# $t\bar{t}H$ analyses in ATLAS

## Part 1: $t\bar{t}H$ with $H \to b\bar{b}, WW, \tau\tau$

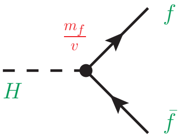
Ximo Poveda (CERN)

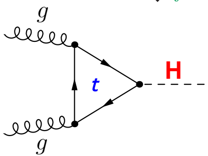
**Top Quark Physics at the Precision Frontier**

Fermilab

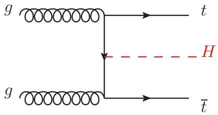
May 16, 2019

# Top quark Yukawa coupling



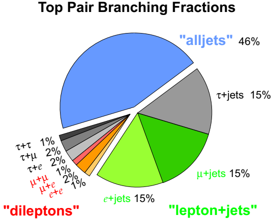

- Yukawa coupling of the Higgs boson and fermions: proportional to fermion mass
  - Top quark is heaviest fermion in the SM $\rightarrow$ Largest Yukawa coupling: $\lambda_t \approx 1$
  - Only coupling that cannot be observed in Higgs decays
- Indirect constraints on the top quark Yukawa coupling extracted from gluon-gluon fusion and $H \rightarrow \gamma\gamma$ decays
  - Resolve the loops, assuming SM contributions only



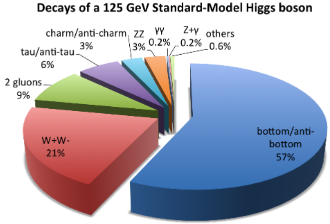

- $t\bar{t}H$ production: best direct way to measure the top quark Yukawa coupling
  - Tree-level process, cross-section proportional to $\lambda_t^2$
- **Complementary approaches, needed disentangle possible BSM effects**
- New Physics could induce deviations from the SM predictions: compositeness, vector-like quarks, SUSY, etc.
- Different higher dimension operators would affect differently ggF and $t\bar{t}H$

# Searching for $t\bar{t}H$ production

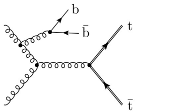
- Combination of $t\bar{t}$+$H$ decays $\rightarrow$ Complex final states, with many objects: jets, $b$-jets, light leptons ($\ell$), hadronic taus ($\tau_{\text{had}}$), photons

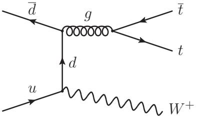



$\otimes$




- $t\bar{t}H$ cross-section: 0.5 pb
- Large irreducible backgrounds:
  - $t\bar{t}b\bar{b}$: $\mathcal{O}(15)$ pb
  - $t\bar{t}W$, $t\bar{t}Z$: $\mathcal{O}(1.5)$ pb



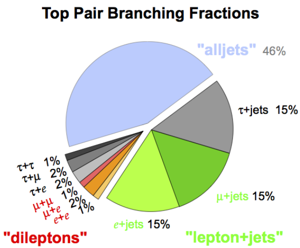

**This talk:**

- $t\bar{t}H(b\bar{b})$ $\ell$+jets (36 fb$^{-1}$): Phys. Rev. D 97 (2018) 072016
- $t\bar{t}H(WW, \tau\tau)$ multilepton (36 fb$^{-1}$): Phys. Rev. D 97 (2018) 072003

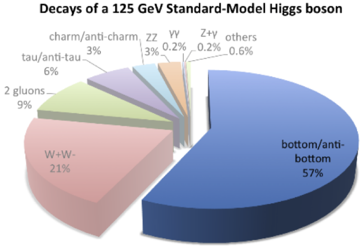
**Next talk:** $t\bar{t}H(\gamma\gamma)$

# $t\bar{t}H$, $H \to b\bar{b}$

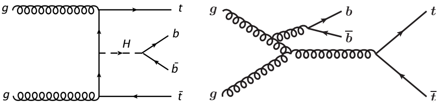
Phys. Rev. D 97 (2018) 072016




$\otimes$



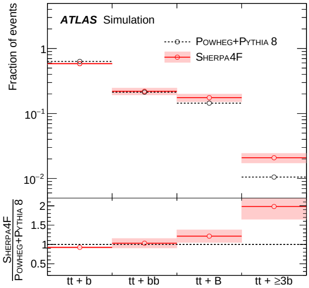
# $t\bar{t}H(b\bar{b})$: Strategy




- Large $H \to b\bar{b}$ branching ratio ($\sim$58%), and leptonic top decays
- Large irreducible background from $t\bar{t}$ + heavy flavour (HF) production

## $\mathbf{t\bar{t}}$ + HF modelling:

- Nominal MC: NLO **Powheg+Pythia8**, modelling studied with 7/8/13 TeV data
- Split in number of HF jets at particle level: $t\bar{t}$+$\geq$1$b$, $t\bar{t}$+$\geq$1$c$, $t\bar{t}$+light




- $t\bar{t}$+$\geq$1$b$: relative sub-components reweighted to $t\bar{t}b\bar{b}$ Sherpa+OpenLoops: NLO, 4-flavour scheme (massive $b$-quarks, $g \to b\bar{b}$ from ME)
- $t\bar{t}$+$\geq$1$b$ and $t\bar{t}$+$\geq$1$c$ normalizations floating
- Background modeling systematics:
  - NLO generator: Sherpa5F vs. nominal
  - Parton shower: Powheg+Herwig7/Pythia8
  - $t\bar{t}$+$\geq$1$b$: Sherpa4F vs. nominal

# Event selection

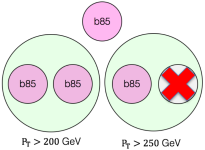
**b-tagging:**

- 4 working points: *loose* (85% eff.), *medium*, *tight*, *very-tight* (60% eff.)
- Rejection factor for *c*-jets [light jets]: 3→35 [30→1500]
- *b*-tagging discriminant built as:

| | none | *loose* | *medium* | *tight* | *very-tight* |
|---|---|---|---|---|---|
| Efficiency | - | 85% | 77% | 70% | 60% |
| Discriminant value | 1 | 2 | 3 | 4 | 5 |

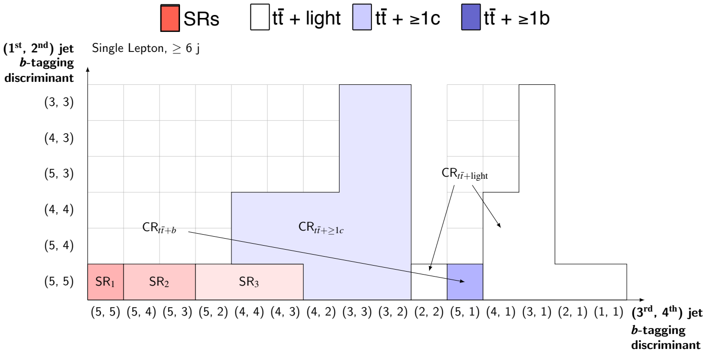
**Event classification:**

- $2\ell$ opposite-sign (OS) channel: $\geq$3 jets and $\geq$2 *medium* *b*-tagged jets




- $1\ell$ channel:
  - High-$p_T$ category:
    - 'Boosted' Higgs and top candidates (large-$R$ reclustered jets), plus a *loose* *b*-tagged jet
    - Higgs candidate ($p_T > 200$ GeV): two *loose* *b*-tagged jets
    - Top candidate ($p_T > 250$ GeV): *loose* *b*-tagged + $\geq$1 non-*b*-tagged jets
  - If failing the 'boosted' selection → 'Resolved' event:
    - $\geq$5 jets and $\geq$2 *very-tight* *b*-tagged jets or $\geq$3 *medium* *b*-tagged jets

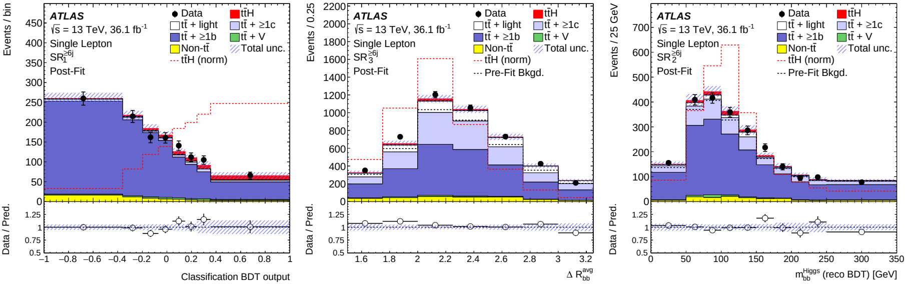
# Signal/Control regions




- Events in the $1\ell$ resolved category and $2\ell$ channel are classified in SR/CR varying the requirements on the *b*-tagging discriminant
- Nine Signal Regions
  - $t\bar{t}H$ signal purity: 1.6%-5.4%
- Ten Control Regions defined for $t\bar{t}+b$, $t\bar{t}+\geq 1c$ and $t\bar{t}$+light loosening the *b*-tagging requirements

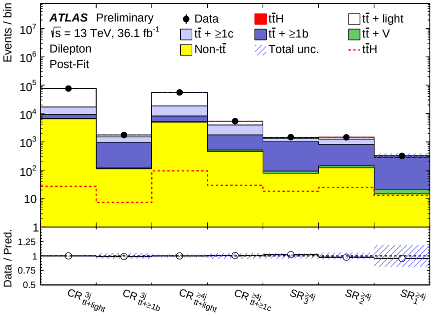
# MVA analysis: Multi-stage BDT

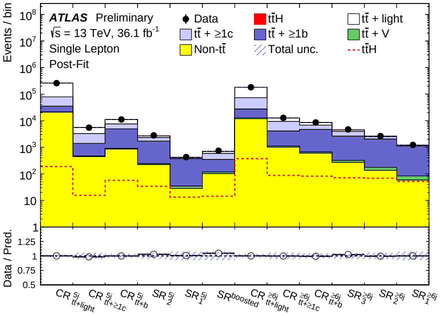
- **'Classification BDT'** to separate signal from background
- Input variables:
  - Discrete *b*-tagging discriminant
  - General kinematic variables
  - 'Reconstruction BDT' (combining jets to reconstruct $t\bar{t}H(b\bar{b})$ system): output score, variables associated to its $H$/top candidates
  - Likelihood and Matrix Element Method discriminators (only in some resolved SRs)



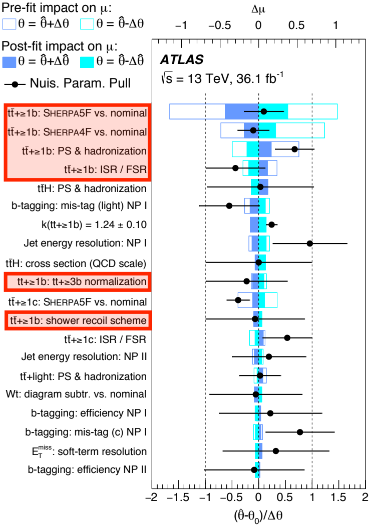
# Fit Model

- Simultaneous profile likelihood fit to all SRs and CRs
  - SRs binned in 'classification BDT'
  - CRs: single bin, except $t\bar{t}+\geq1c$ $1\ell$-CRs (binned in $H_T = \sum_{\text{jet}} p_T^{\text{jet}}$)
- Normalization factors for $t\bar{t}+\geq1b/\geq1c$ constrained in the fit, with no prior uncertainty:
  - $t\bar{t}+\geq1b$: $1.24 \pm 0.10$
  - $t\bar{t}+\geq1c$: $1.63 \pm 0.23$

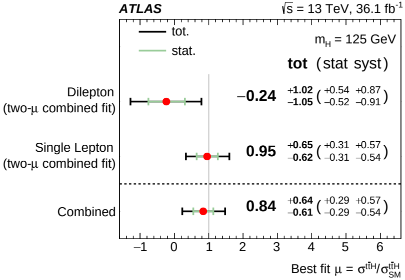


# Results

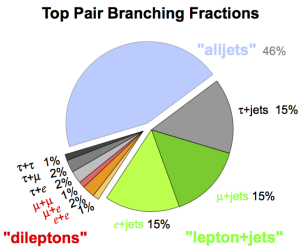



- Analysis dominated by systematics, mainly $t\bar{t}+\geq 1b$ background modelling (46% impact on signal strength $\mu = \sigma/\sigma_{SM}$)
- Constraints for nuisance parameters (NPs) associated to larger variations than observed in data
- Significance: **$1.4\sigma$** (exp: $1.6\sigma$)
- Signal strength: $\mu_{t\bar{t}H} = \mathbf{0.84^{+0.64}_{-0.61}}$

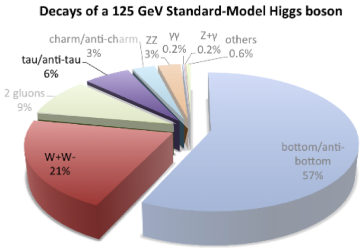


# $t\bar{t}H$ multilepton

Phys. Rev. D 97 (2018) 072003

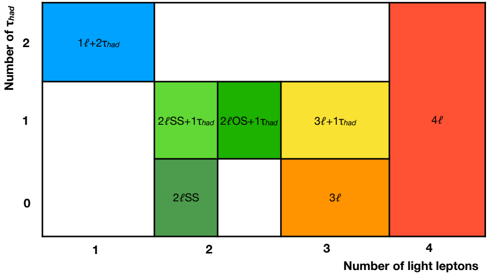


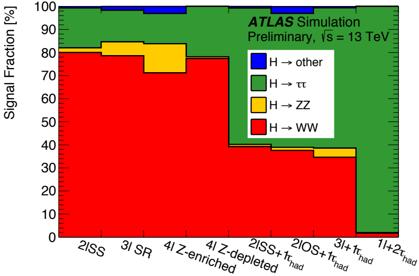

$\otimes$



# $t\bar{t}H$ multilepton: Strategy

- Targeting $H \to WW^*, \tau\tau, ZZ^*$ decay modes, combined with leptonic $t\bar{t}$ decays
- Main background: $t\bar{t} \Rightarrow$ Rely on signature with same-sign (SS) or three leptons $+$ $b$-tagged jets
- Orthogonal SRs varying the number of light leptons ($\ell$) and hadronic taus ($\tau_{\text{had}}$)

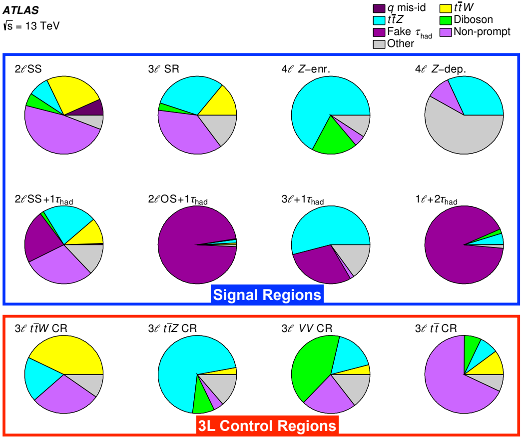



- $\ell$-only channels more sensitive to $H \to WW^*$ decays
- $\tau_{\text{had}}$ channels more sensitive to $H \to \tau\tau$

# $t\bar{t}H$ multilepton: Strategy

- Signal to background ratio ranging from few % to >40% ($4\ell$, $3\ell+1\tau$)
- Very different background contributions:
  - Fake/non-prompt light and $\tau_{\text{had}}$ leptons
  - Irreducible backgrounds: $t\bar{t}W$, $t\bar{t}Z$, other rare SM processes

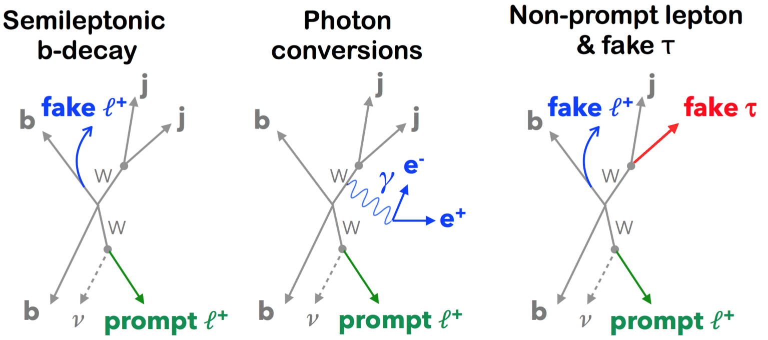



- Sensitivity enhanced with BDTs
- $2\ell SS0\tau$: combination of two BDTs
  - $t\bar{t}H$ vs. $t\bar{t}$
  - $t\bar{t}H$ vs. $t\bar{t}W/Z$
- $3\ell 0\tau$: 5-dimensional multinominal BDTs ($t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, $t\bar{t}$, $VV$) $\to$ 5 categories

# Background estimation

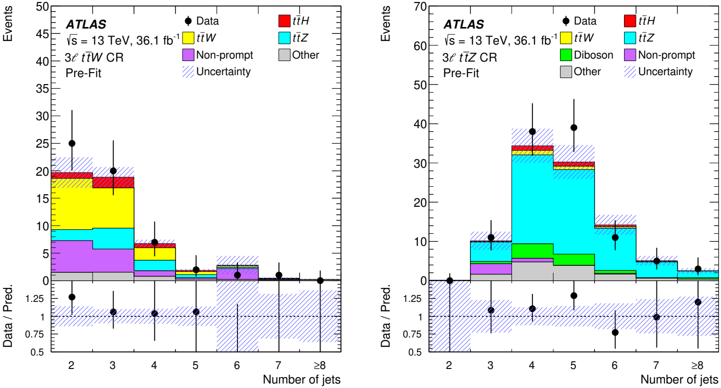
**Main background sources:**

- Processes with all prompt $\ell/\tau_{\text{had}}$ (mainly $t\bar{t}W/Z$) $\rightarrow$ Estimated with MC
- Estimated with data-driven techniques:
  - Events with fake/non-prompt light leptons:
    - Mainly from semileptonic $b$-hadron decays, also from photon conversions
  - Events with fake tau leptons:
    - Mainly from light flavour jets and electrons mis-identified as $\tau_{\text{had}}$



# Background: $t\bar{t}W/Z$

- Estimated using NLO MC samples, with theory/modelling uncertainties
- Validated in several regions, such as $3\ell$ $t\bar{t}W/Z$ CRs

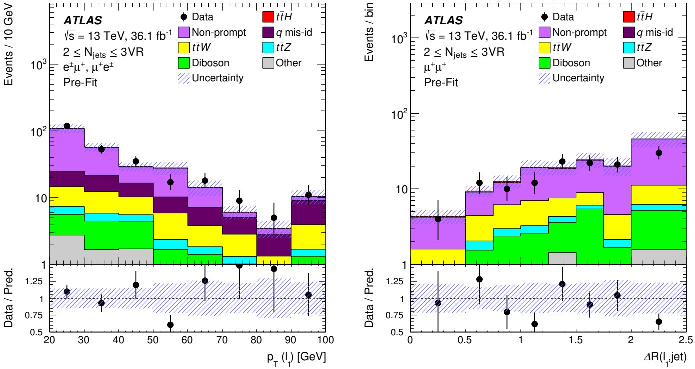



- Free-floating $t\bar{t}W/Z$ normalization: 15% loss of sensitivity for $t\bar{t}H$, $t\bar{t}W/Z$ consistent with SM predictions

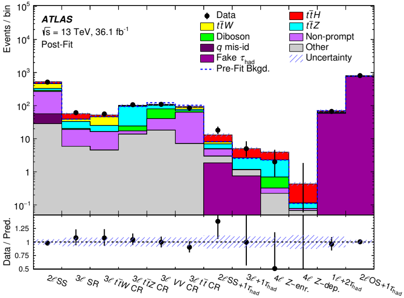
# Background: **Fake/non-prompt light leptons**

**$2\ell SS/3\ell$ channels:**

- Fully data-driven estimate with a loose-to-tight matrix method
- Real/fake efficiencies measured in $t\bar{t}$ data: $e^{\pm}\mu^{\mp}$ (real), $e^{\pm}\mu^{\pm}+\mu^{\pm}\mu^{\pm}$ (fakes)
- Validated in various regions, such as low jet multiplicity

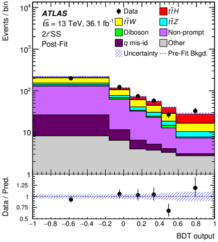


# Fit setup

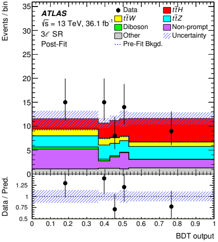



- Simultaneous fit to the 12 signal and control regions
- BDT shape used as discriminant in 5 of the SRs

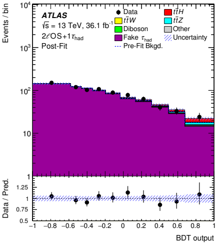
2ℓSS



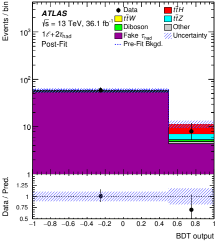

3ℓ




2ℓOS+1$\tau_{had}$

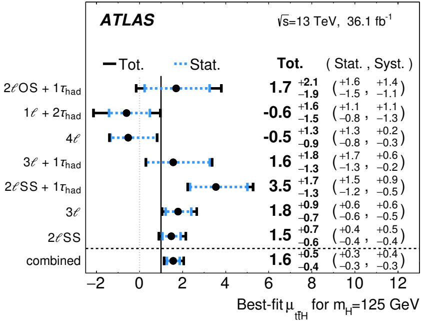



1ℓ+2$\tau_{had}$



# Results

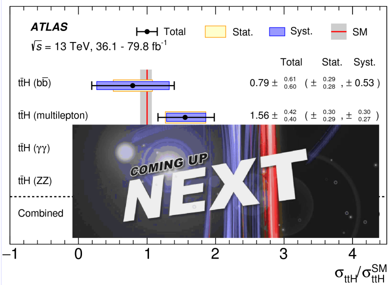
| Uncertainty Source | $\Delta\mu$ | |
|---|---|---|
| $t\bar{t}H$ modelling (cross section) | +0.20 | −0.09 |
| Jet energy scale and resolution | +0.18 | −0.15 |
| Non-prompt light-lepton estimates | +0.15 | −0.13 |
| Jet flavour tagging and $\tau_{\text{had}}$ identification | +0.11 | −0.09 |
| $t\bar{t}W$ modelling | +0.10 | −0.09 |
| $t\bar{t}Z$ modelling | +0.08 | −0.07 |
| Other background modelling | +0.08 | −0.07 |
| Luminosity | +0.08 | −0.06 |
| $t\bar{t}H$ modelling (acceptance) | +0.08 | −0.04 |
| Fake $\tau_{\text{had}}$ estimates | +0.07 | −0.07 |
| Other experimental uncertainties | +0.05 | −0.04 |
| Simulation statistics | +0.04 | −0.04 |
| Charge misassignment | +0.01 | −0.01 |
| Total systematic uncertainty | +0.39 | −0.30 |




- Most relevant uncertainties on the signal strength:
  - Signal modelling (dominated by scale uncertainties)
  - Jet energy scale and resolution
  - Non-prompt $\ell$ estimation (with large contribution from limited CR statistics)
- Significance w.r.t background-only hypothesis: **$4.1\sigma$** (exp: $2.8\sigma$)
- Signal strength: $\boldsymbol{\mu(t\bar{t}H)=1.6\pm0.3(\text{stat.})^{+0.4}_{-0.3}(\text{syst.})}$
- Most of the channels still dominated by statistical uncertainties

# Summary

- Measuring $t\bar{t}H$ is the best way to probe the top-quark Yukawa coupling
- Extensive program of Run-2 $t\bar{t}H$ searches, now turning into measurements
- $t\bar{t}H(b\bar{b})$:
  - Large signal yields but also very large and challenging $t\bar{t}b\bar{b}$ background
  - Sensitivity limited by systematics
- $t\bar{t}H$ multilepton ($H \to WW^*, \tau\tau$):
  - Good compromise between signal and background levels
  - Comparable impact of systematics and statistics for the most sensitive channels ($2\ell SS0\tau$, $3\ell SS0\tau$), other channels are limited by statistics




- $t\bar{t}H(\gamma\gamma)$ (and more): next talk by Haichen