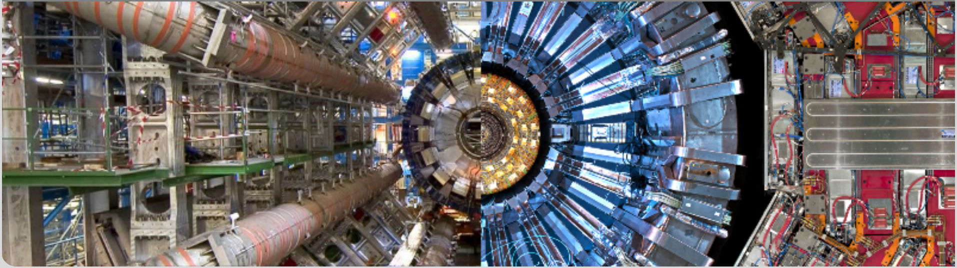
Karlsruher Institut für Technologie

Machine learning at ATLAS and CMS

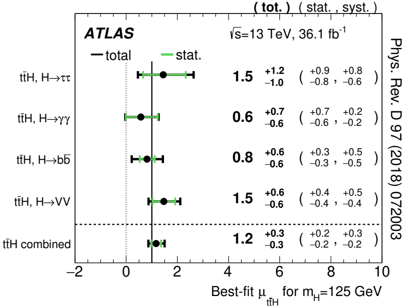
Higgs Toppings Workshop - Probing Top-Higgs Interactions at the LHC (Benasque)

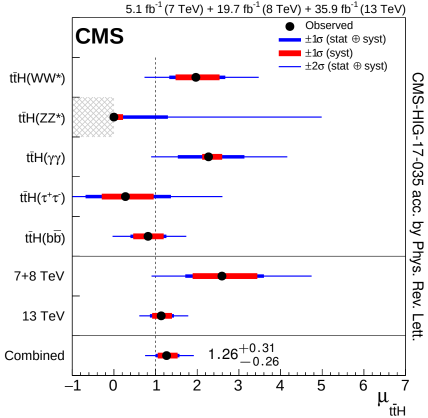
Georges Aad (CPPM, Aix-Marseille Université, CNRS/IN2P3, Marseille, France) and
Matthias Schröder (KIT) | May 31, 2018

INSTITUT FÜR EXPERIMENTELLE TEILCHENPHYSIK (ETP)



$t\bar{t}H$ Status



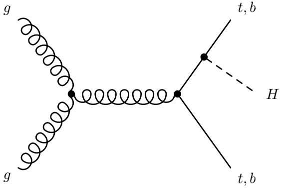


ATLAS Run II (36.1 fb^{-1} @ 13 TeV):
4.2 σ evidence (3.8 expected)

CMS Run I+II (60.7 fb^{-1} @ 7, 8, 13 TeV):
5.2 σ observation (4.2 σ expected)

$t\bar{t}$H Challenges

- **Small $t\bar{t}$H production cross-section** of $\approx 0.5\,$pb at 13 TeV
- Combination of $t\bar{t}$ and H decays: **multitude of possible final states with many objects**
  - Jets and b jets
  - Light leptons and hadronic τs
  - Photons
- Complimentary challenges
  - H $\rightarrow \gamma\gamma$, ZZ: high purity, **tiny rate**
  - H $\rightarrow$ multileptons: intermediate, **small rate, difficult experimental and $t\bar{t}$ + V backgrounds**
  - H $\rightarrow b\bar{b}$: high rates, **difficult $t\bar{t}$ + jets background**



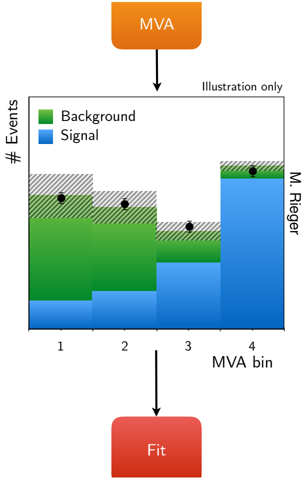
Need **dedicated machine-learning techniques** at various levels

Outline

- Outline
  - Machine learning at the LHC and in $t\bar{t}H$
  - Strategy: categorisation, reconstruction, classification
  - Binary and multi-classification
- Outline and comparison of strategies, no listing of every analysis detail
- Highlighting items for discussion

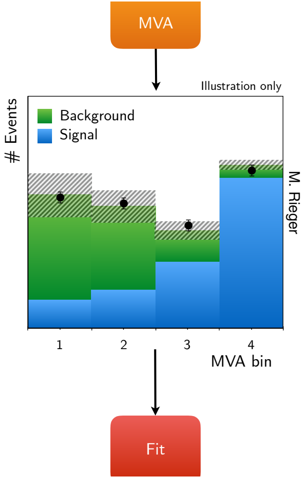
Machine Learning (ML) at the LHC

- Improved sensitivity by *multivariate analysis* (MVA)
  - Combination of various input variables into single output
- Combination with **supervised-learning** = **machine-learning** (ML) techniques
- **Boosted Decision Tree** (BDT)
  - Established ML technique
  - Robust workhorse for binary classification
- Deep **Neural Network** (NN)
  - Rather new ML technique at the LHC
  - Artificial neural network with several hidden layers for multi-classification
- Also: likelihood (LH) technique
  - *Not an ML technique* but important MVA
  - Physics-motivated likelihood ratio discriminant
  - **Matrix-Element-Method** (MEM) and **reco-based LH**



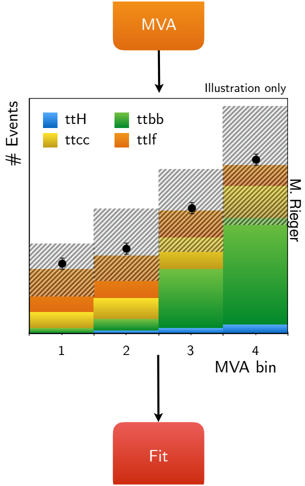
Machine Learning (ML) in $t\bar{t}H$

- ML techniques exploited at **various levels**
- Foremost: final **classification**
  - Classify event as **signal or background**
  - Typically: final fit of ML classifier output distributions
- Sensitivity depends on **ability to constrain background uncertainties**
- Common strategy: **multi-step classification**
  - Categories enriched in signal and different background processes (=uncertainty)
  - Advanced signal vs background separation



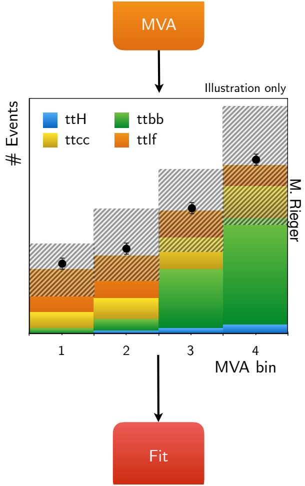
Machine Learning (ML) in $t\bar{t}H$

- ML techniques exploited at **various levels**
- Foremost: final **classification**
  - Classify event as **signal or background**
  - Typically: final fit of ML classifier output distributions
- Sensitivity depends on **ability to constrain background uncertainties**
- Common strategy: **multi-step classification**
  - Categories enriched in signal and different background processes (=uncertainty)
  - Advanced signal vs background separation



Machine Learning (ML) in $t\bar{t}H$

- ML techniques exploited at **various levels**
- Foremost: final **classification**
  - Classify event as **signal or background**
  - Typically: final fit of ML classifier output distributions
- Other important applications
  - **Trigger** level
  - Object **identification**
  - Event **reconstruction**
  - Event **categorisation**

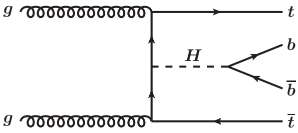


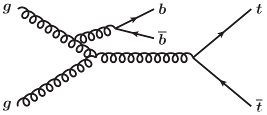
$\mathrm{t\bar{t}H(b\bar{b})}$ Production

- Main example here for ML techniques: search for $\mathrm{t\bar{t}H(b\bar{b})}$ production

  (ATLAS: Phys. Rev. D 97 (2018) 072016, CMS: CMS-HIG-17-026 subm. to JHEP)
  - Applies similarly to other $\mathrm{t\bar{t}H}$ analyses

- Challenging final state
  - **Huge combinatorics** in event reconstruction
  - Large background: $\mathrm{t\bar{t}}$ + jets
  - In particular: **irreducible $\mathrm{t\bar{t}}$ + $\mathrm{b\bar{b}}$ background** (5–10 × signal) with associated large theory uncertainties

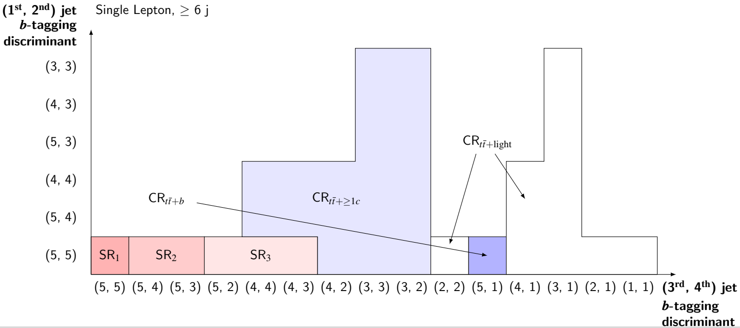




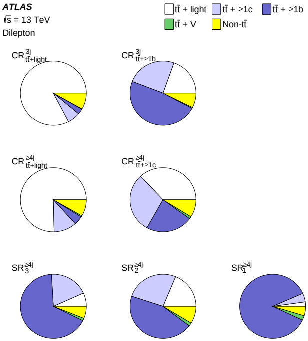
Categorisation

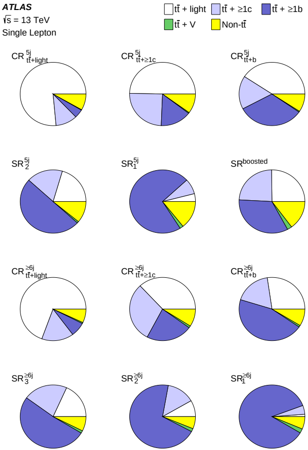
- Goal: **separate signal from $t\bar{t}$ + LF, $t\bar{t} + \geq 1$b, and $t\bar{t} + \geq 1$c** bkgs.
- Common strategy (ATLAS, CMS dilepton channel): **categorisation by jet and b-tagging information**
  - Example from ATLAS

	not tagged	loose	medium	tight	very tight
b jet ID efficiency	—	85 %	77 %	70 %	60 %
discriminant index	1	2	3	4	5



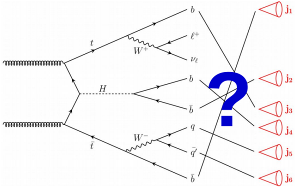
Categorisation



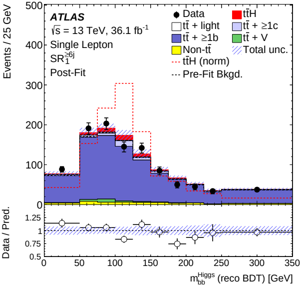


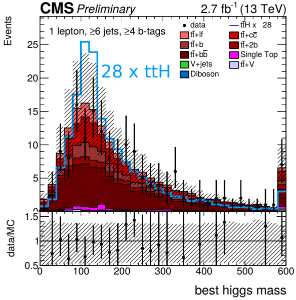
Signal purity $< 6\,\%$: need further separation of signal from background

Event Reconstruction



- **Reconstruct top and Higgs candidates** from final-state objects
  - Additional separating variables for final classification
- **Different techniques** to find best combination
- ATLAS: **BDT**
  - Up to 50 % Higgs reconstruction efficiency (if using m_H)
- CMS: χ^2 **based**
  - $\approx$ 30 % Higgs reconstruction efficiency



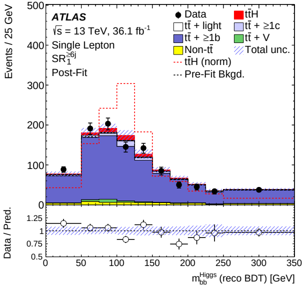


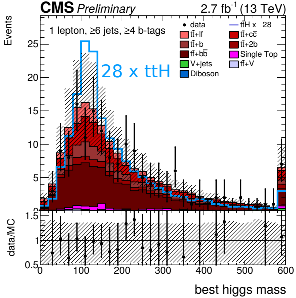
Final Event Classification

- But **Higgs mass: sensitivity not sufficient**
  - Jet energy resolution
  - Combinatorics in jet assignment
- **Several discriminating variables** separating signal from background
  - b-tagging information
  - Jet and lepton kinematics
  - Angular and event-shape variables
  - Invariant masses
  - MEM

ML-based MVA classifiers combining information from several variables

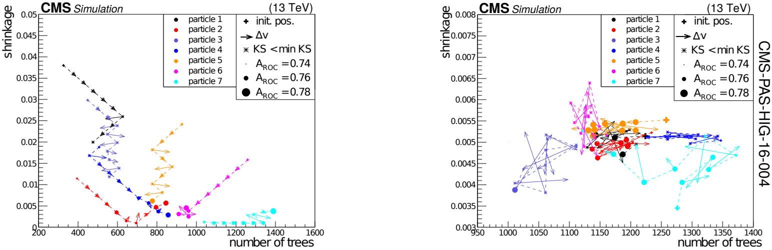
- Possible further improvements?
  - c-tagging information
  - Jet charge





Choice of Configuration and Input Variables

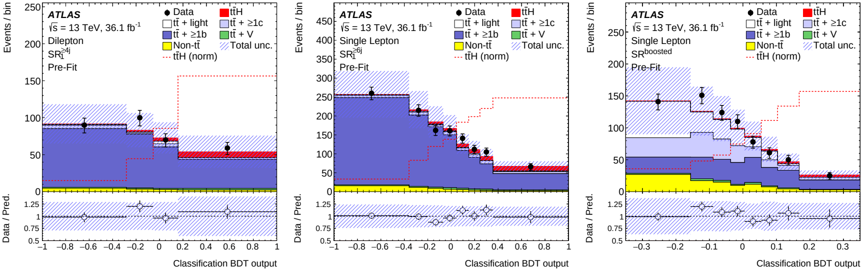
- BDT/NN performance depends on configuration and input variables
- Finding optimal **choice high-dimensional problem**: solution “by-hand” or with algorithm
- Example: solution based on **Particle Swarm Optimisation**[1]
  1. Swarm of candidate BDTs, each initialised at random configuration
  2. Random choice of input variables: train and test performance
  3. BDTs move to new positions in configuration space, depending on their own and the swarms best previous positions



[1] Kennedy and Eberhart, doi:10.1109/ICNN.1995.488968

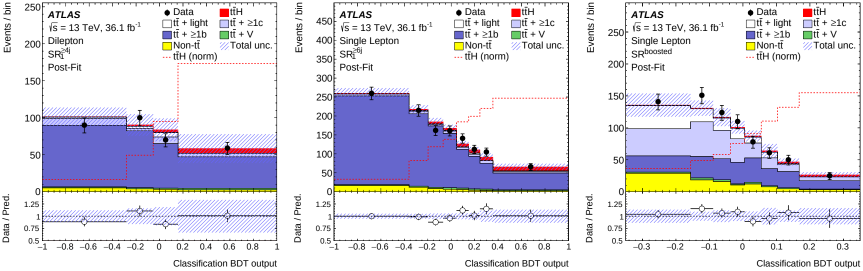
Binary Classification

- Example: **BDT per category to separate signal from background** (ATLAS, CMS dilepton channel)
  - Combination of kinematic variables and output of event reconstruction
  - b-tagging information
    - CMS: continuous b-tagging output improves classification
    - ATLAS: little gain because already used (almost) differentially in categorisation
- Different ML techniques with similar results (after sufficient training)



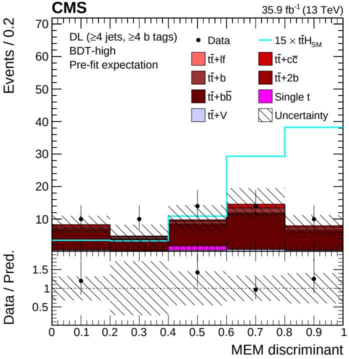
Binary Classification

- Example: **BDT per category to separate signal from background** (ATLAS, CMS dilepton channel)
  - Combination of kinematic variables and output of event reconstruction
  - b-tagging information
    - CMS: continuous b-tagging output improves classification
    - ATLAS: little gain because already used (almost) differentially in categorisation
- Different ML techniques with similar results (after sufficient training)



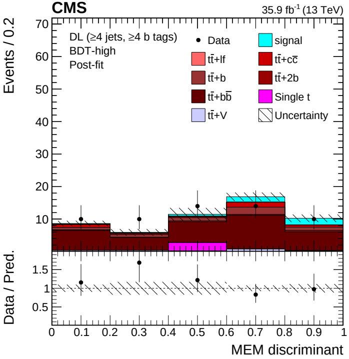
Binary Classification

- Example: **MEM to separate signal from background** (CMS dilepton channel)
  - Pre-classification by BDT
  - MEM in signal-enriched regions **targeting $t\bar{t}H$ vs. $t\bar{t} + b\bar{b}$**
  - Alternatively, MEM as input variable (ATLAS and CMS, single-lepton channel)
- MEM by construction very **powerful against $t\bar{t} + b\bar{b}$**
  - Yields up to 10 % improvement in sensitivity
- Relies on LO calculations and per-jet transfer functions associating reconstructed objects and final-state partons
- CPU intensive



Binary Classification

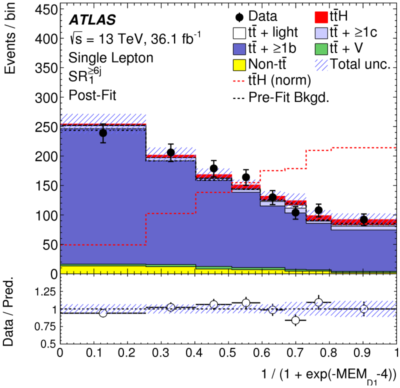
- Example: **MEM to separate signal from background** (CMS dilepton channel)
  - Pre-classification by BDT
  - MEM in signal-enriched regions **targeting $t\bar{t}$H vs. $t\bar{t} + b\bar{b}$**
  - Alternatively, MEM as input variable (ATLAS and CMS, single-lepton channel)
- MEM by construction very **powerful against $t\bar{t} + b\bar{b}$**
  - Yields up to 10 % improvement in sensitivity
- Relies on LO calculations and per-jet transfer functions associating reconstructed objects and final-state partons
- CPU intensive



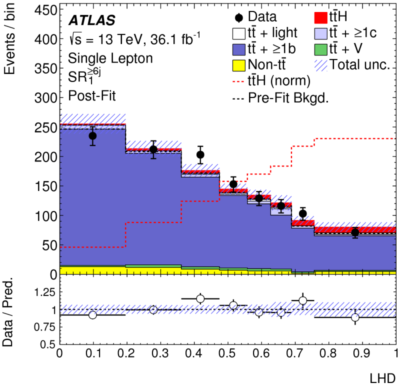
Binary Classification

- Example: **Reco-based LH to separate signal from background** (ATLAS)
  - Less CPU intensive
  - Avoid assumptions of LO and per-jet transfer functions
- Same performance achieved as with MEM

MEM



reco-based LH

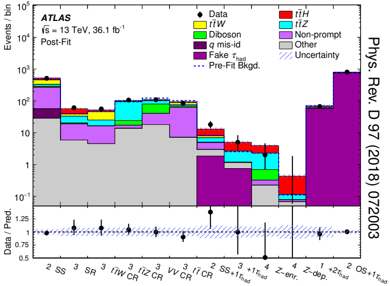


From Binary to Multi-Classification

- Example tt̄H(ML) in ATLAS: different strategies depending on channel

	2ℓSS	3ℓ	4ℓ	$1\ell+2\tau_{\text{had}}$	2ℓSS$+1\tau_{\text{had}}$	2ℓOS$+1\tau_{\text{had}}$	$3\ell+1\tau_{\text{had}}$
Light lepton	2T*	1L*, 2T*	2L, 2T	1T	2T*	2L†	1L†, 2T
τ_{had}	0M	0M	–	1T, 1M	1M	1M	1M
N_{jets}, $N_{b\text{-jets}}$	$\geq 4, = 1,2$	$\geq 2, \geq 1$	$\geq 2, \geq 1$	$\geq 3, \geq 1$	$\geq 4, \geq 1$	$\geq 3, \geq 1$	$\geq 2, \geq 1$

- 2ℓ: **2 BDTs** separating tt̄H from tt̄ $+$ V and non-prompt lepton backgrounds
  - 2D information **mapped into 1D ordered by significance**
- 3ℓ: **multi-class BDT** categorising events as tt̄H or any of 4 main backgrounds
  - Multi-dimensional binning producing 5 regions dominated by the 5 categories[2]
  - Fit signal BDT in tt̄H category
  - 1 bin in background categories

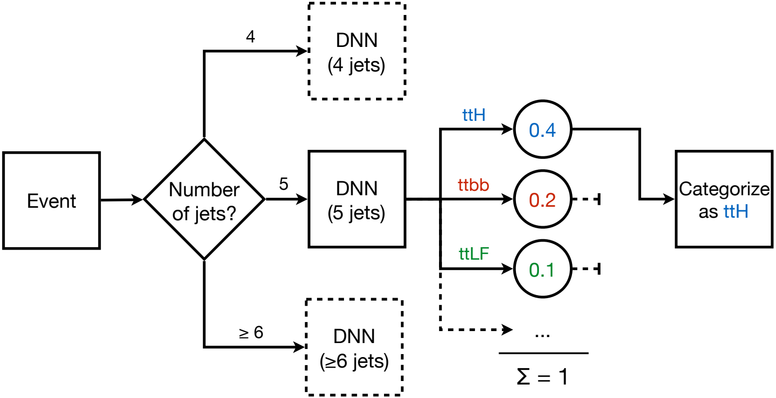


Phys. Rev. D 97 (2018) 072003

[2] "foam" clustering algorithm arxiv:0812.0922

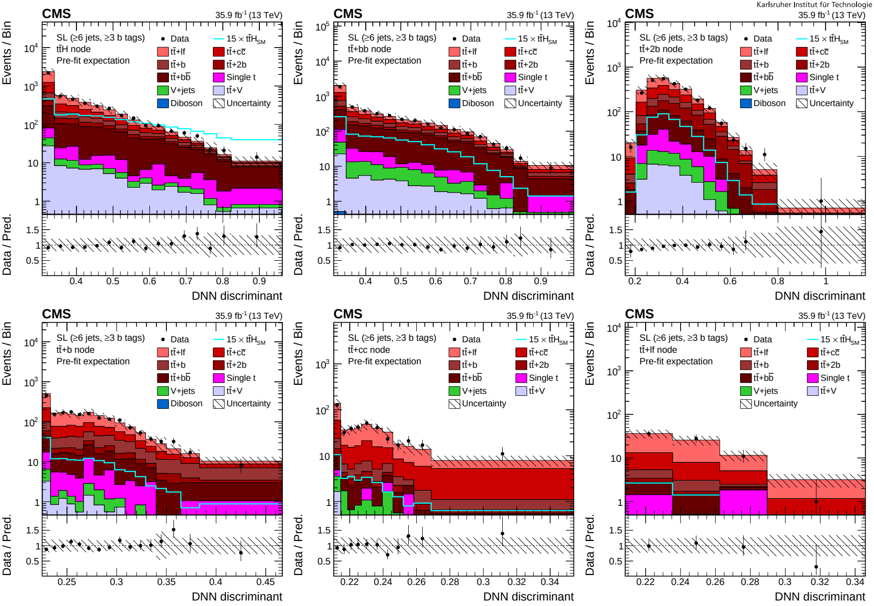
Multi-Classification

- Multi-classification with NN to define categories (CMS single-lepton channel)



Categorisation & classification by same ML classifier

Example: $\geq$ 6 Jets



Binary vs. Multi-Classification

- CMS example: binary classification (with N_{jets}, $N_{\text{b-tags}}$ categorisation) vs. multi-classification: **better precision with multi-classification**
  - Analysis of simulated data using same uncertainty modelling

Channel	Method	Best-fit μ $\pm$tot ($\pm$stat $\pm$ syst)
Single-lepton	BDT+MEM	$1.0^{+0.69}_{-0.66}\left({}^{+0.31}_{-0.30}\;{}^{+0.62}_{-0.59}\right)$
Single-lepton	DNN	$1.0^{+0.58}_{-0.55}\left({}^{+0.30}_{-0.29}\;{}^{+0.50}_{-0.47}\right)$
Dilepton	BDT+MEM	$1.0^{+1.22}_{-1.12}\left({}^{+0.65}_{-0.62}\;{}^{+1.04}_{-0.93}\right)$
Dilepton	DNN	$1.0^{+1.38}_{-1.36}\left({}^{+0.71}_{-0.69}\;{}^{+1.18}_{-1.18}\right)$
Combined	BDT+MEM	$1.0^{+0.60}_{-0.57}\left({}^{+0.28}_{-0.27}\;{}^{+0.53}_{-0.51}\right)$
Combined	DNN	$1.0^{+0.55}_{-0.51}\left({}^{+0.27}_{-0.27}\;{}^{+0.47}_{-0.44}\right)$

- Highly non-trivial to compare performance in general
- Can probably **obtain same sensitivity with almost all techniques** if using sufficient input variables, configuration, and training

Sensitivity Comparison

- Example: S and S/B in most sensitive regions in single-lepton channel

  ATLAS ($SR_1^{\geq 6j}$ + $SR_2^{\geq 6j}$): $S = 143$, $S/B = 3.9\,\%$

  CMS ($\geq$ 6 jets, $\geq$ 3 b-tags, $t\bar{t}H$ node): $S = 142$, $S/B = 2.8\,\%$
  - Same signal efficiency with b-tagging only categorisation
  - ATLAS smaller inclusive background, CMS smaller $t\bar{t} + \geq 1b$ background

ATLAS $\geq$ 6 jets

Sample	$SR_3^{\geq 6j}$ Pre-fit	$SR_3^{\geq 6j}$ Post-fit	$SR_2^{\geq 6j}$ Pre-fit	$SR_2^{\geq 6j}$ Post-fit	$SR_1^{\geq 6j}$ Pre-fit	$SR_1^{\geq 6j}$ Post-fit
$t\bar{t}H$	85 ± 10	71 ± 52	81 ± 10	68 ± 50	62 ± 11	51 ± 38
$t\bar{t}$ + light	750 ± 370	586 ± 98	210 ± 210	96 ± 33	14 ± 10	12.1 ± 5.8
$t\bar{t} + \geq 1c$	880 ± 350	1330 ± 190	350 ± 100	473 ± 99	53 ± 33	44 ± 20
$t\bar{t} + \geq 1b$	2100 ± 420	2290 ± 170	1750 ± 370	1850 ± 130	1010 ± 240	1032 ± 59
$t\bar{t} + V$	51.2 ± 7.4	50.8 ± 5.9	40.8 ± 5.7	40.3 ± 4.8	25.8 ± 3.7	25.3 ± 3.2
Non-$t\bar{t}$	303 ± 82	267 ± 63	155 ± 52	134 ± 46	75 ± 20	58 ± 17
Total	4140 ± 850	4590 ± 110	2550 ± 510	2657 ± 82	1220 ± 250	1223 ± 42
Data	4698		2641		1222	

CMS $\geq$ 6 jets

pre-fit (post-fit) yields

Process	$t\bar{t}H$ node	$t\bar{t}+b\bar{b}$ node
$t\bar{t}$+lf	1982 (1381)	1280 (897)
$t\bar{t}+c\bar{c}$	1150 (1415)	998 (1230)
$t\bar{t}$+b	549 (705)	575 (746)
$t\bar{t}$+2b	306 (233)	282 (215)
$t\bar{t}+b\bar{b}$	834 (769)	1156 (1082)
Single t	110 (116)	146 (145)
V + jets	38 (37)	78 (76)
$t\bar{t}$+V	80 (75)	58 (54)
Diboson	0.9 (0.9)	0.5 (0.5)
Total bkg.	5049 (4733)	4575 (4447)
± tot unc.	±1216 (±186)	±1156 (±142)
$t\bar{t}$H	142 (108)	53 (40)
± tot unc.	±19 (±15)	±8 (±6)

- Full result: stat $\oplus$ syst uncertainty on μ

  ATLAS: 0.29 (35 %) 0.56 (66 %)

  CMS: 0.24 (33 %) 0.38 (53 %)

Non-trivial interplay with final classifier

Systematic uncertainties and control regions play most important role

MC Sample Size

- ML classifiers require **independent data for training** (and validation) and analysis
  - Signal and background class(es) enter training with same weight
  - Increasingly challenging with increasing number of categories/classes
- Mitigation e. g. by cross validation
  - Also, in some cases, train more inclusively than per category (ATLAS)
- Example for $t\bar{t}$ in single-lepton channel: POWHEG+PYTHIA8

ATLAS	single-lepton decays	60 Mio
	$t\bar{t} + \geq$ 1b filtered (6 % efficiency)	10 Mio
CMS	single-lepton decays	120 Mio

- "MC stats" 30–50 % of total uncertainty
  - Negative weights add to challenge

MC **sample size critical factor** in analysis,
can limit application of ML techniques

Input-Variable Modelling

- Good data-MC agreement required for every single input variable
  - BDT/NN exploit deep correlations
- Careful **validation of variables and correlations crucial**
- Difficult to find independent control sample but single variables not very sensitive: can look into signal regions
- *Discussed on Tuesday*

ML techniques require well-modelled variables and correlations

Summary & Conclusions

- **Machine-learning** based MVAs **mandatory in many $t\bar{t}H$ channels**
  - Complex final states (“no clear Higgs mass peak”) and low signal purity
- Different background processes with different (large) uncertainties: **single binary classification not sufficient**
- Common general strategy: **multi-step classification**
  1. Categories enriched in signal and different backgrounds
  2. Separation of signal and backgrounds (allows constraining uncertainties)
  - Can be combined (to some extent) using multi-classification ML techniques
- Main challenge (at least in $H(b\bar{b})$ channel): **control of uncertainties**
  - Improve ML techniques for purer control regions (probably incremental)
  - Explore new ML techniques to **reduce dependence on uncertainties** (adversarial training, ...): explicitly construct classifiers to be insensitive to effects with large uncertainties

Additional Material

Input Variables ATLAS (1/3)

Variable	Definition	$SR_1^{\geq 4j}$	$SR_2^{\geq 4j}$	$SR_3^{\geq 4j}$
General kinematic variables				
m_{bb}^{min}	Minimum invariant mass of a b-tagged jet pair	✓	✓	-
m_{bb}^{max}	Maximum invariant mass of a b-tagged jet pair	-	-	✓
$m_{bb}^{\text{min}\ \Delta R}$	Invariant mass of the b-tagged jet pair with minimum ΔR	✓	-	✓
$m_{\text{jj}}^{\text{max}\ p_{\text{T}}}$	Invariant mass of the jet pair with maximum p_{T}	✓	-	-
$m_{bb}^{\text{max}\ p_{\text{T}}}$	Invariant mass of the b-tagged jet pair with maximum p_{T}	✓	-	✓
$\Delta\eta_{bb}^{\text{avg}}$	Average $\Delta\eta$ for all b-tagged jet pairs	✓	✓	✓
$\Delta\eta_{\ell,\text{j}}^{\text{max}}$	Maximum $\Delta\eta$ between a jet and a lepton	-	✓	✓
$\Delta R_{bb}^{\text{max}\ p_{\text{T}}}$	ΔR between the b-tagged jet pair with maximum p_{T}	-	✓	✓
$N_{bb}^{\text{Higgs 30}}$	Number of b-tagged jet pairs with invariant mass within 30 GeV of the Higgs-boson mass	✓	✓	-
$n_{\text{jets}}^{p_{\text{T}}>40}$	Number of jets with $p_{\text{T}} > 40$ GeV	-	✓	✓
$\text{Aplanarity}_{b\text{-jet}}$	$1.5\lambda_2$, where λ_2 is the second eigenvalue of the momentum tensor [100] built with all b-tagged jets	-	✓	-
$H_{\text{T}}^{\text{all}}$	Scalar sum of p_{T} of all jets and leptons	-	-	✓
Variables from reconstruction BDT				
BDT output	Output of the reconstruction BDT	✓**	✓**	✓
m_{bb}^{Higgs}	Higgs candidate mass	✓	-	✓
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs candidate and $t\bar{t}$ candidate system	✓*	-	-
$\Delta R_{H,\ell}^{\text{min}}$	Minimum ΔR between Higgs candidate and lepton	✓	✓	✓
$\Delta R_{H,b}^{\text{min}}$	Minimum ΔR between Higgs candidate and b-jet from top	✓	✓	-
$\Delta R_{H,b}^{\text{max}}$	Maximum ΔR between Higgs candidate and b-jet from top	-	✓	-
$\Delta R_{bb}^{\text{Higgs}}$	ΔR between the two jets matched to the Higgs candidate	-	✓	-
Variables from b-tagging				
$w_{b\text{-tag}}^{\text{Higgs}}$	Sum of b-tagging discriminants of jets from best Higgs candidate from the reconstruction BDT	-	✓	-

Input Variables ATLAS (2/3)

Karlsruher Institut für Technologie

Variable	Definition	$\mathrm{SR}^{\geq 6j}_{1,2,3}$	$\mathrm{SR}^{5j}_{1,2}$
General kinematic variables			
$\Delta R^{\mathrm{avg}}_{bb}$	Average ΔR for all b-tagged jet pairs	✓	✓
$\Delta R^{\max p_{\mathrm{T}}}_{bb}$	ΔR between the two b-tagged jets with the largest vector sum p_{T}	✓	–
$\Delta\eta^{\max}_{\mathrm{jj}}$	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	✓	–
$m^{\min \Delta R}_{\mathrm{jj}}$	Mass of the combination of any two jets with the smallest ΔR	–	✓
$N^{\mathrm{Higgs}\ 30}_{bb}$	Number of b-tagged jet pairs with invariant mass within 30 GeV of the Higgs-boson mass	✓	✓
$H^{\mathrm{had}}_{\mathrm{T}}$	Scalar sum of jet p_{T}	–	✓
$\Delta R^{\min}_{\ell,bb}$	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 [100] built with all jets	✓	✓
H_1	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	✓	✓
$m_{H,b_{\mathrm{lep\ top}}}$	Mass of Higgs candidate and b-jet from leptonic top candidate	✓	–
$\Delta R^{\mathrm{Higgs}}_{bb}$	ΔR between b-jets from the Higgs candidate	✓	✓
$\Delta R_{H,t\bar{t}}$	ΔR between Higgs candidate and $t\bar{t}$ candidate system	✓*	✓*
$\Delta R_{H,\mathrm{lep\ top}}$	ΔR between Higgs candidate and leptonic top candidate	✓	–
$\Delta R_{H,b_{\mathrm{had\ top}}}$	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 (in $\mathrm{SR}^{\geq 6j}_{1}$ only)	✓	–
Variables from b-tagging (not in $\mathrm{SR}^{\geq 6j}_{1}$)			
$w^{\mathrm{Higgs}}_{b\text{-tag}}$	Sum of b-tagging discriminants of jets from best Higgs candidate from the reconstruction BDT	✓	✓
B^{3}_{jet}	3^{rd} largest jet b-tagging discriminant	✓	✓
B^{4}_{jet}	4^{th} largest jet b-tagging discriminant	✓	✓
B^{5}_{jet}	5^{th} largest jet b-tagging discriminant	✓	✓

Input Variables ATLAS (3/3)

Variable	Definition
Variables from jet reclustering	
$\Delta R_{H,t}$	ΔR between the Higgs-boson and top-quark candidates
$\Delta R_{t,b^{\text{add}}}$	ΔR between the top-quark candidate and additional b-jet
$\Delta R_{H,b^{\text{add}}}$	ΔR between the Higgs-boson candidate and additional b-jet
$\Delta R_{H,\ell}$	ΔR between the Higgs-boson candidate and lepton
$m_{\text{Higgs candidate}}$	Higgs-boson candidate mass
$\sqrt{d_{12}}$	Top-quark candidate first splitting scale [101]
Variables from b-tagging	
$w_{b\text{-tag}}$	Sum of b-tagging discriminants of all b-jets
$w^{\text{add}}_{b\text{-tag}}/w_{b\text{-tag}}$	Ratio of sum of b-tagging discriminants of additional b-jets to all b-jets

Input Variables CMS (1/3)

Variable	Definition	SL (4 jets, ≥ 3 b tags)	SL (5 jets, ≥ 3 b tags)	SL (≥ 6 jets, ≥ 3 b tags)	DL (≥ 4 jets, 3 b tags)	DL (≥ 4 jets, ≥ 4 b tags)
$p_{\mathrm{T}}(\text{jet }1)$	p_{T} of the highest-p_{T} jet	+	+	-	-	-
$\eta(\text{jet }1)$	η of the highest-p_{T} jet	-	+	+	-	-
$d(\text{jet }1)$	b tagging discriminant of the highest-p_{T} jet	+	+	+	-	-
$p_{\mathrm{T}}(\text{jet }2)$	p_{T} of the second highest-p_{T} jet	-	+	-	-	-
$\eta(\text{jet }2)$	η of the second highest-p_{T} jet	+	+	+	-	-
$d(\text{jet }2)$	b tagging discriminant of the second highest-p_{T} jet	+	+	+	-	-
$p_{\mathrm{T}}(\text{jet }3)$	p_{T} of the third highest-p_{T} jet	-	+	-	-	-
$\eta(\text{jet }3)$	η of the third highest-p_{T} jet	+	+	+	-	-
$d(\text{jet }3)$	b tagging discriminant of the third highest-p_{T} jet	+	+	+	-	-
$p_{\mathrm{T}}(\text{jet }4)$	p_{T} of the fourth highest-p_{T} jet	+	+	-	-	-
$\eta(\text{jet }4)$	η of the fourth highest-p_{T} jet	+	+	+	-	-
$d(\text{jet }4)$	b tagging discriminant of the fourth highest-p_{T} jet	+	-	+	-	-
$p_{\mathrm{T}}(\text{lep }1)$	p_{T} of the highest-p_{T} lepton	-	+	+	-	-
$\eta(\text{lep }1)$	η of the highest-p_{T} lepton	+	-	+	-	-
$d_{\mathrm{j}}^{\mathrm{avg}}$	average b tagging discriminant value of all jets	+	+	+	-	-
$d_{\mathrm{b}}^{\mathrm{avg}}$	average b tagging discriminant value of b-tagged jets	+	+	+	+	+
$d_{\text{non-b}}^{\mathrm{avg}}$	average b tagging discriminant value of non-b-tagged jets	-	-	-	+	+
$\sum_b \left(d - d_{\mathrm{b}}^{\mathrm{avg}}\right)$	squared difference between the b tagging discriminant value of a b-tagged jet and the average b tagging discriminant values of all b-tagged jets, summed over all b-tagged jets	+	+	+	-	-
$d_{\mathrm{j}}^{\max}$	maximal b tagging discriminant value of all jets	+	+	+	-	-
$d_{\mathrm{b}}^{\max}$	maximal b tagging discriminant value of b-tagged jets	+	+	+	-	-
$d_{\mathrm{j}}^{\min}$	minimal b tagging discriminant value of all jets	+	+	+	-	-
$d_{\mathrm{j}}^{\min}$	minimal b tagging discriminant value of b-tagged jets	+	+	+	-	-
d_2	second highest b tagging discriminant value of all jets	+	+	+	-	-

Input Variables CMS (2/3)

Variable	Definition	SL (4 jets, ≥ 3 b tags)	SL (5 jets, ≥ 3 b tags)	SL (≥ 6 jets, ≥ 3 b tags)	DL (≥ 4 jets, 3 b tags)	DL (≥ 4 jets, ≥ 4 b tags)
N_{b}(tight)	number of b-tagged jets at a working point with a 0.1% probability of tagging gluon and light-flavour jets	+	+	+	-	-
BLR	likelihood ratio discriminating between 4 b quark jets and 2 b quark jets events	+	+	+	-	-
$\mathrm{BLR}^{\mathrm{trans}}$	transformed BLR defined as $\ln[\mathrm{BLR}/(1.0-\mathrm{BLR})]$	+	+	+	-	-
$\Delta R_{\mathrm{j,j}}^{\min}$	ΔR between the two closest jets	+	+	+	-	-
$\Delta R_{\mathrm{b,b}}^{\min}$	ΔR between the two closest b-tagged jets	+	+	+	-	-
$\Delta R_{\mathrm{j,j}}^{\max}$	ΔR between the two jets furthest apart	-	+	-	-	-
$\Delta R_{\mathrm{b,b}}^{\max}$	ΔR between the two b-tagged jets furthest apart	-	-	+	-	-
$\Delta\eta_{\mathrm{j,j}}^{\max}$	$\Delta\eta$ between the two jets furthest apart in η	-	-	-	-	+
$\Delta\eta_{\mathrm{b,b}}^{\max}$	$\Delta\eta$ between the two b-tagged jets furthest apart in η	-	-	-	+	+
$\Delta\eta_{\mathrm{b,b}}^{\mathrm{avg}}$	average $\Delta\eta$ between b-tagged jets	-	-	+	-	-
$\Delta R_{\mathrm{b,b}}^{\mathrm{avg}}$	average ΔR between b-tagged jets	-	+	+	-	-
$\Delta R_{\mathrm{j,b}}^{\mathrm{avg}}$	average ΔR between jets of which at least one is b-tagged	-	-	-	+	-
$\Delta R_{\mathrm{lep,j}}^{\min\Delta R}$	ΔR between lepton and closest jet	+	+	-	-	-
$\Delta R_{\mathrm{lep,b}}^{\min\Delta R}$	ΔR between lepton and closest b-tagged jet	-	+	+	-	-
$m_{\mathrm{lep,b}}^{\min\Delta R}$	mass of lepton and closest b-tagged jet	+	+	+	-	-
$m_{\mathrm{b,b}}^{\min\Delta R}$	mass of closest b-tagged jets	+	+	+	-	+
$m_{\mathrm{j,b}}^{\min\Delta R}$	mass of closest jets of which at least one is b-tagged	-	-	-	+	-
$m_{\mathrm{b,b}}^{\mathrm{max\ mass}}$	maximal mass of pairs of b-tagged jets	-	-	-	+	+
$p_{\mathrm{T}}{}_{\mathrm{b,b}}^{\min\Delta R}$	combined p_{T} of closest b-tagged jets	-	-	-	+	-
$p_{\mathrm{T}}{}_{\mathrm{j,b}}^{\min\Delta R}$	combined p_{T} of closest jets of which at least one is b-tagged	-	-	-	-	+
$m_{\mathrm{j}}^{\mathrm{avg}}$	average mass of all jets	+	+	+	-	-
$(m^2)_{\mathrm{b}}^{\mathrm{avg}}$	average squared mass of all b-tagged jets	+	-	+	-	-
$m_{\mathrm{b,b}}^{\mathrm{closest\ to\ 125}}$	mass of pair of b-tagged jets closest to 125 GeV	-	+	+	-	-
$N^{\mathrm{j,j}}$	number of pairs of b-tagged jets with an invariant mass within 15 GeV of 125 GeV	-	-	-	+	+
MEM	matrix element method discriminant	+	+	+	-	-

Input Variables CMS (3/3)

Variable	Definition	SL (4jets, $\geq$ 3b tags)	SL (5jets, $\geq$ 3b tags)	SL ($\geq$ 6jets, $\geq$ 3b tags)	DL ($\geq$ 4jets, 3b tags)	DL ($\geq$ 4jets, $\geq$ 4b tags)
H_T^j	scalar sum of jet p_T	-	+	-	+	-
H_T^b	scalar sum of b-tagged jet p_T	+	+	+	-	-
A^j	$\frac{3}{2}\lambda_3$ where λ_i are the eigenvalues of the momentum tensor built with jets [?]	-	+	+	-	-
A^b	$\frac{3}{2}\lambda_3$ where λ_i are the eigenvalues of the momentum tensor built with b-tagged jets [?]	+	+	+	-	-
C^j	H_T^j divided by the sum of the energies of all jets	-	-	+	-	-
C^b	H_T^b divided by the sum of the energies of all b-tagged jets	-	-	+	-	+
S^j	$\frac{3}{2}(\lambda_2 + \lambda_3)$ where λ_i are the eigenvalues of the momentum tensor built with jets [?]	+	+	+	-	-
S^b	$\frac{3}{2}(\lambda_2 + \lambda_3)$ where λ_i are the eigenvalues of the momentum tensor built with b-tagged jets [?]	-	+	+	-	-
S_T^j	$\frac{2\lambda_2}{\lambda_2+\lambda_1}$ where λ_i are the eigenvalues of the momentum tensor built with jets [?]	+	+	+	-	-
S_T^b	$\frac{2\lambda_2}{\lambda_2+\lambda_1}$ where λ_i are the eigenvalues of the momentum tensor built with b-tagged jets [?]	+	+	+	-	-
I^b	a measure of how spherical or linear in $r - \phi$ space b-tagged jets are in the event	-	-	-	+	-
H_2	second Fox–Wolfram moment [?]	-	+	-	-	-
H_3	third Fox–Wolfram moment [?]	+	+	-	-	-
H_3^b	third Fox–Wolfram moment calculated with b-tagged jets [?]	-	-	-	-	+
R_3	ratio of Fox–Wolfram moments H_3/H_0 [?]	-	-	-	+	-
H_4	fourth Fox–Wolfram moment [?]	+	-	+	-	-