Search for the associated production of the Higgs boson with a top quark pair at the ATLAS experiment

Silvia Biondi
Università "Alma Mater Studiorum" & INFN di Bologna

XVI Incontri di Fisica delle Alte Energie
19-21 Aprile 2017, Trieste

Presented analyses

reference

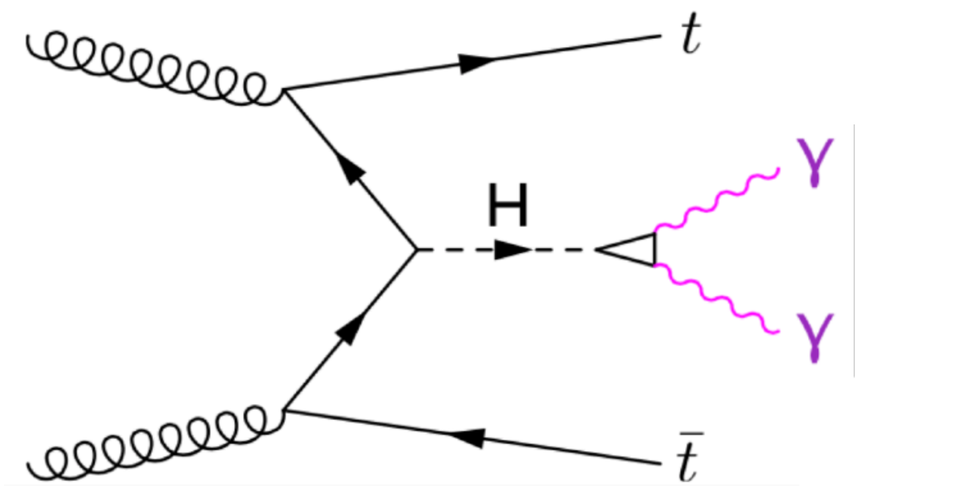
ttH(γγ)

BR ~ 0.23%

2 channels, depending on the tt system decay:

all-hadronic and (semi)leptonic

(covered in H→ γγ)

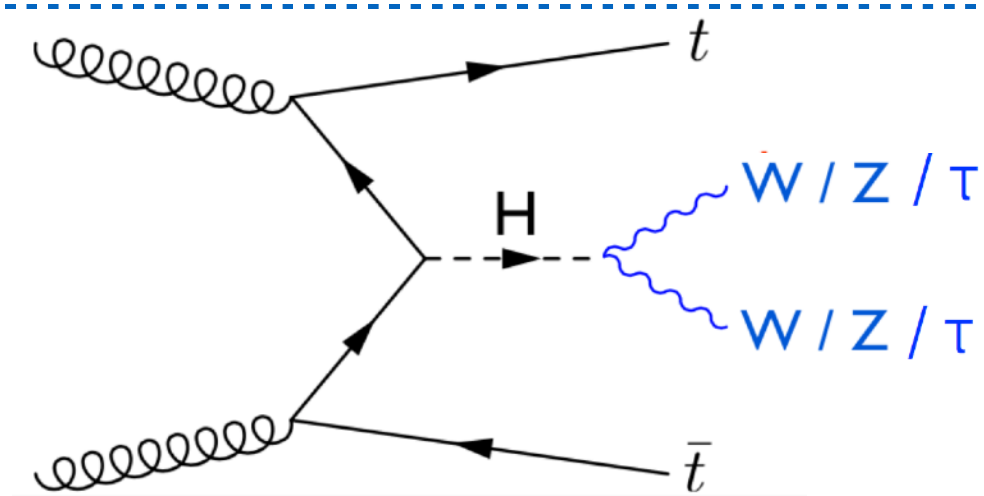


ttH(WW/ZZ/ττ)

BR ~ 30%

4 sub-channels, depending on the number, sign and flavour of leptons:

2lSS+0τ_{had}, 2lSS+1τ_{had}, 3l, 4l



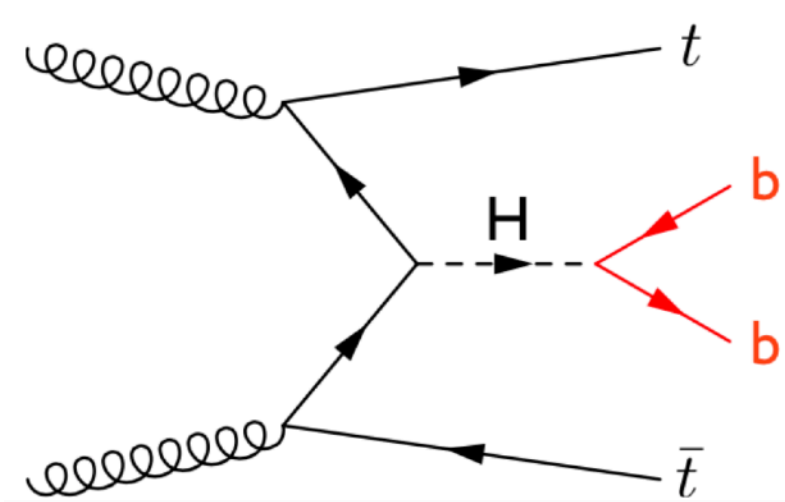
ttH(bb)

BR ~ 58%

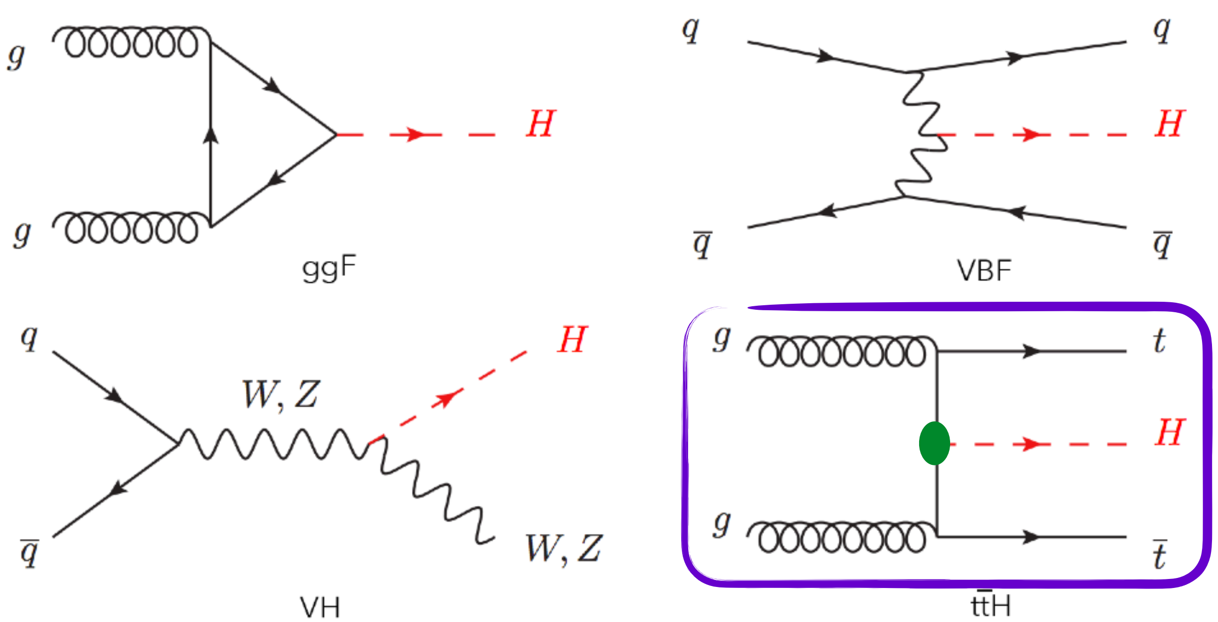
2 channels, depending on the tt system decay:

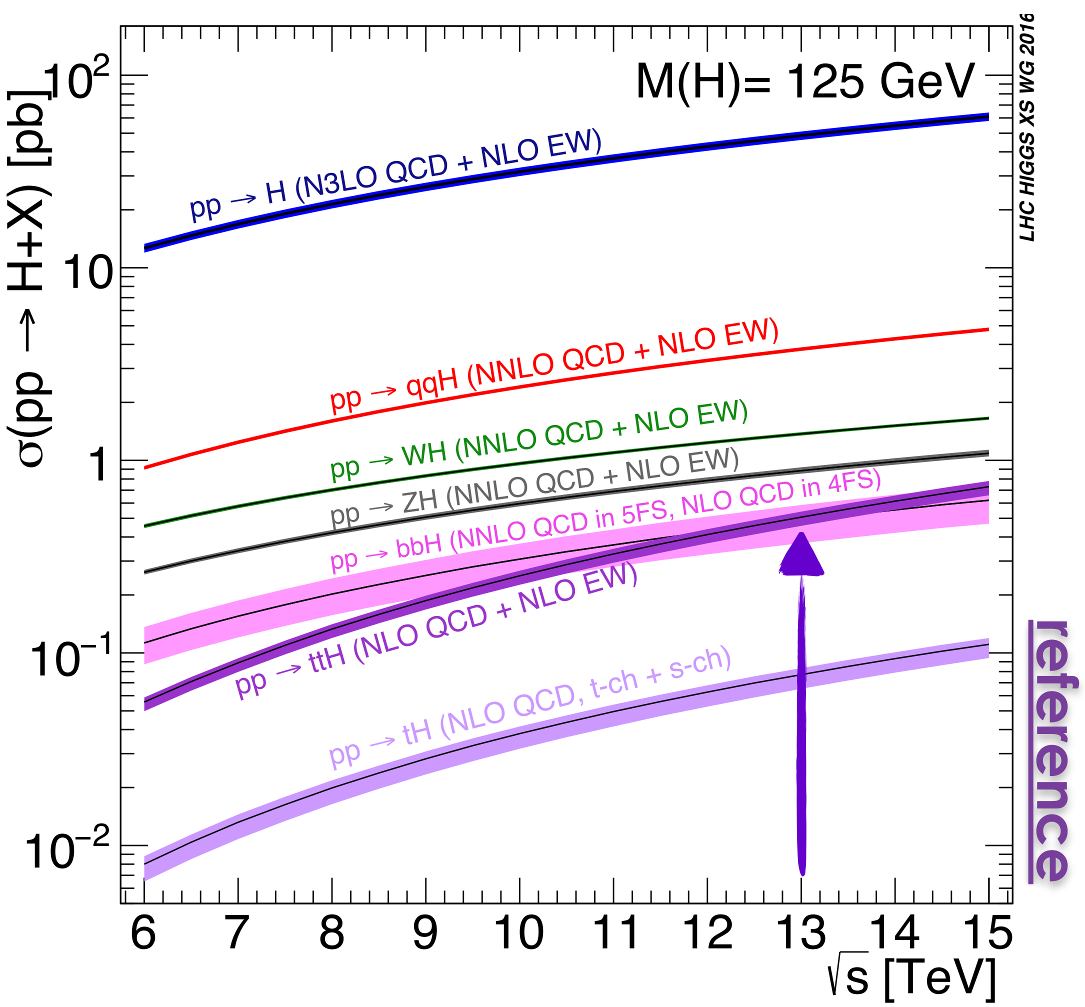
single-lepton - 6 control regions + 3 signal regions (defined by the jets and b-jets multiplicity)

dilepton - control regions + 2 signal regions



The associated Higgs production





Some motivations

- **indirect constraints** on the Top-Higgs Yukawa coupling from **ggF and H→ γγ** (through a loop);
- ttH production allows **direct access to Top-Higgs Yukawa coupling**;
- the **highest cross section increase as a function of energy** wrt other production modes;
- any deviation in the cross-section measurements could be an **hint of new physics**!

σ(**pp**) ~ 10^{11} pb

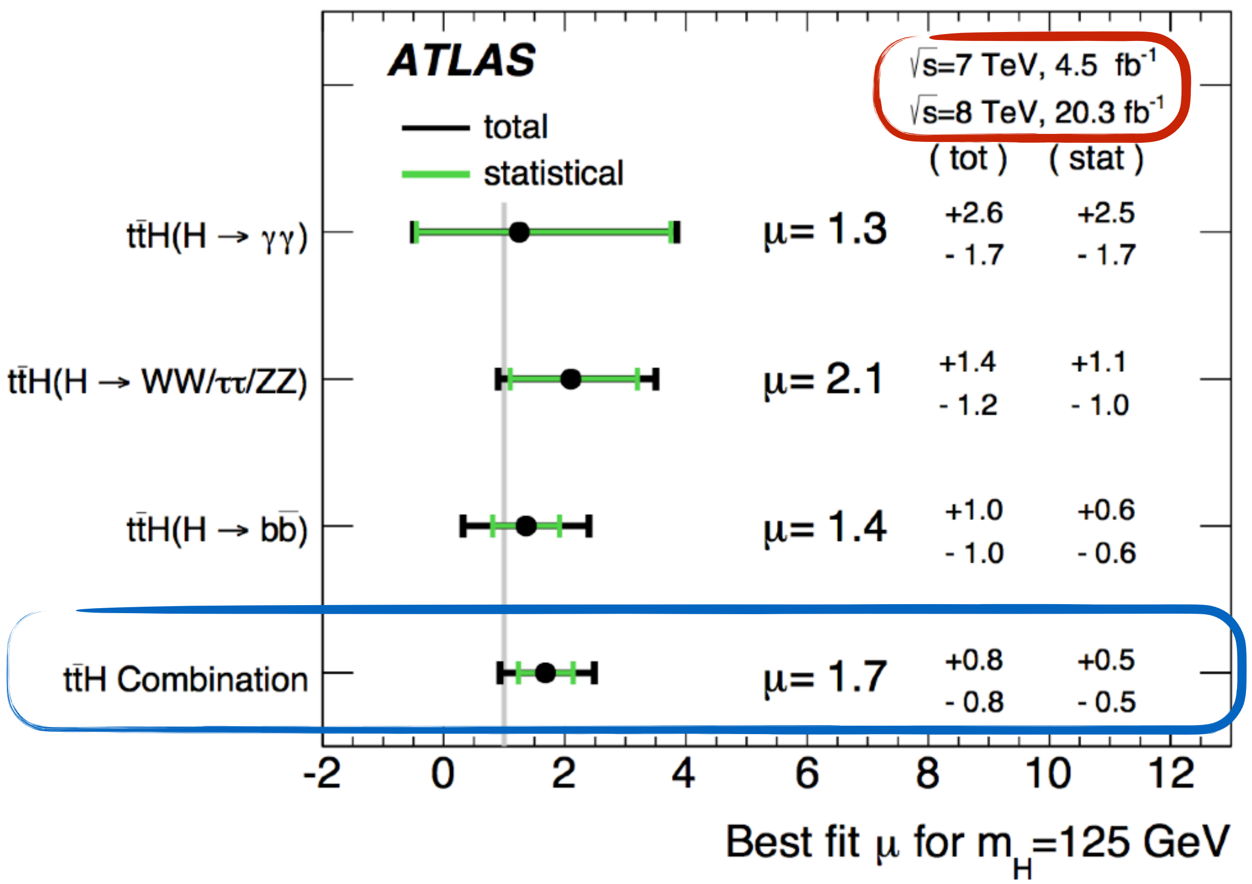
σ(pp->**H+X**) ~ 45 pb

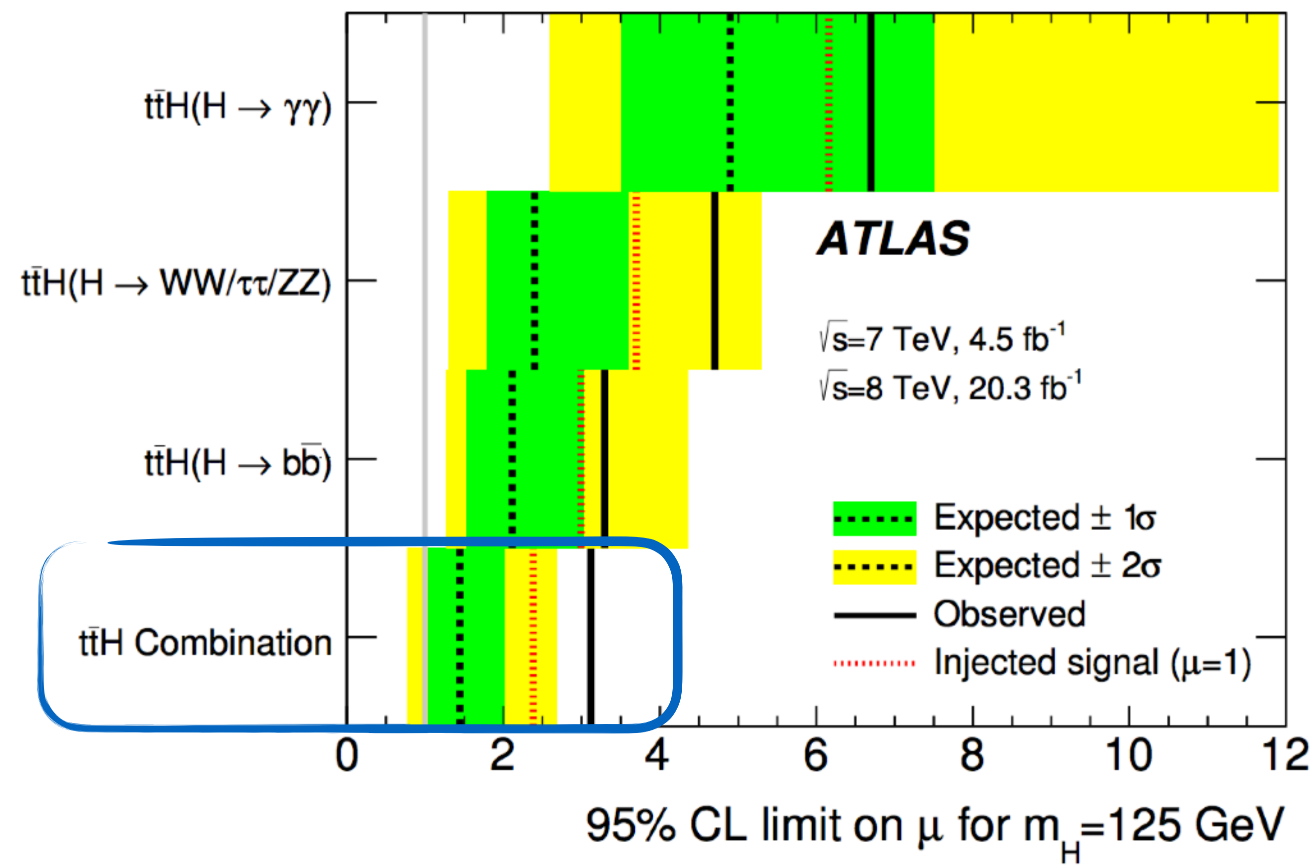
σ(pp->**ttH**) ~ 0.5 pb

@13 TeV

Run-1 results

JHEP05(2016)160





- Many channels sensitive to different final states, depending on the Higgs decay modes: **γγ**, **WW/ZZ/ττ** (multi-lepton) and **bb**;
- a signal strength **μ = 1.7 ± 0.8** has been measured;
- this corresponds to an **observed (expected) significance 2.3 σ (1.5 σ)**;
- **observed (expected) 95% C.L. limit on μ is 3.1 (1.4)**.

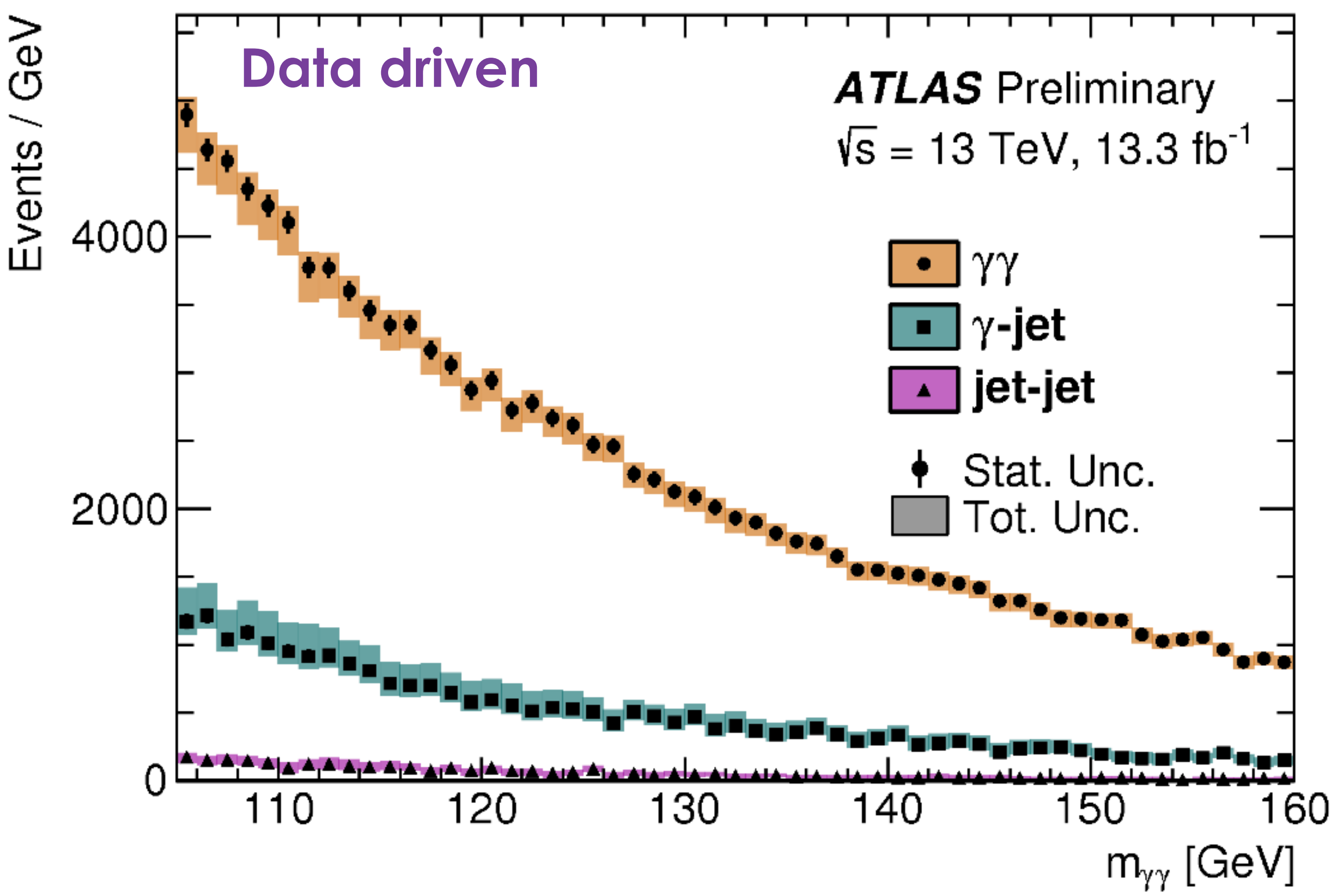
ttH(γγ)

BR ~ 0.23%

ATLAS-CONF-2016-067

- **Clear resonance peak** and low background;
- narrow peak in the di-photon invariant mass spectrum on top of a **smoothly falling background** in the **$m_{\gamma\gamma}$ distribution**;
- due to the narrow width of the Higgs boson, the shape of the distribution is governed by the **resolution of the measured photon energies**.

- To reject hadronic jet backgrounds, the **photon** candidates are required to be **isolated from any other activity in the calorimeter and the tracking detectors**;
- background estimation in **3 control regions**;
- reversing the isolation requirements of the photons.

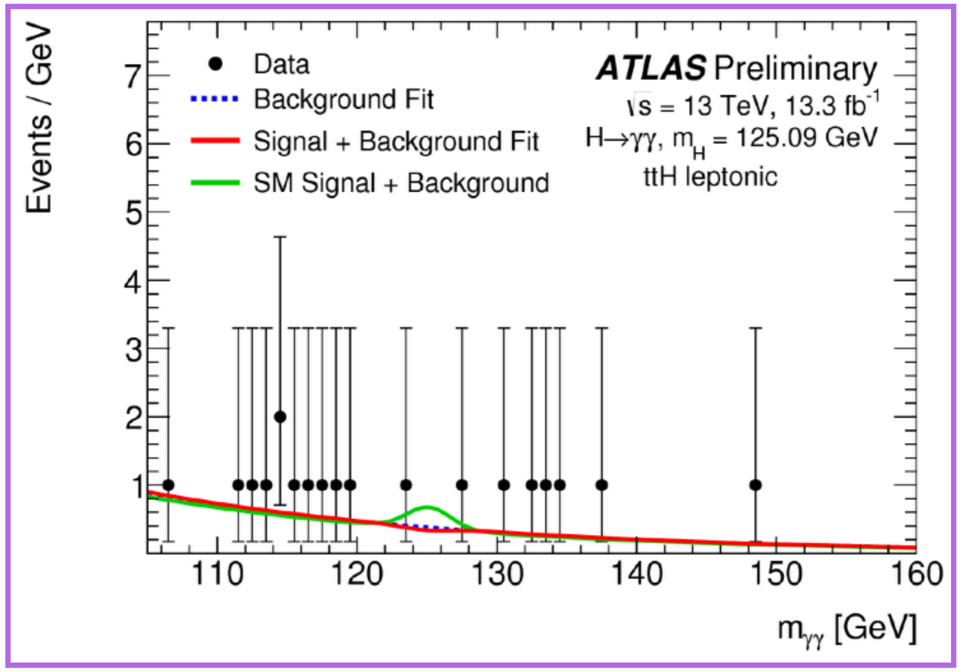


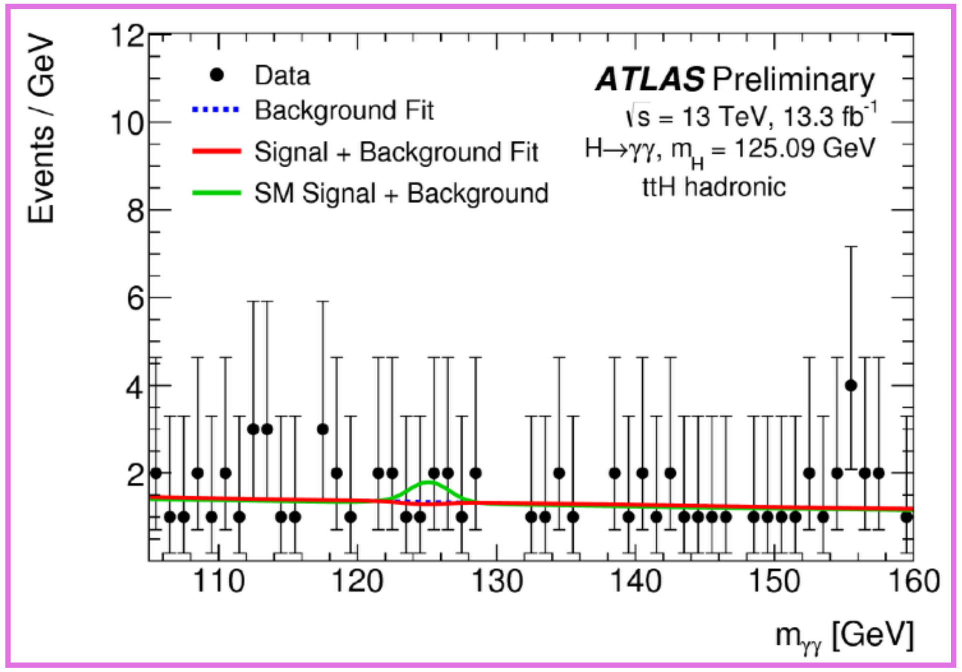
ttH (γγ)

ATLAS-CONF-2016-067

2 signal regions:

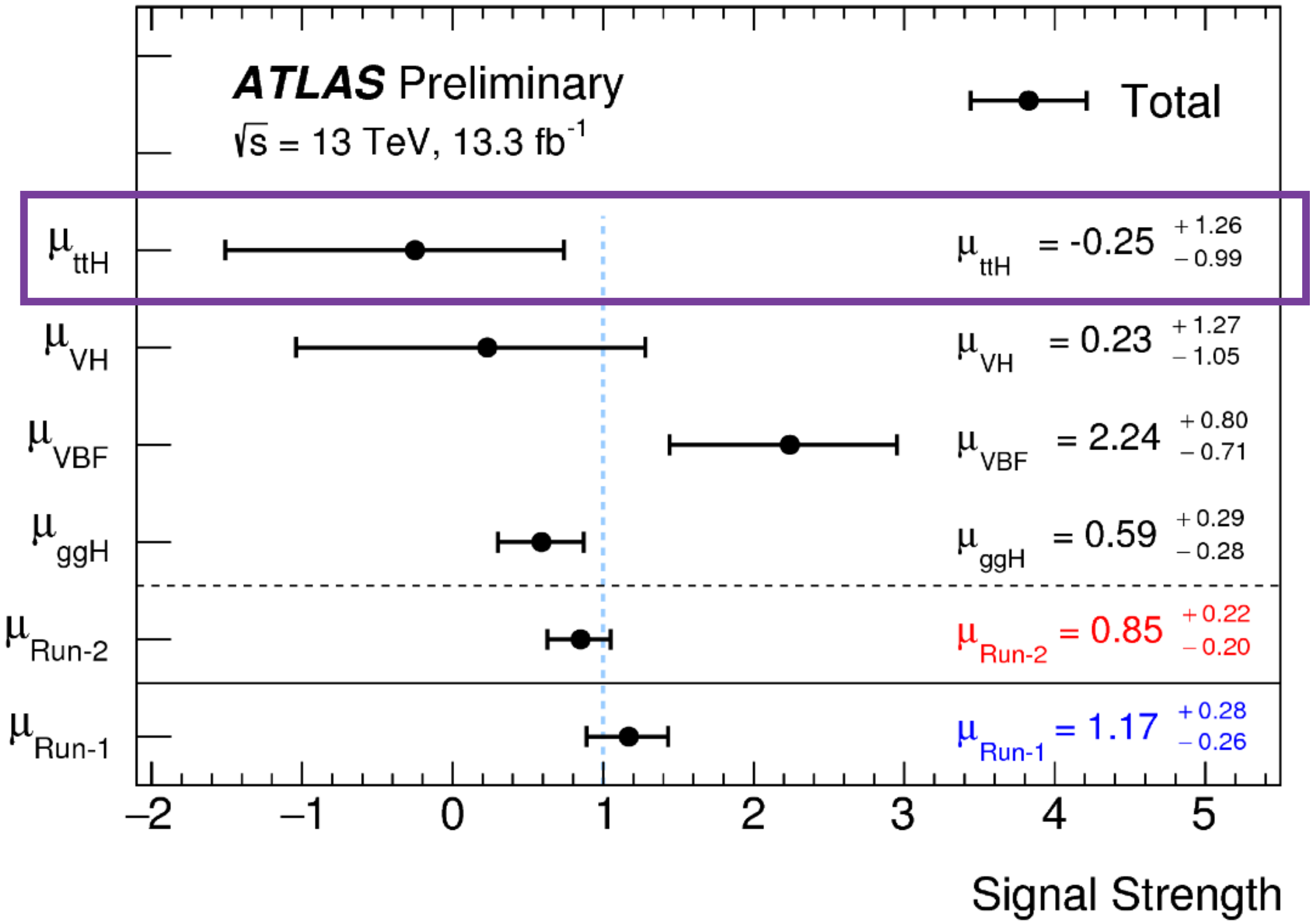
- **leptonic**: 2 isolated γ, ≥ 1 lepton, ≥ 2 jets, ≥ 2 b-jets or ≥ 1 b-jet AND E_T^{miss} > 20 GeV
- **hadronic**: 2 isolated γ, 0 leptons, ≥ 5 jets, ≥ 1 b-jet.





- Event selected and categorised in **regions enriched with different Higgs production mode processes**;
- **combined fit** performed to the categories to determine the signal strengths.

13.3 fb^{-1} @ 13 TeV



ttH(WW/ZZ/ττ)

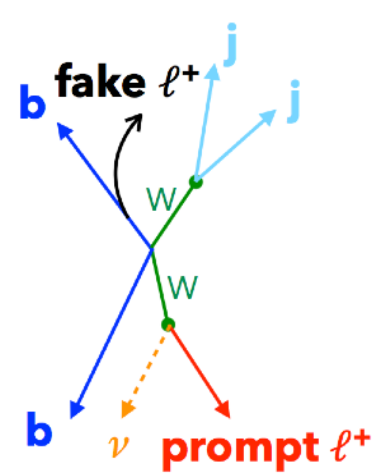
BR ~ 30%

ATLAS-CONF-2016-058

- **Cut-based** analysis;
- several regions depending on the **multiplicity and flavour of leptons in the final state**;
- ensuring a very small background contribution but difficult to estimate due to sensitivity to additional **ttW/Z (hard to control with data)**.

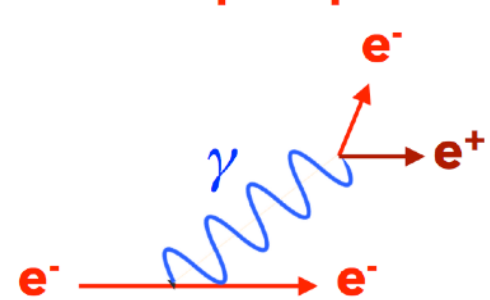
- Irreducible backgrounds: **prompt leptons** (decaying from boson or τ leptons), **ttV** is the main contribution;
- processes with one **non-prompt lepton and fake τ**: mainly from hadrons decay from tt events;

  Data driven

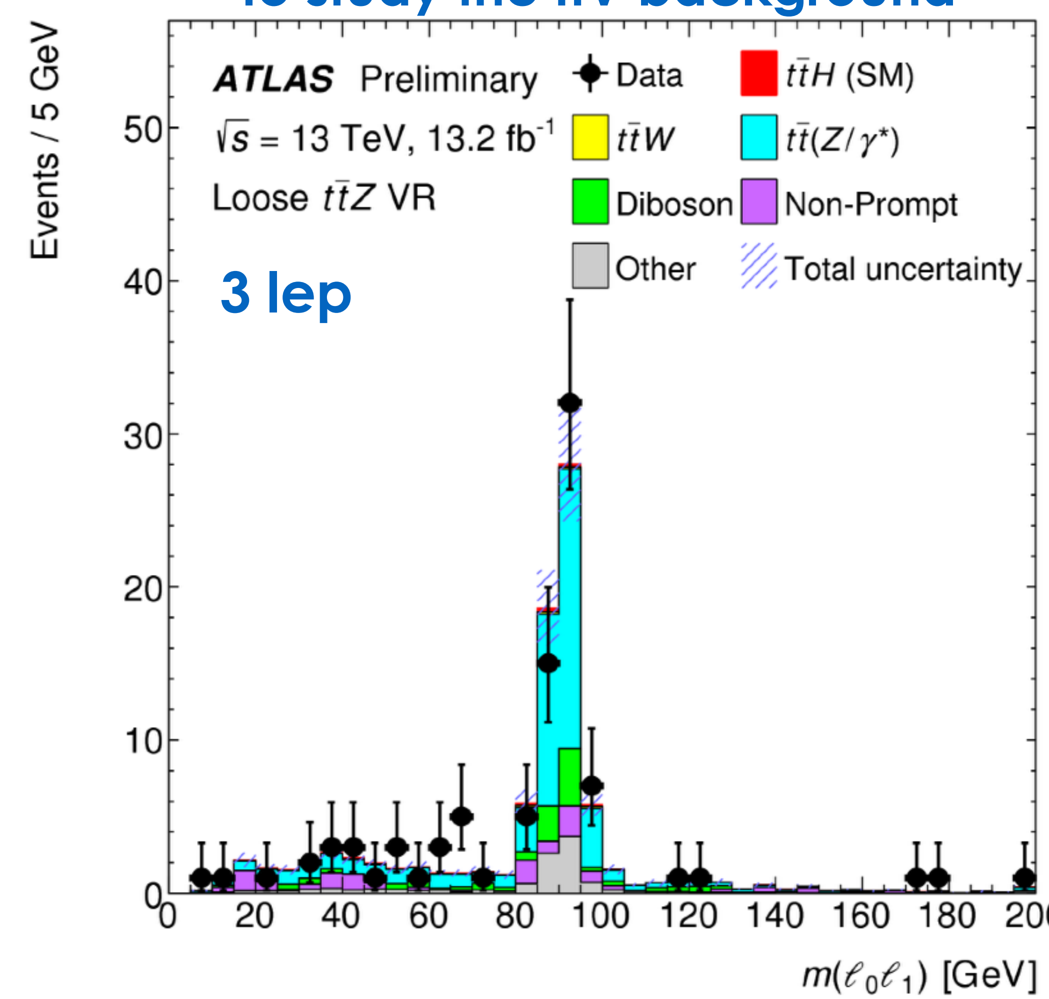


- processes with a **misidentified charged lepton** (especially 2l+1τ and 2l+0τ with electrons).

  Data driven



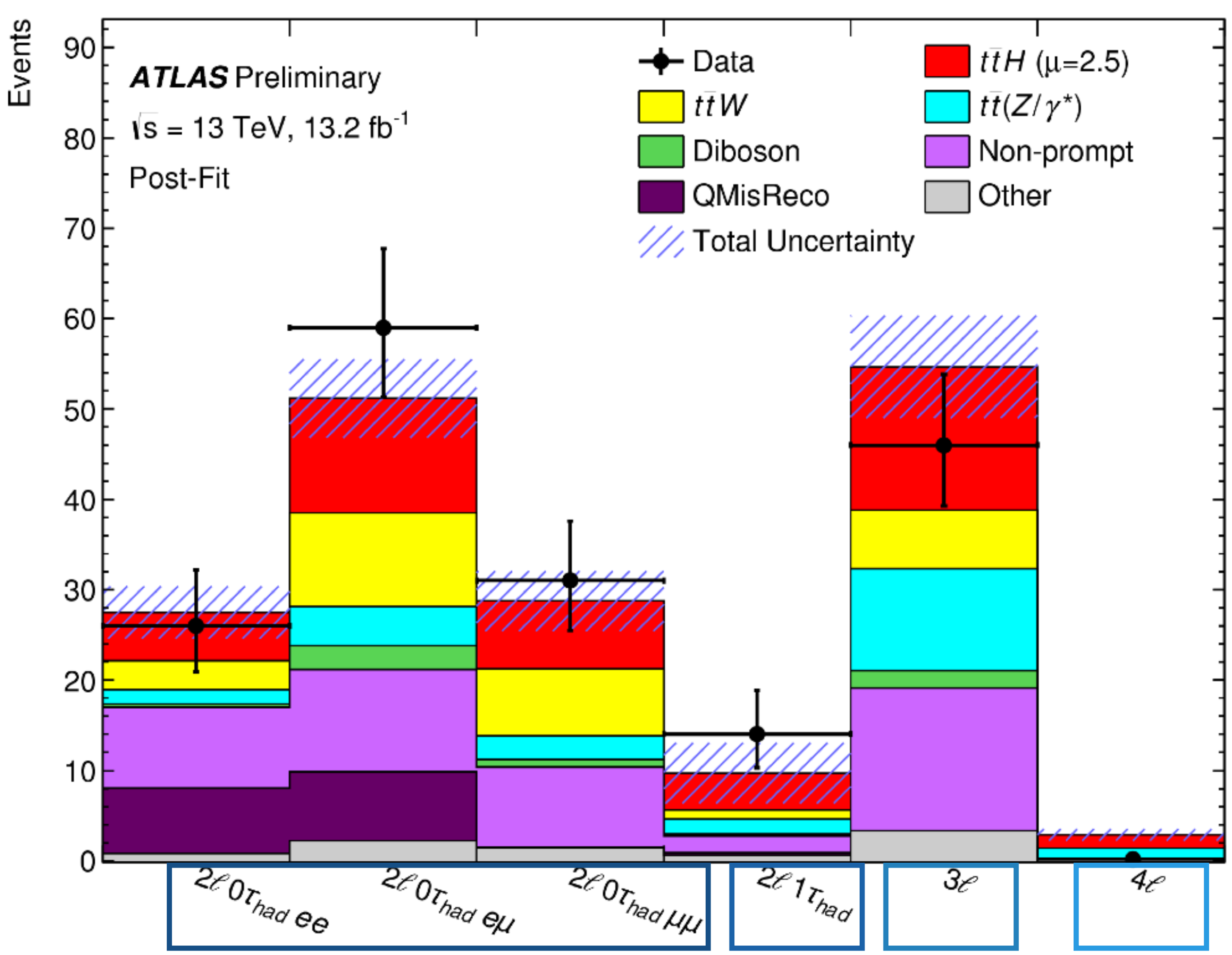
3 Validation Regions (VR) in order to study the ttV background



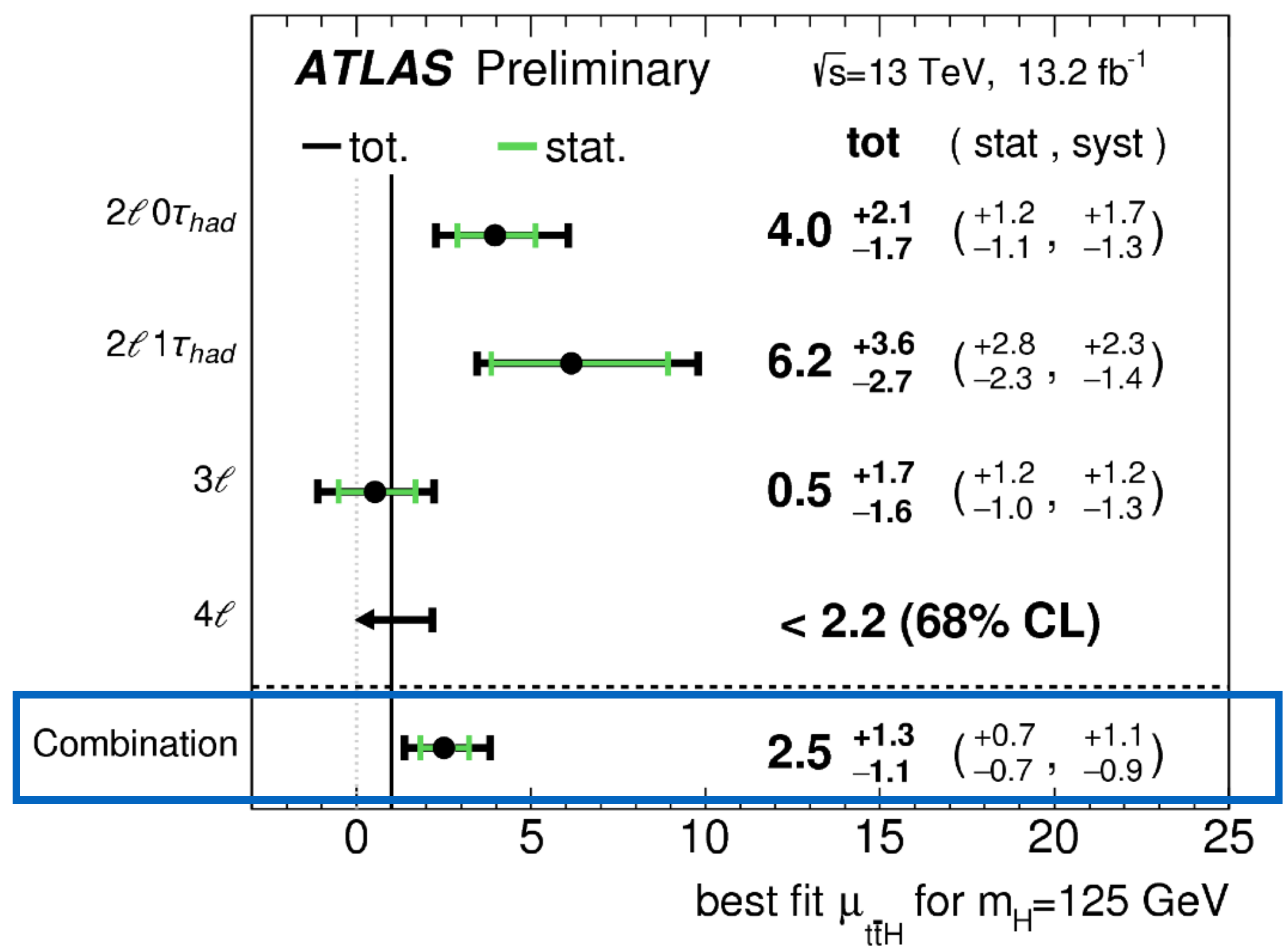
ttH(WW/ZZ/ττ)

ATLAS-CONF-2016-058

- 4 signal regions:
  - **2l0τ**: 2 tight same-sign leptons, 0 τ, ≥ 5 jets, ≥ 1 b-jet;
  - **2l1τ**: 2 tight same-sign leptons, 1 τ (opposite to leptons), ≥ 4 jets, ≥ 1 b-jet;
  - **3l**: 3 no same-sign leptons, ≥ 4 jets and ≥ 1 b-jet or = 3 jets and ≥ 2 b-jets;
  - **4l**: 4 leptons (sum of charges = 0), ≥ 2 jets, ≥ 1 b-jet;



13.2 fb^{-1} @ 13 TeV



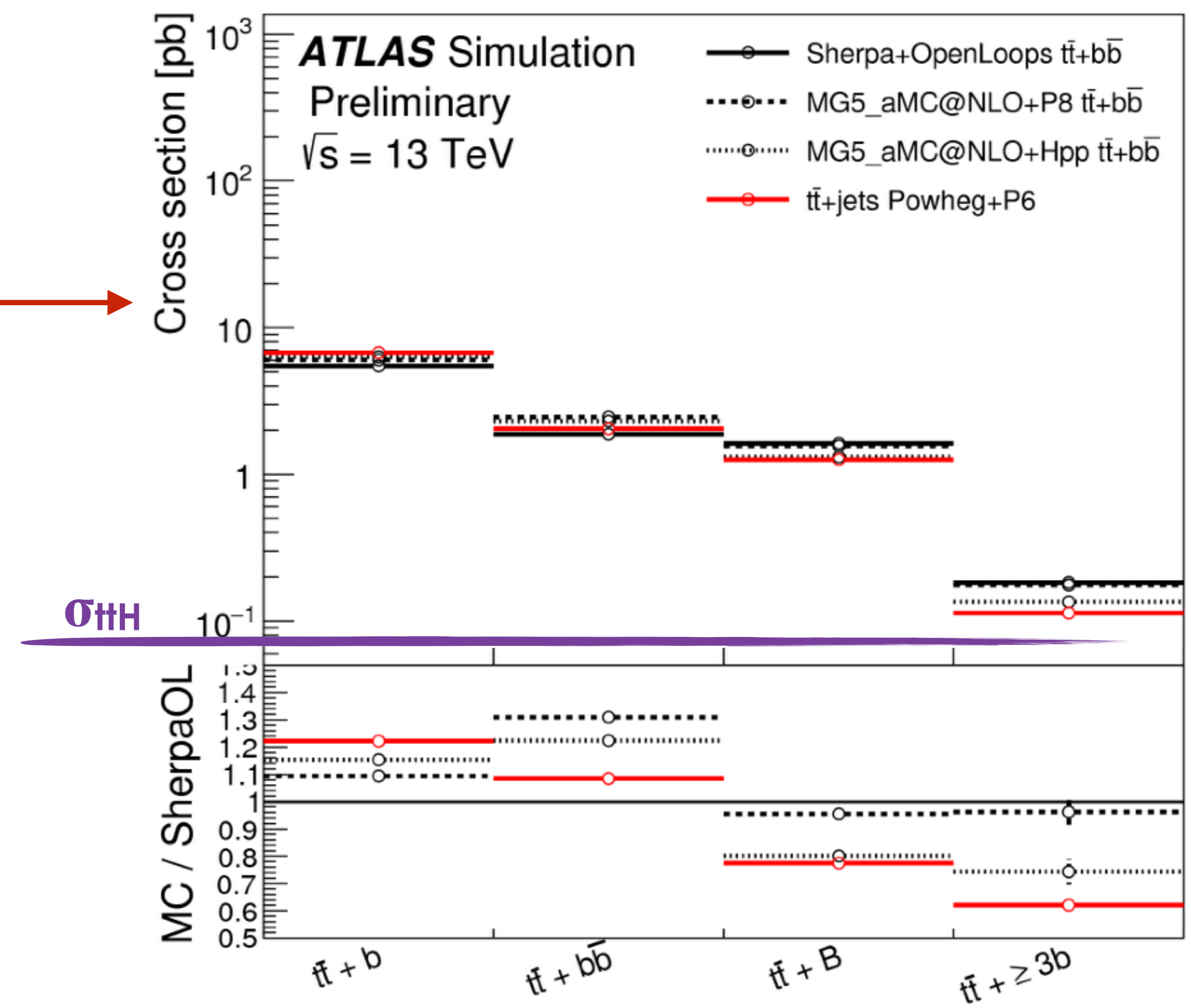
ttH(bb)

BR ~ 58%

ATLAS-CONF-2016-080

- High multiplicity of control and signal regions, depending on the **multiplicity of jets and b-jets**;
- **MultiVariate Analysis** in order to increase the significance and the separation between signal and background;
- most abundant Higgs decay, but **overwhelmed by tt(+HF) jets background** and less easy bb reconstruction.

- Main background: tt+HF, in which the **tt+b-jets show a significant mis-modelling**;
- estimation of the mis-modelling relative to the **generator choice and the PS & hadronization components**;
- **reweighting** is necessary to correct the mis-modelling.

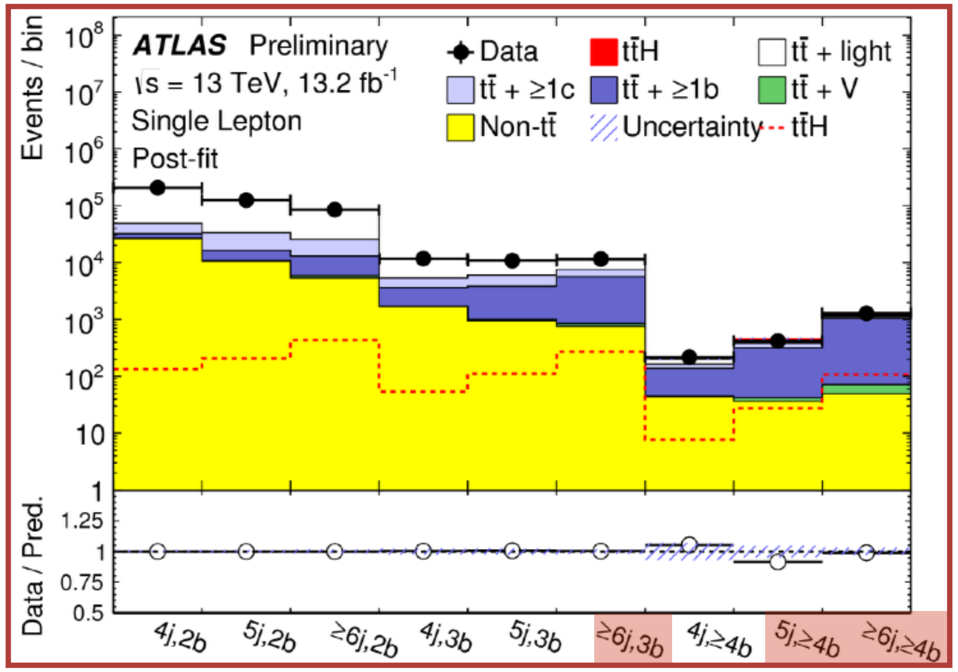


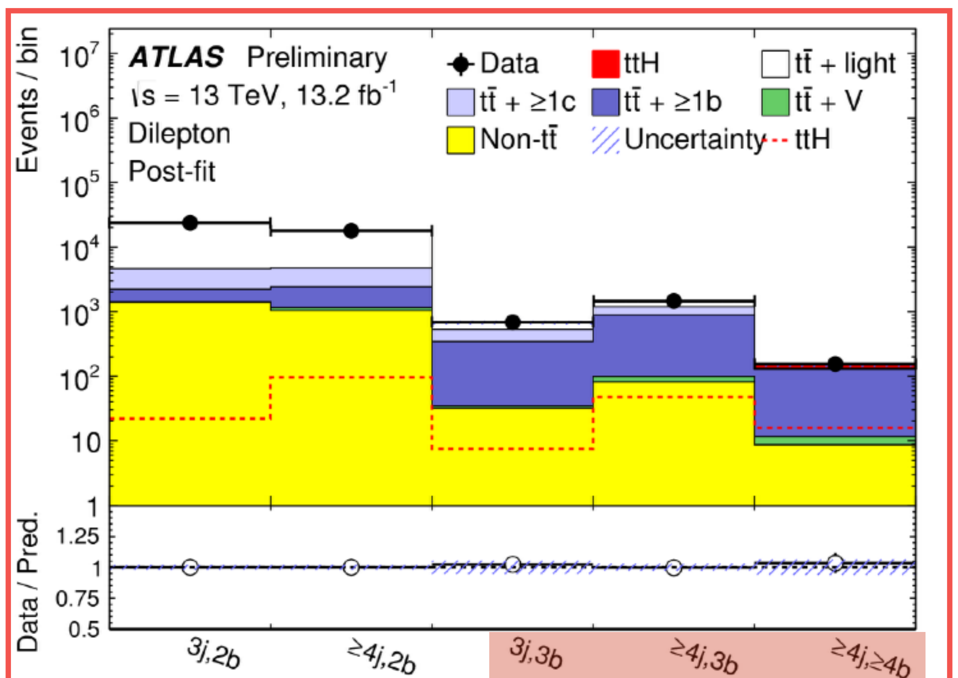
ttH(bb)

2 channels:

ATLAS-CONF-2016-080

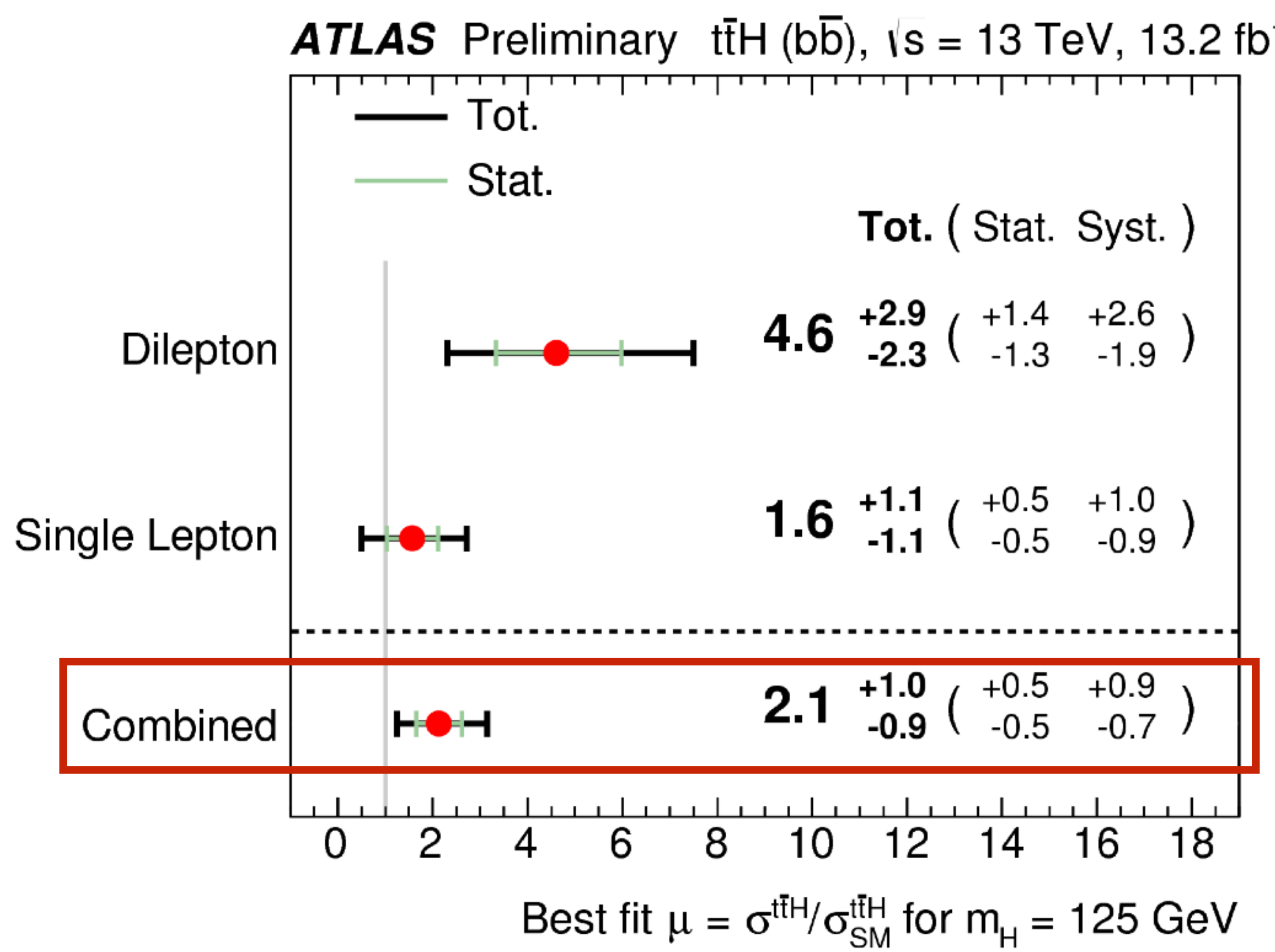
- **single-lepton**: 1 lepton, ≥ 3 jets, ≥ 2 b-jets ⟶ 6 control regions + 3 signal regions;
- **dilepton**: 2 leptons, ≥ 3 jets, ≥ 2 b-jets ⟶ 2 control regions + 3 signal regions.





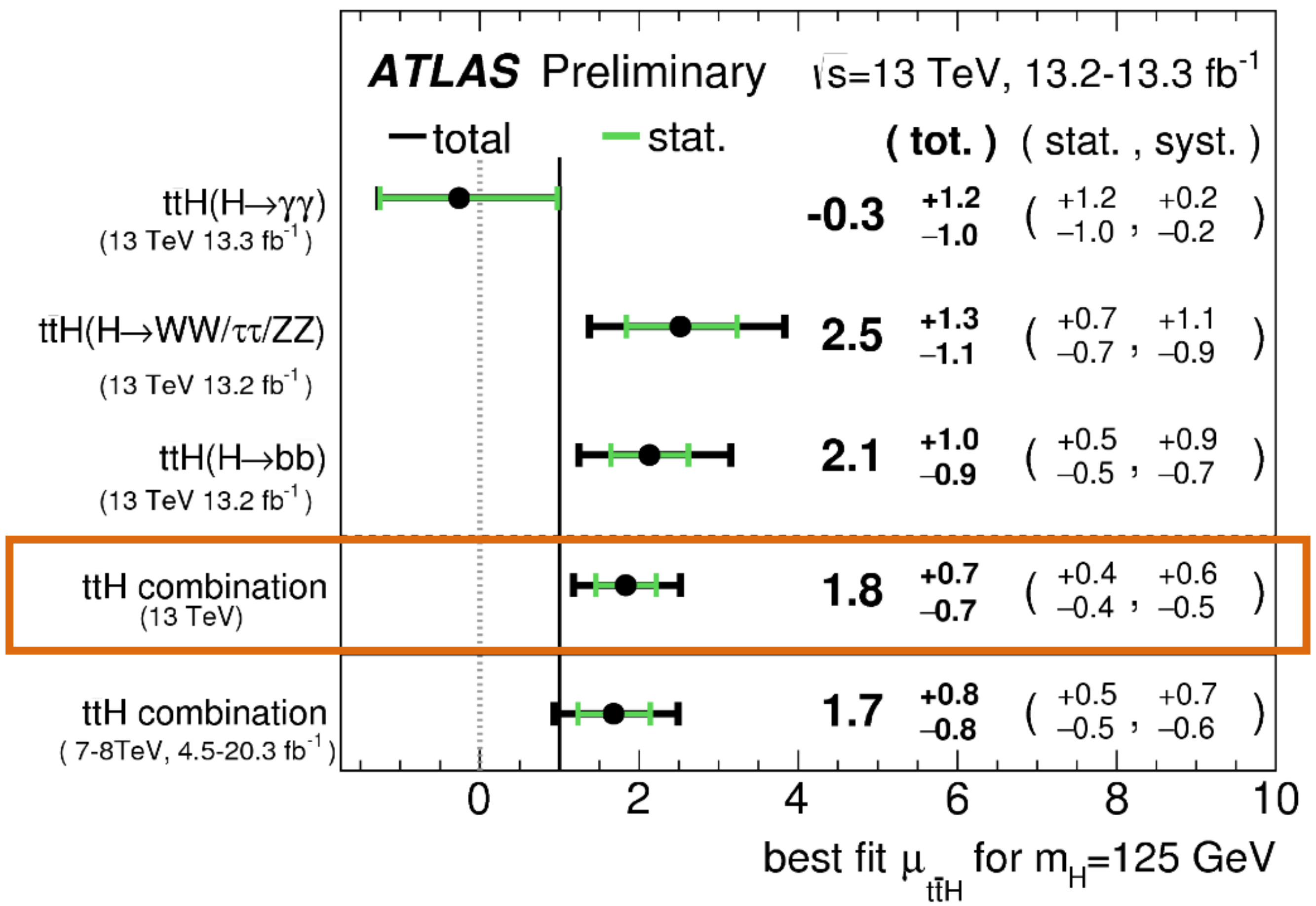
- **discriminant** distributions in CRs and SRs;
- **combined fit** performed to all the regions to determine the signal strengths.

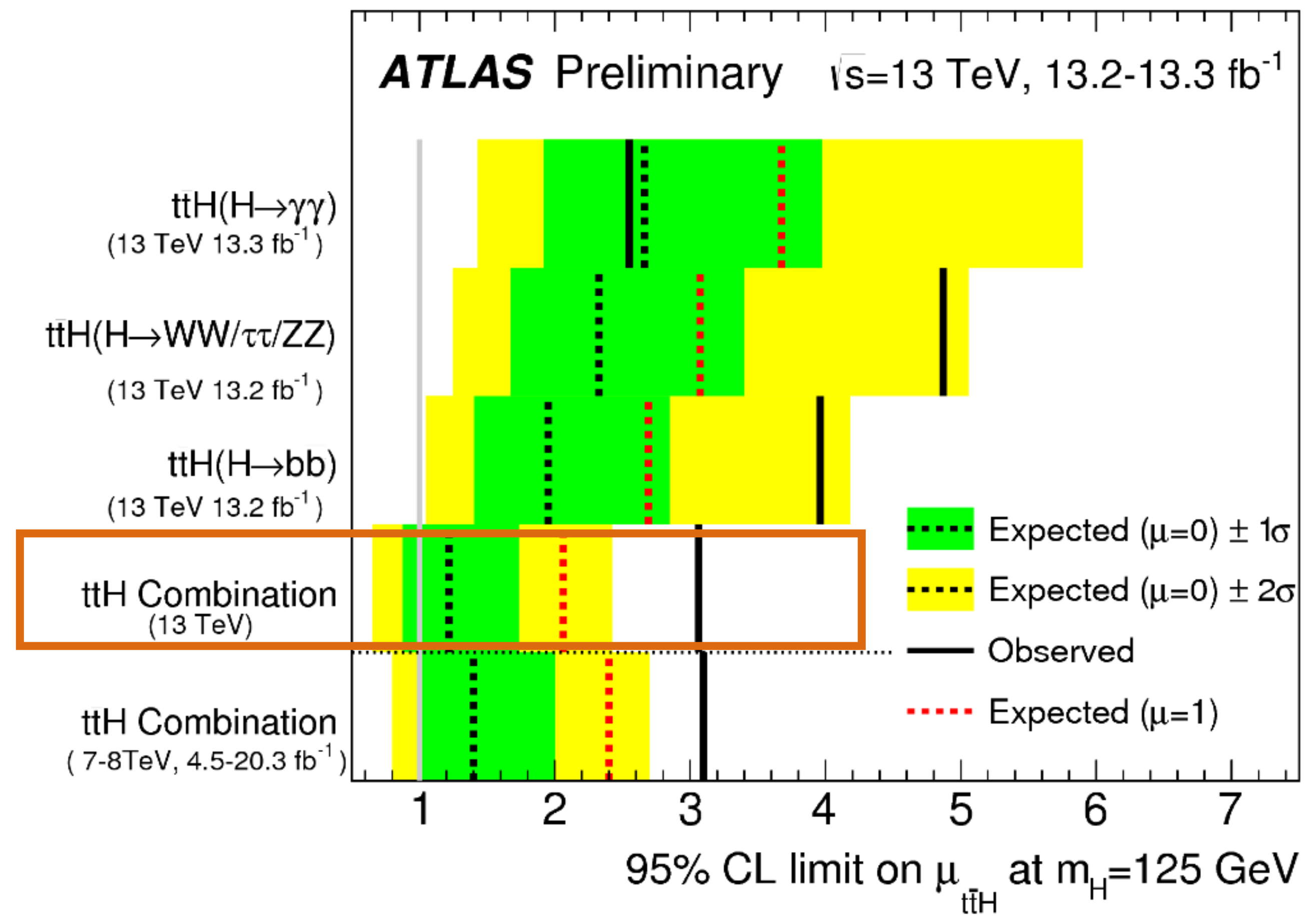
13.2 fb^{-1} @ 13 TeV



Combination and signal strength

ATLAS-CONF-2016-068





- A signal strength **$\mu = 1.8 \pm 0.7$** has been measured;
- this corresponds to an **observed (expected) significance 2.8 σ (1.8 σ)**;
- **observed (expected) 95% C.L. limit on μ is 3.0 (2.1)**.

NB!

- The largest impact in the uncertainty comes from ttH(bb) channel;
- dominated by tt+b/c jets background.

Changes after the publication

ttH(γγ)

- **MVA** technique to get a better separation from QCD.

ttH(WW/ZZ/ττ)

- **MVA** techniques included in the analysis;
- **different division** of phase space in channels.

WORK IN PROGRESS!!

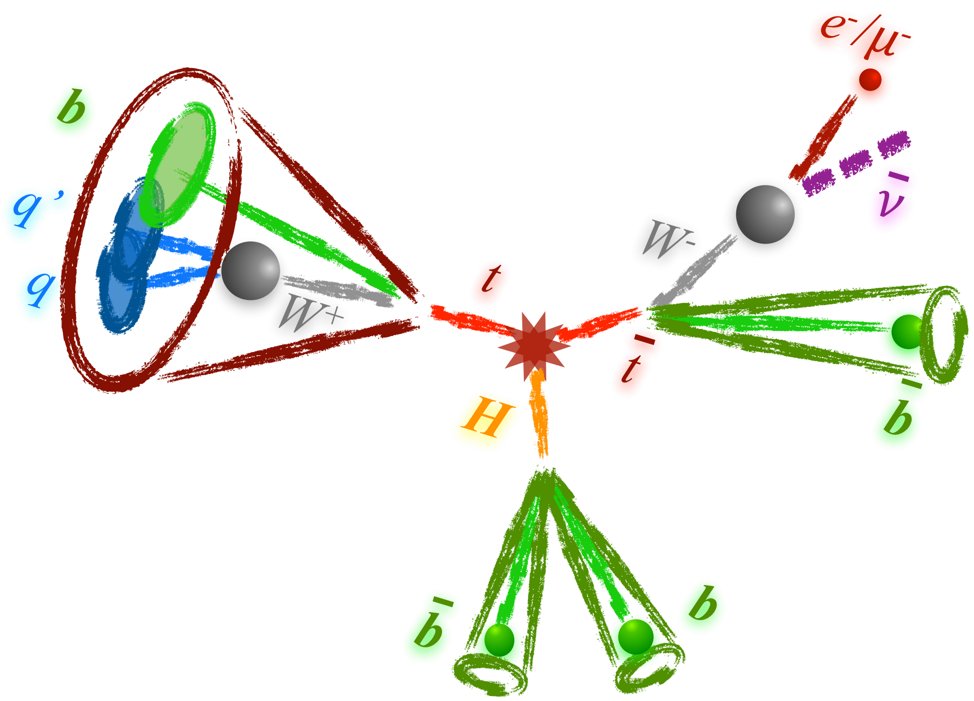
ttH(bb)

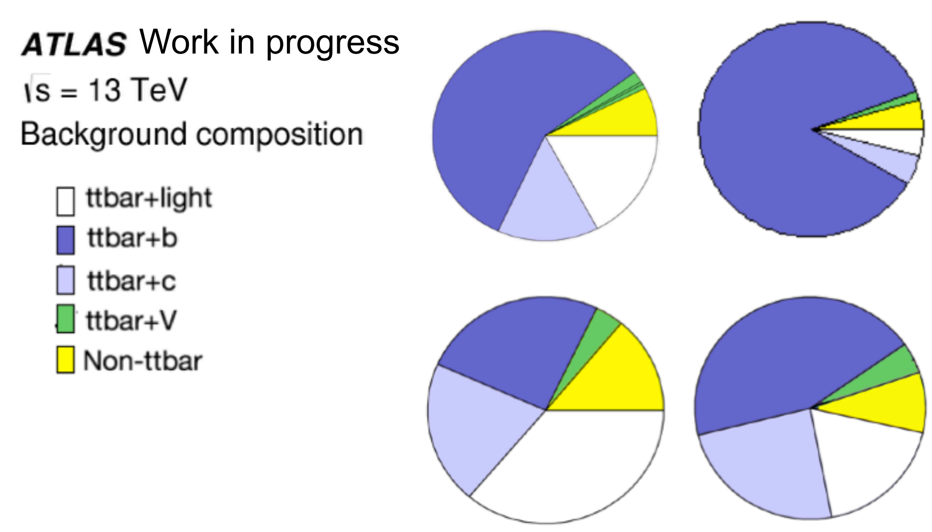
- several regions **options under investigation**;
- improvement in **treatment of tt modelling**.

+

boosted ttH(bb)

- adding **one boosted channel**, sensitive to different kinematics;
- **top/Higgs tagging** techniques;
- right now the strategy is being decided between 4 different SR options (**reclustering jets or large-R jets**).





CMS status

ttH(WW/ZZ/tt) with 35.9 fb^{-1}

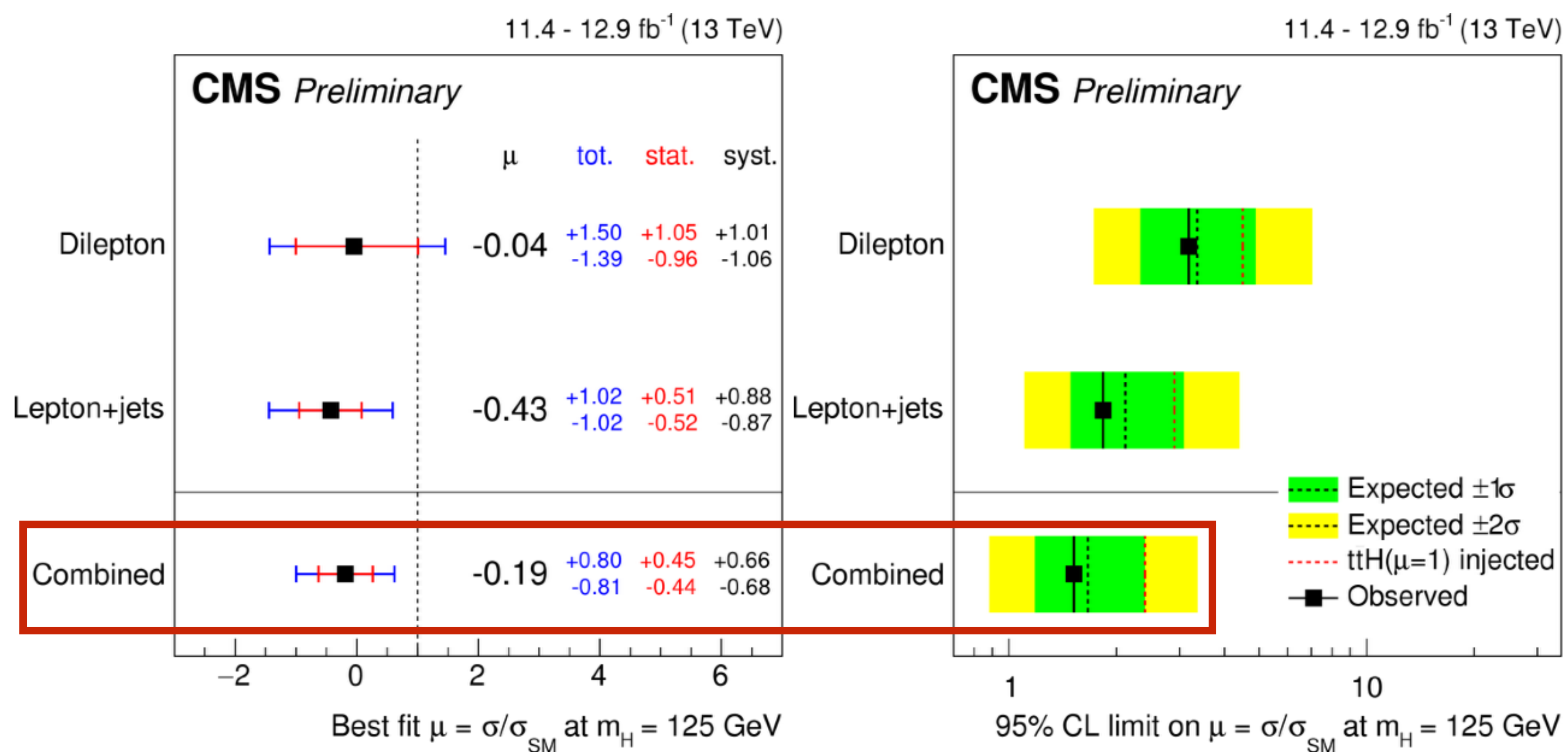
CMS-PAS-HIG-17-004

- **Different analysis strategy:** no channel with τs, different BDT training (against tt and against ttV);
- A signal strength **μ = 1.5 ± 0.5** has been measured;
- this corresponds to an **observed (expected) significance 3.3 σ (2.5 σ)**;
- **observed (expected) 95% C.L. limit on μ is 2.5 (0.8)**.

ttH(bb) with 11.4 -12.9 fb^{-1}

CMS-PAS-HIG-16-038

- **Different analysis strategy:** 2 channels, but different signal regions and different categorisation variables (BDT, MEM);
- A signal strength **μ = -0.19 ± 0.8** has been measured;
- **observed (expected) 95% C.L. limit on μ is 1.5 (1.7)**.



Conclusions

- **"The importance of being ttH"** - (~) Oscar Wilde
  - ttH searches crucial for **Higgs-Top coupling direct measurement**;
  - possible **hint for physics BSM** from indirect measurements;
- **The obtained results:**
  - ATLAS ttH analyses based on **~13fb^{-1} p-p collisions at 13 TeV**;
  - Run-2 results: observed (expected) significance **2.8σ (1.8σ)**;
    - to be compared with: observed (expected) significance **2.3σ (1.5σ) at Run-1**;
  - total uncertainty is **dominated by systematics**.
- **What's next?**
  - **36.5 fb^{-1}** with the full statistics of 2015+2016;
  - **optimisations and big changes** are ongoing in all the analyses;
  - a **boosted ttH channel** will be included in the next combination;
  - preparing **one paper per analysis and one for the combination**.

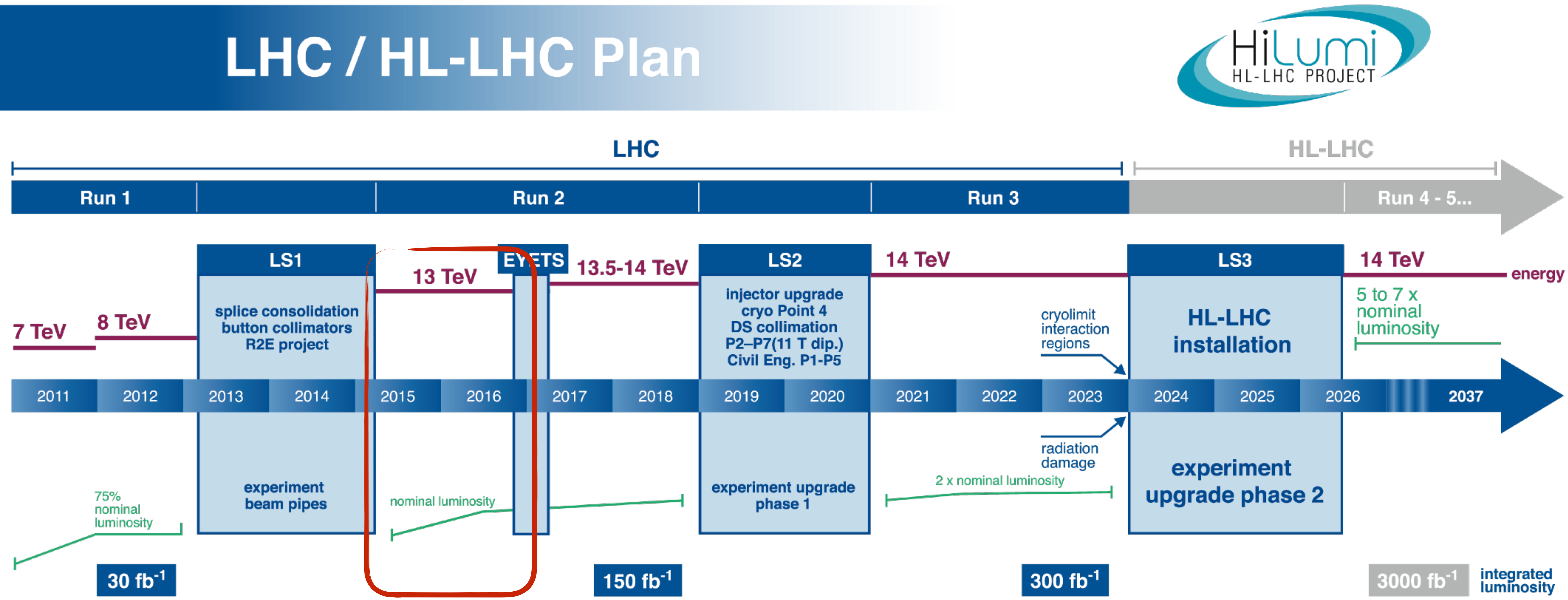
Supporting material

Outline

- ttH production modes and motivations to ttH searches
- Run-1 results: summary
- Different channels in ATLAS:
  - ttH ($\gamma\gamma$);
  - ttH (WW/ZZ/tt);
  - ttH (bb);
- Run-2 early data results
- Changes in the analyses after the publication
- CMS status
- Conclusions

Time schedule of LHC

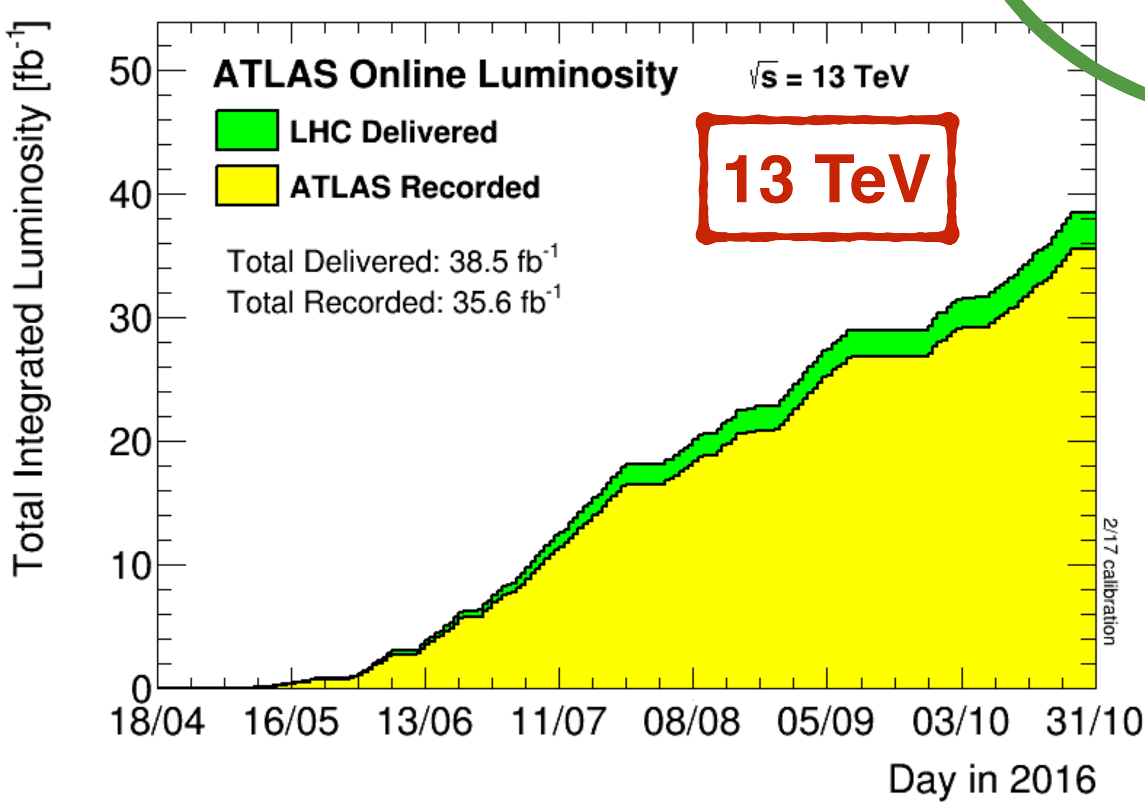
[reference](#)



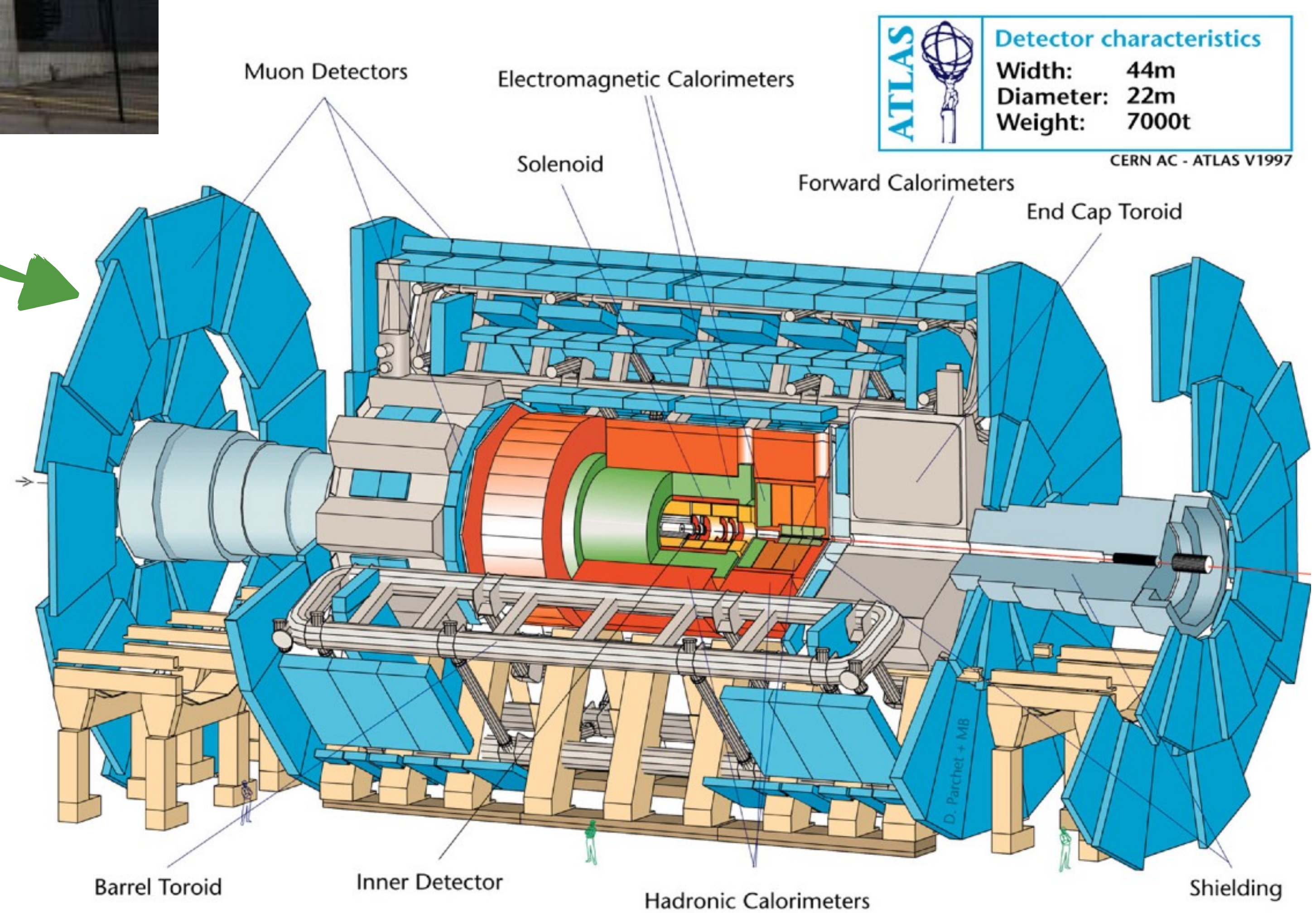
we are here

The ATLAS Experiment

- **Multipurpose experiment** to discover signatures of new physics and to perform precise measurements of Standard Model.
- **Different technologies depending on the kind of physics to be detected**.



reference



Top Quark and Higgs Boson

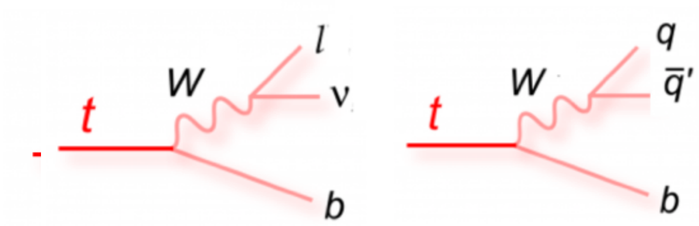
- The last quark discovered, only in **1995**.
- It is the **most massive** fundamental particle known:

[reference]

m_t = 173.34 ± 0.27 (stat) ± 0.71 (syst) GeV

- High m_t implies a **large Yukawa coupling with the Higgs boson** (~1), wrt other couplings ($<10^{-2}$).
- Due to its short lifetime, the top quark **decays before hadronizing** (detected as a **"jet"**: a cone of particles that goes through the detector).
- Unique opportunity to study properties of a bare quark.

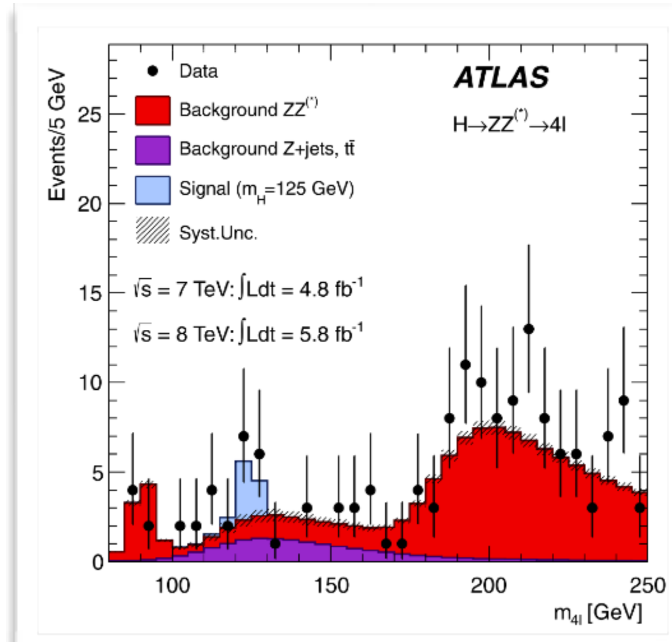
Top decays ~100% of times in **W boson and b quark**

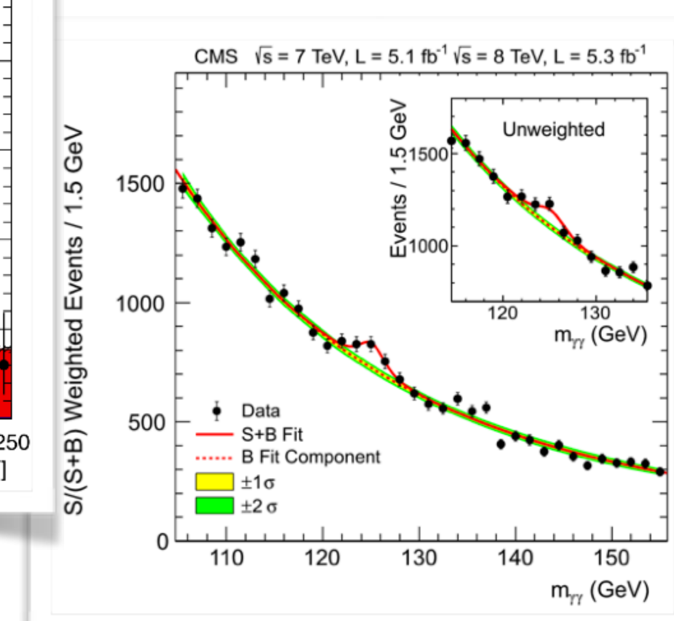


decay time ~ 10^{-24} s

hadronization ~ 10^{-23} s

- Speculated in **1964** by **Higgs**, **Englert** and **Brout**; discovered in 2012 at CERN; **Nobel in 2013**.
- ATLAS and CMS collaborations observed **a neutral scalar particle of mass ~125 GeV**: Higgs boson.



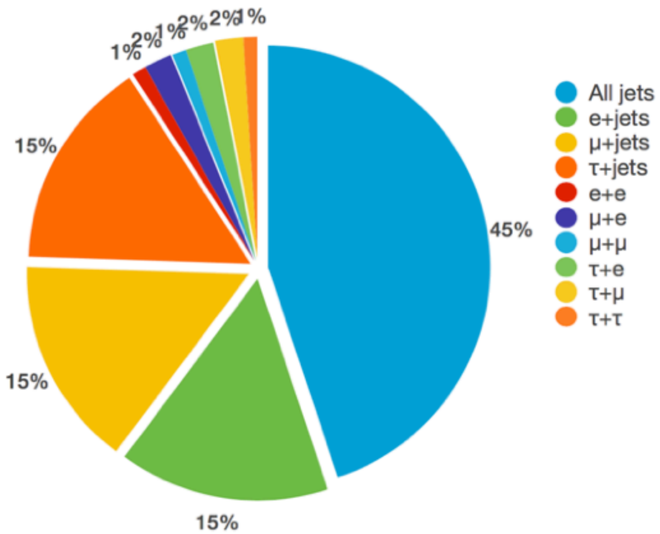


- Still **ongoing studies about its properties** (mass, spin, etc.).

[reference]

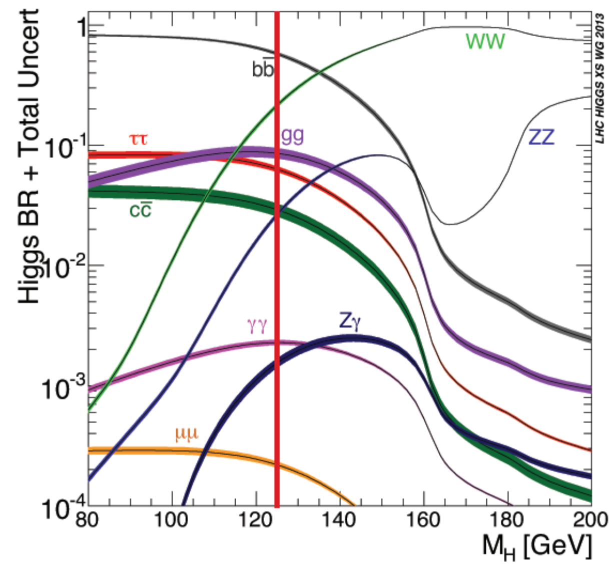
m_H = 125.09 ± 0.21 (stat) ± 0.11 (syst) GeV

The ttH channel



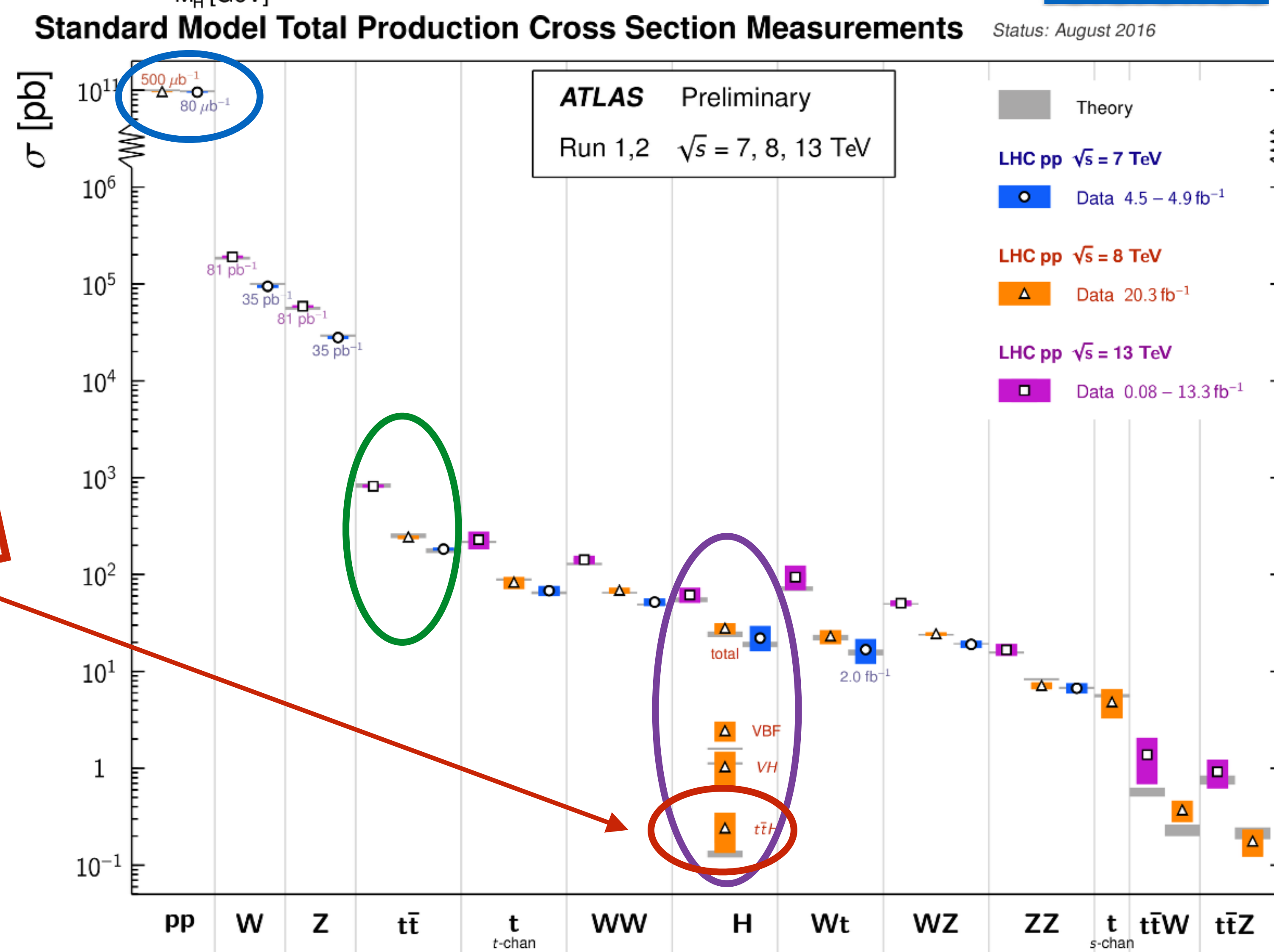
BR (tt->lνqq) ~35%

+



reference

reference



σ(pp->tt) ~ $8\text{x}10^3$ pb

80 events each billion

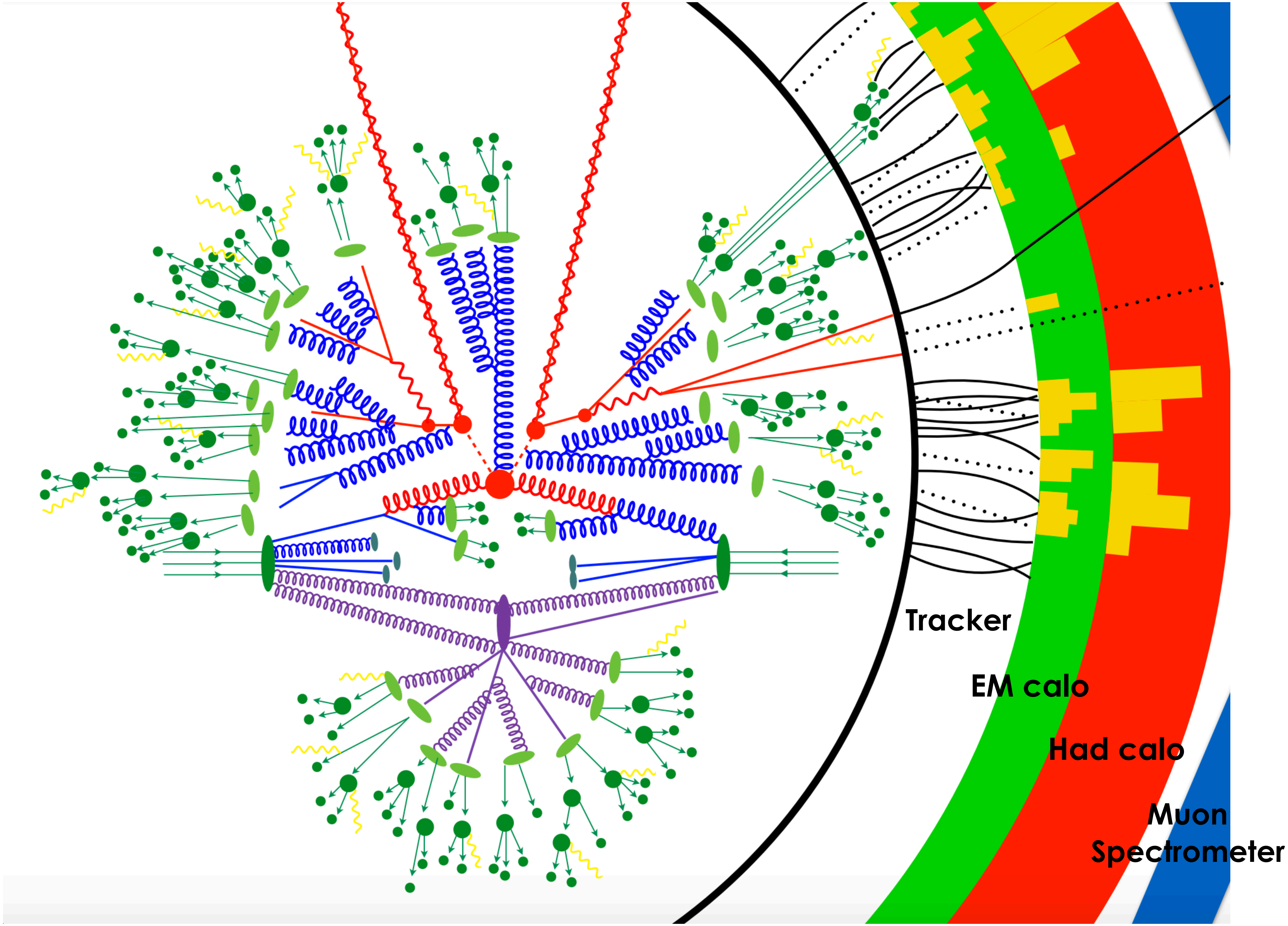
NB! the main background contribution to our channel

Rare event!

σ(pp->**HX**) ~ 45 pb

1 event each 2 billions

Reconstruction



ttH(γγ) - photons isolation

ATLAS-CONF-2016-067

- Reconstruction is seeded in energy clusters in the electromagnetic calorimeter with E_T > 2.5 GeV in a region of $\Delta\eta \times \Delta\phi = 0.075\times0.125$;
- energy is measured from a cluster size of $\Delta\eta \times \Delta\phi = 0.075\times0.175$ in the barrel region of the calorimeter and $\Delta\eta \times \Delta\phi = 0.125\times0.125$ in the calorimeter endcaps;
- The identification of photons is based on the lateral and longitudinal shape of the electromagnetic shower in the calorimeter (2 working points, loose and tight);
- photon candidates are required to deposit only a small fraction of their energy in the hadronic calorimeter and to have a lateral shower shape consistent with that expected from a single electromagnetic shower;
- the info on the shape of the shower in the high granularity first layer is used to discriminate single photons from hadronic jets in which a neutral meson carries most of the jet energy.
- The calorimeter isolation is computed as the sum of transverse energies of positive-energy topological clusters [59] in the calorimeter within a cone of $\Delta R = (\Delta\eta)2 + (\Delta\phi)2 = 0.2$ centered around the photon candidate;
- the track isolation is computed as the scalar sum of the transverse momenta of all tracks in a cone of $\Delta R = 0.2$ with pT > 1 GeV which pass some loose track quality criteria and originate from the diphoton primary vertex.

ttH(γγ) - more info

Signal & Background modeling

Signal model

- Double-sided crystal ball function centered at m_H = 125.09 GeV
- Parameters of the model determined using a fit to MC

Background model

- Modeled with an analytic function
- Chosen from data control regions
  - Minimize bias in extracted signal yield
  - NTNI photons, loosened b-tagging requirement
- Considered functional forms
  - Exponentials, Power Law, Dijet, Bernstein polynomials

ttH(γγ) - more info

Systematics

- Theoretical uncertainties
  - QCD scale
  - PDF
  - Strong coupling constant α_S
  - BR($H \to \gamma\gamma$)
  - PDF acceptance
  - HF content in ttH categories
  - UE & PS
- Experimental uncertainties
  - Yield uncertainties (Luminosity, Trigger, primary vertex selection)
  - Migration uncertainties (JES, JER, flavor tagging, lepton efficiency & ID, pileup reweighting, photon ID & isolation)
- Signal shape uncertainties
  - Scale & resolution
- Background modeling ⇒ Spurious signal

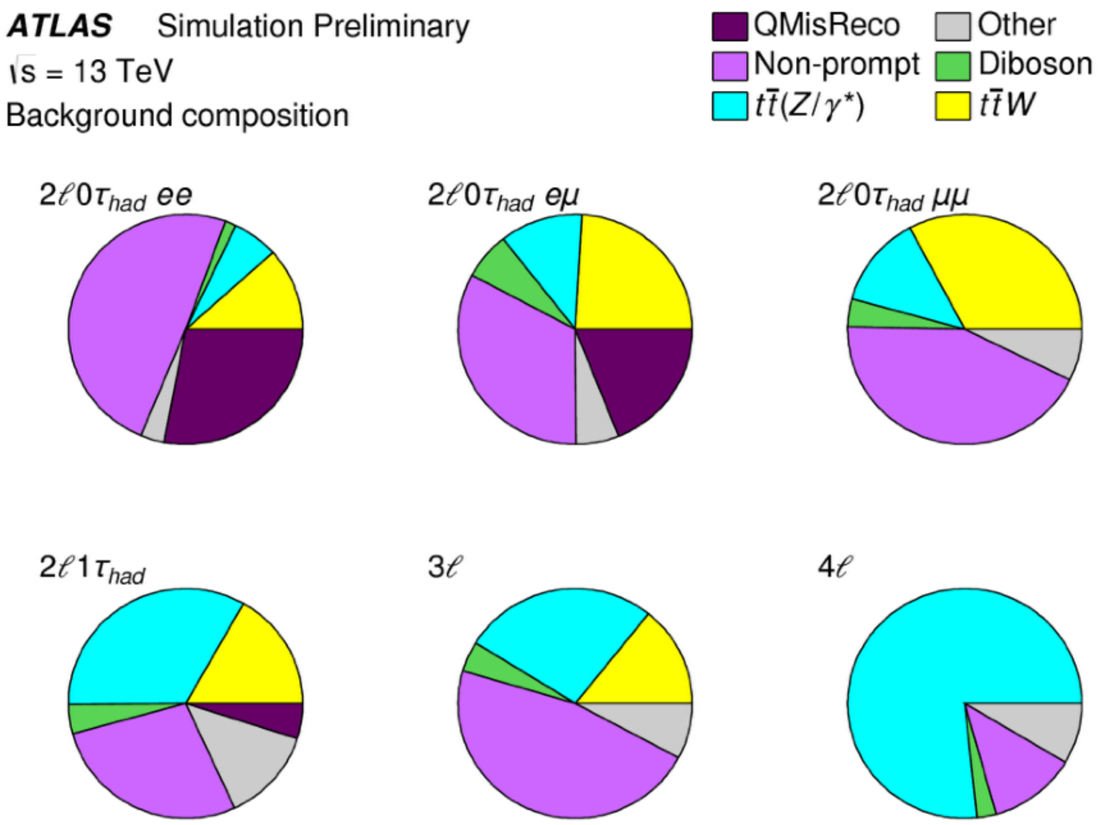
➡ **Most systematics in place except for some theory uncertainties**

ttH(WW/ZZ/ττ)

ATLAS-CONF-2016-058

SR/VR	Channel	Selection criteria
SR	$2\ell 0\tau_{\rm had}$	Two tight light leptons with $p_{\rm T} > 25$, 25 GeV Sum of light lepton charges ± 2 Any electrons must have $\lvert\eta_e\rvert < 1.37$ Zero $\tau_{\rm had}$ candidates $N_{\rm jets} \geq 5$ and $N_{b-{\rm jets}} \geq 1$
SR	$2\ell 1\tau_{\rm had}$	Two tight light leptons, with $p_{\rm T} > 25$, 15 GeV Sum of light lepton charges ± 2 Exactly one $\tau_{\rm had}$ candidate, of opposite charge to the light leptons $\lvert m(ee) - 91.2\ {\rm GeV}\rvert > 10$ GeV for ee events $N_{\rm jets} \geq 4$ and $N_{b-{\rm jets}} \geq 1$
SR	3ℓ	Three light leptons; sum of light lepton charges ± 1 Two same-charge leptons must be tight and have $p_{\rm T} > 20$ GeV $m(\ell^+\ell^-) > 12$ GeV and $\lvert m(\ell^+\ell^-) - 91.2\ {\rm GeV}\rvert > 10$ GeV for all SFOC pairs $\lvert m(3\ell) - 91.2\ {\rm GeV}\rvert > 10$ GeV $N_{\rm jets} \geq 4$ and $N_{b-{\rm jets}} \geq 1$, or $N_{\rm jets} = 3$ and $N_{b-{\rm jets}} \geq 2$
SR	4ℓ	Four light leptons; sum of light lepton charges 0 All leptons pass "gradient" isolation selection $m(\ell^+\ell^-) > 12$ GeV and $\lvert m(\ell^+\ell^-) - 91.2\ {\rm GeV}\rvert > 10$ GeV for all SFOC pairs 100 GeV $< m(4\ell) <$ 350 GeV and $\lvert m(4\ell) - 125\ {\rm GeV}\rvert > 5$ GeV $N_{\rm jets} \geq 2$ and $N_{b-{\rm jets}} \geq 1$
VR	Tight $t\bar{t}Z$	3ℓ lepton selection %and trigger selection At least one $\ell^+\ell^-$ pair with $\lvert m(\ell^+\ell^-) - 91.2\ {\rm GeV}\rvert < 10$ GeV $N_{\rm jets} \geq 4$ and $N_{b-{\rm jets}} \geq 2$
VR	Loose $t\bar{t}Z$	3ℓ lepton selection %and trigger selection At least one $\ell^+\ell^-$ pair with $\lvert m(\ell^+\ell^-) - 91.2\ {\rm GeV}\rvert < 10$ GeV $N_{\rm jets} \geq 4$ and $N_{b-{\rm jets}} \geq 1$, or $N_{\rm jets} = 3$ and $N_{b-{\rm jets}} \geq 2$
VR	WZ + 1 b-tag	3ℓ lepton selection %and trigger selection At least one $\ell^+\ell^-$ pair with $\lvert m(\ell^+\ell^-) - 91.2\ {\rm GeV}\rvert < 10$ GeV $N_{\rm jets} \geq 1$ and $N_{b-{\rm jets}} = 1$
VR	$t\bar{t}W$	$2\ell 0\tau_{\rm had}$ lepton selection %and trigger selection $2 \leq N_{\rm jets} \leq 4$ and $N_{b-{\rm jets}} \geq 2$ $H_{\rm T,jets} > 220$ GeV for ee and $e\mu$ events $E_{\rm T}^{\rm miss} > 50$ GeV and ($m(ee) < 75$ or $m(ee) > 105$ GeV) for ee events

	Higgs boson decay mode				$A \times \epsilon$
Category	WW^*	$\tau\tau$	ZZ^*	Other	$(\times 10^{-4})$
$2\ell 0\tau_{\rm had}$	77%	17%	3%	3%	14
$2\ell 1\tau_{\rm had}$	46%	51%	2%	1%	2.2
3ℓ	74%	20%	4%	2%	9.2
4ℓ	72%	18%	9%	2%	0.88



ttH(WW/ZZ/ττ)

ATLAS-CONF-2016-058

Uncertainty Source	$\Delta\mu$	
Non-prompt leptons and charge misreconstruction	+0.56	−0.64
Jet-vertex association, pileup modeling	+0.48	−0.36
$t\bar{t}W$ modeling	+0.29	−0.31
$t\bar{t}H$ modeling	+0.31	−0.15
Jet energy scale and resolution	+0.22	−0.18
$t\bar{t}Z$ modeling	+0.19	−0.19
Luminosity	+0.19	−0.15
Diboson modeling	+0.15	−0.14
Jet flavor tagging	+0.15	−0.12
Light lepton (e, μ) and τ_{had} ID, isolation, trigger	+0.12	−0.10
Other background modeling	+0.11	−0.11
Total systematic uncertainty	+1.1	−0.9

ttH(bb)

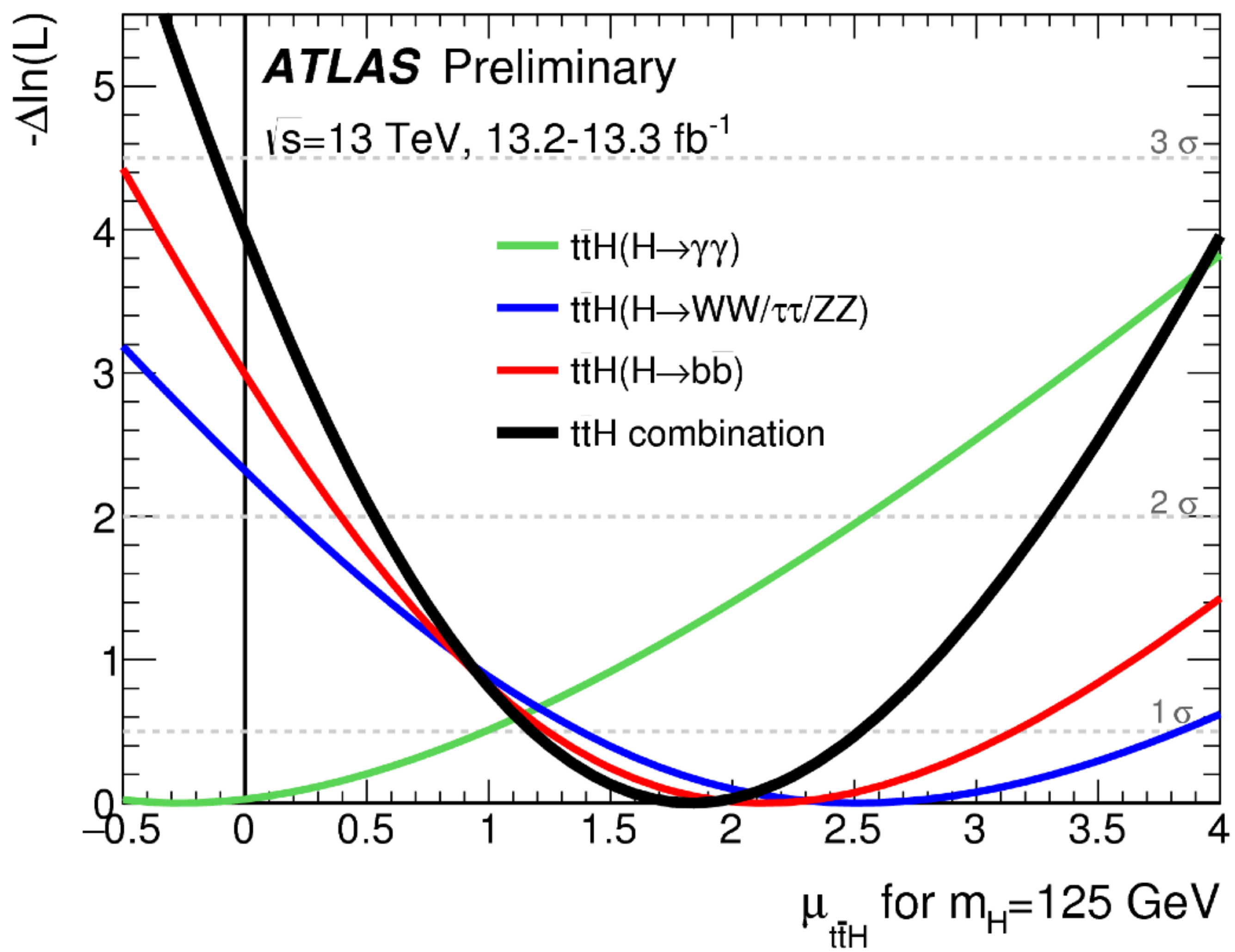
ATLAS-CONF-2016-080

Systematic source	How evaluated	$t\bar{t}$ categories
$t\bar{t}$ cross-section	±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	$t\bar{t}+\geq 1c$

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

Combination and signal strength

ATLAS-CONF-2016-068



Analysis	Observed	$-2\ \sigma$	$-1\ \sigma$	Median ($\mu_{t\bar{t}H} = 0$)	$+1\ \sigma$	$+2\ \sigma$	Median ($\mu_{t\bar{t}H} = 1$)
$t\bar{t}H,\ H \to \gamma\gamma$	2.6	1.4	1.9	2.7	4.0	5.9	3.7
$t\bar{t}H,\ H \to (WW, \tau\tau, ZZ)$	4.9	1.2	1.7	2.3	3.4	5.1	3.1
$t\bar{t}H,\ H \to b\bar{b}$	4.0	1.0	1.4	1.9	2.8	4.2	2.7
$t\bar{t}H$ combination	3.0	0.6	0.9	1.2	1.7	2.4	2.1
$t\bar{t}H$ combination Run-1	3.1	0.8	1.0	1.4	2.0	2.7	2.4

CMS status: ttH(WW/ZZ/ττ)

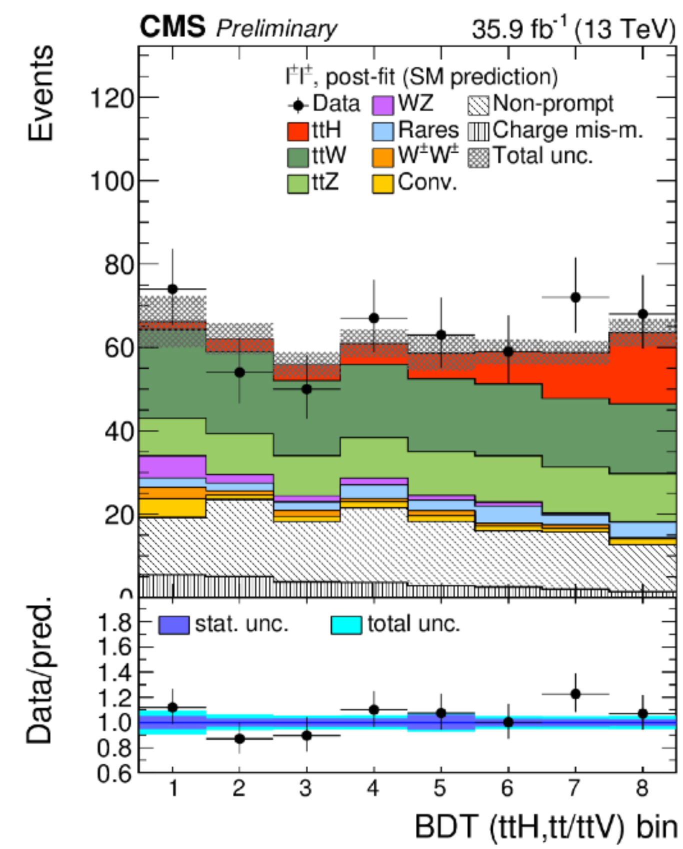
CMS-PAS-HIG-17-004

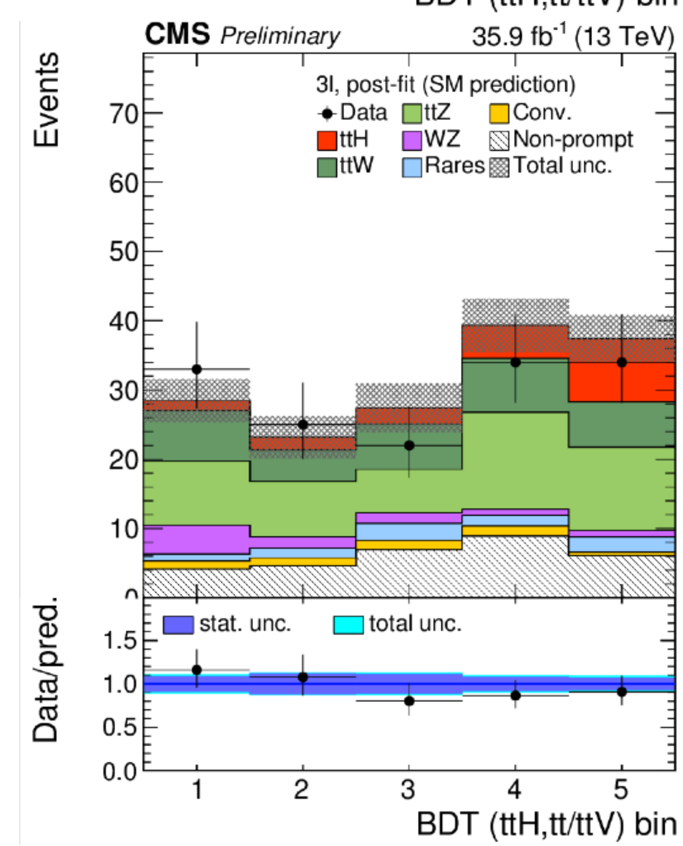
35.9 fb^{-1}

Category	Observed μ fit $\pm 1\sigma$	Expected μ fit $\pm 1\sigma$
Same-sign di-lepton	1.7 (−0.5) (+0.6)	1.0 (−0.5) (+0.5)
Three lepton	1.0 (−0.7) (+0.8)	1.0 (−0.7) (+0.8)
Four lepton	0.9 (−1.6) (+2.3)	1.0 (−1.6) (+2.4)
Combined (2016 data)	1.5 (−0.5) (+0.5)	1.0 (−0.4) (+0.5)
Combined (2015 data) [42]	0.6 (−1.1) (+1.4)	1.0 (−1.1) (+1.3)
Combined (2015+2016 data)	1.5 (−0.5) (+0.5)	1.0 (−0.4) (+0.5)

Category	Observed limit	Expected limit $\pm 1\sigma$
Same-sign di-lepton	2.8	0.9 (−0.3) (+0.4)
Three lepton	2.5	1.4 (−0.4) (+0.7)
Four lepton	5.9	4.9 (−1.7) (+3.1)
Combined	2.5	0.8 (−0.2) (+0.3)

- **Different analysis strategy:** no channel with τs, different BDT training (against tt and against ttV);
- A signal strength **μ = 1.5 ± 0.5** has been measured;
- this corresponds to an **observed (expected) significance 3.3 σ (2.5 σ)**;
- **observed (expected) 95% C.L. limit on μ is 2.5 (0.8).**

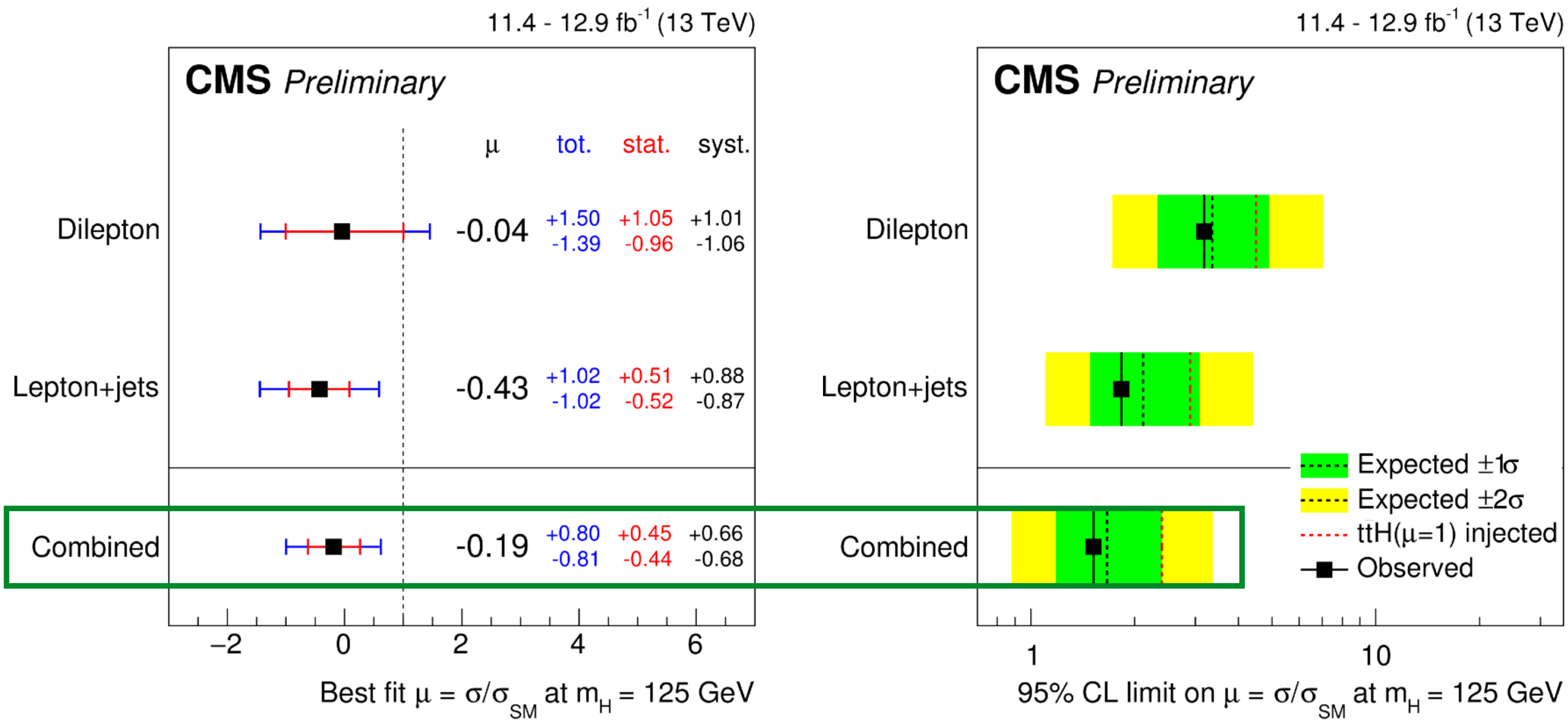




CMS status: ttH(bb)

11.4 -12.9 fb^{-1}

CMS-PAS-HIG-16-038



- **Different analysis strategy:** 2 channels, but different signal regions and different categorisation variables (BDT, MEM);
- A signal strength **μ = -0.19 ± 0.8** has been measured;
- **observed (expected) 95% C.L. limit on μ is 1.5 (1.7)**.

Large-R jets: reconstruction and grooming

both small-R jets and large-R jets

Reconstruction algorithms

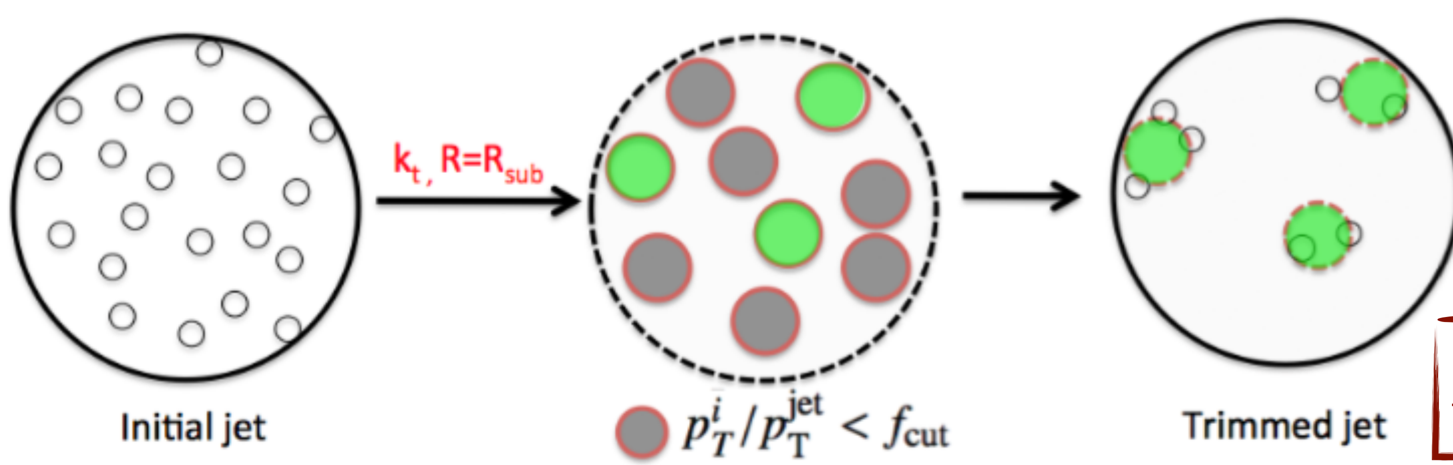
Jet are reconstructed with an iterative algorithm which combines calo deposits inside a given radius R = 1.0

Jets are then cleaned, with "**grooming**" algorithms, from contamination due to the high particles concentration

only large-R jets

Trimming algorithm

Jet constituents are reconstructed again into jets with smaller radius $\mathbf{R_{sub}}$ (subjet). Subjets with lower p_T than a fraction $\mathbf{f_{cut}}$ of initial jet p_T are dropped off.
The final jet is reconstructed using only the remaining subjets.



R_{sub} = 0.2
f_{cut} = 0.05

Anti-k_t and k_t algorithms

- The iterative recombination procedure works by first cleaning a list of all objects (either hadrons, topo-clusters or tracks) in an event.
- The ordering of the list is irrelevant and proto-jets are built from these objects.
- Two distance measures in y-ϕ-space are associated to each member of the list, between the proto-jet and its closest neighbor:

$$\rho_{ij} = \min\left(p_{Ti}^{2p}, p_{Tj}^{2p}\right) \frac{(\Delta R_{ij})^2}{R^2}$$

measure of the opening angle between the two constituents

and between the proto-jet and the beam:

$$\rho_{iB} = p_{Ti}^{2p}$$

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$$

➡ If $\varrho_{iB} < \varrho_{ij}$: the proto-jet is closer to the beam than it is to any other proto-jet in the event, so it is defined as a jet and removed from the list.

➡ If $\varrho_{iB} > \varrho_{ij}$: the two proto-jets i and j are combined into one, thereby forming a new proto-jet. This procedure continues through all proto-jets in the event.

✦ If p = +1 ⟶ k_t algorithm: proto-jets with the smallest p_T tend to be clustered first, so that the highest p_T proto-jets are clustered last.

✦ If p = -1 ⟶ anti-k_t algorithm: proto-jets with the largest p_T are clustered first. A consequence of this is that isolated anti-k_t jets tend to be very close to circular in η-ϕ space, because the axis of the jet is relatively fixed after the first few steps of recombination. This stability makes anti-k_t jets more robust than k_t jets in high multiplicity environments.

Tagging techniques

- Used to exploit all the substructure characteristics of the large-R jets in boosted regimes;
- **Top-tagging:** simple algorithm which provides cuts on two large-R jet substructure variables:

Calibrated mass of the jet

$$(m_{\mathrm{jet}}^{\mathrm{uncalib}})^2 = \left(\sum_i E_i\right)^2 - \left(\sum_i \vec{p}_i\right)^2$$

calibration needed because of low p_T and large angle contributions

Ratio τ_{32} between N-subjettiness variables

$$\tau_N = \frac{1}{d_0} \sum_k p_{\mathrm{T}k} \times \min(\delta R_{1k}, \delta R_{2k}, \ldots, \delta R_{Nk})$$

number of reconstructed subjets

$$d_0 = \sum_k p_{\mathrm{T}k} \times R$$

distance between subjet i and constituent k

$$\tau_{32} = \tau_3 / \tau_2$$

- **Higgs-tagging:** very similar to top-tagging but for the second substructure variable:

Energy correlation D_2

$$D_2^{(\alpha,\beta)} = \frac{e_3^{(\alpha)}}{(e_2^{(\beta)})^{3\alpha/\beta}}$$

e_2, e_3 = energy correlation functions: take into account all p_T, combinations and boost-invariant angle ΔR between constituents

α, β =1

- **Taggers performances** are given by two values, calculated in the same way:

Signal efficiency

$$\epsilon = \left(\frac{N_{tagged}}{N_{total}}\right)_{signal}$$

Background rejection

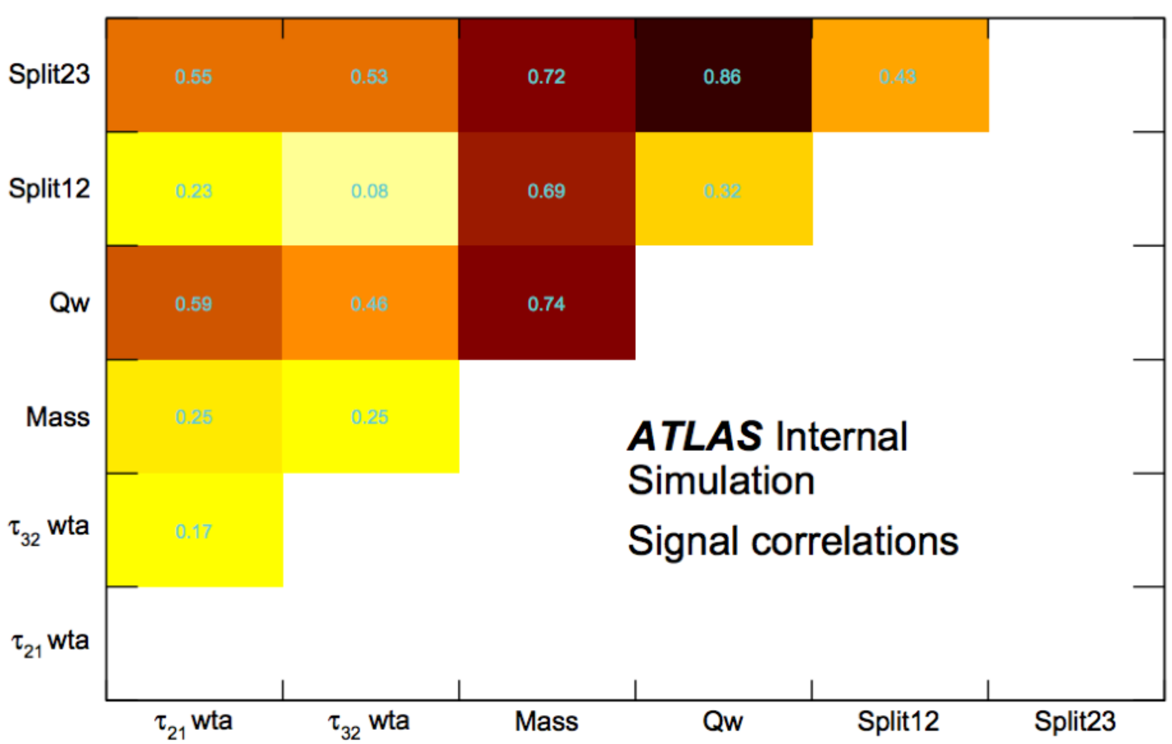
$$r = \left(\frac{N_{total}}{N_{tagged}}\right)_{background}$$

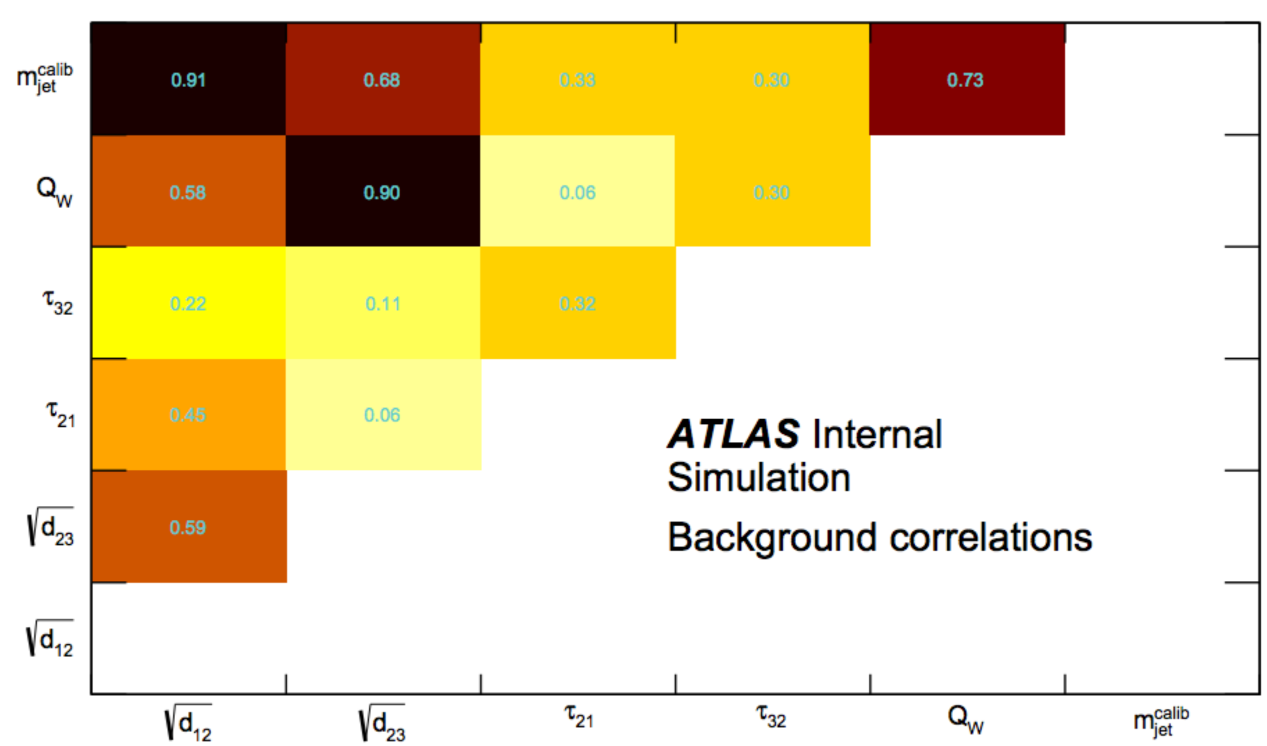
Top Tagging

The two variables $\mathbf{m_{jet}^{calib}}$ and $\boldsymbol{\tau_{32}}$ were chosen from a set of substructure variables, including other N-subjettiness ratio (τ_{21}), splitting scale variables ($\sqrt{d_{12}}$,$\sqrt{d_{23}}$) and the minimum dijet mass from the three subjets (Qw).

- The two chosen variables show a **good background rejection** at 50% and 80% signal efficiency.
- The τ_{ij} variables are **uncorrelated** with respect to the mass and energy scale variables.

This combination of strong performance and lack of correlated behavior motivates the choice of tagging variables.

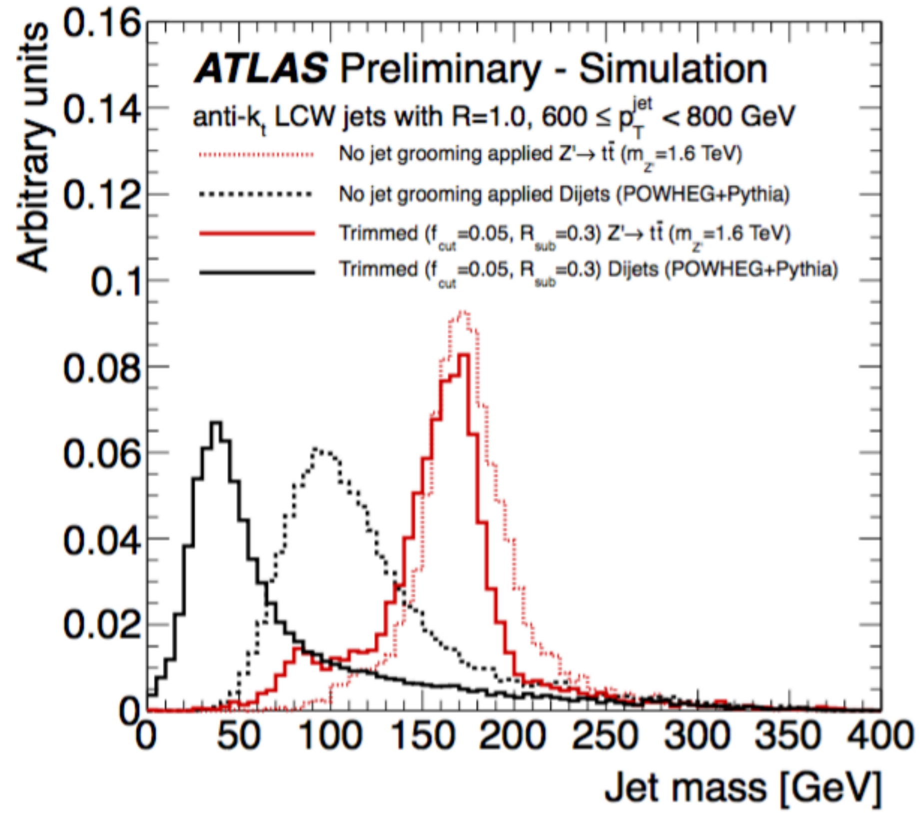


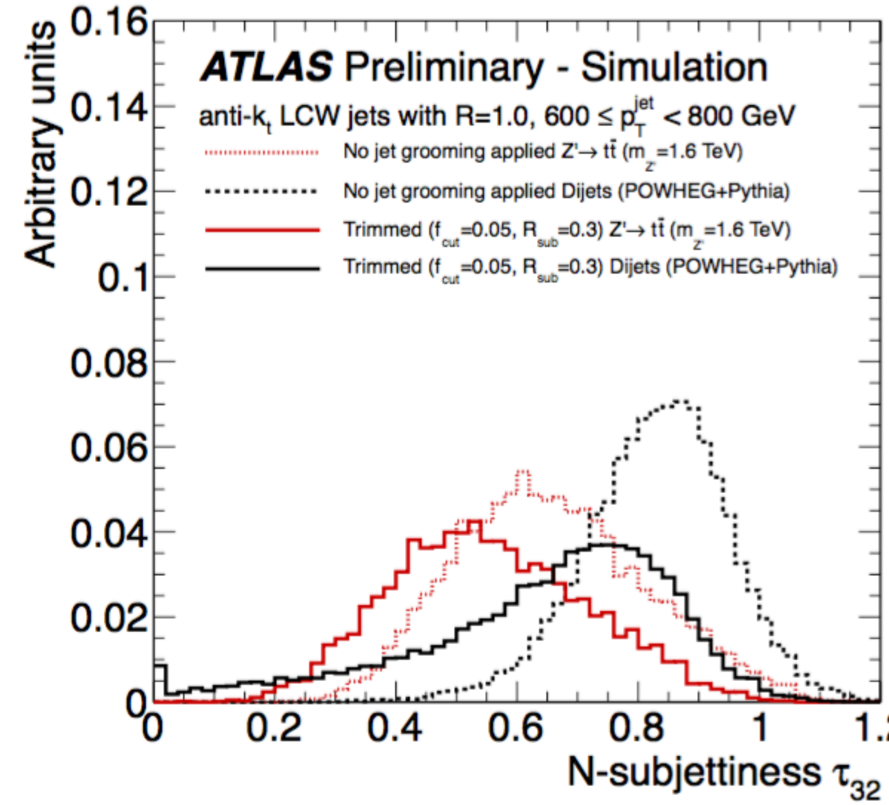


ATL-PHYS-PUB-2015-053

Top Tagging technique

- Used to **exploit all the substructure characteristics of the large-R jets** in boosted regimes;
- substructure variables in the algorithm: **large-R jet mass** and **τ_{32} ratio**.





ATL-PHYS-PUB-2015-053

Performances

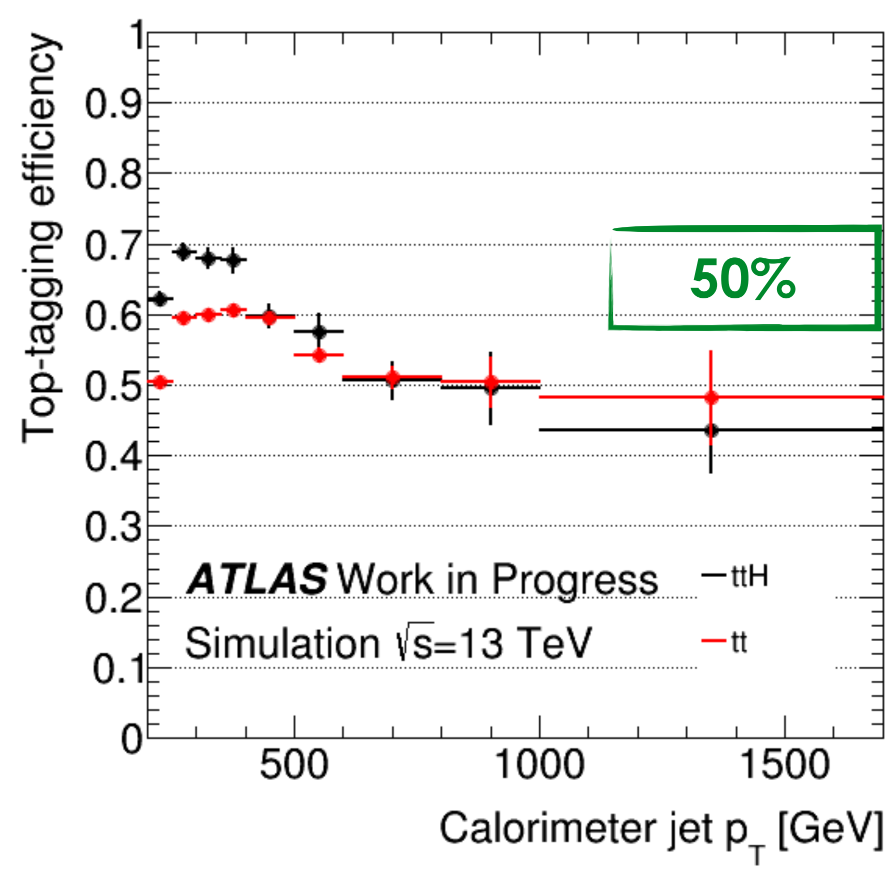
- Comparison between **efficiencies in tt and ttH**;
- looking at:

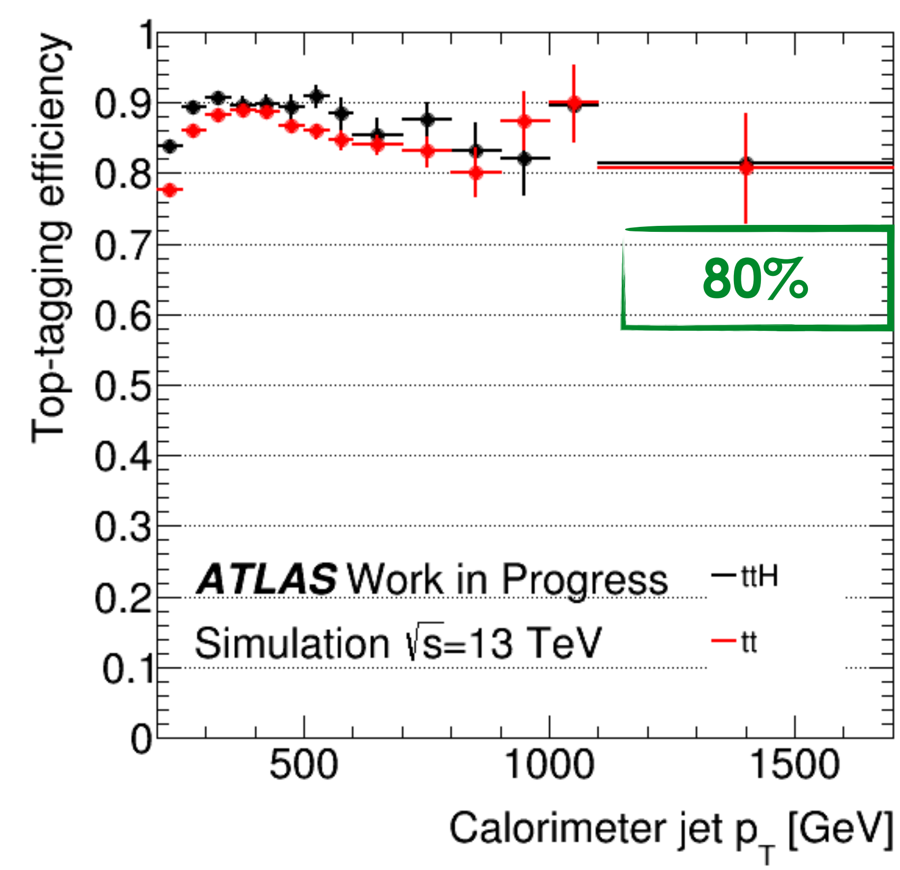
Signal efficiency

$$\epsilon = \left(\frac{N_{tagged}}{N_{total}}\right)_{signal}$$

Background rejection

$$r = \left(\frac{N_{total}}{N_{tagged}}\right)_{background}$$

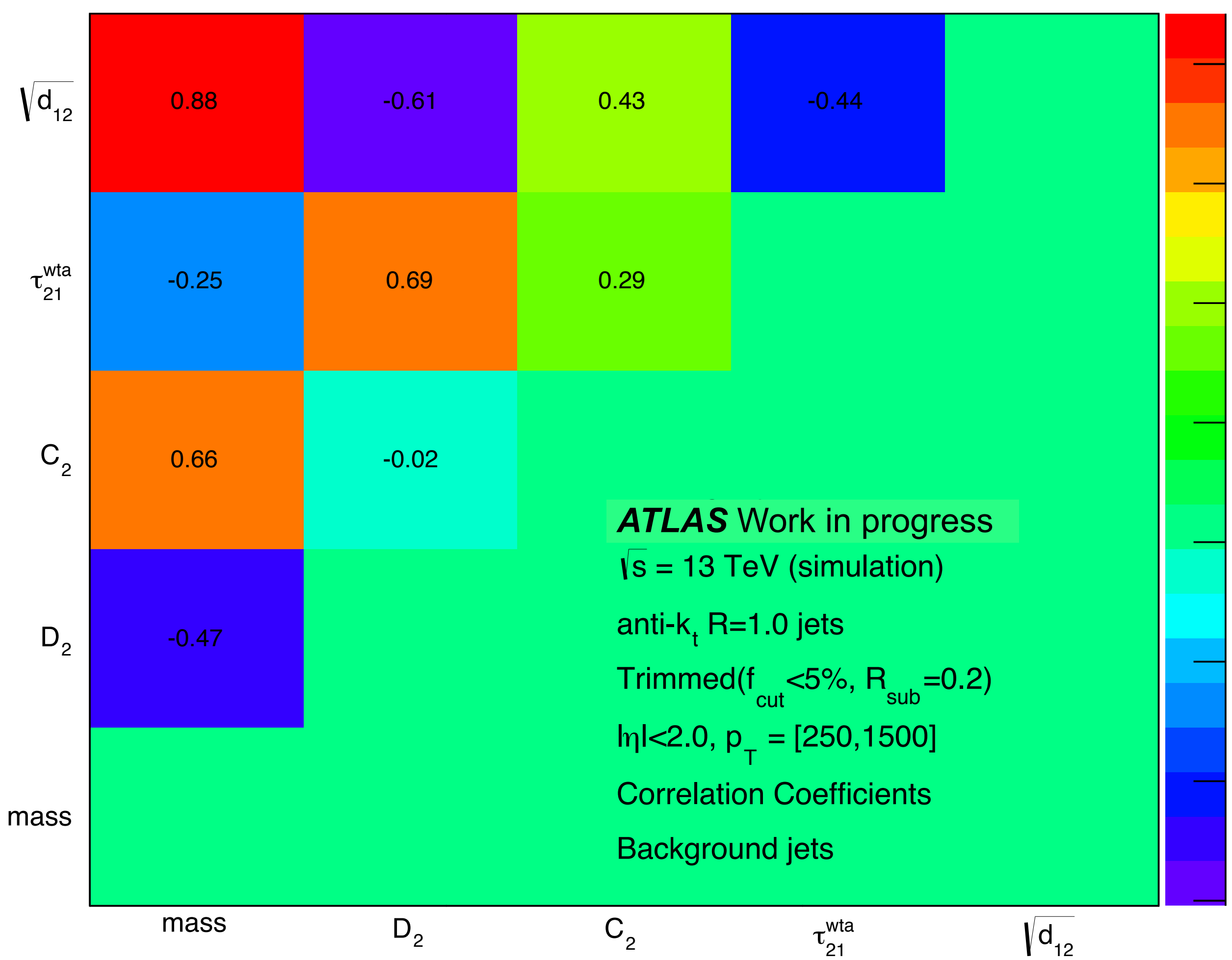




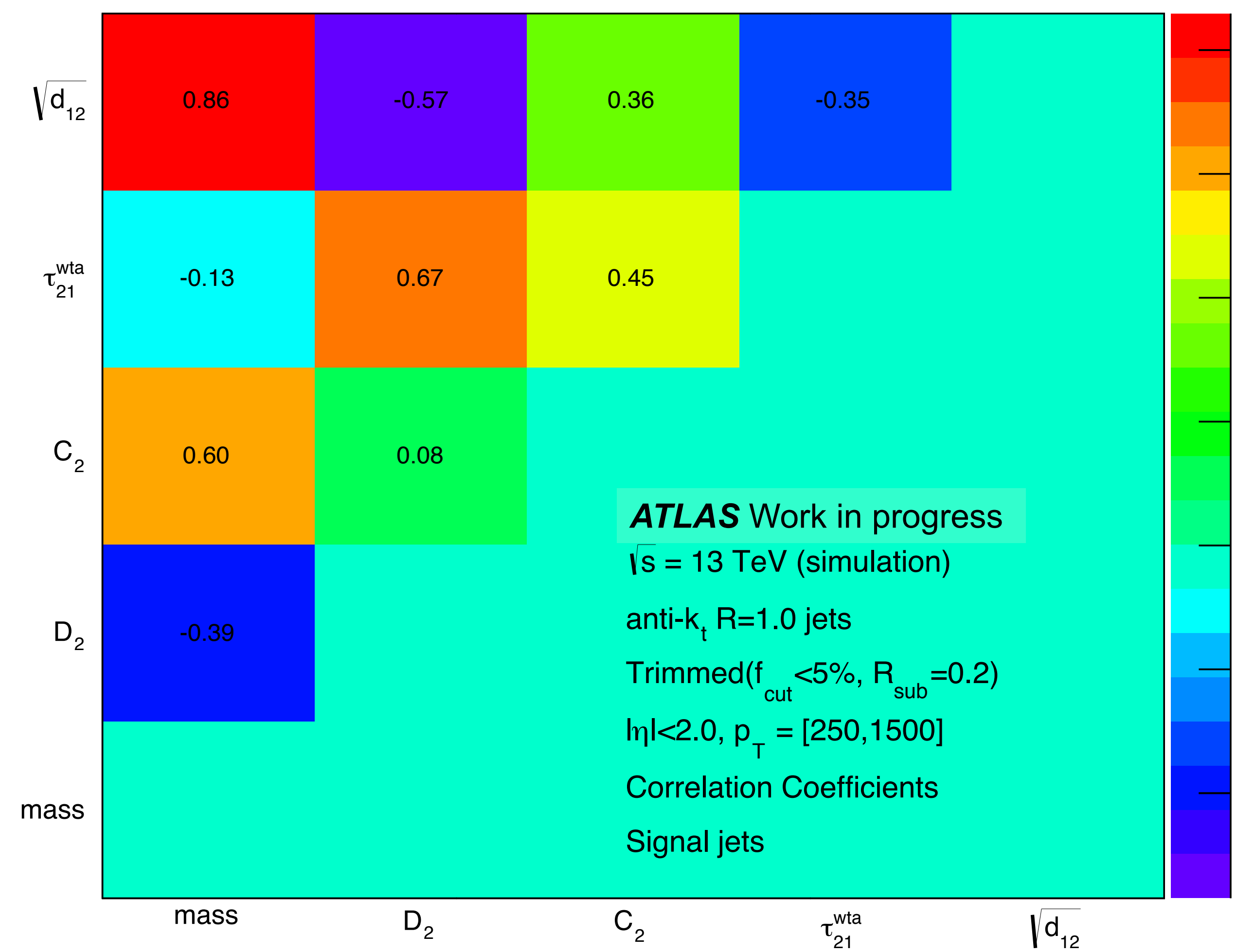
Higgs Tagging

- Correlation matrices:
  - 5 variables studied;
  - considered all the combinations.

Background

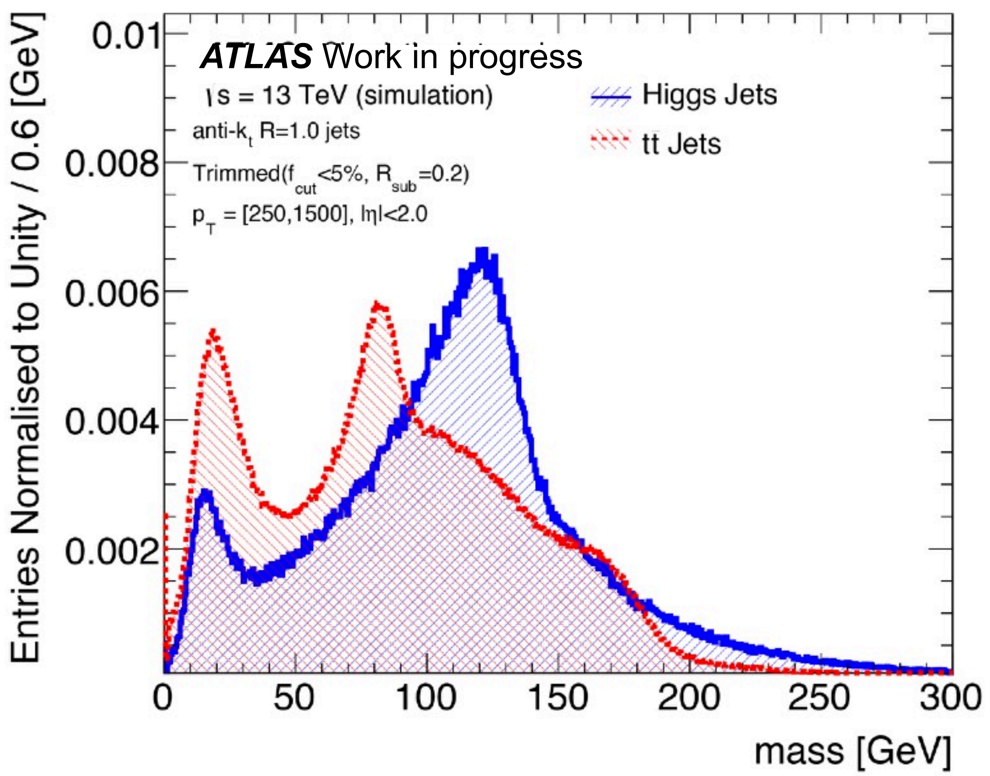


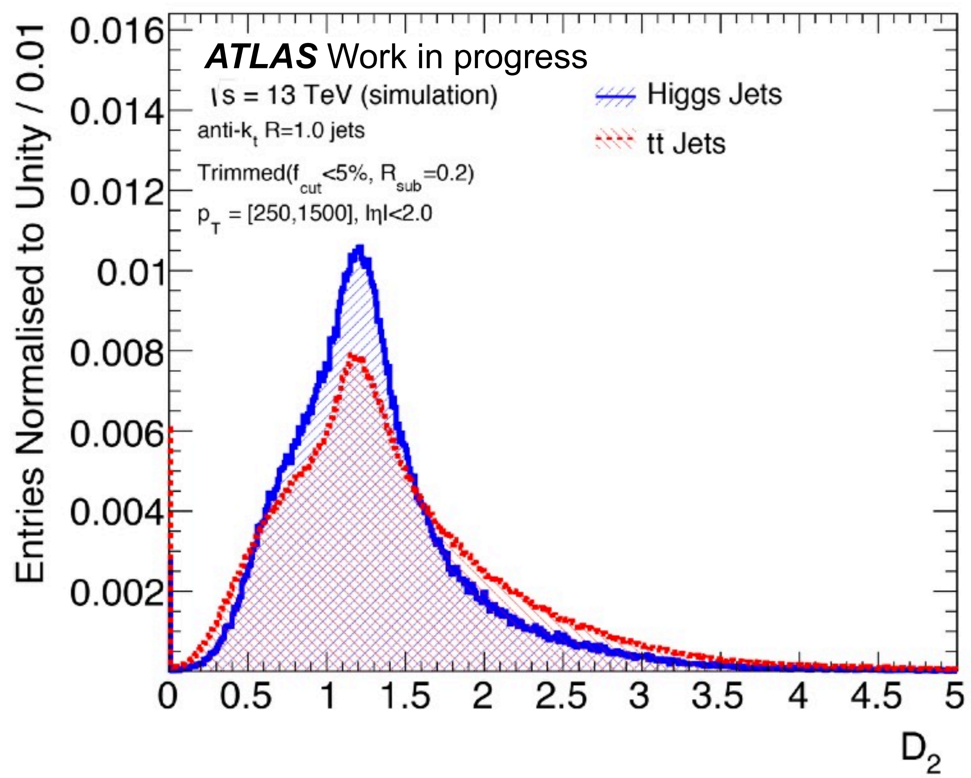
Signal



Higgs Tagging

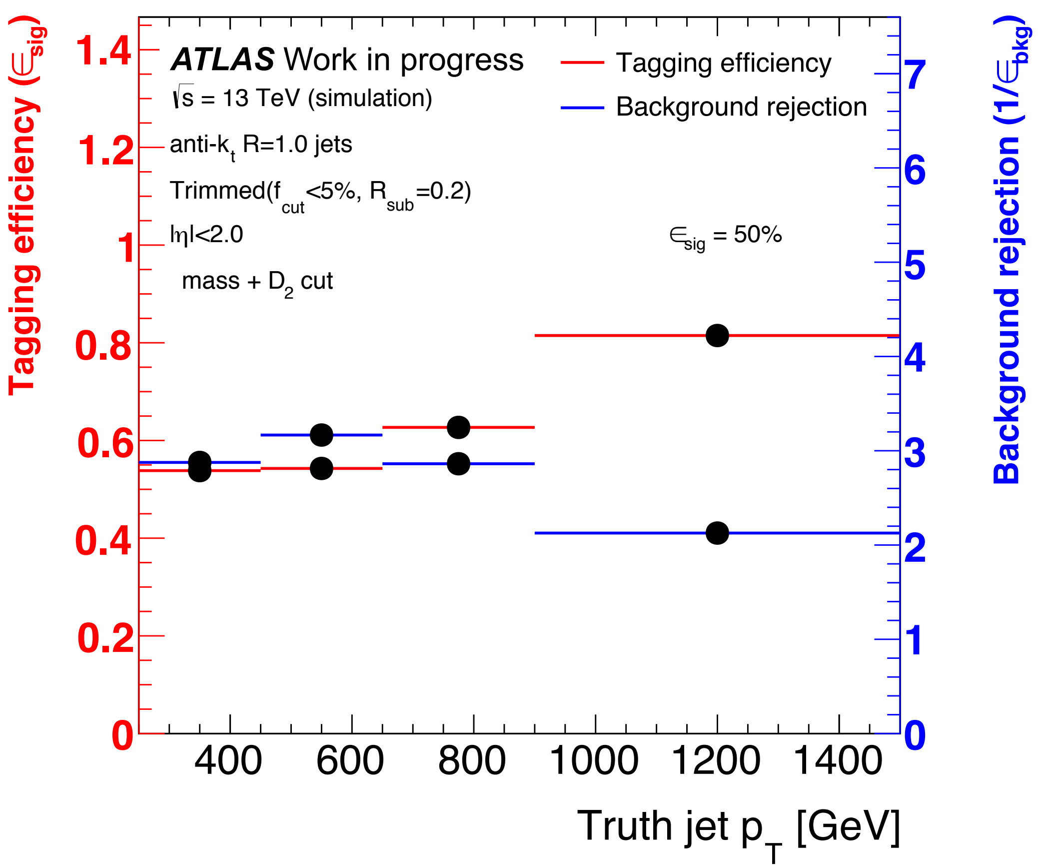
Substructure variables in the algorithm





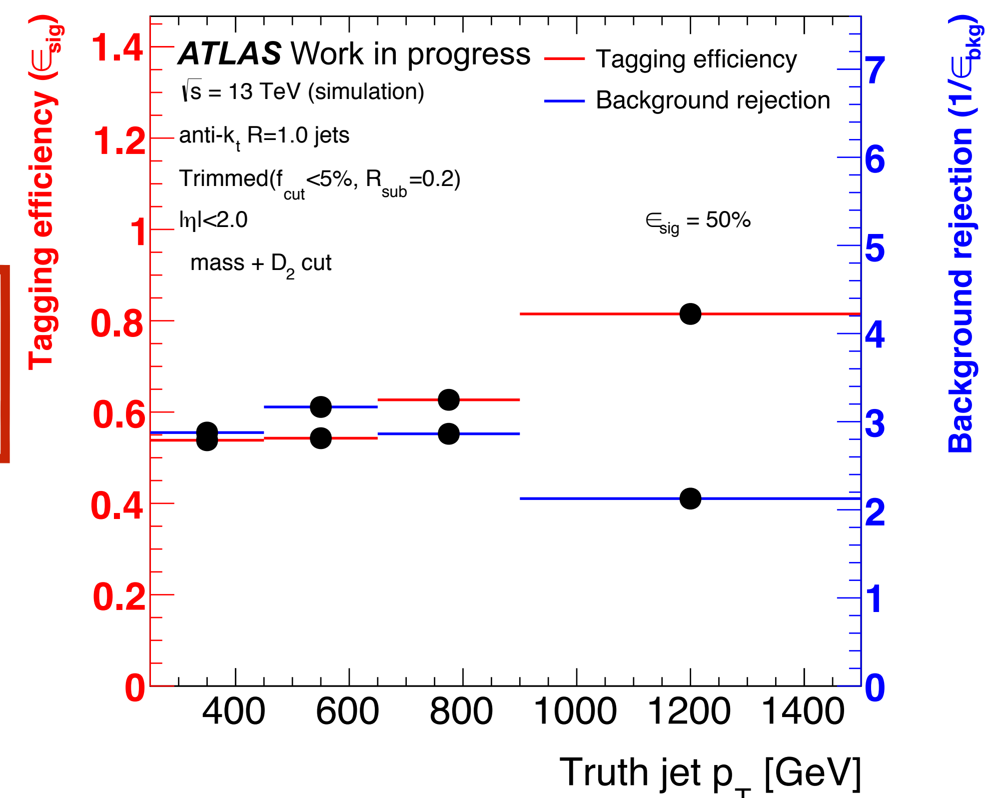
Performances

- Signal efficiency on ttH and background rejection on tt;
- **only one working point**: 50% signal efficiency.

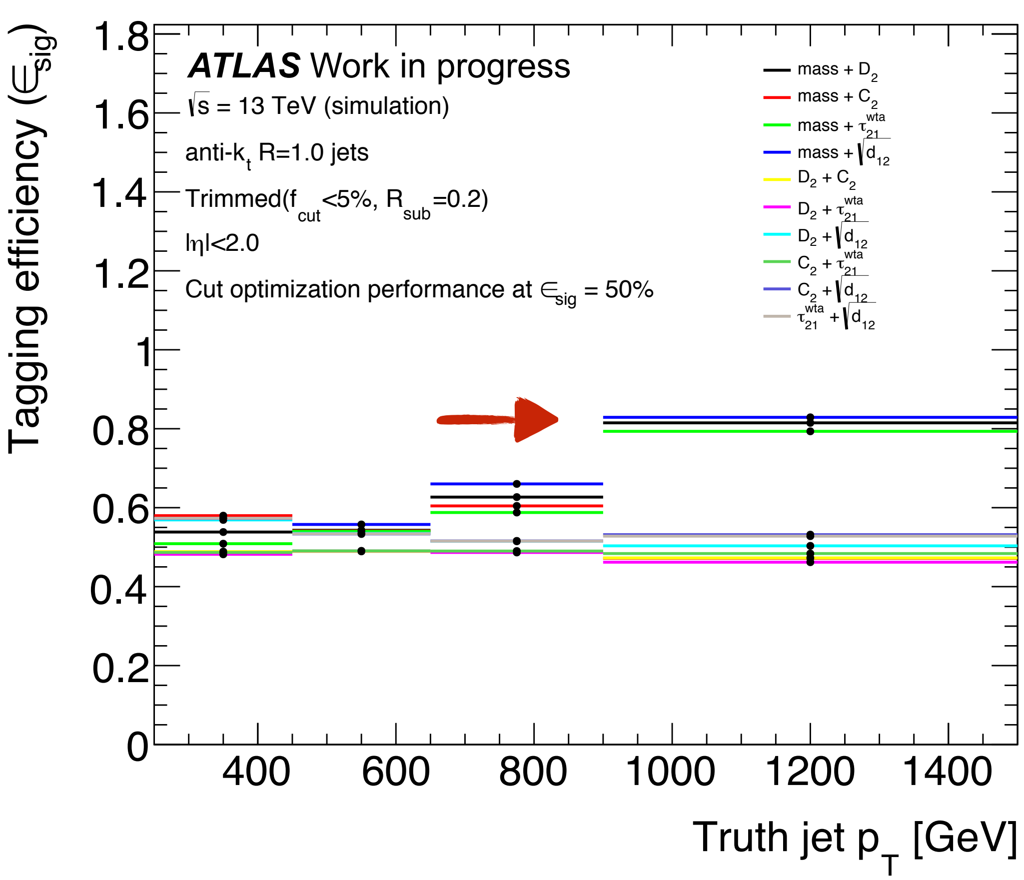


Higgs Tagging performances

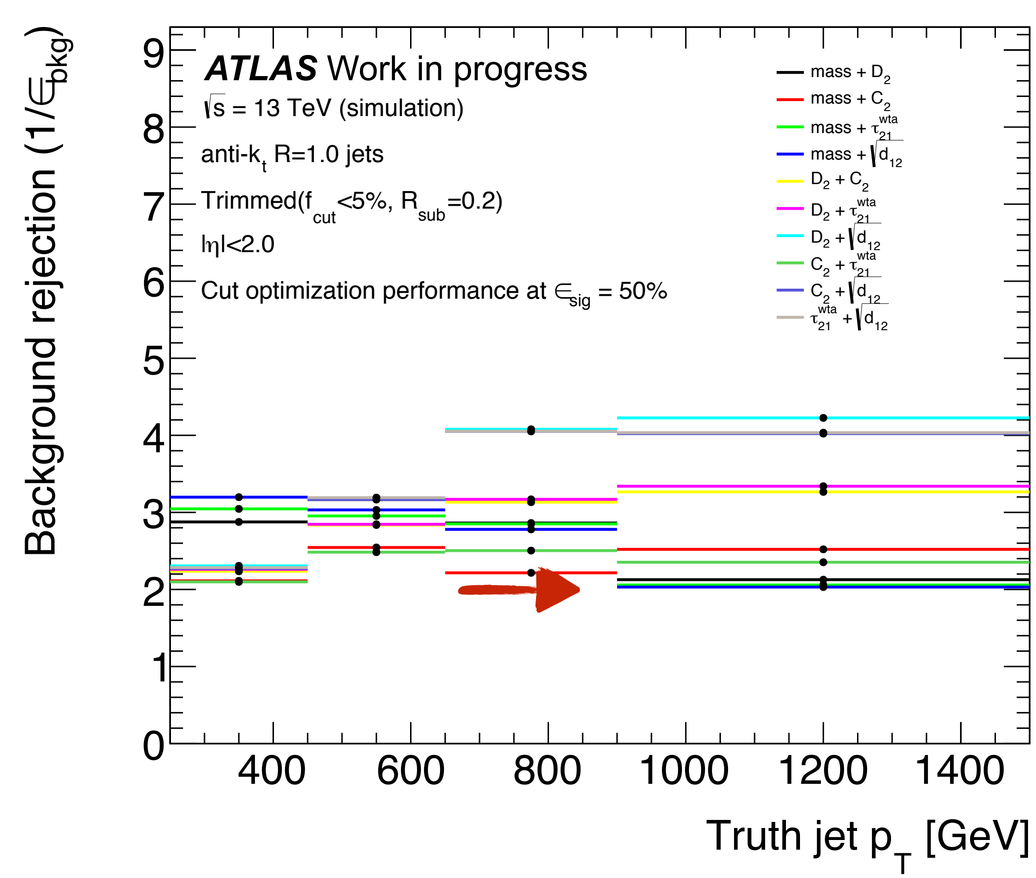
Aim: **discriminate ttH from tt**.



- Different variables combinations studied:
  - correlations, cuts, signal efficiencies and background rejections;
  - **m-D_2** seems the best choice (slightly better than **m-τ_{21}**).



Only one working point (50%).



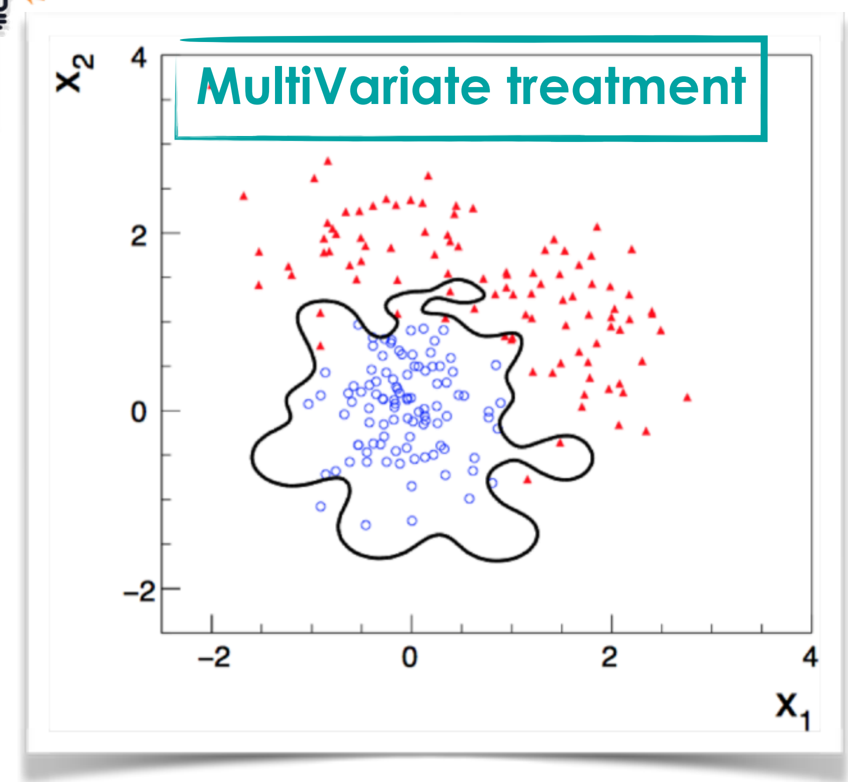
MultiVariate Analysis

Problem

- Analysis aim: to identify **events** that are both **rare** and **overwhelmed** by a wide variety of processes that mimic the signal.
- Conventional approach by using cuts on individual kinematic variables **far from be optimal**!

Solution: MultiVariate Analysis (MVA)

1. choice of **set of variables**, characterising an event;
2. application of **non-linear cuts** on signal and background samples;
3. define a function (**classifier**) that, using the **discriminating variables**, is able to identify each event of the real data belonging to the signal or to the background category.



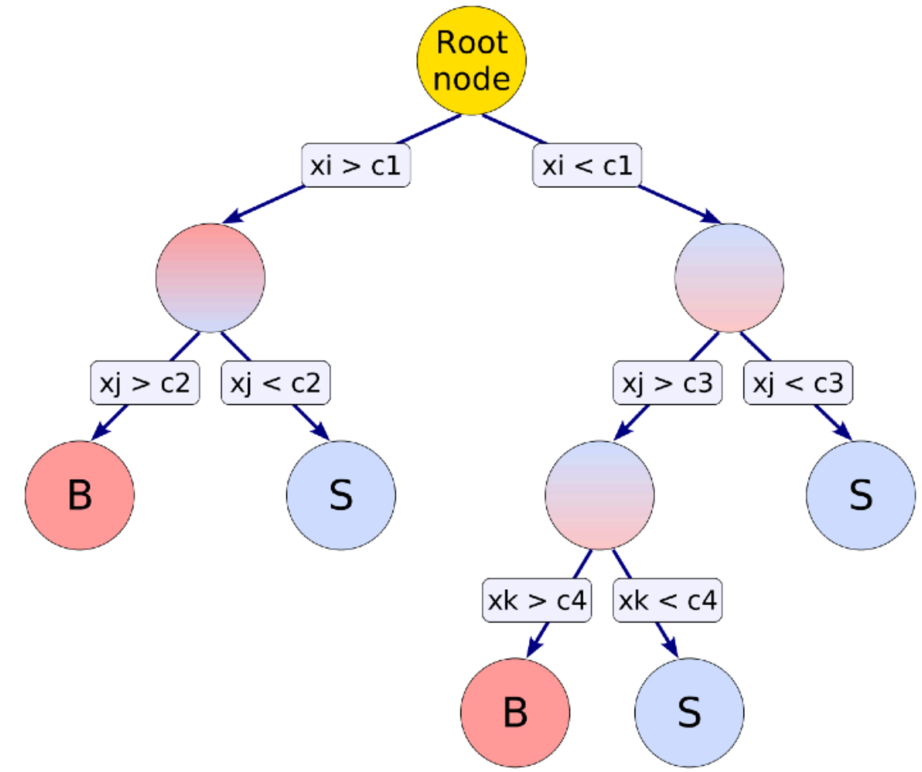
The algorithm "**learns**" signal and background characteristics (**training**) and assigns a weight to each event (~ probability that event is signal or background).

Many different algorithms available (Neural Networks, Boosted Decision Tree, Likelihood, …).

MVA in the ttH channel: boosted

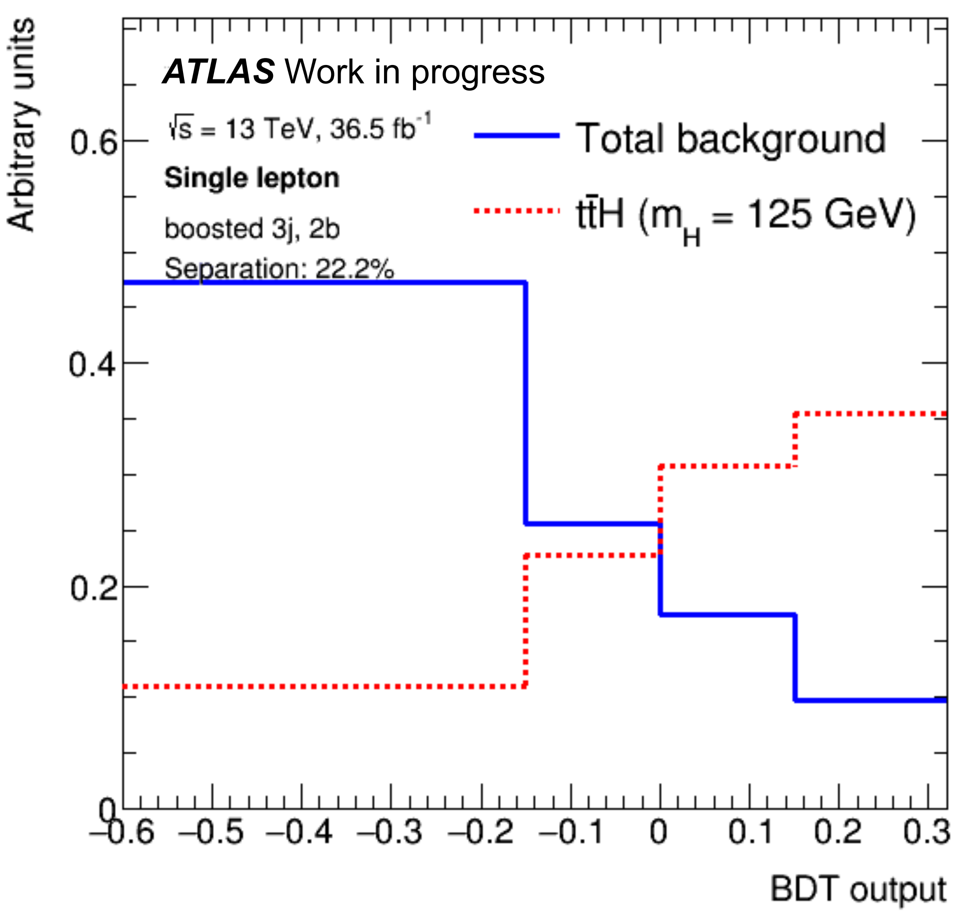
Boosted Decision Tree (BDT)

- **Decision Tree** = **sequence of binary splits** applied to the data, using discriminating variables.
- In order to improve the performance of the algorithm, a "forest" of **binary tree structured classifier** is considered.
- The **final leaves are labelled as signal or background**, depending on the majority of events in the respective node.



Procedure in the boosted channel

1. Choice of a set of **10 kinematic variables**:
    1. optimised from a large set of initial variables, looking at the **separation power**, **importance ranking** and **correlation** amongst them;
    2. **substructure**, **Higgs reconstruction** and **global event topology** related observables;
2. **training** on signal and background events;
3. results: **BDT discrimination output** (gives the best separation between signal and background on the real data as well).



MVA in ttH

- **Separation:** $$< S^2 >= \frac{1}{2}\int \frac{(\hat{y}_S(y) - \hat{y}_B(y))^2}{\hat{y}_S(y) + \hat{y}_B(y)} dy$$
  - y_S and y_B are the signal and background probability density functions of y, respectively;
  - zero for identical signal and background shapes and 1 for shapes with no overlap.

- **Correlation:** $$\rho(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$
  - two random variables X and Y;
  - cov is the covariance and sigma(X) (sigma(Y)) is the variance of X (Y).

- **Importance ranking**:
  - by evaluating the number of times the variables are used to split decision tree nodes;
  - by weighting each split occurrence (by using the same variable) by the separation achieved and by the number of events in the node.

Signal measurement technique

Ingredients

- **BDT** distributions in SRs;
- **H_T** ($= \Sigma_i\, p_{Ti}$) distributions in CRs;
- **P** depends on estimated number of events in each bin (function of μ);
- set of parameters to model the systematics uncertainties (**Nuisance Parameters**);
- **hypothesis**: S+B (μ = 1) or onlyB (μ = 0).

Recipe

In order to **test for signal presence** in the channel:

1. build a **likelihood** as a product of P terms over all the bins of the distributions;
2. perform a **fit** in the signal and control regions;
3. find a best-fit value of the **signal strength $\mu = \sigma/\sigma_{SM}$**;
4. put a **upper limit on μ @ 95% CL**.

Systematics included:

- luminosity (4.1% for 2015+2016);
- JES and JER;
- Jet Flavour Tagging;
- Light leptons;
- Large-R jets;
- Signal modelling;
- Background modelling.

Likelihood and test statistic

- **Likelihood function**:
  - defined as the product of Poisson probabilities for all bins:

$$s_i = s_{tot} \int_{\text{bin } i} f_s(x; \boldsymbol{\theta}_s) dx$$

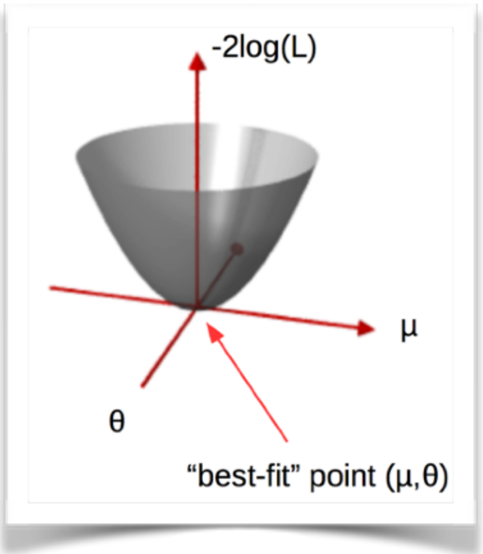
$$b_i = b_{tot} \int_{\text{bin } i} f_b(x; \boldsymbol{\theta}_b) dx$$

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}$$

  - To test a hypothesized value of μ, the **profile likelihood ratio** is considered: $\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\hat{\boldsymbol{\theta}}})}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})}$

- **Test statistic q_μ**:
  - for the purpose of establishing an upper limit on the strength parameter μ, it is defined as

$$q_\mu = \begin{cases} 0, & \mu < \hat{\mu} \\ -2\ln\lambda(\mu), & \mu \geq \hat{\mu} \end{cases}$$



  - Higher values of q_μ represent greater incompatibility between the data and the hypothesized value of μ.

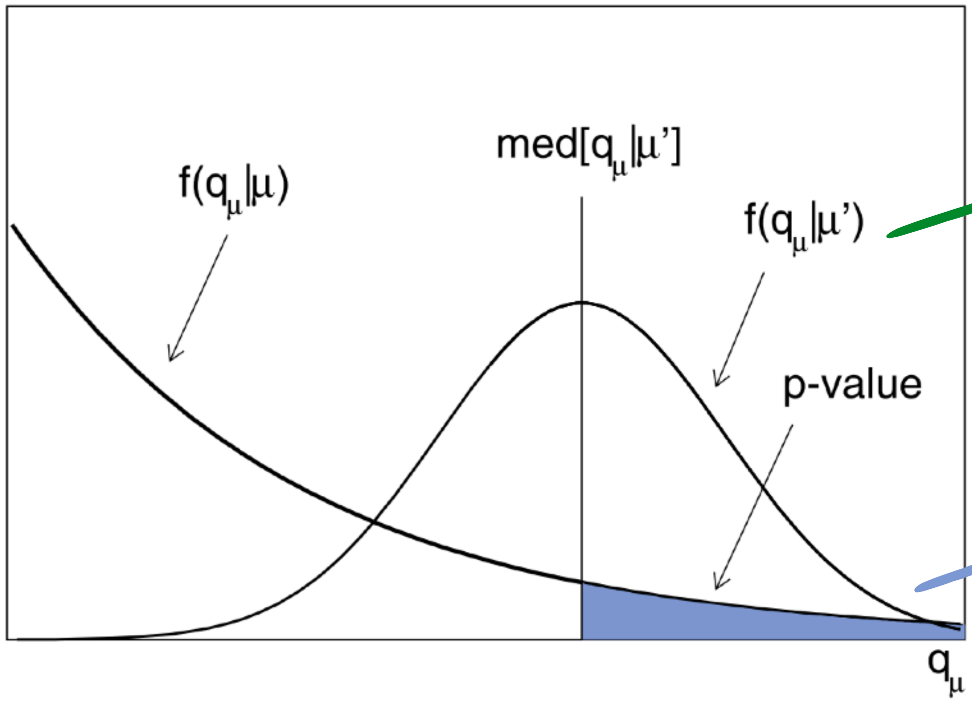
- **P-value**:
  - quantifies the level of agreement between the data and the hypothesized μ.

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu|\mu) dq_\mu$$

  - the value of μ, for which the median p-value is equal to 0.05, gives the median upper limit on μ at 95% confidence level.

Likelihood and test statistic

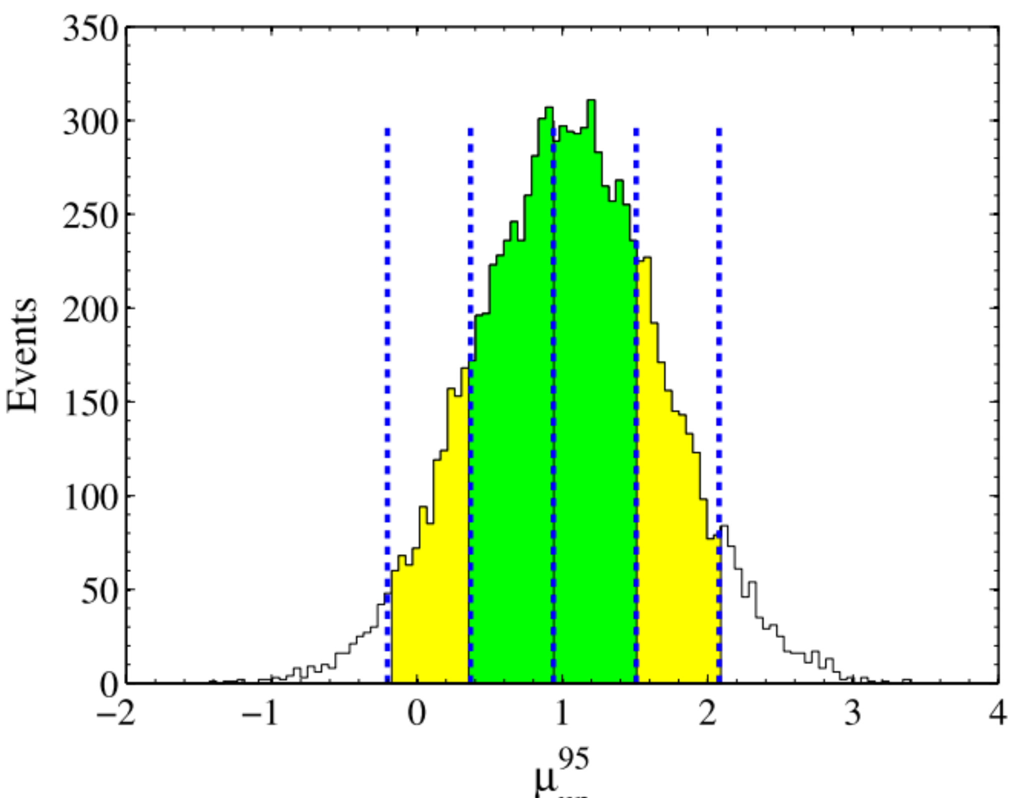
pdf for q_μ assuming both a strength parameter μ and a different value μ'



$f(q_\mu\,|\,\mu')$ is shifted to **higher value of q_μ**, corresponding on average to **lower p-values**

The **sensitivity of an experiment** can be characterised by giving the **p-value corresponding to the median q_μ assuming the alternative value μ'**

The upper limit on μ at a confidence level CL=1–α is the value of μ for which the p-value is p_μ =α.



By simulating the experiment many times with Monte Carlo, it is possible to obtain a histogram of the upper limits on μ at 95% CL.

The ±1σ and ±2σ error bands are obtained from the MC pseudo-experiments.

The **vertical lines** indicate the error bands as estimated directly without Monte Carlo simulation.

CL_s method

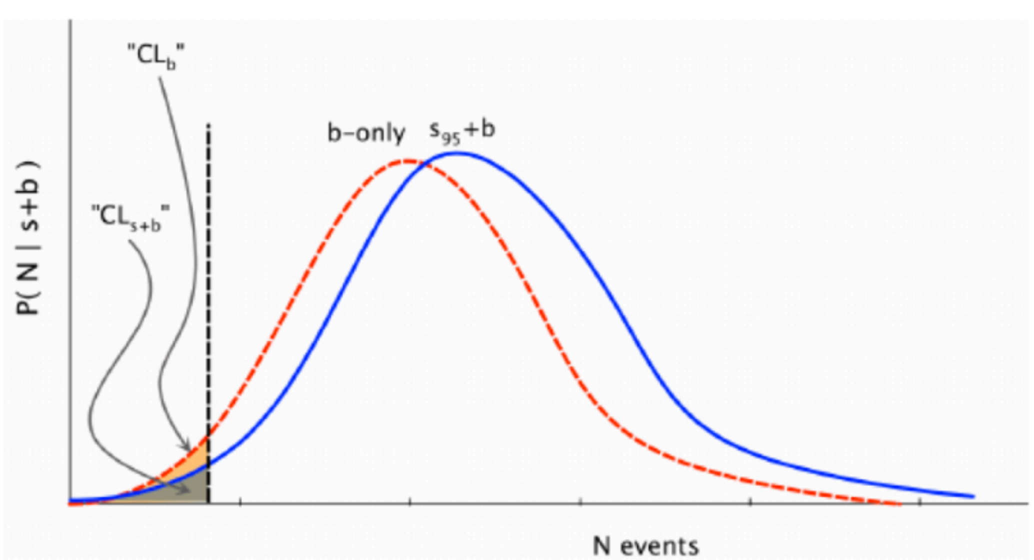
- **Modified Frequentist CL_s method**:
  - confidence level for **excluding the possibility of signal on top of background** (the s+b hypothesis):

$$\alpha_{\mathrm{s+b}} = P_{\mathrm{s+b}}(q_\mu \leq q_{\mu,obs})$$

  - **probability**, assuming the presence of both signal and background at their hypothesized levels, **that the test statistic would be less than or equal to that observed in the data**;
  - confidence level for the background alone:

$$\alpha_{\mathrm{b}} = P_{\mathrm{b}}(q_\mu \leq q_{\mu,obs})$$

  - probability assumes the presence of the background only. This confidence level has been suggested to quantify the confidence of a potential discovery, as it expresses the **probability that background processes would give a number of events smaller than or equal to the number of observed candidates**.



$$CL_{\mathrm{s}} = \alpha_{\mathrm{s+b}}/\alpha_{\mathrm{b}}$$