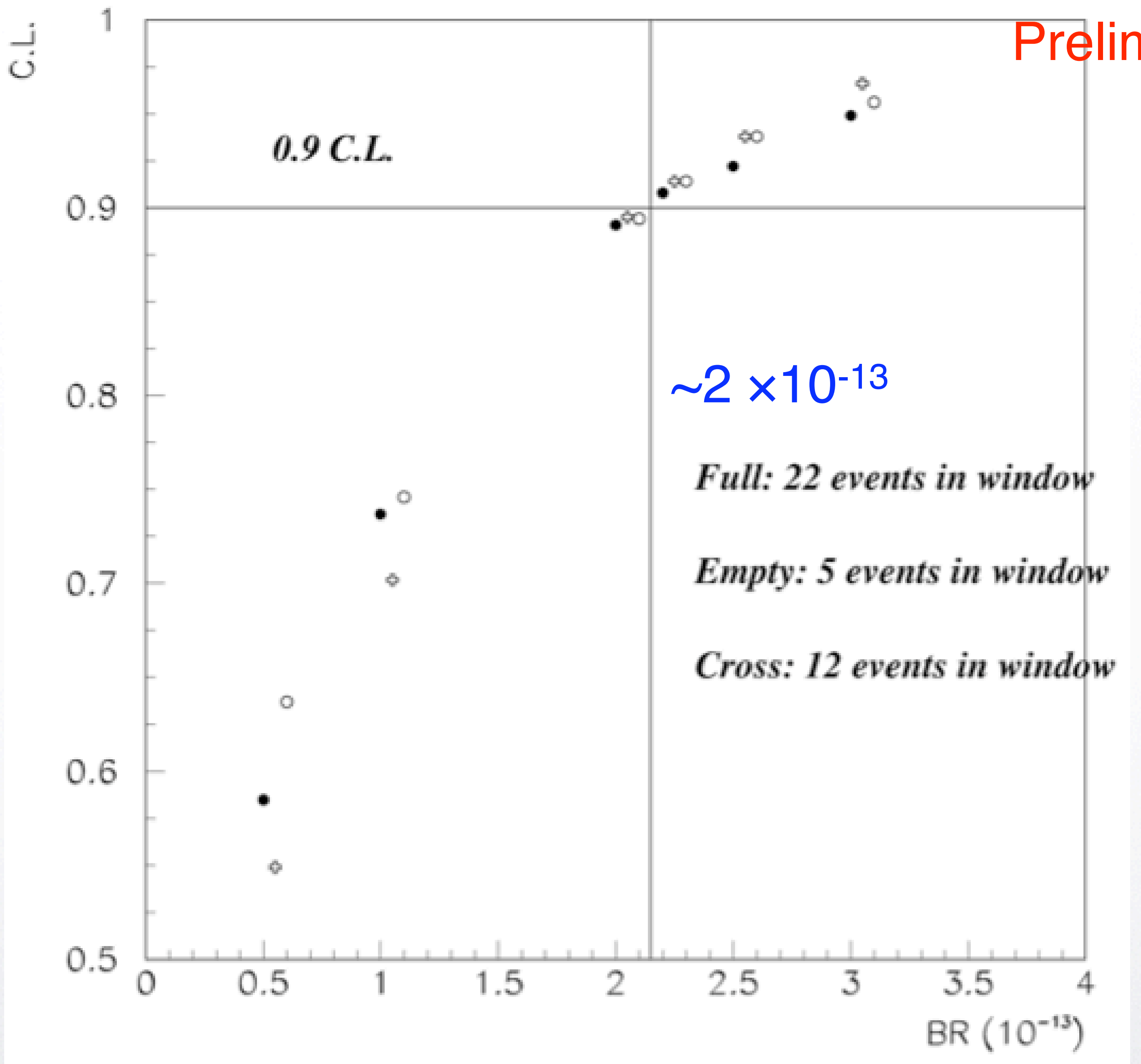
Positron efficiency



- Event generation (interest region)
  - $|\phi|<60$ deg, $0.08<|\cos\theta|<0.35$
- TC hit definition
- "Bar" && "Fiber" && "only primary e^+" && "$\Delta X_{TC} < \pm$ 5cm(Z),±3cm(ϕ)
- Removed each components one by one
- "# of accepted e^+@TC" / "# of generated e^+"

	Efficiency %
Full detector	**64.6**
w/o DC system	88.7
w/o cable duct	66.1
w/o cable	70.2
w/o amp	68.6
w/o frame	77.6
w/o gas, foil	65.1

C.L. of null experiment



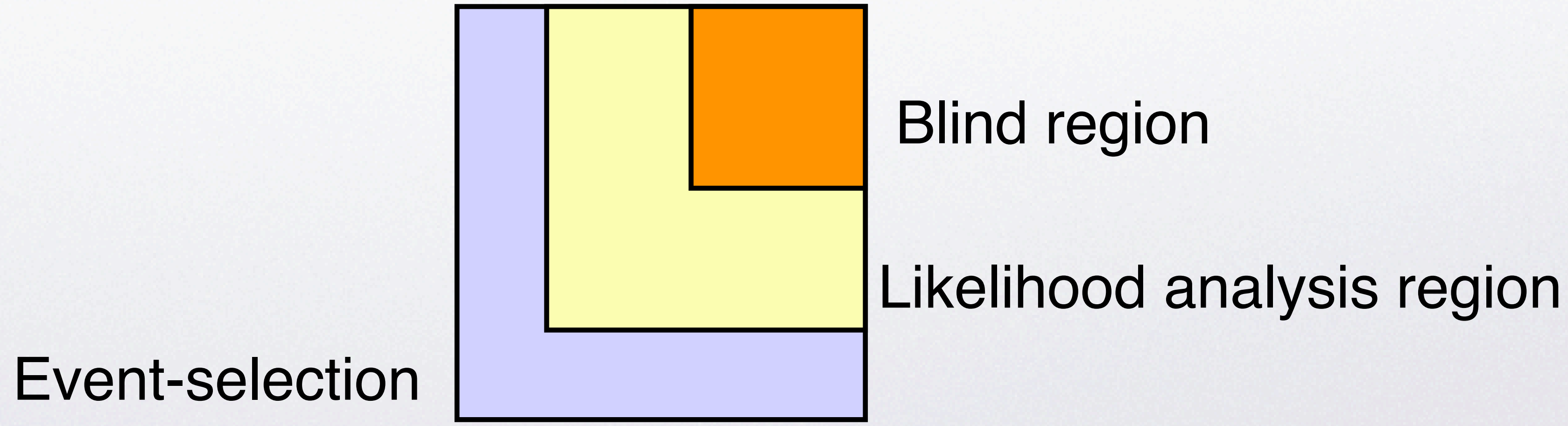
Different circles for different analysis window
N.B. Empty circles & crosses shifted by 0.05 on B.R. axis for clarity.

Blind $\mu \to e\gamma$ Analysis

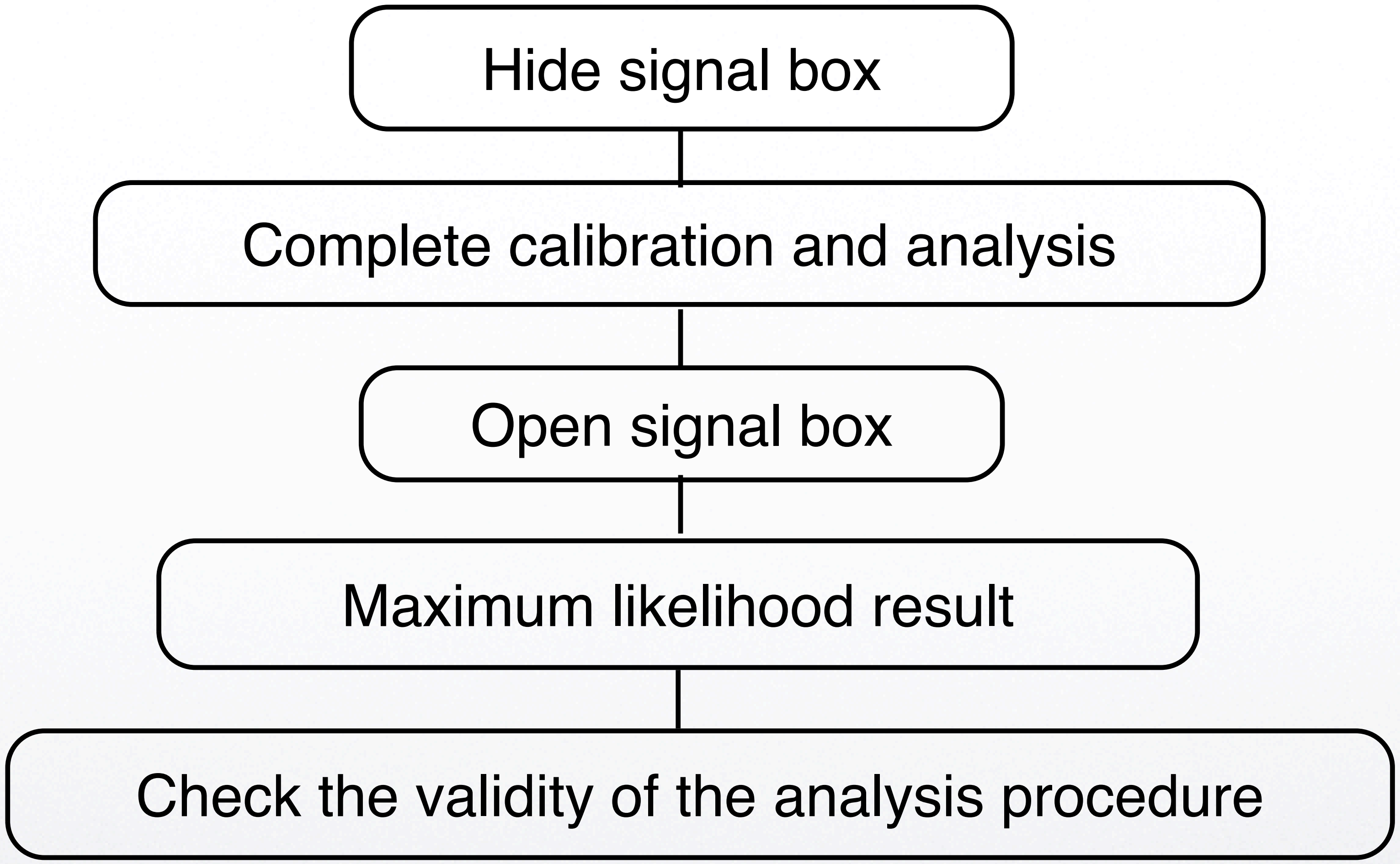
- Possibilities
    - Hidden signal box
    - Hidden offset
    - Divided analysis
    - Adding and removing events
    - ...

Blind $\mu \to e\gamma$ Analysis

- Hidden signal box
  - At least, two out of four measured parameters (E_e, E_γ, $T_{e\gamma}$, $\theta_{e\gamma}$) should be hidden.
  - Which parameters to hide?
    - We should be able to calculate PDF with the parameters hidden.
    - It doesn't matter for signal and BG PDFs since all parameters are independent. We should be careful with RD.
  - Box size? 2-3 sigma?
  - Boxes are nested



Possible Analysis Procedure-1



Possible Analysis Procedure-2

Replace some events with simulated “signal” events randomly.

Hide signal box

Complete calibration and analysis

Open the box for 10-20% of data sample

Check analysis, adjust parameters

Open the remaining 80-90% of data sample

Simple box result | Maximum likelihood result

Check the validity of the analysis procedure

How to Check Analysis Procedure?

- Compare estimated $N_{\mu\rightarrow e\gamma}$ for simulated signal events randomly added in data sample.
- Compare each single spectrum with likelihood fit result.
- Compare result on radiative decay with known branching ratio.
- Re-analyze in subdivided acceptance or subdivided data set.
- Check the distribution of PDF value.
- ...

Computing upgrade

- 20 CPUs → 64 CPUs
  - Need 3 months to become ready after the order
    - 2 months for delivery
    - 1 month for setup
  - Upgrade this year in order to buy same nodes as existing ones
- 26 TBytes → 100 TBytes
  - 20 TB for mass production + 6 TB for use = 26 TB
  - Upgrade next year ?
    - 20 TB might be enough for this year
    - The later the cheaper

Man power

• Online	S.Ritt, R.Sawada, M.Scheneebeli, G.Signorelli
• MC(gem)	P.W.Cattaneo, F.Cei, H.Natori, H.Nishiguchi, Y.Nishimura, W.Ootani, V.Tumakov, S.Yamada
• MC(bartender)	P.W.Cattaneo, Y.Hisamatsu, R.Sawada, V.Tumakov, Y.Uchiyama, S.Yamada
• Analysis	
• DCH	B.Golden, Y.Hisamatsu, F.Ignatov, W.Molzon, D.Nicolo, H.Nishiguchi, M.Schneebeli, C.Topchan, V.Tumakov, F.Yu, F. Xiao
• TC	A.Barchiesi, P.W.Cattaneo, G. Cavoto, S.Dussoni, L.Galli, G.Gallucci, Y.Uchiyama, C. Voena
• XEC	F.Cei, L.Perrozzi, R.Sawada, G.Signorelli, Y.Uchiyama
• Physics	F.Cei, W.Ootani

People who started contribution to software in this one year, or people who is starting contribution.

Summary

- Sub-groups in every stages are actively working
  - DAQ, Simulation, Reconstruction and physics
- People working for software is getting more. Software quality control is getting important
- Data reduction group is implementing analysis in DAQ. Algorithm depends on sub-detectors. Implementation of 3rd level trigger is also considered
- Modifications of simulation for performance, and better comparison with analysis result
- Reconstruction framework of each sub-detectors are getting ready. We are intensively working for improving performance (better resolution, efficiency)
- Study on procedure of physics analysis is ongoing

End