b-Tag Track Studies

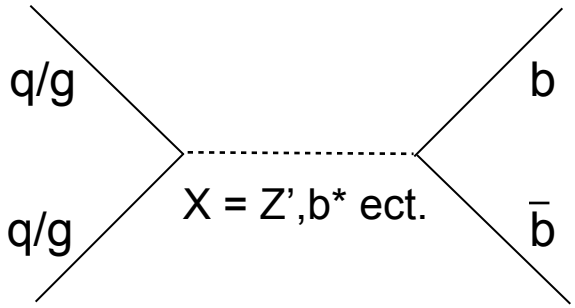
Laurie McClymont
(with Andreas and Tim)

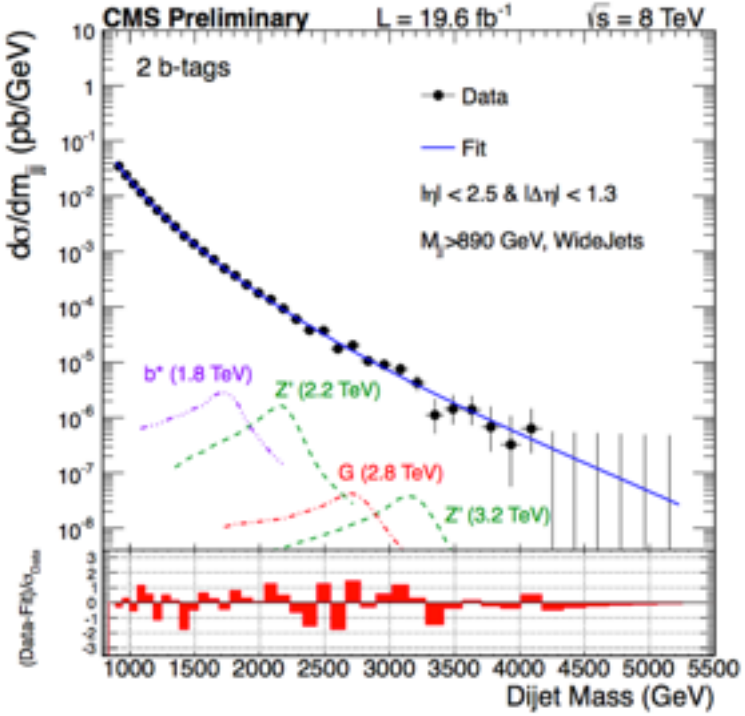
Exotic Dijet Meeting
11/05/15

Aims of Study

- To study the track selection of IP3D, SV1 and Jet Fitter Algorithms when applied to truth b-jets.
- Optimise flavour tagging performance for high P_T by adjusting track selection - Important for di-b-jet resonant studies.
- Use and compare high-P_T Z' bb and QCD samples.
- Explain discrepancy between samples for high-P_T b-tagging efficiency found in previous talk.

Exotic Searches in Di-b-jets



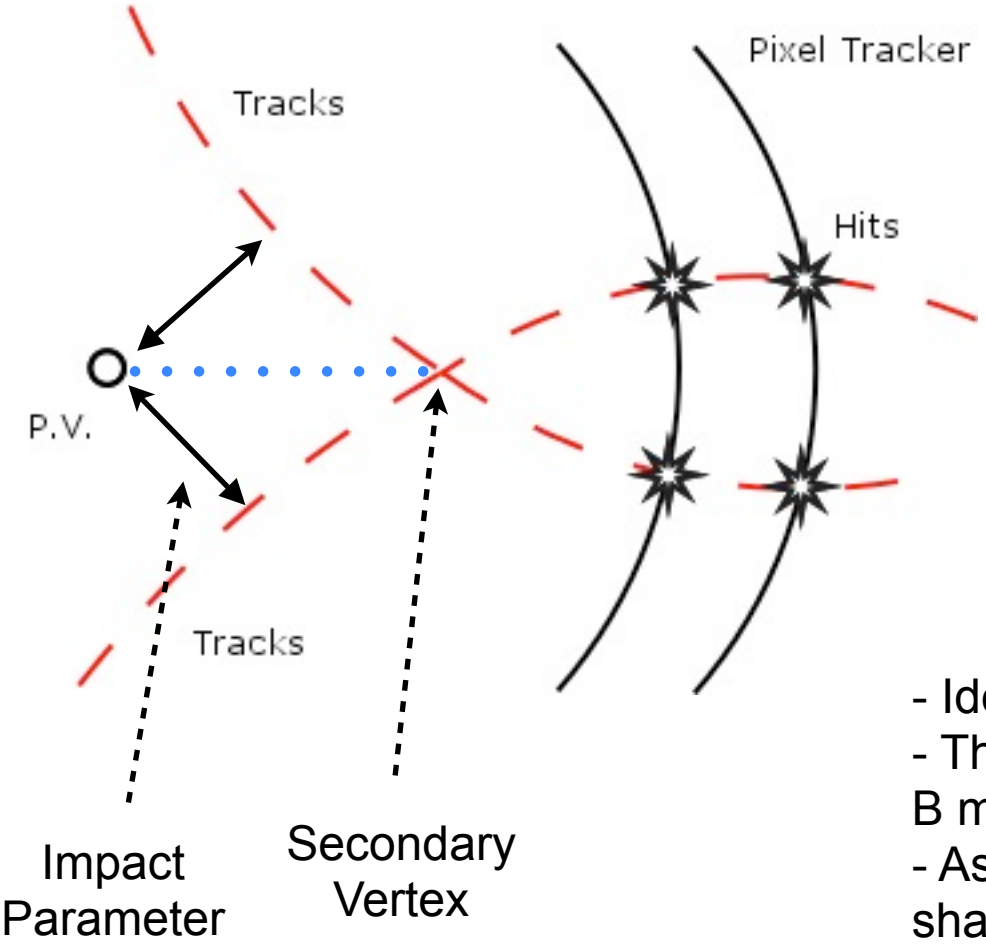


- Search for pairs of b-tagged jets.
- Di-jets is one of the most sensitive probes of new physics at high masses.
- Many BSM models predict resonances that have decays to $b\bar{b}$ as it is a member of the third generation of quarks.

Challenges

- We require effective b-tagging at extremely high-p_T (~1TeV)

B-tagging - IP3D, SV1, JF



IP3D

- Look for tracks with a large impact parameter significance = (I.P. / σ)

SV1

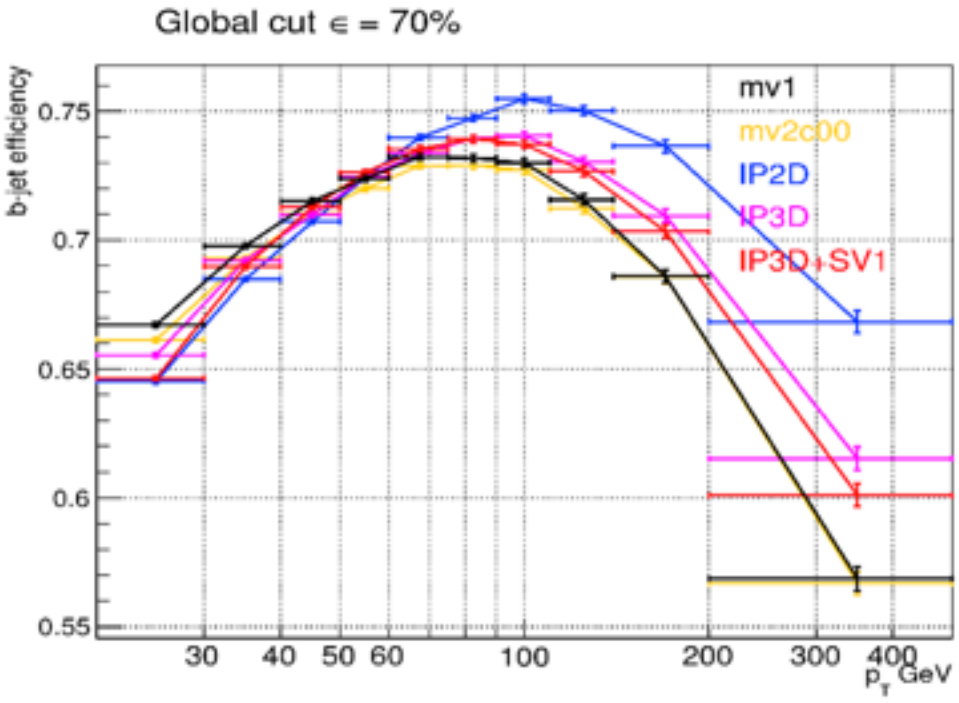
- Identify a secondary vertex
- Look for large flight path significance = (F.P. / σ)

JF

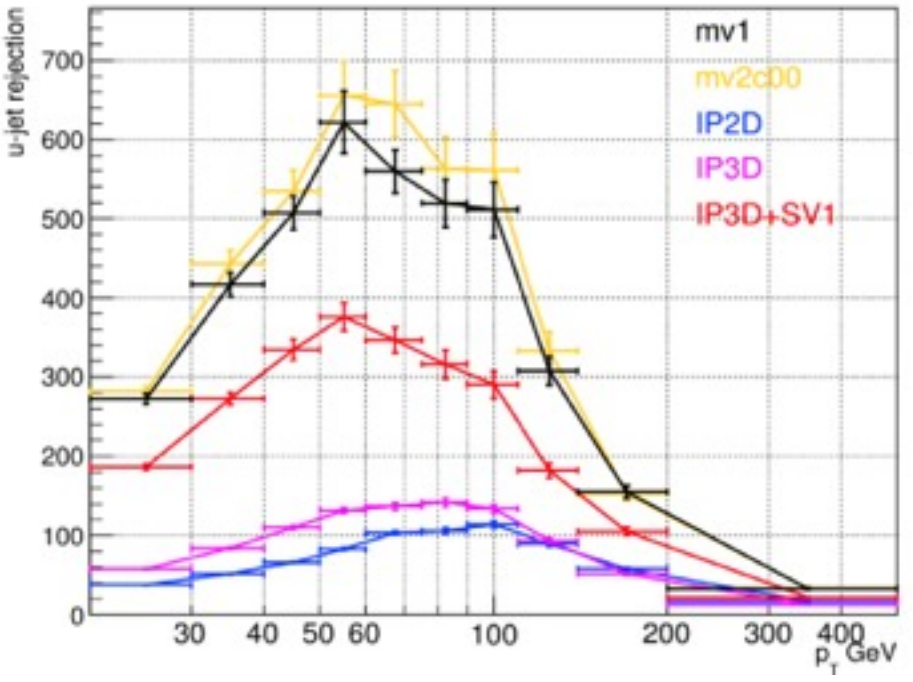
- Identify Second and Tertiary Vertices.
- These vertices correspond to decay of B meson and decay of D meson.
- Assumes these vertices lie on an shared B flight axis

Problems at High P_T

First Look at R20 - Yulia Rodina



- Global Cut Efficiency = 0.7
- For a fixed discriminant cut
- 13TeV ttbar



- Flat Cut Efficiency = 0.7
- 13TeV ttbar

- Longer decay length of B at high P_T
- Higher boosted jets at high P_T

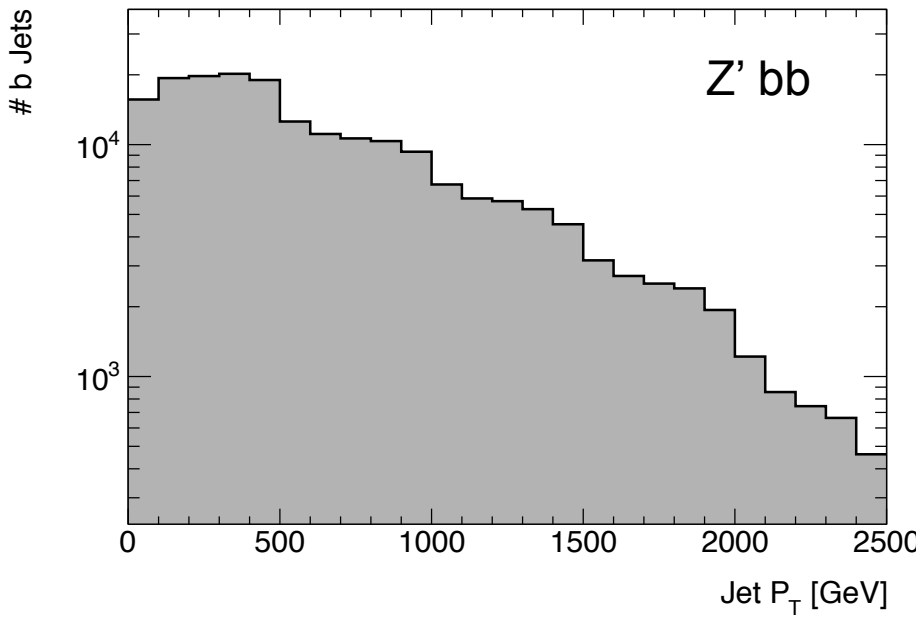
Samples Used for Studies

ZPrimebb

mc14_13TeV.1913.Pythia8_AU2MSTW2008LO_Zprimebb*e3569_s2576_s2132_r6235*

1,2,3,4 + 5 TeV Z'

 # entries = 95,282

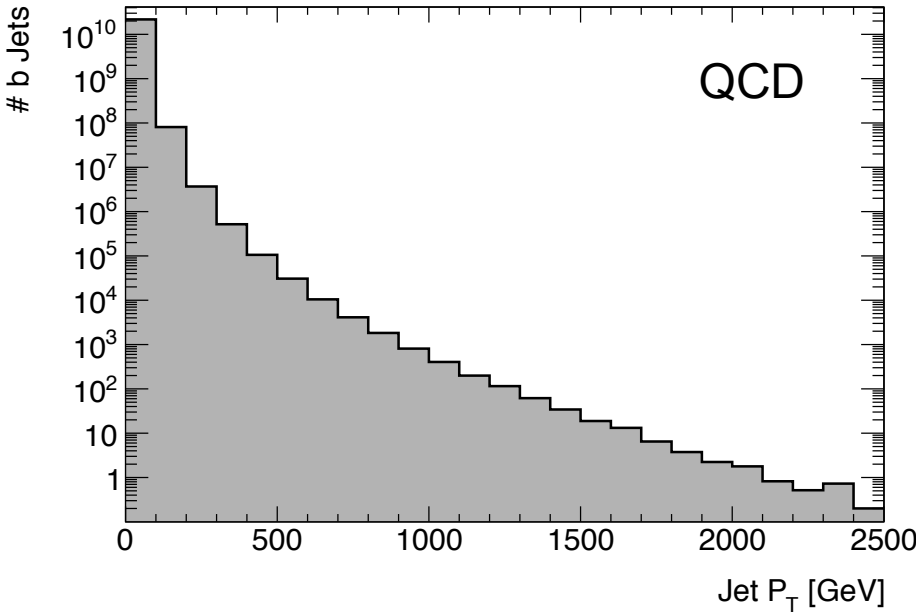


QCD

mc15_13TeV.36100*.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ*.recon.AOD.e3569_s2576_s2132_r6235/

JZ1 20–60 JZ2 60–160
JZ3 160–400 JZ4 400–800
JZ5 800–1300 JZ6 1300–1800
JZ7 1800–2500 JZ8 2500–3200

 # entries = 9,095,282

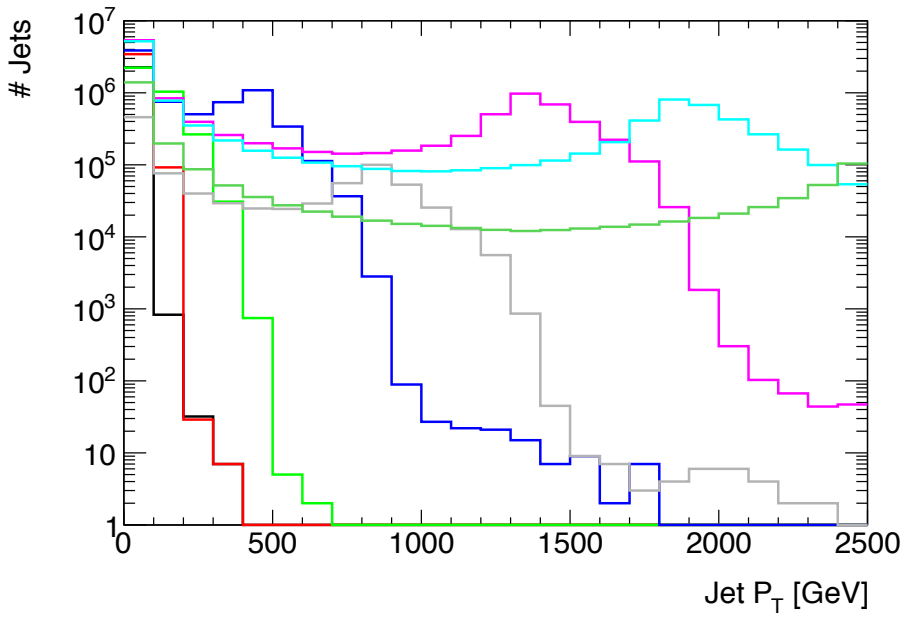


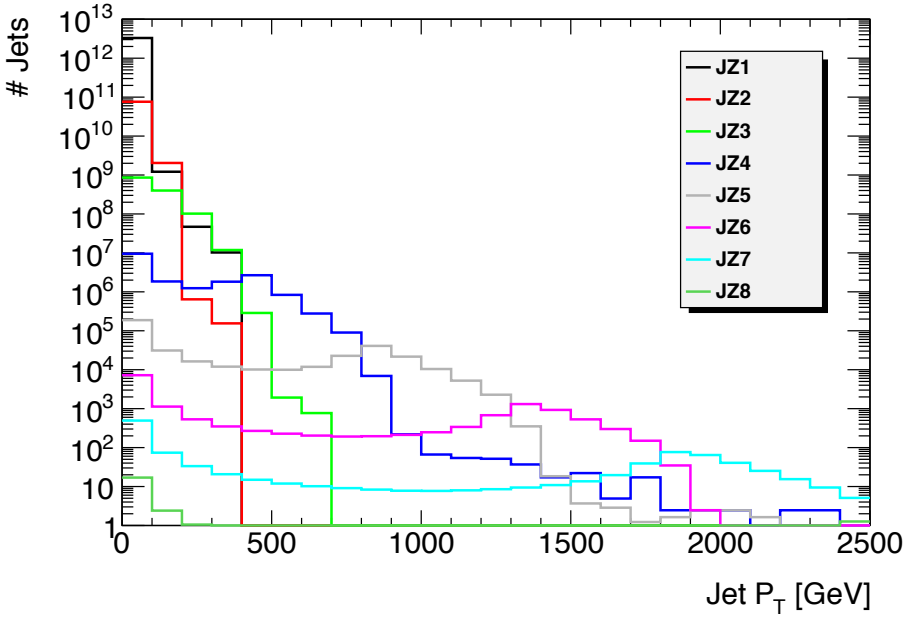
QCD samples weighted

- QCD di-jets have a jet-P_T spectrum that falls off at high-P_T rapidly.

- For MC simulation, we need enough statistical power at high jet-P_T, yet still recreate the correct di-jet spectrum.

- To do this the MC QCD samples are split into jet-P_T cuts each with similar sizes, then re-weighted to get the correct spectrum.





JZ1 20–60	JZ2 60–160
JZ3 160–400	JZ4 400–800
JZ5 800–1300	JZ6 1300–1800
JZ7 1800–2500	JZ8 2500–3200

Discrepancies at High-P_T

- In a previous talk by Ning it had been found that the b-jet tagging efficiency at high-P_T was higher for the QCD sample than the Z' bb sample.

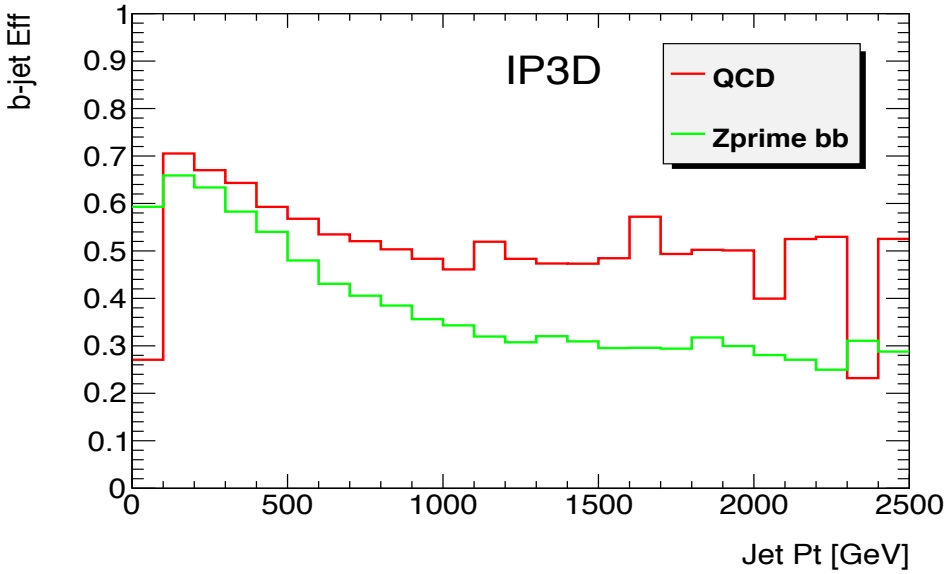
https://indico.cern.ch/event/387897/session/1/contribution/11/material/slides/0.pdf

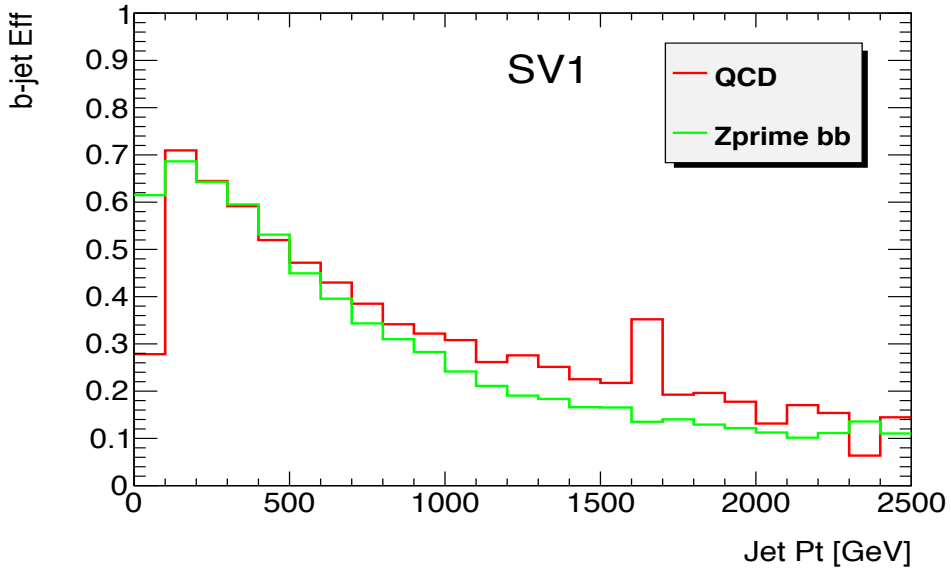
This is shown by plots on right

- Integrated efficiency cut of 70% (Single cut on discriminant)

- Cut on discriminant found using ttbar sample.

- Shows that efficiency falls off faster for Z' bb sample than for QCD sample



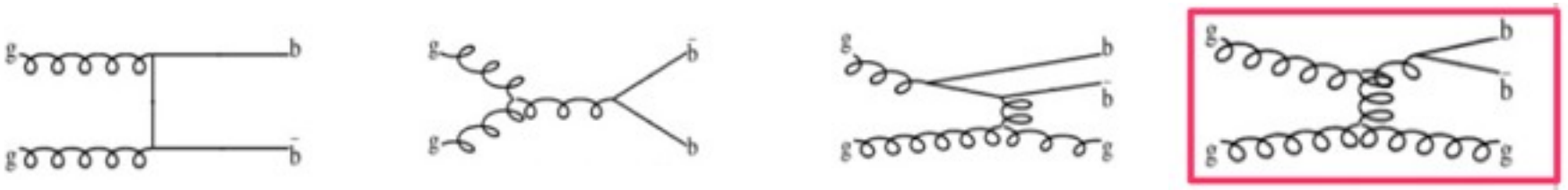


Gluon Splitting in QCD sample?

- The discrepancy could be because in the QCD sample, gluon splitting can occur in high-P_T light/charm jets forming lower-P_T b quarks, which causes a b-jet.

- These jets would have high jet-P_T but lower B-Hadron P_T meaning that the b-tagging performance would be equivalent to a low-P_T jet.

- This would mean that in the QCD sample we have some high-P_T jets with higher b-tagging efficiency, explaining the discrepancy.



Flavour Creation | Flavour Excitation | Gluon Splitting

The effect of the gluon splitting can be seen in an MC sample by...

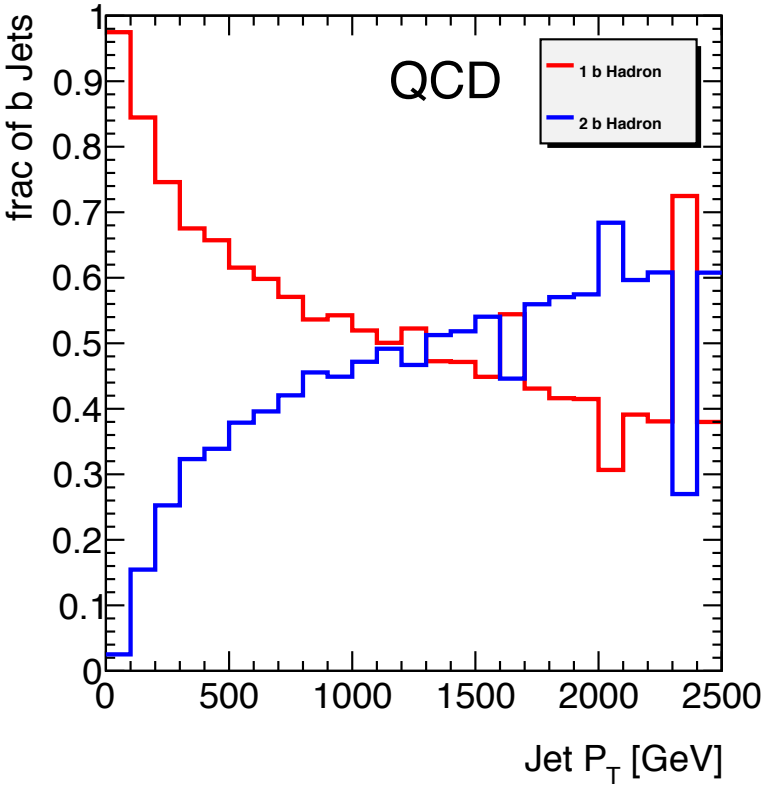
- Plotting against b-Hadron P_T instead of jet-P_T.
- Separating jets that are matched with 1 B-Hadron and 2 B-Hadrons. This is because flavour creation will lead to jets matched with only 1 B-Hadron, where gluon splitting will lead to 2 B-Hadrons.

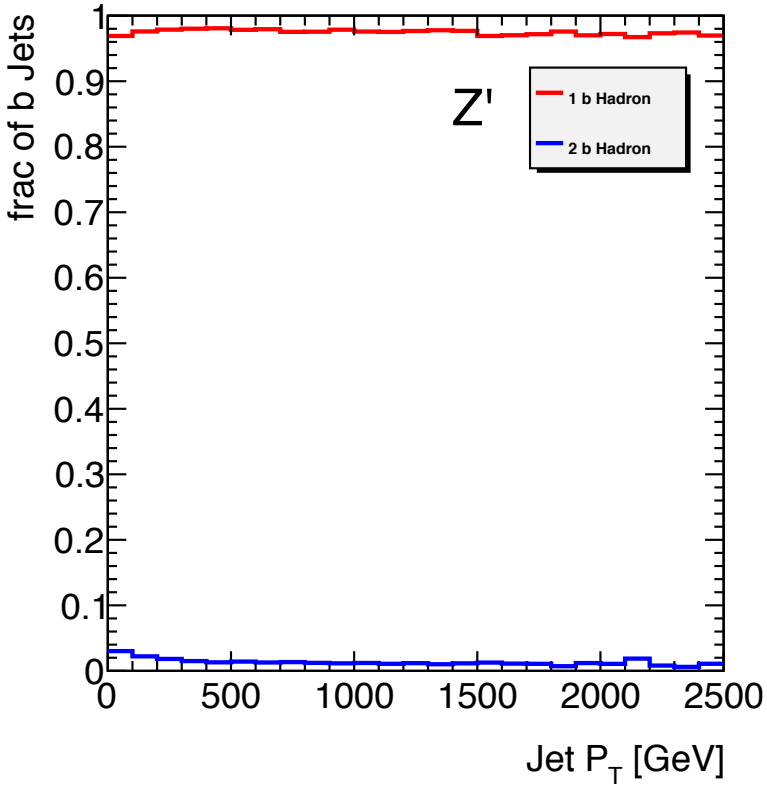
Number of b Hadrons in a b-Jet

- Plots to show number of b-jets that are made contain 1 or 2 b-Hadrons against P_T

- Very few jets with 2 b-Hadrons in Z' bb sample, to be expected.

- Considerable fraction of b-jets consist of 2 b-Hadrons at high jet-P_T in the QCD sample.
- However very few 2 b-Hadron jets at high B Hadron P_T. Suggests gluon splitting occurring.

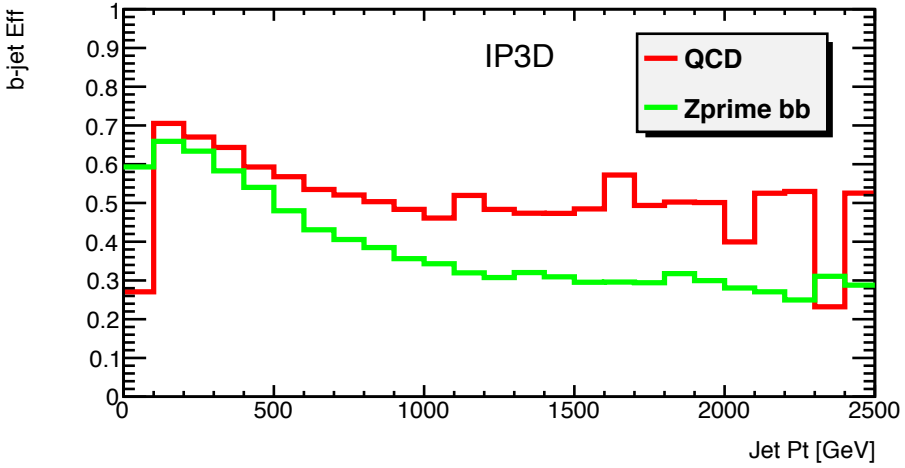


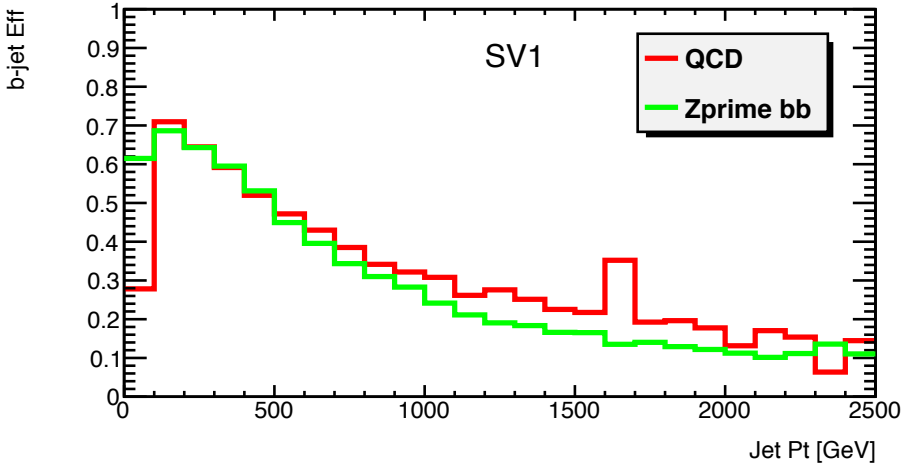


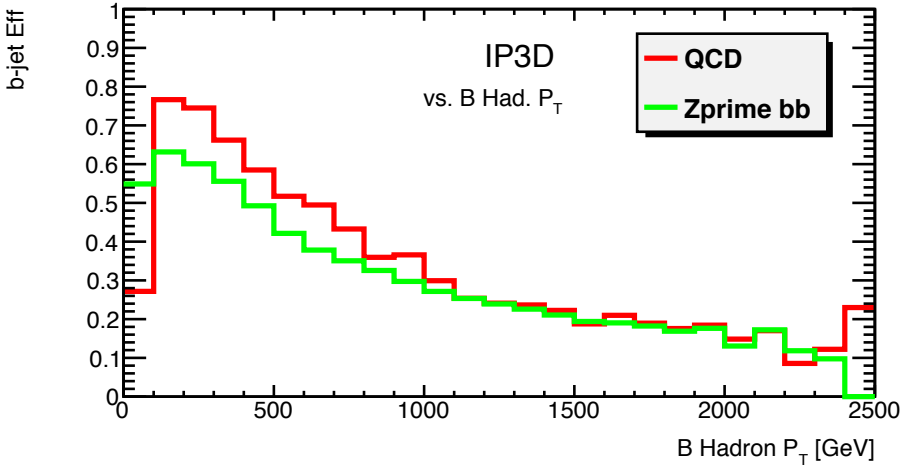
b-Tag Efficiency vs b Hadron P_T

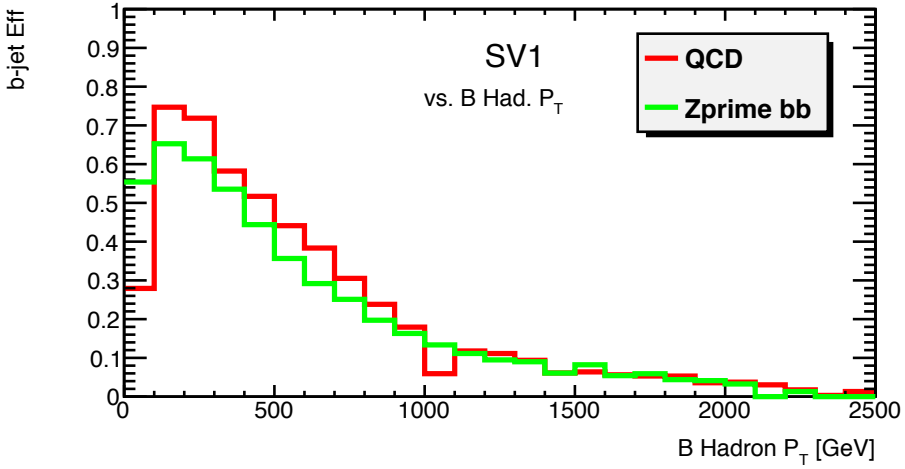
- Discrepancy reduced when plotting against b-Hadron P_T.

- Consist with gluon splitting is responsible for the improved high-P_T b-tag efficiency at high jet-P_T





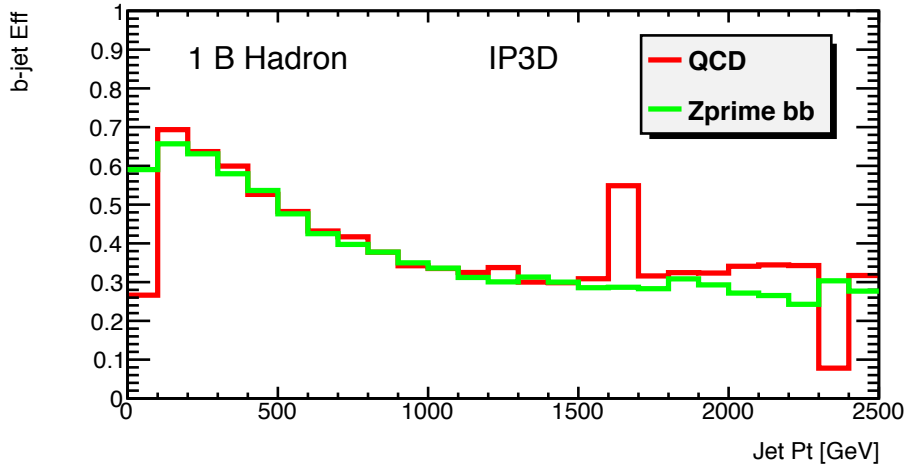


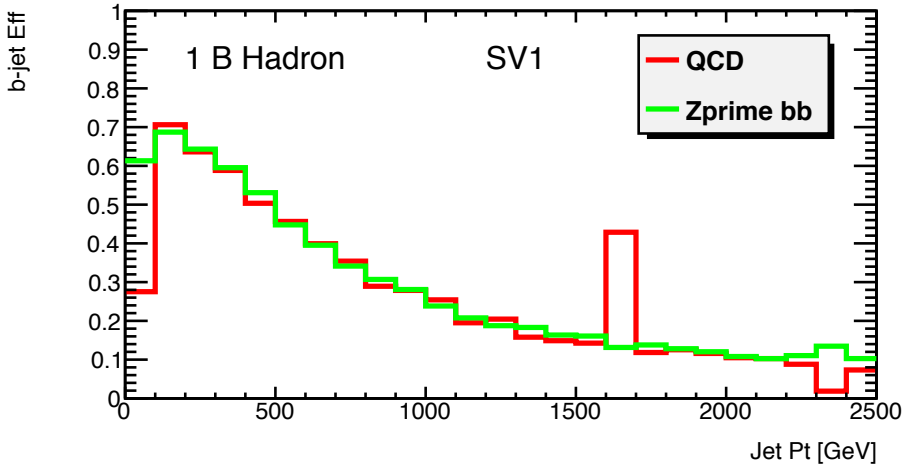


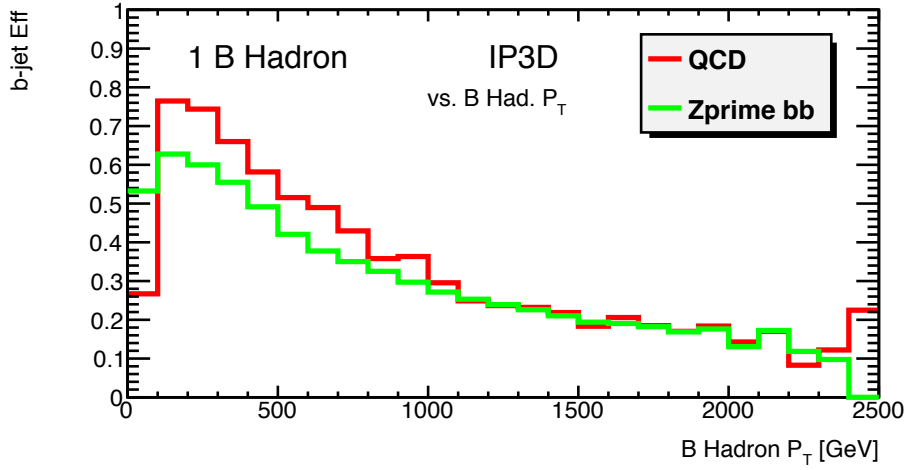
b-Tag Efficiency for 1 B Hadron Jets

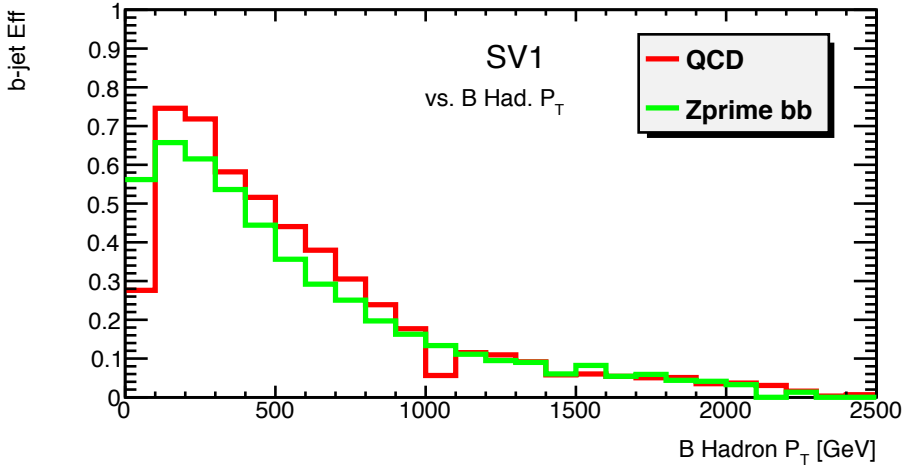
- Discrepancy reduced when plotting only jets containing exactly 1 B Hadron, which will reduce amount of gluon splitting occurring.

- Again consistent with gluon splitting is responsible for the improved high-P_T b-tag efficiency at high jet-P_T

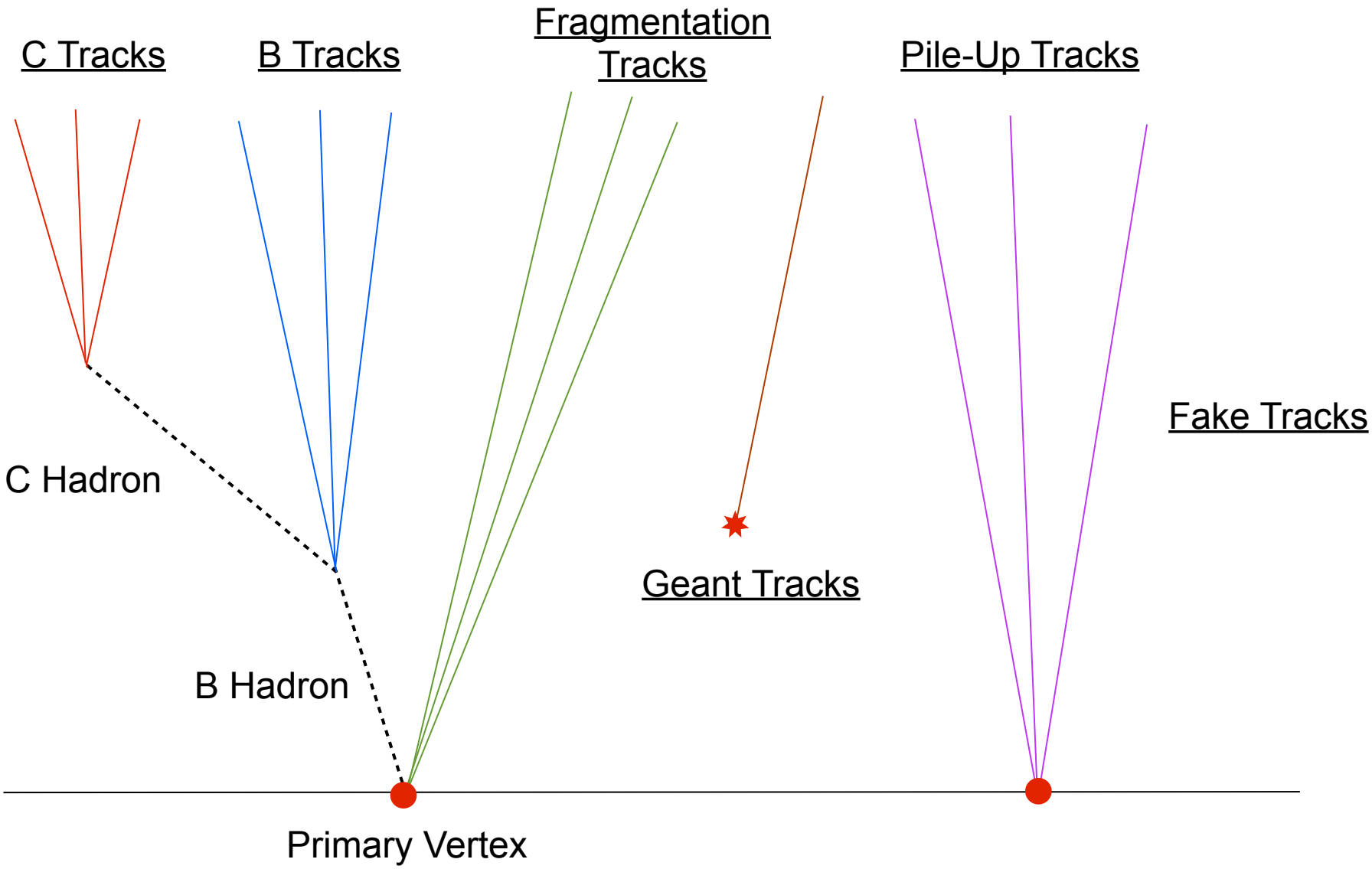








Origins of Tracks in a b-jet



Track Cuts and Some Definitions

Track Cuts

From Talk by R. Zaidan at Flav Tag Workshop 2015

In this study I have applied these cuts manually in my analysis code.

	IP3D	SV1	JFit
$p_T \geq$	1000	700	769.2
$\lvert\eta\rvert \leq$	2.5	2.5	2.5
$N_{SI} \geq$	7	7	7
$N_{SCT} \geq$	-	4	4
$N_{PIX} \geq$	2	1	1
$N_{IBL} \geq$	1	-	-
$N_{IBL} + N_{BL} \geq$	-	-	-
$N^{SH}_{PIX} + \frac{N^{SH}_{SCT}}{2} \leq$	-	-	1
$N^{HOLE}_{SI} \leq$	-	-	-
$N^{HOLE}_{PIX} \leq$	-	-	-
$d_0 \leq$	1	5	3.5
$z_0 * \sin(\theta) \leq$	1.5	25	5
$\sigma(d_0) \leq$	-	1	0.35
$\sigma(z_0) \leq$	-	5	2.5
$\chi 2/NDF \leq$	-	-	3.5

Definitions

From B = Any track associated to the decay of the B or C Hadron

From PuFake = Any track without a well matched truth particle.

From Geant = Any track created by a GEANT interaction.

From Frag = Any track not From B or From Geant

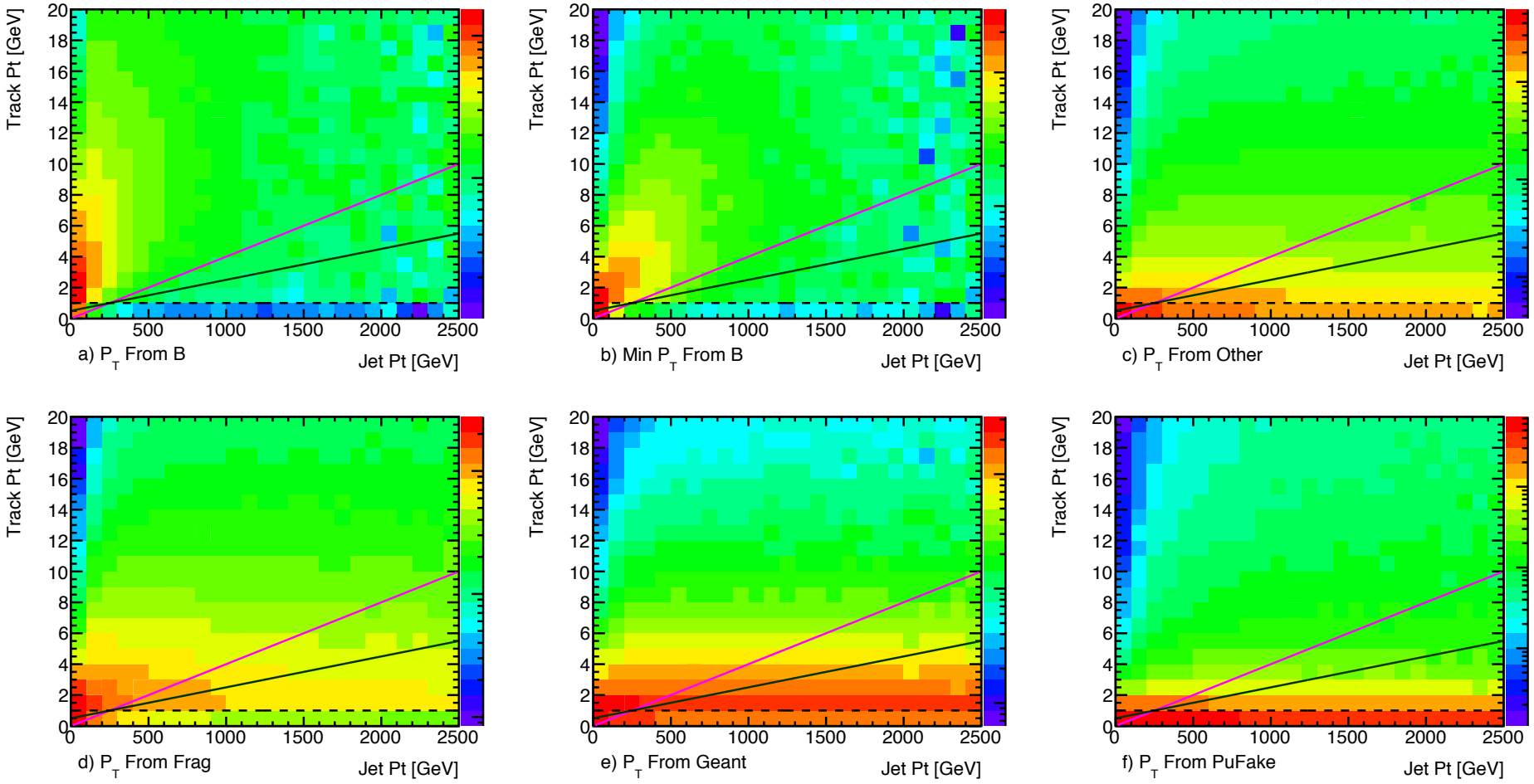
From Other = Any track not From B

Selection Efficiency = $\frac{\text{\# Tracks FromX Selected By a Cut}}{\text{\# Truth Tracks From X}}$

Fraction of Tracks = $\frac{\text{\# Tracks FromX Selected By a Cut}}{\text{Total \# Tracks Selected by a Cut}}$

Track P_T Distributions for a given Jet P_T

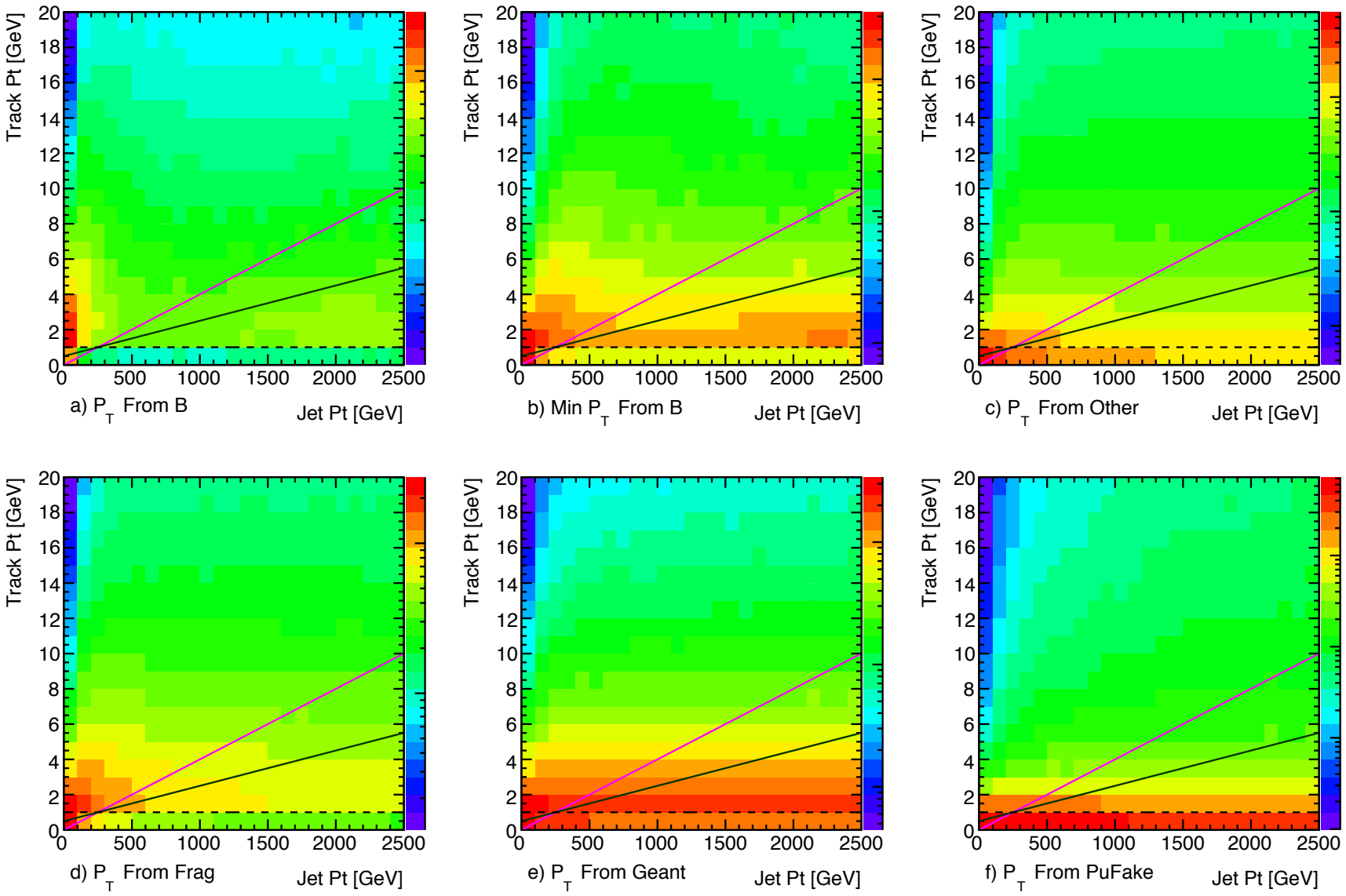
Z' bb sample *mc14_13TeV.1913*.Pythia8_AU2MSTW2008LO_Zprimebb*e3569_s2576_s2132_r6235*



- Track P_T distributions normalised for a given Jet P_T (vertical slices) including overfill bin.
- Evidence that one can use a jet-P_T dependant track cut to reduce contribution from non-B tracks.
- I have tried two variable P_T cuts, vpc.
- vpc1 (magenta line): Track-P_T = (Jet-P_T)*(0.004)
- vpc2 (dark green line): Track-P_T = (Jet-P_T)*(0.002) +0.5
- Existing IP3D cut at 1 GeV (black dotted line).

Track P_T Distributions for a given Jet P_T

QCD Sample

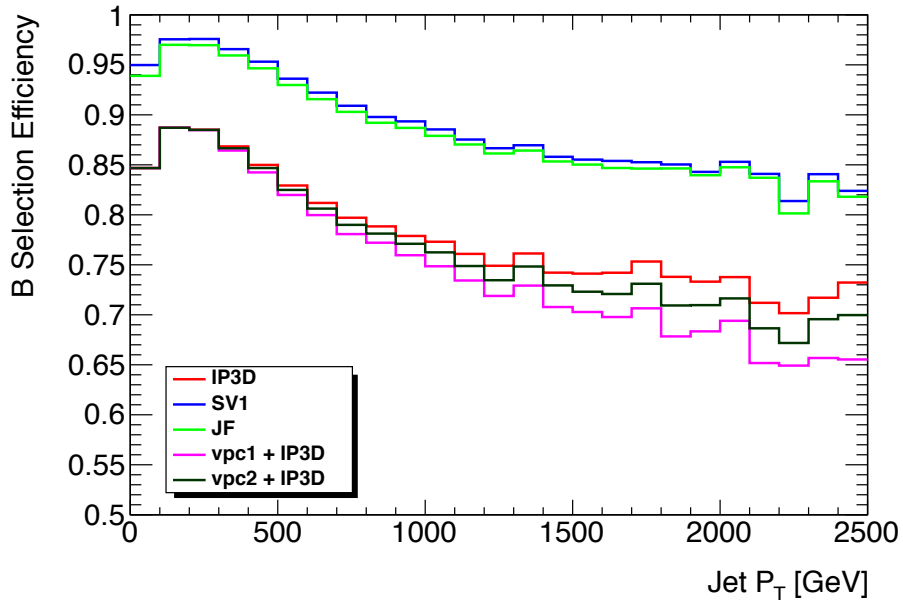


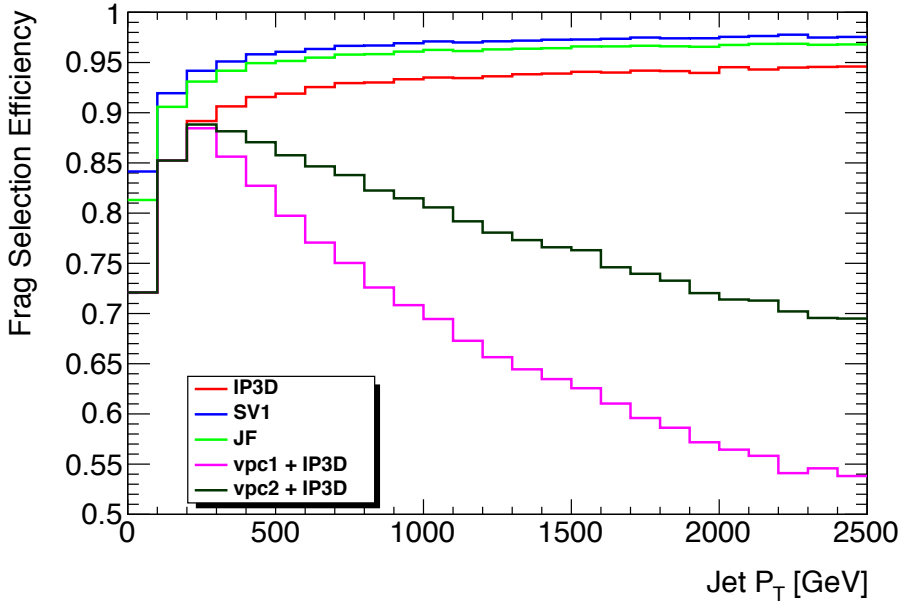
Track P_T distributions normalised for a given Jet P_T (vertical slices) including overfill bin.

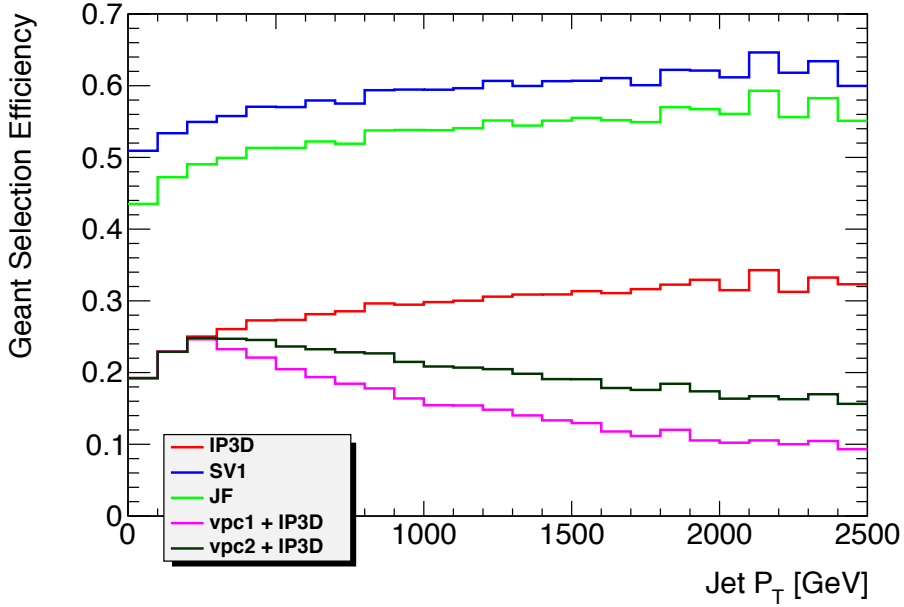
Selection Efficiencies with VarPtCut

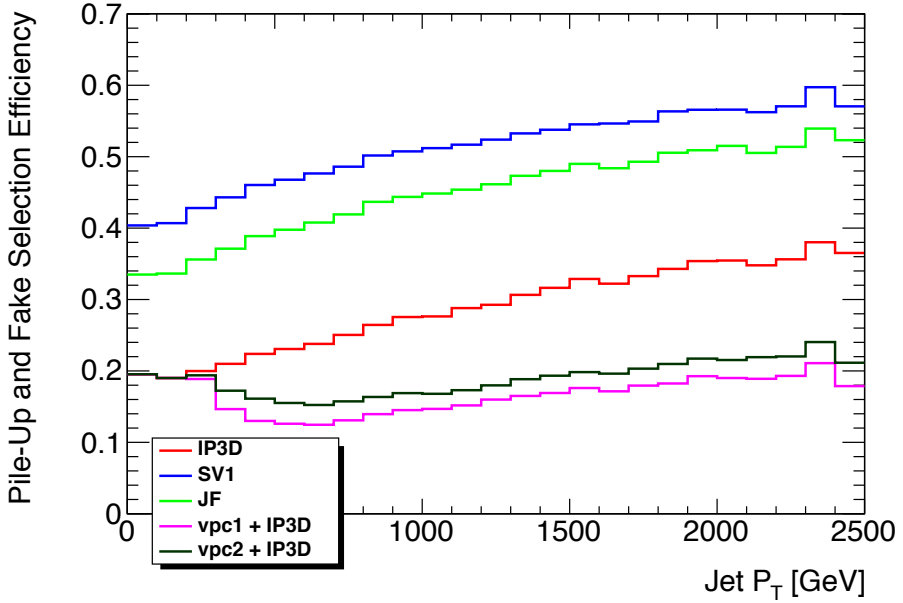
Z Prime bb

Selection on track cuts only - no btagger applied



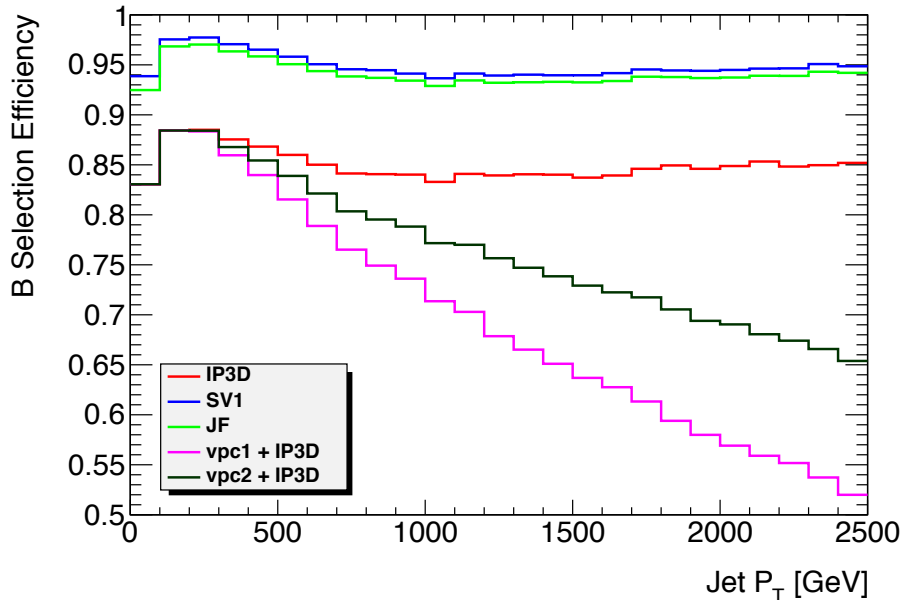




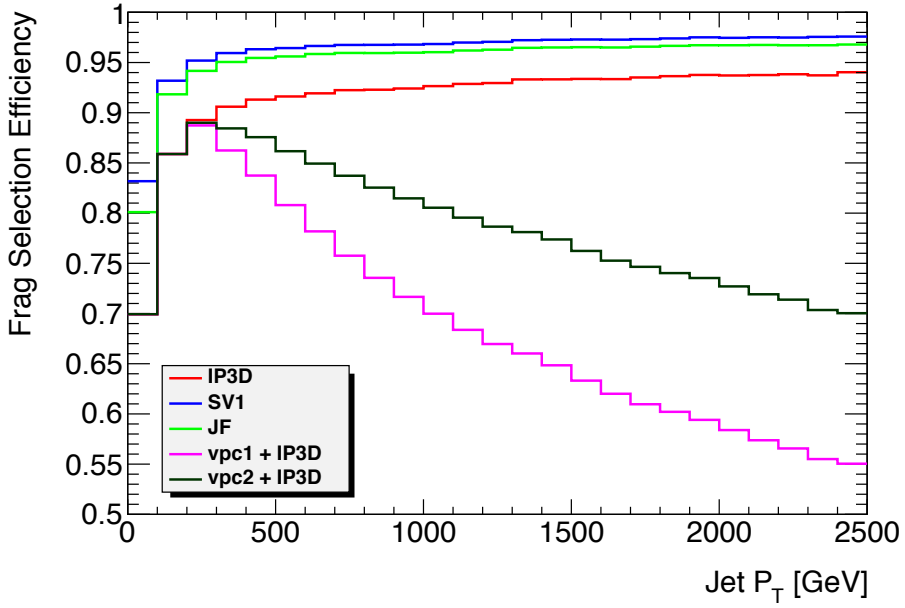


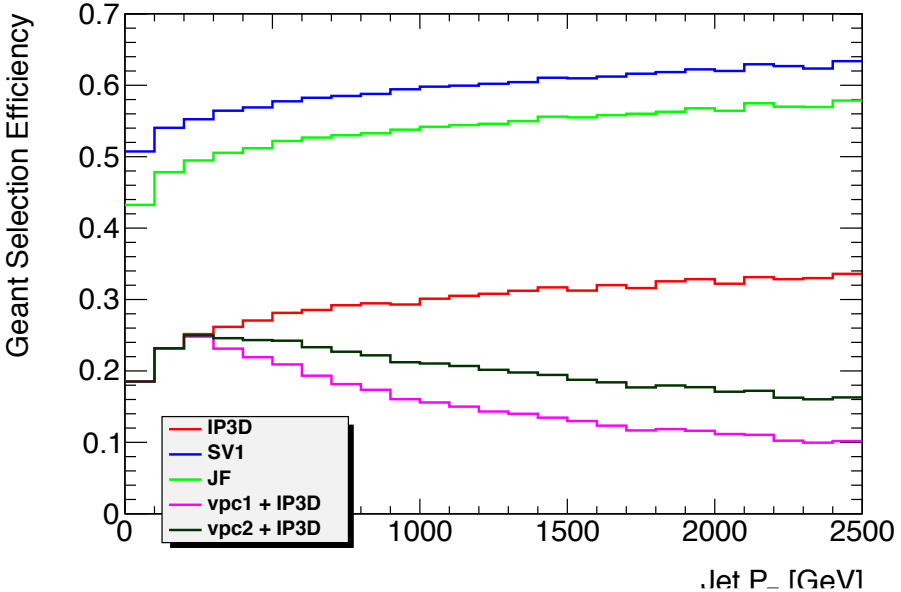
Selection Efficiencies with VarPtCut

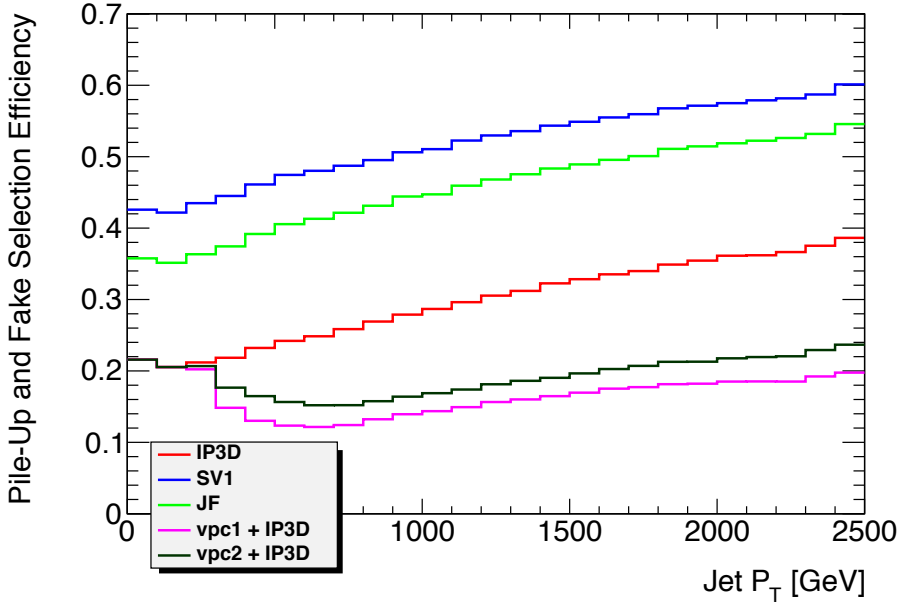
QCD



Selection on track cuts only - no btagger applied



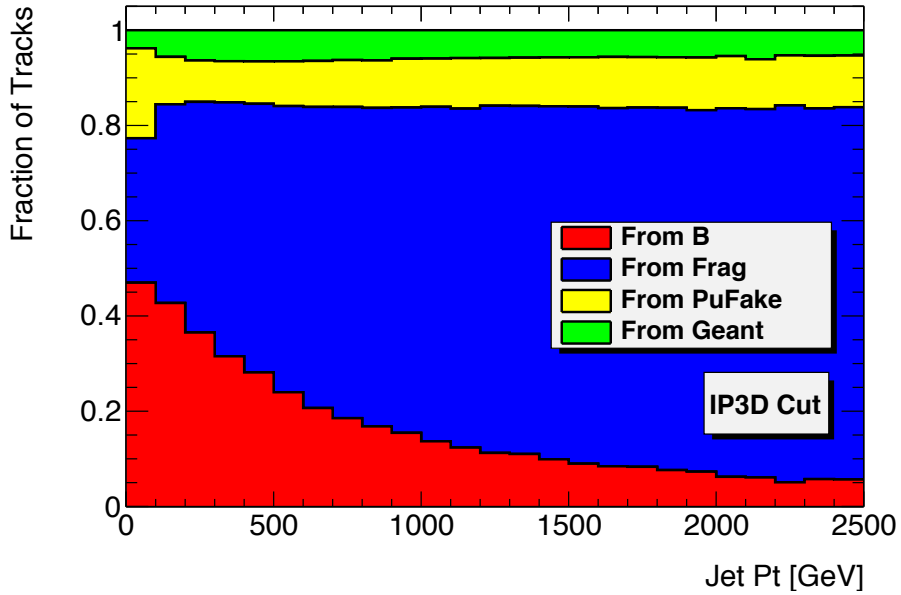


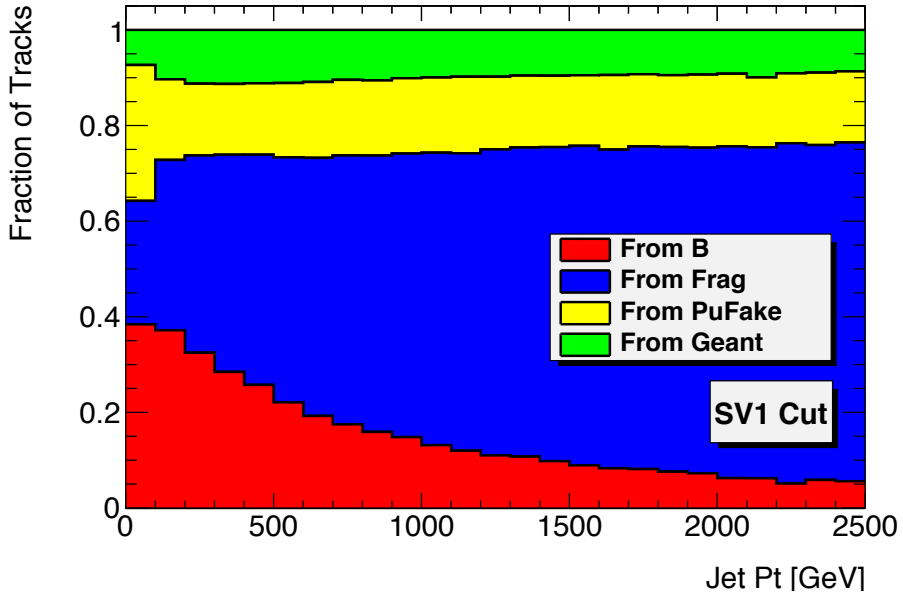


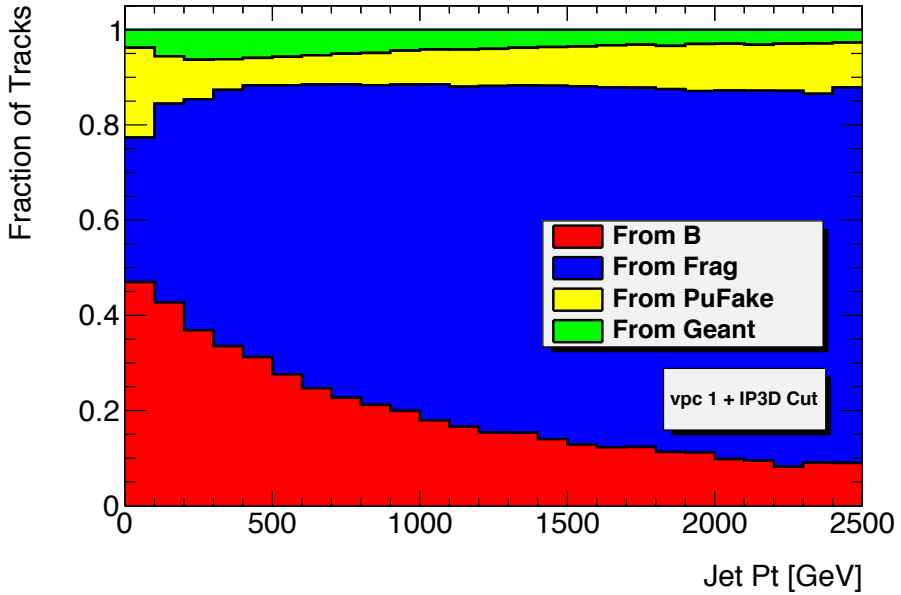
Fractional Make Up of b-Jet tracks

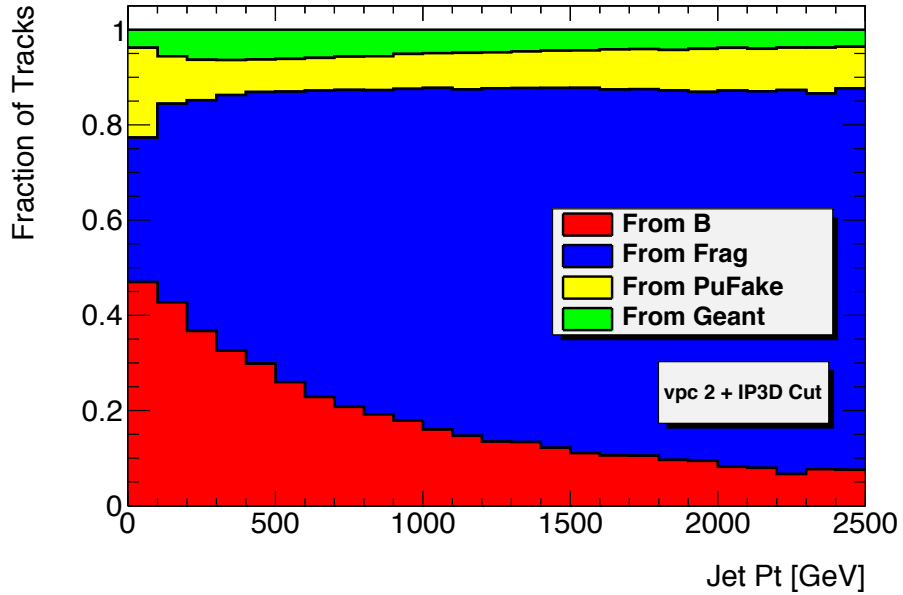
Z Prime bb

Selection on track cuts only - no btagger applied





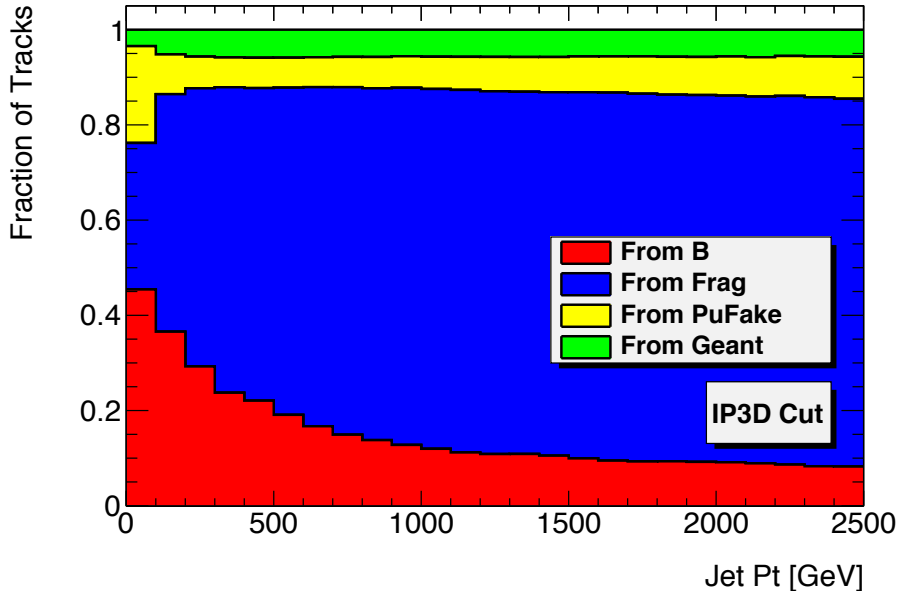


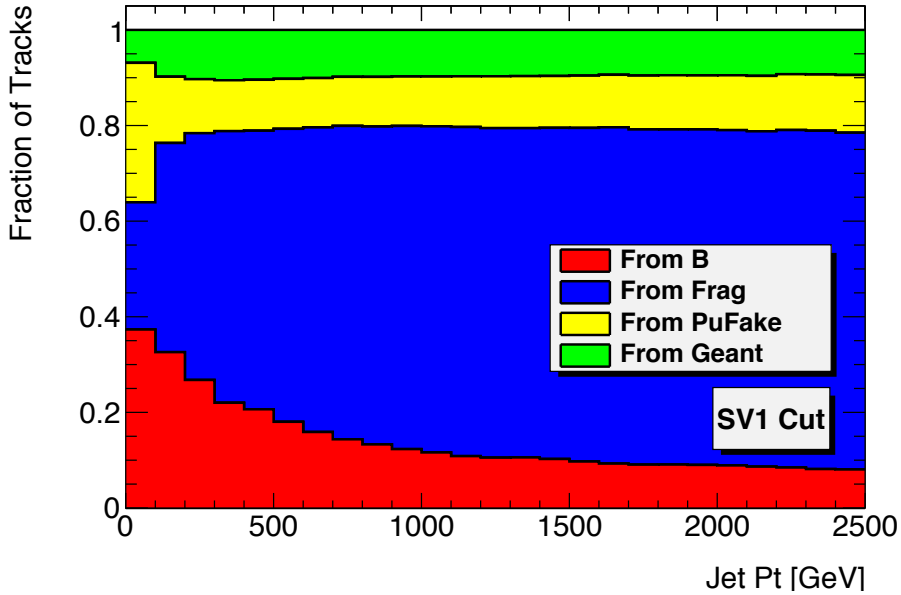


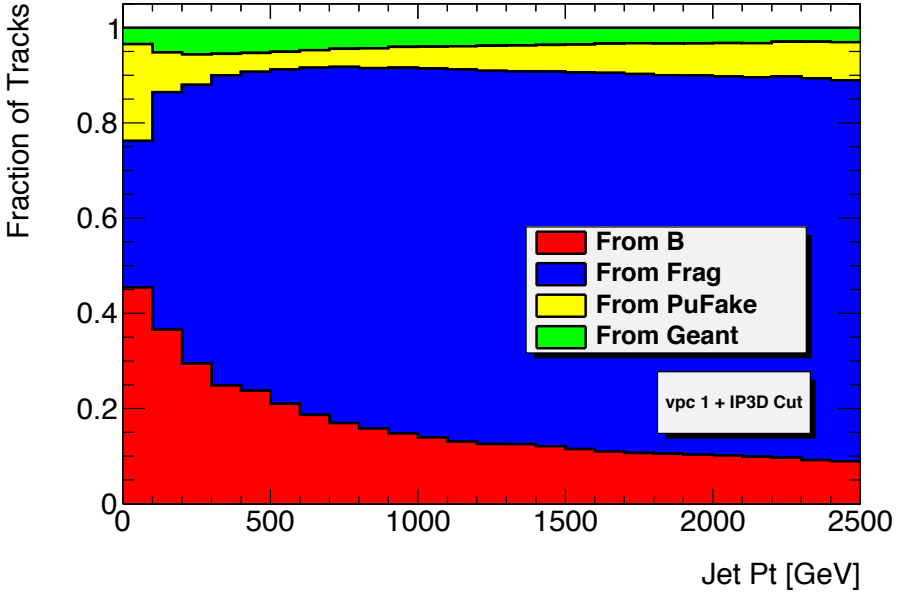
Fractional Make Up of b-Jet tracks

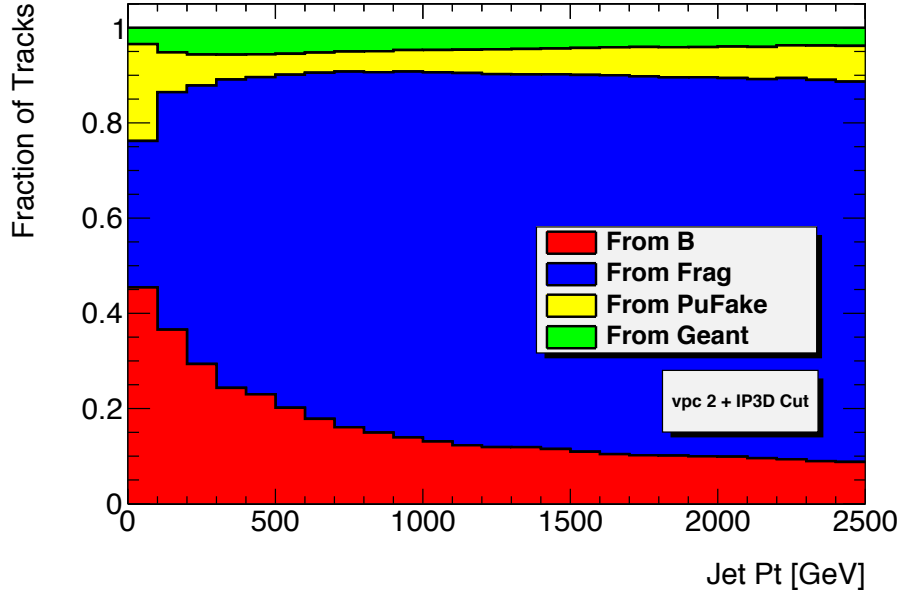
QCD

Selection on track cuts only - no btagger applied



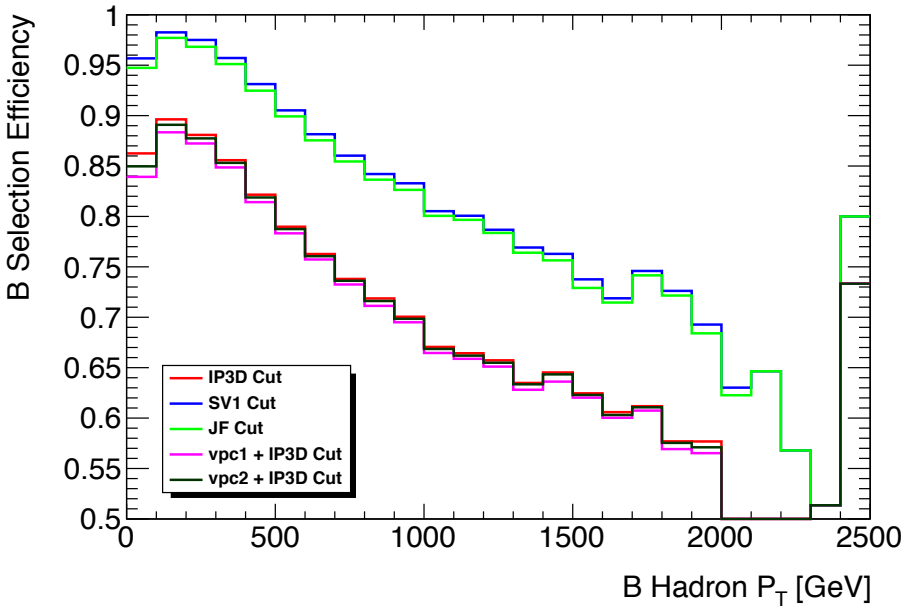




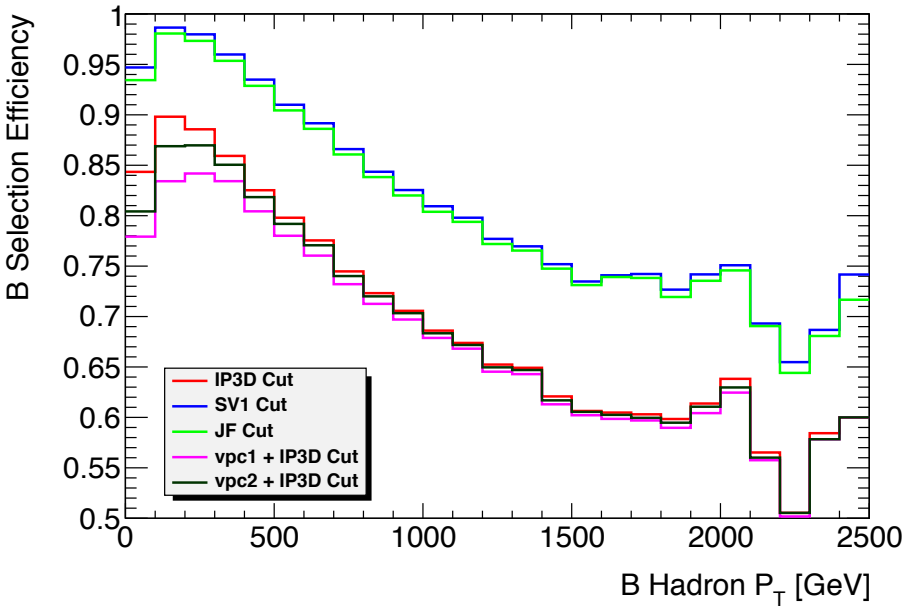


Selection Efficiencies against B-Hadron P_T

ZPrimebb



QCD



- The main discrepancy between the two samples is at low B-Hadron P_T, where the vpc cuts makes a large difference.

- But, the vpc cuts apply strongest at high jet-P_T, so this might show that the vpc cuts is effective at cutting b-jets from gluon splitting.

Conclusions

- We aimed to look at the discrepancy between QCD and ZPrimebb high-P_T b-tagging performances
  - We showed that this discrepancy was consistent with gluon splitting in the QCD sample.

- We aimed to study tracks accepted by IP3D, SV1 and Jet Fitter
  - To optimise track selections for high-P_T.

- There is evidence that a jet-P_T dependant cut can improve b-tagging performance at high-P_T.
  - Need to re-tag to produce ROC curves to show this!

- Perform a similar study to optimise d_0 and z_0 track selection at high-P_T

Back Up!

Number Sanity Check

	No Cut	IP3D Cut	IP3D+VPC1 Cut	IP3D+VPC2 Cut
All	16,589	12,239	7,099	8,880
From Geant	1,979	658	207	336
From PuFake	3,556	1,352	750	855
From Frag	10,075	9,527	5,499	7,008
From B	979	702	643	681
Frac				
From Geant	0.119	0.054	0.029	0.038
From PuFake	0.214	0.110	0.106	0.096
From Frag	0.607	0.778	0.775	0.789
From B	0.059	0.057	0.091	0.077
Eff				
From Geant		0.332	0.105	0.170
From PuFake		0.380	0.211	0.240
From Frag		0.946	0.546	0.696
From B		0.717	0.657	0.696

- Consider one P_T bin: 2300-2400 GeV

- Look at the raw number of tracks that pass different cuts. Does this make sense?