Tagging jets with double B-hadron using multiple secondary vertices

Royer Edson Ticse Torres

Supervisor: Arnaud Duperrin

Thanks to Georges Aad & Laurent Vacavant

CPPM – December 1, 2014

Aix*Marseille université

CENTRE DE PHYSIQUE DES PARTICULES DE MARSEILLE
CPPM

Higgs search & LHC

- **Higgs searches**

  - Observed in:
  $H\to\gamma\gamma$, $H\to ZZ\to 4l$, $H\to WW\to l\nu l\nu$, $H\to\tau\tau$
  consistent with SM Higgs

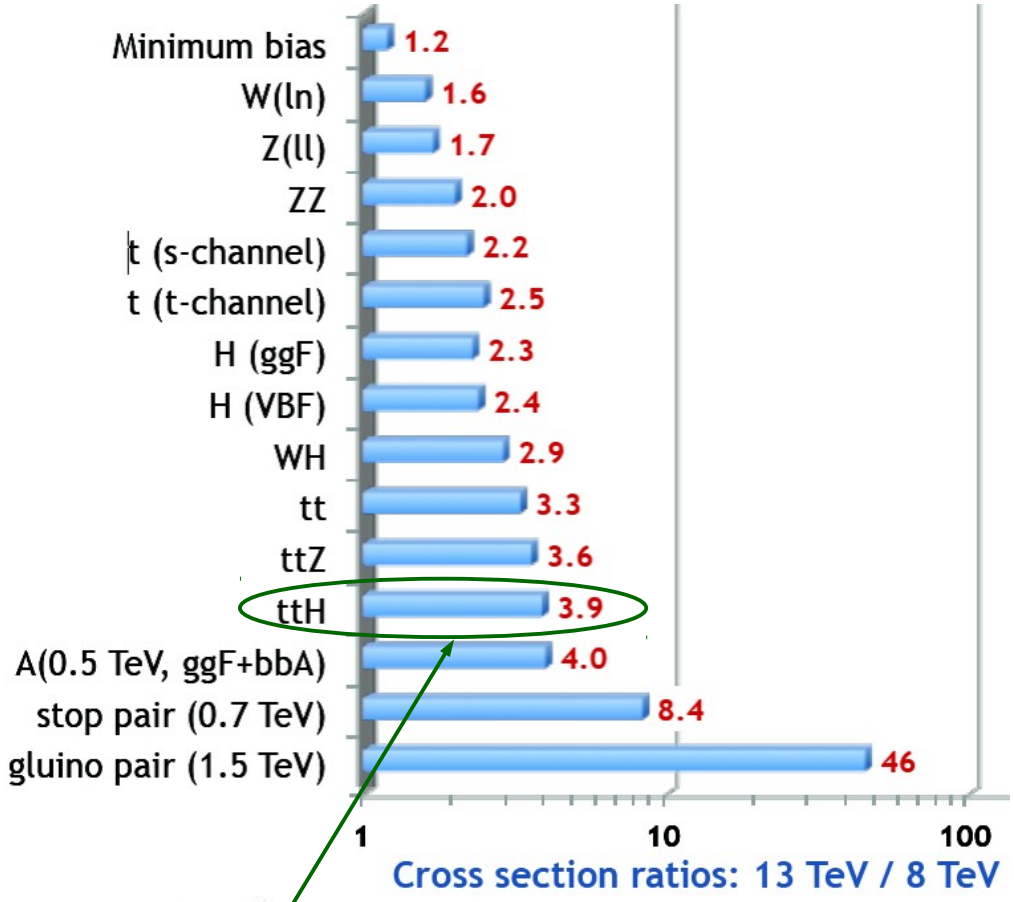
  - Precise measurement of mass, spin / CP, couplings is very important to investigate for possible deviations from the SM.

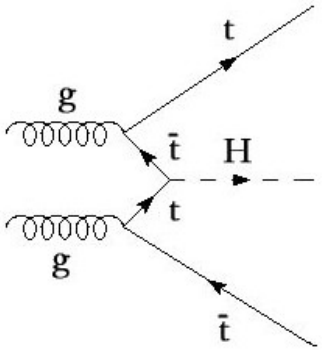
- **LHC Run2 (2015)**

  - Increase center of mass energy to 13TeV, high luminosity at 25 ns bunch spacing

- **Associated prod. with tt**

  - At NLO σ(ttH) = 507 fb to 13TeV
  - decay mode $H\to bb$, (BR=0.58), dominant mode but large background
  - ttH observation and Yukawa coupling measurement will be a highlight of Run 2
  - Subject of my PhD thesis





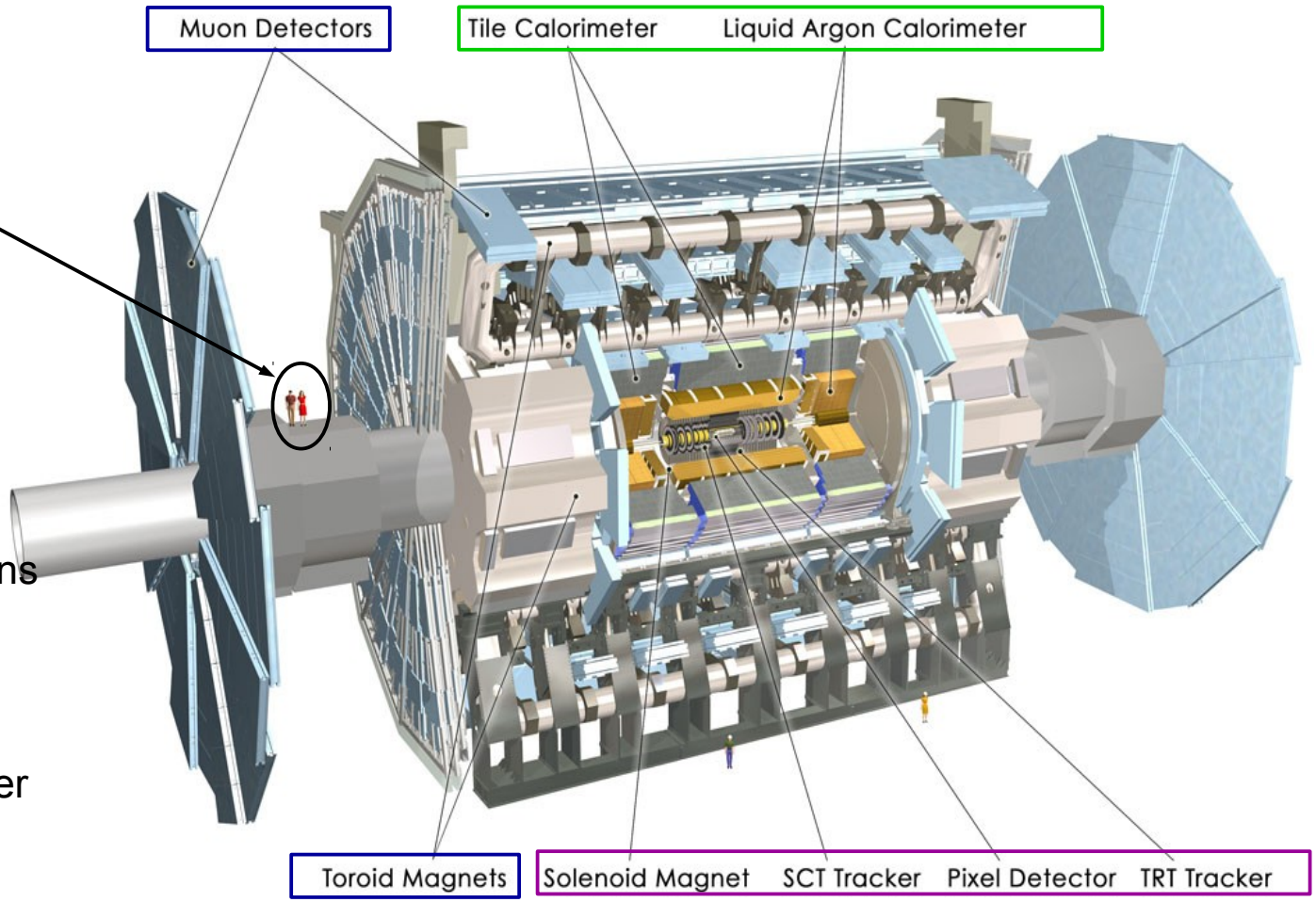
The ttH channel is the only way to directly constrain the top Yukawa coupling at the LHC

ATLAS detector

- **General-purpose detector**
  Cylinder around the beam pipe of
  - Height : 25 m
  - length : 46 m

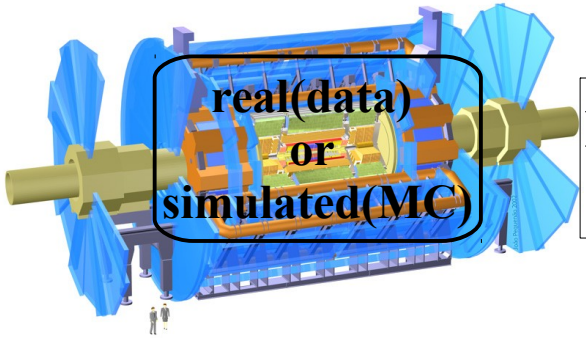
- **Onion structure :**
  - **Tracker + Solenoid magnet**
    tracking and vertexing.
  - **Calorimeters**
    reconstruct electrons and jets
  - **Muon tracker + Toroidal Magnets**
    reconstruct momentum of muons

- **A new subdetector in Run2: Insertable B-Layer(IBL)**,
  -Inserted between the existing pixel detector and a new(smaller radius) beam-pipe.
  -Tracking and vertexing performance are expected to be greatly improved .



One of the most complex scientific instruments ever built !

Atlas analysis model - Run2



RAW/ RDO → reconstruction (Athena) → xAOD → Derivation Reduction Framework (Athena) → derived xAOD → Athena or ROOT analysis → final analysis ntuple

- **New xAOD data model**
  - format readable by both Athena and ROOT
  - faster and compact
  - more object oriented
- **Big migration of the Atlas software**
  - migration of the reconstruction code to the xAOD format
- **Software restructuring in b-tagging**
  - I redesigned the vertexing part in the b-tagging code to xAOD format.
  - I helped in the migration and validation of the b-tagging algorithms

Tagging jets with double B-hadron

- **Motivation: ttH(bb)**

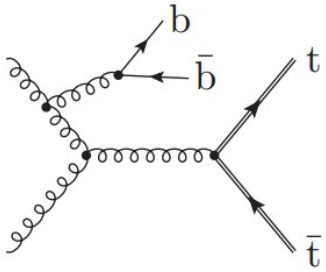
  Main backgrounds:

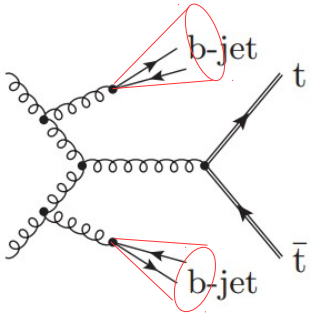
  - Irreducible: tt+bb, give the same final state signature as the signal

  -Reducible: tt+cc and tt+light jets

  b-tagging plays crucially important role!

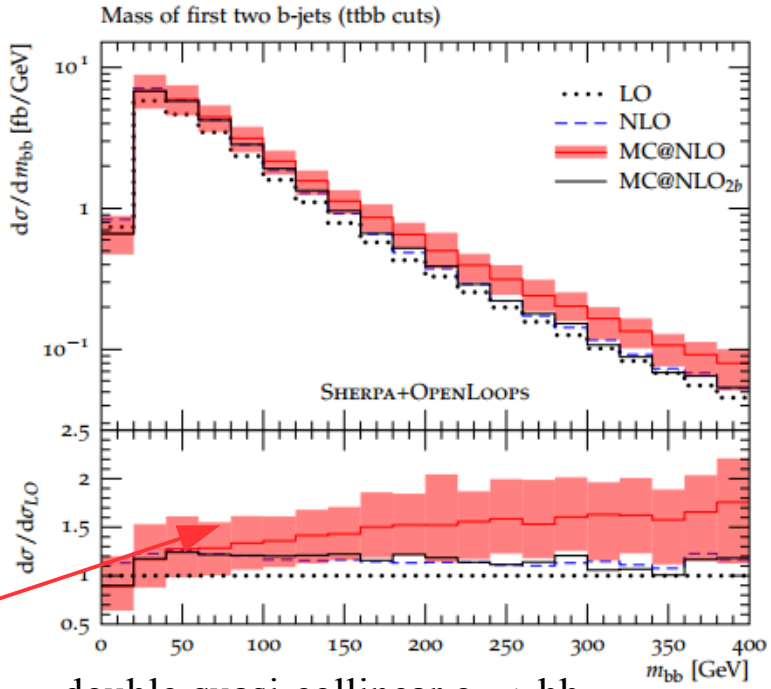
- tt+bb, hard to control, large theory uncertainties. e.g.





g→bb at small angle might be reconstructed as one bb-jet

- Gluon splitting to bb not perfectly modeled by different MC



- double quasi-collinear g → bb (swithed off inPS->black line)
- contribution very relevant for region $m_{bb} > 100$GeV

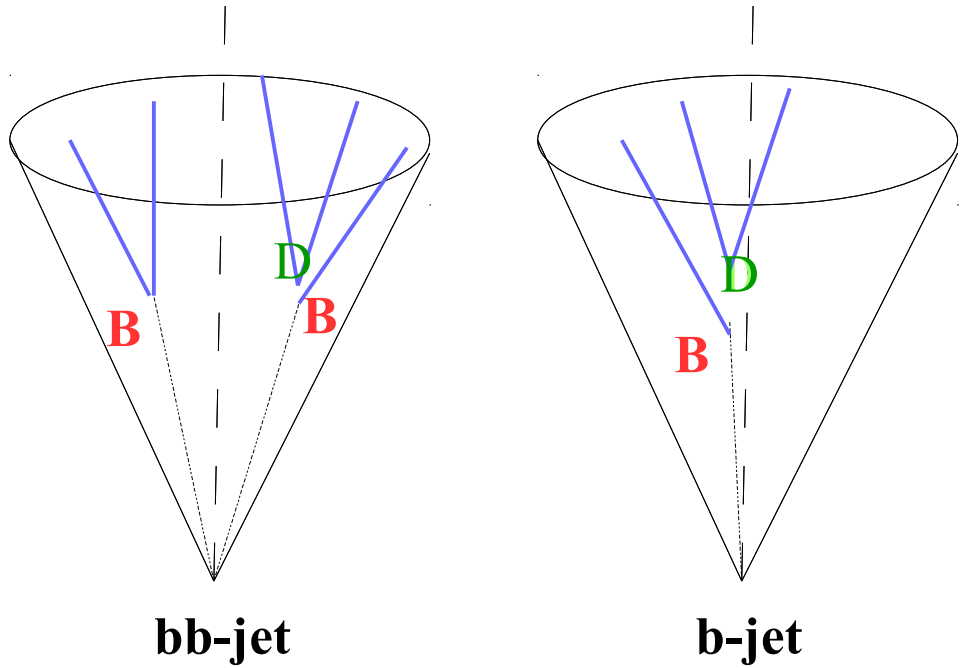
- Need a way to control in data

  - bb-jet tagger

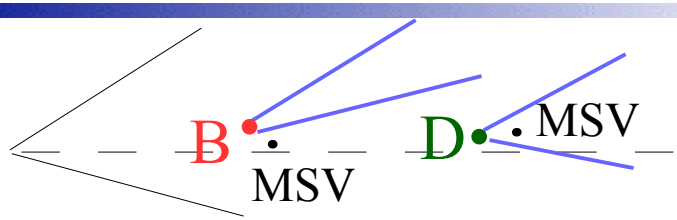
Tagging jets with double B-hadron

- Current b-tagging algorithms do not distinguish between b-jets and bb-jets.
- We use multi secondary vertex finder(**MSV**) to reconstruct multiple vertices within anti-kt R=0.4 jets, with pT>20GeV and |η|<2.5
- Several studies have been done to understand vertexing performance and ambiguous cases(e.g. B/C separation, fakes vertices) in **MSV**.

Goal: Identify jets containing two B-hadrons

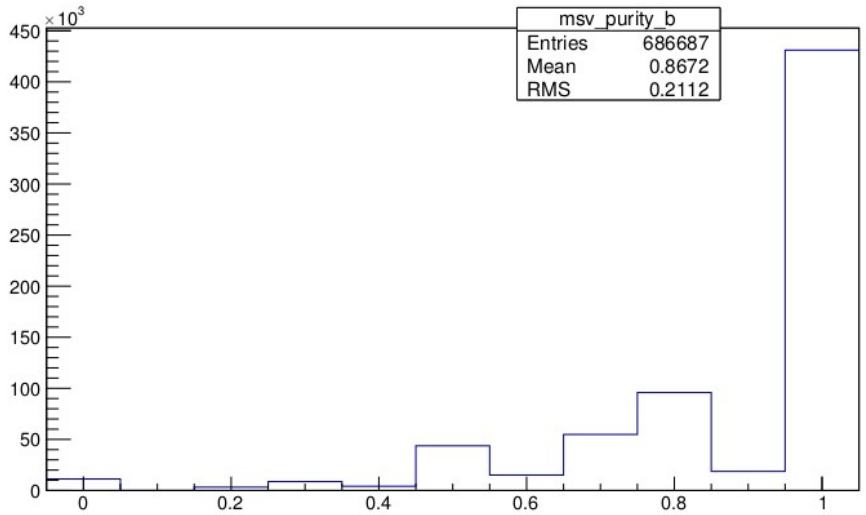


MSV purity & efficiency



- Purity =

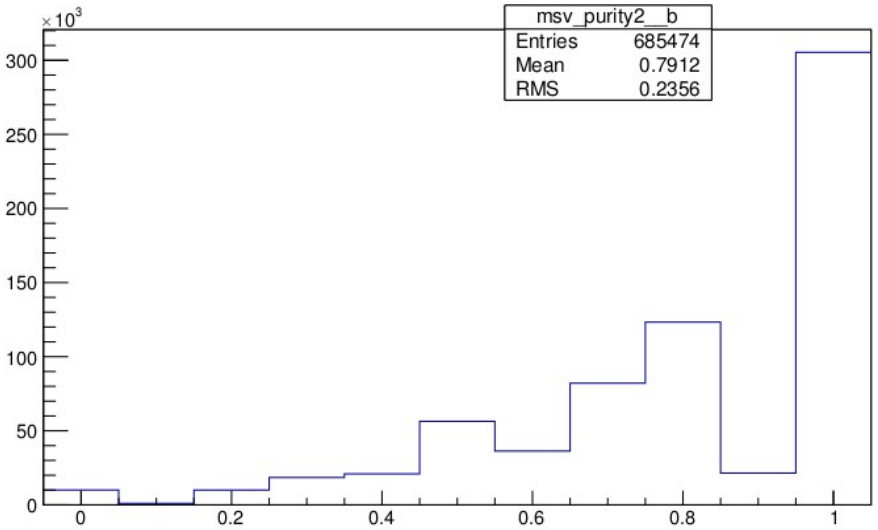
$$\frac{\text{B/C tracks in all vertices in jet}}{\text{all tracks in all vertices in jet}}$$



Purity of the tracks in the vertices

- Efficiency =

$$\frac{\text{B/C tracks in all vertices in jet}}{\text{all B/C tracks in jet}}$$

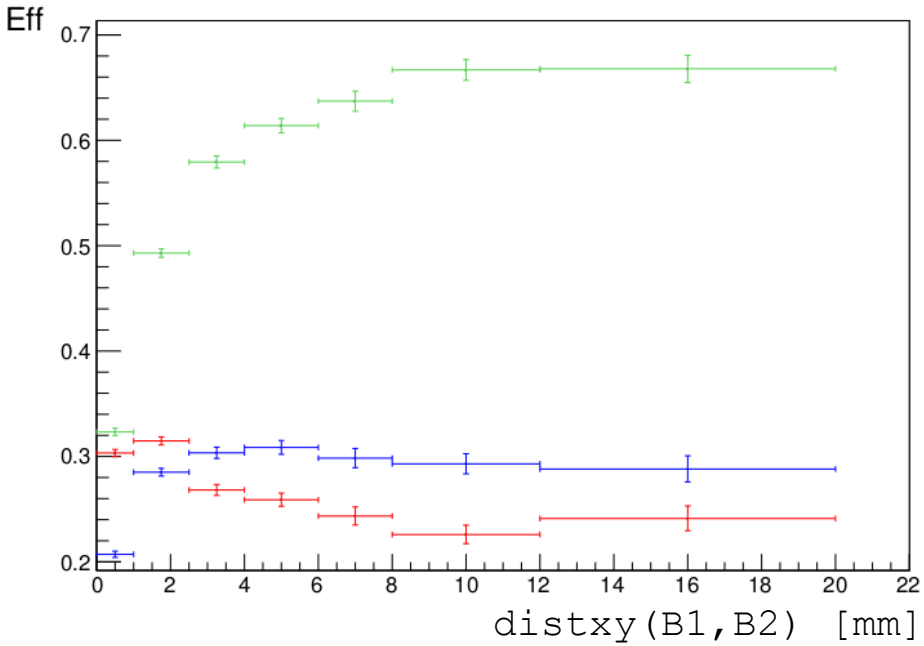


Efficiency to use B/C tracks in the vertices

Multi-vertexing in bb-jets

- Fraction of bb-jets with:

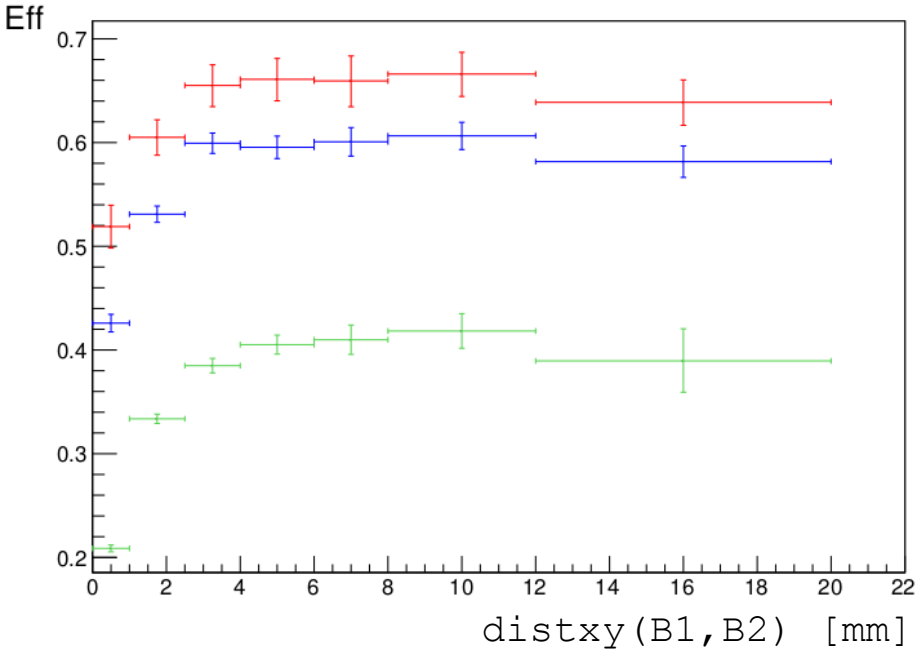
  1 reco vertex, exactly 2 reco vertices, at least 2 reco vertices



around 47% of the bb-jets with at least 2 reco vertices

- Eff = # bb-jets with nvtx≥2 / # bb-jets

  110≤pT<200 GeV
  60≤pT<110 GeV
  20<pT<60 GeV



efficiency increase with the pT

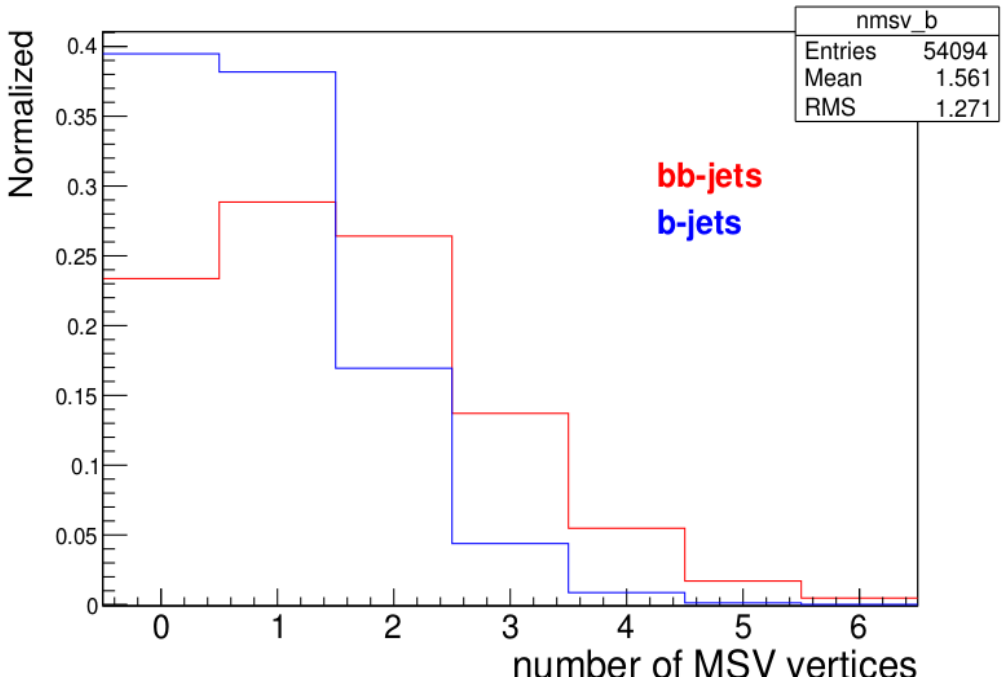
Multivariate Analysis with Boosted Decision Trees

We use a boosted decision trees(BDT) to separate bb-jets from different flavours using multi-vertexing properties.

- Optimized for b-jets rejection while keep light jets rejection at a good rate.
- Training in a mixture of jet flavours(b, c, light and cc-jets)
- Two versions:

MultiSVbb1 (12 variables)
- Use only vertex properties as input variables

MultiSVbb2 (14 variables)
- Include additional topological variables

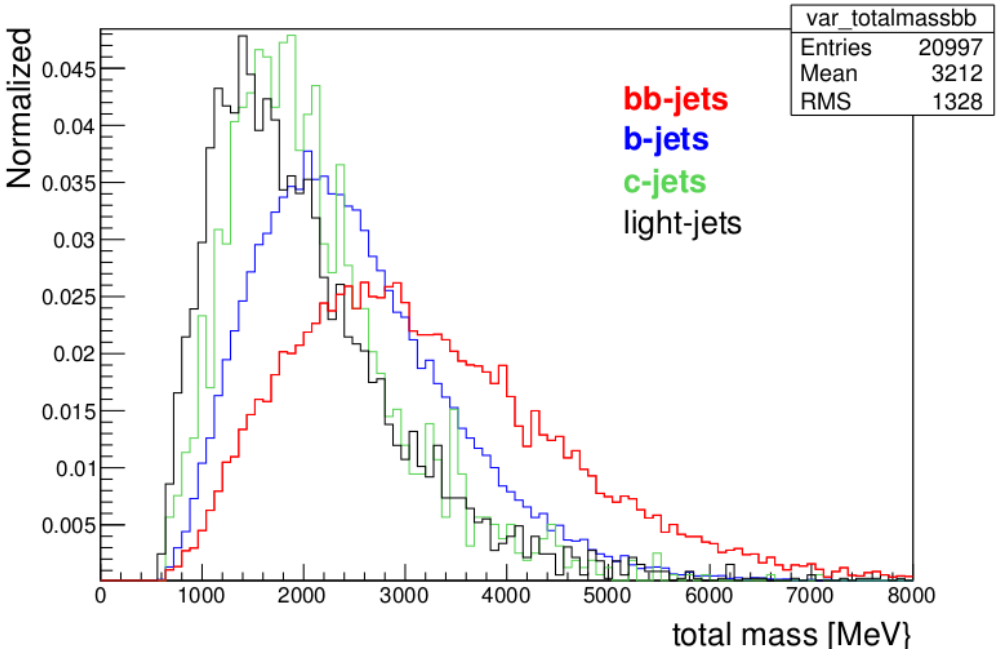


Require at least 2 MSV vertices

Discriminating variables:

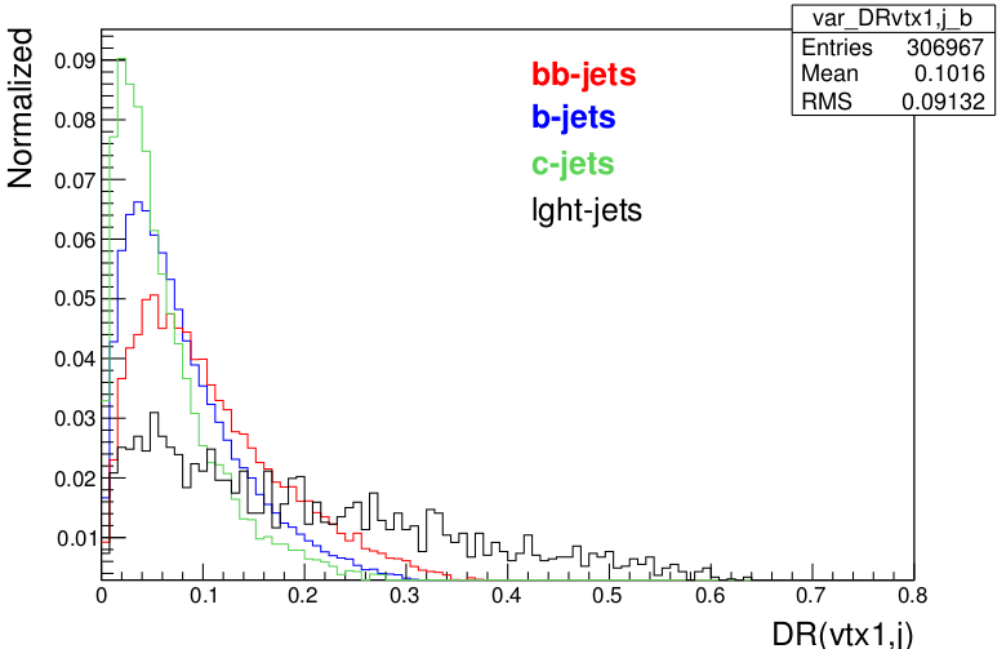
Input variables to train the BDT:

- Total mass of vertices



total mass for bb-jets is greater than other flavours

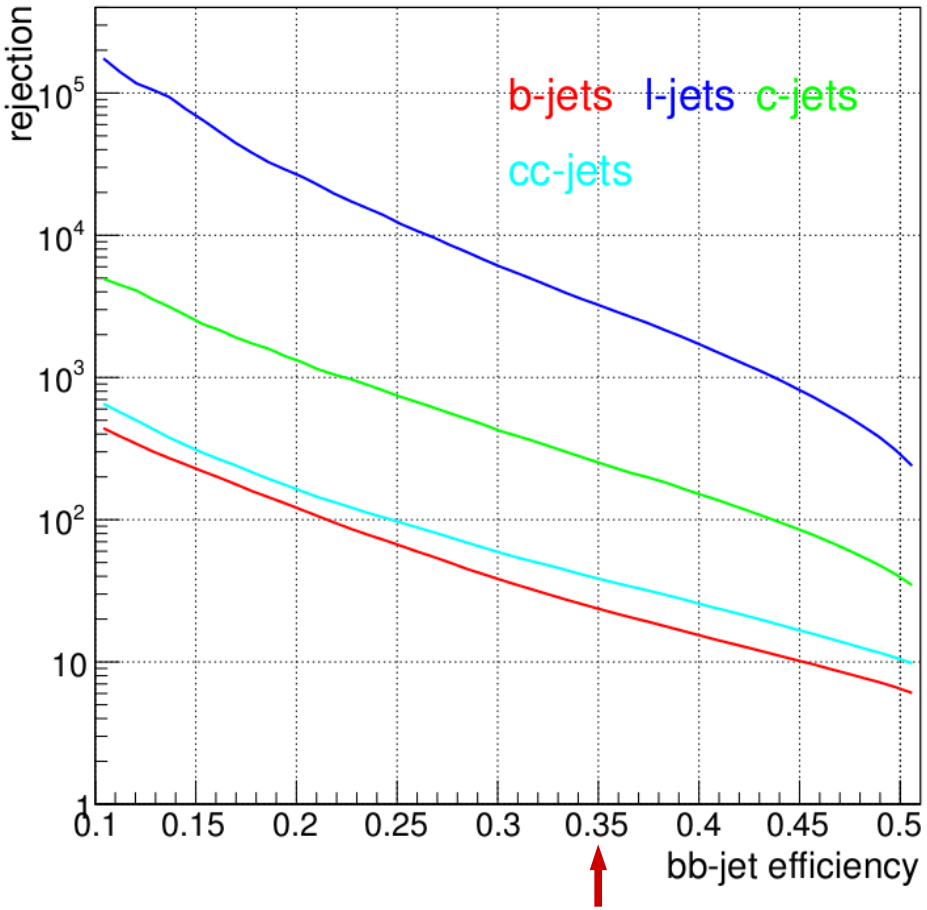
- ΔR between the vertex with maximum mass and jet axis



maximum mass vertex in b-jet and c-jets are close to the jet axis

Performance

- Rejection vs bb-jet efficiency to MultiSVbb2



Rejection: $R_b = 1/\varepsilon_b$

- Rejection at 35% of bb-jet efficiency:

	MV1	MultiSVbb1	MultiSVbb2
b-jets	3	18	23
c-jets	40	200	250
l-jets	10000	2400	3200
cc-jets	40	35	38

- ~7 times better b-jet rejection compare to MV1(default b-tagging algorithm)

Summary

- Successful migration of vertexing in b-tagging to xAOD format.
- Studied multi-vertexing algorithm(MSV): vertexing performance and properties.
- I developed two versions of double b-hadron tagger(MultiSVbb1 and MultiSVbb2) using multi-vertexing properties.
  - Much better b/bb separation than the default algorithm(MV1)
  - Included in the Atlas software to run2

Plans:

- **ttH(bb)** analysis,

  - start work in a common framework code.

  - use the tagger to constrain ttbb.