News on LCFIPlus

Masakazu Kurata, Tomohiko Tanabe
The University of Tokyo
Taikan Suehara
Kyushu University
Jan Strube
PNNL

LCTop16 workshop, 07/08/2016

Publication of DBD LCFIPlus

Nuclear Instruments and Methods in Physics Research A 808 (2016) 109–116

Contents lists available at ScienceDirect

Nuclear Instruments and Methods in Physics Research A

journal homepage: www.elsevier.com/locate/nima

LCFIPlus: A framework for jet analysis in linear collider studies

Taikan Suehara [a,*], Tomohiko Tanabe [b,**]

[a] Department of Physics, Faculty of Science, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan
[b] ICEPP, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

ARTICLE INFO

Article history:
Received 29 June 2015
Received in revised form
11 November 2015
Accepted 11 November 2015
Available online 29 November 2015

Keywords:
Linear collider
Flavor identification
Vertex finding
Jet finding

ABSTRACT

We report on the progress in flavor identification tools developed for a future e^+e^- linear collider such as the International Linear Collider (ILC) and Compact Linear Collider (CLIC). Building on the work carried out by the LCFIVertex collaboration, we employ new strategies in vertex finding and jet finding, and introduce new discriminating variables for jet flavor identification. We present the performance of the new algorithms in the conditions simulated using a detector concept designed for the ILC. The algorithms have been successfully used in ILC physics simulation studies, such as those presented in the ILC Technical Design Report.

1. Introduction

In high-energy collider experiments, the identification of the jet flavor plays an important r
nant decay signatures of t
bottom (b) jets. Many new
supersymmetric theories, ha
Higgs boson $H \to b\bar{b}$, $c\bar{c}$, gg as well as flavor-changing neutral currents in the top sector such as $t \to cZ$ and $t \to cH$. Since charm hadrons have smaller masses and shorter lifetimes compared to

Please refer this: NIM A 808 (2016) 109-116

Flow chart for Flavor tagging in LCFIPlus

Primary Vertex Finder
- Good track selection for vertexing
- Find Primary vertex

Secondary Vertex Finder
- Find secondary vertex candidates

JetFinder(Jet Clustering)
- Cluster PFOs into Jets

Jet Vertex Refiner
- Integrate vertices in jets into upto 2

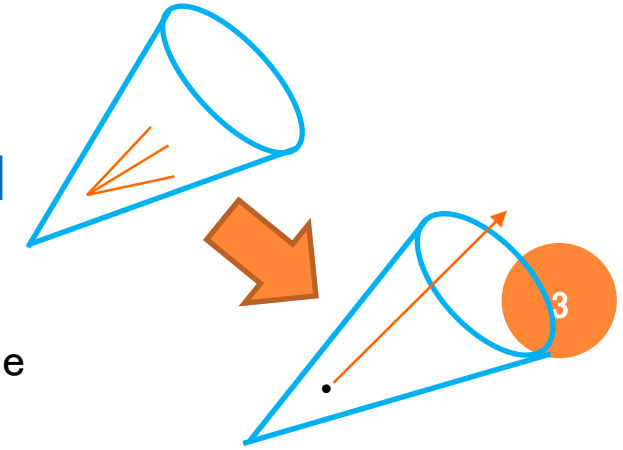
→corresponds to B/D vertices

Flavor Tagging
- Jet flavor is evaluated

→MVA is used for flavor separation

- Vertex finding strategy of LCFIPlus:
  - **Vertex finding first, Jet finding second**
  - All the vertex candidates in a event are checked
  - Jet finding with replaced particles

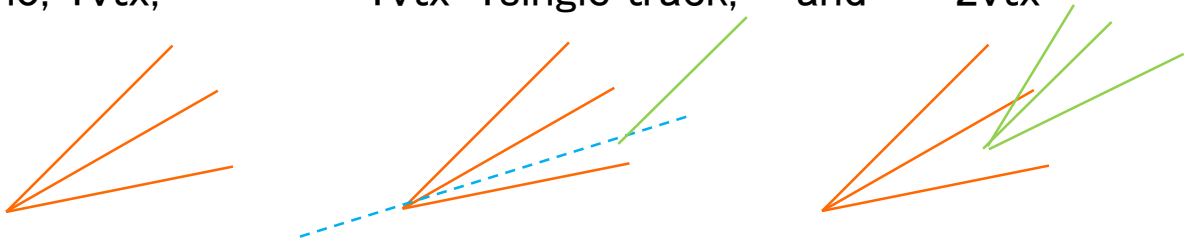
→ Reconstructed vertices are regarded as one particle



FLAVOR TAGGER

- Use Multivariate Analysis(MVA)
  - Classifier: Boosted Decision Tree(BDT)
  - For Flavor separation of b/c/uds jets
  - Categorize using vertex condition in a jet and train independently

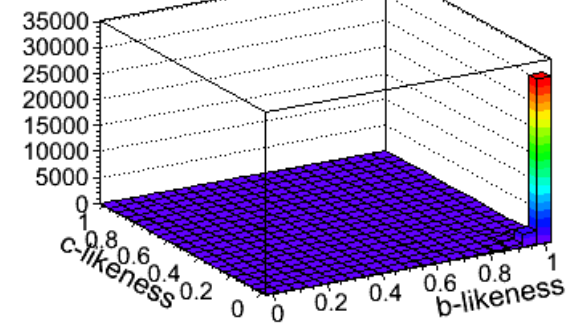
  →This categorization is coming from JetVertexRefiner
    - Vtx: no, 1vtx, 1vtx+1single track, and 2vtx



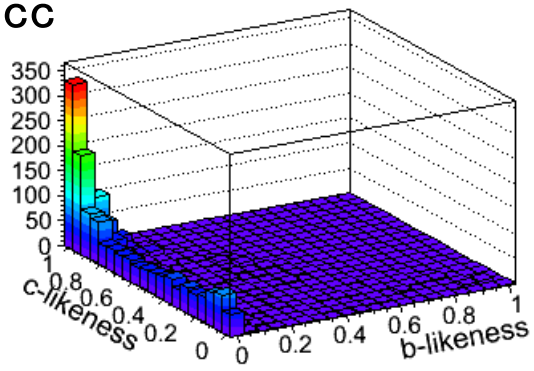
  - 2 output type: b-likeliness and c-likeliness

- Example: distribution of each Higgs decay mode:

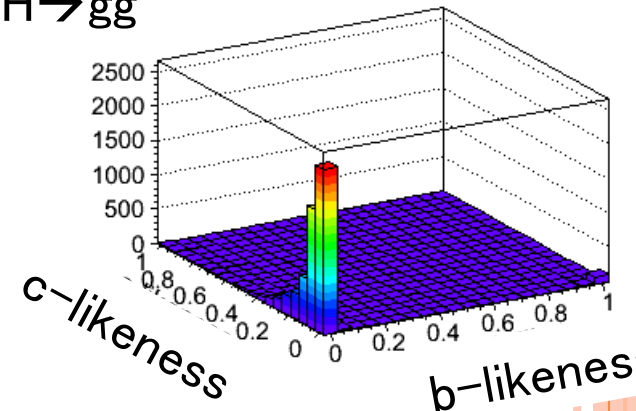
H→bb



H→cc



H→gg



Topics developed since DBD

- Jet clustering algorithms
  - Durham
  - Anti-kT
  - Valencia
  - Beam jet rejection included
- Joint Probability distribution using histograms
- Extract all the input variables available for flavor tagger construction

Jet clustering Algorithms

- Durham
  - the only implementation at DBD ver. LCFIPlus
- Introduce beam induced jets rejection

- $y_{ij} = \frac{\min(E_i^2, E_j^2)(1-\cos\theta)}{E_{vis}^2}$, $y_{beam} = \frac{2E_i^2\alpha^2(1-\cos\theta)}{E_{vis}^2}$

  α : beam rejection parameter

  smaller⟶ beam rejection becomes stronger
  - Particle i with $y_{ij} > y_{beam}$ is discarded

Jet Clustering Algorithms

- kT algorithm
  - Mainly used for beam pileup rejection in CLIC

- $y_{ij} = \frac{\min(Pt_i^2, Pt_j^2)\Delta R^2}{R^2_{parameter}}$, $y_{beam} = Pt_i^2$

  $\Delta R^2 = \Delta\varphi^2 + \Delta\eta^2$

  $R_{parameter}$: beam rejection parameter

  smaller⟶ beam rejection becomes stronger
  - Particle i with $y_{ij} > y_{beam}$ is discarded

Jet clustering Algorithms

- Valencia
  - Intermediate of Durham and kT

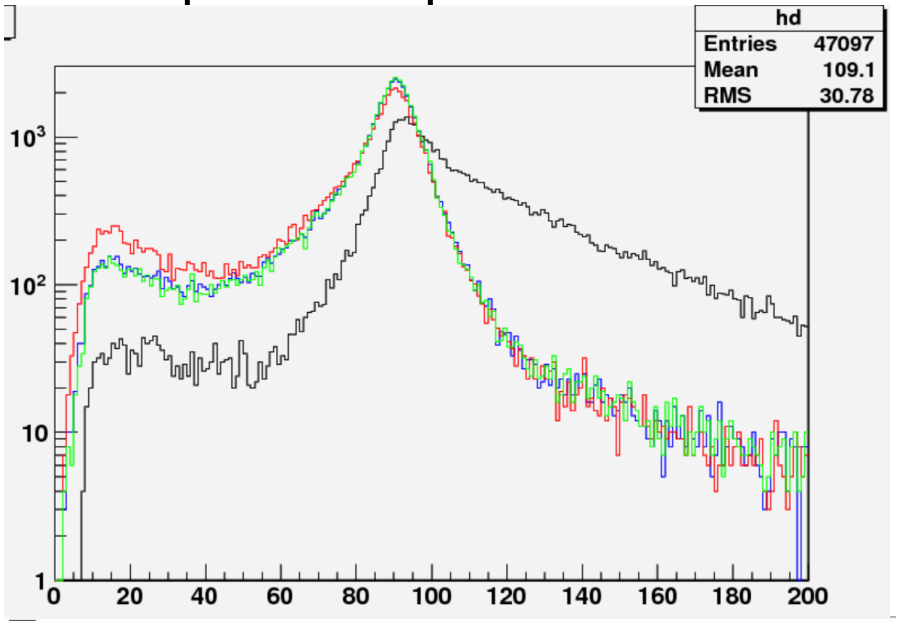
- $y_{ij} = \frac{\min(E_i^{2\beta}, \mathrm{E}_j^{2\beta})(1-\cos\theta)}{R^2_{parameter}}$, $y_{beam} = Pt_i^{2\beta}$

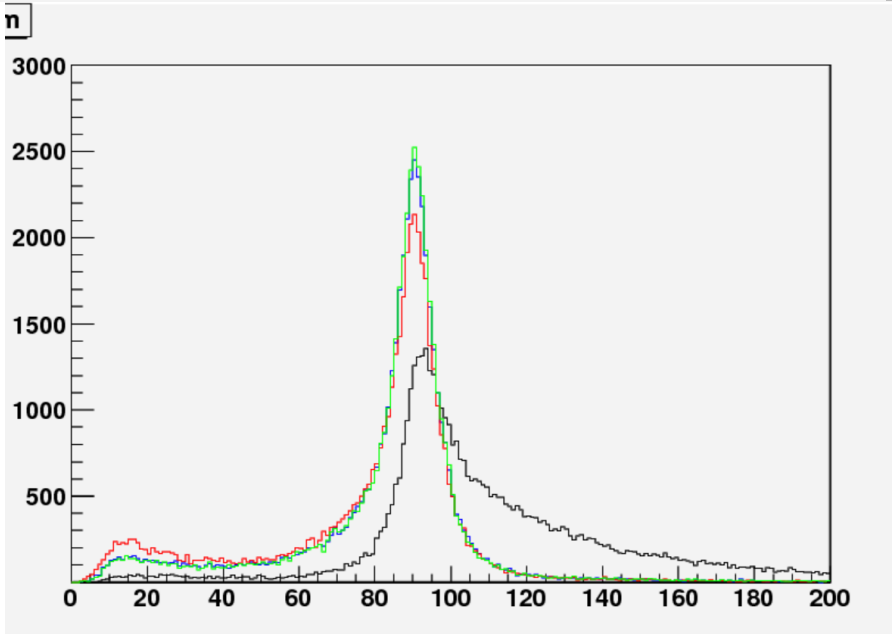
  $R_{parameter}$: beam rejection parameter

  smaller→ beam rejection becomes stronger
  - Particle i with $y_{ij} > y_{beam}$ is discarded

JET CLUSTERING WITH BEAM BACKGROUND REJECTION

- Compare the performance between Durham, Kt and Valencia





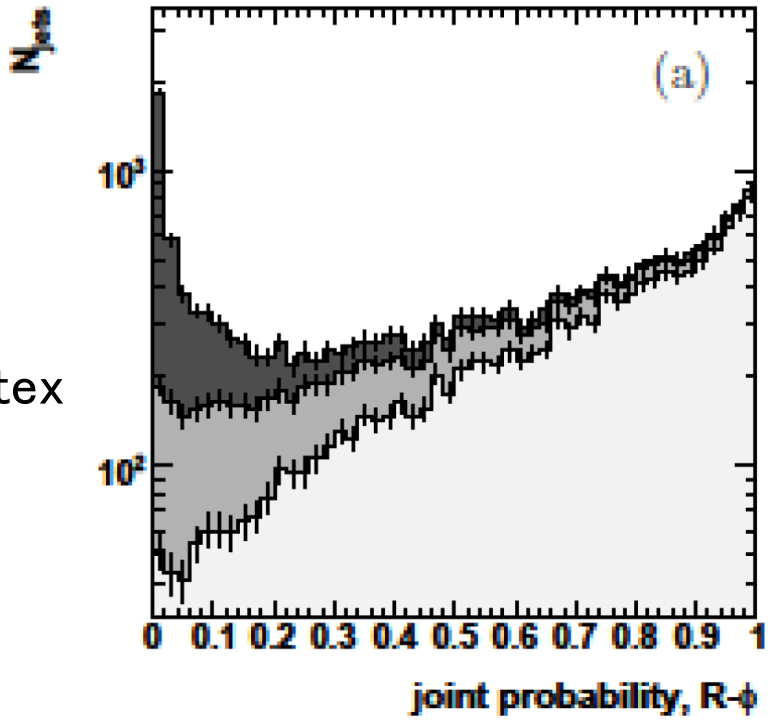
$\nu\ \nu$ Z@500GeV

- 2 jet clustering
- Parameters are tuned for better result

w/o beam b.g. rejection
Kt
Durham
Valencia

JET PROBABILITY

- Joint probability
  - Product of PDF on signed impact parameter(d_0/z_0) significance that the track comes from the primary vertex



- DBD ver. : PDF is formed using hard-coding simple function
  - Optimized within DBD ver. framework
  - Will not be good when vertex configuration is different

→Use histograms

Joint probability histograms

- We added histogram in vertex probability file to be used as Joint probability instead of the hard-coded function
  - Default PDF histograms are already provided
- In addition, we included the macro which can create Joint probability PDF histogram files for any jet situation
- Bug: SEGV with old vertex probability files

  →soon be fixed

LCFIPLUS IMPROVEMENT

- For better flavor tagger, we need
  - Find secondary/tertiary vertices as many as possible
  - Perfect reconstruction of B/D meson mass
    - Perfect attachment of charged particles
    - Recover lost component(especially neutrals)
  - Better variables which can separate jet flavors

- We found that we can acquire flavor tagging improvement:
  - Vertex finding efficiency improvement → introduce a new algorithm
  - Vertex Mass Recovery → using escaping π^0s
  - Better flavor separation for jets of 0 vtx

- Much help is necessary!:
  - Particle ID is one of the key to flavor tagging improvement
    - To classify vertices
  - Pi0 reconstruction(γ pairing) is other key for vertex mass recovery
    - First of all, pi0 is necessary!

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

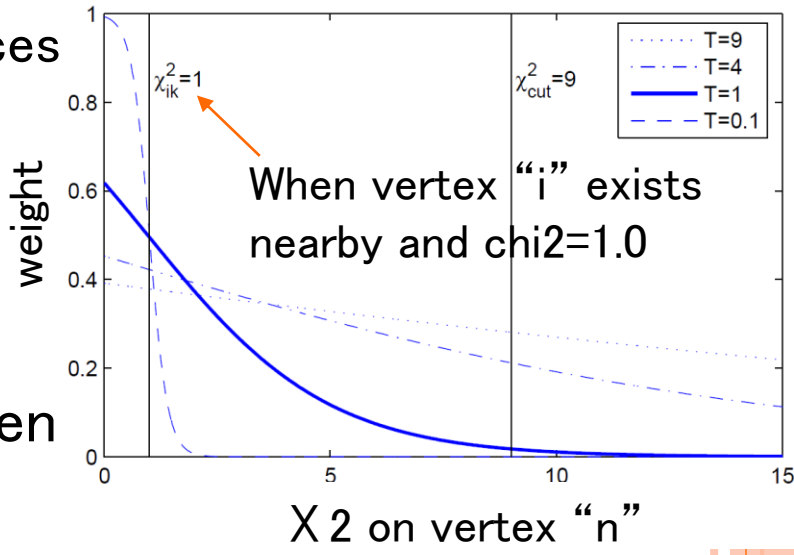
$$w_{nk} = \frac{e^{-\chi^2_{nk}/2T}}{e^{-\chi^2_{\mathrm{cut}}/2T} + \sum_{i=1}^{N} e^{-\chi^2_{ik}/2T}}$$

  - 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 a vertex will degrade

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

→can reject fake tracks well!

Weight of track "k" on vertex "n"



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

→vertex finding eff. will be improved!

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 vertices?
  - Almost same – slightly better!

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 of Adaptive Vertex Fitting for c jets

- Common parameters are set at same values for comparison
- Same event sample(nnH sample@500GeV) 100k events
  - H→cc about 7k events
- 2 jet clustering, jet matching with MCtruth is performed

method	cjet with 2vtx	cjet with 1+1vtx	cjet with 1vtx	total
DBD LCFIPlus	46	153	5987	6186
AVF	64	145	6283	6492

- Total jets with vtx: ～4.8% increased
  - Jets with 2vtx: increased → but too small to say something
  - Jets with 1vtx: ~5% increased → good for uds jet separartion!
- Fake track rate per vtx: how many fake tracks contaminate on vertices?

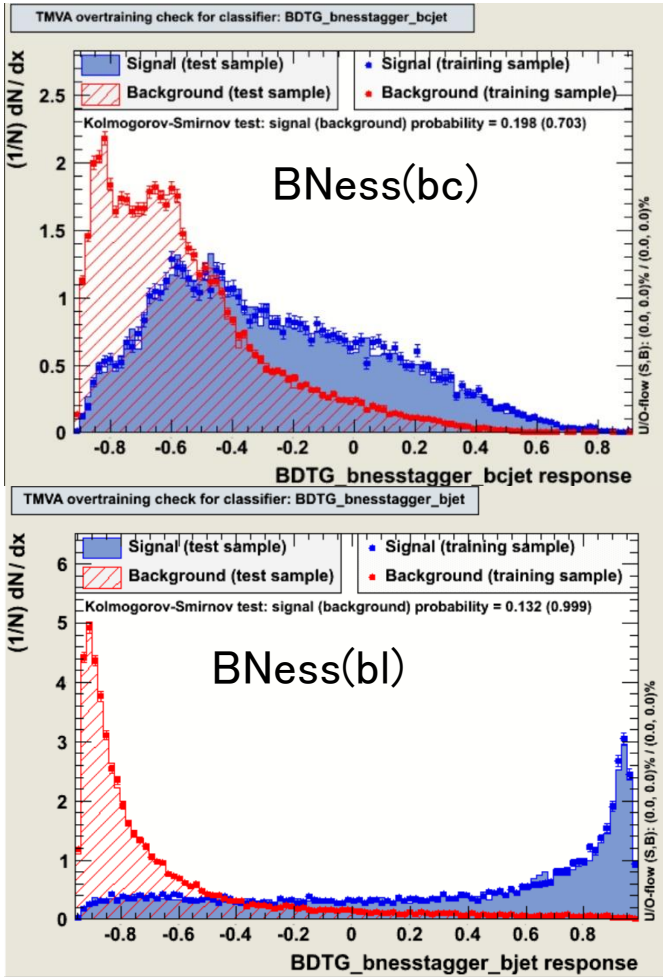
method	bjet with 2vtx	bjet with 1+1vtx	bjet with 1vtx
DBD LCFIPlus	0.00±0.00	0.012±0.006	0.0014±0.004
AVF	0.00±0.00	0.018±0.006	0.0012±0.004

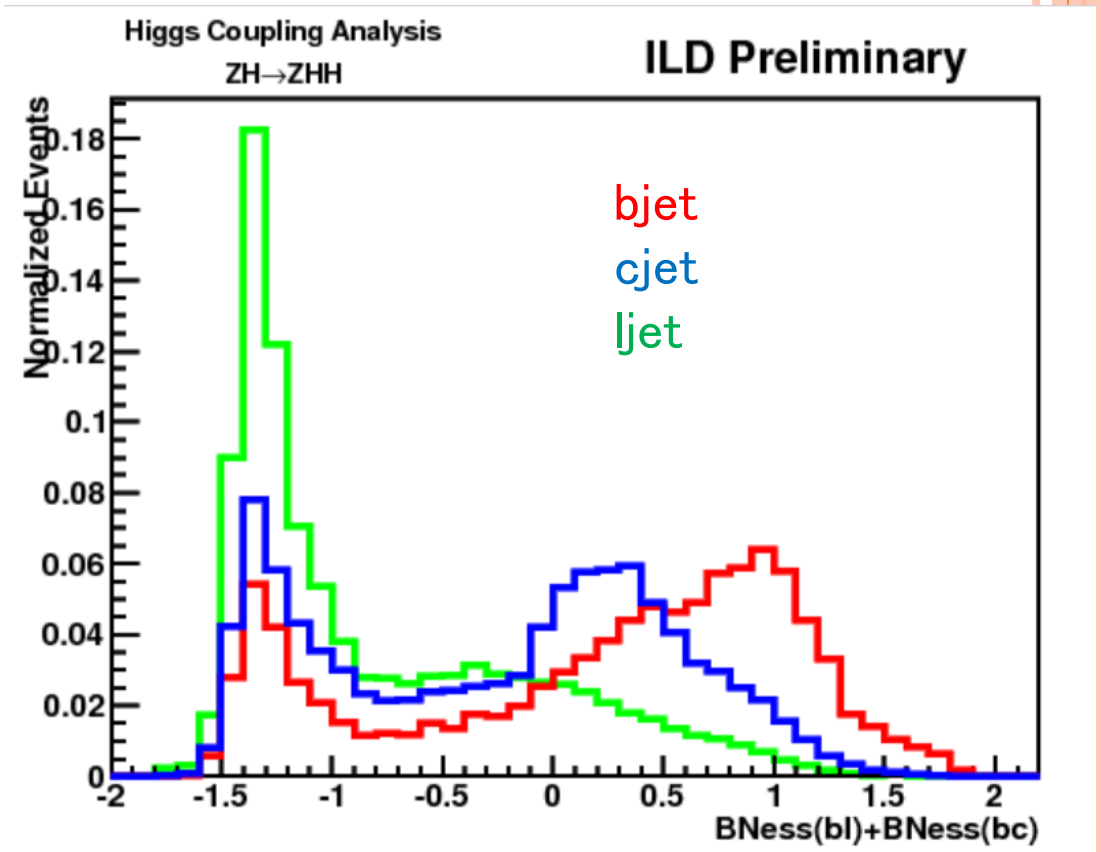
BNESS TAGGER

- Flavor separation of 0vtx jet is most difficult situation
  - Only impact parameter implies the existence of secondary vertices for flavor separation
- BNess tagger will be worth trying in this case!
  - Developed in CDF
  - Focus on individual tracks and evaluate jet flavor only using single track
  - Track's potential for coming from heavy flavor particle(D&B meson and baryons) should be evaluated(using MVA)
- Difficulty in ILC
  - In CDF, it is important to separate b and other flavor → c quark separation is not required
  - In ILC, separation among b, c and other is very important→ bc separation is a key for flavor tagger
- How is bc(& bl) separation using BNess tagger?

BNESS OUTPUT

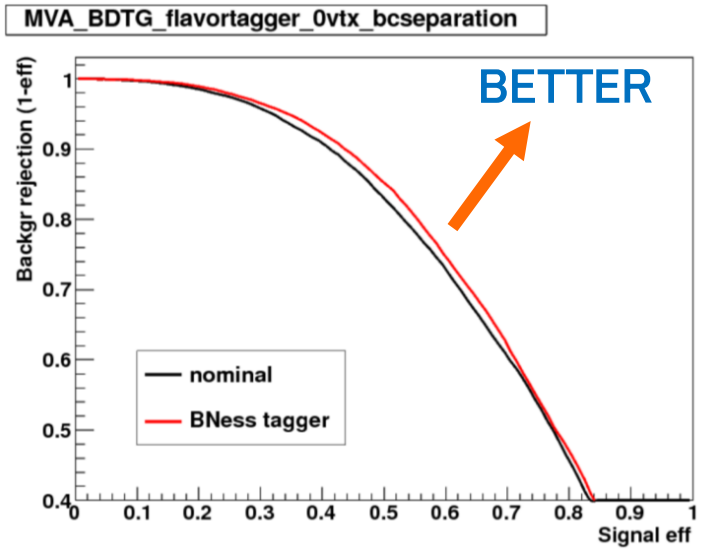
- Collect Highest score BNess track in 0vtx jets
- Final BNess is defined as BNess(bl)+ BNess(bc)
- Well separated between bjets and l jets
- Difference can be seen between bjets and cjets



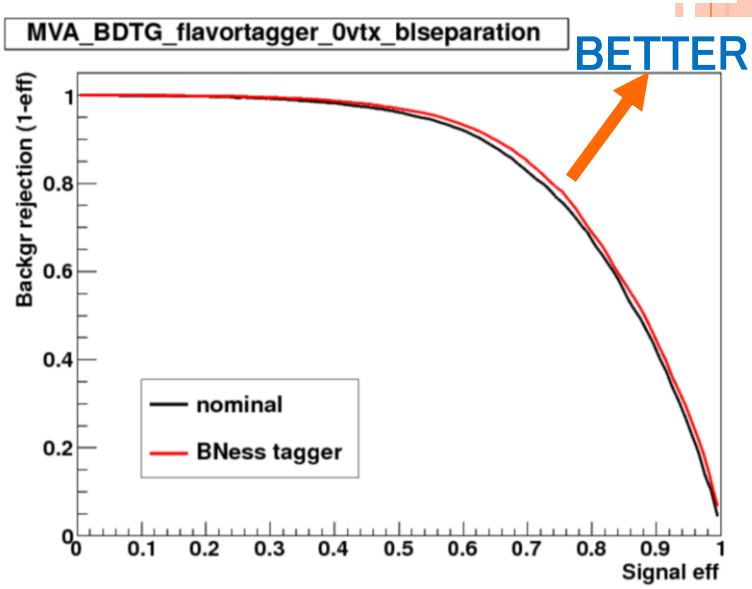


Results of BNess tagger on flavor tagging

- Construct a "toy" flavor tagger
  - Convert nominal input variables to BNesstagger variables
  - Compare with ROC curve

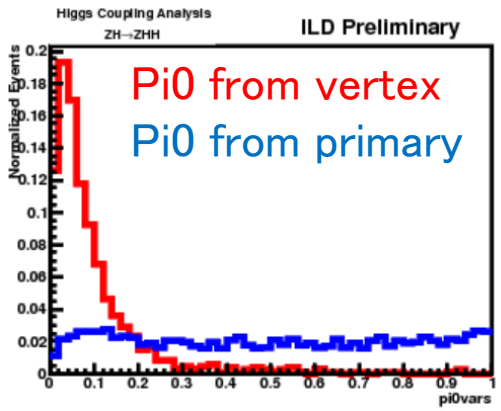


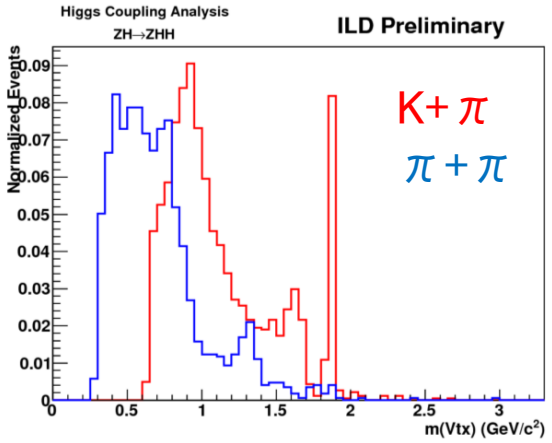
- Both of bc and bl separation cases, some improvement can be obtained
- Need optimization
- Especially, precise study of b–c–l flavor separation is necessary



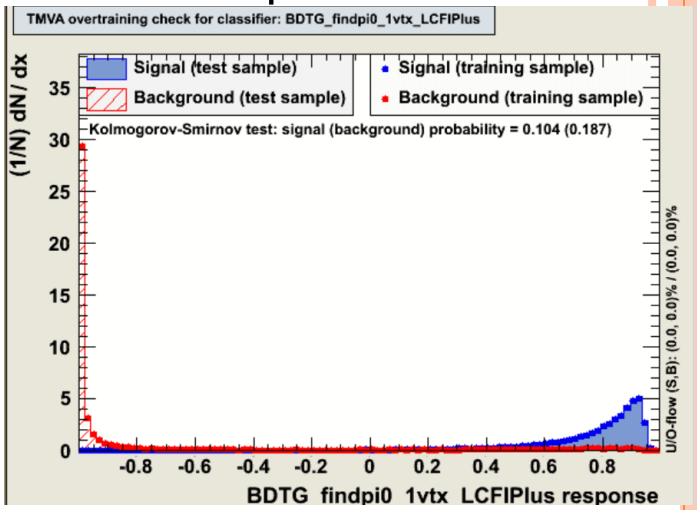
VERTEX MASS RECOVERY

- Using pi0s which escape from vertices
  - Need to choose good pi0 candidates –construct pi0 vertex finder
  - Key issue –pi0 kinematics, very collinear to vertex direction



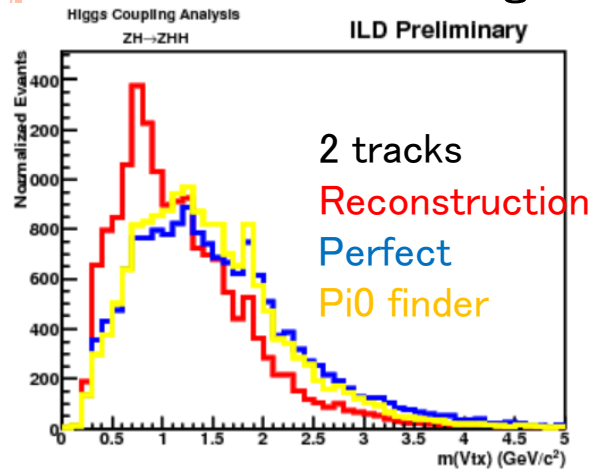


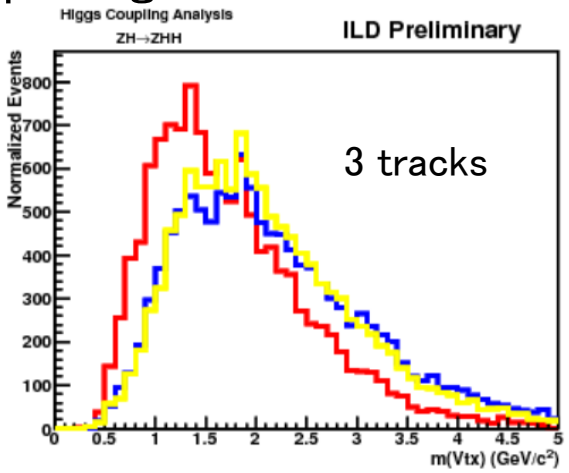
- Particle ID is the other key to classify vertices
  - Different particle patterns have different vertex mass patterns
- Construct Pi0 Vertex finder using MVA
  - Identify which vertex pi0s are coming from

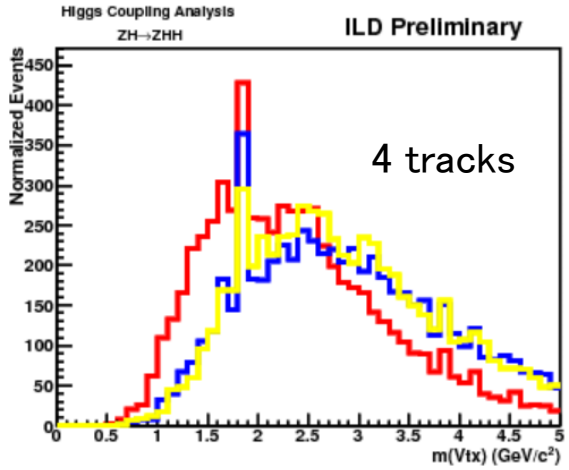


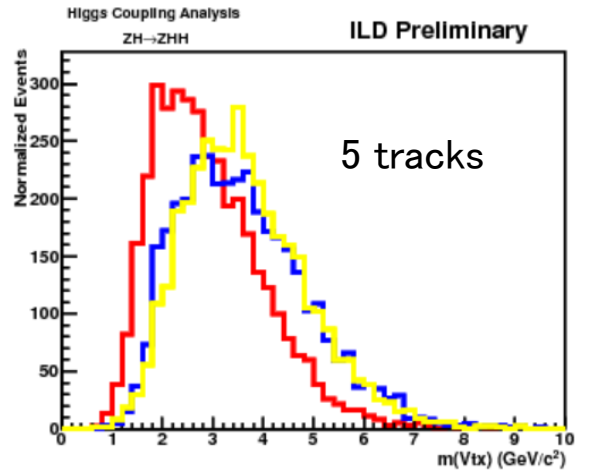
Vtx masses of bjets in double-Higgs process

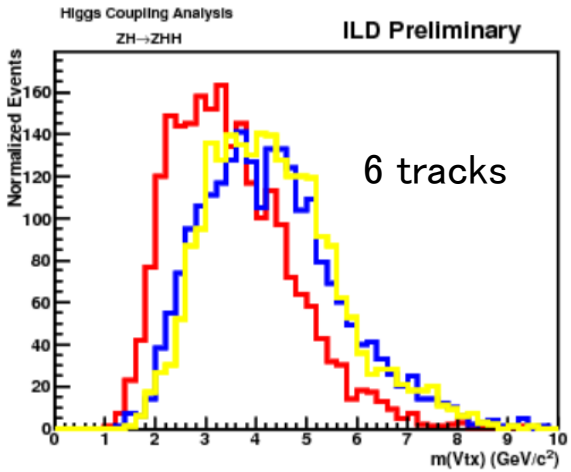
- Vtx mass distributions for each vertex pattern(ntrk)
  - bjets with 1vtx
  - Difference is limited by mis-pairing of gammas(eff. ～50%) and mis-attachment of pi0s
  - Need better gamma pairing!

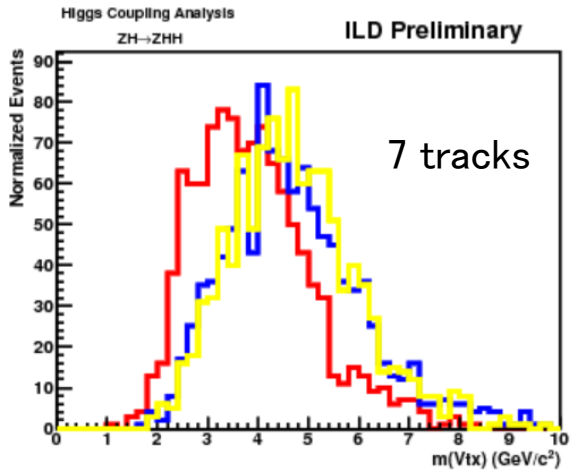




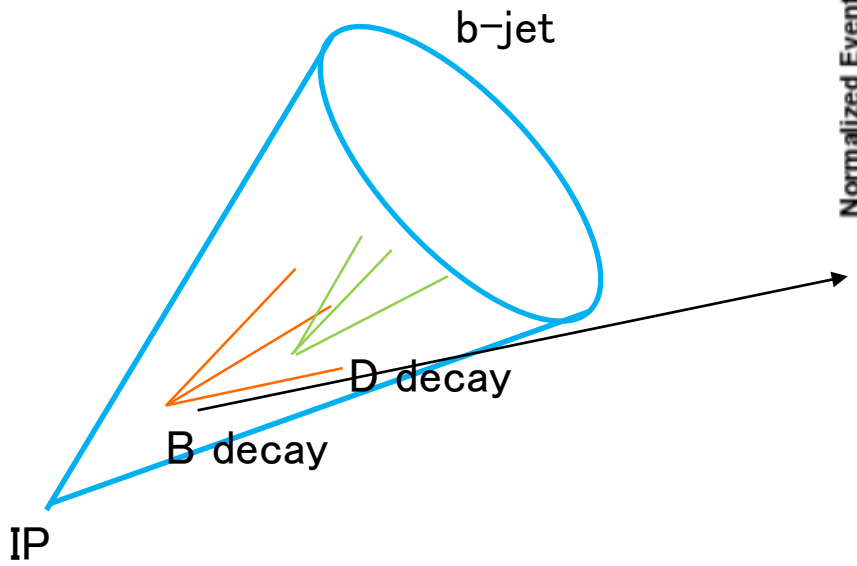


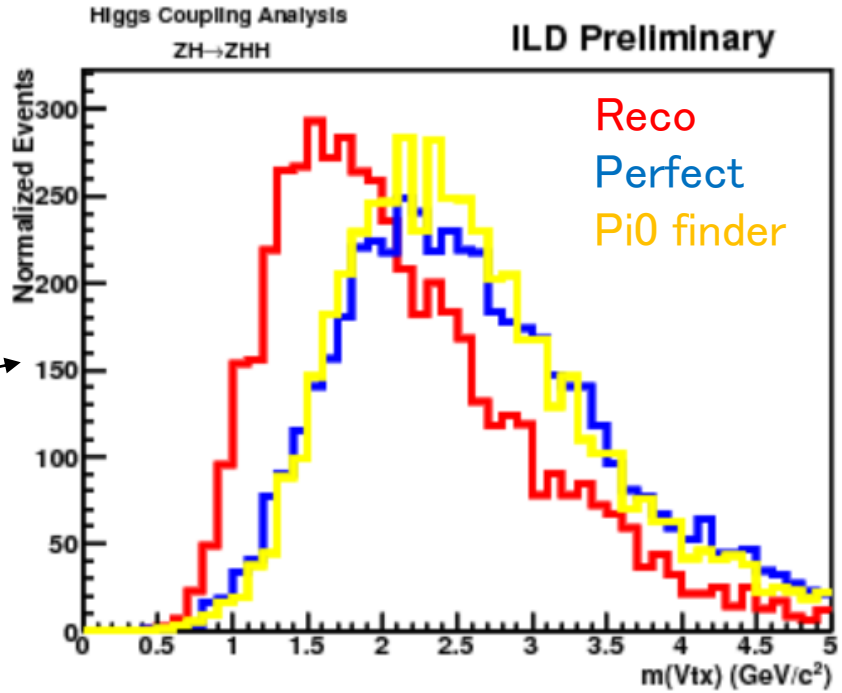




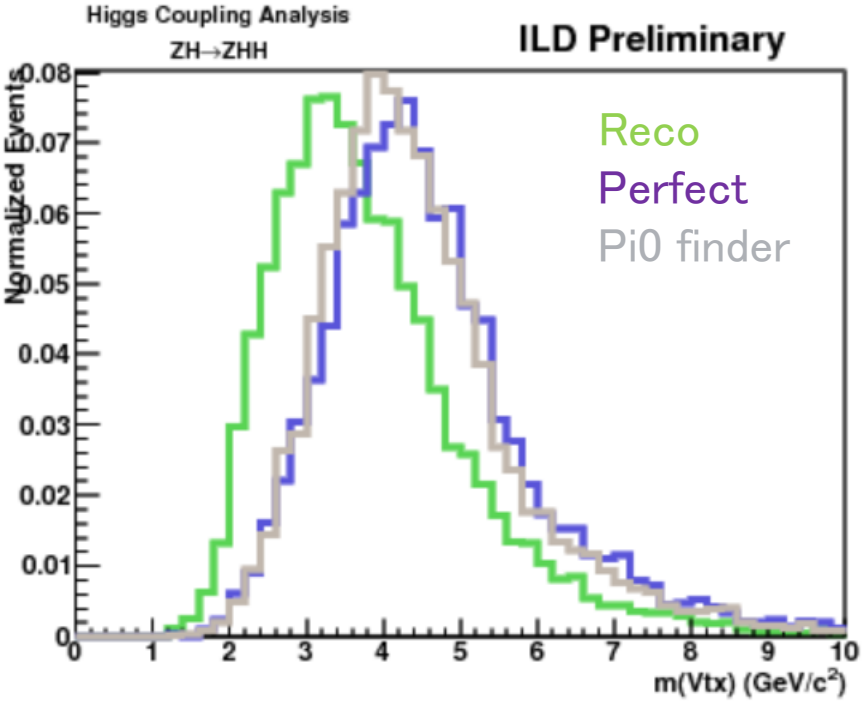


GLANCE AT OTHER CASE



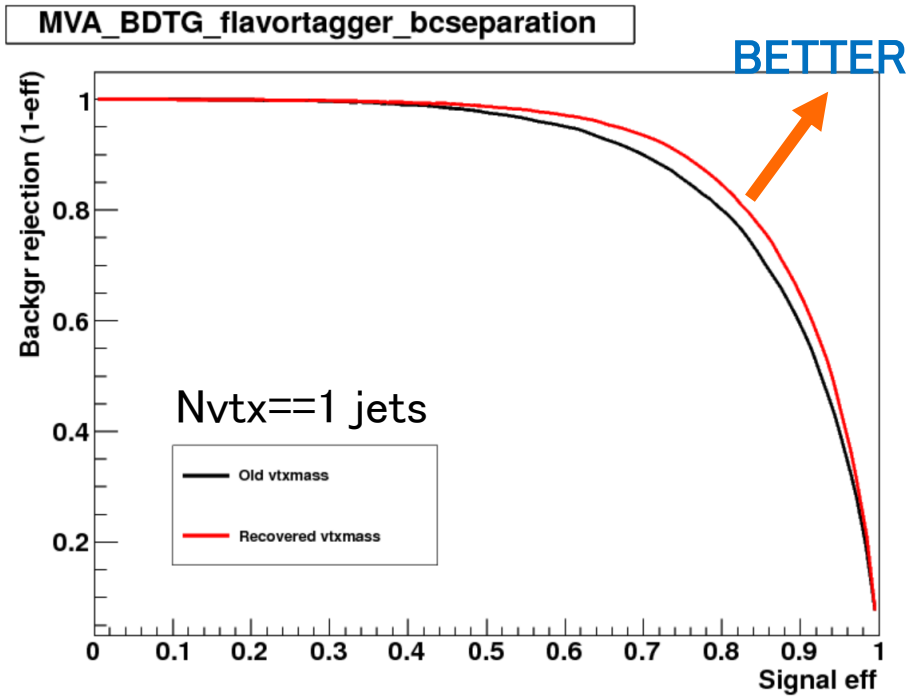


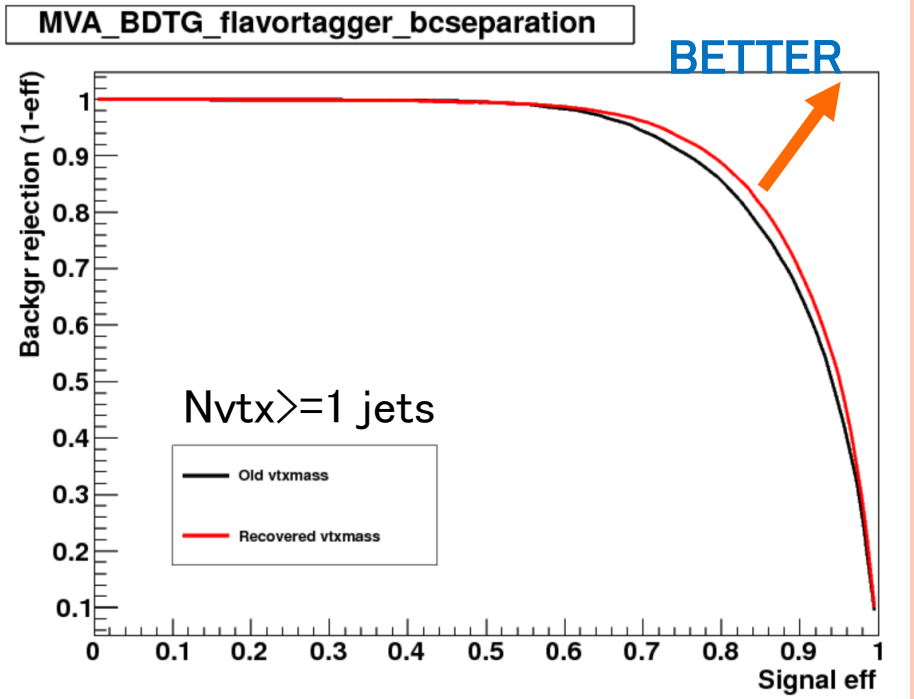
- 2 vertices in bjet
  - Secondary vertex – 4tracks case
  - Merging with tertiary vertex
    - Tertiary vertices allow all the patterns
    - Attach pi0s to both of the vertices using pi0 vertex finder



Vertex mass recovery effect on flavor tagging

- Construct a “toy” flavor tagger
  - Input variables are obtained from LCFIPlus
  - Input variable selection is too primitive!
  - Only vertex mass is replaced to recovered vertex mass
  - Compare with ROC curve





- Vertex is created using DBD LCFIPlus vertex finding
  - need to check AVF case

Summary and prospects

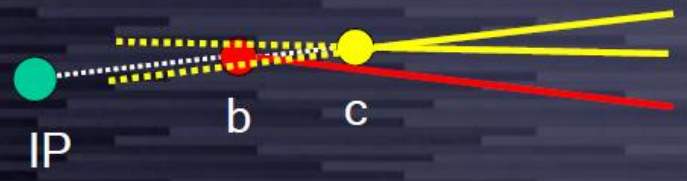
- Valencia jet clustering has been included
- Joint probability can be estimated using PDF histograms instead of hard-coding functions
- For flavor tagging improvement:
  - New vertexing algorithm(AVF) will provide better vertex finding efficiency
  - BNesstagger will give some improvement for 0vtx jet flavor separation
  - There seems hope for attaching pi0s to vertices to recover vertex mass
- So far, AVF will provide ～1.4% improvement of vertex finding in bjets
  - Jets with 2vtx well increased – better for b jet ID!
  - Fake rejection will be same – slightly better!
  - This study will lead to vertex charge assignment improvement
- Vertex mass recovery is reasonable
  - Will provide better flavor tagger using recovered vertex mass
  - Pi0 reco. Improvement will give better vertex mass recovery!

- Finally, incorporate all the ideas and check the final flavor tagging effs.in LCFIPlus!

Back ups

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

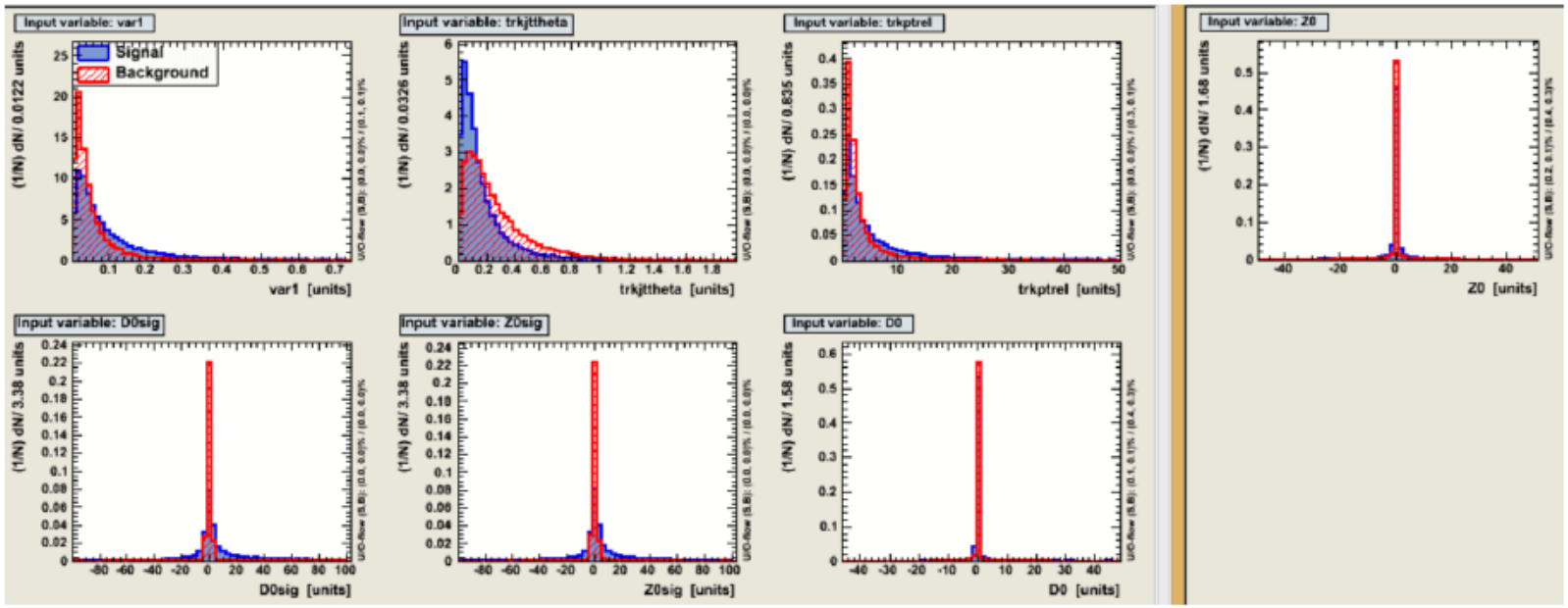
- Treatment of multiple jets inside (Jet substructure ID etc.)
- Treatment of vertices
  - "Concrete" or "Doubtful" vertices
    - based on probability, # of tracks etc.

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: $sig = \sqrt{(\frac{d_0}{\sigma})^2+(\frac{z_0}{\sigma})^2}$

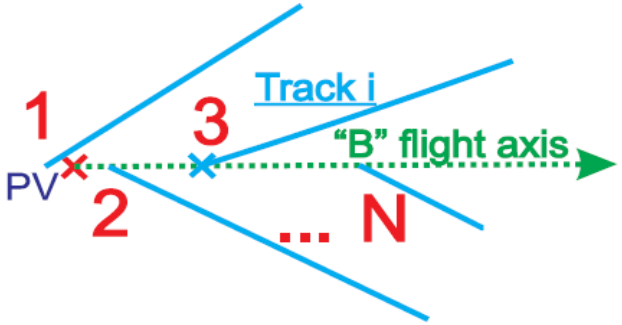


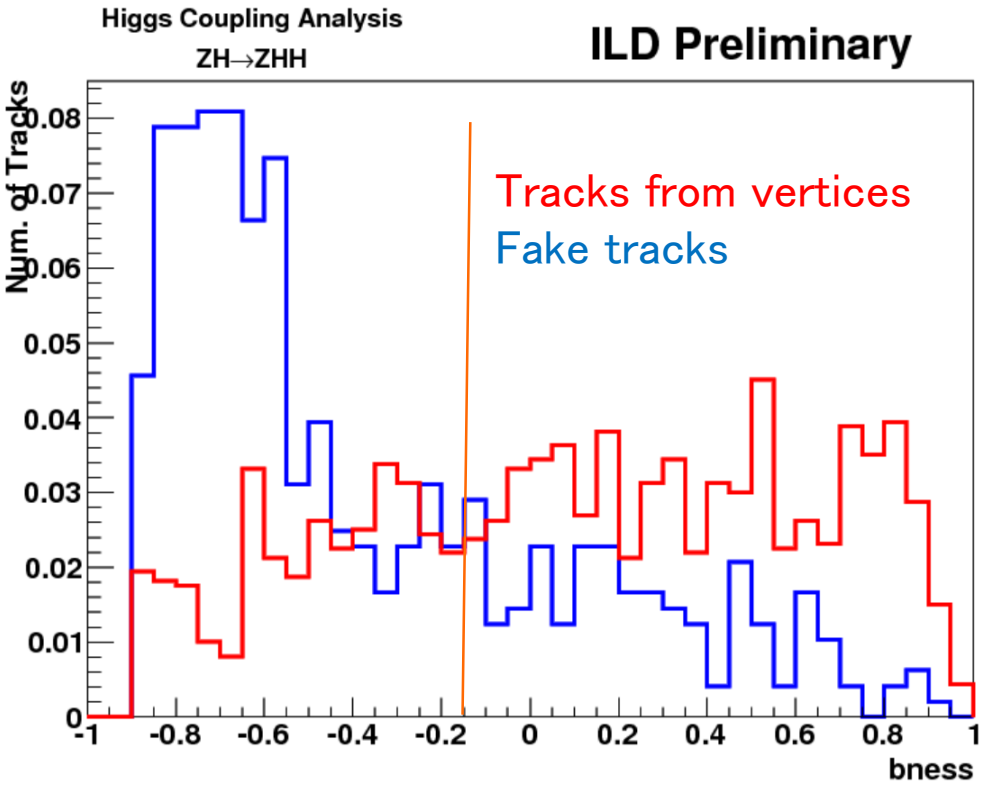
BNESS TAGGER FOR FAKE TRACK REJECTION

- 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?

Example: looking for single track vertex



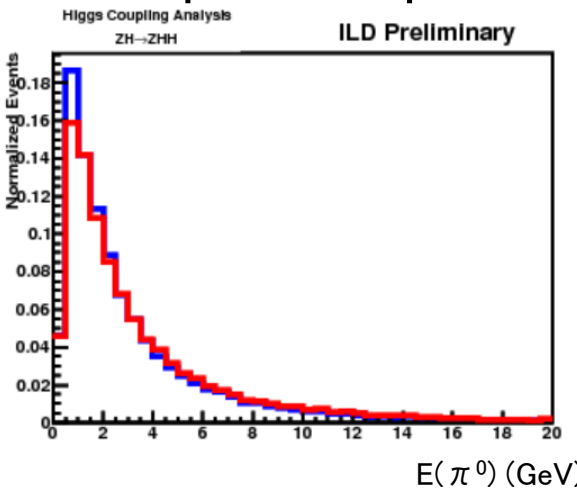


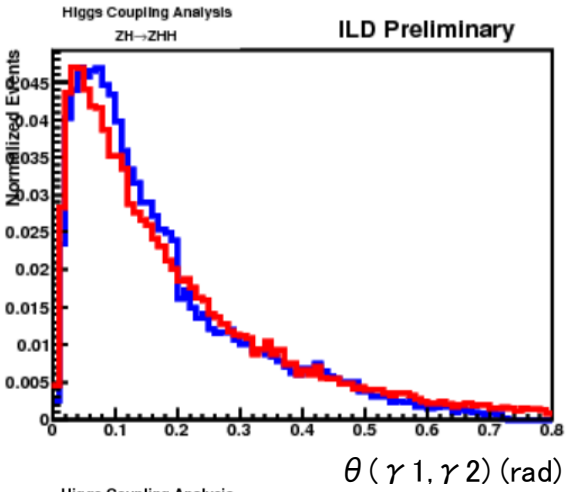
Pi0 reco using Naïve Bayes for vertex mass recovery

- Good pairing eff. & mis-pairing eff.

	Correct pair	Wrong pair
eff. (%)	46.0±0.3	54.0±0.4

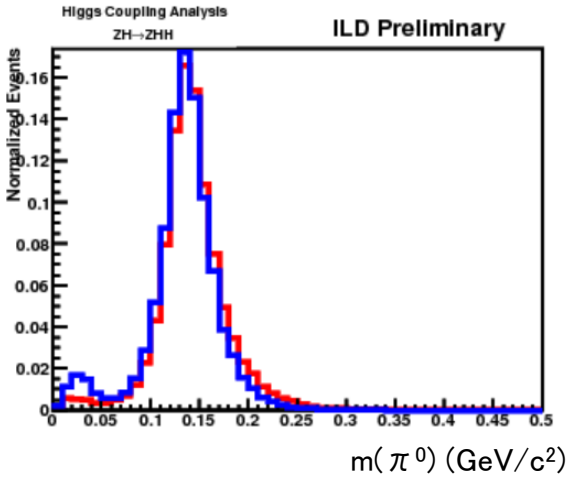
- Kin. plots of pi0 reco. results

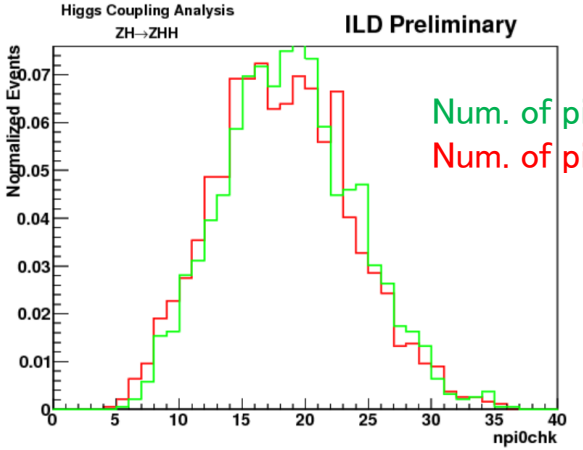




MC truth

Pi0 finder





Num. of pi0s to be reconstructed

Num. of pi0s from pi0 finder

- Integrate pi0 reconstruction?