

Status of LCFIVertex

Tomohiko Tanabe, Taikan Suehara

The Univ. of Tokyo

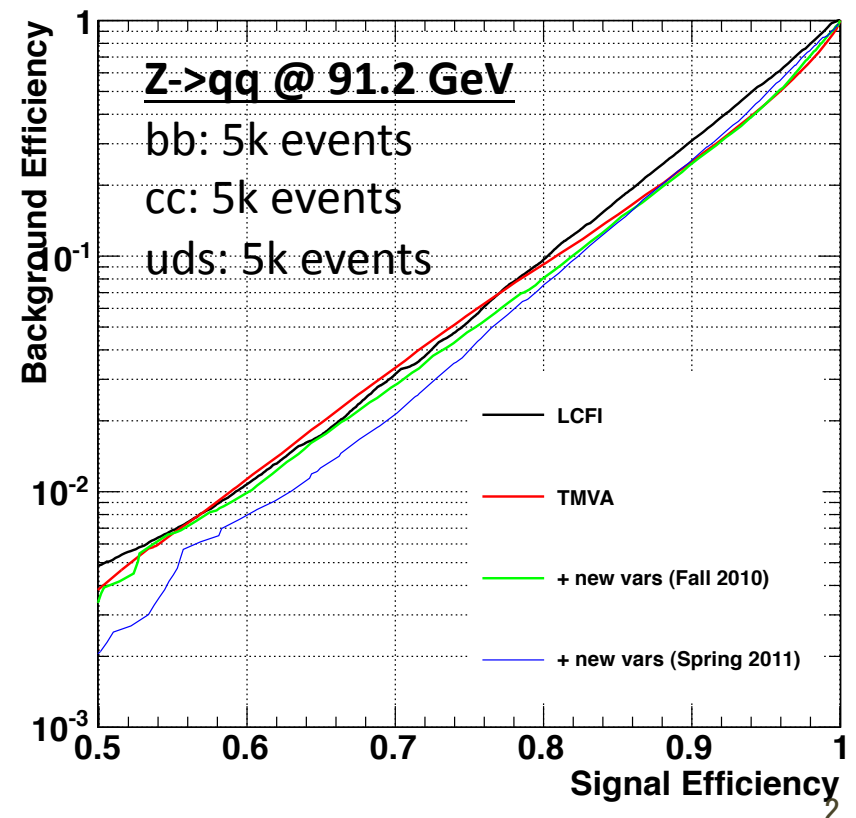
May 22, 2011

ILD Workshop Software Pre-meeting

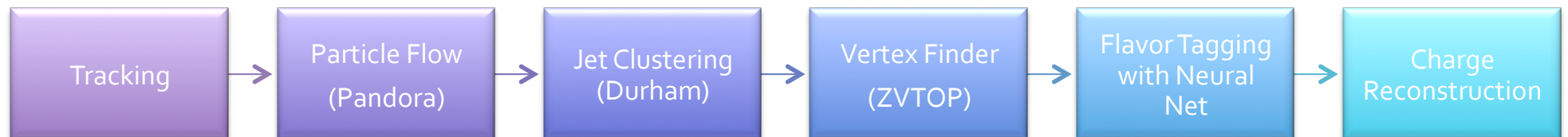
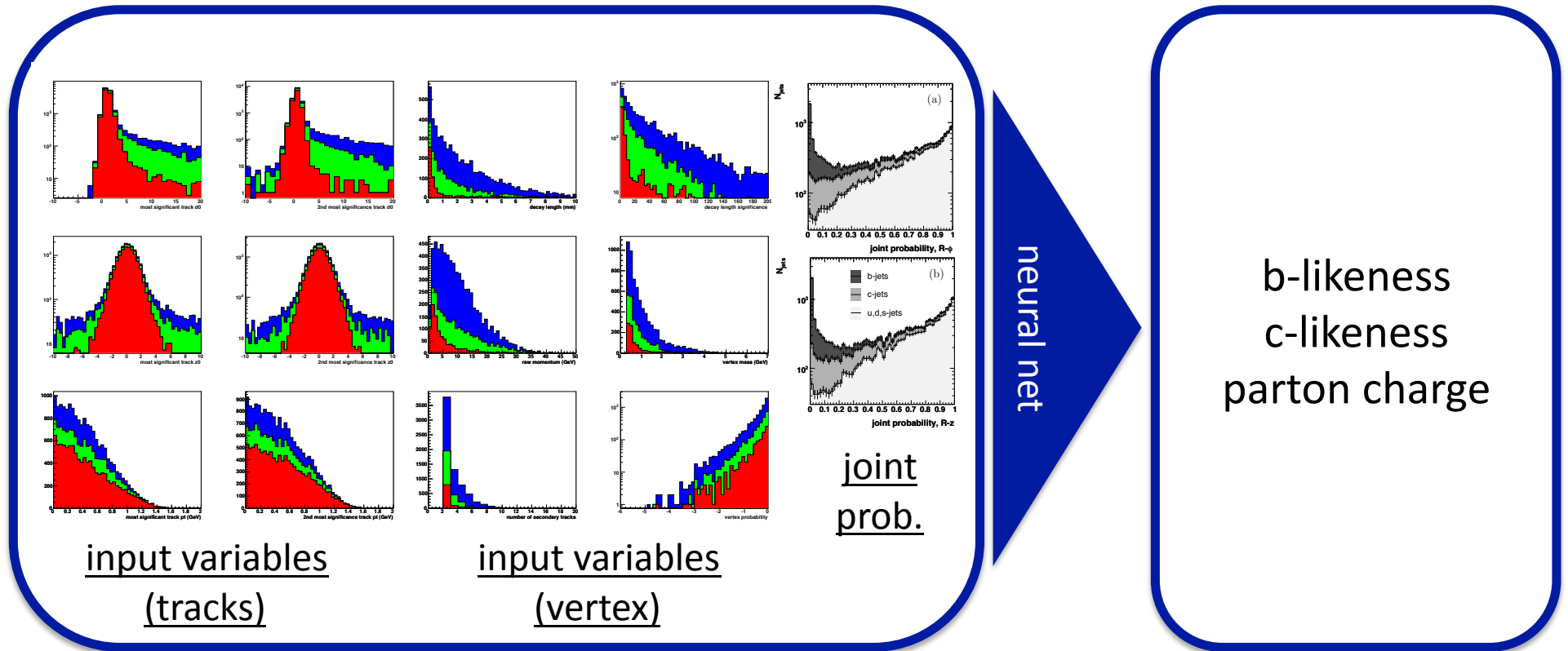
topics

```
<processor name="MyFlavtagProcessor" type="FlavtagProcessor">  
  <parameter name="algorithm" type="stringVec">BuildUpVertex JetClustering</parameter>  
  <parameter name="PFOCollection" type="string" value="PandoraPFOs" />  
  <parameter name="MCPCollection" type="string" value="MCParticlesSkimmed" />  
  <parameter name="MCPFORelation" type="string" value="RecoMCTruthLink" />  
  <parameter name="VertexCollectionName" type="string" value="BuildUpVertex" />  
  <parameter name="JetCollectionName" type="string" value="Suehara_6Jets" />  
</processor>
```

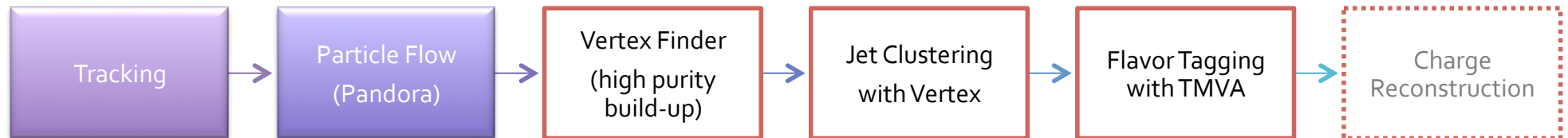
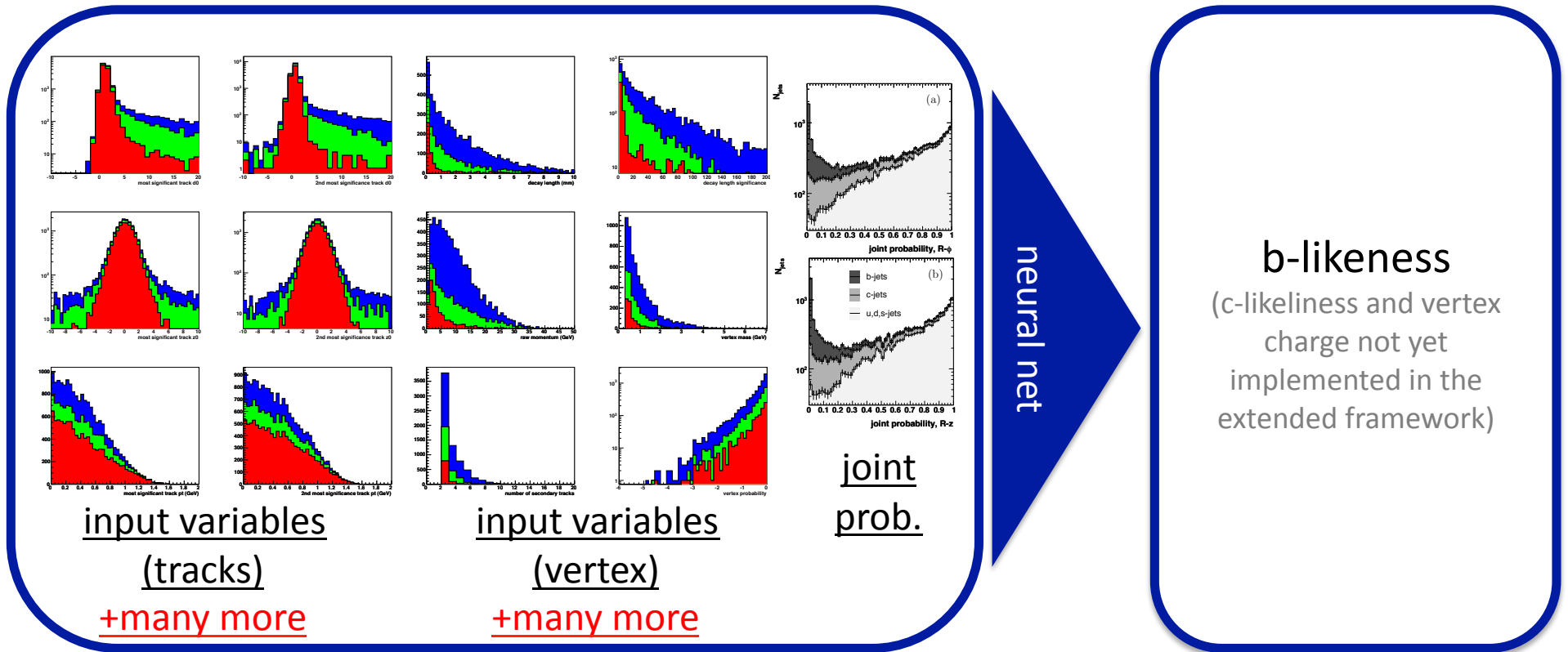
- flavor tagging framework
- improvements in flavor tagging



LCFIVertex (existing)



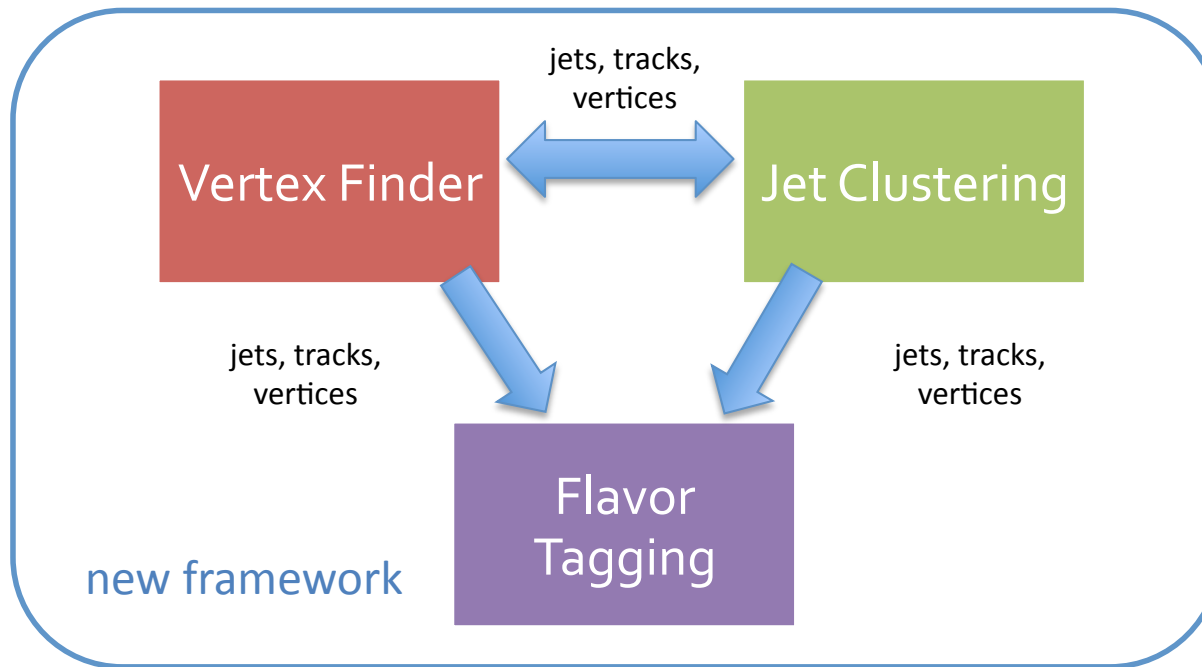
LCFIVertex (extended)



- VF \leftrightarrow JC order switched, critical for many jet events!
- new multivariate analysis framework with TMVA

LCFIVertex framework

- improvements in **vertex finding, jet clustering, flavor tagging** in a unified way
 - creation of a new framework suited to this task
 - data types: event, track, neutral, mcparticle, jet, vertex
 - algorithms: vertex finding, jet clustering, flavor tagging



see Taikan's ALCPG talks for details of VF & JC

ilcsoft integration

- framework runs in different modes
 - standalone executable mode: **done**
 - call from root macro: **done**
 - Marlin processor: **in progress**
 - modular design: algorithms can be turned on/off through XML

define algorithm to run

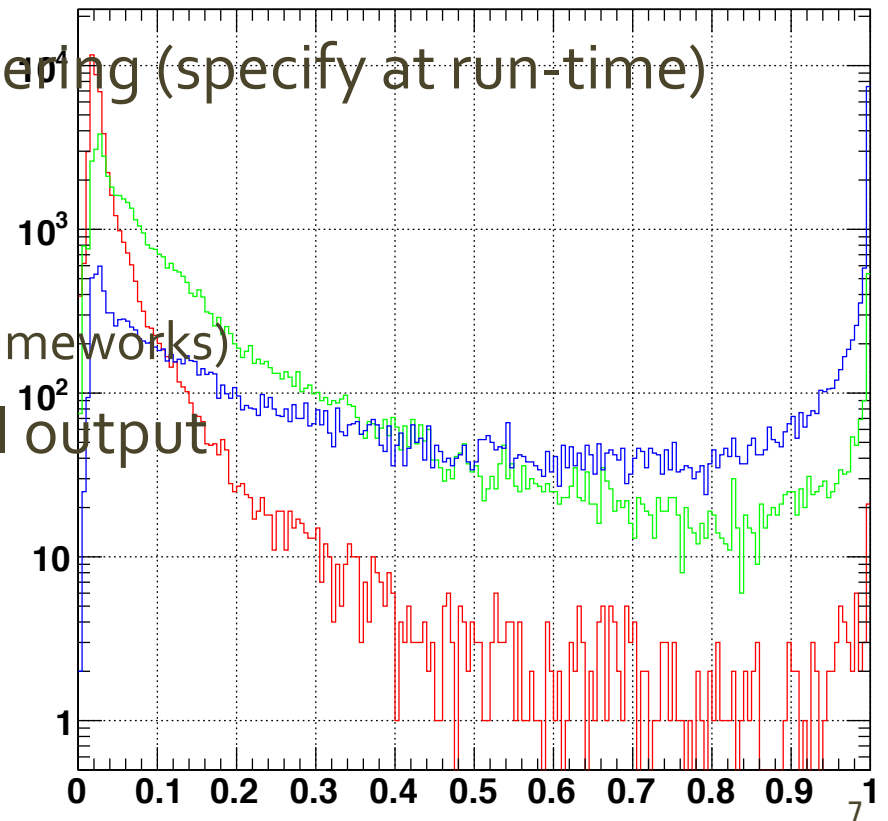
```
<processor name="MyFlavtagProcessor" type="FlavtagProcessor">  
  <parameter name="algorithm" type="stringVec">BuildUpVertex JetClustering</parameter>  
  <parameter name="PFOCollection" type="string" value="PandoraPFOs" />  
  <parameter name="MCPCollection" type="string" value="MCParticlesSkimmed" />  
  <parameter name="MCPFORelation" type="string" value="RecoMCTruthLink" />  
  <parameter name="VertexCollectionName" type="string" value="BuildUpVertex" />  
  <parameter name="JetCollectionName" type="string" value="Suehara_6Jets" />  
</processor>
```

input collection

output collection

flavor tagging procedure

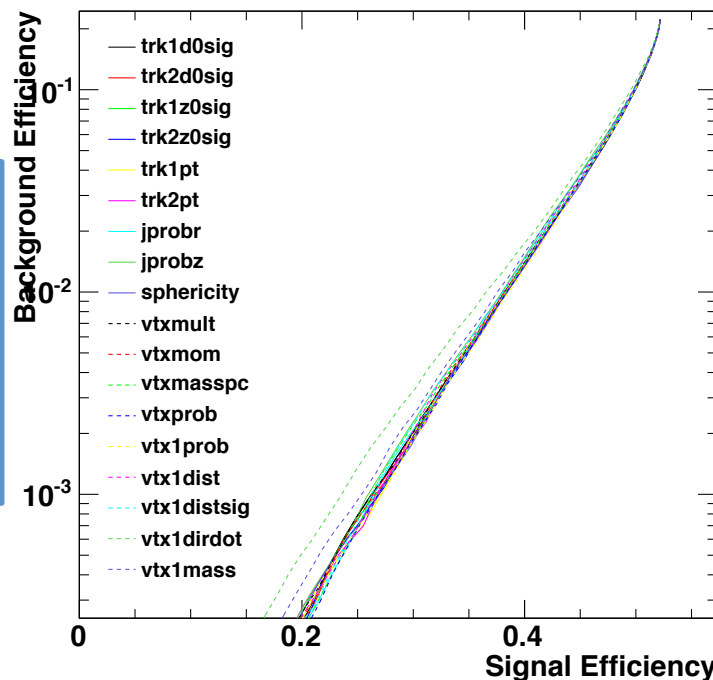
- prepare training and testing samples
- define input variables and training categories
 - need to write code implementing the input variables
- automated training process
 - apply vertex finding and jet clustering (specify at run-time)
 - normalizes of input variables
 - invokes TMVA classifier training
 - (future work: specify other MVA frameworks)
 - creates weight files used for final output
- run physics analysis!



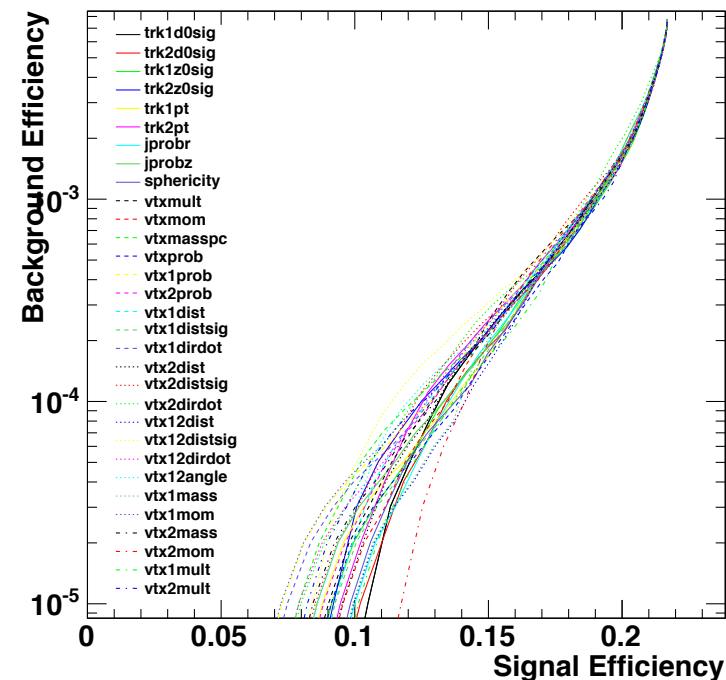
improvements in flavor tagging

- we have already identified new variables (shown at ALCPG11) which improves flavor tagging
 - direction of the vertex momentum (combined from tracks)
 - separation significance of the secondary vertices
 - new variable: look at tracks not combined in any vertex (just count them for now)

$n_{vtx}==1$

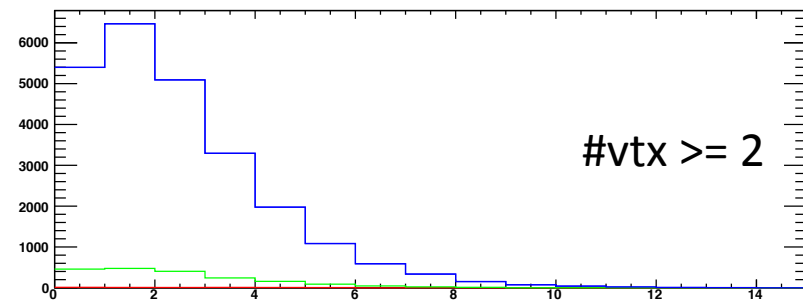
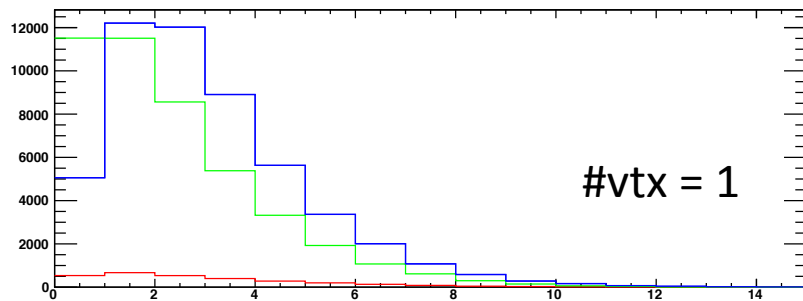
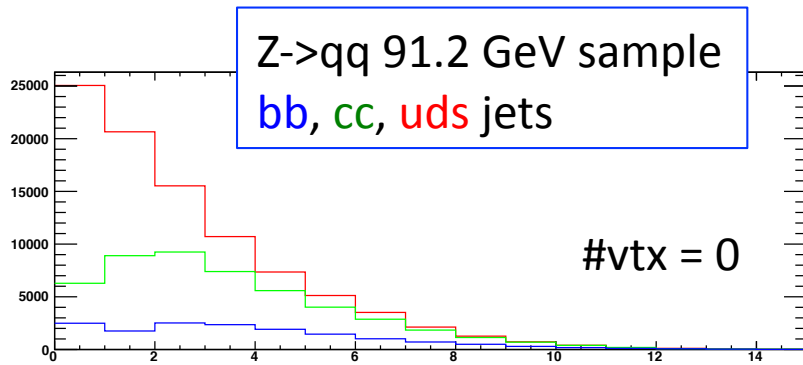


$n_{vtx}>=2$

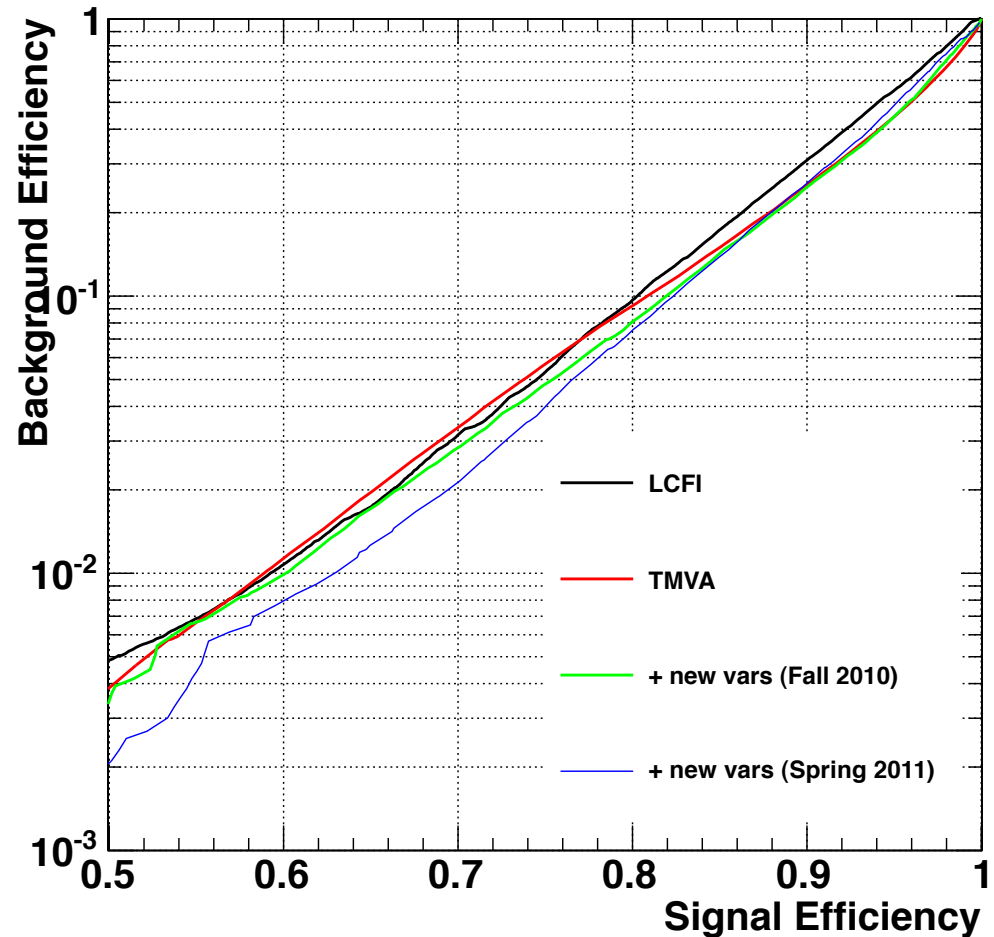


sig vs. bkg efficiency after removing one variable (worse performance means good uncorrelated discrimination power of the variable)

flavor tagging



number of trks *not* used in
primary or secondary vertex



- we do much better in the high purity region
- this demonstrates that there is still information in the jet that we have not previously used
 - should be even better if we use track parameters & errors, e.g. joint probabilities with respect to secondary vertices

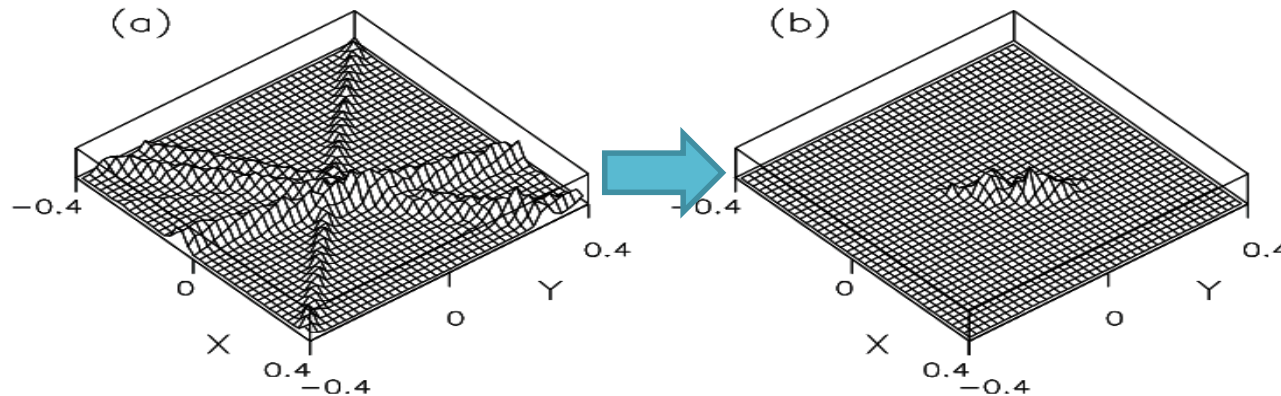
summary and plans

- we have extended the LCFIVertex framework capable of making improvements in vertex finding, jet clustering, and flavor tagging
- we're working towards releasing the framework after Marlin integration is complete (<1 month)
 - TODO: write documentation
- we will establish common development model
 - with close connection with CLIC (Jan, Lucie)
- optimization at higher energies, different detector configurations, inclusion of backgrounds
- lepton ID within jets would be extremely useful
 - need help from PFA expert (Mark?)
- application of new framework in ZHH analysis
 - talk by Taikan in the main meeting

backup

Vertex Finder

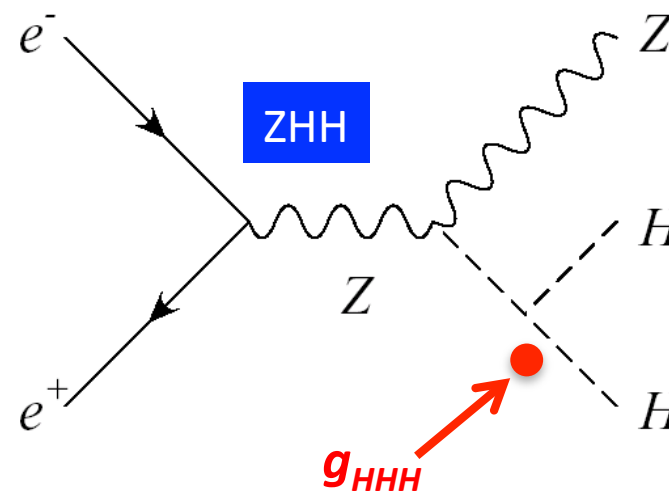
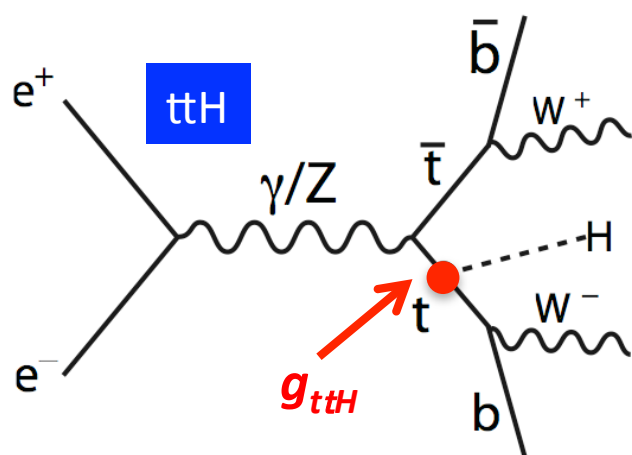
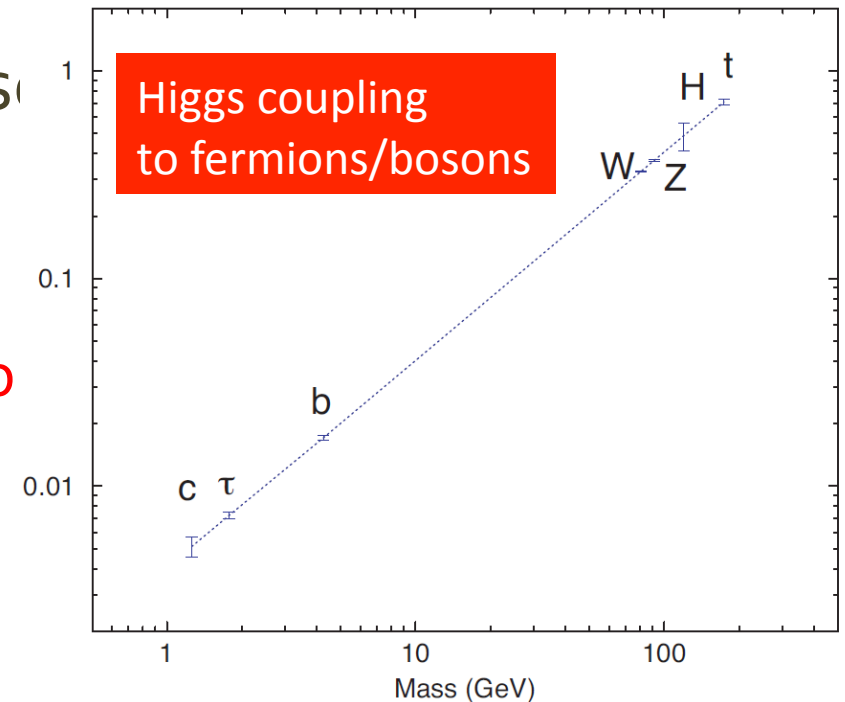
- topological vertex finder (ZVTOP)
track probability tubes
- vertex function: overlapping tubes \rightarrow high values



- can find vertices for arbitrary topology with any number of tracks (must be separated from the primary vertex)
- it takes lots of CPU time due to having to evaluate the vertex function at many points in space
- tear-down
 - start from a set of tracks, remove tracks which are inconsistent (large chi-squared contribution)
 - if the primary tracks are properly removed, vertices can be found with high efficiency
- build-up
 - using track pairs as seed, attach other tracks
 - good seeds lead to good vertices

motivation

- Many important physics processes have multiple heavy flavor jets
 - Higgs BF: $H \rightarrow bb$, $H \rightarrow cc$
 - Higgs self-coupling: $ZHH \rightarrow qqbbb$
 - top-Yukawa: $ttH \rightarrow bWbWbb$
 - top physics: $tt \rightarrow bWbW$



$Z \rightarrow qq$ (70%)
 ll (30%)
 $W \rightarrow qq$ (65%)
 lv (35%)
 $H \rightarrow bb$ (65%)
 $(m_H = 120 \text{ GeV})$

ILD Detector

muon detector

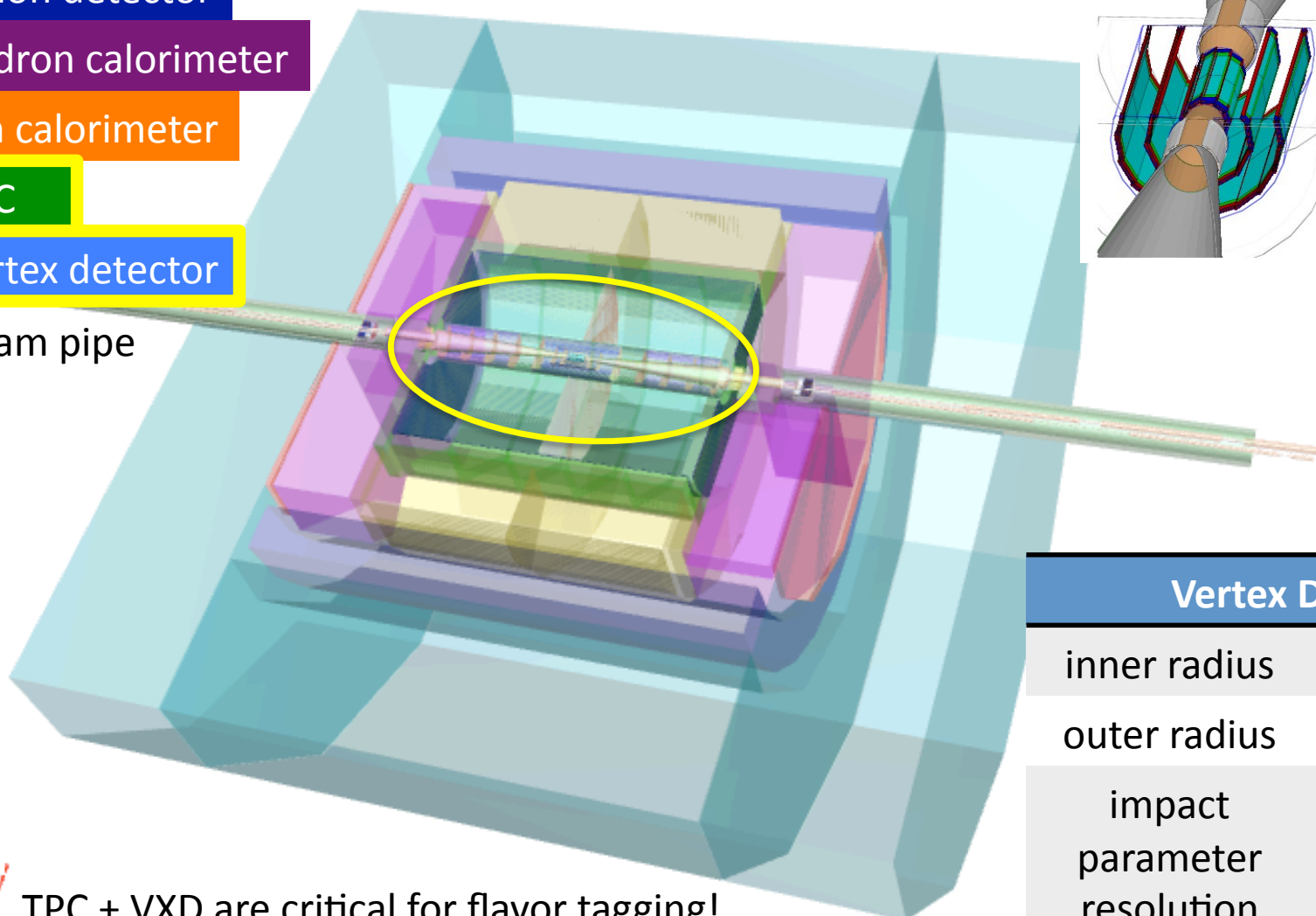
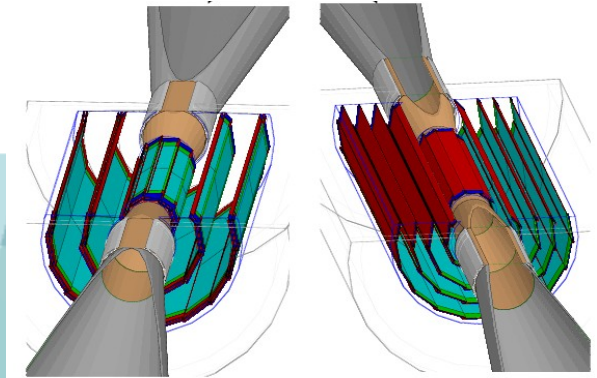
hadron calorimeter

em calorimeter

TPC

vertex detector

beam pipe



Vertex Detector

inner radius 15 mm

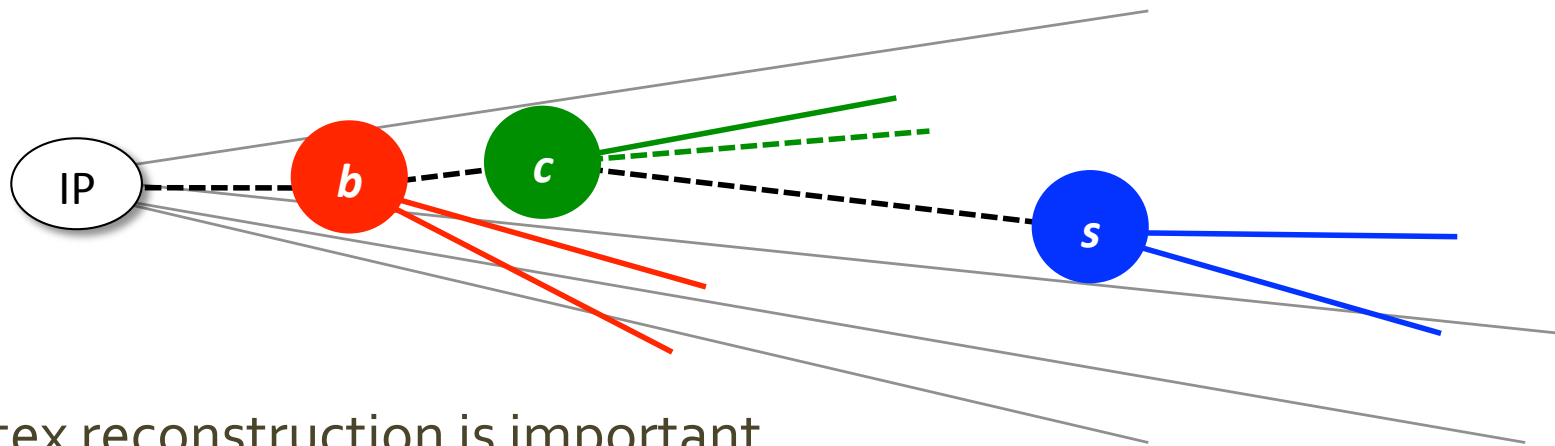
outer radius 60 mm

impact parameter resolution < 5 mm (high momentum)

 TPC + VXD are critical for flavor tagging!

ideal flavor tagging

- reconstruct the entire decay chain ($b \rightarrow c \rightarrow s$) in a jet



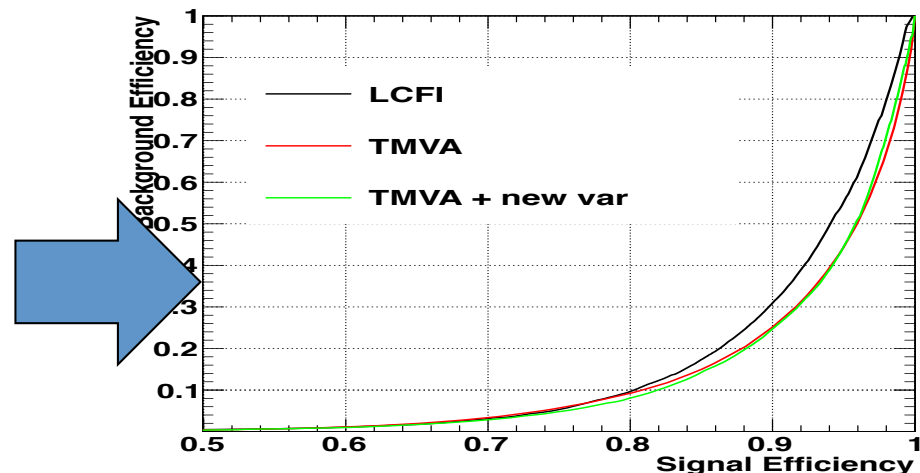
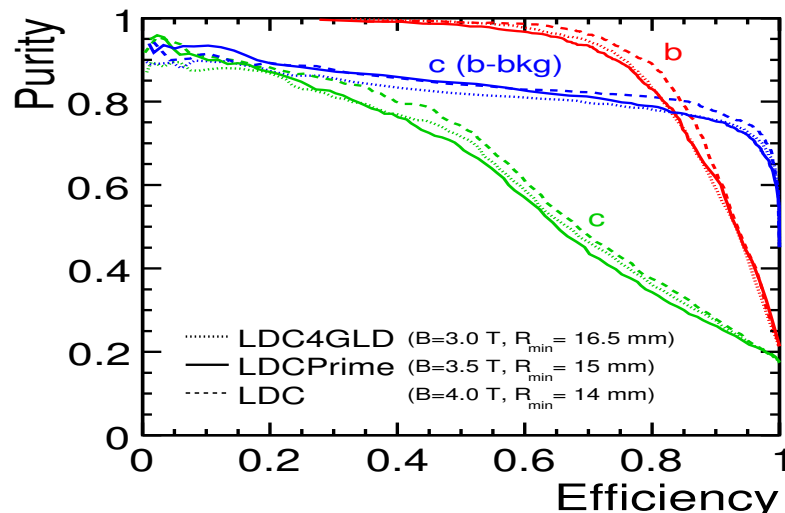
- vertex reconstruction is important
 - but vertices cannot be made from a single track
 - use track measurements (impact parameter)
- presence of neutral particles
 - missing correction by using p_T
- lepton ID: energetic/isolated leptons is a sign of heavy quark decays
- key is variable combination
 - likelihood, multivariate analysis (e.g. ANN)
 - event categorization (discrete variables)

new input variables

- the new variables are trained by BDT
 - for # vertex = 0 (9 variables):
 - d_0 impact parameter (1)
 - d_0 impact parameter (2)
 - z_0 impact parameter (1)
 - z_0 impact parameter (2)
 - track momentum (1)
 - track momentum (2)
 - d_0 joint probability
 - z_0 joint probability
 - boosted sphericity
 - for # vertex = 1 (17 variables):
 - d_0 impact parameter (1)
 - d_0 impact parameter (2)
 - z_0 impact parameter (1)
 - z_0 impact parameter (2)
 - track momentum (1)
 - track momentum (2)
 - d_0 joint probability
 - z_0 joint probability
 - boosted sphericity
 - vertex decay length
 - vertex decay length significance
 - vertex momentum
 - vertex mass (pt-corrected)
 - vertex mass (not pt-corrected)
 - vertex multiplicity
 - vertex probability from the fitter
 - vertex disp/momentum angle
 - for # vertex ≥ 2 (29 variables):
 - d_0 impact parameter (1)
 - d_0 impact parameter (2)
 - z_0 impact parameter (1)
 - z_0 impact parameter (2)
 - track momentum (1)
 - track momentum (2)
 - d_0 joint probability
 - z_0 joint probability
 - boosted sphericity
 - vertex #1 decay length
 - vertex #2 decay length
 - distance between vertex #1 & #2
 - vertex #1 decay length significance
 - vertex #2 decay length significance
 - separation significance between vertex #1 & #2
 - vertex #1 momentum
 - vertex #2 momentum
 - vertex momentum (combined)
 - vertex #1 mass (not pt-corrected)
 - vertex #2 mass (not pt-corrected)
 - vertex mass (combined, pt-corrected)
 - vertex #1 multiplicity
 - vertex #2 multiplicity
 - vertex multiplicity (combined)
 - vertex probability from the fitter
 - vertex #1 disp/momentum angle
 - vertex #2 disp/momentum angle
 - vertex #1/#2 disp/momentum angle
 - vertex #1/#2 angle

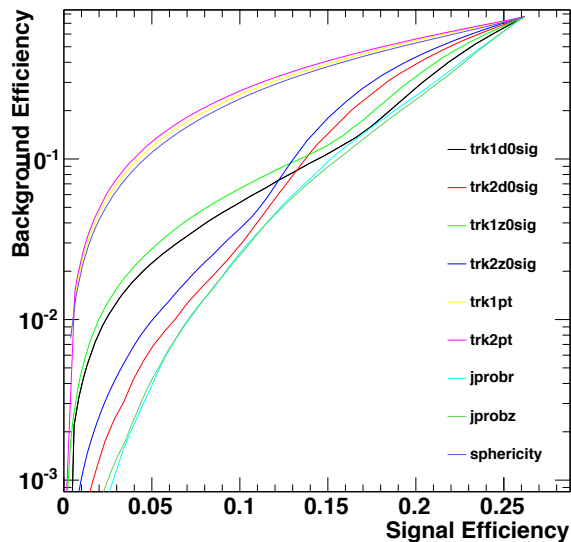
evaluating classifier response

- Lol flavor-tagging evaluation produced purity-efficiency plots
 - but this depends on the fraction of heavy jets, which changes from sample to sample
 - $BF(Z \rightarrow bb) = 15\%$, $BF(H_{120 \text{ GeV}} \rightarrow bb) = 68\%$
- better to use a fraction-independent measure: evaluate using background efficiency versus signal efficiency instead

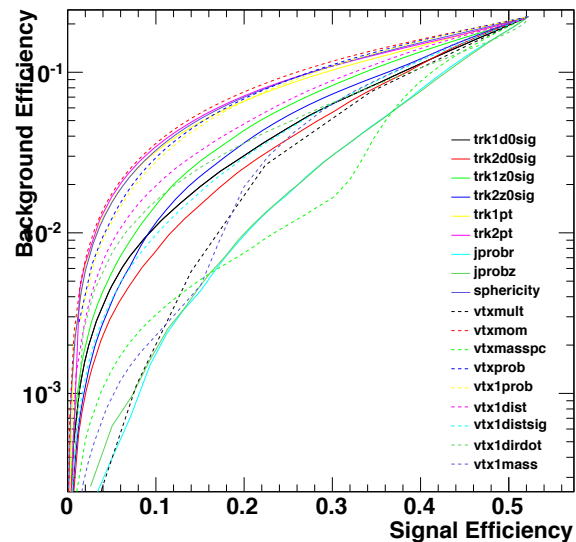


variable ranking

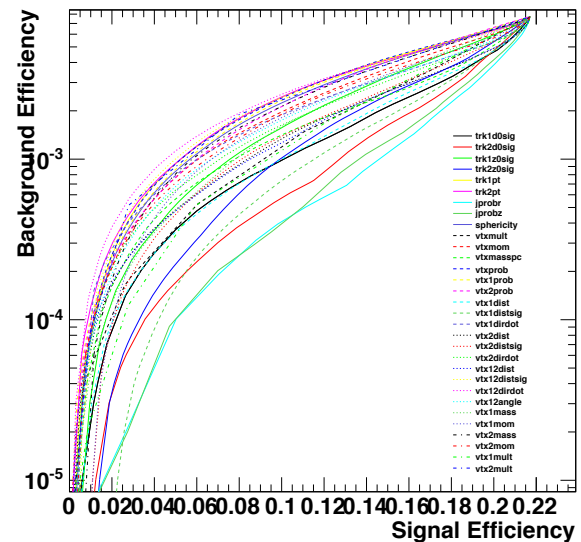
nvtx==0



nvtx==1



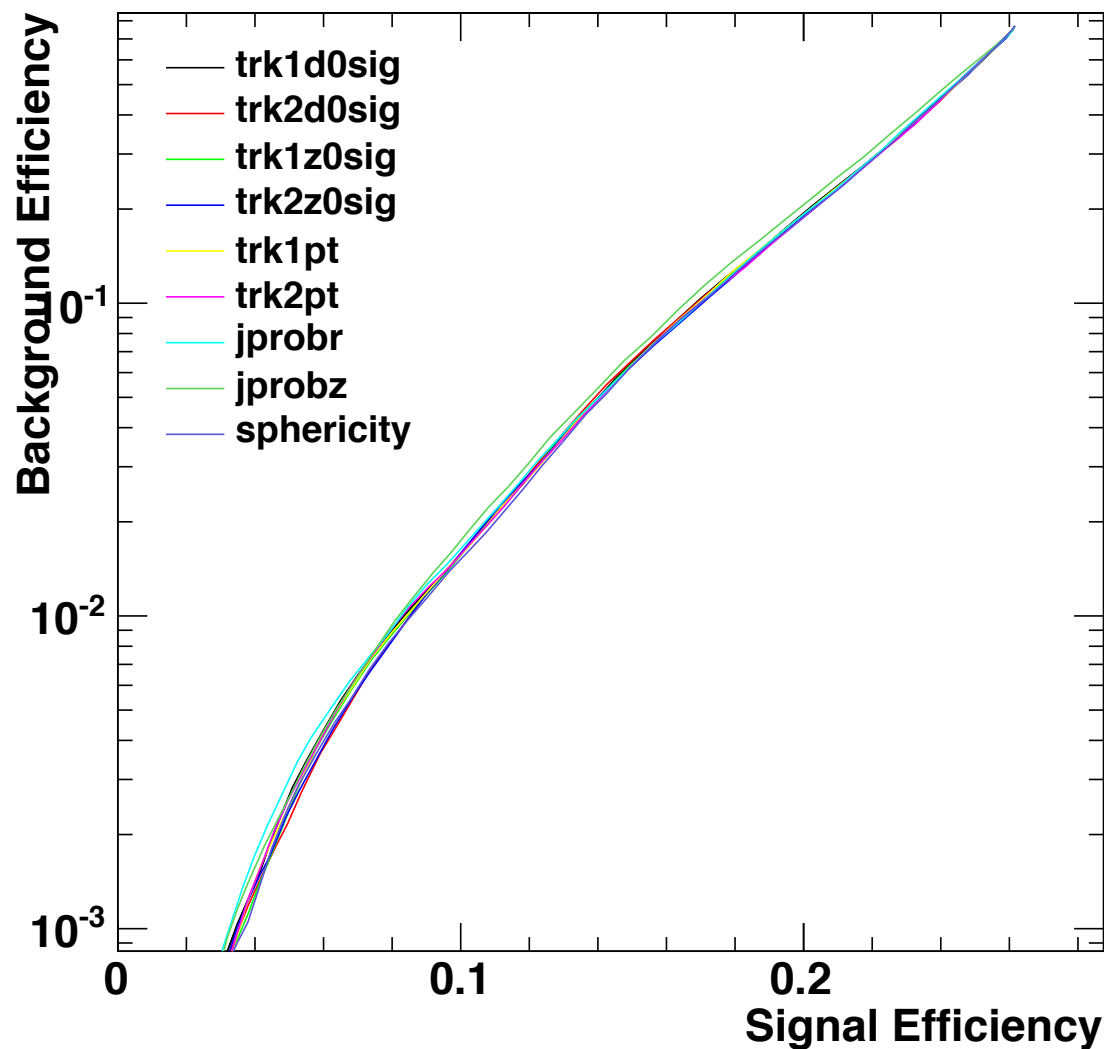
nvtx>=2



- single variable ranking shows the most useful variables (on their own)
 - joint probabilities, vertex mass
- however, any other uncorrelated variable can help; these plots do not show this effect

variable ranking (correlation)

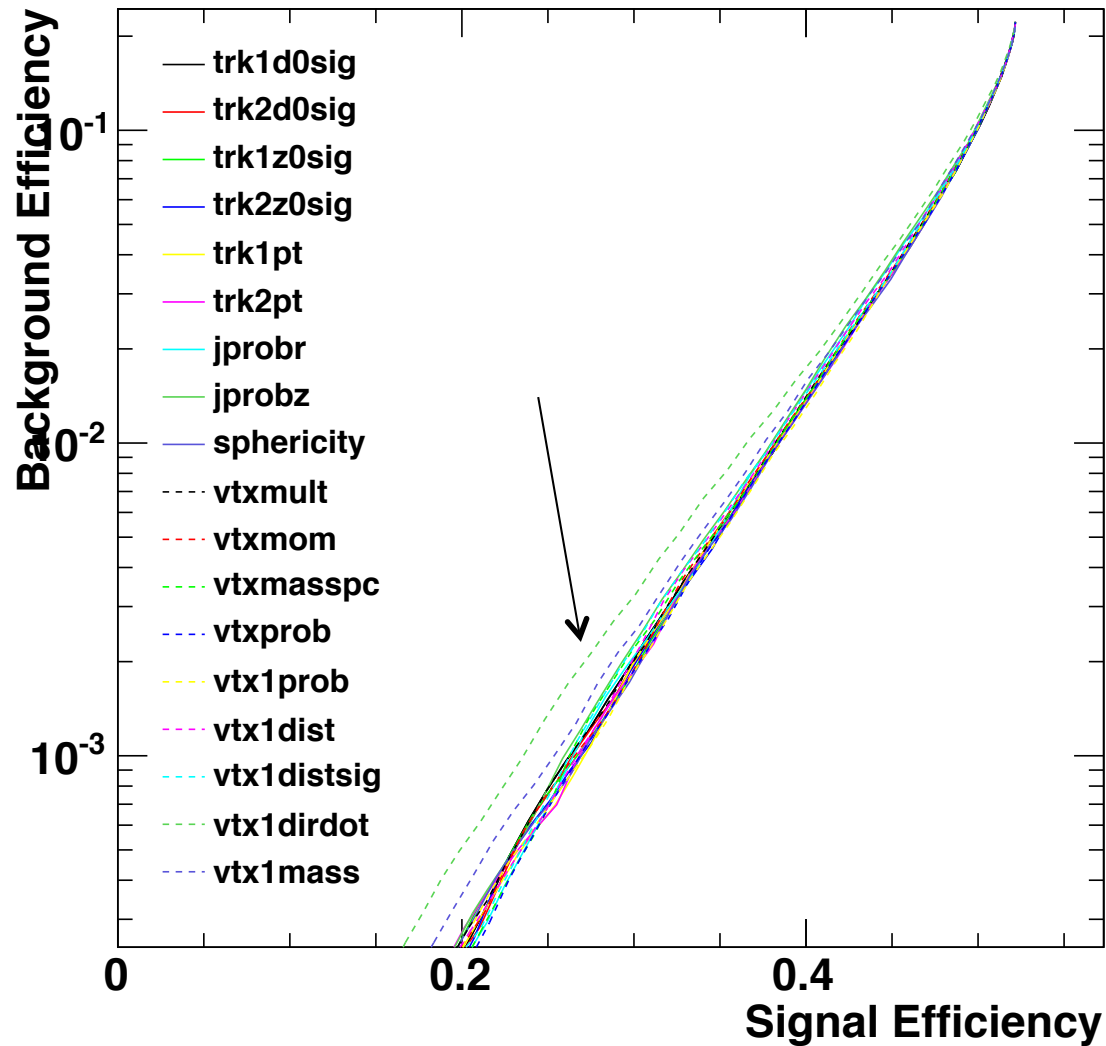
nvtx==0



- result of training after removing a variable
- this shows how “unique” this variable is in terms of uncorrelated classifying power
- significantly worse performance after removing the variable shows that it’s effective
- for nvtx=0, joint probabilities (both d_0 & z_0) are the most powerful as expected

variable ranking (correlation)

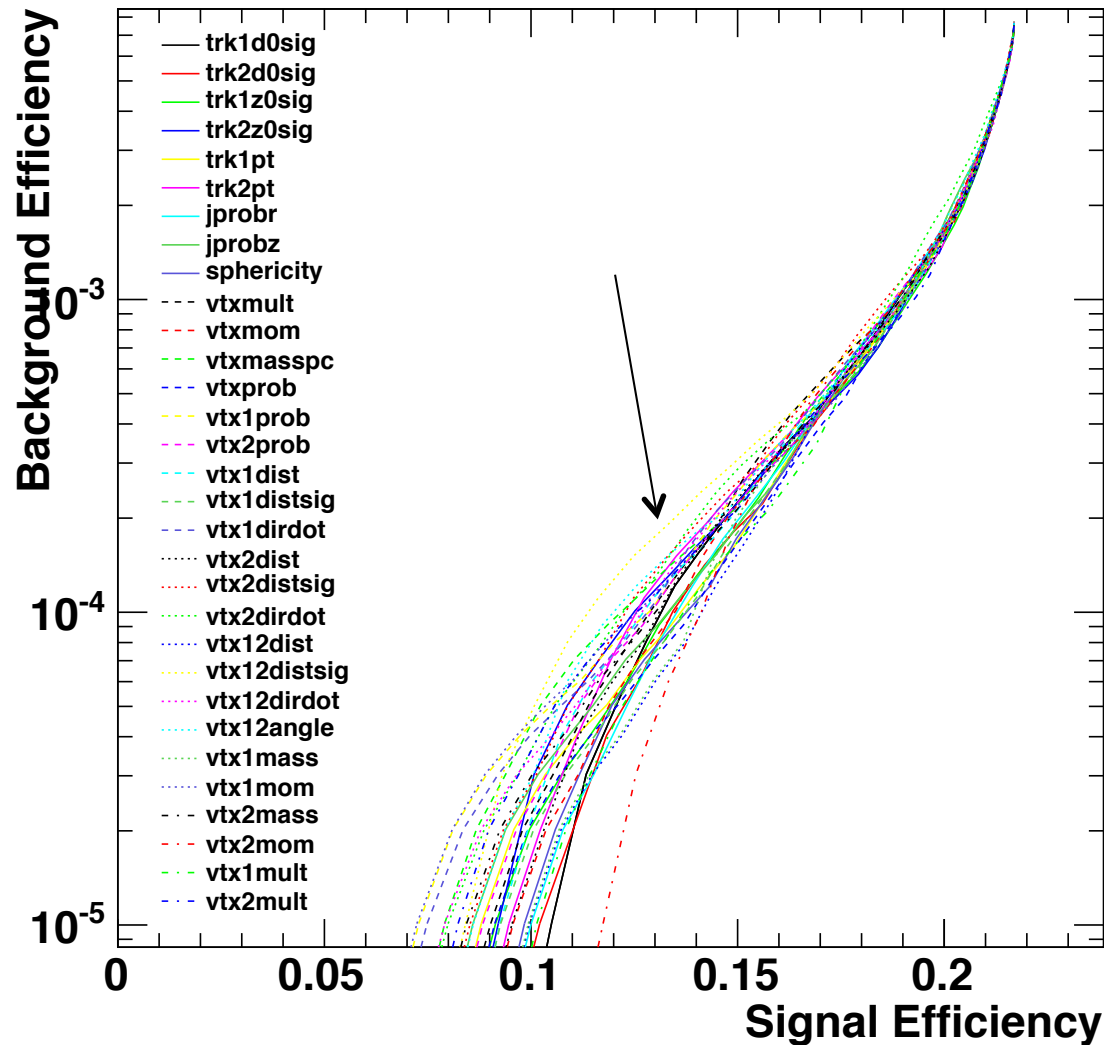
nvtx==1



- for nvtx=1, the most effective variables are:
 - displacement/ momentum angle of the vertex
 - uncorrected mass of the vertex
- newly added variables are shown to be effective!!!

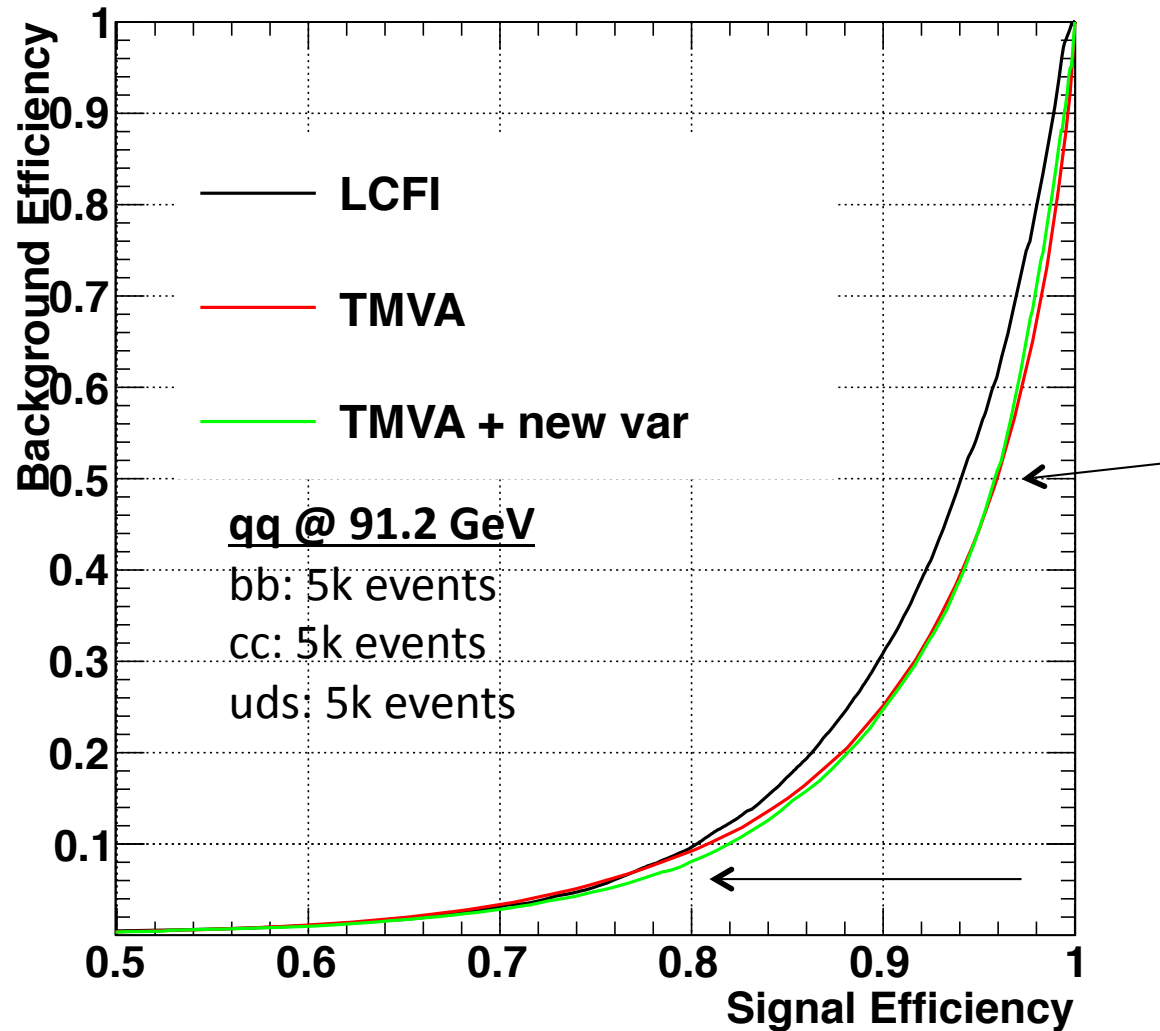
variable ranking (correlation)

nvtx>=2



- for nvtx>=2, the most effective variables are:
 - separation significance between the 1st and 2nd vertices
- AGAIN: newly added variables are shown to be effective!!!

results



background = cc &
uds mixed equally

there is already
improvement
merely by switching
to TMVA
(signal eff > 0.8)

more improvement
by adding new
variables
(signal eff > 0.75)

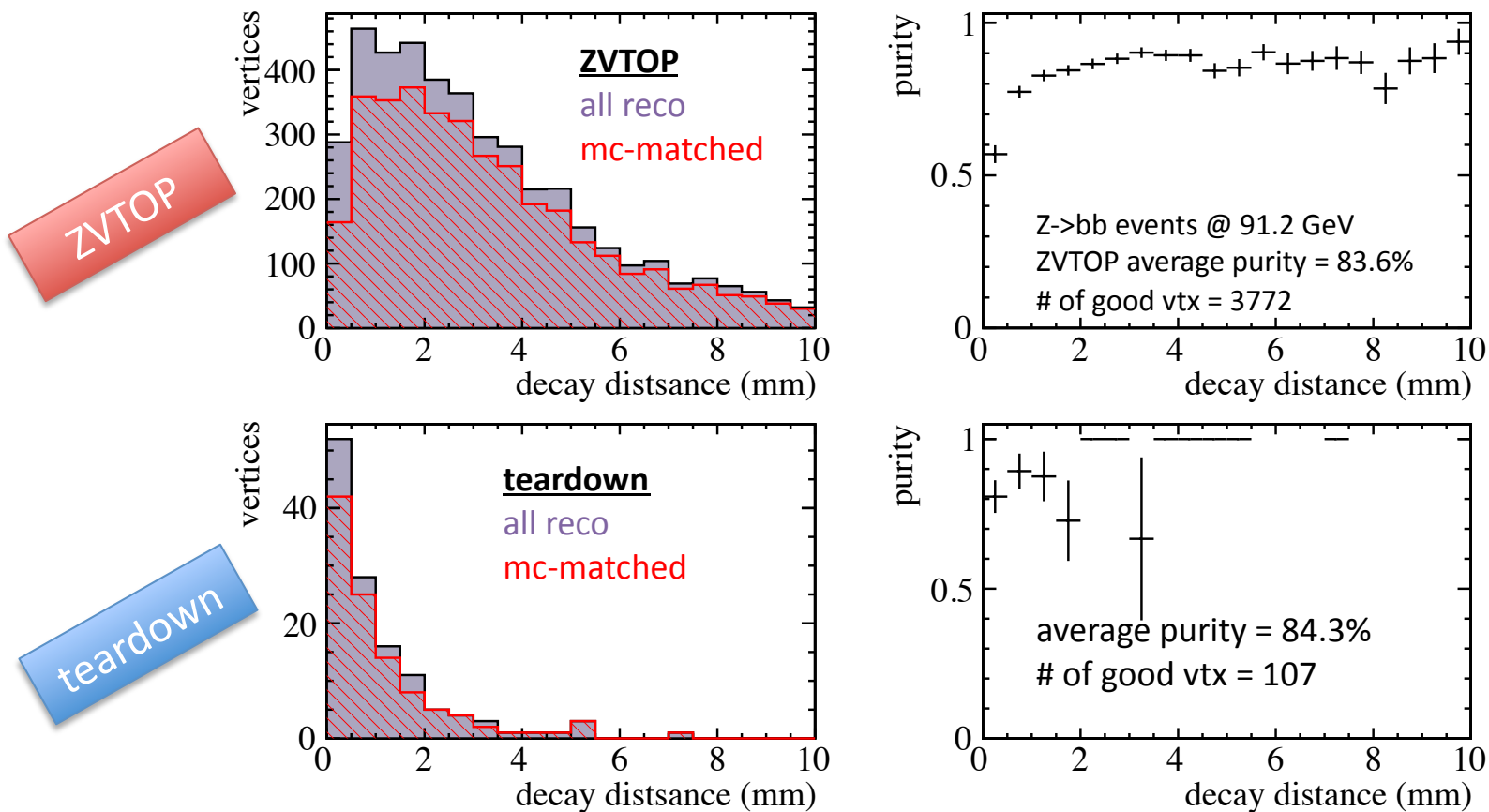
rejection of V^0 particles

- despite having V^0 taggers in the Marlin reconstruction chain, our vertex finders still find V^0 's (K_S , Lambda, conversions) for two-track vertices
- we apply the following cuts to reject V^0 's (reduce uds contamination):
 - cut on the angle θ between the vertex displacement from IP and the V^0 direction
 - mass requirements
 - K_S : $\cos\theta > 0.999$ & mass 15 MeV within PDG value
 - Lambda: $\cos\theta > 0.99995$ & mass 20 MeV within PDG value
 - conversions: $\cos\theta > 0.99995$ & less than 10 MeV for conversion mass, where the mass is geometrically corrected so that it is calculated using the track dip angles

$$m_{\text{conv}}^2 = 2|\vec{p}_1||\vec{p}_2|(1 - \cos \Delta\lambda_{12})$$

	before cut	after cut
K_S	3205	623
Lambda	1482	371
conversions	2544	278
other two-track reco vertices	30747	30333

ZVTOP + teardown



ZVTOP is not good at vertices near the IP.

Teardown is performed using tracks not used by ZVTOP.

After quality cuts (dist, chi2, ntrks), we find +3% efficiency increase for the same purity.

