Current Status of LCFIPlus (and more)

Masakazu Kurata, KEK
Tomohiko Tanabe, The University of Tokyo
Taikan Suehara, Kyushu University
Jan Strube, PNNL
Ryo Yonamine, Tohoku University

LCWS2017, 10/23/2017-10/27/2017

Introduction

- Recent issues and treatments
  - Check sudden crash using v01-19-04
- For flavor tagging improvement
  - Vertex finding and vertex mass recovery
  - Flavor tagging efficiency
- Other high level reconstruction tools
  - Focus on ILD new detector models
  - dE/dx
  - Quick look at PID

Recent issues

Crash in ROOT6(01_v05)

- LCFIPlus is suddenly crashed at somewhere in event creation
- But, sight difference of the environment might affect the crash??
  - Me and Taikan check behavior in ROOT6 environment
    - My side: no crash…(at least flavor tagging and MVA training)
    - Taikan side: LCFIPlus crashes at a point
    - Both use standard environment at KEK computing

      ⇒ MAYBE and SHOULD BE same
    - Where is the difference coming from??
- ILD collaboration starts Re-optimization, so we check this in latest ilcsoft (v01-19-04)
  - Check jet clustering and flavor tagging
  - Check Ntupling and TMVA training

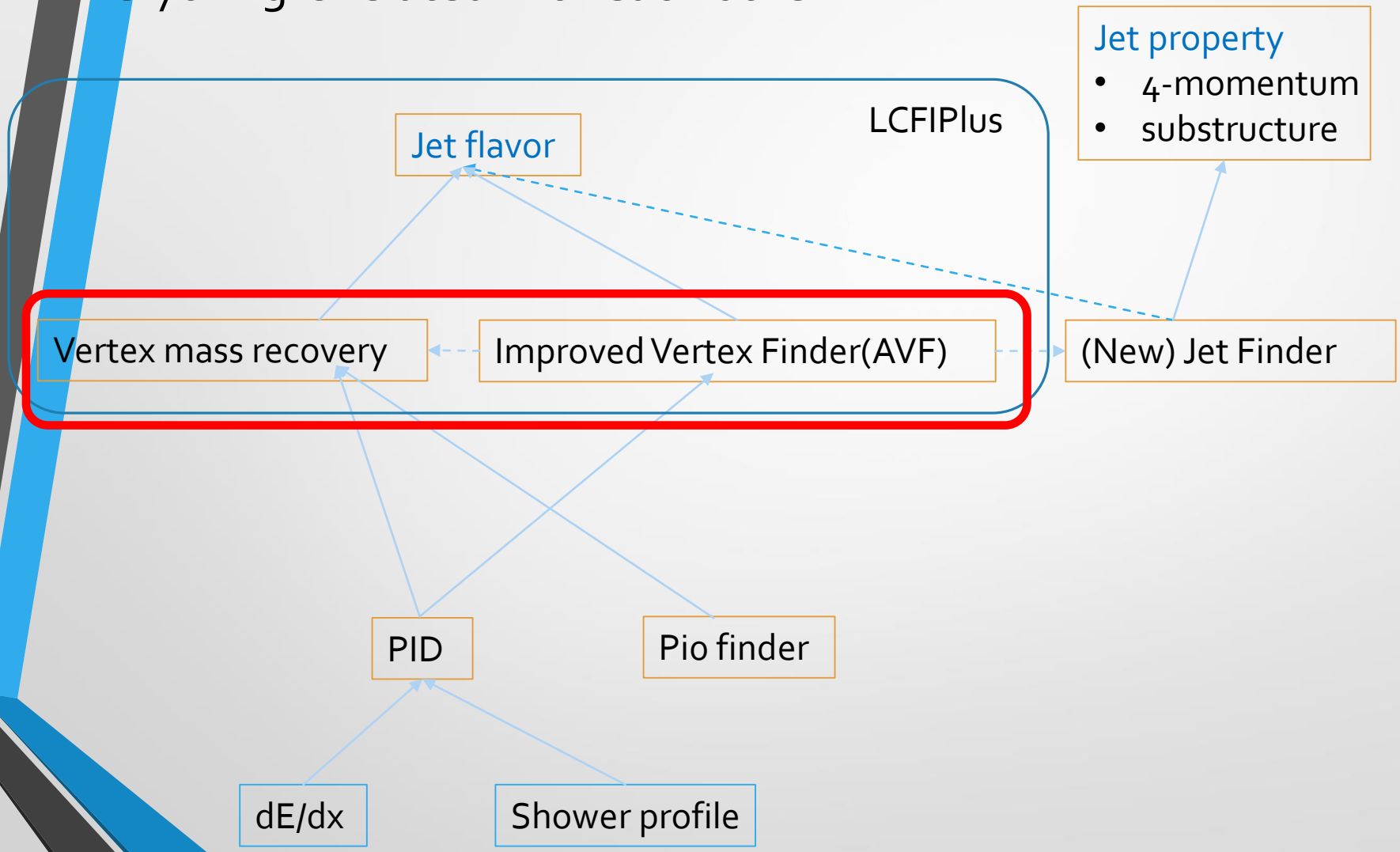
Result so far and plan

- There is no crash in v01-19-04!
  - So as long as using standard environment settings, we can use LCFIPlus as usual
- Plan:
  - Second test file production starts soon
  - We will start flavor tagging training in that round
  - We will ask ~100k test samples
  - Will validate flavor tagging for both detector models

For improvement (01_v05)

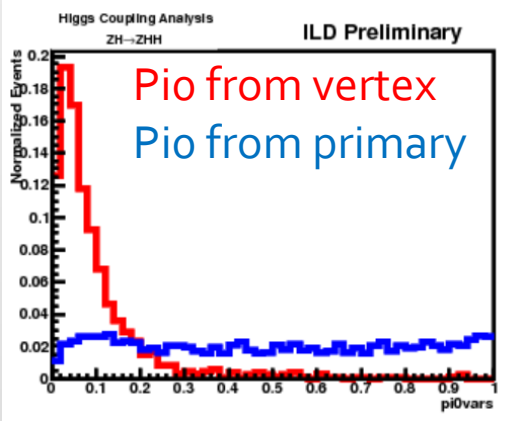
Relationships

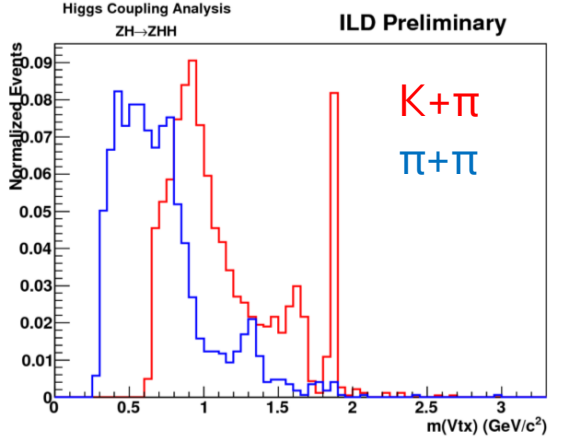
- Everything is related with each other



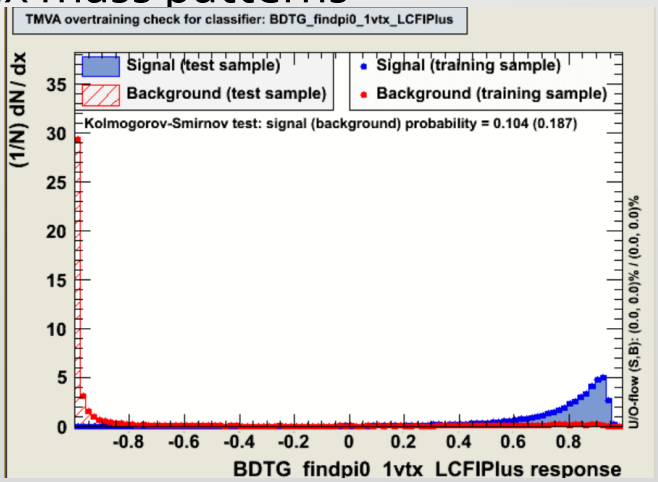
Vertex mass recovery

- Using pios which escape from vertices
  - Need to choose good pio candidates –construct pio vertex finder
  - Key issue –pio kinematics, very collinear to vertex direction





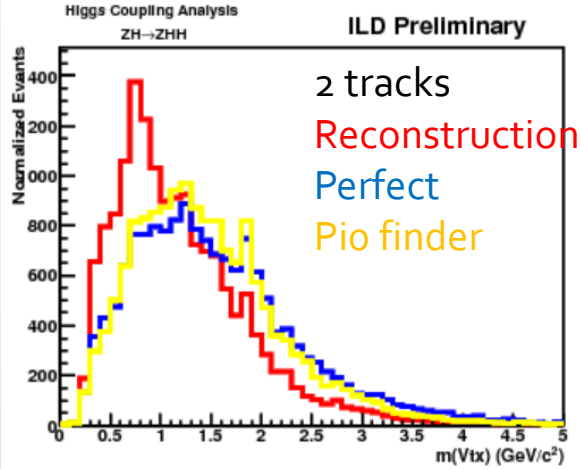
- Particle ID is the other key to classify vertices
  - Different particle patterns have different vertex mass patterns
- Construct Pio Vertex finder using MVA
  - Identify which vertex pios are coming from

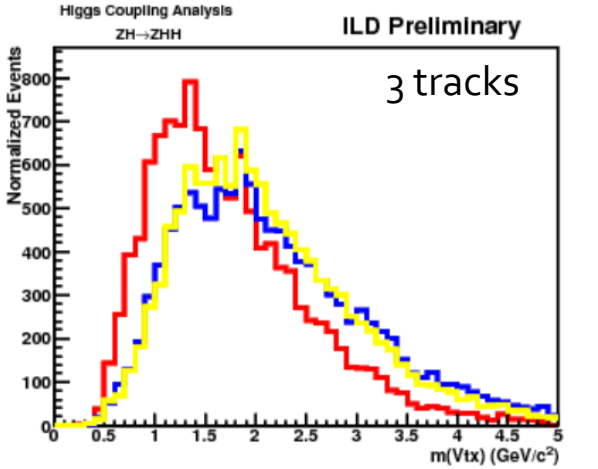


Vtx masses of bjets in double-Higgs process

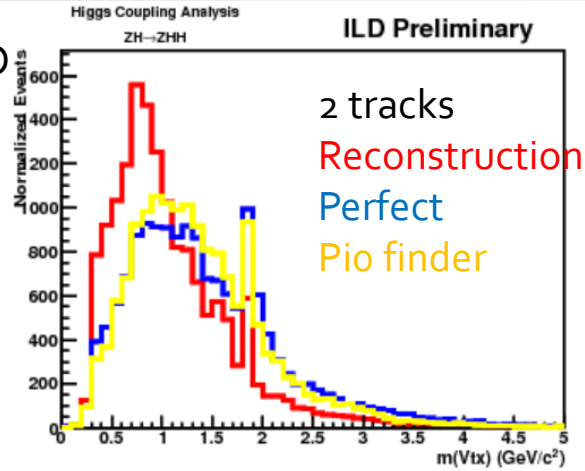
- Vtx mass distributions for each vertex pattern(ntrk)
  - bjets with 1vtx
  - Difference is coming from mis-pairing of gammas(eff. ~50%) and mis-attachment of π^0s

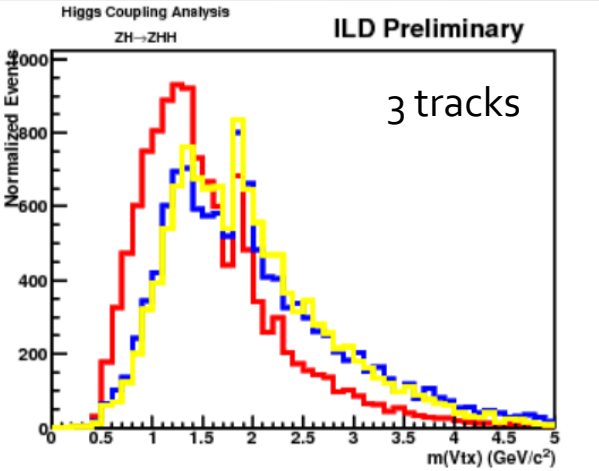
ROOT5 + old PID





ROOT6 + updated PID





Adaptive Vertex Fitting

- To introduce the effect of multi-vertex fitting
  - Introduce weight function to estimate vertex which a track belongs to
  - Weight function definition: k-th track's weight on n-th vertex

$$w_{nk} = \frac{e^{-\chi^2_{nk}/2T}}{e^{-\chi^2_{\mathrm{cut}}/2T} + \sum_{i=1}^{N} e^{-\chi^2_{ik}/2T}}$$

  - Parameter: temperature T
    - If T very small, decision is like χ^2 minimization(almost same as DBD LCFIPlus)
    - If T large, multi-vertex effect becomes large
  - In multi-vertex environment, weight on certain vertex will degrade

    →becomes harder to attach tracks to vertices in multi-vertex environment

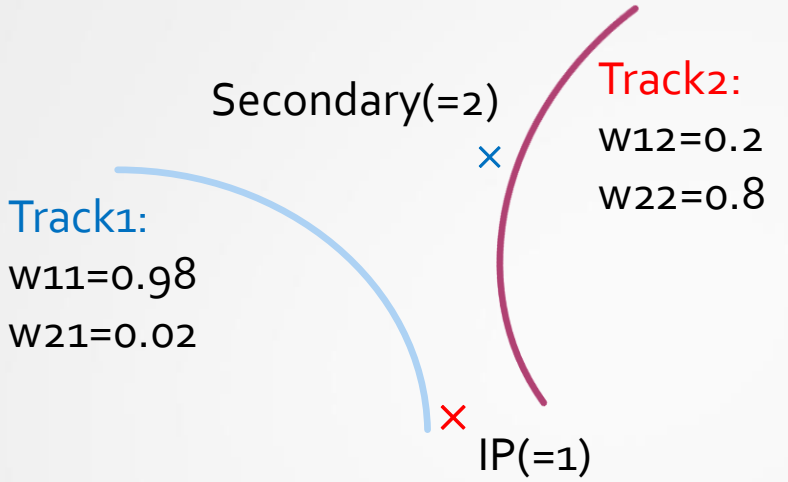
→can reject fake tracks well!

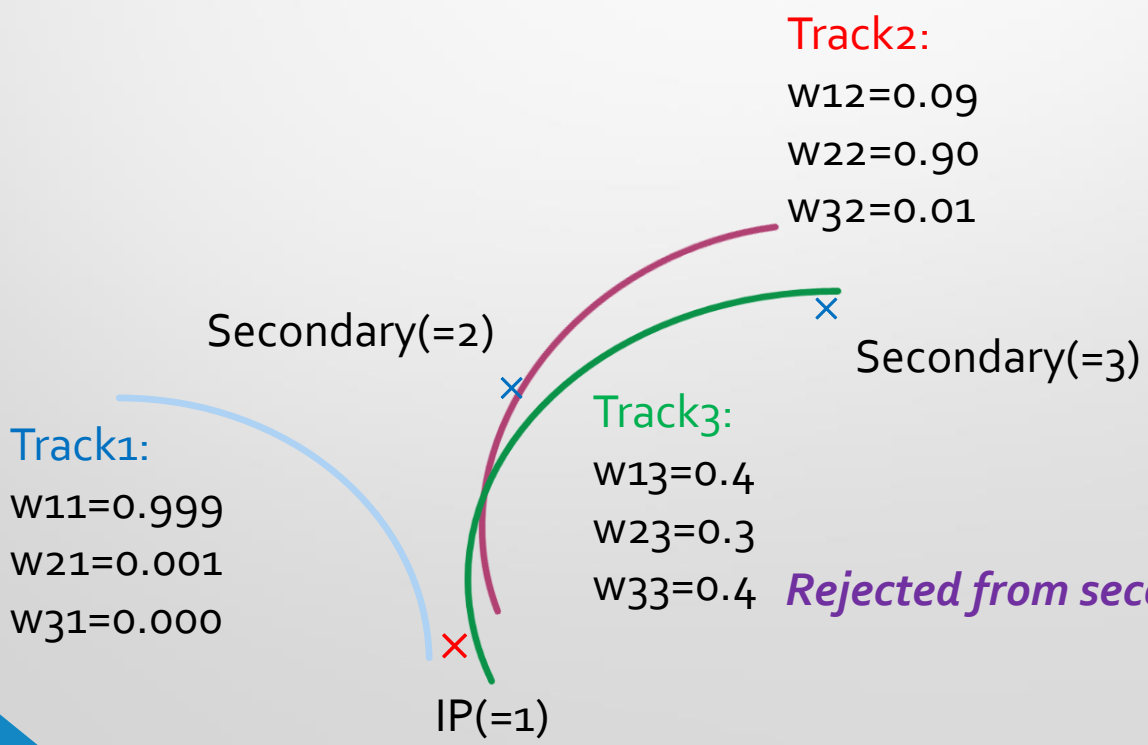
- Thanks to weight function, we can loosen the track quality selection

→vertex finding eff. will be improved!

Examples

- Weight function should be >=0.5





Impact of Adaptive Vertex Fitting

- Common parameters are set at same values for comparison
- Same event sample(qqHH sample@500GeV) 19889 events
- 6 jet clustering, jet matching with MCtruth is performed
- Num. of jets with vertex:

method	bjet with 2vtx	bjet with 1+1vtx	bjet with 1vtx	total
DBD LCFIPlus	10581	9104	12847	32532
AVF	13190	6576	13233	32999

- Total jets with vtx: ~1.4% increased
  - Jets with 2vtx: ~22% increased → good for bjet ID!
  - Jets with 1vtx: ~3% increased → good for uds jet separartion!
- Fake track rate per vtx: how many fake tracks contaminate on

method	bjet with 2vtx	bjet with 1+1vtx	bjet with 1vtx
DBD LCFIPlus	0.029 ± 0.001	0.013 ± 0.0012	0.055 ± 0.002
AVF	0.025 ± 0.001	0.012 ± 0.0013	0.055 ± 0.002

Impact on Flavor tagging efficiency

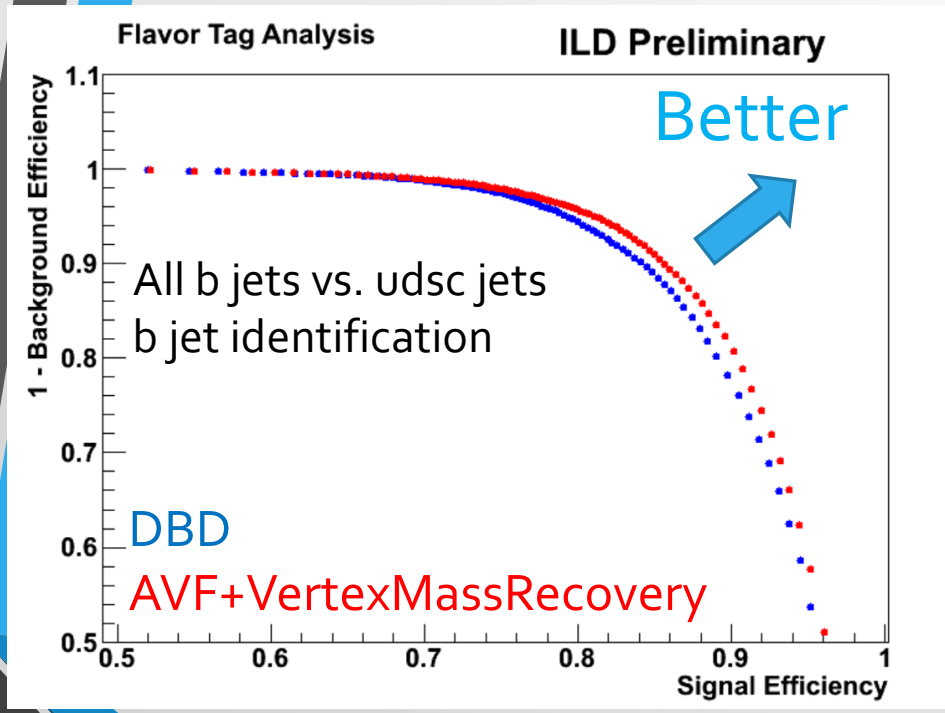
- Need to check final flavor tagging efficiency!
- Using REALISTIC environment of LCFIPlus
  - Classifier: Gradient BDT with multiclass
  - b-likeness, c-likeness, uds-likeness (3-class)
  - 4 categorization with vertices in jets

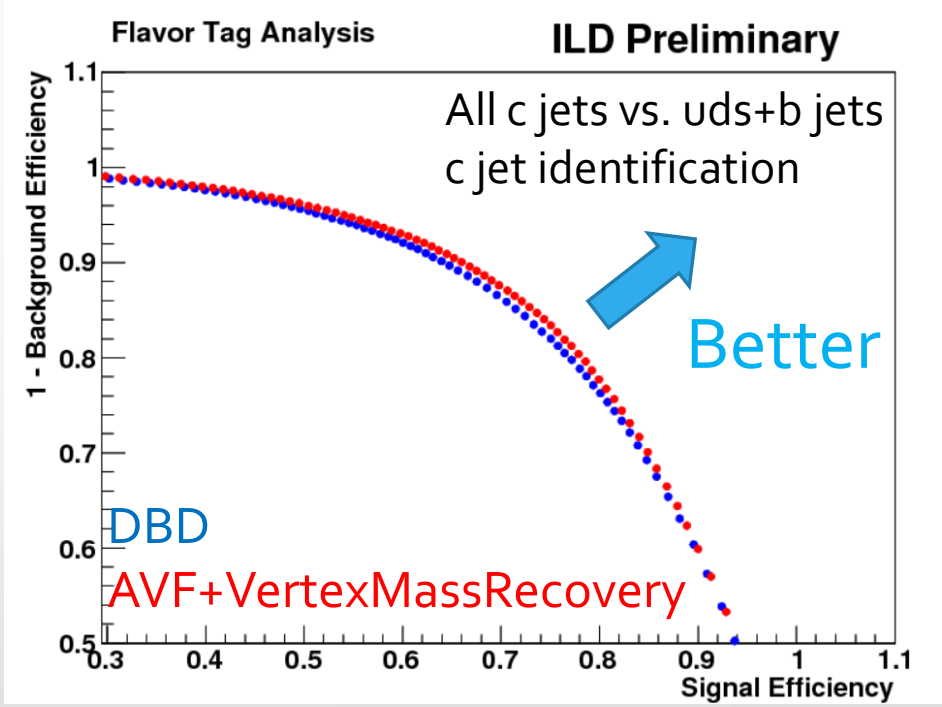
- Train MVA & check b & c-likeliness
  - Samples: 6 fermion samples coming from ZZZ events@500GeV
  - Use AVF & VertexMassRecovery
  - Input variables are add or replaced to their corresponding variables coming from AVF & VertexMassRecovery
  - Check b & c-likeliness of each flavor jet

  Compare ROC curve with DBD ver. LCFIPlus

Impact on Flavor Tagging Efficiency

- 6f samples coming from ZZZ events@500GeV
- Compare with ROC curve

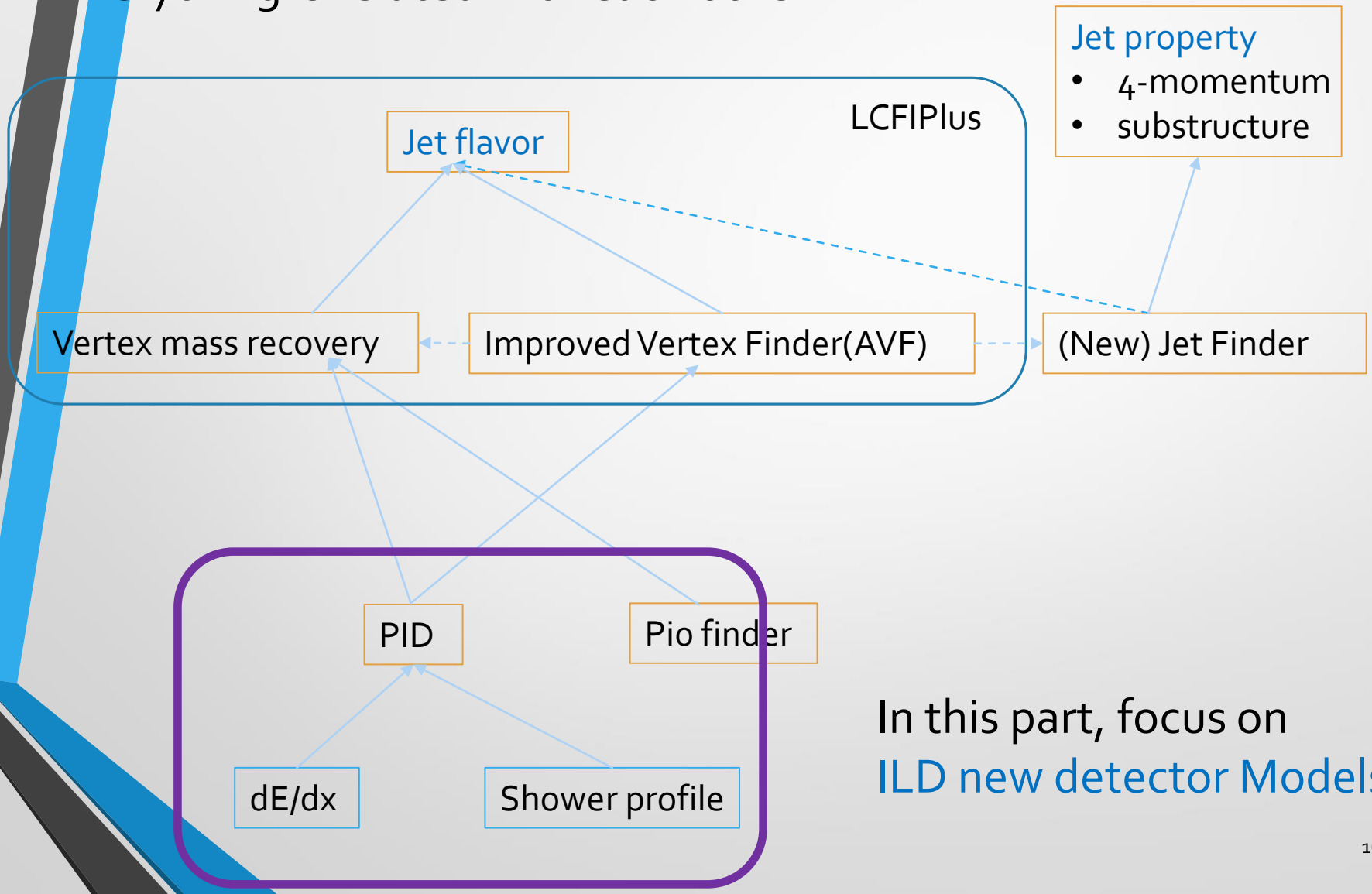




Other Reconstruction tools (New Detector Models)

Relationships

- Everything is related with each other

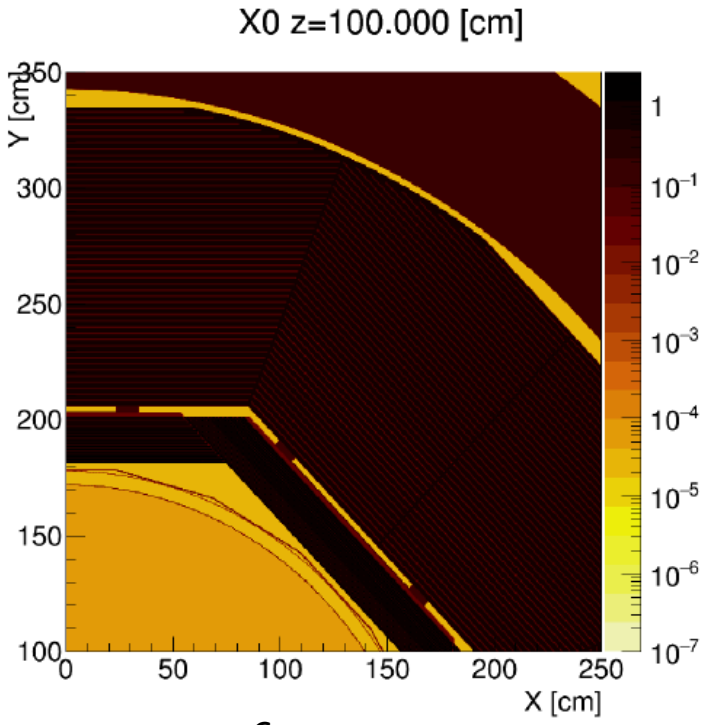


In this part, focus on
ILD new detector Models

2 detector models

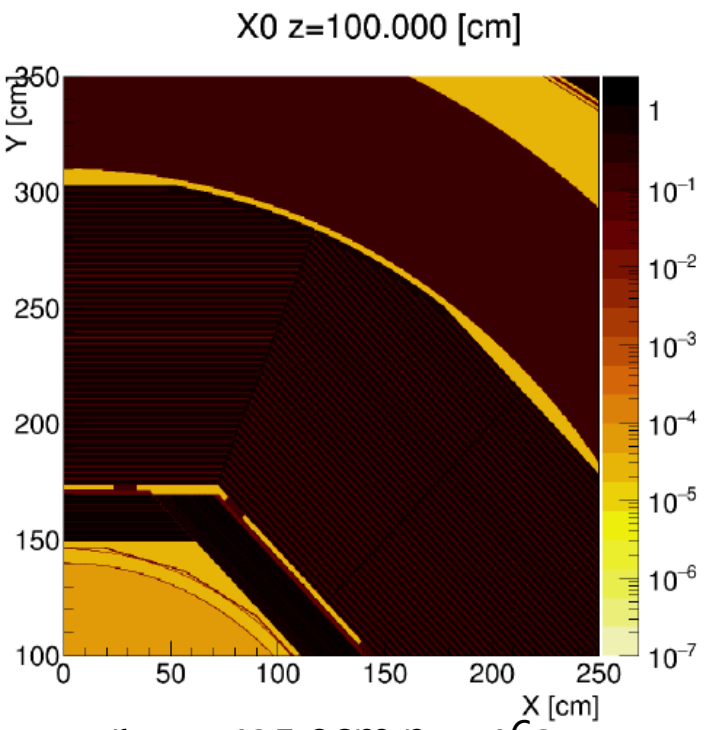
D.Jeans@AWLC2017

large
ILD_l4_v02



r_{TPC}~169.2cm, n_{hit}=220

small
ILD_s4_v02



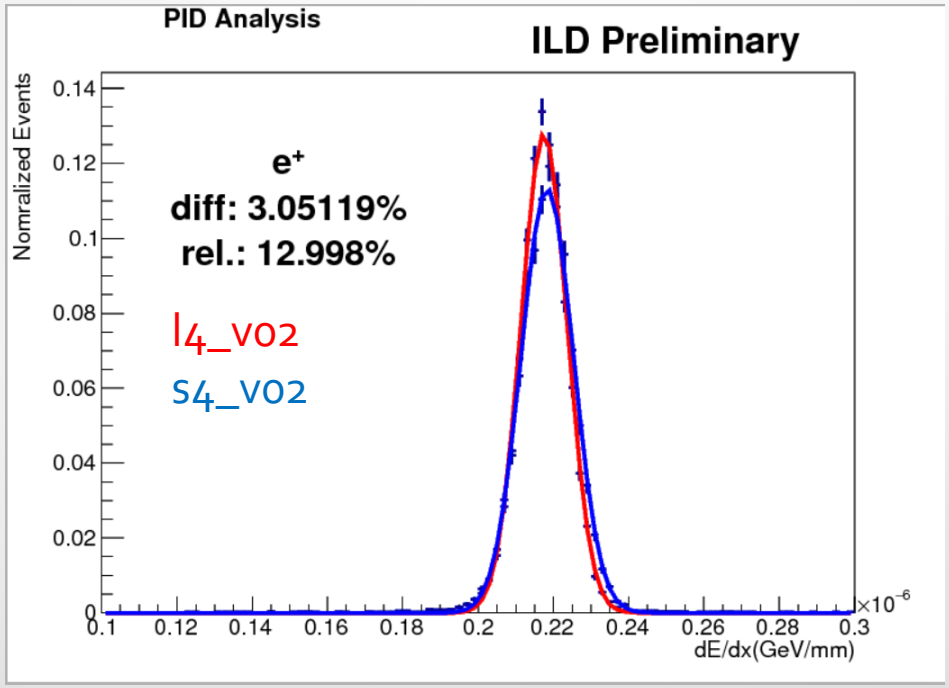
r_{TPC}~135.0cm n_{hit}=163

thicker ECAL, reduced TPC, at different radii

DBD(01_v05): r_{tpc} =171.6cm, n_{hit}=222

New detector models

- dE/dx calculation:
  - Already include ilcsoft standard reconstruction
  - Track dE/dx is calculated using truncation method
    - Upper 30% and lower 8% are discarded
- dE/dx distribution:



- Fluctuation difference(σ/MEAN): ~3%(abs.) ~13%(relative)
  - Natural result
  - Need to check other MIP particles

Additional smearing

- dE/dx resolution of smulation looks too small(Large: 2.8% Small: 3.1%)
  - Need additional smearing

The average of dE/dx resolution

Large	Method① Sampling	Method② Scaling
w/ gating GEM, ϕ= 0°	4.66 ± 0.02%	4.73 ± 0.01 %
w/o gating GEM, ϕ= 0°	4.61 ± 0.02%	4.64 ± 0.01%
w/ gating GEM, ϕ= 20°	4.68 ± 0.02%	4.73 ± 0.01%

The average of dE/dx resolution

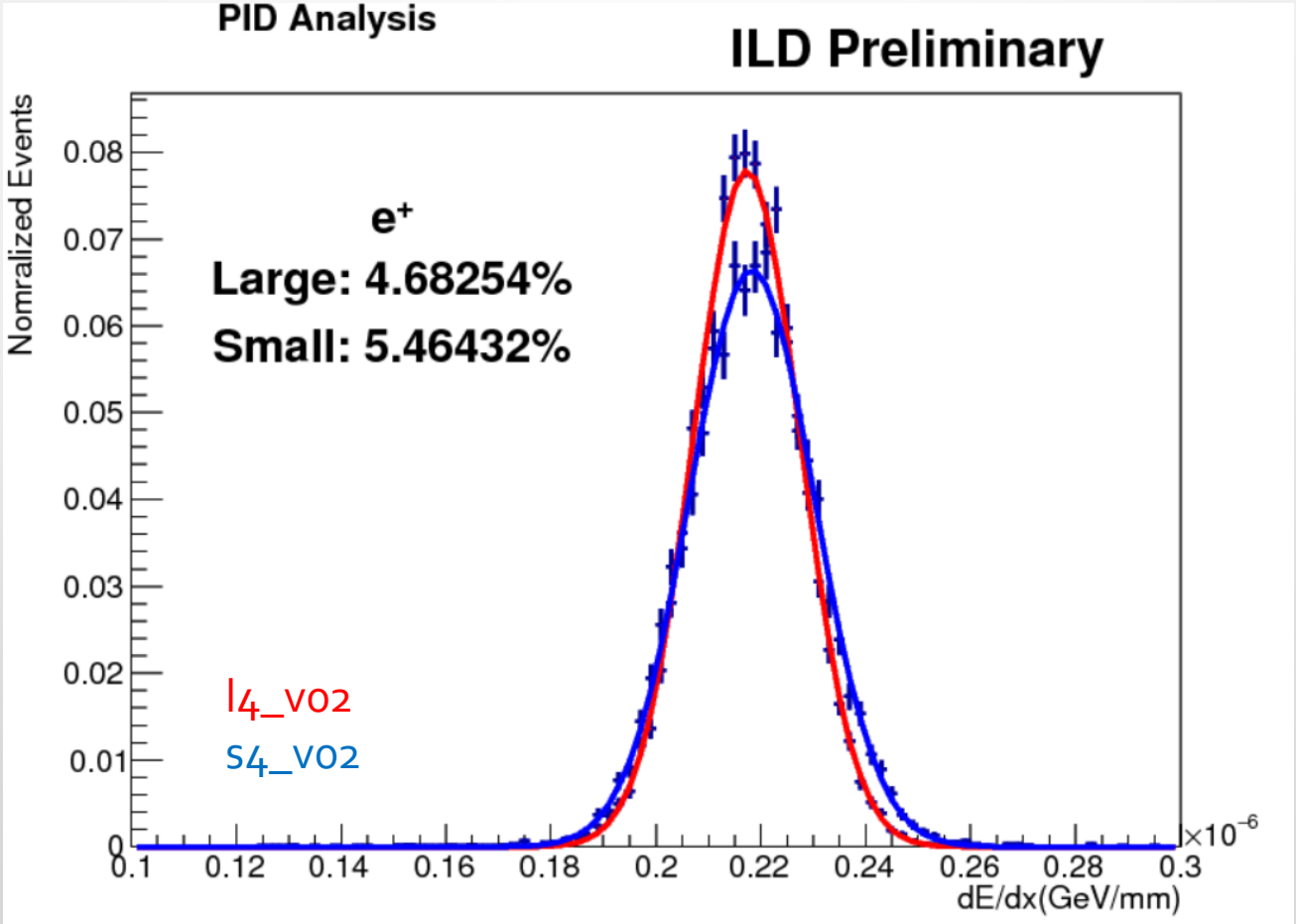
Small	Method① Sampling	Method② Scaling
w/ gating GEM, ϕ= 0°	5.46 ± 0.02%	5.49 ± 0.01 %
w/o gating GEM, ϕ= 0°	5.35 ± 0.02%	5.40 ± 0.01%
w/ gating GEM, ϕ= 20°	5.42 ± 0.02%	5.49 ± 0.01%

A. Shoji

- Add smearing
  - Gaussian noise
  - Target:

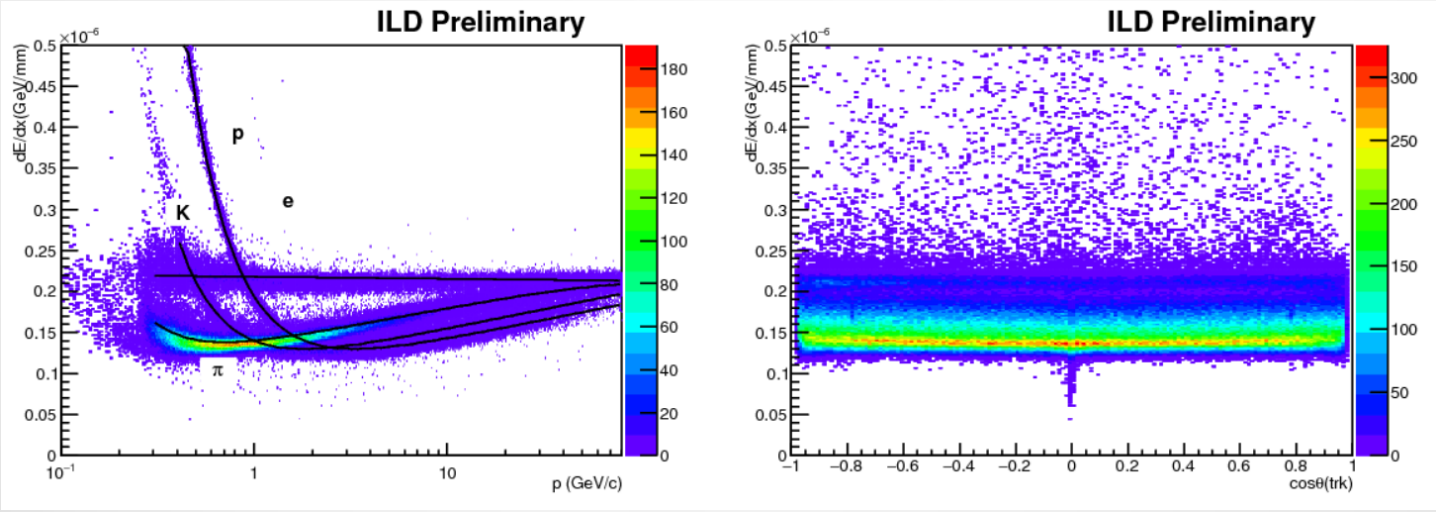
    Large: 4.6%

    Small: 5.4%

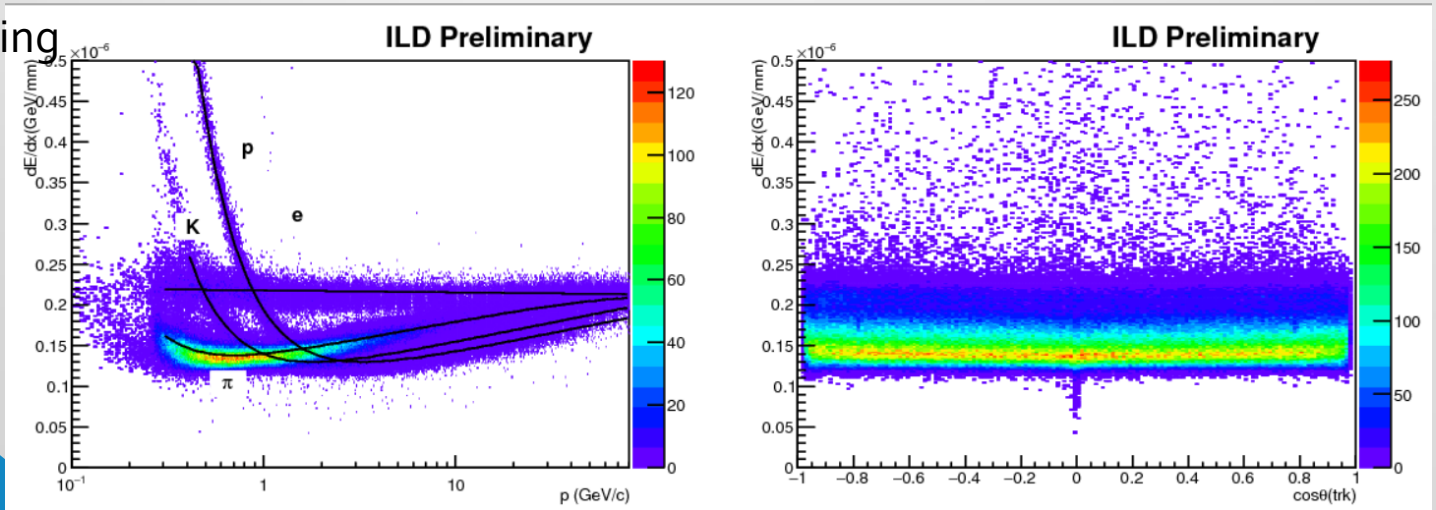


Dependence of dE/dx

- Large detector
  - Use ttbar sample
  - Change angle correction: fit: $\frac{a}{a+b\cos\theta^2}$

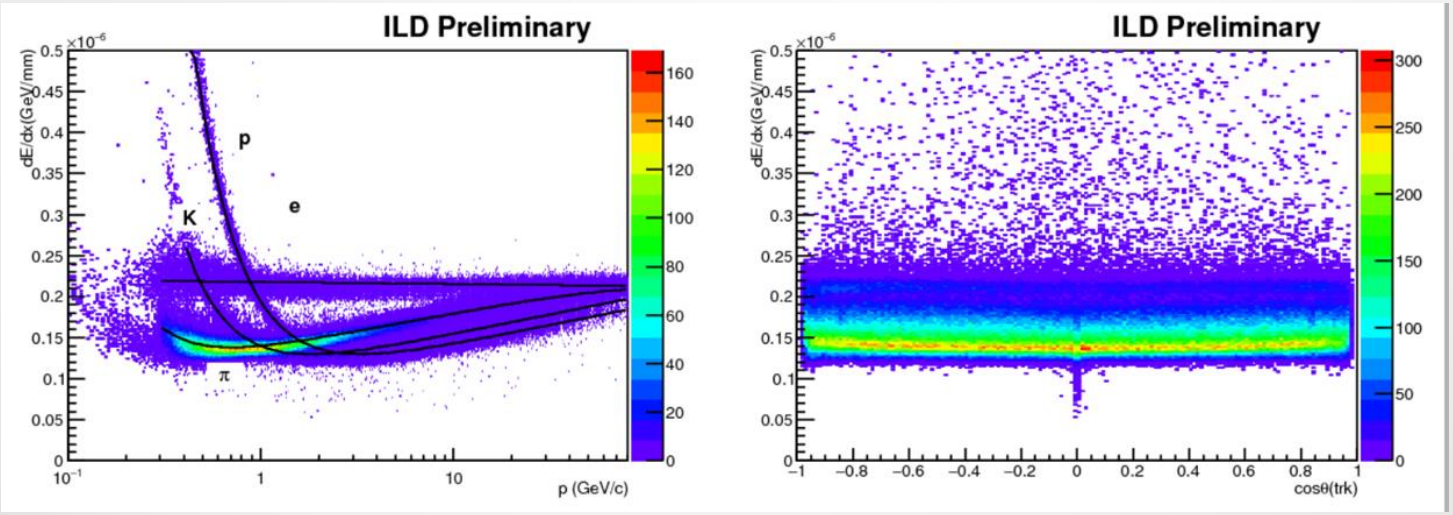


- After smearing

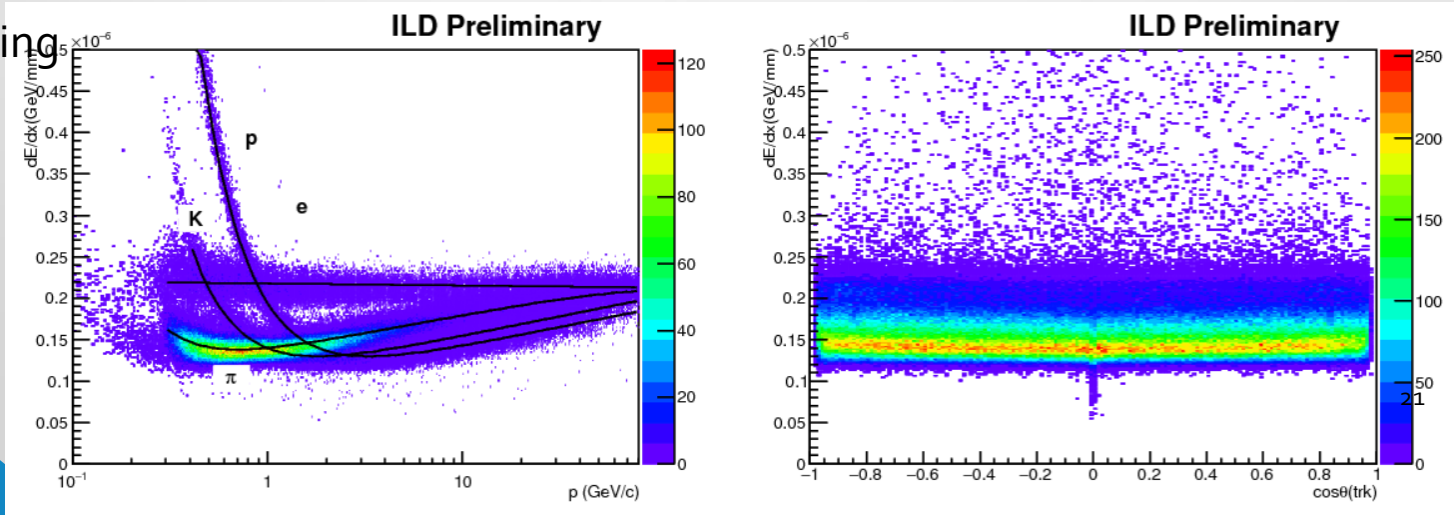


Dependence of dE/dx

- Small detector
  - Use ttbar sample
  - Change angle correction: fit: $\frac{a}{a+b\cos\theta^2}$



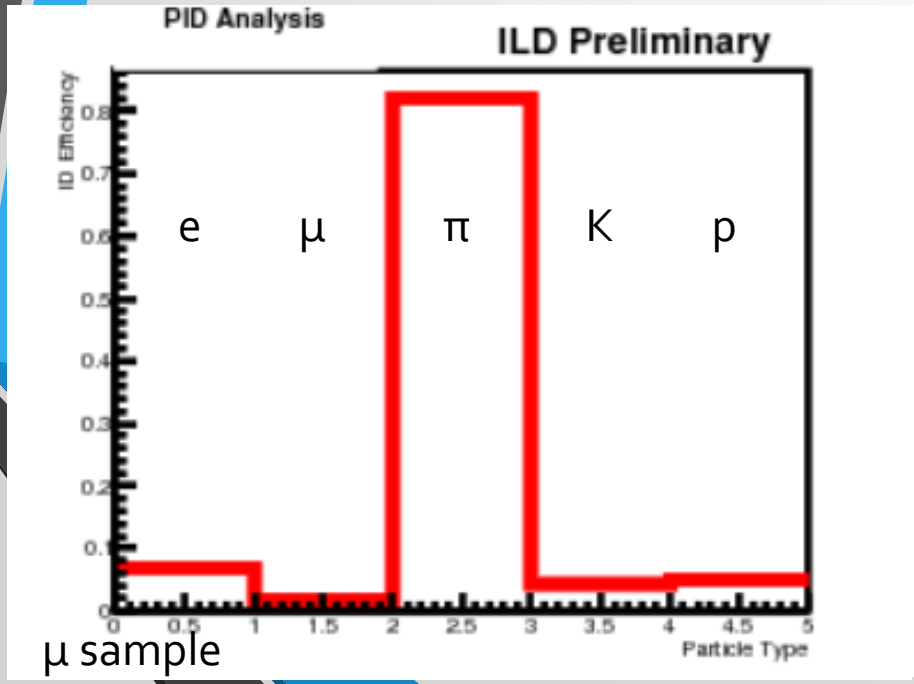
- After smearing



Quick look at PID

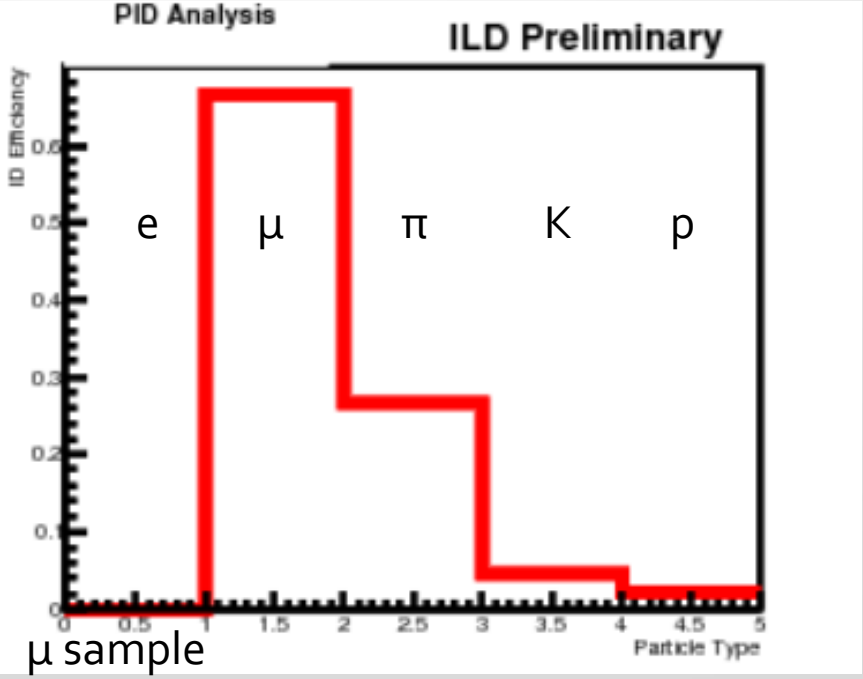
- There is a strange behavior
  - Most of the muons are going to pions…
- Wrong variables referred from PandoraClusters
  - Fix it
  - Looks recovered (but, not muon guaranteed because MCParticle is not available)

v01-19-04



μ sample

correct



μ sample

PID: Plan

- Pull request for change in dE/dx and PID
  - All the existing problems are fixed(dE/dx smearing, angle correction, and strange behavior of PID)
  - Correcting Frank's request now
- In second test sample production, we will check using single track samples
  - Up to 125GeV, both charge
  - 5 fundamental particles
  - Trying to create template file for PID effciency(enough statistics??)

Summary and prospects

- We are going to adjust LCFIPlus to ROOT6 environment
  - We check sudden crash in ROOT 6 environment
    - No crash occurred in standard settings in v01-19-04
- For flavor tagging improvement:
  - Updated PID provides better reconstruction of (charged)vertex mass
  - Vertex mass recovery is still reasonable
- Estimate flavor tagging efficiency using multiclass classifier
  - Efficiency improvement can be seen using AVF+ VertexMassRecovery!
- We adjust dE/dx and PID for ILD new detector models
  - Difference between large/small detector is reasonable, but need additional smearing
  - Strange behavior on PID is fixed, but need more single particle samples for check
- Going to validation in second round!

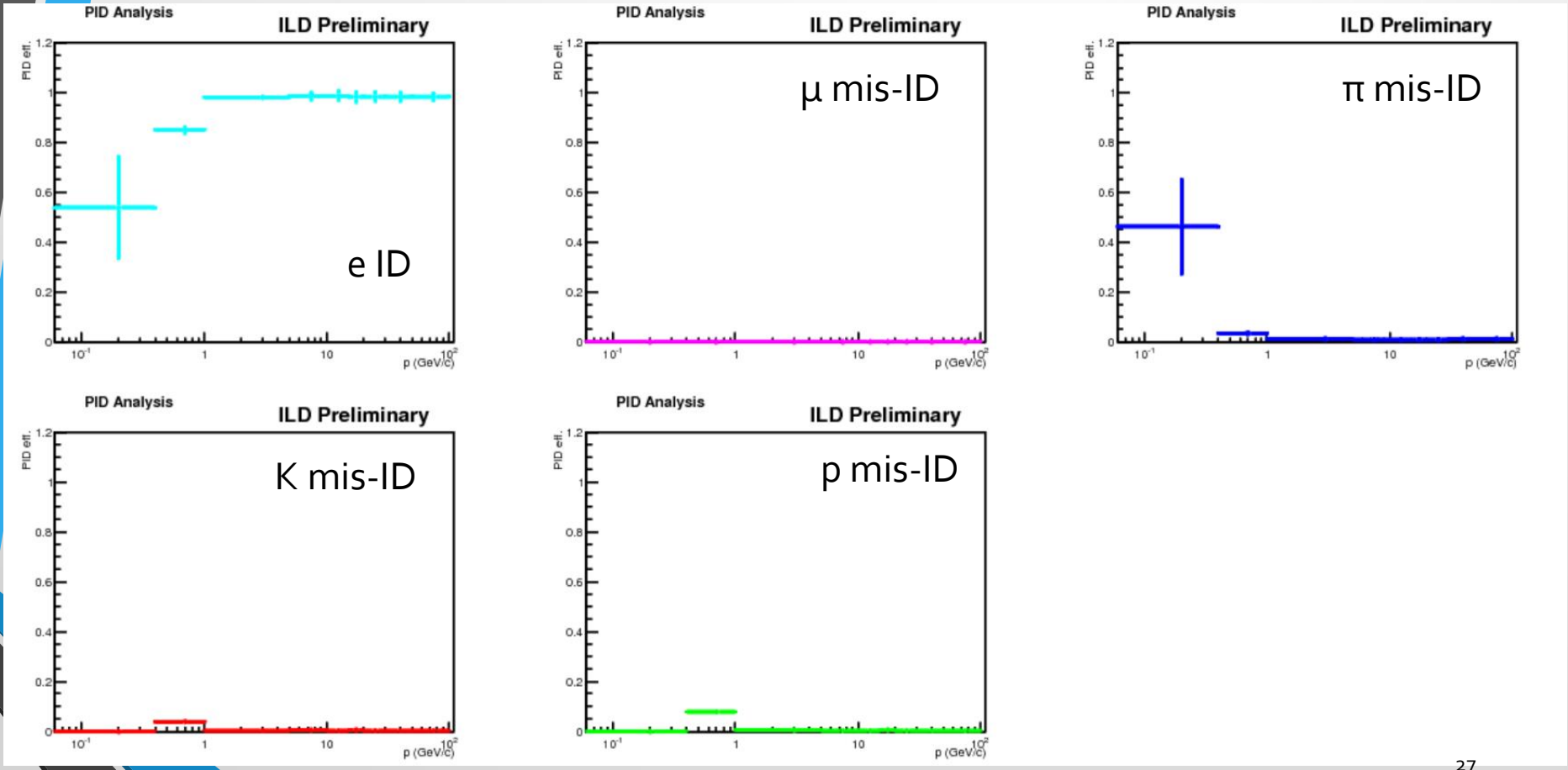
Back ups

PID

- Use dE/dx, Shower shapes in Calorimeters, E/P, etc.
- Construct Bayesian classifier
  - Estimate posterior probability of each particle hypothesis
- μ/π separation for low momentum tracks included
  - Developed For degenerated Higgsino study
  - Use special shower shape information of μ/π at forward calorimeter
  - Using TMVA for separation
  - Effective <2GeV/c tracks
- Simple example: Overall ID & mis-ID efficiency:

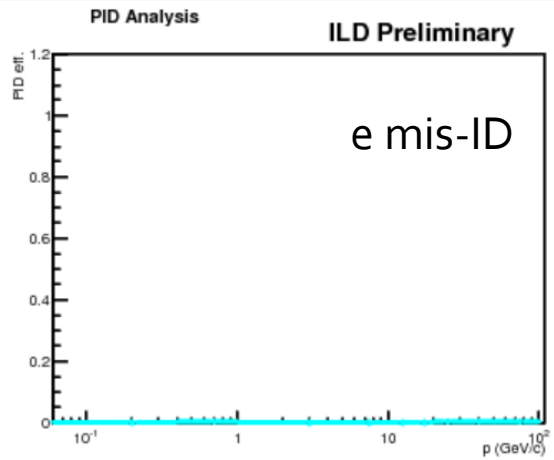
Performance example

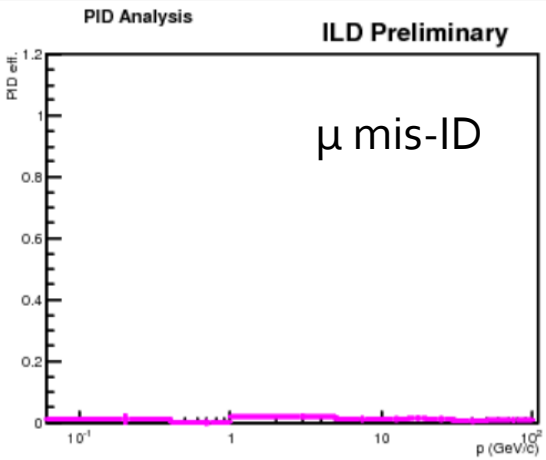
- Single track electron
- Momentum dependence of ID & mis-ID efficiency

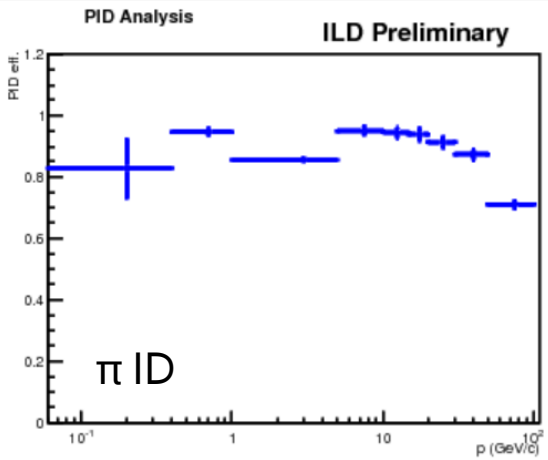


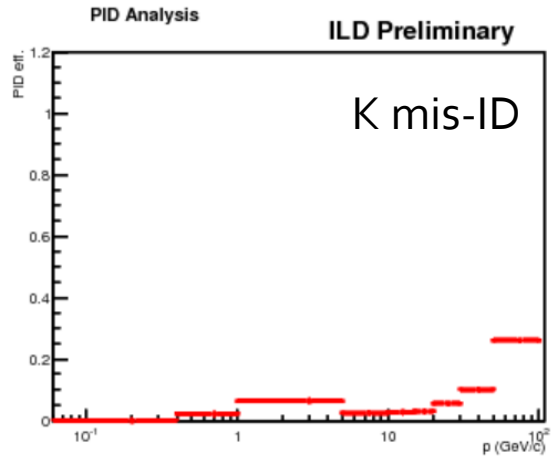
Performance example

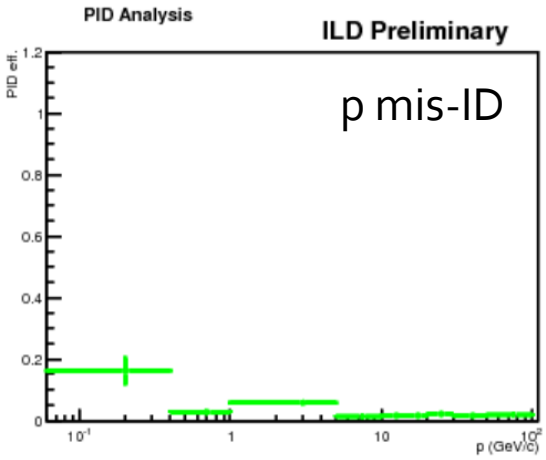
- Single track pion
- Momentum dependence of ID & mis-ID efficiency





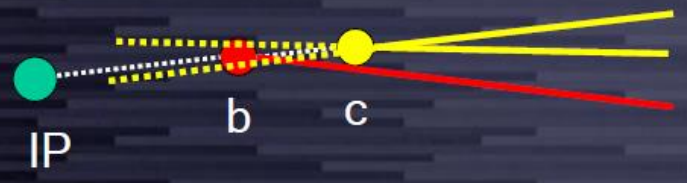






Issues / Vertex finder

- Apply recent developments
- Not robust against background
  - $\gamma\gamma \rightarrow$ hadrons (CLIC performance very bad)
  - pairs
- b/c separation
  - More efficient finder $\leftarrow\rightarrow$ worse b/c separation
  - Should be treated with different vertex finder
- Association of low-energy tracks



Issues / Vertex finder (cont.)

- Refitting tracks
  - may improve the vertex separation
  - Need tracker hits (not available in DSTs)
  - Kalman filter or ...
- Speed of vertex finder
  - Vertex fitter is slow
    - 2 loops of Minuit minimization
  - Vertex finder is also slow
    - trying every pair of tracks

Issues / Jets, Leptons

- Jet clustering
  - Interface to external? eg. Fastjet?
  - Color-singlet? kinematic constraint? More?
- Lepton finder
  - Apply LikelihoodPID instead of simple one
    - Function is already available,
      need training and checking performance
  - Tau finder

Issues / flavor tagging

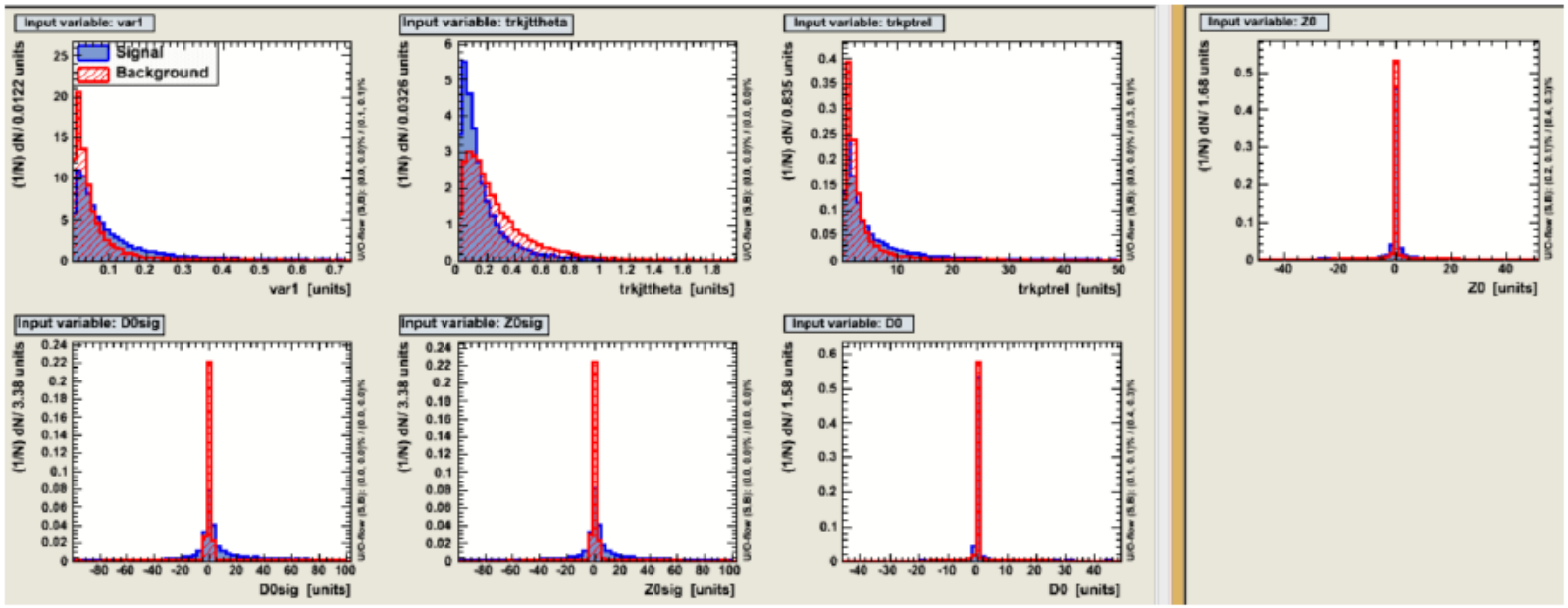
- Treatment of multiple jets inside (Jet substructure ID etc.)
- Treatment of vertices
  - "Concrete" or "Doubtful" vertices
    - based on probability, # of tracks etc.

Issues / others

- Systematic errors from flavor tagging
  - Only 'ballpark estimate' available
  - Should be studied in control samples
  - Application to each physics analysis
- Interface to LCIO/Marlin
  - Some problem from multiple PFO collection
  - External jet clustering
- Documentation!

TRACK MVA(BNESS)

- To identify track which comes from heavy flavor particle
  →using MVA
  - Signal: tracks which come from B mesons or B baryons
  - Background: tracks produced in hadronization process
- Most significant tracks with both plus and minus signed impact parameters in a jet are collected
  - Significance: $sig = \sqrt{(\frac{d_0}{\sigma})^2+(\frac{z_0}{\sigma})^2}$



Bness tagger for fake track rejection

- Loosen the track selection to try to attach as many tracks as possible to vertices
  - Fake track rate will be increased
- To reject fakes, BNess tagger is used
  - So far, just use BNess(bl)
- So far, only BNess

is checked

→some bias for D meson

tracks?

Ex

track