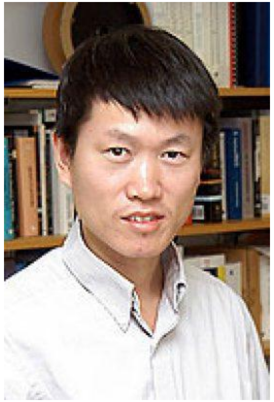
Madgraph Generation of 3-top NLO samples and BDT Optimization

Lynn Rong (University of Michigan)

Who is your advisor / group you are working with?

1. My Advisor and Group

My Advisor and Group



Jianming Qian

His current physics interests include measurements of the properties of the Higgs boson and searches for new physics beyond the Standard Model

My Mentor 👉

👈My Mentor’s Mentor

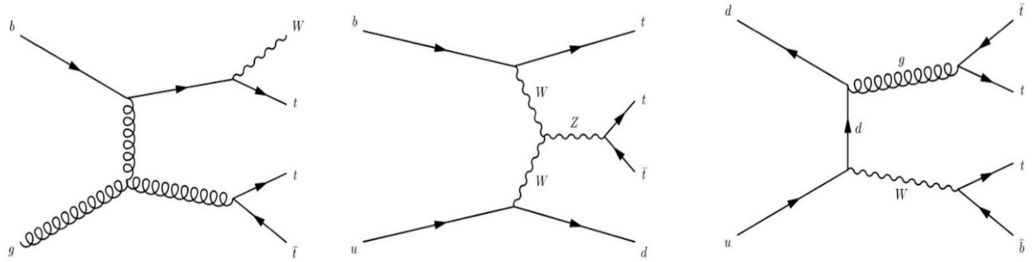
Mengju Tsai

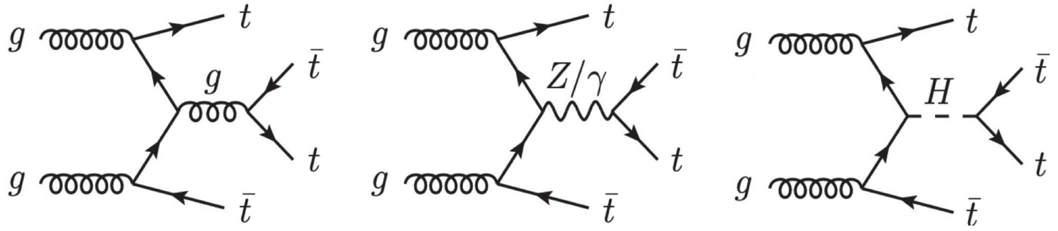
Leading contributor for first observation of SM four-top-quark production in multi-lepton channel

2. 3-top vs 4-top Next Leading Order Studies

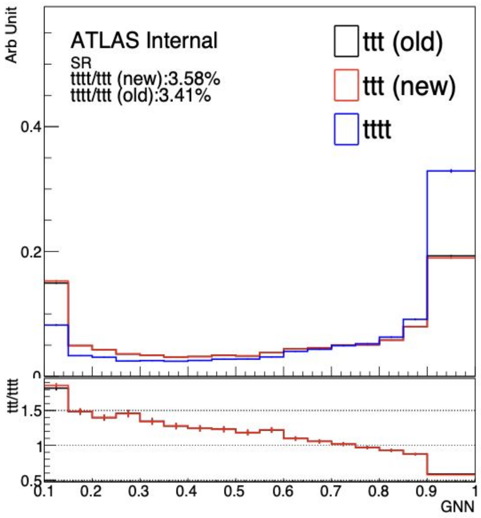
4-top process predicted by Standard Model was detected

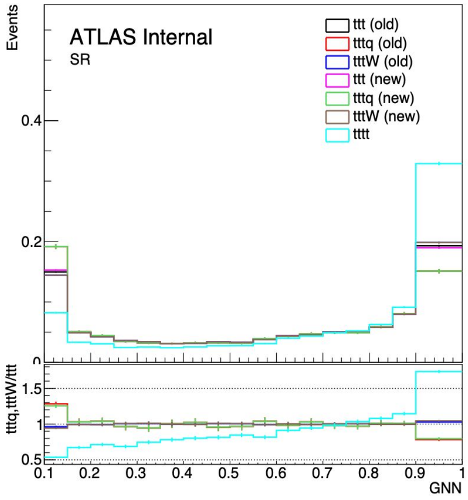
- 4 top generation is a rare process predicted by the Standard Model
- It has been observed by ATLAS
- However, it is hard to validate that the observed process is 4top generation because 4t and 3t have very similar kinematics



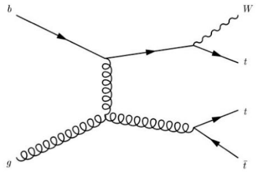


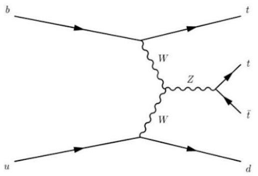
Motivation for 3-top NLO studies:

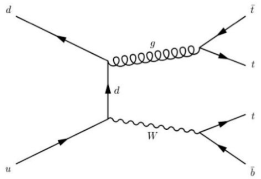




- 3-top process divided into ttt+q (including ttt+b) and ttt+W.
- ttt+W process is larger and expected to be more signal-like due to additional W.
- 3-top and 4-top is very similar -> hard to separate 3-top and 4-top even with the GNN distributions





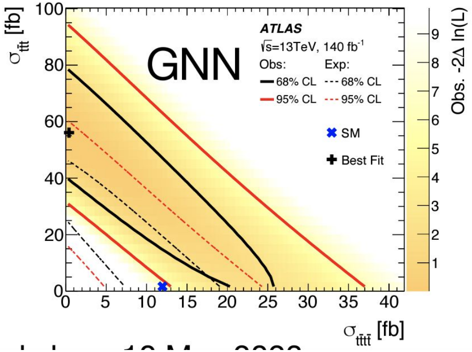


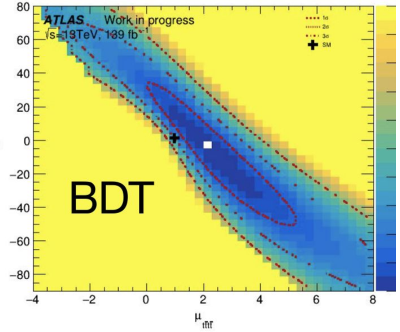
Motivation:

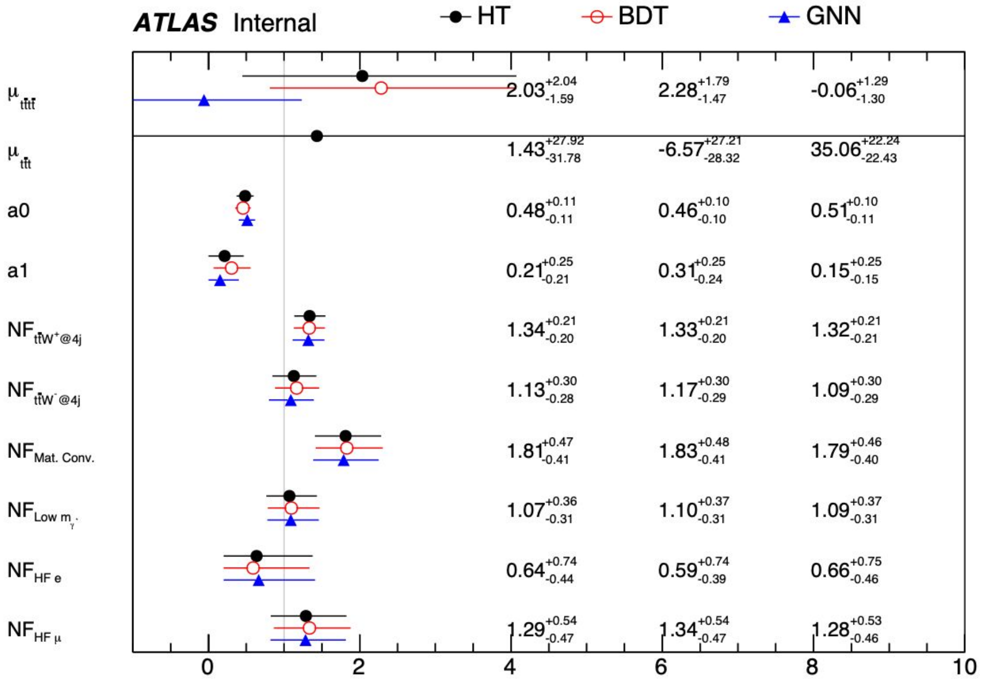
3top NLO sample would benefit the analysis with improved modeling

- Free floating both 4-top and 3-top
- Using HT and BDT, data prefers 4 top-quark production
- However using GNN, data prefers 3 top-quark production

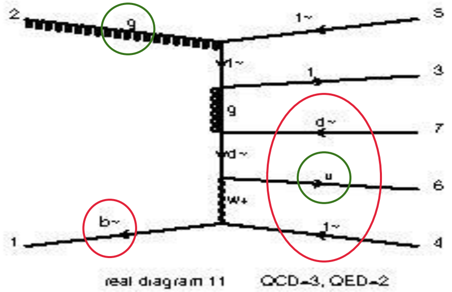
Best fit values away from SM prediction about 2σ



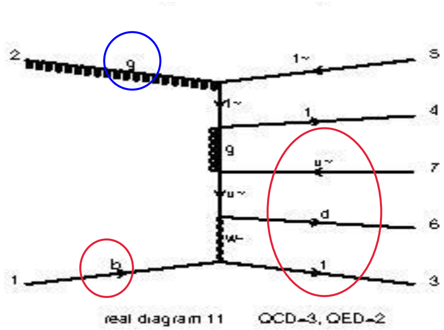




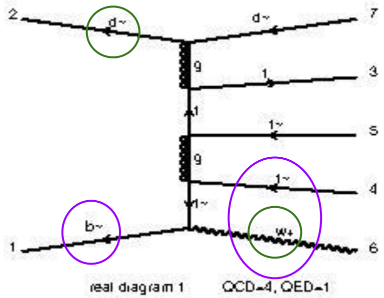
Comparing tttjp, tttjm, tttwp, tttwm Feynman Diagrams



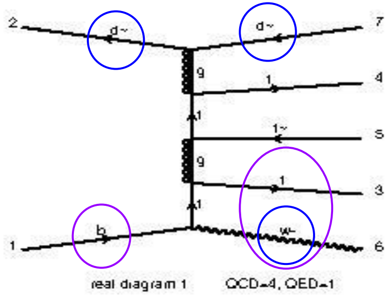
tttjp: P0_bxd_ttxtxu



tttjm: P0_bdx_tttxux



tttwp: P0_bxg_ttxtxwp



tttwm: P0_bg_tttxwm

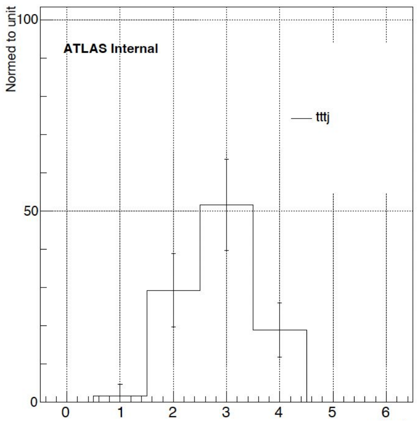
- The difference between tttxp and tttxm is that Jet/W charge is positive for p, and negative for m.
- Between tttj and tttw, one process generates a jet and another generates W boson.

3-top NLO cross sections produced Outside Atlas(MG alone)

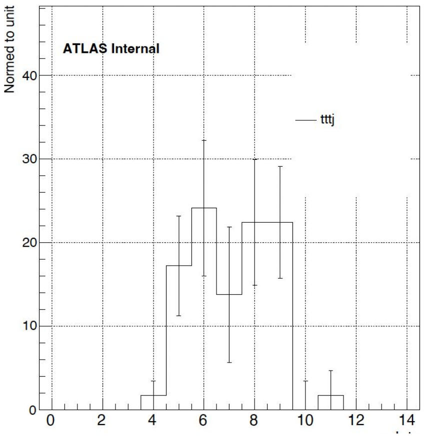
	tttjp	tttjm	tttj	tttwm	tttwp	tttW
NLO (reproduced theorist results)	0.2022 ±2.6e-03 fb	0.4442± 1.9e-03 fb	0.6464 fb	0.5185± 1.5e-03 fb	0.5212 ± 1.6e-03 fb	1.04 fb
NLO (theorist)			0.646 fb			1.02 fb
LO (reproduced theorist results)	0.1132± 2.1e-04 fb	0.2485 ±4.5e-04 fb	0.3617 fb	0.2919 ± 6.1e-03 fb	0.2923 ±6.1e-04 fb	0.5842 fb
LO (theorist)			0.363 fb			0.576 fb

- Cross sections from theorist are reproduced from our study
- tttW has a higher signal compared to tttj

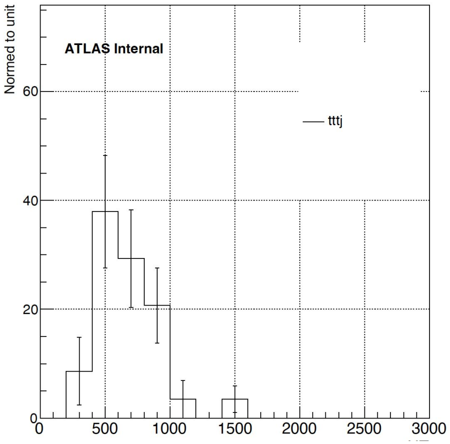
The tttjp sample produced inside ATLAS (10k events)



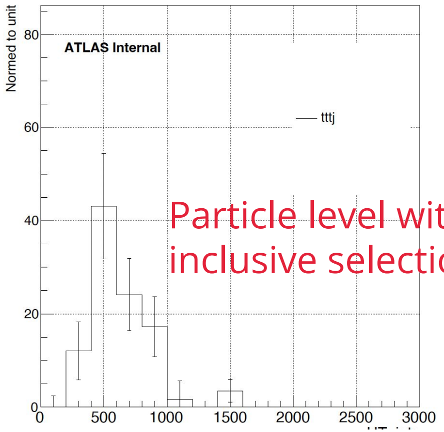
- For B jets, it peaks at 3 as expected since top decays into W+b.



- We expect to have >4 jets since each t decays into W+b, and tttj has an extra jet, so there's at least 4 jets

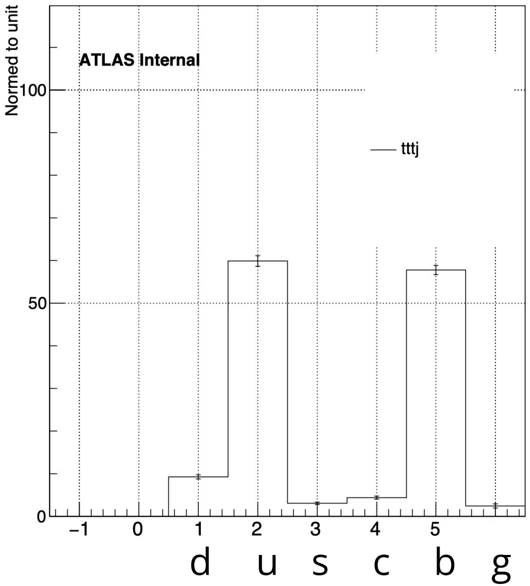


- The scalar sum of transverse momentum of the jets and leptons

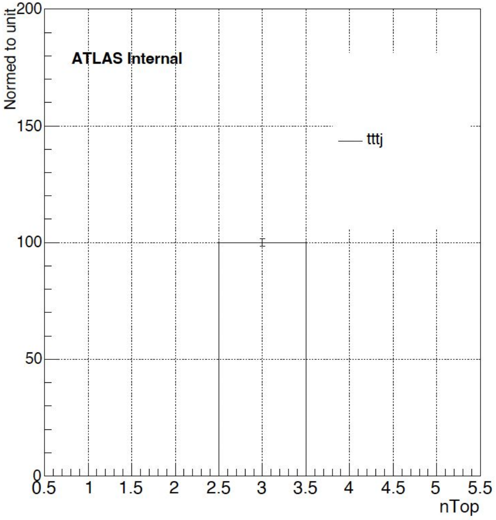


- The scalar sum of transverse momentum of the jets

The tttjp sample produced inside ATLAS (10k events)

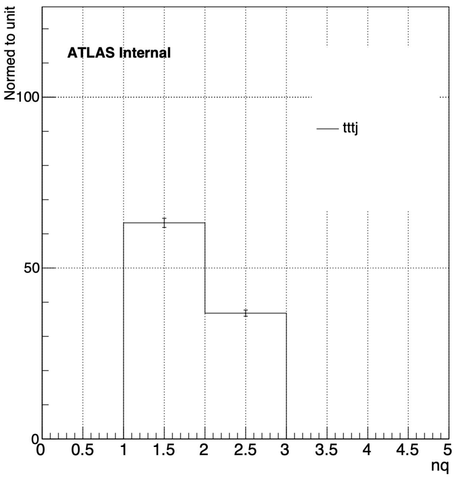


Type of jet or gluon



Number of Tops

- It is 3 as expected because it's 3-top production



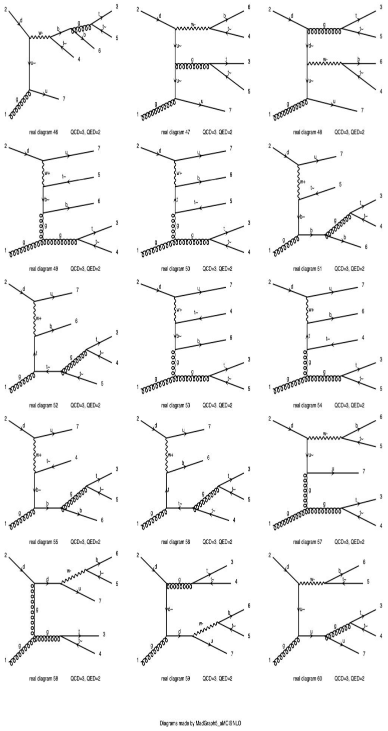
Number of quarks

- Either 1 or 2 as expected. At Leading order, there's only a jet production but at next leading order the jet decays into 4 or 5 quarks/gluon

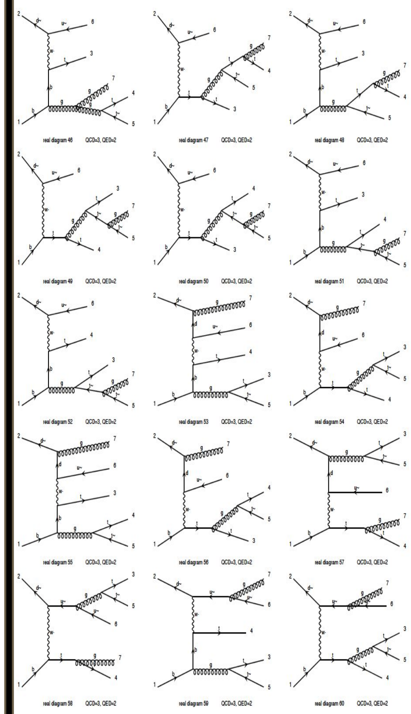
Parton level with inclusive selection

Feynman Diagrams

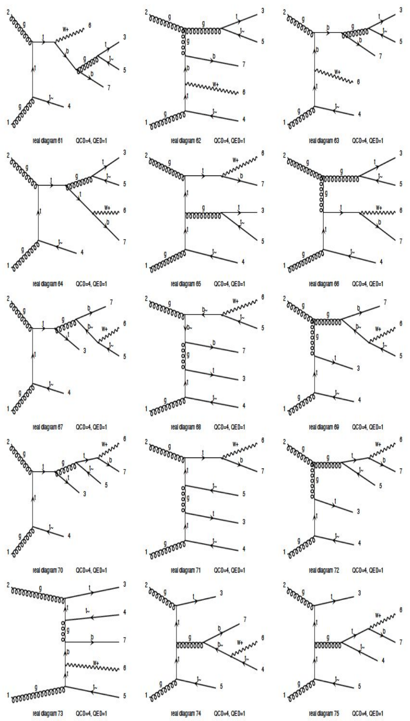
TTTJP



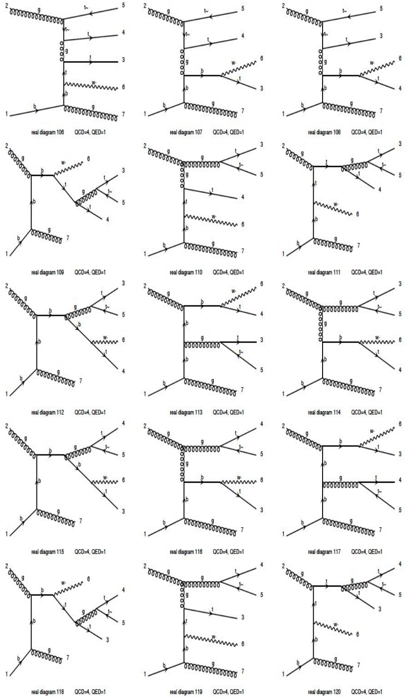
TTTJM



TTTWP



TTTWM



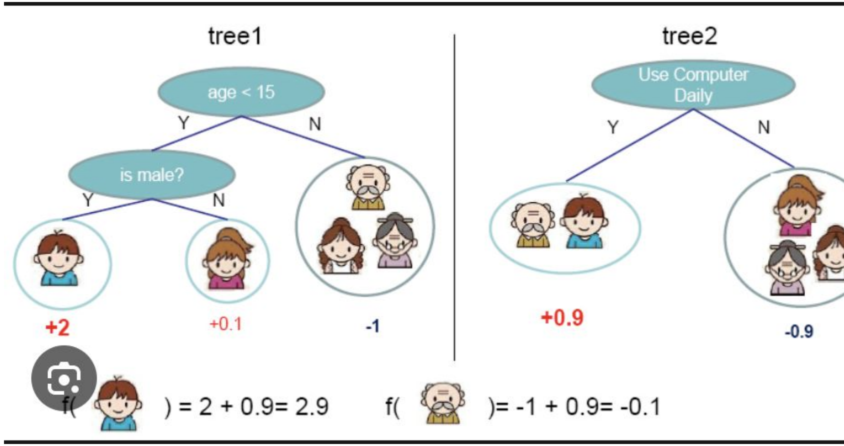
3. BDT OPTIMIZATION

What are Boosted Decision Trees?

- Boosted Decision Trees are a common Machine Learning technique in Particle Physics that separates signal from Background

- Data is recursively split based on their individual features

- An ensemble of decision trees are used, hence the boosting.

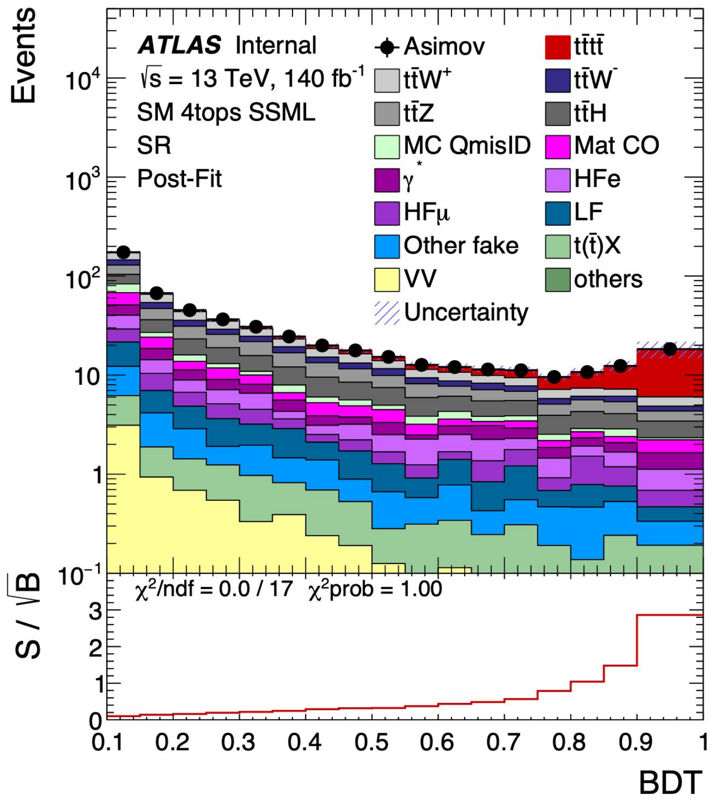


https://xgboost.readthedocs.io/en/stable/tutorials/r
del.html

Motivation

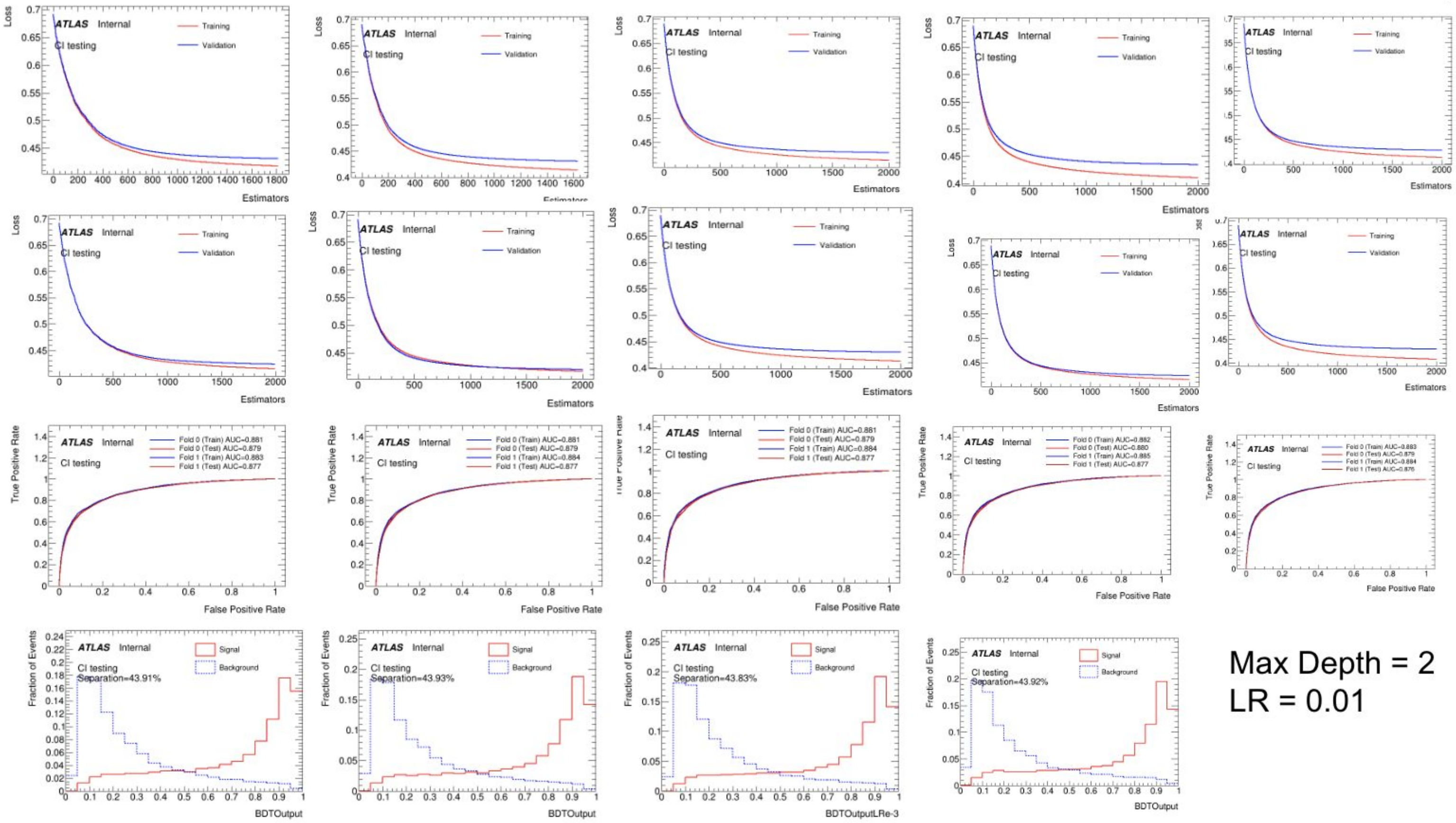
- Using BDT, we are able to distinguish tttt production from its remaining background

- The goal is to separate tttt from ttt, however, we don't have triple top production data yet for R23 ->

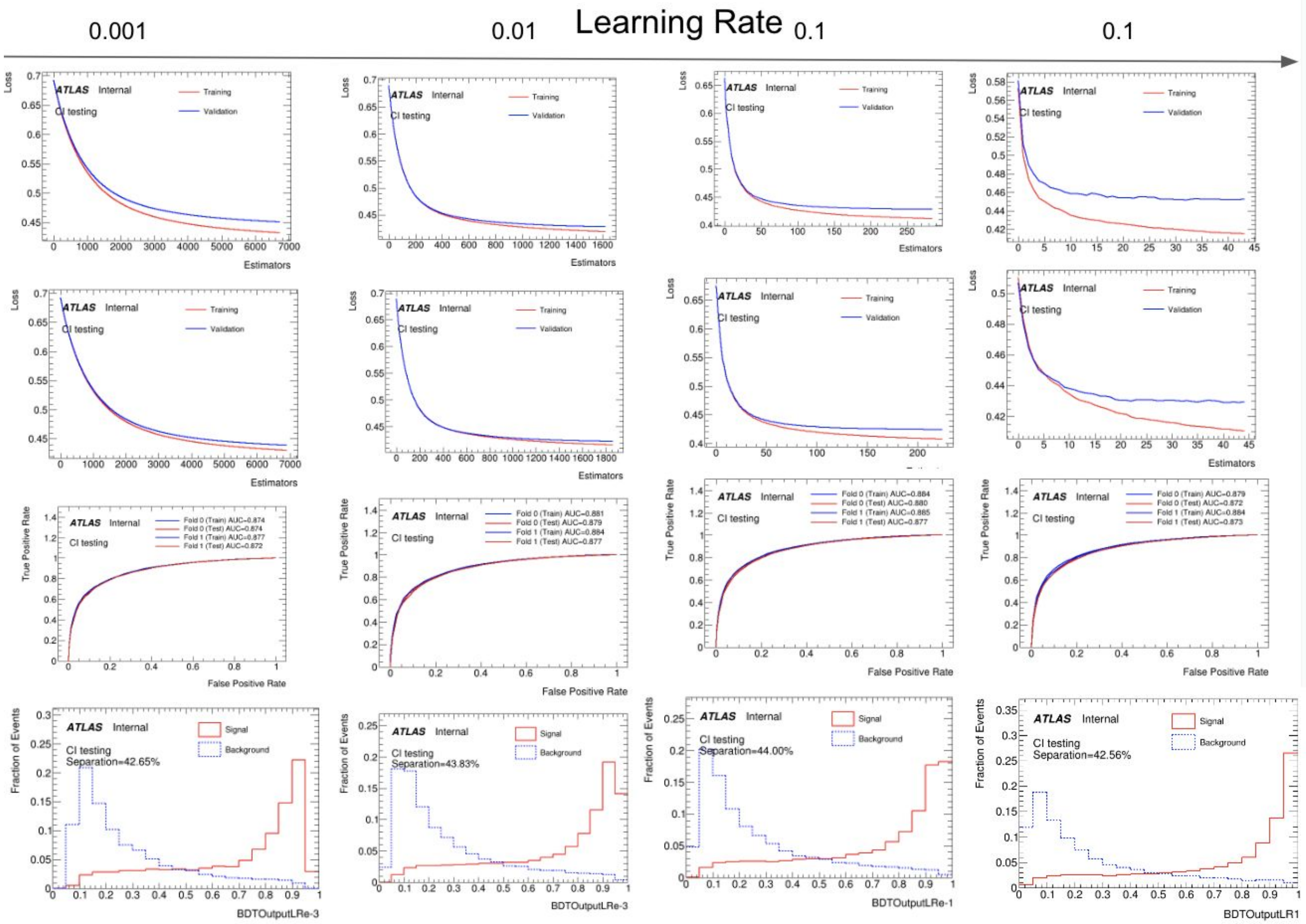


Max Feature

0 2 4 6 8 10



Max Depth = 2
LR = 0.01



Learning Rate
0.001
0.01
0.1
0.1
ATLAS Internal
CI testing
Training
Validation
Loss
Estimators
True Positive Rate
False Positive Rate
Fold 0 (Train) AUC=0.874
Fold 0 (Test) AUC=0.874
Fold 1 (Train) AUC=0.877
Fold 1 (Test) AUC=0.872
Fold 0 (Train) AUC=0.881
Fold 0 (Test) AUC=0.879
Fold 1 (Train) AUC=0.884
Fold 1 (Test) AUC=0.877
Fold 0 (Train) AUC=0.884
Fold 0 (Test) AUC=0.880
Fold 1 (Train) AUC=0.885
Fold 1 (Test) AUC=0.877
Fold 0 (Train) AUC=0.879
Fold 0 (Test) AUC=0.872
Fold 1 (Train) AUC=0.884
Fold 1 (Test) AUC=0.873
Fraction of Events
Signal
Background
Separation=42.65%
Separation=43.83%
Separation=44.00%
Separation=42.56%
BDTOutputLRe-3
BDTOutputLRe-3
BDTOutputLRe-1
BDTOutputLR1

BDT training

- Use MVA framework from ttZ ML group written by Steffen (instead of my private XGBoost BDT framework used in SM 4t)
- BDT is trained in SR (>=2b,>=6j,HT>500GeV) to reproduce similar results as before
- BDT training with same input variable used in SM 4t observation paper
  - DL1r upgraded to DL1d and other variables involving b-jet are tagged at 77% WP with DL1

Variable	Category	Description
$\sum_{i=0}^{3} w_{\mathrm{DL1r}}$	b-tagging	Sum of DL1r pseudo-continuous b-tagging score over leading four jets with highest DL1r scores
N_{jets}	Jet	Jet multiplicity
$\Delta R(\ell,\ell)_{\mathrm{min}}$	Distance	The minimum distance between any lepton pair
$p_{\mathrm{T}}^{\mathrm{jet0}}$	Jet	Transverse momentum of leading jet
$p_{\mathrm{T}}^{\mathrm{b\text{-}jet0}}$	Jet	Transverse momentum of leading b-tagged jet
$p_{\mathrm{T}}^{\mathrm{lep0}}$	Lepton	Transverse momentum of leading lepton
$E_{\mathrm{T}}^{\mathrm{miss}}$	Energy	Missing transverse energy
$\sum \Delta R(\ell,\ell)$	Distance	Sum of the distance between leading and sub-leading leptons in SS or leading, sub-leading and third-leading leptons in 3ℓ
$H_{\mathrm{T}}^{\mathrm{no\ lead\ jet}}$	Energy	Scalar sum of all lepton and jet pT except leading jet
$\Delta R(\ell,b)_{\mathrm{max}}$	Distance	The maximum distance between leptons and b-tagged jets
$p_{\mathrm{T}}^{\mathrm{jet5}}$	Jet	Transverse momentum of 6th leading jet
$\Delta R(j,b)_{\mathrm{min}}$	Distance	The minimum distance between b-tagged jets and jets
$p_{\mathrm{T}}^{\mathrm{jet1}}$	Jet	Transverse momentum of sub-leading jet

Ntree (patience=20)	400
Learning rate	0.025
MaxDepth	2
MaxFeatures	8

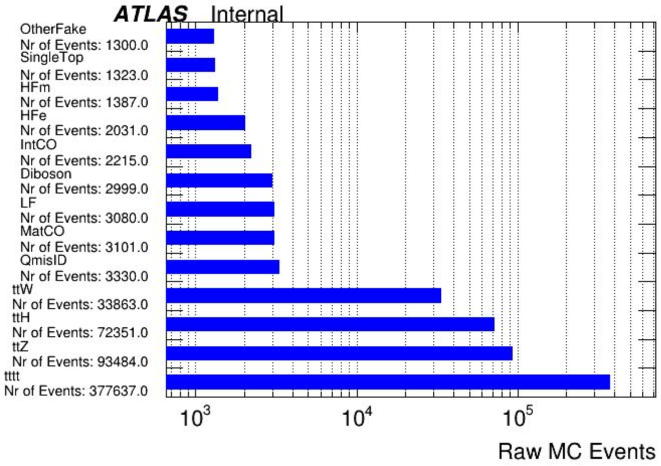
Difference between R21 and R22 (from HuiChi Lin)

	R21	R22
Trigger	Single lepton trigger (SLT) & Di-lepton trigger (DLT)	Single lepton trigger (SLT)
Isolation WP	PLImprovedTight (electrons & muons)	TightTrackOnly_FixedRad (electrons) PflowTight_VarRad (muons)
B-tagging WP	DL1r 77%	DL1d 77%
ttW	Sherpa 2.2.10 ttW QCD+EW (with all sub-leading correction)	Sherpa 2.2.8 NLO QCD
QmisID	Data-driven	MC

Looser PLIV

BDT training

- Samples are trained with the nominal MC weight
- Negative weight treatment: removes negative weights and scales remaining weights down to account for difference (previously just ignore negative weight)
- 2 Fold training depending on the event number with additional validation set from 25% of data (might expect gaining sensitivity with more folds)
- No additional increase of weight for ttW 7-jet and ttW 8-jet, nor increase the whole ttW weight (ttW weight can be further tuned to achieve better separation)

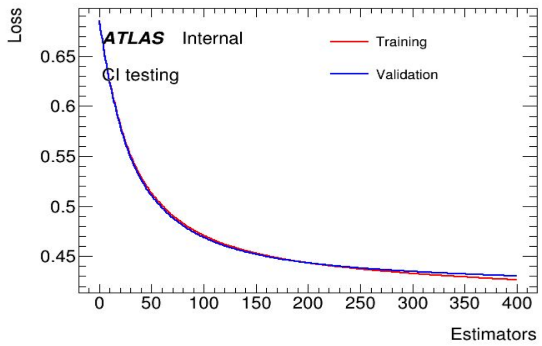


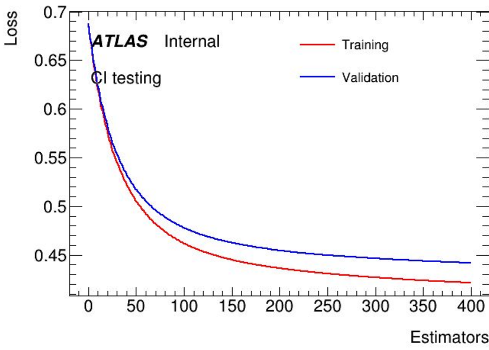
BDT performance

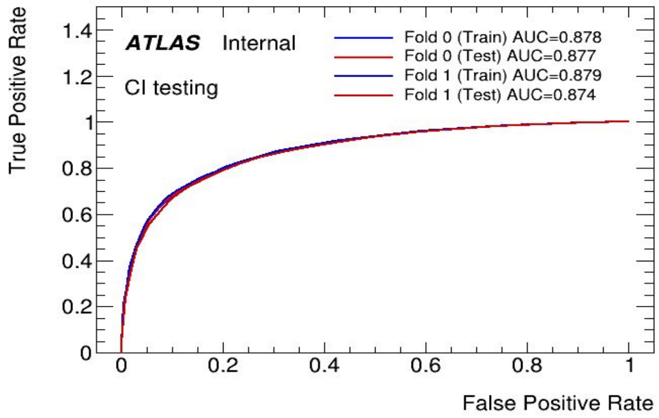
- Model 0 seem to match well, Model 1 could do better
- AUC in similar level as before

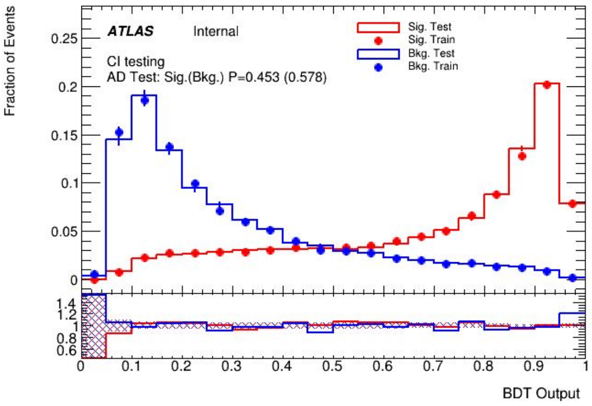
Model 0

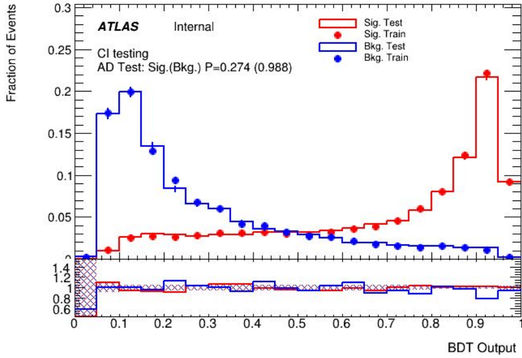
Model 1





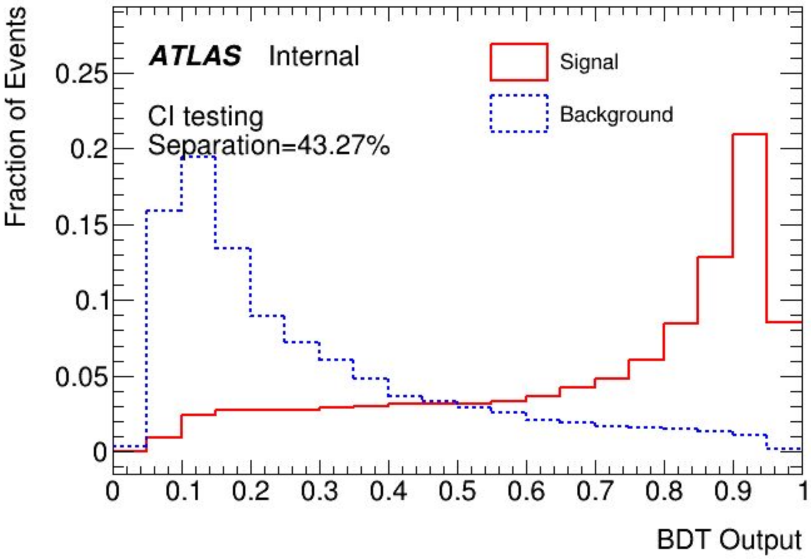


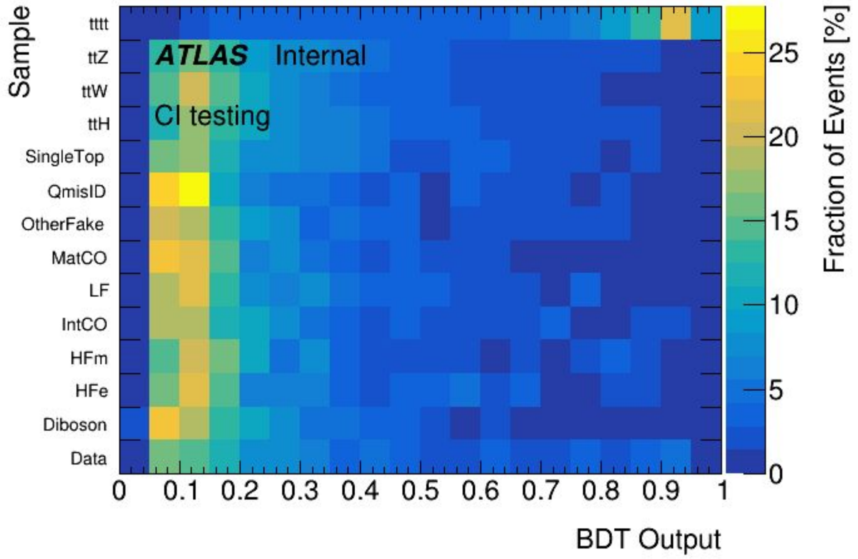




BDT performance

- Separation power seems okay enough to have the preliminary sensitivity

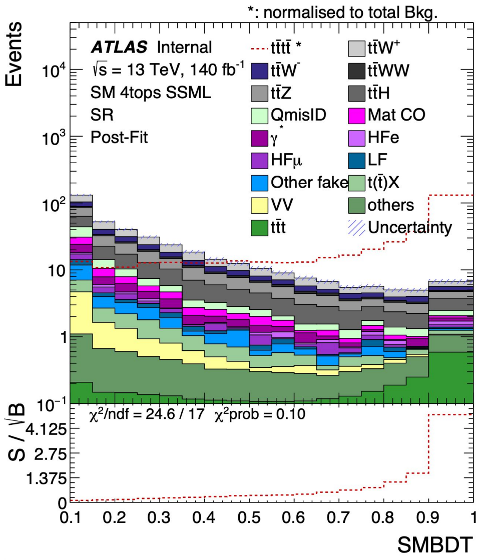




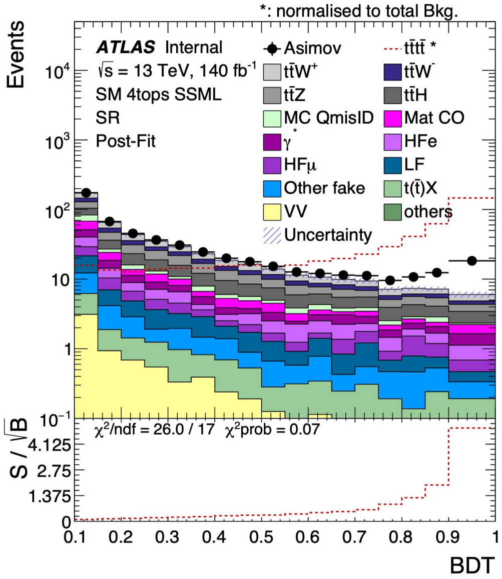
Final MVA shape and S/B

- Only simple optimization for the moment, still have room to improve like increase the ttW weight to seize a better separation

R21 Sig: 4.15



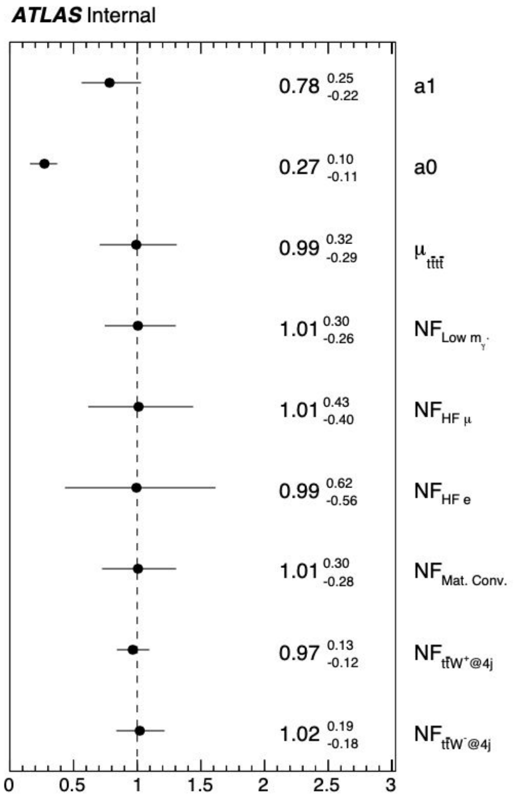
R22 Sig: 4.02 σ

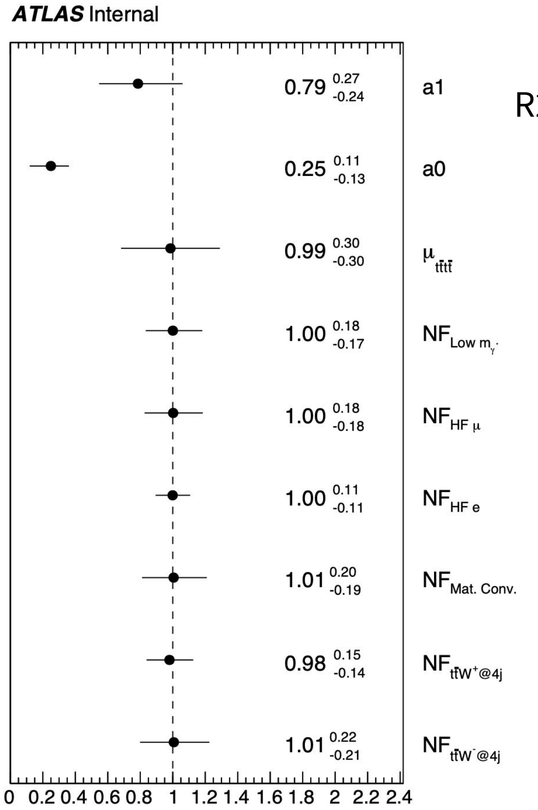


R22	Second to last bin	Last bin
S	5.21	12.23
B	7.20	6.02
S+B	12.41	18.24
S/B	**0.72**	**2.03**
S/sqrt(B)	**1.94**	**4.98**

Final fit results

R21





R22

- R22 result also return to injection. NF stays around 1 and ttw dd have similar scaling factors.

Final Fit Results

R21

ATLAS Internal

	$NF_{t\bar{t}W^-@4j}$	$NF_{t\bar{t}W^+@4j}$	$NF_{Mat.\ Conv.}$	$NF_{HF\ e}$	$NF_{HF\ \mu}$	$NF_{Low\ m_{\gamma^*}}$	$\mu_{t\bar{t}t\bar{t}}$	a0	a1
$NF_{t\bar{t}W^-@4j}$	100.0	43.8	-10.1	-1.2	-37.9	-3.7	-5.0	26.8	-39.0
$NF_{t\bar{t}W^+@4j}$	43.8	100.0	-9.5	-8.5	-24.4	-2.1	-2.3	38.9	-54.8
$NF_{Mat.\ Conv.}$	-10.1	-9.5	100.0	-14.0	1.3	-32.7	0.5	1.0	-1.4
$NF_{HF\ e}$	-1.2	-8.5	-14.0	100.0	-41.6	-5.1	-1.5	-4.0	5.1
$NF_{HF\ \mu}$	-37.9	-24.4	1.3	-41.6	100.0	1.3	3.2	3.3	-1.3
$NF_{Low\ m_{\gamma^*}}$	-3.7	-2.1	-32.7	-5.1	1.3	100.0	0.0	0.4	-2.1
$\mu_{t\bar{t}t\bar{t}}$	-5.0	-2.3	0.5	-1.5	3.2	0.0	100.0	-21.2	11.6
a0	26.8	38.9	1.0	-4.0	3.3	0.4	-21.2	100.0	-95.1
a1	-39.0	-54.8	-1.4	5.1	-1.3	-2.1	11.6	-95.1	100.0

R22

ATLAS Internal

	$NF_{t\bar{t}W^-@4j}$	$NF_{t\bar{t}W^+@4j}$	$NF_{Mat.\ Conv.}$	$NF_{HF\ e}$	$NF_{HF\ \mu}$	$NF_{Low\ m_{\gamma^*}}$	$\mu_{t\bar{t}t\bar{t}}$	a0	a1
$NF_{t\bar{t}W^-@4j}$	100.0	43.4	-11.9	2.3	-35.6	-6.2	-5.0	23.8	-36.0
$NF_{t\bar{t}W^+@4j}$	43.4	100.0	-7.0	2.8	-30.9	-6.6	-4.4	36.5	-51.9
$NF_{Mat.\ Conv.}$	-11.9	-7.0	100.0	-15.7	7.1	-30.0	-0.3	4.2	-6.5
$NF_{HF\ e}$	2.3	2.8	-15.7	100.0	-60.3	-8.6	2.0	-2.8	1.4
$NF_{HF\ \mu}$	-35.6	-30.9	7.1	-60.3	100.0	3.1	1.4	1.7	1.3
$NF_{Low\ m_{\gamma^*}}$	-6.2	-6.6	-30.0	-8.6	3.1	100.0	-1.4	0.7	-1.3
$\mu_{t\bar{t}t\bar{t}}$	-5.0	-4.4	-0.3	2.0	1.4	-1.4	100.0	-30.4	19.2
a0	23.8	36.5	4.2	-2.8	1.7	0.7	-30.4	100.0	-94.7
a1	-36.0	-51.9	-6.5	1.4	1.3	-1.3	19.2	-94.7	100.0

4. Next Steps

Summary and Next Steps

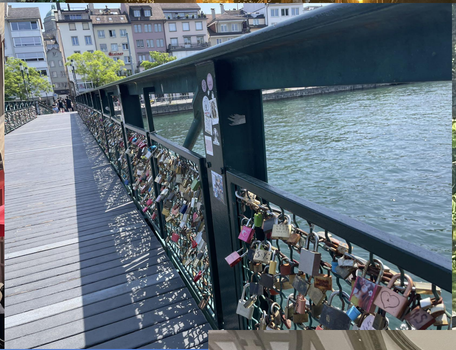
1. 3-top process vs 4-top process
    a. We successfully reproduced theorist cross sections of tttj+, tttj-, tttW+, and tttW- at LO and NLO with dynamic scales (HT/2)
    b. However, we need to reproduce inside atlas samples for tttj-, tttW+ and tttW- which we haven't successfully accomplished yet
2. BDT trained with R22 samples (continued work from Meng-Ju)
    a. Hyperparameter optimization resulted in less loss, better matched training and learning data, and effected prevented overtraining
    b. BDT samples could be better trained
    c. Further tunes with other parameters and set up an optimized BDT model with optimized hyperparameter, k-folding and weight fraction strategy

5. Zurich! (and a lake near france)

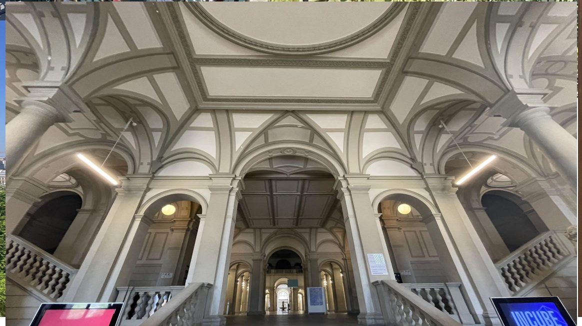






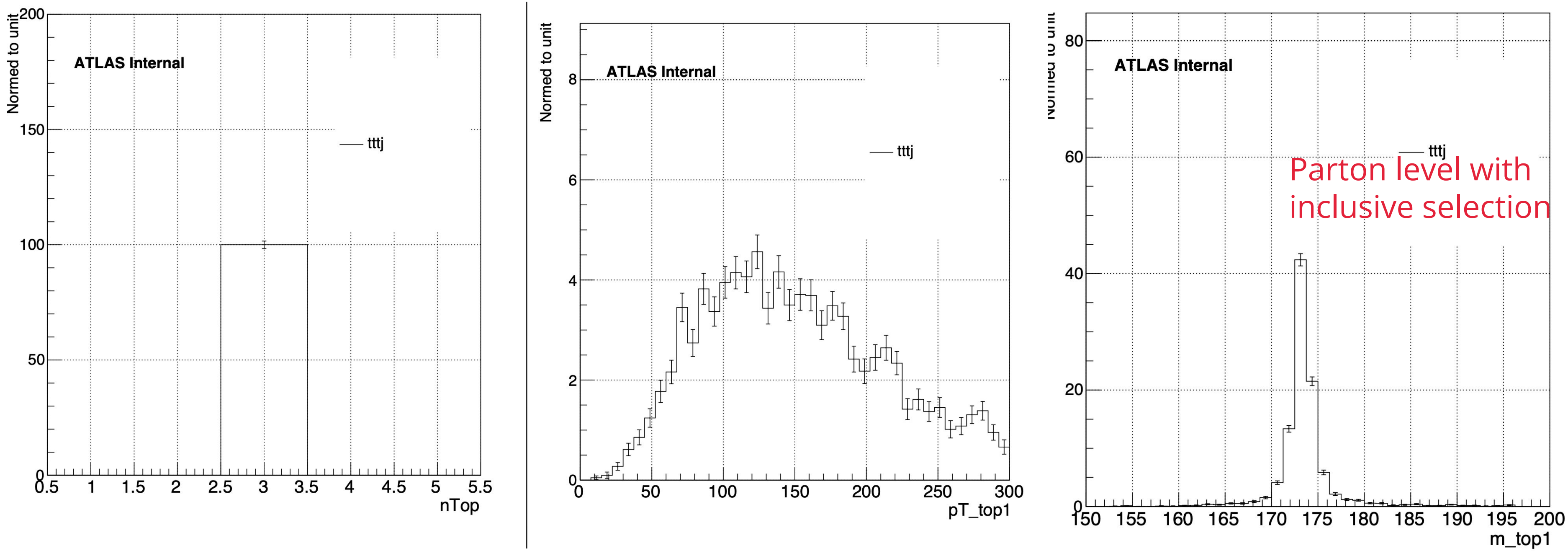





Backup

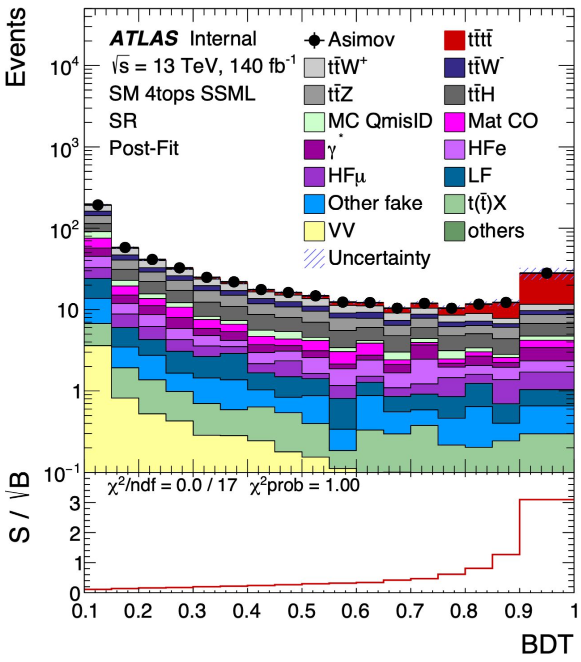
The tttjp sample produced inside ATLAS (10k events)

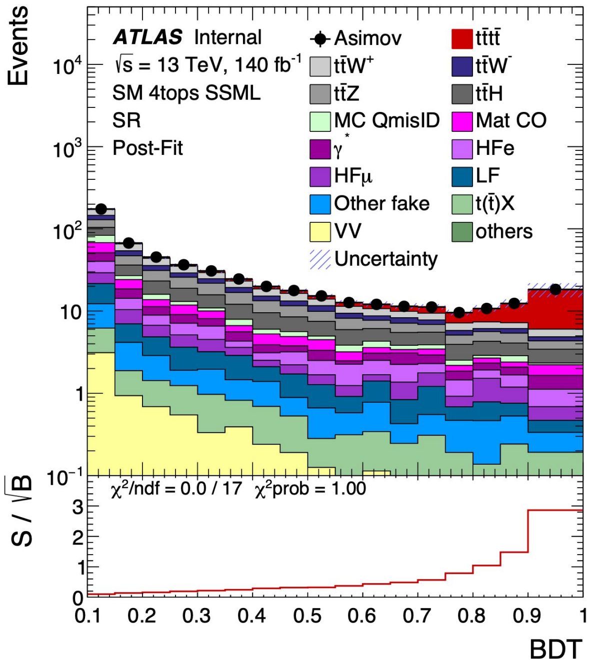


Asimov vs. post-fit

- Good agreement between Asimov and post-fit as expected

MaxDepth = 6





MaxDepth = 2