Data Analysis: 1.5mm Data set

Ani Simonyan
Michel Guidal & Raphael Dupre

Institut de Physique Nucléaire d'Orsay

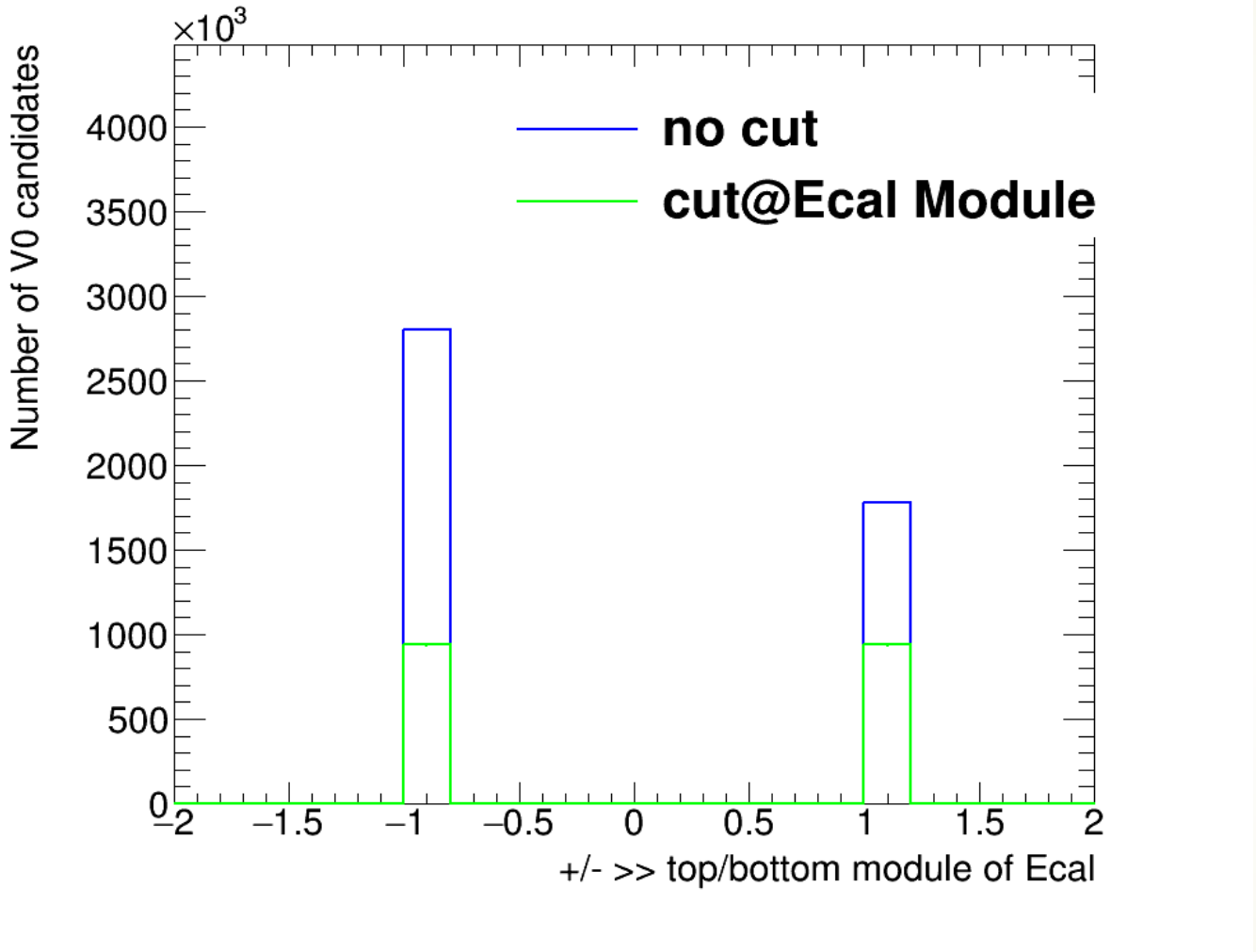
CNRS-IN2P3
Université Paris-Sud

1.5mm Data Analysis

- (e^-e^+)*Pair Selection Strategy*
  - *Event with collection of: Target Constrained V0 Candidates (TCV0)*
  - *V0 candidate have signal in opposite Ecal modules for e+ cluster and e- cluster*
  - *Valid Tracks: Track type > 32, track Chi2 <20 (40 Omar)*
  - *Valid Clusters:*
    - *Min cluster energy >0.15GeV,*
    - *Seed hit in time window (30-55) fADC*
  - *For the selected e+ e- pair:*
    - *2. Cluster seed-hit timing- dT = (-1.643;1.597) (<2ns Omar)*
    - *3. Doesn't contain Coulomb electron* $E_{e^-} < 0.85GeV$ *(0.792 GeV Omar)*
    - *4. Reduce Radiatives* $E_{e^-} + E_{e^+} > 0.8GeV$ *(0.8448 GeV Omar)*
    - *5. Best* χ^2 *<20 for interaction Vertex (if track-cluster matching? Chi2 <10 Omar)*
- *Bump hunt strategy*
  - *Software packages available for stat. analysis, roostats of ROOT*
  - *Statistical analysis on selected data considering invariant mass of* (e^-e^+) *as an observable (very first approach with RooStats::AsymptoticCalculator)*

1. Assume "good" e+ and e- for V0 candidate are selected

Selecting “good ”particles



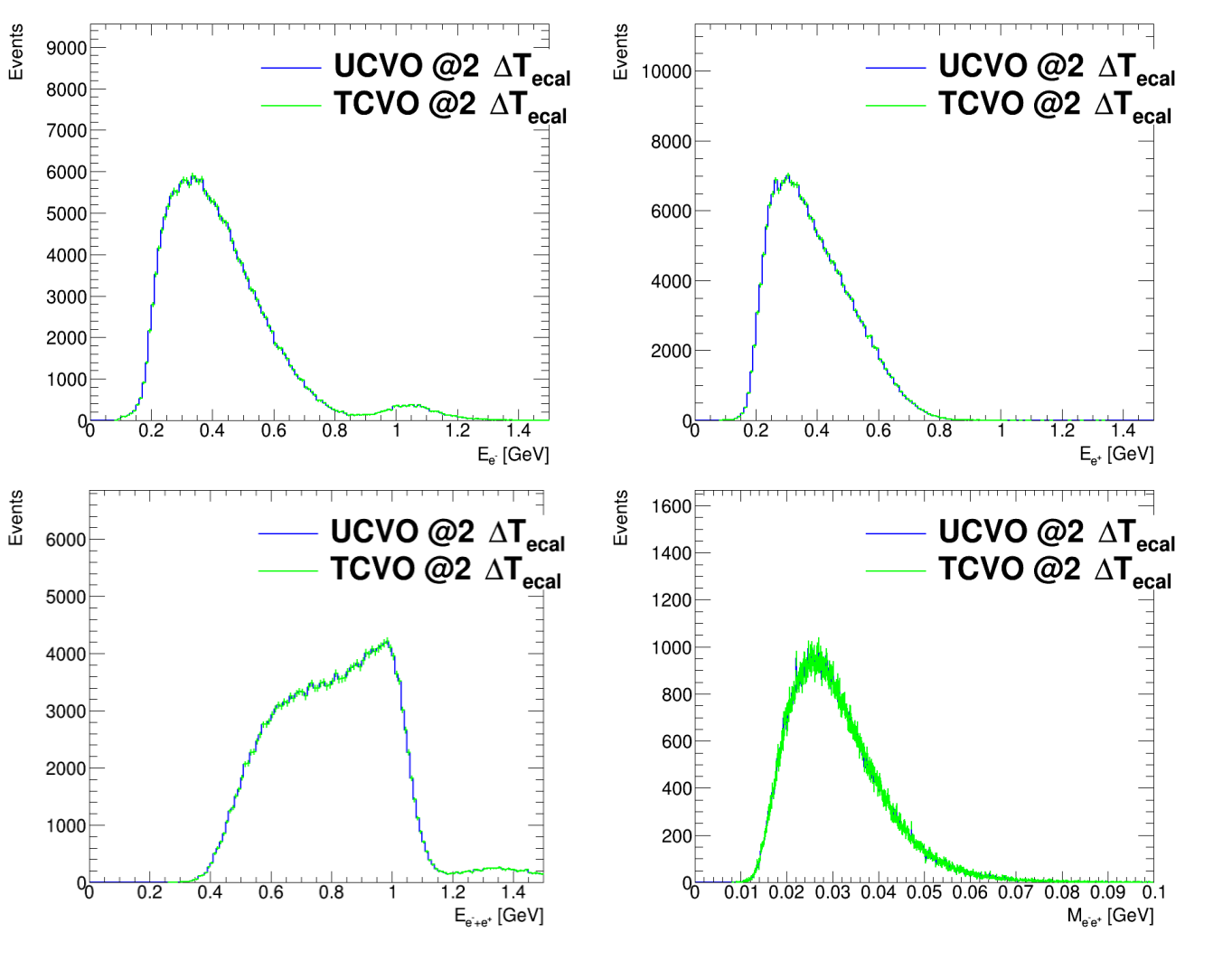
1. *Require cluster in opposite modules of Ecal*

For e+ and e-:

1. *Ecluster>150MeV*
2. *Ecluster timing [30-55] FADC*
3. *Track type > 32*
4. *Track Chi2 < 20*

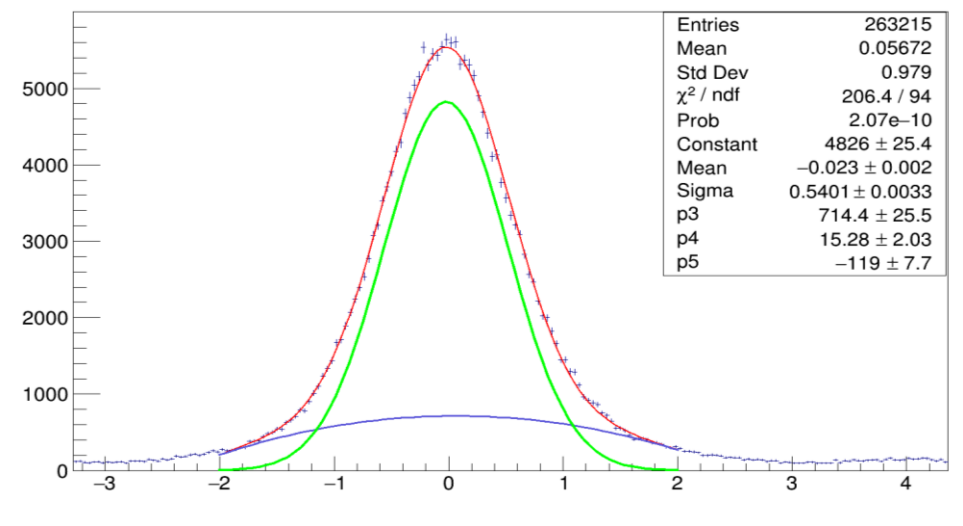
After reconstruction often algorithm forms fake V0 candidates (e+,e- pairs), when both leptons end up in the same ecal module despite on trigger requirement.

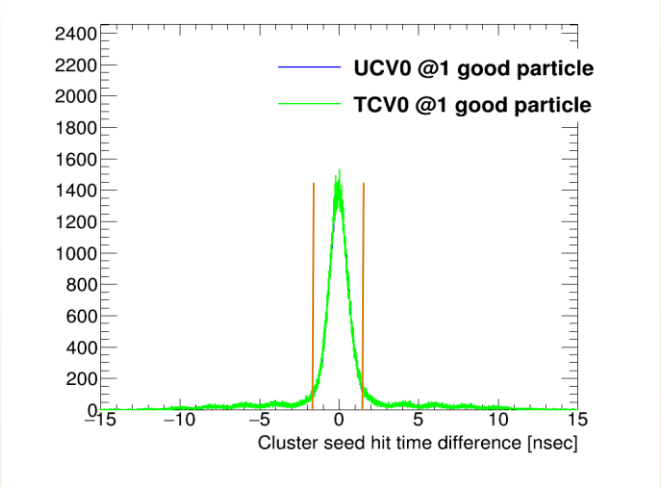
2. Pair selection based on cluster seed hit timing



Seed Hit Time difference fit

Model:gaus+pol2

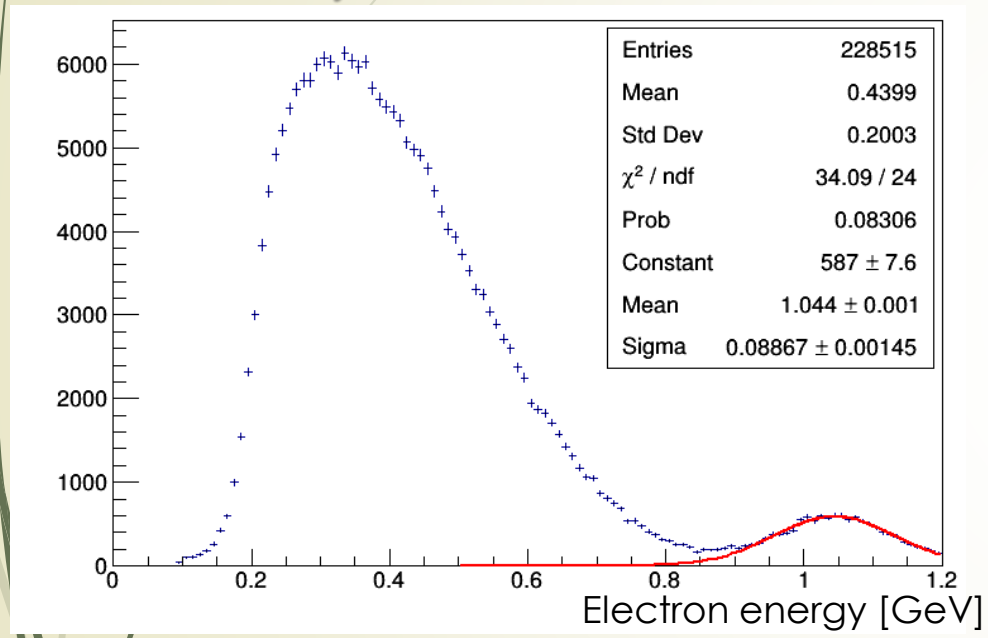




Cut for both for TCV0 & UCV0 Mean+/- 3sigma [-1.643;1.597]

3. Suppressing recoiled beam electrons: (Coulomb electron cut $E_{e^-} < 0.85GeV$)

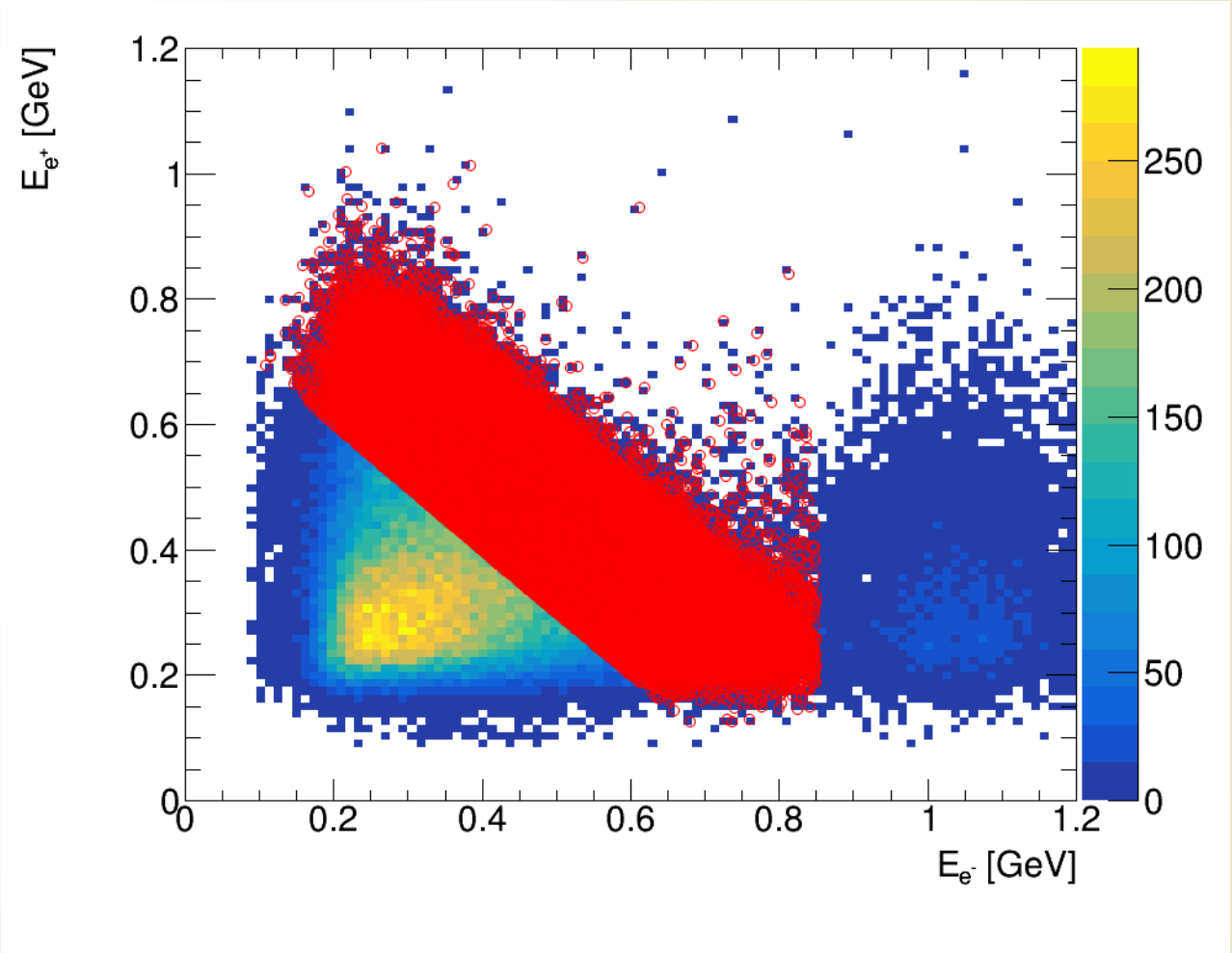
Full energy Electron fit with simple Gaussian



0.85 GeV is chosen as a local minimum , corresponds to ~2.2σ Gaussian .

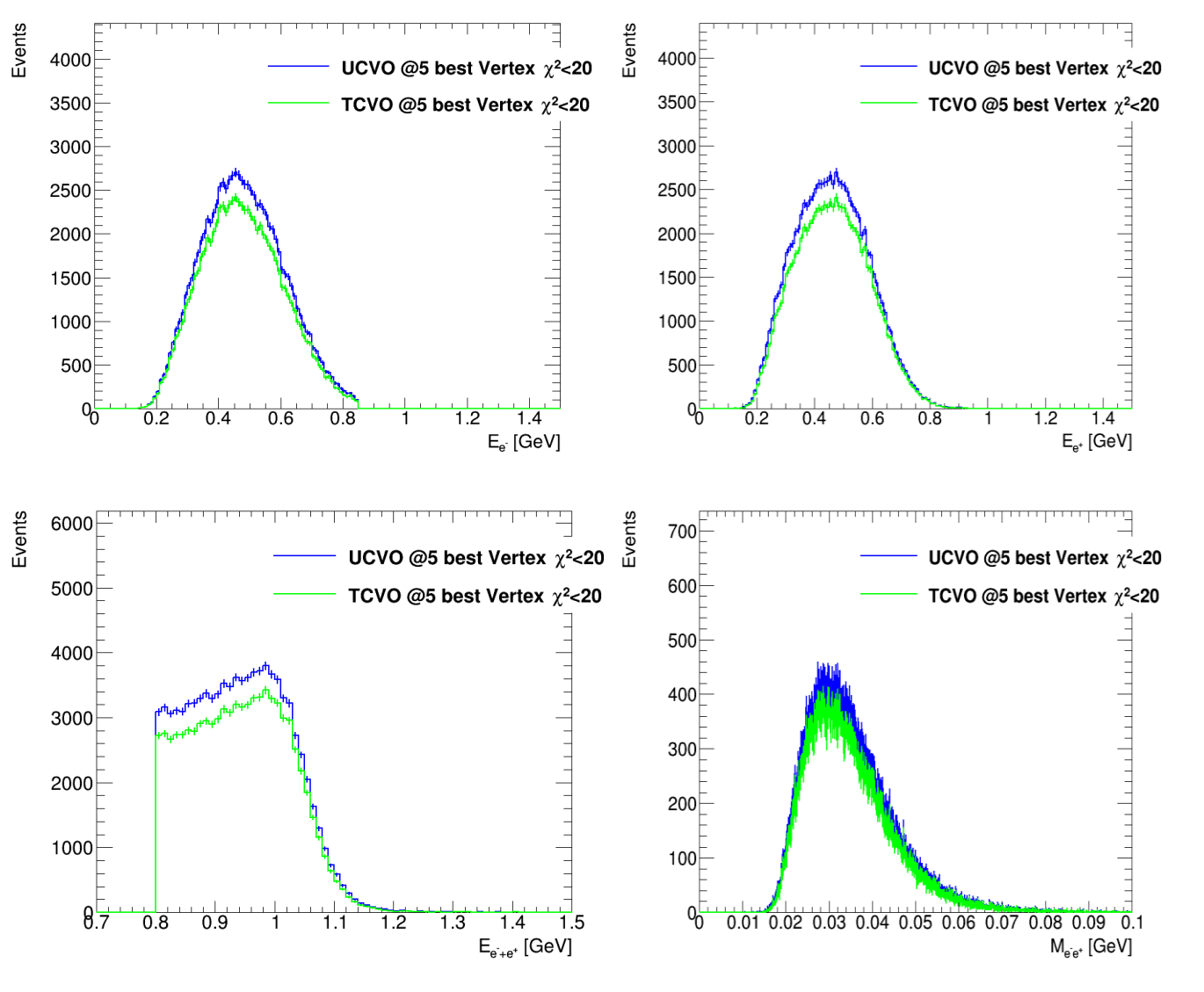
4. Cut on Radiatives :

$$E_{e^-} + E_{e^+} > 0.8GeV$$



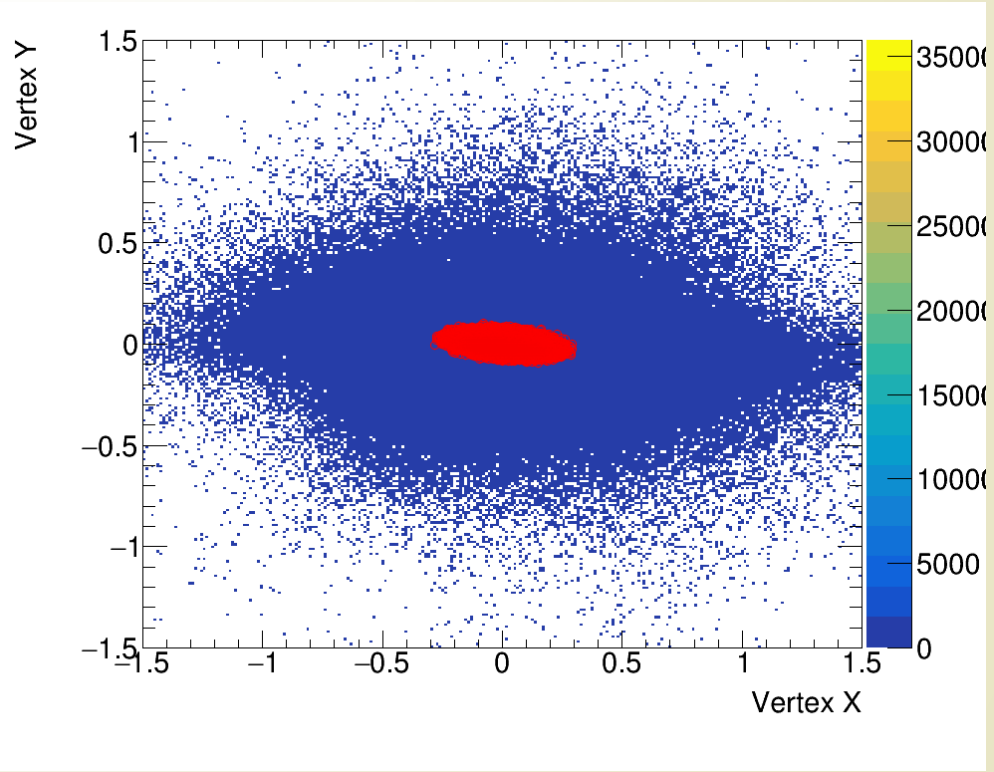
5. Single pair with best χ^2 <20 for interaction vertex

At this stage all (e+,e-) pairs have signal in SVT L1 or L2!

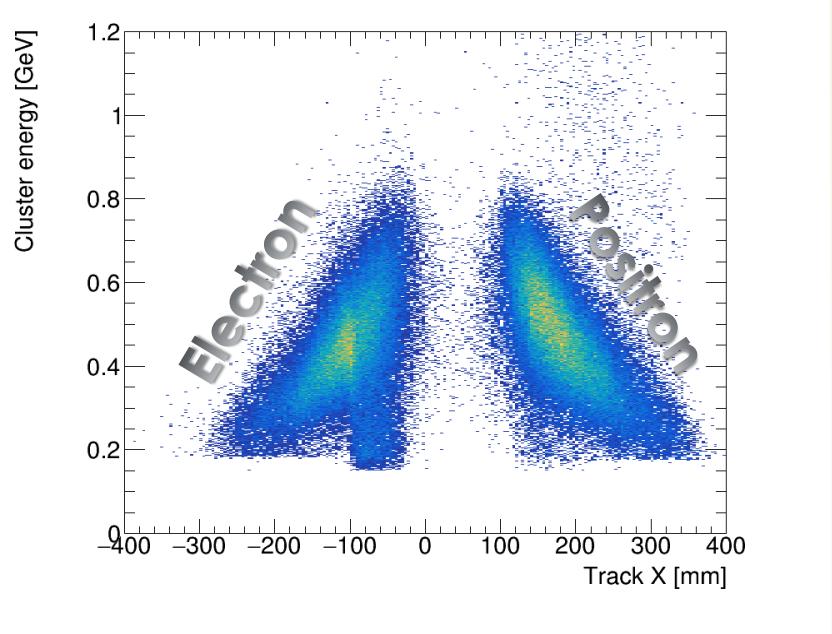


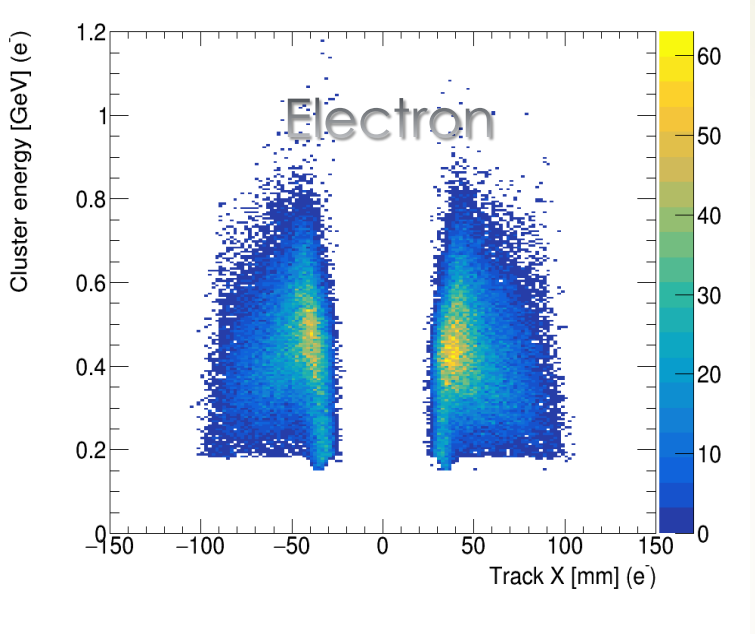
Interaction Vertex [mm]

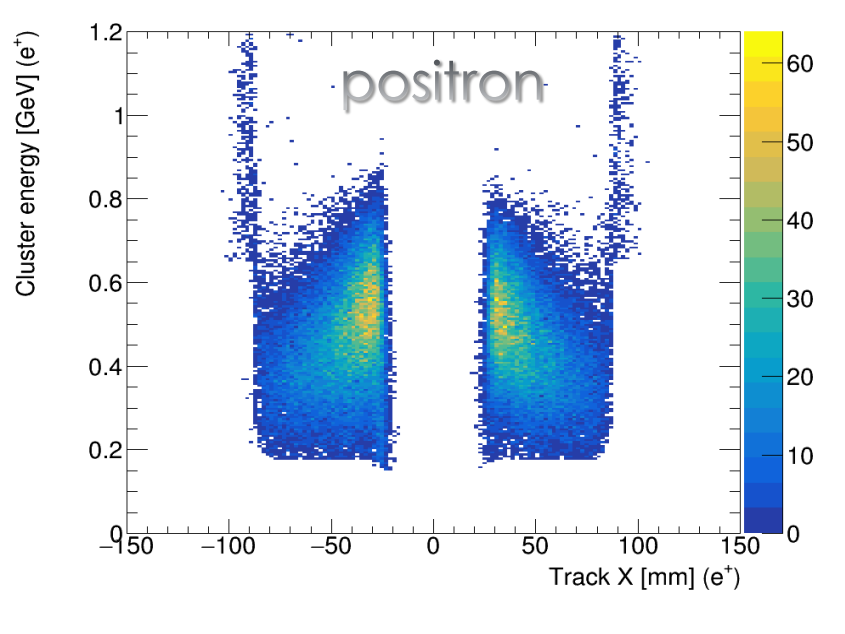
Selected events in red

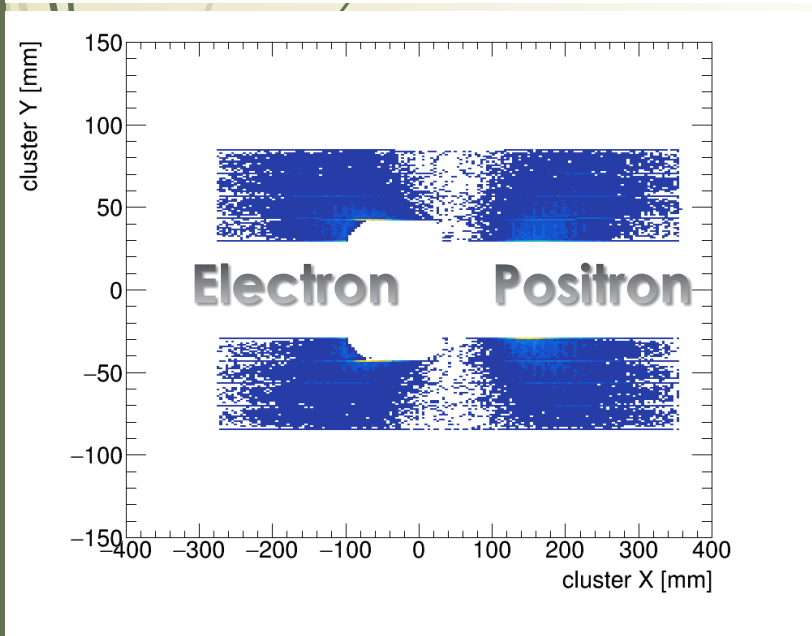


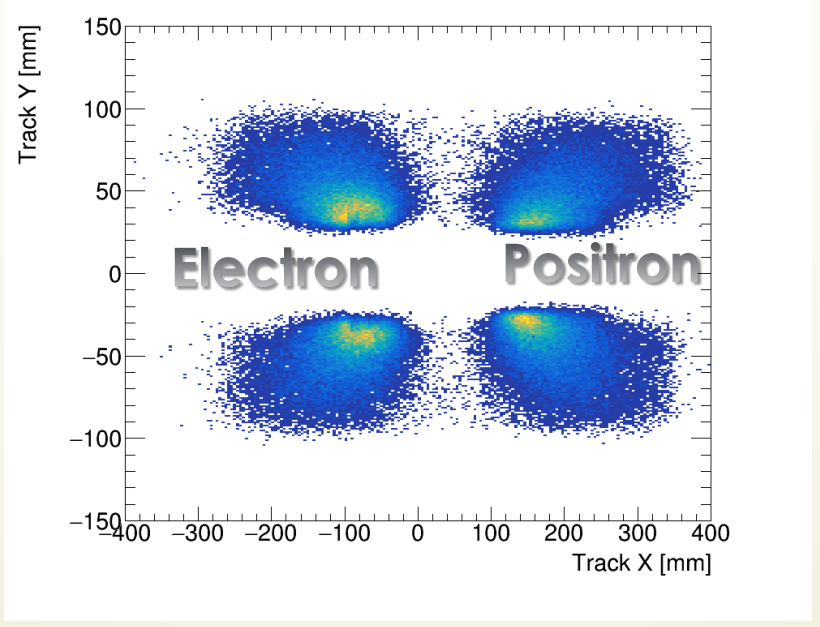
@Best $\chi^2 < 20$ for interaction vertex

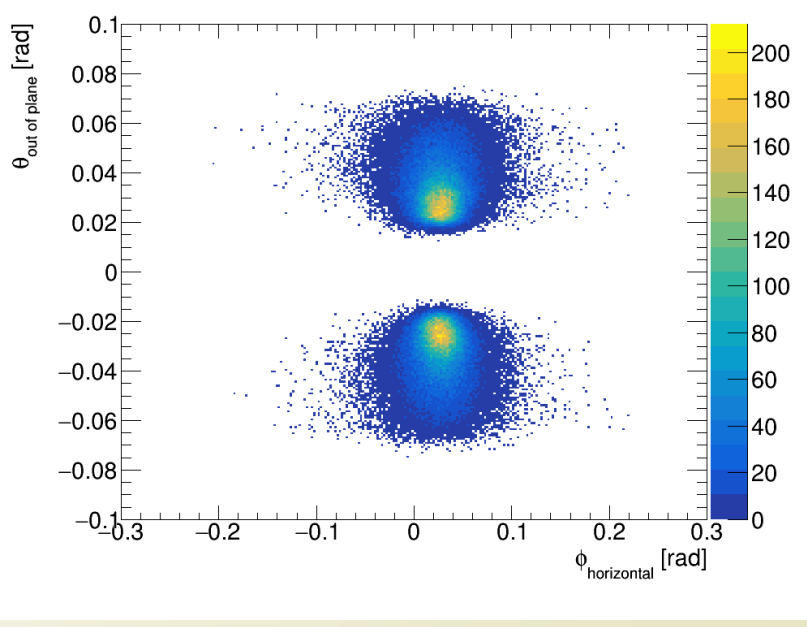




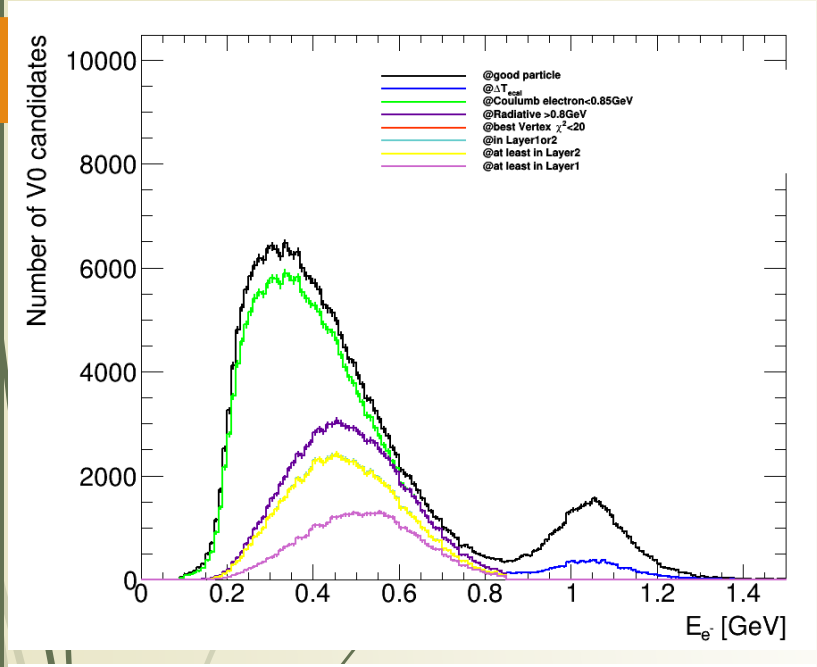


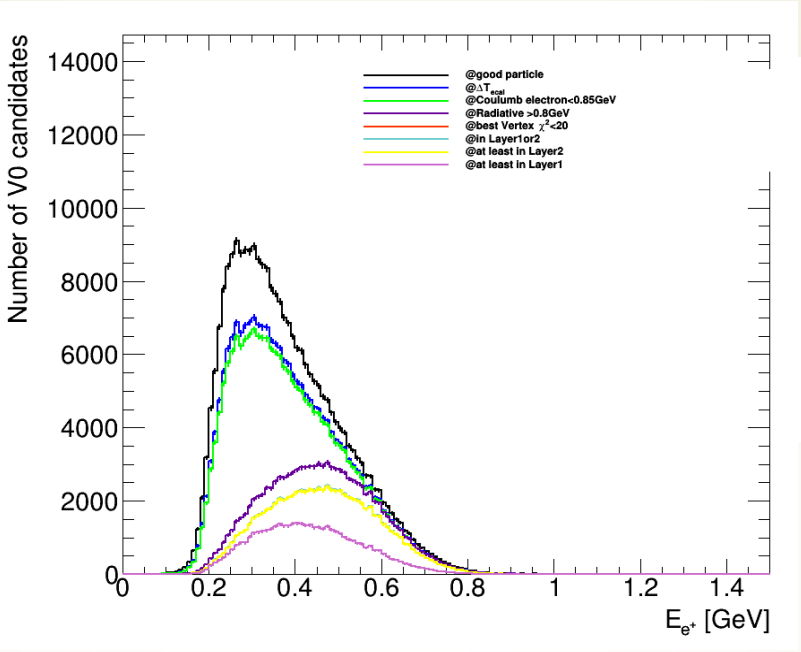


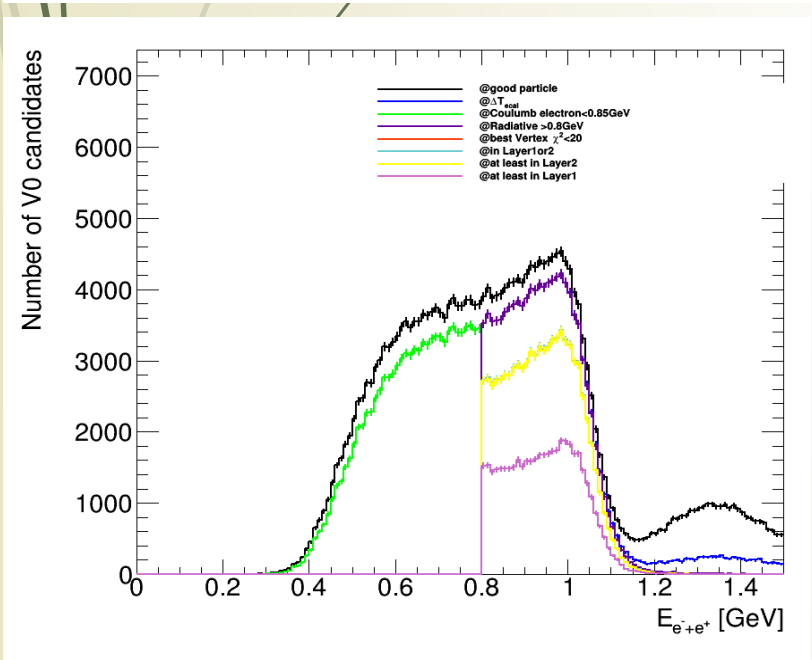


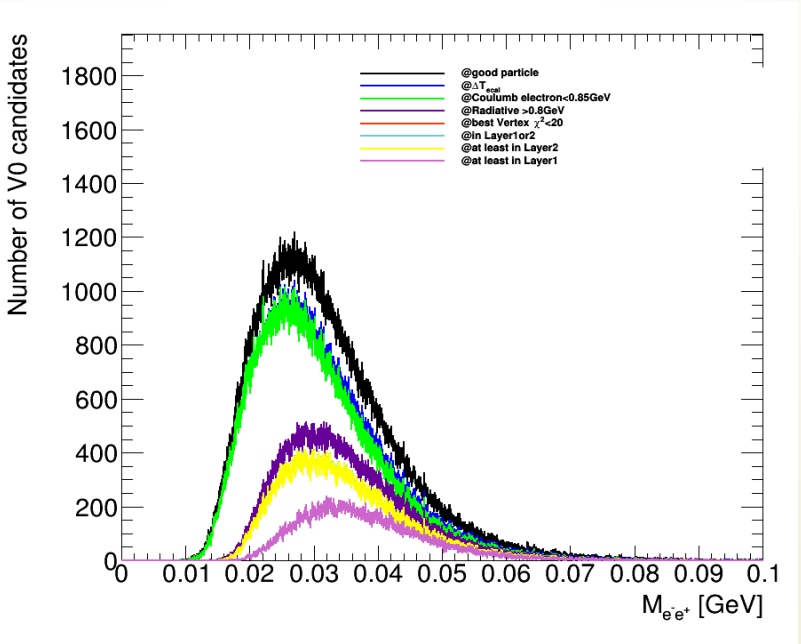


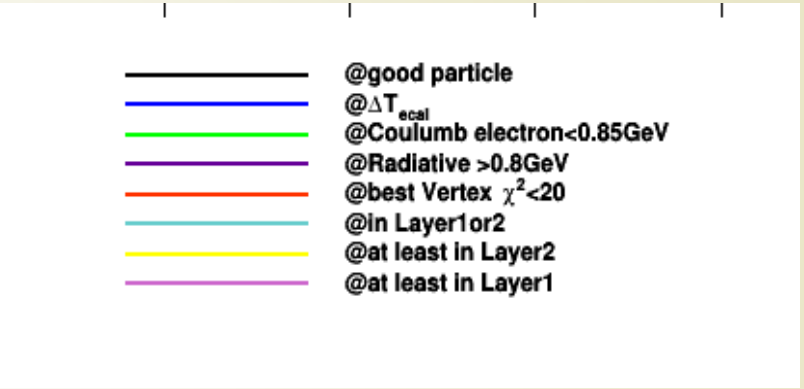
Energy and invariant mass vs cuts



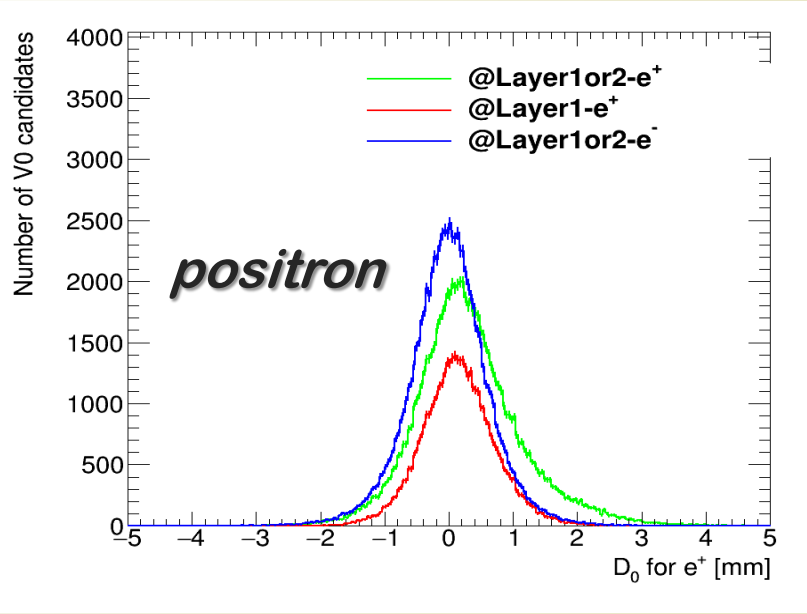




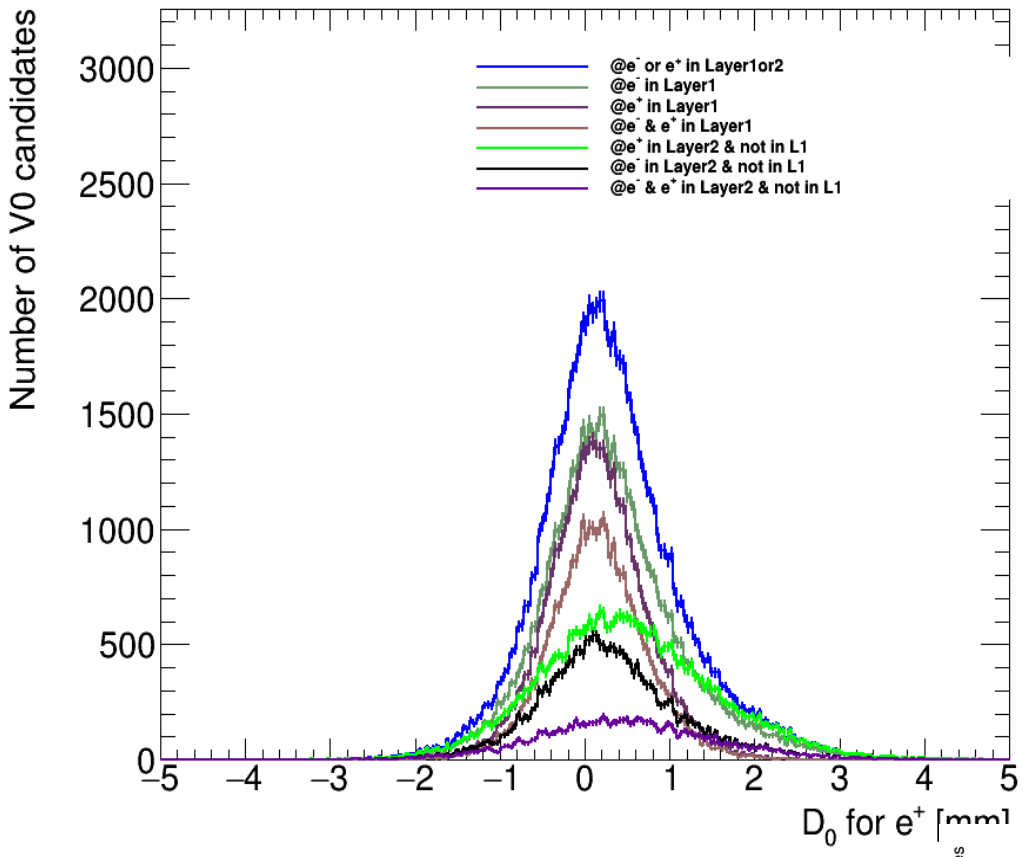


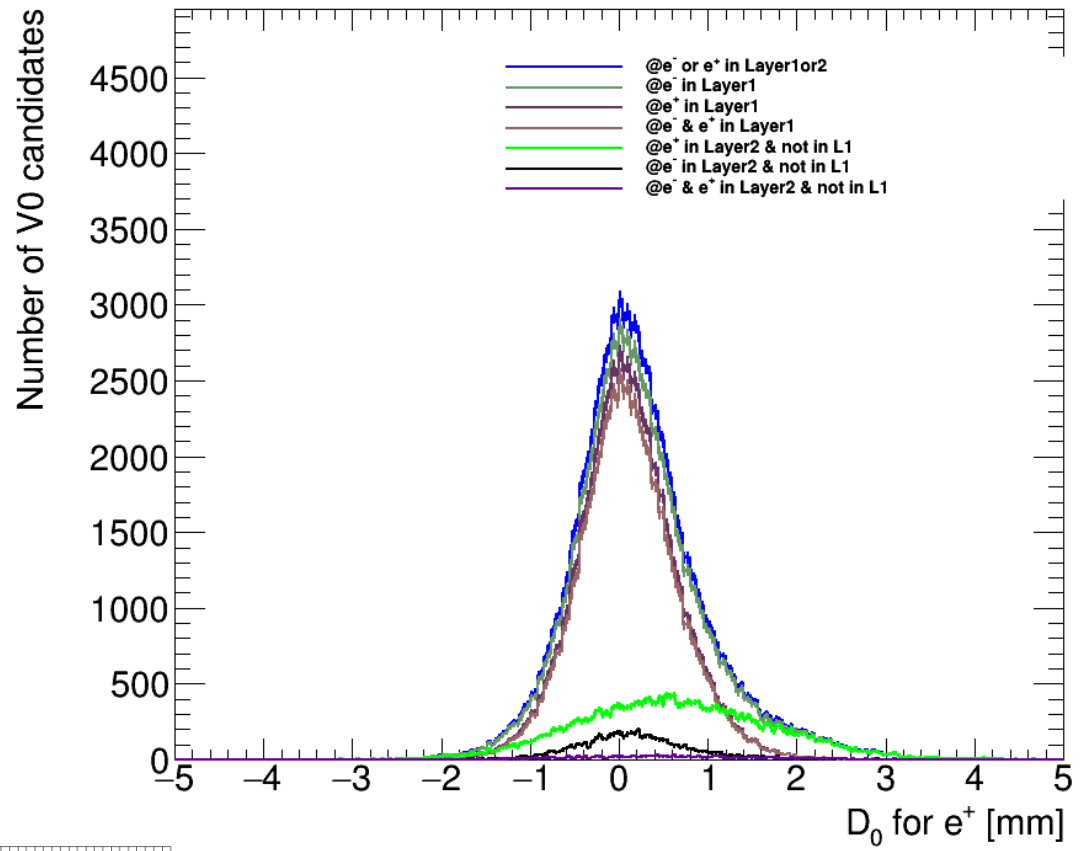


Comparing D0 distribution for positron from SVT L1 and L1or2

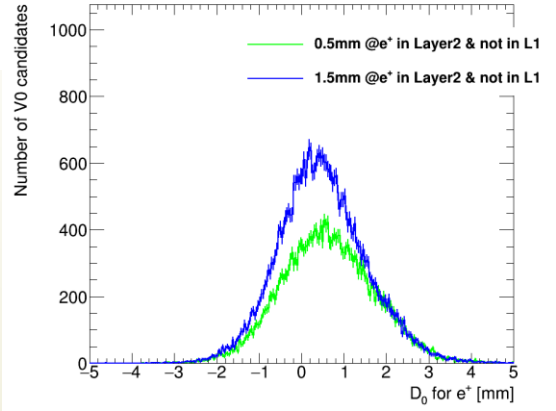


Positron D0 distribution: Comparing with 1.5mm and 0.5mm data



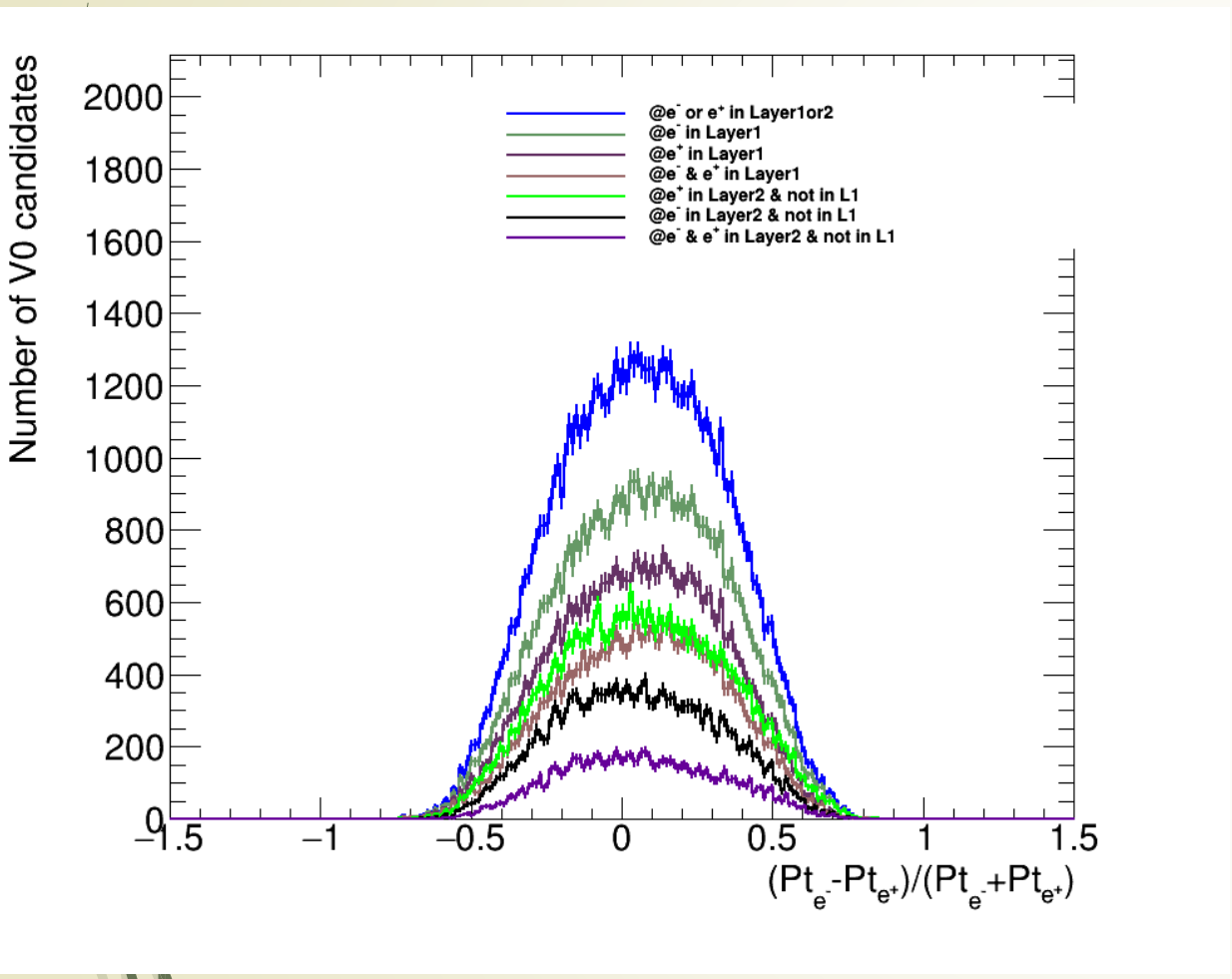


For 1.5mm data

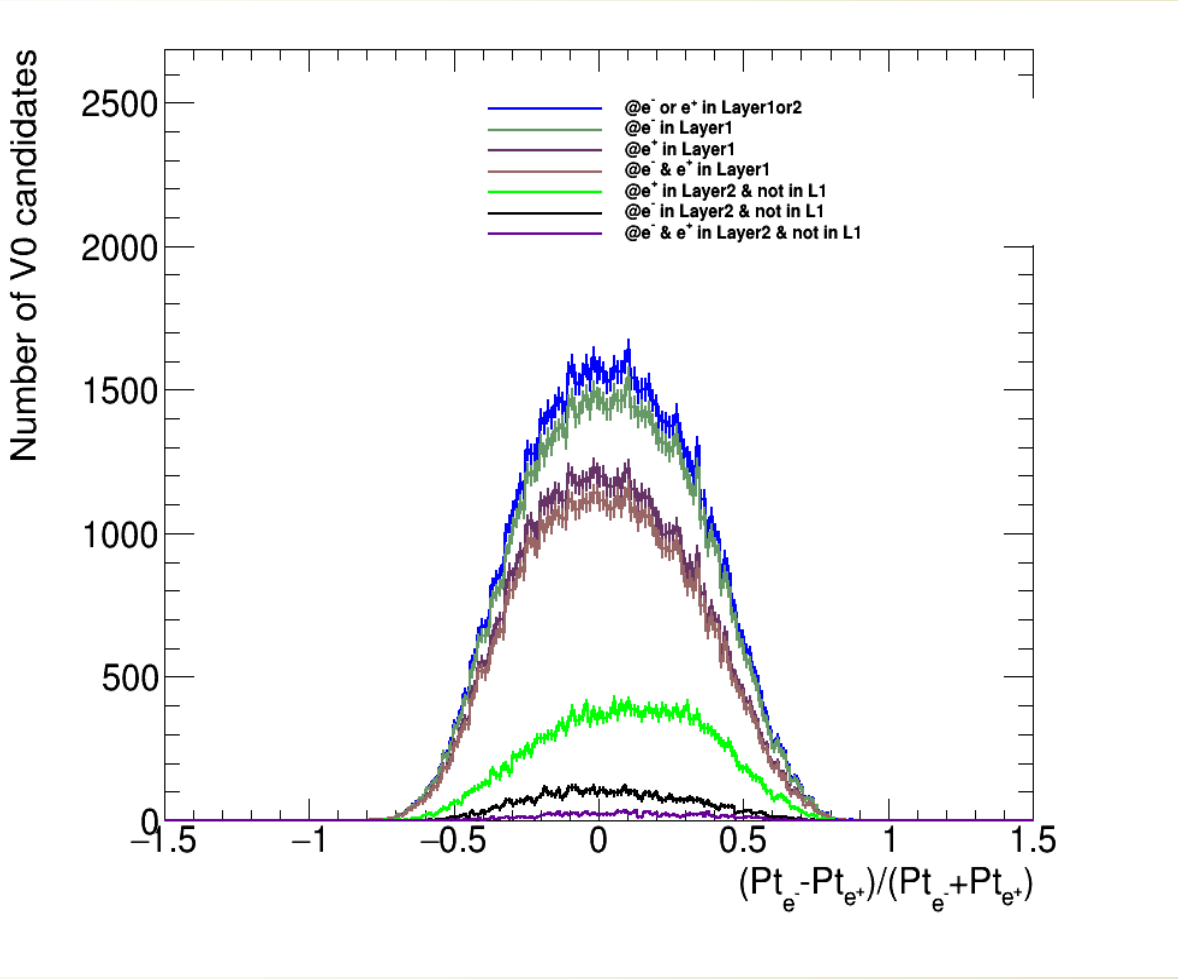


For 0.5mm data

Momentum Balance: *Comparing with 1.5mm and 0.5mm data*



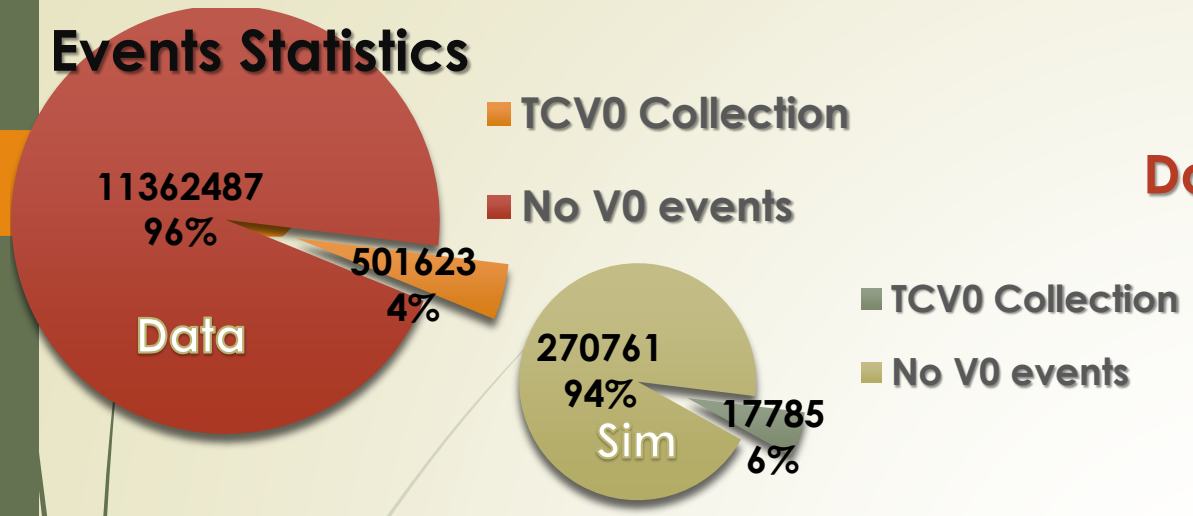
For 1.5mm data

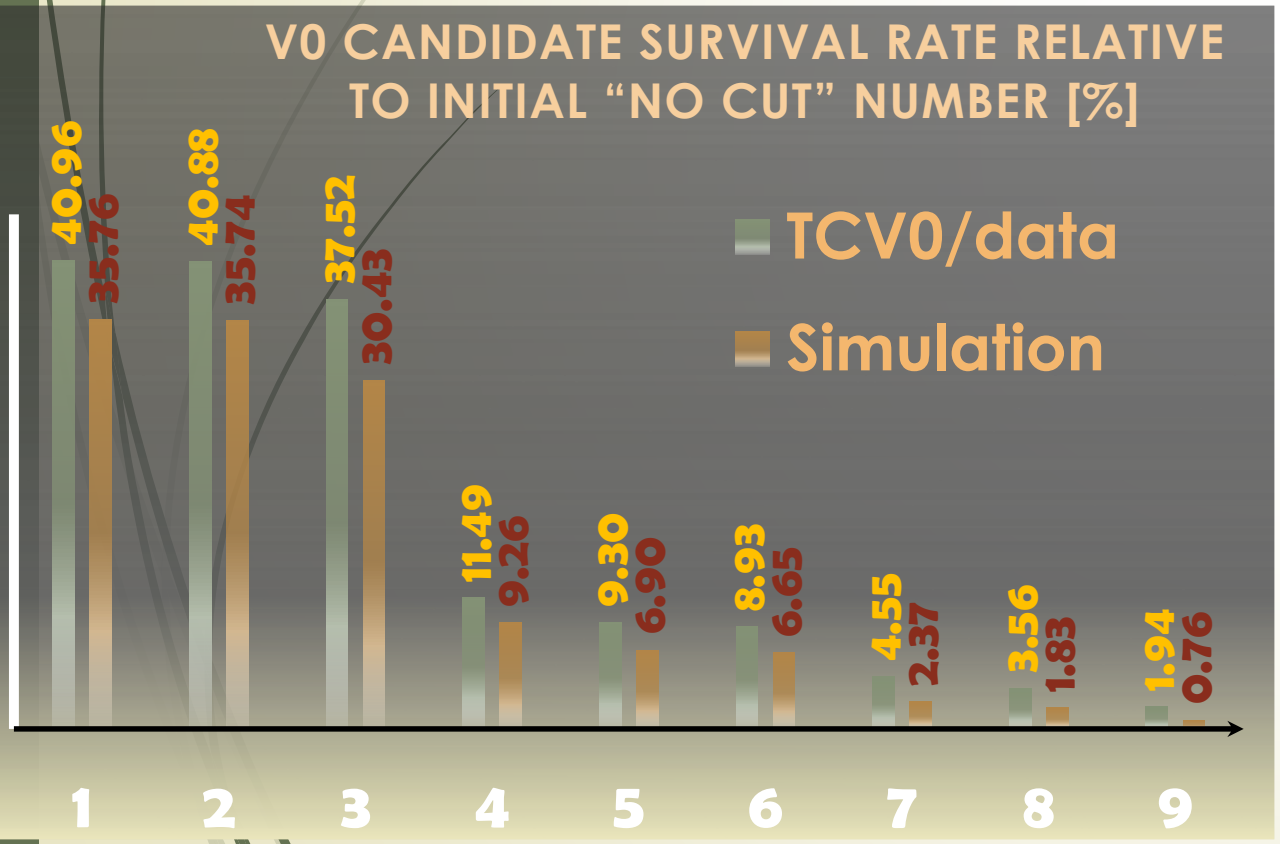


For 0.5mm data

Let's talk numbers...

Data Sample: 45 dst/root files/ pass6/ runN 5578, 5577





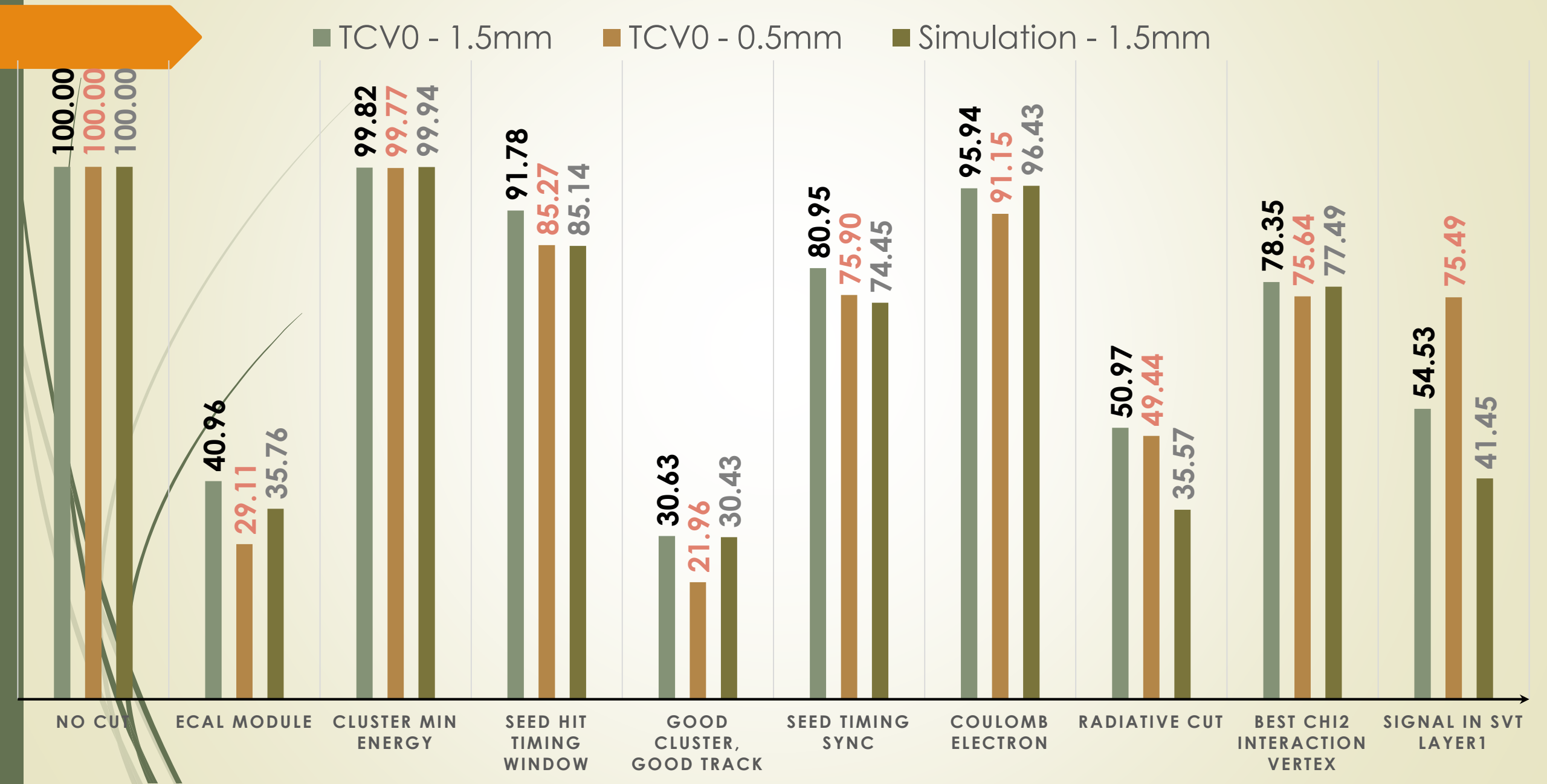
	Cut Name	Number of Candidates TCV0/data	Number of V0 Candidates/simulation	
	No cut	2290313 - 100%	99530	100%
1	Ecal Module	938005 - 40.9%	35596	35.76
2	Cluster min energy	936334 - 99.8%	71152	71.49
3	Seed Hit Timing window	859360 - 91.8%	30290	30.43
4	Good cluster, good track	263215 - 30.6%	9218	9.26
5	Seed timing sync	213069 - 80.9%	6863	6.90
6	Coulomb electron	204422 - 95.9%	6618	6.65
7	Radiative Cut	104194 - 50.9%	2354	2.37
8	Best Chi2 Interaction Vertex	81640 - 78.4%	1824	1.83
9	Positron has signal in SVT Layer 1	44520 - 54.5%	756	0.76

V0 CANDIDATE SURVIVAL RATE RELATIVE TO INITIAL SELECTION [%]

Initial selection:

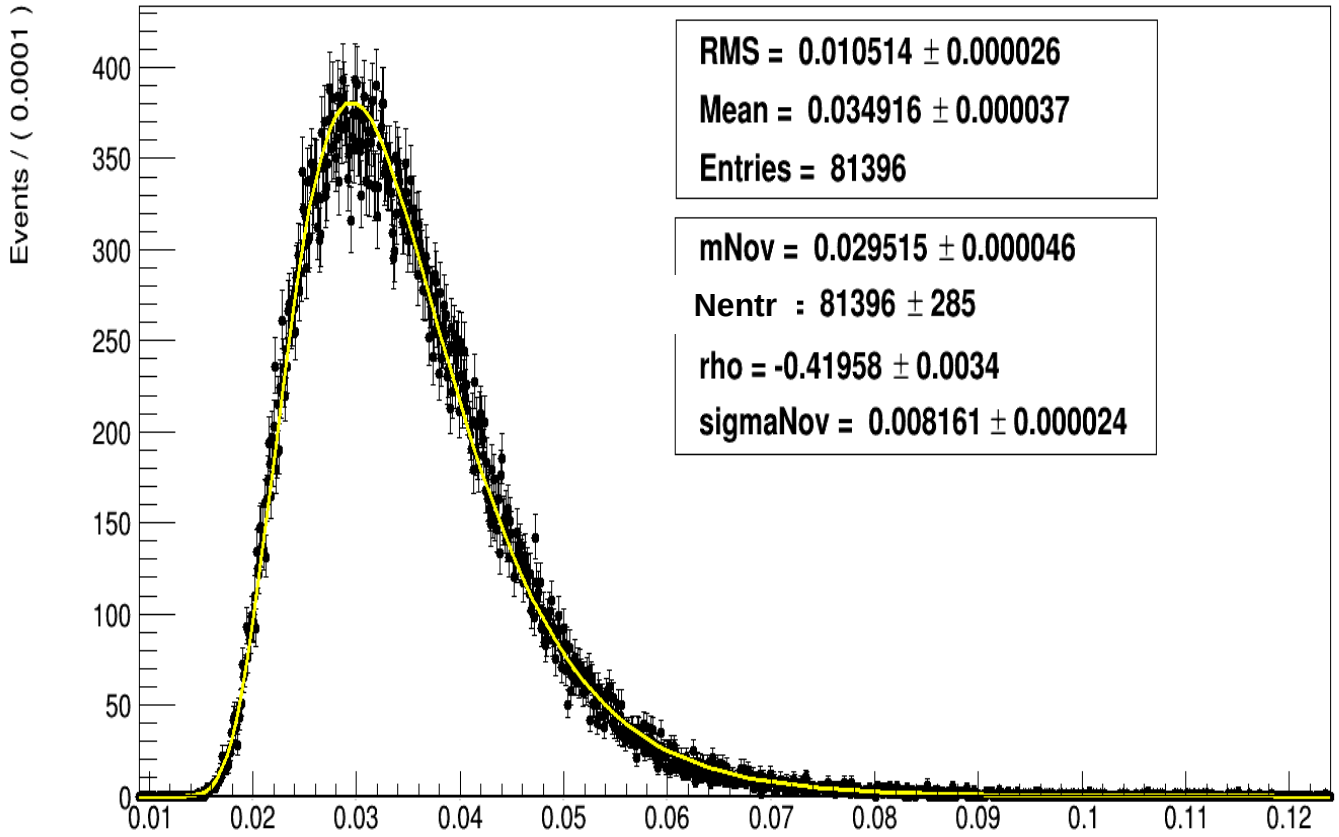
- 1.5mm no cut – 100% - *2,290,313* V0 candidates
- 0.5mm no cut – 100% - *7,676,443* V0 candidates
- Simulation 1.5mm – 100% - *99,530* V0 candidates

V0 CANDIDATE SURVIVAL RATE [% RELATIVE TO PREVIOUS CUT]



Bump hunt in invariant mass distribution

Extended maximum likelihood fit (Yellow line) of invariant mass distribution after event selection .



Invariant mass of electron, positron pair [GeV]

- Novosibirsk probability density function used to fit the data:

$$P(x) = e^{-0.5^{(\ln q_y)^2/\Lambda^2+\Lambda^2}}$$

$$q_y = 1+\Lambda(x-x_0)/\sigma \times \frac{\sinh(\Lambda\sqrt{\ln 4})}{\Lambda\sqrt{\ln 4}},$$

Where $\Lambda \rightarrow \rho$ parameter describes the tail,

σ – width of the distribution

x0 – peak position

- P(x) – background model
- Gaussian – signal model

Statistical analysis: playing with roostats package

Bump hunt in frequentist approach: *probability of the data given the hypothesis*

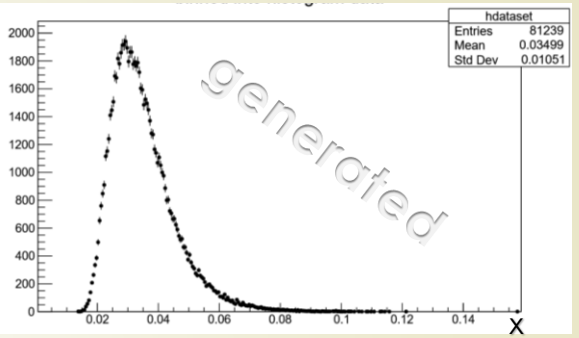
- **Hypothesis Testing:**
  - **H1 - signal with background**
  - **H0 – background only**
- **RooStats :: Asymptotic Calculator – based on profile likelihood test using Asimov* data sets instead of MC toys, therefore relatively fast.**
  ***(Special dataset such that when using it to evaluate the estimators, one obtains the true parameter values.)**
  - **Arguments: data, H1 model, H0 model**
  - **H1 model – Ns*signal +Nb*background,**
    - **Ns= number of signal events (as parameter of interest)**
    - **Nb=number of background (as nuisance parameter)**
  - **H0 model – H1(Ns=0), therefore H0 and H1 are nested**
  - **Signal parameters (mean and sigma for Gaussian) – constant for each mass hypothesis and vary:**
    - **mass - [0.015; 0.1]**
    - **sigma - for each mass calculated from the 3rd order polynomial fit to mass resolution (Omar)**
  - **Output:**
    - **p values for Null hypothesis vs observable distribution plot**

Toy distribution contains tiny signal at 0.07

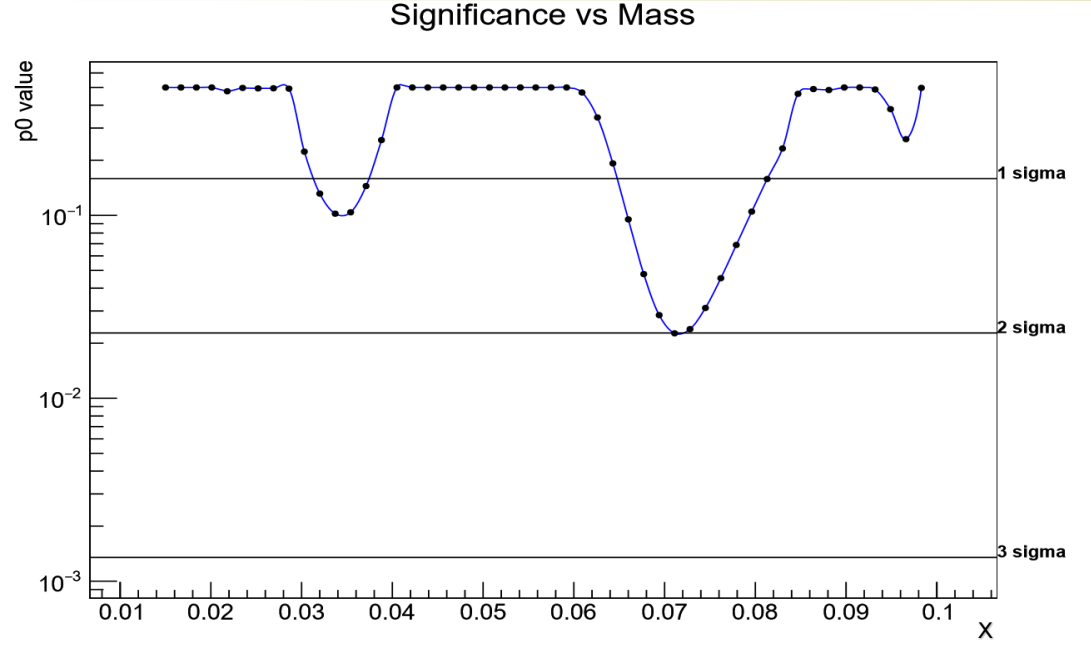
Generated with normalization similar to final distribution from data:

p.d.f = nsig*Gaussian+nbkg*Novosibirsk:

$$P(x) = e^{-0.5^{(\ln q_y)^2/\Lambda^2+\Lambda^2}}$$

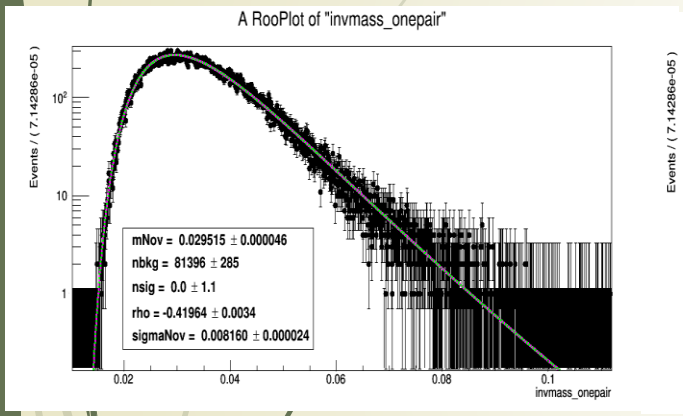


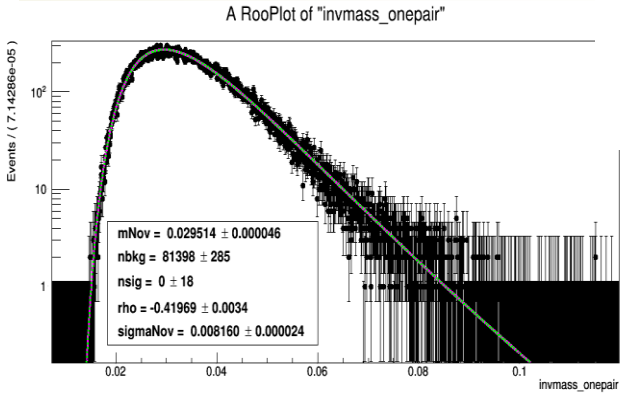
p0 vs mass calculated with asymptotic analysis of generated toy distribution

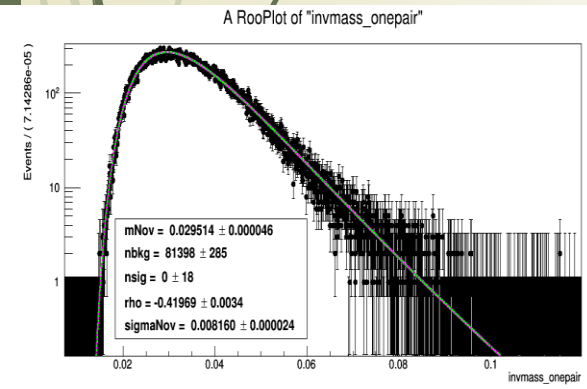


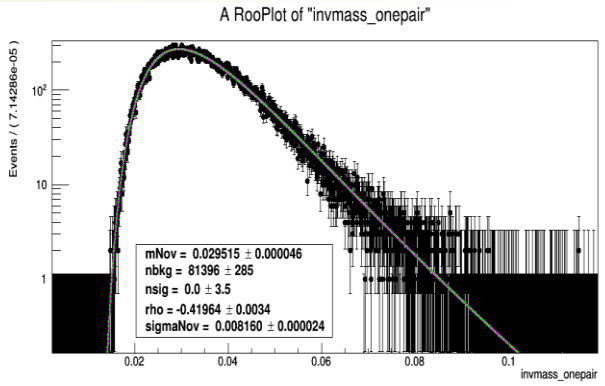
Statistical analysis: IM distribution and roostats package

Estimation of the model parameters from AsymptoticCalculator for hypothesized masses.



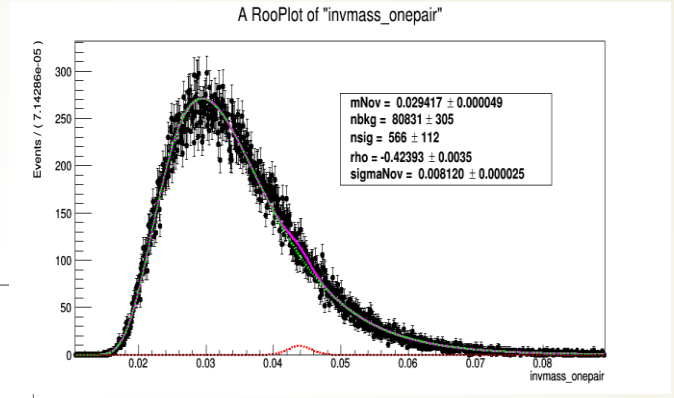


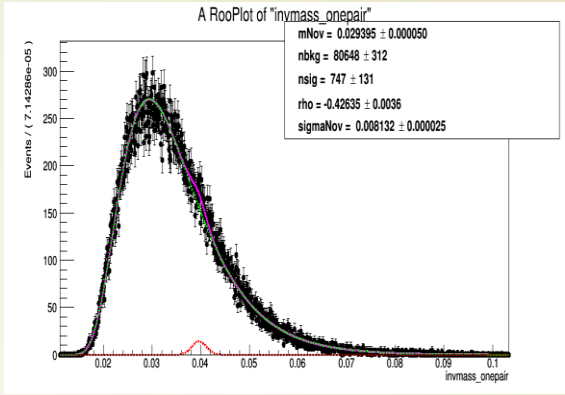


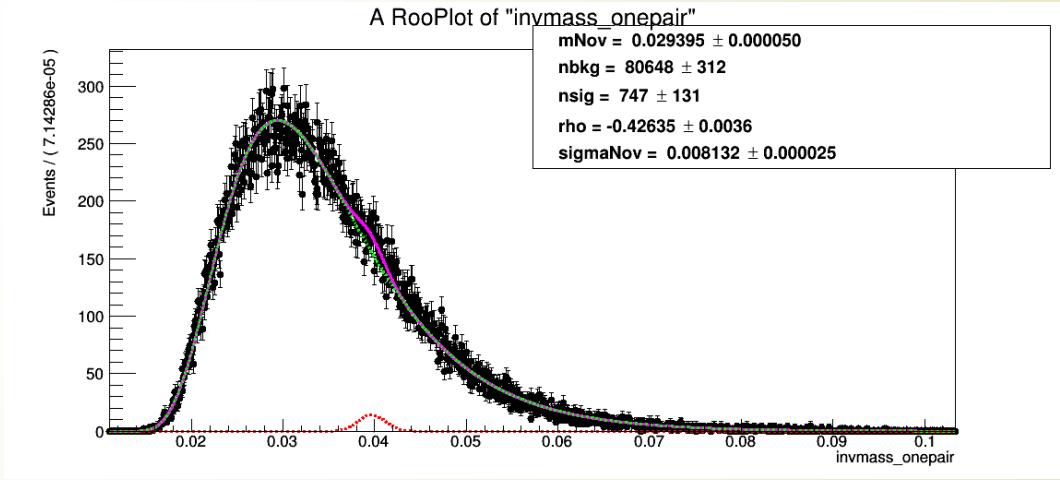


Unbinned extended ML fit of invariant mass (data) with Alt model p.d.f.

(@Note: NOT A HYPOTHESIS TEST)







Extended ML fit succeeds for Null model p.d.f with no signal parameter as well.

Remaining questions

- Fixing the width for the hypothesized mass?
  - Forces for each hypothesized mass search for a peak with certain fixed sigma.
  - The estimation of sigma is based on mass resolution calculated from simulation.
- Expected strength of the signal for alternate Hypothesis?
  - I used small number, as a guess, but this must be well motivated from theory.
- Choice of the p.d.f.-s describing the shape of the background.
- For each mass hypothesis consider full range of the data, or k*sigma window?
- Systematic uncertainties need to be included in the models as nuisance parameters as well.

Summary

- Event Selection:
  - V0 candidate collection choice:
    - target constrained vs unconstrained V0: TCV0 choosen
  - Electron-positron pair cuts are studied and more or less finalized
    - V0 candidate survival rate (V0 Collection, Simulation, vs 0.5mm data with corresponding cuts)
    - Further studies with simulated data for wab contamination
    - Therefore, more simulation needed for 1.5mm data set
- Bump hunt:
  - IM distribution: shape and H0, H1 models
  - Attempt to study IM distribution of (e+,e-) in asymptotic regime
  - And some questions….

Thanks!