Machine learning in the search for $t\bar{t}H(\rightarrow b\bar{b})$ with the ATLAS detector

29th Nov. 2017 @Les jardins d'Anjou

- Ziyu GUO
- Thesis supervisors: Cristinel DIACONU, Thierry ARTIERES, Yann COADOU

Outline:

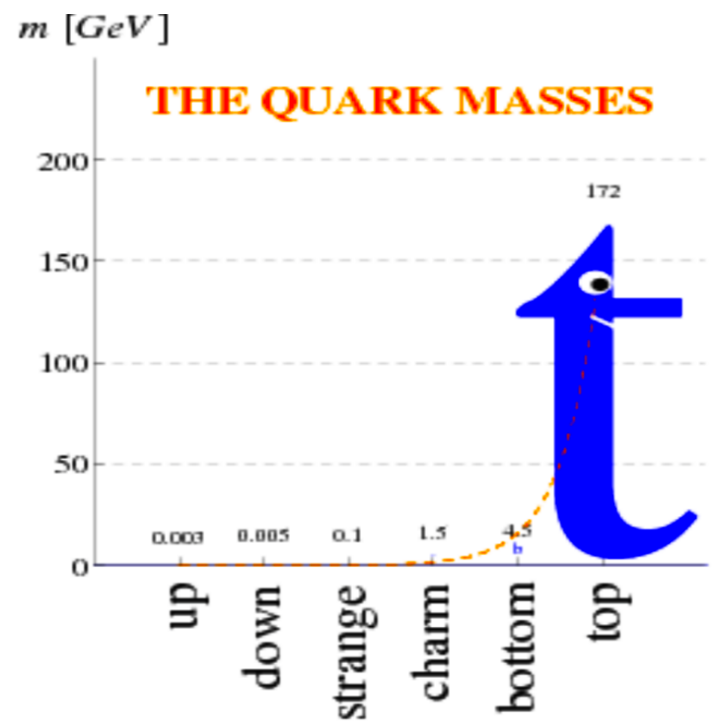
- Motivations for $t\bar{t}H(\rightarrow b\bar{b})$ channel
- Reconstruct the $t\bar{t}H$ signal: BDT
- Classify the signal ($t\bar{t}H$) and backgrounds ($t\bar{t}+jets$): BDT
- 2-dimensional BDT: a new way to categorize the events
- Preliminary tests with deep neural networks
- Summary

Motivations for $t\bar{t}H(\rightarrow b\bar{b})$ channel

- Higgs boson already discovered at the LHC in 2012, (Nobel prize in 2013)
  - Particles get mass by interaction (coupling) with Higgs field

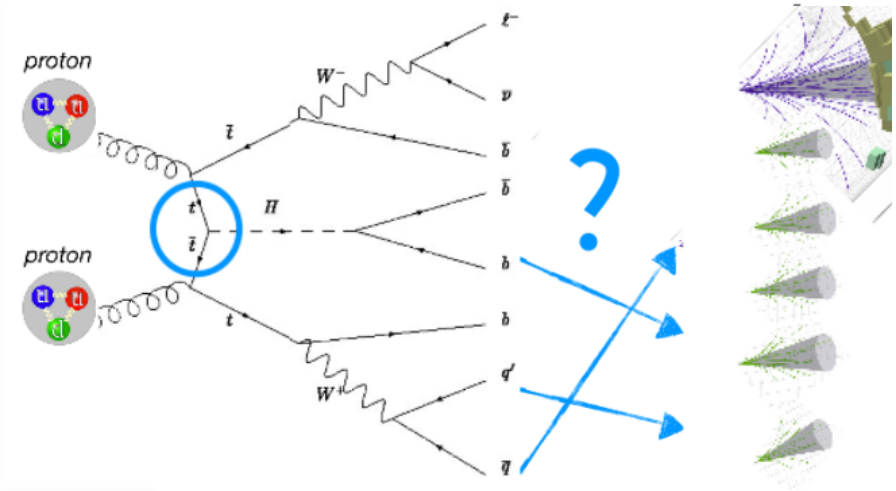
- Still many unknowns, especially intensity of couplings:
  - Higgs-boson coupling ➔ consistent with standard model prediction
  - Higgs-fermion coupling➔ established for leptons, not yet for quarks

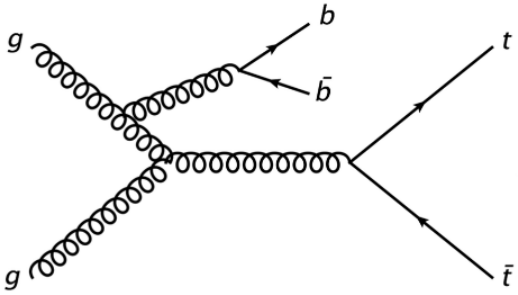
- Higgs-Top quark coupling is the most interesting:
  - Much stronger than other Higgs-quark couplings
  - Not yet directly measured



Analysis motivation

- $t\bar{t}H(\rightarrow b\bar{b})$: direct measurement of Higgs-Top coupling



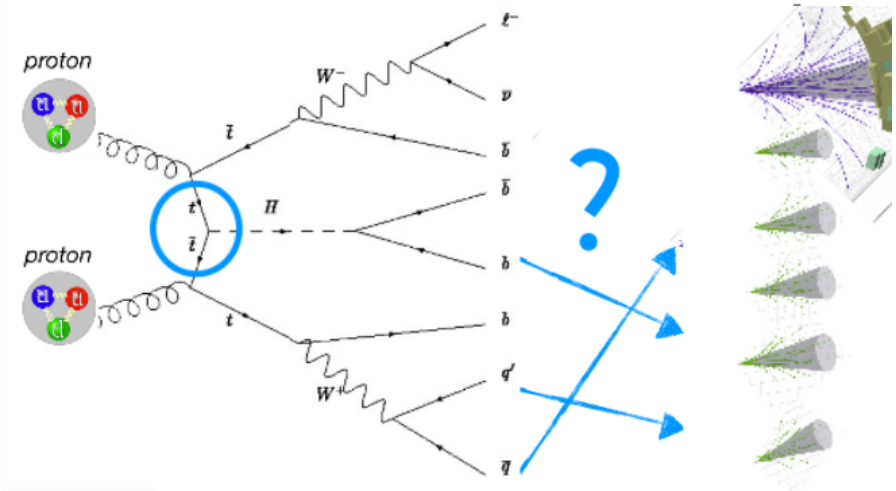


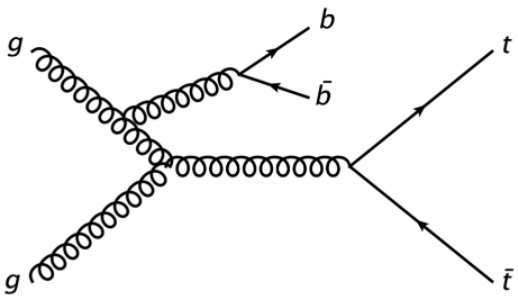
(c)

- Reconstruct signal:
  - Multi-jet final state: 6 jets including 4 b-jets from Higgs and tops, but also other jets from additional radiation and pileup
  - Need to find the correct assignments between observed jets and quarks

Analysis motivation

- $t\bar{t}H(\rightarrow b\bar{b})$: direct measurement of Higgs-Top coupling

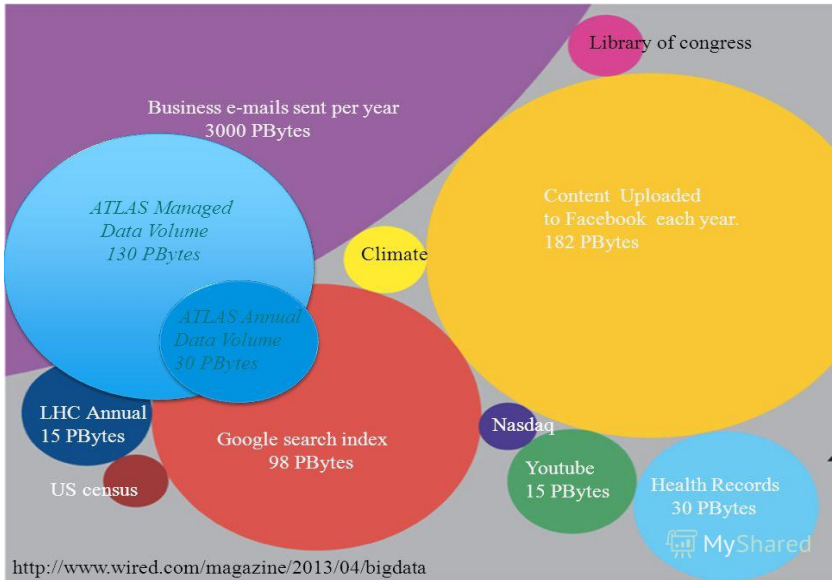




(c)

- Separate $t\bar{t}H$ events from $t\bar{t}$ backgrounds:
  - Small $t\bar{t}H$ x-section w.r.t. massive backgrounds from LHC and ATLAS, mainly $t\bar{t}$+jets
  - $\sim$ 5% s/b ratio in the most signal-rich region

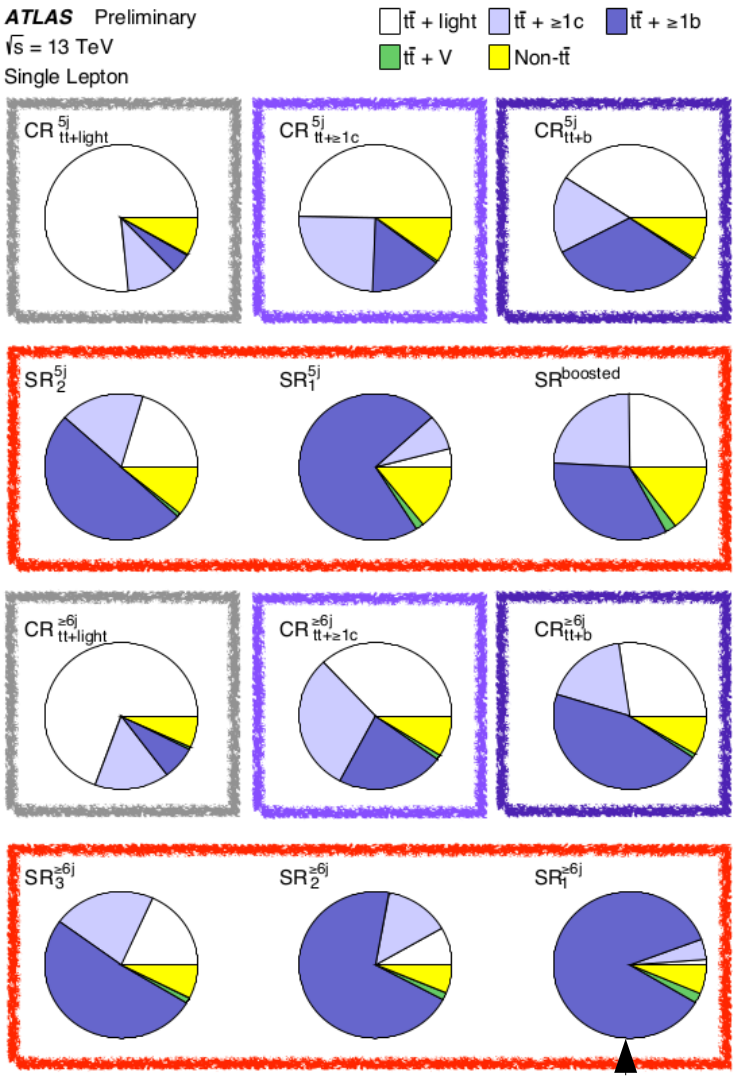
Need big data and associated tools!



$t\bar{t}H(\rightarrow b\bar{b})$ event selection

- Two separate channels depending on the number of light leptons (e, μ) from W decay: single lepton channel, dilepton channel

- Signal rich region:
  - $t\bar{t}H$ purity 1.6%-5.3%
  - Highest purity: 4 very tight b-tags
  - Dominant bkg: $t\bar{t}$+jets, especially $t\bar{t}$+HF
  - Multivariate techniques are used to further separate $t\bar{t}H$ from the bkg.
- Control regions:
  - Define separately for different jet flavors of $t\bar{t}$+jets bkg
  - Constraints on backgrounds



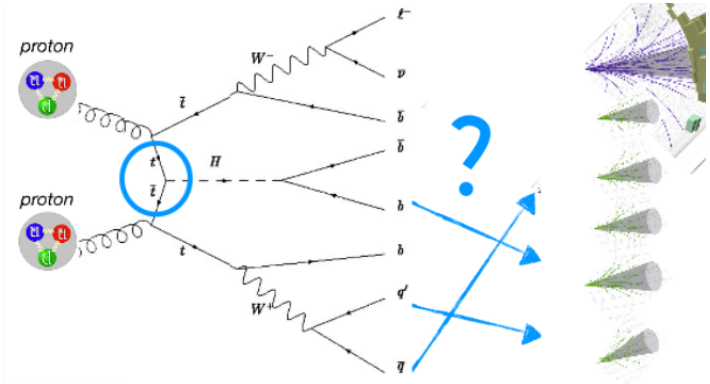
Highest $t\bar{t}H$ purity

Outline:

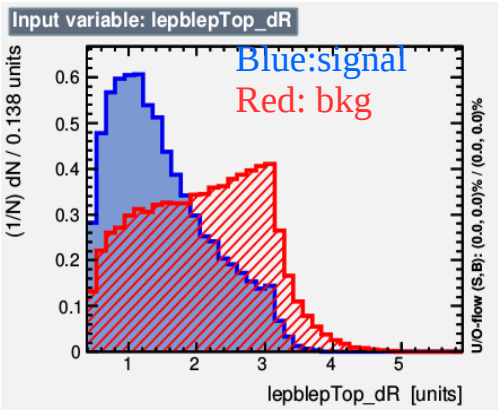
- Motivations for $t\bar{t}H(\rightarrow b\bar{b})$ channel
- Reconstruct the $t\bar{t}H$ signal: BDT
- Classify the signal ($t\bar{t}H$) and backgrounds ($t\bar{t}+jets$): BDT
- 2-dimensional BDT: a new way to categorize the events
- Preliminary tests with deep neural networks
- Summary

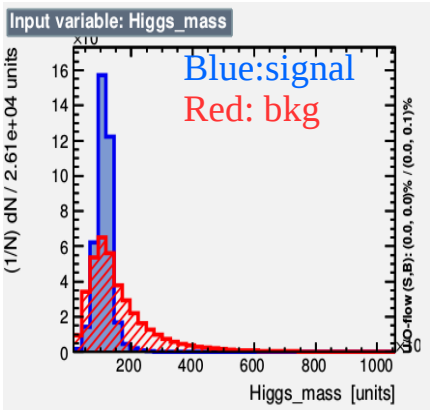
Reconstruction Boosted Decision Trees (BDT)

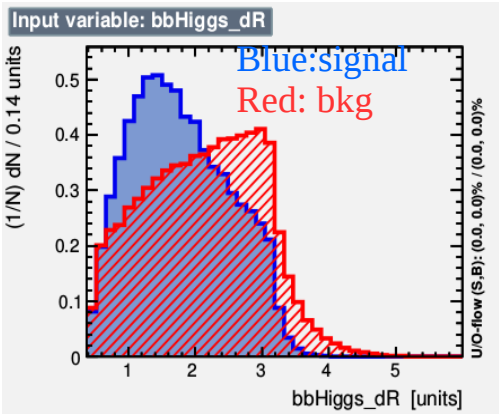
- Train the BDT with the MC simulation of $t\bar{t}H$ signal
- Do combinations of jets and partons
  - Signal: correct assignment
  - Background: all other combinations

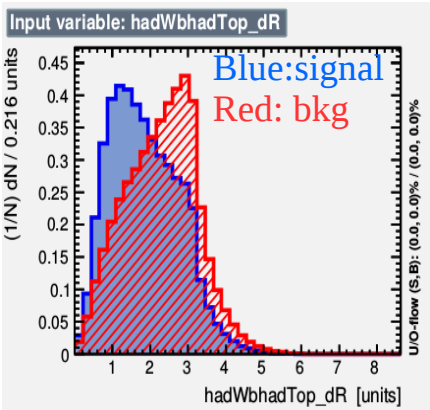


- Two types of BDT are trained:
  - Only $t\bar{t}$ system kinematics as input
  - $t\bar{t}$ system kinematics + variables with Higgs decay products
  - (Will see later) The best performance can be achieved by reco BDT using Higgs info, but will bias the Higgs-related variables of bkg to signal like, losing discriminating power as input for classification later.



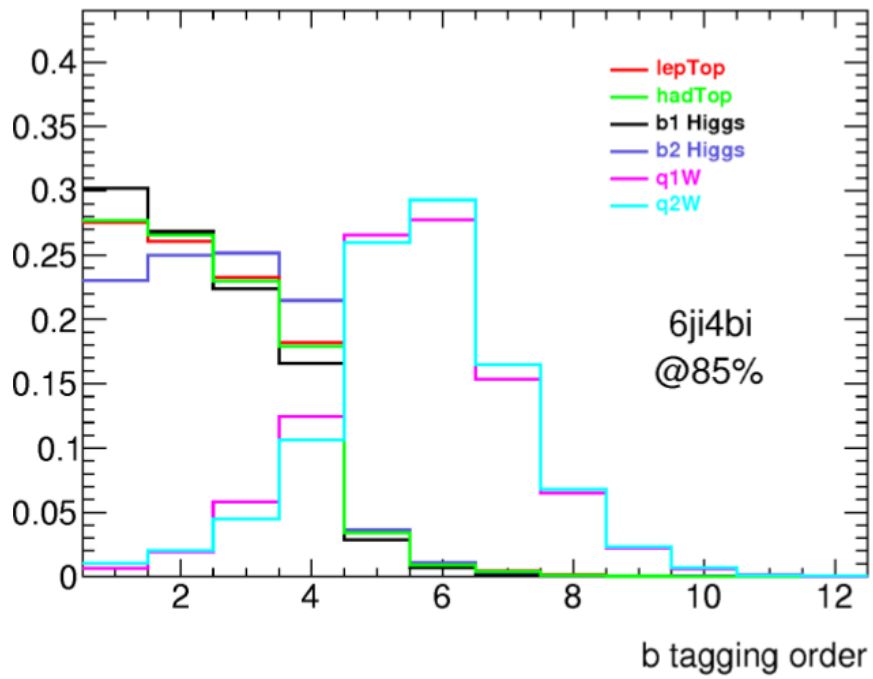






Reconstruction BDT: use b-tagging info

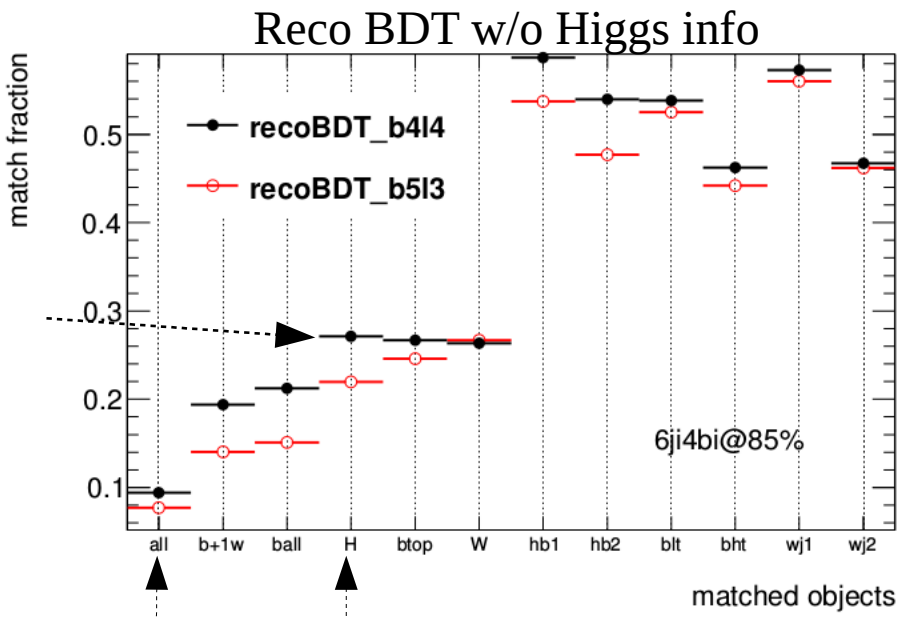
- Relatively often a light quark leads to a jet that has a higher b-tagging probability than some of the real b-jets

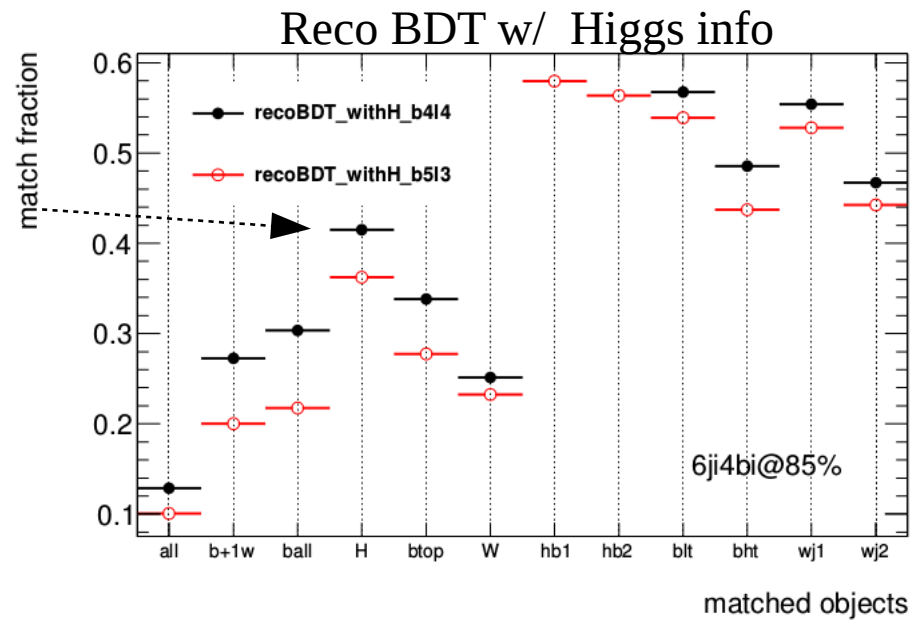


- For reconstruction BDT, to increase the signal (correct matching) and limit the bkg (wrong matching): using b-tagging weight
  - Ranking jets by their MV2 b-tagging weight (probability to be b-tagged)
  - The leading m jets are considered as b-jets, the leading n cannot be used as light jets.
  - Test various (m,n), and (4,4) gives the best results.

Reconstruction BDT: performance

- The fraction of the correctly reconstructed jets by BDT





Event fraction of the two b-jets from Higgs are correctly matched

All the jets are correctly matched

- The best performance: the Higgs boson is correctly reconstructed in 42% w/ Higgs info (28% w/o Higgs info) of the $t\bar{t}H$ events

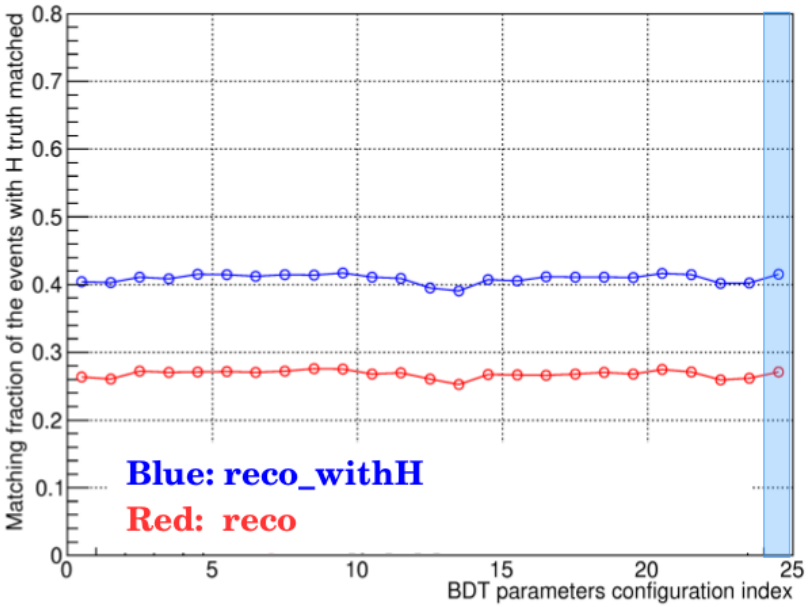
Event Reconstruction: more studies

- Different BDT parameters are tested.
- The blue numbers are the configurations being used

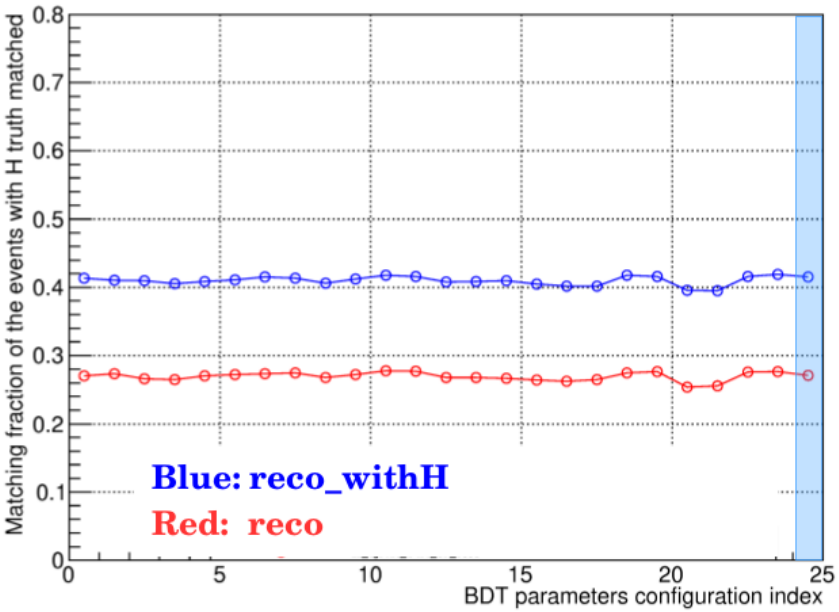
  - BoostType: AdaBoost
  - AdaBoostBeta = 0.15, 0.5
  - Ntrees = 400, 100
  - MinNodeSize = 4%, 1%
  - MaxDepth = 3, 5, 10
  - Ncuts = 80, 10

  - BoostType: Grad
  - Shrinkage = 0.1, 1
  - Ntrees = 400, 100
  - MinNodeSize = 4%, 1%
  - MaxDepth = 3, 5, 10
  - Ncuts = 80, 10

The fraction of the events with the jets from Higgs truth matched, using **AdaBoost**



The fraction of the events with the jets from Higgs truth matched, using **Gradient boost**



Outline:

- Motivations for $t\bar{t}H(\rightarrow b\bar{b})$ channel
- Reconstruct the $t\bar{t}H$ signal: BDT
- Classify the signal ($t\bar{t}H$) and backgrounds ($t\bar{t}+jets$): BDT
- 2-dimensional BDT: a new way to categorize the events
- Preliminary tests with deep neural networks
- Summary

Classification BDT

- Train the BDT with the MC simulation. Signal: $t\bar{t}H$, Bkg: $t\bar{t}$ jets
- To maximize the size of training sample
  - Cross training
  - The absolute value of the event weight is taken, so that events with negative weights are also used
- Discriminating variables (19 variables):
  - General kinematics variables
  - Variables from reconstruction BDT:
    - BDT output of the best combination (the one with highest BDT score)
    - Kinematics: Higgs mass, angular distance of the reconstructed objects corresponding to the best combination
  - Output of two other multi-variate techniques:
    - Likelihood Discriminant (also developed at CPPM also) and Matrix Element Method

Classification BDT

- Training parameters of classification BDT after optimization on it:

BoostType	AdaBoost
AdaBoostBeta	0.15
NTrees	250
MaxDepth	5
nCuts	80
MinNodeSize	4%

TMVA parameters for the classification BDT.

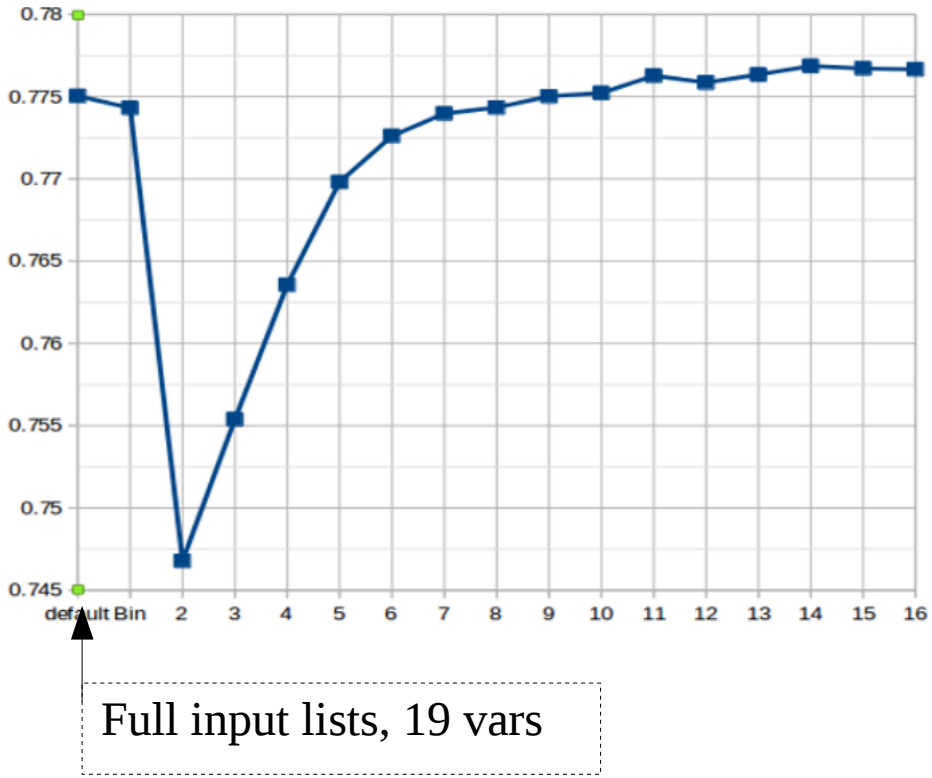
- Many variations on these parameters were tried, but none lead to noticeable improvements.

Classification BDT: variable list optimization

- Goal: to reduce the number of variables used in the classification BDT
- Optimization procedure:
  - Starting from the most discriminant variables: reco BDT + LHD
  - From a list of 29 variables, add each of them one by one
  - The one leading to the best improvement in AUC is retained
  - Then again rescan the rest of 28 variables, and go through the same procedure
  - ...
  - Until the same performance as the original list (19 variables) was achieved

Classification BDT: variable list optimization

- AUC of classification BDT (y-axis) vs. number of input variables (x-axis)



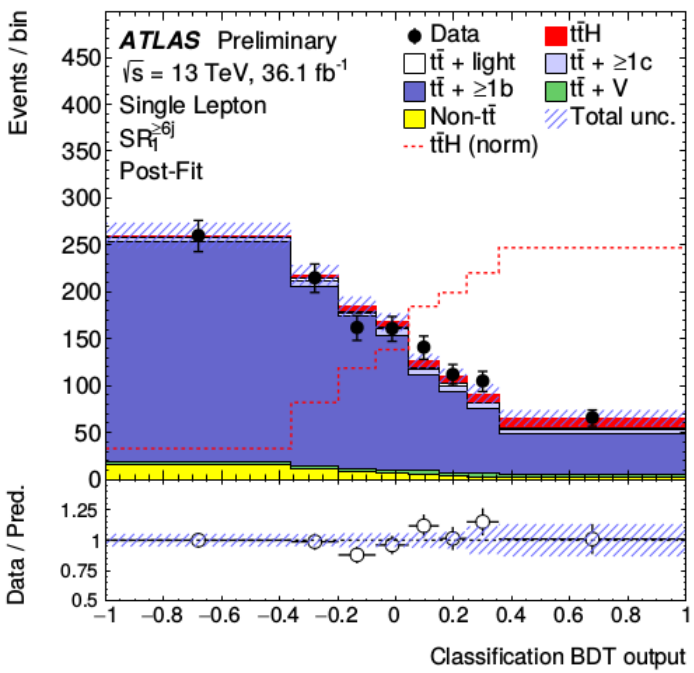
- The optimal list of variables was chosen to be with 7 variables in the very pure signal regions, with which the same performance as with the full list is achieved

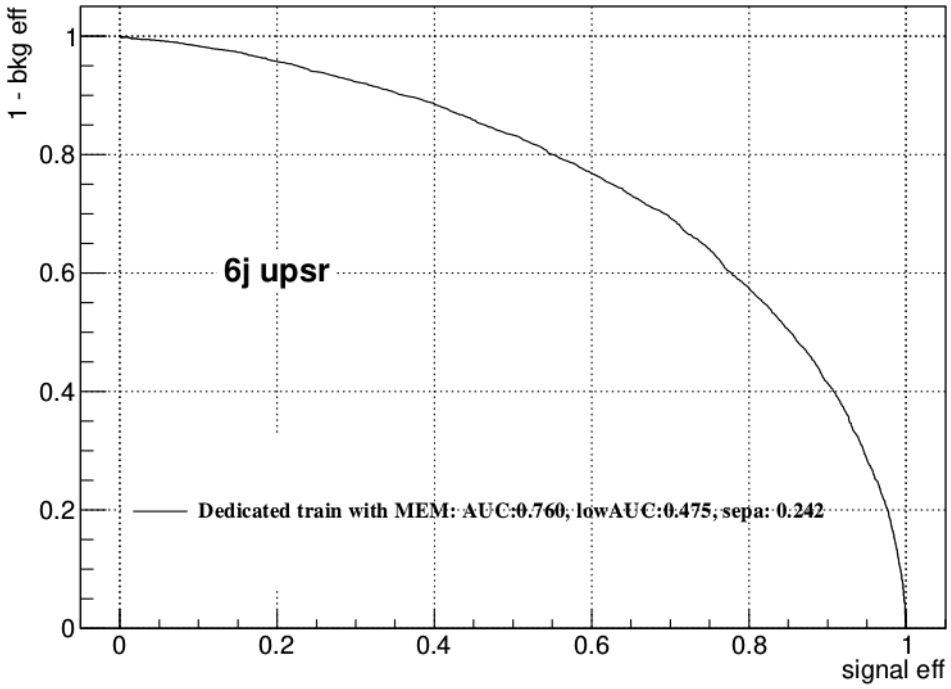
- When the performance was then evaluated with the whole fit model, also including all systematic uncertainties, the reduced lists were found to perform slightly worse than the original full list, so the full list BDT was kept.

Classification BDT

- The BDT output and ROC curves:

ATLAS-CONF-2017-076

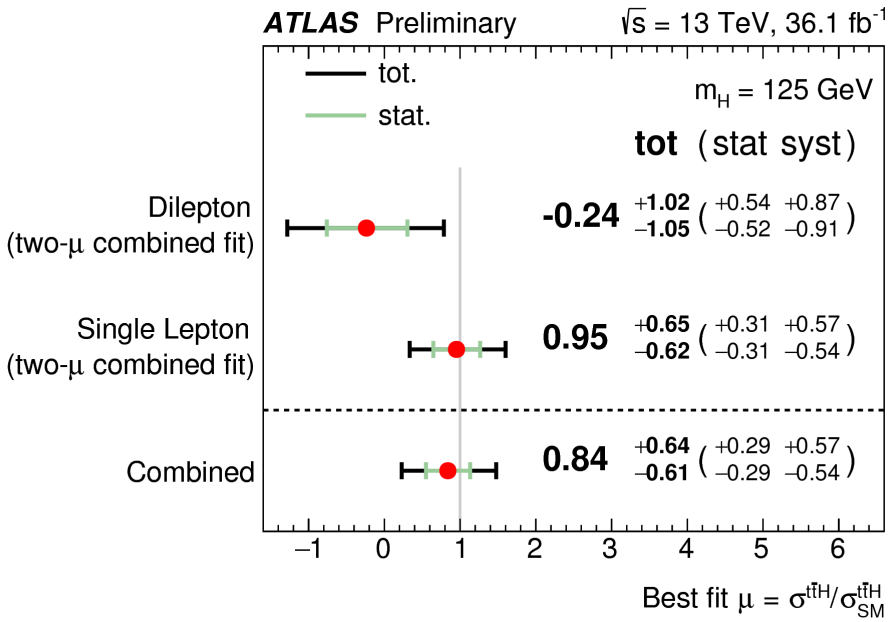


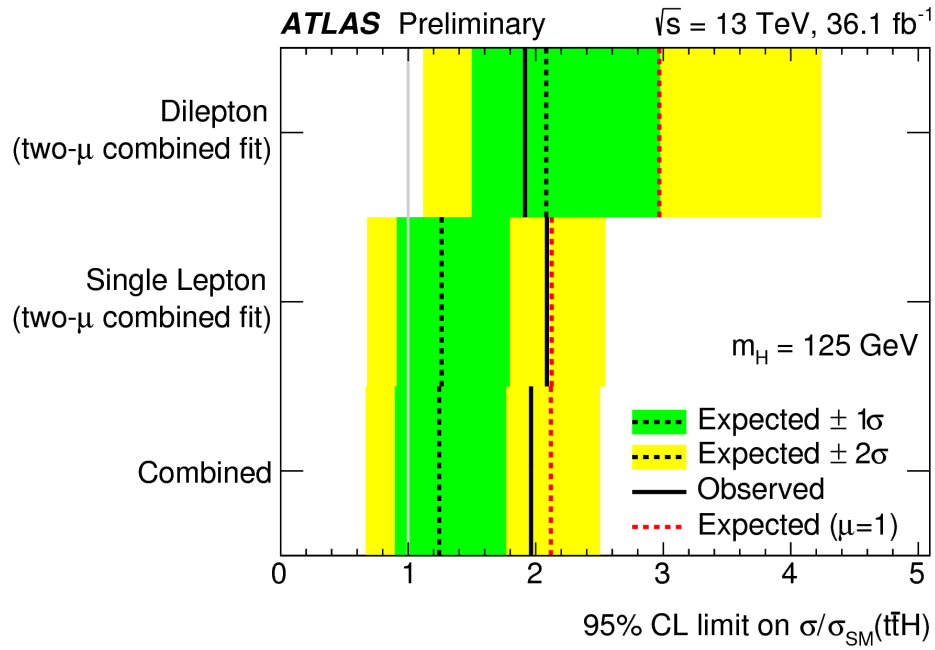


- The distributions of the classification BDTs the signal regions are used as the final discriminants for the profile-likelihood fit.

Analysis result

From ATLAS-CONF-2017-076, with a paper upcoming

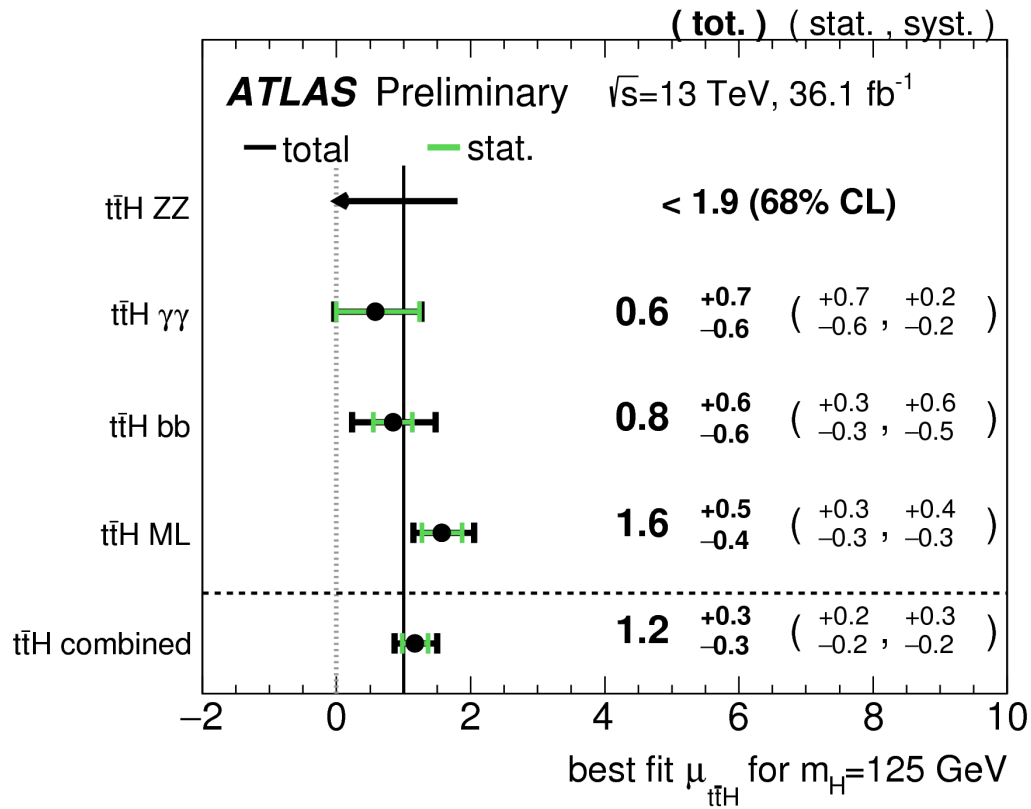




- In $t\bar{t}H(\rightarrow b\bar{b})$ channel:
  - No significant excess for the SM Higgs boson at 125 GeV.
  - The observed limit is 2 times the SM prediction
  - The best-fit signal cross section is $0.84^{+0.64}_{-0.61}$ times SM.

Analysis result

From ATLAS-CONF-2017-077, with a paper upcoming



- Combining $b\bar{b}$, multilepton, γγ, and $ZZ \rightarrow 4l$ channels:
  - The best-fit signal cross section is $1.17 \pm 0.19(\text{stat})^{+0.27}_{-0.23}$ (syst)
  - First evidence for $t\bar{t}H$: significance 4.2 σ

Outline:

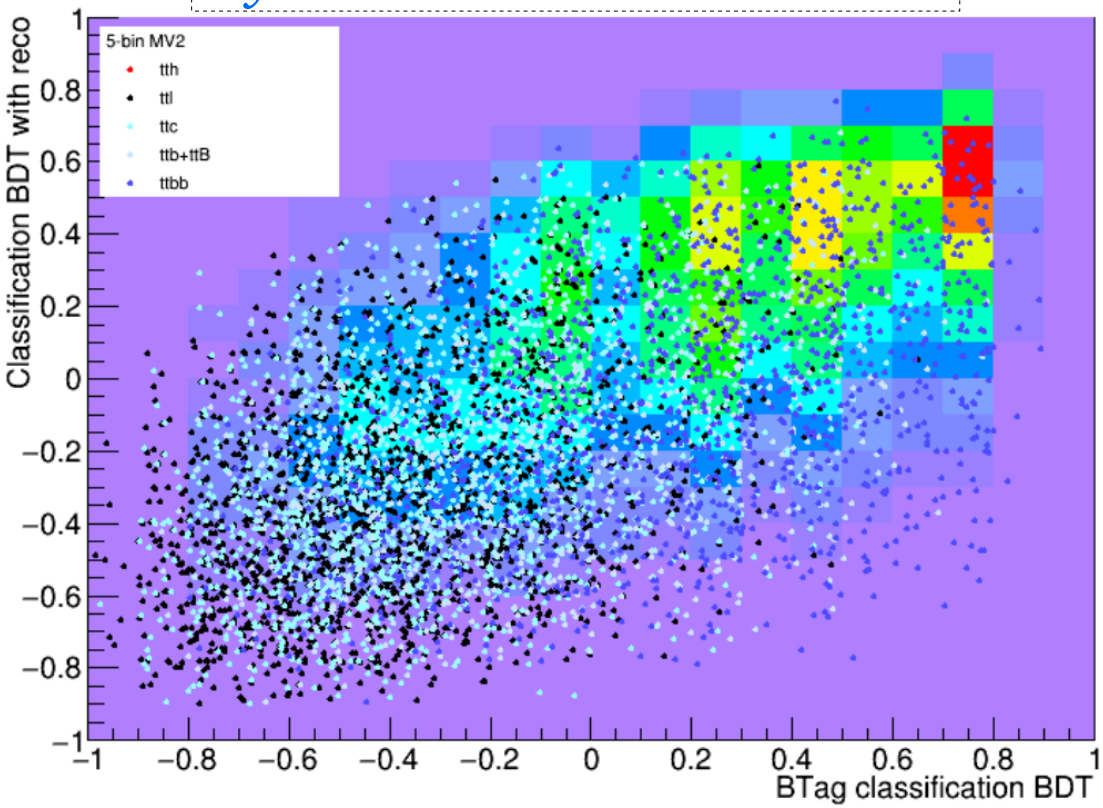
- Motivations for $t\bar{t}H(\rightarrow b\bar{b})$ channel
- Reconstruct the $t\bar{t}H$ signal: BDT
- Classify the signal ($t\bar{t}H$) and backgrounds ($t\bar{t}+jets$): BDT
- 2-dimensional BDT: a new way to categorize the events
- Preliminary tests with deep neural networks
- Summary

2-dimentional BDT: underway studies

- New method: train two orthogonal BDT

- It is thus possible to explore a 2-d binning method to get both sensitive signal-rich regions and bins to control individual backgrounds

y-axis: classification BDT



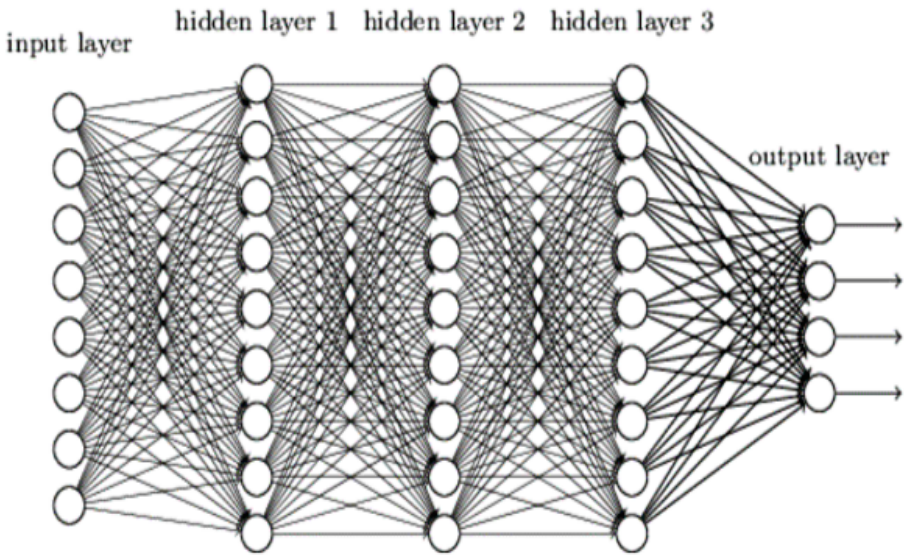
x-axis: Btag BDT using only b-tagging weights as input features

Outline:

- Motivations for $t\bar{t}H(\rightarrow b\bar{b})$ channel
- Reconstruct the $t\bar{t}H$ signal: BDT
- Classify the signal ($t\bar{t}H$) and backgrounds ($t\bar{t}+jets$): BDT
- 2-dimensional BDT: a new way to categorize the events
- Preliminary tests with deep neural networks
- Summary

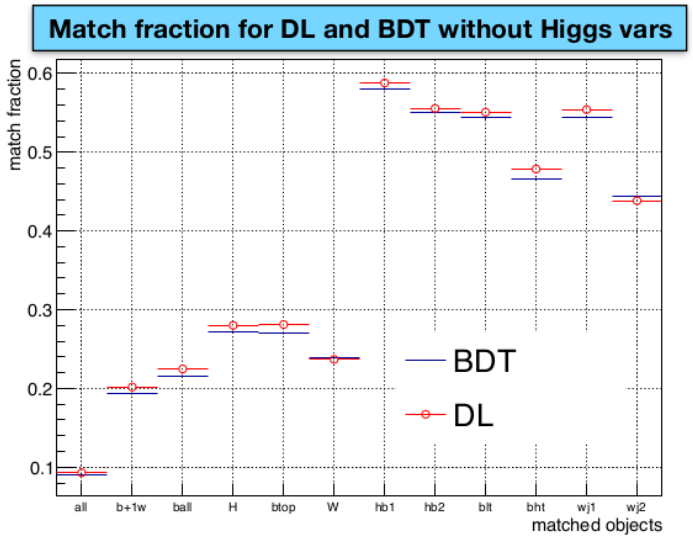
Deep learning

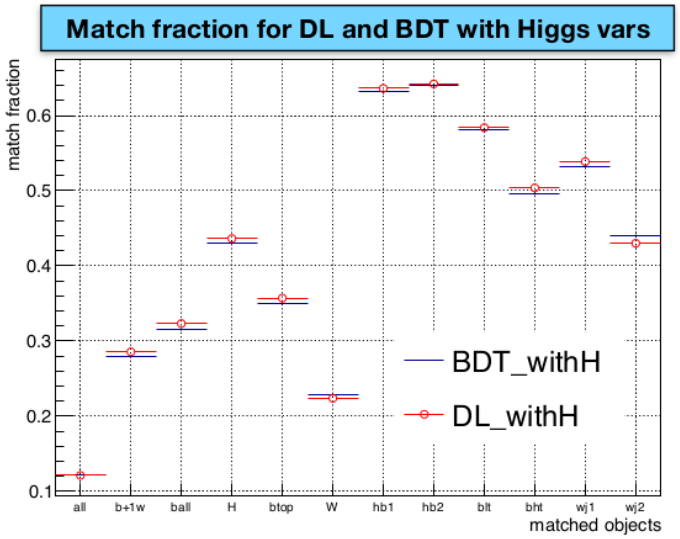
- Deep learning: cutting edge in machine learning
  - Recent breakthroughs in AI (image labeling, speech recognition, natural language processing, AlphaGo)
  - Deep neural network: ≥2 hidden layers
  - Not only do the classification, but also the hierarchical feature extraction:
    - A lot of time was spent on choosing the input features for BDT (high level features).
    - Feature extraction from the low level features (e.g. 4-vectors) automatically?



Deep learning: preliminary results

- Preliminary test with deep neural network:
  - Start with a fair comparison with BDT using exactly the same input features
  - C++ deep learning package, plugged into TMVA
- $t\bar{t}H$ reconstruction:





- **Higgs:**
  - DL: 28.01%
  - BDT: 27.14%

  +3.21%
- **all:**
  - DL: 9.29%
  - BDT: 9.04%

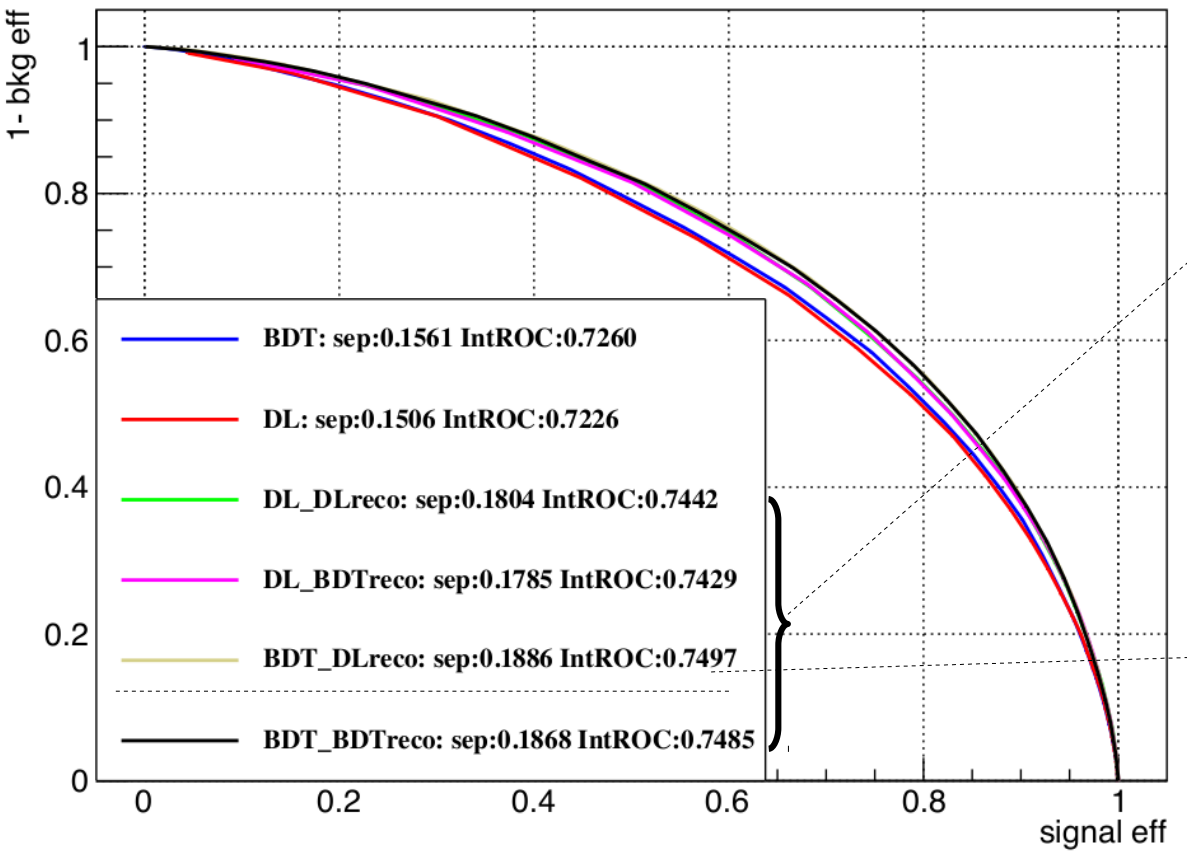
  +2.77%

- **Higgs withH:**
  - DL: 43.63%
  - BDT: 43.00%

  +1.47%
- **all withH:**
  - DL: 12.11%
  - BDT: 12.21%

  -0.82%

Deep learning: preliminary results



- Classification BDT/DL with variables based on reconstruction BDT/DL
  - All the performance are similar
  - BDT classification with DL reconstruction happens to give marginally highest ROC integral and separation

Summary

- BDT is studied for $t\bar{t}H(\rightarrow b\bar{b})$ analysis: (ATLAS-CONF-2017-076, paper upcoming)
  - Reconstruction: associate multiple jets to quarks for $t\bar{t}H$
  - Separate the $t\bar{t}H$ from the main ttbar background
  - First evidence for $t\bar{t}H$: significance 4.2 (ATLAS-CONF-2017-077, paper upcoming)
- Preliminary test with deep neural network:
  - Using the same input features as BDT, the DL already shows comparable performance to BDT.
- 2-dimensional BDT studies underway: a new way to categorize the events

Backup

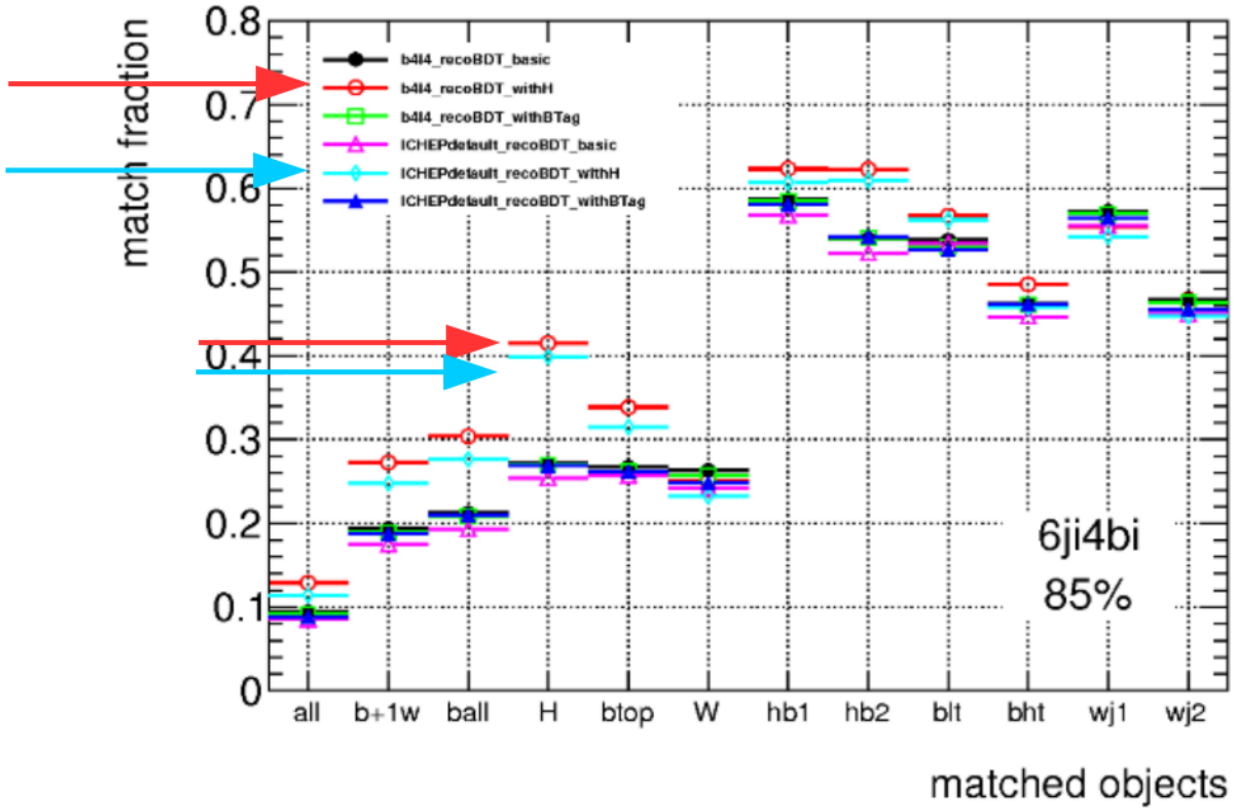
Input list for reconstruction BDT

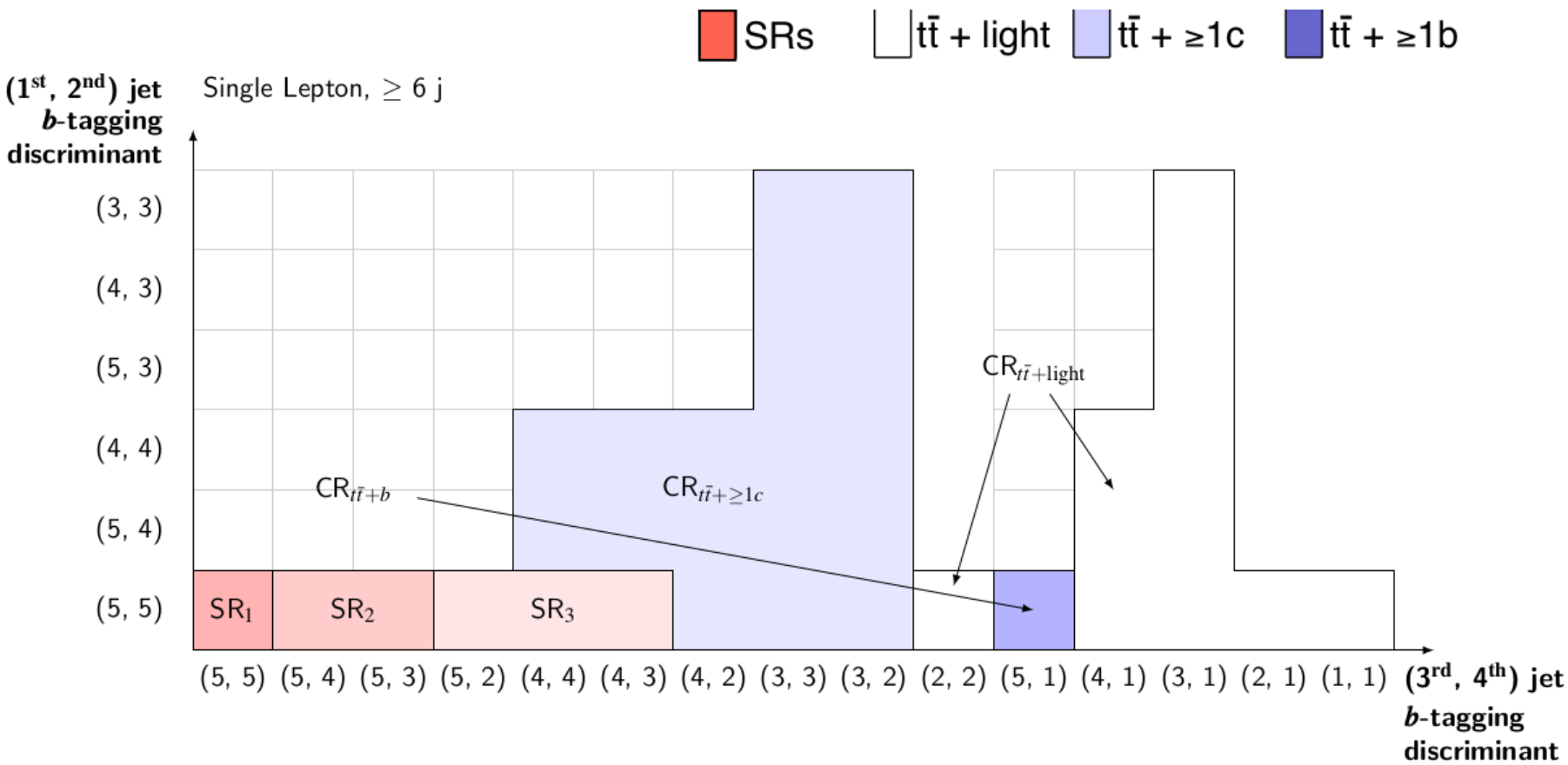
Variable	Definition	≥ 6j		5j	
		UPSR	Incl.	UPSR	Incl.
General kinematic variables					
$\Delta R_{bb}^{\text{avg}}$	Average ΔR for all b-tagged jet pairs	✓	✓	✓	✓
$\Delta R_{bb}^{\text{max } p_{\text{T}}}$	ΔR between the two b-tagged jets with the largest vector sum p_{T}	✓	✓	–	–
$\Delta R_{\text{lep}-bb}^{\text{min } \Delta R}$	Minimum ΔR between the lepton and any pair of b-tagged jets	–	–	✓	✓
$\Delta\eta_{jj}^{\text{max } \Delta\eta}$	Maximum $\Delta\eta$ between any two jets	✓	✓	✓	✓
$m_{bb}^{\text{min } \Delta R}$	Mass of the combination of the two b-tagged jets with the smallest ΔR	✓	✓	✓	✓
$m_{jj}^{\text{min } \Delta R}$	Mass of the combination of any two jets with the smallest ΔR	–	–	✓	✓
$p_{\text{T}}^{\text{jet5}}$	p_{T} of the fifth leading jet (testing full removal)	–	✓	–	–
$\text{N}_{30}^{\text{Higgs}}$	Number of b-jet pairs with invariant mass within 30 GeV of the Higgs boson mass	–	–	✓	✓
Aplanarity	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor [91] built with all jets	–	✓	✓	✓
$\text{Aplanarity}_{b-jet}$	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor built with b-jets	✓	✓	–	–
$H1$	Second Fox–Wolfram moment computed using all jets and the lepton	–	–	✓	✓
Variables from reconstruction BDT output					
BDT	BDT output	✓*	✓*	✓*	✓*
m_H	Higgs boson mass	–	–	✓	✓
$\Delta R_{\text{lep top,had top}}$	ΔR between leptonic and hadronic tops	–	✓*	–	–
$\Delta R_{H,\text{had top}}$	ΔR between Higgs boson and hadronic top	–	✓	–	–
Variable from Likelihood calculation					
D	Likelihood discriminant	✓	✓	✓	✓
Variable from Matrix Element calculation					
MEM_D1	MEM discriminant	✓	–	–	–
Variables from b-tagging					
w_b^H	Sum of binned b-tagging weights of jets from best Higgs candidate	–	✓	–	✓
B_{j^3}	3rd jet binned b-tagging weight (sorted by weight)	–	✓	–	✓
B_{j^4}	4th jet binned b-tagging weight (sorted by weight)	–	✓	–	✓
B_{j^5}	5th jet binned b-tagging weight (sorted by weight)	–	✓	–	✓

Input list for classification BDT

Variable	Definition	$SR^{\geq 6j}_{1,2,3}$	$SR^{5j}_{1,2}$
General kinematic variables			
ΔR^{avg}_{bb}	Average ΔR for all b-tagged jet pairs	✓	✓
$\Delta R^{max\ p_T}_{bb}$	ΔR between the two b-tagged jets with the largest vector sum p_T	✓	–
$\Delta\eta^{max\ \Delta\eta}_{jj}$	Maximum $\Delta\eta$ between any two jets	✓	✓
$m^{min\ \Delta R}_{bb}$	Mass of the combination of two b-tagged jets with the smallest ΔR	✓	–
$m^{min\ \Delta R}_{jj}$	Mass of the combination of any two jets with the smallest ΔR	–	✓
N^{Higgs}_{30}	Number of b-jet pairs with invariant mass within 30 GeV of the Higgs boson mass	✓	✓
H^{had}_T	Scalar sum of jet p_T	–	✓
$\Delta R^{min\ \Delta R}_{\ell,bb}$	ΔR between the lepton and the combination of the two b-tagged jets with the smallest ΔR	–	✓
Aplanarity	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor [97] built with all jets	✓	✓
$H1$	Second Fox–Wolfram moment computed using all jets and the lepton	✓	✓
Variables from reconstruction BDT			
BDT output	Output of the reconstruction BDT	✓*	✓*
m^{Higgs}_{bb}	Higgs candidate mass	✓	✓
$m_{H,b_{lep\ top}}$	Mass of Higgs candidate and b-jet from leptonic top candidate	✓	–
ΔR^{Higgs}_{bb}	ΔR between b-jets from the Higgs candidate	✓	✓
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs candidate and $t\bar{t}$ candidate system	✓*	✓*
$\Delta R_{H,lep\ top}$	ΔR between Higgs candidate and leptonic top candidate	✓	–
$\Delta R_{H,b_{had\ top}}$	ΔR between Higgs candidate and b-jet from hadronic top candidate	–	✓*
Variables from Likelihood and Matrix Element Method calculations			
LHD	Likelihood discriminant	✓	✓
MEM_{D1}	Matrix Element discriminant	✓	–
Variables from b-tagging			
$w^{Higgs}_{b\text{-tag}}$	Sum of b-tagging discriminants of jets from best Higgs candidate from the reconstruction BDT	✓	✓
B^3_{jet}	3rd largest jet b-tagging discriminant	✓	✓
B^4_{jet}	4th largest jet b-tagging discriminant	✓	✓
B^5_{jet}	5th largest jet b-tagging discriminant	✓	✓

The improvement of reco BDT by using b-tagging info.



Event categorization

	none	*loose*	*medium*	*tight*	*very-tight*
Efficiency	-	85%	77%	70%	60%
Discriminant value	1	2	3	4	5