PARTICLE ID STUDY FOR ANALYSIS IMPROVEMENT

Masakazu Kurata
The University of Tokyo

ILD Meeting, 09/08/2014

For analysis improvement

- All the analyses are saturated within the present framework
  - Needs new idea
- Fundamental new variables might provide improvements of analysis tools
  - dE/dx in TPC
  - Shower profiles in the calorimeters
- Those variables have already boosted lepton ID efficiency
  - Improvement can be obtained!
  - Show that later
- Will those variables give improvements to other analysis components?
  - Particle ID will be available using those variables
  - Energy correction
  - Flavor tagging? →looks hopeful!
  - Hope for jet clustering?

→it is necessary to study them

dE/dx

- For improvement, using dE/dx is one of the powerful tools
  - Particle ID for each track will give a large impact to the analysis
  - Application to general analysis component is very wide
    - Lepton ID
    - Track energy correction
    - B-tagging
    - Jet clustering?
- Important factor to use dE/dx is: fluctuation
  - TDR: measurement resolution is **5%**
  - So, fluctuation from simulation is within 5% without detector effect

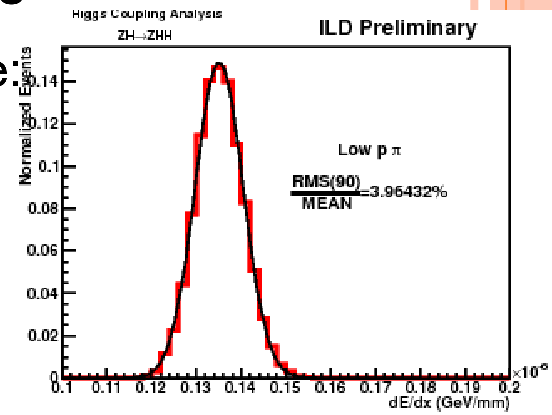
- dE/dx definition:
  - $\frac{dE}{dx} = \frac{energy\ deposit}{flight\ path\ in\ the\ hit(TPC)}$
  - dE/dx can be calculated at any hit point
  - Truncated mean is calculated as track dE/dx

    $\left\langle \frac{dE}{dx} \right\rangle = \frac{1}{n} \sum_i^n \frac{dE_i}{dx_i}$ upper 30%, lower 8%(important!) hits are discarded to avoid Landau tail

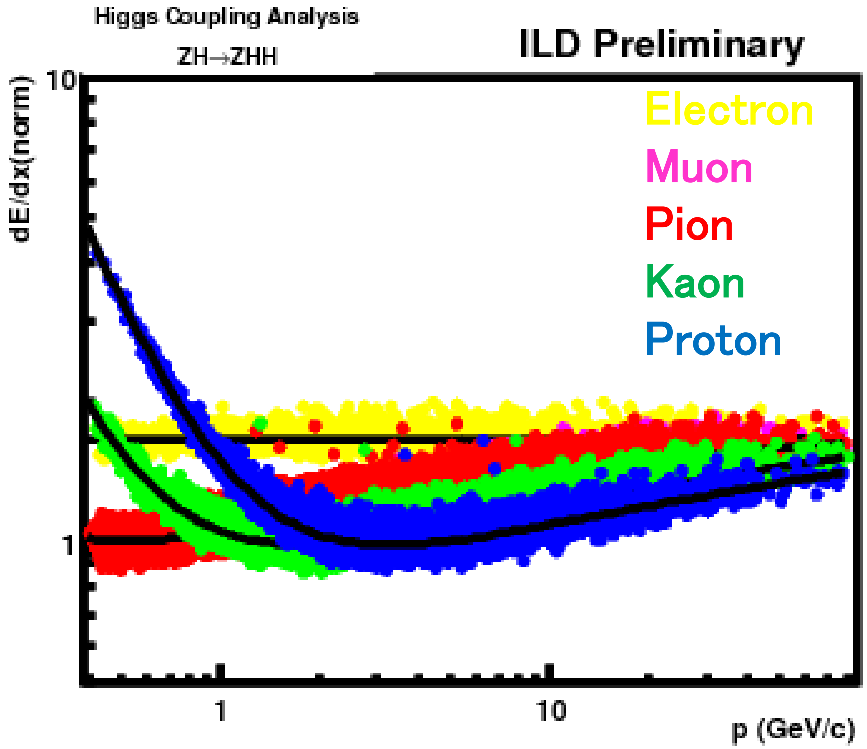
    →optimization is necessary

dE/dx Fluctuation and Distribution

- Fluctuation of dE/dx using various type of tracks
  - Fluctuations of each particle/each momentum range:
    3 – (<5)%!!
  - Including detector effect is necessary



- Momentum dependence of dE/dx for each particle
  - Polar angle dependence corrected
  - Num. of Hits dependence corrected
  - Scale to $\left\langle \frac{dE}{dx} \right\rangle = 1.0$ for MIP pion



Shower profile

- Shower shapes in the calorimeter are different between electron/photon/muon/hadrons
  - So characters of the clusters will be a good tool to distinguish tracks
  - Especially, electromagnetic shower shape is well known
  - Grabbing those information will boost leptonID efficiency/fake rejection efficiency

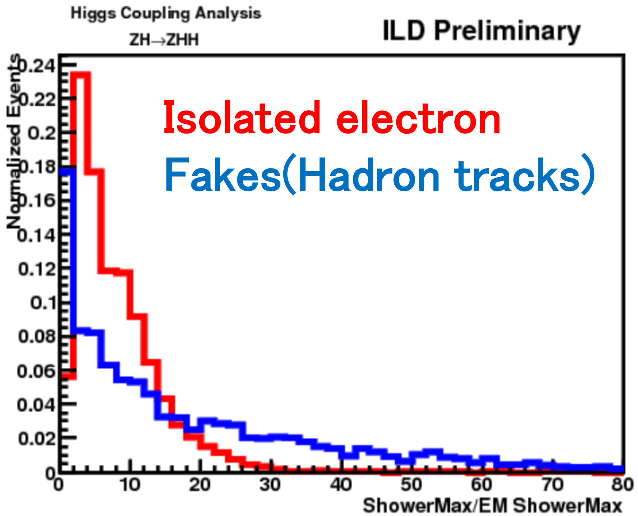
- Information extraction is based on the fitting:
  - Well-known EM shower profile

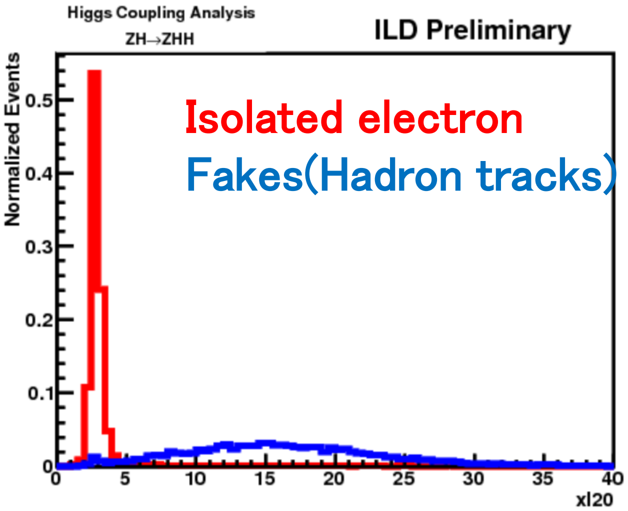
$$f(x_l, x_t) = ac\frac{(c(x - x_{l0}))^{b-1} \cdot \exp(-c(x - x_{l0})) \cdot \exp(-dx_t)}{\Gamma(b)}$$

- In addition, hit based variable is introduced to identify shower start
  - Xl20 – length from cluster start to 20% of total energy deposit

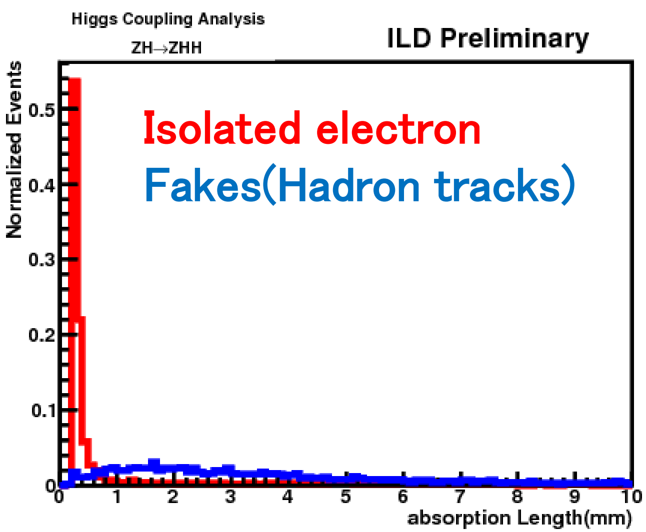
SHOWER PROFILE

- Longitudinal information – shower Max. & shower start position





- Transverse information – Absorption length



Application – Lepton ID

- Lepton ID for single lepton – using likelihood method
  - Lepton likeliness: $L = \frac{\prod s}{\prod s + \prod b}$,
- Signal detection efficiency – set almost same efficiency
- Signal is HH→(bb)(WW*)→(bb)(l ν jj)

method	Cut based	Likelihood_old	Likelihood_new
Signal(%)	98.1	98.1	97.8

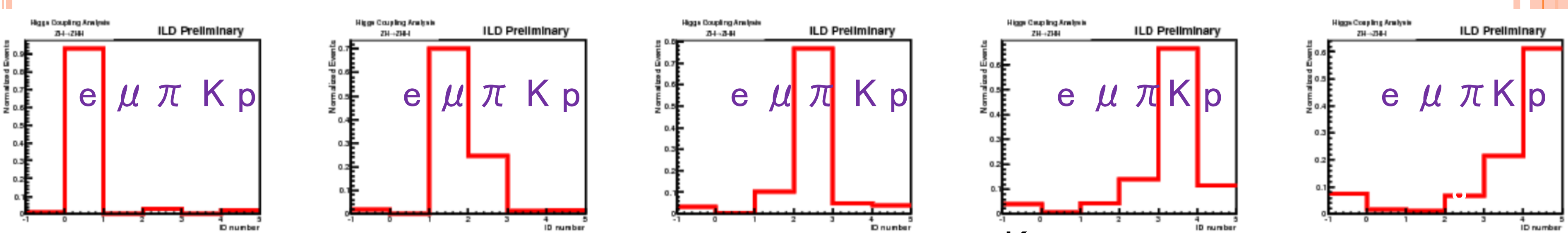
- Background rejection efficiency:

method	Cut based	Likelihood_old	Likelihood_new
ttbar – lep+jets(%)	62.2	-	62.4
ttbar – allhad(%)	7.9	3.1	2.3
ttbar – dilepton(%)	47.2	-	17.9
HH→(bb)(bb) (%)	-	2.3	1.0

  - Improvement of all hadronic event rejection: ～30%
  - Note: lepton energy threshold is loosened on likelihood_new
    - From E(lep)>15GeV → E(lep)>10GeV

Particle ID

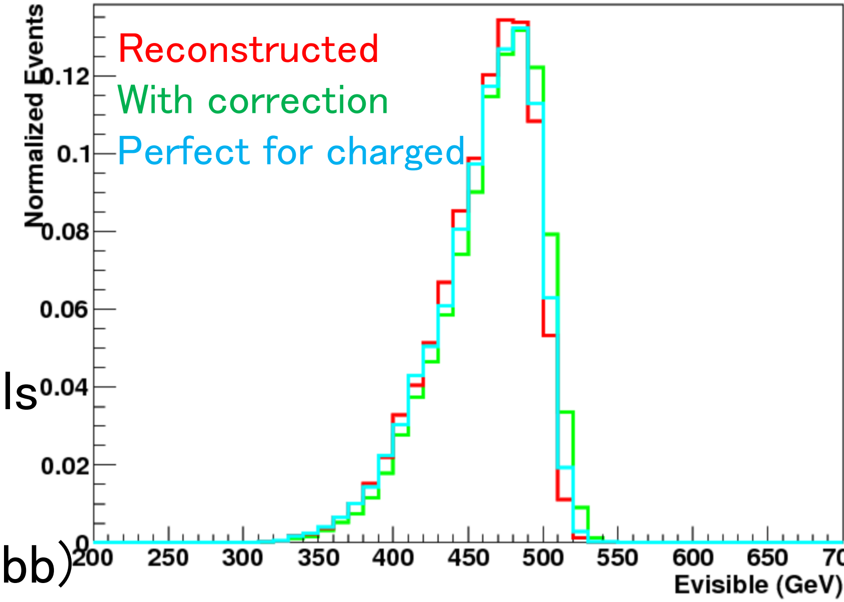
- New variables make Particle ID available
  - How are particles identified as each particle type?
- Construct Particle ID algorithm:
  - Based on maximum likelihood method(Bayesian approach)
  - Make “rejected” category:
    - Track is rejected if its likelihood(posterior) is below threshold
    - Those tracks are moved to pions
- ID efficiency:
  - Electron can be identified almost perfectly(>90%)
  - Muon ID eff. is ～70% →due to low energy muons(μ/π separation)
  - Hadron ID effs. are 62%～75%

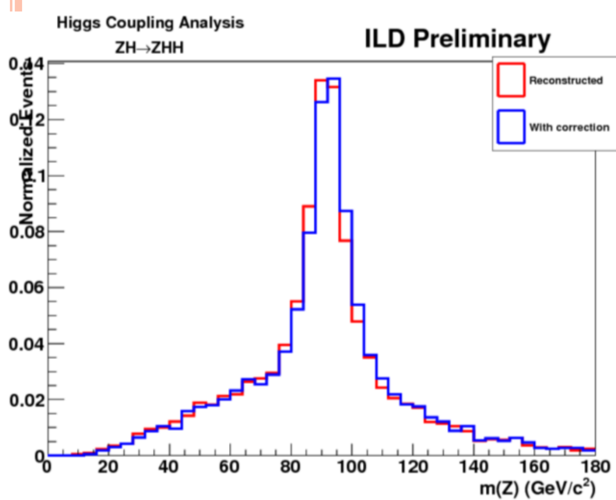


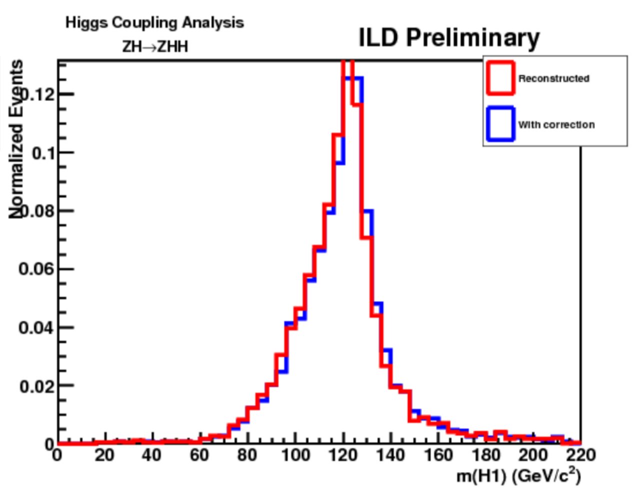
e μ π K p

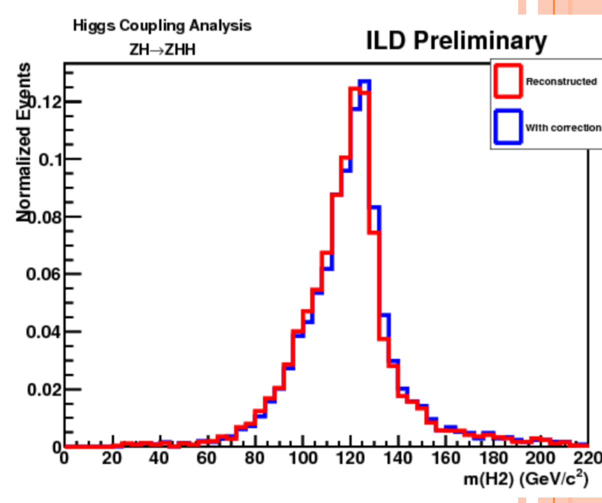
TRACK ENERGY CORRECTION

- Track energies are corrected using momentum & mass
  - Using particle ID to identify tracks
- Visible energy
  - Using qqHH→qq(bb)(bb)
  - So far, overestimated due to misID
  - Correction effect is small due to neutrals
- Mass distribution
  - Checking Z(Z→qq, q is light) and H(H→bb)
  - Jet matching with MC truth is applied
  - Effect is small too due to neutrals



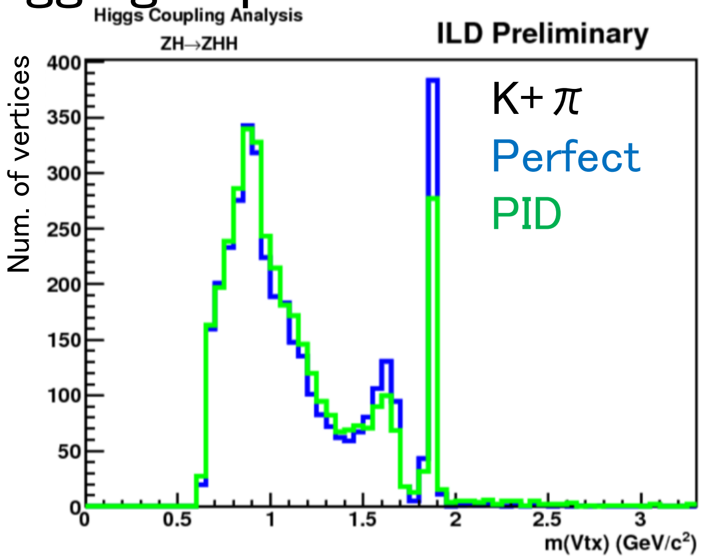




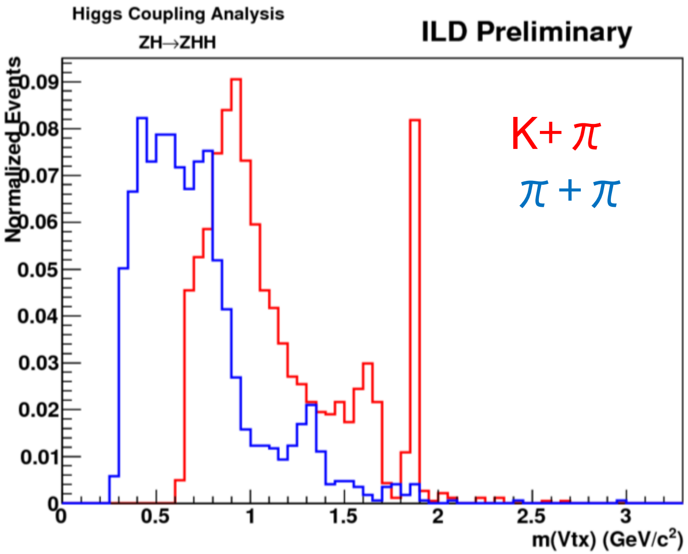


Vertex classification

- Can Particle ID be used for flavor tagging improvement?
  - Checking vertex mass distribution
  - Vertex is from LCFIPlus
  - How much effect on vertex mass?

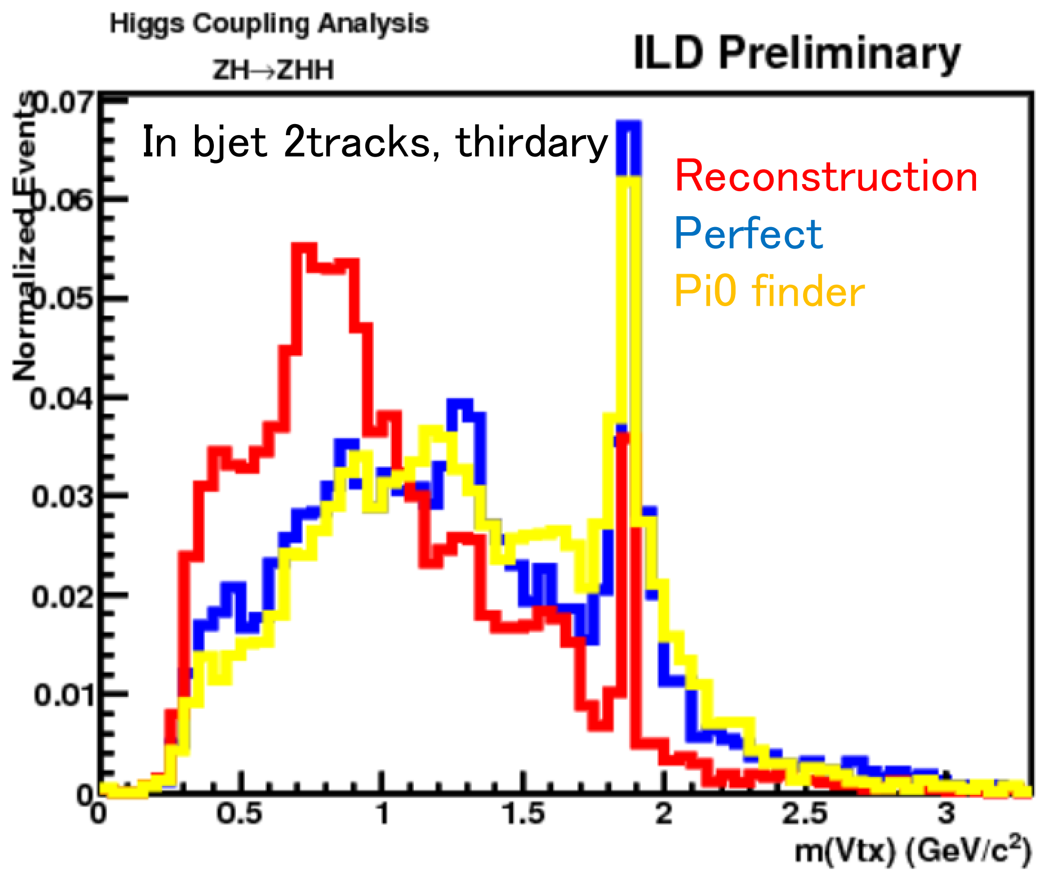


- Classifying vertices with particle type using particle ID
  - Different vertex pattern has different vertex mass pattern
  - e.g.) K+π v.s. π+π



VERTEX MASS RECOVERY

- Can the vertex mass be recovered?
  - Possibility of attaching pi0s which escape from vertices
  - Particle type on vertices is of course the key point! → particle ID
  - D meson mass peak will be a landmark for the study
  - Looking for gammas from neutral particles → gamma finder
  - Constructing pi0s from 2 gammas → pi0 finder
  - Looking for pi0 candidates



  - Study on going
  - How is the effect on flavor tagging?

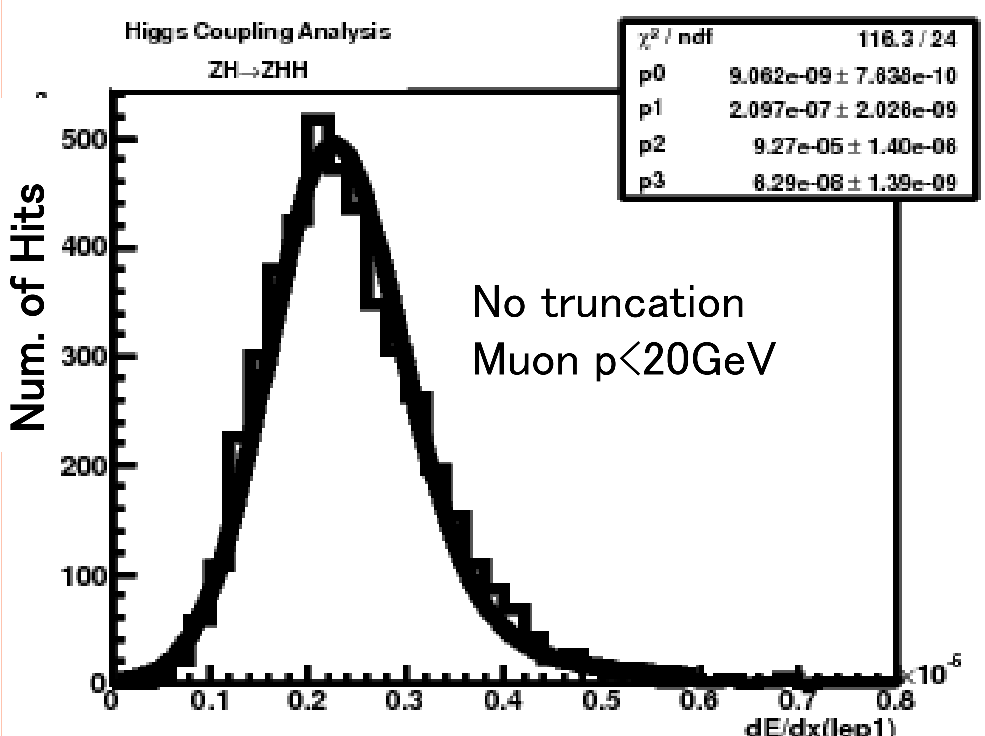
Summary and Todo

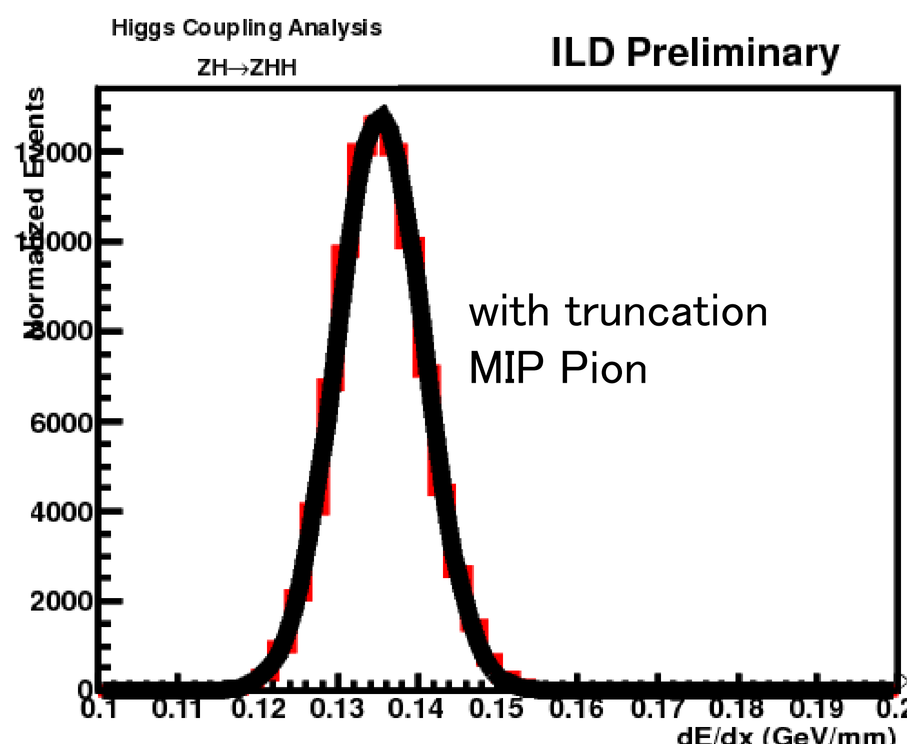
- Explore some fundamental variables for analysis improvement
  - dE/dx in TPC and shower profile
- dE/dx and shower profile information provide ～30% improvement for Isolated lepton ID
- Studying particle ID:
  - Hadron ID eff. is 62%～75%
  - Energy correction effect is small, but going to good direction
  - Vertex mass recovery is hopeful using particle ID
- Todo:
  - Particle ID optimization
  - Check vertex recovery effect on flavor tagging
  - Is there room for flavor tagging improvement in 0 vertex jet case?
    - Kaon tag is the key?

BACKUPS

Effect of Landau tail

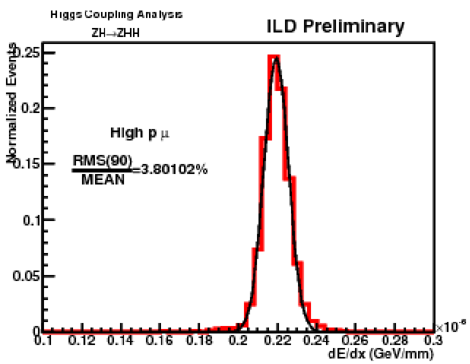
- Landau tail effect – muon tracks
  - dE/dx distribution of track hits
  - fitting – convolution of Gaussian and Landau
  - Tail can be seen in the case of no truncation
- Truncated mean distribution – MIP pion(0.3GeV/c<p<0.6GeV/c)
  - Good Gaussian shape

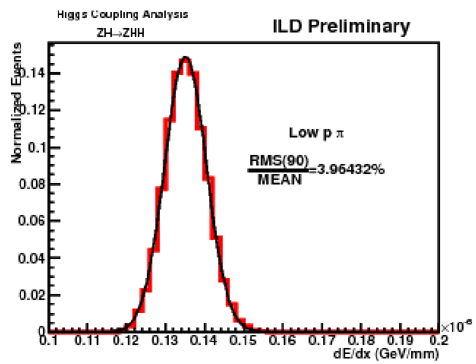


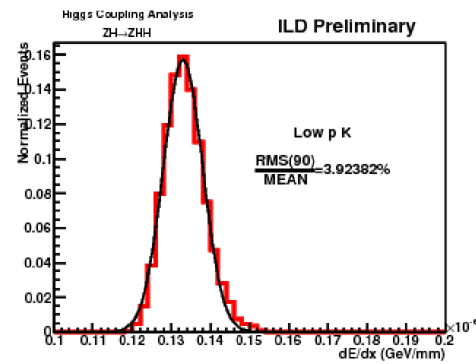


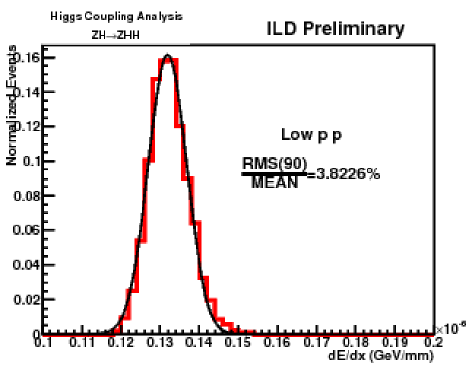
dE/dx FLUCTUATION

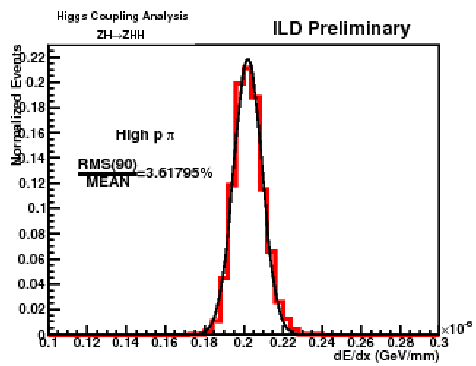
- Estimation of RMS(90)/MEAN

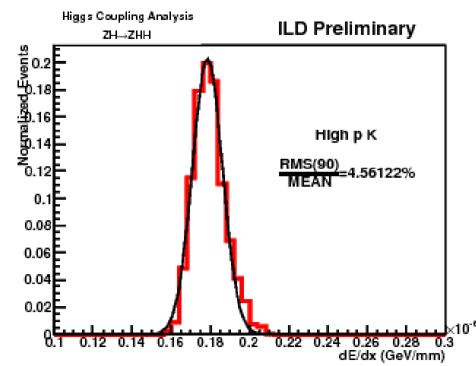


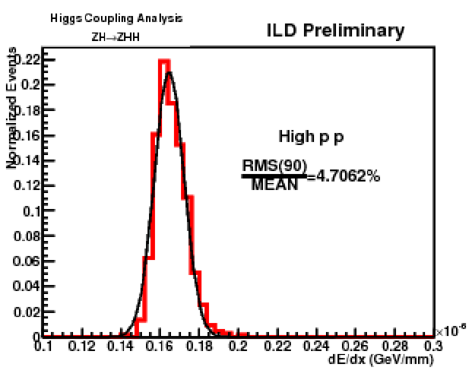












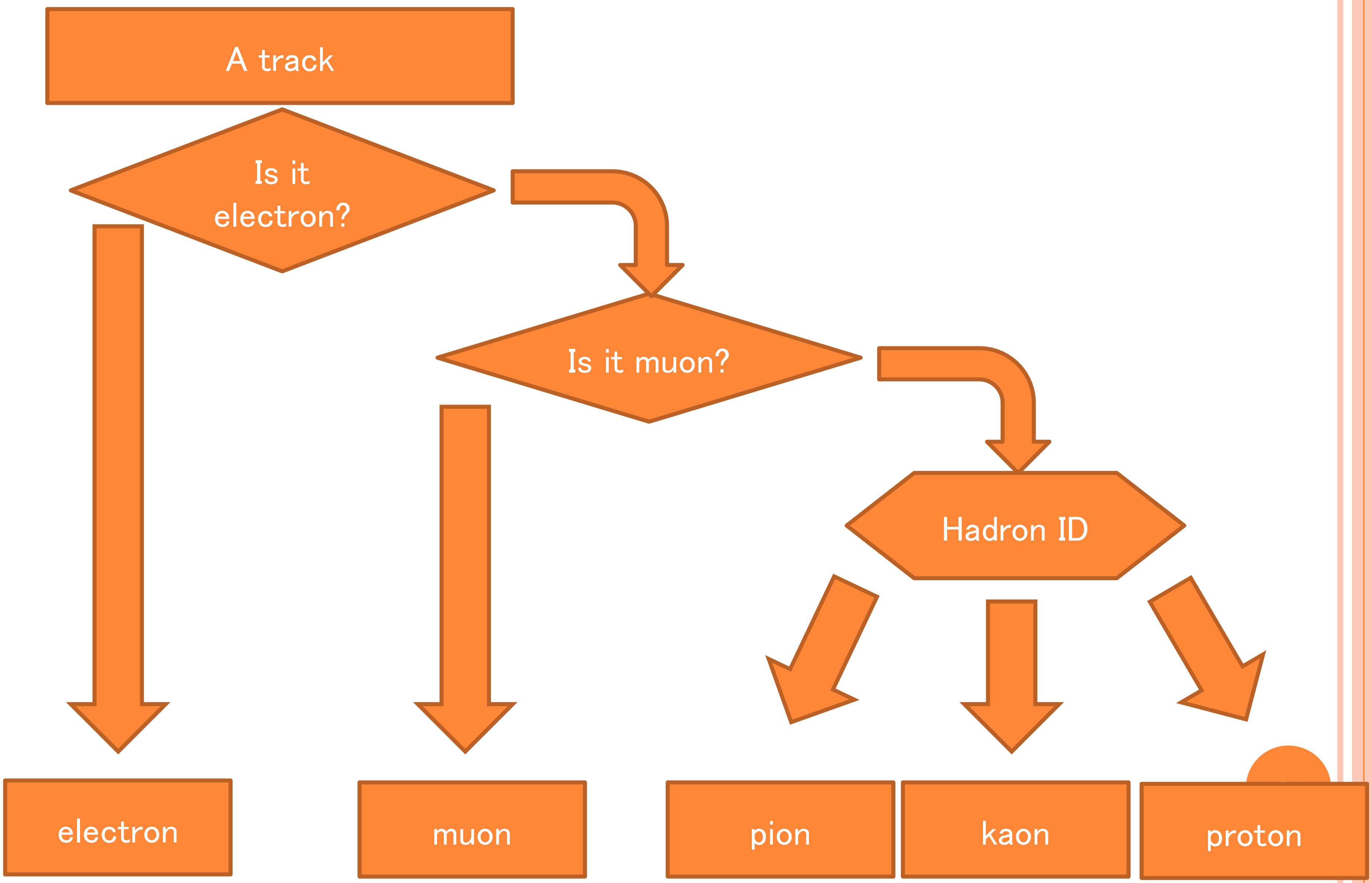
High: p >20GeV/c

Low: π 0.3GeV/c<p<0.6GeV/c

K 1.0GeV/c <p<3.0GeV/c

p 2.0GeV/c<p<4.0GeV/c

Application – Particle ID strategy



Basic idea

- Lepton ID using likelihood is introduced:
  - Lepton selection imposing just one cut
- Target is to find the leptons from W boson as Higgs daughter
  - In some case, lepton energy is so small
  - Form general lepton ID to make the analysis easier
    - Want to apply it to Z lepton finding too

- Likelihood definition:
  - Isolated lepton likeliness

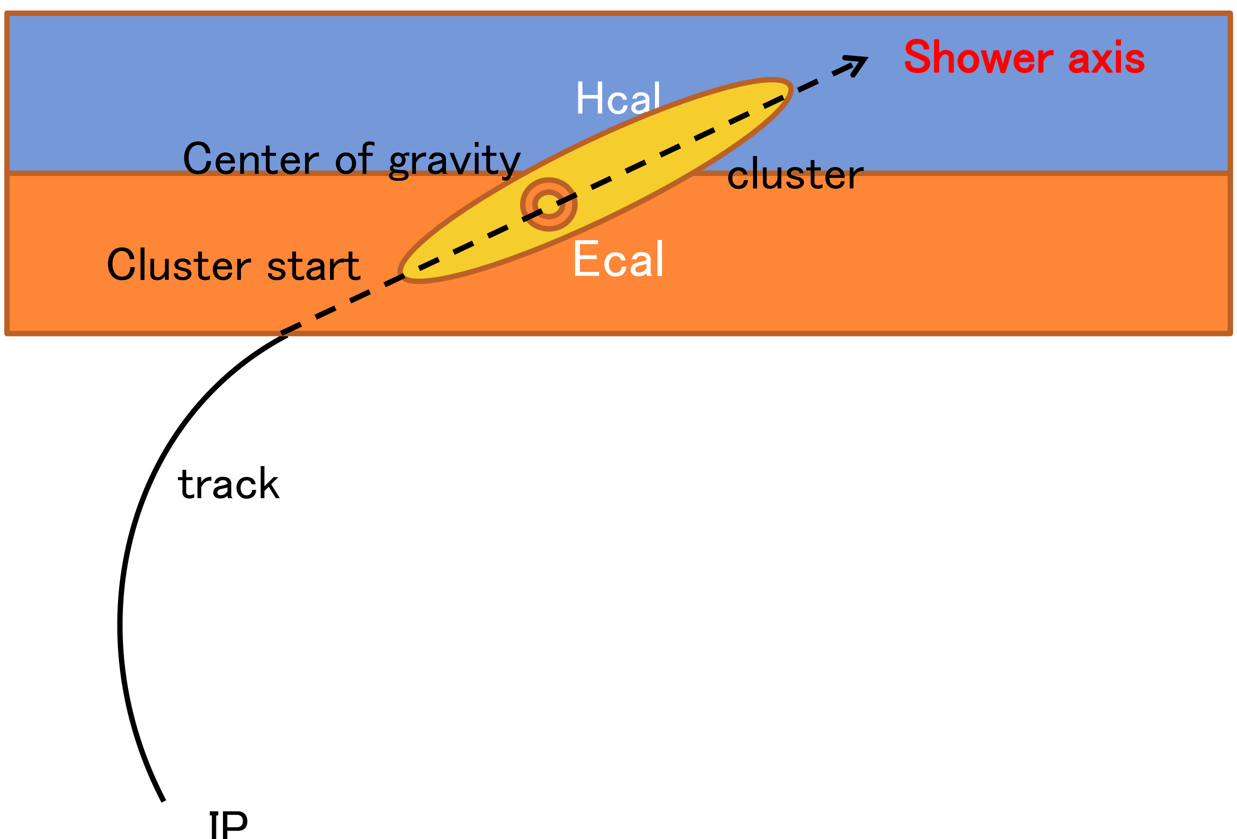
$$L = \frac{\prod s}{\prod s + \prod b},$$

s:pdfs of signal variables

b:pdfs of background variables

Definition of the shower axis

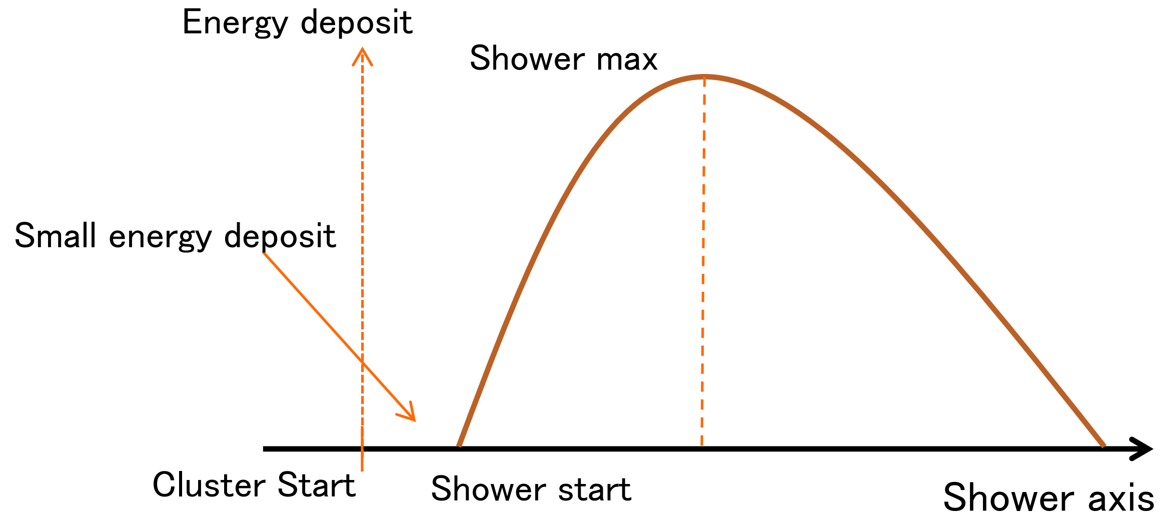
- Shower axis is the direction of the track intruding into calorimeter
  - This correction will change the shower start distribution from last talk
- All the hit points(x,y,z) are converted to longitudinal and transverse components along to the shower axis



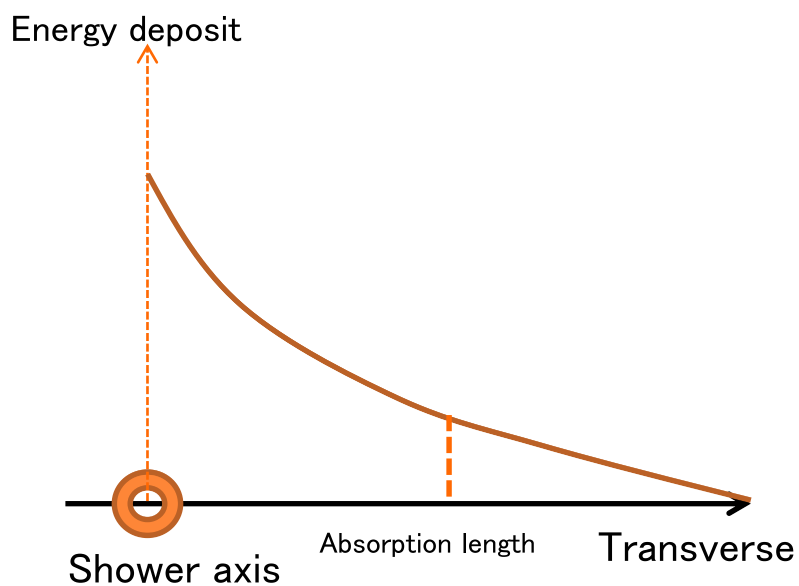
SHOWER PROFILE –STRUCTURE IN THE CLUSTER

Hit points in the cal. are converted from (x,y,z) to (xl, xt)

longitudinal



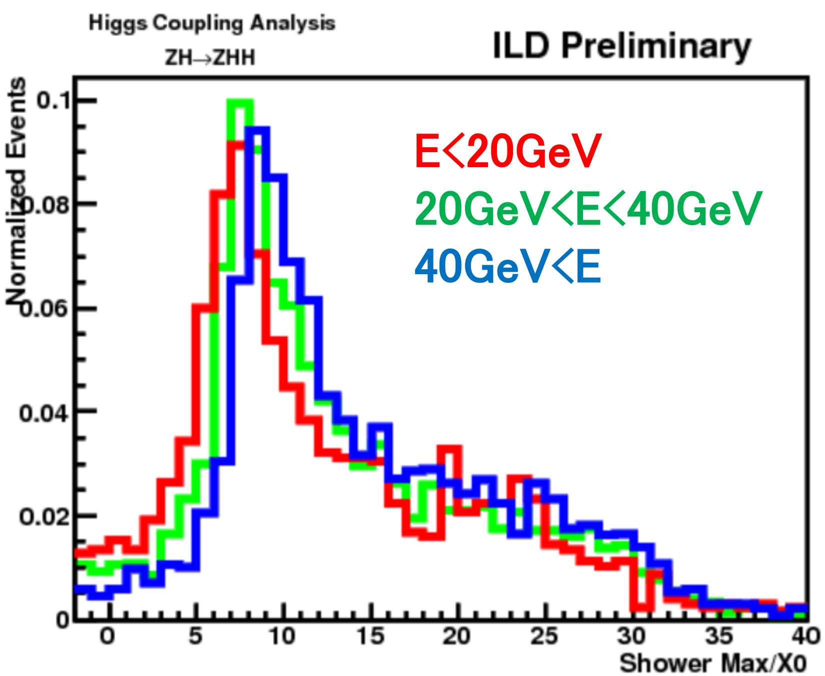
transverse

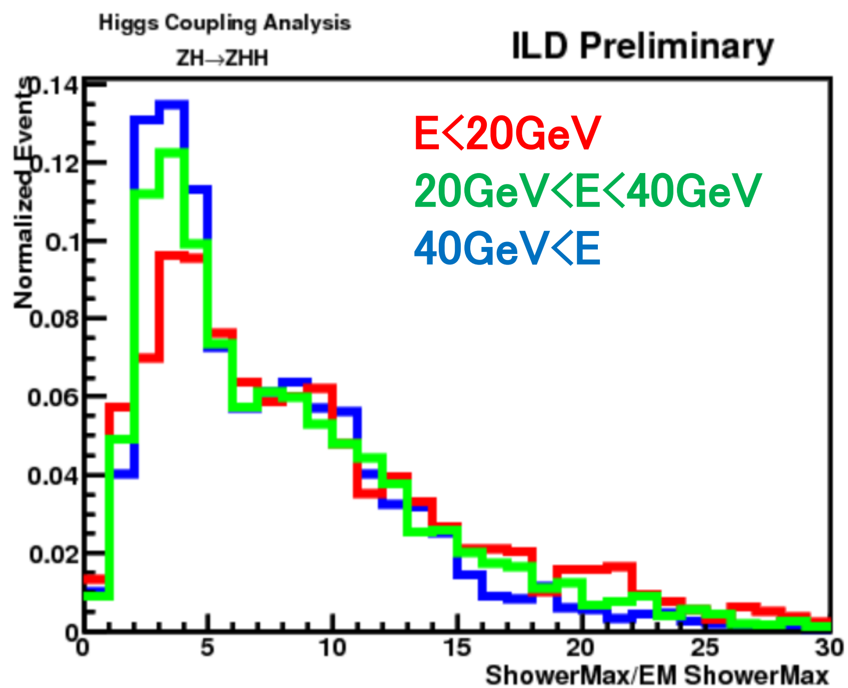


Longitudinal information

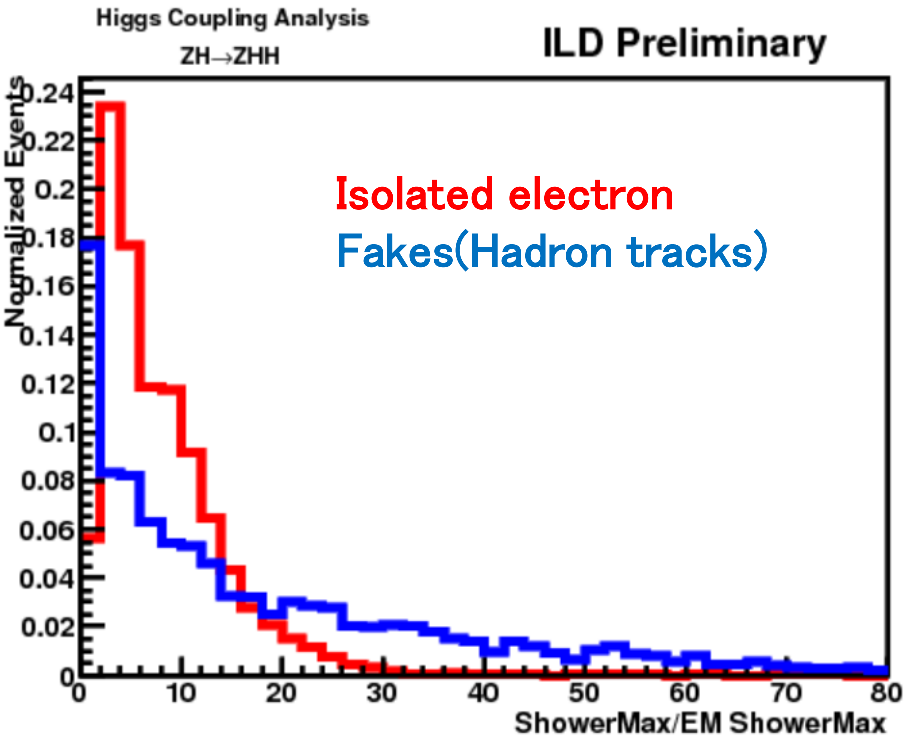
- Length from calorimeter surface to the point which has maximum energy deposit
  - Of course, there is an energy dependence
  - But, the dependence is logarithmic
  - Taking ratio with Expected shower Max

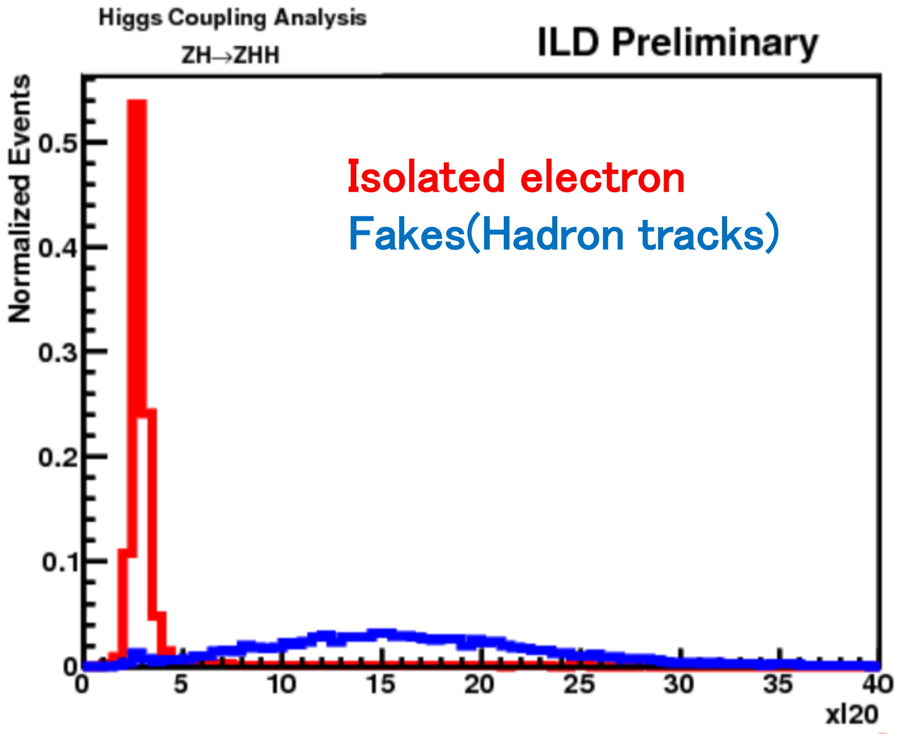
$$\text{Exp. Shower Max} = 1.0\left(\log\frac{E_0}{E_c} - 0.5\right), \quad E_c = 0.021\frac{X_o}{Rm}\text{GeV}$$





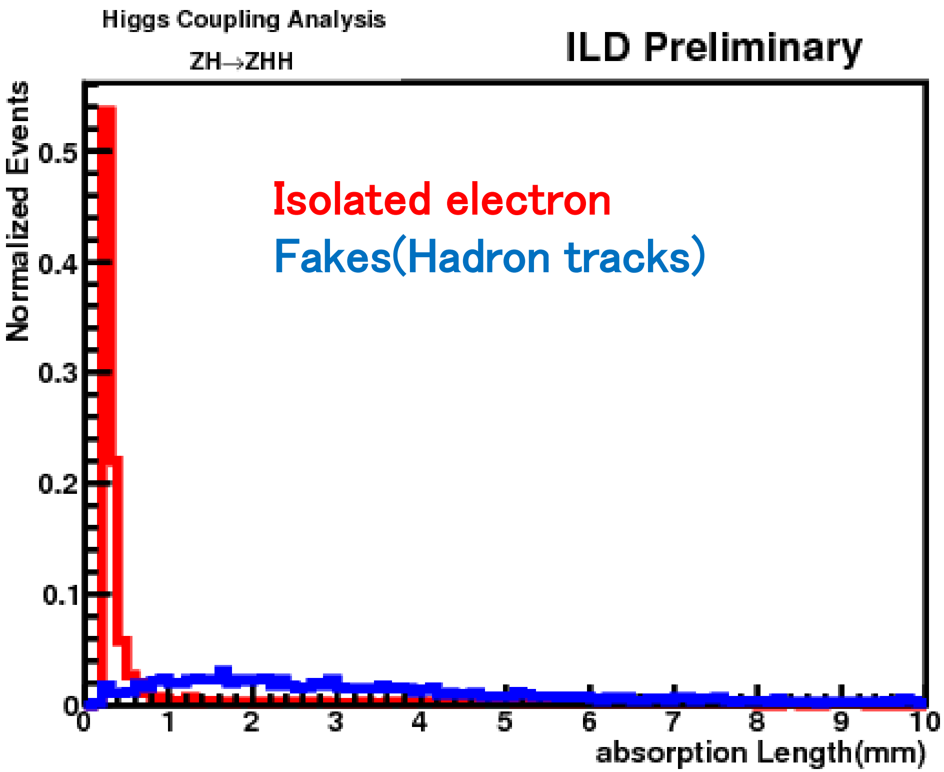
Difference between electron/Fakes

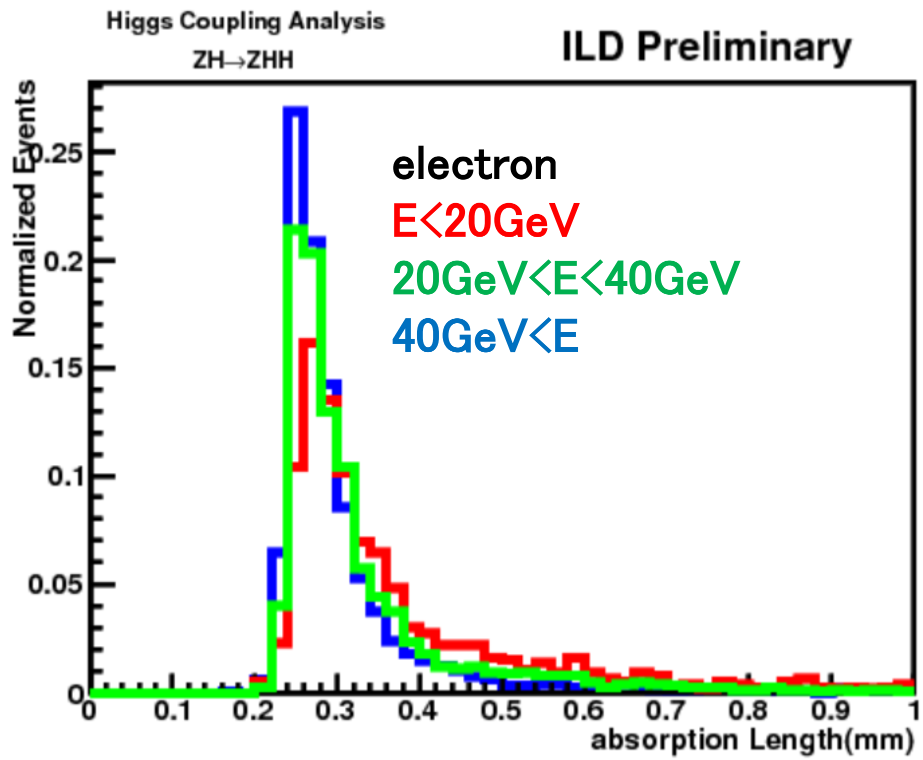




Transverse information

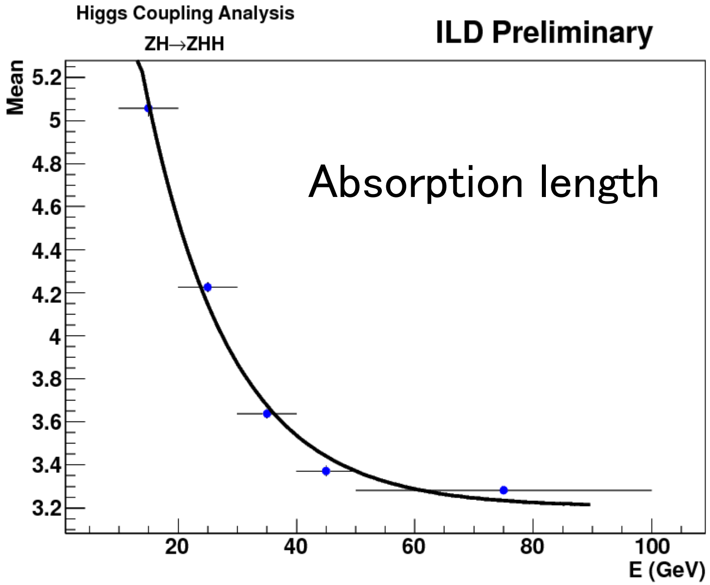
- Transverse shower profile is characterized by absorption length
  - EM shower spread is very small – 90% energy within Moliere Radius
  - Hadron shower spread is wide
  - There is an energy dependence of course, but the effect is small in the case of electron

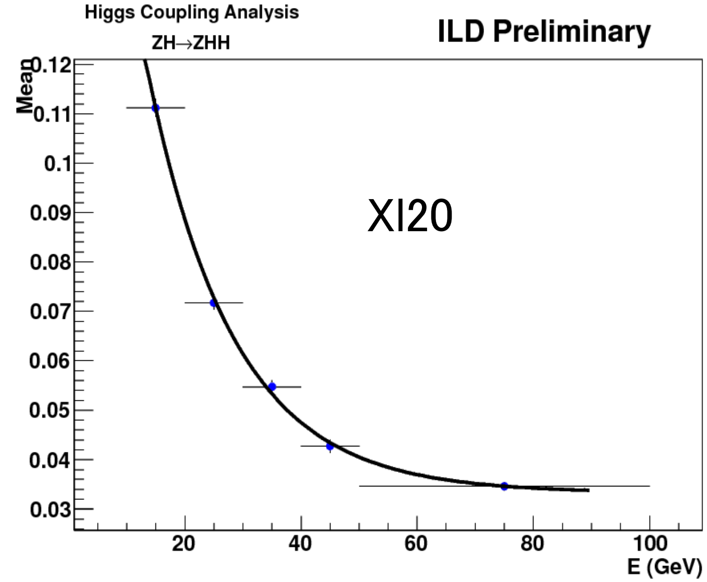




Correction

- Mean is corrected to reduce the momentum dependence





- After correction: much better. Dependence reduced