L2 b-tagging
Using IP Chi-Square Probability

Ariel Schwartzman
(SLAC)

SLAC ATLAS Forum, 18-Apr-2007

Outline

Introduction.

Primary vertex reconstruction.

Impact parameter studies and RoI angular resolution.

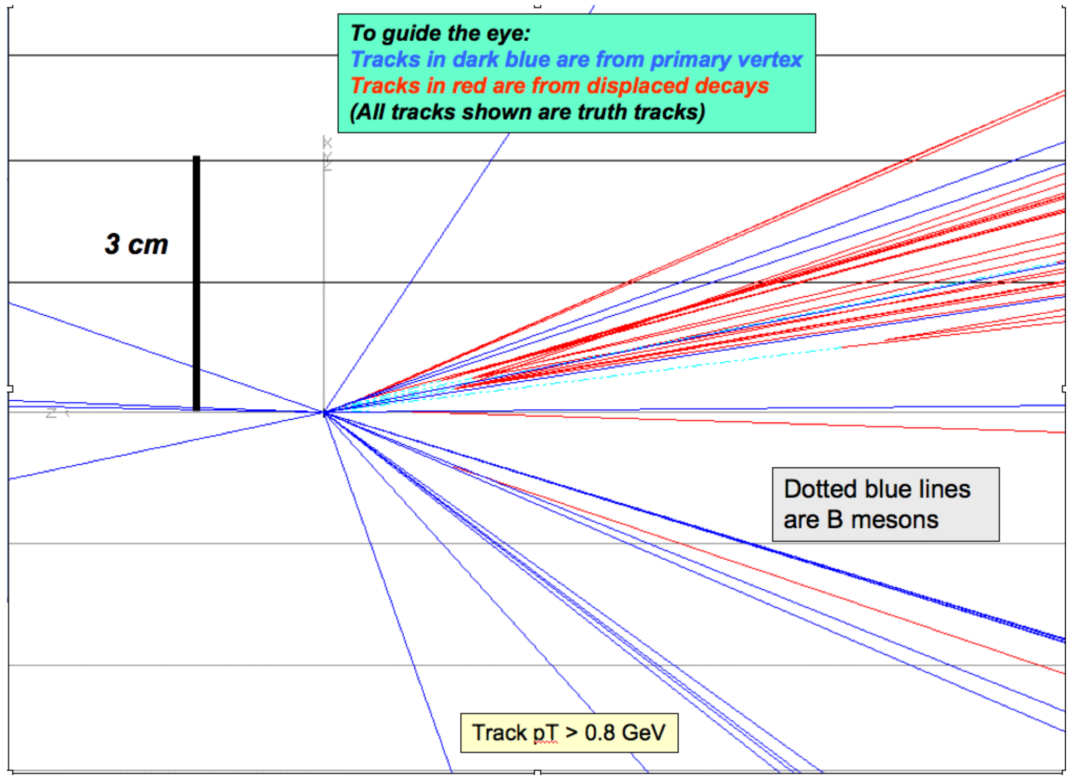
Track Jet reconstruction at L2.

Chi-Square probability b-tagging algorithm.

Performance.

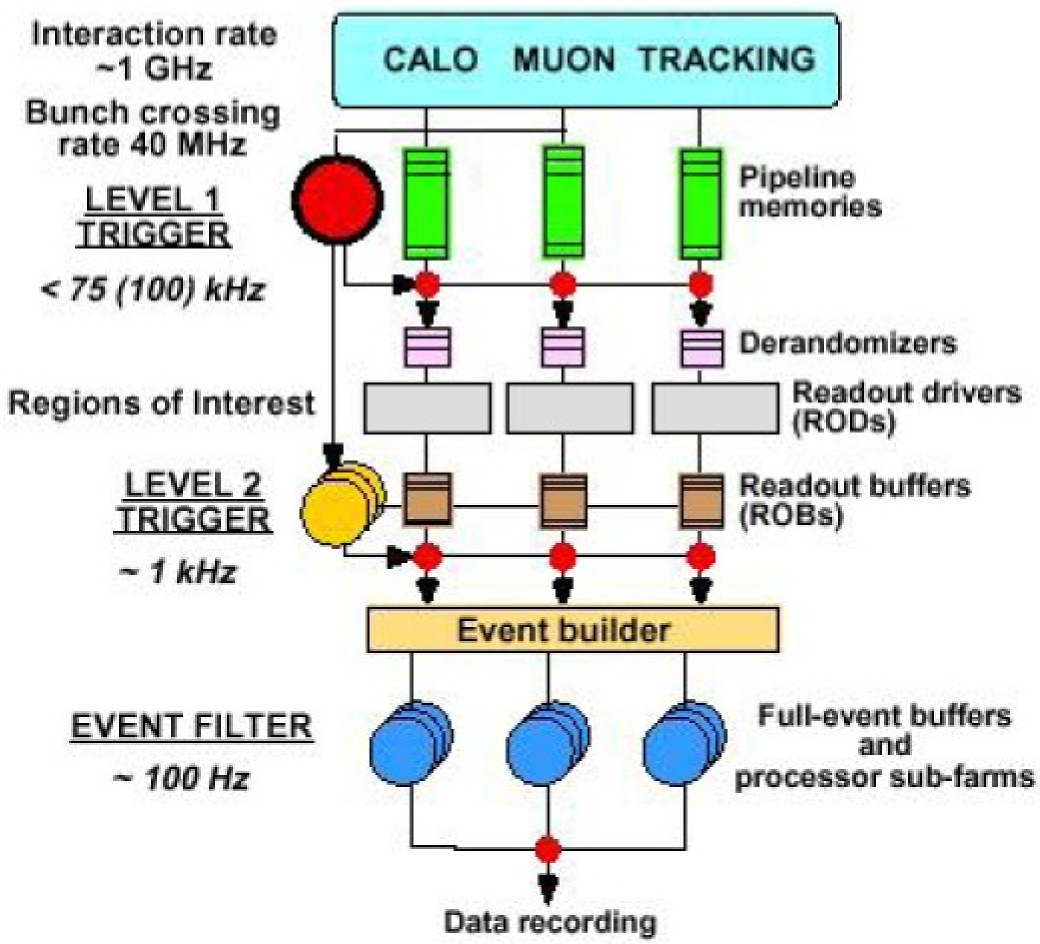
Introduction

Physics signatures with many b-jets (ttH)
Long lived particles decaying into b bbar (displaced jets)
Multiple vertices, often more than 1 vertex per jet
Allow to lower jet pT threshold.



Matthew Strassler (UW/Rome)

Track Reconstruction at Level 2 (I)



Inner Detector data readout only at L2.

L1: hardware trigger

L2 starts from Region of Interests (RoIs) defined by L1.
Latency of 10ms.

Tracking algorithms at LV2 should be fast and access the smallest amount of data needed.

Track Reconstruction at Level 2 (II)

SiTrack Algorithm:

Form space points from a given RoI, starring from B layer.

Form track seeds made of 2 space points, fitted with an straight line (track segments) (pt>2GeV tracks)

Estimate the primary vertex z position by histogramming the z intersection of the seed with the beam line. Keep more than one candidate.

Extend track seeds with a third SP, and remove ambiguities based on the fit quality.

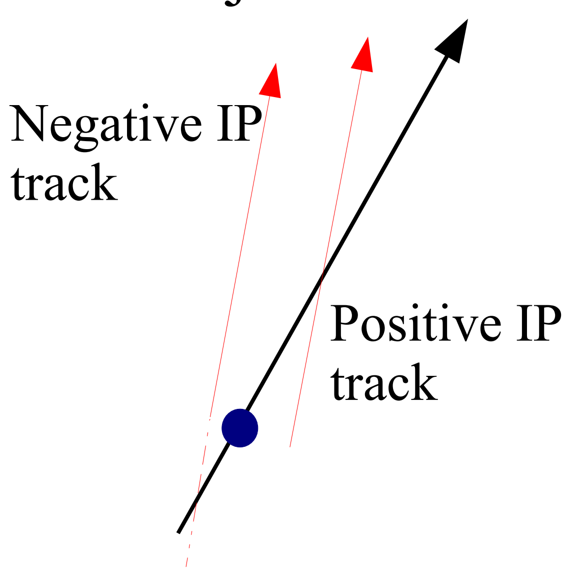
3D fit of candidate tracks, extended in roads to SCT.

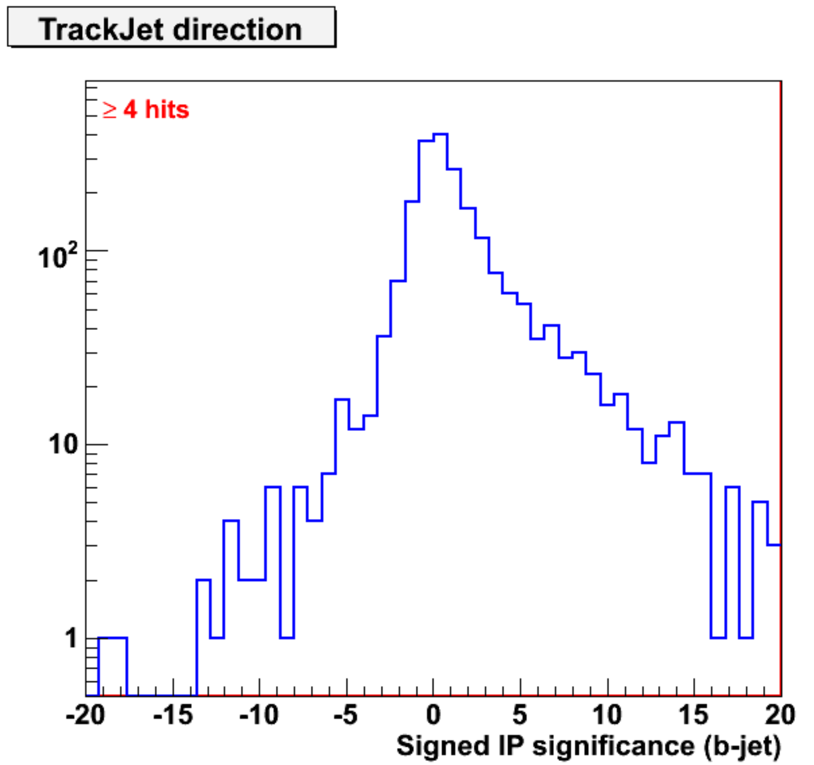
Method originally developed in ALEPH, and used by Dzero and CDF.

IP Chi-Square Probability b-Tagging

Tag b-quarks based on the Impact parameter of tracks pointing to jets

Fake displaced tracks from light jets have a symmetric distribution of IP around 0 (due to resolution), whereas tracks from b-jets will have a positive tail

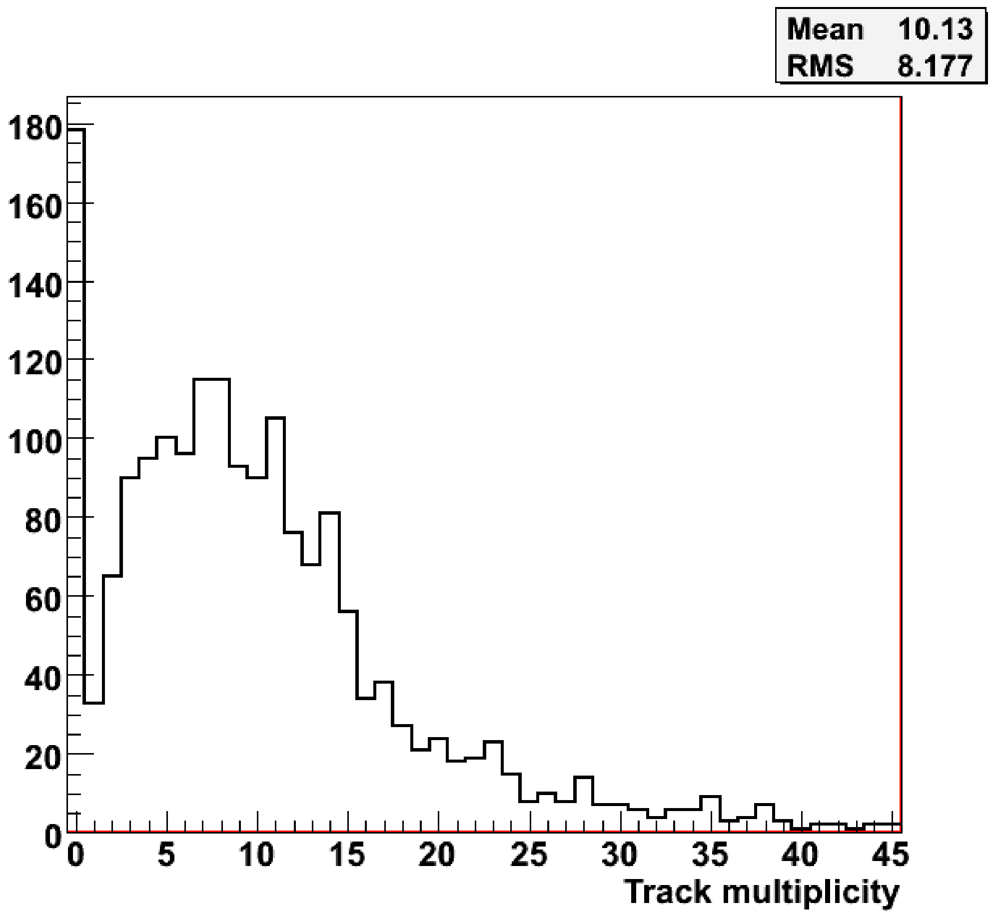


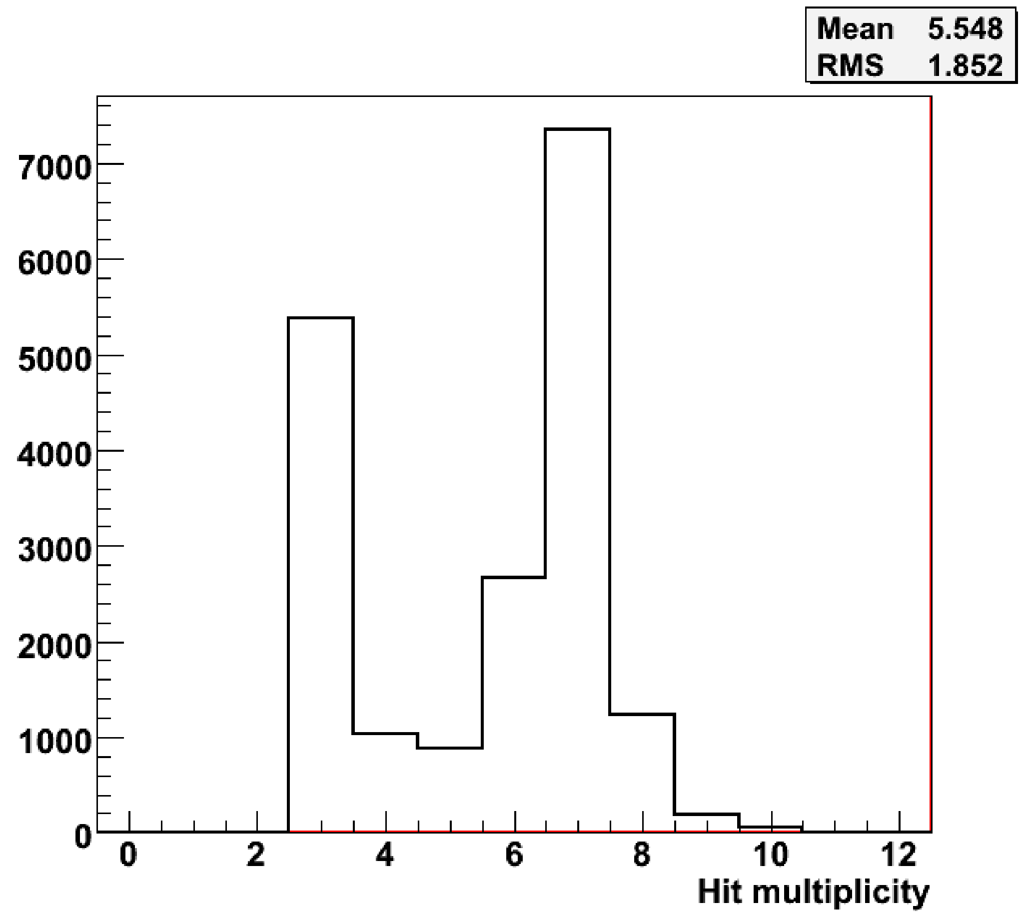


Use the negative side of the IP distribution to compute a track and jet IP chi-square probability: probability for a track (jet) to originate from the primary vertex.

Keys: 1- Good angular resolution of tracks and jets (to determine sign)
2- Good IP resolution (L2 tracking and primary vertex)

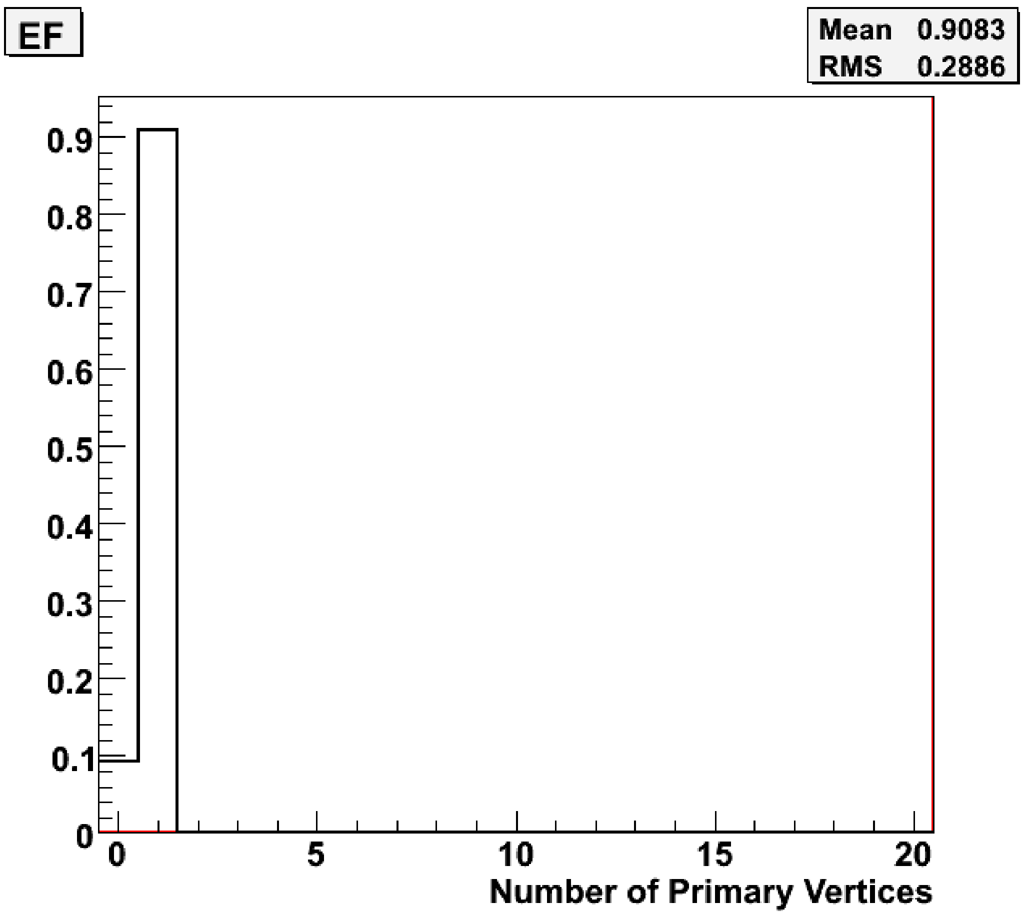
SiTrack Kinematics (I)

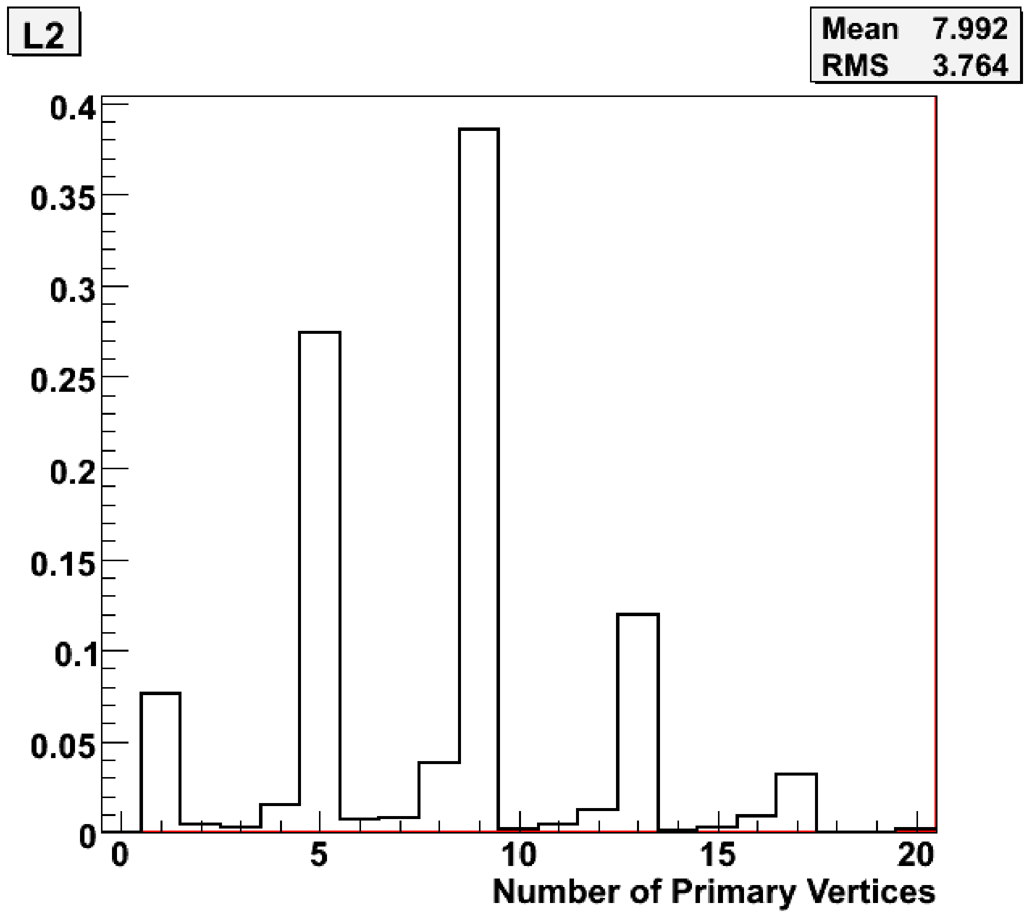




Distribution for all L2 tracks in the event (independently of RoI direction)
Hit multiplicity includes both Pixel and SCT spatial points.

Primary Vertex Studies





Whereas vertexing at EF finds one vertex, L2 hit-based vertexing reconstructs many primary vertex candidates, for each RoI.

New L2 Primary Vertex Algorithms

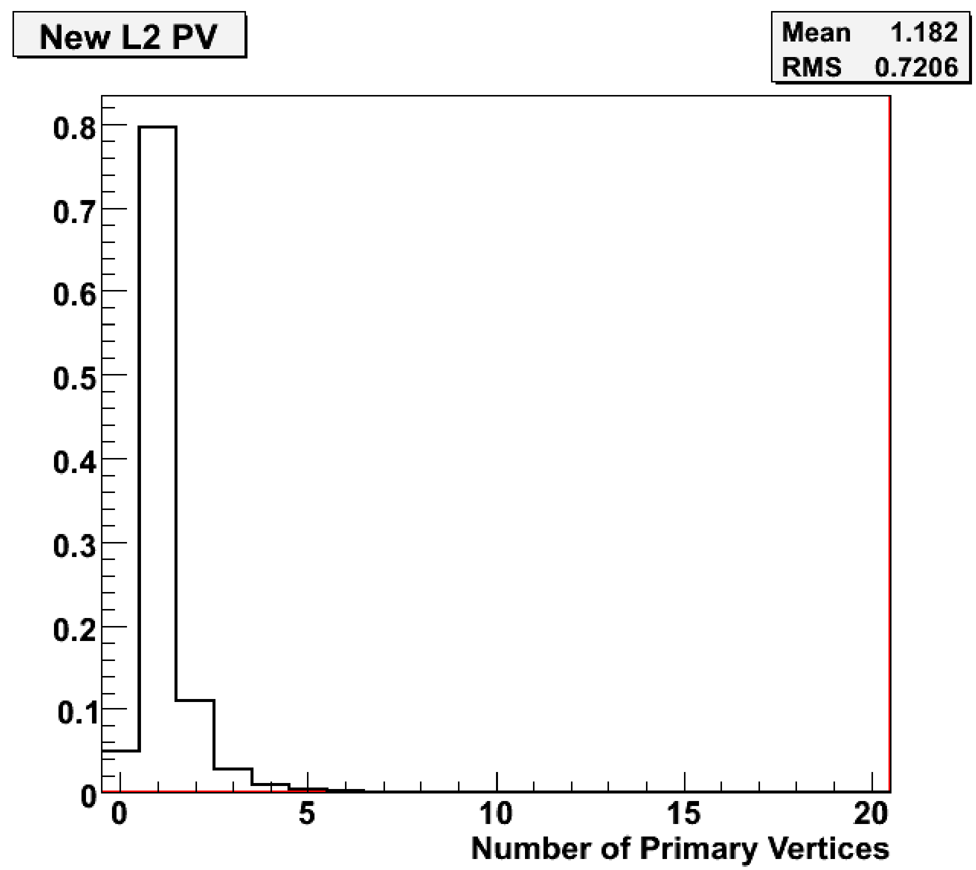
1- Histogramming L2 Primary Vertex (Andrea/Fabrizio)

Uses a Histogramming technique with sliding window.
Vertex ordering based on the track multiplicity.

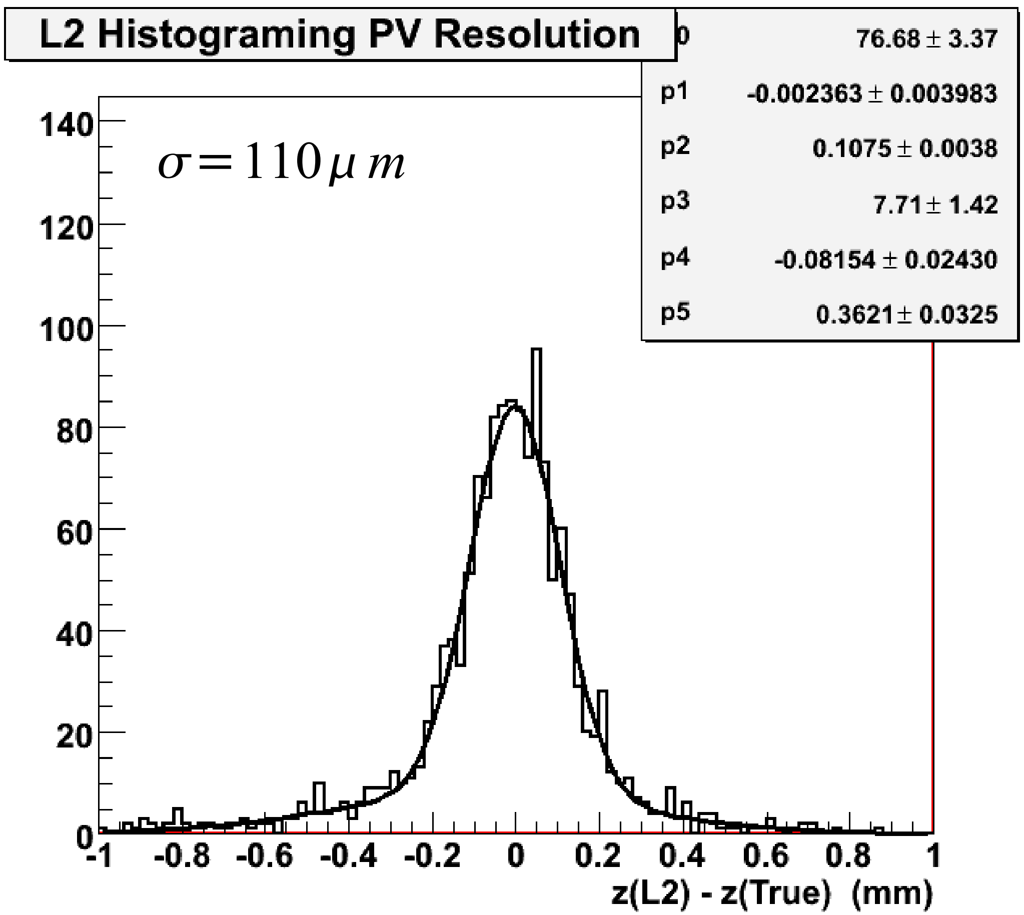
2- Z-Clustering:

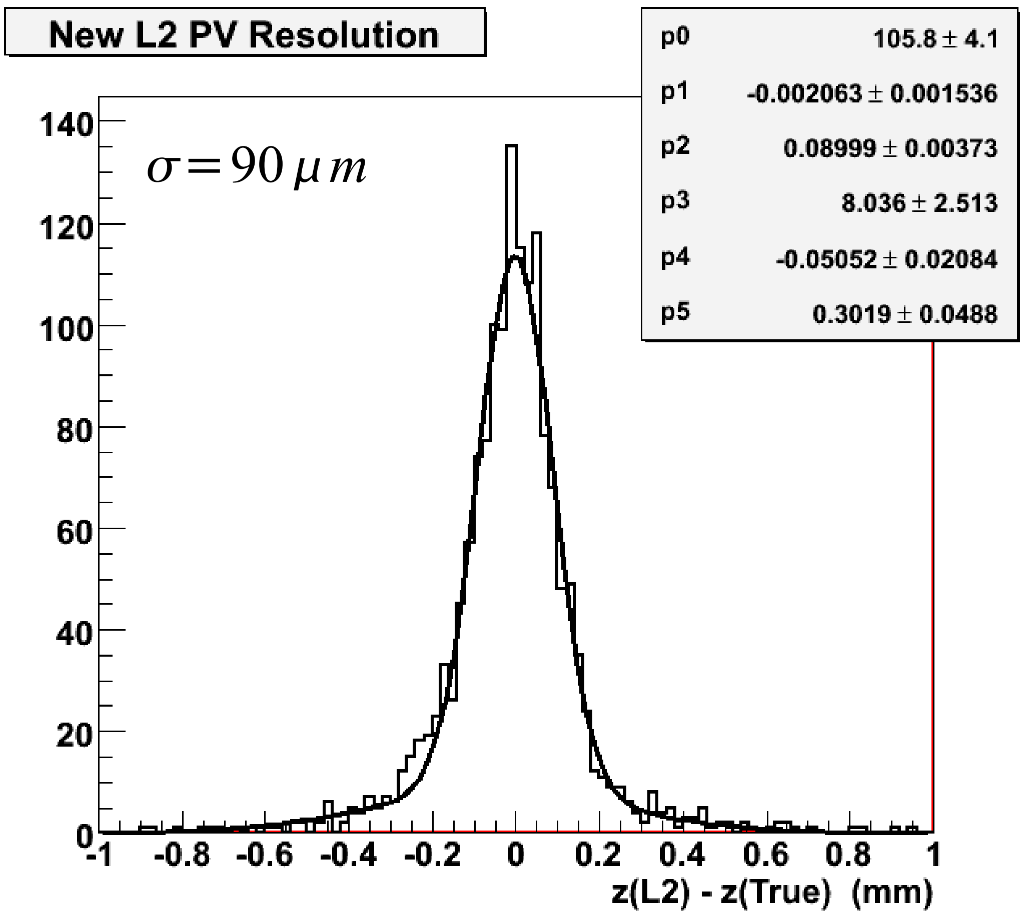
Clustering of tracks in the z plane, pT weighted determination of the z vertex position.
Selection based on PTsum.

The plan is to compare both techniques and possibly merge them into an improved algorithm.



New L2 Primary Vertex Algorithm (II)

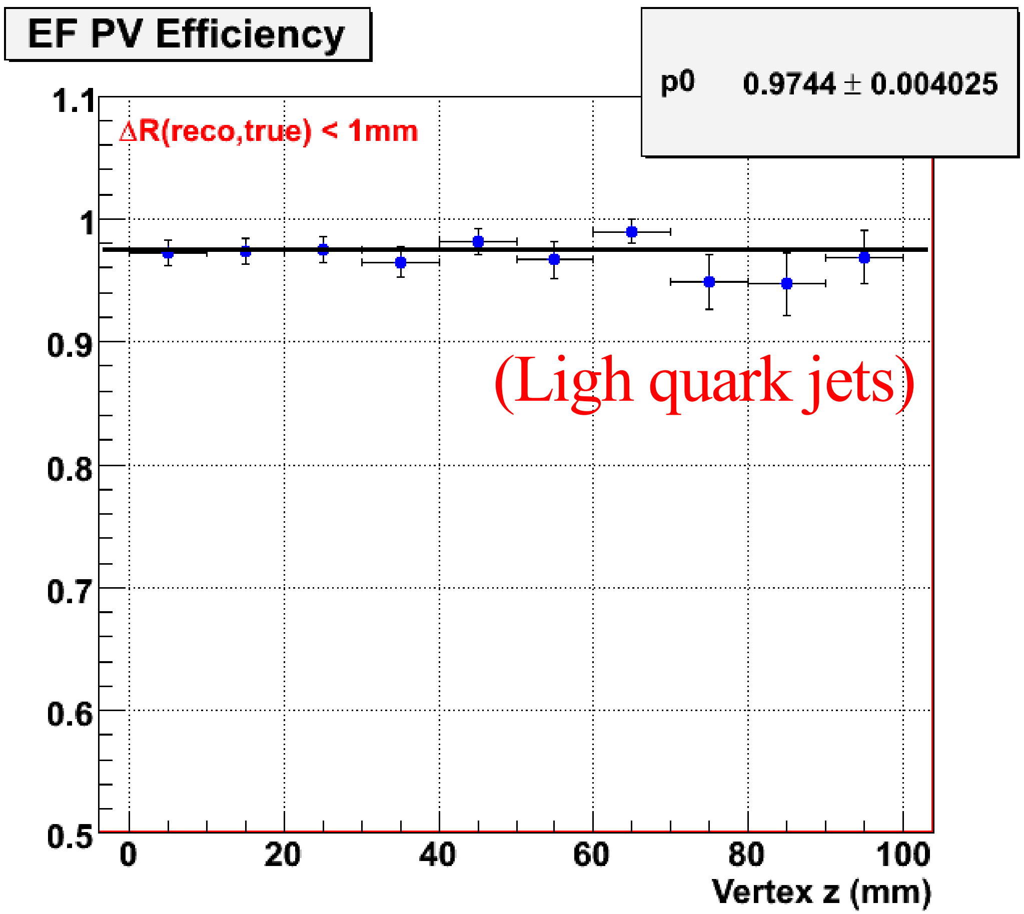


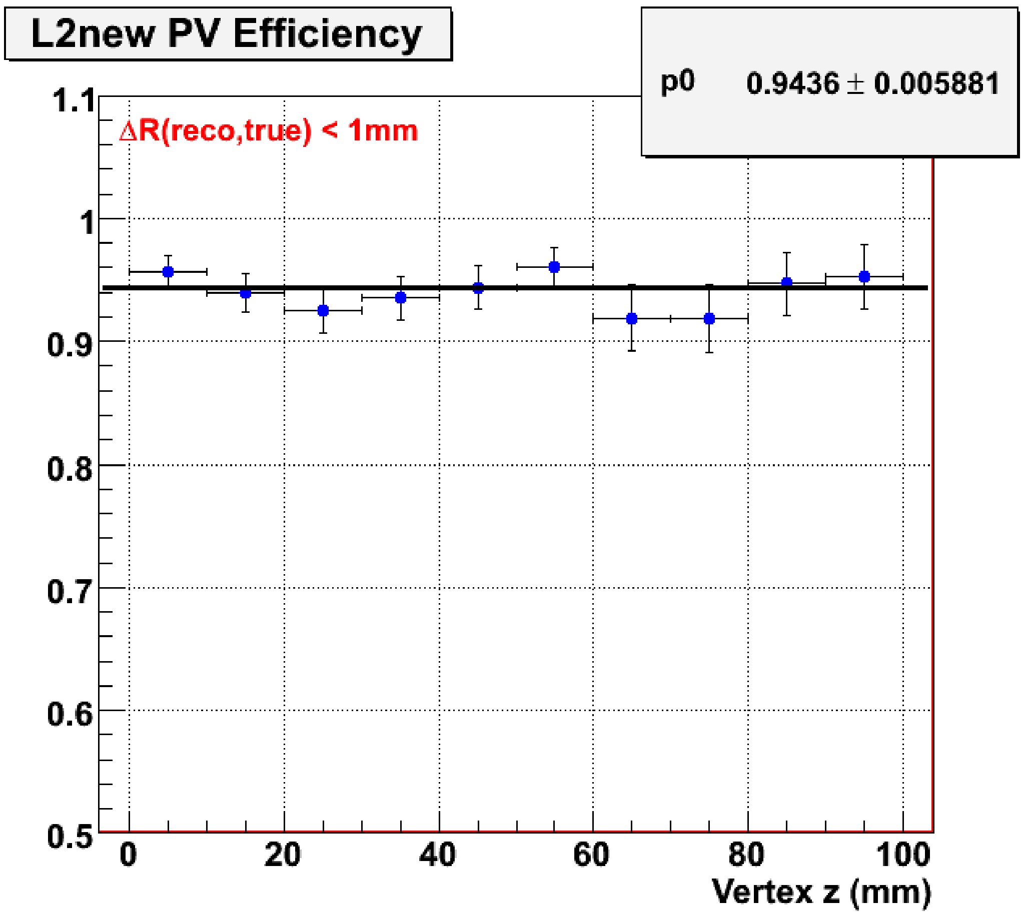


Z-Clustering resolution is smaller because it uses all tracks in the event. Expect similar improvement from Histograming.

20% Resolution improvement with respect to default (hit-based) PV.

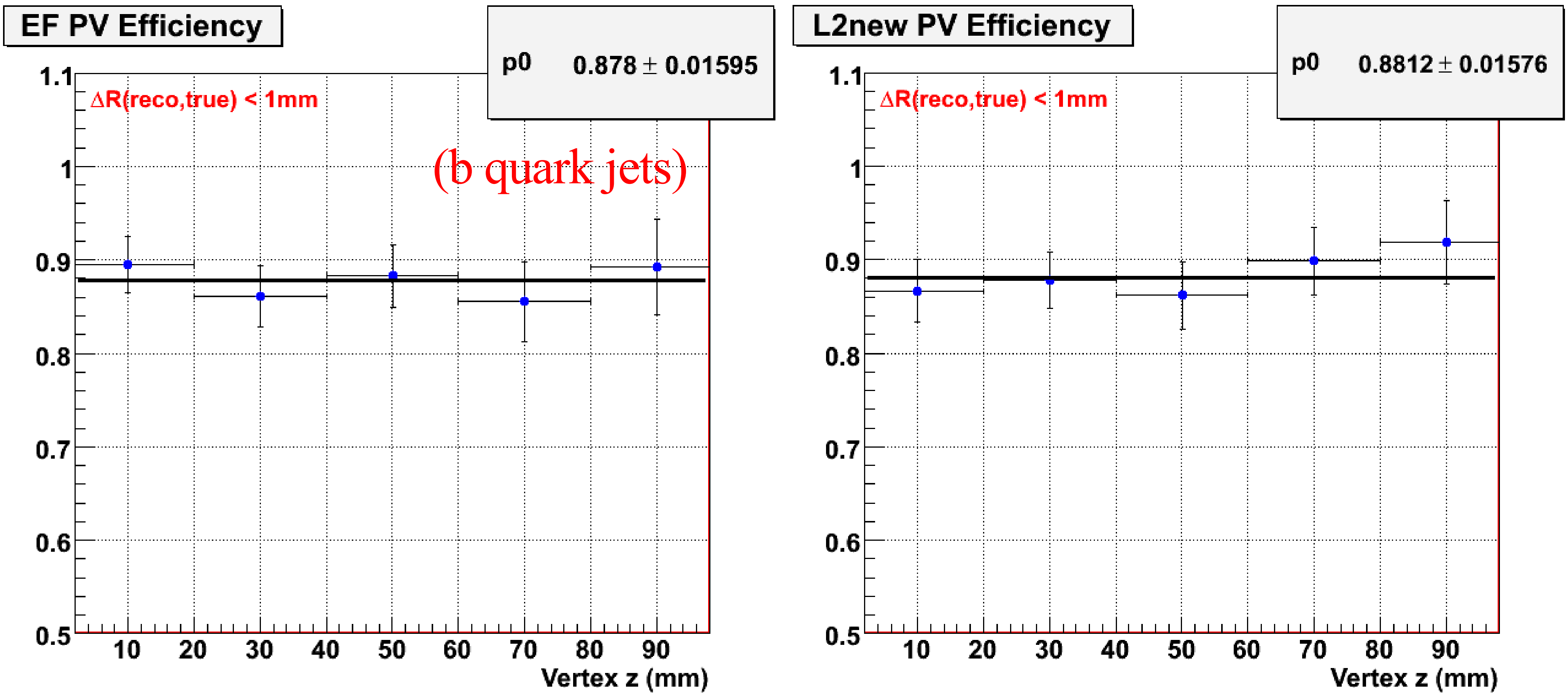
New L2 Primary Vertex Algorithm (III)





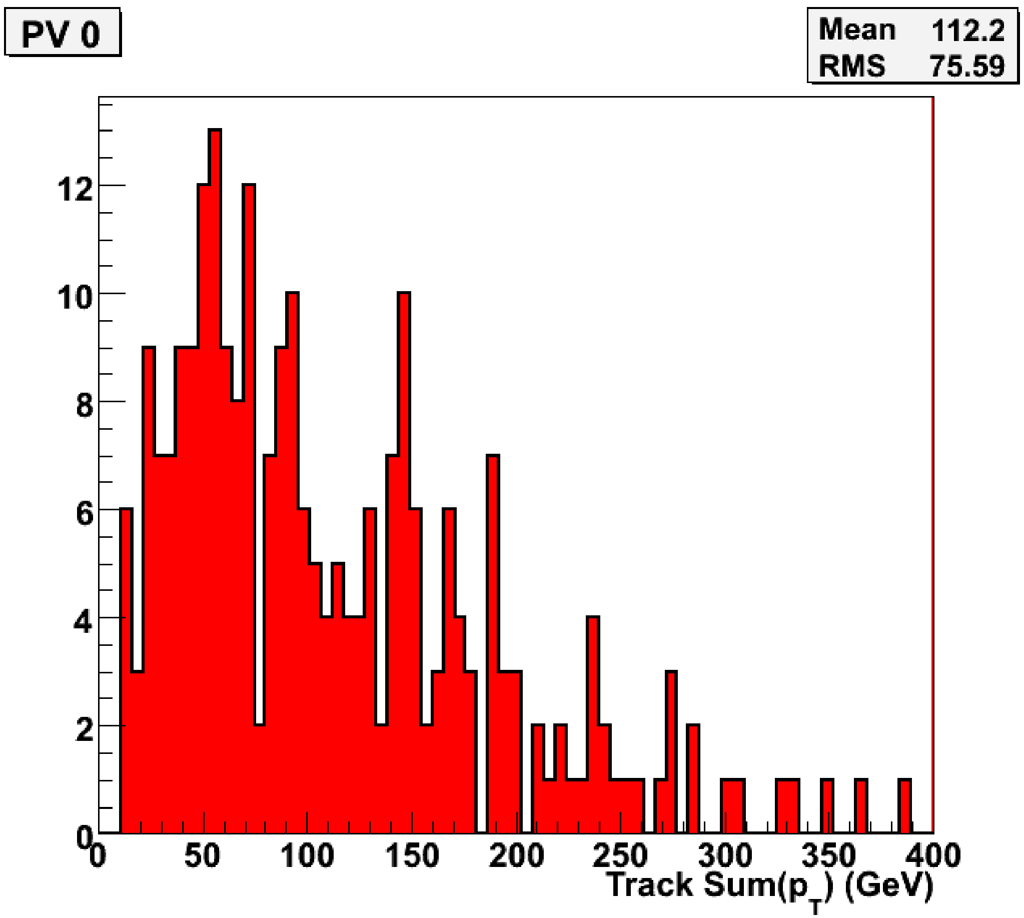
L2 PV efficiency close to 95%.

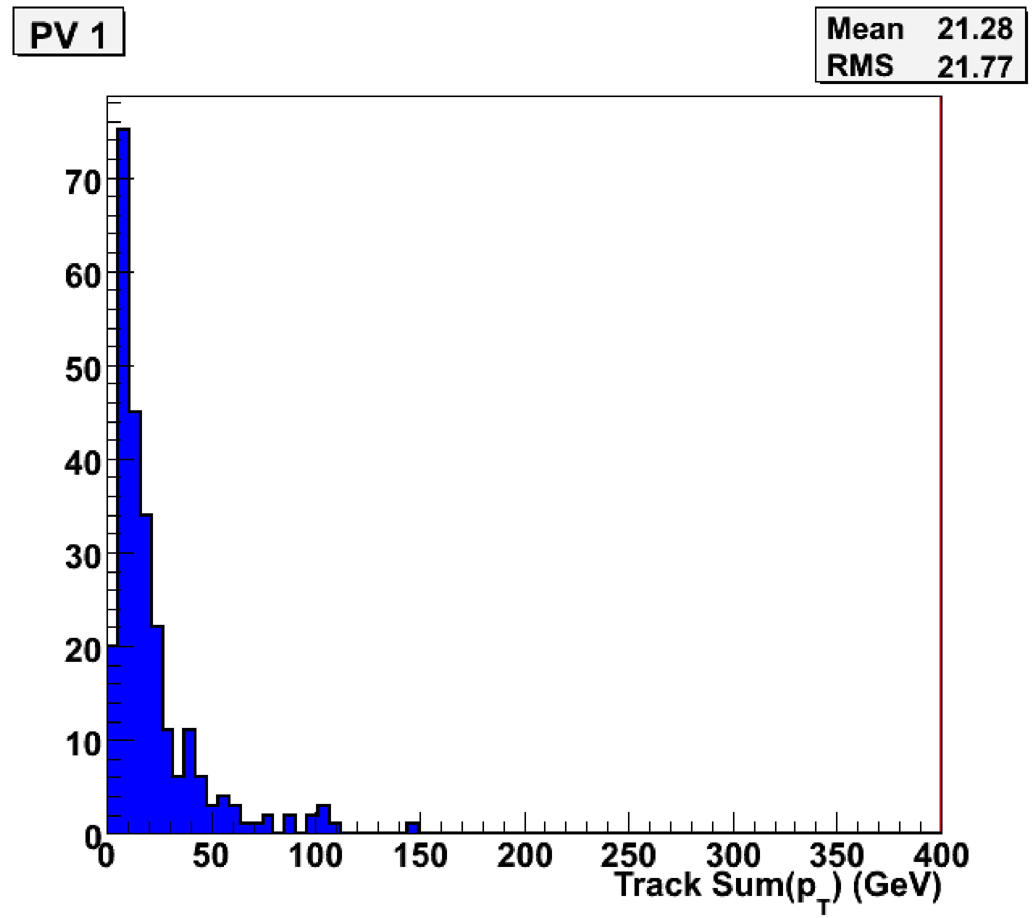
New L2 Primary Vertex Algorithm (III)



L2 PV efficiency close to 90%. Both EF and L2 have lower efficiency in a sample with 2 b-jets

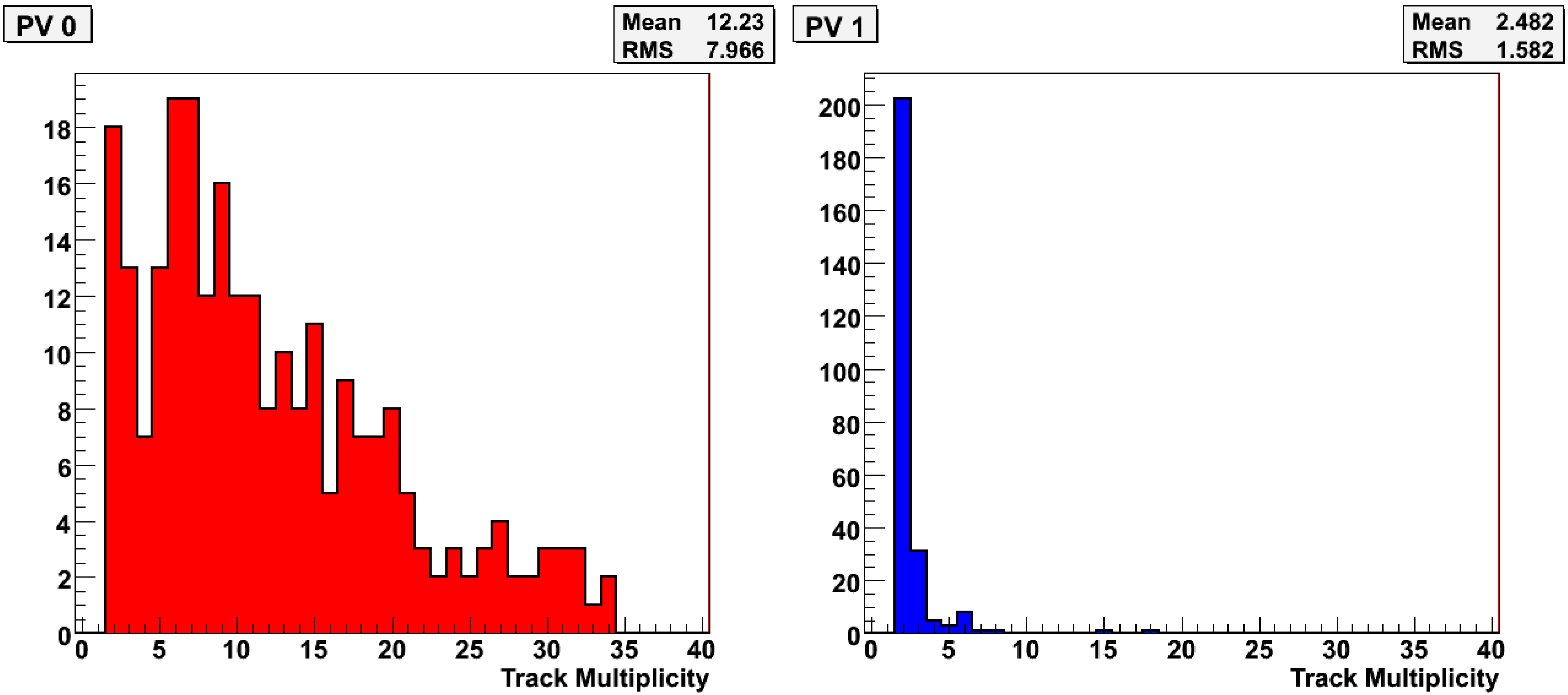
Primary Vertex Selection (I)





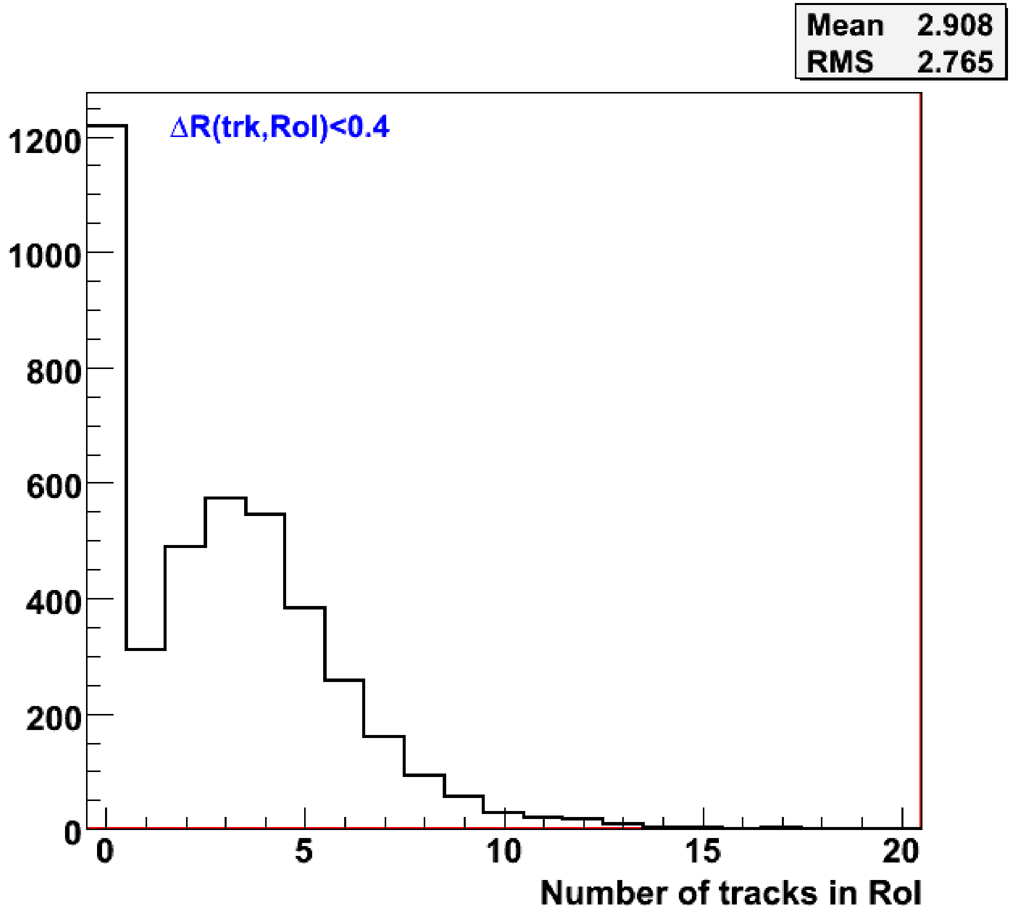
Order vertices by PTsum. Choose the vertex with the largest PTsum as the hard scatter vertex.

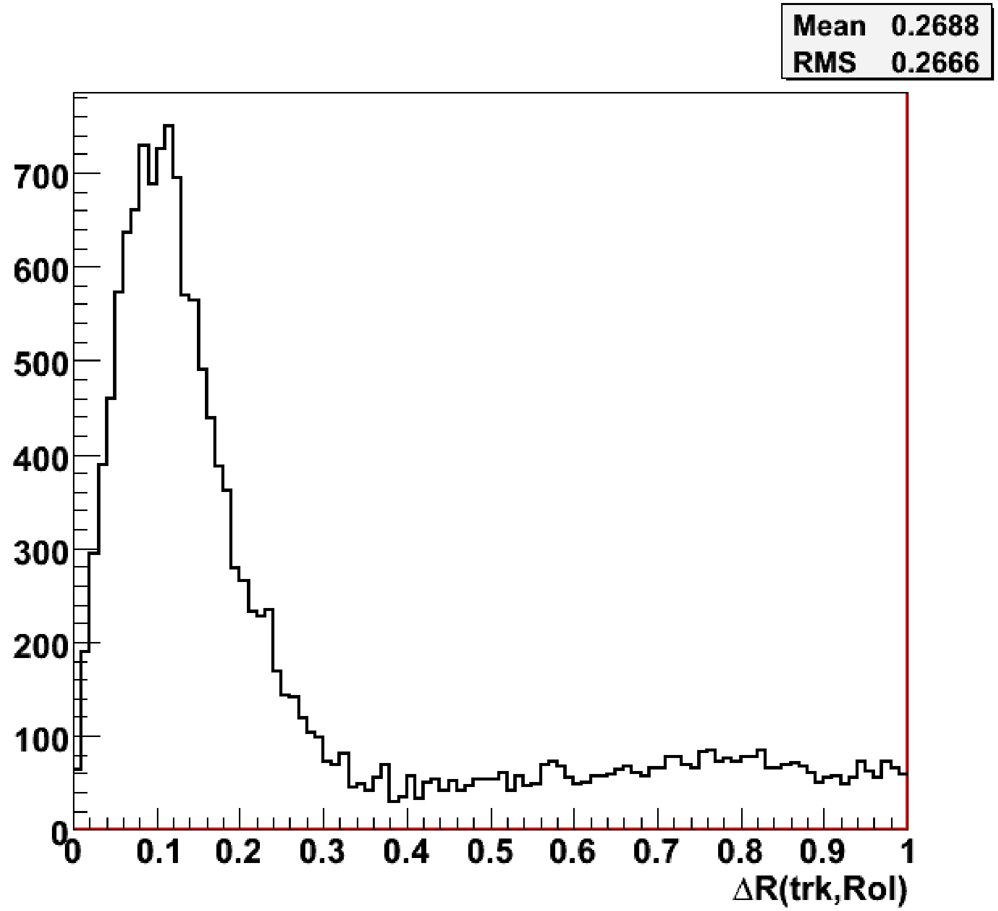
Primary Vertex Selection (II)



Additional vertices are mostly made of 2 tracks.
Can be improved by requiring tighter quality cuts to tracks.

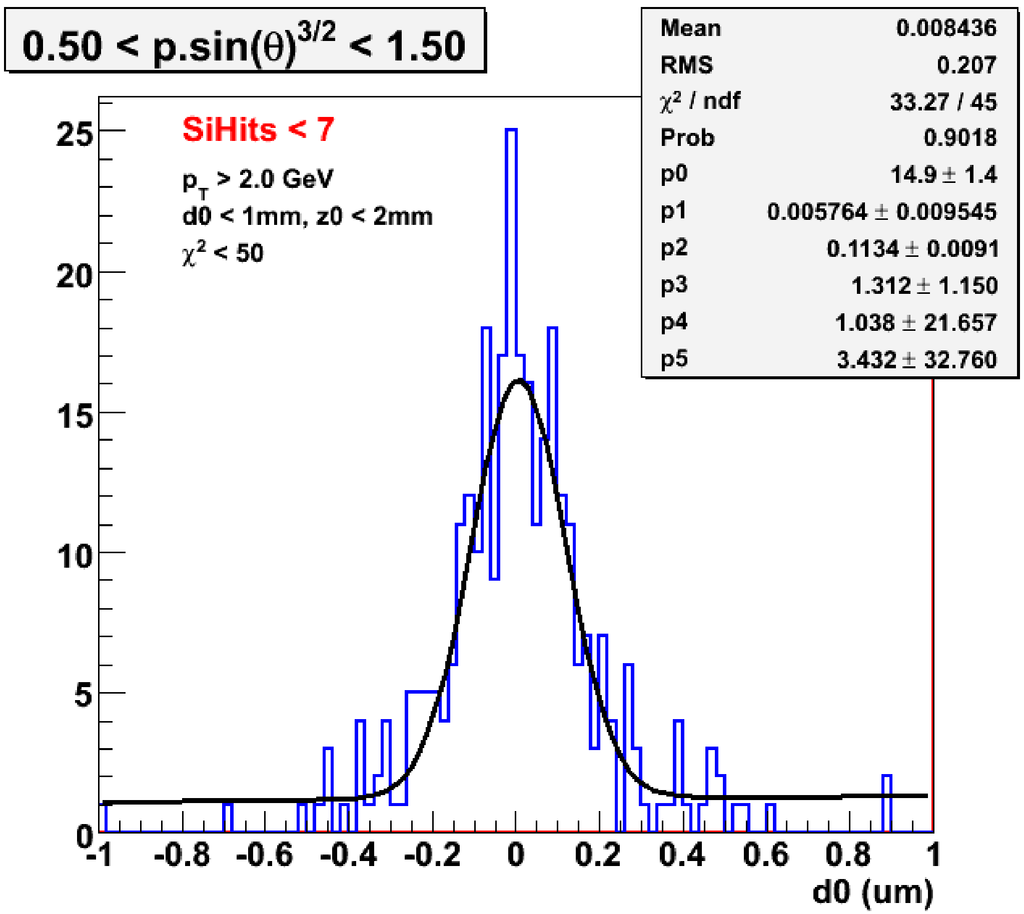
Matching Tracks to RoIs (I)

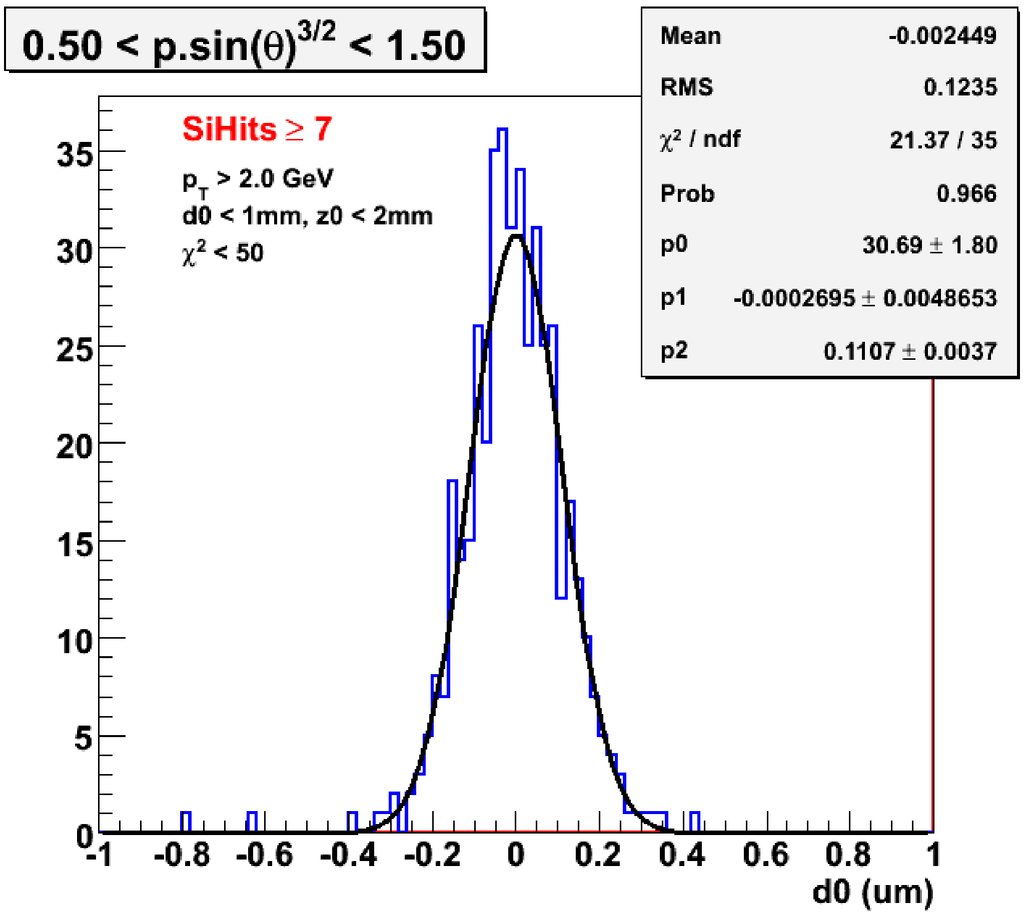




Track pT>2.0GeV, |d0|<1mm, DZ(trk,PV)<1mm.

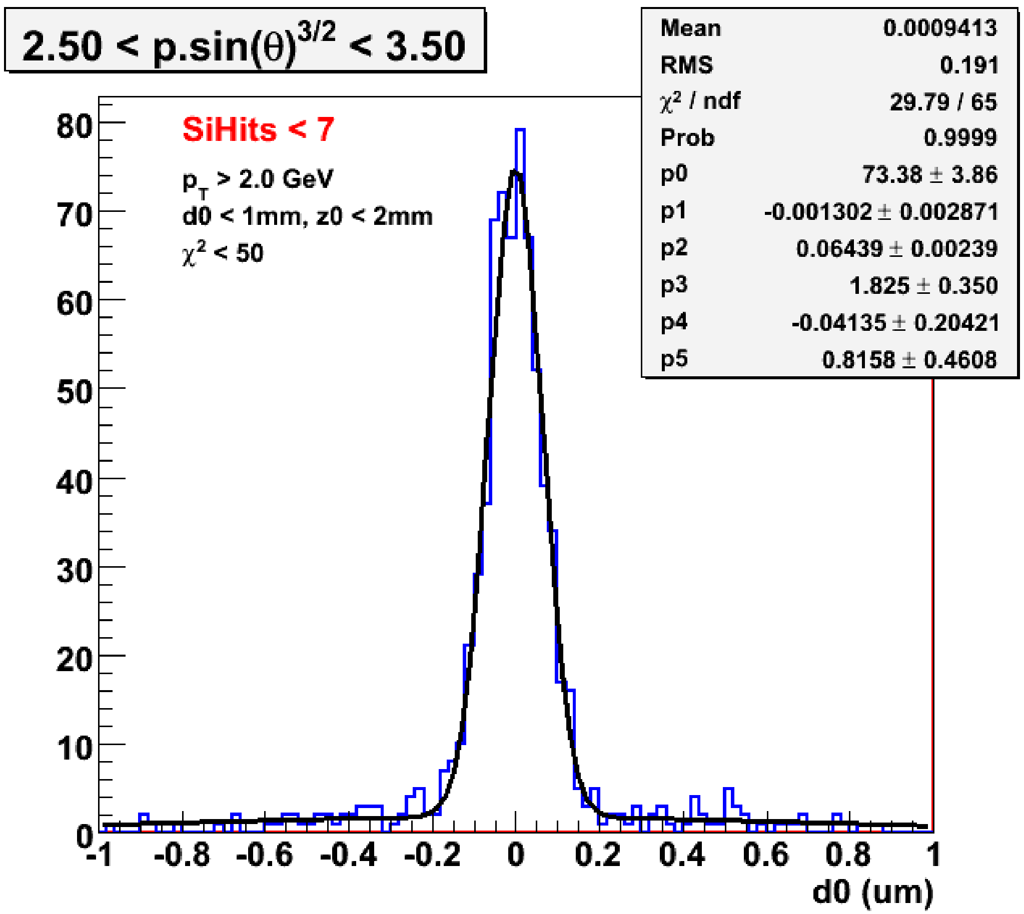
Parameterization of Transverse IP Error

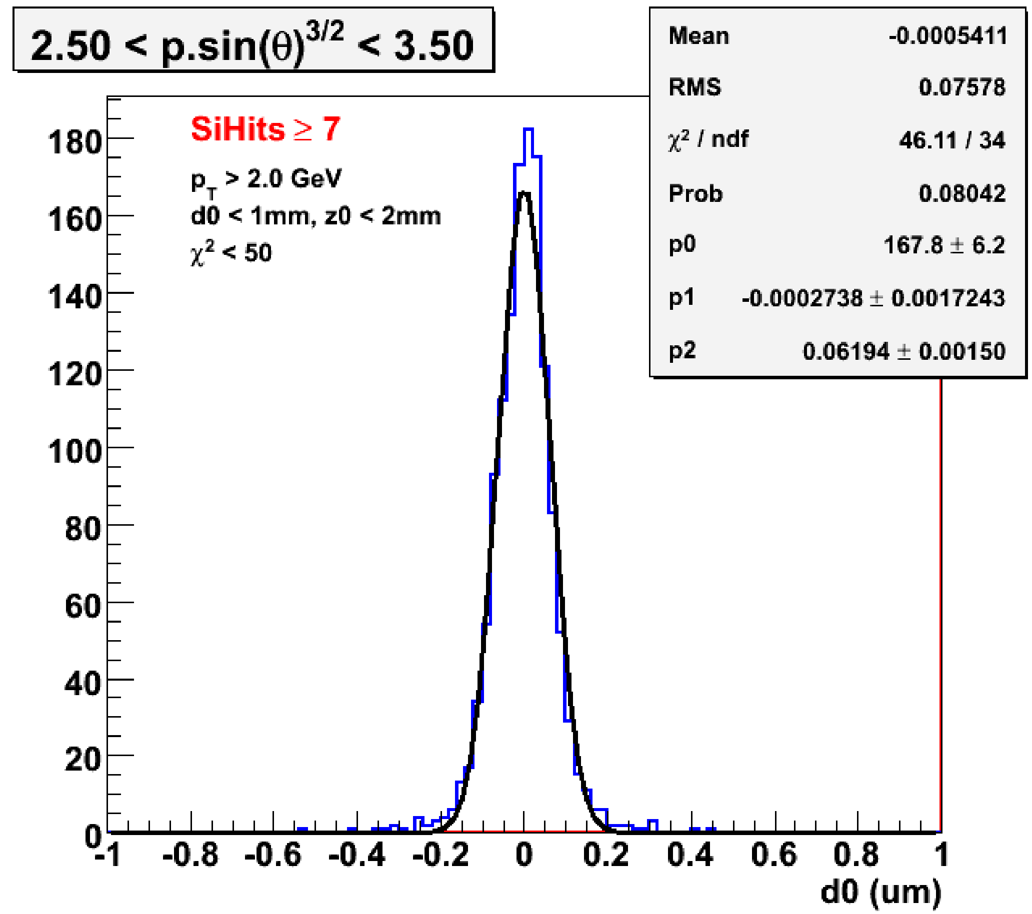




Multiple scattering is inversely proportional to pT=p.sin(theta), and proportional to the squared root of the distance traveled by the track: sqrt[sin(theta)]. Parameterized as a funcion of pScat=p.sin(theta)^3/2

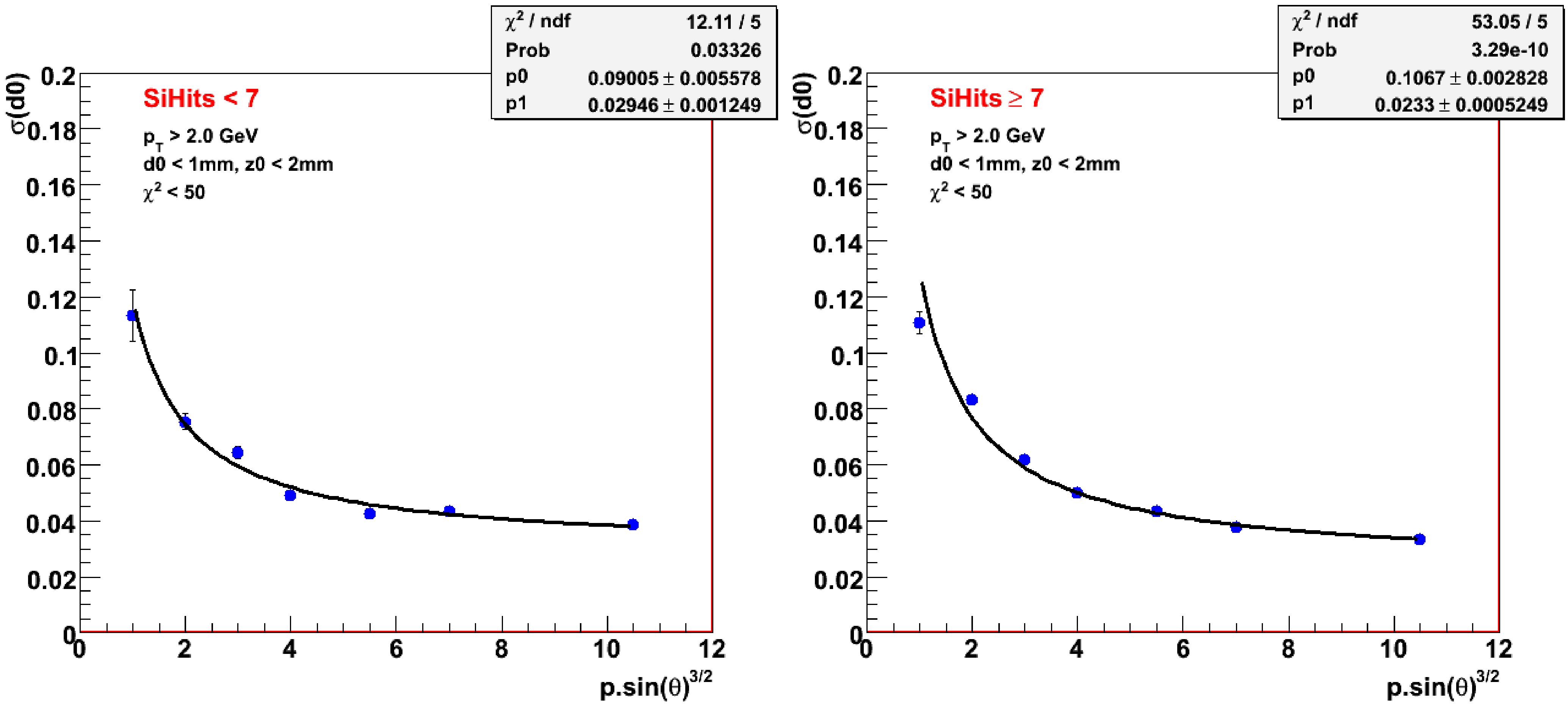
Parameterization of Transverse IP Error





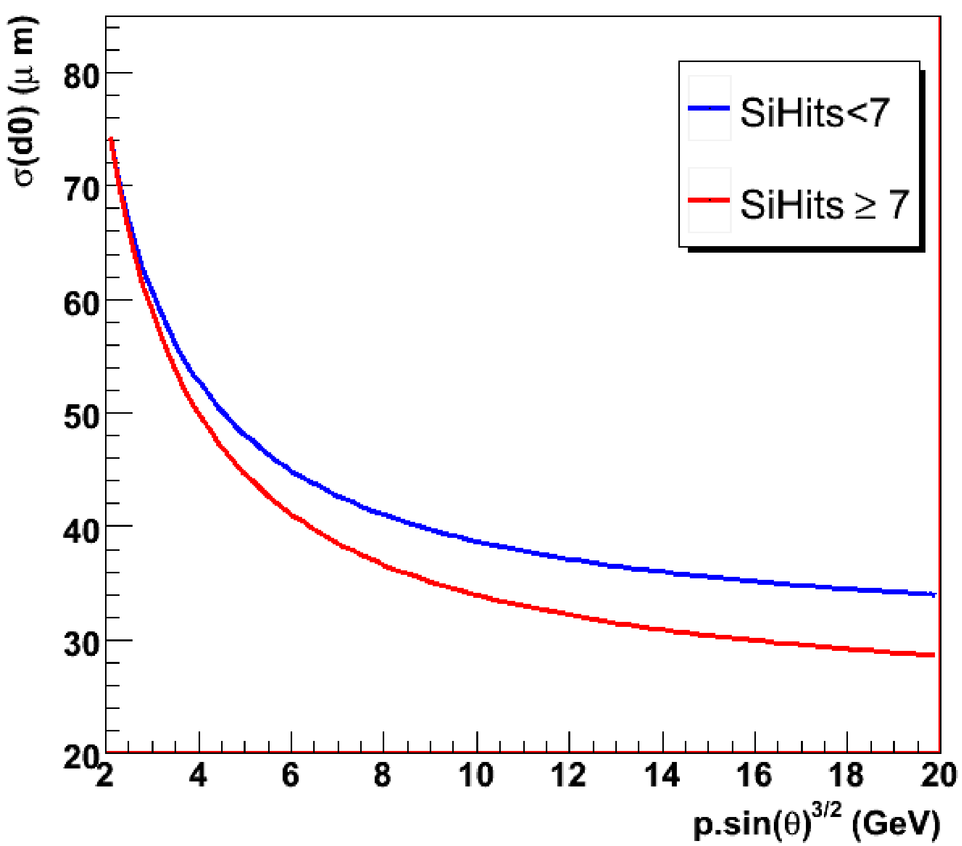
Tracks with 7 or more hits and low pscat have almost no tails, but the central resolution is almost the same as tracks with less than 7 hits.

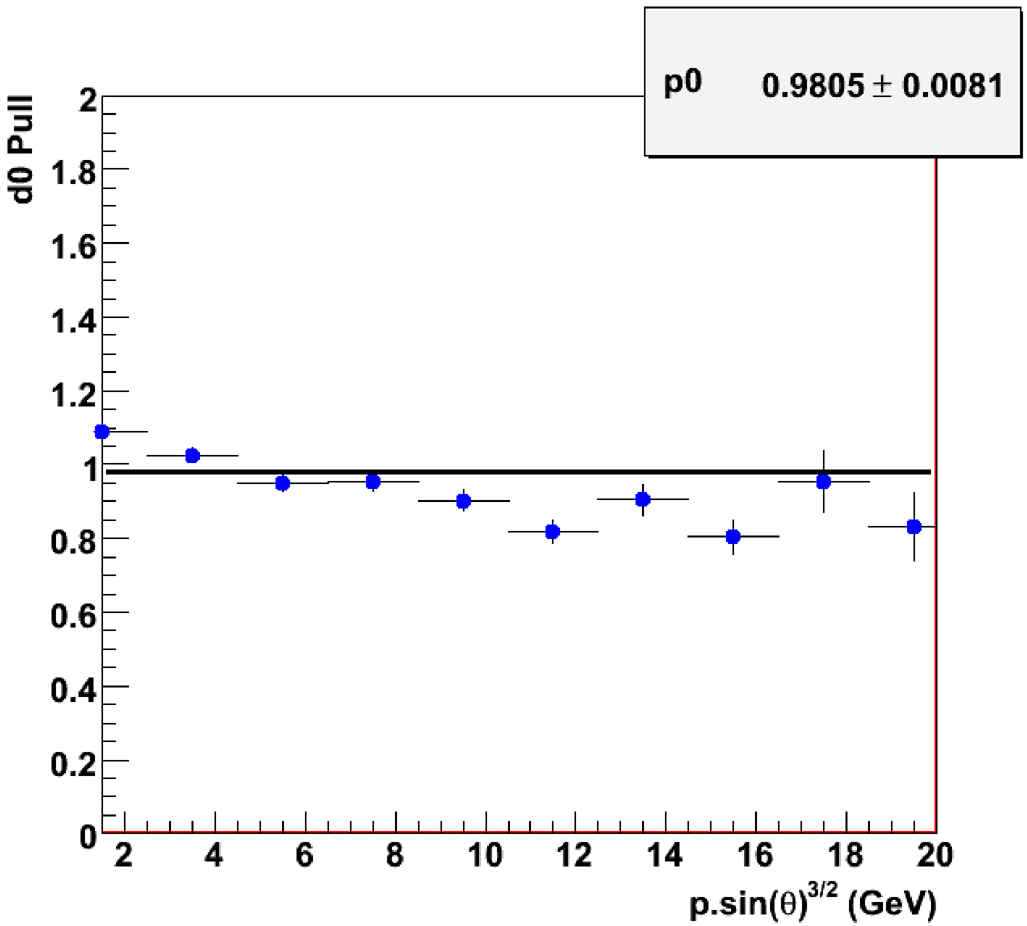
Parameterization of Transverse IP Error



IP resolution constant term ~ 30-40um.

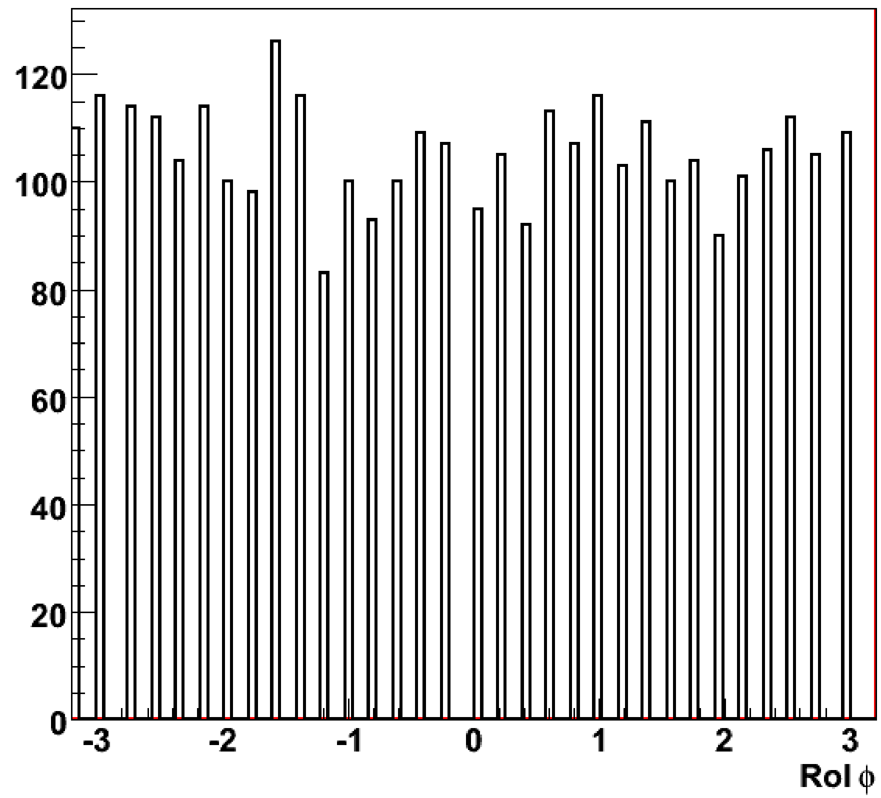
Parameterization of Transverse IP Error

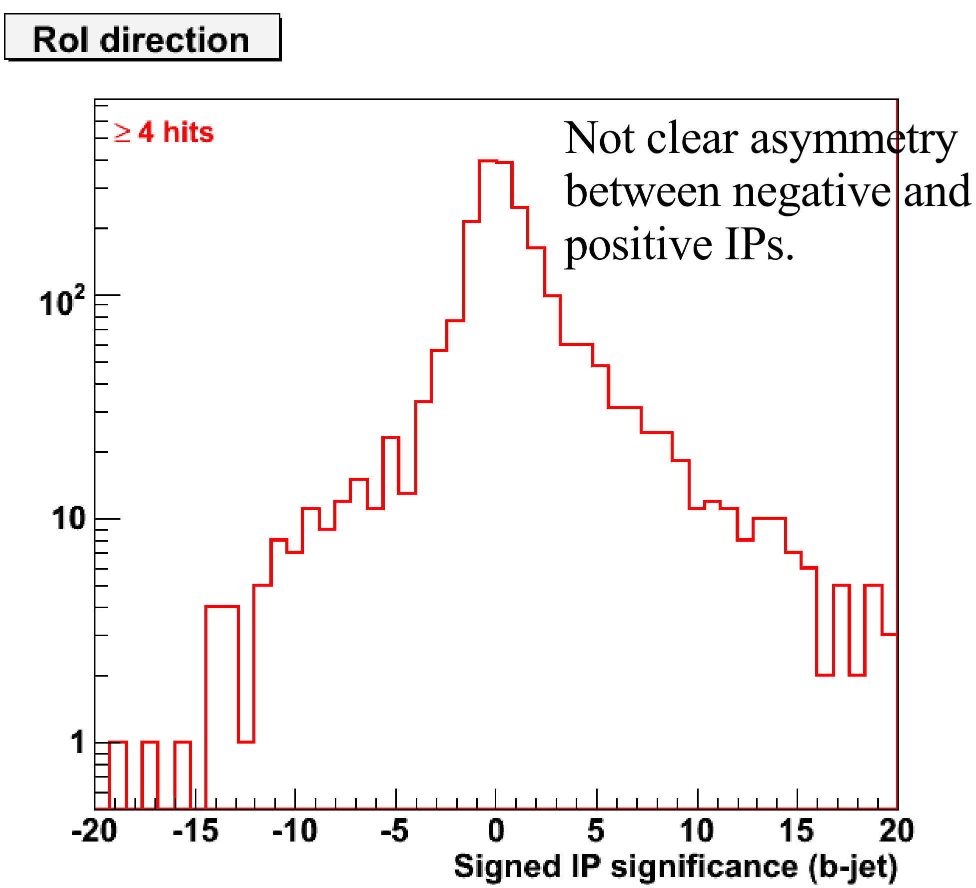




Signed Impact Parameter

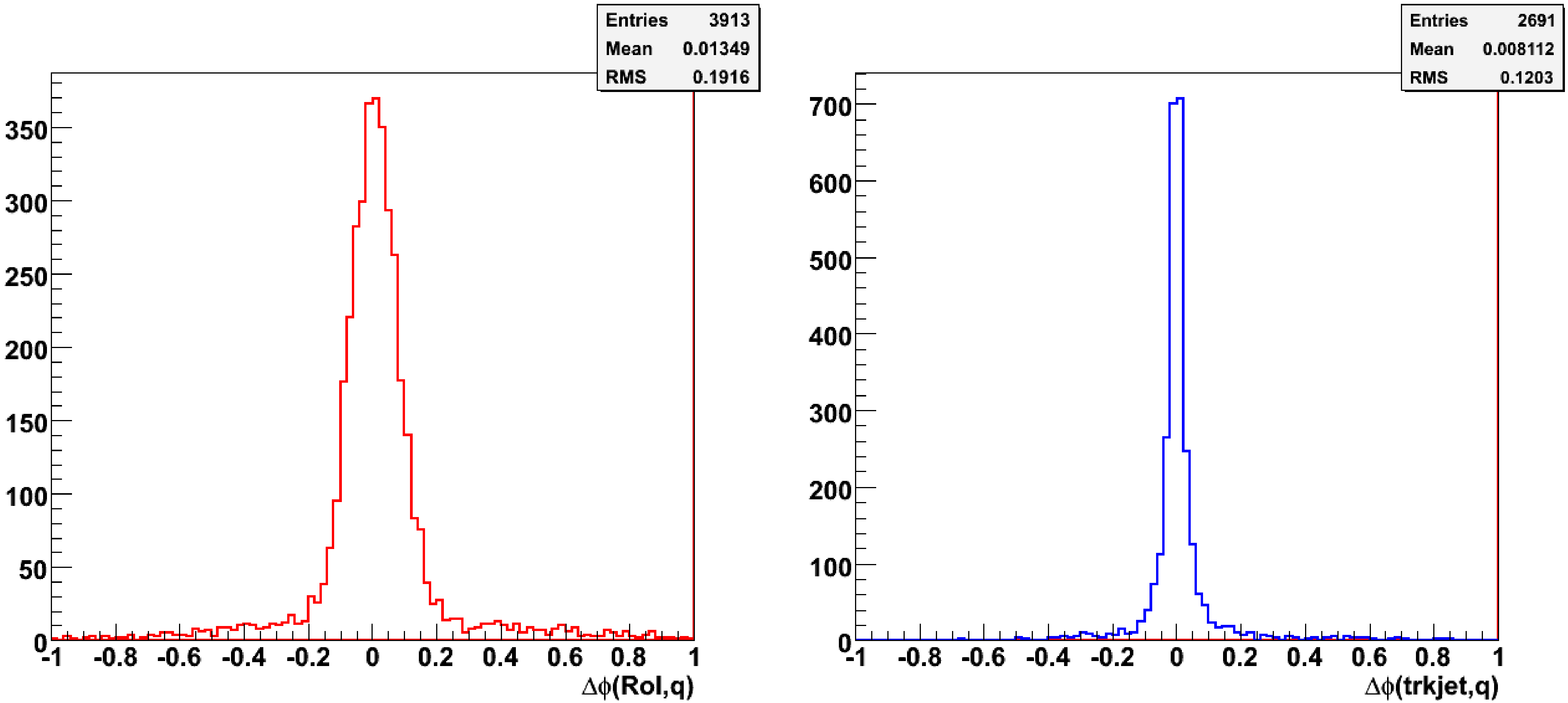
Chi2 probability b-tagging algorithm is based on the sign of the track IP. Does the RoI direction provide good phi resolution? Sign misidentification results in excess of b-jets with negative IP, and light jets with positive IP: degradation in performance.





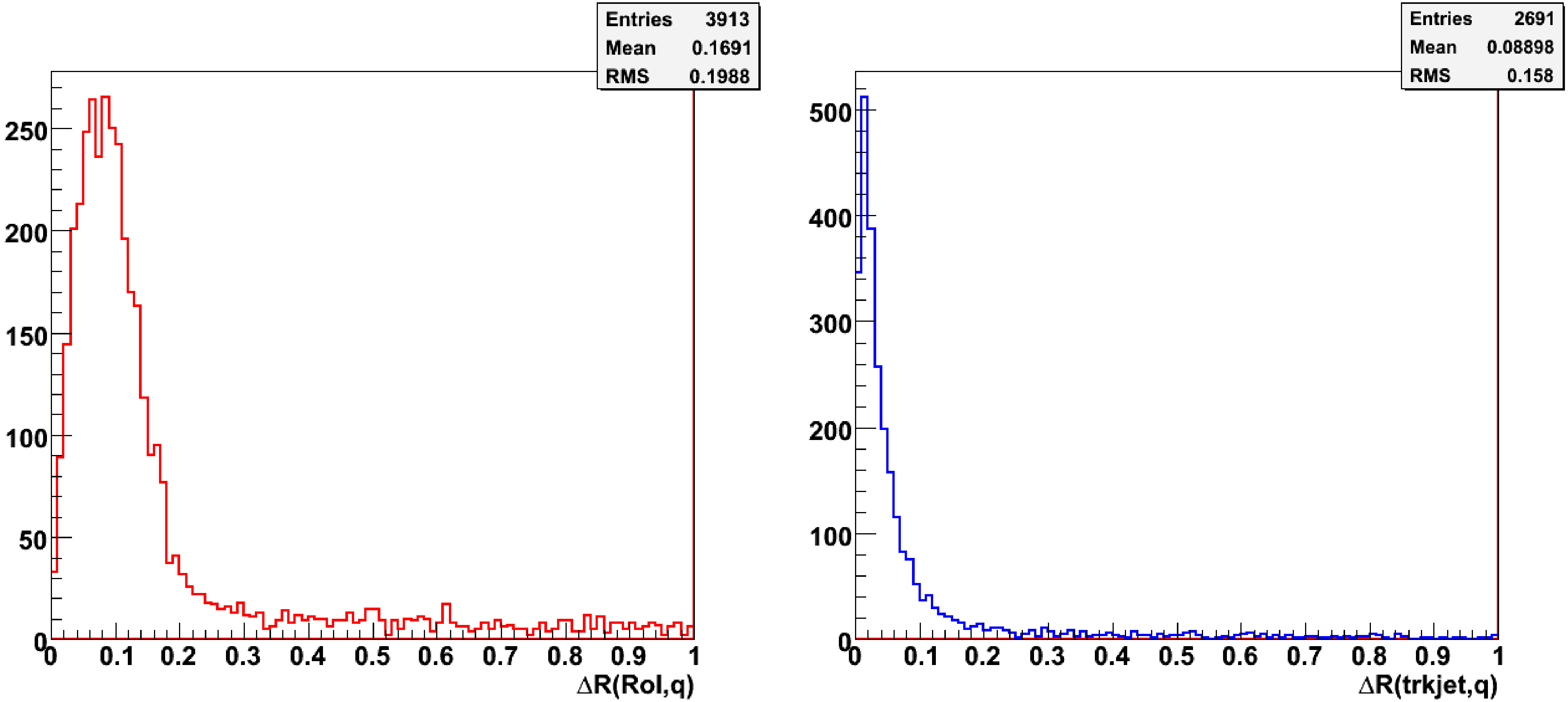
Improving the Sign Resolution (I)

Use ***Track-Jets*** to reconstruct the jet direction and improve RoI phi resolution. Use Simple Cone algorithm with size 0.4



37% improvement in phi resolution.

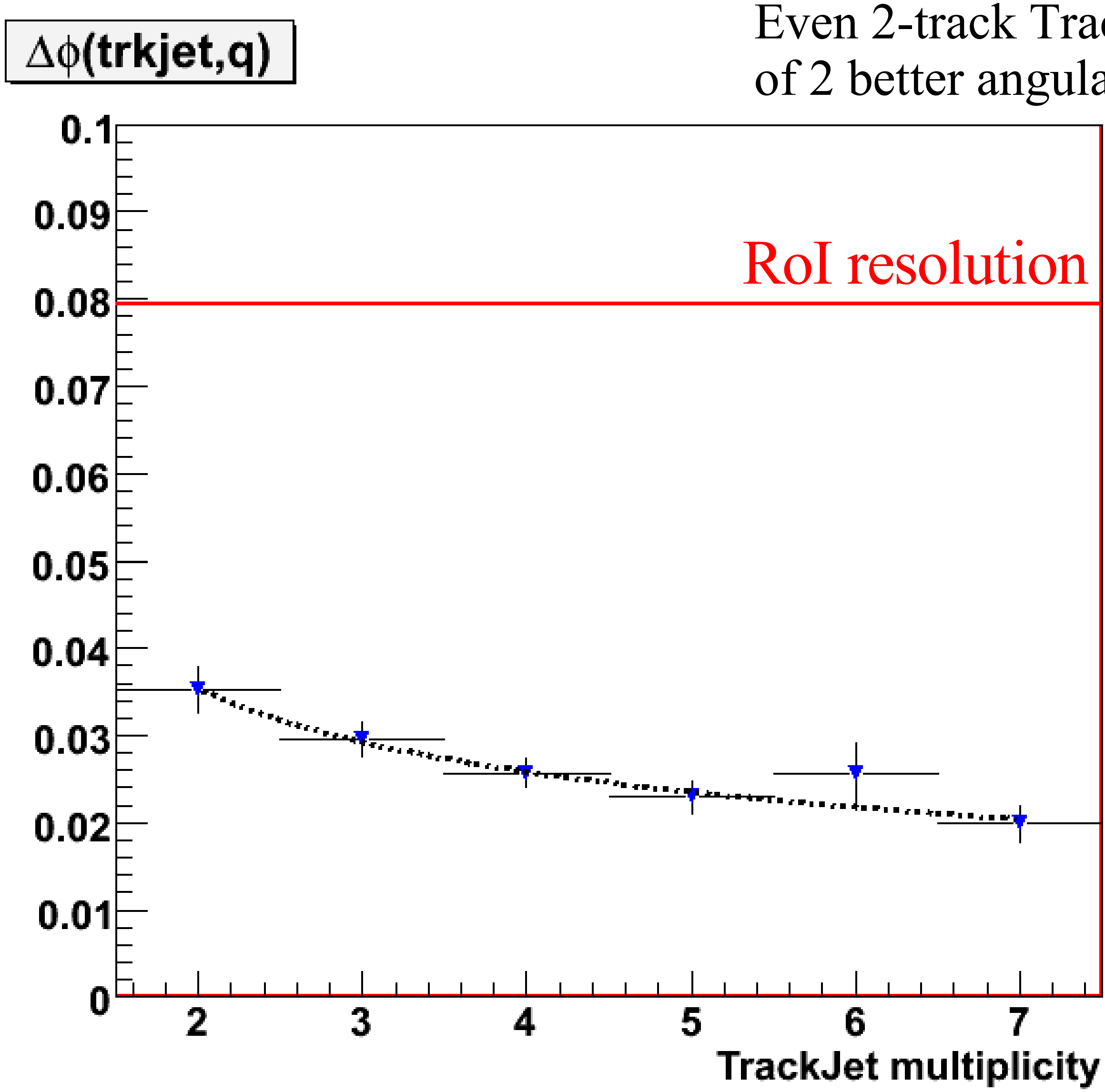
Improving the Sign Resolution (II)



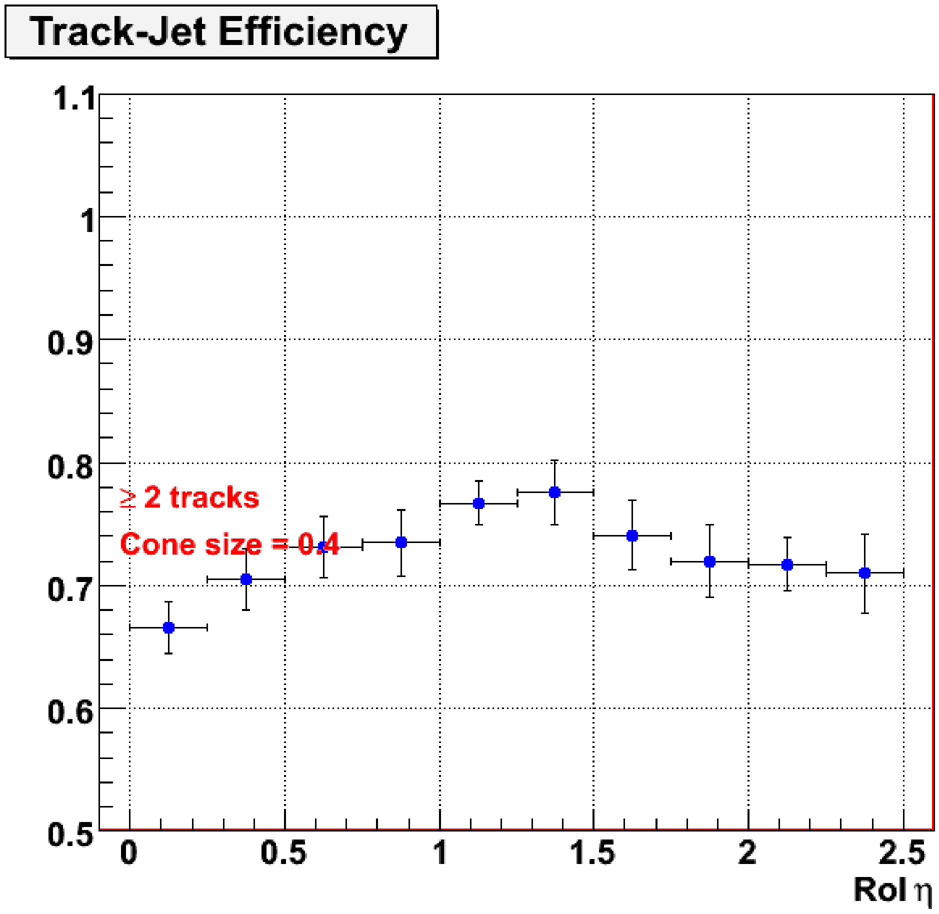
Significant improvement of TrackJets over L1jets in both eta and phi resolution. Could this idea be used to seed L2jets? Another approach is to use L2jets.

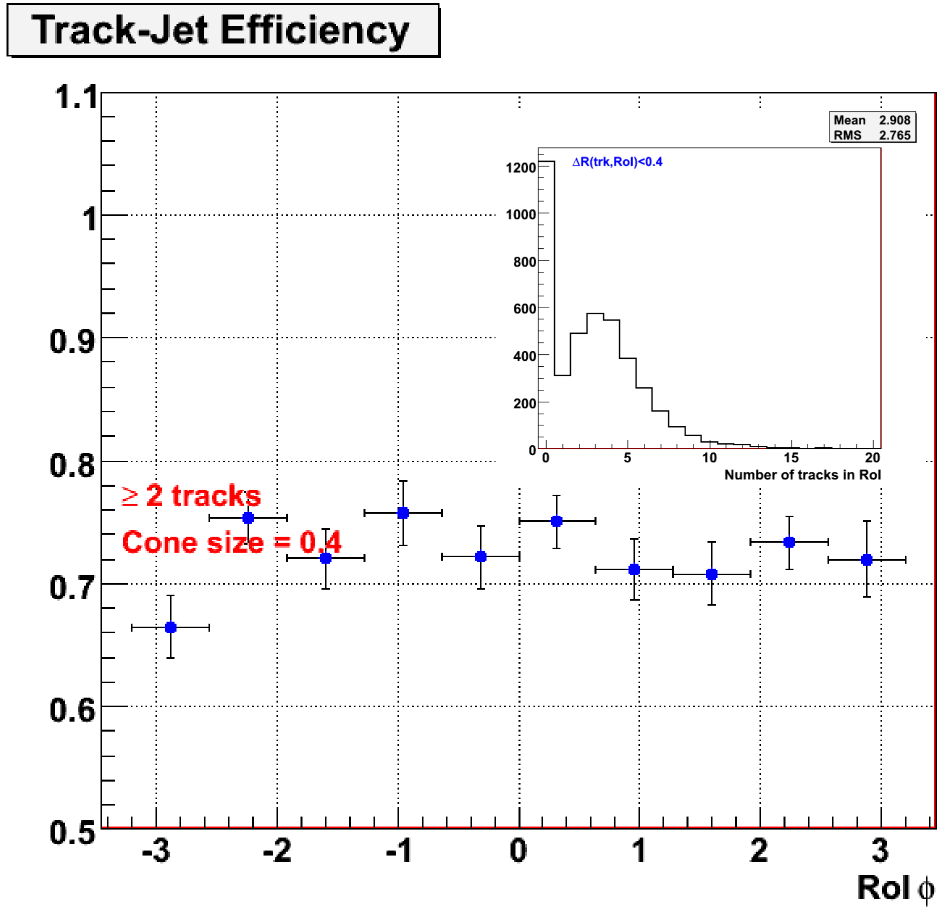
TrackJet Angular Resolution

Even 2-track TrackJets provide a factor of 2 better angular resolution than RoIs



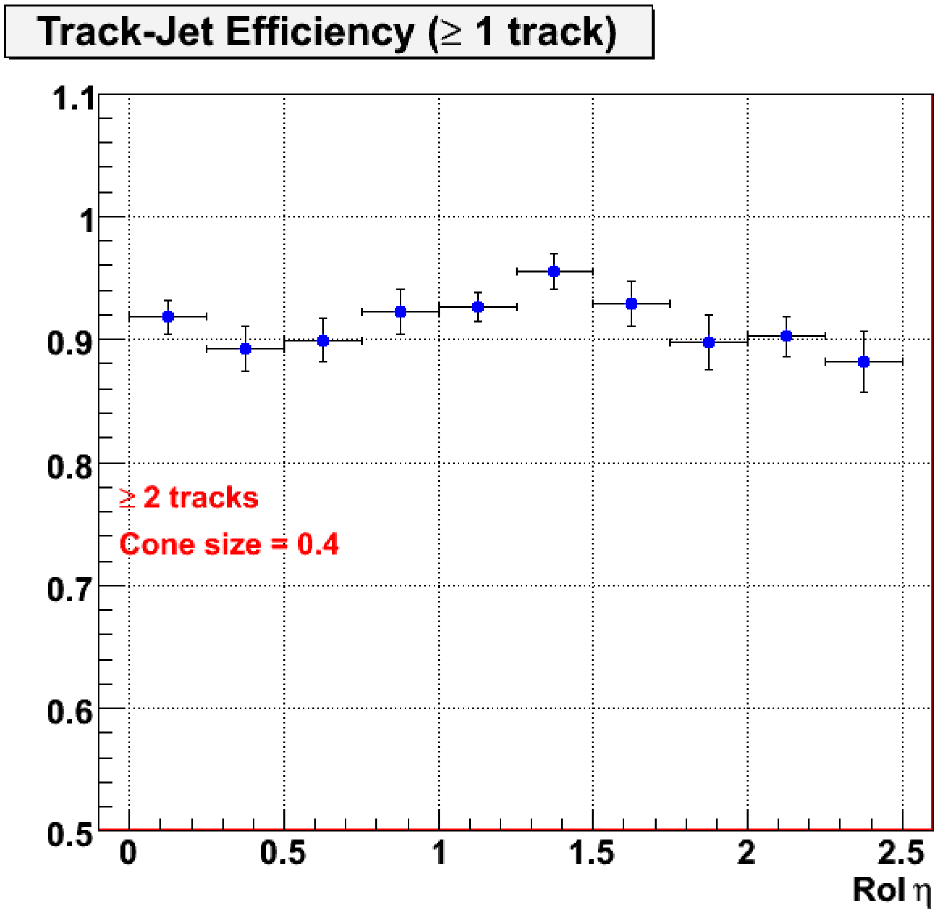
TrackJet Efficiency (I)

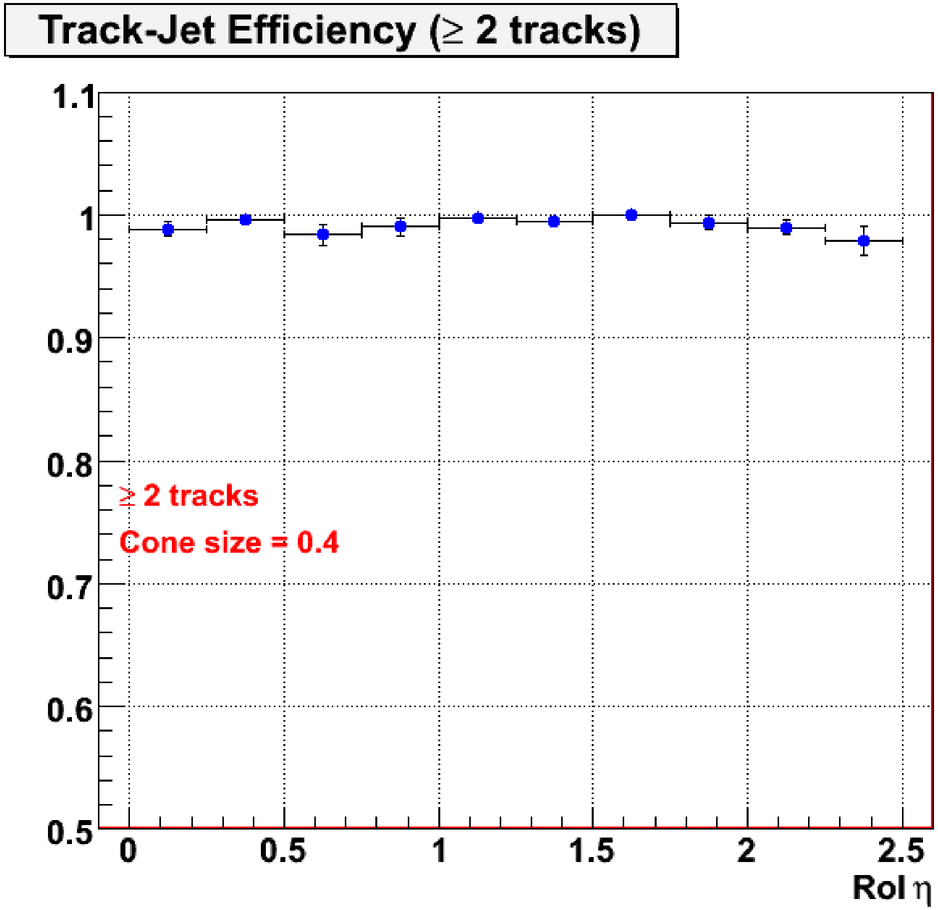




On average, a TrackJet (of at least 2 tracks) is found in 70% of RoIs. The inefficiency is mostly due to the lack of tracks pointing in the RoI direction.

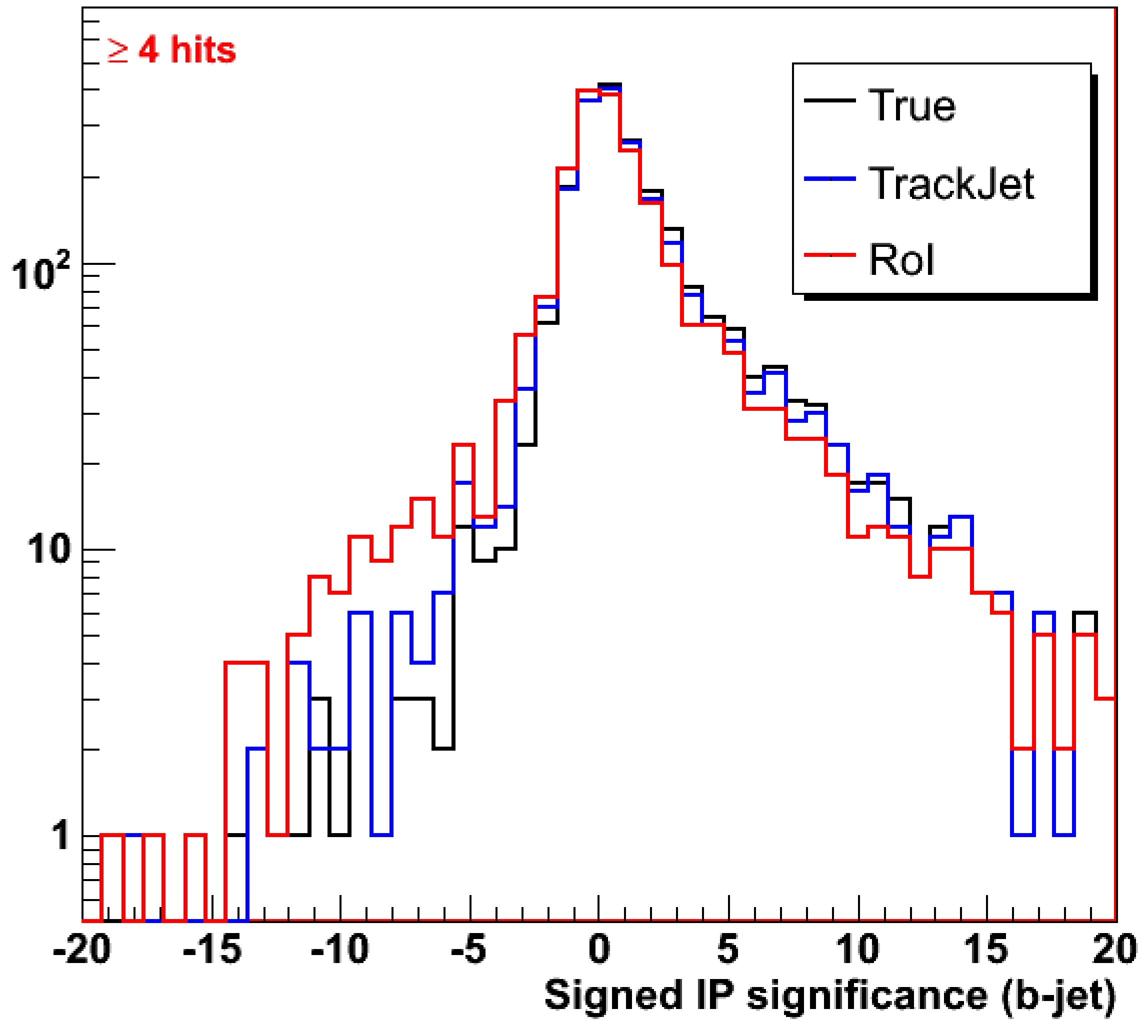
TrackJet Efficiency (II)



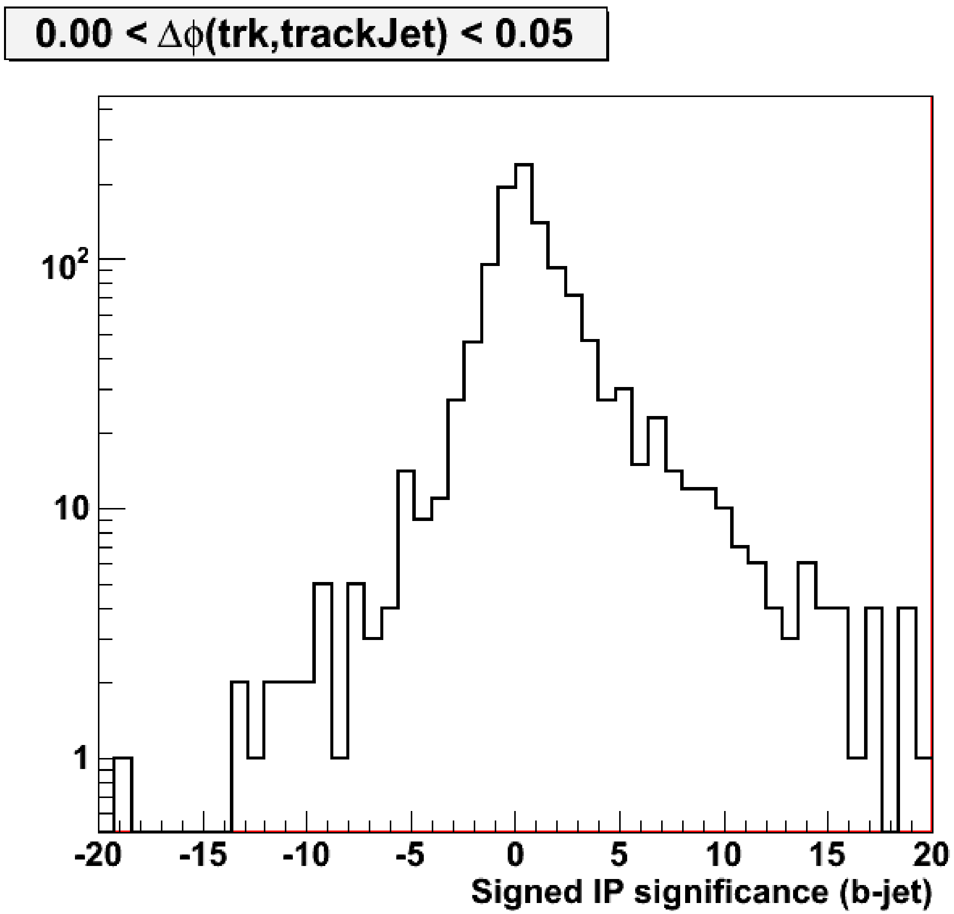


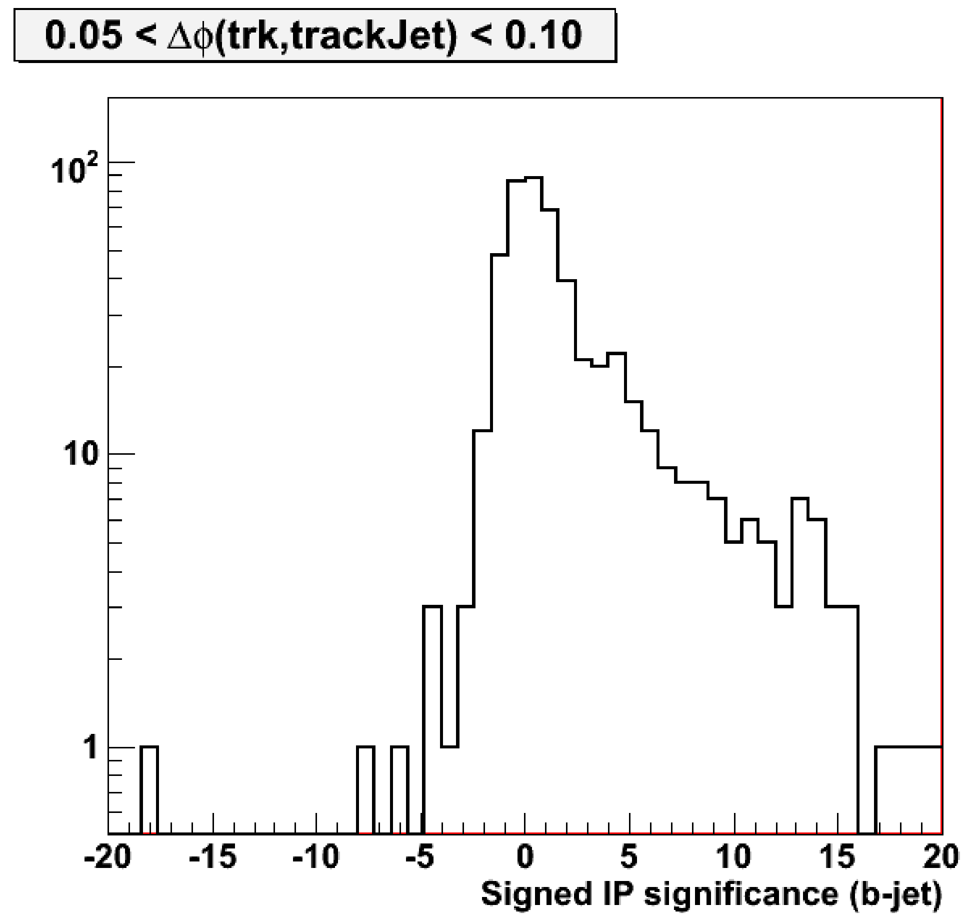
When there is at least 1 track pointing in the RoI direction, a trackJet can be found 90% of the time.

Signed Impact Parameter Resolution



Sign Misidentification

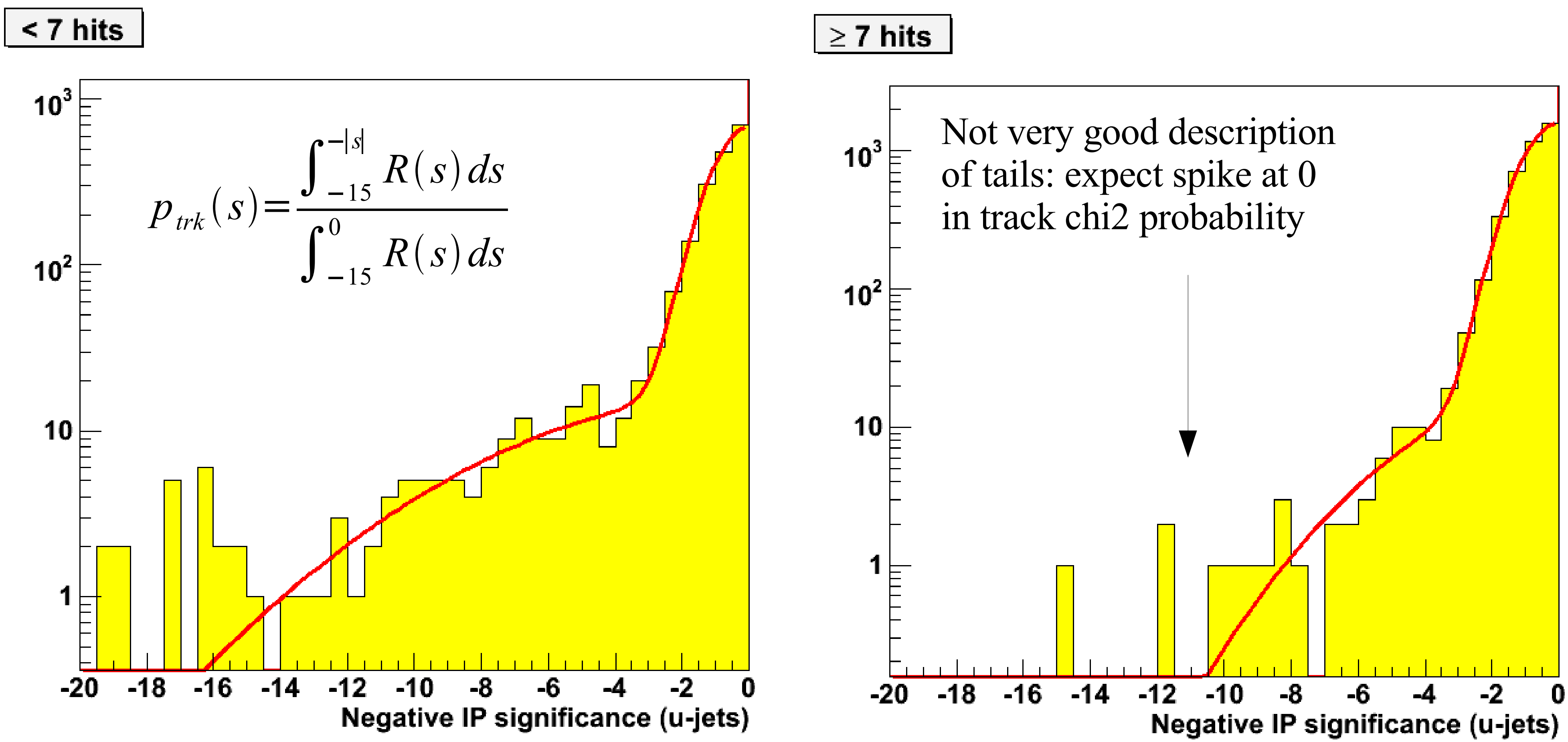




The sign of the IP is not well defined when DeltaPhi(trk,jet) is small (due to IP and phi resolution fluctuations)

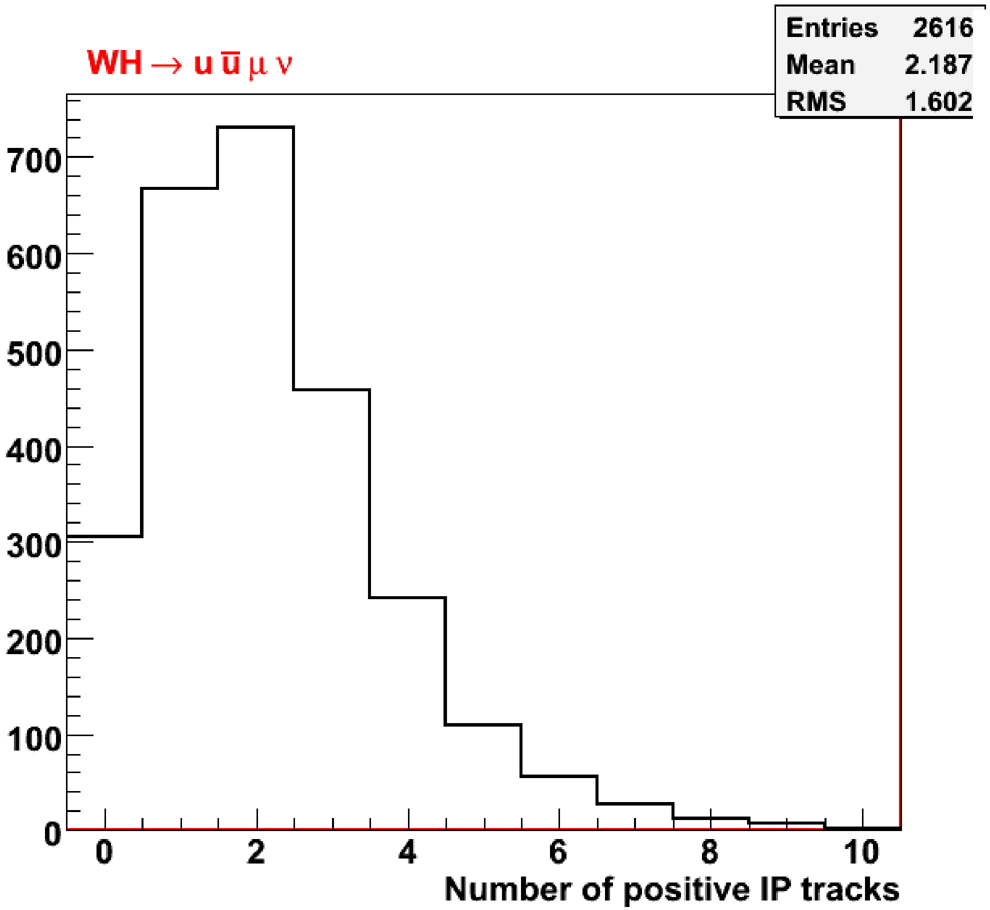
Investigate the possibility of adding Dphi(trk,jet) to the b-tag algorithm.

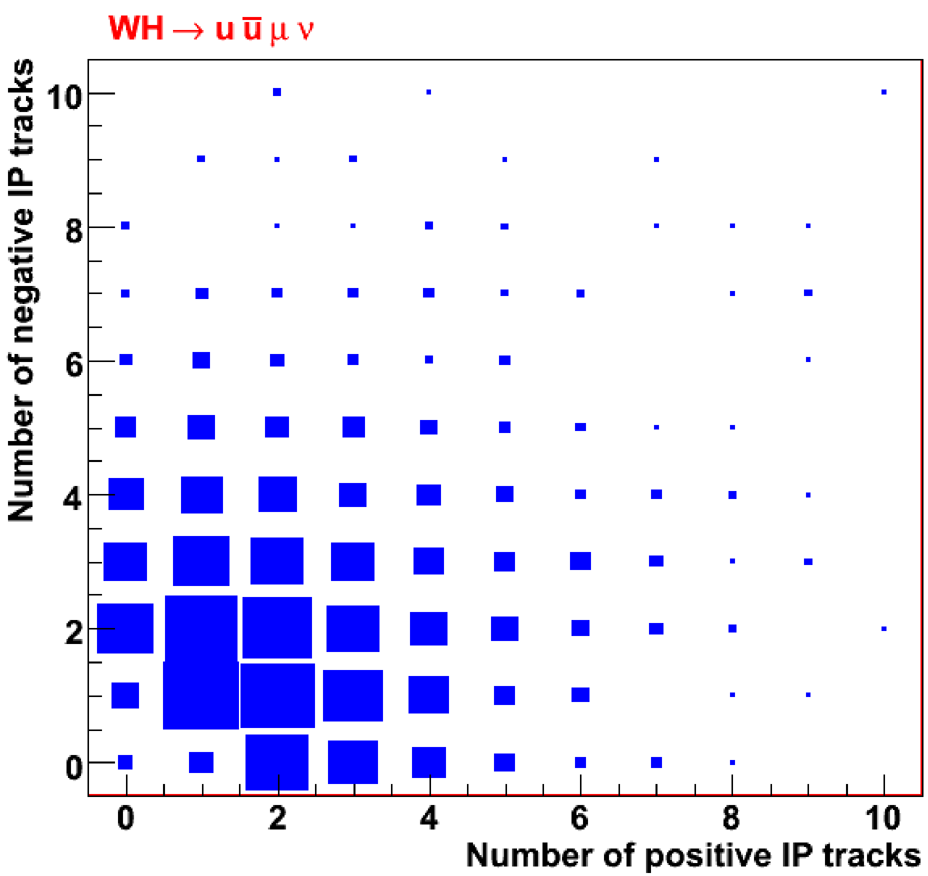
IP Resolution Functions



Parameterize the negative IP significance as a function of the number of hits. Need more statistics to correctly describe tails: this is the key to improve the btag performance of this method.

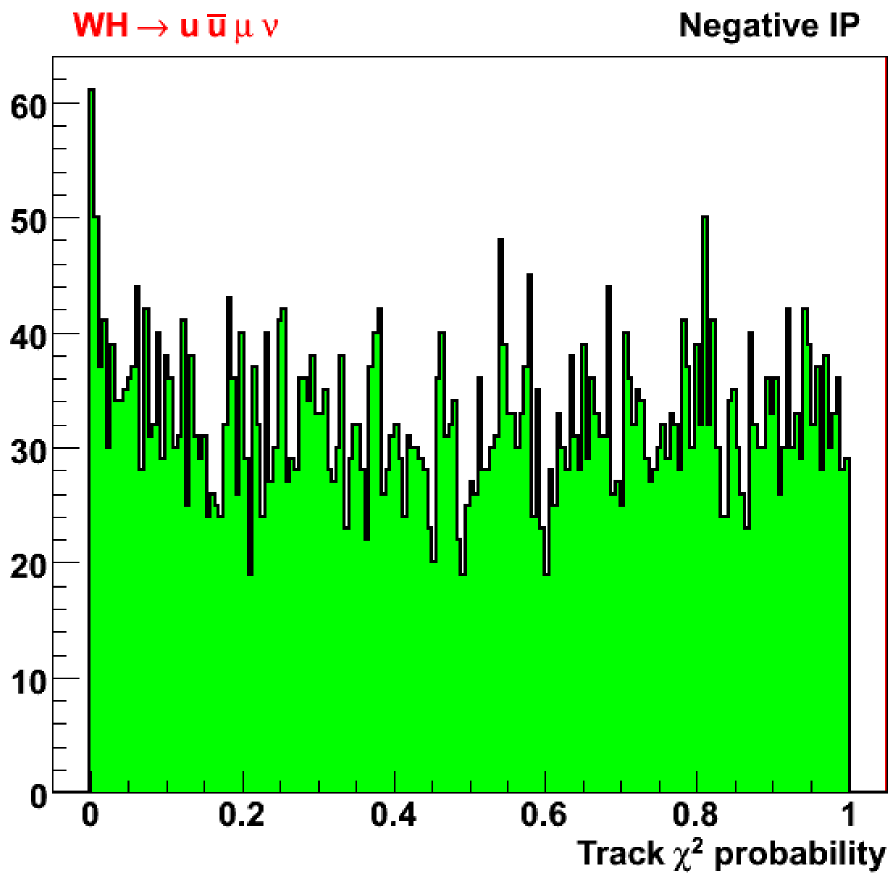
Chi2 Probability Tagging Algorithm

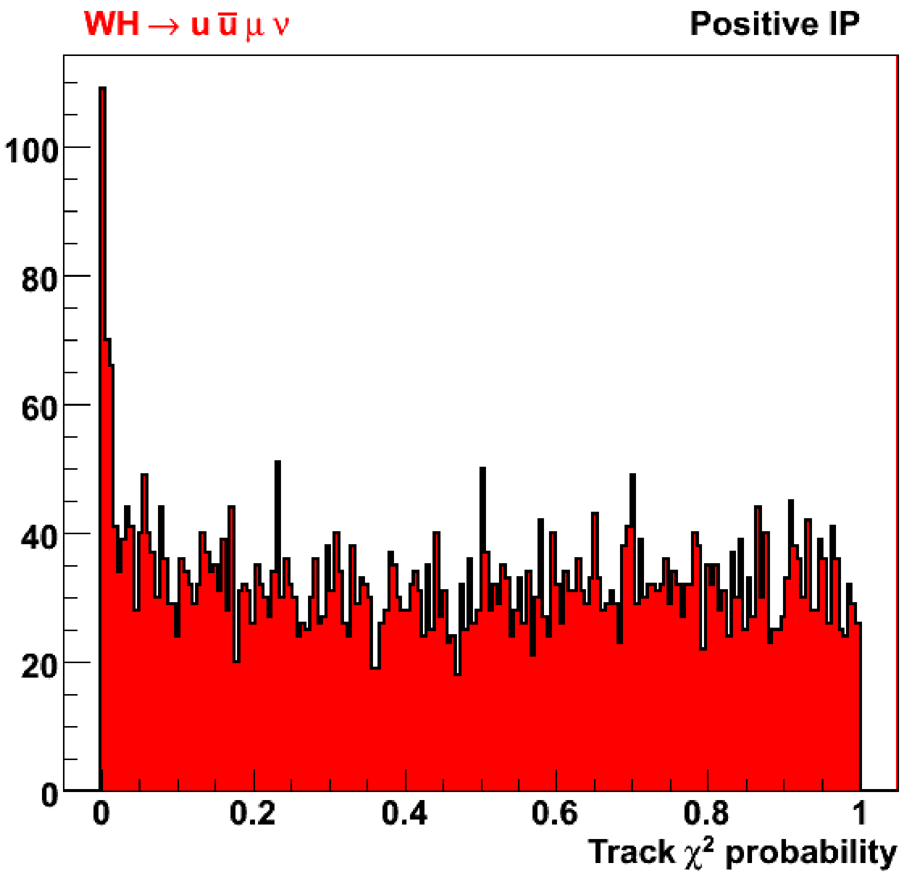




For each trackJet, calculate the jet chi2 probability for positive and negative IP tracks. Consider jets with at least 1 positive or negative track.

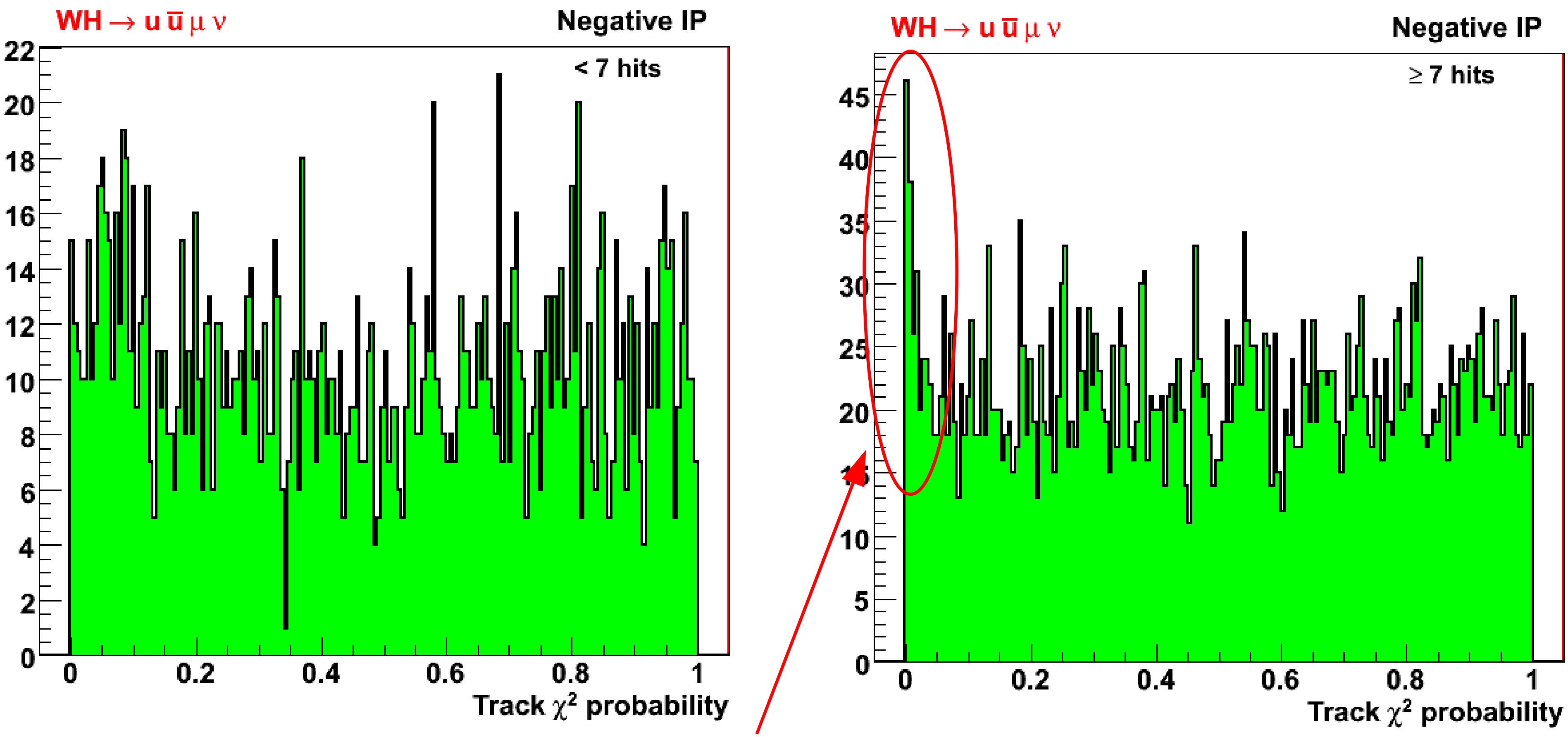
Track Chi-Square Probability (I)





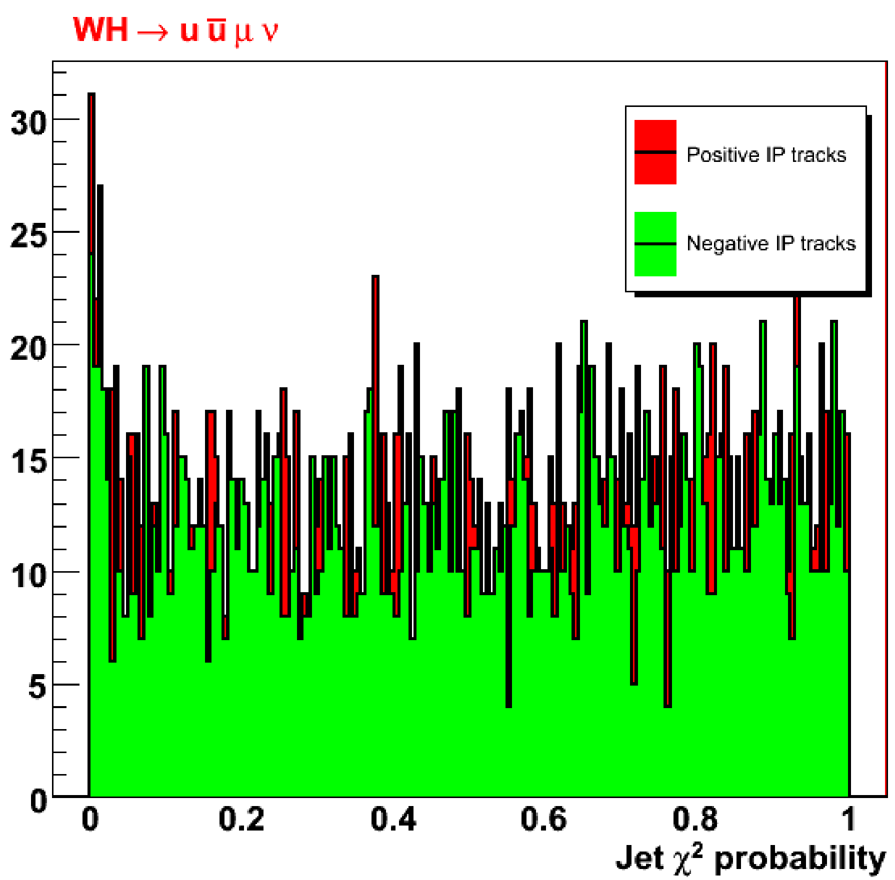
Spike at 0 in negative IP tracks: non-Gaussian tails.
Spike at 0 in positive IP tracks: Long-lived particles.

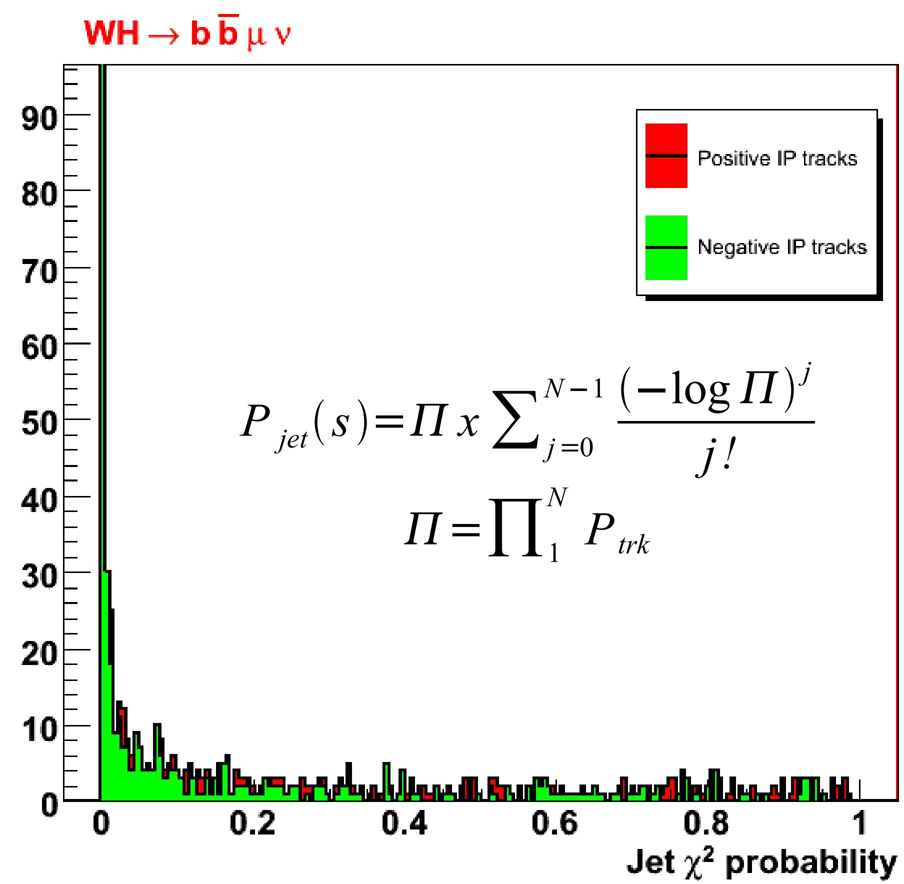
Track Chi-Square Probability (II)



The spike at 0 comes from the parameterization of tracks with 7 or more hits, for which the tails are not well described: room for improvement!

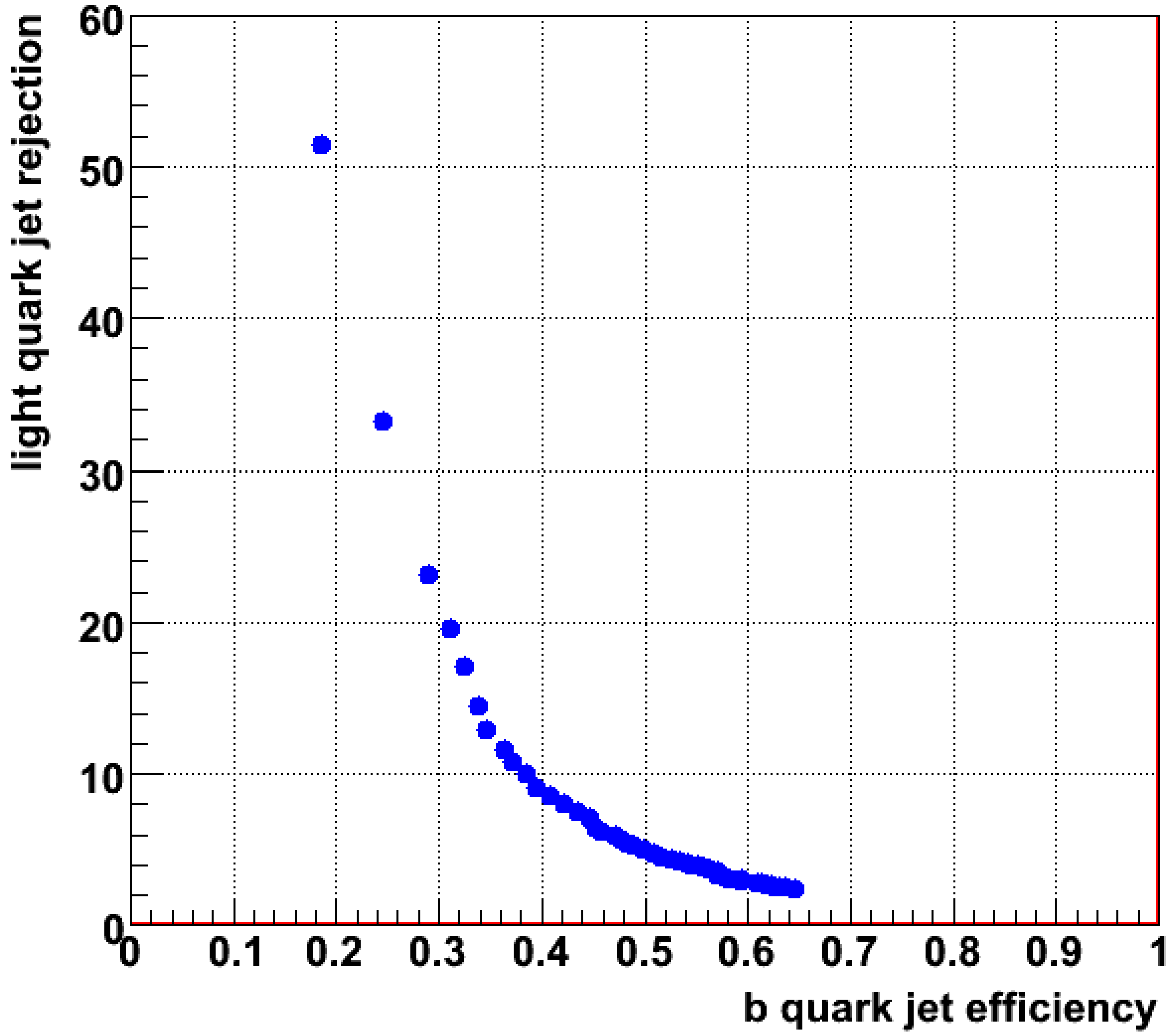
Jet Chi-Square Probability (I)





Significant spike at 0 for negative IP tracks in b-jets: probably due to sign misidentification: Consider the use of Dphi(trk,jet) to improve b-jet efficiency.

Performance in WH Simulated Events



Summary and Conclusions

First implementation of chi-square IP probability b-tagging hypothesis algorithm:

- Improved PV reconstruction and selection at L2, using SiTracks (20% better z resolution)

- IP error parametrization as a function of number of hits.

- Use of Track-Jets to improve the RoI direction:
  Crucial for signed IP definition.
  Track-Jets provide more than a factor of 2 better angular resolution. 90% efficiency for RoIs containing at least 1 track.
  More robust to multiple interactions (z information)
  May play an important roll in some new physics scenarios.

Summary and Conclusions (II)

- IP significance templates derived for 2 track categories, based on the number of hits:

  Need more statistics to improve parameterization and add additional parameters (exponential)

- Large fraction of b-jets have large spike at 0 for negative jet chi-square probability:

  Due to sign misidentification (dca and phi0 resolution)
  Explore the use of DeltaPhi(jet,trackjet) to downweight tracks with small Dphi for which the sign of the IP is not well defined:
  For instance: tracks with Dphi<0.05 are always consider as having both positive and negative sign.

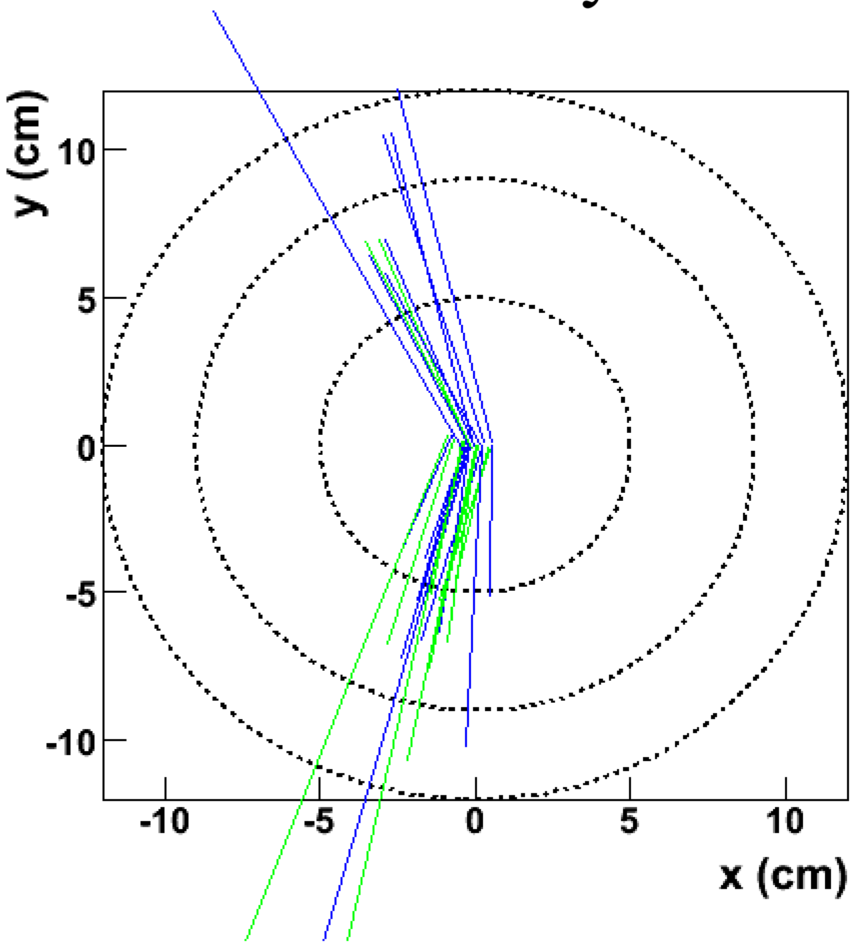
- Room for improvements: 3D IP, track selection, ...

Plans: L2 Secondary Vertex b-tagger

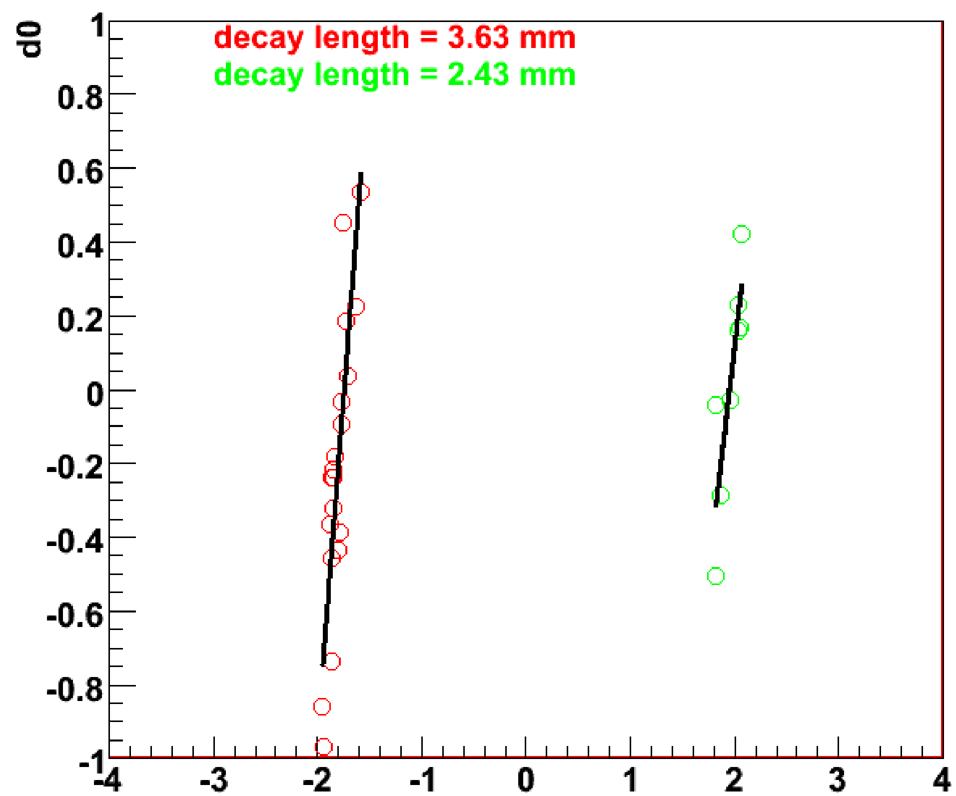
For tracks belonging to a secondary vertex, IP and phi are correlated:

$$dca = L\sin(\phi_{trk} - \phi_{SV})$$

Search for secondary B vertices by fitting the dca-phi distribution.



Technique originaly developed by CDF.



Work in progress: Algorithm has been implemented, need to measure its performance.