

# Multivariate Analysis using a Neural Network to improve sensitivity in the ATLAS $t\bar{t}H$ ( $b\bar{b}$ )

Simon Kirchhof

DESY Hamburg – ATLAS group  
University of Heidelberg

September 10, 2015



# Outline

## ① Introduction to $t\bar{t}H$

# Outline

- ① Introduction to  $t\bar{t}H$
- ② Event Categorisation

# Outline

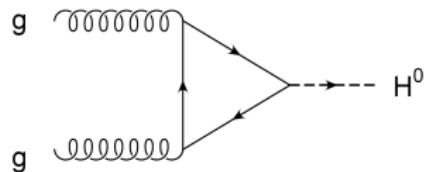
- ① Introduction to  $t\bar{t}H$
- ② Event Categorisation
- ③ Multivariate Analysis

# Outline

- ① Introduction to  $t\bar{t}H$
- ② Event Categorisation
- ③ Multivariate Analysis
- ④ Results

# Motivation for $t\bar{t}H$

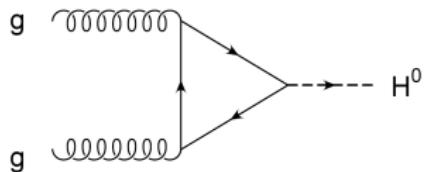
Higgs processes confirmed @LHC:



- 
- |     |   |
|-----|---|
| $H$ | $\longrightarrow \gamma\gamma$                  |
| $H$ | $\longrightarrow ZZ^{(*)} \longrightarrow 4l$   |
| $H$ | $\longrightarrow WW^{(*)} \longrightarrow llvv$ |
| $H$ | $\longrightarrow \tau\tau$                      |
-

# Motivation for $t\bar{t}H$

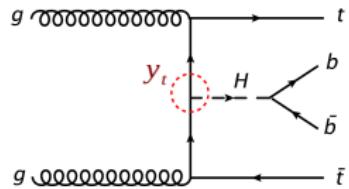
Higgs processes confirmed @LHC:



$$\begin{aligned} H &\rightarrow \gamma\gamma \\ H &\rightarrow ZZ^{(*)} \rightarrow 4l \\ H &\rightarrow WW^{(*)} \rightarrow llvv \\ H &\rightarrow \tau\tau \end{aligned}$$

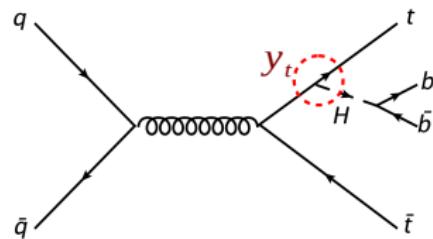
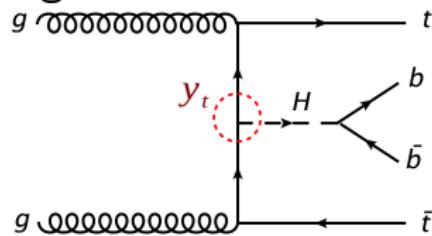
## Interesting: Yukawa coupling $q\bar{q}H$

- Not directly observed so far
- $t\bar{t}H$  ( $b\bar{b}$ ) is most probable:
  - $Y_t \propto m_t$
  - BR ( $H \rightarrow b\bar{b}$ ) high
- associate top production "proofs" coupling

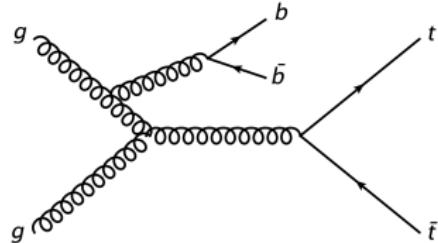


# Signal & Background

## Signal:



## Dominant background:



## Other backgrounds:

- $t\bar{t}V$
- non- $t\bar{t}$

# Outline

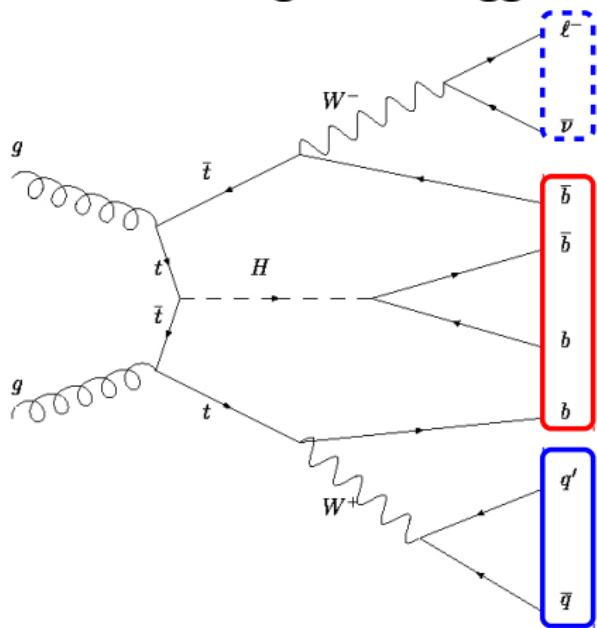
## ② Event Categorisation

Pre-Selection

Categorisation

# Basic Event Selection

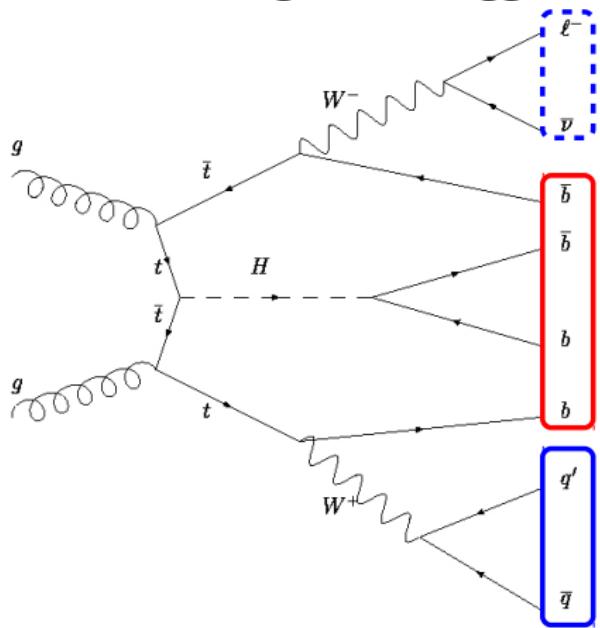
Considerations: **good to trigger & signal-like**



- focus on **single-lepton** events:
  - $e/\mu$  with  $p_T > 25$  GeV

# Basic Event Selection

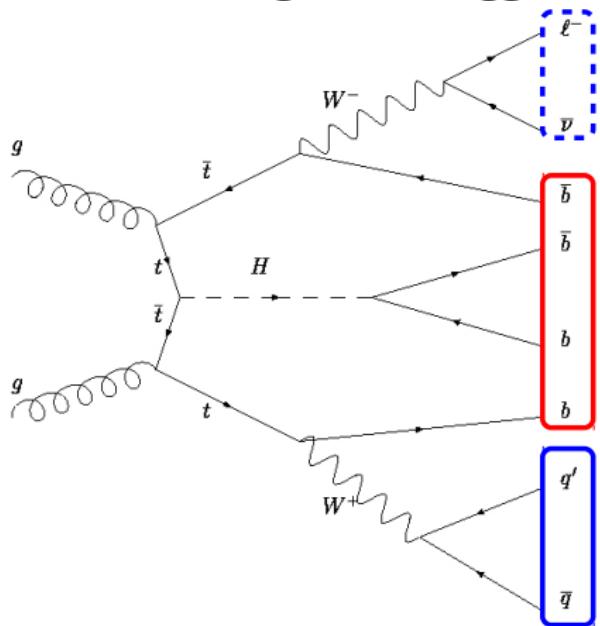
Considerations: **good to trigger & signal-like**



- focus on **single-lepton** events:
  - $e/\mu$  with  $p_T > 25$  GeV
- at least 4 *reconstructed jets*
- at least 2 *reconstructed b-jets*

# Basic Event Selection

Considerations: **good to trigger & signal-like**



- focus on **single-lepton** events:
    - $e/\mu$  with  $p_T > 25$  GeV
  - at least 4 *reconstructed jets*
  - at least 2 *reconstructed b-jets*
- 
- classify events according to...
    - number of **jets**
    - number of **b-jets**

# Categorisation to Analysis Regions

Table: Nine analysis regions.

	2 b-jets	3 b-jets	$\geq 4$ b-jets
4 jets			
5 jets			rich
$\geq 6$ jets		rich	rich

"Signal rich":

- $\frac{S}{B} > 1\%$
- $\frac{S}{\sqrt{B}} > 0.3$

Remark:

- technical definition of b-jet is ambiguous

# Outline

## ③ Multivariate Analysis

Discriminating Variables

Motivation for Multivariate Analysis

Neural Network

# Discriminating Variables

Find differences between **signal** and **background**, possible discriminators:

## Event-shape variables

- Centrality
- Aplanarity
- ...

# Discriminating Variables

Find differences between **signal** and **background**, possible discriminators:

## Event-shape variables

- Centrality
- Aplanarity
- ...

## Global event variables

- scalar sum of the jets  $p_T s$
- ...

# Discriminating Variables

Find differences between **signal** and **background**, possible discriminators:

## Event-shape variables

- Centrality
- Aplanarity
- ...

## Global event variables

- scalar sum of the jets  $p_T s$
- ...

## Object pair properties

- $\Delta R$  of jet pair with ...
  - ... largest  $\sum p_T$
  - ... largest invariant mass
- ...

# Discriminating Variables

Find differences between **signal** and **background**, possible discriminators:

## Event-shape variables

- Centrality
- Aplanarity
- ...

## Global event variables

- scalar sum of the jets  $p_T s$
- ...

## Object pair properties

- $\Delta R$  of jet pair with ...
  - ... largest  $\sum p_T$
  - ... largest invariant mass
- ...

There are  $\sim 20$  in the analysis. I'll only show two.

# High transverse momentum jets (1)

## Definition

$$N_{40}^{\text{jet}} := N_{\text{jet}}(|p_T| > 40 \text{ GeV})$$

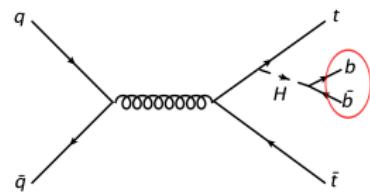
# High transverse momentum jets (1)

## Definition

$$N_{40}^{\text{jet}} := N_{\text{jet}}(|p_T| > 40 \text{ GeV})$$

## Signal:

- Two b-jets from **higgs**  $\implies M_{b\bar{b}} = M_H$
- $|p_T|$  likely to be high



# High transverse momentum jets (1)

## Definition

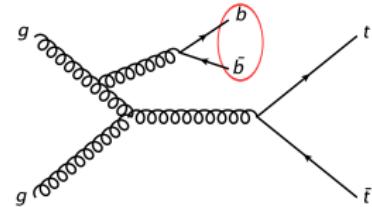
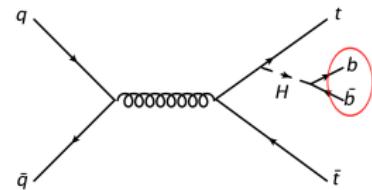
$$N_{40}^{\text{jet}} := N_{\text{jet}}(|p_T| > 40 \text{ GeV})$$

## Signal:

- Two b-jets from **higgs**  $\implies M_{b\bar{b}} = M_H$
- $|p_T|$  likely to be high

## Background:

- Two b-jets from **gluon**
- gluon radiation tends to be "soft"
- $|p_T|$  likely to be lower



# High transverse momentum jets (1)

## Definition

$$N_{40}^{\text{jet}} := N_{\text{jet}}(|p_T| > 40 \text{ GeV})$$

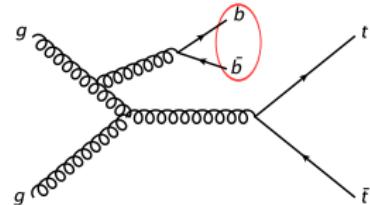
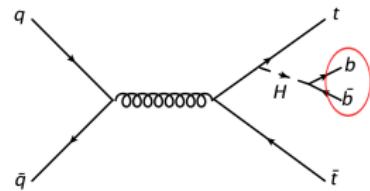
## Signal:

- Two b-jets from **higgs**  $\implies M_{b\bar{b}} = M_H$
- $|p_T|$  likely to be high

## Background:

- Two b-jets from **gluon**
- gluon radiation tends to be "soft"
- $|p_T|$  likely to be lower

$\implies$  Signal has on average more high transverse momentum jets!



# High transverse momentum jets (1)

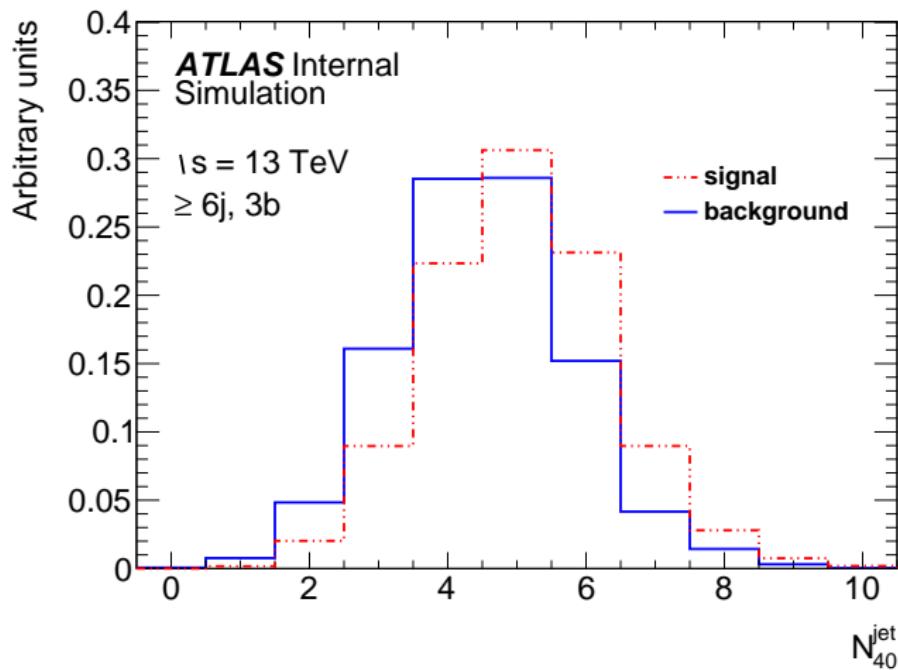


Figure:  $N_{40}^{\text{jet}}$  in the  $\geq 6$  jets 3 tight b-jets region.

# Motivation for Multivariate Analysis

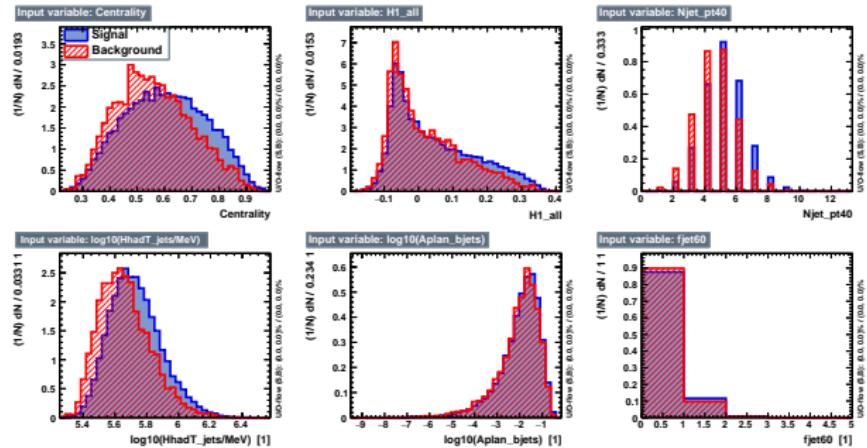


Figure: Some of the "better" variables

# Motivation for Multivariate Analysis

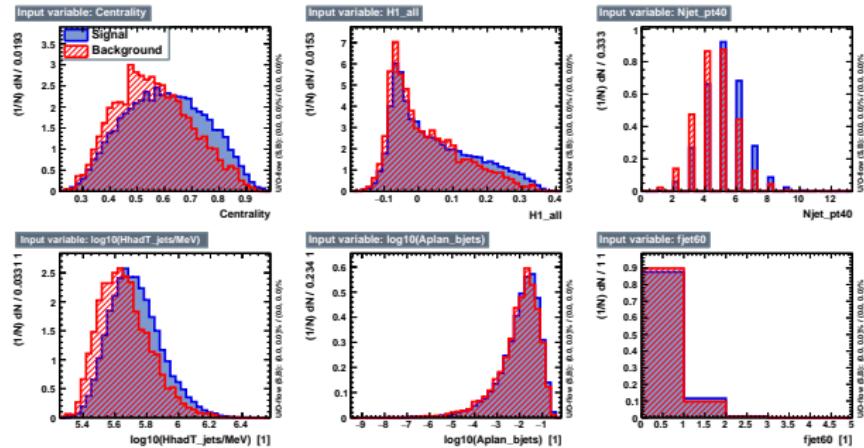


Figure: Some of the "better" variables



Figure: Me when I look at them.

# Motivation for Multivariate Analysis

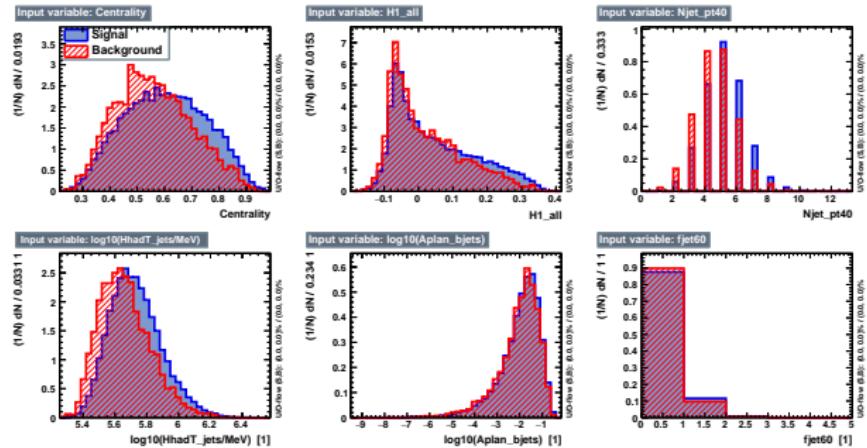


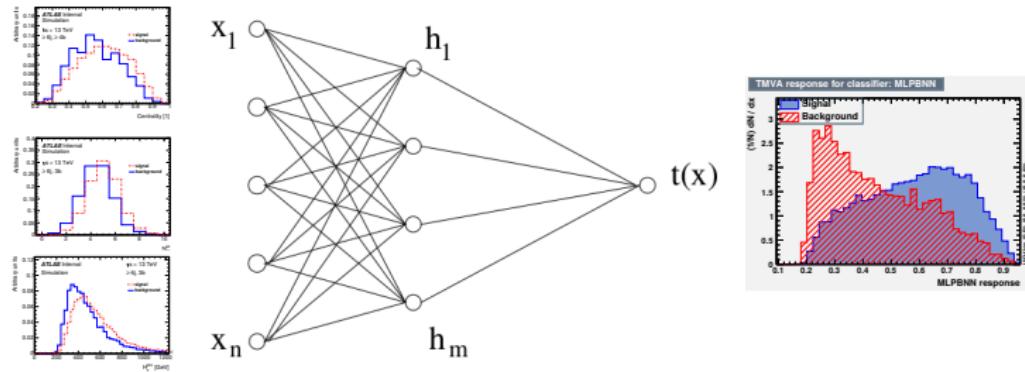
Figure: Some of the "better" variables



Figure: Me when I look at them.

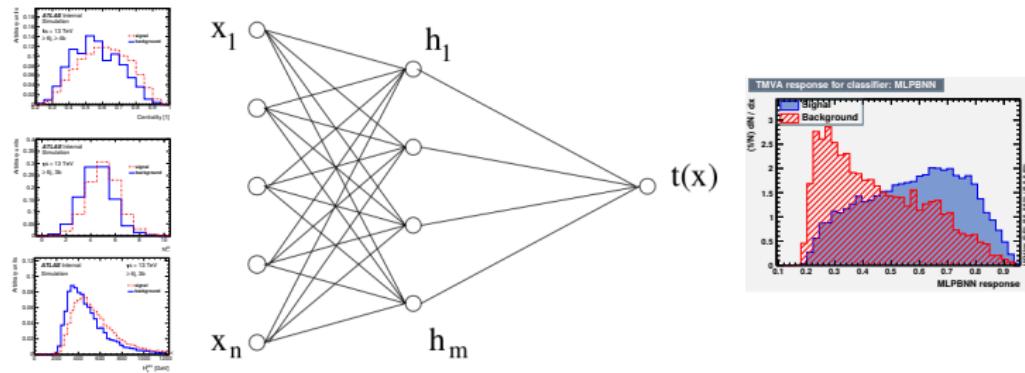
Individual cuts won't do the job. We definitely need more complex methods!

# Artificial Neural Network



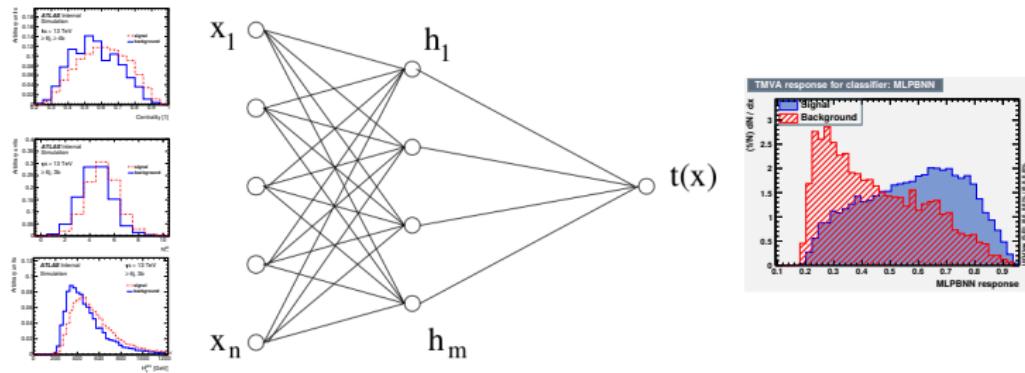
- Inspired by biology (i.e. the brain)
- Process information through Multilayer Network
  - input layer: variables
  - "hidden layer": neurons
  - output layer:  $t$

# Artificial Neural Network



- Inspired by biology (i.e. the brain)
- Process information through Multilayer Network
  - input layer: variables
  - "hidden layer": neurons
  - output layer:  $t$
- **parameters ("weights")**  $\vec{w}$ : strength of connections

# Artificial Neural Network



- Inspired by biology (i.e. the brain)
- Process information through Multilayer Network
  - input layer: variables
  - "hidden layer": neurons
  - output layer:  $t$
- **parameters ("weights")**  $\vec{w}$ : strength of connections
- **Important:**  $t(\tilde{x})$  is non-linear in general!

# Artificial Neural Network - training

Define error function  $\epsilon$  for an input vector  $\vec{x}$ , e.g.

$$\epsilon(\vec{x}, \vec{w}) = (t(\vec{x}, \vec{w}) - \hat{t})^2$$

$$\hat{t}(\text{sig}) = 1$$

$$\hat{t}(\text{bkg}) = -1$$

# Artificial Neural Network - training

Define error function  $\epsilon$  for an input vector  $\vec{x}$ , e.g.

$$\begin{aligned}\epsilon(\vec{x}, \vec{w}) &= (t(\vec{x}, \vec{w}) - \hat{t})^2 \\ \hat{t}(\text{sig}) &= 1 \\ \hat{t}(\text{bkg}) &= -1\end{aligned}$$

Iterative training with Monte-Carlo samples:

- start with random/guessed weights  $\vec{w}$
- calculate error for training event
- adjust weights  $\vec{w}$  to minimize  $\epsilon$

# Outline

## ④ Results

Overview

Neural Network training studies

Sensitivity of ttH

# Overview of my work

## Software:

- 
- **TMVA** (4.2.1), comes with **ROOT** (6.04/02)
- Toolkit for **Mu**lti**V**ariate **A**nalysis
- Set up a training testing framework for NNs

# Overview of my work

## Software:

- TMVA (4.2.1), comes with ROOT (6.04/02)
- Toolkit for MultiVariate Analysis
- Set up a training testing framework for NNs

## Variables & training settings:

- adapted from LHC run 1 analysis
- began optimisation for run 2

# Overview of my work

## Software:

- TMVA (4.2.1), comes with ROOT (6.04/02)
- Toolkit for MultiVariate Analysis
- Set up a training testing framework for NNs

## Variables & training settings:

- adapted from LHC run 1 analysis
- began optimisation for run 2

## Impact on Sensitivity of ttH

- Calculate signal significance for estimated luminosity

# NN with variables and settings from run 1

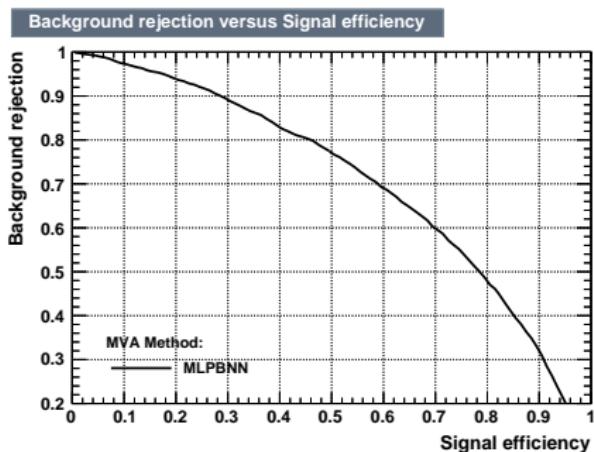
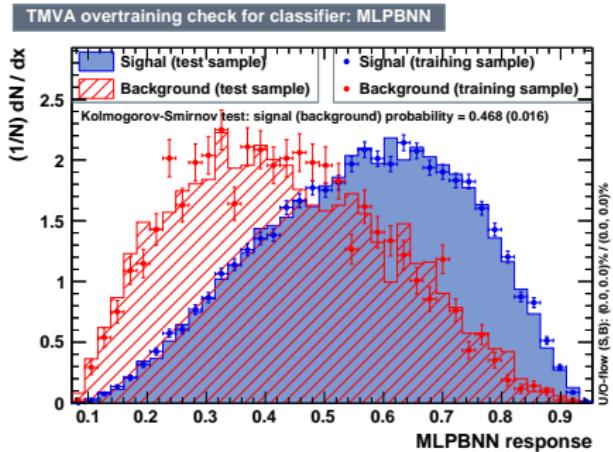


Figure: Neural Network output in the  $\geq 6$  jets 3 tight b-jets region

**Receiver Operating Characteristic** := Curve on the right, telling how good different cuts are.

# Signal Significance: Category + NN

"New" category:  $\geq 6$  jets,  $\geq 3$  tight b-jets,  $\geq 4$  loose b-tags

- $\frac{S}{B} = 1.7\%$
- $\frac{S}{\sqrt{S+B}} = 0.9$

For a Luminosity of  $\mathcal{L} = 5 \text{ fb}^{-1}$ :

- Signal: 45 events
- Background: 2577 events

After applying optimal Neural Network cut:

- $\frac{S}{\sqrt{S+B}} = 1.0$

## Outlook

- analysis will include **all different background samples**

## Outlook

- analysis will include **all different background samples**
- Neural Network training with **more MC statistics**

## Outlook

- analysis will include **all different background samples**
- Neural Network training with **more MC statistics**
- **new MVA-variables** will possibly be added

# Outlook

- analysis will include **all different background samples**
- Neural Network training with **more MC statistics**
- **new MVA-variables** will possibly be added
- Inefficient: Cut in Neural Network output

# Outlook

- analysis will include **all different background samples**
- Neural Network training with **more MC statistics**
- **new MVA-variables** will possibly be added
- Inefficient: Cut in Neural Network output
- Better: **Likelihood fit to Neural Network** output distribution

Special thanks to my supervisors:

- **John Stakely Keller**
- **Judith Katzy**

Thanks to the whole ATLAS group here at DESY!!!

**THANK YOU FOR YOUR ATTENTION!!!**

Questions?

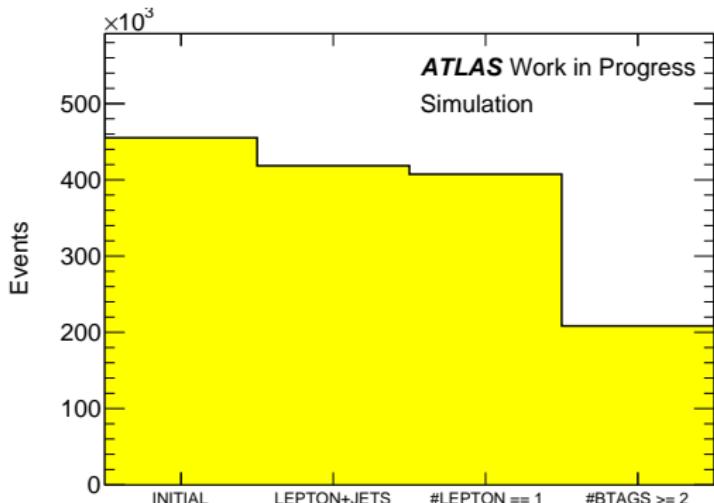
# Appendix: Simulated Data

## Two Monte-Carlo samples

Type	Description	Generator	Events
$t\bar{t}$ (bkg.)	semi-/dileptonic	Powheg+Pythia	~ 2 million
$t\bar{t}H$ (sig.)	inclusive higgs decay, semileptonic top decay	aMcAtNlo+Herwig++	~ 2 million

- minor bkg. contributions missing ( $t\bar{t}V$ , non- $t\bar{t}$ )
- Number of available events (after skimming):
  - signal: ca. 400000
  - background: ca. 450000

# Appendix: Pre-Selection Cuts



- $N_{\text{jets}} \geq 4$
- One lepton with  $|p_T| > 25 \text{ GeV}$
- Exactly one Lepton
- Require  $N_{\text{b-tags}} \geq 2$

Figure: Cutflow of skimmed **bkg.** sample.

# Appendix: categories used for Neural Network

## Reminder:

- categorisation according to number of jets & number b-jets
- different b-tagging working points possible
- loose: 85% efficiency
- tight: 77% efficiency

## The most promising regions:

- good signal purity ( $\frac{S}{B} > 1\%$ ) & good signal significance ( $\frac{S}{\sqrt{B}} > 0.3$ )
- as much events as possible

## Raw MC events in examined categories

	old		new (Hyoyin)
	tight b-tags: 3	tight b-tags: $\geq 4$	tight b-tags: $\geq 3$ loose b-tags: $\geq 4$
$\geq 6$ jets	<b>sig:</b> $\sim 46000$ <b>bkg:</b> $\sim 7800$	<b>sig:</b> $\sim 22600$ <b>bkg:</b> $\sim 900$	<b>sig:</b> $\sim 35600$ <b>bkg:</b> $\sim 2300$

## Appendix: MVA in formulas

- Each event (**signal** or **background**) has  $D$  measured variables  $\vec{x}$
- "Feature space" in  $\mathbb{R}^D$ :
  - $x_0 = \text{Centrality}$
  - $x_1 = N_{40}^{\text{jet}}$
  - ...
- Find mapping:

$t(\vec{x}) : \mathbb{R}^D \longrightarrow \mathbb{R}$
- $t$  is new "classifier", desired properties:
  - good separation power, i. e.  $\langle t_{\text{sig}} \rangle \neq \langle t_{\text{bkg}} \rangle$  and  $\sigma_{t_{\text{sig}}}, \sigma_{t_{\text{bkg}}}$  small
  - good generalization properties when applied to "unknown" events

# Appendix: Artificial Neural Network

## Feed-Forward Network

$$h_i(\vec{x}) = s \left( w_{i0} + \sum_{i=1}^n w_{ij} x_i \right)$$

$$t(\vec{x}) = s \left( a_0 + \sum_{i=1}^n a_i h_i \right)$$

## Our Artificial Neural Network (ANN):

- Architecture: MLP
- Information propagation: feed-forward

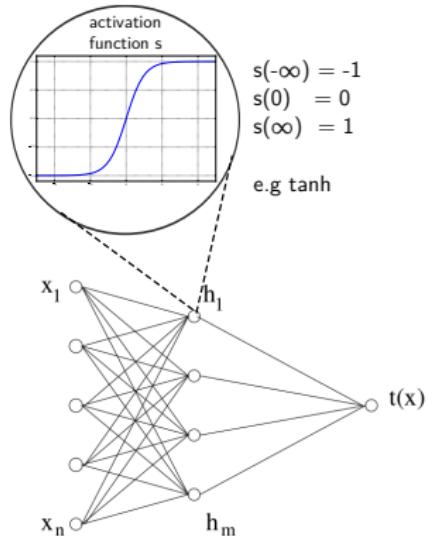


Figure: MultiLayer Perceptron with one hidden layer.

# NN with variables and settings from run 1

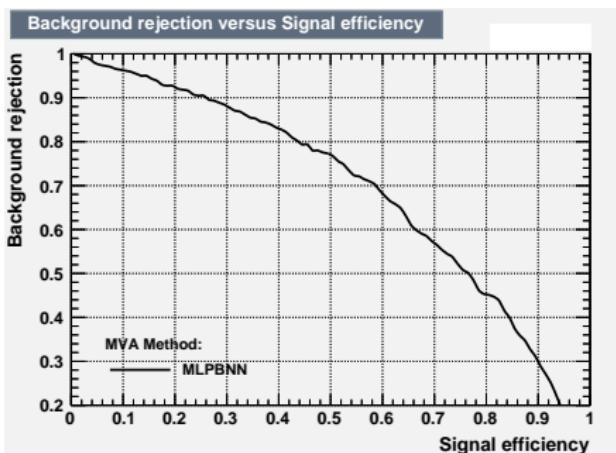
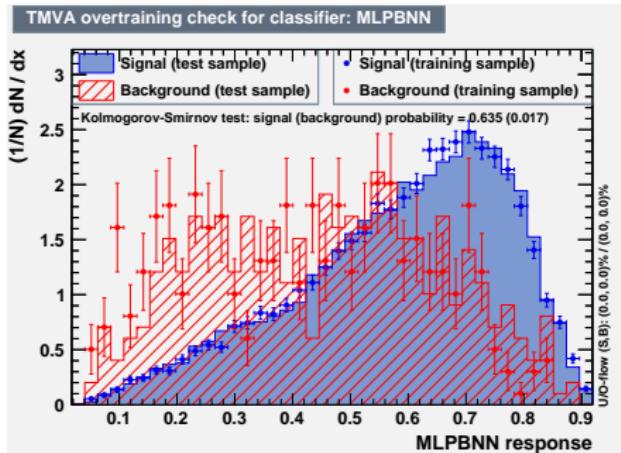


Figure: Neural Network output in the  $\geq 6$  jets  $\geq 4$  tight b-jets region

# NN with variables and settings from run 1

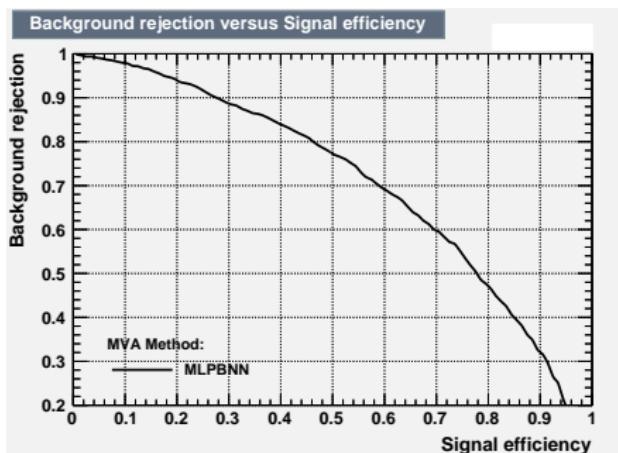
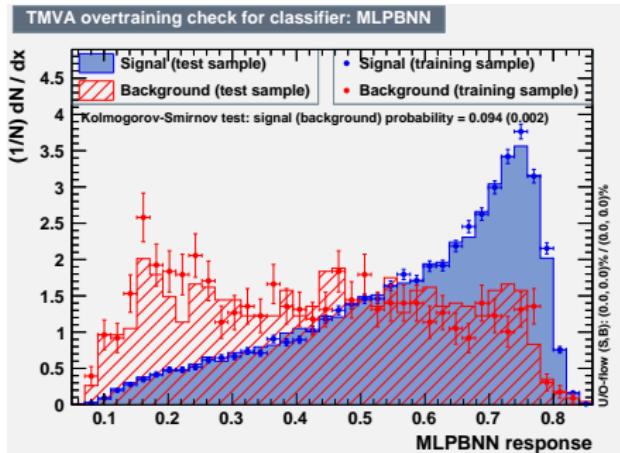


Figure: Neural Network output in the  $\geq 6$  jets  $\geq 3$  tight &  $\geq 4$  loose b-jets region

# Convergence of the Neural Netork

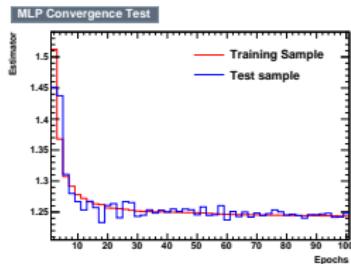


Figure: Error Estimator vs training cycles (6j 3b).

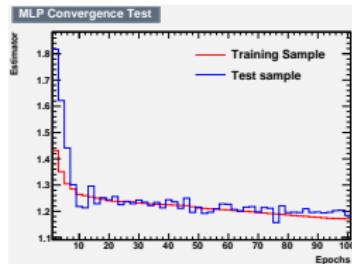


Figure: Error Estimator vs training cycles (6j 4b).

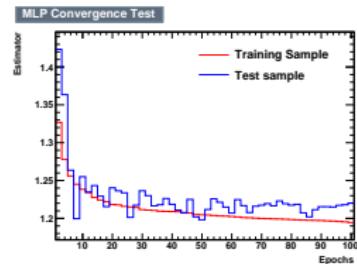


Figure: Error Estimator vs training cycles (6j 3/4b).

# Ranking of Variables – naive

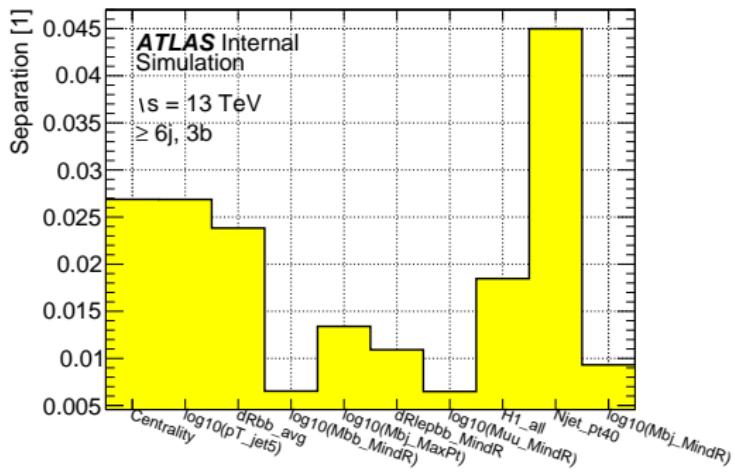


Figure: Separation power of all variables used in the  $\geq 6$  jets 3 tight b-jets region.

Separation power of a discriminating variable  $\xi$ :

$$\int \frac{(N_{\text{sig}}(\xi) - N_{\text{bkg}}(\xi))^2}{(N_{\text{sig}}(\xi) + N_{\text{bkg}}(\xi))} d\xi$$

Ranking according to separation

- ①  $N_{40}^{\text{jet}}$
- ② Centrality
- ③  $p_T^{\text{jet5}}$
- ④ ...

## Ranking of Variables

- Train Neural Network 10 times
- leave out one variable each time
- see how quality of NN "suffers"

# Ranking of Variables

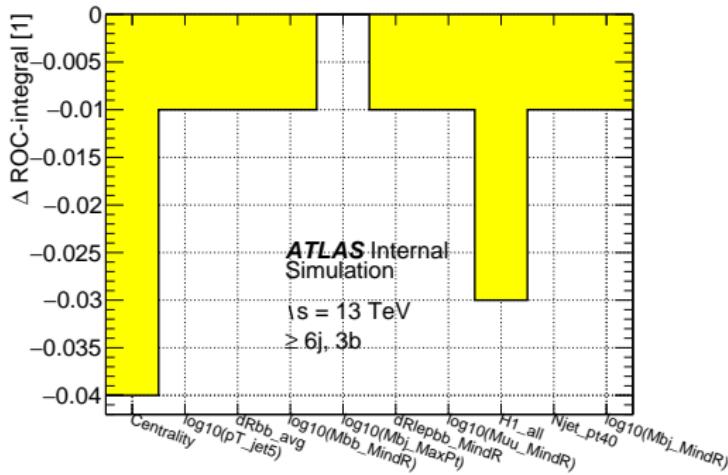


Figure: Separation power of all variables used in the  $\geq 6$  jets 3 tight b-jets region.

- Train Neural Network 10 times
- leave out one variable each time
- see how quality of NN "suffers"

Ranking according to separation

- ① Centrality
- ②  $H_1$  (Fox-Wolfram)
- ③  $N_{40}^{\text{jet}}$
- ④ ...

# Ranking of Variables – Run 1 vs. Run 2 simulation (6j3b)

## Run 1

- ①  $N_{40}^{\text{jet}}$
- ② Centrality
- ③  $H_1$  (Fox-Wolfram moment)
- ④  $M_{\text{bj}}^{\min \Delta R}$
- ⑤  $\Delta R_{\text{bb}}^{\text{avg}}$
- ⑥  $p_T^{\text{jet5}}$
- ⑦  $M_{\text{bj}}^{\max p_T}$
- ⑧  $M_{\text{uu}}^{\min \Delta R}$
- ⑨  $\Delta R_{\text{lepb}}^{\min \Delta R}$
- ⑩  $M_{\text{bj}}^{\min \Delta R}$

## Run 2 simulation

- ① Centrality
- ②  $H_1$  (Fox-Wolfram moment)
- ③  $N_{40}^{\text{jet}}$
- ④  $\Delta R_{\text{bb}}^{\text{avg}}$
- ⑤  $p_T^{\text{jet5}}$
- ⑥  $M_{\text{bb}}^{\min \Delta R}$
- ⑦  $M_{\text{bj}}^{\max p_T}$
- ⑧  $\Delta R_{\text{lepb}}^{\min \Delta R}$
- ⑨  $M_{\text{bj}}^{\min \Delta R}$
- ⑩  $M_{\text{uu}}^{\min \Delta R}$