New Analysis Techniques in ttH

Top LHC France 2017

Georges Aad - CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France
For the ATLAS Collaboration

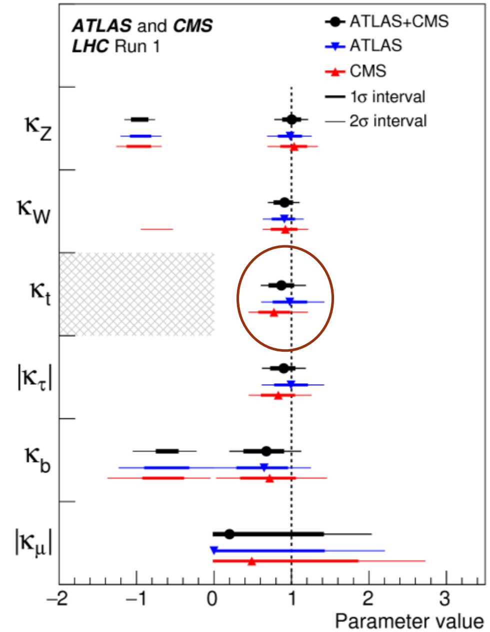
Nicolas Chanon - IPHC Strasbourg, CNRS/IN2P3
For the CMS Collaboration

Introduction

$$\kappa_t^2 = \frac{\Gamma^t}{\Gamma^t_{SM}}$$

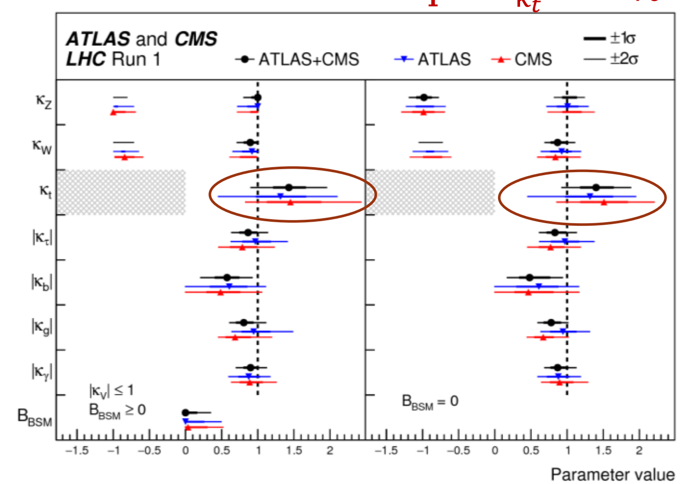
- The top quark is special vis-à-vis of the Higgs boson
  - Largest Yukawa coupling, order of one
  - Just a coincidence or there is something deeper?
- Only coupling that can't be directly measured through Higgs decay
  - Indirect access through loops in Higgs gluon fusion production and $H\rightarrow\gamma\gamma$
  - Decent constraint in run I assuming the SM particle content
- Need ttH to constrain top-Higgs coupling with a (more) model-independent assumptions
  - tH help resolve the top-Higgs coupling sign (with respect to H-W)
- ttH observed for the first time in LHC run II?
  - 5σ discovery very challenging in run II
  - Constantly improving analyses techniques to separate the signal and improving background modeling

No BSM in loops: $\sigma_{\kappa_t} \sim 15\%$



JHEP 08 (2016) 045

Possible BSM in loops: $\sigma_{\kappa_t} \sim 30\%$



JHEP 08 (2016) 045

ttH Channels Overview

- Challenging channel with small cross section
  - Exploit every accessible experimental signature
- A wide variety of final states
  - Not always corresponding to a specific production/decay mode
    - Reconstructing the event is not always possible
- 3 main groups of channels
  - $H\rightarrow bb$
  - Multiplepton (electron, muon)
    - Mainly from $H\rightarrow WW$ and $H\rightarrow ZZ$ and $\tau\rightarrow e/\mu$
    - Also exploiting channels with τ_h
  - $H\rightarrow\gamma\gamma$

Decay	BR (125 GeV)
$H\rightarrow bb$	57.7%
$H\rightarrow WW^*$	21.5%
$H\rightarrow ZZ^*$	2.6%
$H\rightarrow\tau\tau$	6.3%
$H\rightarrow\gamma\gamma$	0.2%

Don't forget subsequent W/Z decays

- Rich phenomenology of Higgs coupling
  - However hard to exploit with the low statistics in most of these channels
- More exclusive selection/splitting is available with more statistics in run II
  - Split further and target better S/B

ttH what to do

- Complex techniques targeting complex final state with low purity
  - ttH reconstruction with MVAs
  - Matrix Element Method
  - Likelihood discriminants
  - Advanced signal vs background MVAs (mainly BDTs, DNN should follow)
- Work also at object and trigger level
  - Complex MVAs at object level
    - Improve b and τ identification MVAs
    - Dedicated lepton isolation MVAs in ttH environment
  - Improvements in jet and b-jets triggers
- Adding more challenging final states
  - More channels with hadronic τ
  - ttH(bb) with all hadronic final state
- Better techniques for background estimation (not covered)

General Strategy

H→γγ and H→4l
Fit H invariant mass
Stat limited
Not covered

General tendency:
- Looser cuts
- Include more info in MVAs
- Split into regions/bins

Preselected Events

Split according to nleptons, njets, nbjets, nhadtaus, ...

Region1 | Region2 | Region3

Build discriminant variables
Combine into a final MVA

- Kinematic variables
- Signal reconstruction
- MEM
- Likelihood discriminant
- Object based MVAs (b-tag, lepton Iso, ...)

Discriminant1 | Discriminant2

Region1[1] | Region1[2]

Fit a different discriminant shape

Fit shape (several bins)

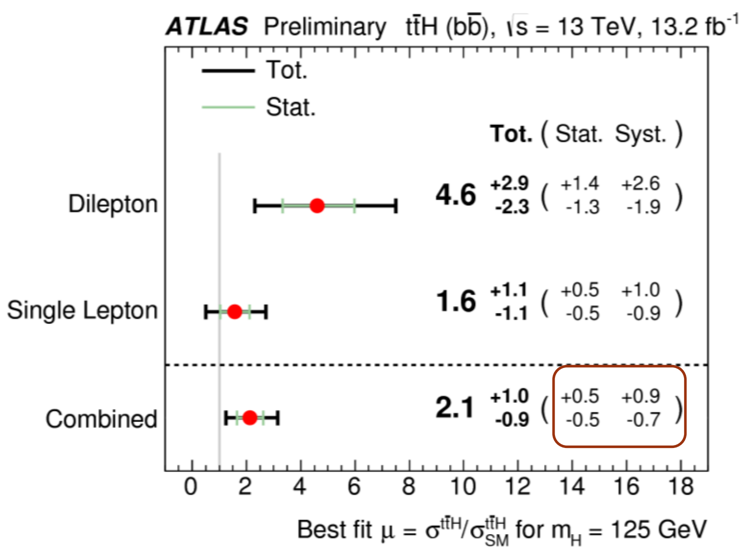
Fit norm (one bin)

ttH(bb) Results

- Main uncertainties from ttb background modeling
  - "Irreducible" background
- Need to improve signal/bkg separation
  - Increase S/B
  - Advanced MVA techniques
  - Can't just rely on the luminosity increase
- And of course need better modeling of ttbb

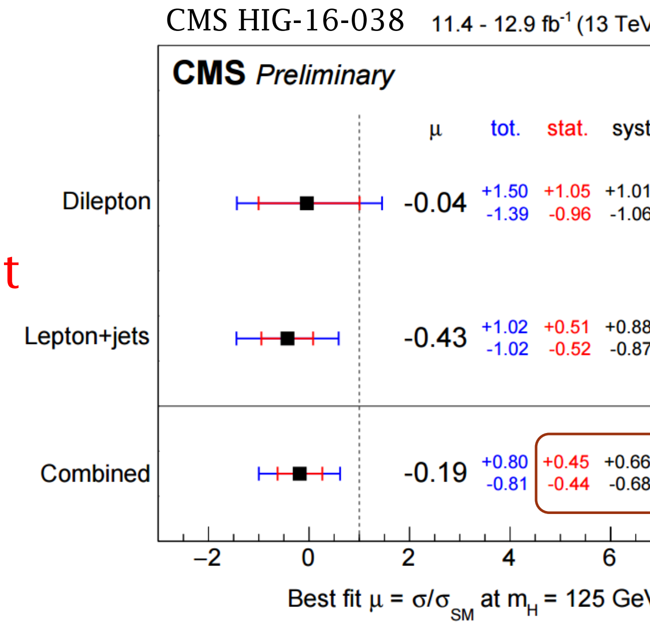
Not reaching 2σ with ~13fb^{-1}
But should do better
Improvements are on the way

ATLAS-CONF-2016-080



Results already dominated by syst uncertainties

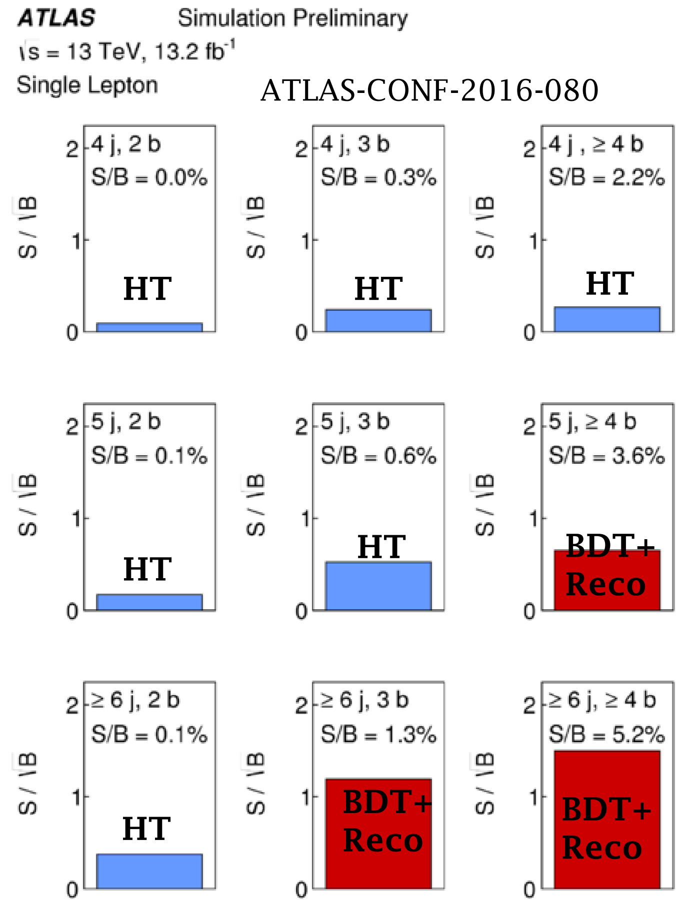
CMS HIG-16-038



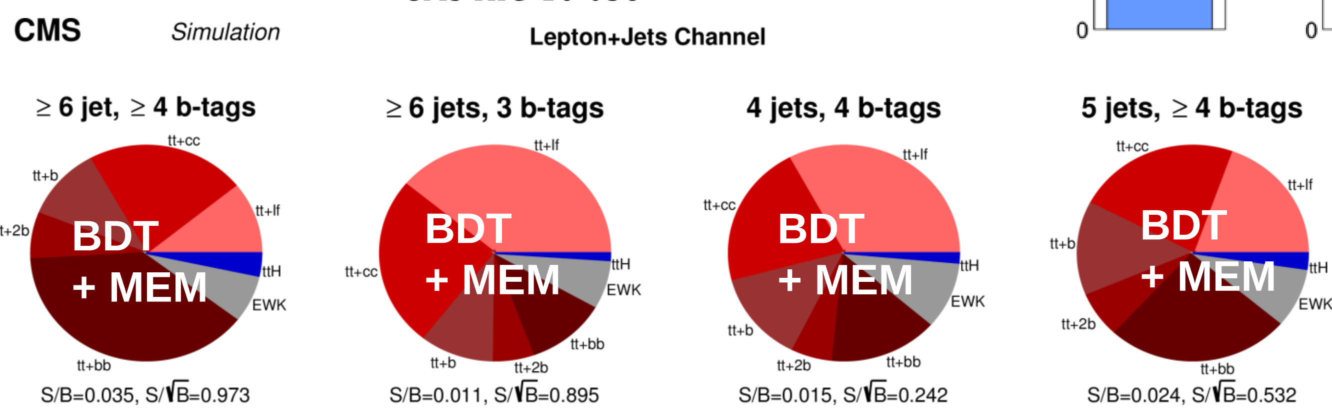
ttH(bb) Strategy

- Basic ttbar selection
  - Then require additional jets (b-jets)
- Split and conquer strategy
  - Divide sample according to the number of (b)jets
  - Exploit regions with different BKG composition
  - But still with low purity (Max: 3%-6%)
- Advanced MVA techniques in signal enriched regions
  - Signal depleted regions to control bkg systematics

ATLAS-CONF-2016-080



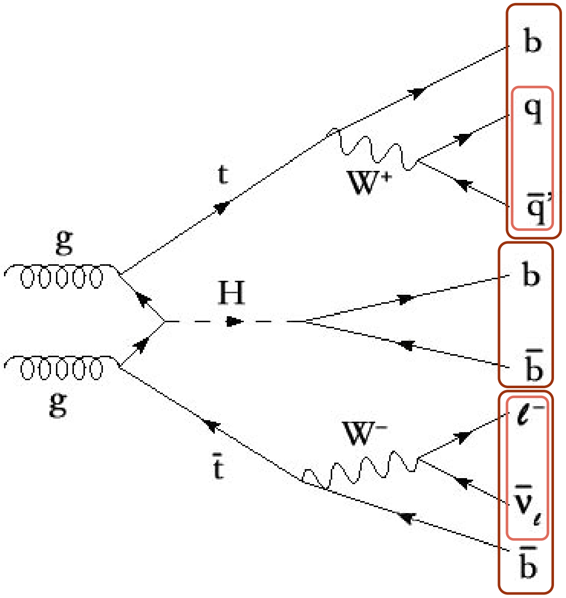
CMS HIG-16-038



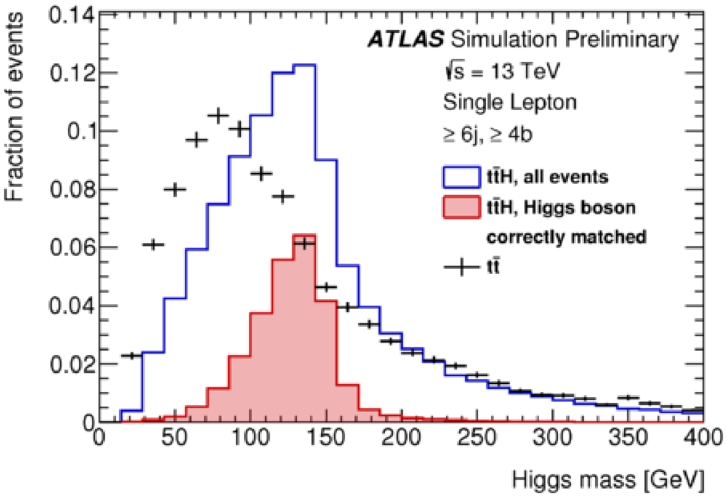
Dilep categories in backup

ttH(bb) Reconstruction

- Main difference between ttH(bb) and ttbb
  - "additional" bb-system
- Differences diluted in combinatorics
  - Many (b)jets in the final state
- Try to reconstruct the ttH system using an MVA
  - BDT trained on ttH to select the "correct" combination
  - Can reach ~40% purity for finding the correct Higgs boson candidate
  - Reconstruction BDT output and reconstructed ttH kinematics to separate signal and background
    - Input to final discriminant



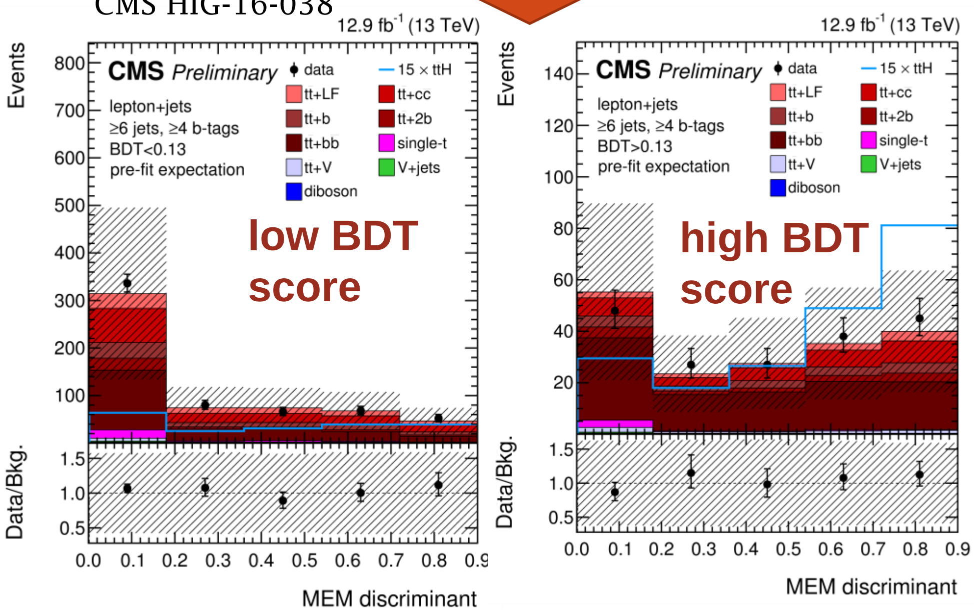
ATLAS-CONF-2016-080



ttH(bb) MEM and Likelihood discriminant

- Powerful technique that take advantage of full matrix element information
  - Maybe also the most "elegant"
- However very complex in a process like ttH(bb)
  - In many cases the jets in the final state do not correspond to the simple LO view
  - In addition to the usual problems with C(G)PU time and transfer functions)
- Can be used together with BDTs to increase the performance
  - Include MEM as a variable in a final BDT
  - Fit in "2D" with BDT

CMS HIG-16-038

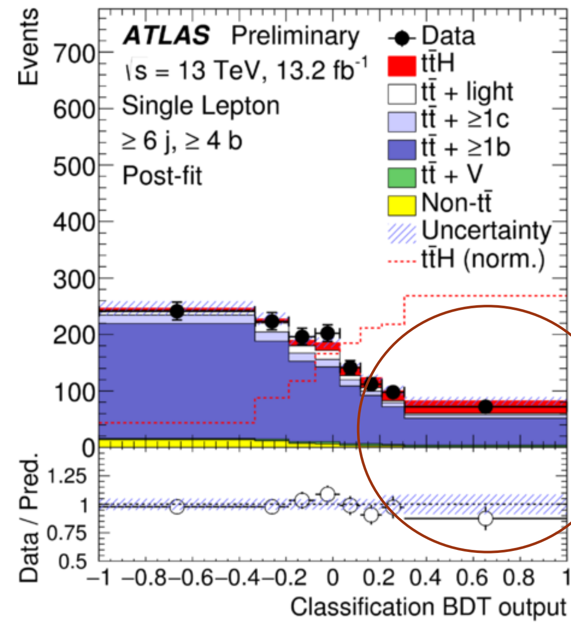


- Can also build a similar method at reconstructed object level
  - Likelihood using few kinematic variables (masses, angles)
- Build probabilities out of all permutations like for the MEM
  - Add also b-tag probabilities
- Successfully tested for ttH-like final states e.g. in arxiv:1509.06047

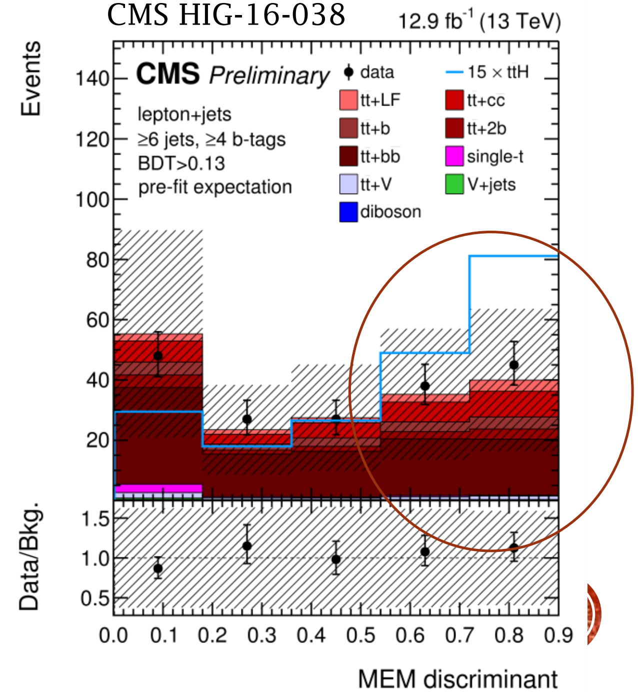
Final discriminant

- Using BDTs as final discriminant or BDT/MEM combination
- BDTs including 3 categories of variables
  - Usual kinematic variables
  - MVA outputs from different discriminants
    - MEM, likelihood, reconstruction BDT, …
  - Object identification variables
    - Most importantly jet b-tagging discriminant which is itself an MVA
- Final discriminant is used to categorize events into bins/regions before the fit to data
  - Aim to have bins with largest S/B with reasonable statistics
  - Most of the analysis power is in the last few bins of the final discriminant

ATLAS-CONF-2016-080

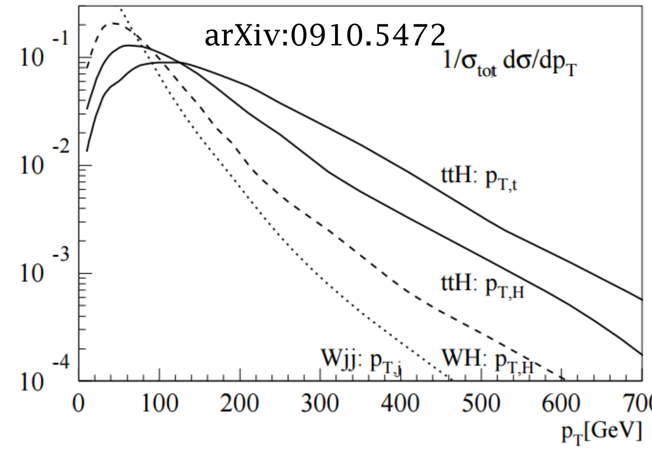


CMS HIG-16-038



ttH(bb) boosted

- Interesting category especially with increased LHC energy in run 2 and increased luminosity
  - However low stat with current luminosity
  - Analyses basically using pseudo-boosted regime
- Still not reaching "resolved" analysis performance
  - but can improve in combination with resolved
- Will definitely improve with more data
- Very interesting in some BSM scenarios

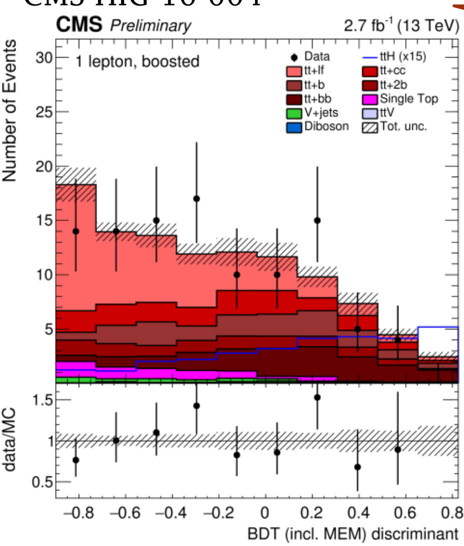


Boosted fractions @14 TeV

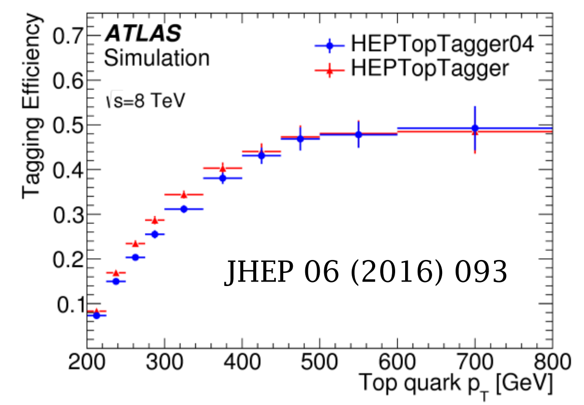
	>200 GeV	>400 GeV
Higgs	14%	1.5%
top	31%	5%

Analysis with 2.7fb^{-1} (Moriond2016, 2015 data)
Not included in latest CMS results

CMS HIG-16-004



Category	Observed	Expected
4 jets, 3 b-tags	14.5	$18.6^{+8.2}_{-5.5}$
4 jets, ≥ 4 b-tags high BDT output	35.7	$25.6^{+13.4}_{-8.1}$
4 jets, ≥ 4 b-tags low BDT output	86.6	$84.2^{+41.3}_{-25.8}$
5 jets, 3 b-tags	16.0	$12.3^{+5.5}_{-3.6}$
5 jets, ≥ 4 b-tags high BDT output	7.5	$10.3^{+5.6}_{-3.4}$
5 jets, ≥ 4 b-tags low BDT output	35.2	$31.9^{+16.1}_{-9.9}$
≥ 6 jets, 2 b-tags	25.4	$41.1^{+21.1}_{-13.1}$
≥ 6 jets, 3 b-tags	9.6	$7.6^{+3.3}_{-2.2}$
≥ 6 jets, ≥ 4 b-tags high BDT output	9.2	$8.3^{+4.4}_{-2.7}$
≥ 6 jets, ≥ 4 b-tags low BDT output	15.4	$18.3^{+9.6}_{-5.8}$
≥ 4 jets, ≥ 2 b-tags, boosted	7.5	$10.7^{+5.9}_{-3.5}$
lepton+jets combined	4.0	$4.1^{+1.8}_{-1.2}$

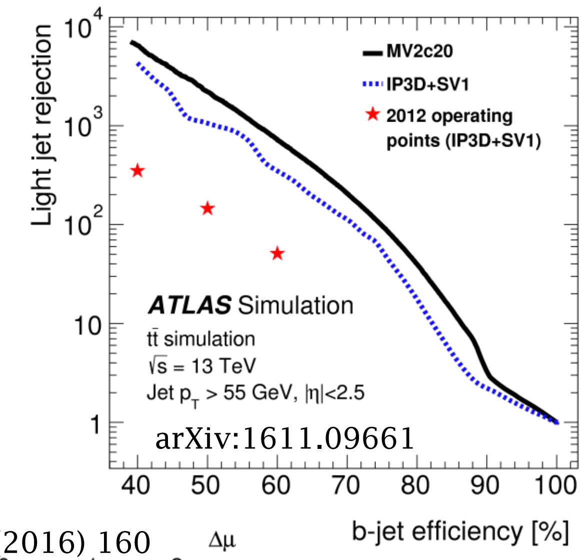


JHEP 06 (2016) 093

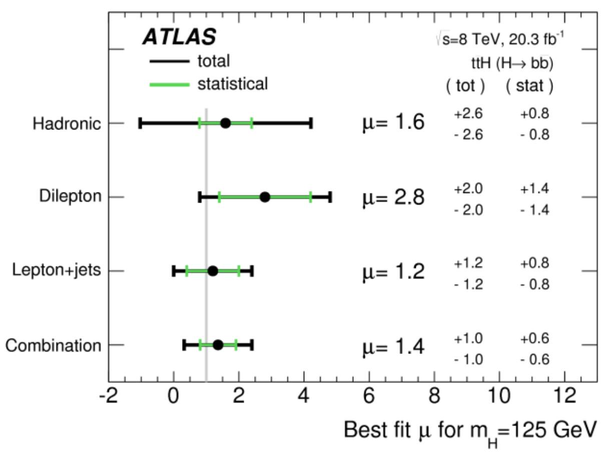
3rd most powerful individual category
Using Fat jet substructure (C/A R=1.5), pT>200 GeV
MEM using sub-jets

ttH(bb) All Hadronic Channel

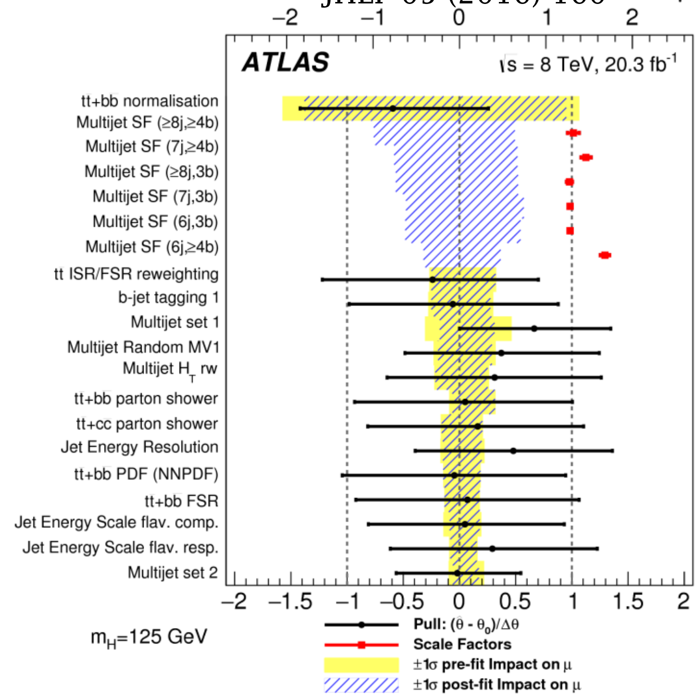
- Difficult channel due to the overwhelming multi-jet background
  - Considered only by ATLAS in run I
- Triggering is one of the key points for this analysis
  - Important benefit from improvements in (b)jets triggers
- Largest systematic impact from ttbb background
  - Even if multi-jet background is largely dominant
  - Should benefit from techniques used in other channels to separate ttbb



JHEP 05 (2016) 160



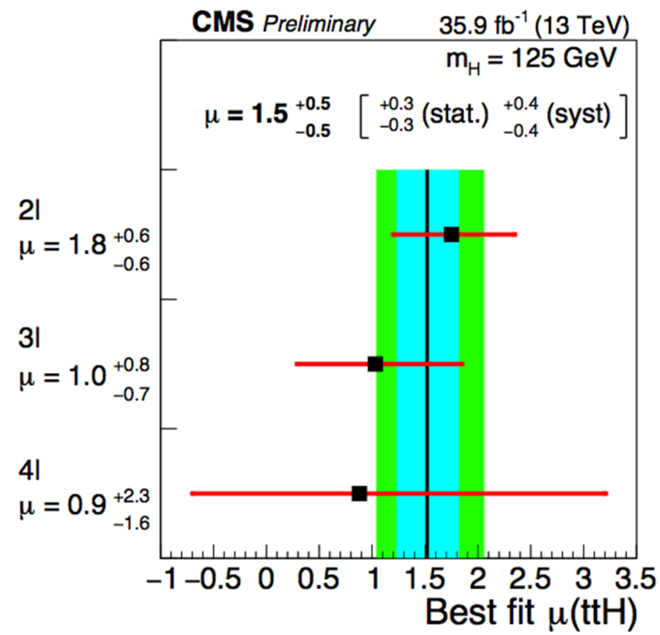
JHEP 05 (2016) 160



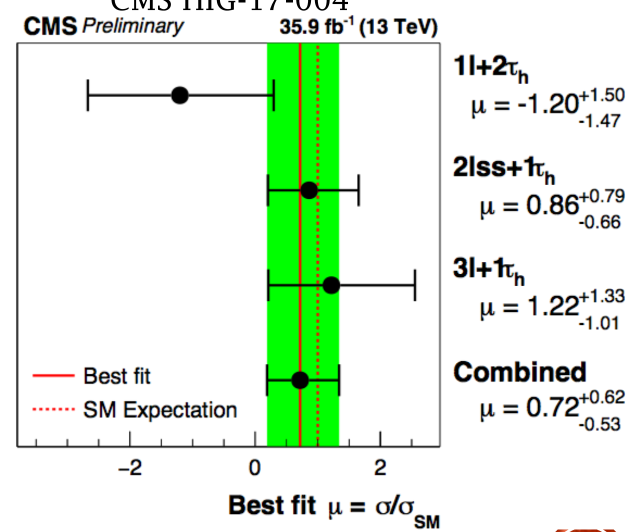
ttH multilepton situation

- Main uncertainties from fake backgrounds
  - Reducible but hard to model
- Need to improve signal/bkg separation
  - Channel with lower stat than ttH(bb)
  - But reached enough stat for advanced MVA techniques
- Also include more final states
  - Mainly more channels with hadronic τ_h
- And of course need better bkg modeling
  - Improved data driven methods
- ATLAS: 1σ expected sensitivity (13.2fb^{-1})
- CMS: 2.5σ expected sensitivity (35.9fb^{-1})
  - Not only due to increased luminosity
  - But also due to advanced techniques

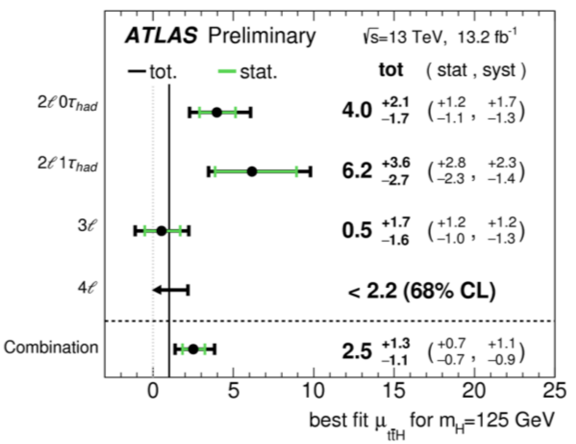
CMS HIG-17-003



CMS HIG-17-004



ATLAS-CONF-2016-058



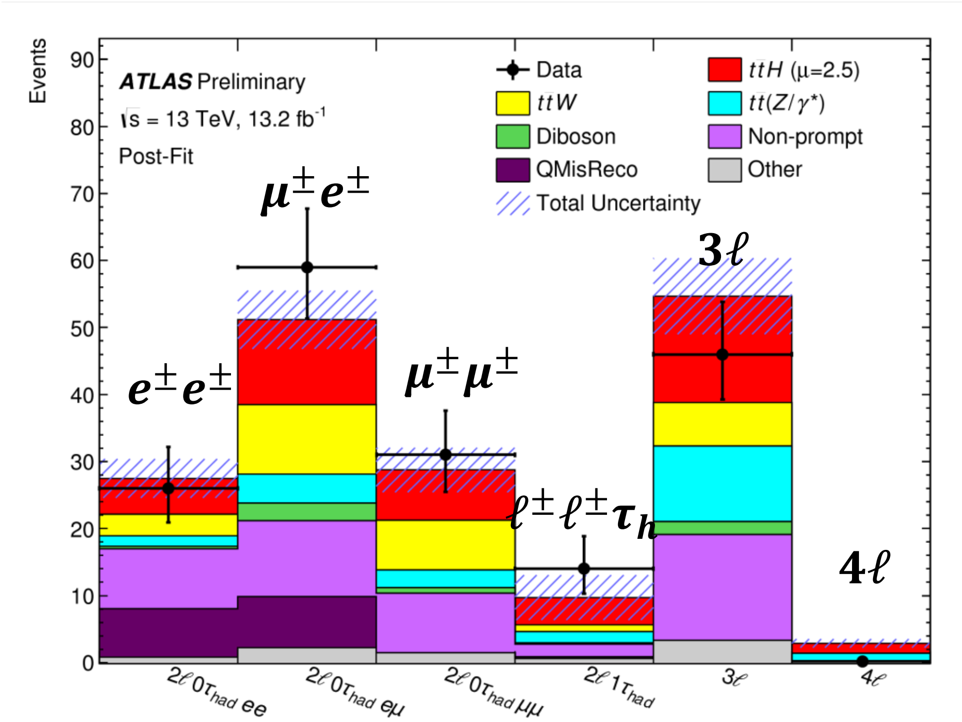
Mainly dominated by syst uncertainties

ATLAS-CONF-2016-058

ATLAS ICHEP Analysis

- ATLAS chose to start with a simple analysis for ICHEP 2016
  - Cut and count in 6 exclusive signal categories
  - No use of MVAs
- The main focus was on understanding backgrounds
- ATLAS is moving now towards more advanced techniques as already done by CMS
  - Stay tuned

Will focus on advanced techniques from CMS in what follows



CMS Categories

- Total of 19 categories
  - 15 categories for channels without τ_h
  - 4 categories for channels with τ_h
- MVAs are used in most categories
  - Not in 4l due to very low stat
- Exclude events compatible with ttH (H→4l) selection
  - Dedicated analysis

$\mu^\pm\mu^\pm$: b tight (+ −), b loose (+ −)

$\mu^\pm e^\pm$: b tight (+ −), b loose (+ −)

$e^\pm e^\pm$: (+ −)

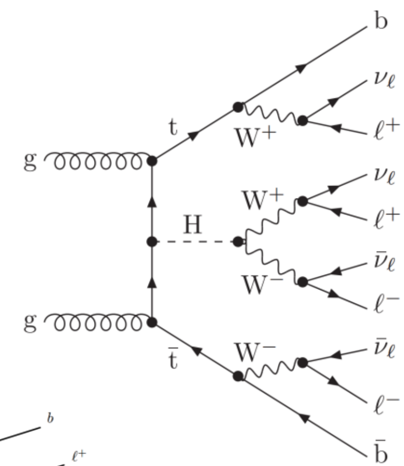
3ℓ: b tight (+ −), b loose (+ −)

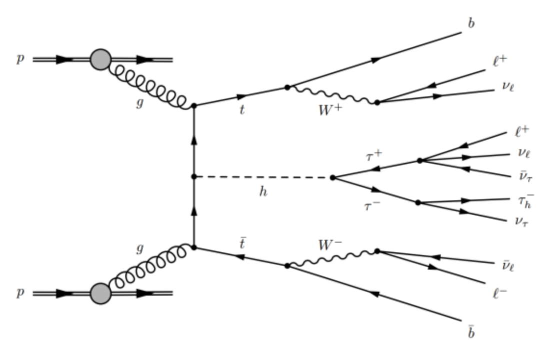
4ℓ

$1\ell 2\tau_h$

$\ell^\pm\ell^\pm\tau_h$: missing-jet, no-missing-jet

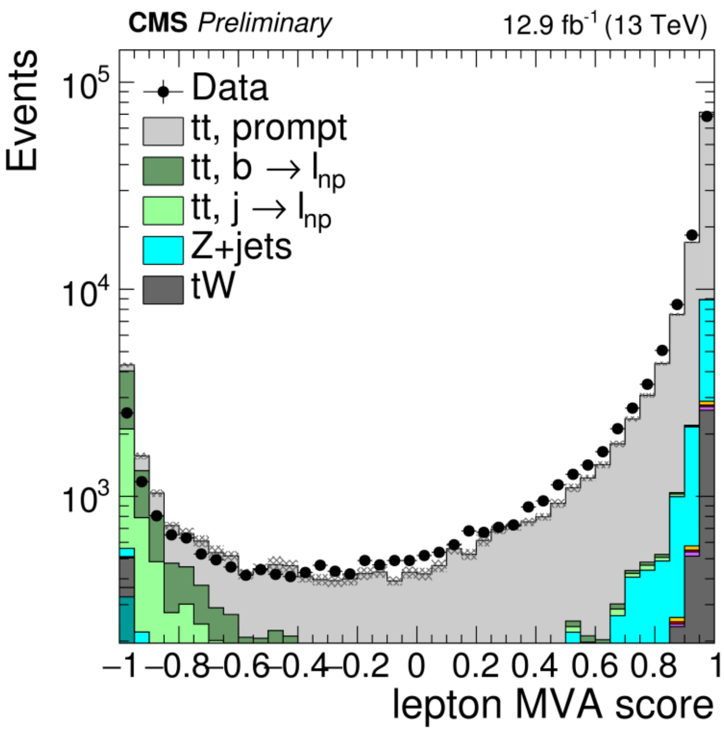
$3\ell 1\tau_h$





Lepton ID/ISOLATION

- Misidentification of leptons (electrons,muons) is one of the main problems in ttH multilepton analysis
- Background composition enriched with muons/electrons from semi-leptonic b/c hadron decays
- Dedicated MVA targeting non-prompt leptons in ttH final state

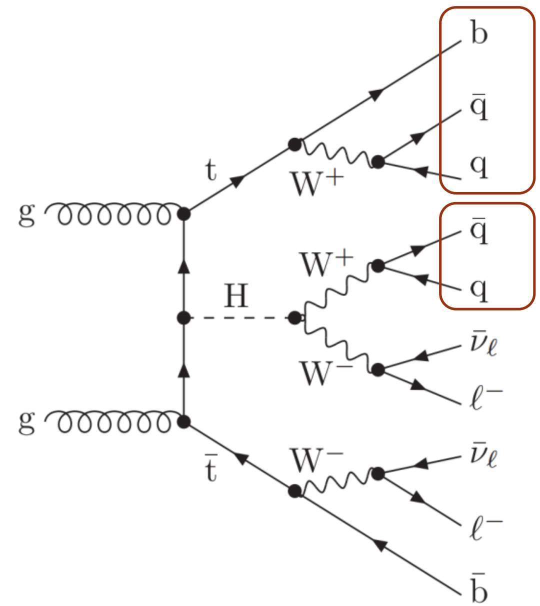


BDT variables:

- From lepton object itself
- Isolation information
- Overlapping jets and their probability to be b-jets

Partial Event Reconstruction

- Complex final state with many neutrinos
  - Very hard to reconstruct
- However partial reconstruction is possible
  - Identify jets from the Higgs or hadronic top
  - Use this information to separate signal and ttV/ttbar backgrounds
- Used for 2lss final state

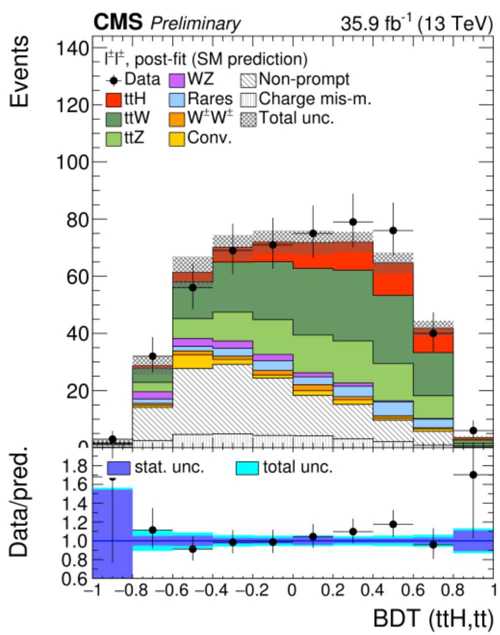


- Hadronic top decay tagger
  - Identify jets from hadronic top decay
  - BDT trained against incorrect permutation in ttH
- Tagging jets from Higgs decay
  - Exclude jets compatible with hadronic top decay
  - Identify the presence of jets from Higgs decay
  - Trained against ttV backgrounds

CMS HIG-17-004

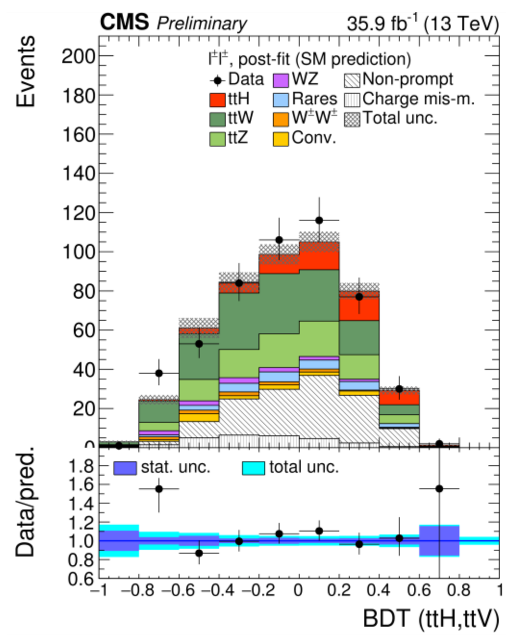
ttH 2lss BDT discriminants

- Train 2 kinematic BDTs, against ttbar and ttW/Z
- Map 2D into 1D (add bins with similar S/B)



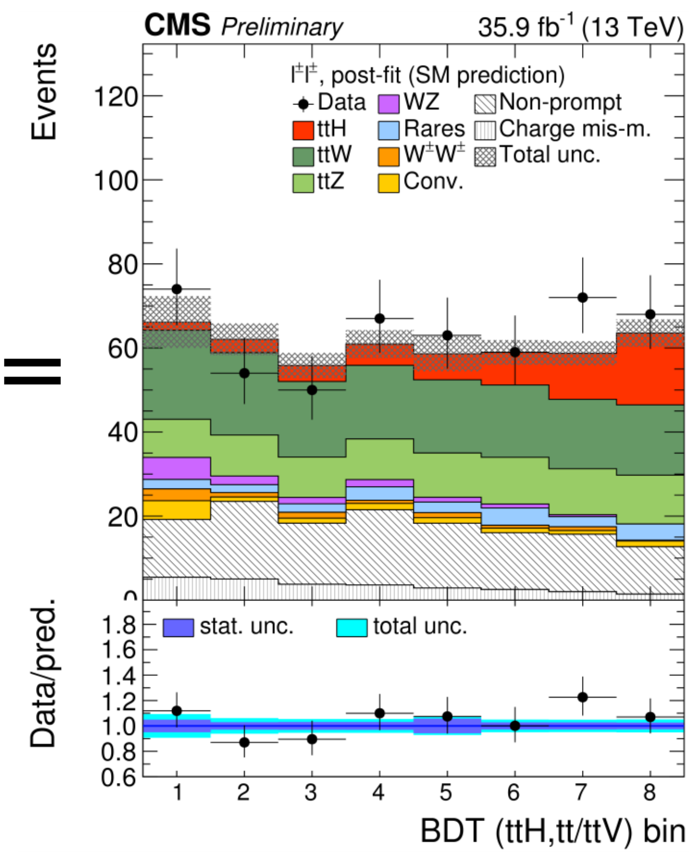
2ℓss, vs ttbar:
Includes **hadronic top tagger**

+

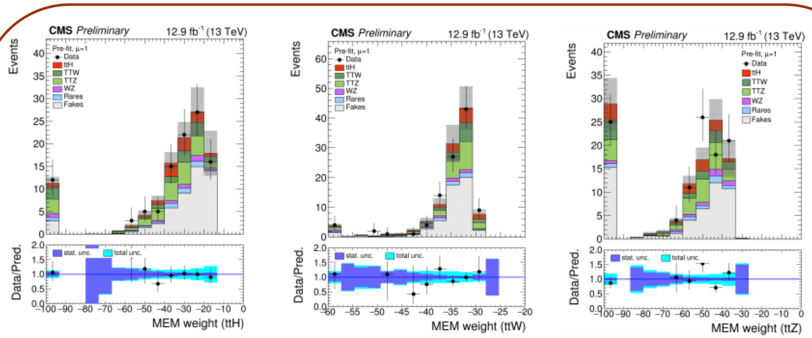


2ℓss, vs ttW/Z:
Include tagging of **jets from Higgs**

=

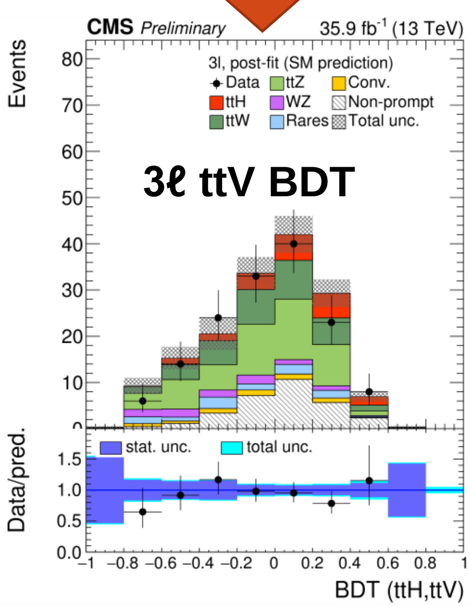


ttH 3l BDT discriminant

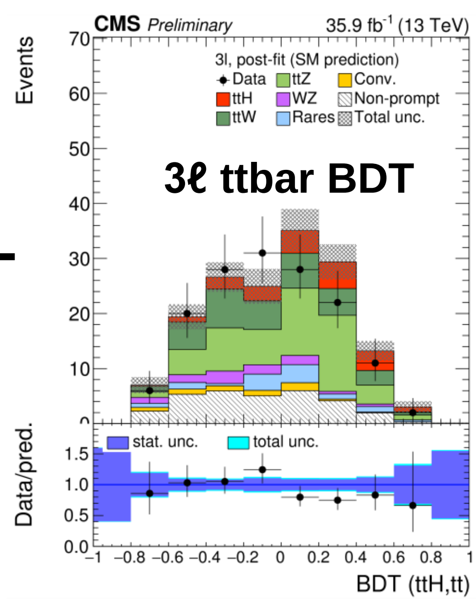


$$-log\left(\frac{\sigma_{t\bar{t}Z}w_{t\bar{t}Z} + k\cdot\sigma_{t\bar{t}W}w_{t\bar{t}W}}{\sigma_{t\bar{t}H}w_{t\bar{t}H} + \sigma_{t\bar{t}Z}w_{t\bar{t}Z} + k\cdot\sigma_{t\bar{t}W}w_{t\bar{t}W}}\right)$$

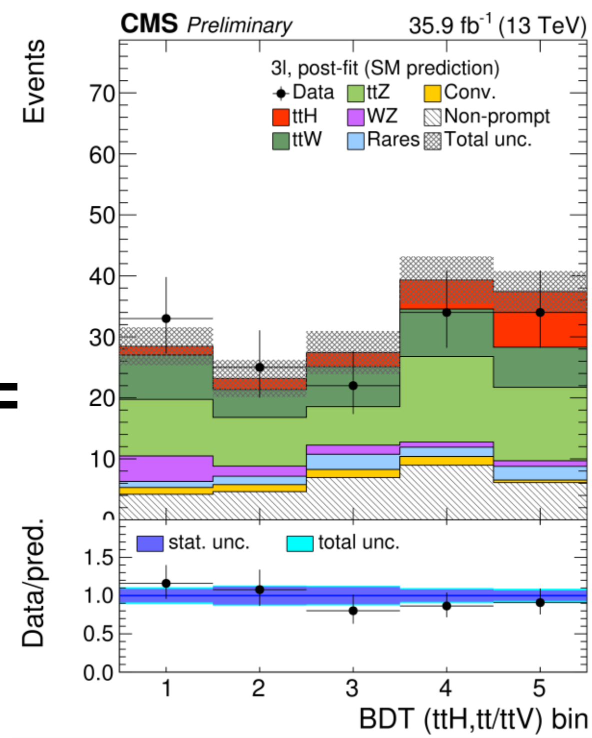
- Evaluate MEM weights under ttH, ttW, ttZ hypotheses
  - Build likelihood ratio of ttV vs ttH+ttV
- MEM weight included in ttH vs ttV BDT
- Another BDT to discriminate ttH and ttbar
- Mix both BDTs
  - Adding bins with similar S/B



+



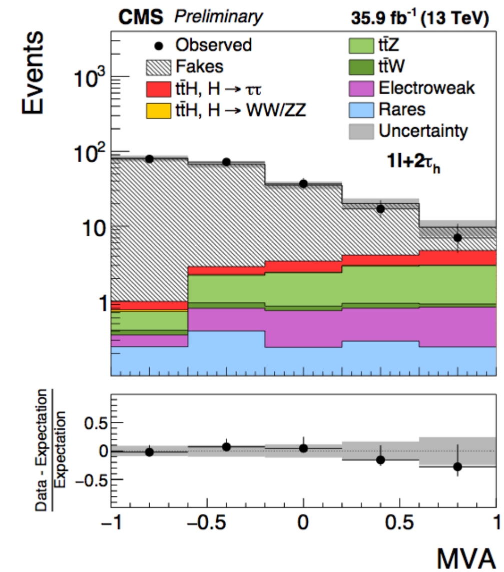
=

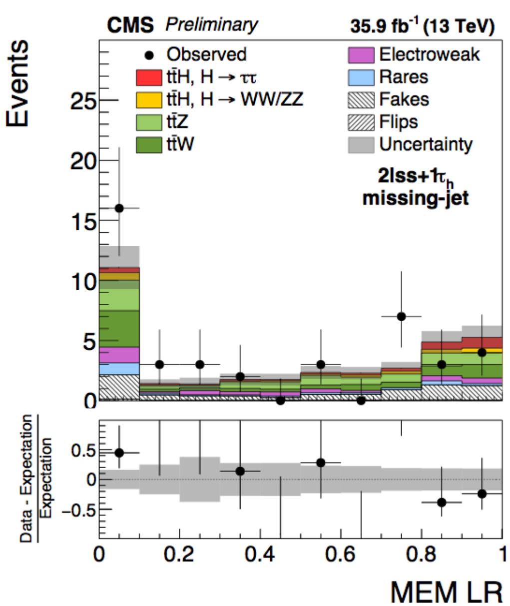


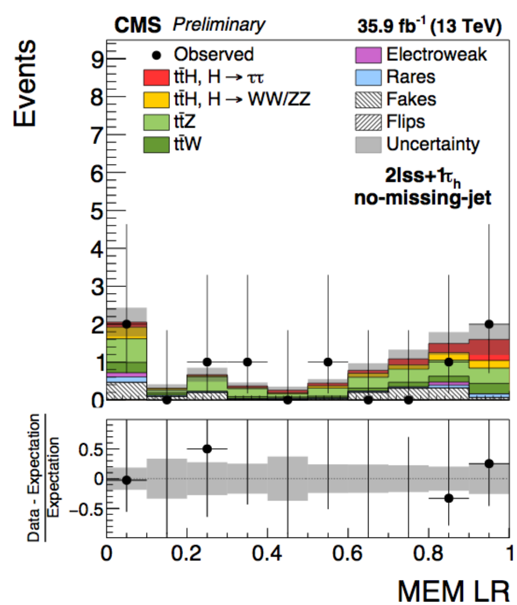
ttH, H→$\tau\tau$ discriminants

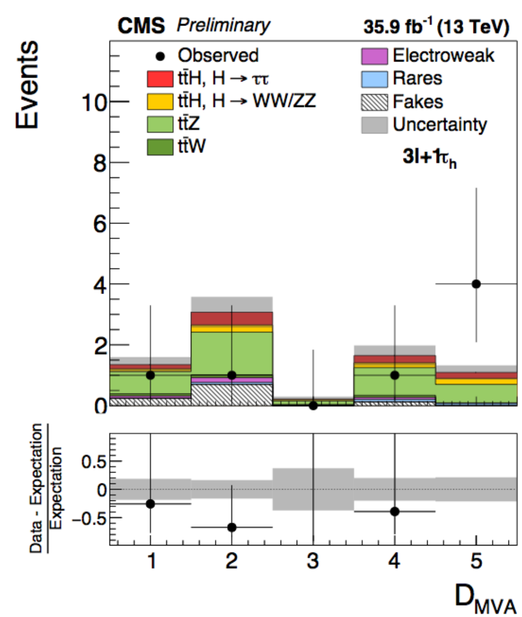
CMS HIG-17-003

- Split into 3 channels with leptons and τ_h in the final state
  - Following similar techniques as ttH with leptons and muons
- $2\ell ss+1\tau_h$
  - MEM likelihood ratio with ttH vs ttZ and ttbar hypotheses
  - Further split according to the presence of two jets compatible with a W boson decay
- $1\ell+2\tau_h$
  - BDT trained against ttbar
- $3\ell+1\tau_h$
  - 2 BDTs: against ttV and ttbar
  - 1D bin mapping according to S/B (D_{MVA})







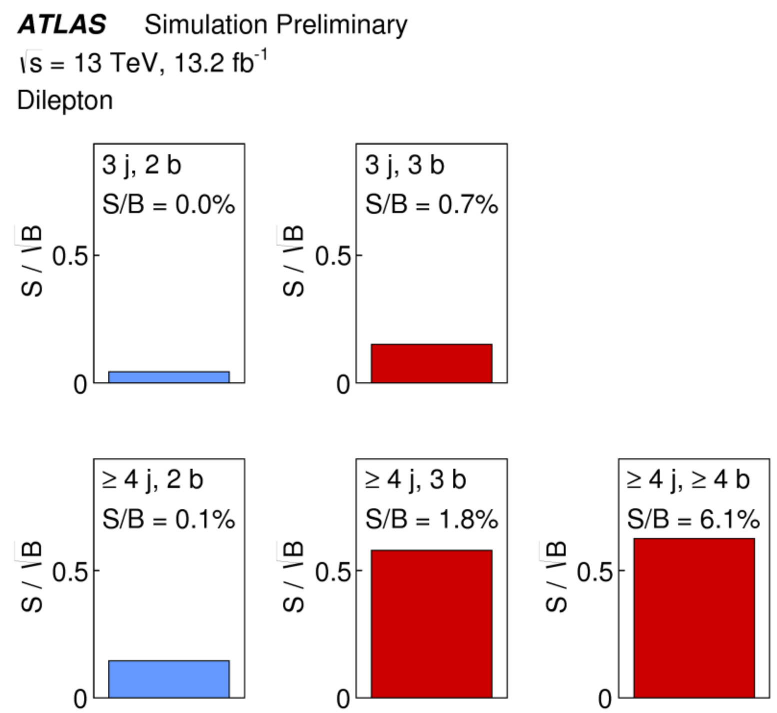


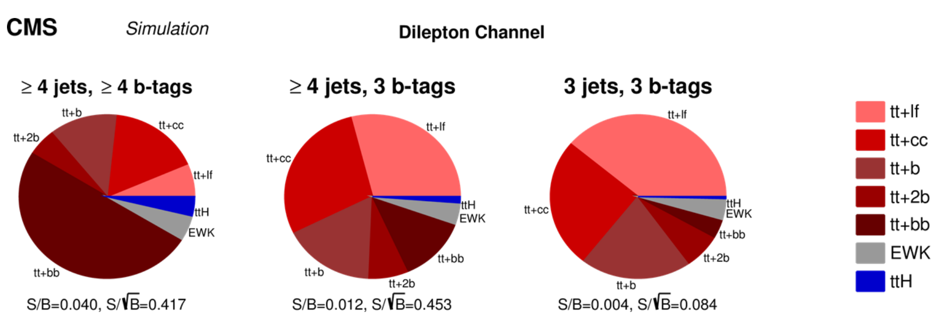
Conclusion

- ttH channel one of the hot topics of the LHC run 2 physics program
  - However observing the ttH with run 2 data will be challenging
  - Current results dominated by systematic uncertainties
- Very low stat channels ($H\rightarrow\gamma\gamma$ and $H\rightarrow 4l$) with large purity will become more important at the end of run II
  - But most probably not enough alone with run II expected luminosity
- Need to combine as many channels as possible
- Need to use advanced techniques to increase signal/bkg separation
  - Complex final state leaves more room for ideas
- Both ATLAS and CMS are constantly improving their analyses techniques
  - BDTs, MEM and reconstruction techniques are now widely used
    - More powerful techniques like DNN are being investigated
    - Usage of object level MVAs (b-tag, lepton iso, …)
  - Adding new methods and including new channels
  - More categories are included with the increasing collected luminosity
- The other important front is of course to reduce systematics
  - Especially related to background modeling

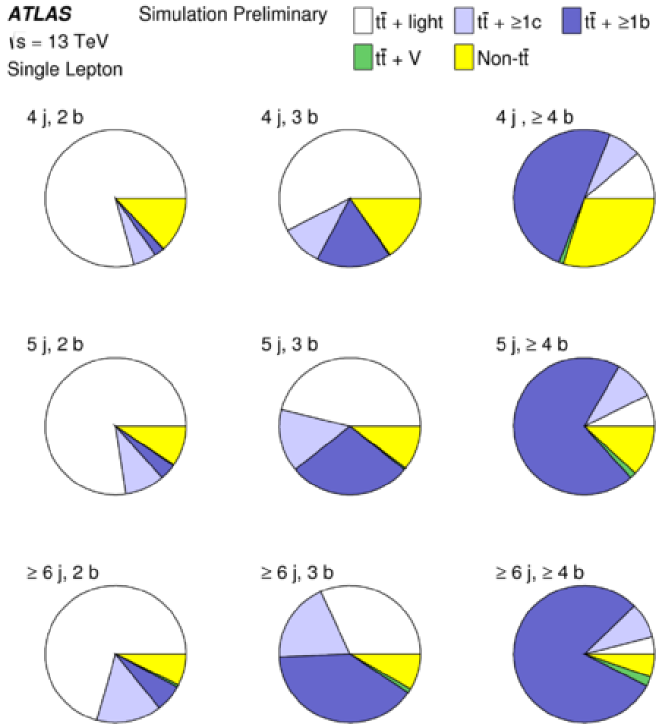
Backup

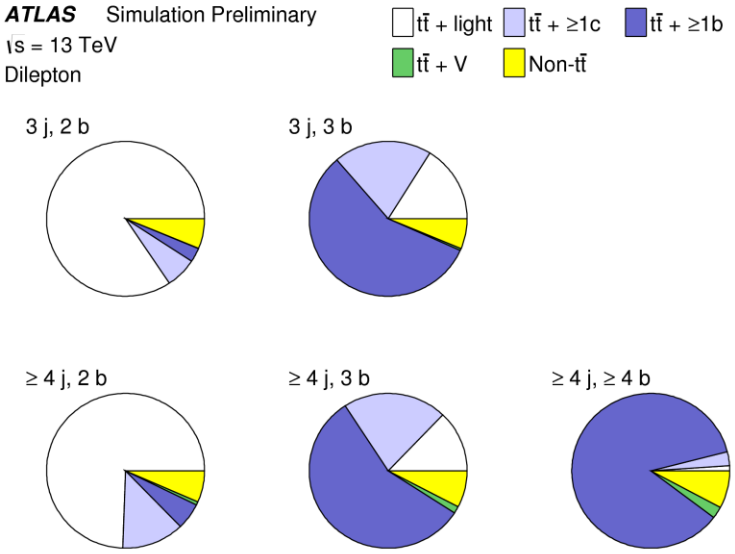
ttH(bb) dilep categories





ttH(bb) ATLAS BKG Composition





ttH(bb) All Hadronic Channel

JHEP 05 (2016) 160