Search for the Higgs boson in the $t\bar{t}H(H \to b\bar{b})$ channel and the identification of jets containing two B hadrons with the ATLAS experiment

Royer Edson Ticse Torres

PhD defense

September 29, 2016

Overview

Theoretical introduction

The ATLAS experiment

My work:

- **Part 1: Identification of jets containing two B hadrons**
- **Part 2: Search for the Higgs boson in the single lepton $t\bar{t}H(H \rightarrow b\bar{b})$ channel**

Conclusion

Theoretical introduction

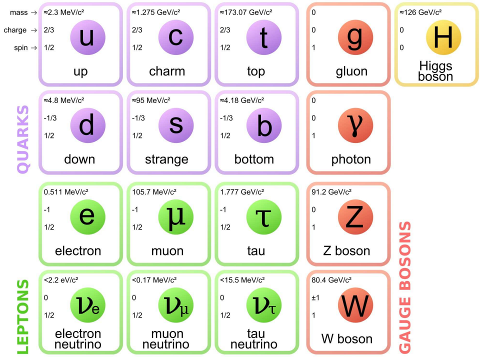
- The Standard Model (SM) is a gauge theory based on the gauge group:

$$\mathbf{SU}(3)_{\mathbf{C}} \times \mathbf{SU}(2)_{\mathbf{L}} \times \mathbf{U}(1)_{\mathbf{Y}}$$

- It describes the elementary constituents of matter and their interactions.
- A crucial ingredient is the concept of spontaneous symmetry breakdown (SSB).

$$SU(2)_L \times U(1)_Y \longrightarrow U(1)_{EM}$$

The SM elementary particles



The Brout-Englert-Higgs mechanism

- The Higgs field is introduced as a complex scalar doublet: $\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$, and the Higgs Lagrangian is written as:

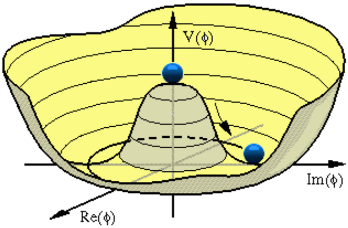
$$\mathcal{L}_{Higgs} = (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi^\dagger \phi),$$

with **the scalar potential** in the form:

$$V(\phi^\dagger \phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2,$$

- For $\mu^2 < 0$ and $\lambda > 0$:
  - Local maximum at $\phi = 0$
  - **Nonzero minimum** at $\phi_0 = \sqrt{\frac{-\mu^2}{2\lambda}} = v/\sqrt{2}$, which breaks the $SU(2) \times U(1)$ symmetry.

The Higgs potencial



- The masses of the W and Z bosons can be written as:

$$\mathbf{m_W = \frac{1}{2} g_2 v}$$

$$\mathbf{m_Z = \frac{1}{2} \sqrt{g_2^2 + g_1^2} v}$$

$\mathcal{L}_{Yukawa}$ and fermion masses

- The Higgs boson mass term is:

$$\mathbf{m_h} = \sqrt{-\mathbf{2}\mu^2} = \sqrt{\mathbf{2}\lambda}\mathbf{v}$$

where the vacuum expectation value $v \approx 246$ GeV.

- Finally, $\mathcal{L}_{\mathbf{Yukawa}}$ gives the mass term for fermions:

$$\mathbf{m_f} = \mathbf{y_f}\frac{\mathbf{v}}{\sqrt{\mathbf{2}}}$$

- The top quark Yukawa coupling is considered particularly interesting since it is of the order of 1:

$$\mathbf{y_{top}} = \sqrt{\mathbf{2}}\frac{\mathbf{m_{top}}}{\mathbf{v}} \approx \mathbf{1}$$

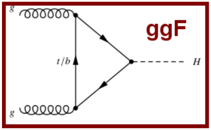
The SM particles

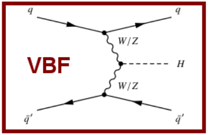
Particle	Q	mass [v]
$W^{\pm}$	± 1	$\frac{1}{2}g_2$
Z	0	$\frac{1}{2}\sqrt{g_2^2 + g_1^2}$
A	0	0
g	0	0
h	0	$\sqrt{2\lambda}$
e, μ, τ	-1	$y_{e,\mu,\tau}/\sqrt{2}$
ν_e, ν_μ, ν_τ	0	0
u, c, t	$+2/3$	$y_{u,c,t}/\sqrt{2}$
d, s, b	-1/3	$y_{d,s,b}/\sqrt{2}$

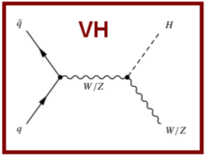
The Higgs mass cannot be predicted from the theory. Experimental value: $m_h = 125.09 \pm 0.21 \pm 0.11$ GeV

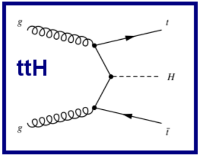
The SM Higgs boson

Production processes in hadron colliders





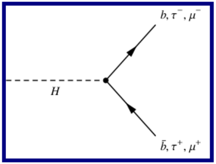


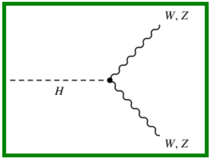


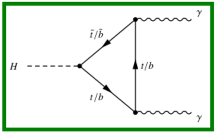
- Cross section [pb], at $\sqrt{s}$=13 TeV:

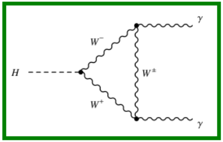
ggF	48.6
VBF	3.78
WH	1.37
ZH	0.884
ttH	**0.507**

Higgs decays









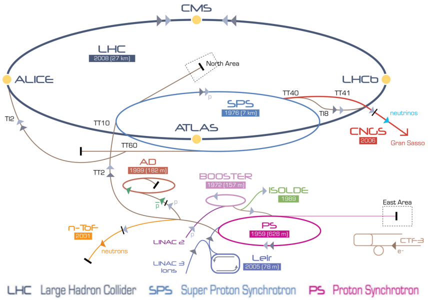
- Branching ratio [%]:

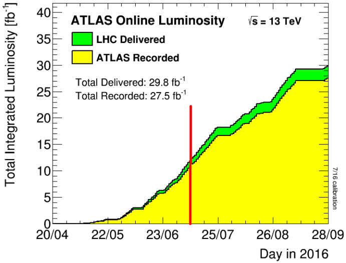
$H \to b\bar{b}$	**58.2**
$H \to WW$	21.4
$H \to c\bar{c}$	2.89
$H \to ZZ$	2.62
$H \to \gamma\gamma$	0.227

CERN & LHC

- The **Large Hadron Collider** (LHC) at CERN is the world's largest and most powerful particle accelerator.
- The aim of the LHC and its experiments is to **test the SM** or reveal the **physics beyond the SM**.
- **The Higgs boson** particle was **discovered** by the ATLAS and CMS collaboration in **July 2012**, forty years after its prediction.
- The LHC machine **collides protons** at a centre-of-mass energy of **13 TeV in Run 2** (7-8 TeV in Run 1).
- Expected $\sim$35 fb^{-1} of data in 2016.

CERN accelerator complex

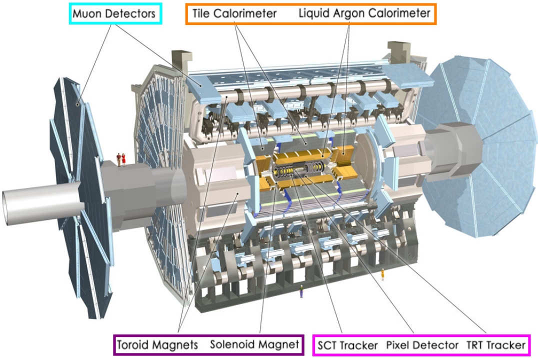




The ATLAS experiment

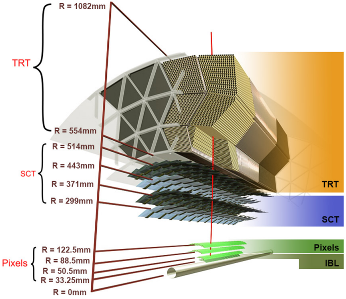
- The ATLAS experiment is a **general purpose** particle physics **experiment**. It investigates a wide range of physics, from the search for the Higgs boson to new physics
- The ATLAS detector consists of four major components:
  - Inner Detector
  - Calorimeters
  - Muon spectrometer
  - Magnet system: one solenoid and three air-core toroids.

ATLAS detector



The ATLAS Detector

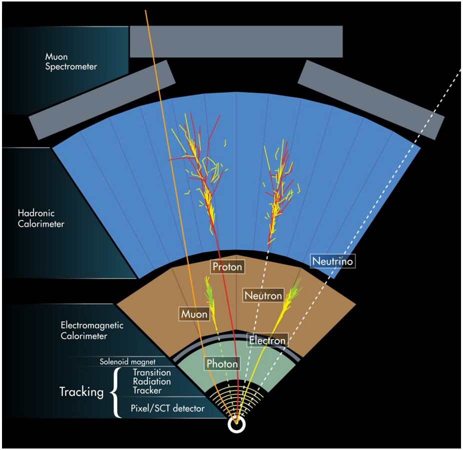
- The Inner Detector consists of 3 independent sub-detectors:
  - The Pixel detector
    - The Insertable B-Layer (IBL), added for Run 2
  - The silicon microstrip (SCT)
  - The transition radiation tracker (TRT)
- The electromagnetic calorimeter
  - Based on a highly granular liquid-argon technology (LAr)
- The Hadronic calorimeter:
  - Based on Tile and LAr technology.



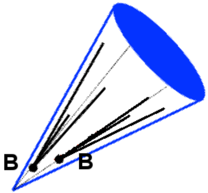
- The muon chambers:
  - Monitored Drift Tube chambers (MDTs), Cathode Strip Chambers (CSCs), Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC)

The ATLAS event reconstruction

- The Inner Detector (ID) is responsible for the **charged particle tracks** and the **primary vertex** reconstruction.
- **Muons** are reconstructed using the information from the muon spectrometer, ID and the calorimeter
- **Electrons** are reconstructed using information from the ID and the electromagnetic calorimeter
- **Jets**, in this thesis, are reconstructed using the anti-k_T algorithm with a radius parameter of $R = 0.4$

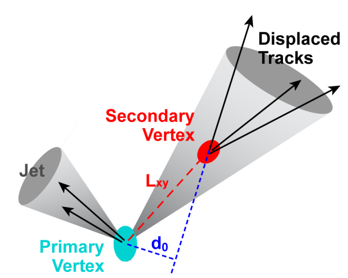


Part 1: Identification of jets containing two B hadrons



Identification of *b*-jets in ATLAS

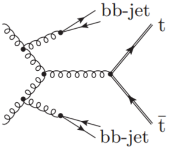
- The procedure to identify jets originating from *b*-quarks is called *b*-tagging.
- Based on the *b*-hadron relatively large lifetime properties such as a **displaced decay secondary vertex** and **large impact parameter tracks**.
- Basic ATLAS *b*-tagging algorithms:
  - Impact parameter (IP)
  - Single Secondary Vertex Finding (SSVF)
  - JetFitter
- Information from IP, SSVF and JetFitter are combined using multivariate analysis (MVA).
  - In Run 1: MV1
  - In Run 2: MV2



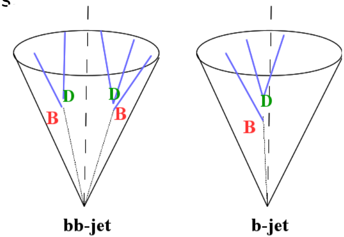
- As qualification work, I was involved in the migration of the ATLAS *b*-tagging software to the new format for Run 2, mainly in the new design of the *b*-tagging interface to secondary vertex algorithms.

Identification of jets containing two B hadrons

- Separate *bb*-jets from *b*-jets is important for several physics processes:
  - $g \rightarrow b\bar{b}$ at small angle might be reconstructed as one *bb*-jet
  - Measurement of gluon splitting properties (test pQCD, parton shower description)
  - Important to control $t\bar{t}b\bar{b}$ (background to $t\bar{t}H$)
  - Identifying boosted signal processes ($H \rightarrow b\bar{b}$)



- No information on the number of B hadrons inside a jet is given by usual *b*-tagging algorithms

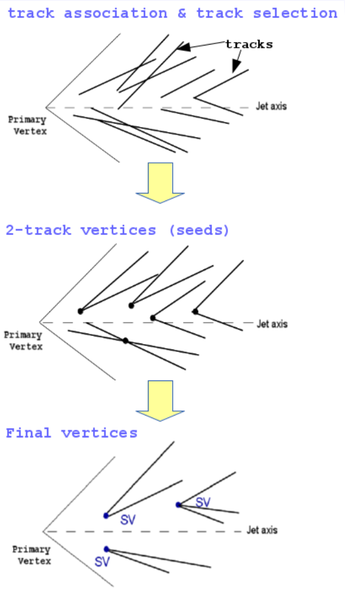


Goal: identify jets containing two B-hadrons

Multi Secondary Vertex Finder algorithm

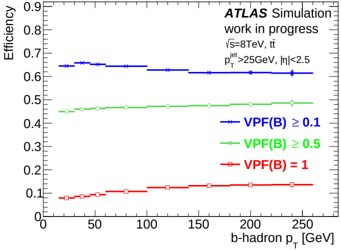
- Association of tracks to jets
  - Tracks are associated to each jet based on the ΔR distance between the jet and tracks: $\Delta R(p_T) = 0.315 + e^{-0.367-1.56.10^{-5} \cdot p_T}$.
- Track selection
  - Designed to select well-measured tracks, and to reject poorly reconstructed tracks or tracks from pileup interactions.

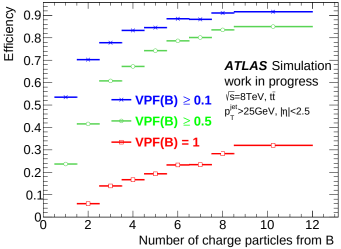
    p_T>700MeV, $|d_0|$<5mm, number of Pixel hits≥1, etc
- 2-track vertices (seeds)
  - The algorithm form two-track vertex candidates. The vertices compatible with V_{0s}: K_s or Λ, photon conversions or hadronic interactions with the detector material are rejected.
- Merge vertices and apply quality cuts on them and finally we have final vertices.



MSVF performance in single-b jets

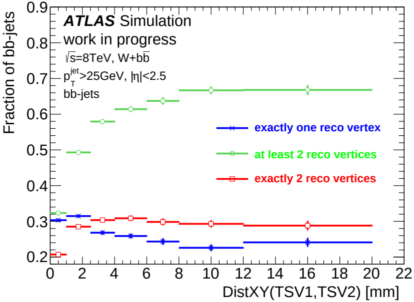
- Studied the **purity of the reconstructed vertices** in jets containing single b-hadron.
- Several metrics were defined to quantify the performance of the MSVF algorithm.
- **VPF(B)= 1**: 100% of the tracks in the vertex are coming from the b-hadron.
- Low efficiency for the VPF(B)=1 category, $\sim$ 20%.





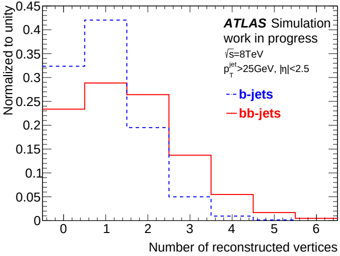
MSVF performance in *bb*-jets

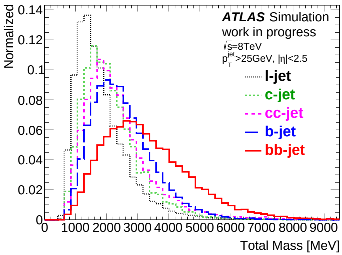
- About **48%** of the *bb*-jets have **at least two reconstructed vertices**. Good performance in *bb*-jets.
- Fraction of *bb*-jet with one, two or more vertices as a function of the transverse distance between the truth secondary vertices:

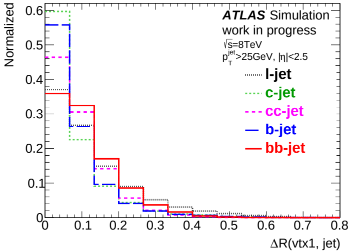


Multi-vertexing properties

- The differences between *b*-jets and **bb-jets** are expected to arise from the presence of two *b*-hadrons in *bb*-jets leading to a **higher number of reconstructed vertices**.
- In addition, different kinematic and topological variables from the reconstructed vertices were investigated.





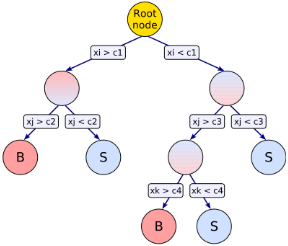


Developed of MultiSVbb taggers

A boosted decision tree (BDT) is used to separate bb-jets form different flavours using multi-vertexing properties.

- **Decision Tree**, splits recursively events into two branches using cuts on some discriminating variables, until a stopping condition is satisfied. In each split the best separation variable is used.
- **Boosting**, averaging several trees for better stability.

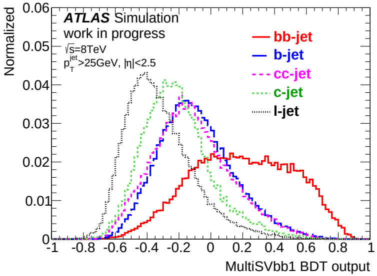
Decision Tree

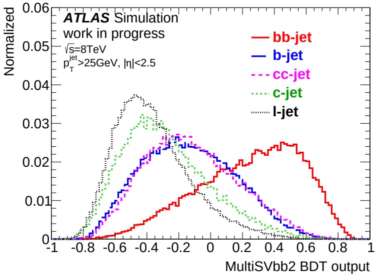


- **BDT training**
  - Optimized for b-jet rejection while keeping light jet rejection at a good rate
  - Training in a mixture of jet flavours (b-, c-, light and cc-jets)

BDT output

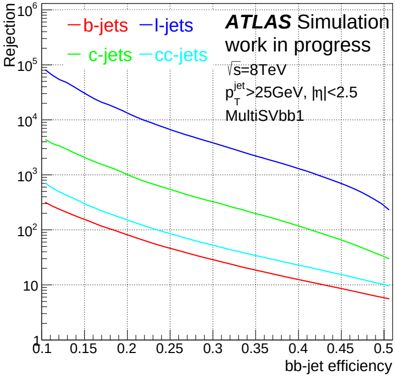
- Two versions were developed:
  - **MultiSVbb1** (12 variables): Use only vertex properties as input variables (simple and robust)
  - **MultiSVbb2** (14 variables): Include topological variables (more model dependent, more efficient)
- The MultiSVbb taggers provide **better separation** between *bb*-jets and other flavour jets **than any individual discriminating variable**.





Performance

- Rejection vs *bb*-jet efficiency to MultiSVbb2 ($r_b = \frac{1}{\epsilon_b}$)



- Rejection at 35% of *bb*-jet efficiency:

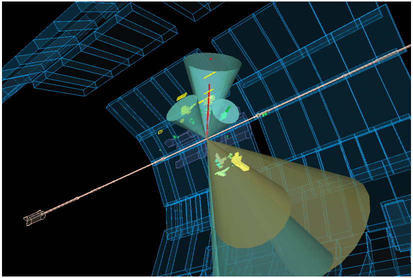
Rejection	MV1	MultiSVbb1	MultiSVbb2
b-jets	3	18	23
c-jets	40	200	250
l-jets	10000	2400	3200
cc-jets	40	35	38

- MultiSVbb2 performs ∼**7 times** better *b*-jet rejection compare to the default Run 1 tagger (MV1).
- MV1 tagger is not tuned to separate *b*- and *bb*- jets; 33% *b*-jet efficiency at 35% *bb*-jet effciency.

Summary of the part 1

- Performance of the multi-secondary vertex algorithm in single-b jets and bb-jets was studied.
- A new b-tagging tool (MultiSVbb) was developed to identify jets containing two b-hadrons (bb-jets).
- Two configurations are retained (MultiSVbb1 and MultiSVbb2) with the best one (MultiSVbb2) performing 7 times better b-jet rejection than the default b-tagging algorithm in ATLAS Run 1 (MV1).

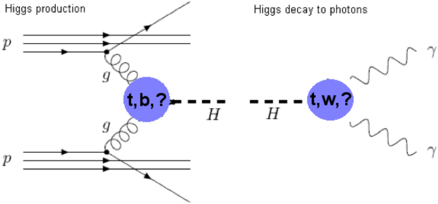
Part 2: Search for the Higgs boson in the single lepton $t\bar{t}H(H \to b\bar{b})$ channel

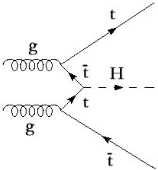


The Higgs boson production in association with a pair of top quarks $t\bar{t}H$

- Precision measurements of the **Higgs coupling to the top and bottom** quark is a **major goal for precision physics** in the scalar sector.

- **Indirect constrain on the top Yukawa coupling**
  - Measurements of the Higgs boson productions via gluon fusion are consistent with SM within experimental uncertainties.
- **Direct measurement**
  - The associated production of a Higgs boson with a top quark is the only way.
  - A measurement of the rate of **ttH production** provides a direct test of this coupling.



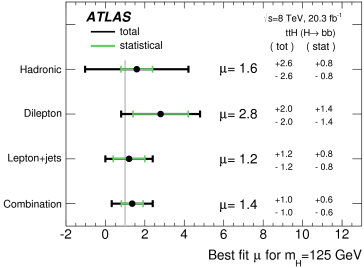


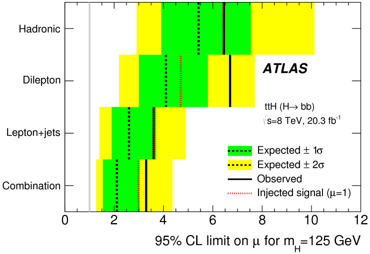
Search for the Higgs boson in the $t\bar{t}H(H \to b\bar{b})$ channel Run 1

ATLAS Run 1

- A search for $t\bar{t}H(H \to b\bar{b})$ by the ATLAS collaboration with **20.3 fb**$^{-1}$ at $\sqrt{s} =$ **8 TeV** was published. A neural network is used to discriminate between signal and background events.

  (no Higgs reconstruction)
- A combined signal strengths $\mu = \sigma/\sigma_{SM}$ of 1.4 ± 1.0 is observed.
- The ATLAS and CMS Run 1 combination: **measurement (expected) significance of 4.4σ (2.0σ)** and a combined signal strengths, $\mu = \sigma/\sigma_{SM}$, of $2.3^{+0.7}_{-0.6}$.

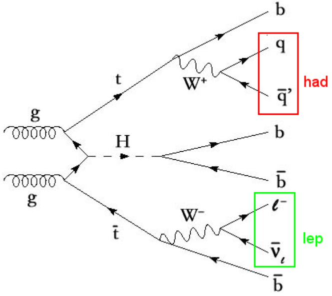




Search for the Higgs boson in the single lepton $t\bar{t}H(H \rightarrow b\bar{b})$ channel

- **Run 2**: increase of the $t\bar{t}H$ ($t\bar{t}$) cross-section by a factor of 3.8 (3.3) at $\sqrt{s}$ = **13 TeV**.
- $t\bar{t}H(H \rightarrow b\bar{b})$ channel has largest BR but large background.
- The analysis is performed with **13.2 fb^{-1}** at $\sqrt{s} = 13$ TeV recorded with the ATLAS experiment in 2015 and between April and July 2016 (shown in **ICHEP**).
- The single lepton $t\bar{t}H(H \rightarrow b\bar{b})$ channel produces **6 jets**, **4** of them **b-jets** and **1 lepton** and **1 neutrino**.

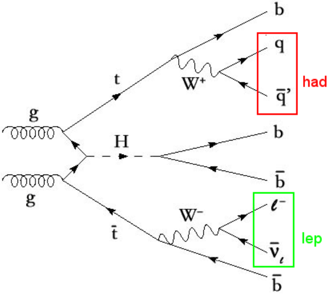
Single lepton channel



Event selection

- The triggers used are single electron and single muon triggers.
- Object selection:
  - **electrons**:
    $p_{\mathrm{T}} > 25$ GeV, $|\eta| < 2.47$, `TightLH`, and required to be isolated.
  - **muons**:
    $p_{\mathrm{T}} > 25$ GeV, $|\eta| < 2.5$, `medium` quality and required to be isolated.
  - **jets**:
    $p_{\mathrm{T}} > 25$ GeV, $|\eta| < 2.5$, anti-pileup cut is applied using information from the associated tracks.
  - ***b*-jets**:
    MV2 tagger at 70% *b*-jet efficiency.

Single lepton channel

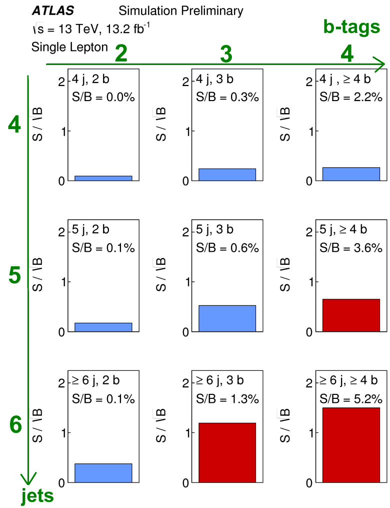


Event categorization

- Events required:

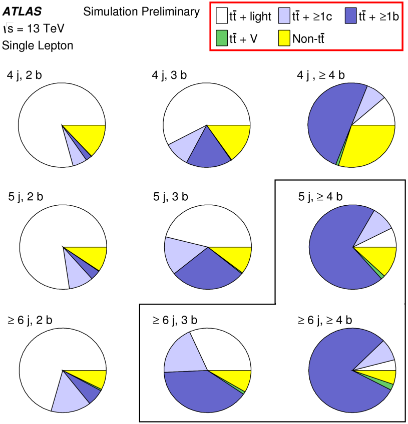
  exactly one lepton and at least 4 jets, 2 of them *b*-tagged jets.
- Selected events are **classified** based on the number of **jets** and **b-tagged jets**.
- As a baseline and following the Run 1 analysis, a total of nine independent regions are considered.
- The regions with a relative large signal-to-background ratio S/B and $S/\sqrt{B}$ are referred to as **signal-rich regions** The remaining regions are referred to as **control regions**.

9 independent regions



Background modelling

Background composition

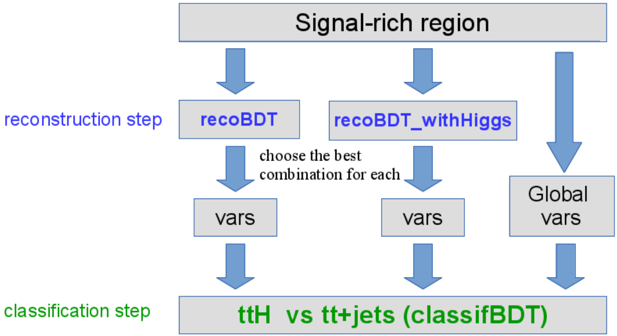


- The main background contribution in the signal regions corresponds to $t\bar{t}+\geq 1b$

- $t\bar{t}$+jets background modelling (Powheg+Pythia6 NLO)
  - $t\bar{t}$+light and $t\bar{t}+\geq 1c$: reweighting to NNLO theory prediction.
  - $t\bar{t}+\geq 1b$: reweighting to SherpaOL NLO $t\bar{t}b\bar{b}$.
- The W/Z + jets, $t\bar{t}V$, single top (s-, t- and Wt-channel) and diboson backgrounds are estimated from MC simulations
- Misidentified lepton background
  - A data-driven method is used (Matrix Method).

Analysis strategy

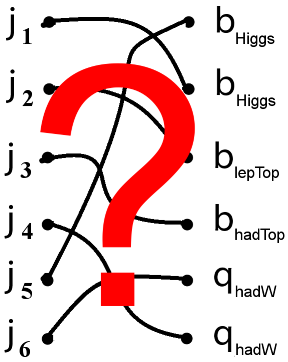
- Multivariate techniques (MVA) is used to further discriminate the signal process from the background.
- **New in Run 2**: Two MVAs have been developed in order to increase the signal-to-background separation.



Multivariate Analysis: Reconstruction BDT

- There are many ways to **associate jets to the partons** of the hard-scattering
- The **reconstruction BDT** is used to find the best match between the observed jets and the final-state quarks from the $t\bar{t}H(H \rightarrow b\bar{b})$ system.

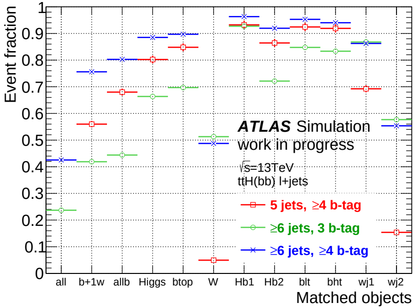
Jet assignment



Multivariate Analysis: Reconstruction BDT

- Jets are identified with the corresponding quarks from the hard scattering process using MC generator level truth matching. $(\Delta R(\text{jet}, \text{quark}) < 0.3)$
- **All products** from the $t\bar{t}H(H \rightarrow b\bar{b})$ system **might not be present** in the selected event.
- In the $\geq$ 6jets, $\geq$ 4b-tags region: Only **42%** of the selected events **have all the products** from the $t\bar{t}H(H \rightarrow b\bar{b})$ decay matching to the reconstructed jets.

Truth matching



Multivariate Analysis: Reconstruction BDT

- Jets are assigned to the quarks from $t\bar{t}H(H \to b\bar{b})$ decay and combinations of jets are used to reconstruct the objects
- **Reconstruction of the leptonic W boson**
  Constraining the mass of the neutrino-lepton system by the $M_W = 80.385$ GeV one can compute:

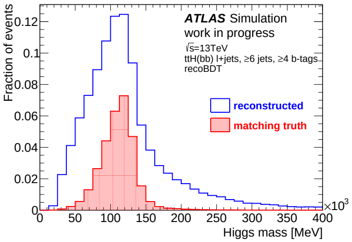
$$p_{z\nu}^{\pm} = \frac{1}{2} \frac{p_{zl}\beta \pm \sqrt{\Delta}}{E_l^2 - p_{zl}^2},$$

  in cases of two solutions, two different leptonic W bosons are considered.
- **Reconstruction of the hadronic W boson**
  The hadronic W is reconstructed using all combinations of 2 jets that are not considered as b-tagged jets
- **Reconstruction of the Top quarks**
  Top quarks are reconstructed by association of one W boson and one b-tagged jet
- **Reconstruction of the Higgs boson**
  The remaining b-tagged jets are used to reconstruct the Higgs boson.

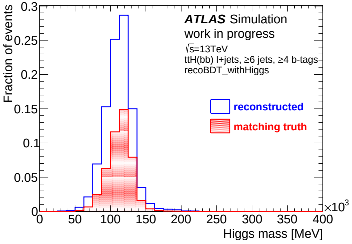
Multivariate Analysis: Reconstruction BDT

- **BDT technique for combinatorial solving**
  Signal: correct combination
  Background: all different jets combinations.
- Two versions of the reconstruction BDT:
  - **recoBDT**, variables depending only on the $t\bar{t}$ system are used.
  - **recoBDT_withHiggs**, it includes variables correlated with the Higgs boson
- recoBDT_withHiggs **bias the reconstructed Higgs mass** distribution reducing thus its discriminant power

recoBDT

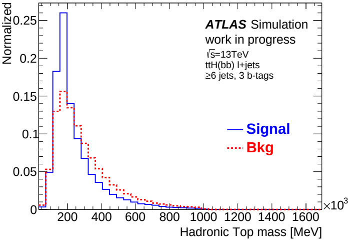


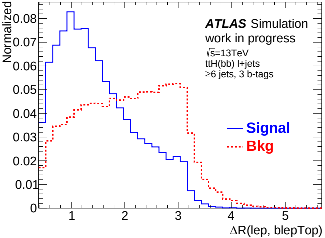
recoBDT_withHiggs

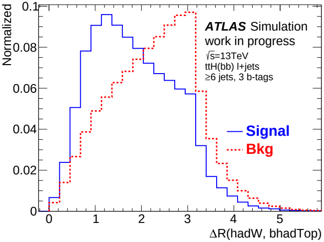


Multivariate Analysis: Reconstruction BDT

- **Input variables**: masses and angular separation which reveal particular kinematic characteristics of the correct and wrong jet combinations.



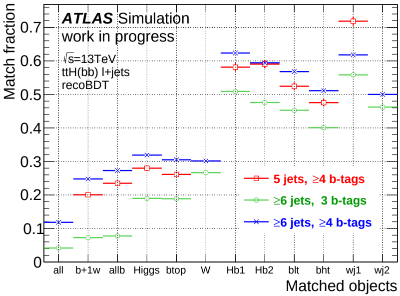




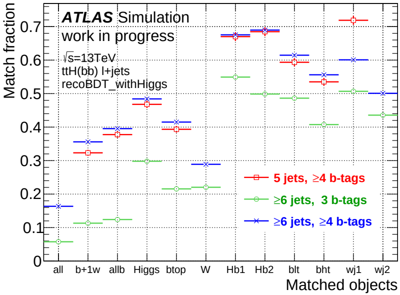
Multivariate Analysis: Reconstruction BDT

- The trained BDT is evaluated for each jet combination in the event, and the jet combination with **the highest BDT score is selected**
- A reconstruction **Higgs matching efficiency** of up to **48%** is obtained in the ($\geq$6 jets, $\geq$4 *b*-tags) region for recoBDT_withHiggs (maximum achievable matching efficiency of $\sim$89%).

recoBDT



recoBDT_withHiggs



Multivariate Analysis: Classification BDT

- The classification BDT provides the final discrimination between the $t\bar{t}H$ signal and the $t\bar{t}$+jets background.
- Three sets of discriminating variables are used.

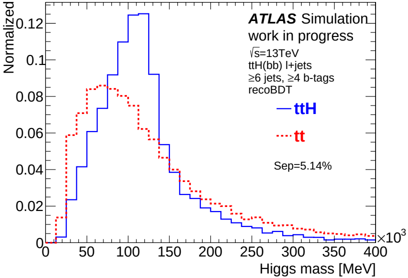
- Variables using the jet assignment from **recoBDT**:
  - Higgs mass
  - ΔR(b1Higgs, b2Higgs)
  - mass(Higgs+blepTop)
  - ΔR(Higgs, lepTop)
- Variables using the jet assignment from **recoBDT_withHiggs**:
  - Highest BDT score
  - ΔR(Higgs,$t\bar{t}$)
  - ΔR(Higgs, bhadtop)

- Global variables: (as used in Run 1)
  - Object kinematic: $p_{\mathrm{T}}^{\mathrm{jet5}}$
  - Event kinematic variables: $H_{\mathrm{T}}^{\mathrm{had}}$
  - Event shape variables: Centrality
  - Object pair properties: $\Delta R_{\mathrm{bb}}^{\mathrm{avg}}$

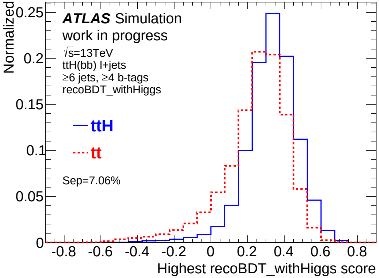
Multivariate Analysis: Classification BDT

- The highest BDT score in the event and Higgs mass are two of **the most important variables**.

Higgs candidate mass



Highest BDT score

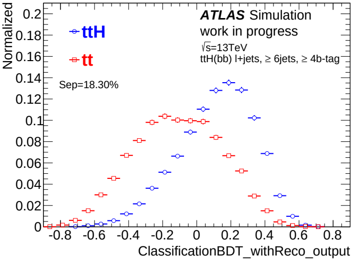


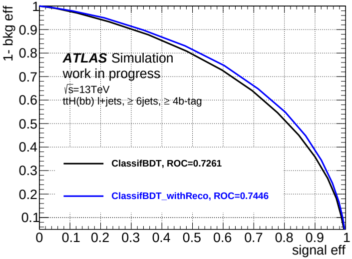
- TMVA separation defined as: $\text{Sep} = \frac{1}{2}\sum_i^{bins} \frac{(N_i^S - N_i^B)^2}{N_i^S + N_i^B}$

Multivariate Analysis: Classification BDT

- In all signal-regions, the classification BDT with reconstruction variables have better separation than the classification BDT with purely global kinematic variables.

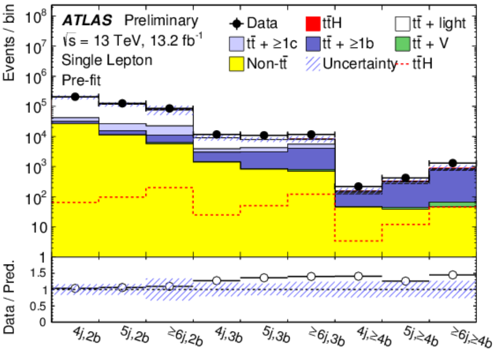
TMVA separation (%)	classificationBDT without Reco	classificationBDT with Reco	Gain
≥6 jets, ≥4 b-tags	15.68	18.30	16.7%
5 jets, ≥4 b-tags	18.10	19.88	9.8%
≥6 jets, 3 b-tags	12.99	13.94	7.3%





Data/MC prediction

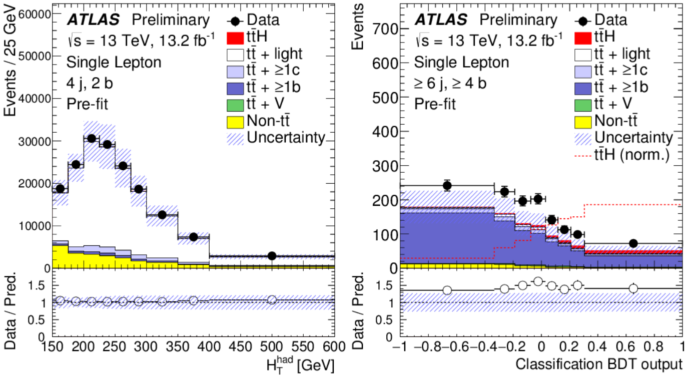
- Discrepancy observed in $t\bar{t}$+HF-enriched regions due to an underestimation of the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ prediction.



- Uncertainties on the normalisation of $t\bar{t}+\geq 1b$ or $t\bar{t}+\geq 1c$ are not included in the pre-fit plots.

Data/MC prediction

- Discrepancy observed in $t\bar{t}$+HF-enriched regions due to an underestimation of the $t\bar{t}+\geq1b$ and $t\bar{t}+\geq1c$ prediction.



- Uncertainties on the normalisation of $t\bar{t}+\geq1b$ or $t\bar{t}+\geq1c$ are not included in the pre-fit plots.

Fit model

- The **signal strength** modifier $\mu = \sigma/\sigma_{SM}$ is determined in a simultaneous binned maximum-likelihood fit to data in all the regions.
- **Normalisation** of $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ backgrounds taken as **free parameters** in the fit to data.
- **Classification BDTs** are used in the profile likelihood fit for the **signal regions**. In the other regions, the scalar sum of the jet p_T (H_T^{had}) is used.

Region	2 b-tags	3 b-tags	4 b-tags
4 jets	$H_{\mathrm{T}}^{\mathrm{had}}$	$H_{\mathrm{T}}^{\mathrm{had}}$	$H_{\mathrm{T}}^{\mathrm{had}}$
5 jets	$H_{\mathrm{T}}^{\mathrm{had}}$	$H_{\mathrm{T}}^{\mathrm{had}}$	BDT
$\geq$ 6 jets	$H_{\mathrm{T}}^{\mathrm{had}}$	BDT	BDT

- The uncertainties are taken into account via nuisance parameters in the fit procedure.

Systematic Uncertainties

Systematic uncertainty	Type	Components
Luminosity	N	1
Pileup reweighting	SN	1
Reconstructed Objects		
Electron trigger+reco+ID+isolation	SN	4
Electron energy scale+resolution	SN	2
Muon trigger+reco+ID+isolation	SN	6
Muon momentum scale+resolution	SN	3
Jet vertex Tagger	SN	1
Jet energy scale	SN	18
Jet energy resolution	SN	1
Missing transverse momentum	SN	3
b-tagging efficiency	SN	5
c-tagging efficiency	SN	4
Light-jet tagging efficiency	SN	14
High-p_{T} tagging	SN	2
Background Model		
$t\bar{t}$ cross section	N	1
$t\bar{t}+b\bar{b}$: NLO Shape	SN	10
$t\bar{t}+c\bar{c}$: NLO Shape	SN	1
$t\bar{t}+\geq 1b$ modelling: (residual) Radiation	SN	1
$t\bar{t}+\geq 1b$ modelling: (residual) NLO generator	SN	1
$t\bar{t}+\geq 1b$ modelling: (residual) parton shower+hadronisation	SN	1
$t\bar{t}$+light, $t\bar{t}+\geq 1c$ modelling: Radiation	SN	2
$t\bar{t}$+light, $t\bar{t}+\geq 1c$ modelling: NLO generator	SN	2
$t\bar{t}$+light, $t\bar{t}+\geq 1c$ modelling: parton shower+hadronisation	SN	2
$t\bar{t}$+light, $t\bar{t}+\geq 1c$ NNLO reweighting	SN	4
W+jets normalisation	N	6
Z+jets normalisation	N	1
Single top cross section	N	2
Wt modelling	SN	3
Diboson normalisation	N	1
$t\bar{t}V$ cross section	N	4
Fakes normalisation	N	6
Signal Model		
$t\bar{t}H$ cross section	N	2
$t\bar{t}H$ branching ratios	N	3
$t\bar{t}H$ model	SN	2

- Experimental uncertainties (65)
  - Luminosity, pile-up, leptons (ID, trigger, isolation), jets (JVT, **JES**, JER), MET, **b-tagging**
- Uncertainties on the background modelling (48)
  - $t\bar{t}$, complete set of $t\bar{t}$ modelling
    - decorrelated between $t\bar{t}$+light, $t\bar{t}+\geq 1c$, $t\bar{t}+\geq 1b$: parton shower and hadronisation, MC generator, ISR/FSR radiation.
    - $t\bar{t}+\geq 1b$: SherpaOL variations, alternative generator and parton shower.
- Uncertainties on the signal modelling (7)
  - XS $t\bar{t}H$ QCD, XS $t\bar{t}H$ PDF, $t\bar{t}H$ **PS and hadronisation** and uncertainties on the Higgs BR ($b\bar{b}$, WW and others).

Results

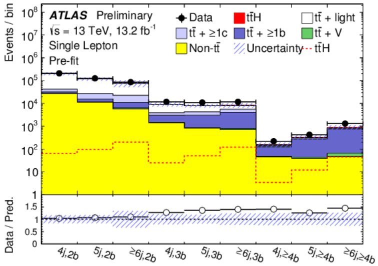
- The best-fit value of $\mu = \sigma/\sigma_{SM}$ is:

$$\mu = 1.6^{+0.5}_{-0.5}(\mathrm{stat.})^{+1.0}_{-0.9}(\mathrm{syst.}) = 1.6^{+1.1}_{-1.1}$$

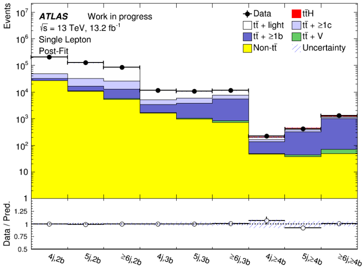
  dominated by systematics.

- Post-fit $t\bar{t} + \geq 1b$ normalisation is 1.24 and $t\bar{t} + \geq 1c$ normalisation is 1.37, compatible with Run 1 results.

Pre-fit



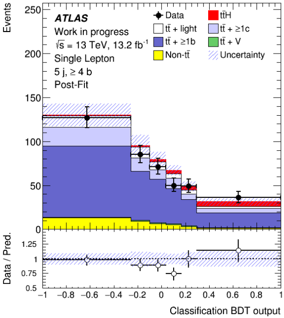
Post-fit



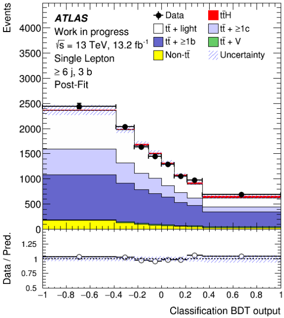
Results

- A good agreement between data and MC simulation is observed.

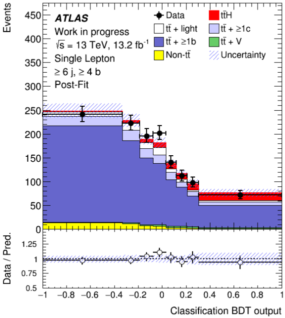
5 jets, $\geq$4 b-tags



$\geq$6 jets, 3 b-tags

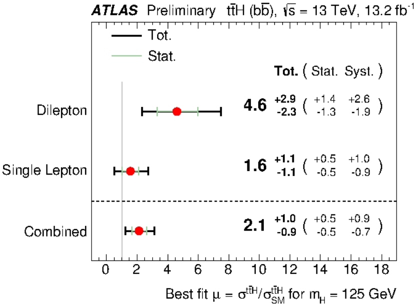


$\geq$6 jets, $\geq$4 b-tags

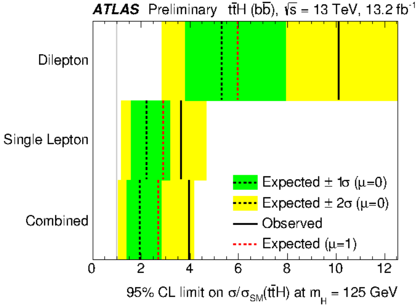


Combination with the dilepton channel

Signal strength: $2.1^{+1.0}_{-0.9}$



Upper limits: 4.0 (1.9) obs (exp)



- Sensitivity is improved with respect to Run 1.

	$\sqrt{s}$ [TeV]	Lumi [fb^{-1}]	μ	Significance obs (exp) [σ]
ATLAS Run 1	8	20.3	1.5 ± 1.1	1.4 (1.1)
ATLAS Run 2	13	13.2	$2.1^{+1.0}_{-0.9}$	2.4 (1.2)

Summary of the part 2

- Performed a search for the Higgs boson in the single lepton $t\bar{t}H(H \rightarrow b\bar{b})$ channel with 13.2 fb^{-1} of data at 13 TeV.
- The analysis has been carried out in event categories based on the number of jets and *b*-tagged jets: six signal-depleted regions and three signal-rich regions.
- In the signal-rich regions a method to reconstruct the $t\bar{t}H(H \rightarrow b\bar{b})$ system was implemented, up to 48% to correctly reconstruct the Higgs boson.
- Variables from the MVA reconstruction improve the signal-to-background separation by about 16% in the most sensitive region ($\geq$6 jets, $\geq$4 *b*-tags).
- The best-fit value of μ is found to be 1.6 ± 1.1. An observed (expected) 95% confidence level upper limit of 3.6 (2.2) times the SM cross section is obtained.

Conclusion

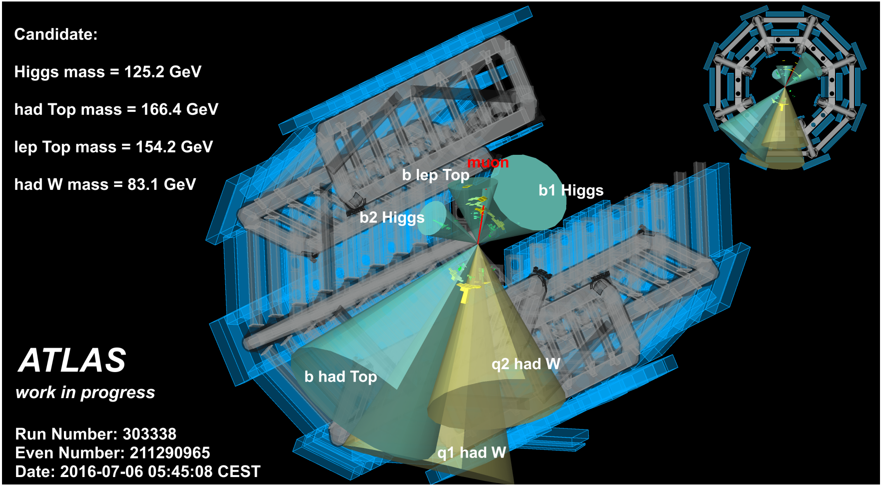
This thesis presented two major studies:

- **The identification of jets containing two B hadrons (bb-jets).**
  - Developed of a new *b*-tagging tool for the identification of *bb*-jets. The proposed method provides an increase of about 7 times in the separation power between *bb*-jet and *b*-jet compared to the default *b*-tagging algorithm in ATLAS Run 1.
  - Prospects: combined with other methods for a better performance and start thinking in the calibration.

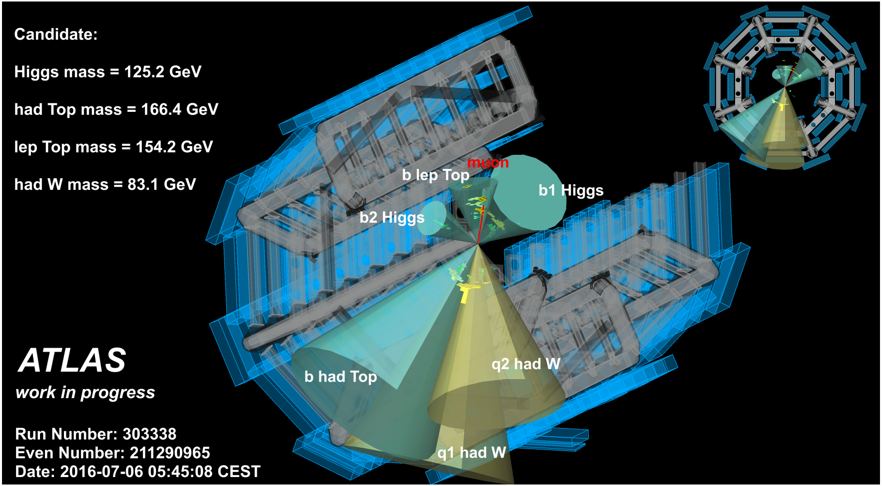
Conclusion

- **The search for the Higgs boson in the $t\bar{t}H(H \to b\bar{b})$ single lepton channel.**
  - First ATLAS Run 2 results at 13 TeV with 13.2 fb^{-1} were showed. Better sensitivity than Run 1 analysis, in agreement with the SM. (ATLAS-CONF-2016-080)
  - Main contribution: a sequence of BDTs in order to improve the signal-to-background separation was developed.
  - Prospects for ATLAS Run 2:
    - Improve the analysis techniques (MVAs, *b*-tagging, etc).
    - Understand and improve $t\bar{t}$ modeling.
    - Top-Higgs coupling should be accessible via associated ttH production in Run 2.

Single lepton $t\bar{t}H(H \to b\bar{b})$ event candidate



Single lepton $t\bar{t}H(H \to b\bar{b})$ event candidate



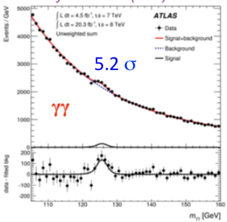
Thank you for your attention

Backup

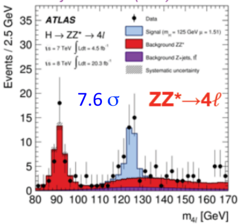
Higgs Run 1

"Higgs boson" also discovered in several single channels in Run 1

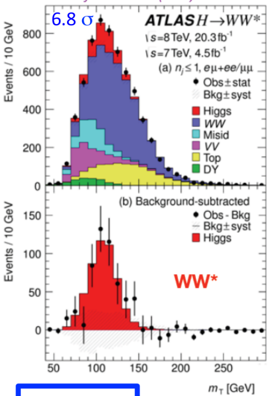
Phys. Rev. D90 (2014) 112015



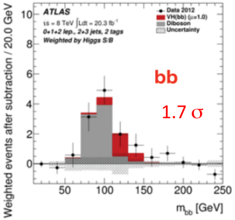
Phys. Rev. D91 (2015) 012006



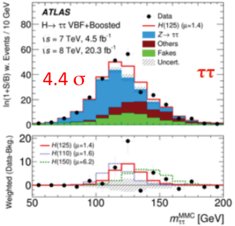
Phys. Rev. D92 (2015) 012006



JHEP 01 (2015) 069



JHEP 04 (2015) 117



Zγ < 11 x SM
μμ < 7 x SM

04/08/16 B. LAFORGE - Higgs Studies with ATLAS 5

MultiSVbb: Discriminating variables:

MultiSVbb1 (12 variables)

Use only vertex properties as input:

Properties per jet:

- jet p_T.
- number of reco vertices.
- total mass of vertices.
- total number of tracks.
- diffntrkSSVF = totalntrk - total number of track from SSVF. (in vertices)
- normDist, significance with a common(weighted) vertex position.

Properties of vertex with maximum(vtx1) and second maximum(vtx2) mass:

- mass of vtx1 ,
- mass of vtx2,
- energy fraction of vtx1,
- energy fraction of vtx2,
- significance of vtx1,
- significance of vtx2.

MultiSVbb2 (14 variables)

Include additional topological variables:

Properties per jet:

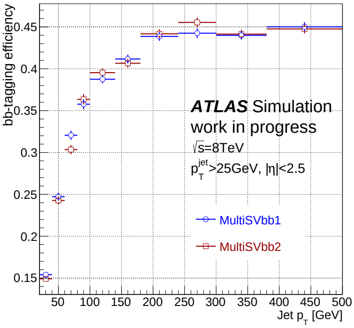
- jet p_T.
- number of reco vertices.
- maximum energy fraction.
- total mass of vertices.
- total number of tracks.
- diffntrkSSVF = totalntrk - total number of track from SSVF. (in vertices)
- normDist, significance with a common(weighted) vertex position.

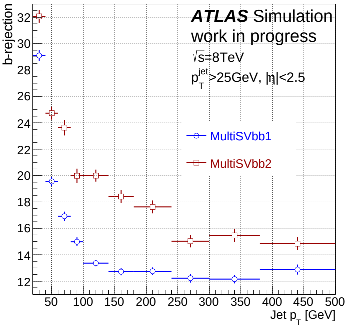
Properties of vertex with maximum(vtx1) and second maximum(vtx2) mass:

- energy fraction of vtx2,
- significance of vtx1,
- ΔR between vtx1 and jet axis,
- ΔR between vtx2 and jet axis,
- distance xy between vtx1 and vtx2,
- ΔR between vtx1 and vtx2,
- Angle between vtx1 and vtx2.

MultiSVbb performance: Global @35% Eff

- Efficiency as function of the jet p_T
- b-rejection as function of the jet p_T





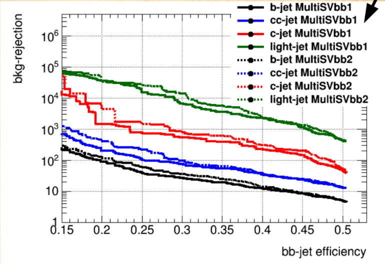
MultiSVbb taggers in Run 2

MultiSVbb performances and optimization for Run 2.

T. Calvet (CPPM)

10

Caution : comparison between mc12 and mc15. Different p_T spectrum.



Performances:

- Bkg-rejections VS bb-jet efficiency.
- Bkg-rejections at 35% efficiency and increase with respect to 8TeV train/test.

@35% eff	MultiSVbb1	MultiSVbb2
b-rej/gain	19 / +5%	24 / +4%
cc-rej/gain	55 / +57%	63 / +66%
c-rej/gain	390 / +100%	600 / x2.5
light-rej/gain	3600 / +50%	5500 / +70%

Overall Few % gain in b-jet rejection and high gain in other rejection at 35% bb-jet efficiency.
Better than MV2 to differentiate bb-jets from b-jets.

Signal and background modelling

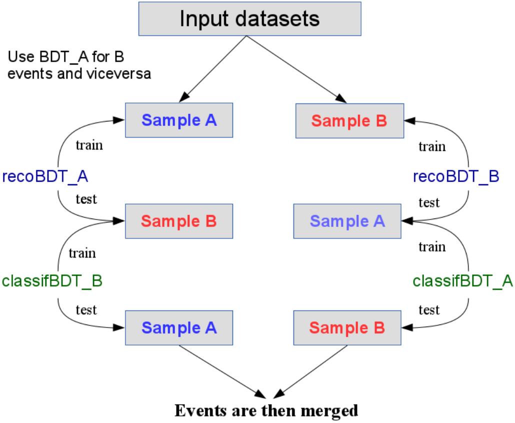
MC samples

Process	Generator	Shower
	Signal	
$t\bar{t}H$	MG5_aMC	Pythia 8.210
	Top-quark	
$t\bar{t}$	Powheg-Box	Pythia 6.428
t-channel single top	Powheg-Box	Pythia 6.428
s-channel single top	Powheg-Box	Pythia 6.428
Wt-channel single top	Powheg-Box	Pythia 6.428
	V + jets	
W + jets	Sherpa	Sherpa 2.1.1
Z + jets	Sherpa	Sherpa 2.1.1
	$t\bar{t}V$	
$t\bar{t}V$	MG5_aMC	Pythia 8.210
	Diboson + jets	
WW + jets	Sherpa	Sherpa 2.1.1
WZ + jets	Sherpa	Sherpa 2.1.1
ZZ + jets	Sherpa	Sherpa 2.1.1

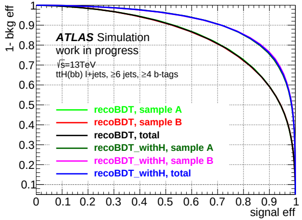
- $t\bar{t}$ + jets background modelling
  - $t\bar{t}$+light and $t\bar{t}+\geq 1c$: reweighting to NNLO theory prediction. Sequential $p_{\mathrm{T}}(t\bar{t})$ and $p_{\mathrm{T}}(\mathrm{top})$.
  - $t\bar{t}+\geq 1b$: reweighting to Sherpa OL NLO $t\bar{t}b\bar{b}$.
- Misidentified lepton background
  - A data-driven method known as the Matrix Method is used.
- The W/Z + jets, $t\bar{t}V$, single top (s-, t- and Wt-channel) and diboson backgrounds are estimated from MC simulations

MVA: cross-training

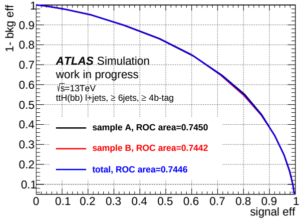
Crooss-training is used to use the full MC statistics.



recoBDT



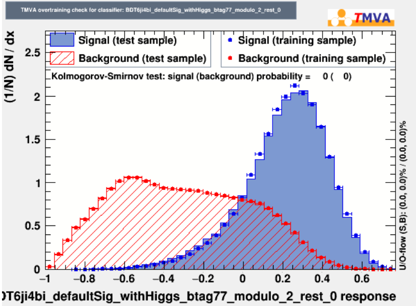
classifBDT

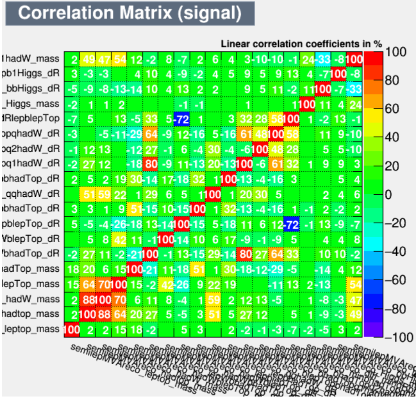


MVA: recoBDT discriminating variables

Variable	Region		
	≥6 jets, ≥4 b-tags	≥6 jets, 3 b-tags	5 jets, ≥4 b-tags
Topological information from $t\bar{t}$:			
Leptonic Top mass	✓	✓	✓
Hadronic Top mass	✓	✓	–
Incomplete hadronic Top mass	–	–	✓
Hadronic W mass	✓	✓	–
Mass of hadW and blepTop	✓	✓	–
Mass of hadW and blepTop	–	–	✓
Mass of lepW and bhadTop	✓	✓	✓
ΔR(hadW, bhadTop)	✓	✓	–
ΔR(qhadW, bhadTop)	–	–	✓
ΔR(hadW, blepTop)	✓	✓	✓
ΔR(qhadW, blepTop)	–	–	✓
ΔR(lep, blepTop)	✓	✓	✓
ΔR(lep, bhadTop)	✓	✓	✓
ΔR(blepTop, bhadTop)	✓	✓	✓
ΔR(q1hadW, q2hadW)	✓	✓	–
ΔR(bhadTop, q1hadW)	✓	✓	–
ΔR(bhadTop, q2hadW)	✓	✓	–
$\Delta R^{\min}$(bhadTop, q$_i$hadW)	✓	✓	–
$\Delta R^{\min}$(bhadTop, q$_i$hadW) - ΔR(lep, blepTop)	✓	✓	–
ΔR(bhadTop, qhadW) - ΔR(lep, blepTop)	–	–	✓
Topological information from Higgs :			
Higgs mass	✓	✓	✓
ΔR(b1Higgs, b2Higgs)	✓	✓	✓
ΔR(b1Higgs, lep)	✓	✓	✓
Mass of Higgs and q1hadW	✓	✓	✓
ΔR(b1Higgs, bleptop)	–	✓	✓
ΔR(b1Higgs, bhadtop)	–	✓	✓

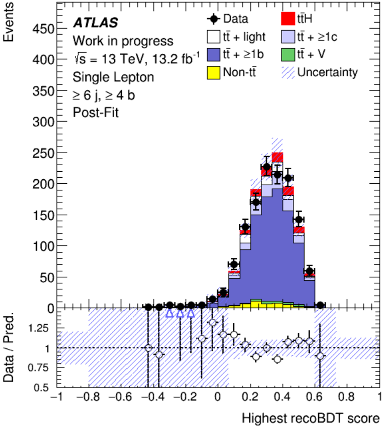
MVA: recoBDT output

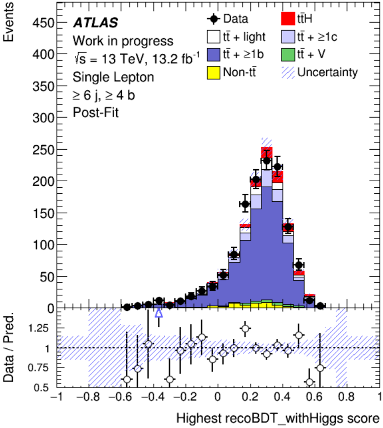




MVA: recoBDT Data/MC

$\geq$6 jets, $\geq$4 b-tags

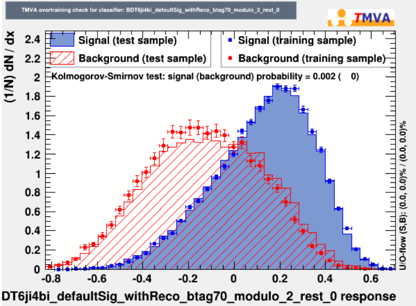


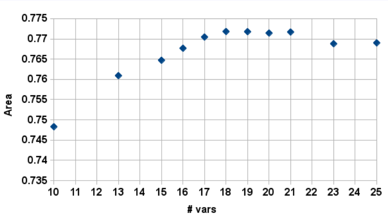


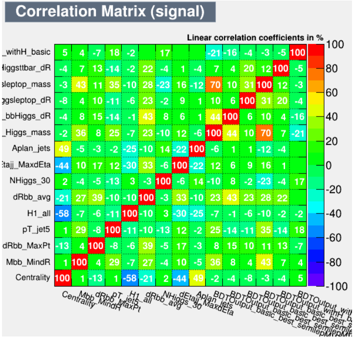
MVA: ClassifBDT discriminating variables

Variable	Region		
	≥6 jets, ≥4 b-tags	≥6 jets, 3 b-tags	5 jets, ≥4 b-tags
Global variables:			
Centrality	✓	✓	✓
$\Delta\eta_{\mathrm{jj}}^{\max\ \Delta\eta}$	✓	✓	✓
$H1$	✓	✓	✓
$p_{\mathrm{T}}^{\mathrm{jet5}}$	✓	✓	✓
$\Delta R_{\mathrm{bb}}^{\mathrm{avg}}$	✓	✓	✓
Aplan	✓	✓	✓
$\mathrm{N}_{30}^{\mathrm{Higgs}}$	✓	–	✓
$m_{\mathrm{bb}}^{\min\ \Delta R}$	✓	✓	–
$m_{\mathrm{bj}}^{\max\ p_T}$	–	✓	–
$\Delta R_{\mathrm{bb}}^{\max\ p_T}$	✓	–	–
$\Delta R_{\mathrm{lep-bb}}^{\min\ \Delta R}$	–	–	✓
N_{40}^{jet}	–	✓	–
$H_{\mathrm{T}}^{\mathrm{had}}$	–	✓	✓
$m_{\mathrm{jj}}^{\min\ \Delta R}$	–	–	✓
Variables from `recoBDT`:			
Higgs mass	✓	✓	✓
ΔR(b1Higgs, b2Higgs)	✓	✓	✓
Mass(Higgs+blepTop)	✓	–	–
ΔR(Higgs, lepTop)	✓	–	–
Variables from `recoBDT`_withHiggs:			
Highest BDT score	✓	✓	✓
ΔR(Higgs, $t\bar{t}$)	✓	✓	✓
ΔR(Higgs, bhadtop)	–	✓	✓

MVA: classifBDT output





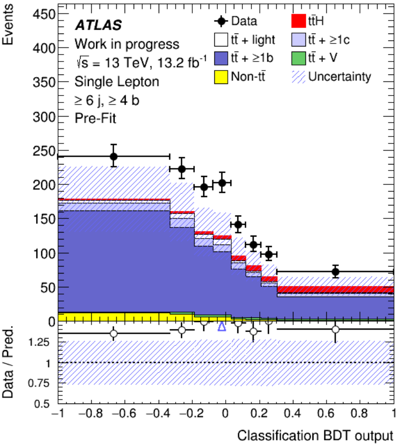


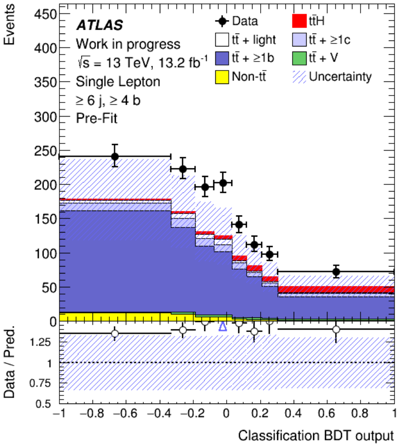
$t\bar{t}$ modelling systematics

- An uncertainty of $\pm 6\%$ is assumed for the inclusive $t\bar{t}$ cross-section.
- **NLO MC generator**, derived by comparing two alternative predictions, Powheg-Box and MG5_aMC.
- **PS and hadronisation model**, derived by comparing events produced by Powheg-Box interfaced with Pythia6 or Herwig++.
- **Initial and final state radiation (ISR/FSR)**, obtained by comparing two alternative radiation variation samples of Powheg+Pythia6.
- $t\bar{t}+\geq 1b$ **NLO Sherpa OL**, evaluated by varying the Sherpa settings (PDF and scales choises).
- $t\bar{t}+\geq 1b$ **NLO MC generator (reweighting)**, derived by comparing the prediction from SherpaOL (4FS) and MG5_aMC+Pythia8.
- $t\bar{t}+\geq 1b$ **PS and hadro (reweighting)**, taken from the difference between MG5_aMC showered with Pythia8 or Herwig++ (4FS).
- $t\bar{t}+\geq 1c$ **NLO generator**, obtained from the comparison with a dedicated NLO $t\bar{t}+c\bar{c}$ sample generated with MG5_aMC+Herwig++ (3F).
- **NNLO reweighting $p_{\mathrm{T}}(t\bar{t})$ and $p_{\mathrm{T}}(\mathrm{top})$**, take the largest difference between NNLO and all considered $t\bar{t}$ samples.

DC/MC prediction

Pre-fit plot after the $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$ SFs applied (right)





Systematic uncertainties: impact on μ

Uncertainty Source	$\Delta\mu$	
$t\bar{t}+\geq 1b$ modelling	+0.53	−0.53
Jet flavour tagging	+0.26	−0.26
$t\bar{t}H$ modelling	+0.32	−0.20
Background model statistics	+0.25	−0.25
$t\bar{t}+\geq 1c$ modelling	+0.24	−0.23
Jet energy scale and resolution	+0.19	−0.19
$t\bar{t}$+light modelling	+0.19	−0.18
Other background modelling	+0.18	−0.18
Jet-vertex association, pileup modelling	+0.12	−0.12
Luminosity	+0.12	−0.12
$t\bar{t}Z$ modelling	+0.06	−0.06
Light lepton (e,μ) ID, isolation, trigger	+0.05	−0.05
Total systematic uncertainty	+0.90	−0.75
$t\bar{t}+\geq 1b$ normalisation	+0.34	−0.34
$t\bar{t}+\geq 1c$ normalisation	+0.14	−0.14
Total statistical uncertainty	+0.49	−0.49
Total uncertainty	+1.02	−0.89

Fits with alternative MVA

Differences in the fit results are expected from both the difference in separation power and the difference in systematic uncertainties.

S+B hypothesis, fit to real data

Channel	$\mu(t\bar{t}H)$	$k(t\bar{t}+\geq 1b)$	$k(t\bar{t}+\geq 1c)$
	Boosted Decision Trees		
Single Lepton	1.57 +1.15/-1.06	1.24 +0.23/-0.21	1.37 +0.70/-0.60
Dilepton	4.6 +2.9 / -2.3	1.30 +0.29 / -0.29	2.29 +0.84 / -0.70
Combined	2.1 +1.0/-0.9	1.33 +0.18/-0.17	1.31 +0.18/-0.17
	Neural Network		
Single Lepton	0.91 +1.12 / -1.13	1.27 +0.24 / -0.21	1.40 +0.74 / -0.62
Dilepton	1.27 +3.20 / -2.56	1.53 +0.34 / -0.33	1.92 +0.87 / -0.78
Combined			

Table 43: Fitted values and post-fit uncertainties of the signal strength $\mu(t\bar{t}H)$, and of the normalization factors for $t\bar{t}+\geq 1b$ and $t\bar{t}+\geq 1c$, from the unblinded fit to data with the signal-plus-background hypothesis.

The difference in the shape and normalisation of the $t\bar{t}+\geq 1b$ residual generator uncertainty can have a large impact on the fitted value of the signal strength.

Fin