Commissioning of ATLAS Flavour Tagging Algorithms Using Run-2 Data

Laurie McClymont

Supervisor: Andreas Korn

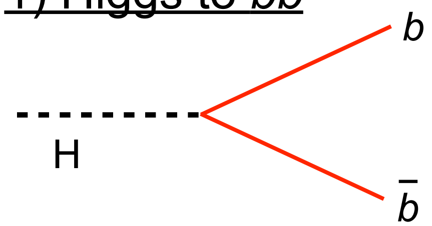
First Year Talk, UCL
28/09/2015

Contents

b-quarks at ATLAS

- *b*-quarks feature in many important processes for the ATLAS experiment.

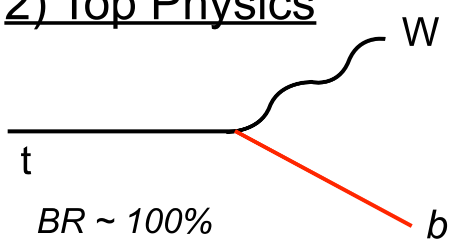
1) Higgs to $b\bar{b}$



Branching ratio ~ 57.7%

- Higgs coupling to *b*-quark has not been directly observed.
- Large mass of *b*-quark means strong coupling with Higgs.
- Higgs to $b\bar{b}$ is an important search at ATLAS in Run-2.

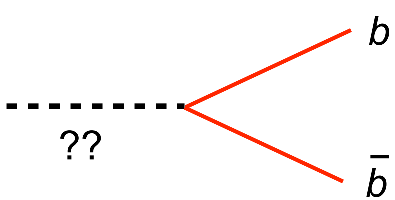
2) Top Physics



BR ~ 100%

- Top physics is an important part of the ATLAS physics program.
- The top decays to a W and *b* almost all the time.

3) Exotic Physics



BR ~ ??

- The *b*-quark has many interesting properties;
  - Large mass (m_b ~ 5 GeV)
  - Part of third generation of quarks.
- Hence the *b*-quark is important in many BSM searches.
  - More on this later...

- The ability to identify *b*-quarks is important! This is where flavour tagging comes in...

What is Flavour Tagging?

- A jet is a spray of particles formed from interactions that contain quarks or gluons in final state by their radiation and hadronisation.

- Jets are tagged in three different flavours:
  > *b*-jets = Contains a B-hadron.
  > *c*-jets = Contains a C-hadron (and no B-hadron).
  > light-flavoured jets = No B-hadron or C-hadron (everything else)

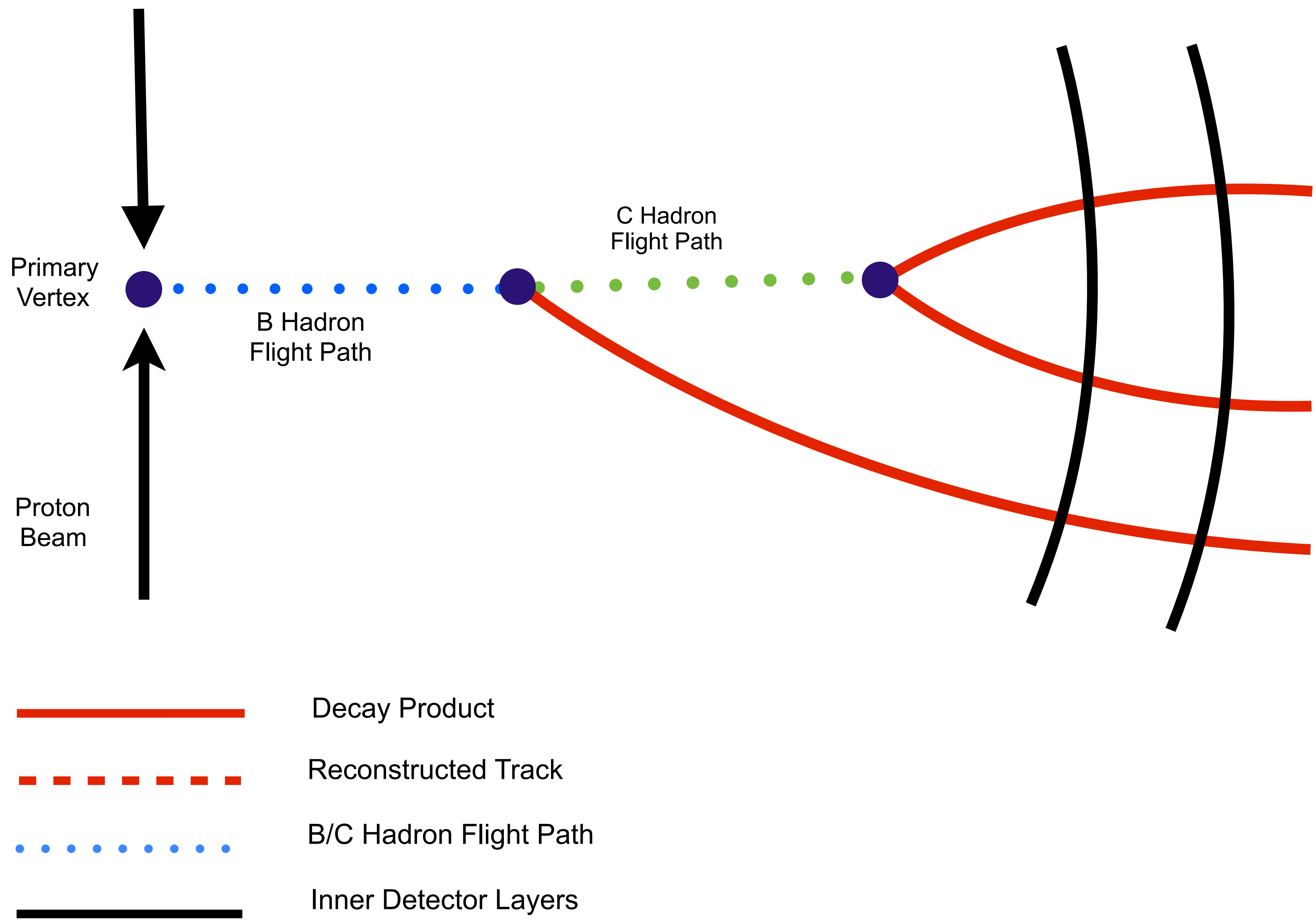
How is it done?

- Both the decay chain of B-hadrons and C-hadrons contains at least one flavour changing weak decay.

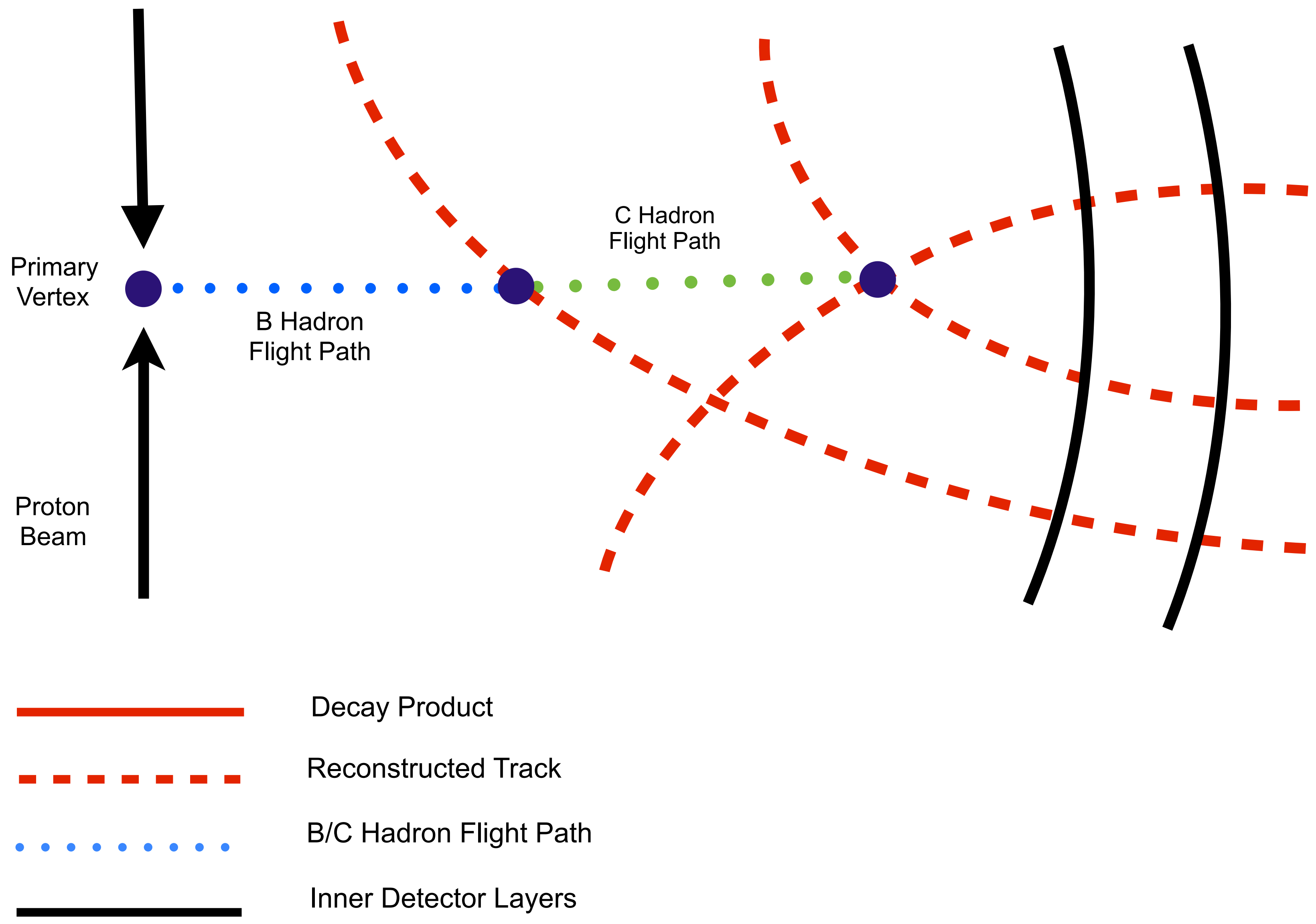
- As a result, B-hadrons and C-hadrons have a finite lifetime
  - $c\tau$ ~500 μm for B_0, B_+ and B_-

- Hence, heavy hadrons travel a finite distance before decaying.
  - A B_0 meson with a P_T of 50 GeV will travel ~5mm.
  - Offset decay vertices are the key flavour discriminating feature utilised.

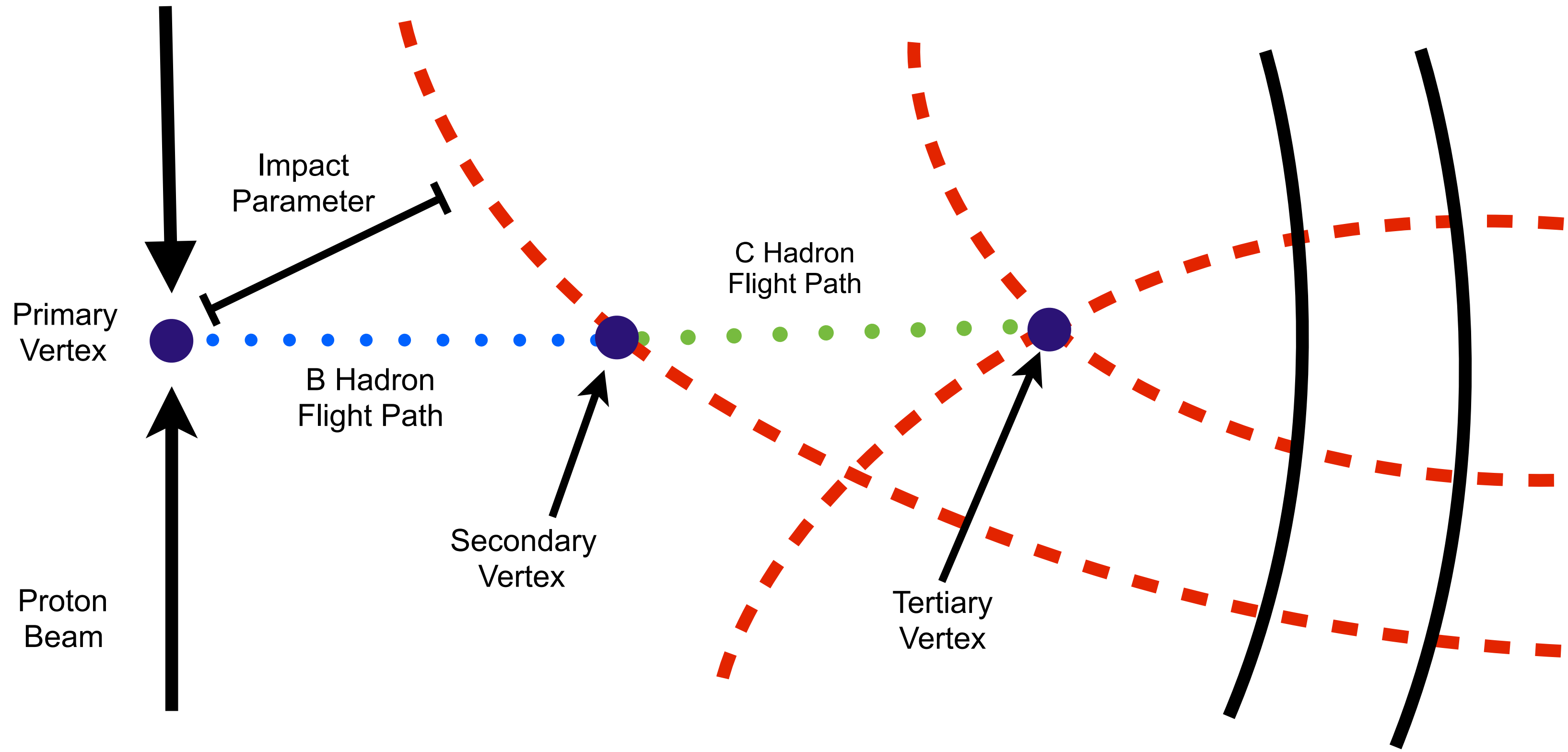
Flavour Tagging Algorithms at ATLAS



Flavour Tagging Algorithms at ATLAS



Flavour Tagging Algorithms at ATLAS



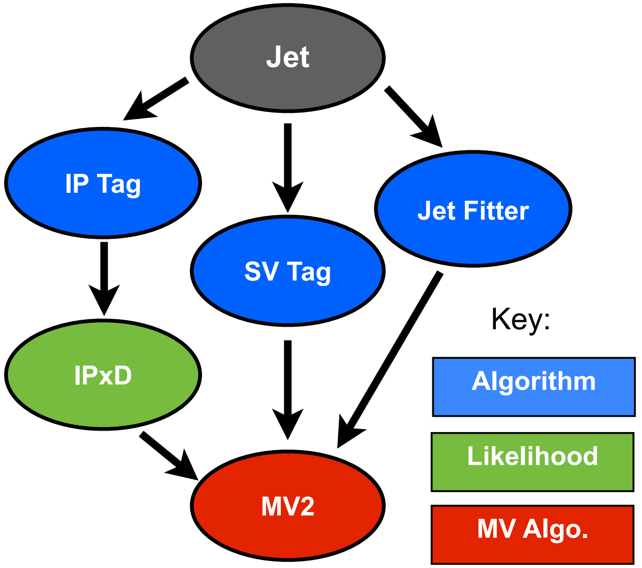
1. **Impact Parameter Based (IPxD)**
 - Uses the impact parameter distributions of tracks.

2. **Secondary Vertex Tagger (SV)**
- Reconstructs a secondary vertex from the crossing of two or more tracks
- Discriminating variables are SV mass, flight path length, number of tracks...

3. **Decay Chain Reconstruction [Jet Fitter (JF)]**
- Reconstructs many vertices along a flight-path axis.
- Discriminating variables are number of vertices, vertex mass, flight path length...

Multivariate Tagger (MV2)

- MV2 uses a Boosted Decision Tree (BDT) to combine the three base algorithms.
- MV2 gives the optimal tagging performance.
- MV2c20 is the recommended tagger for Run-2
  - Trained on sample containing 20% charm.
  - This produces an output that can reject both charm and light flavours effectively.

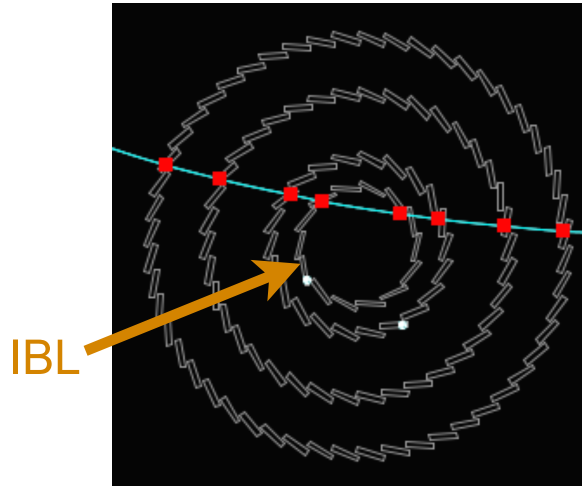


Contents

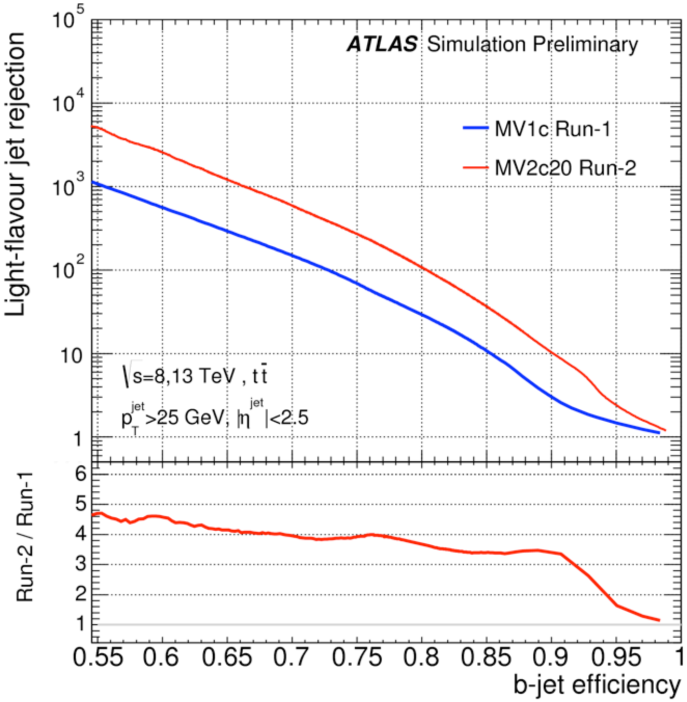
Motivations for Data Comissioning

Improvements to Flavour Tagging for Run-2

1) Insertable B-layer (IBL).
   - New innermost tracking layer located 33mm from the the beam line
2) Upgraded the multi-variate tagger to MV2
3) Improvements to the base tagging algorithms.



These changes have been tested and optimised using Monte Carlo (MC) Simulation
- See Andy's talk...



ATL-PHYS-PUB-2015-022

Data Commissioning

- However, we need to show that we the flavour tagging performance found in MC is also seen in data.
- This can be done by comparing Monte-Carlo Simulation to Data for key flavour tagging variables.

Samples and Event Selection

Samples

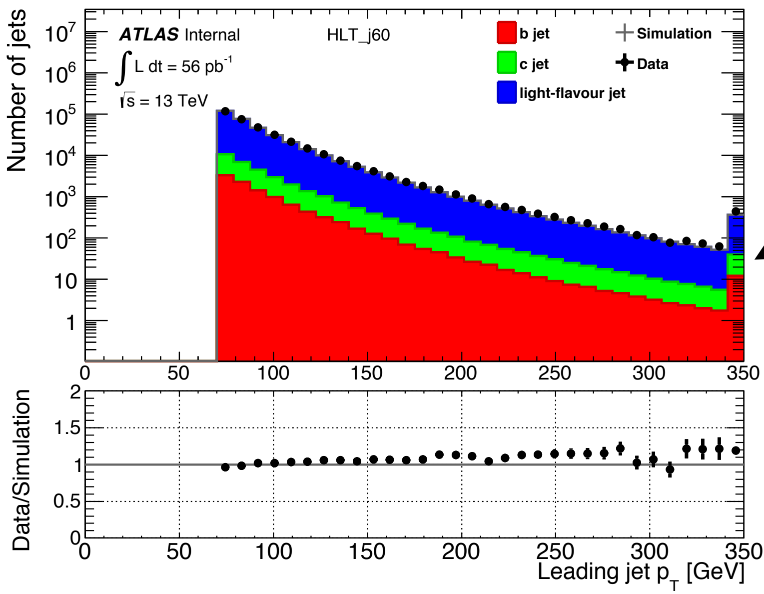
- Data - 56 pb^{-1} of early Run-2 data taken May-July 2015
  - Only use data if stable beam and calorimeters and inner detector working.
  - 13 TeV collision energy with a 50ns bunch spacing.
- Simulation - Monte Carlo simulation of QCD multi-dijet events
  - 50ns bunch spacing and 13 TeV collision energy.
  - Pythia 8 used to simulate the events, EVTGEN used to decay heavy hadrons.
  - GEANT4 model of ATLAS detector used to model detector response.
  - Reweight average number of interactions per beam crossing, $\langle\mu\rangle$, to match data.

Object and Event Selection

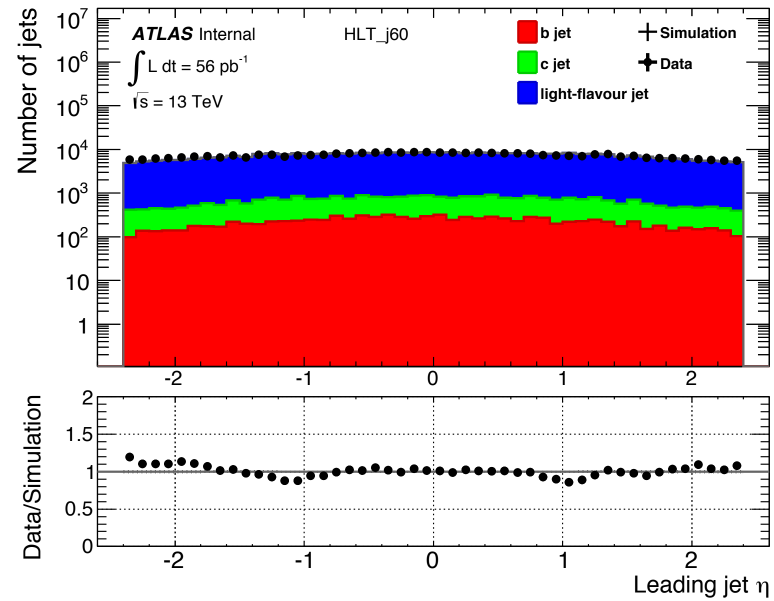
- Jet - Reclustered using the anti-k_T algorithm with R = 0.4
  - $P_T > 35$ GeV, $|\eta| < 2.4$
  - Jet Vertex Tagger (JVT) used to reduce pile-up jets.
- Trigger: Look for one high-P_T jet in event (HLT_j60)
- General Cuts
  - 2 jets in event (dijet event)
  - Leading jet-$P_T > 70$ GeV
- Monte-Carlo Cuts
  - Average P_T of two highest P_T jets < 1.4 * Truth Leading Jet P_T

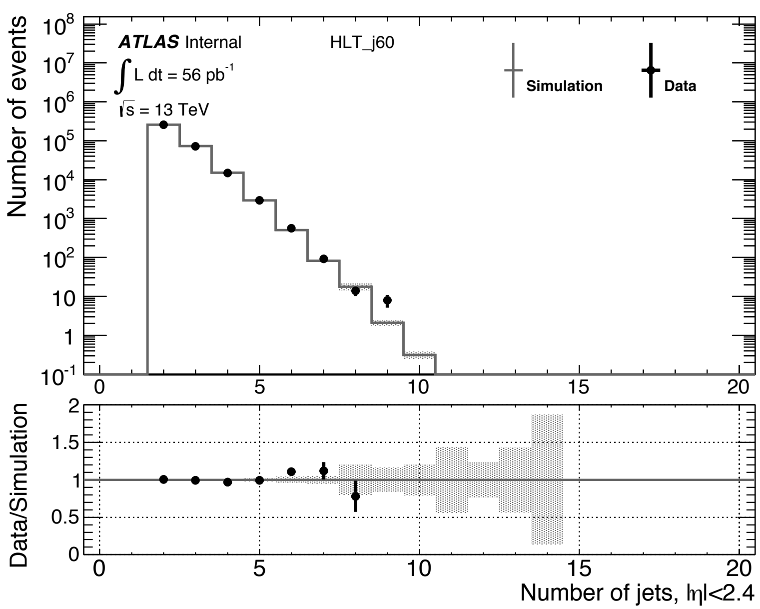
Jet Kinematic Distributions

Leading Jet:



Overfill Bin





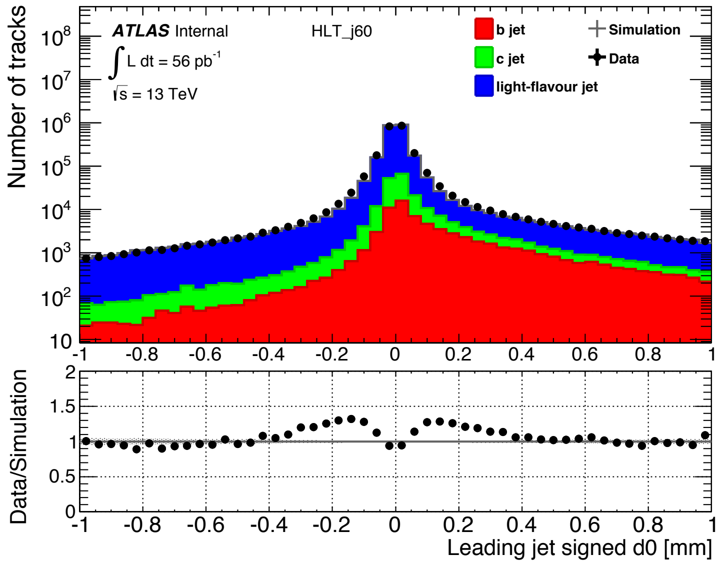
- Jet Kinematics important for flavour tagging.
- Data shown as solid points
- Simulation shown as a stack of its flavour fractions.
- Integral of simulation normalised to data.

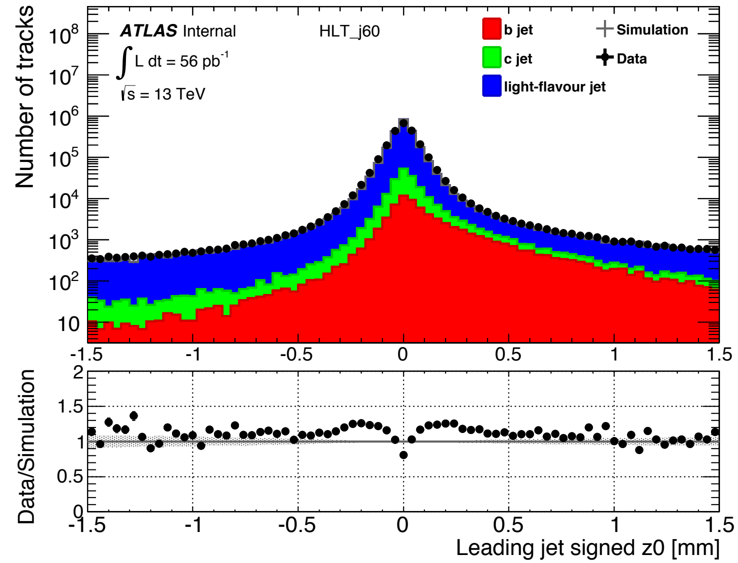
13 Impact Parameter Distributions

Transverse (d0)

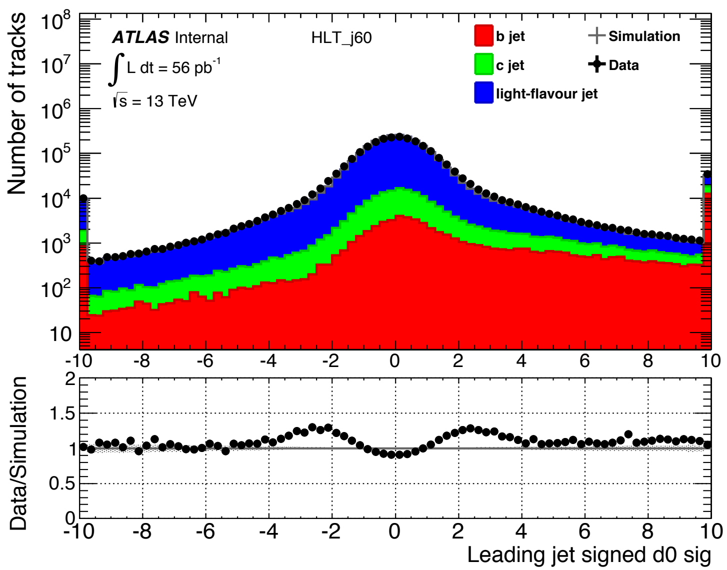
Longitudinal (z0)

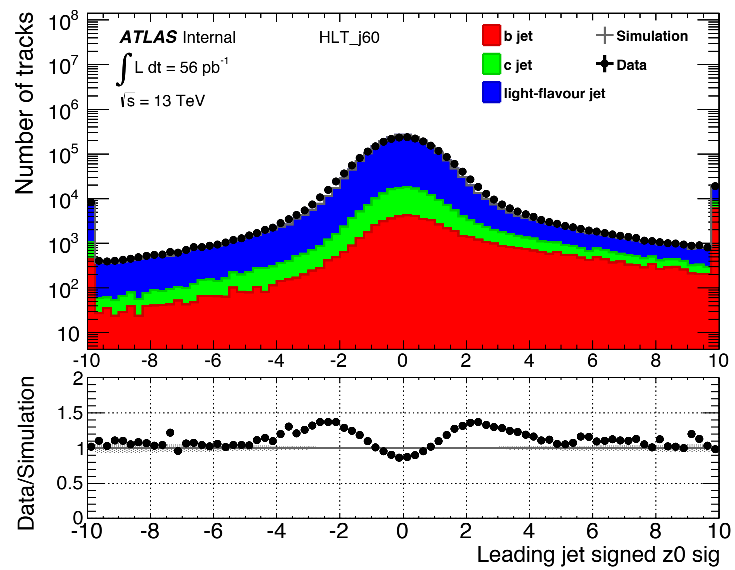
Impact Parameter





Impact Parameter Significance

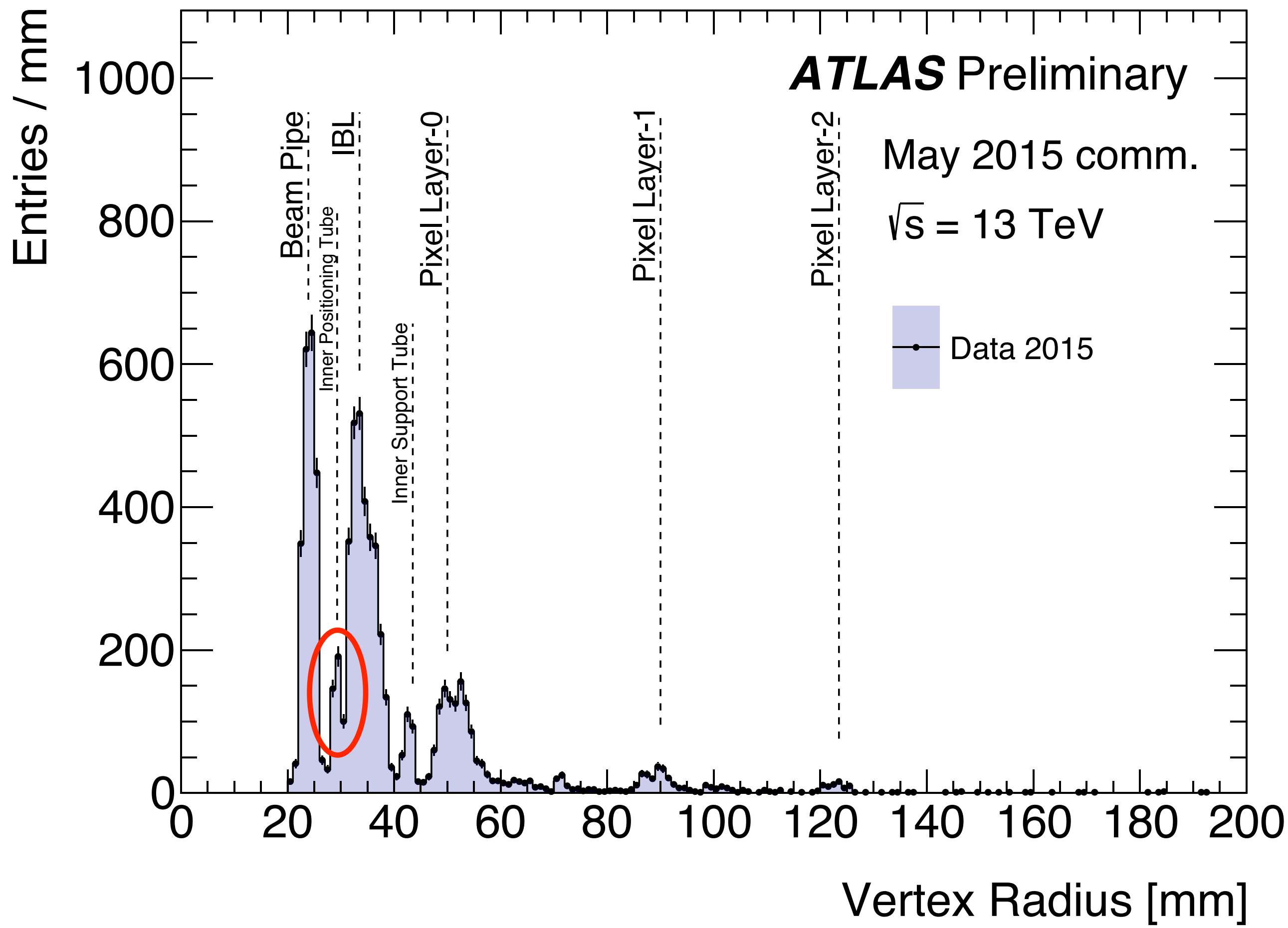




Why The Discrepancies?

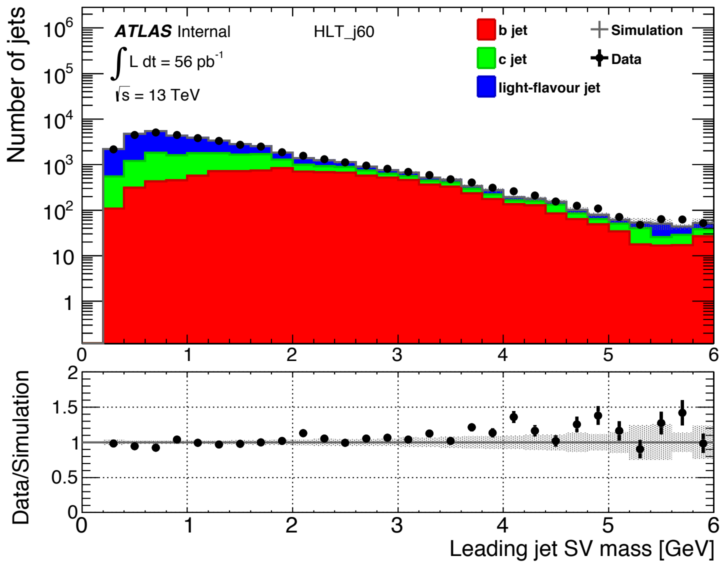
- Problems likely due to material mismodelling in simulation.
  - Believed that IBL material is underestimated.

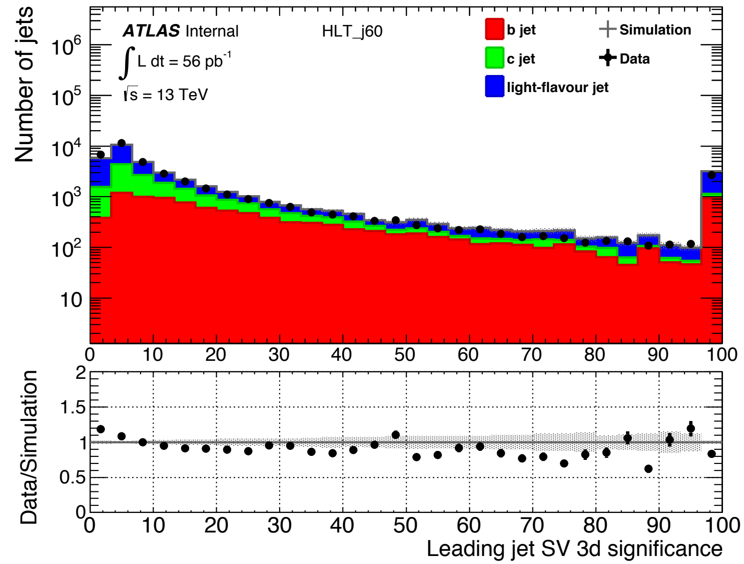
- Working with tracking experts to solve this problem.
  - New geometry model undergoing validation.



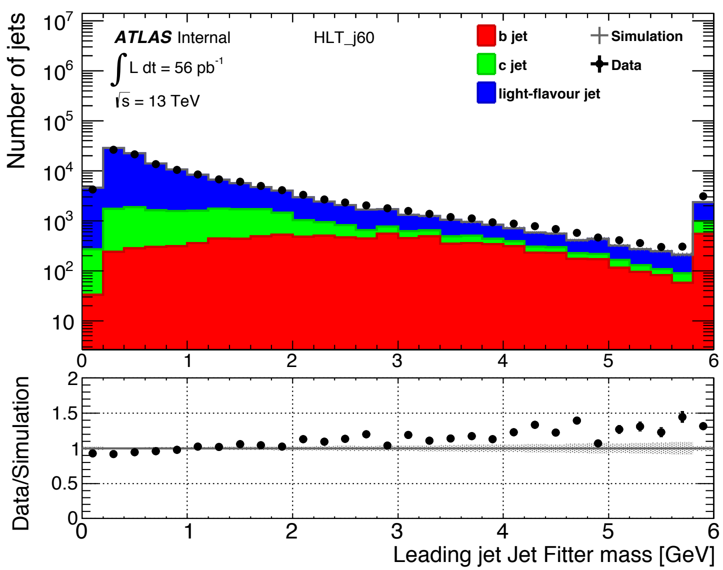
SV1 and Jet Fitter

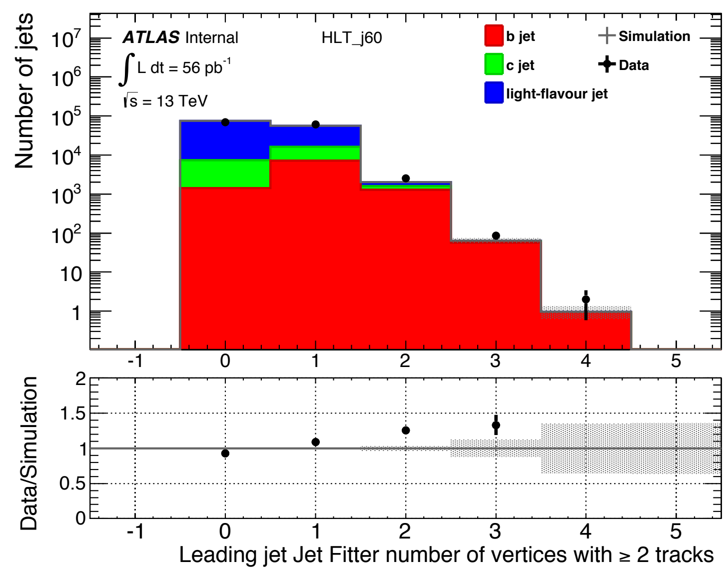
Secondary Vertex





Jet Fitter

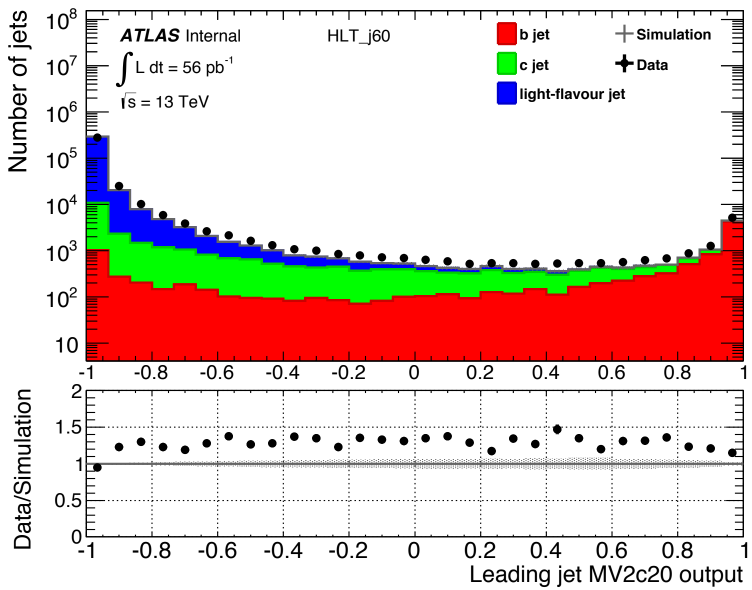




MV2c20

- MV2c20 is the output of the BDT that combines the three base taggers.

-1 = Unlikely to be a b-jet ⟷ 1 = Highly Likely to be a b-jet



- This is the tagger that analyses use, hence most important to understand.

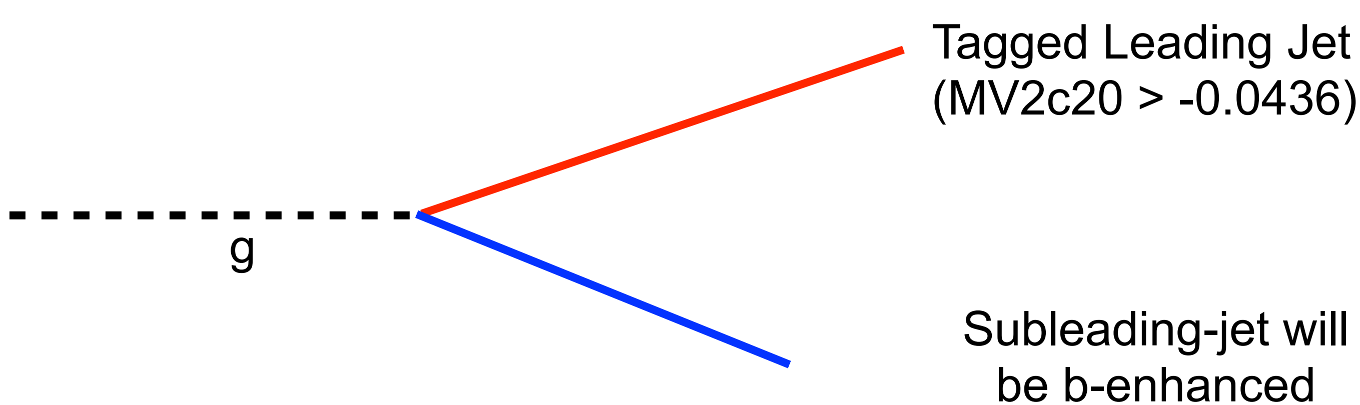
Unbiased b-Enhanced Sample

- As b-jets are most interesting to us we want to test our modelling in an sample that has an increased fraction of b-jets.

- However using b-tagging tools to increase fraction b-jets biases the sample
  - This is because we would only choose jets with "good" properties
  - i.e. A reconstructed SV.

- Hence what we do is **tag the leading jet**, and **study the subleading jet**.
  - Leading jet tagged at 70% efficiency point (MV2c20 > -0.0436)
  - Flavour correlation in QCD then means that the **subleading jet forms an unbiased b-enhanced sample**

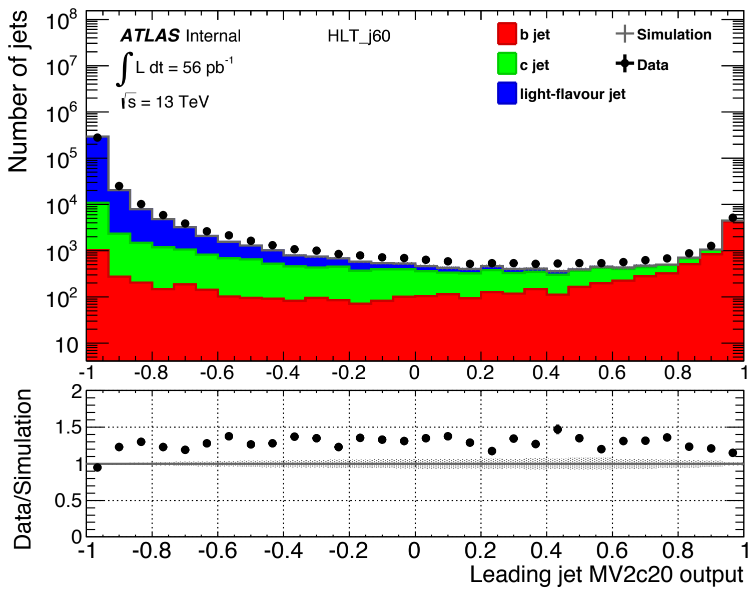
- **Fraction of *b*-jets increases from ~3% to ~14%**



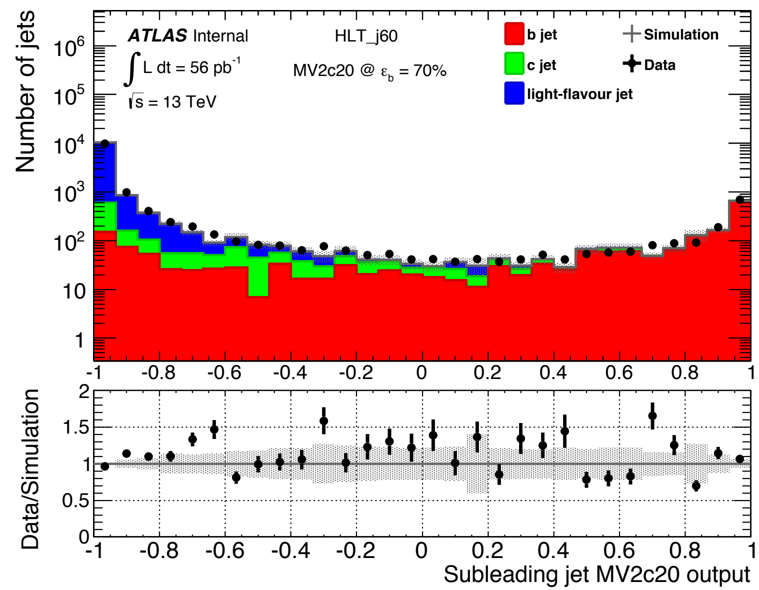
MV2c20 with b-enhanced sample

- MV2c20 is the output of the BDT that combines the three base taggers.
  -1 = Unlikely to be a b-jet ⟷ 1 = Highly Likely to be a b-jet

Leading jet sample:



b-enhanced sample.



- Clear discrepancy in lowest bin in leading jet sample.
- Why? Not sure...
  - Improved primary vertex selection.
  - Pile-up jets being mismodelled.
  - Mismodelling of flavour fractions?
- Further investigation required.

Contents

1) Introduction To Flavour Tagging

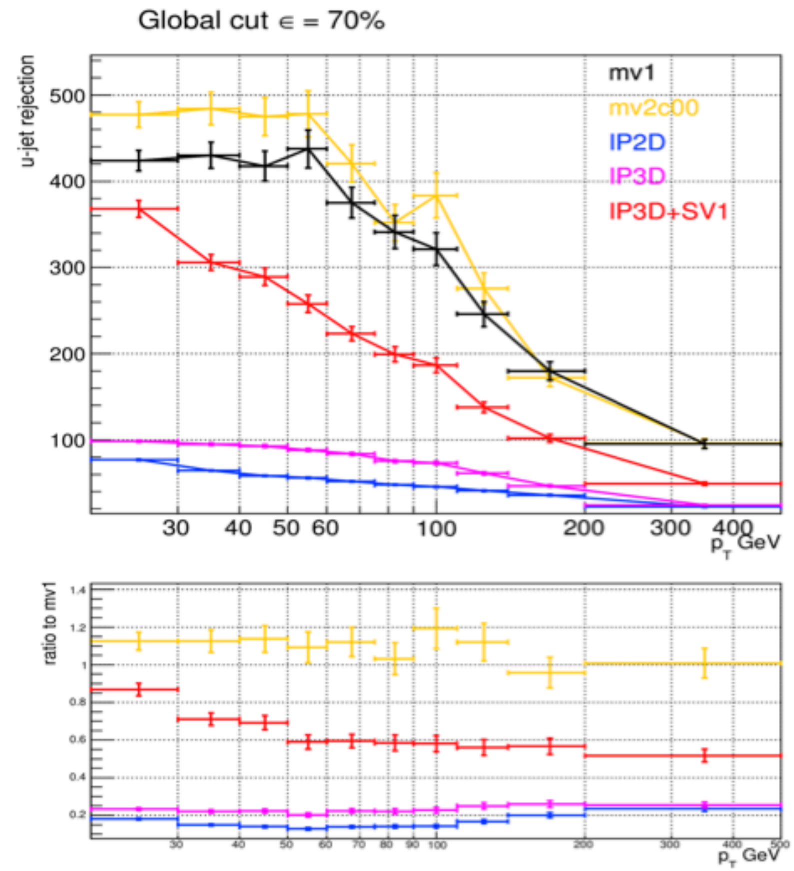
2) Data Commissioning

3) High-P_T *b*-tagging Improvements

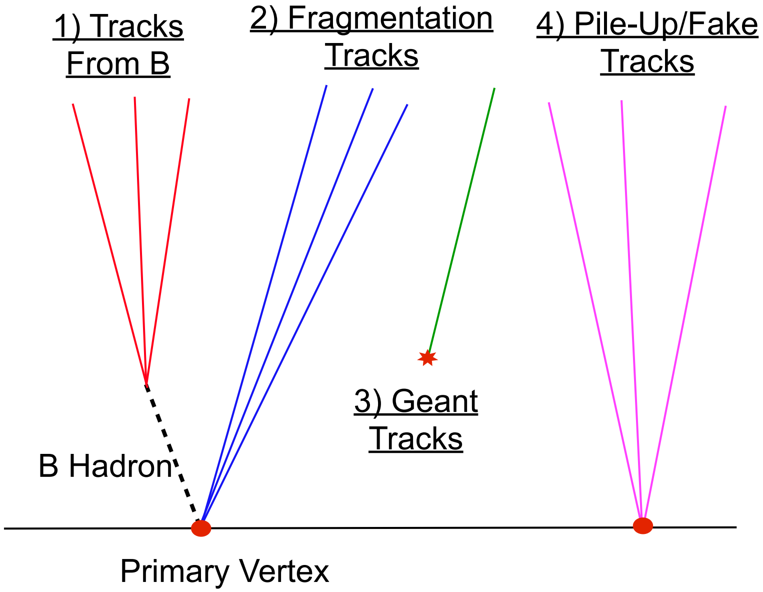
4) Exotic Di-b-jet Analysis

High P_T *b*-Tagging

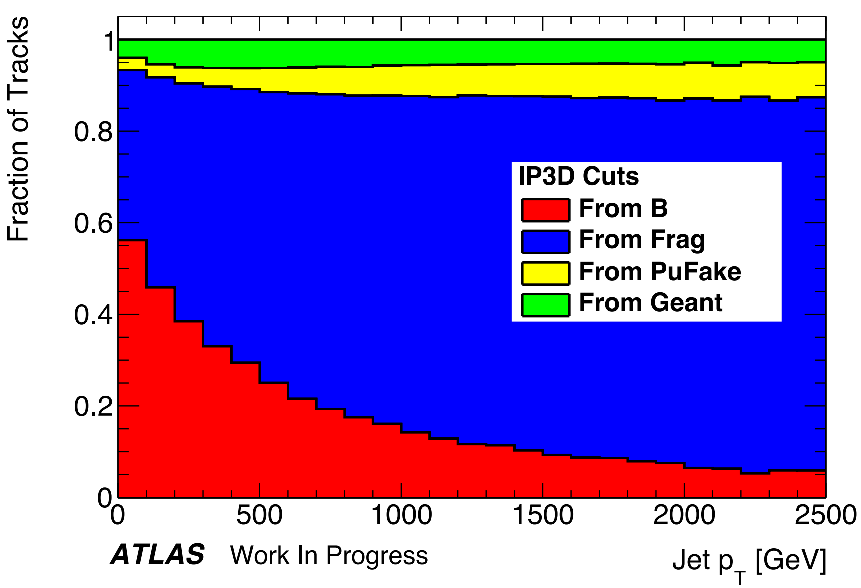
- Many analyses, in particular exotics, require high-P_T *b*-tagging.
- However, at high P_T there is a large drop in *b*-tagging performance.
- This is for a number of reasons.
  - More collimated jets, hence higher density of tracks.
  - Larger tracking errors associated with high-P_T tracks.
  - B-hadron can decay on the far side of the IBL.
  - Taggers are optomised for lower P_T jets.



Track Origins and Track Cuts



Z' bb sample:



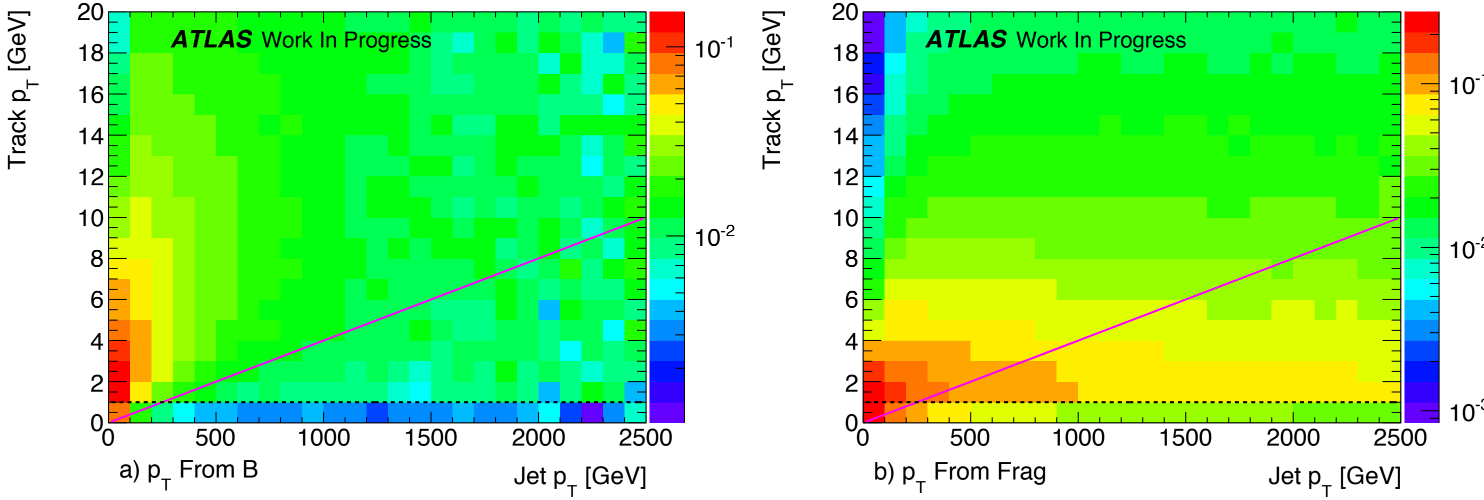
- Tracks in a *b*-jet can have several origins.
  - Tracks From B are the only ones that show the discriminating properties we look for.
- To ensure the quality of these tracks each algorithm uses a set of track cuts.
- IP3D Track Cuts include
  - $P_T > 1$ GeV
  - Impact parameter cuts
  - Inner detector hit requirements
- As P_T of jet increase, fragmentation begins to dominate.
  - Higher collimation of jets.
  - Higher energy quarks radiate more.
- Maybe we need a jet-p_T dependant set of track cuts?

Track Optimisation for High-P_T b-Tagging

Track P_T vs. Jet-P_T

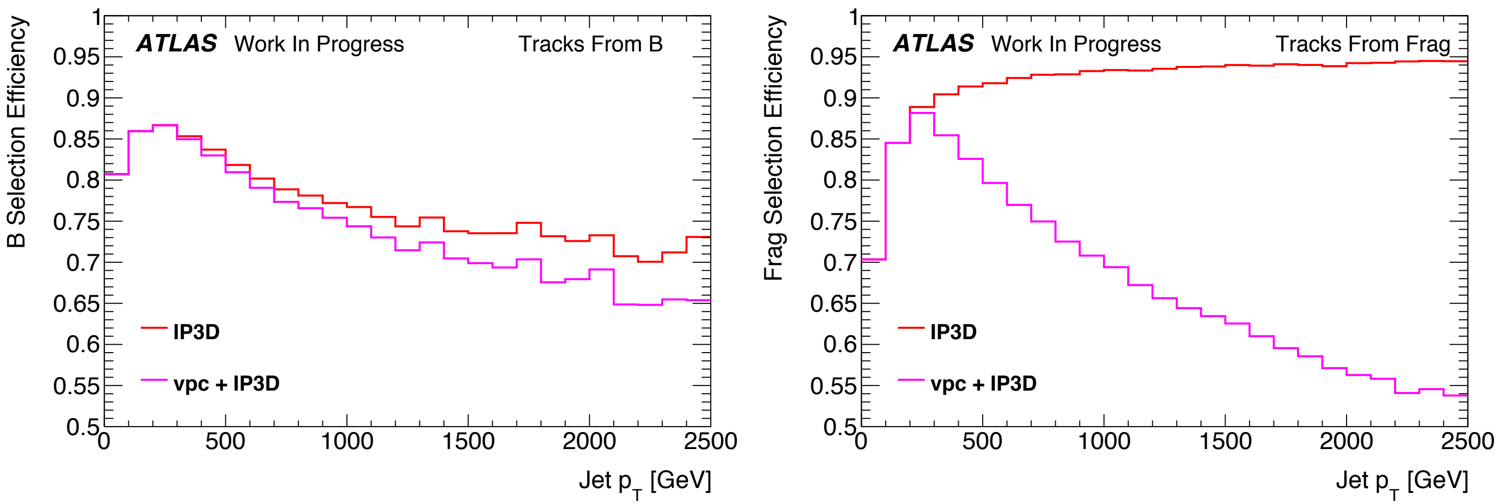
IP3D cut (Black)
- 1 GeV flat cut

VPC (Magenta)
- 0.4% Jet-P_T



- Track-P_T of tracks from B are more dependant on jet-P_T than tracks from other origins.
- Try a jet-P_T dependant cut (variable p_T cut or VPC) = Track P_T < 0.4% Jet-P_T
  - This cut is shown in magenta.
- By applying this cut
  - A small reduction in the selection efficiency of tracks from B
  - A large reduction in the selection efficiency of tracks from Frag.

Tracks From B



Tracks From Frag.

Contents

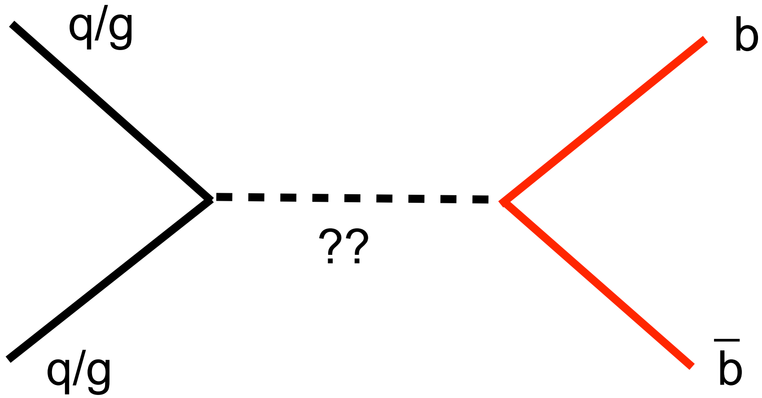
1) Introduction To Flavour Tagging

2) Data Commissioning

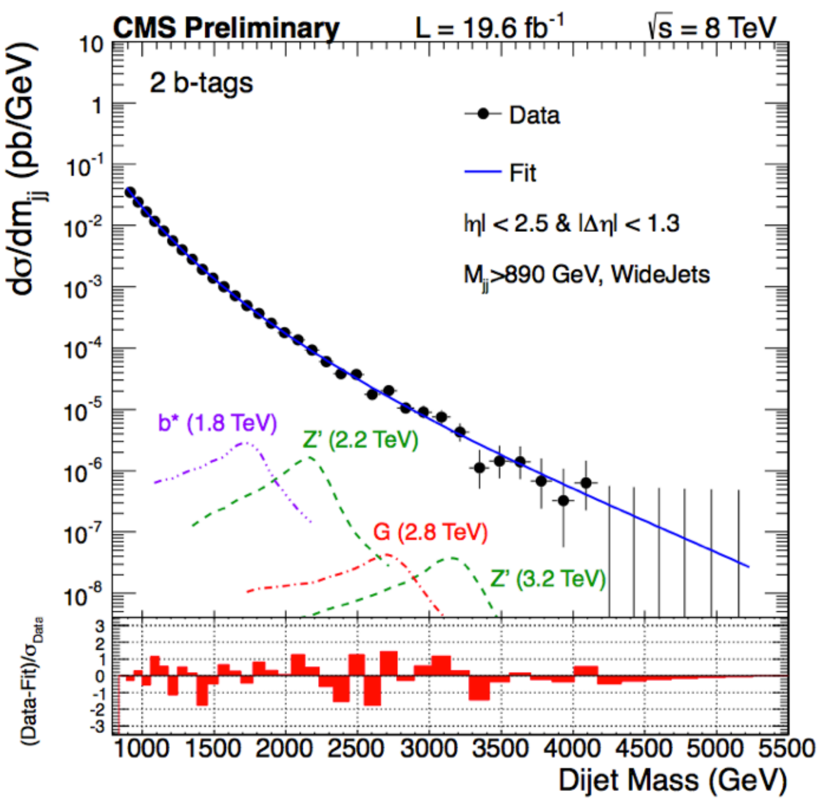
3) High-P_T *b*-tagging Improvements

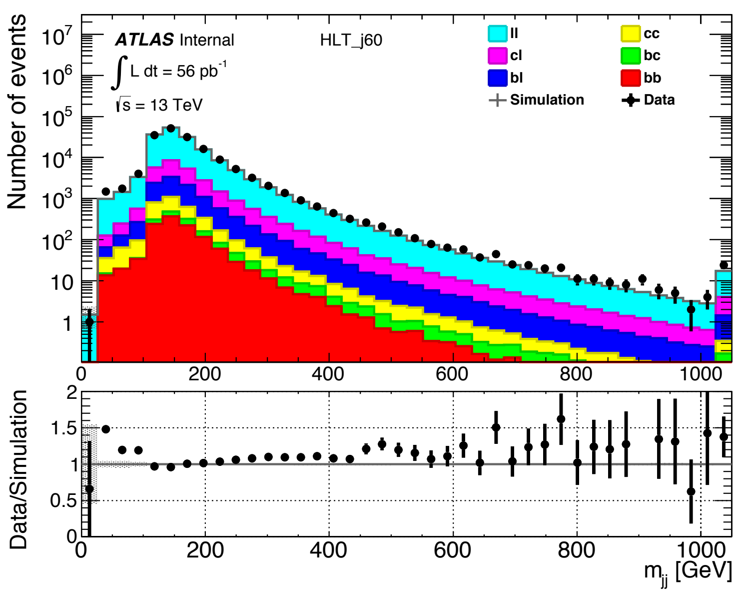
4) Exotic Di-b-jet Analysis

Exotic Di-b-Jet Analysis



- Di-jets is one of the most sensitive probes of new physics at high masses.
- Many BSM models predict resonances that can decay to bb
- Use of *b*-tagging increases sensitivity to these models.
  - This requires extremely high-P_T b-tagging.
- I have studied effect of VPC on the background
  - Appears to reduce effect of gluon splitting.





No b-tagging applied.

Conclusions

1) Data Commissioning
- Some good agreement (Jet Kinematics, SV and JF variables).
- Working with tracking experts to improve I.P. distributions.
- Need to understand discrepancy in the MV2 distributions.

2) High-P_T *b*-tagging Improvements
- Studied the effect of jet-P_T dependant track-P_T cut
- Great promise here, but need to demonstrate effectiveness and optimise the cut choice.
- This can be done by studying ROC curves.

3) Exotic Di-b-jet Analysis
- Analysis beginning to ramp up as new data is taken.
- Exciting opportunity to discover new physics!

Thanks for Listening

Any Questions?

Backup

Samples

- MC Sample:

- **Full xAOD**
- 50ns dijet MC sample data
- Split into 4 slices and the re-weighted (see backup) JZ1W-JZ4W - No JZ0W used.
- ~ 8M Events.

“mc15_13TeV.361021.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1W.merge.AOD.e3569_s2576_s2132_r6630_r6264/”
“mc15_13TeV.361022.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2W.merge.AOD.e3668_s2576_s2132_r6630_r6264/”
“mc15_13TeV.361023.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3W.merge.AOD.e3668_s2576_s2132_r6630_r6264/”
“mc15_13TeV.361024.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4W.merge.AOD.e3668_s2576_s2132_r6630_r6264/”

- Data Sample:

- 50ns data from stable beam collisions.
- **FTAG** derivation
- ~4M Events from 6 Runs: 270806, 270953, 271048, 271421, 271516 and 271595

“data15_13TeV.00270806.physics_Main.merge.DAOD_FTAG1.f611_m1463_p2375/”
“data15_13TeV.00270953.physics_Main.merge.DAOD_FTAG1.f611_m1463_p2375/”
“data15_13TeV.00271048physics_Main.merge.DAOD_FTAG1.f611_m1463_p2375/”
“data15_13TeV.00271421.physics_Main.merge.DAOD_FTAG1.f611_m1463_p2375/”
“data15_13TeV.00271516.physics_Main.merge.DAOD_FTAG1.f611_m1463_p2375/”
“data15_13TeV.00271595.physics_Main.merge.DAOD_FTAG1.f611_m1463_p2375/”

- We are using NTuples created using Run2BTagOptimisationFramework

Details and Cuts

- 20.1.5.3 with all tags recommended by CP group
- Running xAOD fix on full xAOD

- **HLT_j60 Trigger for MC and Data with Leading Jet P_T > 70 GeV.**

- AntiKt4EMTopoJets
- Run1LooseBadCuts and "ugly" jet removal.
- Jet Calibration:

```
- calibfile  ="JES_MC15Prerecommendation_April2015.config"
- calSeg     ="JetArea_Residual_Origin_EtaJES_GSC" (_Insitu for data)
```

- GRL = *"data15_13TeV.periodAllYear_DetStatus-v63-pro18-01_DQDefects-00-01-02_PHYS_StandardGRL_All_Good.xml"*

Select event if:

- njets ≥ 2 with $P_T > 35$ GeV, $|\eta| < 2.4$ and JVT > 0.641 if $P_T < 50$ GeV

- Leading jet, $P_T > 70$ GeV, $|\eta| < 2.4$ and JVT > 0.641 if $P_T < 50$ GeV

Then plot subleading if subleading jet has:

- $P_T > 35$ GeV
- $|\eta| < 2.5$
- JVT > 0.641 if $P_T < 50$ GeV

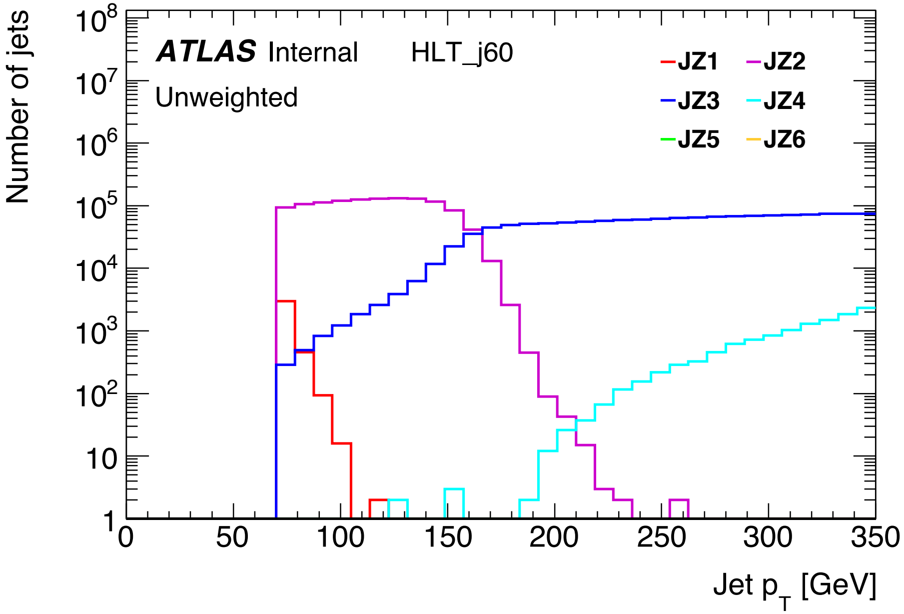
Monte Carlo Cuts

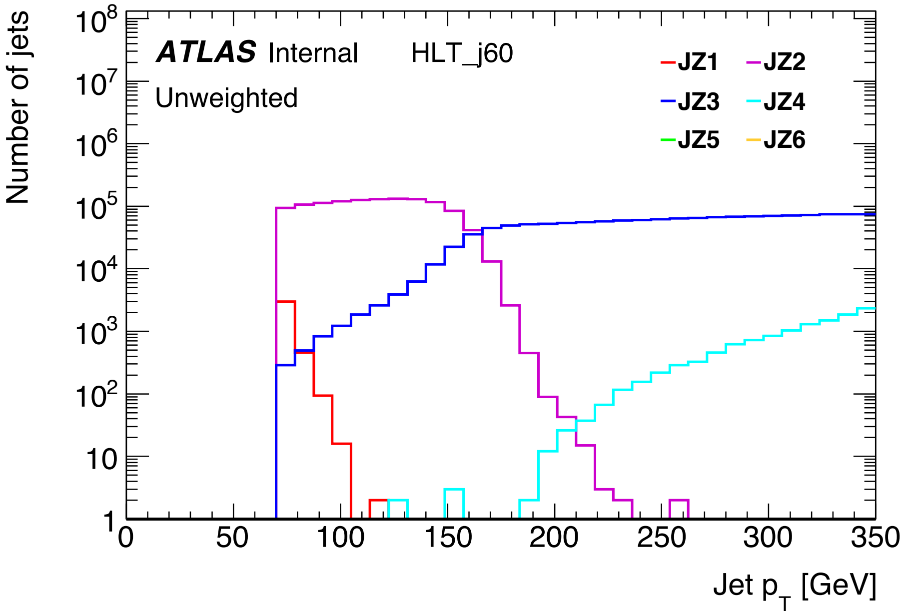
Truth Dijet Test applied to MC to clean sample

- (Lead P_T +Sublead P_T)/2 < 1.4* Truth Lead P_T

- LabDr_HadF truth matching.

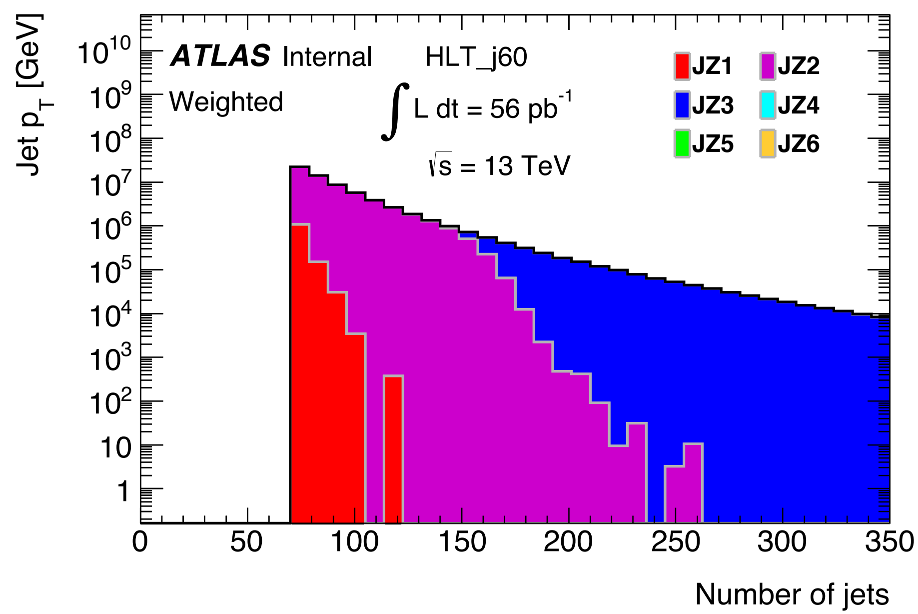
Di-jet sample re-weighting





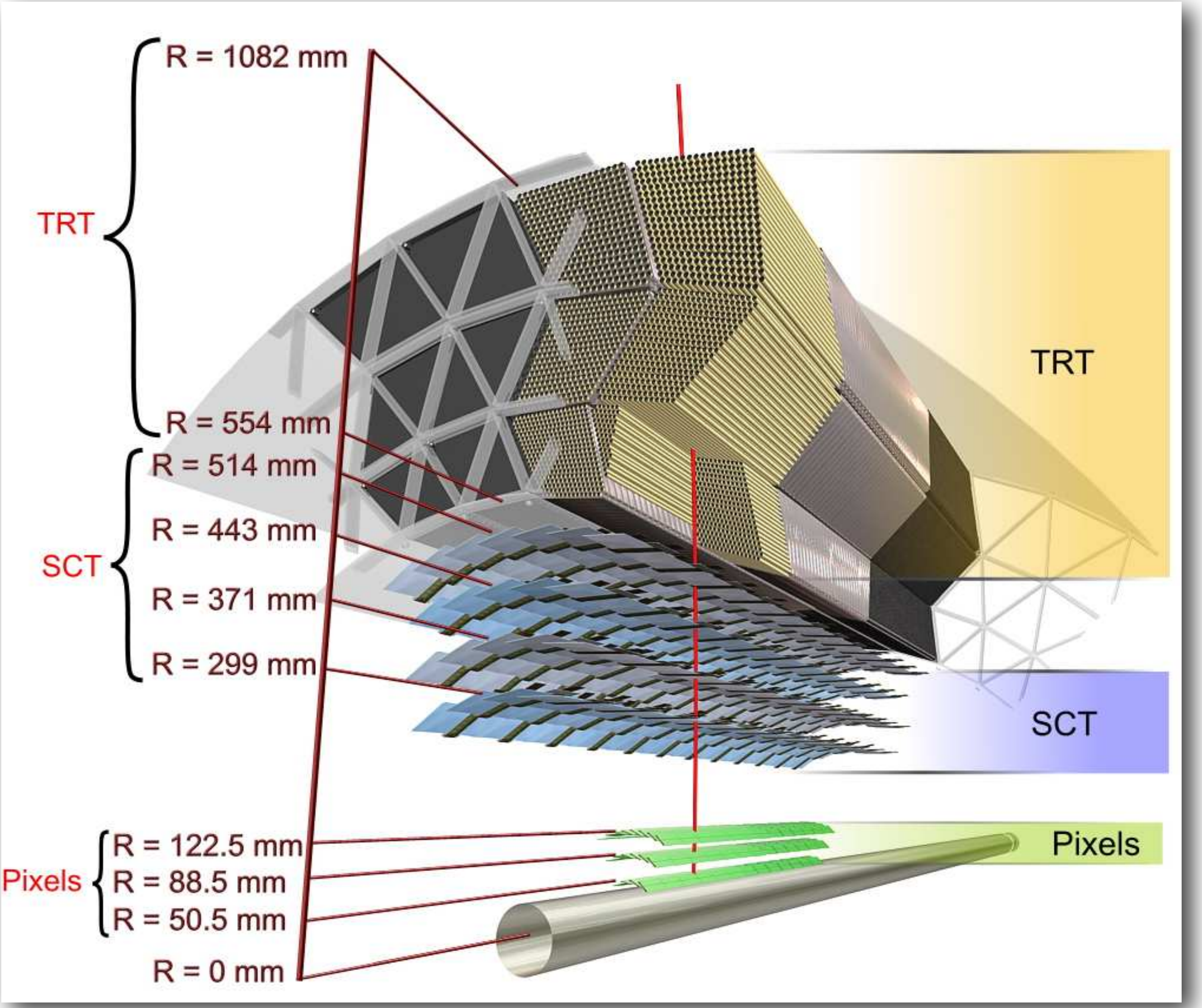
$$\textbf{Total Weight} = \frac{\text{mcwg*(Filter Eff.)*(CS[fb])*(Lumi[fb}^{-1}\text{])}}{\text{(\# Events)}}$$

Xs(fb)	Eff.	Slice and Energy
7.8420E+13	6.7198E-04	#JZ1W 20-60 GeV
2.4334E+12	3.3264E-04	#JZ2W 60-160 GeV
2.6454E+10	3.1953E-04	#JZ3W 160-400 GeV
2.5464E+08	5.3009E-04	#JZ4W 400-800 GeV



Average P_T of two highest P_T jets
< 1.4 * Truth Leading Jet P_T

Inner Detector

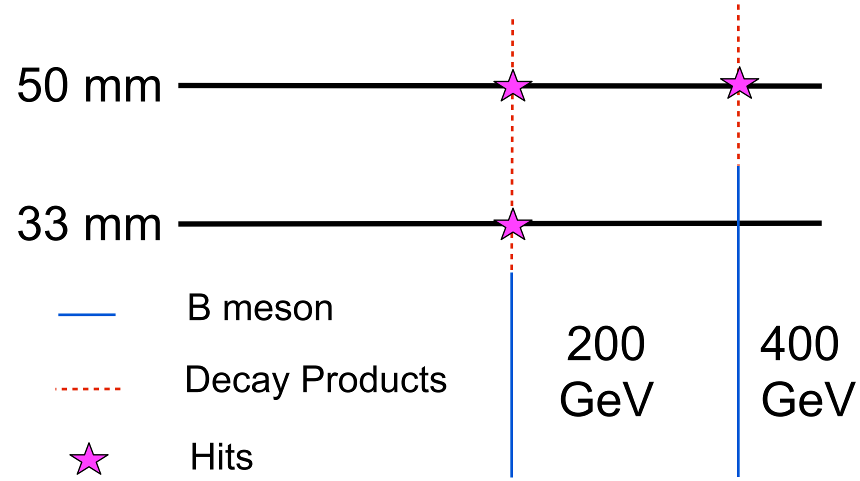


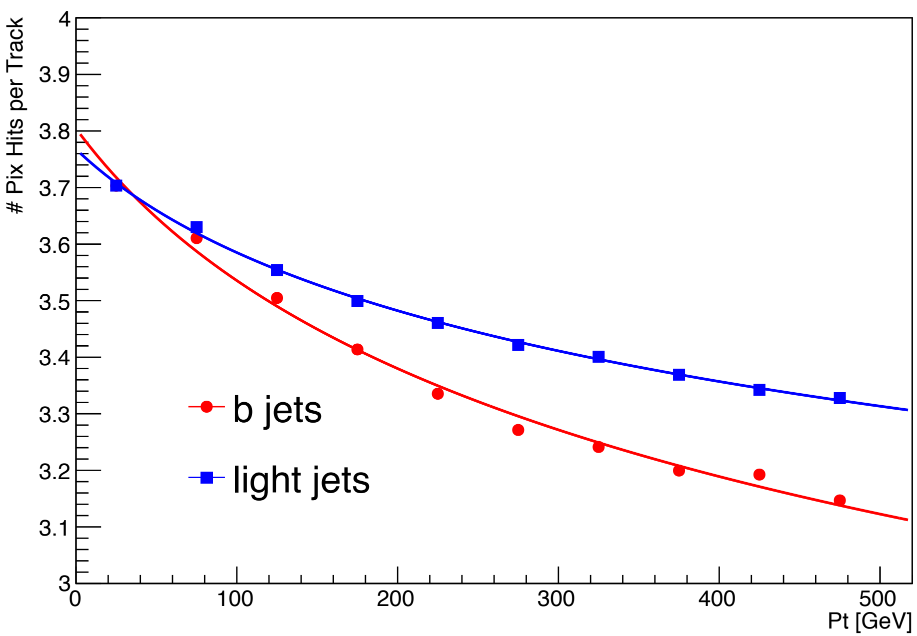
Flavour Fractions

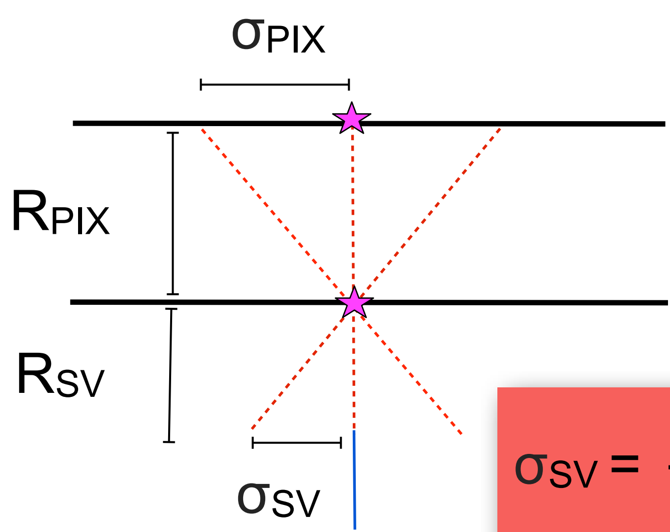
Working Point	Leading(L) or Subleading(SL) Jet?	Number of Jets	Fraction of b-jets	Fraction of c-jets	Fraction of Light Jets
100	L	6.50E+07	0.02962 +/- 0.0007	0.06323 +/- 0.0005	0.9072 +/- 0.0001
	SL	6.00E+07	0.03198 +/-0.0007	0.06122 +/- 0.0005	0.9068 +/-0.0001
70	L	2.20E+06	0.6508 +/-0.0008	0.2760 +/- 0.0013	0.0732 +/- 0.0024
	SL	2.00E+06	0.1369 +/-0.0019	0.0687 +/- 0.0027	0.7944 +/- 0.0007

Problems at High P_T - Lifetime

B_0 Meson	$c\tau \sim 450\mu m$
$P_T \sim 200$ GeV	$vt \sim 20mm$
$P_T \sim 400$ GeV	$vt \sim 40mm$







$$\sigma_{SV} = \frac{\sigma_{PIX} R_{SV}}{R_{PIX}}$$

Problems at High P_T - Boost

- Each track has a trajectory error associated with it from reconstruction.

➡ These errors can be visualised as a “tube” around the track.

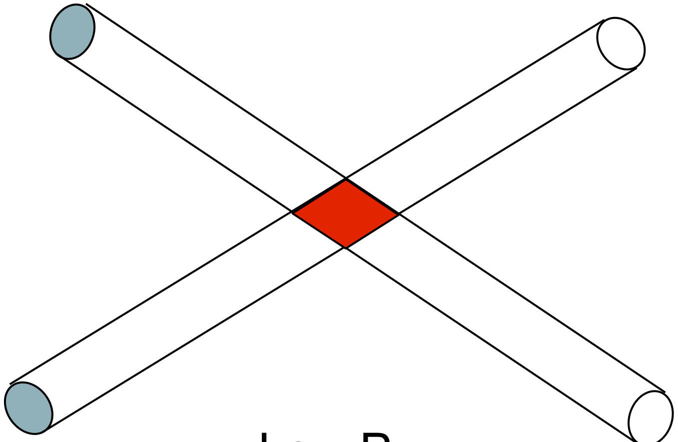
- Verities are constructed from two tracks crossing

➡ The error on the position of the vertex is the overlap of the two “tubes”.

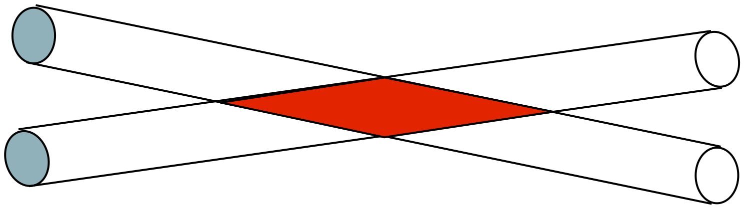
- At high P_T the jet becomes boosted, and tracks lie closer together

➡ This leads to larger errors on the vertices.

➡ Also we will get more fake vertices.



Low P_T



High P_T