# $b$-tagging in the ATLAS experiment

Eloi Le Quilleuc,
on behalf of the Heavy Flavour Performance Group

CEA Saclay

May 19, 2016

# Introduction

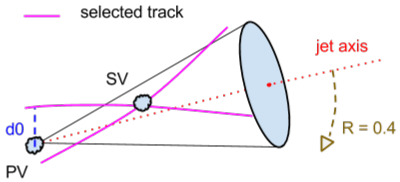
*b*-tagging identification of jets containing a *b*-hadron

**Motivation**

An important information for high energy physics, especially for SM (top, H$\rightarrow b\bar{b}$), SuSy (stop) and Exotics (VLQ) analysis

**$b$-tagging inputs**

- calorimeter jet
- associated tracks




**$b$-tagging is based on**

- impact parameters (IP) of associated tracks
- properties of reconstructed secondary vertices (SV)

# Outline

Expected Performance in Run 2

- Basic Algorithms
- Final Discriminant
- Comparison Run 1/Run 2

Commissioning in Early Run 2 Data

- Jet Selection
- Data/Prediction Comparison

**References**

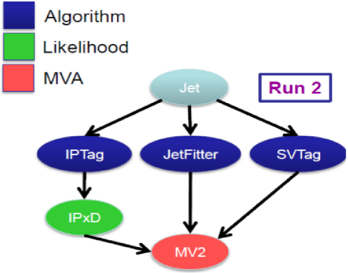
▸ Performance ▸ Commissioning

# Outline

# Overview of Algorithms in Run 2

## Algorithms

- 3 basics taggers IP3D, JetFitter, SV1
- MV2 BDT combines basic taggers information



# Impact Parameter based Algorithm : IP3D

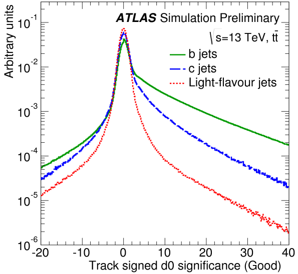
## Discriminant

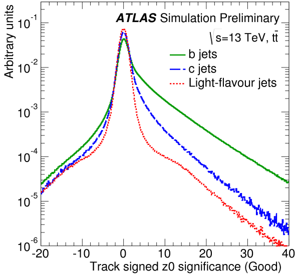
based on transerve/longitudinal IP significances of tracks associated to the jet

## Main feature of IPs : high tails

due to tracks from V0s decays, $b/c$-hadrons, hadronic interactions, $\gamma \rightarrow e^+e^-$

Run 2 improvement : better handling of the track quality information (*i.e* track hits in the silicon layers)



# Secondary Vertex Finding Algorithm : SV1

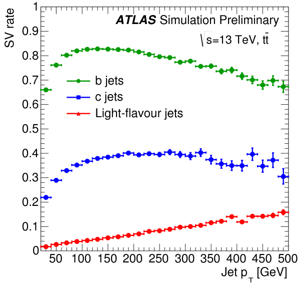
Reconstruct a single SV within the jet

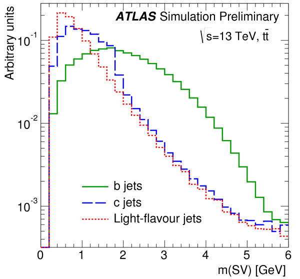
## Algorithm

- remove tracks from V0s, hadronic interactions, $\gamma \rightarrow e^+e^-$ vertices
- reconstruct a SV with remaining tracks, with outlier tracks iteratively removed

## SV properties

8 BDT input variables : mass, decay length, number of tracks, energy fraction, *ect*



# Decay Chain Multi-Vertex Algorithm : JetFitter

Aim to reconstruct the full
$PV \to b \to c$-hadron decay chain

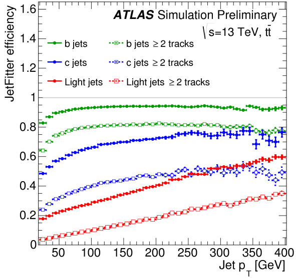
## Vertex finding strategy

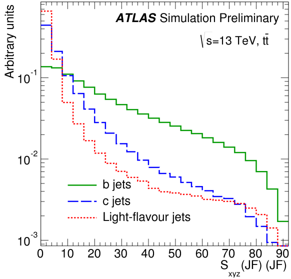
- constrain all tracks to intersect the same $b$-flight axis
- compatible vertices are merged, high significance SV are kept

Possible 1-track SV(s) in the jet

## SV properties

8 BDT input variables : mass, decay length, number of tracks, energy fraction, *ect*



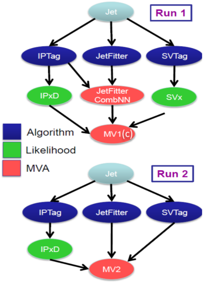



# Multivariate Algorithm : MV2

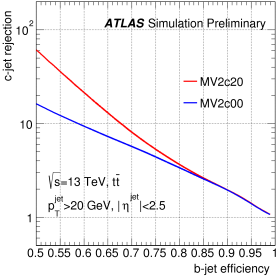
## Final discriminant

MV1 (Run 1) evaluations of basic taggers were combined into a NN

MV2 (Run 2) output variables of basic taggers combined into a BDT




- simplified combination
- better handling of correlations between the basic taggers



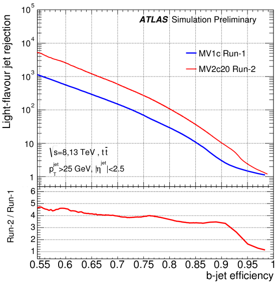

training background : 80% (100%) of light-jet, 20% (0%) of $c$-jet for MV2c20 (MV2c00)

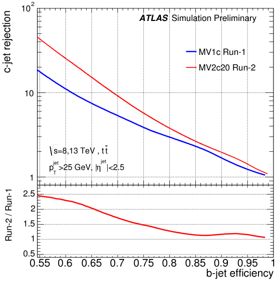
## Run 2 conditions

- addition of the IBL
- algorithm updates

## 70% $b$-jet efficiency in Run 2

- light-jet rejection $\sim$ 600, factor 4 increase
- $c$-jet rejection $\sim$ 10, factor 1.7 increase



# Outline

# Commissioning with $t\bar{t}$ events

## Motivation

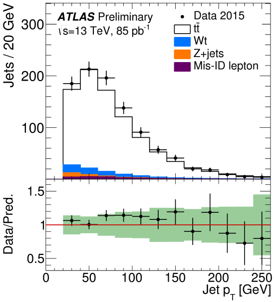
Study the $b$-tagging modelling at high $b$-jet purity

### Event selection

1 $e$, 1 $\mu$ with opposite charge, $\geq$ 2 jets

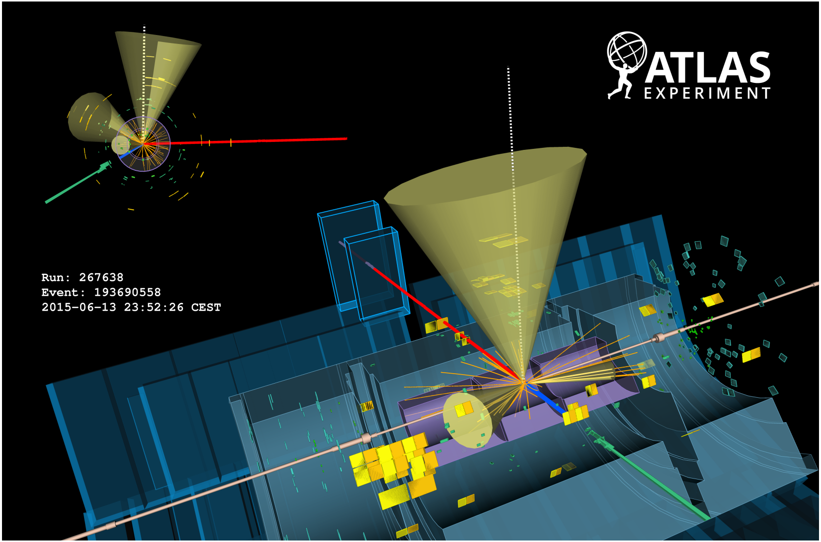
### Data sample, early Run 2

$\sqrt{s_{pp}} = 13$ TeV, $\int Ldt = 85$ pb$^{-1}$,
50 ns bunch-spacing




| | Fraction of Jets [%] | | |
|---|---|---|---|
| Flavour | Leading Two Jets | Jets with a Secondary Vertex | $b$-tagged Jets |
| $b$ | 69.0 ± 6.0 | 96.2 ± 0.9 | 98.8 ± 0.3 |
| Light | 28.9 ± 6.2 | 2.5 ± 0.8 | 0.5 ± 0.2 |
| $c$ | 2.0 ± 0.5 | 1.2 ± 0.3 | 0.8 ± 0.2 |

# Display[1] of a $t\bar{t}$ Event Candidate, $\sqrt{s} = 13$ TeV




[1] muon path (red line), electron path (blue line), jets (yellow cones), calorimeter energy desposits (green and yellow bars)

# Data/MC Comparison in $t\bar{t}$ *dilepton* Final State

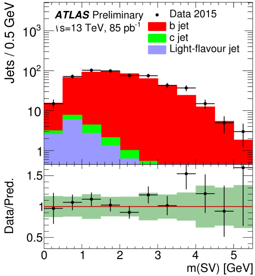
## Distributions

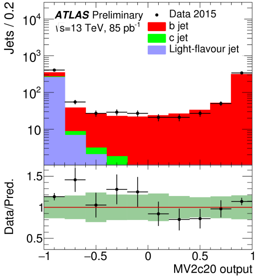
| | |
|---|---|
| m(SV) | secondary vertex mass |
| MV2c20 | final discriminant |

## Main systematic uncertainty

$t\bar{t}$ modelling — affect the predicted flavour composition

- Good agreement data/simulation
- Ongoing data-based calibrations of flavour tag rates



# Conclusion

## Significant improvements in Run 2

- addition of the IBL
- algorithm updates

## Commissioning of *b*-tagging algorithms in early Run 2

good data/simulation agreement for basic tagger output variables and MV2c20

## Perspective

- very high light-jet rejection makes the calibration challenging
- *c*-jet rejection uncertaintty is becoming even more important

Backup

# Track Categories used by IP3D

| # | Description | Rate [%] | | |
|---|---|---|---|---|
| | | $b$ jets | $c$ jets | light jets |
| 0 | No hits in first two layers; exp. hit in L0 and L1 | 1.5 | 1.6 | 1.6 |
| 1 | No hits in first two layers; exp. hit in L0 and no exp. hit in L1 | 0.1 | 0.1 | 0.1 |
| 2 | No hits in first two layers; no exp. hit in L0 and exp. hit in L1 | 0.03 | 0.03 | 0.03 |
| 3 | No hits in first two layers; no exp. hit in L0 and L1 | 0.03 | 0.03 | 0.02 |
| 4 | No hit in L0; exp. hit in L0 | 2.4 | 2.3 | 2.1 |
| 5 | No hit in L0; no exp. hit in L0 | 0.9 | 0.9 | 0.9 |
| 6 | No hit in L1; exp. hit in L1 | 0.5 | 0.5 | 0.5 |
| 7 | No hit in L1; no exp. hit in L1 | 2.4 | 2.4 | 2.3 |
| 8 | Shared hit in both L0 and L1 | 0.01 | 0.01 | 0.04 |
| 9 | Shared pixel hits | 2.1 | 1.6 | 1.8 |
| 10 | Two or more shared SCT hits | 2.4 | 2.2 | 2.2 |
| 11 | Split hits in both L0 and L1 | 1.2 | 1.1 | 0.8 |
| 12 | Split pixel hit | 2.1 | 1.6 | 1.1 |
| 13 | Good: a track not in any of the above categories | 84.3 | 85.5 | 86.6 |

Figure : Description of the track categories used by IP2D and IP3D along with the fraction of tracks in each category for the $t\bar{t}$ sample. The categories are constructed with respect to the track quality, which is defined by the clusters (hits), from the silicon layers of the Inner Detector, used in the track reconstruction.

# MV2 Inputs Variables

| Input | Variable | Description |
|---|---|---|
| Kinematics | $\mathrm{p}_T(jet)$ | Jet transverse momentum |
| | $\eta(jet)$ | Jet pseudo-rapidity |
| IP2D, IP3D | $\log(P_b/P_{\mathrm{light}})$ | Likelihood ratio between the $b$- and light jet hypotheses |
| | $\log(P_b/P_c)$ | Likelihood ratio between the $b$- and $c$-jet hypotheses |
| | $\log(P_c/P_{\mathrm{light}})$ | Likelihood ratio between the $c$- and light jet hypotheses |
| SV | $m(\mathrm{SV})$ | Invariant mass of tracks at the secondary vertex assuming pion masses |
| | $f_{\mathrm{E}}(\mathrm{SV})$ | Fraction of the charged jet energy in the secondary vertex |
| | $N_{\mathrm{TrkAtVtx}}(\mathrm{SV})$ | Number of tracks used in the secondary vertex |
| | $N_{\mathrm{2TrkVtx}}(\mathrm{SV})$ | Number of two track vertex candidates |
| | $L_{xy}(\mathrm{SV})$ | Transverse distance between the primary and secondary vertices |
| | $L_{xyz}(\mathrm{SV})$ | Distance between the primary and secondary vertices |
| | $S_{xyz}(\mathrm{SV})$ | Distance between the primary and secondary vertices divided by its uncertainty |
| | $\Delta R(\mathrm{jet}, \mathrm{SV})$ | $\Delta R$ between the jet axis and the direction of the secondary vertex relative to the primary vertex |
| Jet Fitter | $N_{\mathrm{2TrkVtx}}(\mathrm{JF})$ | Number of 2-track vertex candidates (prior to decay chain fit) |
| | $m(\mathrm{JF})$ | Invariant mass of tracks from displaced vertices assuming pion masses |
| | $S_{\mathrm{xyz}}(\mathrm{JF})$ | Significance of the average distance between the primary and displaced vertices |
| | $f_{\mathrm{E}}(\mathrm{JF})$ | Fraction of the charged jet energy in the secondary vertices |
| | $N_{\text{1-trk vertices}}(\mathrm{JF})$ | Number of displaced vertices with one track |
| | $N_{\geq\text{2-trk vertices}}(\mathrm{JF})$ | Number of displaced vertices with more than one track |
| | $N_{\mathrm{TrkAtVtx}}(\mathrm{JF})$ | Number of tracks from displaced vertices with at least two tracks |
| | $\Delta R(\vec{p}_{\mathrm{jet}}, \vec{p}_{\mathrm{vtx}})$ | $\Delta R$ between the jet axis and the vectorial sum of the momenta of all tracks attached to displaced vertices |

Figure : The 24 input variables used by the MV2c00 and MV2c20 $b$-tagging algorithm