Tagging jets with double B-hadron using multiple secondary vertices

Royer Edson Ticse Torres

CPPM ATLAS Group

Journées Jeunes Chercheurs 2014
December 10, 2014

Higgs search & LHC

- **Higgs searches**

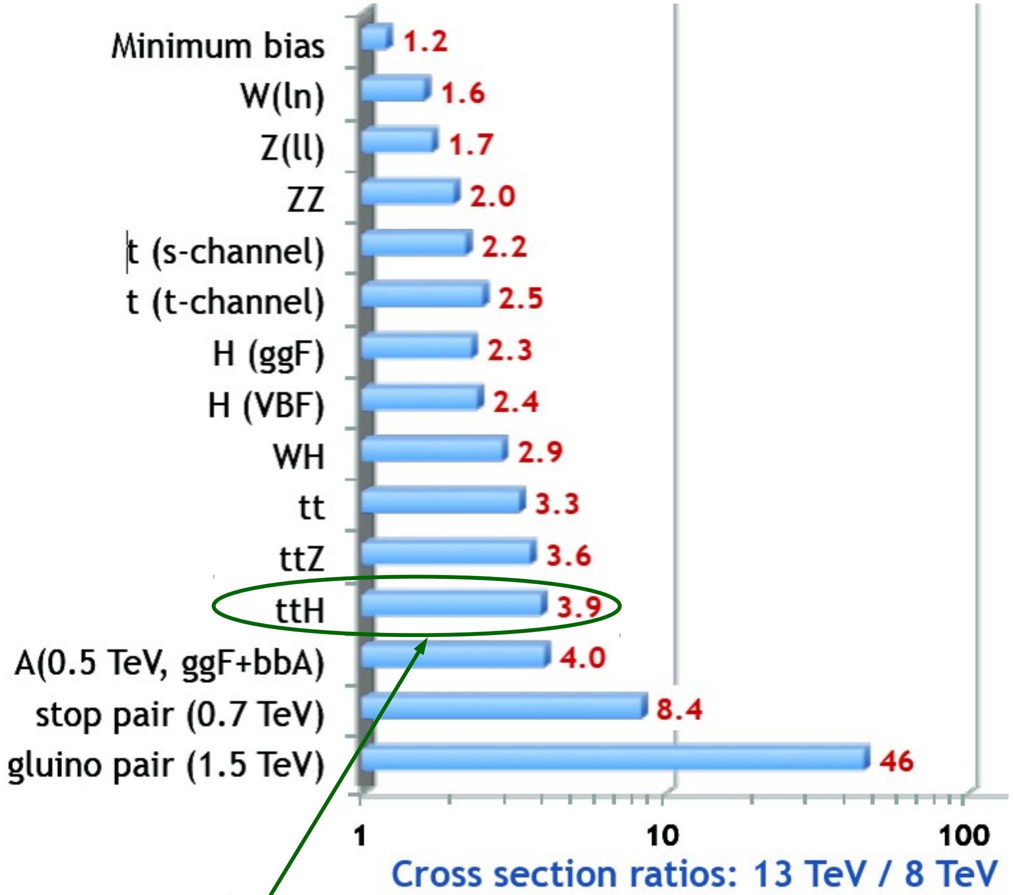
  - Observed in:
  H→γγ , H→ZZ→4l, H→WW→lνlν, H→ττ
  consistent with SM Higgs

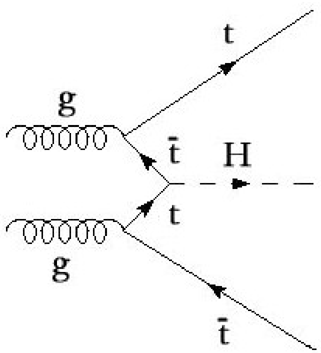
  - Precise measurement of mass, spin / CP, couplings is very important to investigate for possible deviations from the SM.

- **LHC Run2 (2015)**

  - Increase center of mass energy to 13TeV, high luminosity ~$10^{34}cm^{-2}s^{-1}$ and 25 ns bunch spacing

- **Associated prod. with tt**
  - It will be accessible for the first time after the LHC upgrade in 2015
  - ttH observation and Yukawa coupling measurement will be a highlight of Run2
  - Subject of my PhD thesis

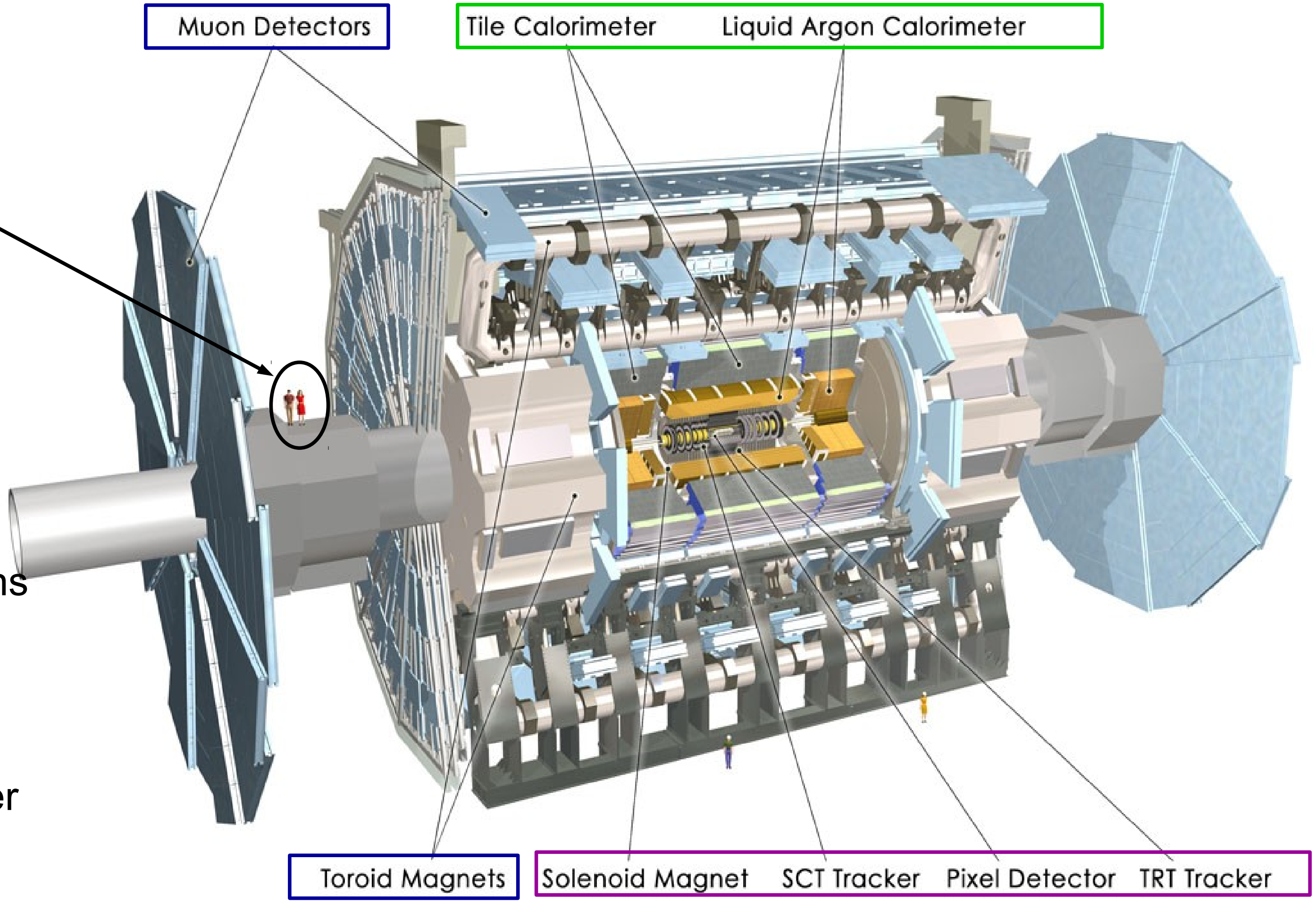




The ttH channel is the only way to directly constrain the top Yukawa coupling at the LHC

ATLAS detector

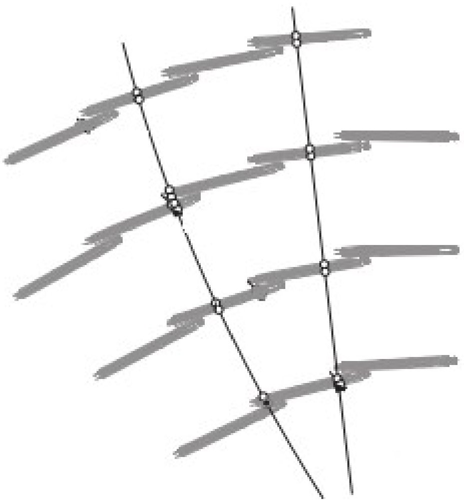
- **General-purpose detector**
  Cylinder around the beam pipe of
  - Height : 25 m
  - length : 46 m
- **Onion structure :**
  - **Tracker + Solenoid magnet**
  tracking and vertexing.
  - **Calorimeters**
  reconstruct electrons and jets
  - **Muon tracker + Toroidal Magnets**
  reconstruct momentum of muons
- **A new subdetector in Run2: Insertable B-Layer(IBL),**
  -Inserted between the existing pixel detector and a new(smaller radius) beam-pipe.
  .

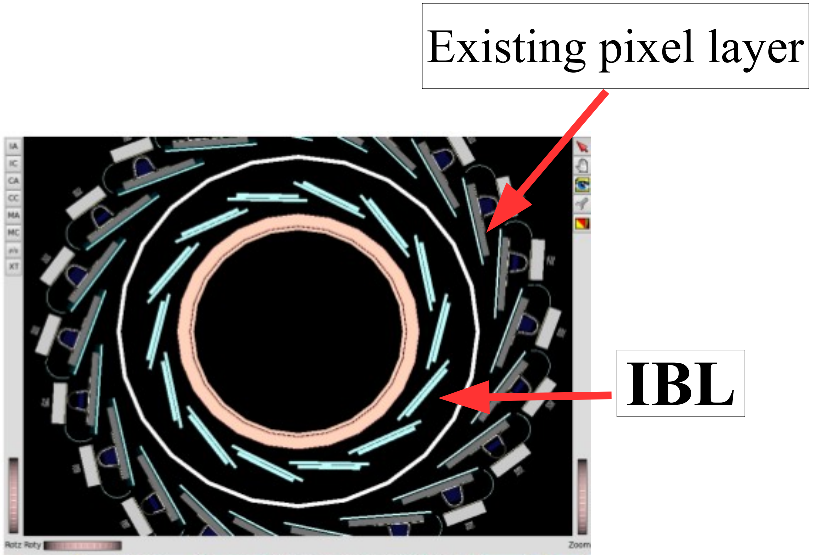


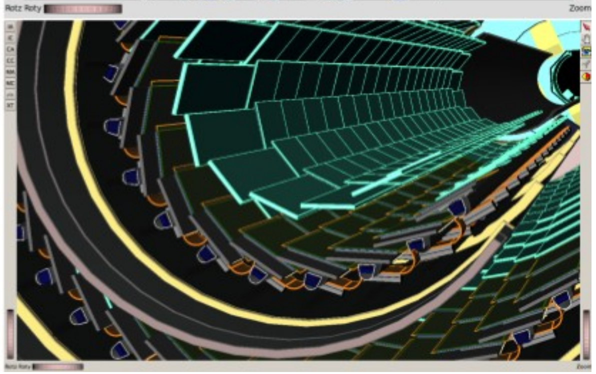
One of the most complex scientific instruments ever built !

IBL

- New pixel layer is inserted inside the existing pixel detector together with new beam pipe.
  -R_{in} = 31mm
  -R_{out} = 38.2mm
- Tracking and vertexing performance are expected to be greatly improved







-very significant improvements in b-tagging performance with IBL:

b tagger	Without IBL	With IBL	Ratio
IP3D	83 ± 1.5	147 ± 3.4	1.8
IP3D+SV1	339 ± 12	655 ± 32	1.9

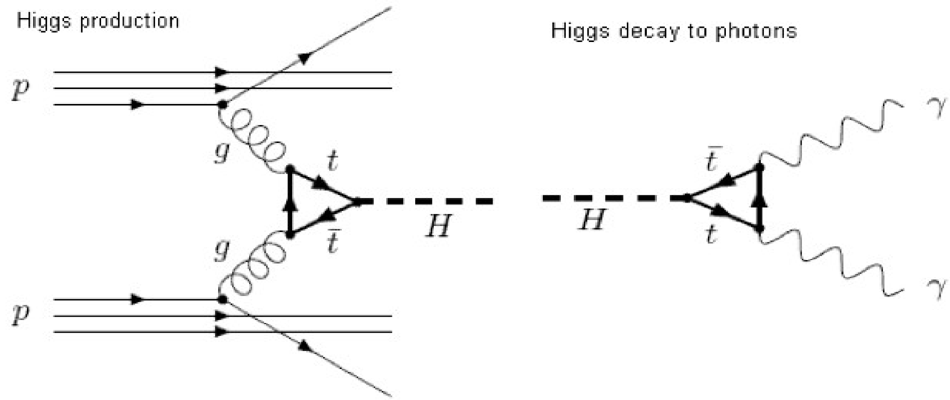
Rejection of light jets in tt events without pileup for b efficiency of 60%.

ttH motivation

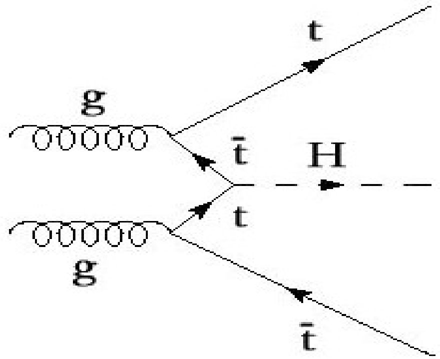
- **Top quark is the most strongly-coupled SM particle($Y_t \approx 1$)**

- **Indirect constrain on the top-Higgs Yukawa coupling Y_t**
  -Current measurements of Higgs boson productions via gluon fusion are consistent with SM within experimental uncertainties.

- **Direct measurement of Y_t in ttH production**
  - A measurement of the rate of ttH production provides a direct test of the coupling.
  - Observation of a significant deviation in the ttH production rate would be an indirect indication of unknown phenomena. e.g. several new physics scenarios predict the existence of heavy top-quark partners, that would decay into a top quark and a Higgs boson



Higgsstrahlung from top quarks



$\sigma(ttH) \propto g^2_{ttH}$

$Y_t = \sqrt{2}\, m_t/v$ $\quad g_{ttH} = m_t/v \quad$ *where* $v = (\sqrt{2}G_F)^{-1/2} \approx 246\,\mathrm{GeV}$

ttH production and decay

- **Production**
  - σ(ttH) increase with the center of mass energy
  - ≈2600 events in 20.3 fb-1 at 8TeV
- **Main background**
  - tt+X

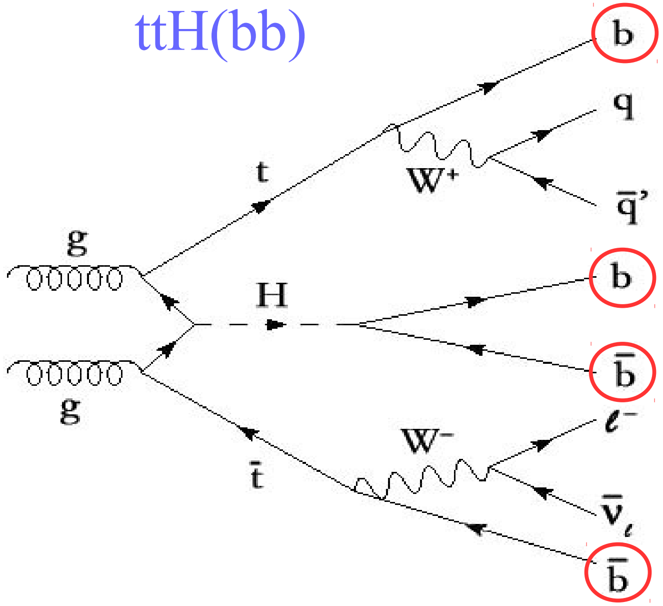
√s (TeV)	7	8	14
ttH(m_H=125GeV)(fb)	86	130	611
tt (pb)	177	253	950

Dominant background for ttH(bb)
2000 times larger than signal

Deacy mode	B
H→bb	0.58
H→γγ	2.3×10^{-3}
H→WW,ZZ,ττ	0.3

→ Dominant mode but large background
→ tiny but has clean resonant signature
→ multi-lepton final states

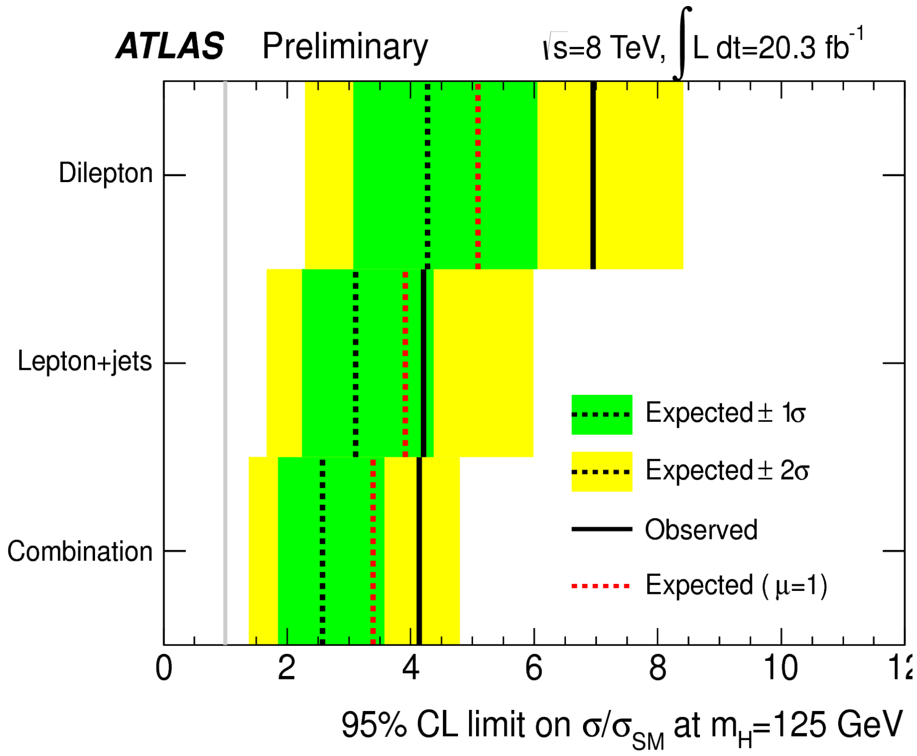
ttH(bb)

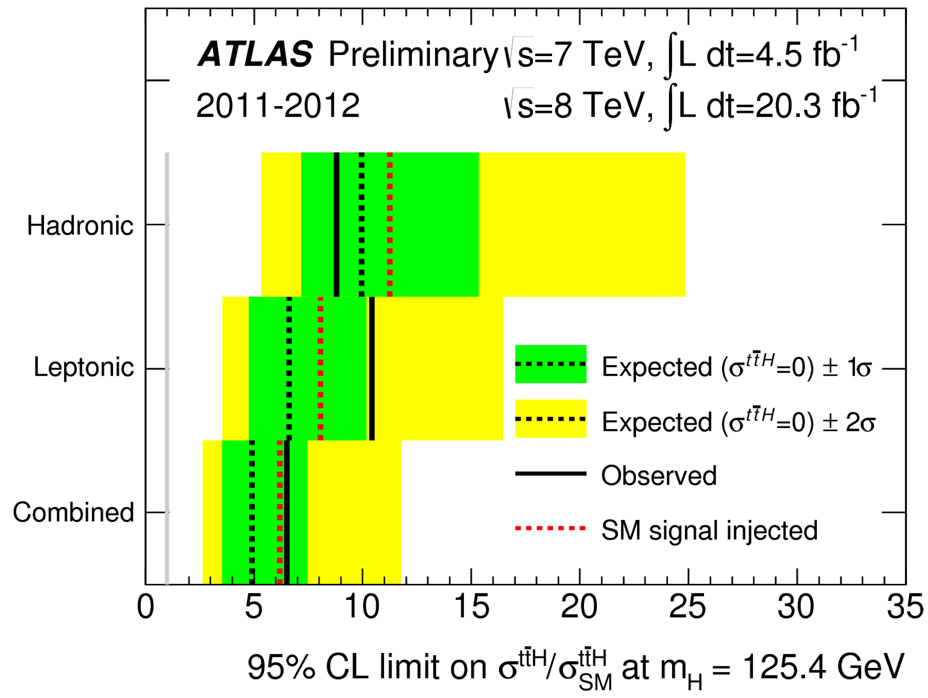


lepton+jets channel

ttH results Run 1

- Observed and expected 95% CL upper limits on the ttH production cross section times BR(H → bb) and BR(H → γγ)





no significant excess of events above the background expectation is found in Run1

ttH(bb) main background

- **Irreducible background:**
  - **tt+bb**, give the same final state signature as the signal
- **Reducible background:**
  - **tt+cc** and **tt+light** jets

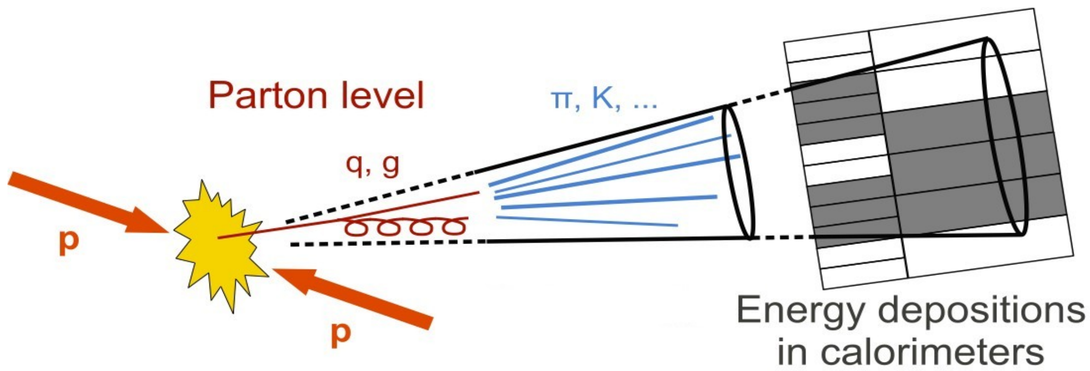
  b-tagging plays crucially important role!.

- Other backgrounds are:
  ttZ(Zbb)
  W/Z +jets
  Dibosons
  Singletop

- **Physics objects used in analysis**

  Jet: narrow cone of hadrons and other particles produced by the hadronization of a quark or gluon.

  b-tagging: Identify a jet as originated from a b-quark
  - One of the most powerful probe in energy frontier.
  Higgs physics, SUSY, top physics, etc.

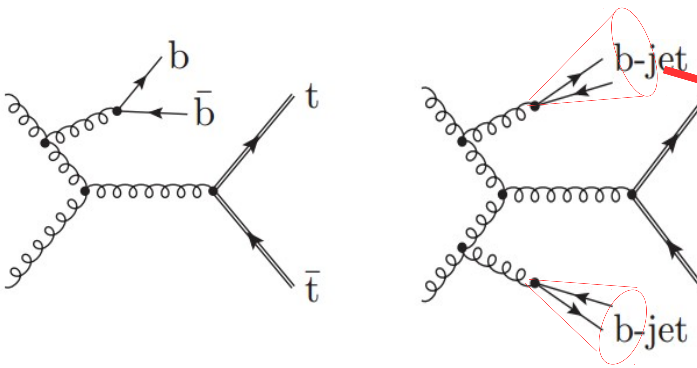


Jets are reconstructed from clusters built from energy deposits in the calorimeters, using the anti-kt algorithm with a radius parameter R = 0.4.

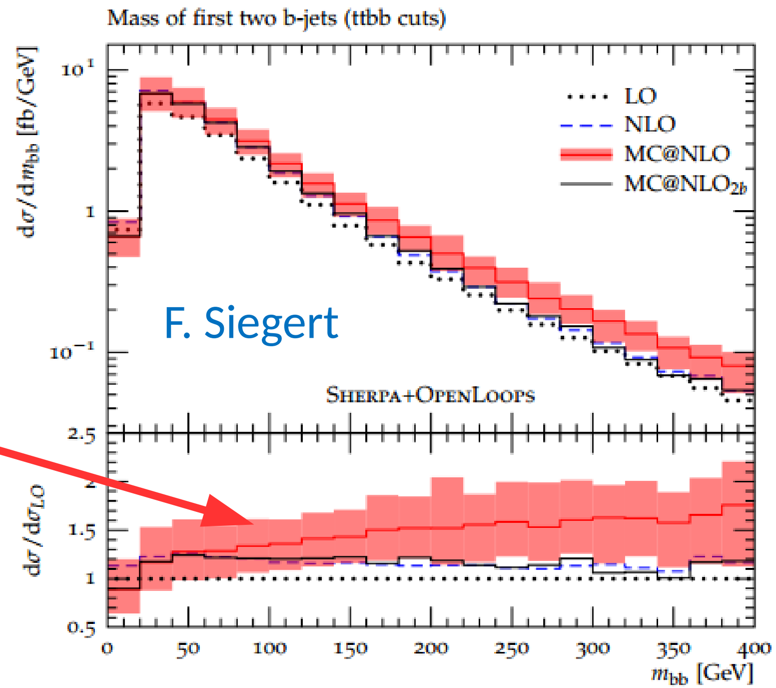
Tagging jets with double B-hadron

- **Motivation: ttH(bb)**

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



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



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

https://indico.cern.ch/event/346303/contribution/1/material/slides/0.pdf

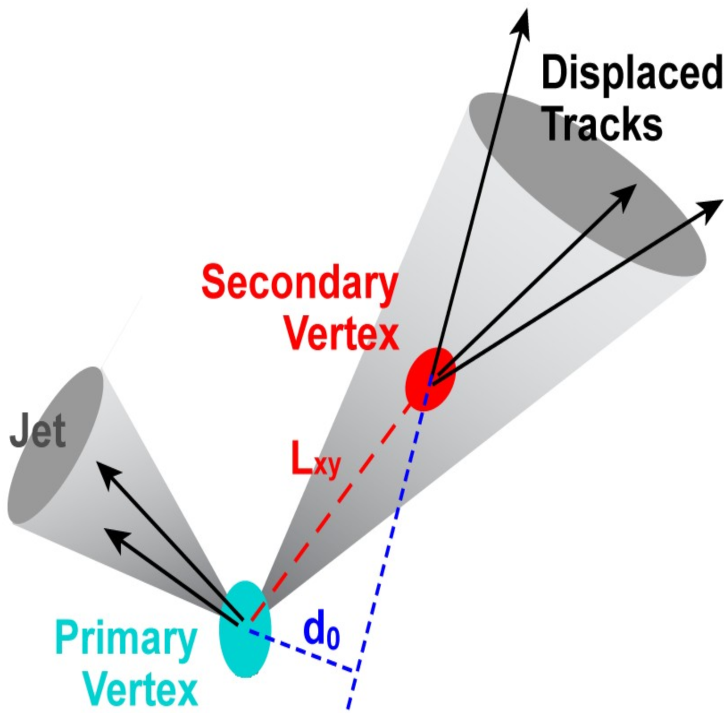
- Gluon splitting to bb not perfectly modeled by different MC
- Need a way to control in data

  - bb-jet tagger

- Close-by double b-hadrons production is the main background for boosted analysis

b-tagging

- Identification based on the b-jet features listed below:



- **b-jet**
  - B hadron decay lifetime: cτ~490μm
  - has secondary vertex(SV) of B hadron decay
  - has high track multiplicity at SV
  - has higher SV mass coming from B hadron mass, m_B~5GeV
  - contain tracks which are displaced from primary vertex. Tracks with large impact parameter(IP)

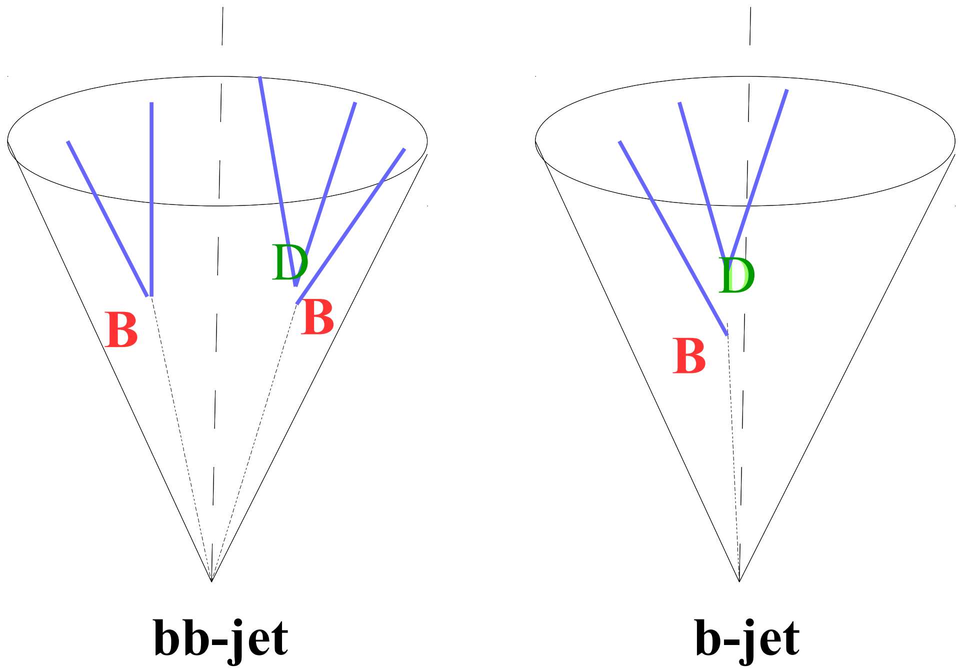
- **c-jet**
  - has very similar but weaker features compared to b-jet
  - has SV of C-hadron decay lifetime: cτ~310μm
  - has SV mass from C-hadron: m_C~2GeV

- **light-jet**
  jet originates from light quarks(u, d, s) and gluons.
  - absence of SV or fakes vertices with lower mass

Tagging jets with double B-hadron

Goal: Identify jets containing two B-hadrons

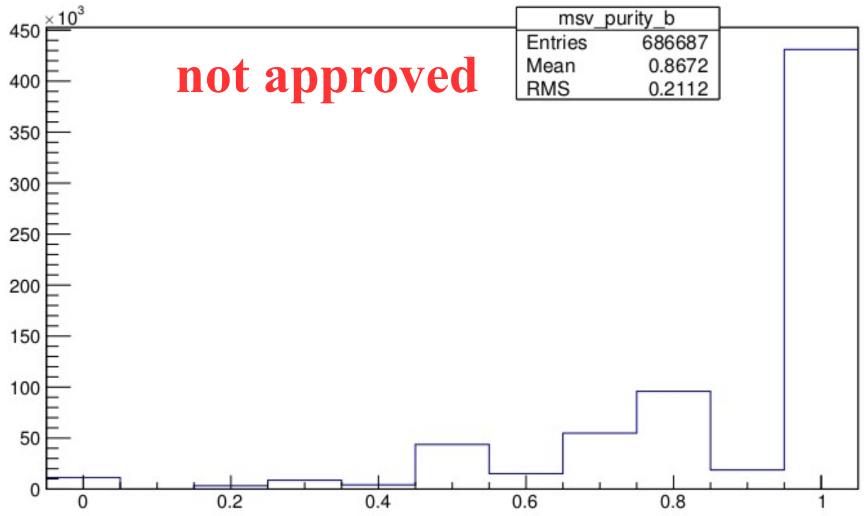
- Current b-tagging algorithms do not distinguish between b-jets and bb-jets.
- We use multi secondary vertex finder(**MSV**) to reconstruct multiple vertices within anti-kt R=0.4 jets, with pT>20GeV and |η|<2.5
- **MSV** algorithm is based on Kalman filtering method for vertex position estimation inside the jet.



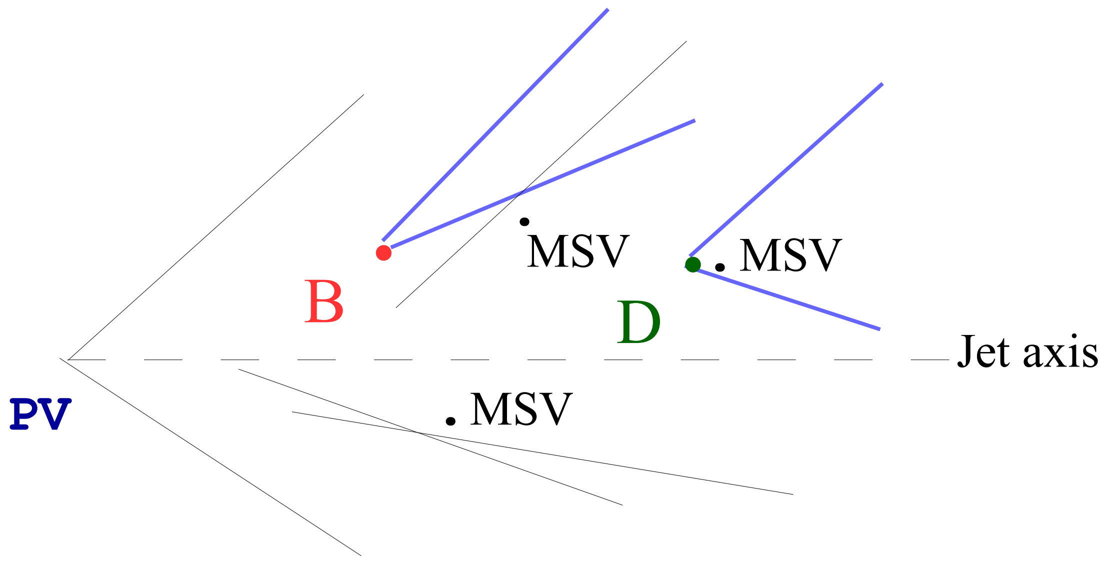
MSV purity

- 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

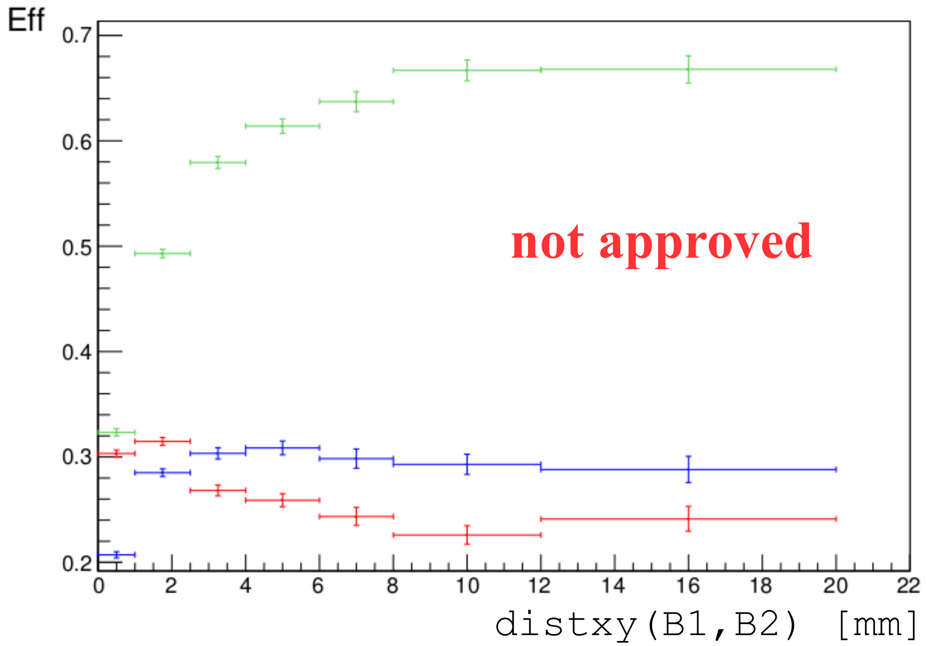


- Several studies have been done to understand vertexing performance and ambiguous cases(e.g. B/C separation, fakes vertices) in **MSV**.

Multi-vertexing in bb-jets

- Fraction of bb-jets with:

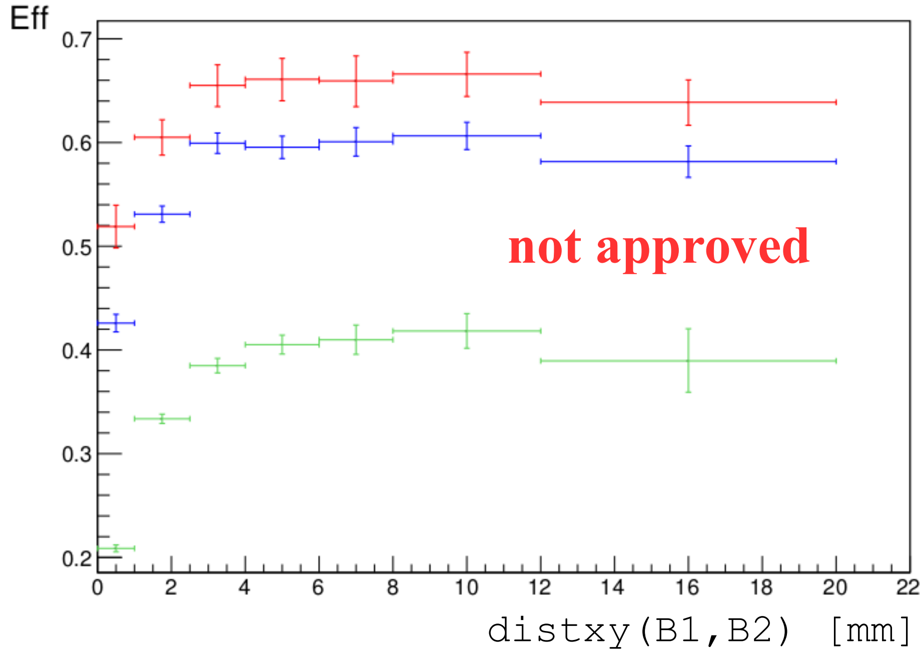
  1 reco vertex, **exactly 2 reco vertices**, **at least 2 reco vertices**



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

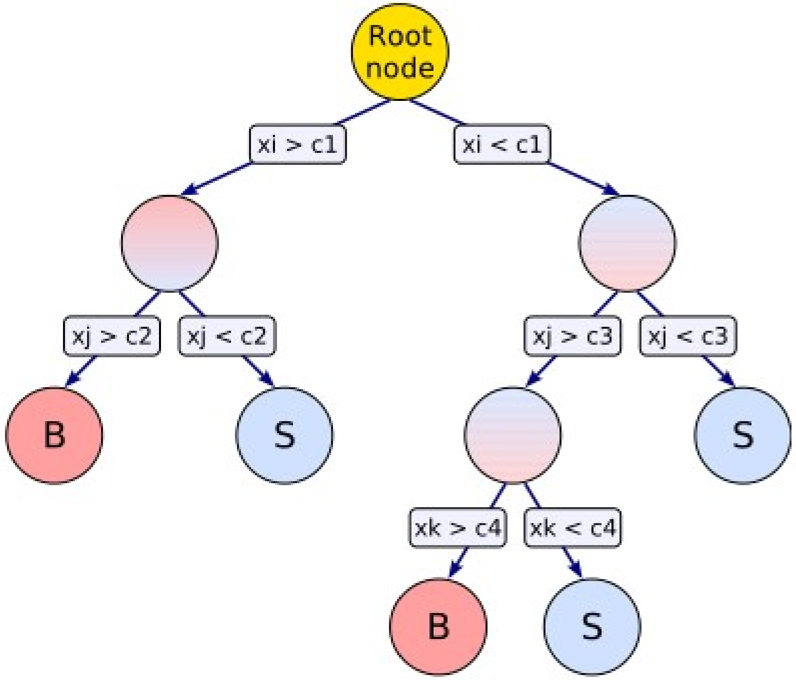


efficiency increase with the pT

Multivariate Analysis with Boosted Decision Trees

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

- The strategy to identify double b hadrons in jet is to exploit properties of multiple vertices inside the jet using a multivariate analysis.
- Boosted Decision Trees exploits the differences between jets containing two b hadrons and the single b jets
- Increase the discrimination power between jet with two b hadrons from single b jets, c jets and light jets.

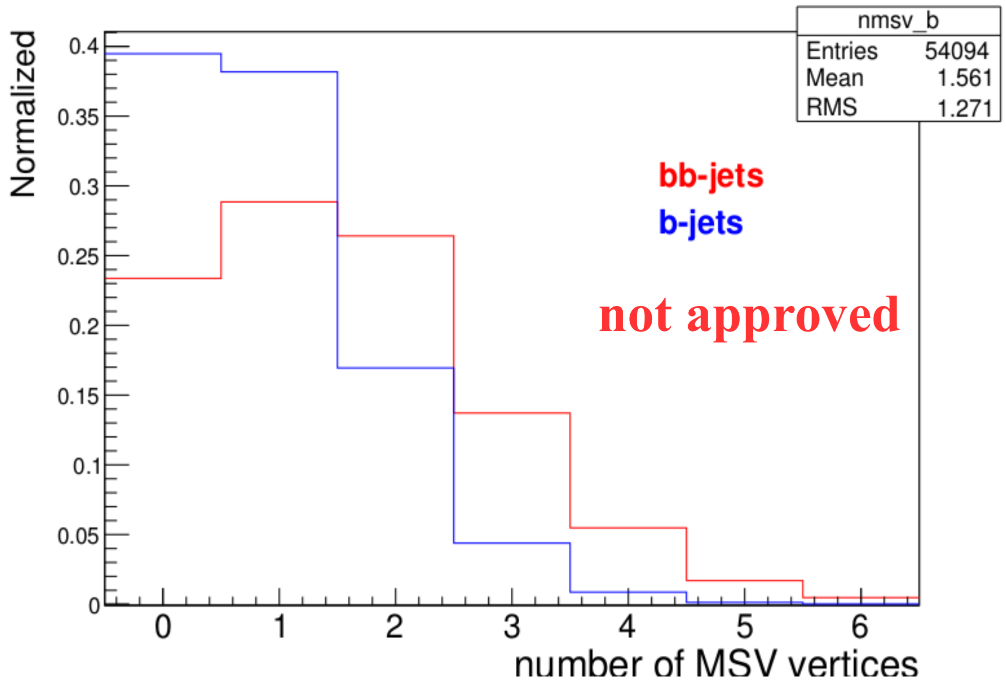


Training

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

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

  MultiSVbb2 (14 variables)
  - Include additional topological variables

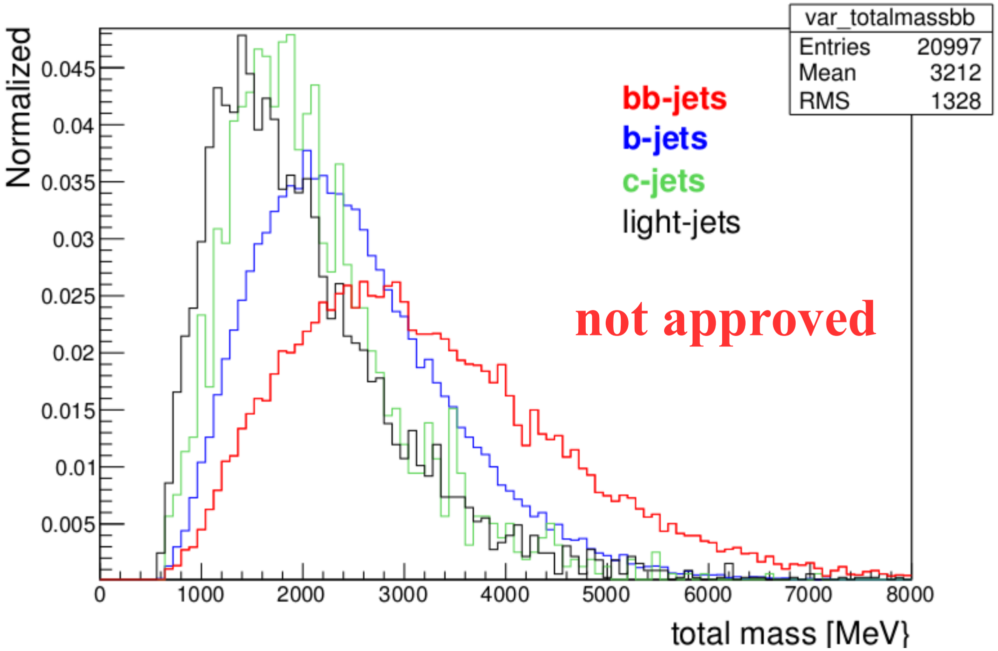


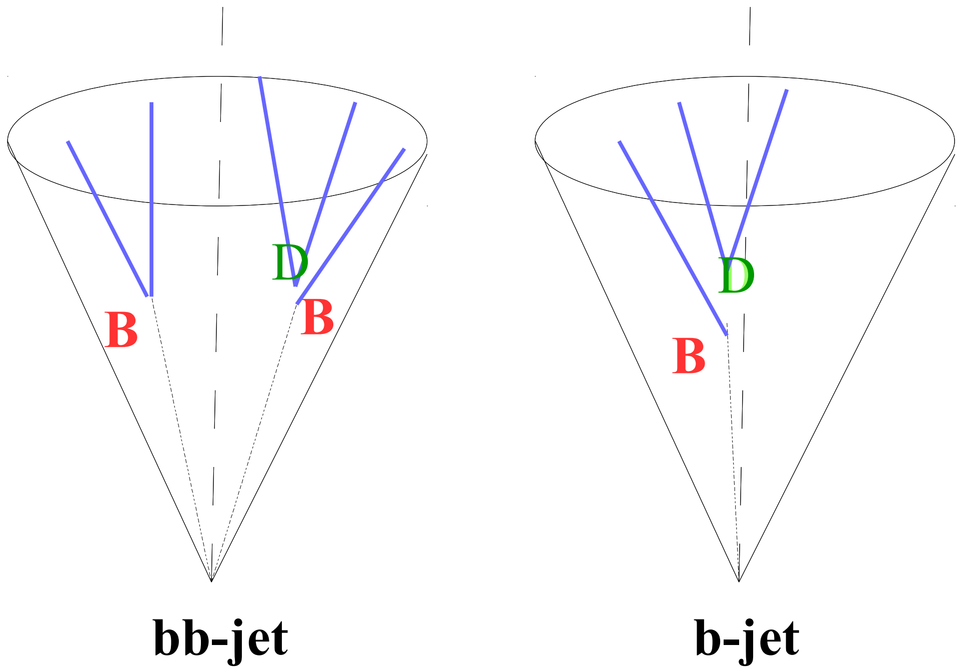
Require at least 2 MSV vertices

Discriminating variables:

Input variables to train the BDT:

- Total mass of vertices



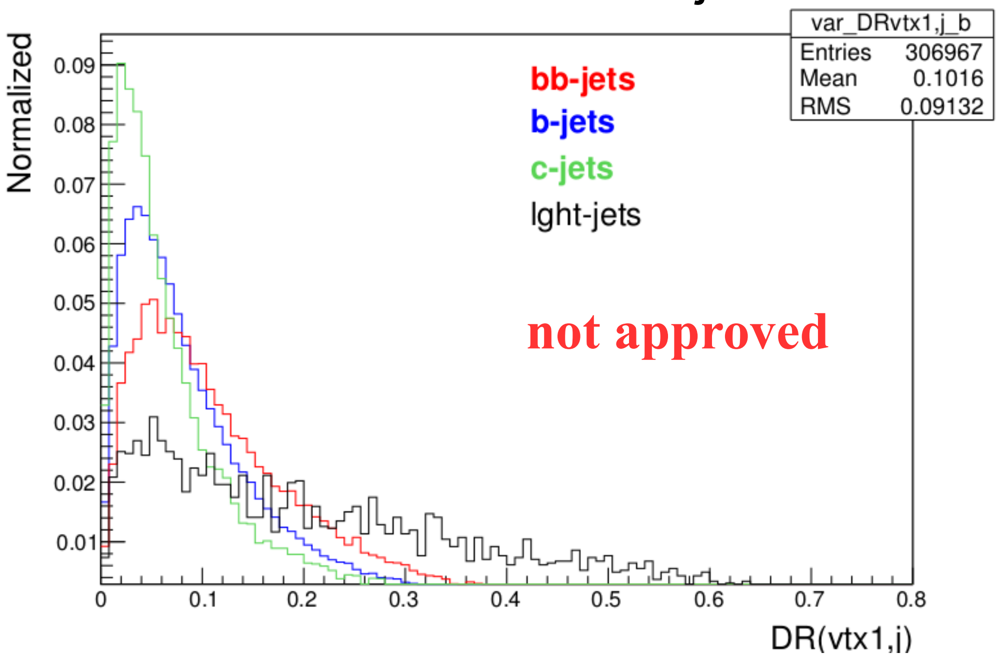


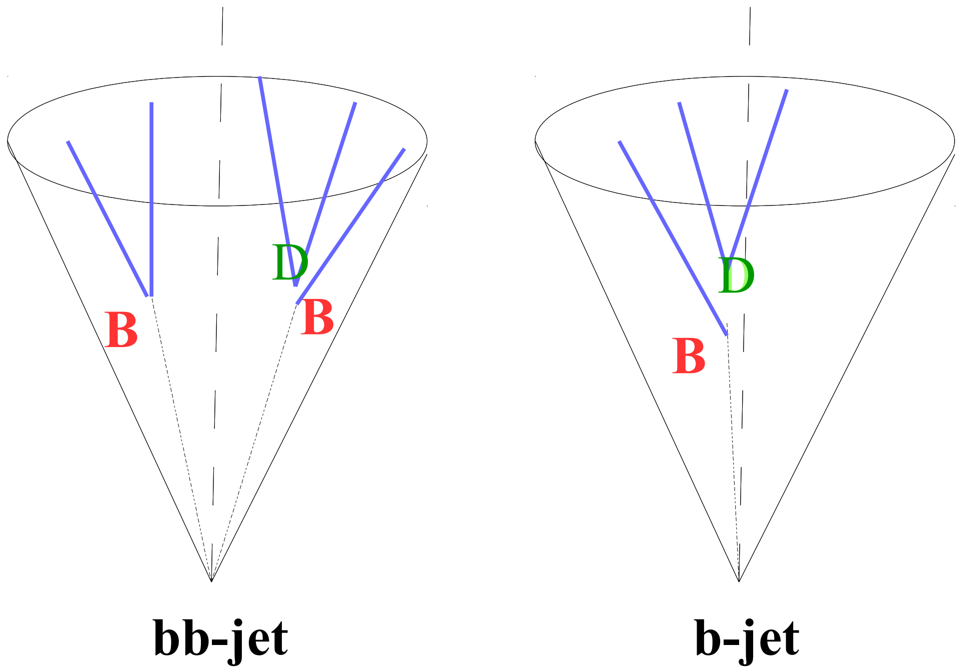
total mass for bb-jets is greater than other flavours

Discriminating variables:

Input variables to train the BDT:

- ΔR between the vertex with maximum mass and jet axis

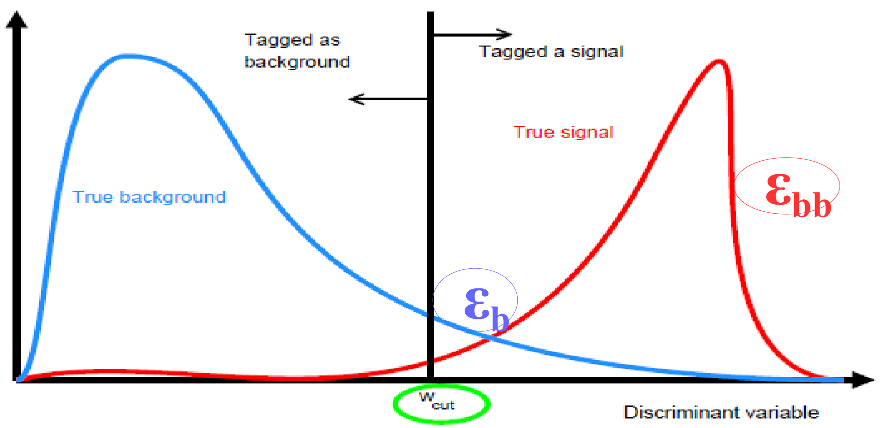




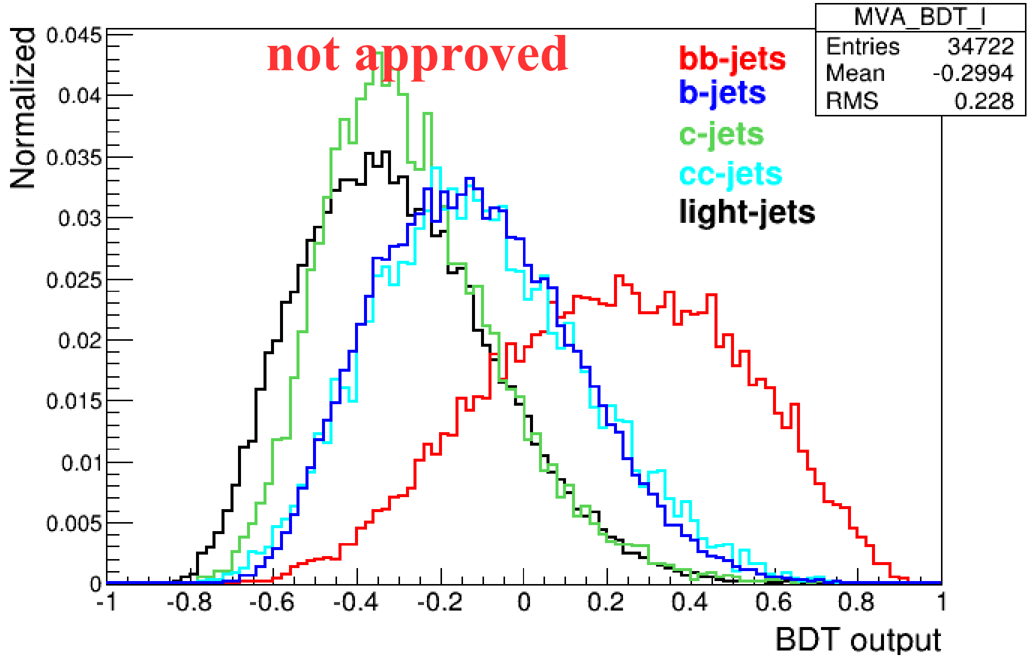
maximum mass vertex in b-jet and c-jets are close to the jet axis

BDT output

- Typically we chose a weight cut and calculate efficiency and rejection for the algorithm
- Performing cut (w_{cut}):
  - bb-jet efficiency ε_{bb}
  - b-jet mis-rate ε_b or rejection $R_b = 1/\varepsilon_b$
  - p_T and η dependence

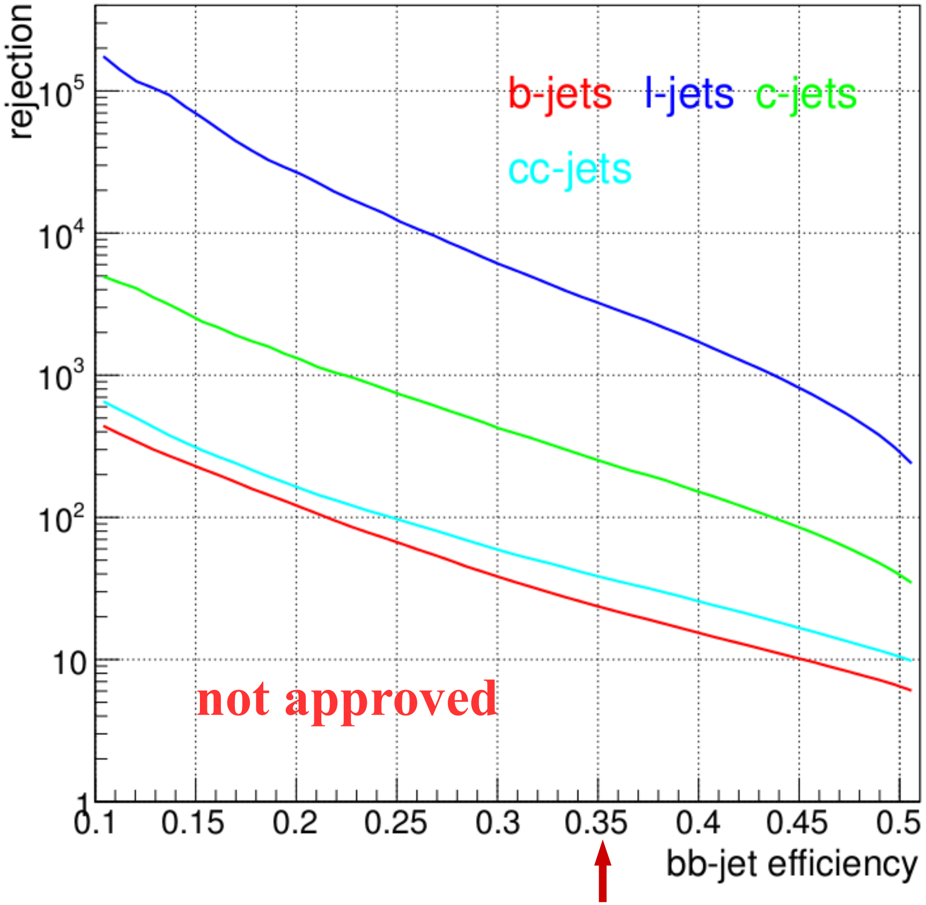


Distribution of the BDT output



Performance

- Rejection vs bb-jet efficiency to MultiSVbb2



Rejection: $R_b = 1/\varepsilon_b$

- Typically the working point in b-tagging is 70%, by the requirement of two b (ε_{bb}~ε_bxε_b) we can use 35% bb-jet efficiency.

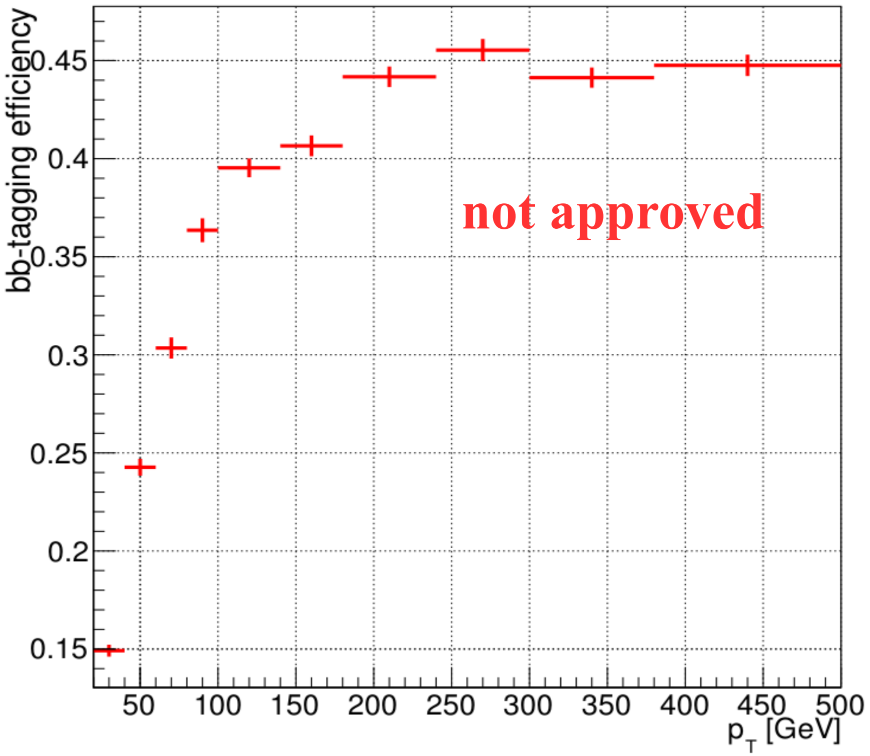
- Rejection at 35% of bb-jet efficiency:

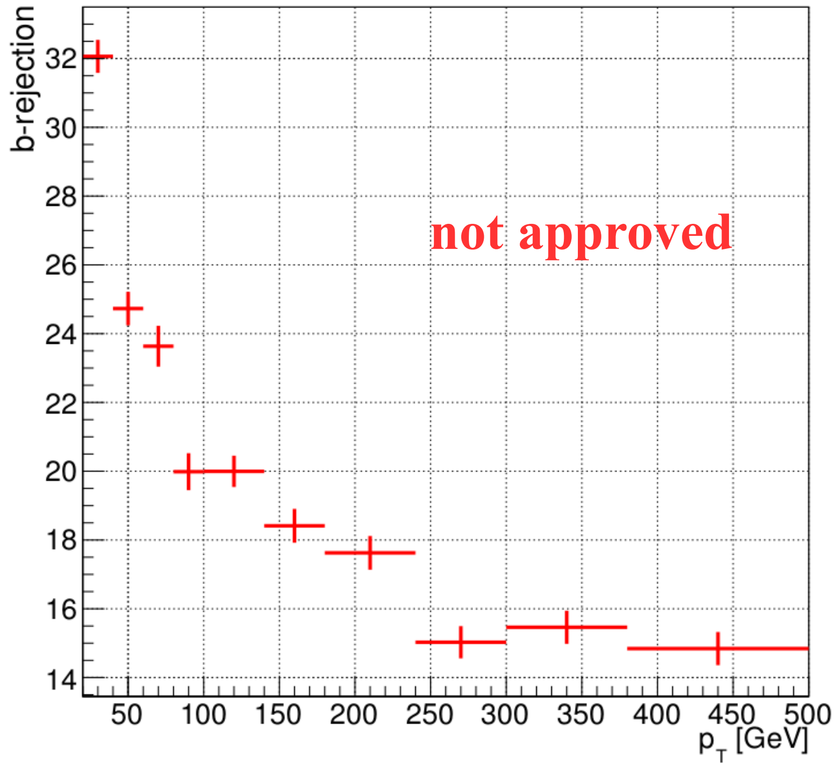
	MV1	MultiSVbb1	MultiSVbb2
b-jets	3	18	23
c-jets	40	200	250
l-jets	10000	2400	3200
cc-jets	40	35	38

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

Performance

- bb-tagger performance with the globally fixed efficiency at 35%





The efficiency increase with pt, in high pt we have more good tracks consequently better secondary vertex reconstruction, and the b-rejection is the opposite, it fall downs as pt increase

Summary

- 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
- Improvements in performance are expected with the IBL. We need calibrate the taggers with data

Plans:

- **ttH(bb)** analysis,

  - use the tagger to constrain ttbb.