# Search for $t\bar{t}H(bb)$ in LHC Run-2 data

# b-tagging optimisation for Run-2: IP2D/IP3D taggers

**Yulia Rodina**

Supervisors: **Laurent Vacavant,** CPPM, Aix-Marseille Université and CNRS/IN2P3, Marseille
and
**Aurelio Juste Rozas**, IFAE, Universitat Autònoma de Barcelona and ICREA, Barcelona

PhD Day, CPPM, 25 November 2015

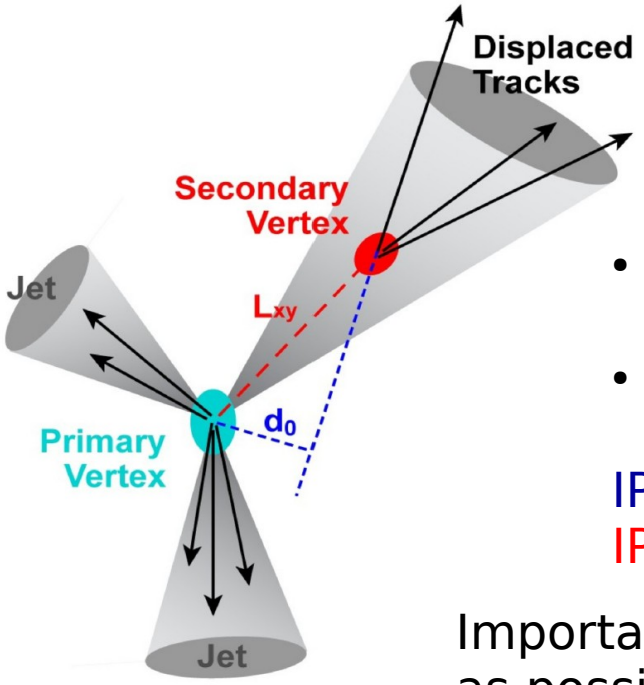
# b-tagging

For the first year of my PhD I work on b-tagging
Goal of b-tagging: identify jets stemming from hadronisation of b-quarks
B-tagging used in many analysis domains:

- Top physics
- Higgs physics
- Beyond Standard Model physics

b-tagging algorithms rely on b-hadron properties:

- High mass (~5 GeV)
- Relatively long lifetime (~1.5 ps)




b-tagging algorithms:

- Spatial tagging
  - Impact parameter based: IP2D, IP3D
  - Secondary vertex based: SV0, SV1, JetFitter
  - Multivariate: MV1, MV2
- Soft lepton tagging

## Impact parameter based taggers

- Use lifetime signed impact parameter (IP) significance of tracks matched to a jet
- IP - distance of closest approach to the primary vertex (PV)

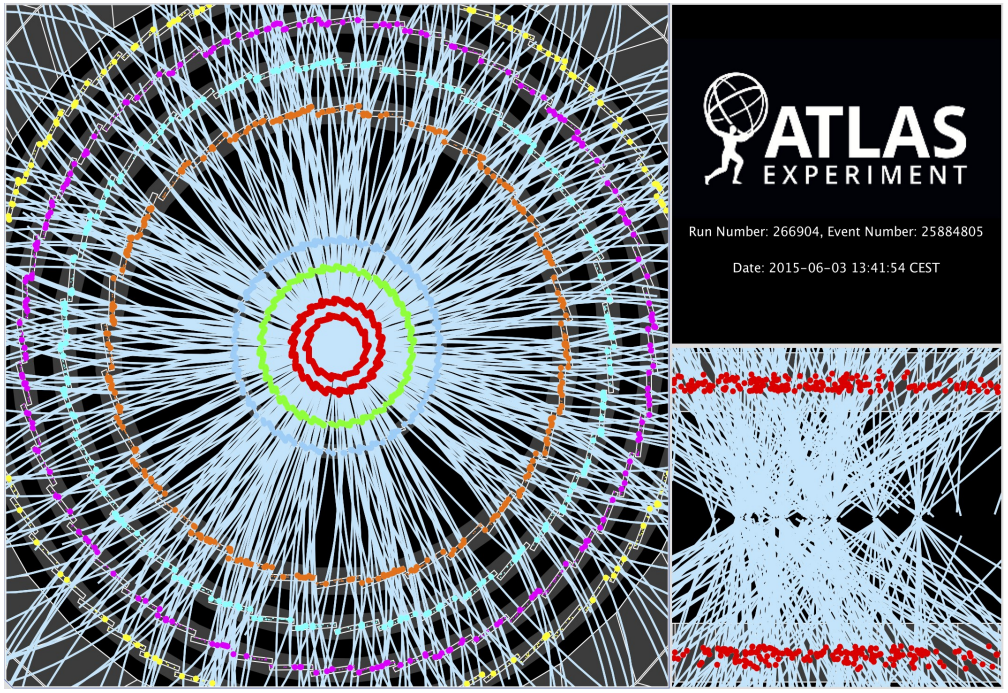
IP2D: only use transverse IP in 1D likelihood
IP3D: combine with longitudinal IP in 2D likelihood

Important to measure particle as close to interaction point as possible

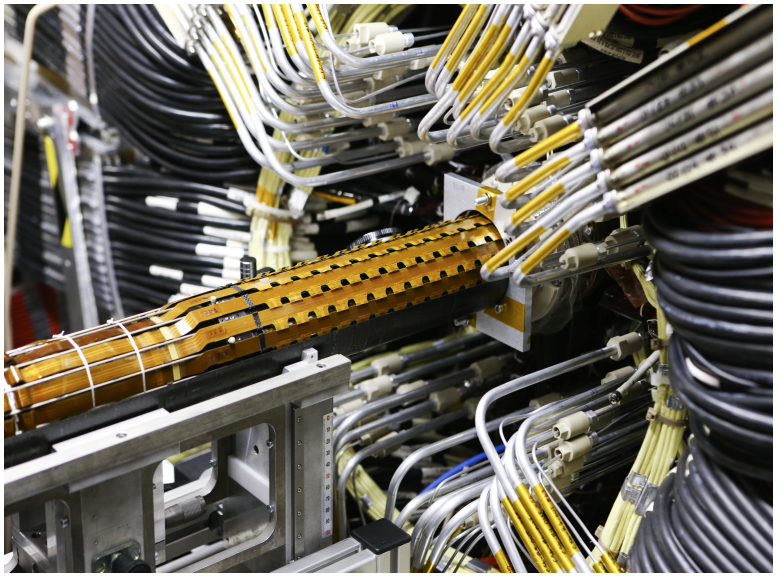
# Insertable B-Layer (IBL)

Major ATLAS inner detector upgrade for LHC Run-2: addition of Insertable B-layer (IBL) as fourth innermost pixel layer

- Radius (IBL) ~ 3.3cm
- Radius (former Layer 0) ~ 5cm




Collision event collected by full ATLAS tracker (June 2015)



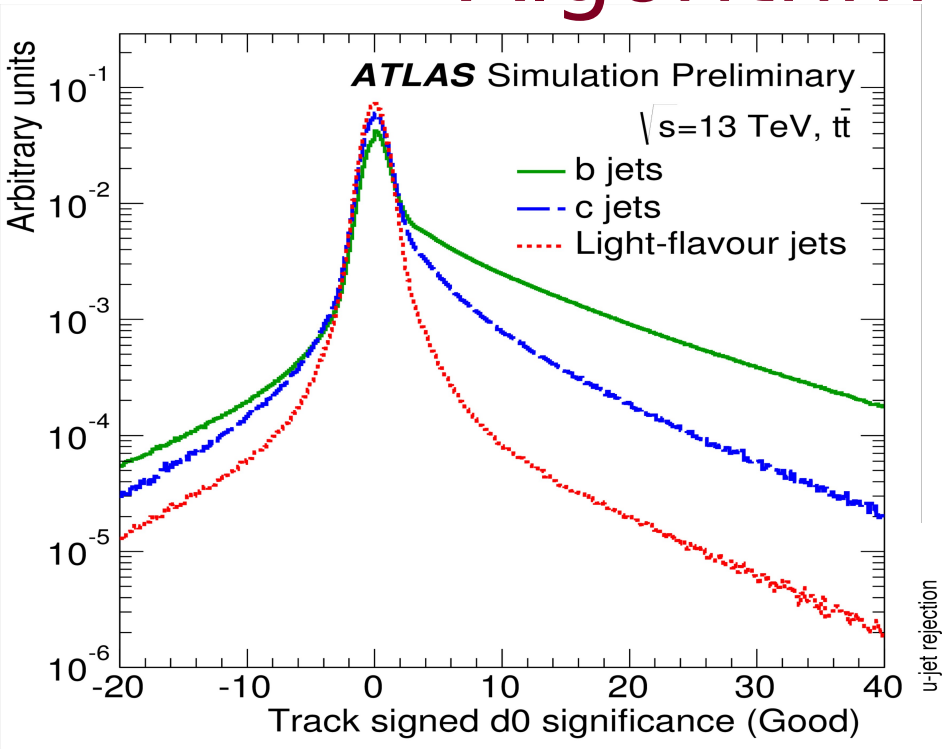
IBL installation (May 2014)

Advantages of IBL:

- Proximity to the interaction region
- Higher granularity (50μm x 250μm instead of 50μm x 400μm the case of former Layer 0)
- Increase pixel measurements of a track from 3 to 4

The key goal of IP3D algorithm optimization for Run-2 is to take advantage of the presence of IBL

# Algorithm optimisation

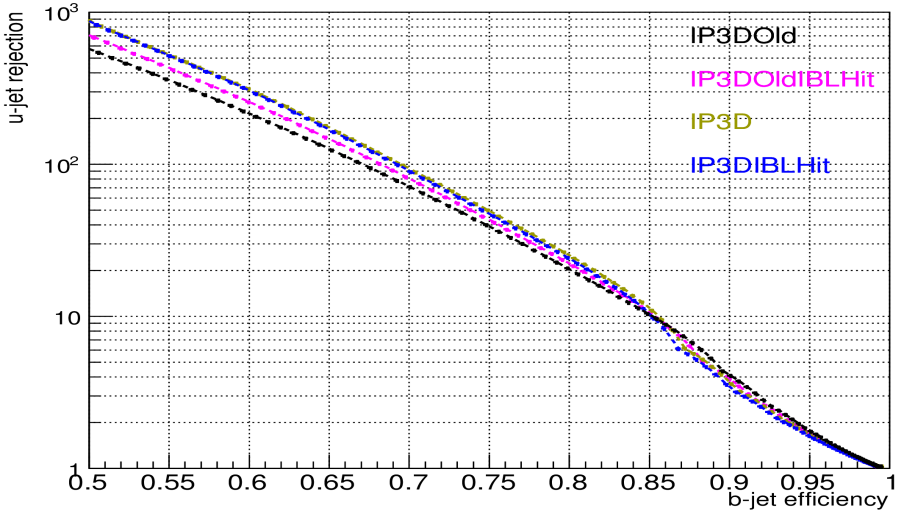


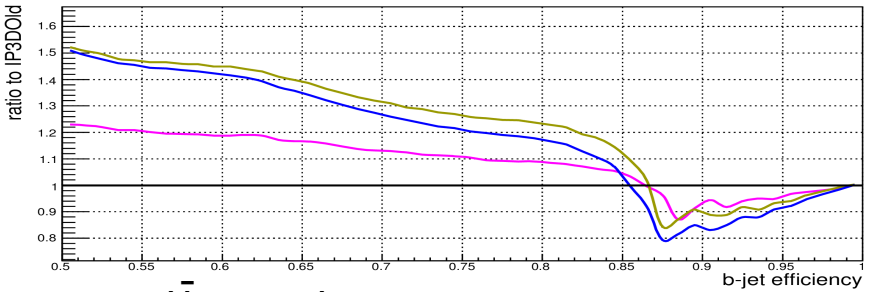

Some tracks give worse resolution:

- those with missing hits in some tracker layers
- with ambiguities in pattern recognition

What we need – divide track into categories and compute IP3D weight for each of them separately:

$$w_{track} = \frac{p_b}{p_{light}}; \qquad w_{jet} = \sum_{tracks} \log\left(w_{track}\right).$$




New track grading for Run-2 conditior was developed, taking into consideration new track variables available: 14 track categories instead 5 used in Run-1

IP3D versions with new grading give better performance than with old grading for up to ~86 efficiency region

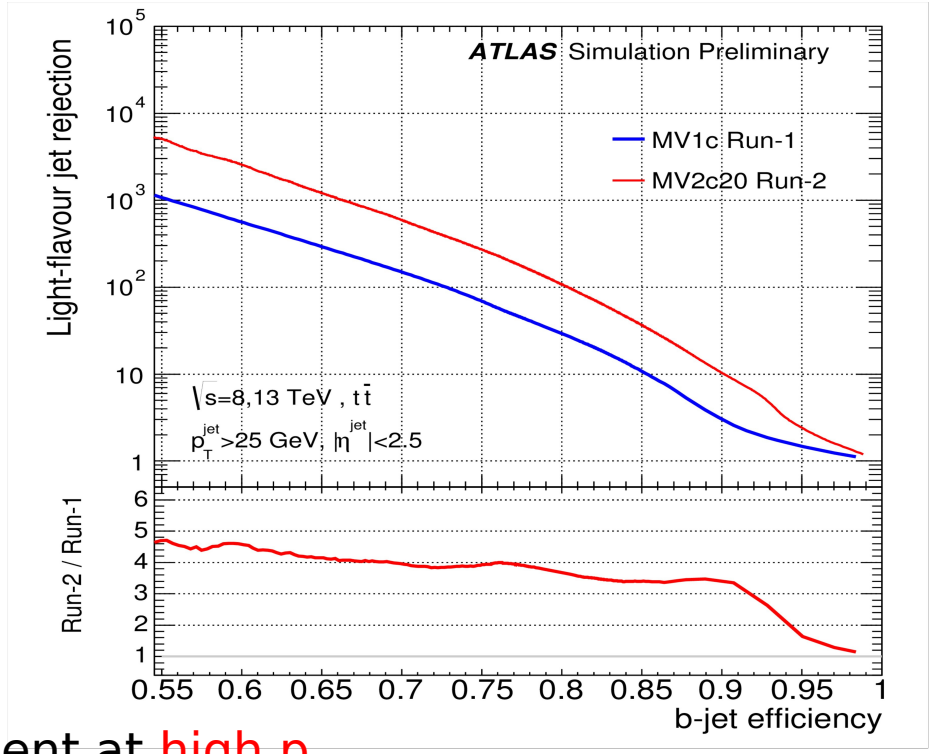
$t\bar{t}$ events

# b-tagging performance for Run-2

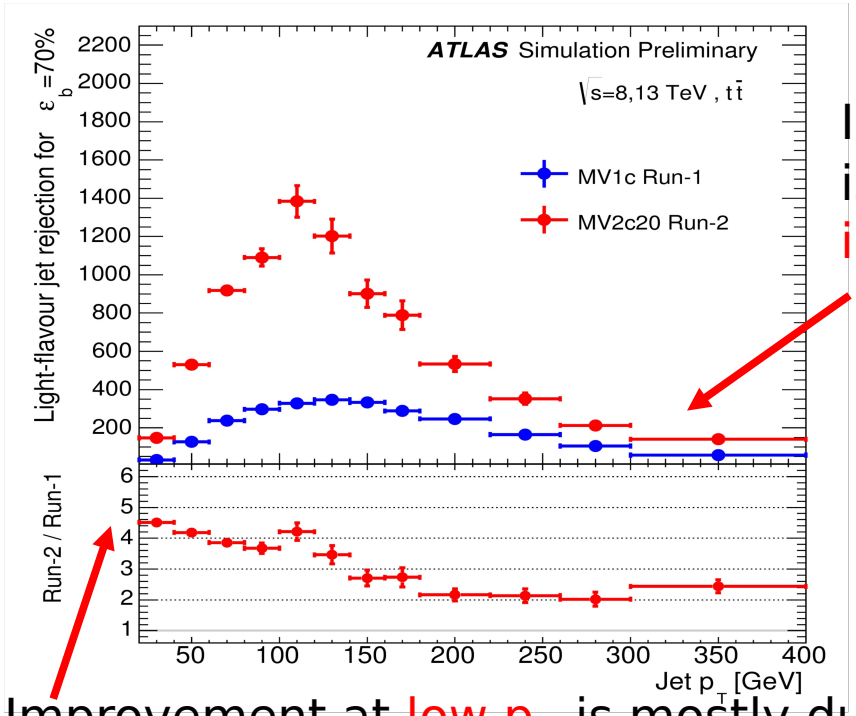
A b-tagging performance improvement is expected to be achieved in Run-2 due to:

- Addition of the IBL
- Many algorithmic updates in track reconstruction and b-tagging

Comparison of the default Run-2 b-tagging algorithm MV2c20 and the equivalent Run-1 b-tagging algorithm MV1c.




Light jet rejection as a function of jet $p_T$ given a fixed b-jet efficiency of 70% in each bin:



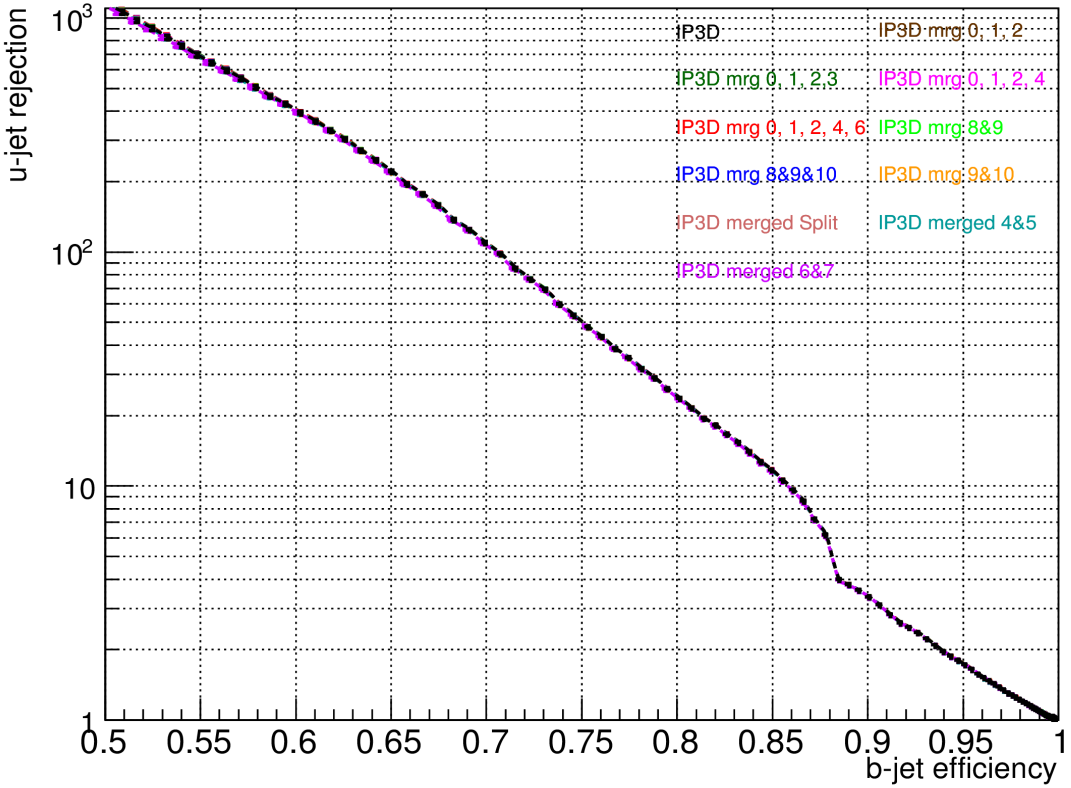

Improvement at high $p_T$ is due to algorithm improvements

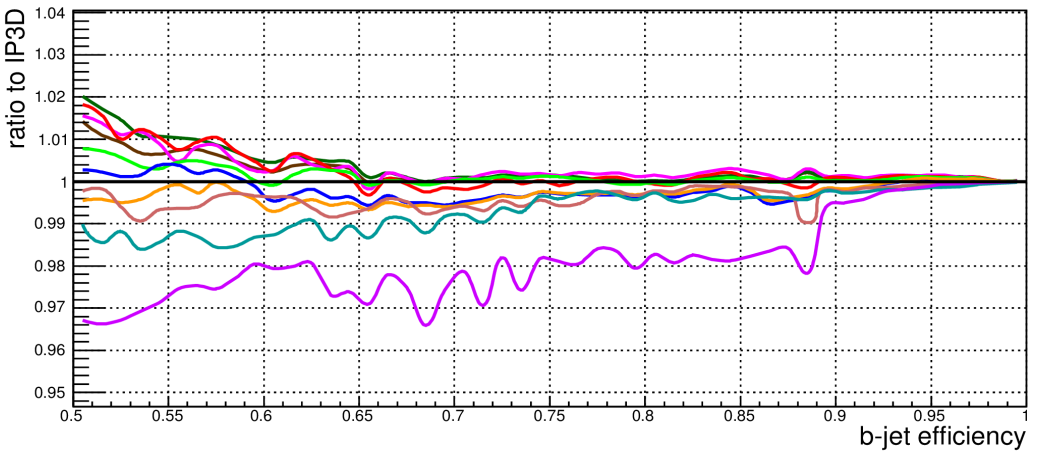
At 70% efficiency light-flavour jet rejection in Run-2 is improved inclusively by a factor of 4 compared to Run-1. This corresponds to a relative improvement of ~10% in efficiency at a constant light-jet rejection!

Improvement at low $p_T$ is mostly due to addition of the IBL

Thanks to this expected improvement on the signal yield for ttH(bb): ~ 46%

# Merging categories




0) No hit in 2 first layers; expected in both

1) No hit in 2 first layers; exp hit in IBL, not exp in BLayer

2) No hit in 2 first layers; not exp in IBL, exp in BLayer

3) No hit in 2 first layers; not expected in both

4) No hit in IBL; exp in IBL

5) No hit in IBL, not exp in IBL

6) No hit in BLayer, exp in BLayer

7) No hit in BLayer, not exp in BLayer

8) Shared hits in both 2 first layers

9) Shared pixel hits (other layers)

10) Shared SCT hits>1

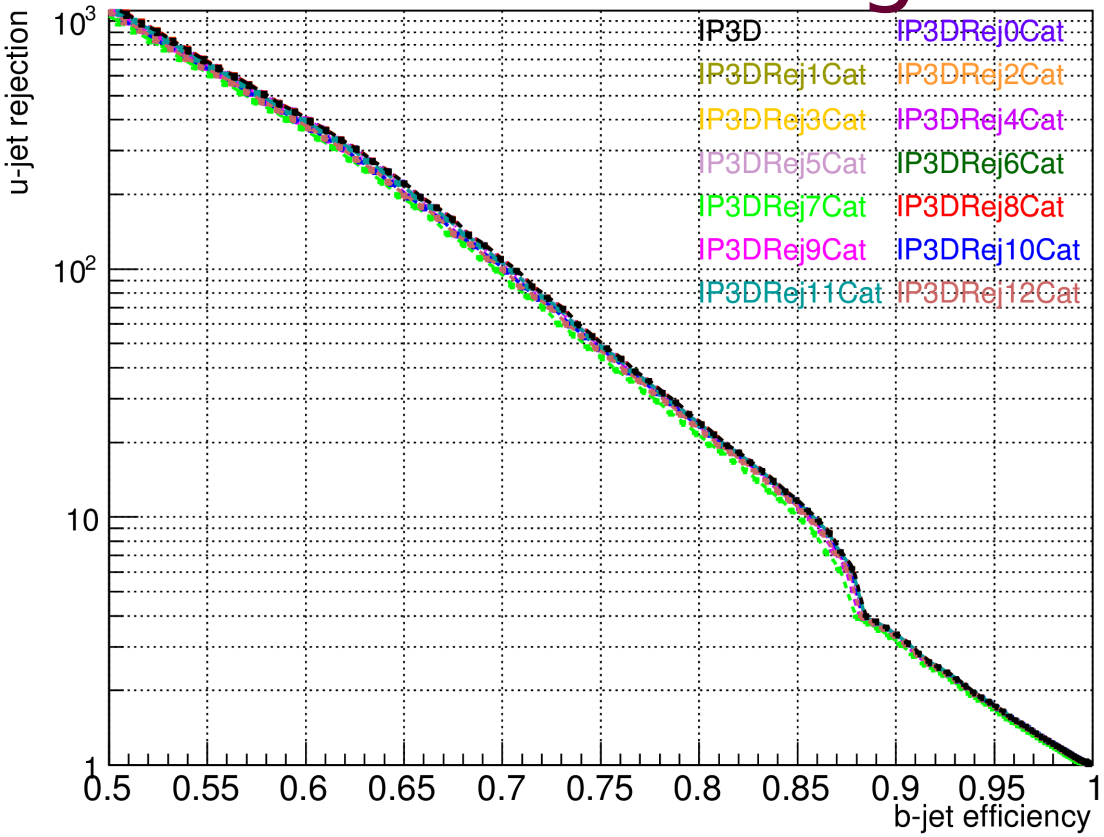
11) Split hits in both 2 first layers

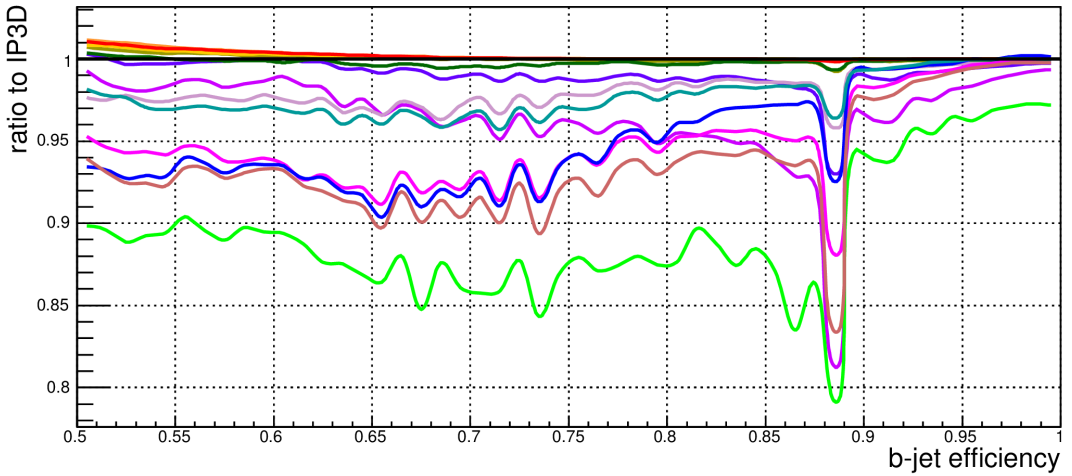
12) Split pixel hits (other layers)

13) Good

- Merging of some categories is possible as giving just slight fluctuation of the performance, while merging others can decrease performance
- Also checked pT dependance, it is not changing much for different pT, see backup

# Removing categories




0) No hit in 2 first layers; expected in both

1) No hit in 2 first layers; exp hit in IBL, not exp in BLayer

2) No hit in 2 first layers; not exp in IBL, exp in BLayer

3) No hit in 2 first layers; not expected in both

4) No hit in IBL; exp in IBL

5) No hit in IBL, not exp in IBL

6) No hit in BLayer, exp in BLayer

7) No hit in BLayer, not exp in BLayer

8) Shared hits in both 2 first layers

9) Shared pixel hits (other layers)

10) Shared SCT hits>1

11) Split hits in both 2 first layers

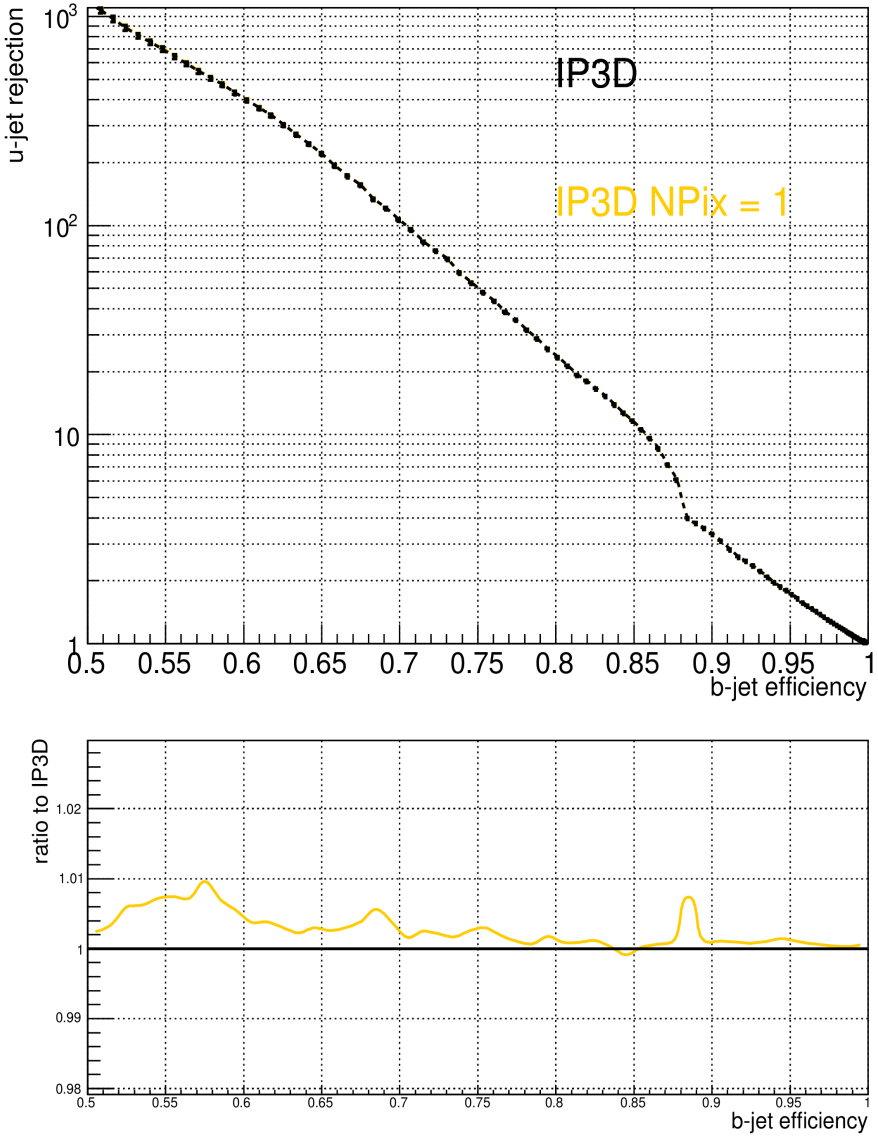
12) Split pixel hits (other layers)

13) Good

Removing any of the categories either decrease the performance of don't change it, but impacts of the categories are different

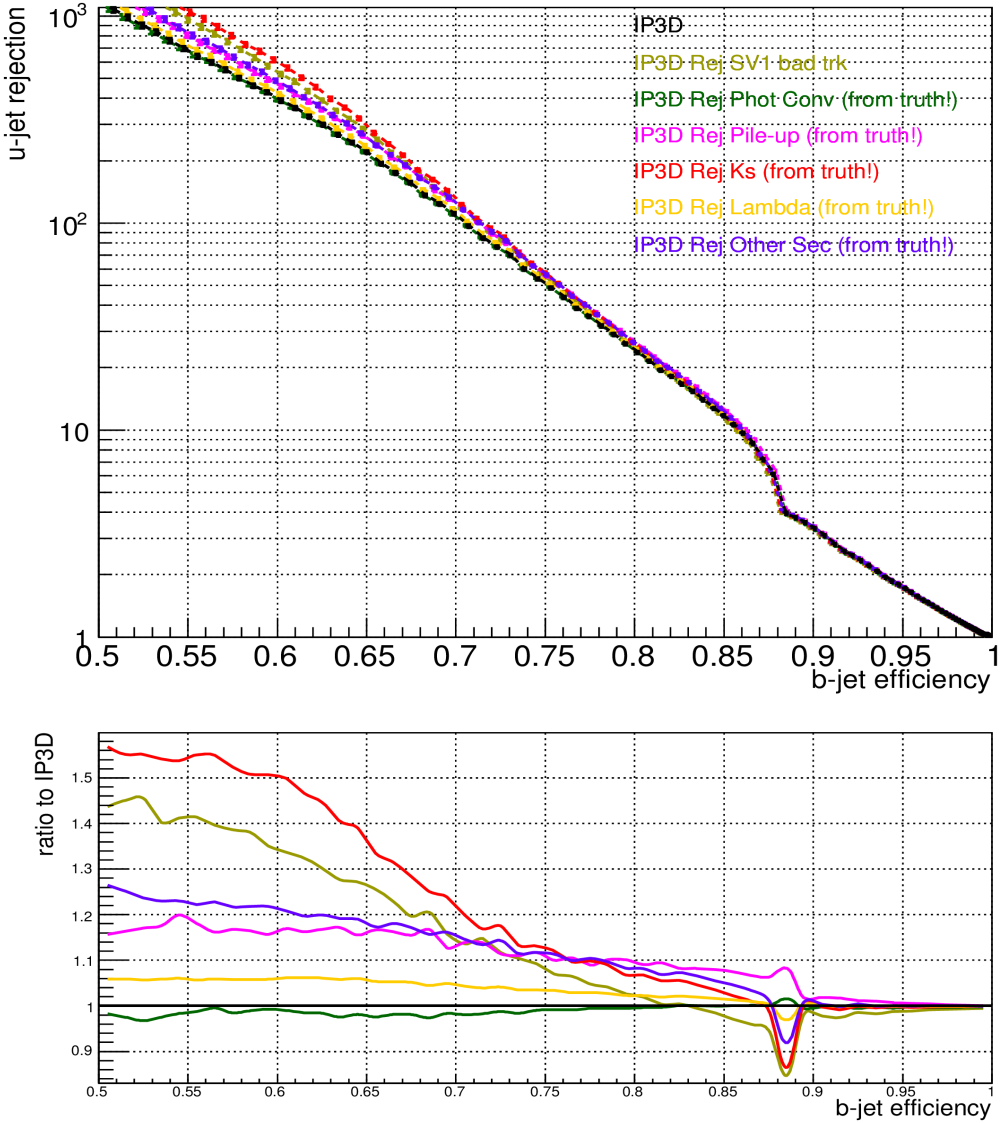
# IP3D track selection optimisation

## N Pix Hit >= 1




- Doesn't seem to be an impact while using requirement >= 1 pixel hit instead of >= 2 pixel hit

## “Bad” track rejection




- Tool removing “bad” tracks (coming from decay of long-lived particles, such as Ks or Lamda)

# $t\bar{t}H(bb)$

LHC Run-1: Standard Model like Higgs boson was discovered

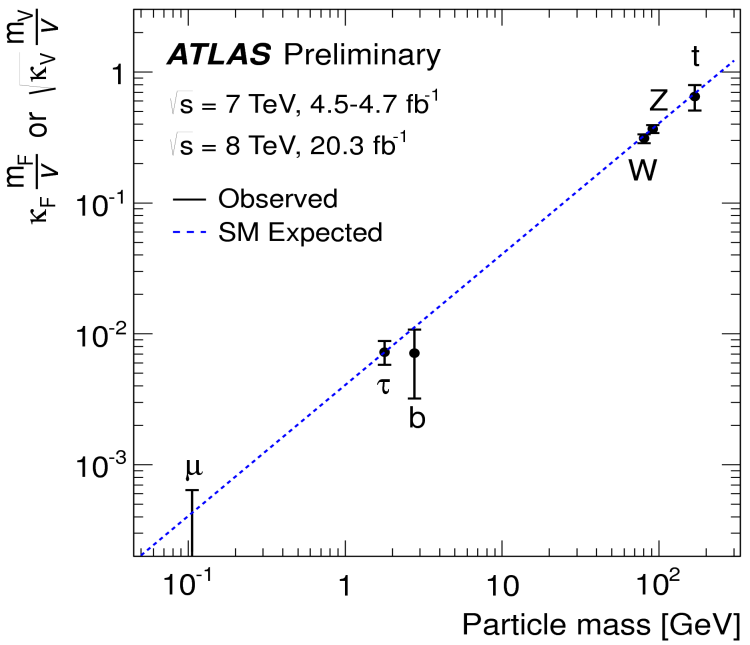
$m_H$ = 125.09±0.21(stat.)±0.11(syst.) GeV

But not yet observed in $t\bar{t}H$ channel!

Run-1 $t\bar{t}H$ search result:

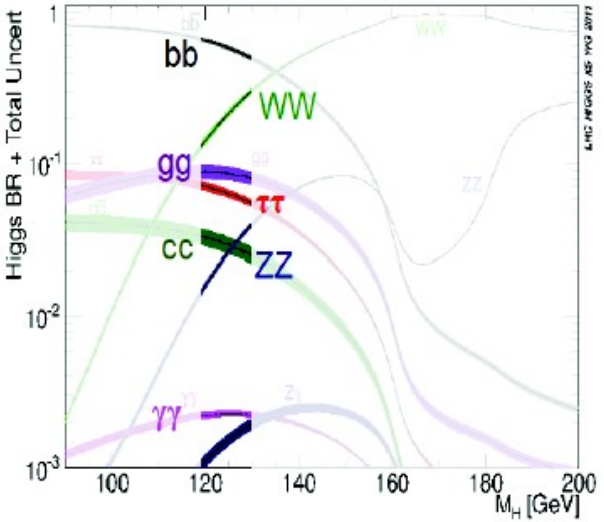
$$\mu = \frac{\sigma}{\sigma_{SM}} = 1.5 \pm 1.1$$

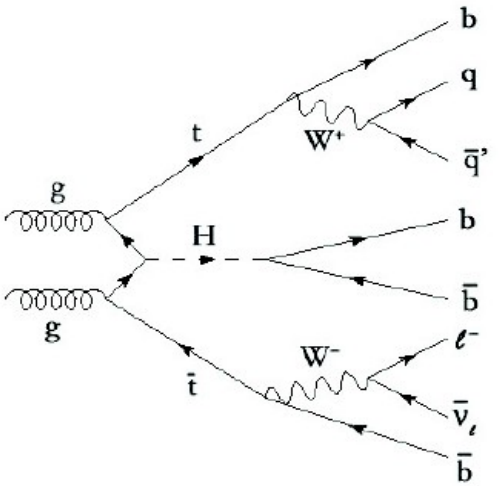
- $t\bar{t}H$ has the lowest cross section of Higgs production at LHC
- with higher energies the impact is becoming larger




Why important to study $t\bar{t}H$:
gives direct access to the top Yukawa coupling → may shed a light on the origin of the mechanism of the EW symmetry breaking

H→bb branching ratio - the largest in the SM

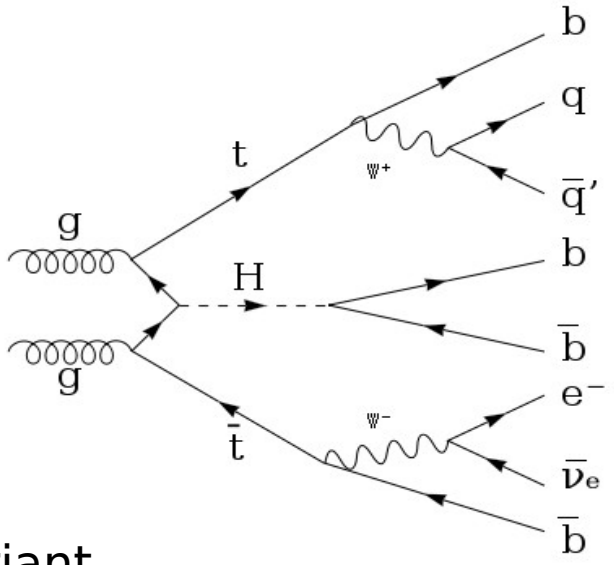



4 b-jets in the final state → important to identify them

# Work on $t\bar{t}H(bb)$ analysis

Second year of PhD: starting with actual data analysis for the $t\bar{t}H(bb)$ search (leptons+jets channel)




Working on discriminant method for distinguishing signal ($t\bar{t}H$) and background ($t\bar{t}$+jets)

Idea of method: probability to be signal can be expressed using PDFs of the reconstructed invariant masses in the event:

$$P^{ttH} = P^{ttH}_{M_{l\nu b_l}}\left(l,\nu,b_l\right)\cdot P^{ttH}_{M_{b_1 b_2}}\left(b_1,b_2\right)\cdot P^{ttH}_{M_{q_1 q_2 b_h}}\left(q_1,q_2,b_h\right)\cdot P^{ttH}_{M_{q_1 q_2}}\left(q_1,q_2\right).$$

The final discriminant between signal and background:

$$D = \frac{P^{ttH}}{P^{ttH} + P^{ttbb}}$$

# Summary and outlook

- IP2D/IP3D b-tagging algorithms optimization for Run-2 and IBL
  - New track grading for Run-2 conditions, using new track variables: 14 track categories instead of 5 in Run-1
  - Applying mechanism of removal of tracks coming from long-lived particles, which allows to increase b-tagging performance significantly
  - Optimisation of track selection cuts

- Now working on software for the ttH(bb) analysis
- The further aim is applying new discriminant method for distinguishing signal and background

# Backup material

# Tracking categories

## Run 1

1) B-layer hits = 0
2) Shared B-layer hits >0
3) Shared pixel hits (other layers)>0
4) Shared SCT hits>1
5) Good

## Run 2

1) No hit in 2 first layers; exp hit in IBL>0 & exp hit in BLayer>0
2) No hit in 2 first layers; exp hit in IBL >0 & exp hit in BLayer=0
3) No hit in 2 first layers; exp hit in IBL =0 & exp hit in BLayer>0
4) No hit in 2 first layers; exp hit in IBL =0 & exp hit in BLayer=0
5) No hit in IBL, exp. hit in IBL>0
6) No hit in IBL, exp. hit in IBL=0
7) No hit in BLayer, exp. hit in BLayer>0
8) No hit in BLayer, exp. hit in BLayer=0
9) Shared hits in both 2 first layers
10) Shared pixel hits (other layers)
11) Shared SCT hits>1
12) Split hits in either IBL or BLayer
13) Split pixel hits (other layers)
14) Good