Search for ttH production ($H \rightarrow b\bar{b}$) at the ATLAS experiment

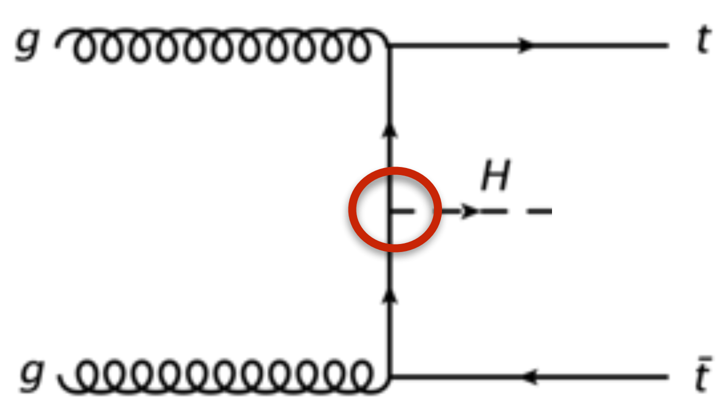
Nedaa-Alexandra Asbah (DESY)
On behalf of the ATLAS collaboration

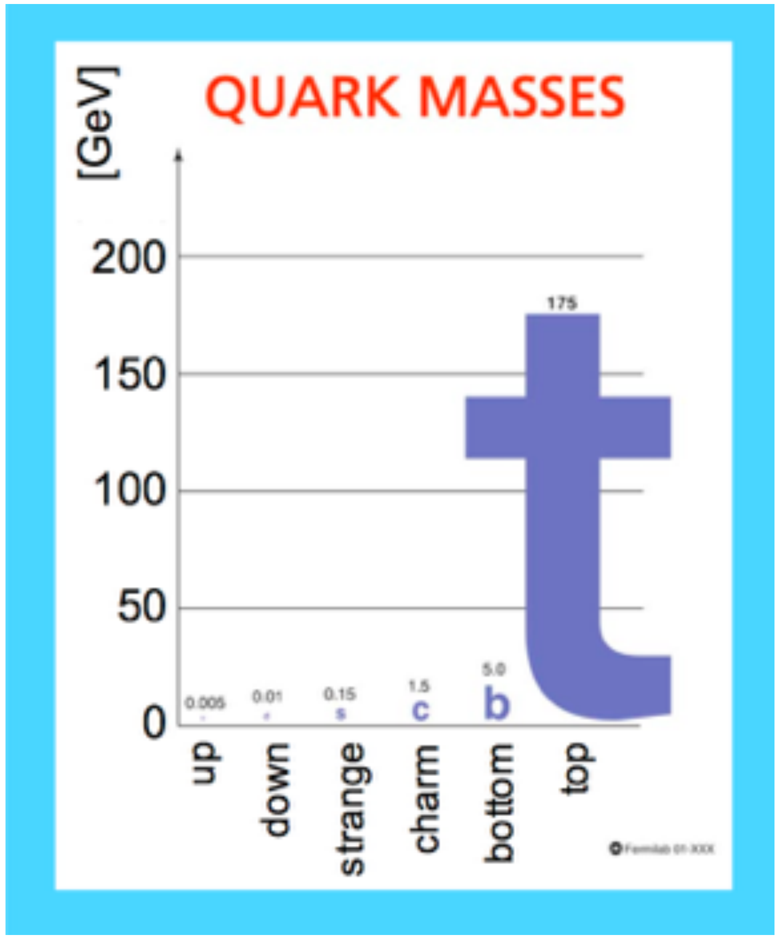
Epiphany2017

ATLAS-CONF-2016-080

Higgs coupling to Top

- **Why Higgs coupling to the Top quark is interesting ?**
  - Top is the heaviest particle in the SM
  - **Yukawa coupling** to Higgs is responsible for generating masses of fermions in the SM and it is proportional to the fermion mass
    - Top coupling to Higgs is the strongest ~1
  - Measuring the Top Yukawa coupling is an essential test of the SM

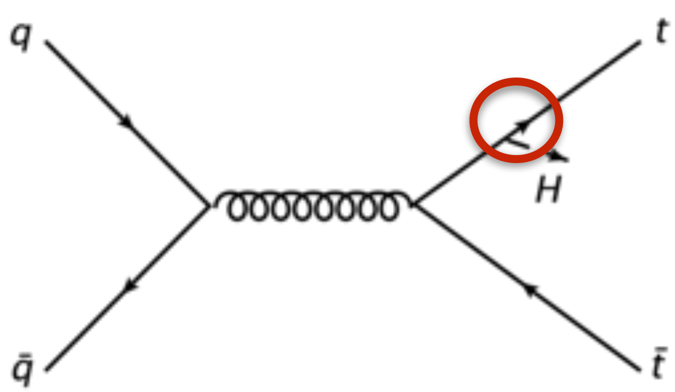


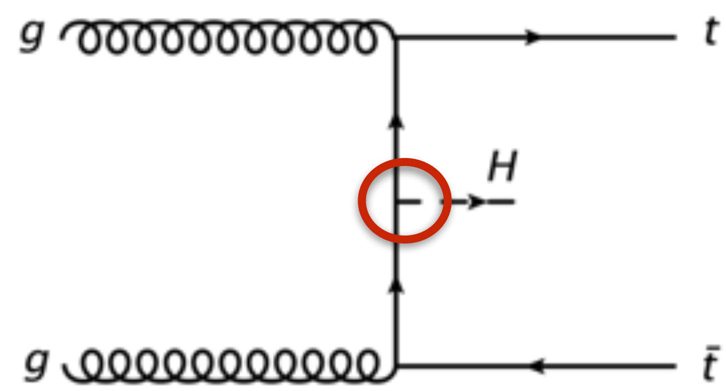


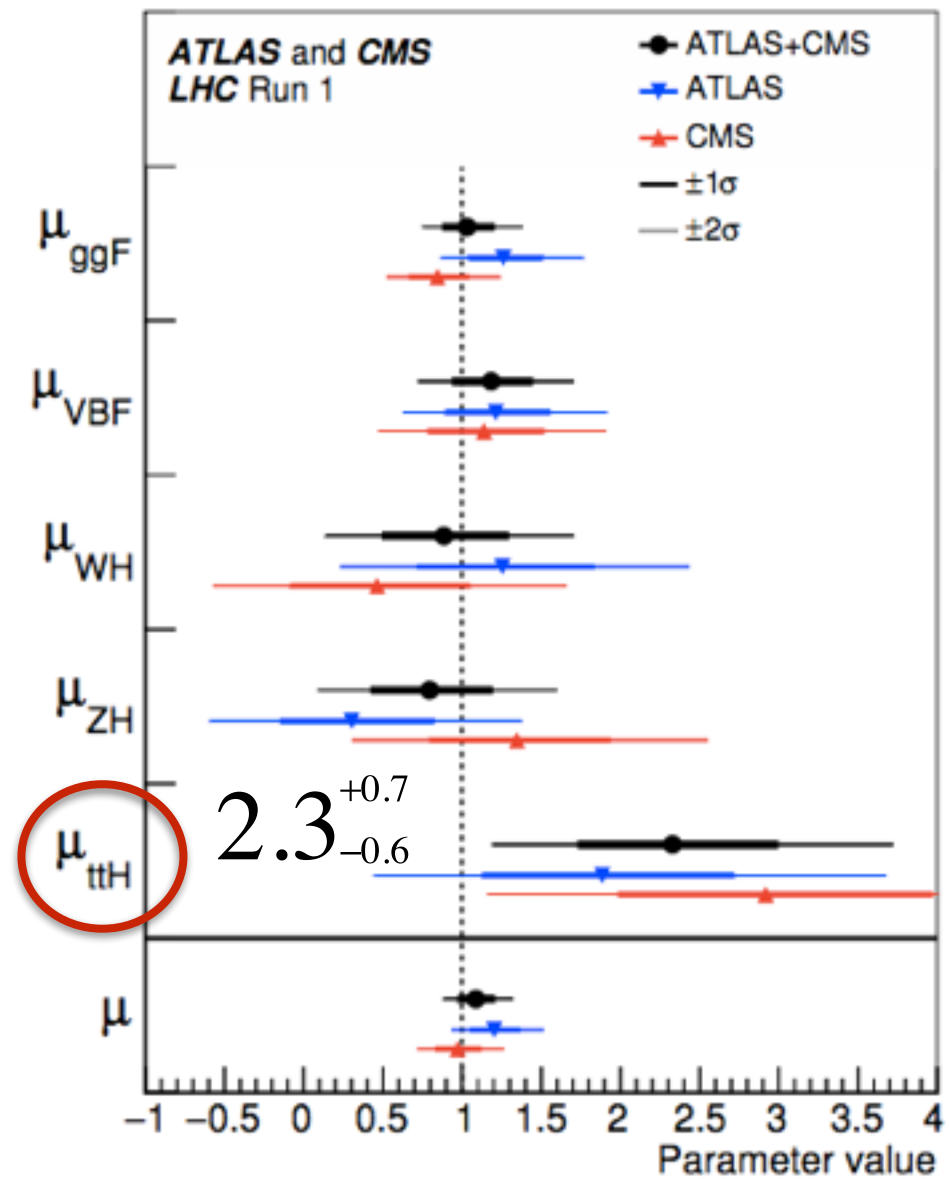
ttH: direct access to top-Higgs coupling

Any deviation might be hint for new physics

ttH Production at the LHC







- LHC Run 1: ttH signal strength “μ_{ttH}” has been measured.
- 2.0σ is the expected significance over background of observing the SM ttH process
- The observed significance was found to be **4.4σ**
- The observed ttH production rate is 2.3σ higher than the SM prediction, **still compatible and maybe a hint of something interesting**
- **LHC Run 2: Large increase in the ttH cross section**

8 TeV	13 TeV
0.13 pb	0.5 pb

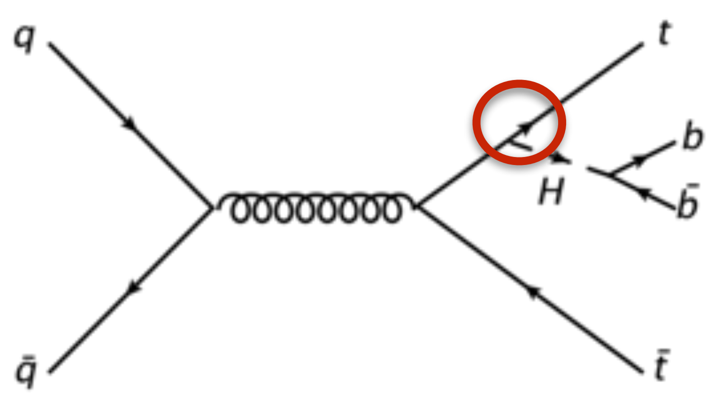
$$\mu_i = \frac{\sigma_i}{(\sigma_i)_{SM}}$$

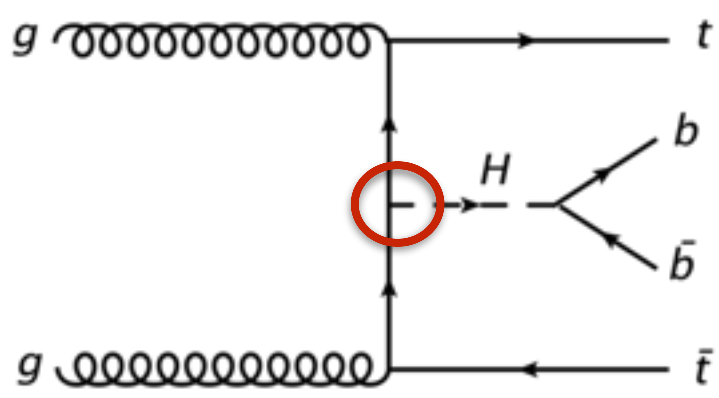
Observed

Signal Strength

Theoretical

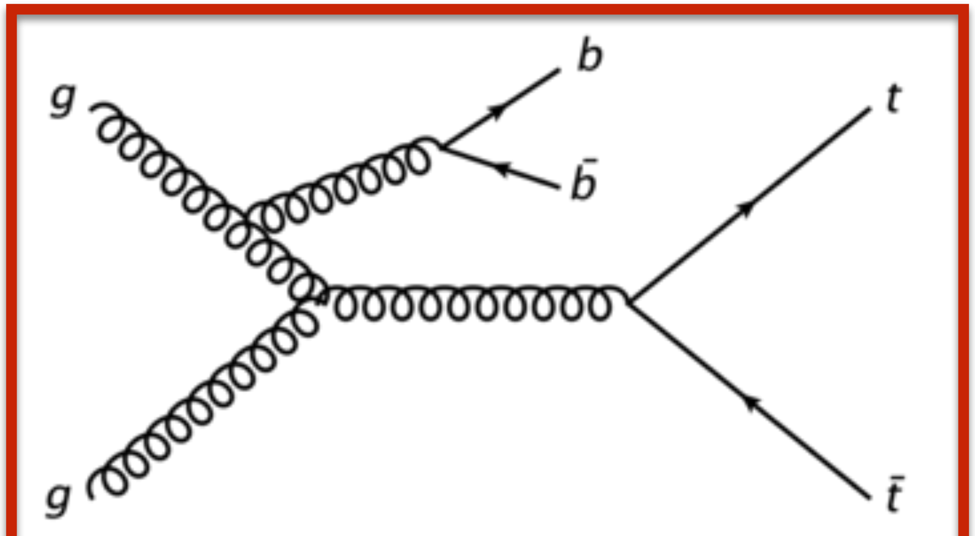
Higgs Branching Ratios





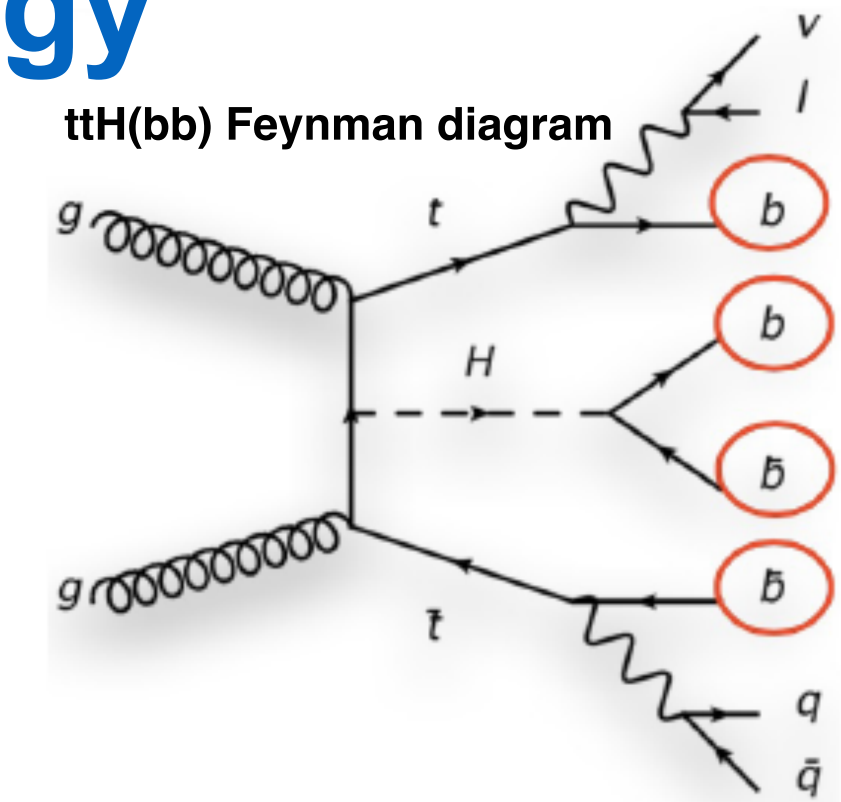
- H →bb has the largest branching ratio
  - Large tt+jets background
    - Similar final state to tt+bb
  - H→ bb gives access to coupling to b's

Higgs Decay Mode	Branching ratios
H → γγ	0.23%
H → ZZ	2.6%
H → ττ	6.3%
H → WW	21.5%
H → bb	**58.51%**



Event Topology

Events triggered by single lepton triggers



Single Lepton channel

- One leptonic W decay
- One electron or one muon
- At least 4 jets
- At least 2 b-tagged jets

Di-lepton channel

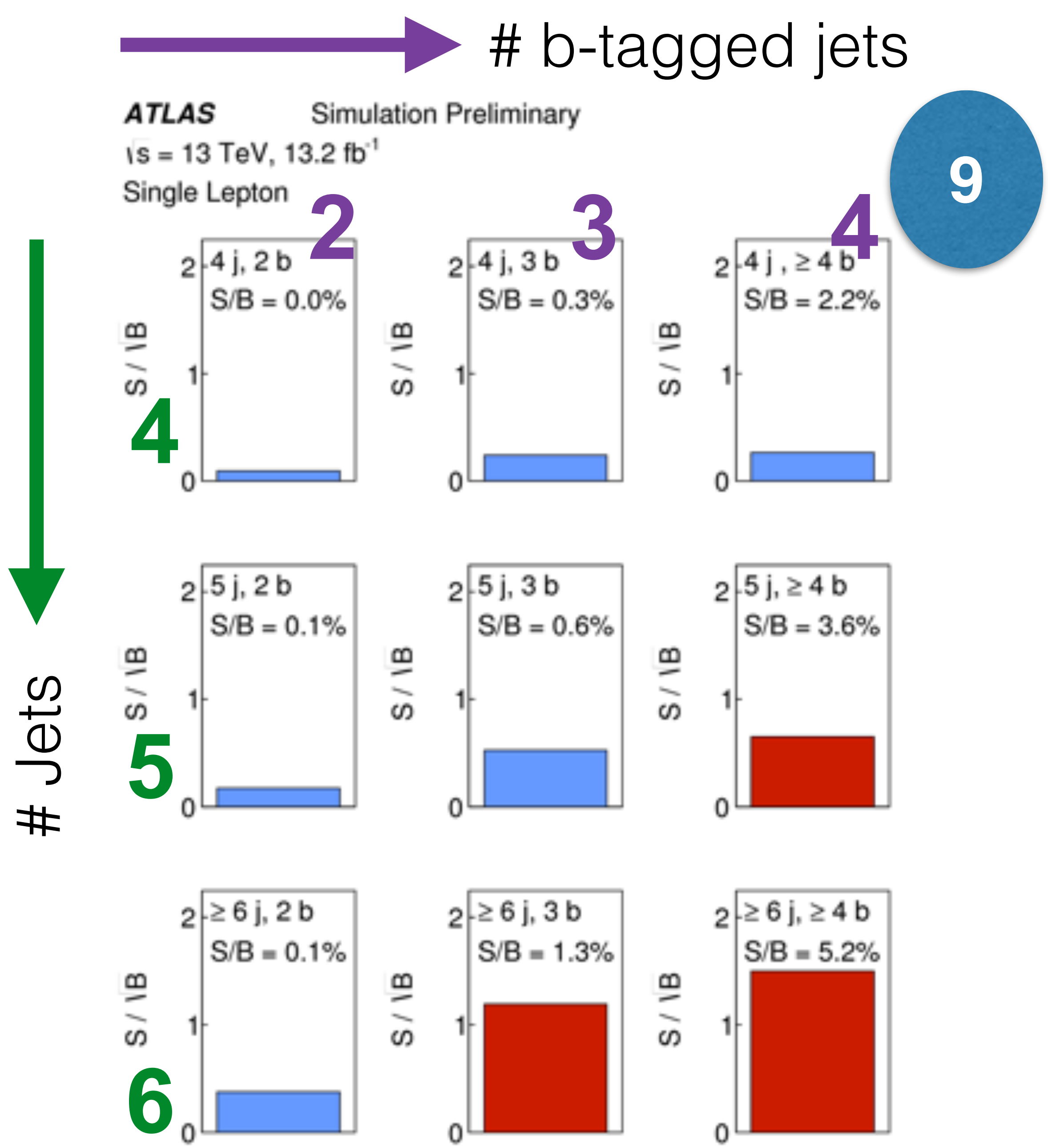
- Two leptonic W decays
- Two opposite charge light leptons (e,μ)
- At least 3 jets
- At least 2 b-tagged jets

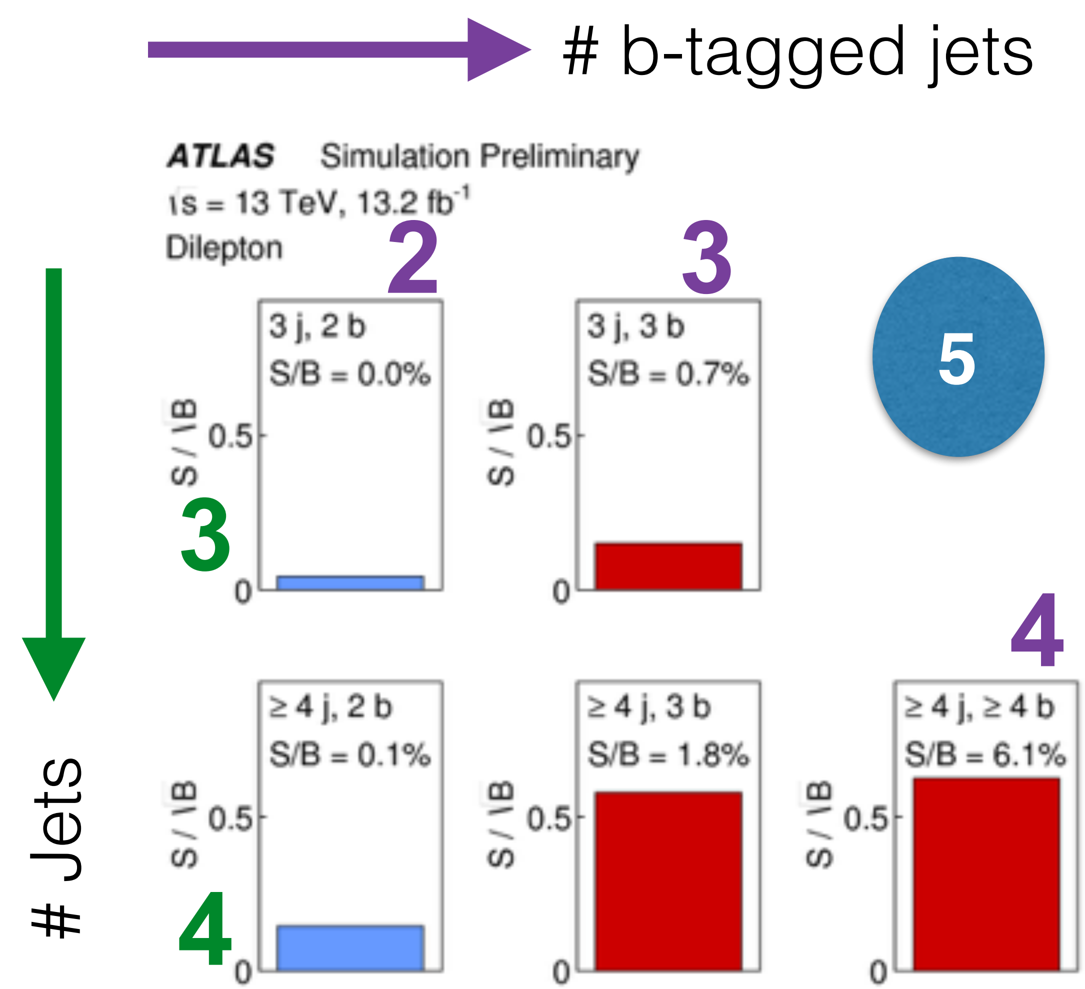
- Very challenging analysis :
  - 4 b-jets in final state
  - Large background from $t\bar{t}$+jets
  - Strategy : **Divide** into different regions

- This dataset used 3.2 fb^{-1} from 2015 & 10.0 fb^{-1} from 2016.

Signal & Background

Events are categorised according to # of jets and # of b-tagged





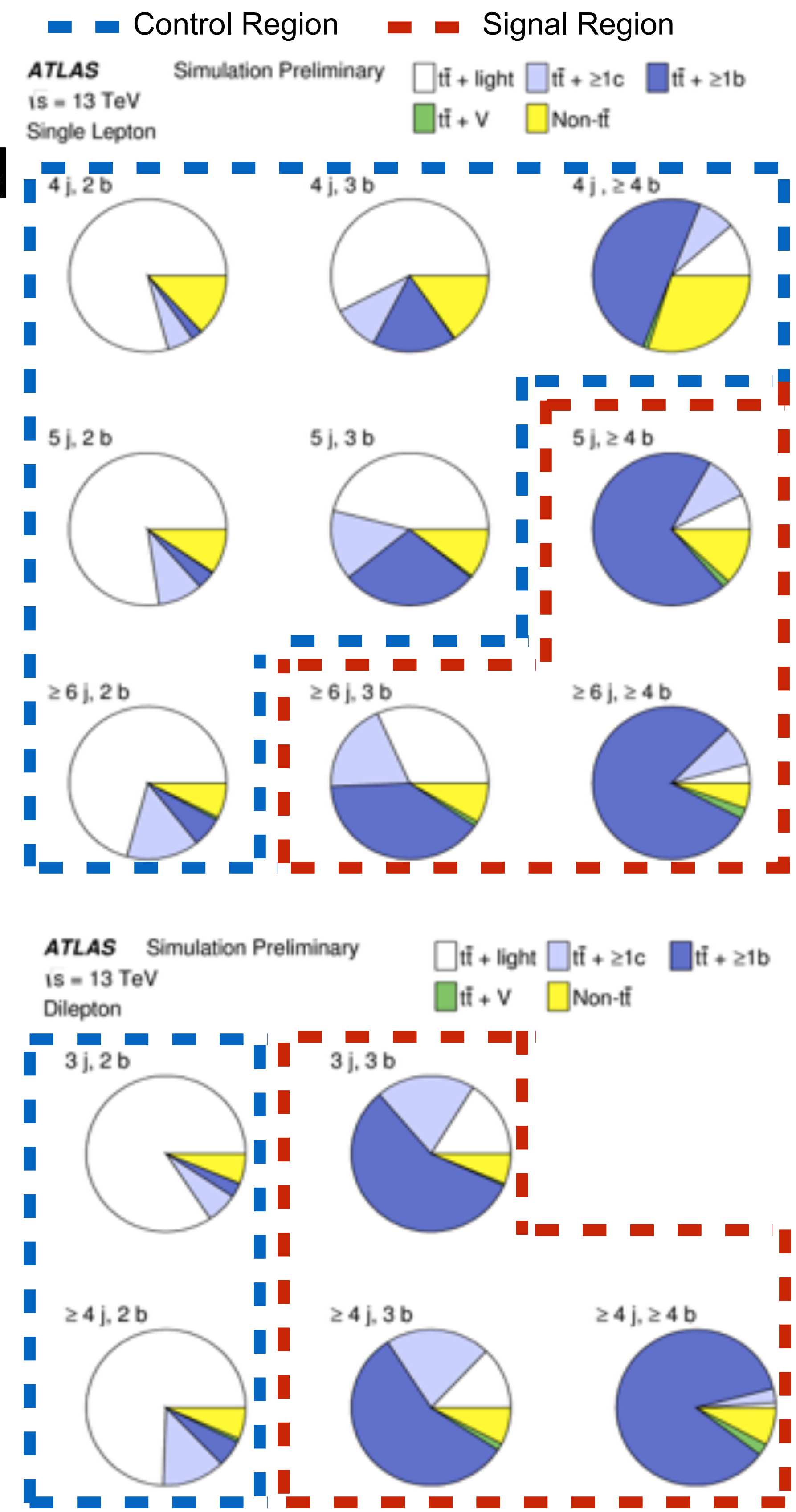
Signal- rich region

Control regions

(constrain systematic uncertainties on the background)

Signal & Background Composition

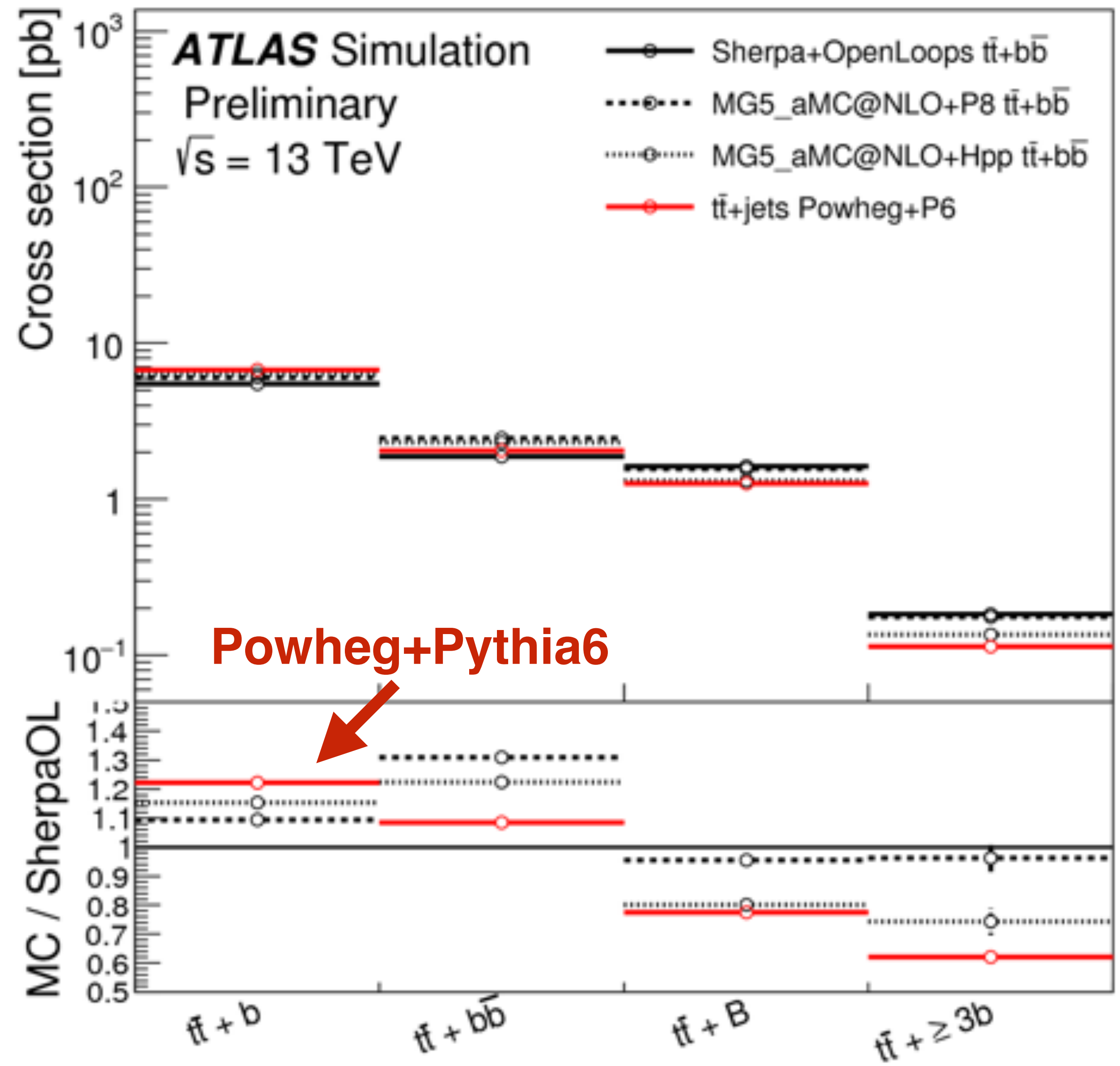
- The ttH signal process is modelled using MadGraph5_aMC@NLO+Pythia8
- Dominating background ($t\bar{t}$+jets)
  - $t\bar{t}$+ ≥ 1 b jets
  - $t\bar{t}$+ ≥ 1 c jets Powheg+Pythia6
  - $t\bar{t}$+ light jets
- Other backgrounds
  - $t\bar{t}$+V
  - Non-$t\bar{t}$
    - Single Top, W/Z+jets, Diboson
    - Multi-jet (Fakes and non-prompt)



$t\bar{t}$+ Heavy Flavour (HF) background modelling

- Estimating $t\bar{t}$+ HF is a critical & challenging part of the analysis
- Powheg+Pythia6 NLO, $p_T(\bar{t})$ & $p_T(t\bar{t})$ spectra corrected to NNLO calculation
- $t\bar{t}$+ ≥ 1 b-jet corrected to Sherpa +OpenLoops NLO 4-flavour scheme calculation in the different sub-categories

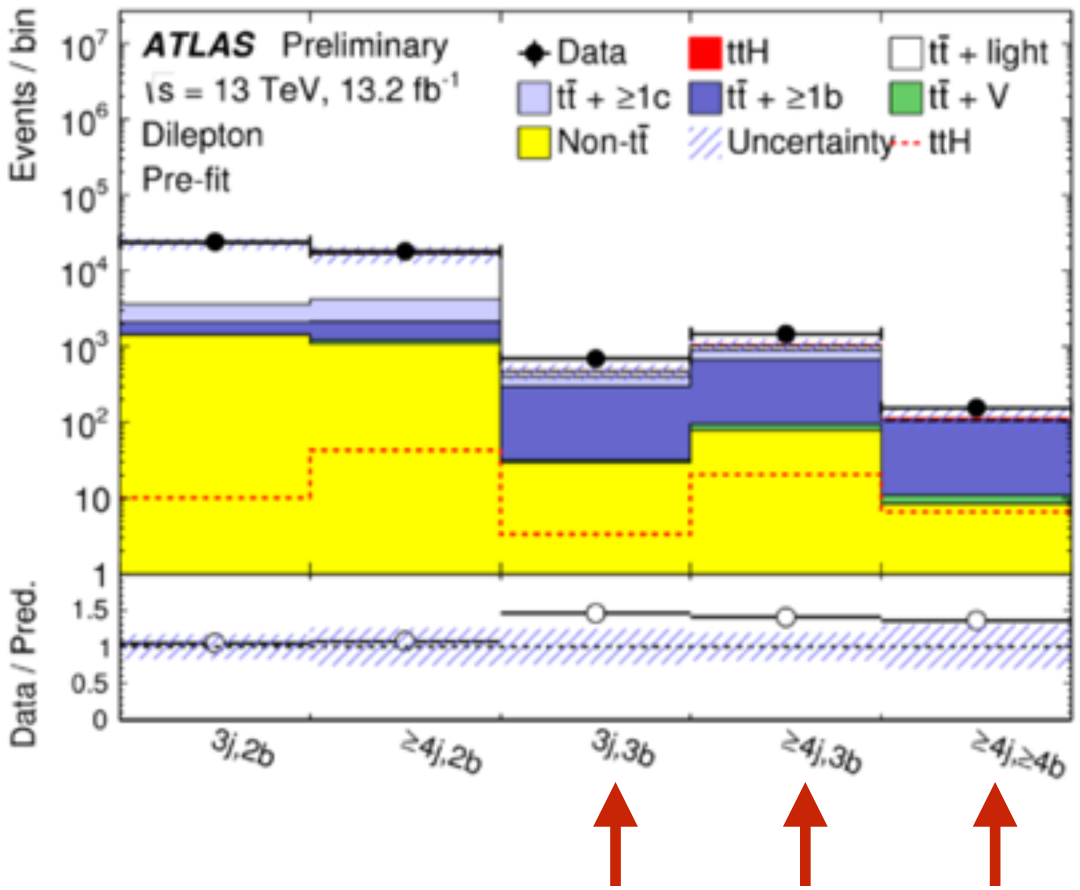
Predicted cross-sections for the $t\bar{t}$ + ≥1b sub-categories

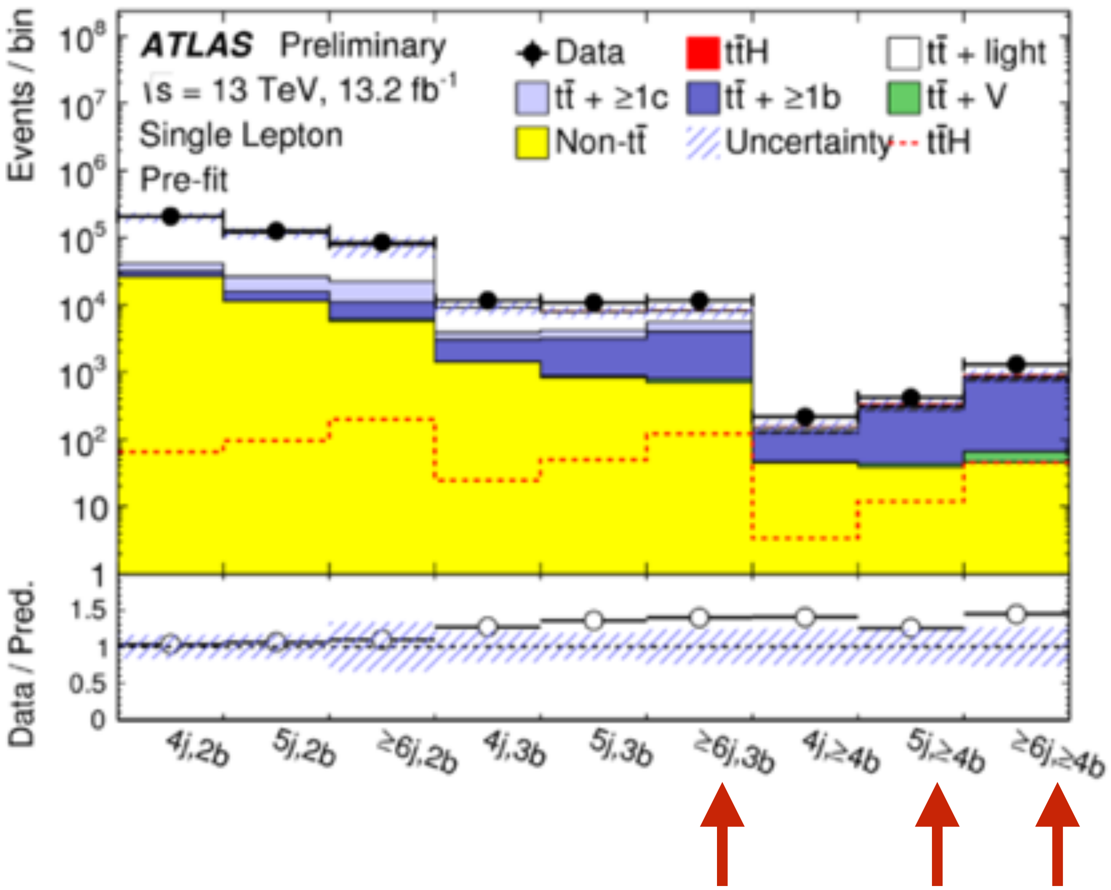


The inclusive Powheg+ Pythia 6 prediction is compared to 4-flavour $t\bar{t}$+bb calculations from SherpaOL and MG5_aMC with different parton showers

Data/ MC yields in different regions

Comparison of predicted and observed event yields

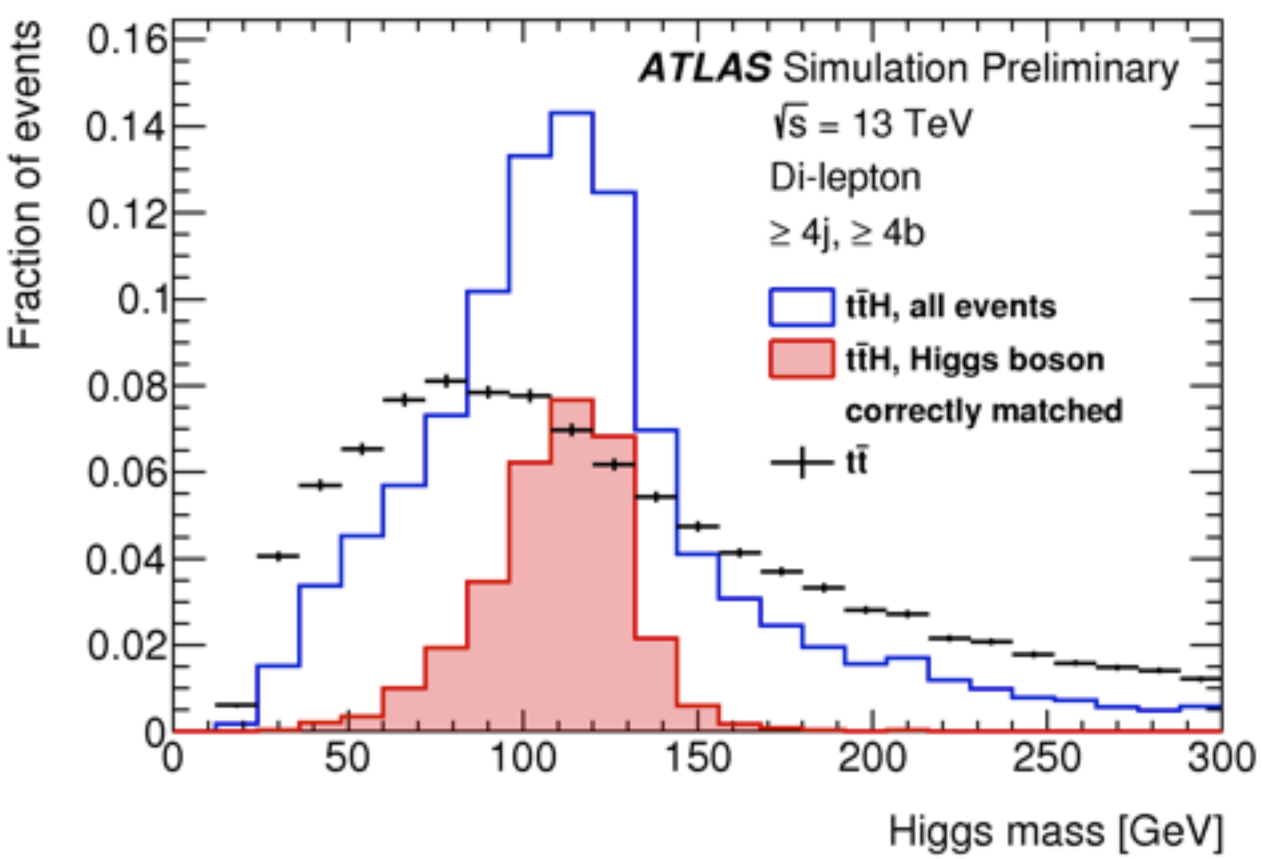


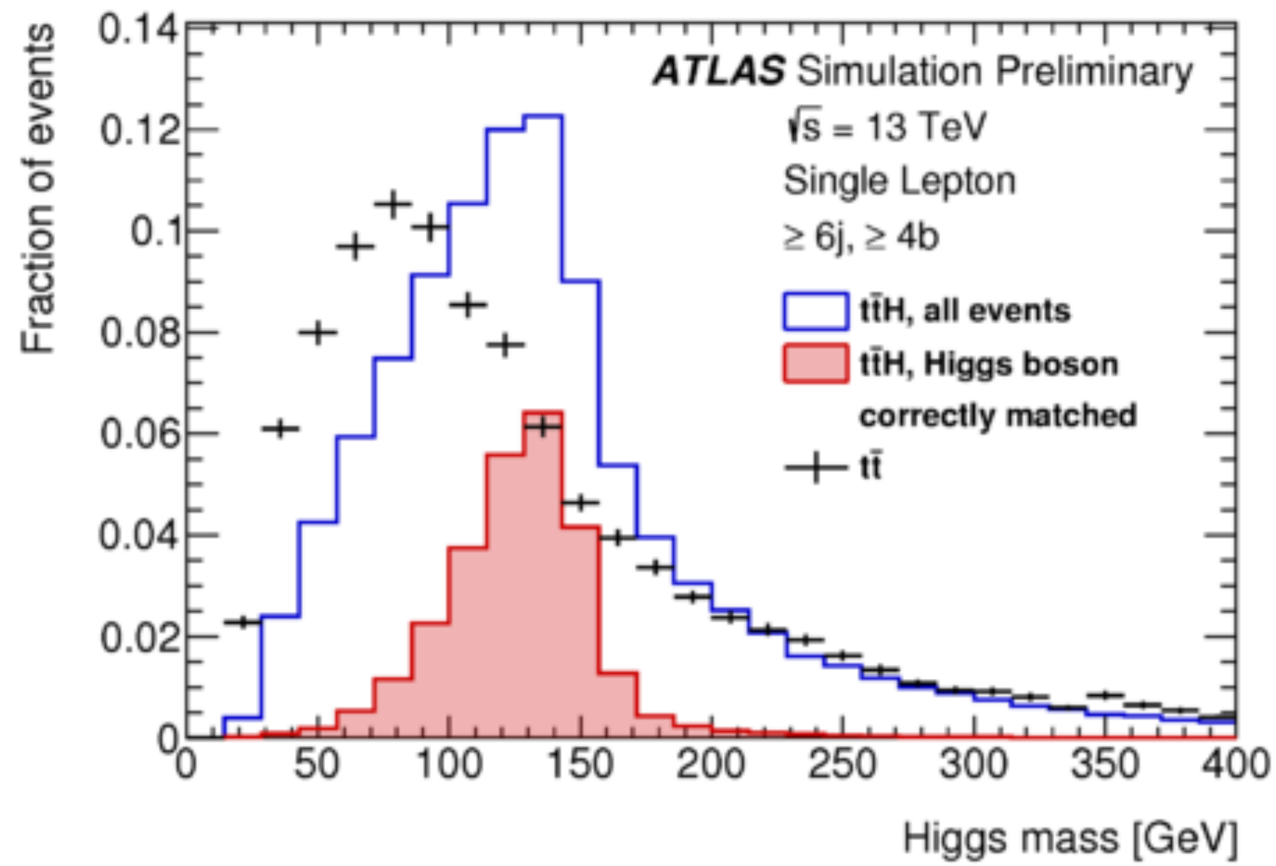


- Data overshoots MC at high b-tag multiplicity
- $t\bar{t}$+HF ($t\bar{t}$+ ≥ 1b/c-jet) component is underestimated in MC with large theoretical uncertainties

Reconstruction of Higgs and Top

- A two-stage multivariate technique (MVA) is used to separate the signal from the background
- First **Reconstruction BDT** to assign jets to top and Higgs decay products and reconstruct the signal





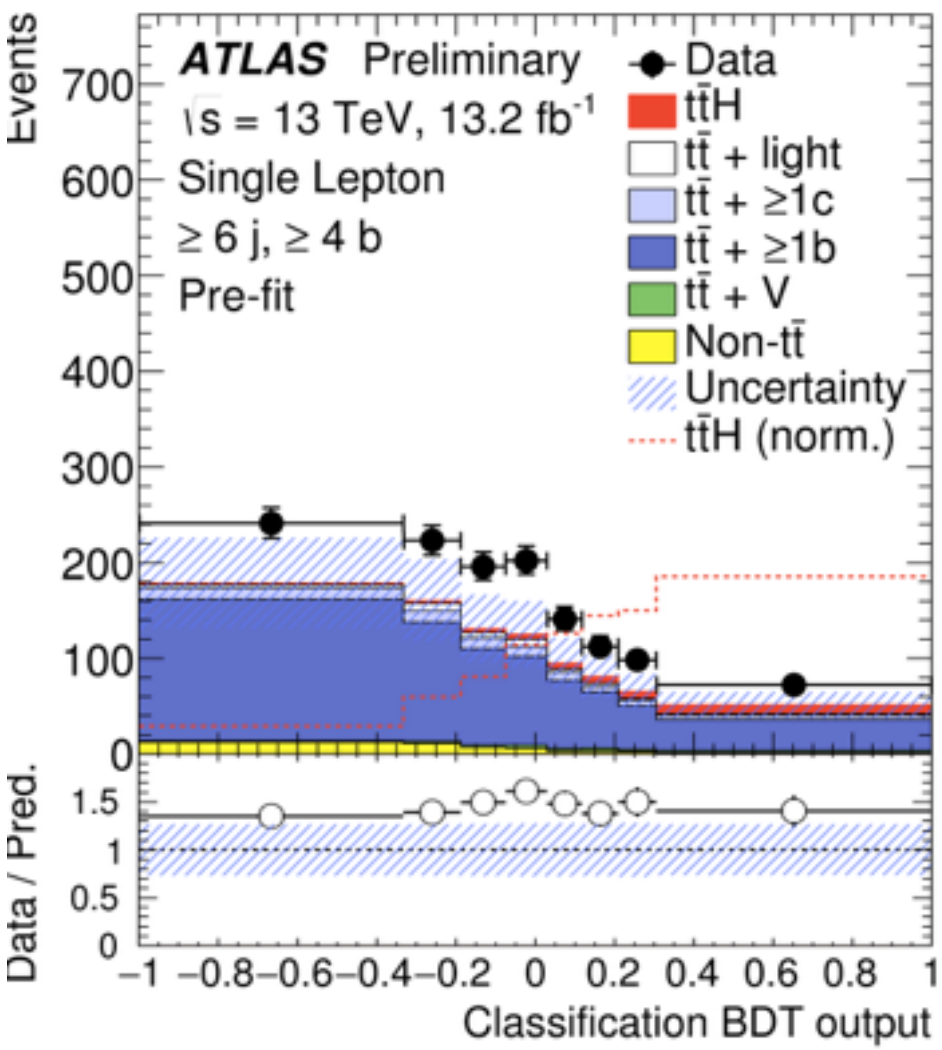
The reconstructed Higgs boson invariant mass

Event Classification

- **Second step**
  - Output of this is used with other variables as input to **classification BDT** : classify events as more signal- like or background- like

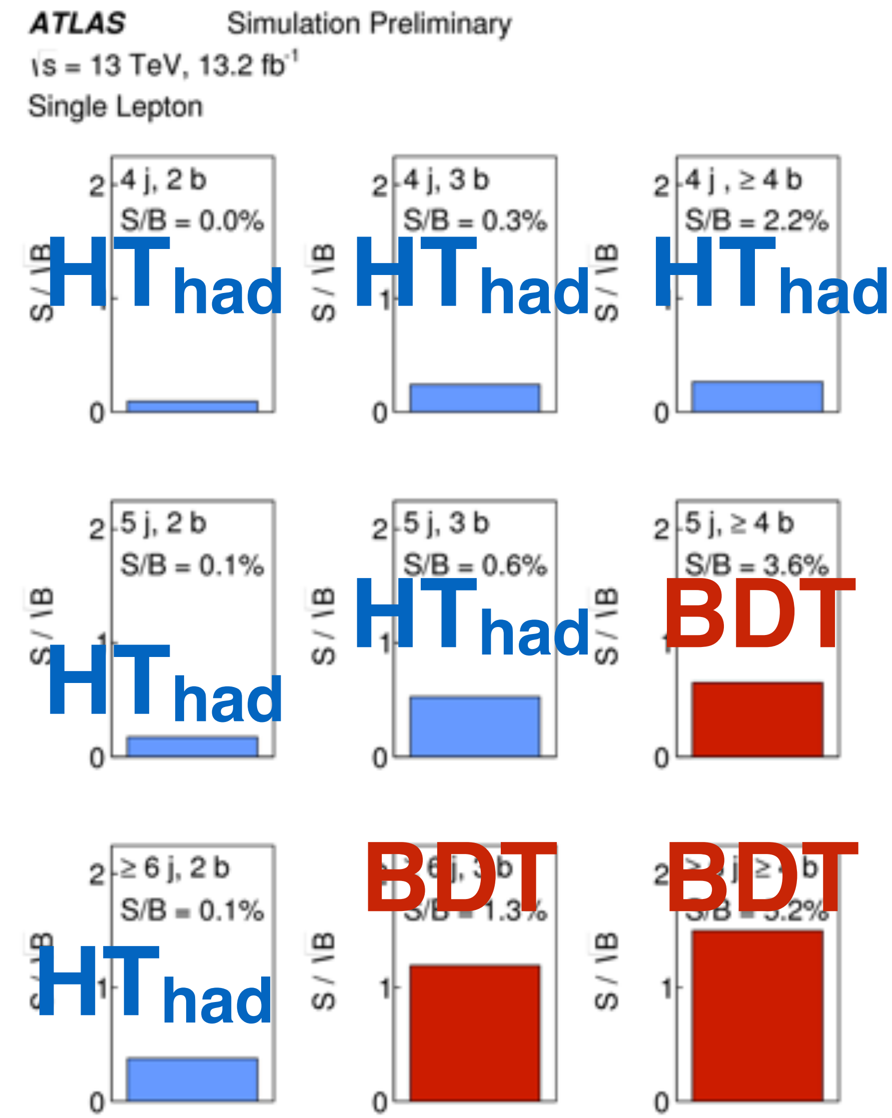
Signal Region

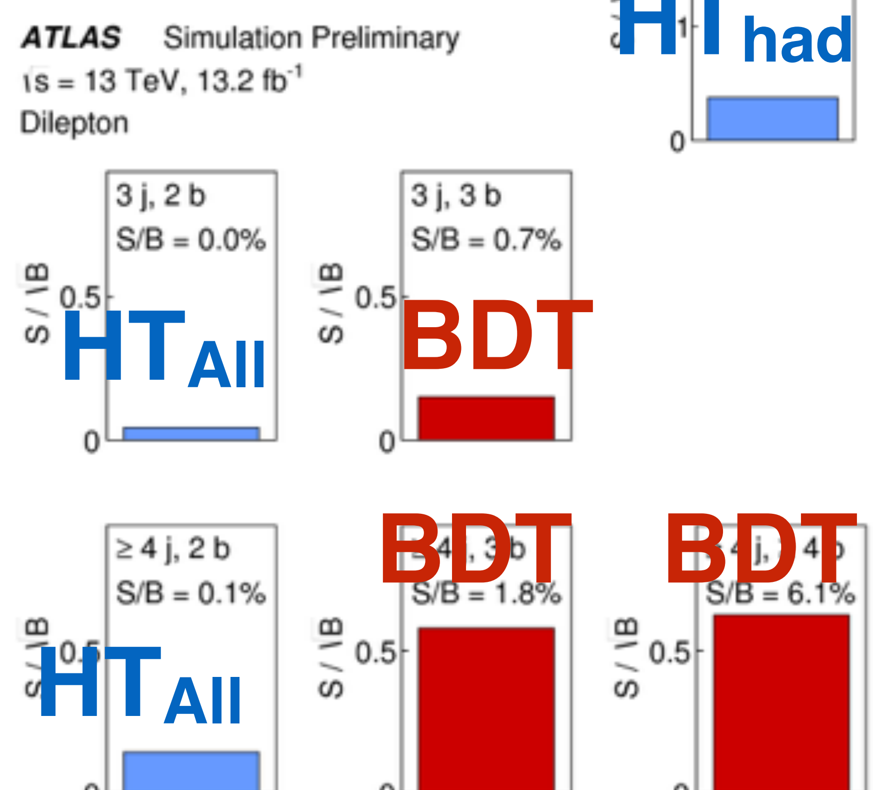
(≥ 6jets, ≥4b-tagged)



Fit input variables

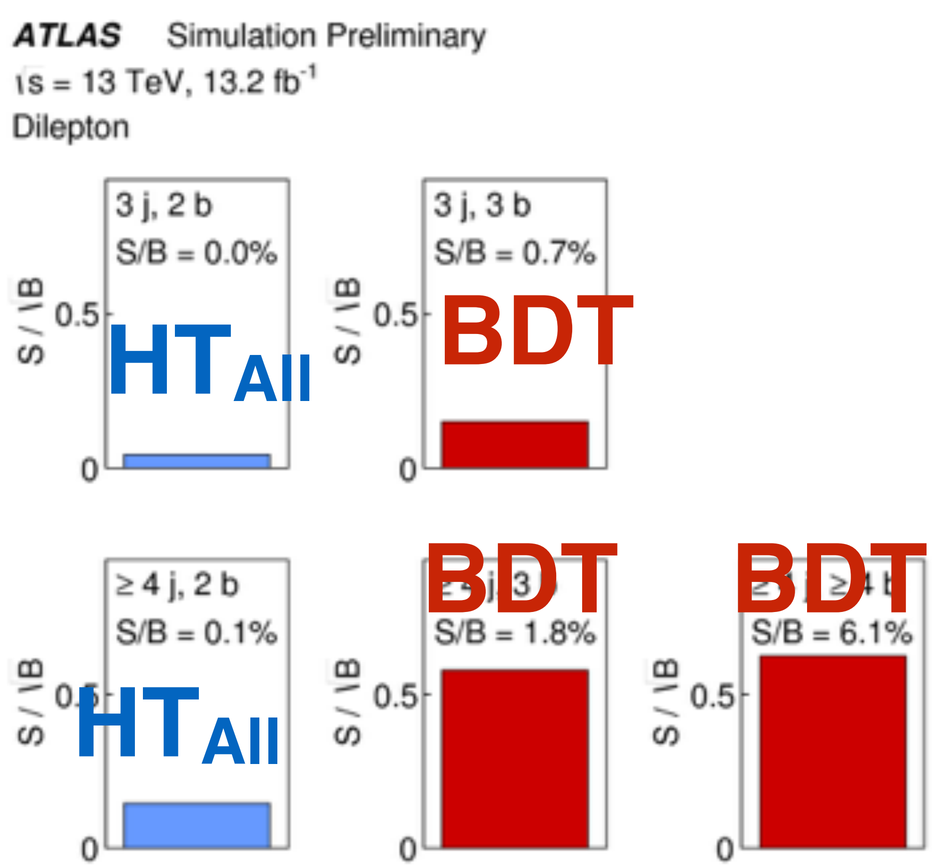
- Fit is performed in all 14 regions
- Control regions:
  - The scalar sum of all jets (and the leptons) H^{T}_{had} and (H^{T}_{all}) is used as a discriminating variable.
- Signal regions:
  - The output of the classification BDT is used as the variable in the fit
- Normalizations of $t\bar{t}+ \geq$ 1b/c-jet were free parameters in the final fit

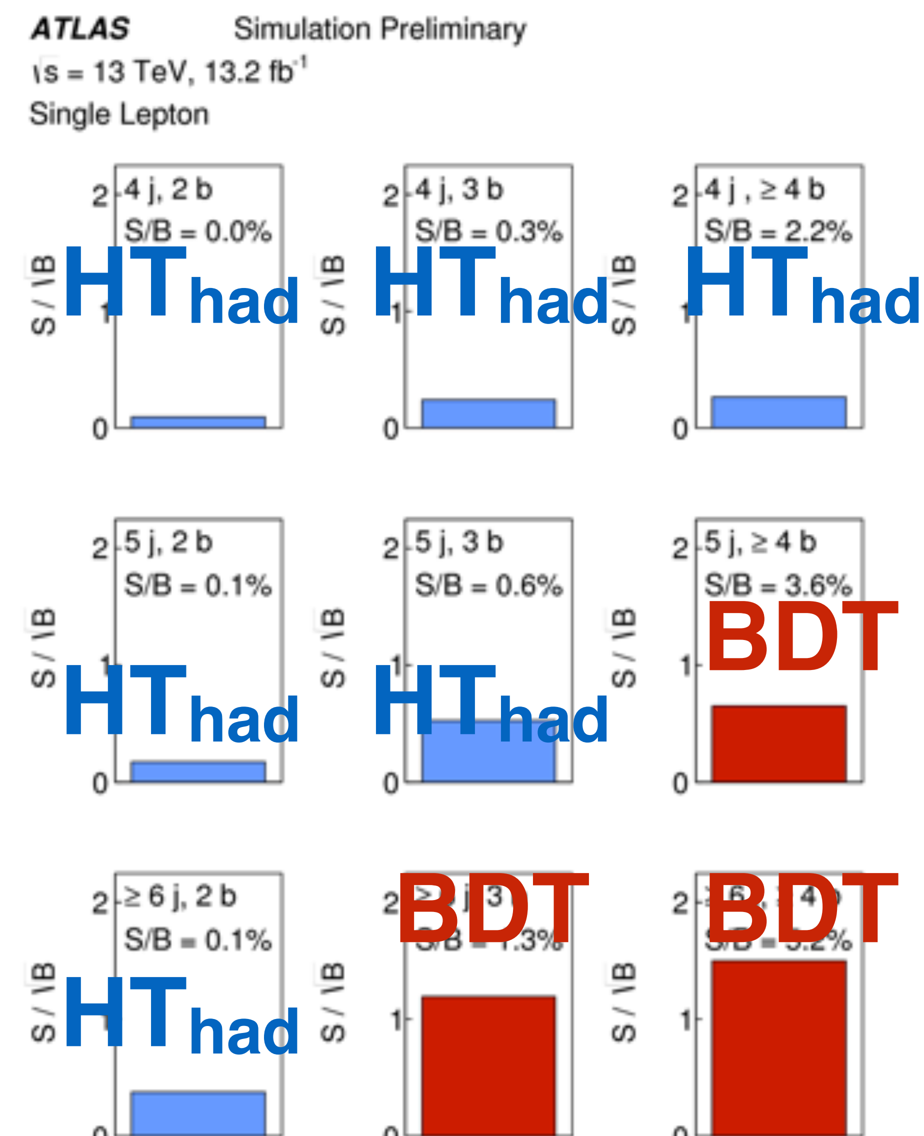




Fit input variables

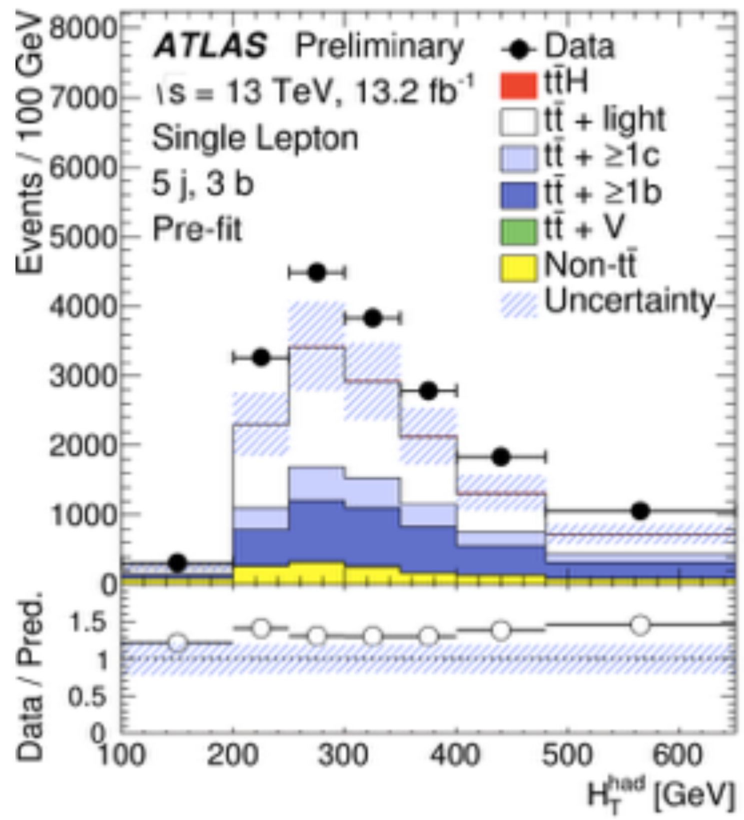
- The distributions of the discriminants from each of the regions are combined to test for the presence of a signal.
- The statistical analysis is based on a likelihood function $\mathcal{L}(\mu,\theta)$
  - The Poisson probability depends on the predicted number of events in each bin (function of the signal-strength $\mu = \sigma/\sigma_{SM}$)
  - "θ" The set of nuisance parameters which encodes the effect of the systematic uncertainties

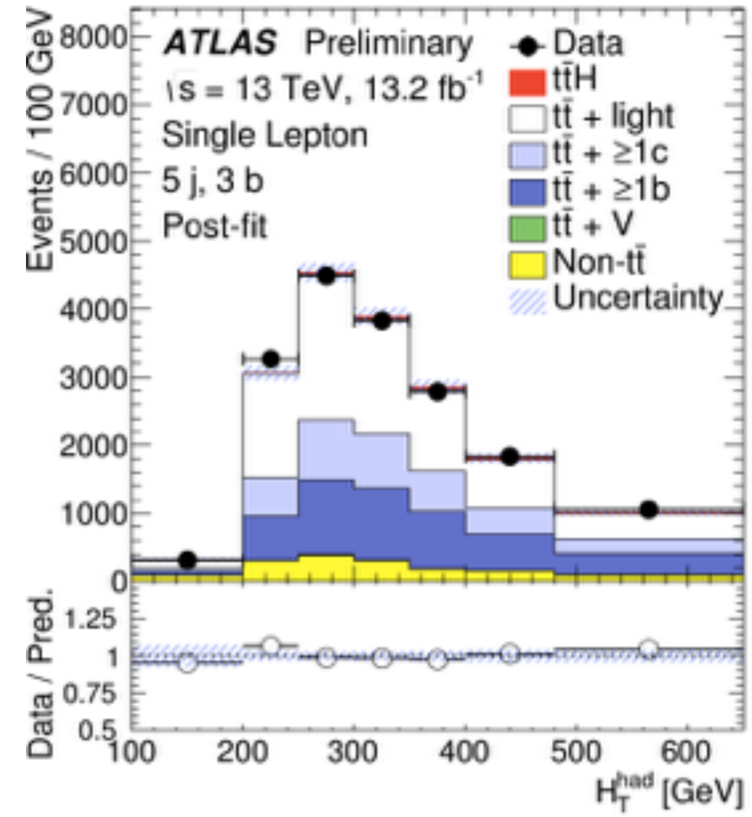




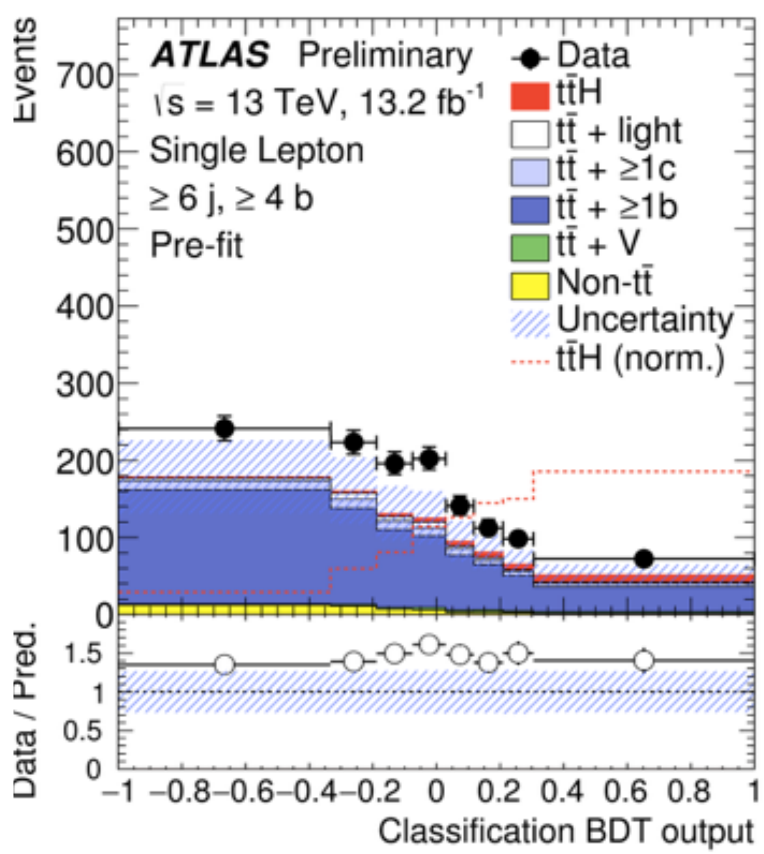
Perform a Fit

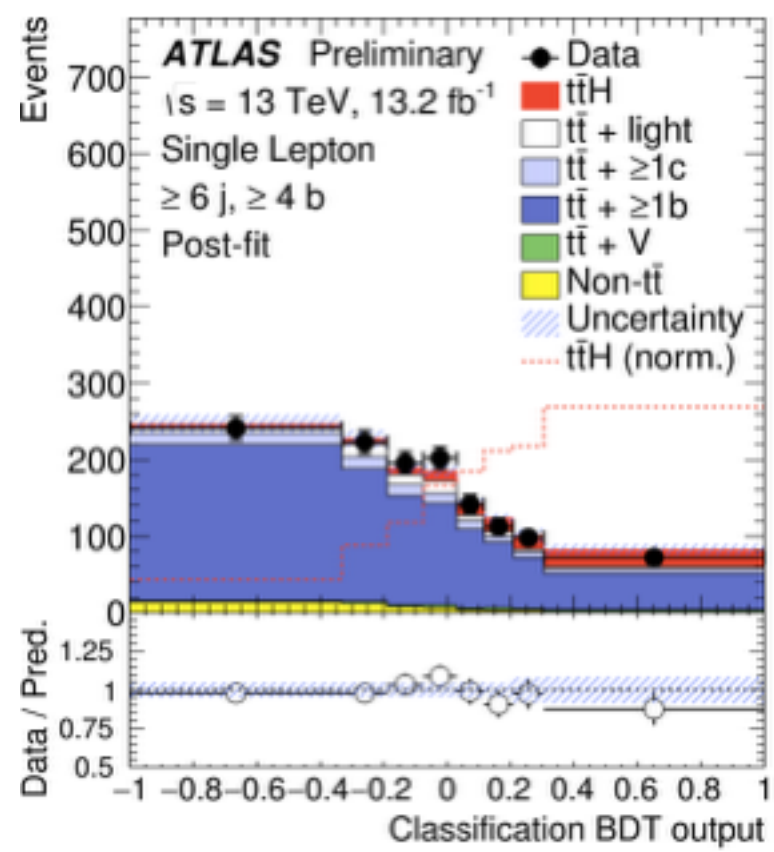
Distributions of the discriminating variable H_T^{had} (scalar sum of the p_T of all jets) in CR: 5jets, 3b-tagged





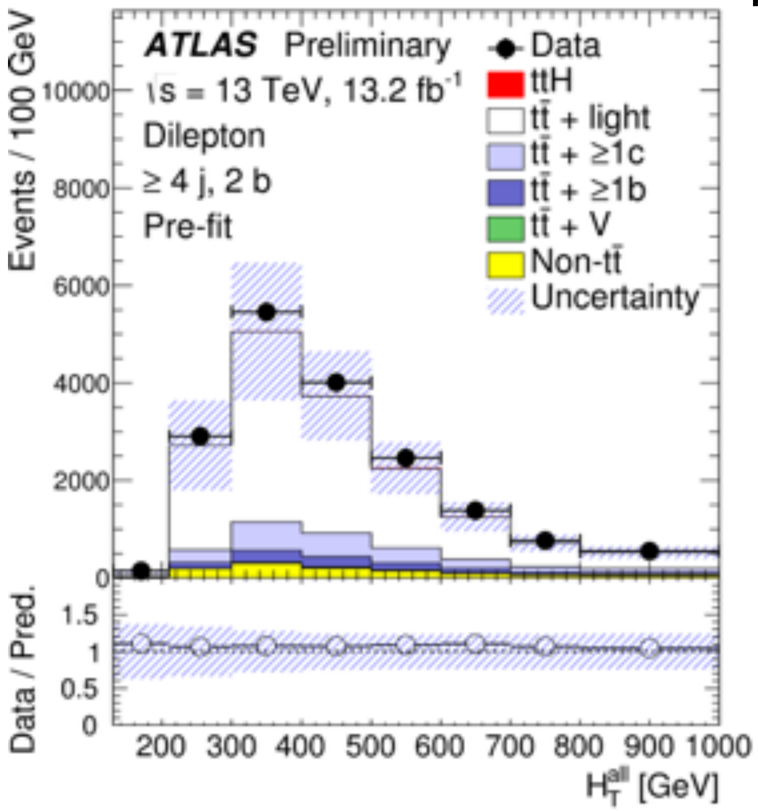
Comparison between data and prediction for the BDT discriminant in SR: ≥ 6jets, ≥4b-tagged

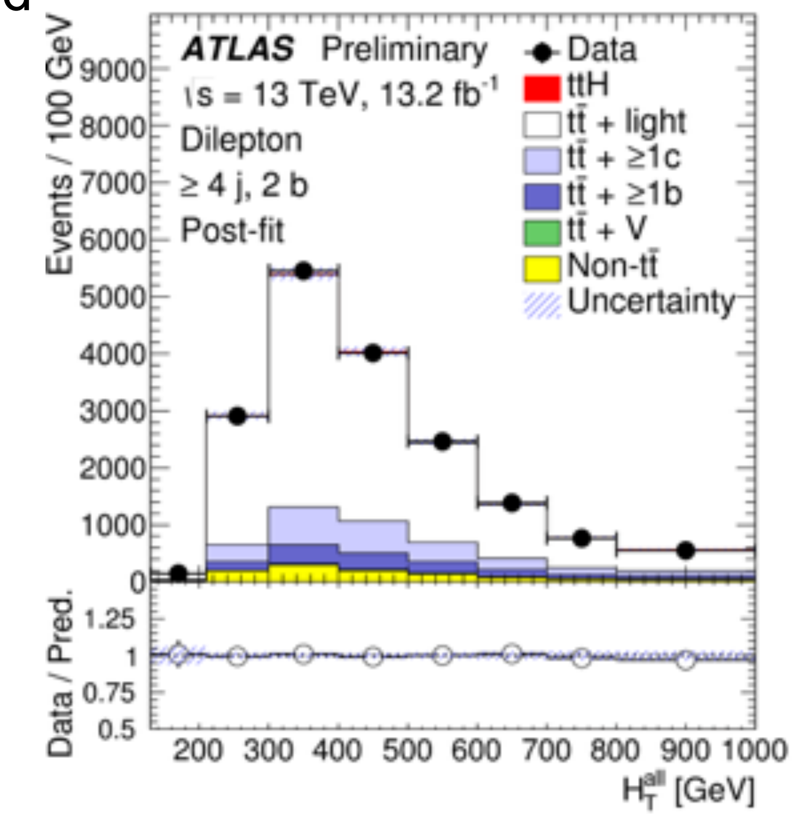




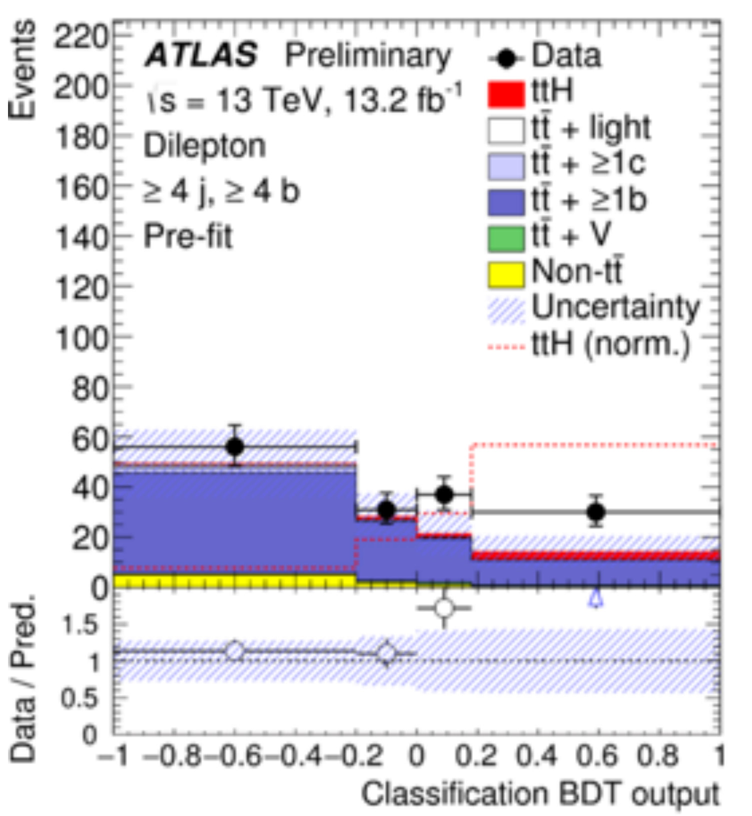
Perform a Fit

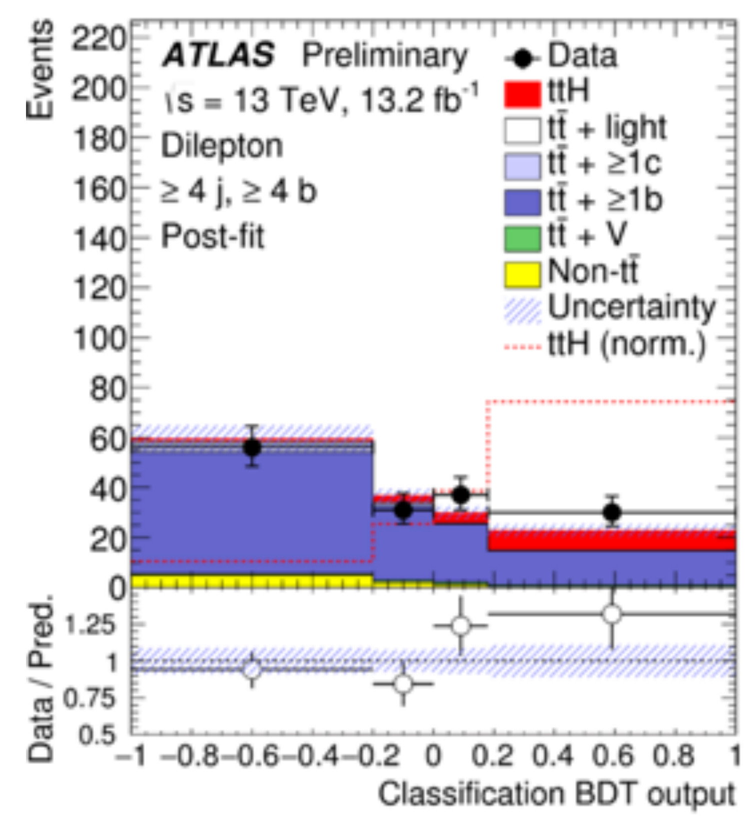
Distributions of the discriminating variable $H_T^{\,all}$ (scalar sum of the p_T of all jets and the leptons) in CR: ≥ 4jets, 2b-tagged





Comparison between data and prediction for the BDT discriminant in SR: ≥ 4jets, ≥4b-tagged



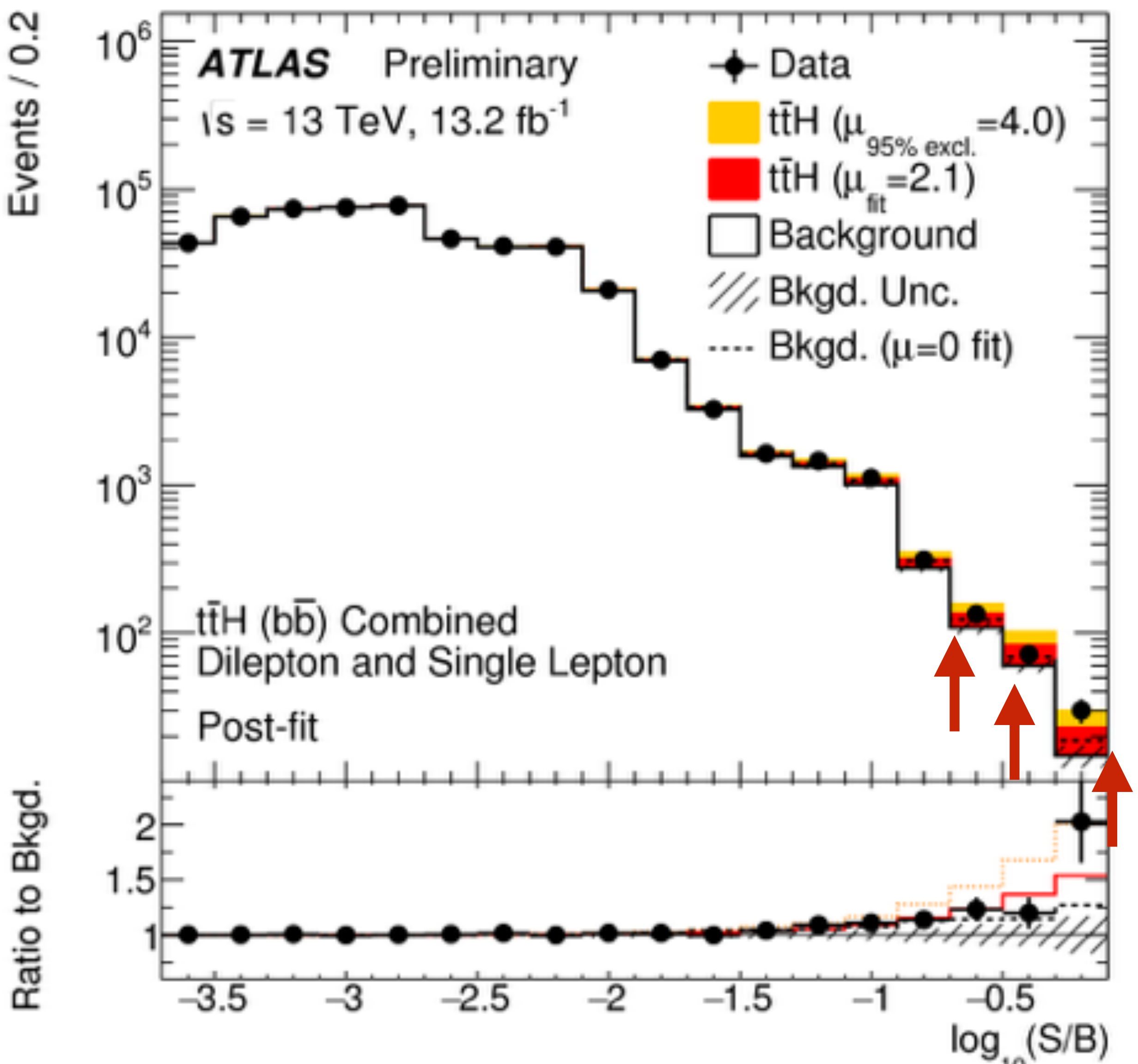


Results

- Data vs. prediction in all analysis bins, ranked by S/B

- Various sources of uncertainty were considered: choice of generator, parton shower and hadronisation model, PDF, and initial and final- state radiation

Post-fit yields of signal and total background per bin which are ordered by log(S/B)



Summary of the effects of the systematic uncertainties 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
Statistical uncertainty	+0.49	−0.49
Total uncertainty	+1.02	−0.89

Results

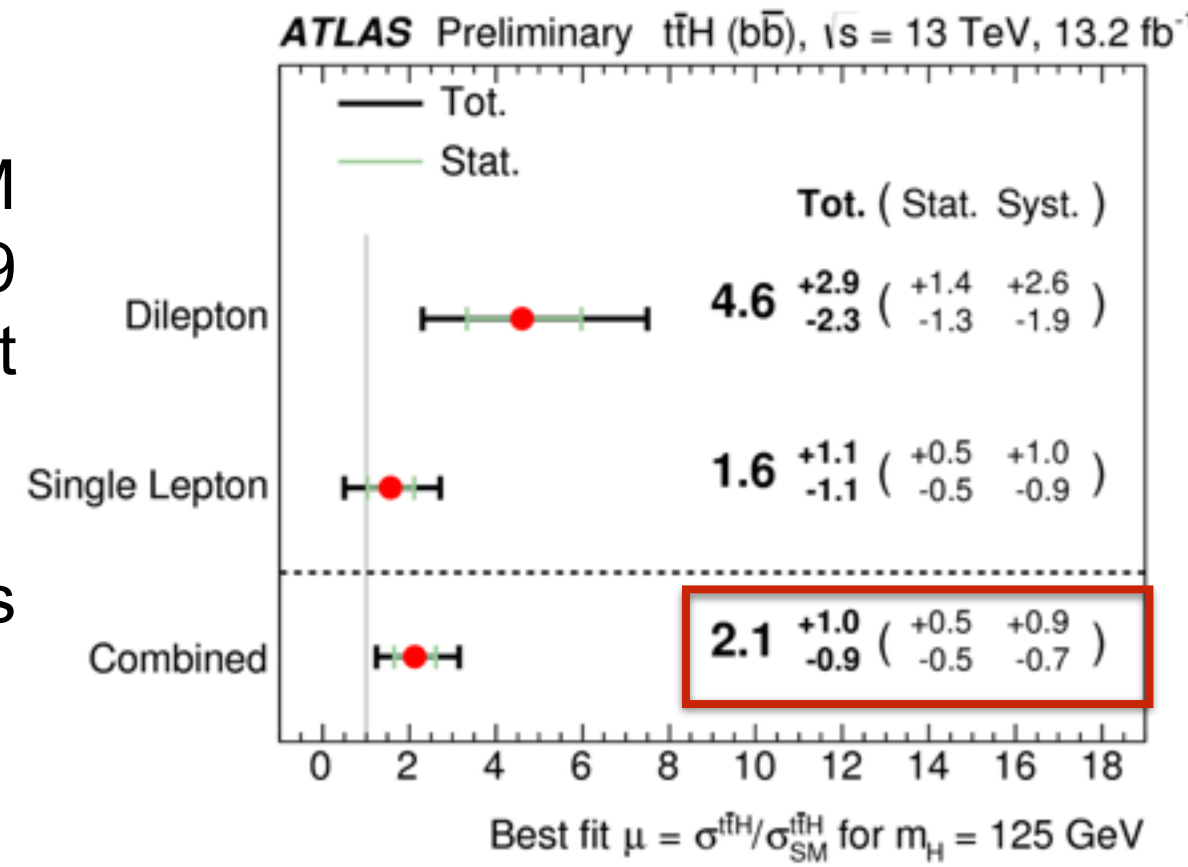
- The best combined fit was found to be

$$\mu_{ttH}=2.1^{+1.0}_{-0.9}$$

- This result is consistent with the SM prediction & signals larger than 3.9 times the SM prediction ruled out at 95% C.L.

- Stay tuned for the updated results with the full 2016 dataset!

Summary of the signal strength measurements in the individual channels and for the combination



Observed and expected 95% CL upper limits on the signal strength

	Observed	Expected ($\mu = 0$) Median	+/-1σ	+/-2σ	Expected ($\mu = 1$)
Dilepton	10.1	5.3	[3.8, 7.9]	[2.8, 12.6]	6.0
Single lepton	3.6	2.2	[1.6, 3.2]	[1.2, 4.7]	2.9
Combined	4.0	1.9	[1.4, 2.8]	[1.0, 4.2]	2.7

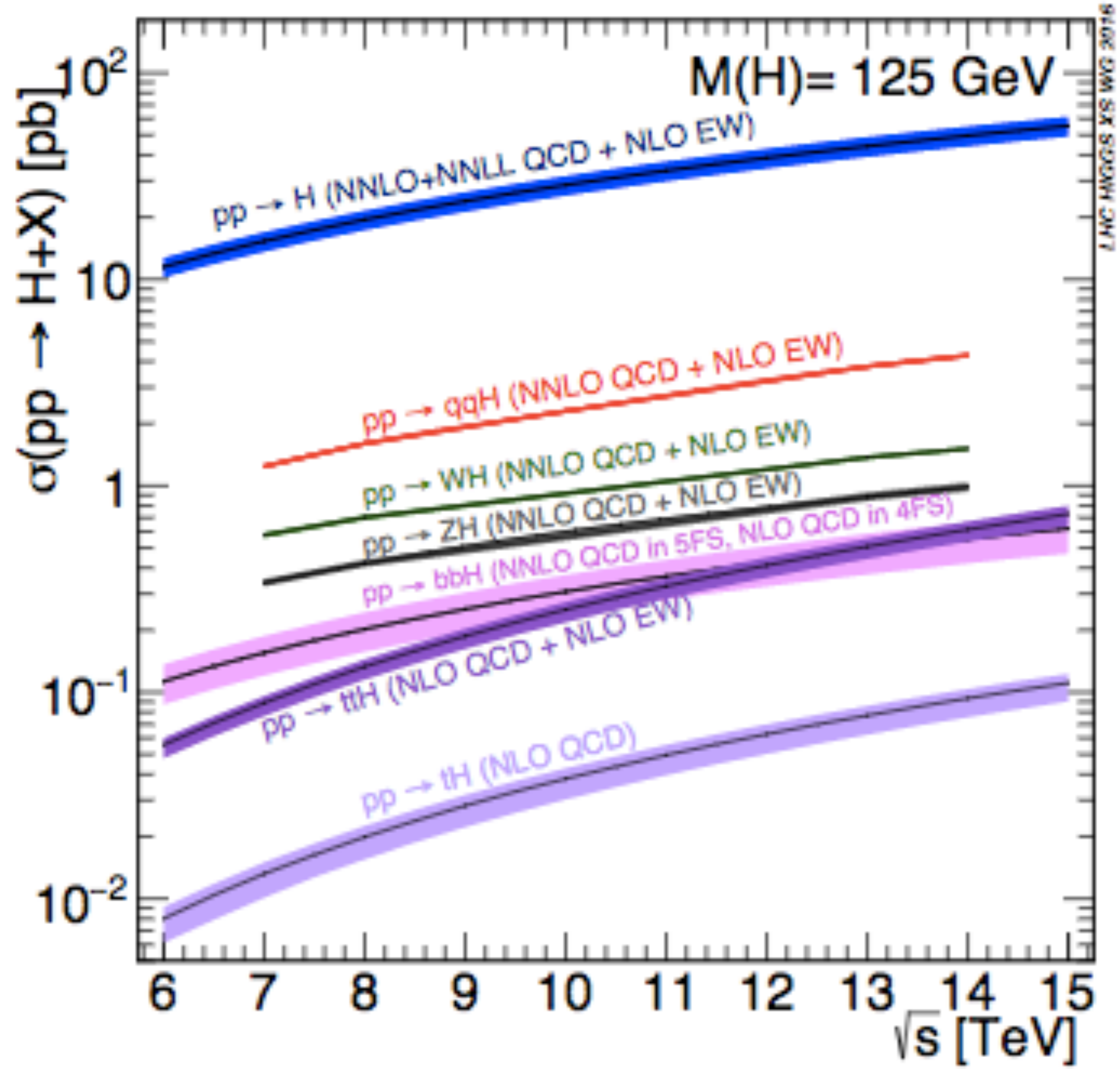
$$\mu_i = \frac{\sigma_i}{(\sigma_i)_{SM}}$$

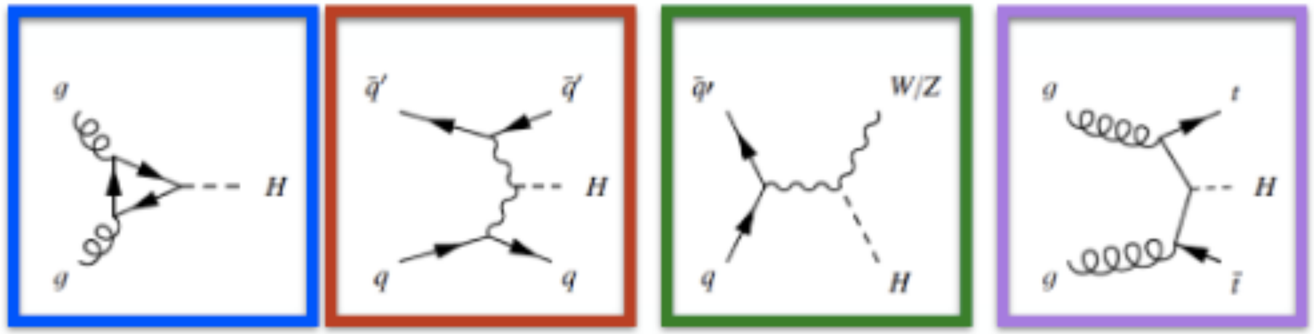
This is an exciting time at the LHC and more data to look at

Thank you

Backup

ttH Production at the LHC





Summary of the settings used for the simulation of the inclusive tt samples

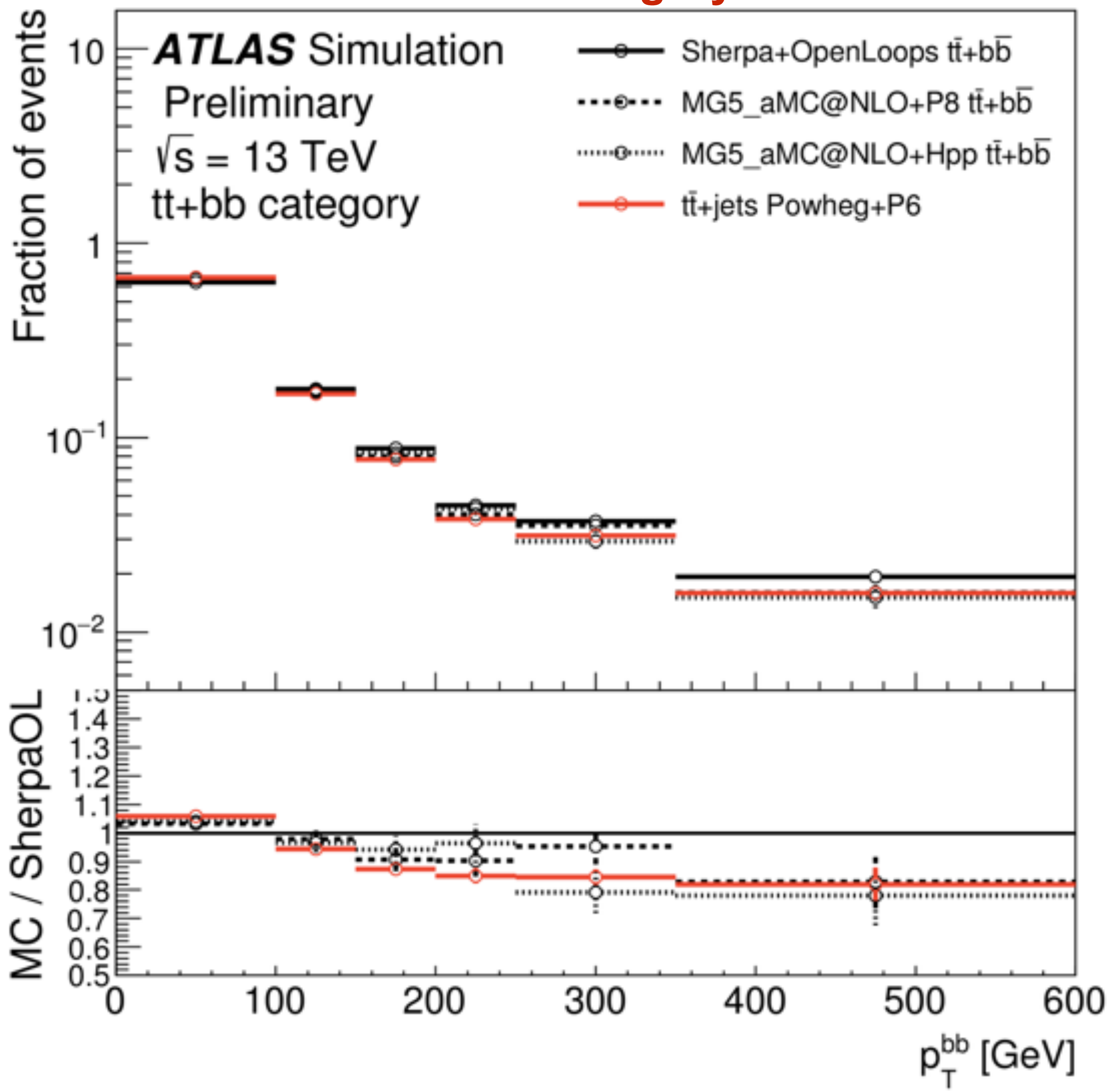
ME gen. PS/UE gen.	Powheg-Box Pythia 6.428	Powheg-Box Herwig++2.7.1	MG5_aMC Herwig++2.7.1	Powheg-Box Pythia 6.428	Powheg-Box Pythia 6.428
Ren. scale	$\sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$\sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$\sqrt{m_t^2 + \frac{1}{2}(p_{\mathrm{T,t}}^2 + p_{\mathrm{T,\bar{t}}}^2)}$	$\frac{1}{2} \cdot \sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$2 \cdot \sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$
Fact. scale	$\sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$\sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$\sqrt{m_t^2 + \frac{1}{2}(p_{\mathrm{T,t}}^2 + p_{\mathrm{T,\bar{t}}}^2)}$	$\frac{1}{2} \cdot \sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$	$2 \cdot \sqrt{m_t^2 + p_{\mathrm{T,t}}^2}$
hdamp	m_t	m_t	–	$2 \cdot m_t$	m_t
ME PDF	CT10	CT10	CT10	CT10	CT10
PS/UE PDF	CTEQ6L1	CTEQ6L1	CTEQ6L1	CTEQ6L1	CTEQ6L1
Tune	P2012	UE-EE5	UE-EE5	P2012 radHi	P2012 radLo

Summary of the settings used for the simulation of tt+bb 4FNLO samples

ME gen.	MG5_aMC	MG5_aMC	SherpaOL
PS/UE gen.	Herwig++ 2.7.1	Pythia 8.210	Sherpa
Renorm. scale	μ_{CMMPS}	μ_{CMMPS}	μ_{CMMPS}
Fact. scale	$H_{\mathrm{T}}/2$	$H_{\mathrm{T}}/2$	$H_{\mathrm{T}}/2$
Resumm. scale	$f_{\mathrm{Q}}\sqrt{\hat{s}}$	$f_{\mathrm{Q}}\sqrt{\hat{s}}$	$H_{\mathrm{T}}/2$
ME PDF	NNPDF3.0 4F	NNPDF3.0 4F	CT10 4F
PS/UE PDF	CTEQ6L1	NNPDF2.3	
Tune	UE-EE-5	A14	Author's tune

tt+ jets background modelling

Predicted cross-sections for the tt +bb sub-category



Systematic uncertainties

Summary of the effects of the systematic uncertainties 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
Statistical uncertainty	+0.49	−0.49
Total uncertainty	+1.02	−0.89

Systematic uncertainties

Systematic source	How evaluated	$t\bar{t}$ categories
$t\bar{t}$ cross-section	$\pm 6\%$	All, correlated
NLO generator (*residual*)	Powheg-Box + Herwig++ vs. MG5_aMC + Herwig++	All, uncorrelated
Radiation (*residual*)	Variations of μ_R, μ_F, and *hdamp*	All, uncorrelated
PS & hadronisation (*residual*)	Powheg-Box + Pythia 6 vs. Powheg-Box + Herwig++	All, uncorrelated
NNLO top & $t\bar{t}$ p_T	Maximum variation from any NLO prediction	$t\bar{t}+\geq 1c$, $t\bar{t}$ +light, uncorr.
$t\bar{t}+b\bar{b}$ NLO generator *reweighting*	SherpaOL vs. MG5_aMC+ Pythia8	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ PS & hadronis. *reweighting*	MG5_aMC + Pythia8 vs. MG5_aMC + Herwig++	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ renorm. scale *reweighting*	Up or down a by factor of two	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ resumm. scale *reweighting*	Vary μ_Q from $H_T/2$ to μ_{CMMPS}	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ global scales *reweighting*	Set μ_Q, μ_R, and μ_F to μ_{CMMPS}	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ shower recoil *reweighting*	Alternative model scheme	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ PDF *reweighting*	CT10 vs. MSTW or NNPDF	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ MPI	Up or down by 50%	$t\bar{t}+\geq 1b$
$t\bar{t}+b\bar{b}$ FSR	Radiation variation samples	$t\bar{t}+\geq 1b$
$t\bar{t}+c\bar{c}$ ME calculation	MG5_aMC + Herwig++ inclusive vs. ME prediction	$t\bar{t}+\geq 1c$

BDT input variables

Variable	≥6j, ≥4b	≥6j, =3b	=5j, ≥4b
Topological information from $t\bar{t}$:			
leptonic Top mass	✓	✓	✓
hadroinc Top mass	✓	✓	–
incomplete hadronic Top mass	–	–	✓
hadronic W mass	✓	✓	–
(hadW+blepTop)_mass	✓	✓	–
(qhadW+blepTop)_mass	–	–	✓
(lepW+bhadTop)_mass	✓	✓	✓
ΔR(hadW, bhadTop)	✓	✓	–
ΔR(qhadW, bhadTop)	–	–	✓
ΔR(hadW, blepTop)	✓	✓	–
ΔR(qhadW, blepTop)	–	–	✓
ΔR(lep, blepTop)	✓	✓	✓
ΔR(lep, bhadTop)	✓	✓	✓
ΔR(blepTop, bhadTop)	✓	✓	✓
ΔR(qhadW, qhadW)	✓	✓	–
ΔR(bhadTop, q1hadW)	✓	✓	–
ΔR(bhadTop, q2hadW)	✓	✓	–
min_ΔR(bhadTop, qhadW)	✓	✓	–
ΔR(lep, blepTop) - min_ΔR(bhadTop, qhadW)	✓	✓	–
ΔR(lep, blepTop)-ΔR(bhadTop, qhadW)	–	–	✓
Topological information from Higgs :			
Higgs mass	✓	✓	✓
ΔR(b1Higgs, b2Higgs)	✓	✓	✓
ΔR(b1Higgs, lep)	✓	✓	✓
(Higgs+q1hadW)_mass	✓	✓	✓
ΔR(b1Higgs, bleptop)	–	✓	✓
ΔR(b1Higgs, bhadtop)	–	✓	✓

Reconstruction BDT input variables

Variable	≥4j, ≥4b	≥4j, 3b
Topological information from $t\bar{t}$:		
ΔR(bTop,lepTop)	✓	✓
ΔR(bAntitop,lepAntitop)	✓	✓
Mass(bTop,lepTop)	✓	✓
Mass(bAntitop,lepAntitop)	✓	✓
p_T(bTop,lepTop)	✓	✓
p_T(bAntitop,lepAntitop)	✓	✓
ΔR(bTop,bAntitop)	✓	✓
$\lvert\Delta R$(bTop,lepTop) - ΔR(bAntitop,lepAntitop)$\rvert$	✓	✓
$min\Delta R(b$ from top,lep)	✓	✓
$max\Delta R(b$ from top,lep)	✓	✓
Topological information from Higgs :		
Higgs Mass	✓	✓
$\Delta R(b_1$ Higgs,b_2 Higgs)	✓	✓
ΔR(Higgs, $t\bar{t}$)	✓	✓
$\lvert\Delta R(b_1$ Higgs,b_2 Higgs) - ΔR(bTop,bAntitop)$\rvert$	✓	✓
$min\Delta R$(Higgs,lep)	✓	✓
$max\Delta R$(Higgs,lep)	✓	✓
$min\Delta R$(Higgs, b from top)	✓	✓
$max\Delta R$(Higgs, b from top)	✓	✓

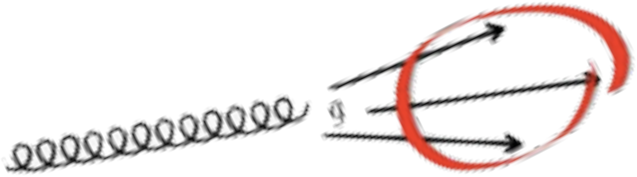
Multi-jet backgrounds

- Multi-jet Backgrounds are “fake and non-prompt leptons”
- Non- prompt leptons :
  - Semi-leptonic decays from c- and b-quarks
  - Photon conversion
  - Kaon decay (usually small)

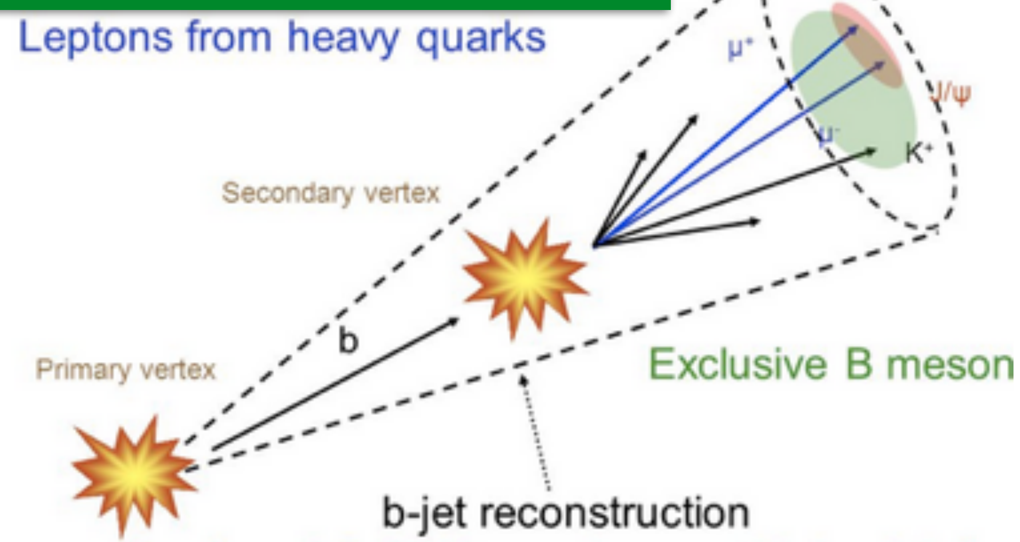
- Fake leptons
  - Jets can be misidentified as a reconstructed electron

- A fully data-driven estimate using Matrix Method in Single lepton channel & MC (corrected in SS control region) in di-lepton channel

Gluon jet fakes a lepton



non-prompt lepton



Results from CMS

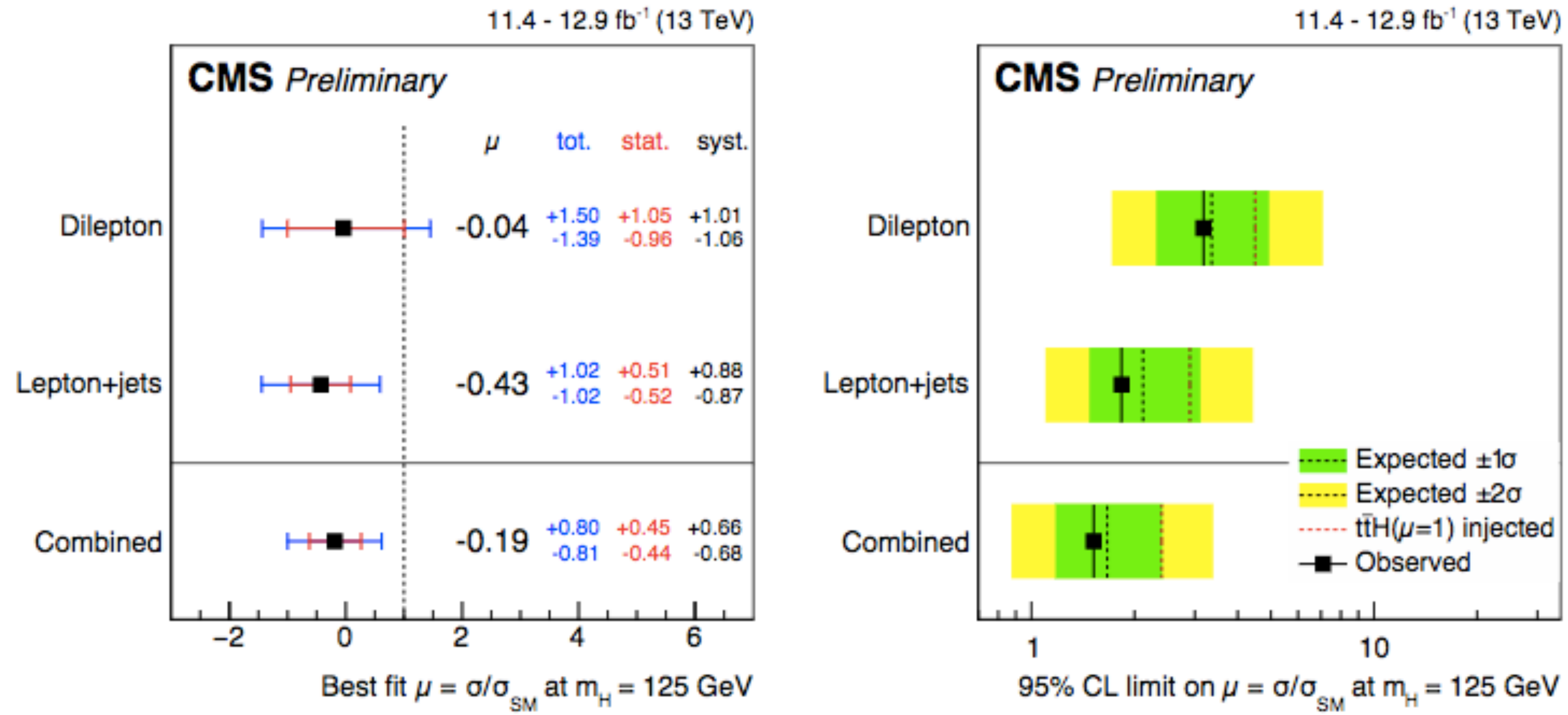


Figure 10: Best-fit values of the signal strength modifiers μ with their $\pm1\sigma$ confidence intervals, also split into their statistical and systematic components (left), and median expected and observed 95% CL upper limits on μ (right). The expected limits are displayed together with $\pm1\sigma$ and $\pm2\sigma$ confidence intervals. Also shown are the limits in case of an injected signal of $\mu = 1$.