ttH Analysis in Multilepton at ATLAS

Chao Wang

FCPPL Workshop

27-03-2017

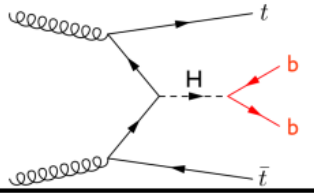
Overview of ttH Analysis at ATLAS

HTop Analysis
168 people
(authors of the current analysis)

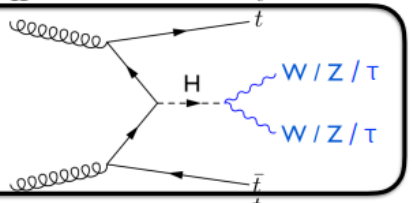
58 authors of the current analysis
18 institutions
1) 2015 Run1 results
2) 2016 ICHEP Conf. Note
3) 2017 EPS Paper

CPPM covers
ttH->bb,
ttH->2l, 3l, 4l analysis
SDU mainly in 3l

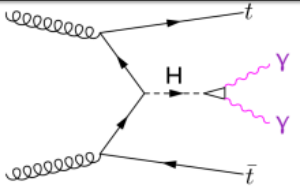
$t\bar{t}H$ ($H \to b\bar{b}$)
$t\bar{t}$→(lepton+jets, dilepton)



$t\bar{t}H$ ($H \to WW/ZZ/\tau\tau \to \geq 1\ell$)
$t\bar{t}$→(lepton+jets, dilepton)
--- "multilepton (ML)" ---



$t\bar{t}H$ ($H \to \gamma\gamma$)
$t\bar{t}$→(all-had, lepton+jets&dilepton)
--- covered in $H \to \gamma\gamma$ ---



Marseille, towards EPS paper Last Week

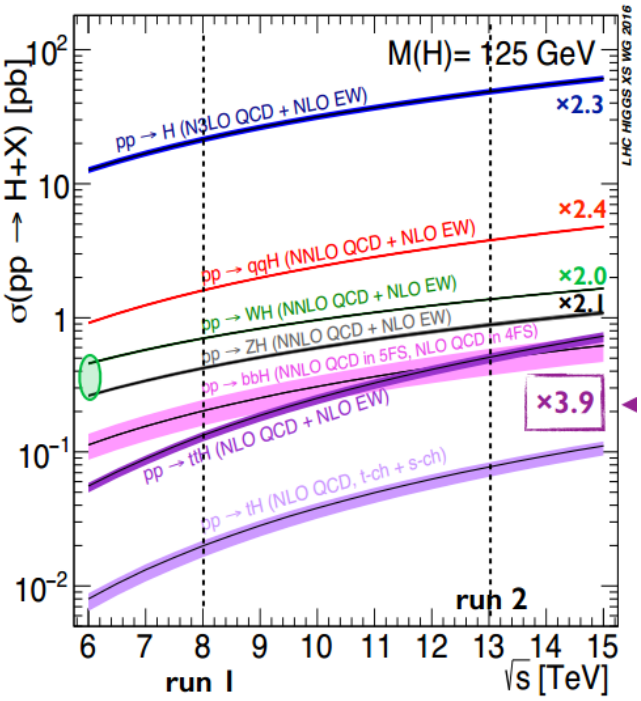
Outline

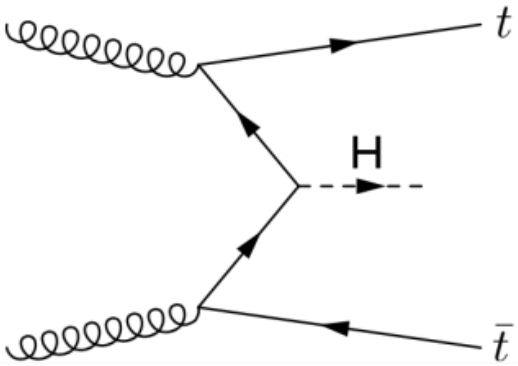
ICHEP Results Review

Status of Analysis in Multi-lepton Channel

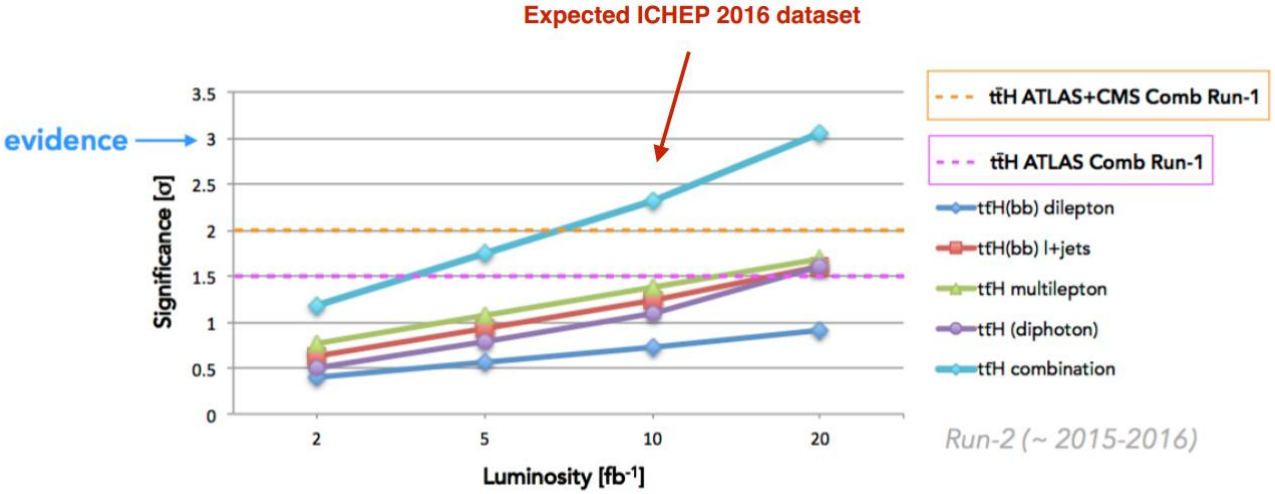
Summary and To-do

ICHEP Review





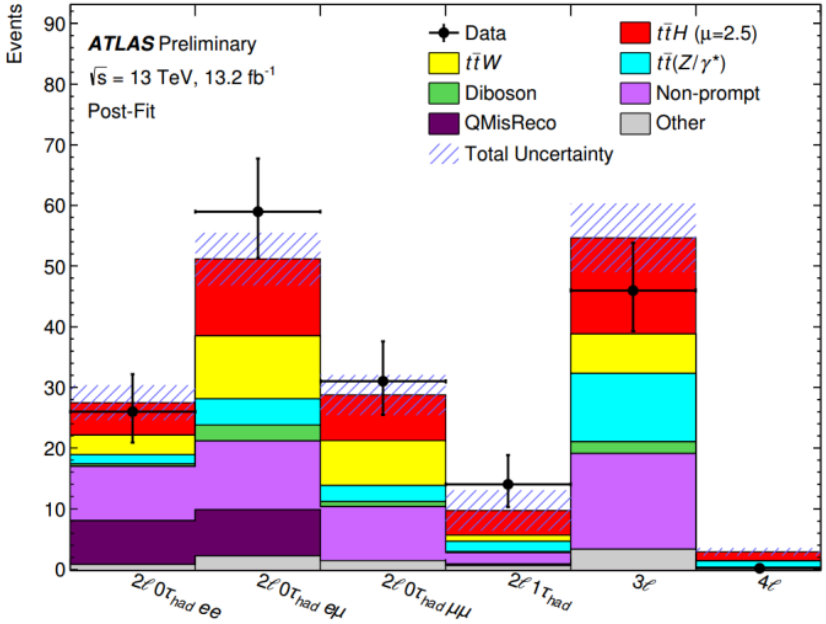
ttH Cross Section
8 TeV: 0.13
13 TeV: 0.5085



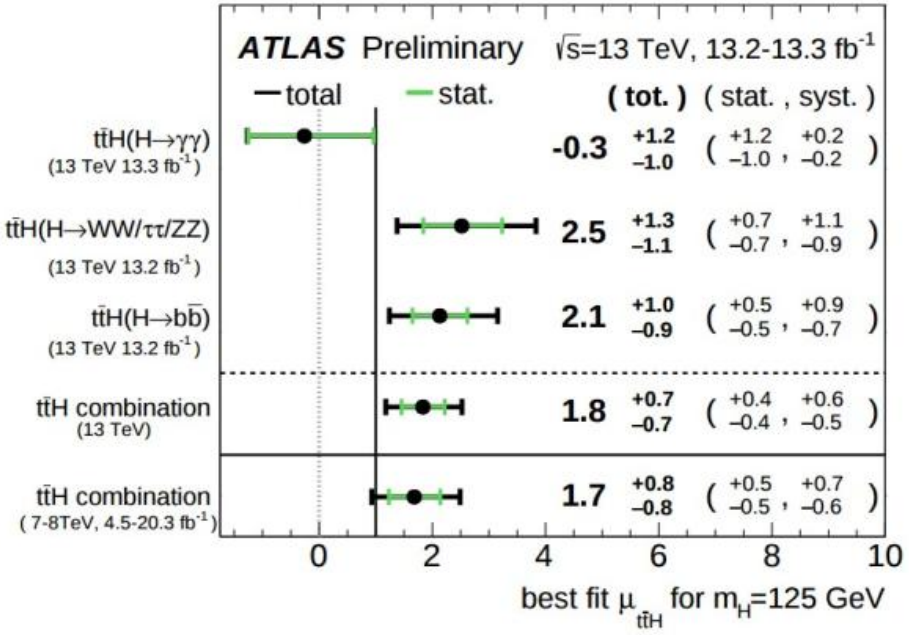
ICHEP Review

13.3 fb^{-1} of pp collision data at $\sqrt{s}$ = 13 TeV

Data vs Predicted postfit events in the different channels



Best fit values of the ttH signal strength



Channel	Significance	
	Observed [σ]	Expected [σ]
$t\bar{t}H, H \to \gamma\gamma$	−0.2	0.9
$t\bar{t}H, H \to (WW, \tau\tau, ZZ)$	2.2	1.0
$t\bar{t}H, H \to b\bar{b}$	2.4	1.2
$t\bar{t}H$ combination	2.8	1.8

The best fit value of the ttH signal strength is 1.8 ± 0.7

Observed significance: 2.8 sigma (1.8 expected from SM). 95% CL upper limit on ttH signal strength: 3.1 (1.4 expected from bkg. only)

ICHEP Review

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

	significance obs (exp) [σ]	μ
ATLAS run 1	2.33 (1.53)	1.7 ± 0.8
ATLAS run 2	2.8 (1.8)	1.8 ± 0.7
CMS run 1	3.4 (1.2)	$2.8^{+1.0}_{-0.9}$
ATLAS+CMS run 1	4.4 (2.0)	$2.3^{+0.7}_{-0.6}$

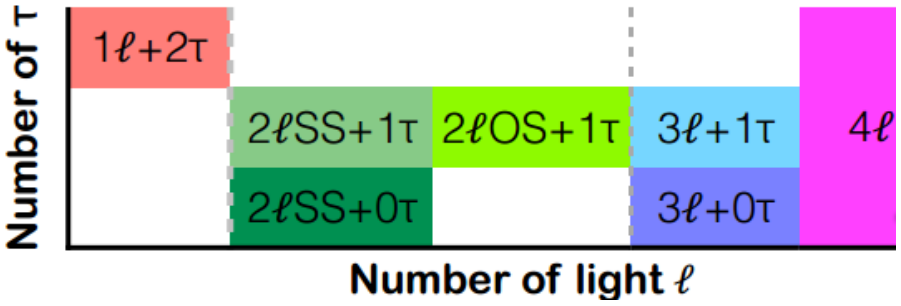
Comparison with Projection from Run1

Exp. Significance* (σ)	Proj (2015)	ICHEP (2016)
ttHbb dil	0.8	0.6 (-30%)
ttHbb l+j	1.4	1.1 (-30%)
ttHML	1.5	1.0 (-50%)
ttHyy	1.3	0.9 (-50%)
ttHComb	2.7	1.8 (-50%)

Evidence(3σ) with 2015+2016 data set will not be easy

Significant improvements in all analysis and all channels is necessary

Multilepton Overview

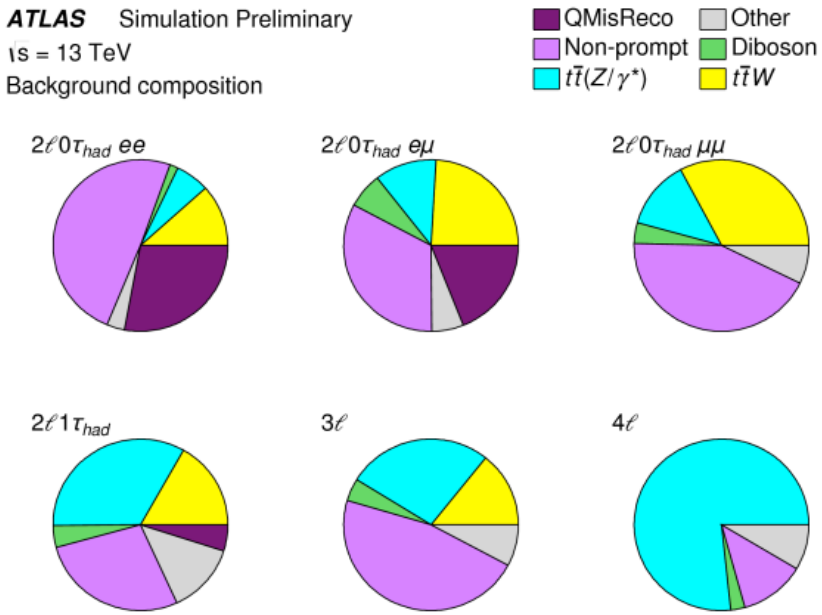


Category	Higgs boson decay mode WW^*	$\tau\tau$	ZZ^*	Other	$A \times \epsilon$ $(\times 10^{-4})$
$2\ell 0\tau_{\text{had}}$	77%	17%	3%	3%	14
$2\ell 1\tau_{\text{had}}$	46%	51%	2%	1%	2.2
3ℓ	74%	20%	4%	2%	9.2
4ℓ	72%	18%	9%	2%	0.88

Objections towards EPS

1. **New trigger strategy (single OR dilepton)**
2. **Overlap removal:**
   based on the Run-1 procedure
3. **Jets and bjets:**
   1. Pass jet clean criteria
   2. pT > 25 GeV ; |η| < 2.5
   3. remove jets with |JVT| < 0.59
   4. |η| < 2.4, and pT < 60 GeV
   5. MV2c10_70
4. **QMisID MVA is used to reduce the charge flip**
5. **New loose and tight lepton definition based on a MVA method**

Run-2 ICHEP16



New lepton ID

	Loose		Tight	
	e	μ	e	μ
Isolation	-	-	PromptLeptonIso < -0.5 AND IsoLoose (*)	PromptLeptonIso < -0.5 AND IsoLoose (*)
ID working point	Loose	Loose	Tight	Loose
Charge misID BDT	-	-	> 0.0670415	-
$\lvert d_0 \rvert / \sigma_{d_0}$	< 5	< 3	< 5	< 3
$\lvert \Delta z_0 \sin\theta_\ell \rvert$	< 0.5 mm	< 0.5 mm	< 0.5 mm	< 0.5 mm

Remove isolation requirement at loose level To increase the statistics for the Matrix method

Move to new LepID MVA cut from: ~~Electron: FixCutTightOnly~~ ~~Muon:FixCutTightTrackOnly~~

- **Improvement with new tight lepton for instance**

Event	Lepton selection	S	B	S/B	$S/\sqrt{B}$	$S/\sqrt{B+\sigma_B^2}$	$t\bar{t}$
$p_T^{1,2} > 20$ GeV	FixedCutTight(TrackOnly) isolation	19.8	80.4	0.25	2.21	1.84	14.8
	PromptLeptonIso < -0.50	16.9	62.9	0.27	2.13	2.06	5.2
$p_T^{1,2} > 15$ GeV	FixedCutTight(TrackOnly) isolation	23.1	104.0	0.22	2.26	1.46	30.0
	PromptLeptonIso < -0.50	19.9	74.7	0.27	2.30	2.11	9.4

New tight LepID(BDT with b-tagging info. from the track jet that contains the lepton track and isolation **) suppress the main ttbar background efficiently**

Strategy towards EPS

- Event MVA in: 2l, 3l, 4l
  - New channels like; 2lOS1tau, 1l2tau will also employ the MVA

- Fakes estimation:
  - 2l: Matrix Method, extended fake factor,
  - 3l: Matrix Method, extended fake factor, MC template
  - 4l: Fake factor

Matrix Method
Rely on regions to measure the real and fake rate

Fake factor
Rely on this region to measure the fake factor

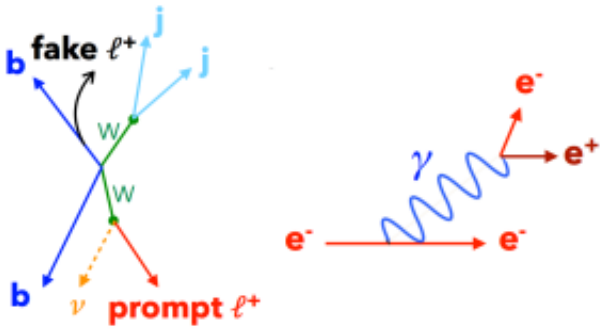
MC template fit
Rely on regions to extract norm. factor

Topological fit method
Weak correlation to this region since using MVA to select fake enriched region

- Fit shapes in relaxed SR categories in high stat. channels, add CRs to fit to have a handle on backgrounds

Fakes estimation

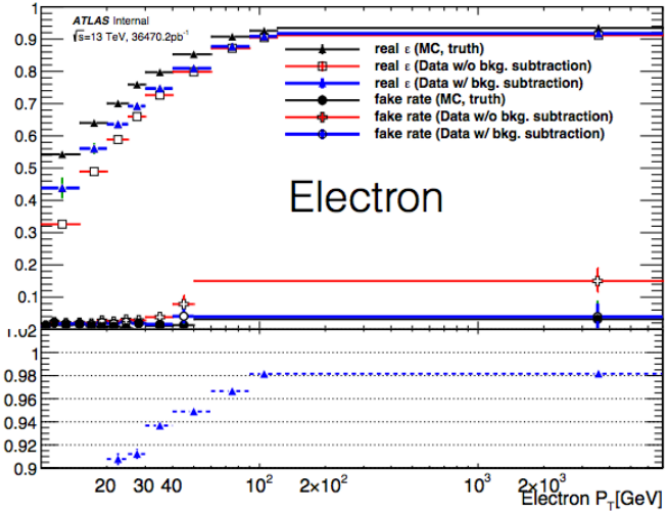
Fake Estimation in 2l



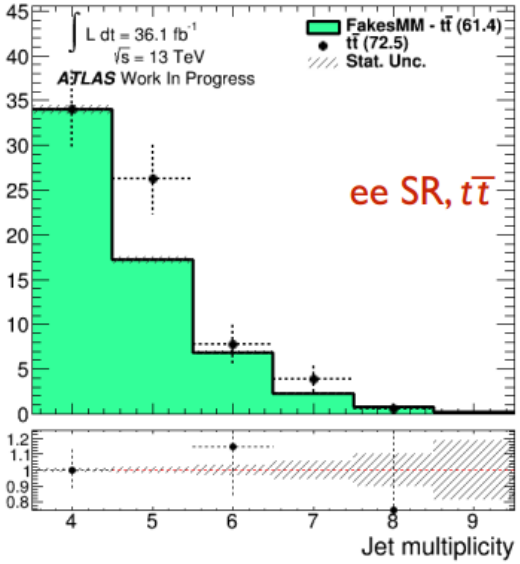
Major sources of "fake" lepton backgrounds in 2ℓSS and 3ℓ:

Matrix Method:

Measure the efficiency in data for real/fake leptons to pass "Tight" selection in dedicated CRs, and obtain the total number of fakes in SR from "Anti-Tight ($\bar{T}$)" sidebands via a matrix



Matrix Method Modelling

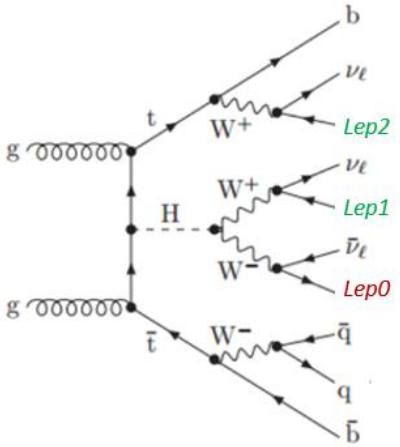


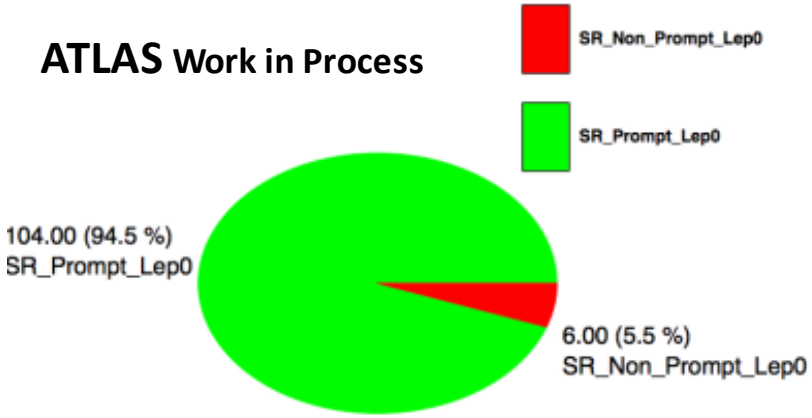
Fakes	elec.-elec.	muon-muon	elec.-muon
central value	76.2	185.8	247.3
stat. error	1.4	6.0	5.4
syst. error	14.7	-	78.7
ttbar-PP8 pred.	69.1 +/- 6.8 (76.36)	106.7 +/- 8.4 (106.96)	179.5+/-10.3 (200.57)
DD / MC ratio	1.10	1.74	1.38

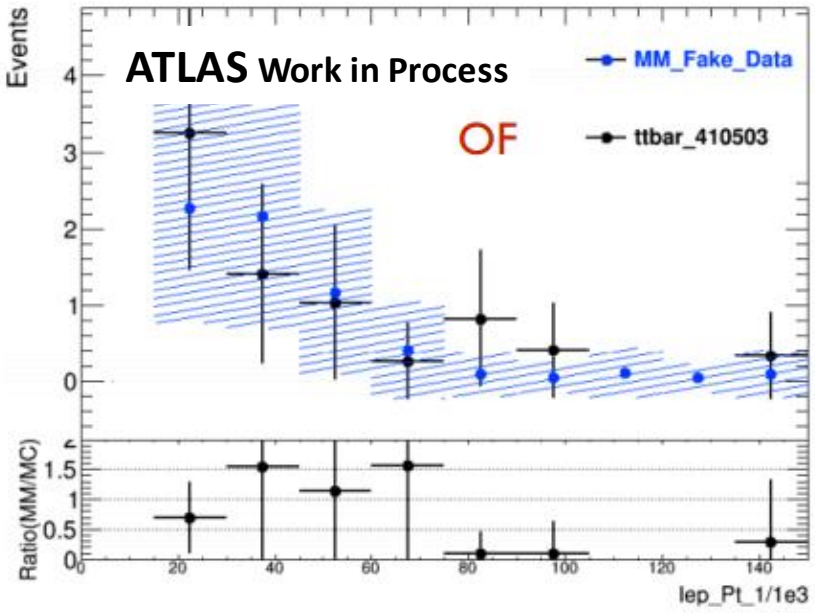
Fake estimates in SR-like region

Fake Estimation in 3l

Assumption on lep0 : with very low possibility to be the fake and checked by MC
If so, a simplified matrix method can be used in 3l as well







	ee	Mumu	elmu
MM(ttbar)	1.52+/- 0.09	1.7+/- 0.56	3.06 +/- 0.15
Expected(ttbar)	1.33+/- 0.40	1.62+/- 0.14	2.83+/- 0.65

Shape Modelling is good with MC

Full sets of systematics is ongoing

Fakes in 3l is also checked and predicted is fine

Multivariate Analysis

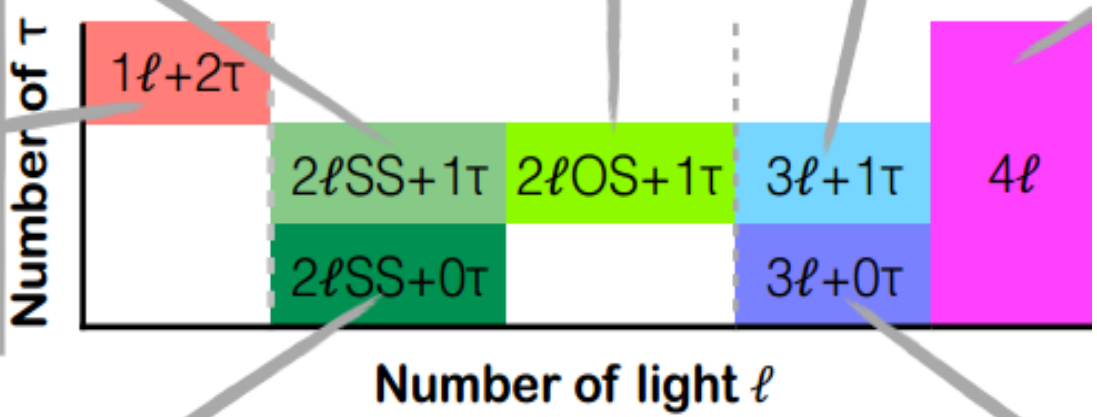
- Two proposals:
  - BDTG ($t\bar{t}H$-vs-$t\bar{t}$)
  - Categorisation based on event variables

- BDTG ($t\bar{t}H$-vs-$t\bar{t}$); $t\bar{t}$ contributing with 1 fake tau
- Cut&count based on BDTG cut?

- Two proposals:
  - BDTG ($t\bar{t}H$-vs-$t\bar{t}Z$, $t\bar{t}H$-vs-VV)
  - Categorisation based on event variables

- Cut&count

- BDTG ($t\bar{t}H$-vs-$t\bar{t}$); $t\bar{t}$ contributing with 2 fake taus
- Cut&count based on BDTG cut?



- Two main proposals using event variables:
  - BDTG ($t\bar{t}H$-vs-$t\bar{t}V$, $t\bar{t}H$-vs-$t\bar{t}$)
  - Multinominal classification
- Additionally, event reconstruction discriminants ($t\bar{t}H$-vs-$t\bar{t}$) can enter as input to the above methods:

- Several proposals which can be summarised as:
  - BDTG ($t\bar{t}H$-vs-$t\bar{t}V$, $t\bar{t}H$-vs-$t\bar{t}$)
  - Multinominal classification
- Additionally, event reconstruction discriminants ($t\bar{t}H$-vs-$t\bar{t}Z$/ttX) can enter as input to the above methods:

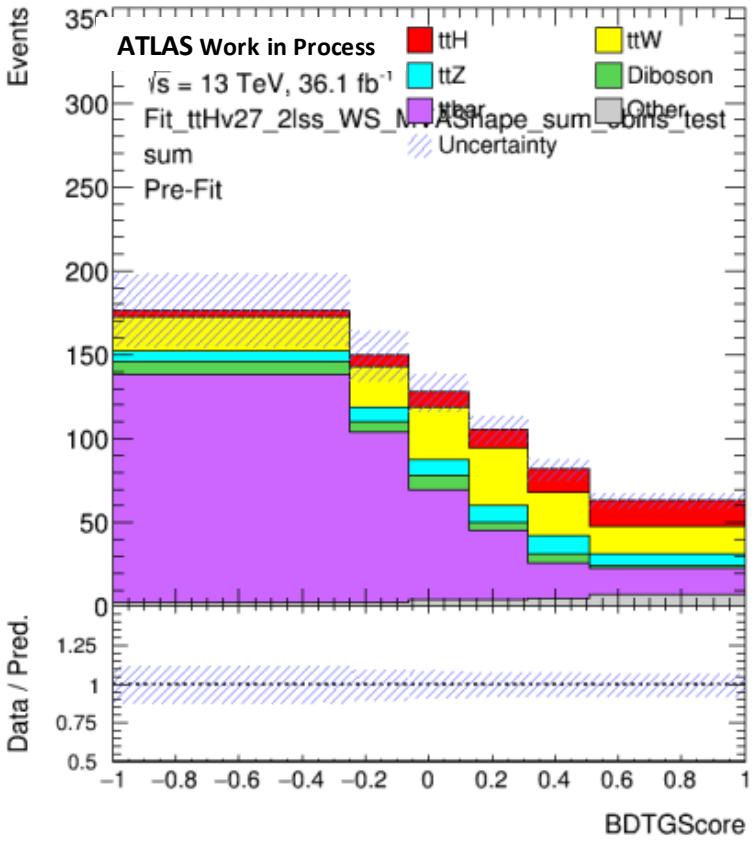
MVA Study in 2lSS

MVA trained versus two main backgrounds:
- ttH vs ttV = ttW + ttZ
- ttH vs ttbar (MC only)
- neglecting all other smaller backgrounds such as dibosons

Finally fit the BDT shape

Selection: Relaxed SR : SLT||DLT, pT(lep) as low as the trigger threshold, 3 < nJets < 8, nBJets >=1, new lepton tight definition

Variables : 9 kinematic variables rather good modeling seen for all variables, also with data driven fakes



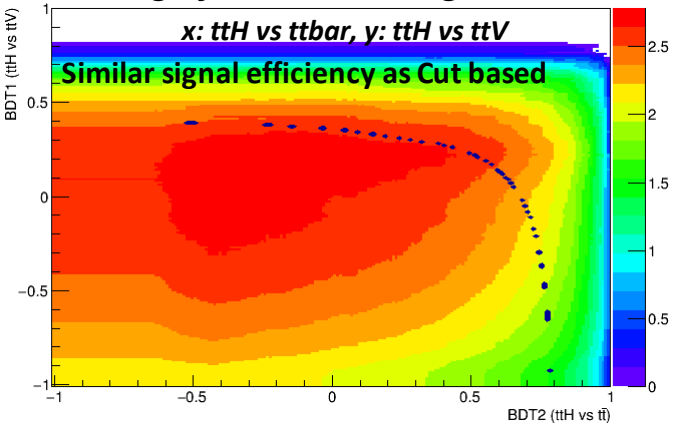
2 given BDTs (ttH-ttbar and ttH-ttV), different configurations were tested
All fit the shape of a discriminant using 6 bins with auto-binning (also 4 and 8 were tested)

Multivariate Analysis with BDT in 3l

Aiming at separating the ttV and ttbar from the ttH signal

- ✓ BDTG is trained for ttV and ttbar with sets of kinematic variables under a looser ICHEP SR region (looser jets requirement, w/o z veto, lower pt(15GeV for SS))
- ✓ Scan BDT score in 2D for all backgrounds to get the best sensitivity and do the shape fitting in future

Significance Scanning in 2D



Expected results with 36.5 fb^{-1}

Yields	ttH	ttbar	ttV	Total	S(Cowan)
Cut Based	16.3±0.3	12.26±1.4	46.3±0.6	88.44±4	1.85±0.51
Best with BDT Cut*	23.3±5.4	13.1±3.62	50.3±7.3	104±24.2	2.50±0.61

***BDTG_Best: the best significance with BDT with new tight lepton**

- A better performance with BDT compared to cut based
- Re-optimize input variables and consider fakes with data driven method
- BDT shape fit with sets of systematics error

Summary and To-dos

Generally, new changes after ICHEP bring the improvements to all channels, but are still under testing

- New tight lepton gives the good ttbar suppression
  - Expected to have better performance in MVA or MM(ongoing)
  - lepton MVA calibration (ongoing)

- Data-vs-MC comparisons in various bkg. control regions

- Matrix method
  - Check fake composition on 2l and 3l channels (extra systematics for 3l)

- MVA is used in almost all channels
  - Show the promising results
  - Need more testing on the method, input variables
  - Full sets of systematics errors

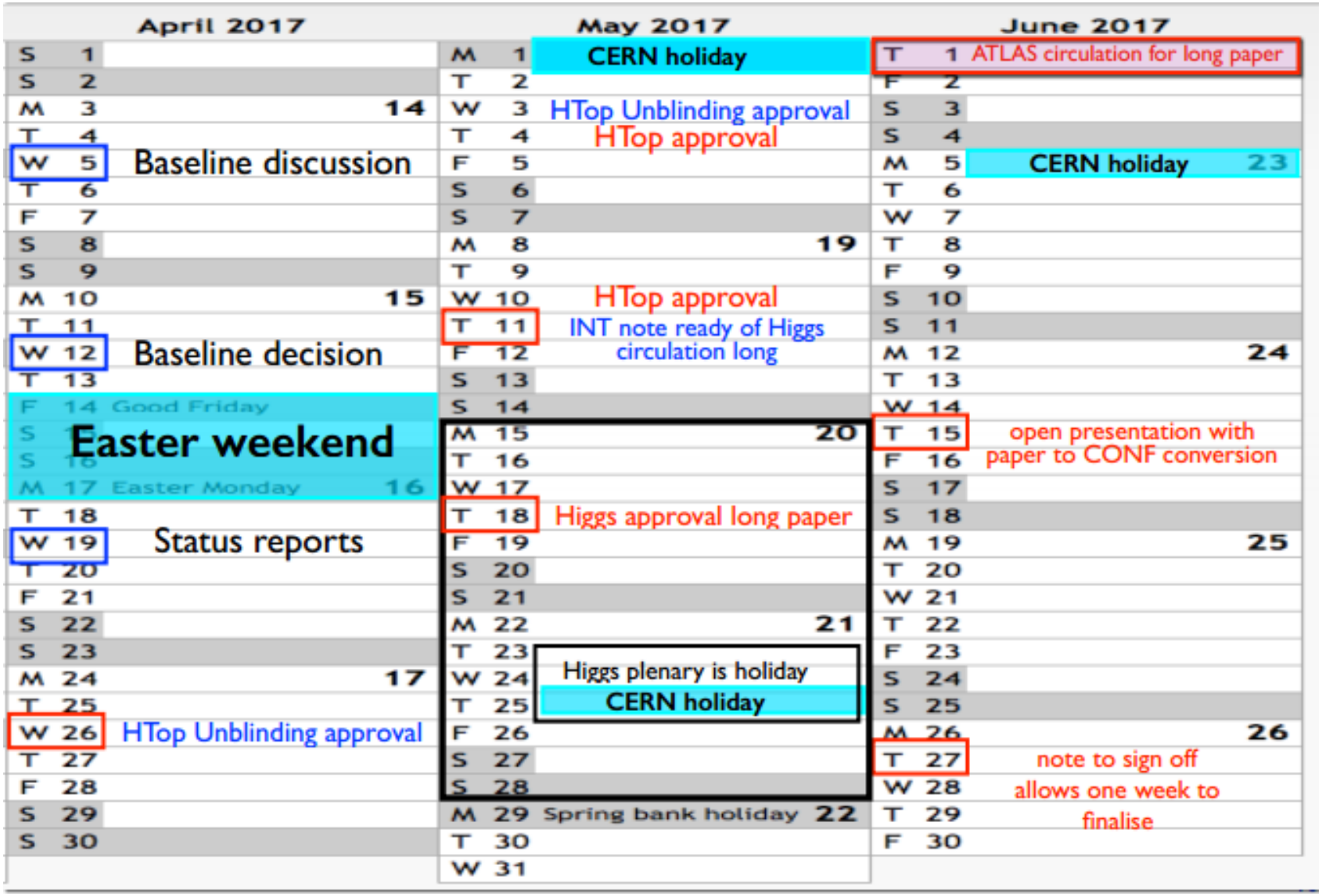
Conclusions

- ✓ **New changes bring the improvements to 3l**
- ✓ **Multivariate analysis is employed after ICHEP**
- ✓ **Preliminary studies show promising improvement**
- ✓ **A lot of things targeting at the significance improvements are going to be tested and will be implemented soon**


Thank you for your attention!

Backup

EPS Schedule



Overlap Removal

ICHEP-2016 overlap removal

Keep	Remove	Cone size (Δ R) or track
electron	tau	0.2
muon	tau	0.2
electron	CaloTagged muon	shared track
muon	electon	shared track
electron	jet	0.2
jet	electron	0.4
muon	jet	(0.2 or ghost-matched to muon) and (numJetTrk $\leq$2)
jet	muon	0.4
tau	jet	0.2

Run 1 overlap removal

Keep	Remove	Cone size (Δ R)
electron	electron (low p_{T})	0.1
muon	electron	0.1
electron	jet	0.3
jet	muon	**Run-1:** $0.04 + 10$**[GeV]**$/p_{\mathrm{T}}$ **(muon)**
		modified Run-1: min(0.4, $0.04 + 10$**[GeV]**$/p_{\mathrm{T}}$ **(muon))**
electron	tau	0.2
muon	tau	0.2
tau	jet	0.3

CPPM- BDT Sample and Selection

- **Signal** : ttH(Pythia8)
- **Background** :
  - 1) ttV (ttW, ttZ)
  - 2) ttbar(410009-dilepton ttbar, pythia6) and 410503(pythia8) is also used for testing

Selection(ICHEP basic but with some changes):

Loosing some selections based on the cut based 3l SR to increase the statistics

No changes:

1) Event cleaning
2) Charge, number of lepton
3) Tight ID on Lep1 and Lep2

Changes:

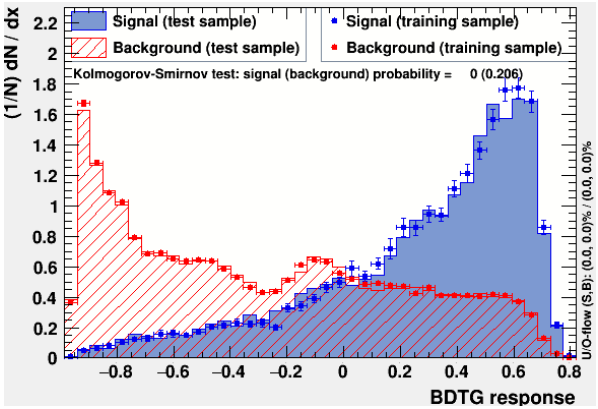
1) Pt: 10GeV, 15GeV, 15GeV
2) Trigger: single-lepton || di-lepton
3) jets>1 and bjets >0 to let more statistics in
4) Loose requirement on impact parameter
5) No Z veto on invariant mass of mll
6) Tight isolation or (Ele and muon: PromptLeptonIso_TagWeight < -0.5) on lep1 and lep2

CPPM- BDT

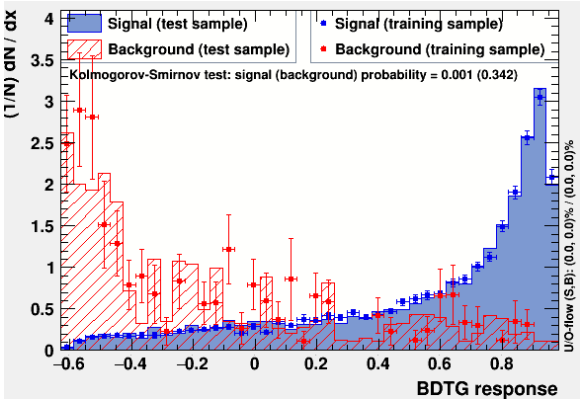
BDTG is trained:
ttH vs **ttbar**
ttH vs **ttV**

Selection(ICHEP basic but with some changes)
Loosing some selections based on the cut based 3l SR to increase the statistics:
Lower pt, S||D trigger, looser jets(2-1), loose impact parameter, no Z veto on invariant mass of mll, new tight isolation on lepton is tested.

Samples are split into odd and even two parts during the training and testing stage



ttH vs ttV



ttH vs ttbar