B-tagging in ATLAS Experiment

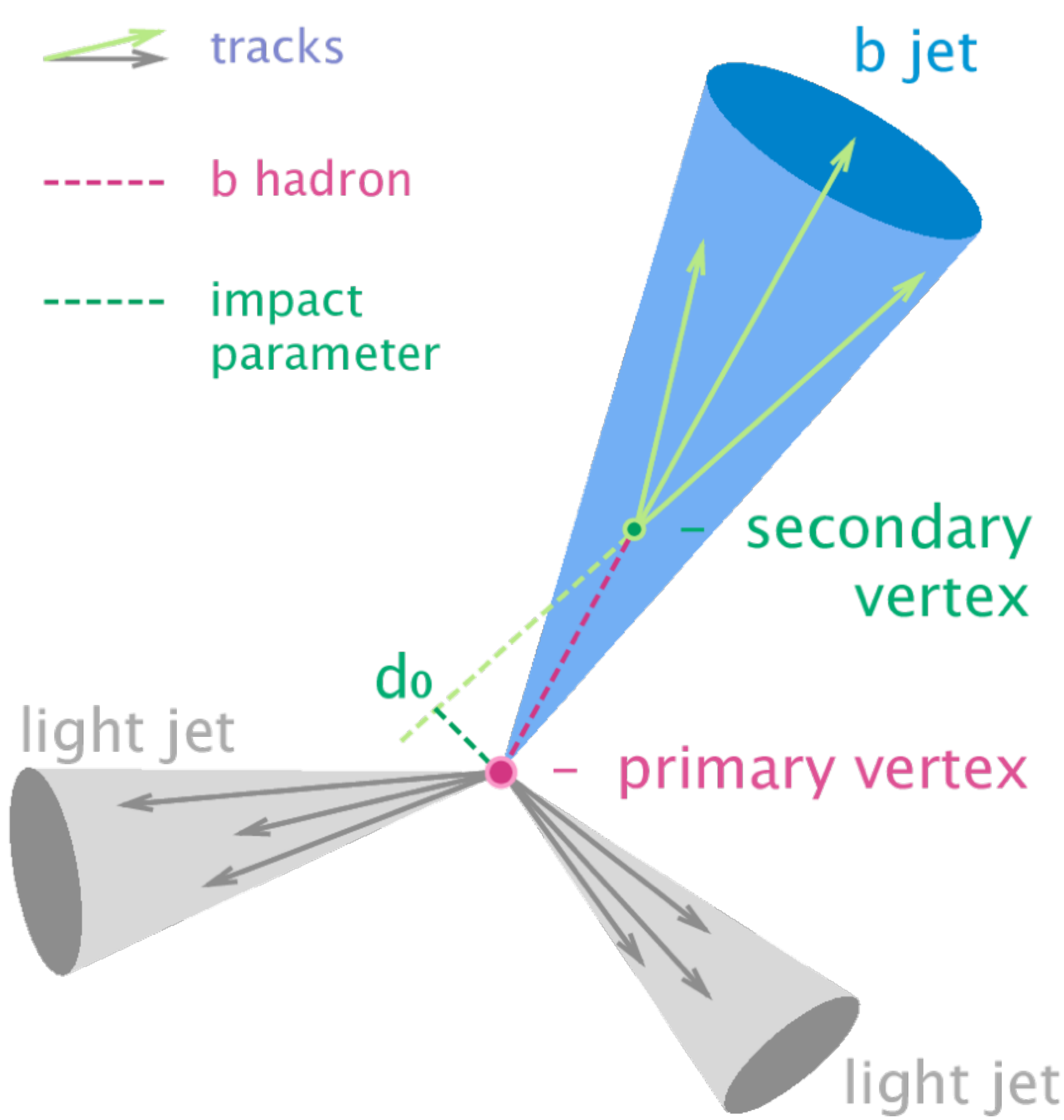
Changqiao LI[1][2]

[1]University of Science and Technology of China
[2]Laboratoire de Physique Nucléaire et de Hautes Energies

About me

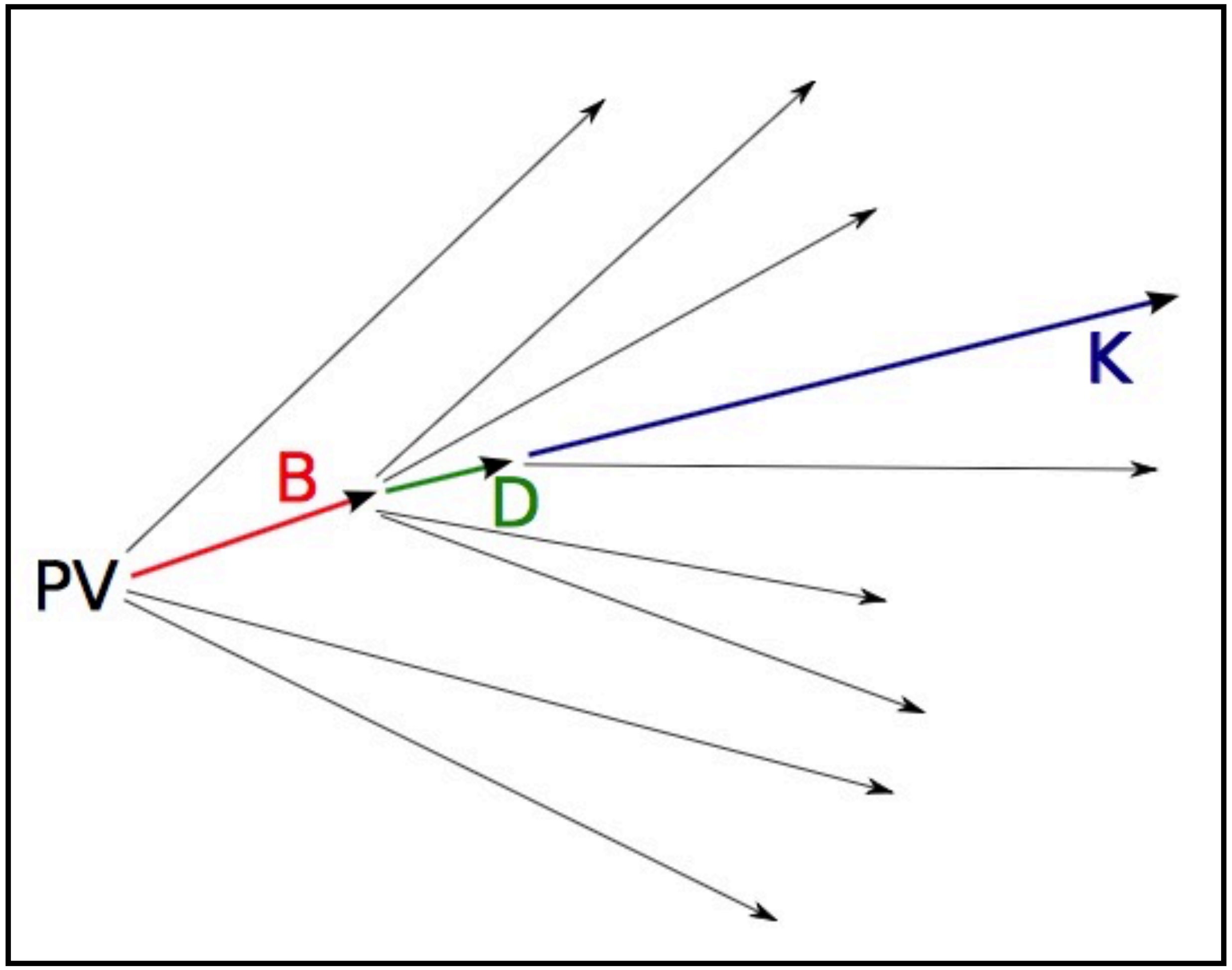
- Chang-qiao
  (“qiao” pronounced similar like “ciao”)
- Co-Ph.D. of LPNHE and USTC(China)
- Work on the ATLAS Experiment
- Qualification work:
  - **B-tagging calibration with Tag-and-Probe Method**
- Thesis topic:
  - SM VH($\rightarrow$bb) analysis
    (Introduced in the next talk by Charles)

What is B-tagging of jet?



- Identification of jets originating from b-quarks
- Plays a vital role in many ATLAS analysis:
  - Higgs→b-quarks
  - top physics
  - search for new physics
- Hadrons containing bottom quarks have sufficient lifetime that they travel some distance before decaying

How to tag?



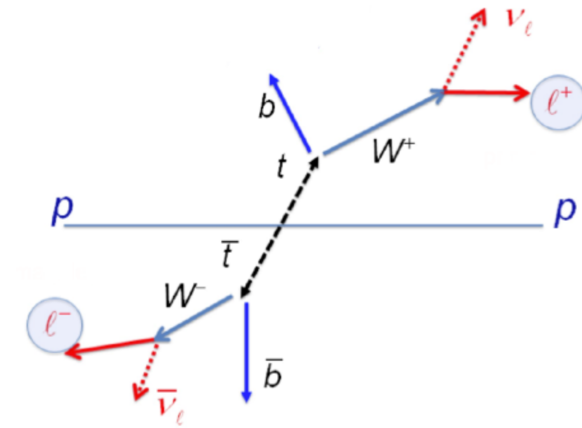
Procedure:

1. Select a jet
2. Match tracks by ΔR matching
3. Build vertices
4. compute discriminant with:
    - track-based variables associated with the jet
    - large mass
    - high decay multiplicity

Tagging algorithms:

- impact parameter-based (IP)
- inclusive secondary vertex reconstruction (SV)
- decay chain multi-vertex reconstruction (JetFitter)
- combination of above in a multivariate discriminant (MV)

How to calibrate b-tagging?



- Using dileptonic ttbar events
- b-jet calibration method:
  - Aim is to extract b-tagging efficiency for supported working points in data to determine Pt dependent scale factors:

$$\kappa_{\epsilon_j}^{data/sim} = \frac{\epsilon_j^{data}}{\epsilon_j^{sim}}$$

  - ttbar Probability Distribution Function (PDF) method:
    Measure the b-jet tag weight distribution from a Maximum Likelihood fit in eμ+2/3 jets or ee/μμ+2/3 jets
  - ttbar tag-and-probe method:
    Measure the b-jet tag efficiency by subtracting the noneb-jets contamination in probe jets from ttbar in eμ+2 jets
- Systematics:
  - Detector related systematics
  - MC modelling: matrix element, hadronisation, showering
  - MC statistical uncertainty

Probability Distribution Function

- Use a likelihood formalism which exploits per event jet flavour correlations
  - Extract b-jet tagging efficiency from data in each jet pt bin
  - Large gain in precision when compared to determining efficiency from individual jets

Likelihood for 2 jet case:

$$\begin{aligned}\mathcal{L}(p_{T,1}, p_{T,2}, w_1, w_2) = & [f_{bb}\mathcal{P}_{bb}(p_{T,1}, p_{T,2})\mathcal{P}_b(w_1|p_{T,1})\mathcal{P}_b(w_2|p_{T,2}) \\ & + f_{bj}\mathcal{P}_{bj}(p_{T,1}, p_{T,2})\mathcal{P}_b(w_1|p_{T,1})\mathcal{P}_j(w_2|p_{T,2}) \\ & + f_{jj}\mathcal{P}_j(p_{T,1}, p_{T,2})\mathcal{P}_j(w_1|p_{T,1})\mathcal{P}_j(w_2|p_{T,2}) \\ & + 1 \leftrightarrow 2]/2,\end{aligned}$$

$\mathcal{P}_{f_1 f_2}(p_{T,1}, p_{T,2})$	2D PDFs for jets of flavour f1, f2 to have $p_{T,1}$, $p_{T,2}$
$\mathcal{P}_b(w\|p_T)$ $\mathcal{P}_j(w\|p_T)$	PDFs for b-tagging discriminant for b-jets (from data) and non-b-jets (MC)
$f_{bb}, f_{bj}, f_{jj} = 1 - f_{bb} - f_{bj}$	Flavour fractions in 2-jet case

Tag-and-Probe

- Using eμ+2 jets:
  - tagging requirement: tag weight passing tagging working point
  - 1tag events: non tagged jet defined as probe jet
  - 2tag events: both jets defined as probe jets

Step 1

Non-$t\bar{t}$ events are subtracted:

$$\varepsilon_{data}^{Uncorr} = \frac{\boxed{N_{data}^{pass}} - \boxed{N_{non-t\bar{t}}^{pass}}}{\boxed{N_{data}} - \boxed{N_{non-t\bar{t}}}}$$

Step 2

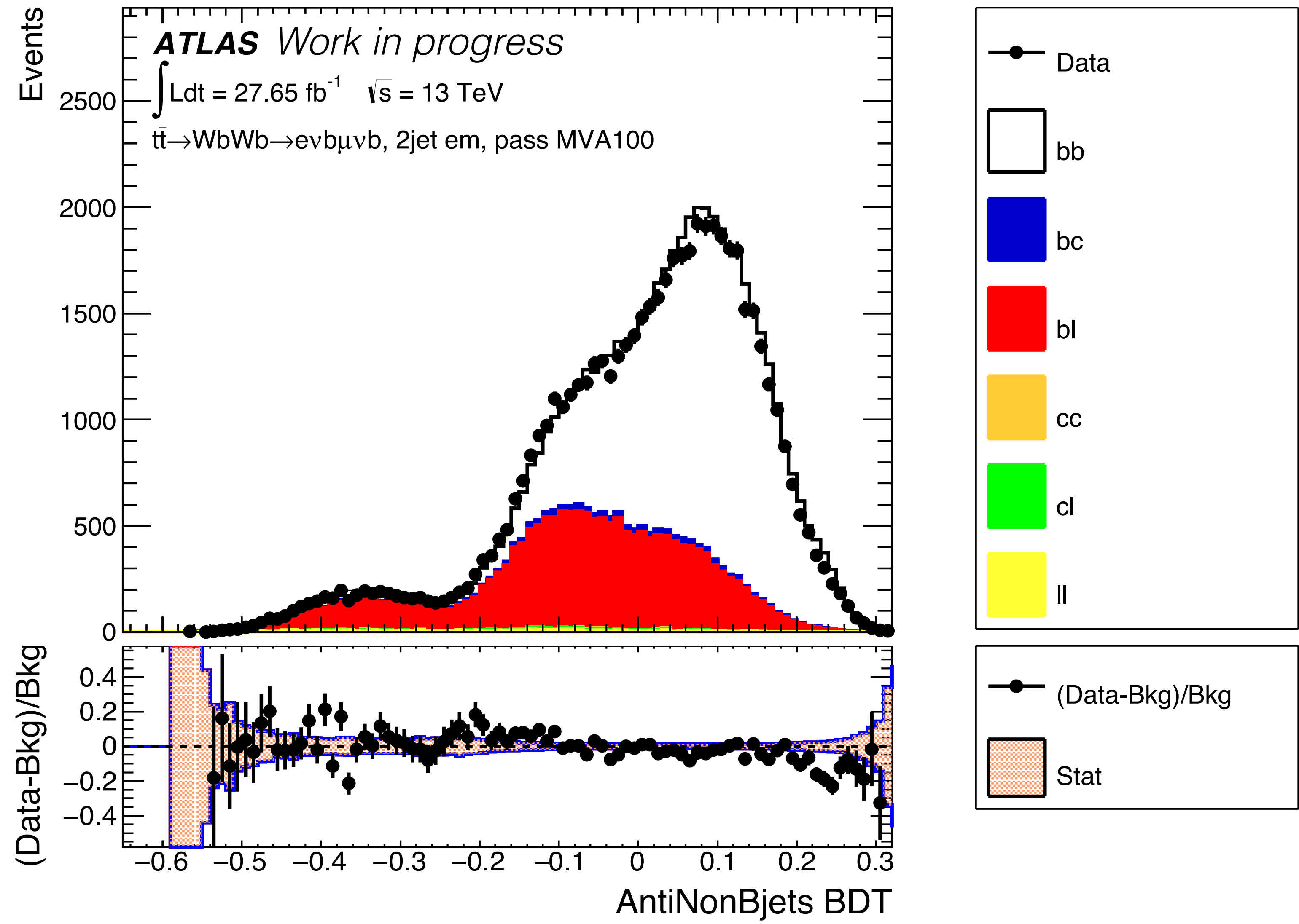
Non b-labelled probes in $t\bar{t}$ events are subtracted, based on the b-labelled jet fraction $f_b^{t\bar{t}\ MC}$ in $t\bar{t}$ and efficiency for non b-labelled jets in $t\bar{t}$ $\varepsilon_{non-b}^{t\bar{t}\ MC}$:

$$\varepsilon_{data} = \frac{\varepsilon_{data}^{Uncorr} - (1 - \boxed{f_b^{t\bar{t}\ MC}}) \times \boxed{\varepsilon_{non-b}^{t\bar{t}\ MC}}}{\boxed{f_b^{t\bar{t}\ MC}}}$$

Tag-and-Probe

AntiBL BDT

	description
$\Delta\phi$(j1,j2)	azimuthal angle between two jets
$\Delta\eta$(j1,j2)	pseudo-rapidity difference between two
ΔR(j1,j2)	angular separation between two jets
j1 pt	leading jet pt
j2 pt	sub-leading jet pt
RpT	Ratio of two jets pt
lep1 pt	pt of leading charge lepton
lep2 pt	pt of sub-leading charge lepton
met	Missing ET
nFowJet	Number of forward jets
Mlj	Minimum average invariant mass of two possible cases of lep and jet pairs

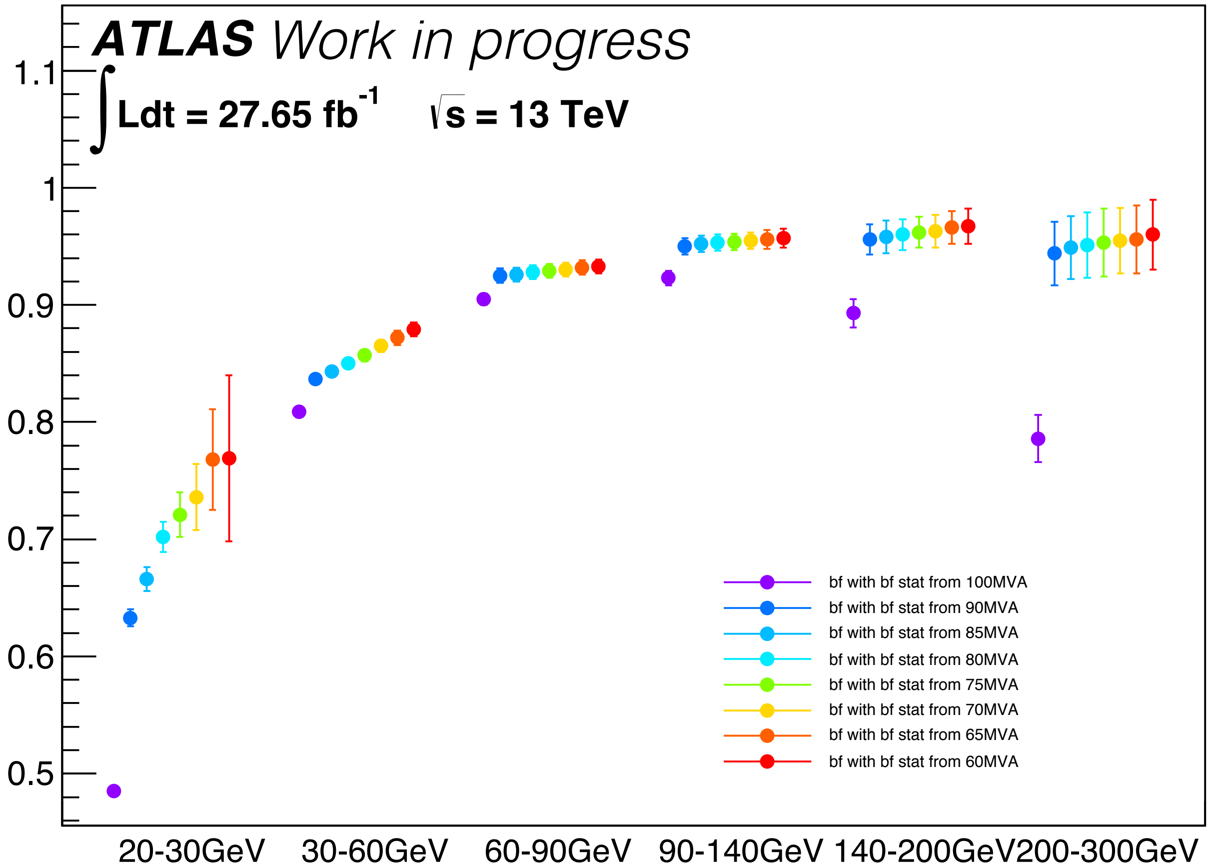


bb efficiency	60	65	70	75	80	85	90	100
mva cut	0.051	0.037	0.021	0.003	-0.019	-0.043	-0.073	-

Tag-and-Probe

BDT cut effect and optimisation

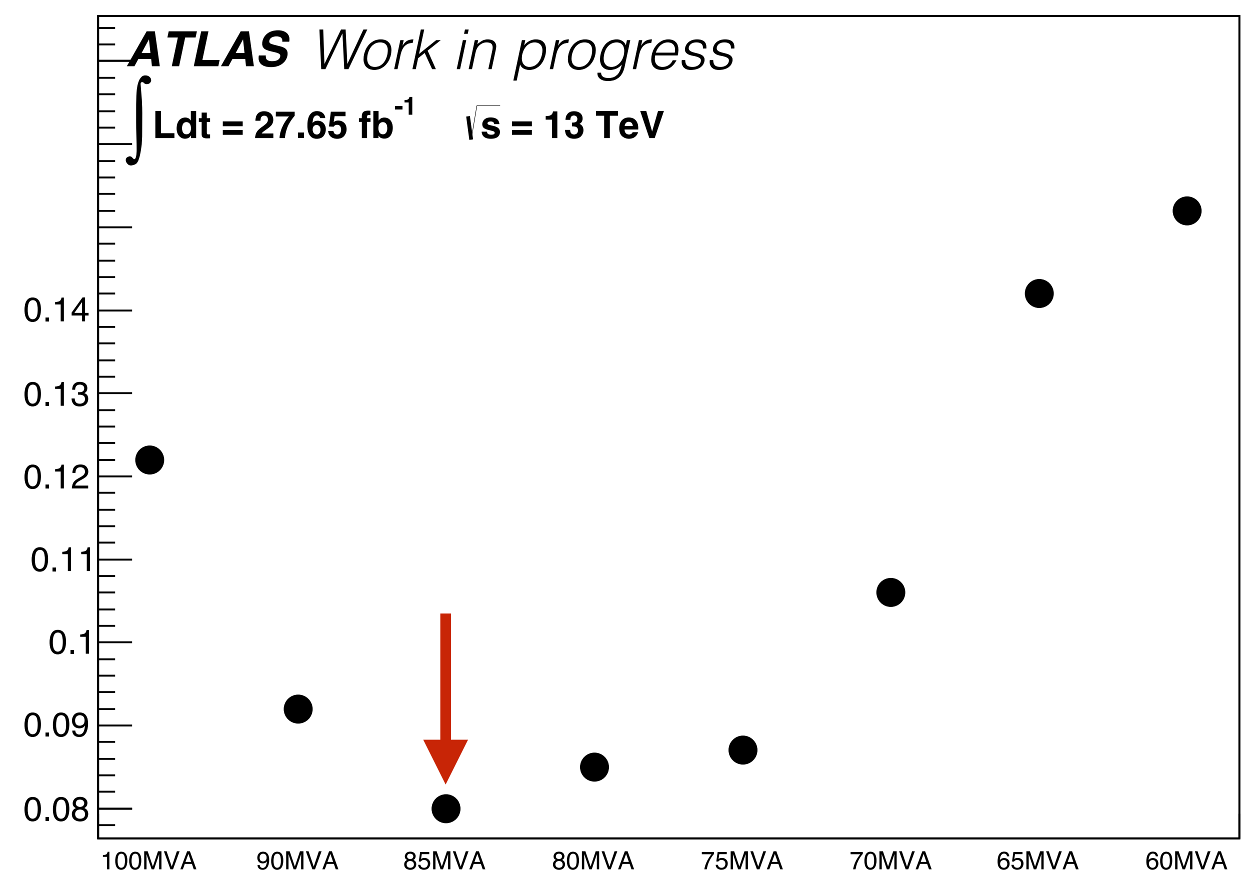
Comparison across MVA cuts at 77 WP of Calo Jets Calibration



B-jets purity in ttbar

Total uncertainty in lowest pt bin

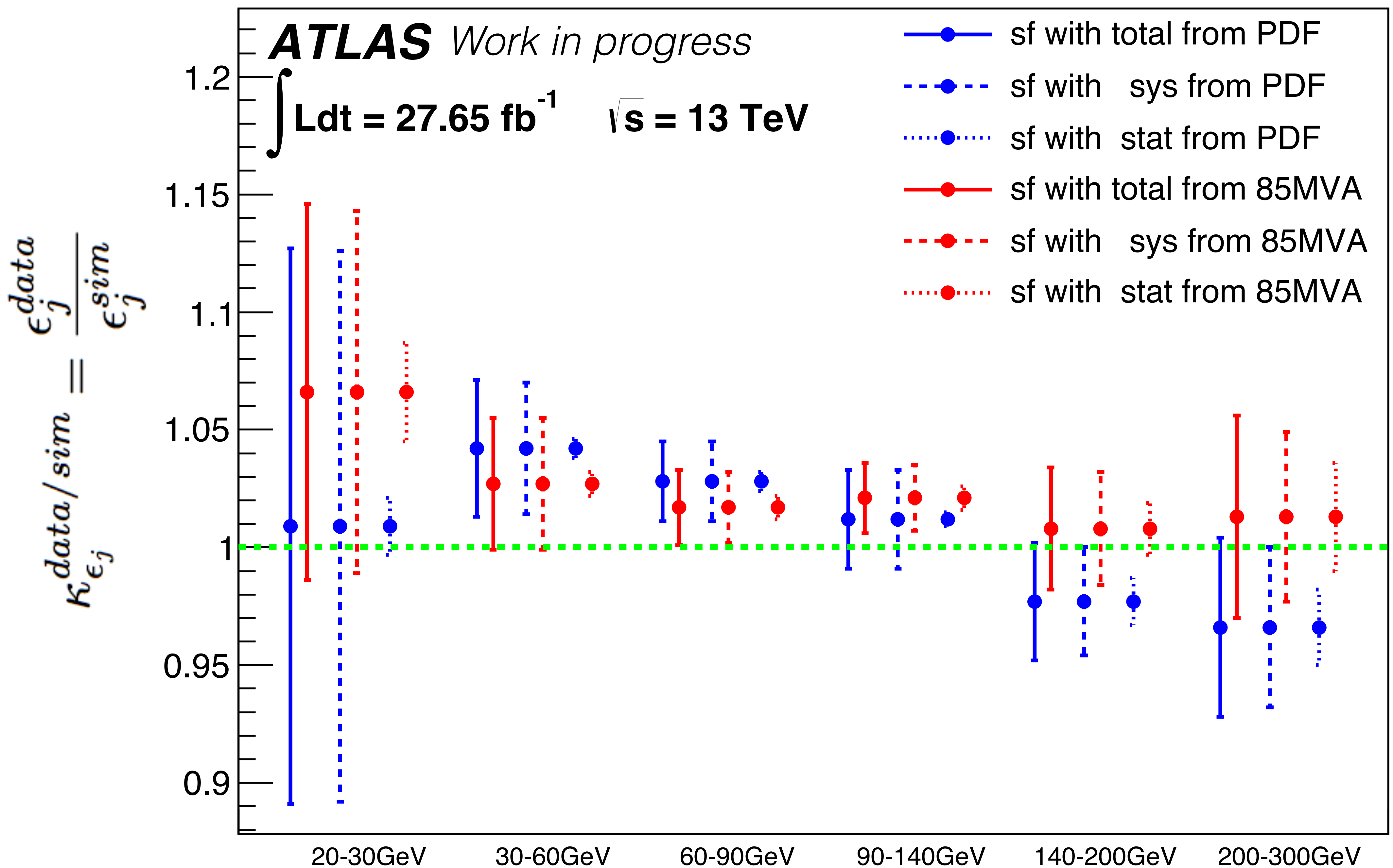
Comparison across MVA cuts at 77 WP of Calo Jets Calibration of total in 1 ptbin



Comparison with PDF method

Calo-jets
PDF v.s. **TP(85MVA)**

Comparison across MVA cuts at 77 WP of Calo Jets Calibration



Summary

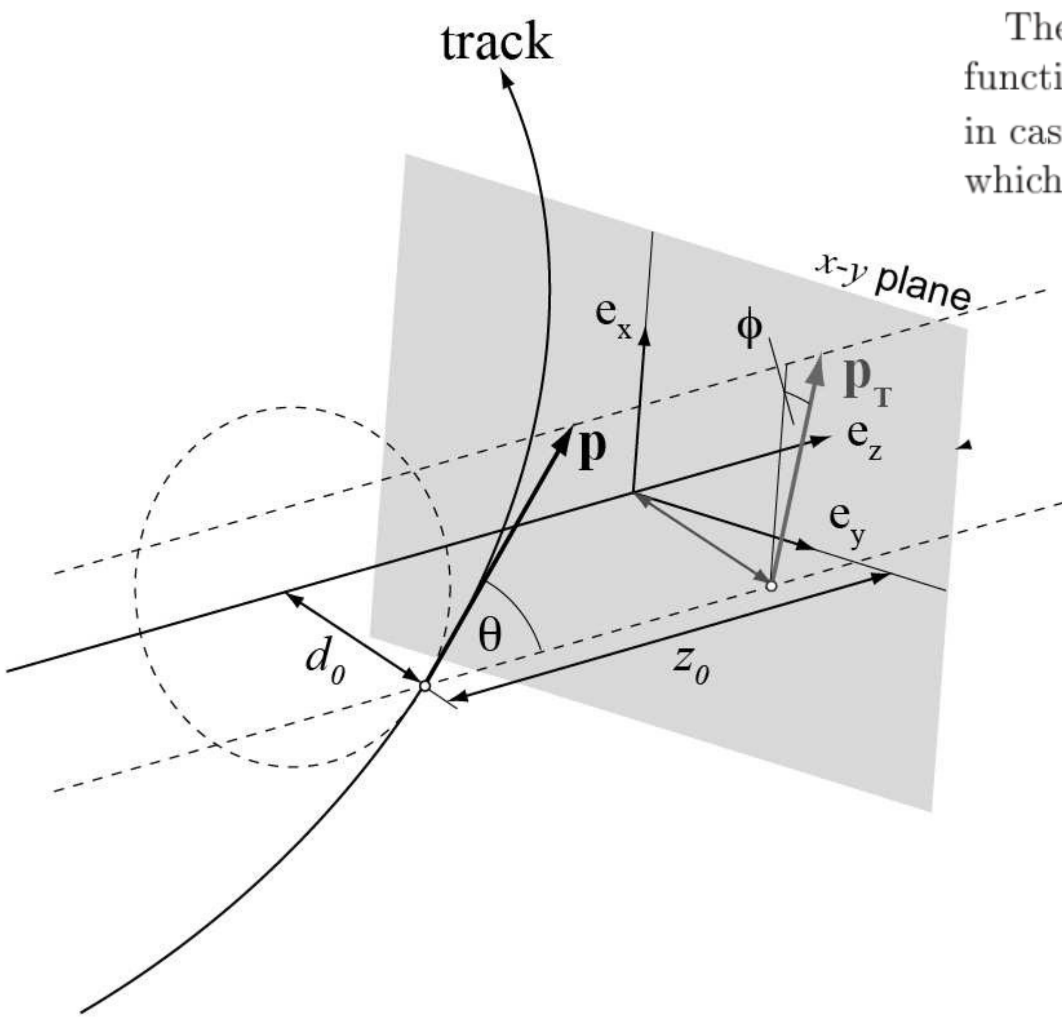
- Identification of jets originating from b-quarks is very important for many analysis in ATLAS
- Three basic algorithms are used to perform b-jet tagging
- To achieve a better discrimination than any of the basic algorithms can achieve individually, a Boosted Decisions Tree (BDT) algorithm is employed. It combines the outputs of the basic taggers.
- Introduce ttbar PDF method and Tag-and-Probe methods.
- Tag-and-Probe: The worst uncertainty is in the lowest jet pt bin and smaller than 10%, for intermediate jet pt bins, it's smaller than 5%.
- Both methods gives compatible calibration results.

BACKUP

Backup

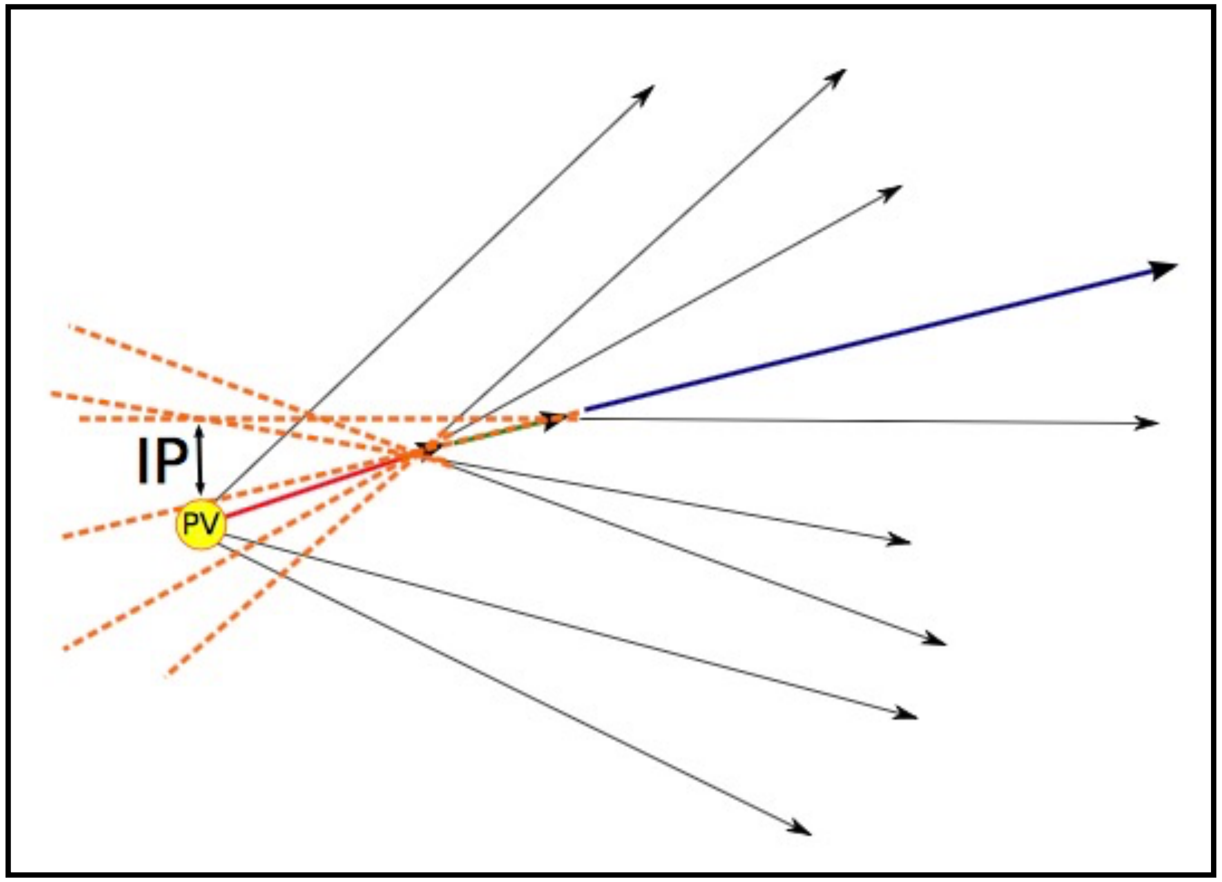
ATLAS uses a right-handed coordinate system with its origin at the nominal interaction point (IP) in the centre of the detector and the z-axis along the beam pipe. The x-axis points from the IP to the centre of the LHC ring and the y-axis points upwards. Cylindrical coordinates (r,ϕ) are used in the transverse plane, being ϕ the azimuthal angle around the z-axis. The pseudo-rapidity is defined in terms of the polar angle θ as η= -ln tan$(\theta/2)$ while $\Delta R^2 = \sqrt{\Delta\eta^2 + \Delta\phi^2}$.

$$\chi^2(\vec{r}) = \sum_{i=1}^{N_{trk}} (\vec{r} - \vec{r}_i(\hat{\phi}_{p,i}))^T \mathrm{cov}^{-1}_{3\times3,i}(\hat{\phi}_{p,i})(\vec{r} - \vec{r}_i(\hat{\phi}_{p,i})) \qquad (5.4)$$



The easiest way to estimate the vertex position $\vec{r}$ and its related error is to minimise this χ^2 function numerically. The auxiliary parameters $\vec{\phi}_p$ can be dealt with by repeating the fit again in case some of the initial parameters $\vec{\phi}_p$ correspond to positions along the tracks trajectories which are significantly displaced with respect to the final estimated vertex position $\vec{r}$.

(Tracks') Impact parameter-based



The track selection:

- track pT above 1 GeV
- |d0|<1mm |z0sinθ|<1.5mm
- silicon hits requirements

Probability density functions (PDF) obtained from reference histograms for IP significances:
For a jet, given associate tracks:

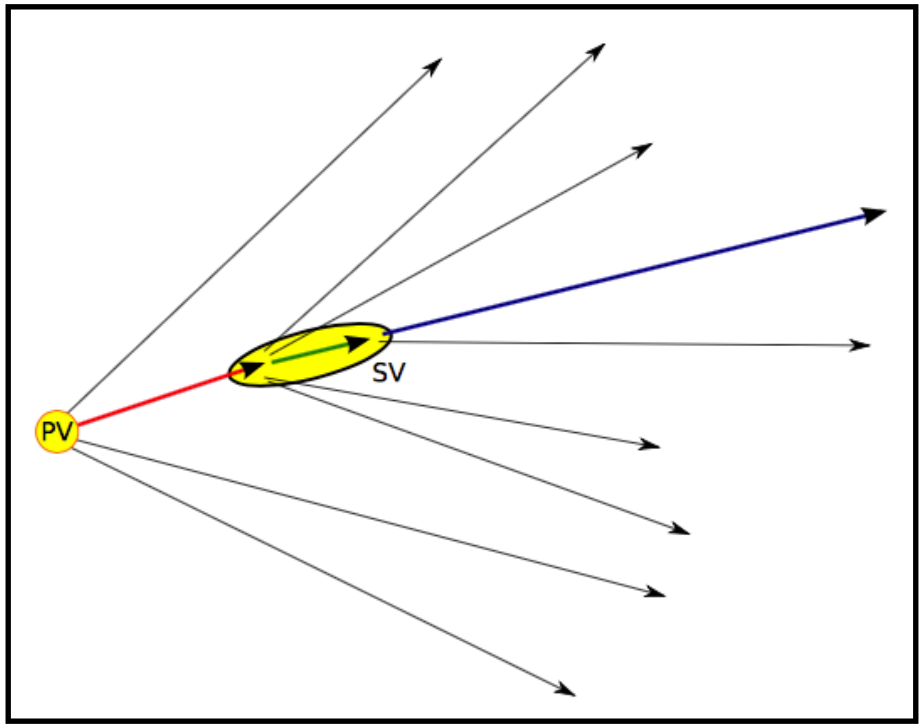
- Separated into exclusive categories depending on the hit pattern
- In each category, build PDF for jet-flavour hypothesis (b,c,light)

$$P_f = \prod_{\text{trk}} \mathcal{L}_f(S_{d_0}, S_{z_0}, \text{grade})$$
$$\text{with } f \in \{b, c, \text{light}\}$$

- Transverse impact parameter significance, d0/σd0
- Longitudinal impact parameter significance, z0*sinθ /σ(z0sinθ)
- A log-likelihood ratio (LLR) discriminant is computed as the sum of the per-track contributions,(N is the number of tracks of jet)

$$\sum_{i=1}^{N} \frac{\log p_b}{\log p_u}$$

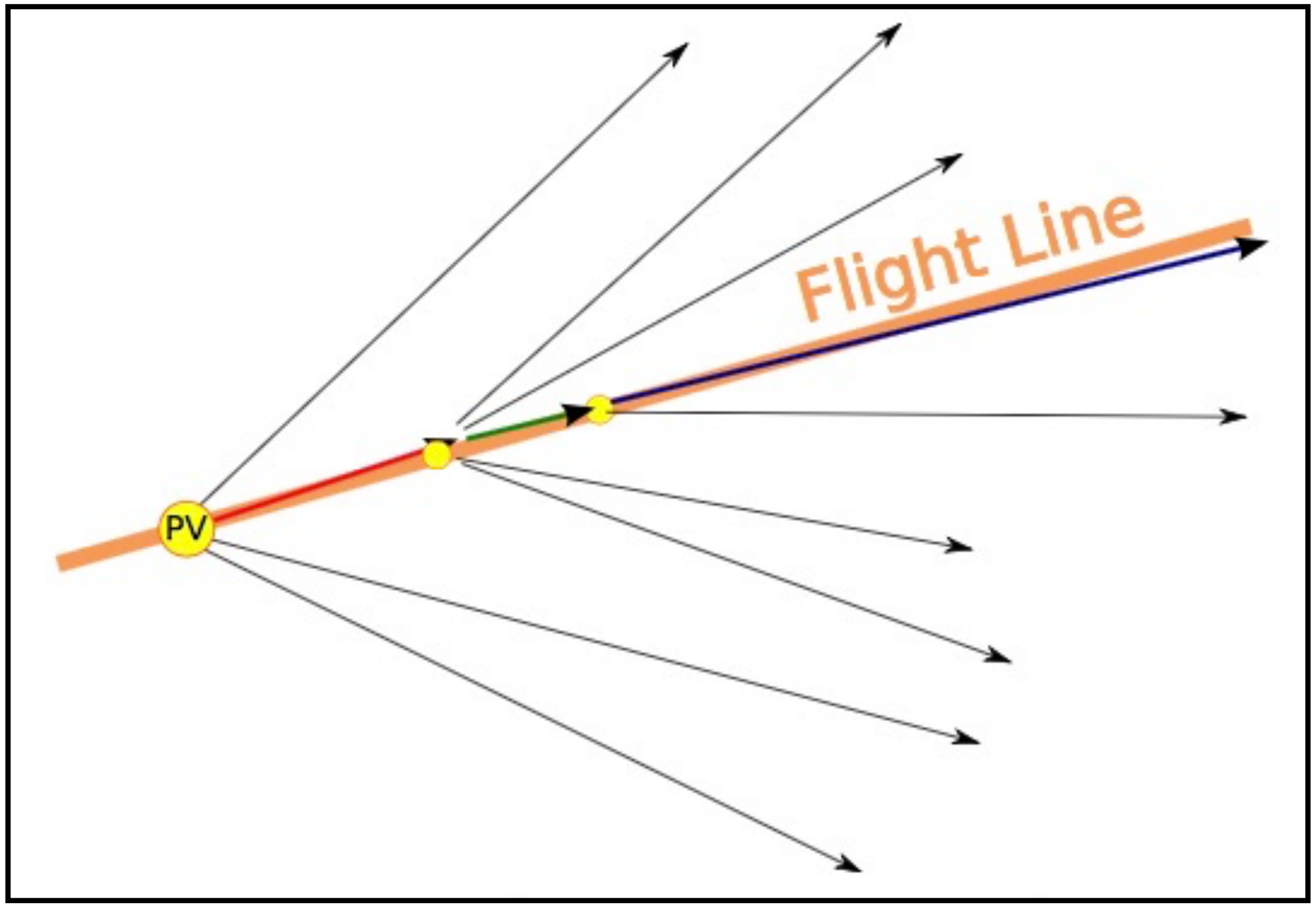
Inclusive secondary vertex reconstruction



SV	$m(SV)$	Invariant mass of tracks at the SV assuming π masses
	$f_E(SV)$	Fraction of the charged jet energy in the SV
	$N_{TrkAtVtx}(SV)$	Number of tracks used in the SV
	$N_{2TrkVtx}(SV)$	Number of two track vertex candidates
	$L_{xy}(SV)$	Transverse distance between the PV and the SVs
	$L_{xyz}(SV)$	Distance between the PV and the SVs
	$S_{xyz}(SV)$	Distance between the PV and SVs divided by its uncertainty
	$\Delta R(jet, SV)$	ΔR between the jet axis and the direction of the SV relative to the PV

- All selected tracks used to form all possible two-track vertices.
- Selection perform on these two-track vertex candidates.
- All tracks corresponding to the remaining accepted two-track vertices are used to determine a single secondary vertex.
- Calculate chi-2 of the vertex, if Prob(chi-2) is very small, remove the track with the highest contribution until Prob(chi-2) acceptable

Decay chain multi-vertex reconstruction

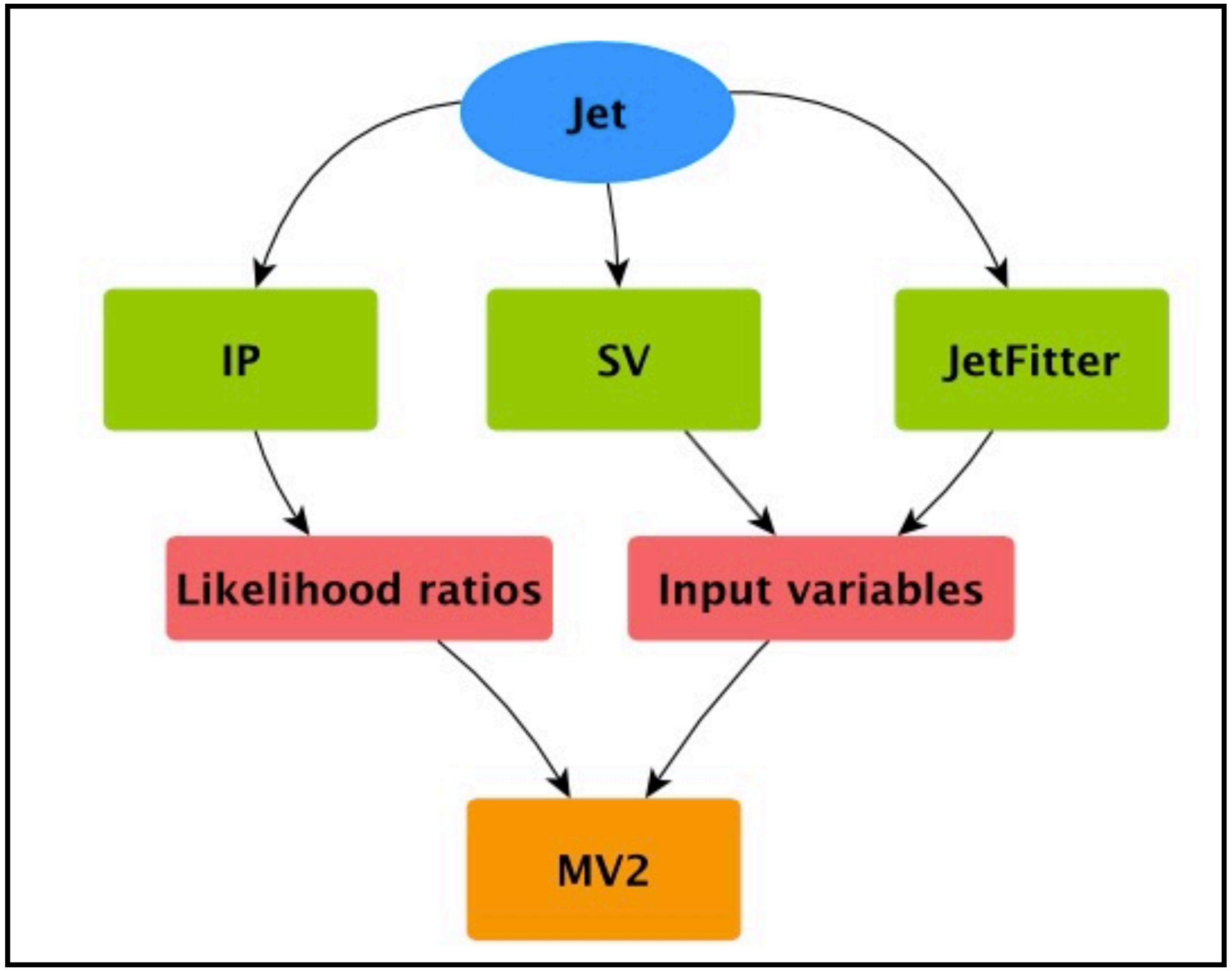


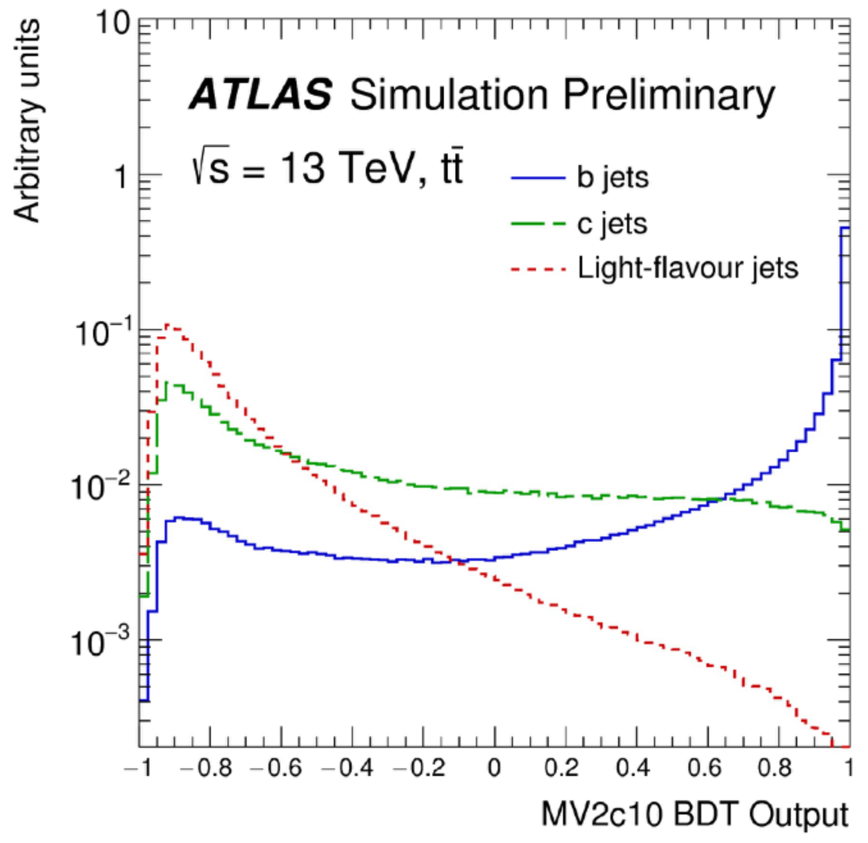
Jet Fitter	$N_{2TrkVtx}(JF)$	Number of 2-track vertex candidates
	$m(JF)$	Invariant mass of tracks from displaced vertices assuming π masses
	$S_{xyz}(JF)$	Significance of the average distance between the PV and displaced vertices
	$f_E(JF)$	Fraction of the charged jet energy in the SVs
	$N_{1-trkvertices}(JF)$	Number of displaced vertices with one track
	$N_{\geq 2-trkvertices}(JF)$	Number of displaced vertices with more than one track
	$N_{TrkAtVtx}(JF)$	Number of tracks from displaced vertices with at least two tracks
	$\Delta R(\vec{p}_{jet}, \vec{p}_{vtx})$	ΔR between the jet axis and the vectorial sum of the momenta of all tracks attached to displaced vertices

- Exploits the topological structure of weak b- and c-hadron decays inside the jet.
- Tries to reconstruct the full PV→ b → c-hadron decay chainwith the Kalman filter.
- Using similar variable as SV1.
- The variables left are used as input nodes in an artificial neural network.
- It has three output nodes, corresponding to the b-, c- and light-flavour-jet hypotheses.

Multivariate discriminant (default algorithm used for ATLAS)

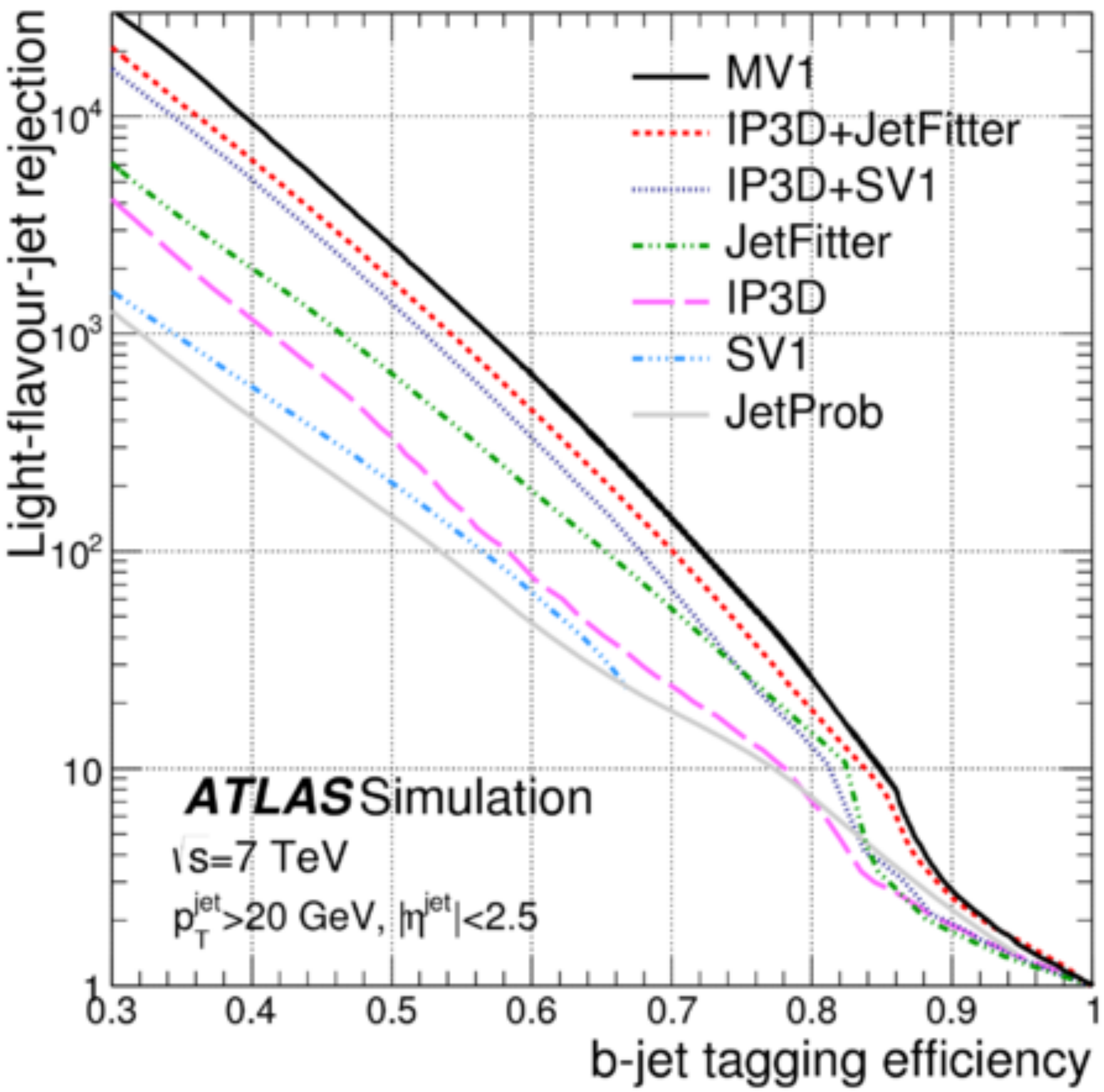
- To achieve a better discrimination
- Combine IP, SV and JetFitter in a BDT
- Training settings:
  1000 Trees, Max depth of 30,
  no cuts below 5% of sample





BDT Cut Value	b-jet Efficiency [%]	c-jet Rejection	Light-jet Rejection	τ Rejection
0.9349	60	34	1538	184
0.8244	70	12	381	55
0.6459	77	6	134	22
0.1758	85	3.1	33	8.2

Backup



#	Category	Fractional contribution [%] b-jets	c-jets	light-jets
0	No hits in first two layers; expected hit in IBL and b-layer	1.9	2.0	1.9
1	No hits in first two layers; expected hit in IBL and no expected hit in b-layer	0.1	0.1	0.1
2	No hits in first two layers; no expected hit in IBL and expected hit in b-layer	0.04	0.04	0.04
3	No hits in first two layers; no expected hit in IBL and b-layer	0.03	0.03	0.03
4	No hit in IBL; expected hit in IBL	2.4	2.3	2.1
5	No hit in IBL; no expected hit in IBL	1.0	1.0	0.9
6	No hit in b-layer; expected hit in b-layer	0.5	0.5	0.5
7	No hit in b-layer; no expected hit in b-layer	2.4	2.4	2.2
8	*Shared* hit in both IBL and b-layer	0.01	0.01	0.03
9	At least one *shared* pixel hits	2.0	1.7	1.5
10	Two or more *shared* SCT hits	3.2	3.0	2.7
11	*Split* hits in both IBL and b-layer	1.0	0.87	0.6
12	*Split* pixel hit	1.8	1.4	0.9
13	*Good*	83.6	84.8	86.4

Table 1: Description of the track categories used by IP2D and IP3D together with the fraction of tracks in each category for jets in $t\bar{t}$ events. The order of the layers is explained in the text. The categories further down in the list can be more inclusive than the first ones because, when a category is not fulfilled, the next one is evaluated.

Backup

To be reliable any b-tagging algorithm must be calibrated on data. To facilitate the calibration and to reduce the necessary amount of data the chosen variables have been transformed (for details see [12]):

- Invariant mass: $M' = \dfrac{M}{M+1}$;
- Energy ratio: $R' = R^{0.7}$;
- Number of good two-track secondary vertices: $N' = \log N$.

Due to the efficiency to reconstruct a secondary vertex inside a jet not reaching 100%, the probability density functions (PDF) of the vertex based variable have to contain a δ–function [12].

$$PDF = (1-\varepsilon)\cdot\delta(M',F',N') + \varepsilon\cdot ASH(M',F',N')$$

with ε being the efficiency to reconstruct a secondary vertex inside a jet. The continuous probability density function of the vertex variables is constructed from multidimensional calibration histograms using the ASH smoothing method [13].

Two slightly different taggers based on the *BTagVrtSec* algorithms are available in ATLAS, denoted $SV1$ and $SV2$. They use exactly the same variables but handle them in a different way. $SV1$ treats M' and R' jointly and adds N' as independent variable (2+1 decomposition), whereas $SV2$ uses joint three-dimensional probability density functions.