Search for Higgs production in association with top quark pair in multilepton final state with 80fb^{-1} data collected by ATLAS detector

Based on conf note: ATLAS-CONF-2019-045

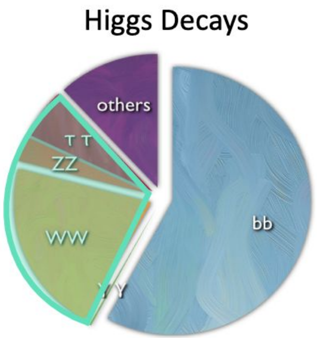
Speaker: Xuan Yang
Shandong University
CLHCP 2019
Dalian

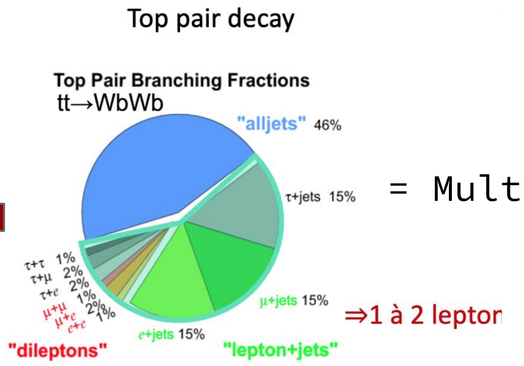
Outline

- Motivation
- Analysis overview
- Background estimate
- Result

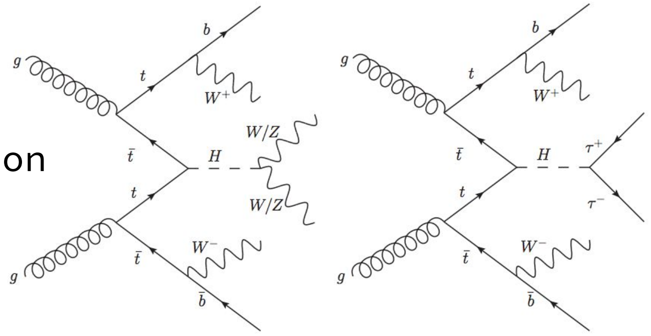
Motivation

- Direct measurement of the Yukawa coupling between top quark and Higgs boson at tree level via ttH cross section measurement
- ttH prodution cross section at 13 TeV is ~1 % of total Higgs production cross section
- Multilepton signatures at decay modes with >= 1leptons (WW, ττ, ZZ) from both top quark and Higgs boson decays

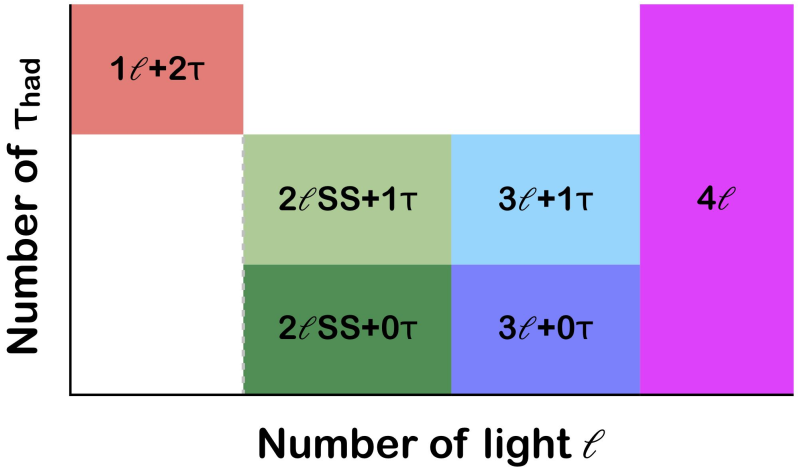




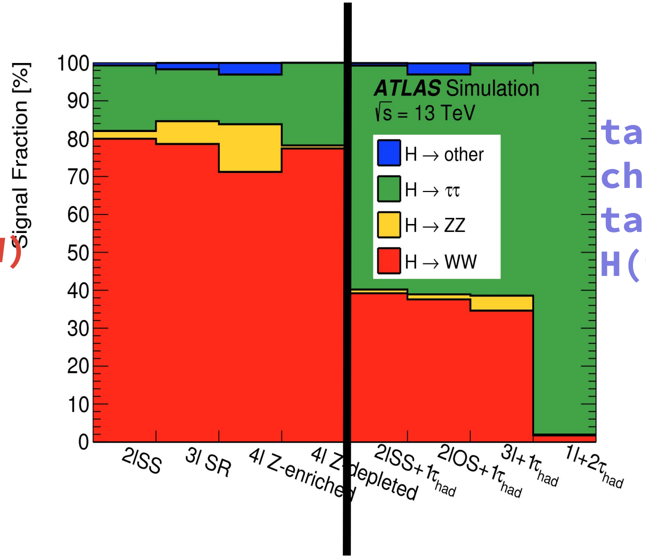
= Multilepton



Overview



0-tau: target at H(WW)



tau channels: target at H(ττ)

What we have in multilepton:

- 2/3/4 leptons (e/μ)
- 1 or 2 hadronic taus (τ_had)
- several jets (usually >=2)
- in which some are b-tagged (usually >=1)

The selections among different sub-channels are **orthogonal at object level** to avoid overlaps and to allow combination.

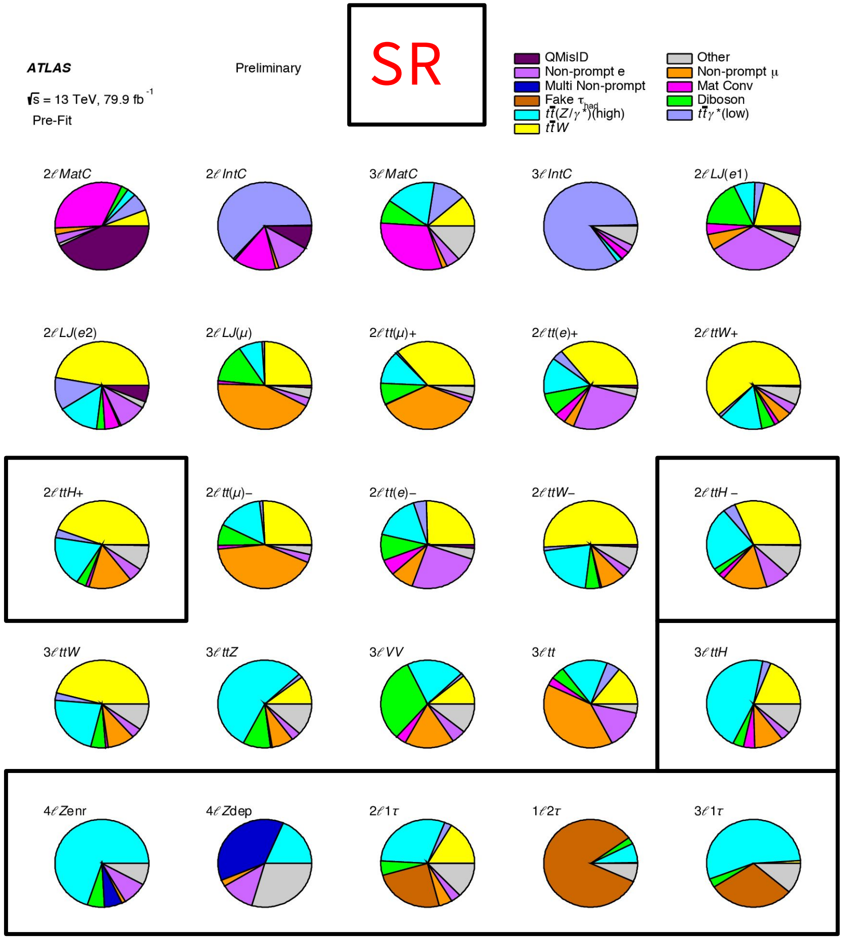
Backgrounds

- Reducible backgrounds: Events with non-prompt leptons
  - Non-prompt leptons
  - Fake tau
  - Charge flip

Important bkg in most channels

- Irreducible background: Can really mimic the signal topology, with prompt leptons
  - ttW
  - ttZ
  - Diboson
  - ...

Basically from simulation, with some corrections from CR



Non-prompt tau

- Mainly arise from ttbar/ttV events with a jet mis-identified as hadronic tau
- 2lss1tau/3l1tau: Estimate from a dedicated CR: requiring OS lepton pair, >=3 jets, >=1 bjets, 1 hadronic tau
  - A scale factor (Data/MC) is derived to correct possible mismodelling of the fake tau rate in simulation

Number of Tracks	25-45(50) GeV	45(50) - 70(75) GeV	> 70(75) GeV
1 prong	$1.05 \pm 0.04 \pm 0.05$	$0.94 \pm 0.08 \pm 0.21$	$0.64 \pm 0.10 \pm 0.07$
3 prong	$1.25 \pm 0.10 \pm 0.41$	$1.30 \pm 0.32 \pm 0.72$	$0.52 \pm 0.30 \pm 0.64$

- 1l2tau: measured from a 1l2tau CR with SS tau pair (OS in SR)
  - Jets have identical chance to be reconstructed as positve or negative charged tau
  - The estimation is taken from the SS data with small corrections from simulation samples (truth tau contribution)

Charge Mis-assignment

- The charge of electrons can be mis-identified due to:
  - Wrong assignment of EM cluster
  - Slightly curved track that induces measurement error

$$N^{os} = (1 - 2\varepsilon + 2\varepsilon^2)$$

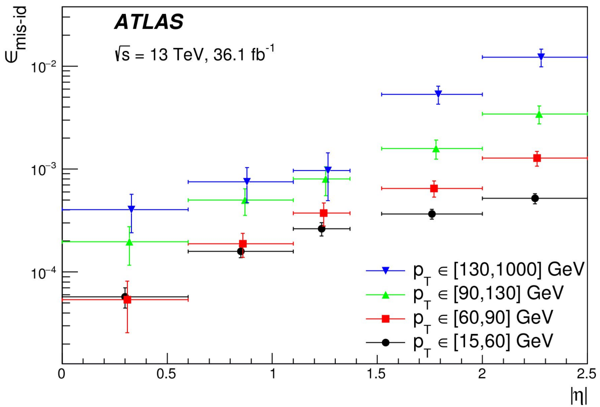
$$N^{ss} = 2\varepsilon(1 - \varepsilon)$$

→

$$N^{ss} = \frac{\varepsilon_i + \varepsilon_j - 2\varepsilon_i\varepsilon_j}{1 - \varepsilon_i - \varepsilon_j + 2\varepsilon_i\varepsilon_j} N^{os} \text{ for the } ee \text{ channel,}$$

$$N^{ss} = \frac{\varepsilon}{1-\varepsilon} N^{os} \text{ for the } e\mu \text{ channel,}$$

ε is the rate of a single electron to be mis-identified, measured from Z->ee events.
N is the total number of true OS events



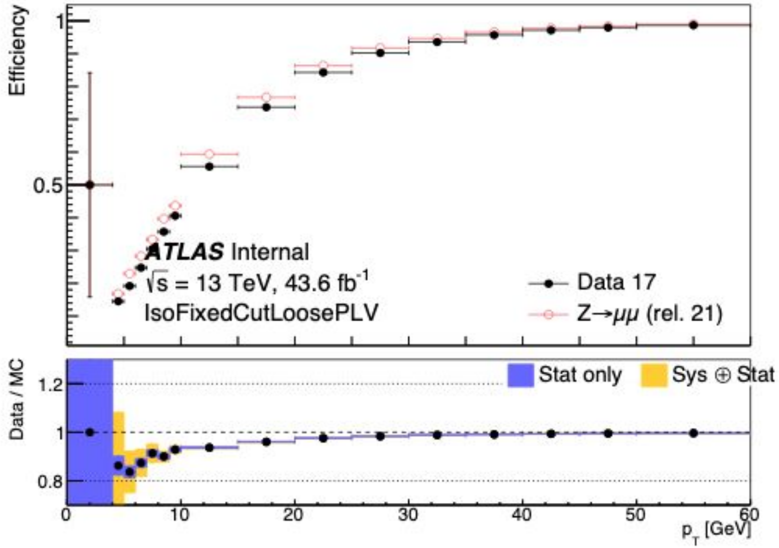
Leptons MVA

- To suppress the most important background: non-prompt leptons from semi-leptonic b decay
- Implement a new variable to reject non-prompt leptons -> PromptLeptonVeto(PLV)

PLV is a BDT trained with:

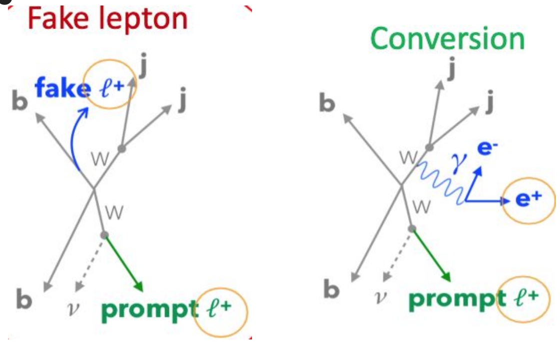
- lepton and overlapping track jets properties
- lepton track/calorimeter isolation variables

Details in Fudong He's talk yesterday!



Selection eff. of PLV for muon. Ratio plot shows Data/MC agreement

Non-prompt light leptons



- Mainly arise from:
  - Hadrons: semi-leptonic hadron decay
  - Material photon conversion (MatCO)
- In last round, Data-driven method (Matrix method/Fake factor) was used for fake lepton estimation.
- Change to a semi-DD method, **template fit**, to estimate fake leptons.
  - Shape from MC simulation
  - Normalization from fit to Data
- Hence re-define the analysis strategy

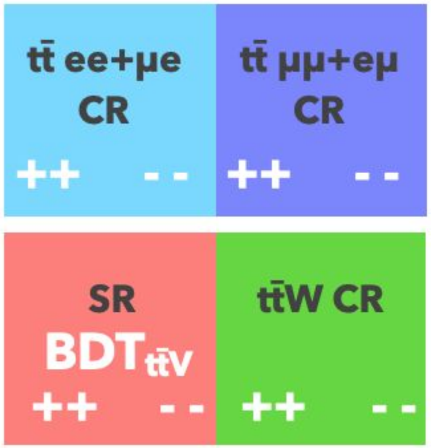
Event categorization (for 2l/3l)

- Events are categorized to control:
  - ttW enriched regions
  - Internal photon conversion (IntCO) -> decay radius < 20 mm
  - Material photon conversion -> decay radius > 20 mm
  - Non-prompt leponts from semi-leptonic hadron decay
  - Signal regions
- Each of those components is assigned by a normalization factor, and determined in a simultaneous fit to data
- ttW is also left free-float in the fit. A factor of 1.2 is applied in addition to the normal ttW cross section to account for the missing 1j@NLO and EW corrections in YR4

Event categorization (for 2l/3l)

- "High Njet region":
  - 2l: Based on a combination of 2 BDTs (vs. ttV/ vs. ttbar)
  - 3l: a multi-dimensional BDT (vs. ttW/fakes/ttZ/VV)

2ℓSS0τ [≥4j, ≥1bj]

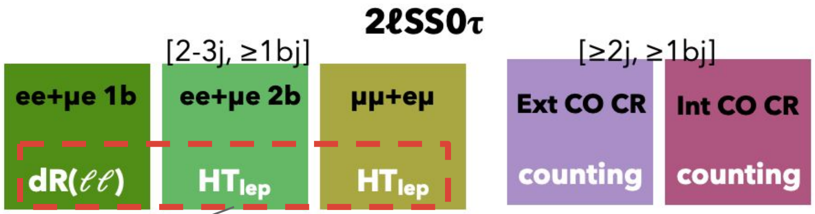


3ℓ0τ [≥2j, ≥1bj]

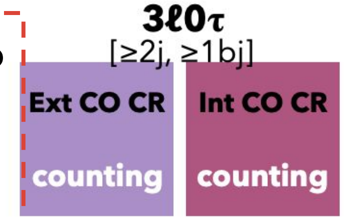
Categories based on nD BDT space, split by **lepton charge** and **lepton flavor**

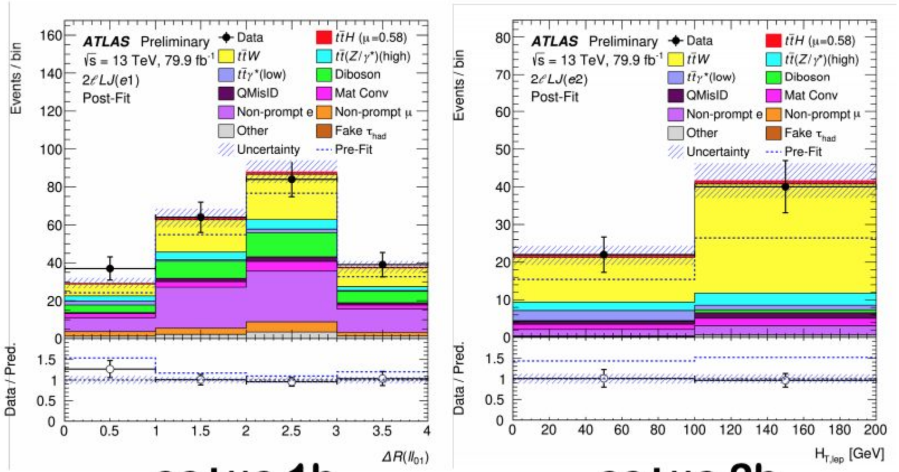
Event categorization (for 2l/3l)

- “Low Njet region”:
  - 2-3 jets (>=4 in High Njet), enriched in non-prompt leptons and ttW
- Conversion CRs:
  - >= 1 electron passing material / internal conversion selection



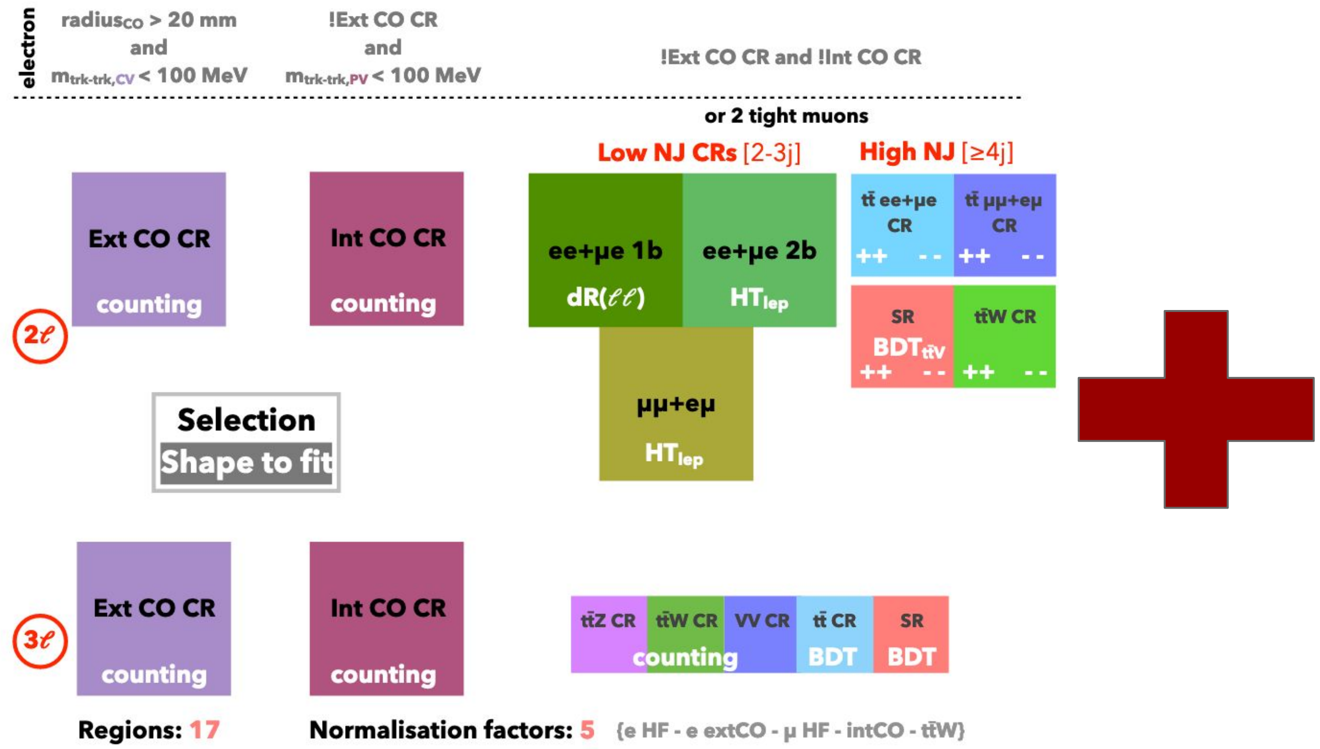
“Template” to estimate specified type of fake leptons!





ee+μe 1b ee+μe 2b

Final fit setup



Other multilepton channels:

2lss1tau
3l1tau
1l2tau
4l

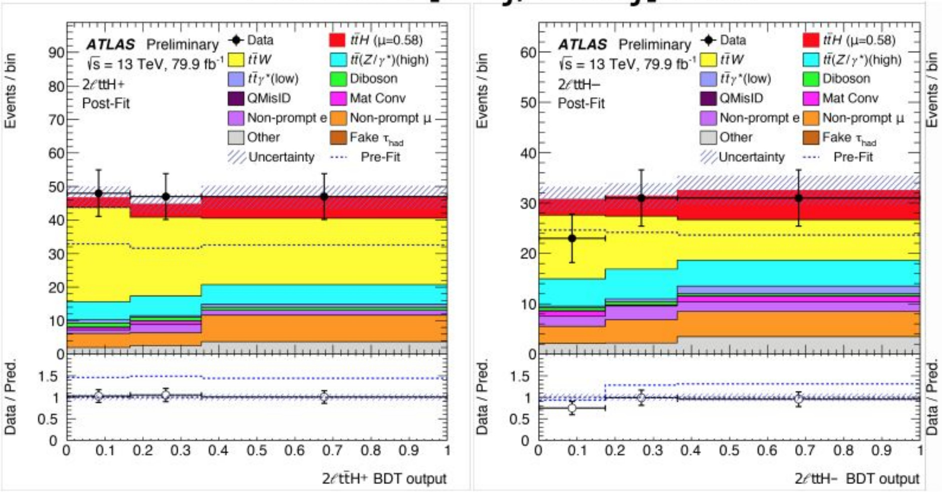
Signal extraction

- For channels with high statistics, using MVA methods
  - 2lss, 3l, 1l2tau
- For channels with limited statistics, using event counting
  - 2lss1tau, 3l1tau, 4l (One BDT is trained to define SR)

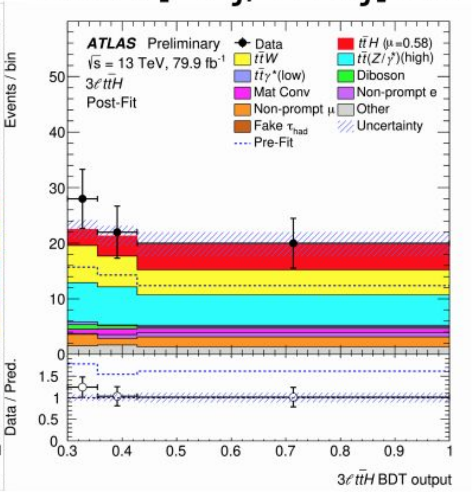
	Non-tau channels			Tau channels		
	2ℓSS	3ℓ	4ℓ	1ℓ+$2\tau_{\text{had}}$	2ℓSS+$1\tau_{\text{had}}$	3ℓ+$1\tau_{\text{had}}$
BDT trained against	Fakes and $t\bar{t}V$	$t\bar{t}$, $t\bar{t}W$, $t\bar{t}Z$, VV	$t\bar{t}Z$ / -	$t\bar{t}$	-	-
Discriminant	2D BDT	5D BDT	Event count	BDT	Event count	Event count
Number of bins in SR	4	5	1 / 1	3	1	1
Control regions	6	4	-	-	-	-

MVA distribution

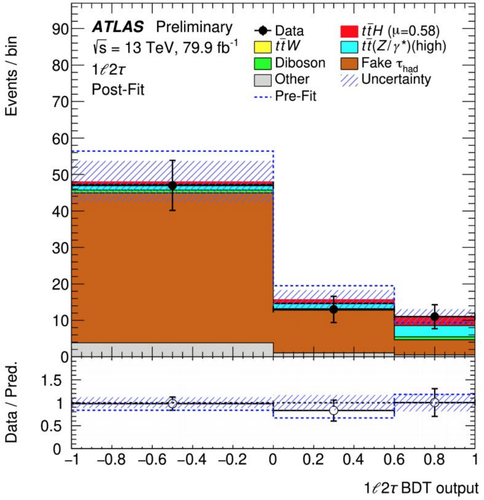
2ℓSS0τ [≥4j, ≥1bj] SRs



3ℓ0τ [≥2j, ≥1bj] SR



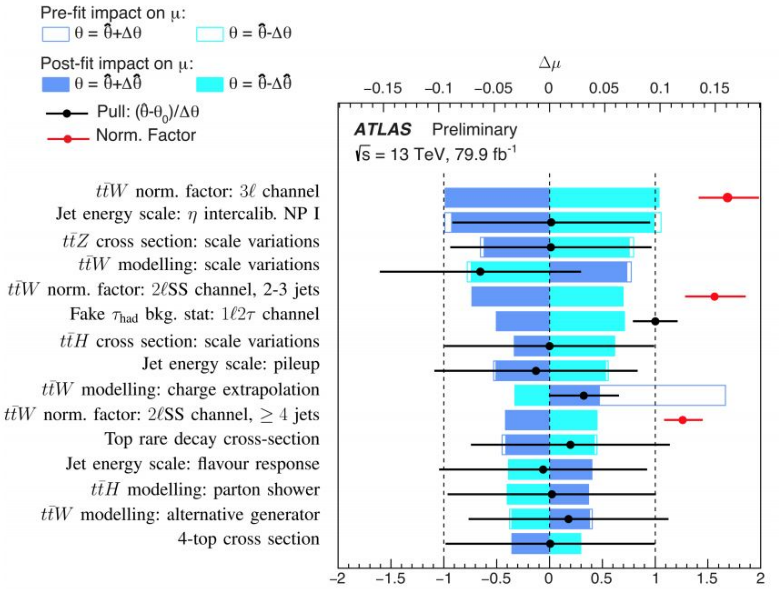
1ℓ2τ



Good agreement after fit to data. Prefit distribution is shown in dash line.

Systematic uncertainties

Uncertainty source	$\Delta\hat{\mu}$	
Jet energy scale and resolution	+0.13	−0.13
$t\bar{t}(Z/\gamma^*)$ (high mass) modelling	+0.09	−0.09
$t\bar{t}W$ modelling (radiation, generator, PDF)	+0.08	−0.08
Fake τ_{had} background estimate	+0.07	−0.07
$t\bar{t}W$ modelling (extrapolation)	+0.05	−0.05
$t\bar{t}H$ cross section	+0.05	−0.05
Simulation sample size	+0.05	−0.05
$t\bar{t}H$ modelling	+0.04	−0.04
Other background modelling	+0.04	−0.04
Jet flavour tagging and τ_{had} identification	+0.04	−0.04
Other experimental uncertainties	+0.03	−0.03
Luminosity	+0.03	−0.03
Diboson modelling	+0.01	−0.01
$t\bar{t}\gamma^*$ (low mass) modelling	+0.01	−0.01
Charge misassignment	+0.01	−0.01
Template fit (non-prompt leptons)	+0.01	−0.01
Total systematic uncertainty	+0.25	−0.22
Intrinsic statistical uncertainty	+0.23	−0.22
$t\bar{t}W$ normalisation factors	+0.10	−0.10
Non-prompt leptons normalisation factors (HF, material conversions)	+0.05	−0.05
Total statistical uncertainty	+0.26	−0.25
Total uncertainty	+0.36	−0.33

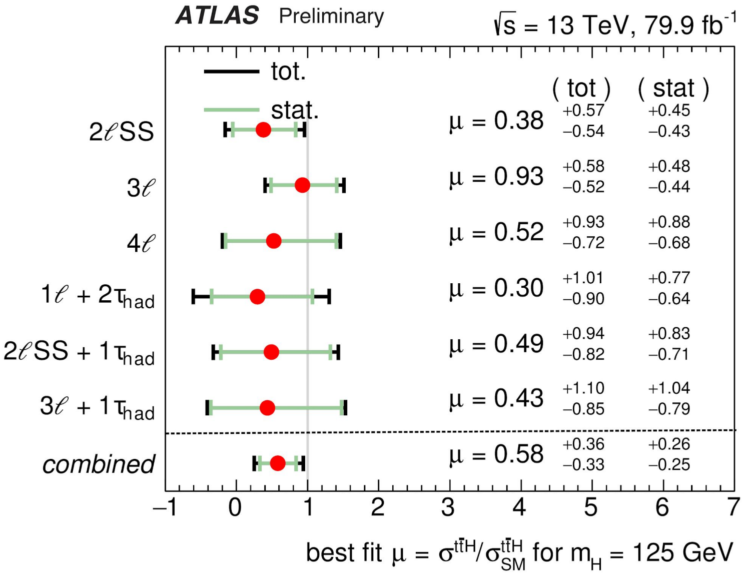


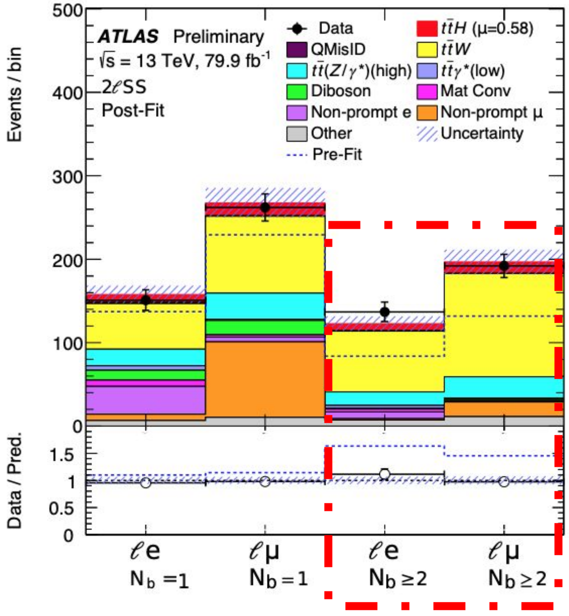
- ❖ largest uncertainties from ttW/ttll modelling
- ❖ reduced impact from fake
- ❖ JES/JER still important

Results

NF(ttW) -> already scaled by 1.2:

- 2l LowNj: $1.56^{+0.30}_{-0.28}$
- 2l HighNj: $1.26^{+0.19}_{-0.18}$
- 3l: $1.68^{+0.30}_{-0.28}$





Data Still prefer to pull up ttW even after taking into account missing scale factor of 1.2 from YR4!

Observed (expected) significance w.r.t bkg only hypothesis: **1.8σ (3.1σ)**

Conclusion

- Updated results with 80/fb data collected by ATLAS is published recently (ATLAS-CONF-2019-045)
- Re-designed event categorization and new semi-DD method for fake lepton estimation
- Also free-float ttW at the same time, and measured NF > 1
- Measured ttH cross section agree with SM prediction within uncertainty

$$\hat{\sigma}(t\bar{t}H) = 294^{+132}_{-127}\ (\text{stat.})^{+94}_{-74}\ (\text{exp.})^{+73}_{-56}\ (\text{bkg. th.})^{+41}_{-39}\ (\text{sig. th.})\ \text{fb} = 294^{+182}_{-162}\ \text{fb}.$$

V.S.

$$\sigma(t\bar{t}H) = 507^{+35}_{-50}\ \text{fb}$$

Exp. **Theory**

Thanks For Your Attention!

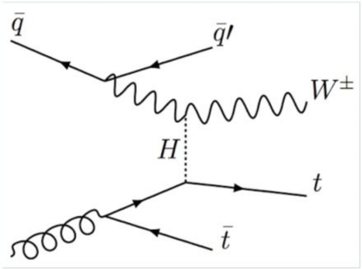
BACKUP

$t\bar{t}W$: higher order QCD and EW corrections

- A normalisation factor of **1.2** applied on top of the YR4 cross section for $t\bar{t}W$
- Origin of the correction factor:
  - Factor **1.11** to account for missing QCD corrections in higher order XS
    - ttW+0j@NLO ➜ $t\bar{t}$W+0,**1j**@NLO
    - estimated using dedicated samples generated with Sherpa 2.2.1 using the MEPS@NLO prescription, and cross-checked with the NLO generator MadGraph5_aMC@NLO 2.2.1 using the FxFx prescription
  - Factor **1.09** to account for missing EW corrections
    - [1711.02116] shows "subleading" NLO EWK corrections, not included in YR4 XS, can be large
    - primarily because of the large **NLO3 term** driven by the $t\bar{t}$W+1-jet diagrams with a Higgs boson exchanged in the t-channel

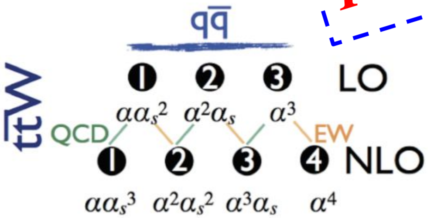
From Tamara's slide

https://indico.cern.ch/event/792576/contributions/3405586/attachments/1913568/3162767/09242019_ttH_ATLASCMS_Top2019_TVS.pdf

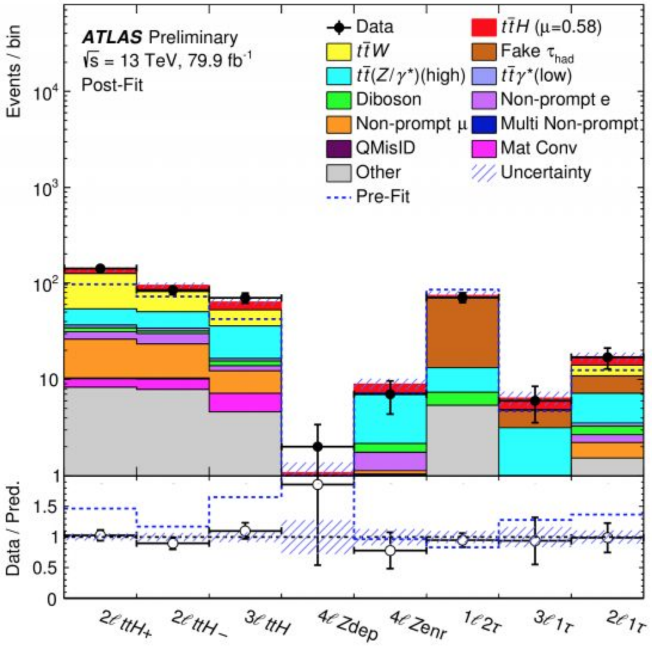


13 TeV

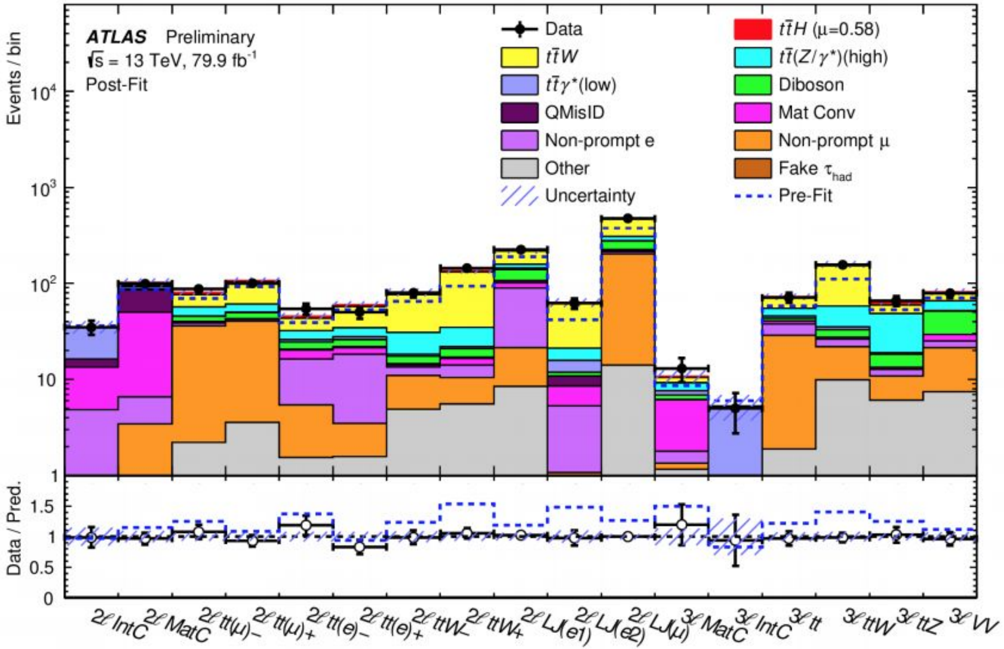
$\delta[\%]$	$\mu = H_T/2$
LO_2	-
LO_3	0.9
NLO_1	50.0 (25.7)
NLO_2	−4.2 (−4.6)
NLO_3	12.2 (9.1)
NLO_4	0.04 (−0.02)



Post-fit distributions



SR



CR

$t\bar{t}H$ (multilepton): object definition summary

	e				μ		
	L	L*	T	T*	L	L*	T/T*
Identification	Loose		Tight		Loose		Medium
Isolation	No	Yes			No	Yes	
Non-prompt lepton veto	No		Yes		No		Yes
Charge misidentification veto	No		Yes		N/A		
Material/internal conversion veto	No			Yes	N/A		
Lepton $\|\eta\|$	< 2.47			< 2	< 2.5		
$\|d_0\|/\sigma_{d_0}$	< 5				< 3		
$\|z_0 \sin\theta\|$	< 0.5 mm						

From Tamara's slide

https://indico.cern.ch/event/792576/contributions/3405586/attachments/1913568/3162767/09242019_ttH_ATLASCMS_Top2019_TVS.pdf

τ_{had}
Medium BDT ID to reject jets (1M, 1T in $1\ell+2\tau$)
$p_T > 25$ GeV
BDT to reject el faking τ
τ-μ overlap removal
b-jet veto
τ_{had} vertex is PV

Jets $p_T > 25$ GeV
BJets MV2c10 70% WP

- **Material conversion candidates** have a reconstructed displaced vertex with radius r > 20 mm that includes the track associated with the electron
  - The invariant mass of the associated track and the closest (in $\Delta\eta$) opposite-charge track reconstructed in the silicon detector, calculated at the conversion vertex, is required to be < 100 MeV.
- **Internal conversion candidates** are required to fail the requirements for material conversions, and the di-track invariant mass, this time calculated at the primary vertex, is also required to be < 100 MeV.
- **Very tight** electron candidates are tight electrons that fail the internal conversion and material conversion requirements, and have $|\eta| < 2$.
  - The latter requirement rejects a small fraction of electrons with a *large charge misidentification rate* because of the limited number of hits used in the track reconstruction.

Event selection

Channel	Selection criteria
Common	$N_{\text{jets}} \geq 2$ and $N_{b\text{-jets}} \geq 1$
2ℓSS	Two same-charge (SS) very tight (T*) leptons, $p_{\text{T}} > 20$ GeV No τ_{had} candidates $m(\ell^+\ell^-) > 12$ GeV for all SF pairs **13 categories**: enriched with $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}$, mat. conv, int. conv., split by lepton flavour, charge, jet and b-jet multiplicity
3ℓ	Three loose (L) leptons with $p_{\text{T}} > 10$ GeV; sum of light-lepton charges $= \pm 1$ Two SS very tight (T*) leptons, $p_{\text{T}} > 15$ GeV One OS (w.r.t the SS pair) loose-isolated (L*) lepton, $p_{\text{T}} > 10$ GeV No τ_{had} candidates $m(\ell^+\ell^-) > 12$ GeV and $\lvert m(\ell^+\ell^-) - 91.2$ GeV$\rvert > 10$ GeV for all SFOS pairs $\lvert m(3\ell) - 91.2$ GeV$\rvert > 10$ GeV **7 categories**: enriched with $t\bar{t}H$, $t\bar{t}W$, $t\bar{t}Z$, VV, $t\bar{t}$, mat. conv, int. conv
4ℓ	Four loose-isolated (L*) leptons; sum of light lepton charges $= 0$ $m(\ell^+\ell^-) > 12$ GeV and $\lvert m(\ell^+\ell^-) - 91.2$ GeV$\rvert > 10$ GeV for all SFOS pairs $m(4\ell) < 115$ GeV or $m(4\ell) > 130$ GeV **2 categories**: Zenr (Z-enriched;1 or 2 SFOS pairs) or Zdep (Z-depleted; 0 SFOS pairs)
$1\ell 2\tau_{\text{had}}$	One tight (T) lepton, $p_{\text{T}} > 27$ GeV Two OS τ_{had} candidates At least one tight τ_{had} candidate $N_{\text{jets}} \geq 3$
2ℓSS$1\tau_{\text{had}}$	2ℓSS selection, except: One medium τ_{had} candidate $N_{\text{jets}} \geq 4$
$3\ell 1\tau_{\text{had}}$	3ℓ selection, except: One medium τ_{had} candidate, of opposite charge to the total charge of the light leptons Two SS tight (T) leptons