New b-tagging algorithms for $H \rightarrow b\bar{b}$ searches at ATLAS

2013 Dec 16th Mon.
CPPM student seminar

Kazuya Mochizuki

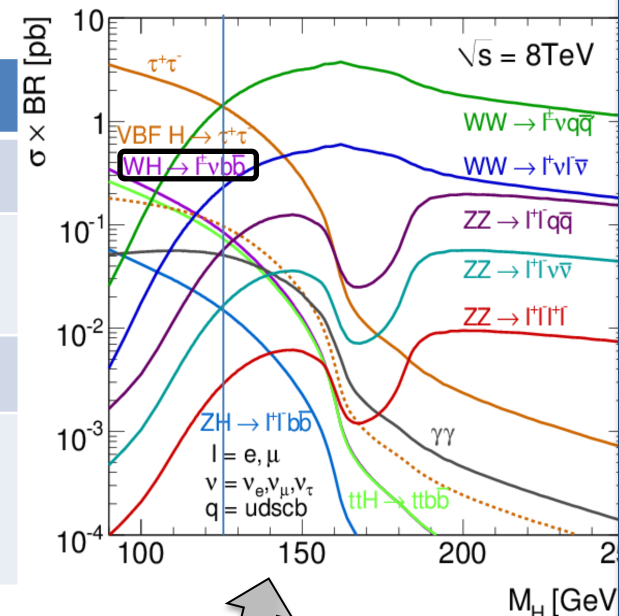
Supervisor: Laurent Vacavant

Co-supervisor: Yann Coadou

CPPM ATLAS Group

Introduction

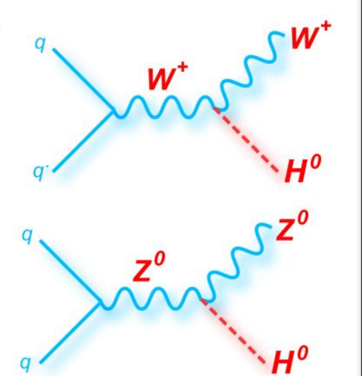
LHC passage with Higgs boson	
2012 Jul	Observation of a new Higgs-like particle at ~125 GeV
2013 Mar	evidence for J=0 with positive parity strongly preferred However fermion coupling was not yet confirmed
2013 Oct	Nobel Prize in Physics to Higgs & Englert
2013 Nov	ATLAS observed an evidence of $H \to \tau\tau$ with 4.1(3.1 expected) standard deviation First direct evidence of the Higgs coupling to the fermions!



Does the Higgs mechanism give mass to also the quarks?

- Search for $H \to q\bar{q}$
- b-quark has the largest branching ratio: 58% for $m_H = 125$ GeV
- Search for $H \to b\bar{b}$?
  - ⇒ Overwhelming QCD background ($pp \to b\bar{b}$)
  - ⇒ Looking for associated production with some probes (e.g. WH/ZH/ttH)
- B-tagging plays crucially important role

  Today's main topic



LHC & ATLAS

- **Physics Goals**

  Higgs boson, New Physics(SuSy, Extra dimension, etc.)
- Proton-proton collision
- $\sqrt{s} = 8(7)$ TeV in 2012(2011)
- $\int L\,dt \sim 20(5)$ fb^{-1}

Inner detector (tracker)

The most important detector for b-tagging because we need accurate tracking and vertexing.
$\sigma(d_0) \sim 35\ \mu$m for typical track with $p_T = 5$ GeV

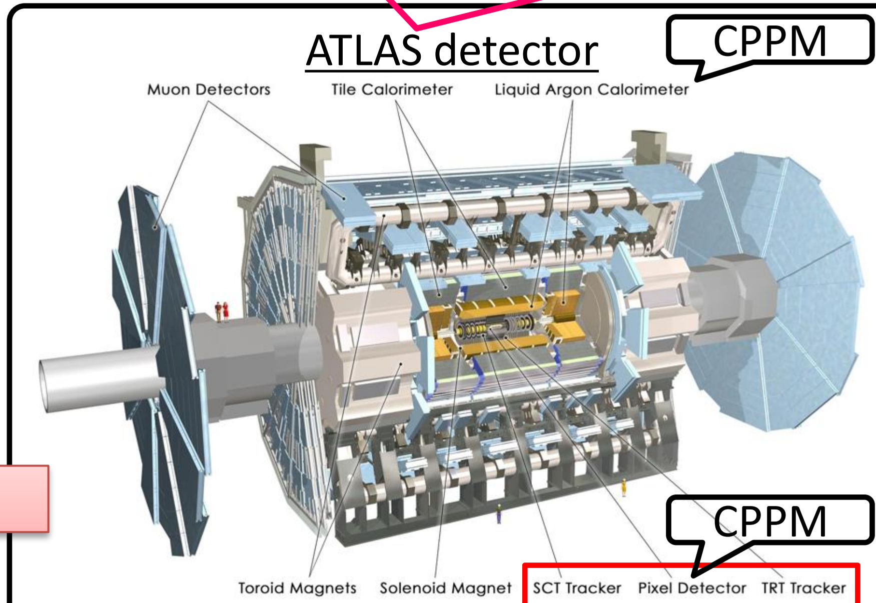
Calorimeter

Reconstruct electrons and jets. It is composed of electromagnetic and hadron calorimeters.

Muon spectrometer

Reconstruct momenta of muons which goes through detectors inside.

ATLAS detector

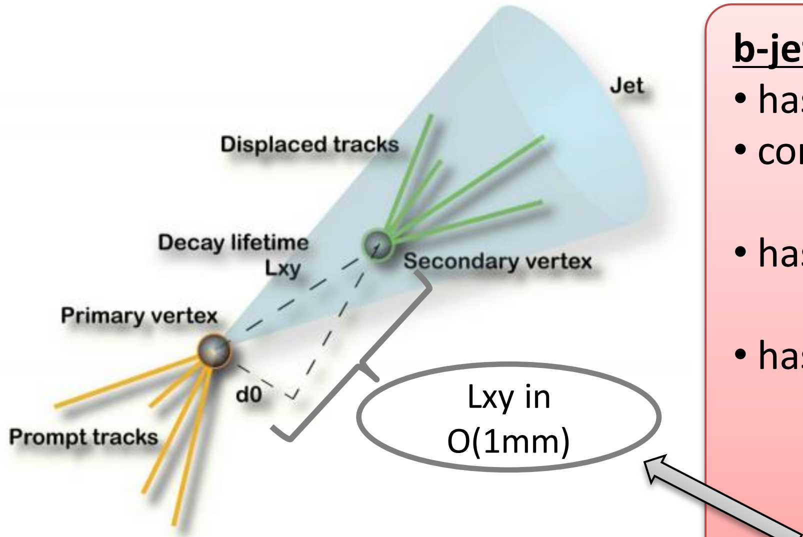


Reconstruction

4 momenta of physics objects
- Electrons
- Muons
- Tau jets
- Jets
- Transverse missing energy (MET)

b-tagging

- Goal: identify a b-jet from light jets firstly and possibly from c-jets
- One of the most powerful probe in energy frontier
  Higgs physics, SuSy, top physics, etc.
- Identification based on the b-jet features listed below.



b-jet
- has high track multiplicity
- contains tracks which are displaced from primary vertex
  Tracks with large impact parameter(IP)
- has higher jet invariant mass coming from B hadron mass
  $m_B \sim 5$ GeV
- has secondary vertex (SV) of B hadron decay
  - SV mass
  - track multiplicity at SV
  - direction
  - lifetime: $c\tau \sim 490\ \mu$m

Light jet
jet originates from light quarks (u, d, s) and gluons
- Lower jet mass
- Absence of SV

c-jet
- has very similar but weaker features compared to b-jet
- has jet mass of C hadron: $m_C \sim 2$ GeV
- has SV of C hadron decay
  - Lifetime: $c\tau \sim 310\ \mu$m

B-tagging at ATLAS

IP3D CPPM
Track based algorithm. Using 2D histogram of transverse & longitudinal IP of the tracks in order to use their correlation

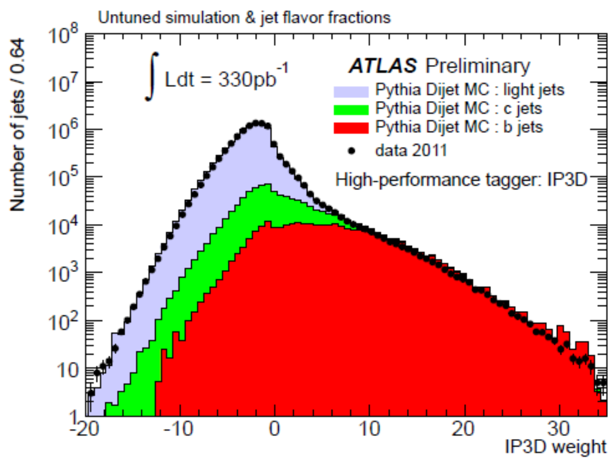
SV1 CPPM
Secondary vertex fitting algorithm. Using SV information of mass, direction, track multiplicity, etc.

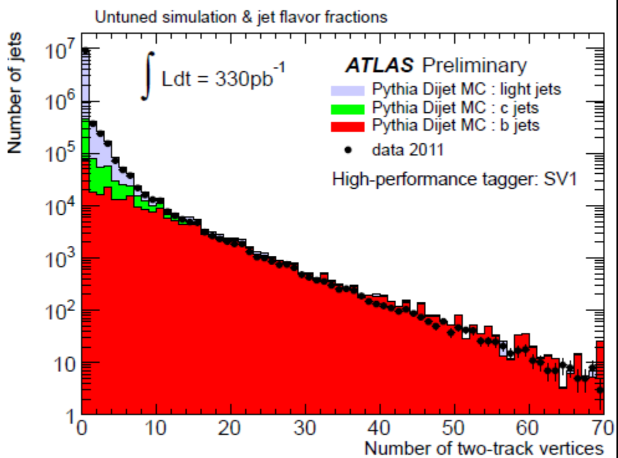
JetFitter
Fit vertices along b flight direction which is assumed same as jet axis. Using SV informations.

MV1 CPPM, L.Vacavant
- Neural network combination of 3 taggers' information
- Best performing b-tagging algorithm in ATLAS
- Recommended tagger in the Run 1 analyses

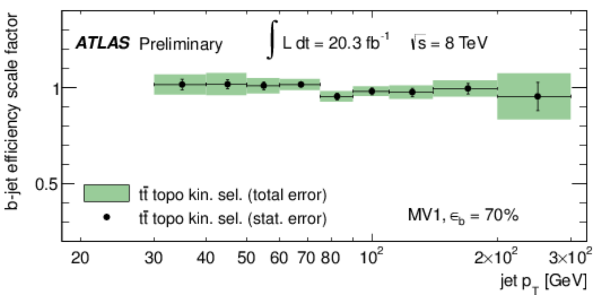
MV1c CPPM, M.Ughetto
- A successor of MV1 but trained against also c-jets
- Therefore it has better c-jet rejection than MV1
  With the modest cost in light jet rejection
- Now used in H->bb analyses

Commissioning of the ATLAS high-performance *b*-tagging algorithms (CPPM, N.Bousson)





b-tag efficiency data-to-simulation scale factor for MV1 using the tag & probe method on ttbar simulation sample.



MV2 blueprint

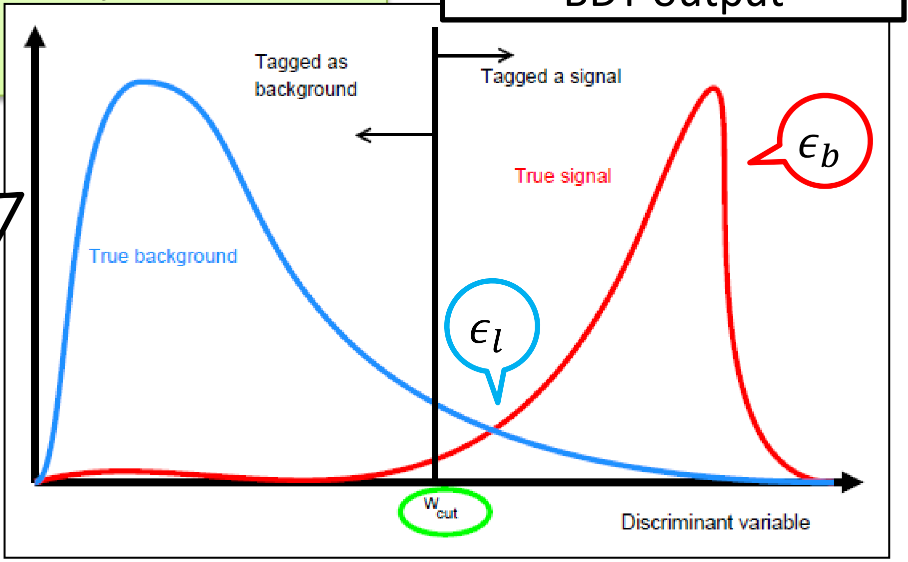
MV2 prototype
- BDT based
  - Using 25 input variables from basic taggers
- Pt dependence
  - B-tagging strategy in low and high pt range should be very different
  - => Binned multi BDTs in pt ranges
- Flavor discrimination
  - Effort to achieve 2D cuts, 3BDT discrimination of b/u, b/c, c/u

Results were interesting but incompatible with limited computing resources. It used too many BDTs.

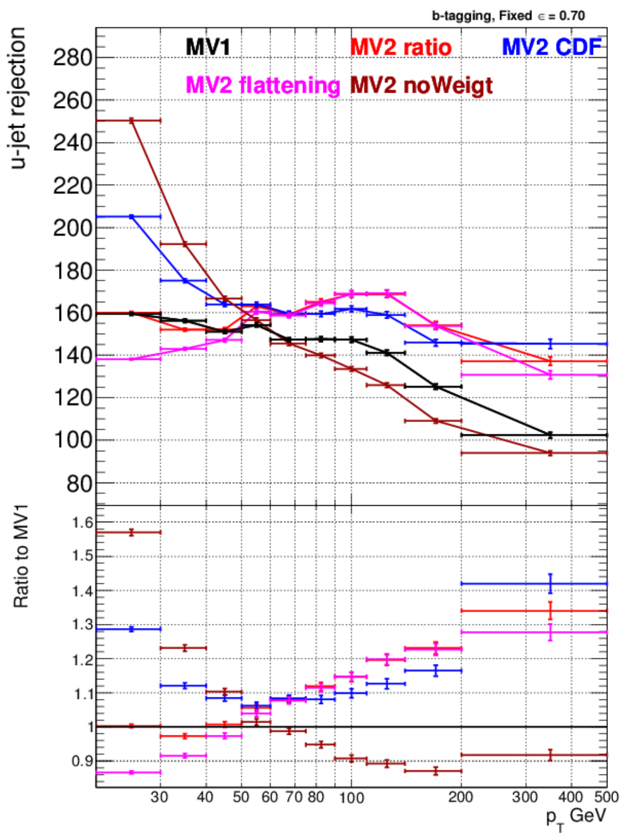
MV2
- BDT based, same input variables to prototype, using only one BDT
- Pt dependence
  - Can we achieve desired pt dependence of the tagger's output?
- Flavor discrimination
  - Instead of 3 flavor discrimination, aiming further rejection of c-jets
  - Using the same training scheme to MV1c

When performing cut, we have to care about
- b-jet efficiency ϵ_b
- Light jet mis-tag rate ϵ_l or rejection $R_l = 1/\epsilon_l$
- Correlation to other variables
  - **p_T dependence**
  - η dependence

BDT output



Result I: Pt dependence of output



B-tagging performance with the globally fixed efficiency @70% which is typically used in the analyses.

Pt dependence

To avoid undesired pt dependence, several weighting ways are tested.

- The problem of no weighting is l-rejection falls down as pt increases.

  BDT is trained with pt difference of b and light jets.

  Focusing at only low pt

- Flattening/ratio weightings focus at nowhere

  they treat high and low pt equally

  But the fact is that we have light rich region in low pt and we can't ignore it.

- CDF* weighting takes advantages from both of them

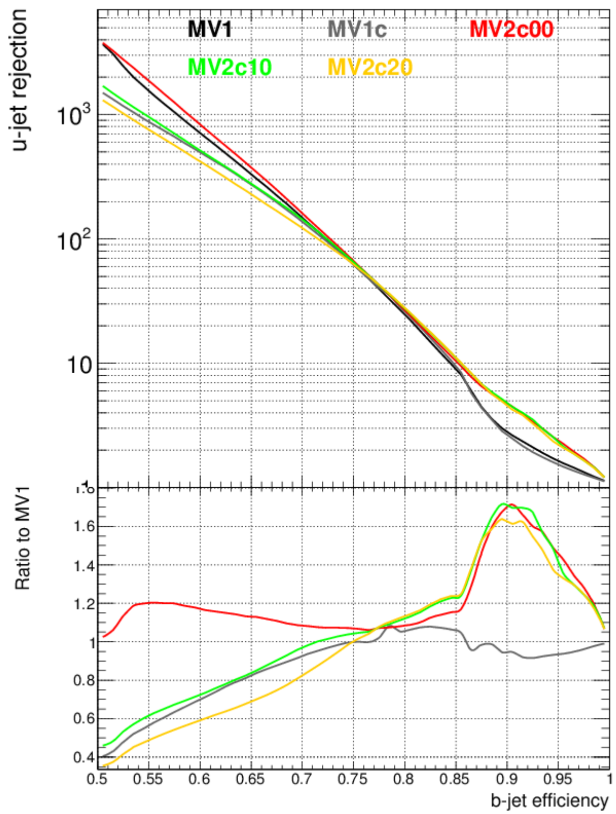
  build CDF in (pt, eta) 2D histograms for each flavor

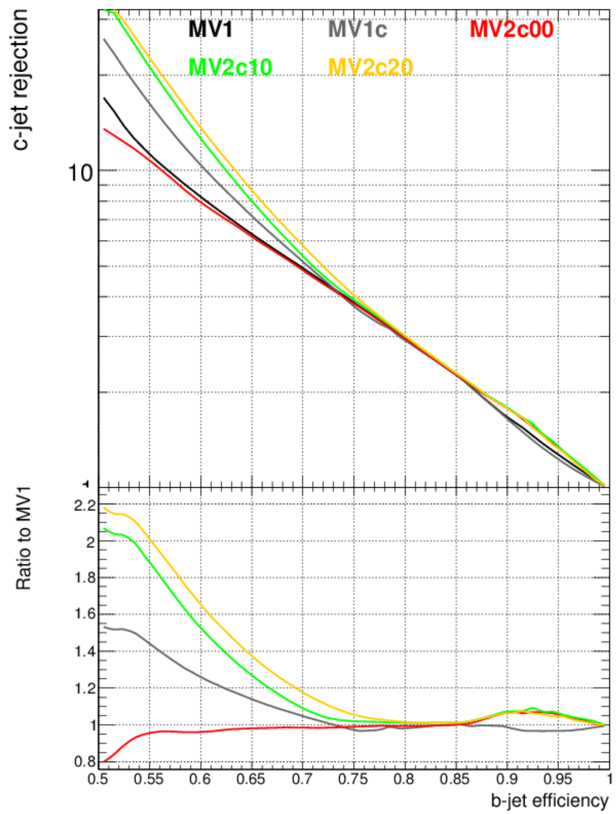
  apply inverted CDF value

  in each flavor depending on jet's (pt, eta).

  * Cumulative distribution function.

Result II: Trade-off between charm and light jet rejection





C-rejection

For further c-jet rejection, use training a la MV1c. Several c-fractions are tested.

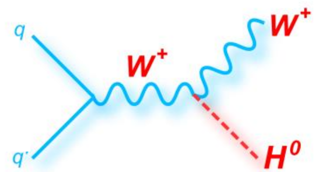
- MV2c: a charm tuned MV2 which is trained for b against light and c-jets just like MV1c but controlling the fraction of c-jets added in the training.
- Following number shows the c-fraction wrt number of b jets in % in training sample. Number of b jets and light jets are always kept at 1M.
  e.g. MV2c20 means training with b:(u+c) = 1M:(1M+200k)
- Gain of c-jet rejection up to factor 2 to MV1!!

$WH \rightarrow \ell\nu b\bar{b}$ analysis

Current $WH \rightarrow \ell\nu b\bar{b}$ analysis in ATLAS ($\ell = e, \boldsymbol{\mu}$)

- Exactly 1 isolated lepton with $p_T > 25$ GeV
- 5 bins in p_T^W

  0-90, 90-120, 120-160, 160-200, 200- (GeV)
  In each bin, background contribution changes
  => different selection strategy
- Using MV1c for b-tagging



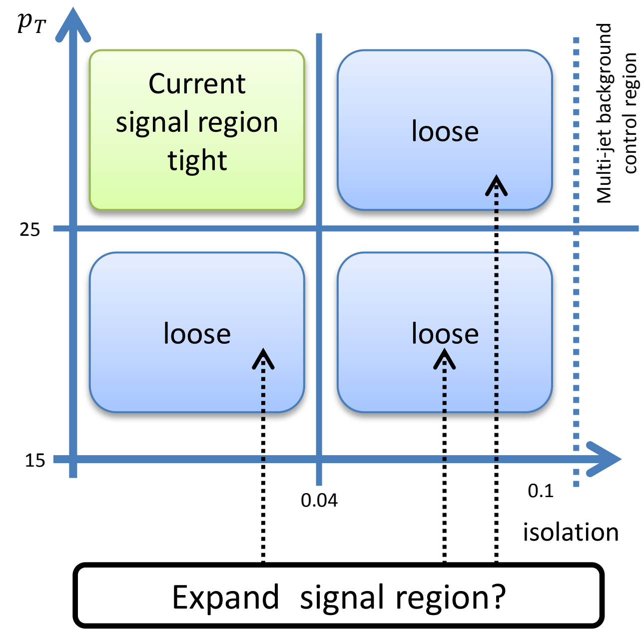
My work: Expansion of the signal region using

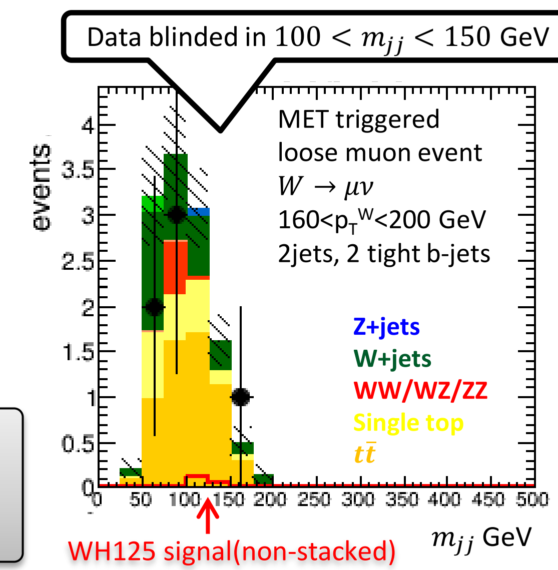
- Looser leptons

  lepton is categorized in tight and loose depending on it's p_T and isolation requirement. Currently loose lepton is rejected in the analysis. However there is signal acceptance gain there.

- MET trigger

  The main trigger for this analysis is single lepton trigger. However for muon channel, MET trigger is also helpful because of the reduced muon chamber coverage and less efficiency for high p_T muons. Furthermore, we still have some room to expand MET trigger applying region for further signal acceptance gain.





Preliminary study shows <u>a few percent gain</u> of $S/\sqrt{B}$ significance. No golden channel exists in VH analyses and the statistics is limited. Therefore even with a few percent, improvement from each contribution is really important for combined result!

Summary

Conclusion

1. New b-tagging algorithms using BDT are successfully developed.
   - better c-jet rejection up to 25% compared to MV1c
     while keeping the same level of light jet rejection.
   - MV2c00/MVc10/MV2c20 under validation of official sample production.
     MV2 is likely to replace MV1/MV1c for Run 2 analyses.

– Effort for $WH \rightarrow \ell\nu b\bar{b}$ analysis

Test study shows there is significance gain of a few percent level
by signal region expansion and this few percent is important.

Plan

- Keep working in WH analysis aiming at Moriond 2014
  $\Rightarrow$ will become one of my thesis subjects
- ttH simulation study for Run 2 analysis preparation.

Merci ☺


BACKUP