Vertexing and Flavour tag performance ~Ideas for IDR~

Ryo Yonamine
Tohoku University

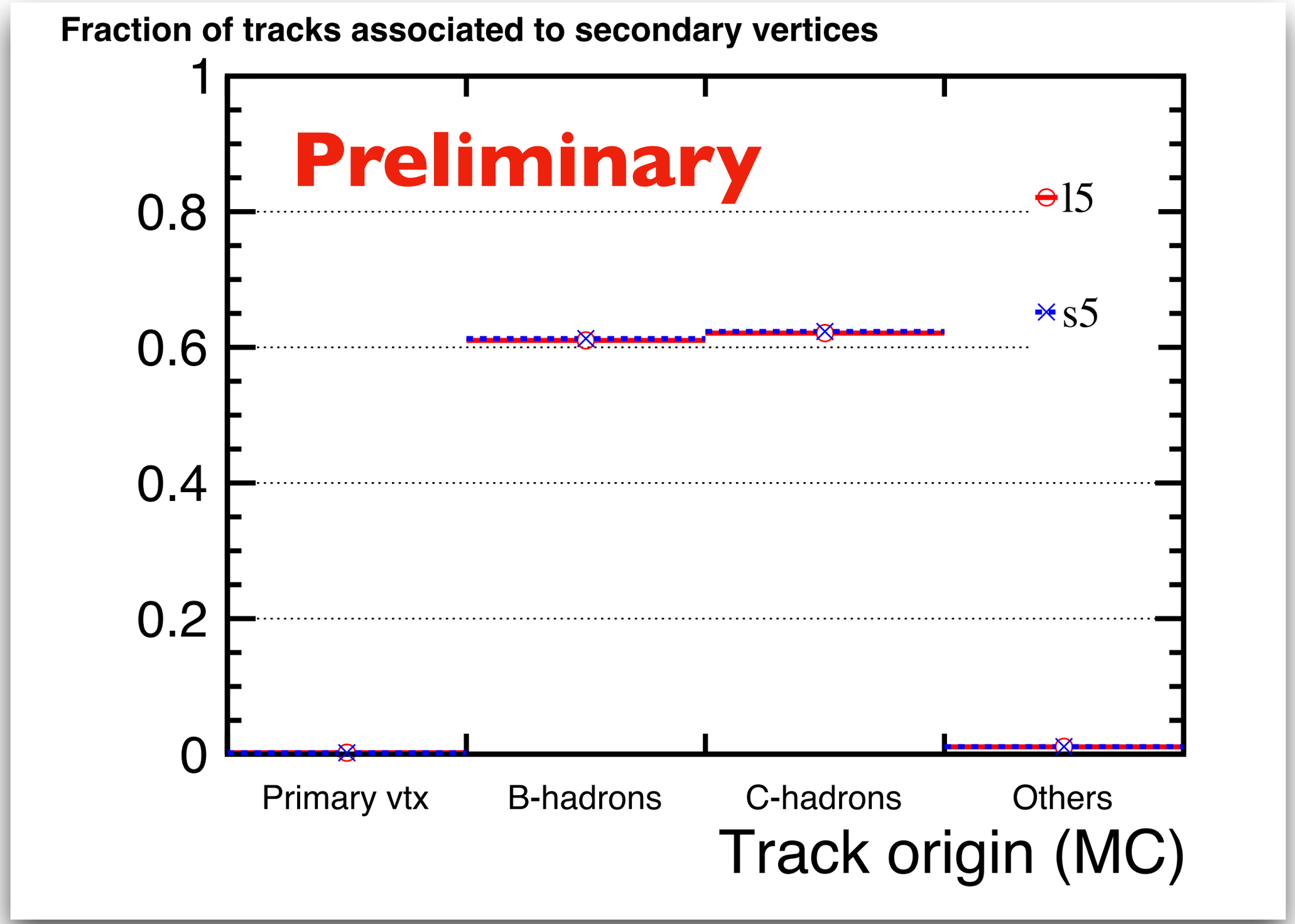
Plots shown in this presentation has not yet been discussed in the LCFIPlus developers. Will be done during this meeting.

ILD Benchmarking Days II@ KEK, 23rd Feb. 2019

Condition to produce the plots in this talk

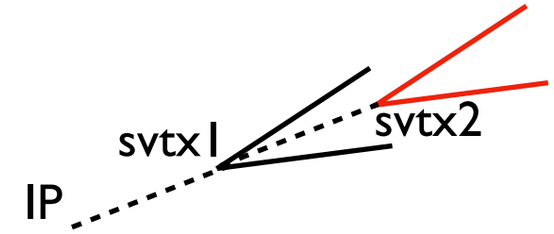
- **ilcsoft v02-00-02 + LCFIPlus v00-07**
  - Reprocess primary + secondary vertexing with LCFIPlus.
- **Samples (6b, 6c, 6q 500GeV)**
  - w/ beam bkg, I5 500GeV,
    - /hsm/ilc/grid/storm/prod/ilc/mc-opt.dsk/ild/dst/500-TDR_ws/flavortag/ILD_I5_o1_v02/v02-00-01/00010586
  - w/ beam bkg, s5 500GeV,
    - /hsm/ilc/grid/storm/prod/ilc/mc-opt.dsk/ild/dst/500-TDR_ws/flavortag/ILD_s5_o1_v02/v02-00-01/00010585

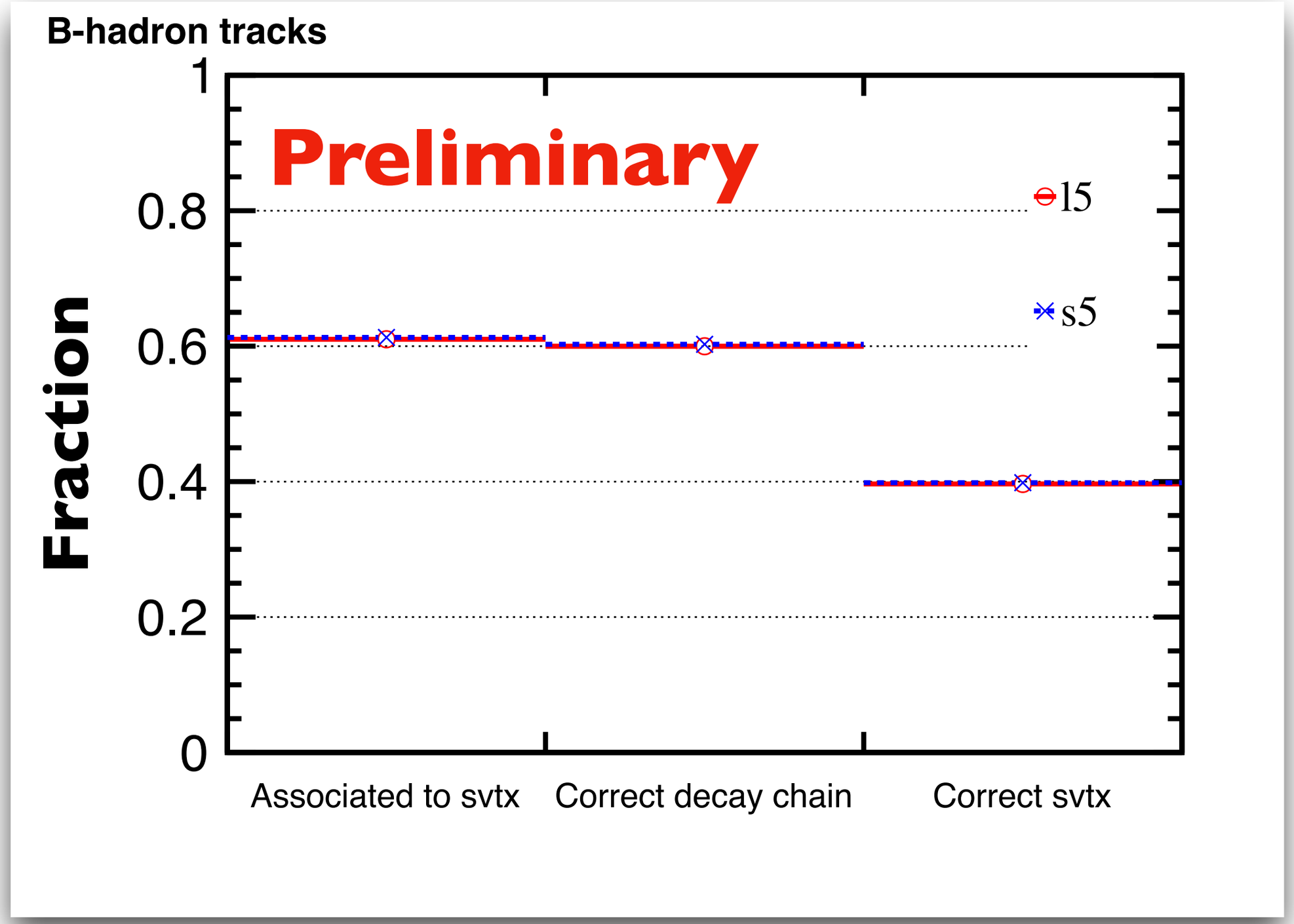
Vertexing performance



6b, 500 GeV, w/ beam bkg., w/ IP smearing.
Consistent with the earlier result with 2b, 91 GeV, w/o IP smearing.

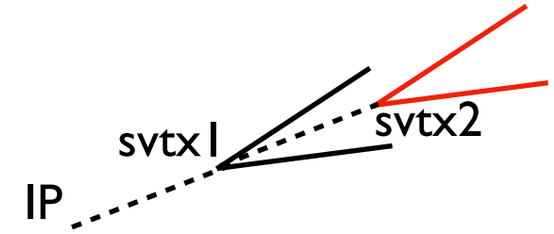
Vertexing performance

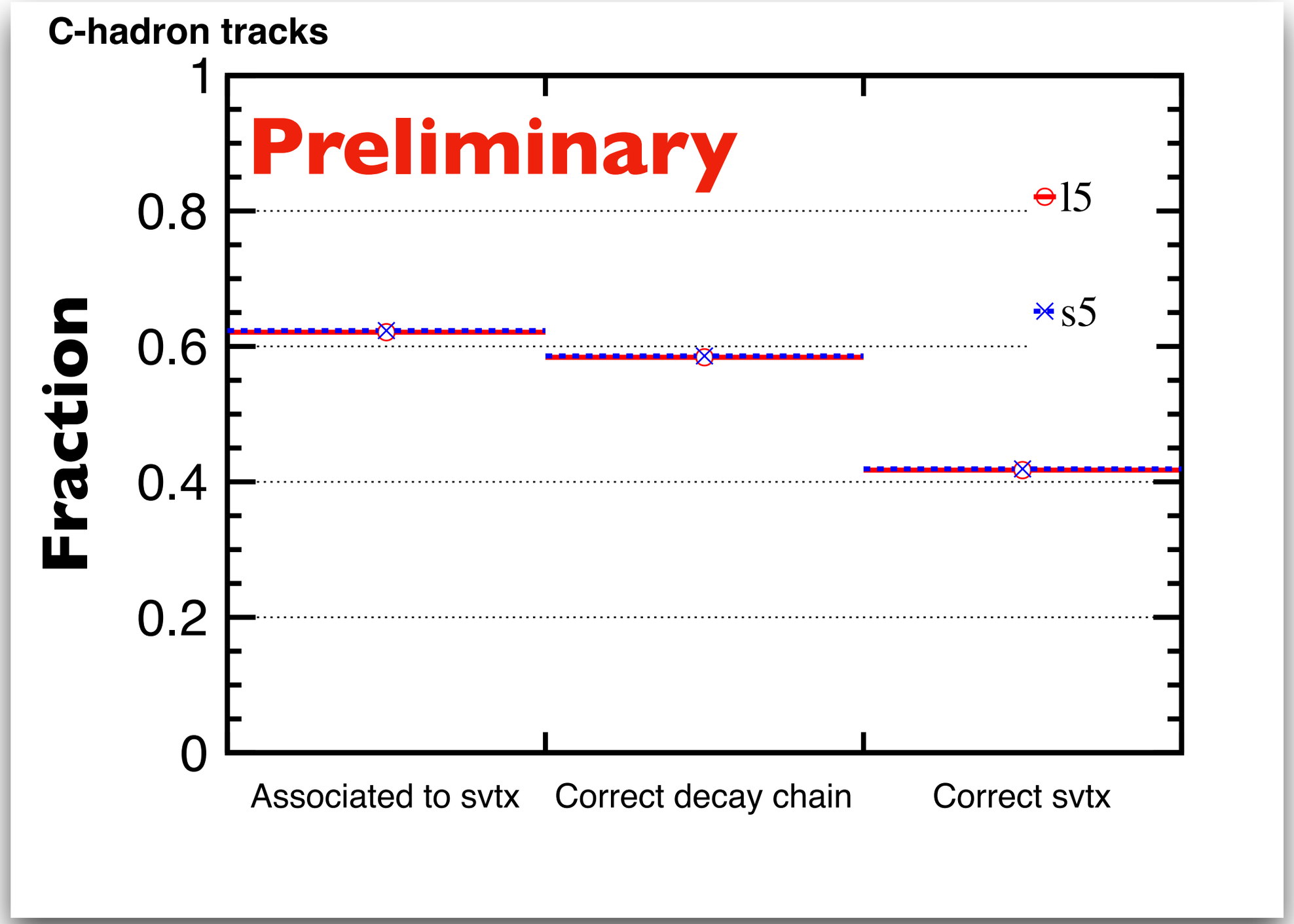




6b, 500 GeV, w/ beam bkg.,
consistent with the earlier result with 2b, 91 GeV.

Vertexing performance

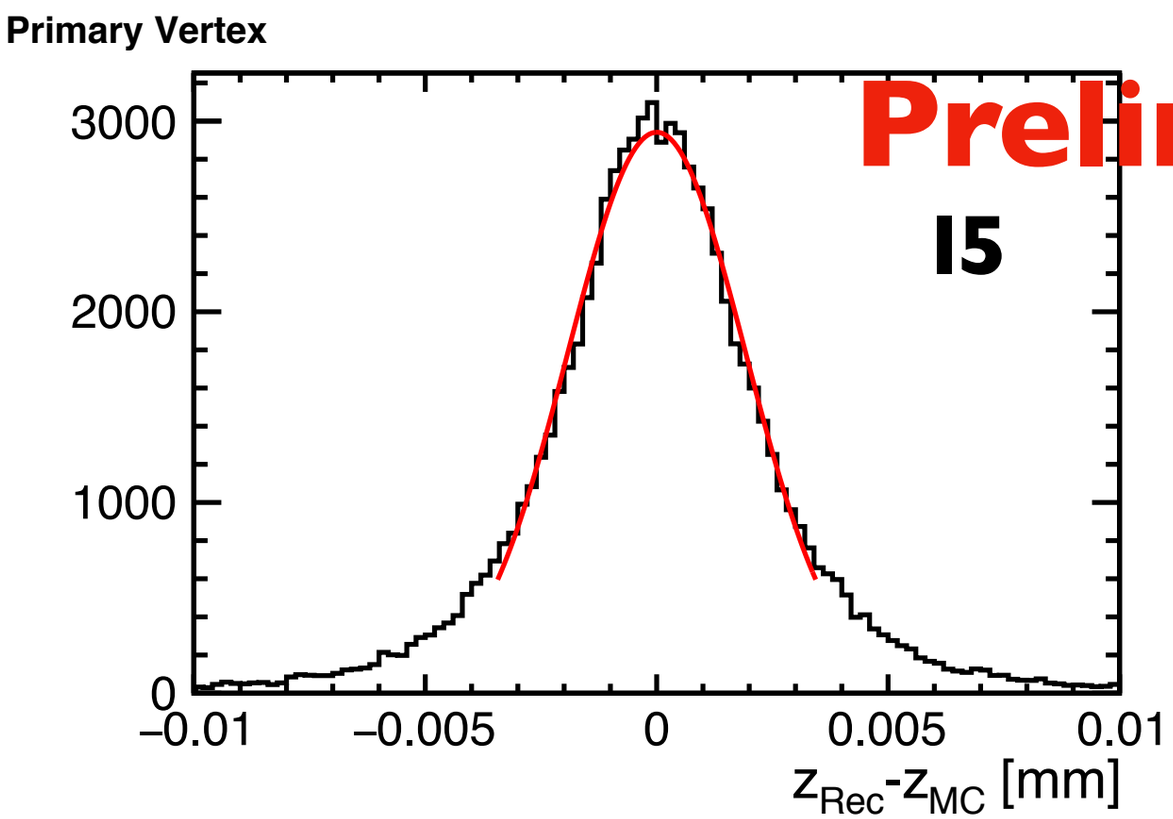


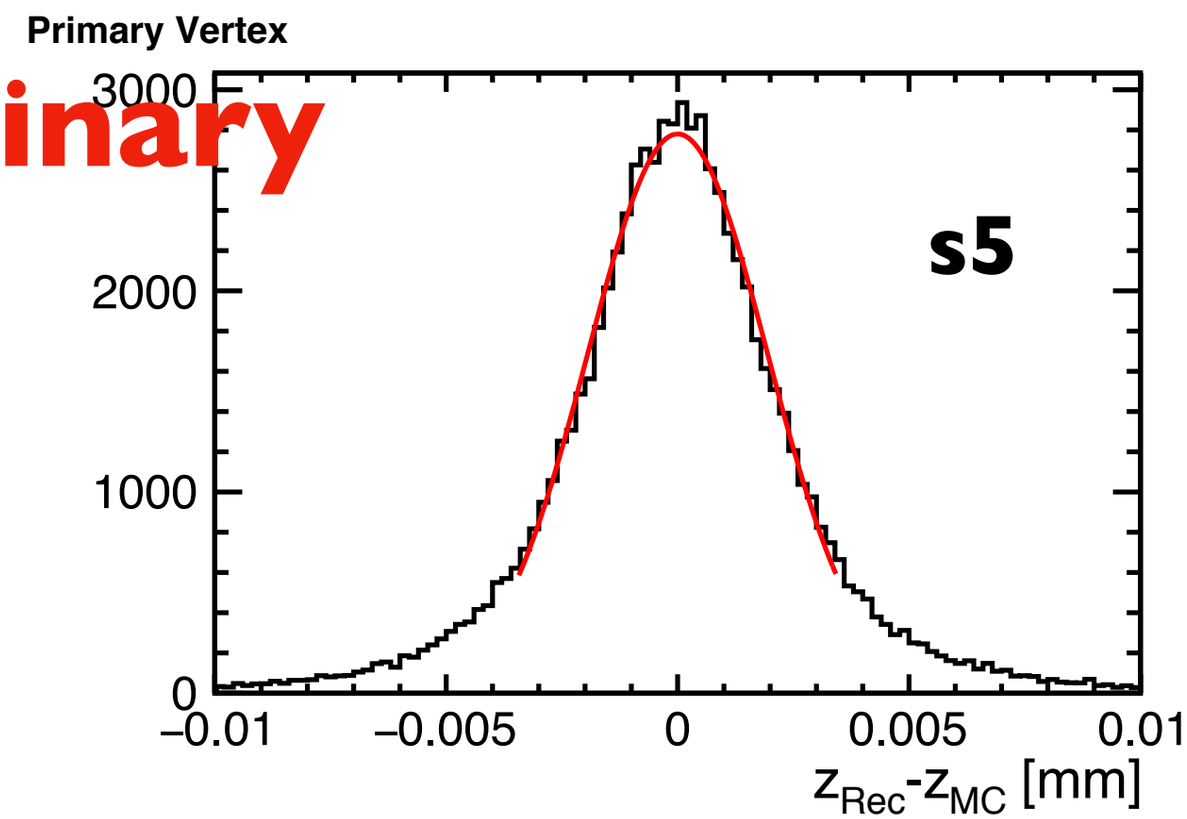


6b, 500 GeV, w/ beam bkg.,
consistent with the earlier result with 2b, 91 GeV.

Vertexing performance (Extra)

Primary vertex position resolution (z)



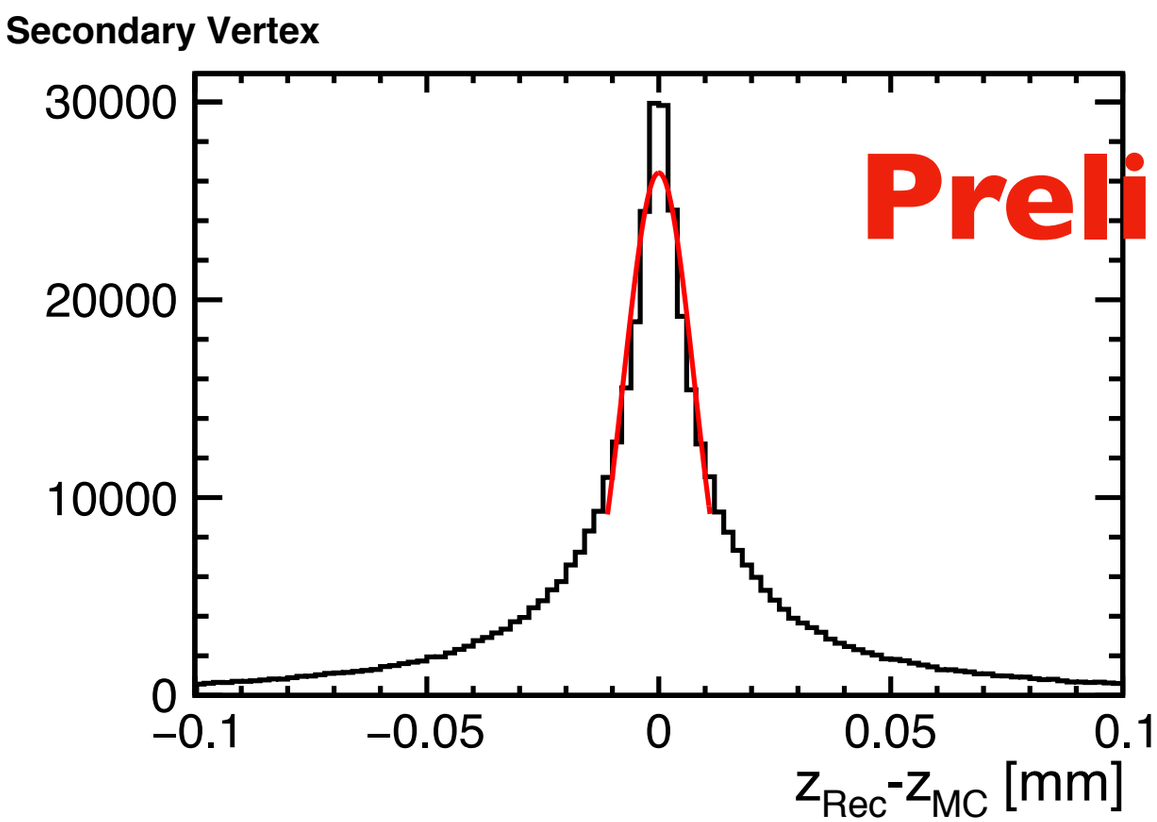


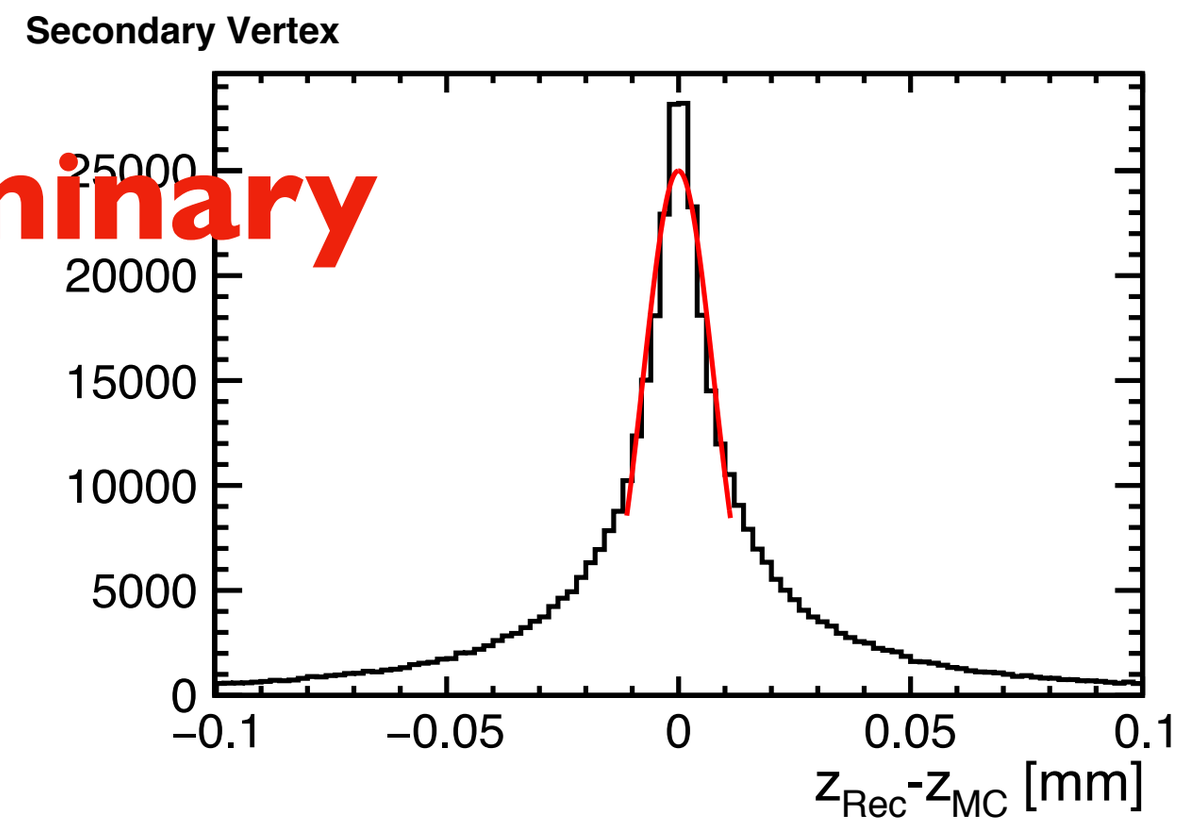
6b, 500 GeV, w/ beam bkg., w/ beam spot constraint (σz~200μm)

sigma(gauss fit) ~ 2μm

Vertexing performance (Extra)

Secondary vertex position resolution (z)



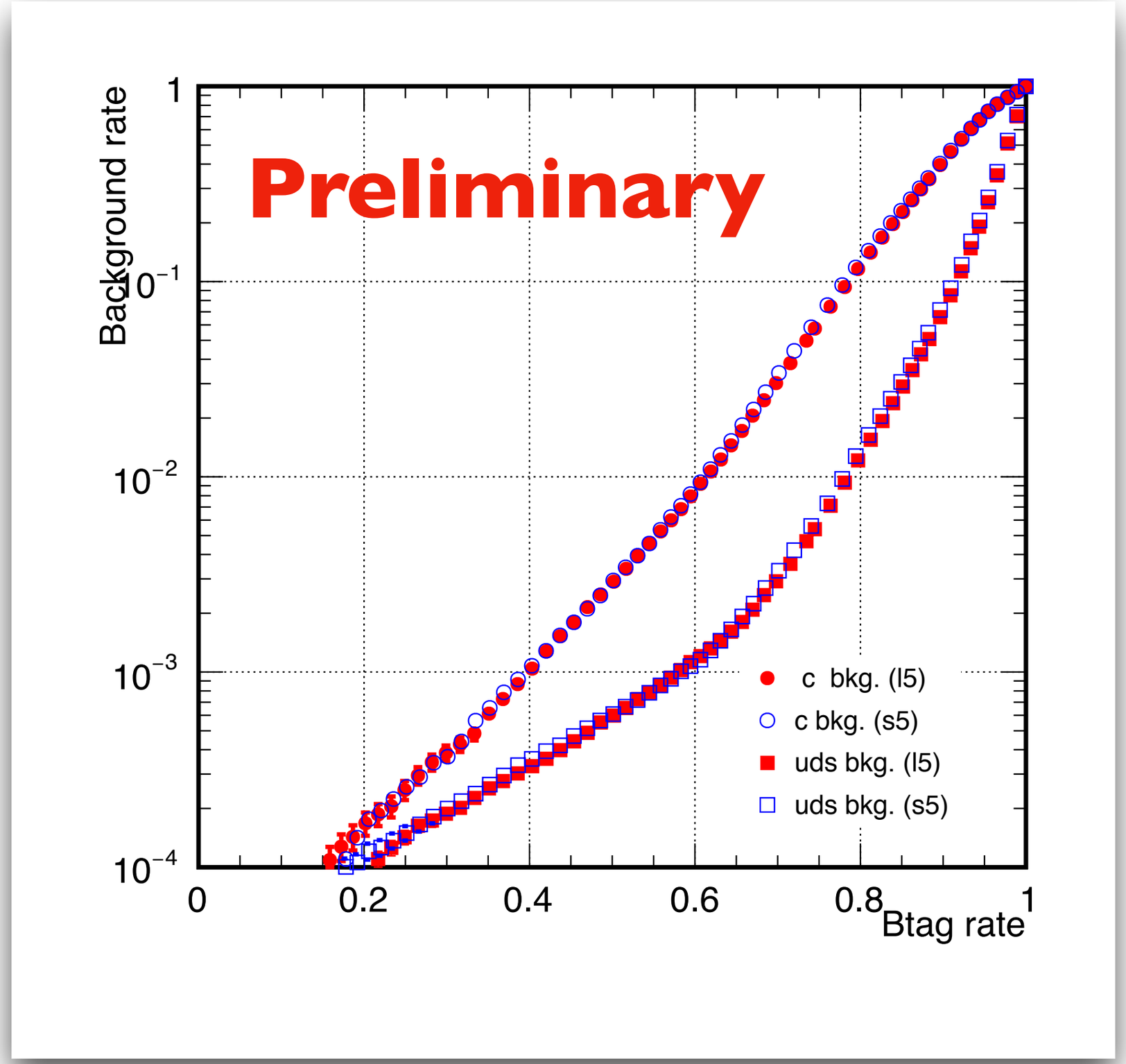


Preliminary

Doesn't look Gaussian... sigma(gauss fit) ~ 10μm

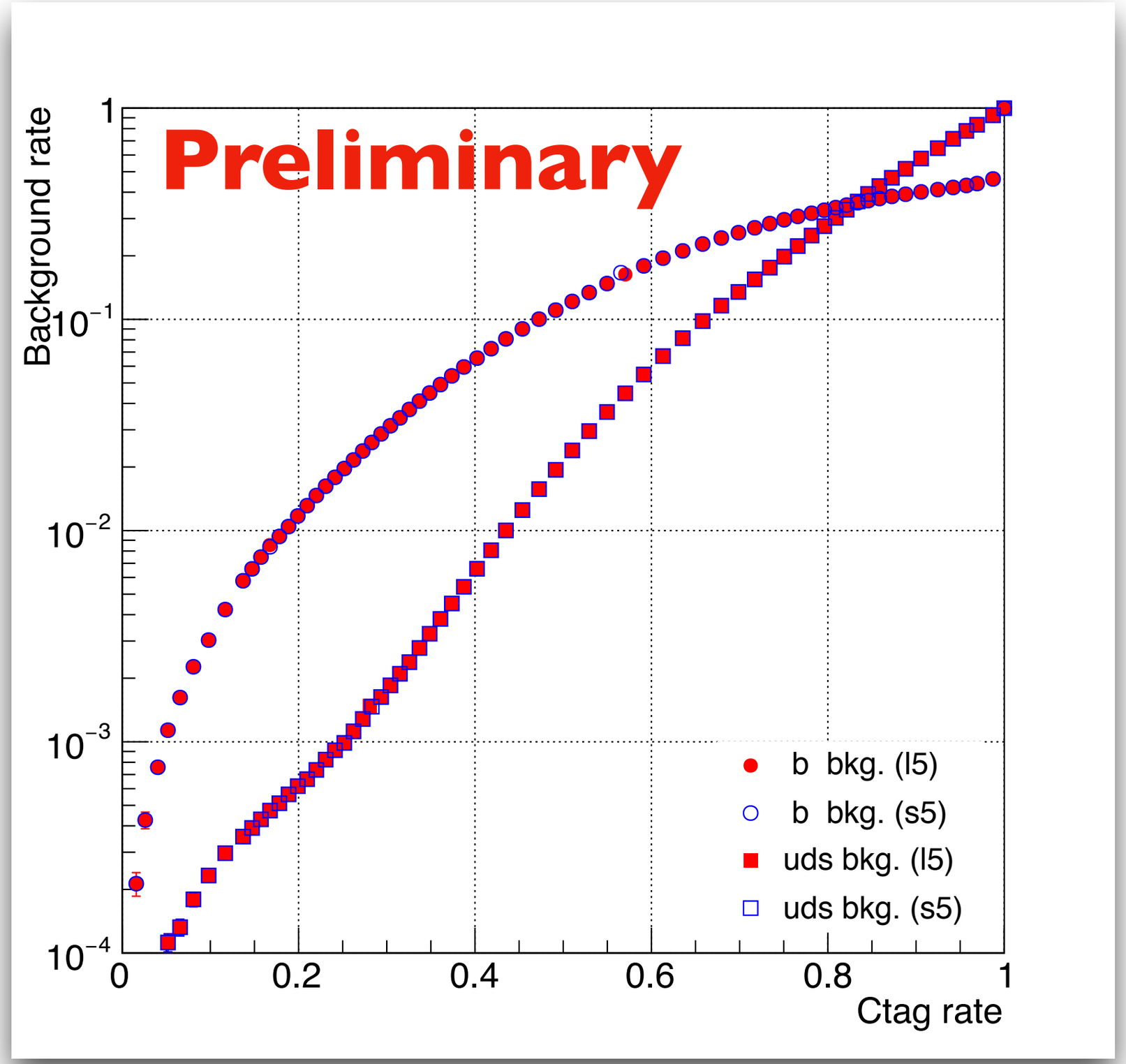
6b, 500 GeV, w/ beam bkg., w/ beam spot constraint

Flavour Tag performance



6b, 500 GeV, w/ beam bkg. **BDT parameter: NMaxDepth=6**

Flavour Tag performance



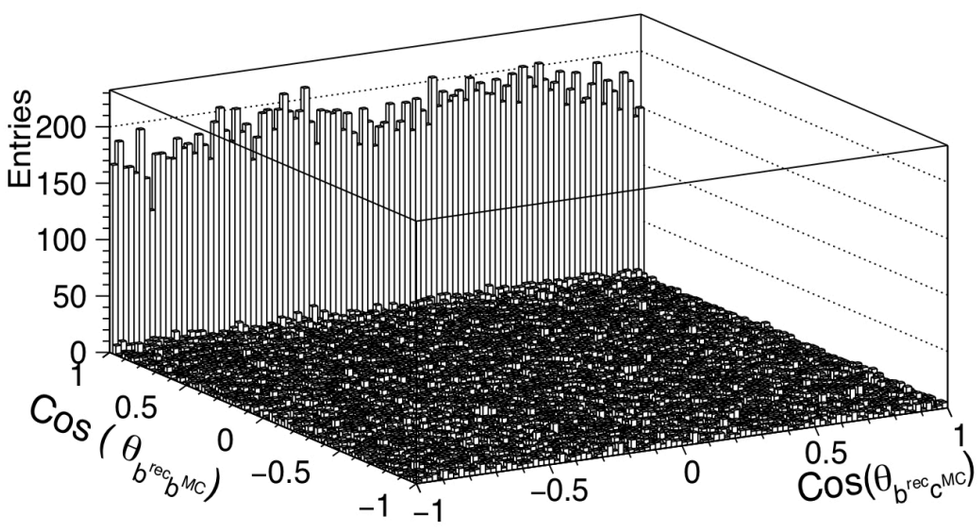
6b, 500 GeV, w/ beam bkg.

New issue

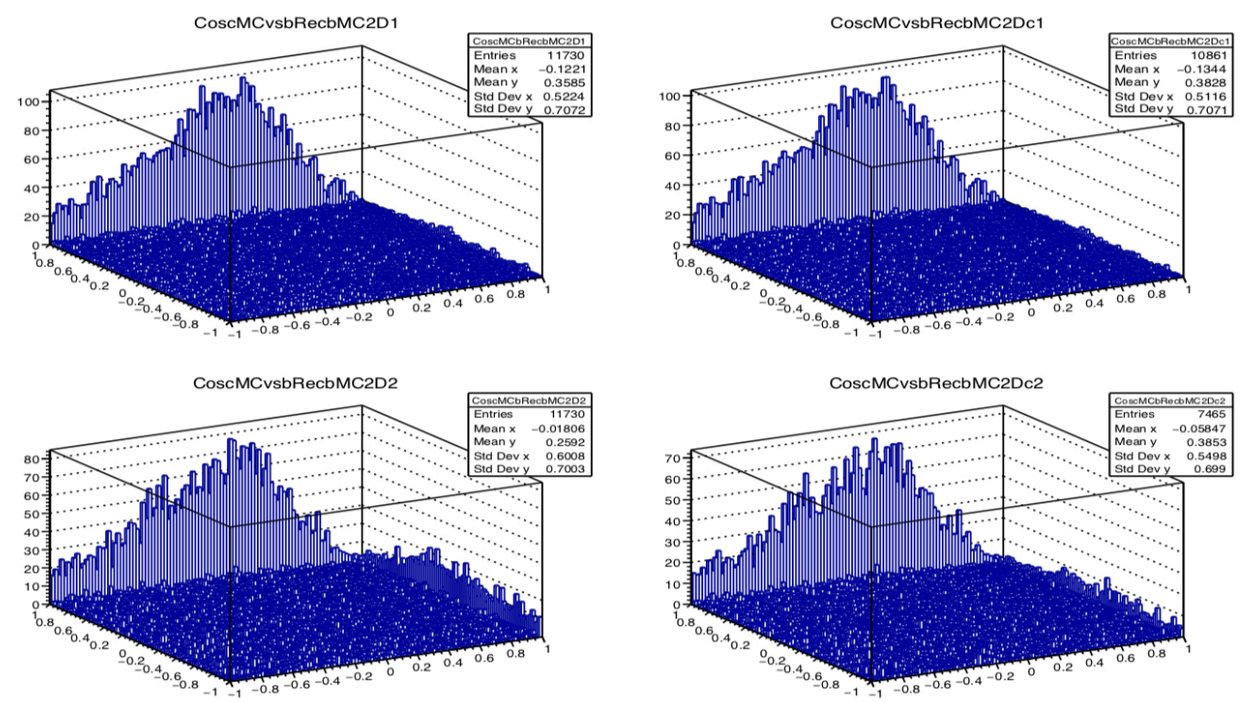
Sohail has reported that he observes a clear discrepancy between DBD and IDR samples (ttbar samples).

b/c mis-tagging (DBD)

A 3D Projection of comparison of the cosine of the angle between reconstructed b-jets and generate b quarks, vs cosine of the angle between reconstructed b-jets and generated c quarks, to see the mis-tagging of c jets as b jets.



After Reprocessing Vertexing - eLpR



I hope we can efficiently get better understanding of this issue making use of this opportunity that the experts have got together.

Conclusion

- **The difference between l5 and s5 is negligible.**

- **New issue reported by Sohail must be understood and solved.**
  - No problems found with the flavour samples.
  - More careful checks are necessary with ttbar samples.
  - Personally I need to study what Sohail does in his code to get any clues.

Backup

Vertex finding

- **Starts from track selection**
  - Define unreliable tracks and will not try to associate them to any vertices.
- **Use Beam spot constraint for Primary vertex finding**
  - Beam spot constraint is powerful to distinguish non-primary vertices.
  - Beam spot size must be specified to use this constraint.
- **Use TearDown algorithm for Primary vertex finding**
  - Make a vertex using all tracks passed the track selection.
  - Compute chi2s from distances between the vertex and each track.
  - Remove tracks that give the highest contribution to the chi2.
  - Repeat until all the tracks satisfy a user-defined chi2 requirement.
- **For Secondary vertex finding, use tracks that are not associated to primary vertex.**
  - Make all possible track pairs, and requiring its invariant mass being less than 10GeV and sum of both track energies.
  - Apply V0 selection (vertex mass, vertex position etc.)
  - Attach additional tracks to the vertices if possible.

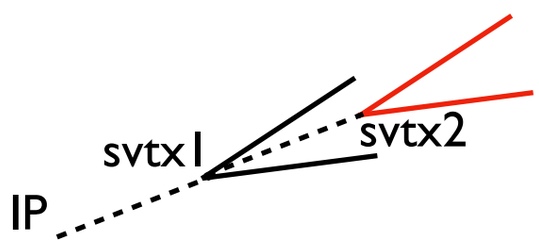
Vertex finding performance

Fractions of tracks associated to three types of secondary vertices

	Track origin	Primary	Bottom	Charm	Others
	Total number of tracks	496897	258299	247352	56432
cat.1	Tracks in secondary vertices	0.6%	57.5%	64.3%	2.5%
cat.2	... from the same decay chain	—	56.6%	63.4%	1.9%
cat.3	... from the same parent particle	—	32.2%	38.9%	1.2%

ILD sample of bbbar events with sqrt(s)=91.2GeV.

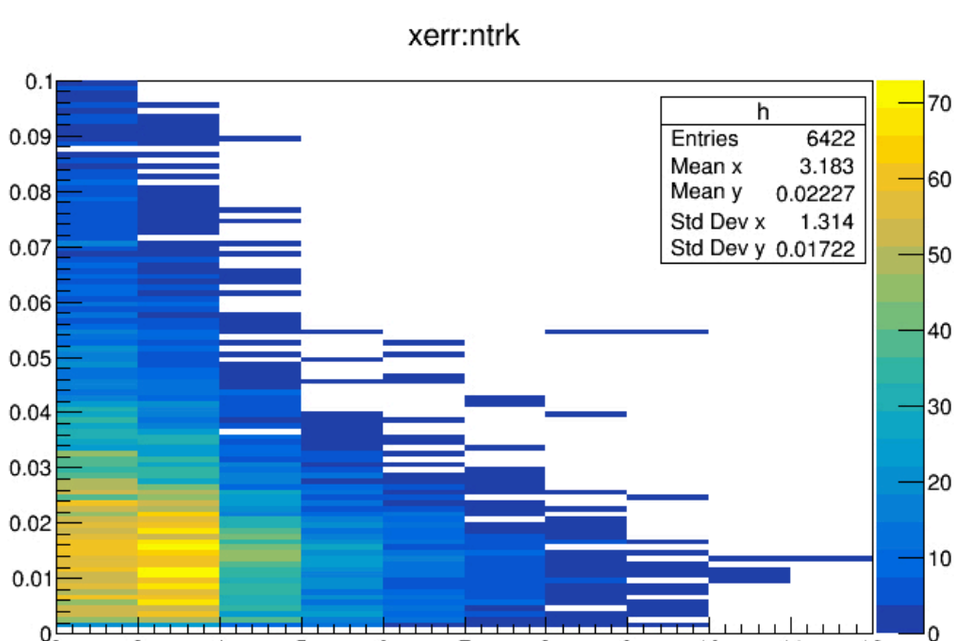
T. Suehara, T. Tanabe, "LCFIPlus: A Framework for Jet Analysis in Linear Collider Studies", NIM A 808 (2016) 109-116



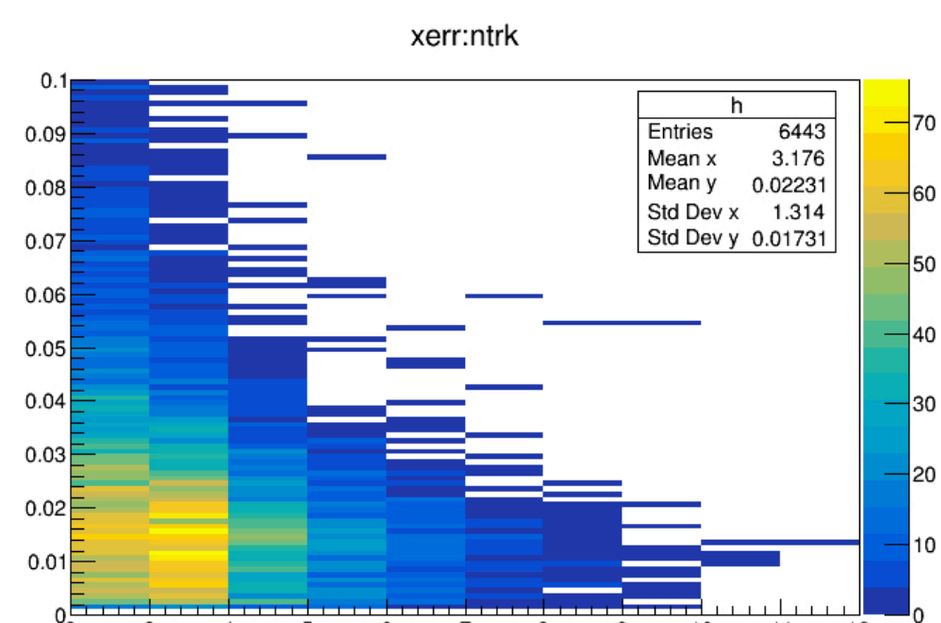
If a red track is associated to svtx1 or svtx2, this track is categorized into cat.2.
if a red track is associated to svtx2, this track is categorized into cat.3.
—> A drop from cat.2 to cat.3 indicates confusion of these two vertices.

Position (x) resolution w.r.t # of vtx tracks (Secondary vertex)

(w/o beam bkg overlay)



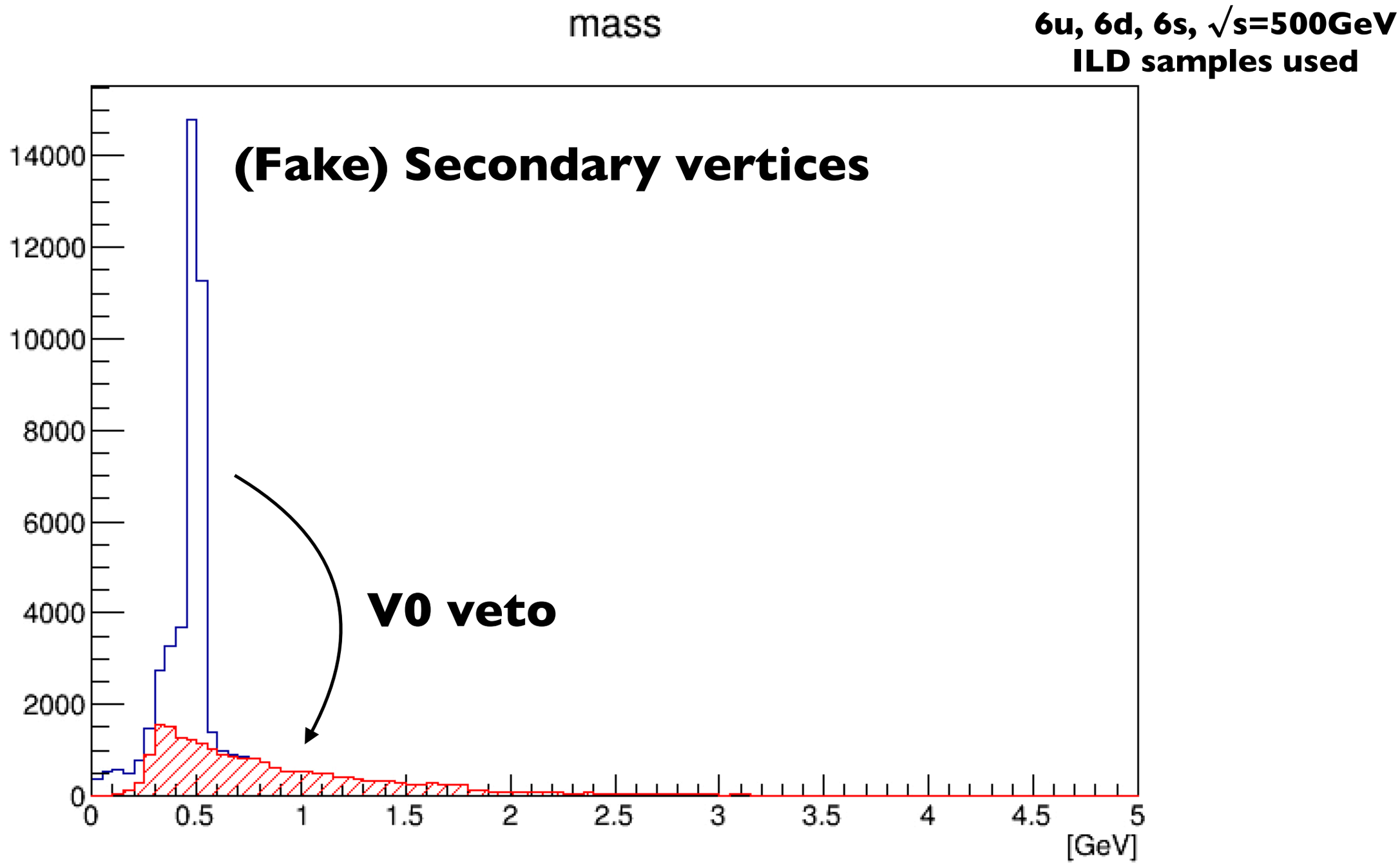
(w/ beam bkg overlay)



/s=500GeV, 6b sample

V0 rejection

We remove two-track vertices that are consistent with K_s, Λ^0, photon conversions (V0 vertices), because V0 vertices mimic B- or C-hadron vertices.



Jet clustering

- **Define jet cores by secondary vertices or leptons, and combine nearest jet cores until the required number of jets are obtained.**
  - We do not want to merge the jet cores any further. Will set α = 100 when 2 jet-cores are being combined in (modified) clustering algorithms.
- **Attach remaining tracks and neutral particles to one of the jet cores by using following jet algorithms.**

- **Built-in jet algorithms in LCFIPlus**
  - Durham
  - Kt
  - Valencia

  } **standard version**

  - DurhamVertex
  - KtVertex
  - ValenciaVertex

  } **modified version**

  Intend to protect jet core structures —> Effective for multi-jet events

Modified y value (DurhamVertex) :

$$Y(i,j) = \frac{2\min(E_i, E_j)^2 (1 - \cos\theta_{ij})}{Q^2} + \alpha$$

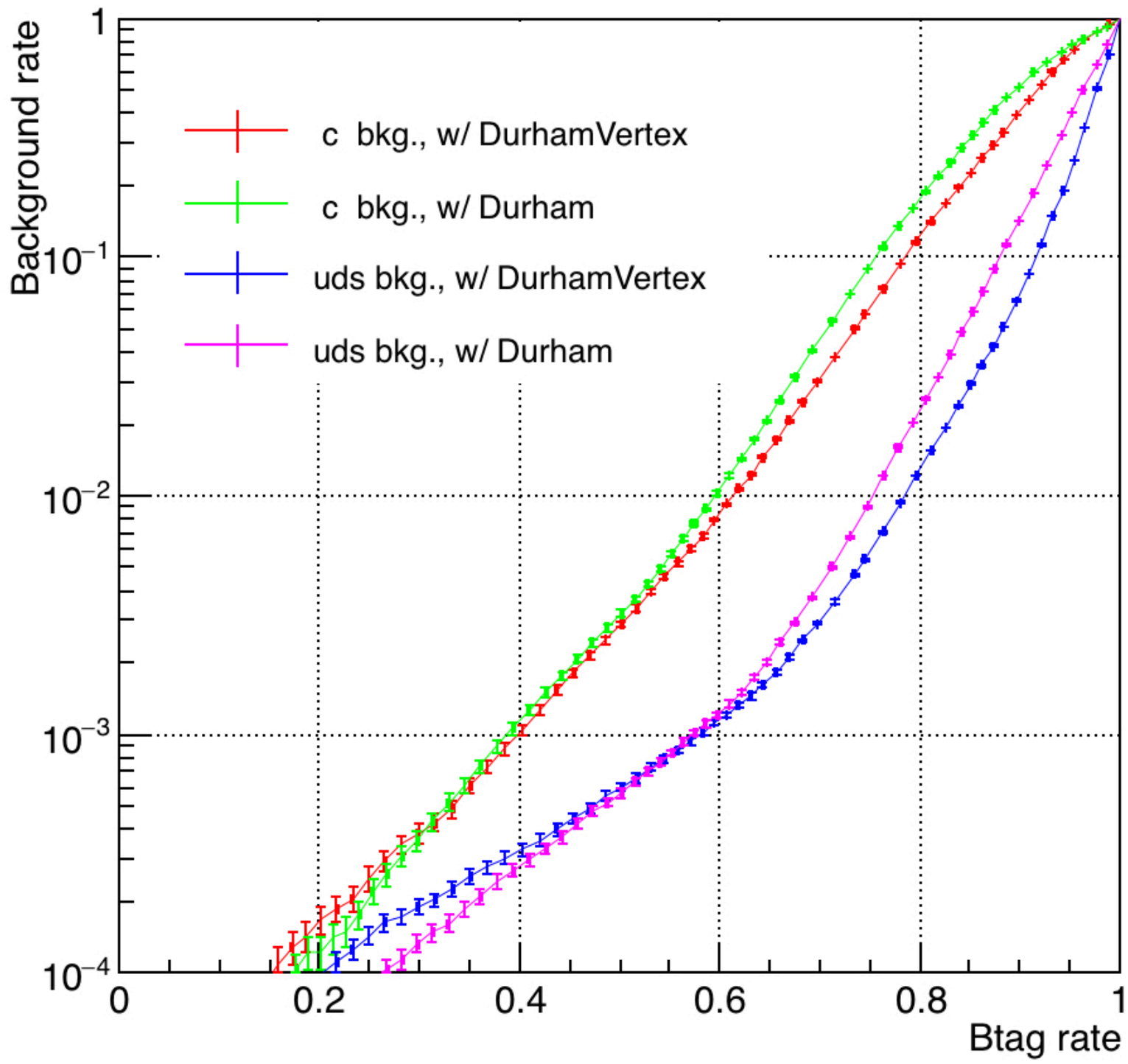
E_i,E_j : Jet energy
θ_{ij} : angle b/w Jets
Q : $\sqrt{s}$
α=0, 100

- **Jet collections produced by external packages can also be used instead of using jet clustering in LCFIPlus.**

However built-in jet algorithms that use vertex information are recommended.

Jet clustering effect on flavour tagging

Comparison between Durham and DurhamVertex

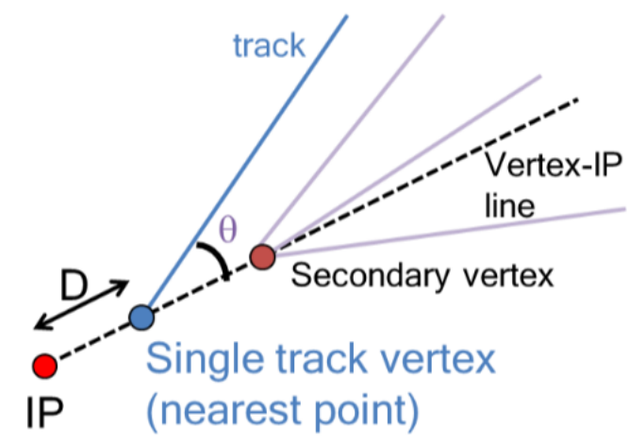


6b, 6c, 6q
$\sqrt{s}$=500GeV
ILD(I5) sample used

Jet vertex refiner

- **Re-vertexing but now using jet information**
  - More than one secondary vertex in a jet implies a b-jet.
  - Useful for b-c separation.
  - Try to improve the efficiency of secondary vertex reconstruction.

- **Pseudo vertex : Single track vertex**
  - If one secondary vertex is found in a jet and if there is a track whose trajectory comes near a point collinear to the primary and secondary vertices, it is defined as pseudo vertex, unless the track is tagged as a primary track.



- **For each vertex in a jet, compute chi2 again to all tracks and check if there is any possibility to refine vertex reconstruction.**

Jet vertex refiner performance

w/o vertex refiner
(w/ vertex refiner)

	(# of vtx , # of pseudo-vtx)	6 b	6 c	6 q
b-jet signature	(2 , 0)	22.35 % (23.92%)	0.45% (0.42%)	0.05% (0.06%)
	(1 , 1)	2.18% (17.78%)	0.15% (1.42%)	0.00% (0.06%)
c-jet signature	(1 , 0)	53.09% (36.84%)	42.67% (41.94%)	0.97% (1.20%)

6b, 6c, 6q $\sqrt{s}$=500GeV
ILD(I5) sample used

The main effect of vertex refining is the recovery of vertices and pseudo-vertices: as a result, b jets migrate into the b-like category (2,0) and (1,1) from the c-like category (1,0), which helps with b/c separation.

Flavour tagging

- **Based on multi-variate analysis ("BoostedDecisionTrees")**
  - input variables : impact parameters, track multiplicity, vertex mass, etc.

- **For efficient training, 4 jet-categories are used.**
  - 0 vertex jet —> light flavour like
  - 1 vertex jet —> c like
  - 2 vertex jet (pseudo vertex = 1) —> b like
  - 2+ vertex jet —> b like

- **We typically offer training samples for different energies and different jet multiplicities. For the best performance, the analyst should compare the different weight files.**

  examples:
  - 91GeV, 2b, 2c, 2q sample
  - 500GeV, 6b, 6c, 6q samples
  - 1TeV, 6b, 6c, 6q samples

Update list

1. Migrated to ROOT6. ROOT >= 6.08.00 required.

2. Adaptive Vertex Fitting Under testing

- currently relatively strict track selection is applied to prevent spoiling vertex reconstruction with fake tracks.
- try to loosen the track selection while keeping fake track rate low by introducing a weight.

3. BNess for better track selection Under testing

- identifying tracks from B-hadron using MVA.

4. Vertex Mass Recovery for better B/C separation Under testing

- Pi0 reconstruction

5. Fix related to the IP smearing.

- Some MVA variables assumed IP=(0,0,0).

6. Position errors on primary vertex.

- Fit parameters for primary vertex slightly modified (—> fraction of fitting failures on primary vertex was reduced. No changes for secondary vertex finding.).