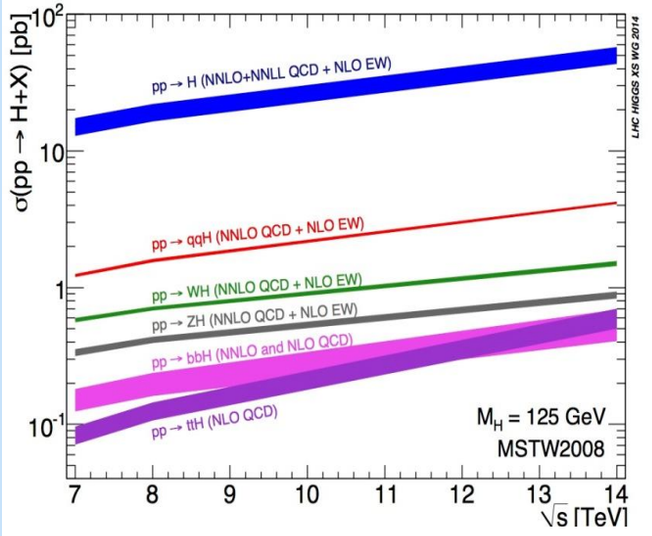
Study of the Higgs boson production with a single top quark in ATLAS experiment

IGOR BOYKO, NAZIM HUSEYNOV, OKSANA KOVAL, ANASTASIA TROPINA, IVAN YELETSKIKH

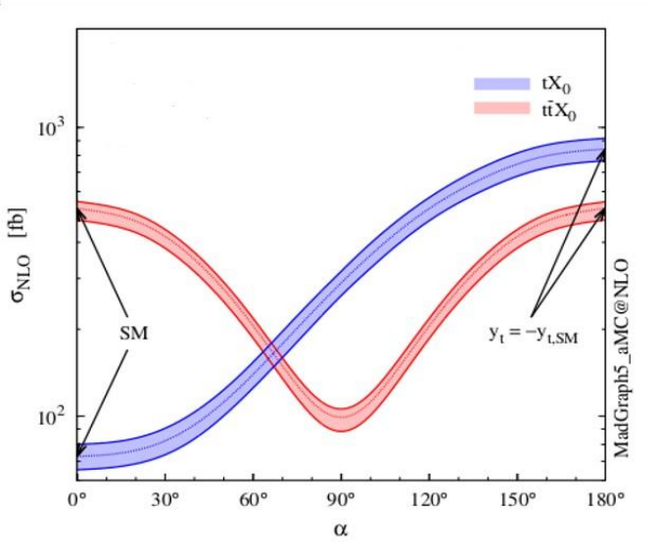
The XXVI International Scientific Conference of Young Scientists and Specialists (AYSS-2022)
25th of October 2022

$pp \rightarrow tH$ theory

- Top Yukawa coupling is directly measured in **ttH** events
- However, **ttH** is only sensitive to square (i.e. absolute value) of y_t
- **tH** sensitive to magnitude and sign of Higgs-top-Yukawa coupling, y_t (or, more generally the phase between ttH & WWH)
- Gives us a chance to direct probe of the sign of y_t



[1] Higgs production in association with a single top quark at the LHC [arXiv:1504.00611]



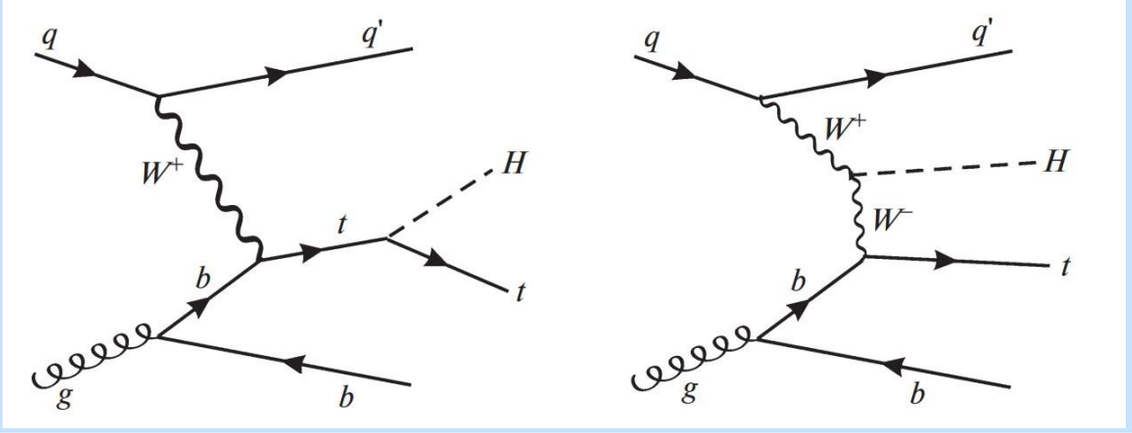
[2] Higgs Physics, C. Grojean [arXiv:1708.00794]

In Standart Model y_t =1 and there is destructive interference between **ttH** & **WWH**

=> σ(tHq)SM~74fb on 13TeV

If y_t =-1 (Inverted Top Coupling) then interference is constructive

=> σ(tHq)BSM~13×σ(tHq)SM



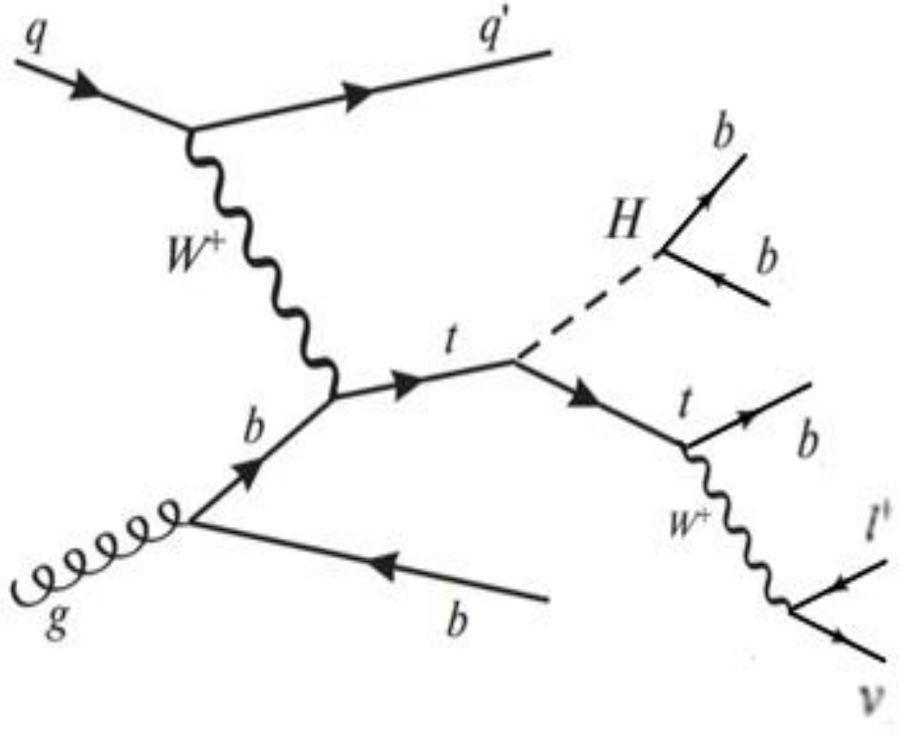
Preselection region

Topology-motivated requirements:

- Ntight lep = 1 (PLIV-Tight isolation)
- Nbjets ≥ 3
- ET miss > 25 GeV

Other requirements:

- Nτ = 0 — orthogonality with tau channels
- (Njets ≥ 5 && Nbjets ≥ 4) – orthogonality condition with ttH(bb)



$pp \rightarrow tH\mathrm{bb}$

Events Yields

The tH(bb) deals with the range of backgrounds, the most important are:

- $t\bar{t} + jets$
- *SingleTop*
- *non-prompt*
- $W + jets$

In total on preselection level:

- *Signal events – 72.6*
- *Total background events - 227650*

Stat-only uncertainty

	PR
tH	72.6 ± 1.3
tWH	35.7 ± 0.5
$t\bar{t}$ + ≥1b	61 610 ± 90
$t\bar{t}$ + ≥1c	34 240 ± 70
$t\bar{t}$ + light	108 920 ± 120
ttH	1733.6 ± 1.5
$t\bar{t}$ + Z	923 ± 6
$t\bar{t}$ + W	276.6 ± 1.5
tZq	169.8 ± 1.3
tWZ	2.18 ± 0.22
Wt channel	6017 ± 29
t channel	3177 ± 14
s channel	257.1 ± 2.8
W + jets	4023 ± 34
Z + jets	674 ± 13
VV	289 ± 6
other Higgs	26 ± 4
Rare top	100.43 ± 0.29
Non-prompt	5090 ± 70
Total	227 650 ± 190
Data	244 167

Will define clearly in a few slides!

	SR
tH	49 ± 7
tWH	8.6 ± 0.6
$t\bar{t}$ + ≥1b	11 500 ± 2400
$t\bar{t}$ + ≥1c	6100 ± 3500
$t\bar{t}$ + light	23 000 ± 4000
$t\bar{t}H$	280 ± 40
$t\bar{t}$ + Z	100 ± 40
$t\bar{t}$ + W	27 ± 8
tZq	93 ± 9
tWZ	0.22 ± 0.16
Wt channel	1000 ± 400
t channel	1160 ± 310
s channel	60 ± 40
W + jets	790 ± 330
Z + jets	160 ± 60
VV	60 ± 31
other Higgs	5 ± 6
Rare top	2.71 ± 0.04
Non-prompt	1300 ± 600
Total	46 000 ± 6000
Data	49 033

89%

4.5%

3%

The Multiclass BDT

5 hypotheses:

- Signal (1L tH(bb) only)
- $t\bar{t}$ +≥ 1b, $t\bar{t}$ +≥ 1c, $t\bar{t}$ +≥ 0 *lights*
- *Others (all remaining backgrounds)*

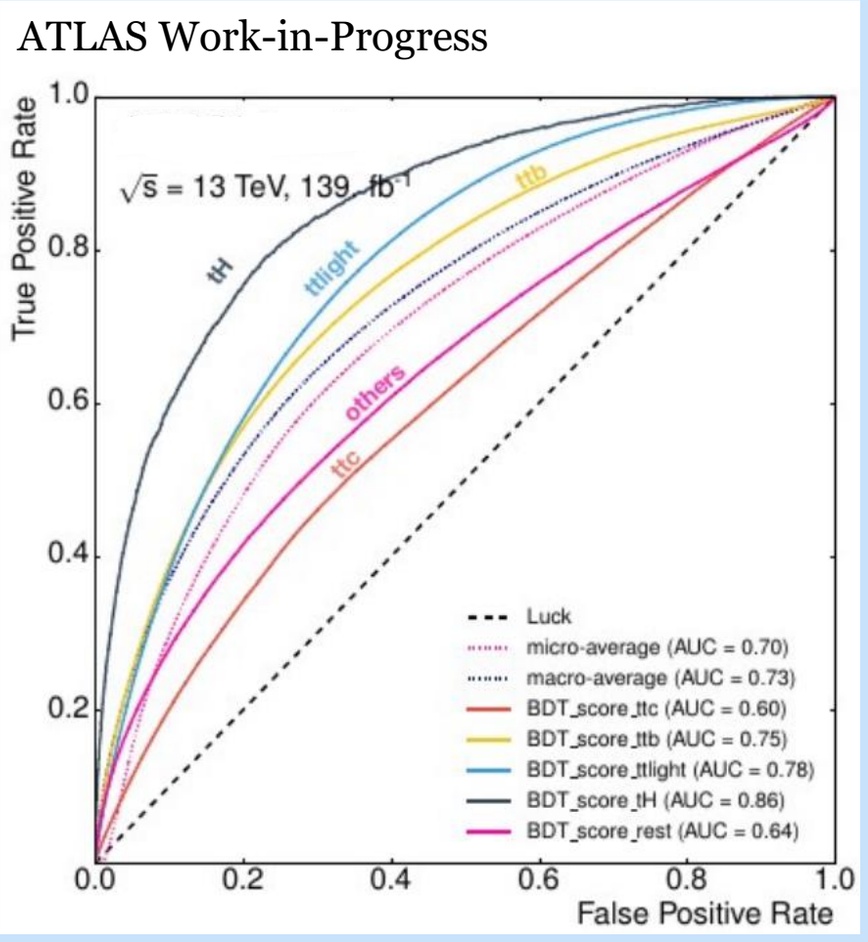
ROC curves are derived:

- *Best performing tH*
- *Least performing* $t\bar{t}$ +≥ 1c
  - *Expected topological similarity to dominant* $t\bar{t}$ +≥ 0 *lights*

BDT trained with 26 variables with early stopping criteria (up to 4k epochs)

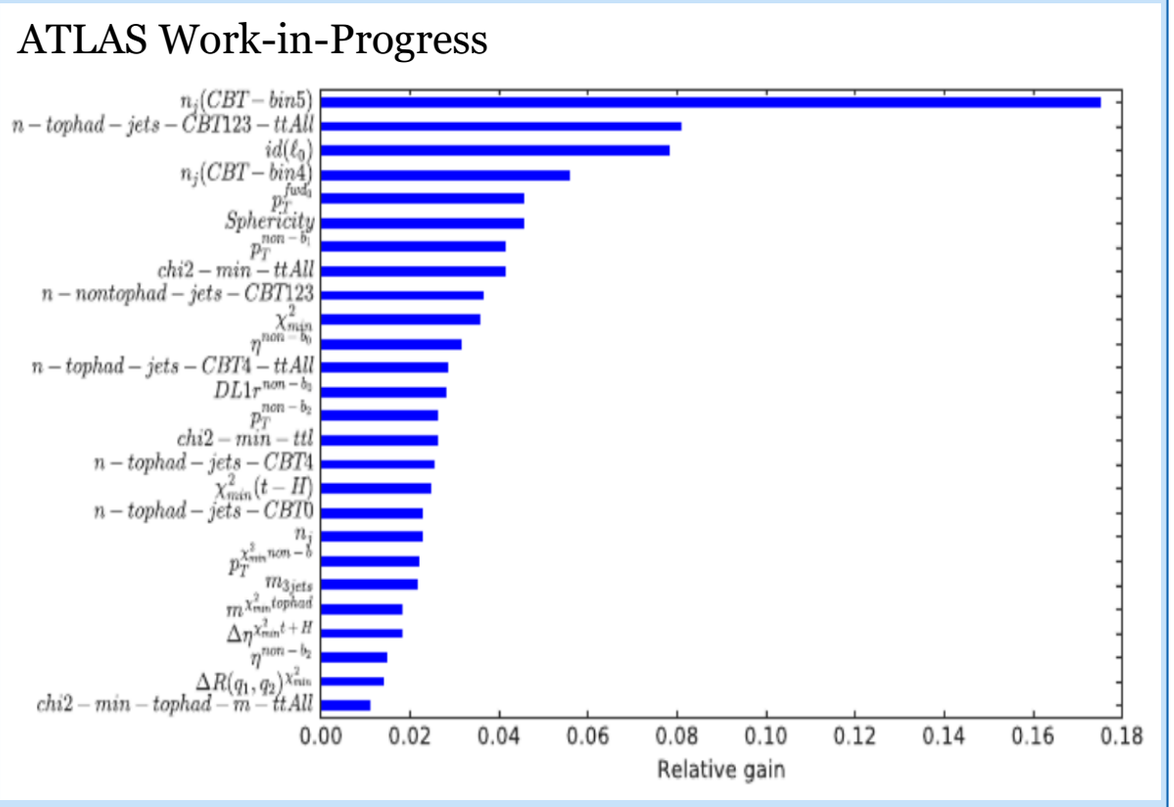
Five BDT scores are returned for each event (1 per hypothesis)

Neural-network is used to cross-check BDT performance



BDT variables

- n_j (CBT – binX): number of jets that fall into pseudo-continuous b-tagging (PCBT) bin number X
- "chi2-min" variables: from reconstruction $t\bar{t} + jets$ events using χ^2-minimization
- n-tophad-jets-CBTX: jets from hadronic top-quark decay which fall in PCBT bin X
- n-tophad-jets-CBTX: jets NOT from hadronic top-quark decay which fall in PCBT bin X, NLO + leptonic top-quark decay jets
- Sphericity: a measure of summed p_t^2 of all jet energy clusters with resprct to jet axis; back-to-back sub-jets: S=0, isotopic sub-jets: S=1



Implementation of new sensitive variables

In tHbq signal we expect a very forward "tagging jet" from "spectator quark":

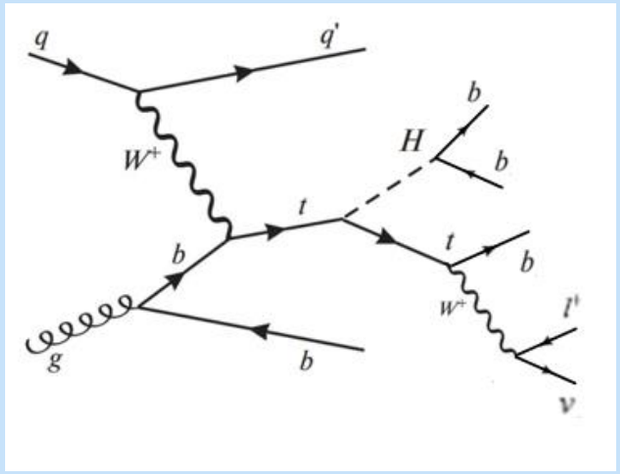
- It should have large rapidity gap with the rest of the event
- For ttbar background there is no reason to have this forward jet
- Until now, the "tagging jet" was the jet with

$|Eta|>2$

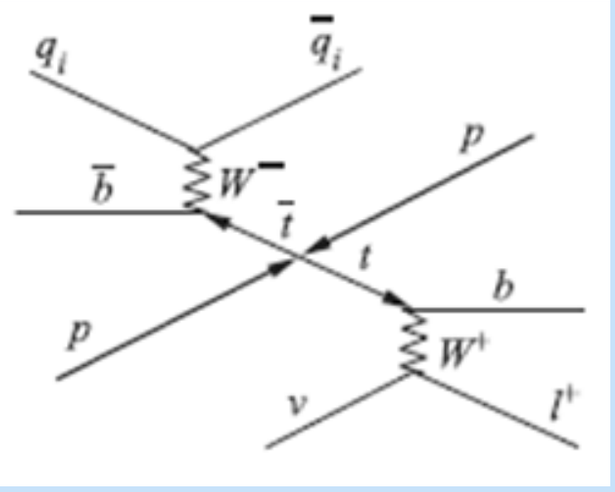
(if several, highest P_t was taken)

However, ttbar events have many additional non-b jet

=> There is high probability that one of them will be in the forward with |Eta|>2



$pp \rightarrow tH$



$pp \rightarrow tt$

Implementation of new sensitive variables

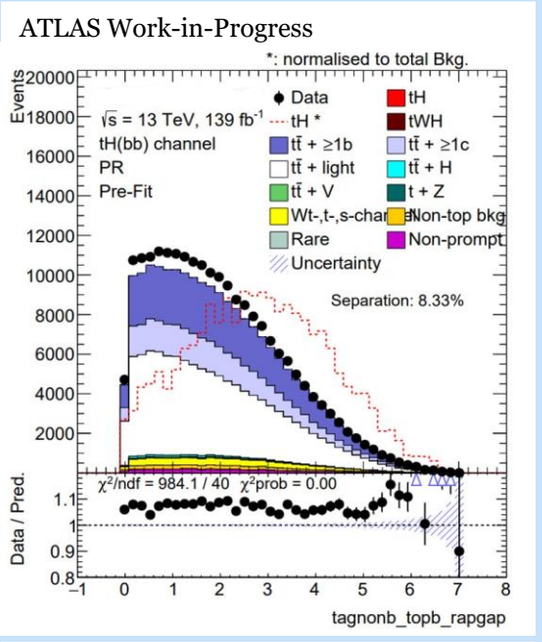
New algorithm for finding "tagging jet" was suggested:

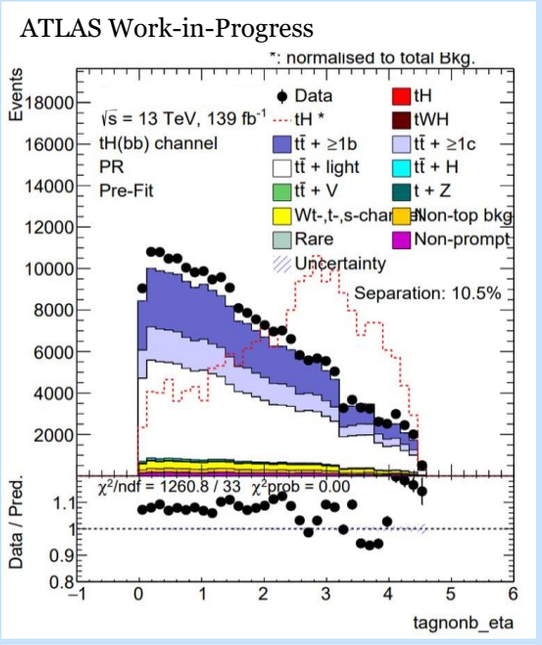
- Was found b-jet in the event related to top quark decay
- For all non-b jets in the event was found invariant mass with b-jet from top quark
- Non b-jet with maximum value of invariant mass with b-jet from top quark, was considered to be *the* "tagging *jet*"

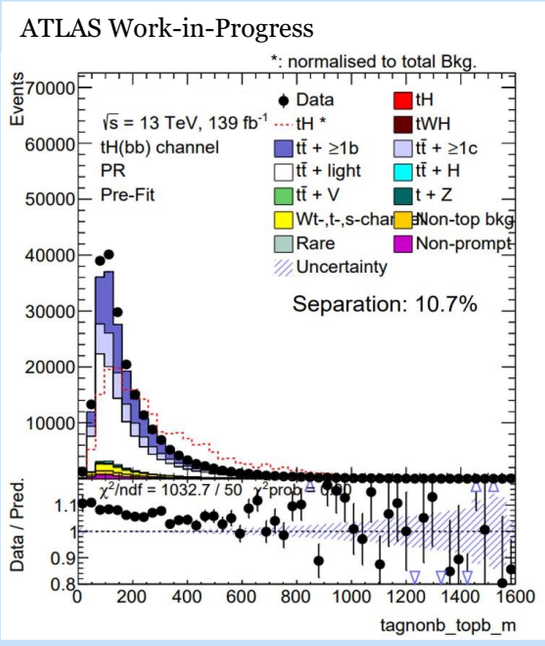
Variables constructed:

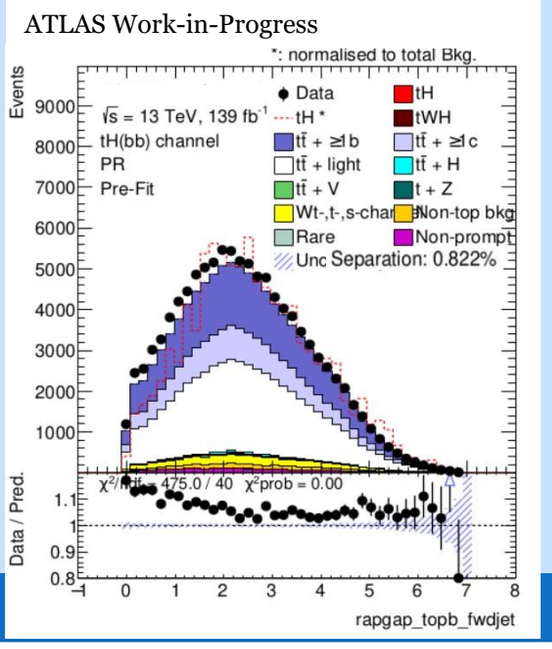
- *Invariant mass of* tagging *jet and b from top decay*
- Tagging *jet's eta*
- *Rapidity gap between* tagging *jet and b from top decay*

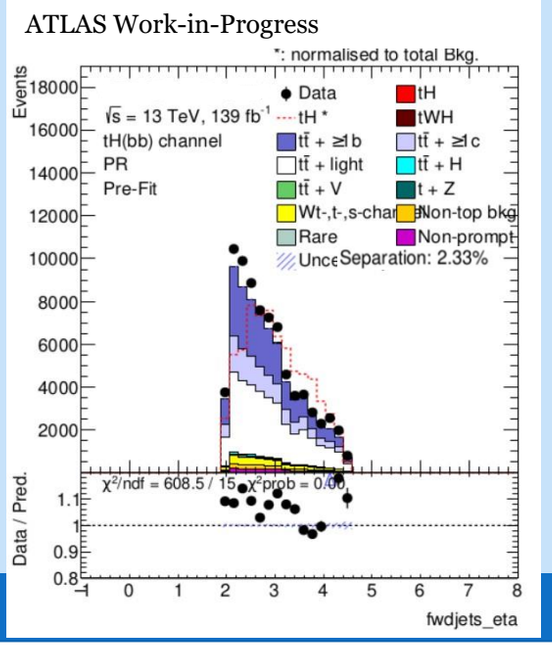
Distributions of implemented variables (SM)











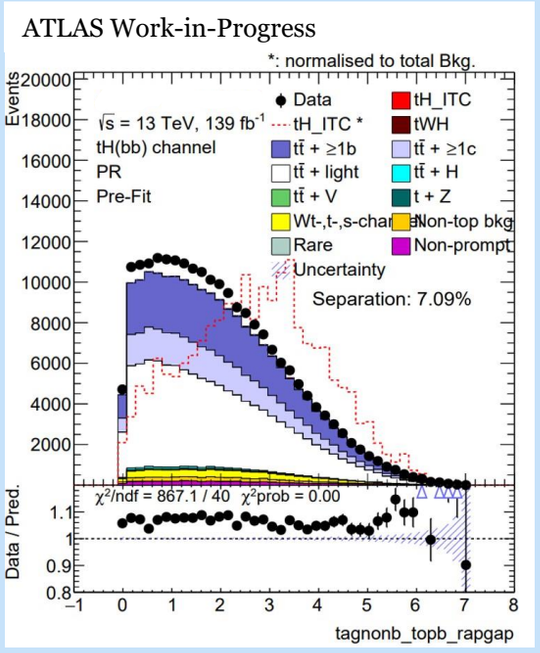
Top row– distributions of new variables ;

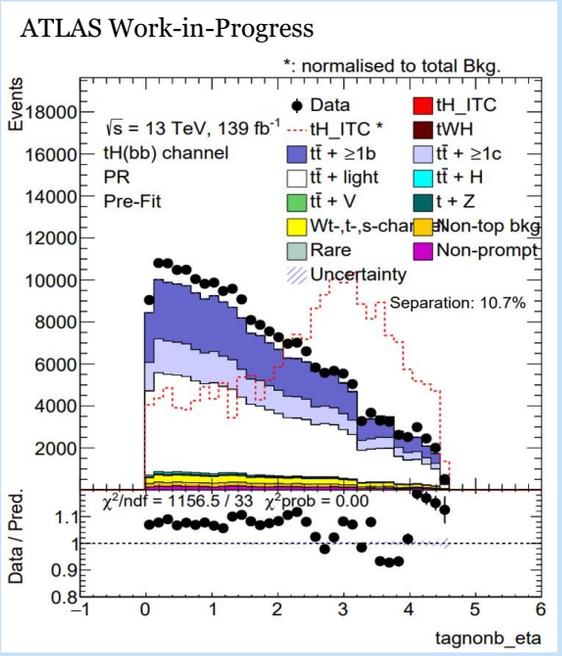
Bottom row – distributions of old variables;

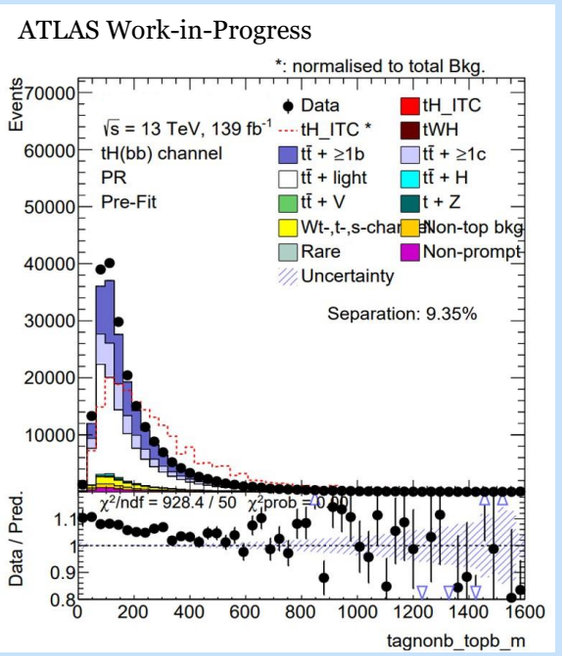
Separation power is defined as:

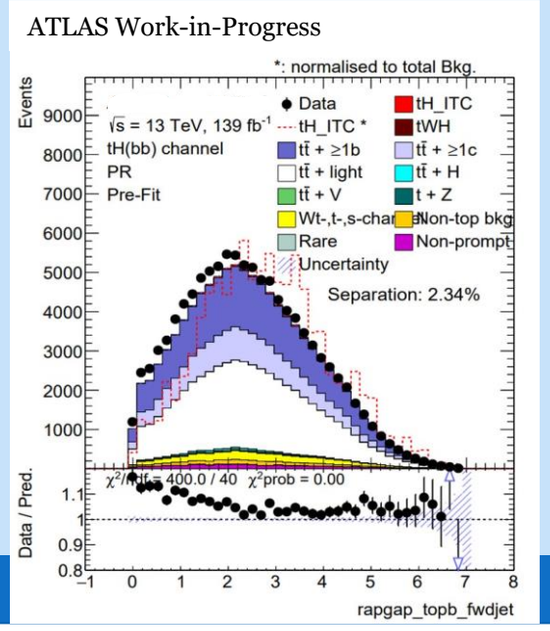
$$Sep = \frac{1}{2}\left(\sum_{i=1}^{nbins} \frac{(s_i - b_i)^2}{s_i + b_i}\right) \times 100$$

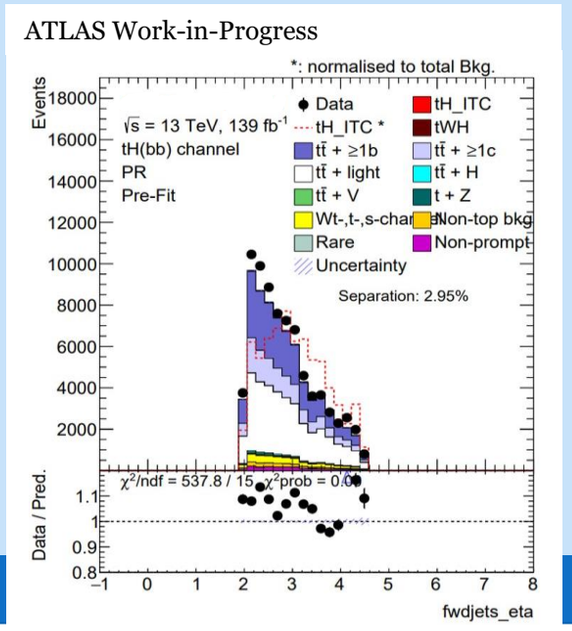
Distributions of implemented variables (ITC, y_t =-1)











Top row– distributions of new variables ;

Bottom row – distributions of old variables;

Separation power is defined as:

$$Sep = \frac{1}{2}\left(\sum_{i=1}^{nbins} \frac{(s_i - b_i)^2}{s_i + b_i}\right) \times 100$$

Cut&Count analysis

SM tH: significance=0.224 (BDT result: 0.227)

95% C. L. s. upper limit				
-2σ	-σ	exp. median	+σ	+2σ
5.00	6.71	9.32	12.96	17.38

ITC tH: significance=3.633

95% C. L. s. upper limit				
-2σ	-σ	exp. median	+σ	+2σ
0.31	0.42	0.59	0.82	1.10

Results presented without systematic uncertainty

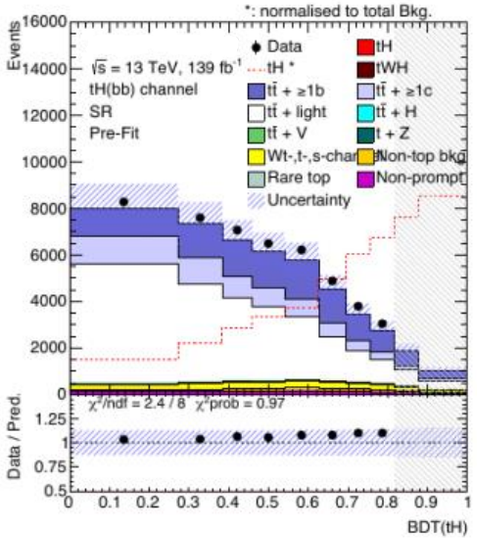
Fit on the BDT scores distributions

ATLAS Work-in-Progress

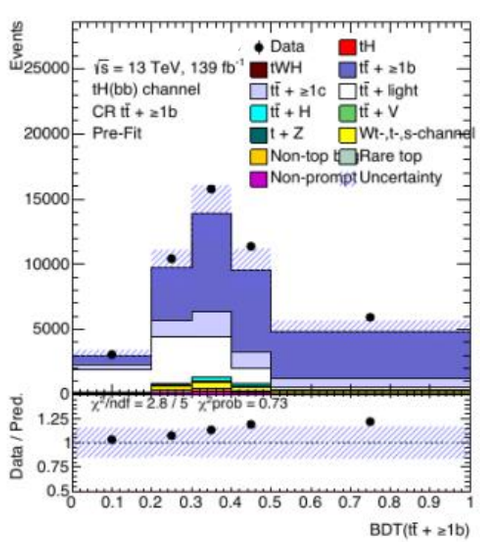
Pre-Fit results

The BDT scores distribution

SR



CR



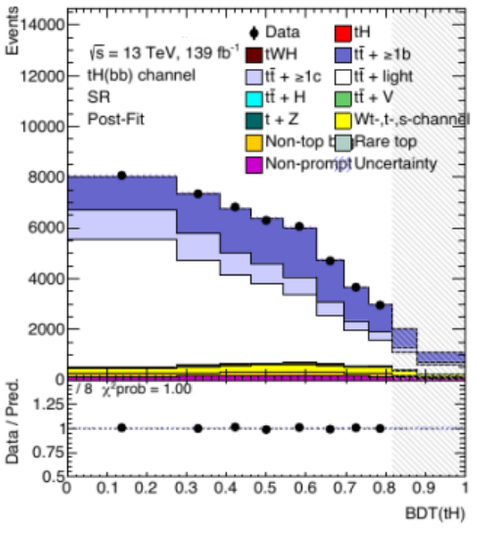
Expected mis-modeling in HF backgrounds
Good Data/MC ratio for all other distributions

ATLAS Work-in-Progress

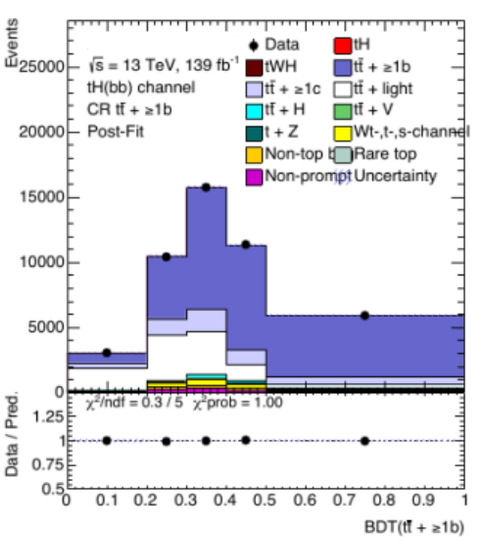
Post-Fit results

The BDT scores distribution

SR



CR

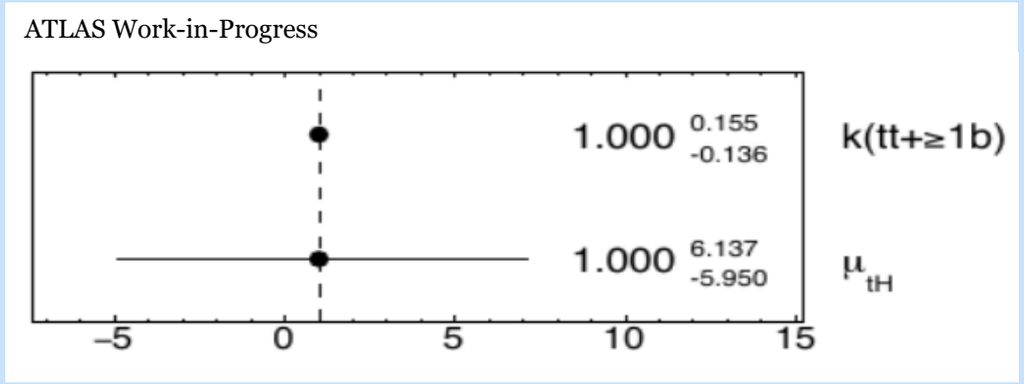


Good Post-fit agreement in fitted regions

Asimov Fit – limits and ranking plot

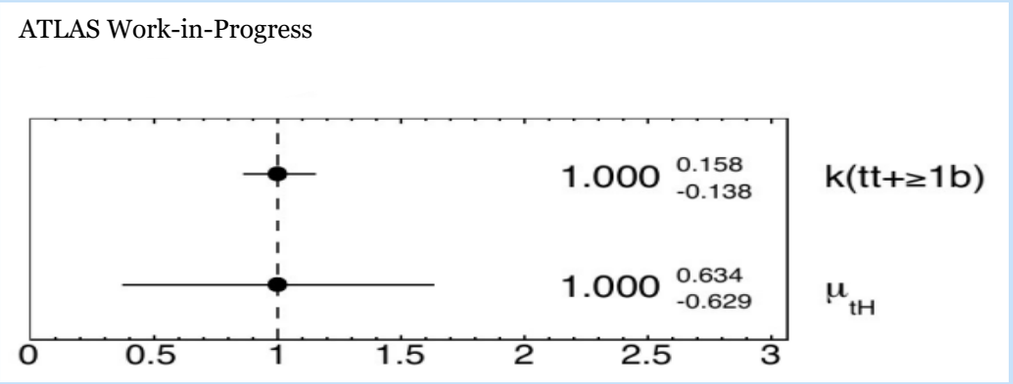
SM tH

95% C. L. s. upper limit				
-2σ	-σ	exp. median	+σ	+2σ
6.7	8.9	12.4	17.8	25.1



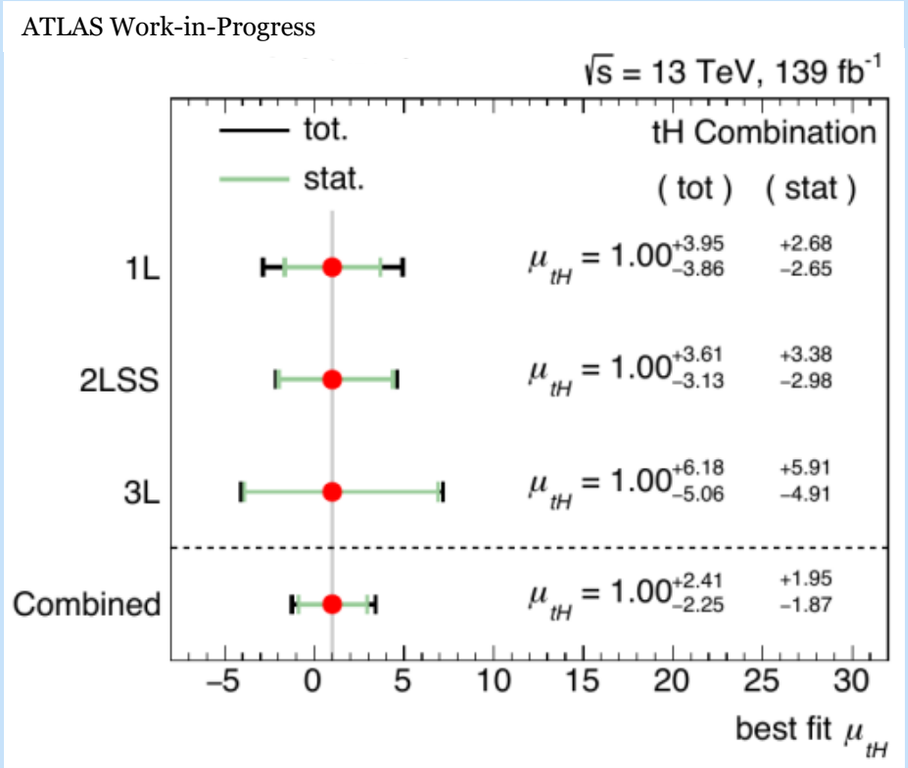
ITC tH

95% C. L. s. upper limit				
-2σ	-σ	exp. median	+σ	+2σ
0.66	0.88	1.23	1.71	2.3



- Mostly $t\bar{t}$ modeling systematics are high-ranked
  - Large statistics and big fluctuations
- The MC Statistics in the last bin of SR also high-ranked
  - Last bin is where tH purity is highest — low background stats expected
- Upper limit at 95% CL stands at μ = 12.4
  - Uncertainty on μ ~ ± 6

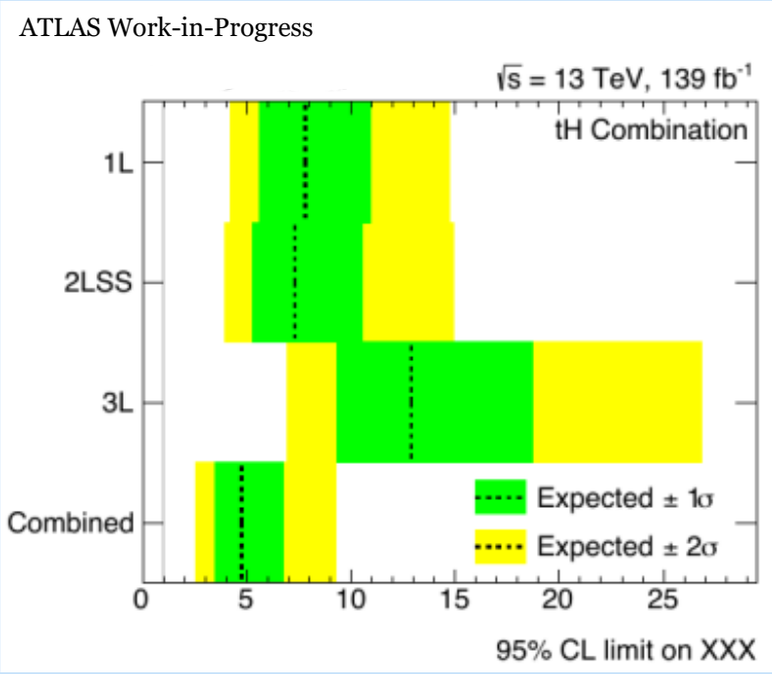
Complete Fit



- *Combination of all channels allows to discover ITC signal after unblinding*
- *SM signal requires Run 3 data to observe*

95% C.L.s upper limit

Expected	-2σ	-1σ	μ_{tH}	+1σ	+2σ
SM (y_t=+1)	**2.56**	**3.44**	**4.77**	**6.74**	**9.26**



Alternative Monte Carlo samples & analysis

- Alternative Monte Carlo samples were produced at generator level:

Process	tH (SM)	tH (ITC)	tt	$ttbb$	ttH	ttZ	tZ
Cross-section, fb	10.8	97	306 000	3600	72	23	18
Number of events	20 000	20 000	410 000	40 000	37 000	37 000	20 000

- Cut and Count analysis was applied to these samples:

Criteria	tH (SM)	tH (ITC)	tt	$ttbb$	ttH	ttZ	$tZbq$
$p_T^{\text{lead}}, p_T^{\text{sublead}}$	55.8%	53.1%	49.8%	46.2%	45.0%	44.7%	53.5%
p_T^{miss}	49.4%	47.1%	44.4%	41.3%	40.6%	40.5%	47.9%
N_b	29.9%	28.9%	2.7%	10.9%	28.1%	27.5%	25.6%
p_T^{fwd}	19.9%	19.1%	0.9%	4.7%	11.4%	11.6%	17.3%
$\lvert\Delta\eta\rvert$	16.3%	12.6%	0.5%	2.9%	6.8%	6.4%	11.2%
M_H, m_t	10.9%	8.0%	0.4%	0.9%	2.0%	1.8%	7.2%
Number of events	77	1519	59 000	4400	206	58	182

SM tH: significance=0.30

ITC tH: significance=6.0

Conclusions

- Fast generator-level analysis of tH channel was developed
- Results were applied to analysis of ATLAS experimental data
- C&C analysis was completed on private samples
- BDT analysis was completed (Roma group)
- Cross-check NN analysis performed
- New variables implemented for BDT
- C&C analysis was completed
- Preliminary Fit result produced (Manchester group)
- ITC signal can be observed with existing ATLAS data, SM signal can only be observed with Run 3 data

Thanks for your attention.