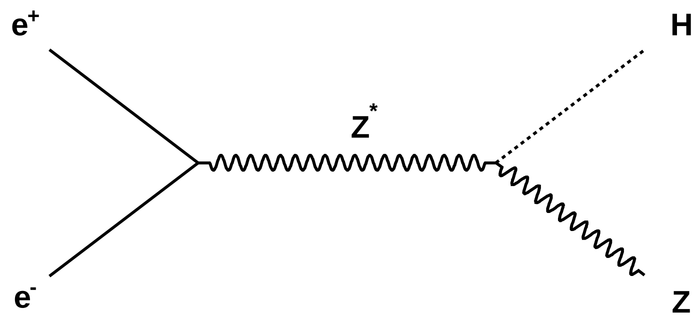
$ZH \longrightarrow q\bar{q}b\bar{b}$ Study With a Neural Network

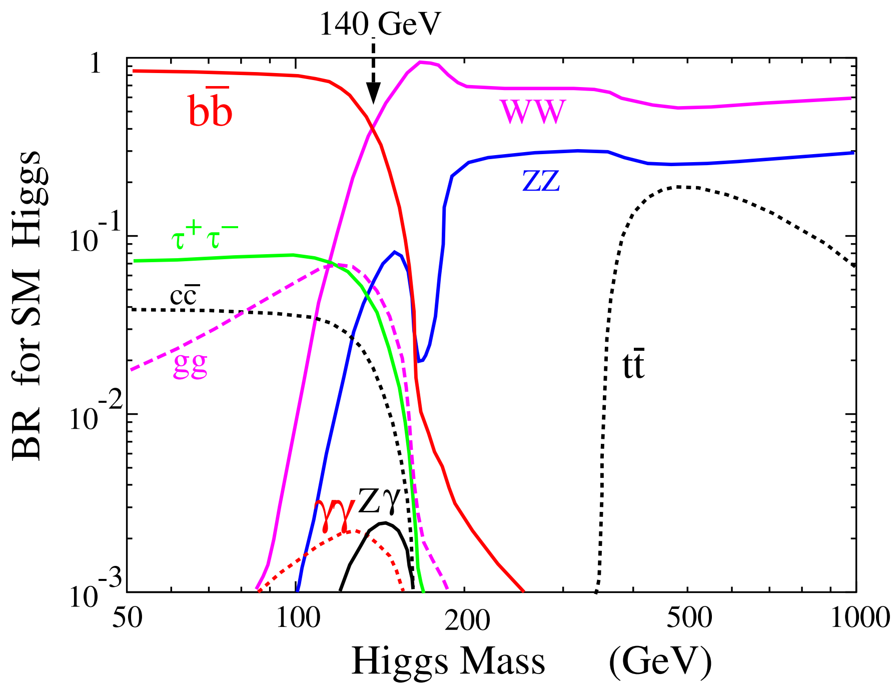
David Ward and Wenbiao Yan

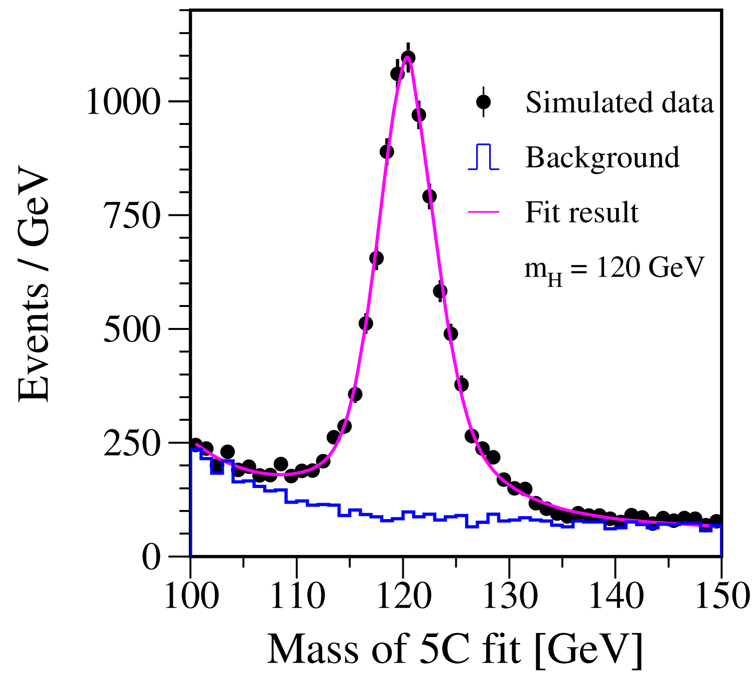
UNIVERSITY OF CAMBRIDGE



- **ZH selection with a neural network**
- **Higgs mass reconstruction via $M_{b\bar{b}}$**

Higgs decay





- Br($h \to b\bar{b}$) $\sim$ 68% @ $M_h = 120$ GeV
- Pandora-pythia v3.3 Monte Carlo
- $e^+e^- \to ZH \to q\bar{q}b\bar{b}$ @ 350 GeV
- integrated luminosity of 500fb^{-1}

- TESLA fast simulation: EPJ C44(2005) 481 by P. García-Abia, W. Lohmann, A. Raspereza
- $ZH \to q\bar{q}q'\bar{q}'$ @ 350 GeV $\Longrightarrow \Delta(m_H) =$ 45 MeV
- goal: full detector simulation and reconstruction
- goal: compare different PFAs

Event reconstruction

- **Detector simulation: LDC00Sc detector model @ Mokka v6.2;**
- **Marlin v00-09-09; MarlinReco v00-05; MarlinUtil v00-05**
- **Jet finder: KtJet package v1.08 (C++)**
- **Kinematic fitting: KinFit package @ CMS (C++)**
- **TMVA package v3.8.13 for neural network**
- **Flavour tagging: LCFIVertex package v00-02-02**
  - **use training sample in b-tag package**
  - **is it OK for LDC00SC at Mokka 6.2 ??? calibration constant and PFA**
  - **we do not show results by b-tag in this talk**

$e^+e^- \to ZH \to q\bar{q}b\bar{b}$ event selection

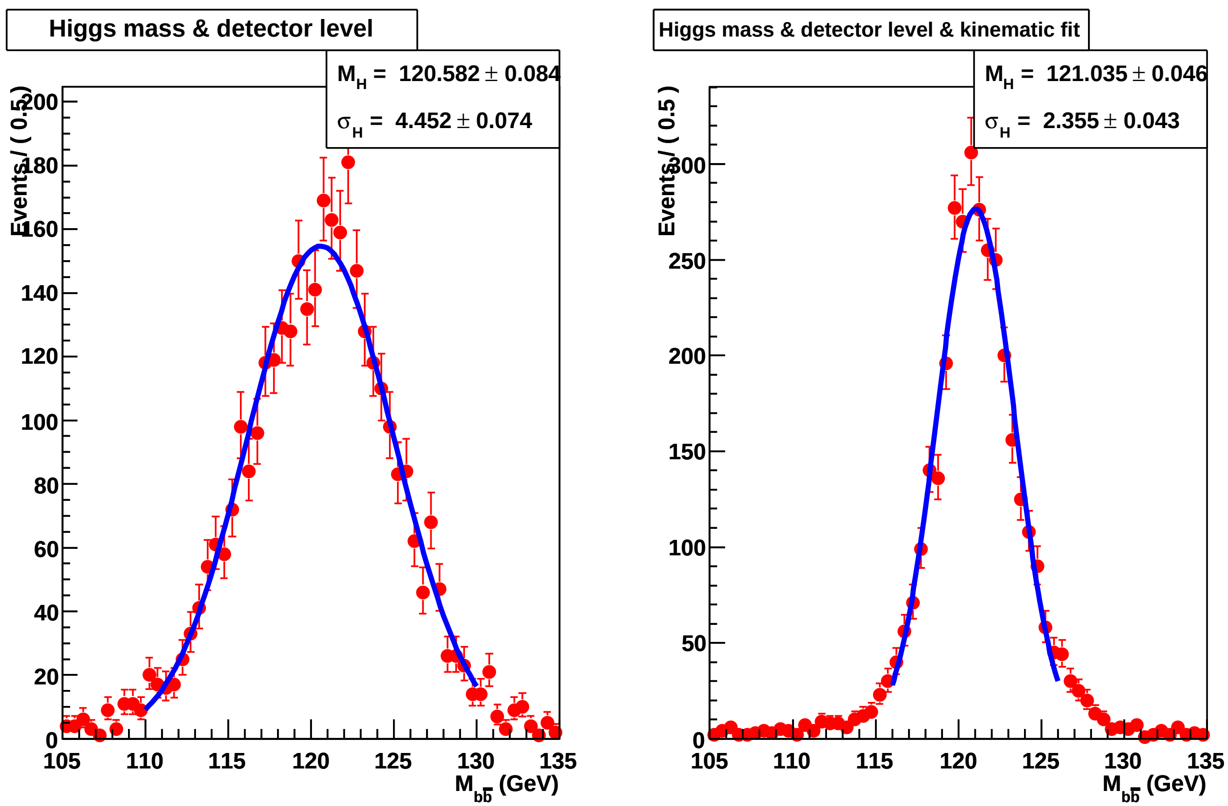
- **We follow the paper EPJ C44(2005) 481 to select four jets event**
  - **total visible energy:** $E_{visible} >= 0.8 * 350$ **GeV**
  - **particle plow object number:** $N_{PFOs} >= 40$
  - **event shape parameter T (thrust):** $T < 0.85$ **and** $|\cos(\theta_T)| < 0.80$
  - **force events to have 4 jets, and parameter** $\log_{10}(1/y_{34}) < 5.0$
    - **jet energy:** $E_{jet} > 10.0$ **GeV**
    - **jet theta:** $|\cos(\theta)| < 0.99$

- **Kinematic fitting: 5C fitting**
  - χ^2 **probability:** $P(\chi^2) > 0.05$
  - **use smallest** χ^2 **of kinematic fitting to choose jet pairing**

- **MC data @ 250 GeV: not yet** $\Longrightarrow$ **analysis codes @ 350 GeV with local data samples**

Higgs mass fitting (without background events)

- **Kinematic fitting improves mass resolution: $\Delta(m_H) = 46$ MeV**
- **TESLA fast simulation with background events: $\Delta(m_H) = 45$ MeV**



MC data samples @ 350 GeV

- $ZH \rightarrow q\bar{q}b\bar{b}$: **Pandora-pythia** $\sim 32K$ **(signal)**
- $ZH \rightarrow q\bar{q}b\bar{b}$: **Pythia** $\sim 51K$ **(training & test @ Neural network)**
- $WW \rightarrow q_1\bar{q}_1q_2\bar{q}_2$: **Pythia** $\sim 92K$
- $ZZ \rightarrow q_1\bar{q}_1q_2\bar{q}_2$: **Pythia** $\sim 127K$
- **QQ: Pythia** $\sim 99K$
- **event preselection**
  - **total visible energy:** $E_{visible} >= 0.6 * 350$ **GeV**
  - **particle plow object number:** $N_{PFOs} >= 40$
  - **four good jets**

Why neural network ?

- **event selection with cuts: thrust, thrust theta, visible energy, y_{34}, PFA number and χ^2 probability**
- **event selection with a neural network: SAME variables; default neural network settings and architecture N:(N+1):N:1 for $N = 6$**
- **both case: χ^2 probability $P(\chi^2) > 0.05$ for good reconstructed events**

	ZH	WW	ZZ	QQ	B/S
cuts	4350	704	11623	33571	10.5514
neural network	4350	183	5920	22768	6.6338

- **neural network could improve signal/background ratio**

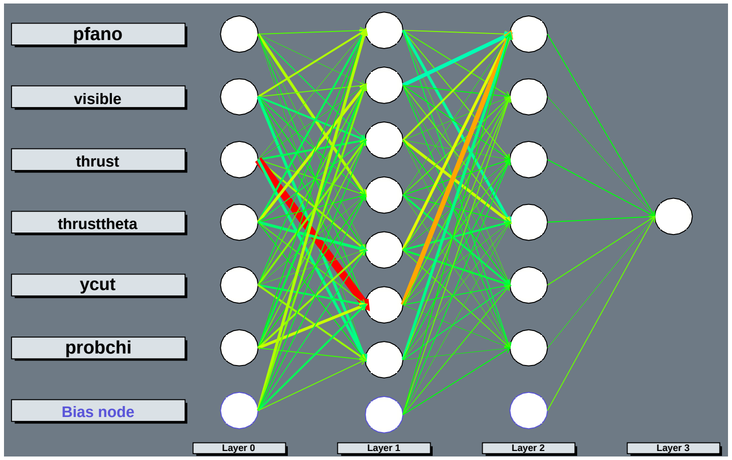
Try to use neural network, NOT final answer

Useful variables

- **total visible energy; PFA number;** y_{34}
- **thrust; theta of thrust axis; sphericity; aplanarity**
- **Fox-Wolfram moments** h_{30} **and** h_{40}
- **minimum jet-jet angle;** E_{jet}^{min}; $E_{jet}^{max} - E_{jet}^{min}$
- χ^2 **of 5C fitting;** Z^0 **mass**
- j_{mom}, j_{ang}, **modified Nachtmann-Reiter angle @ OPAL CERN-EP/98-167**
  - **sort jets by jet energy** $E_{j1} \geq E_{j2} \geq E_{j3} \geq E_{j4}$
  - $j_{mom} = \frac{|\vec{p_1}|+|\vec{p_2}|-|\vec{p_3}|-|\vec{p_4}|}{\sqrt{s}}$
  - $j_{ang} = \frac{E_4}{\sqrt{s}}(1 - \cos\theta_{12}\cos\theta_{13}\cos\theta_{23})$
  - $|\cos\theta_{N-R}| = \frac{(\vec{p_1}-\vec{p_2})\cdot(\vec{p_3}-\vec{p_4})}{|\vec{p_1}-\vec{p_2}|\cdot|\vec{p_3}-\vec{p_4}|}$

Neural network architecture

- how many variables at input layer ? one output nodes
- how many hidden layers ? and how many nodes at hidden layers ?
  - In principle, one hidden layer is sufficient. In practice, two hidden layers with a small number of neurons may work better (and/or learn faster) than a network with a single layer.
- A neural network MLP default architecture N:N+1:N:1



Compare different neural network architectures

- *separation*: **TMVA package; neural network response y**

$$\frac{1}{2}\int\frac{(y_s - y_b)^2}{y_s + y_b}dy \qquad (1)$$

separation **is zero for identical signal and background shapes, and it it one for shapes with no overlap.**

- $S/\sqrt{S+B}$ **with "0.5-criterion"**
  - **Signal (NN's response > 0.5);**
  - **Background (NN's response < 0.5)**

- **background-to-Signal ratio B/S with "0.5-criterion"**

- **gaussian significance $S/\sqrt{B}$ with "0.5-criterion"**

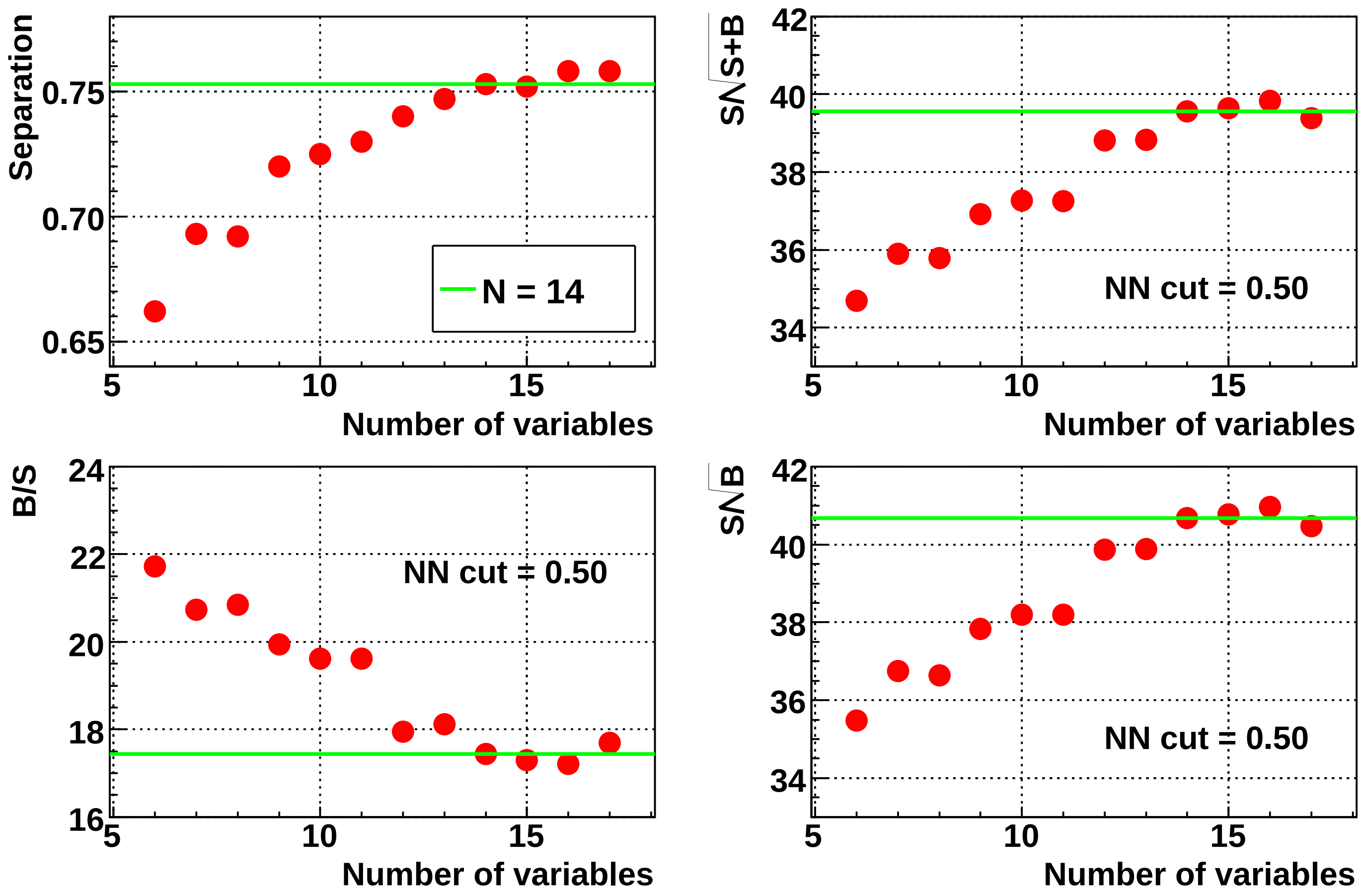
Number of variables at input layer

- **N variables from total 17 variables: C_{17}^{N} choices;**
  e.g. $C_{17}^{10} = 17 * 16 * 15 * 14 * 13 * 12 * 11 = 98017920$!!!
- **remove variable with smallest $Importance$ at each step**
  $Importance$: sum of the weights-squared of the connections that leave input variable

```
--- MLP          : Ranking result (top variable is best ranked)
--- MLP          : -----------------------------------------------------------------
--- MLP          : Rank : Variable             : Importance
--- MLP          : -----------------------------------------------------------------
--- MLP          :    1 : aplanarity           : 9.111e+01
--- MLP          :    2 : jang                 : 6.002e+01
--- MLP          :    3 : h30                  : 1.918e+01
--- MLP          :    4 : thrusttheta          : 1.174e+01
--- MLP          :    5 : sphere               : 8.744e+00
--- MLP          :    6 : thrust               : 7.013e+00
--- MLP          :    7 : h40                  : 2.858e+00
--- MLP          :    8 : chi2                 : 2.721e+00
--- MLP          :    9 : ycut                 : 2.195e+00
--- MLP          :   10 : jmom                 : 2.130e+00
--- MLP          :   11 : visible              : 1.953e+00
--- MLP          :   12 : pfano                : 1.832e+00
--- MLP          :   13 : minjetenergy         : 4.206e-01
--- MLP          :   14 : nrang                : 3.874e-01
--- MLP          :   15 : massz0               : 1.636e-02
--- MLP          :   16 : smallangle           : 1.457e-02
--- MLP          :   17 : jetenergydifference  : 8.297e-04
```

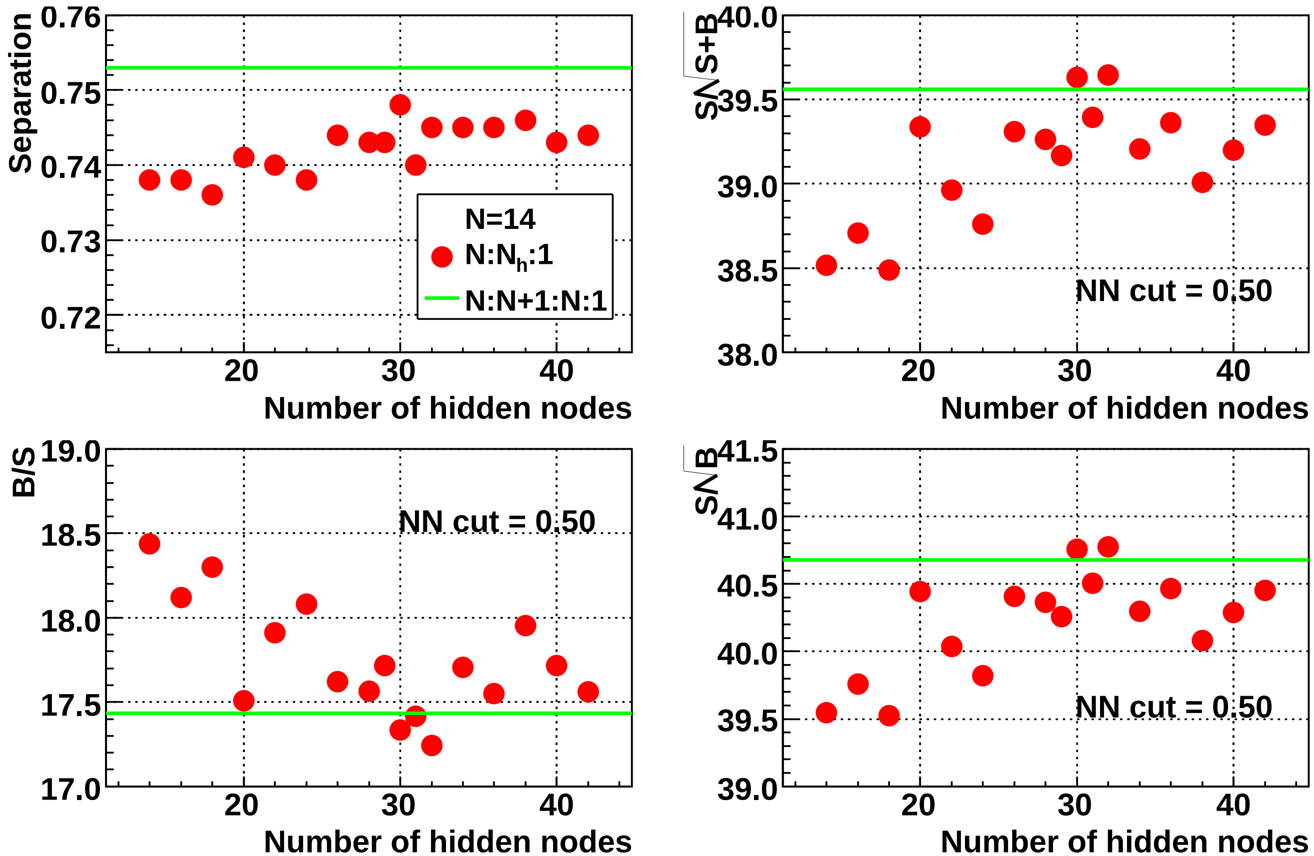
Number of variables at input layer: N:N+1:N:1

- **at input layer: 14 variables** √

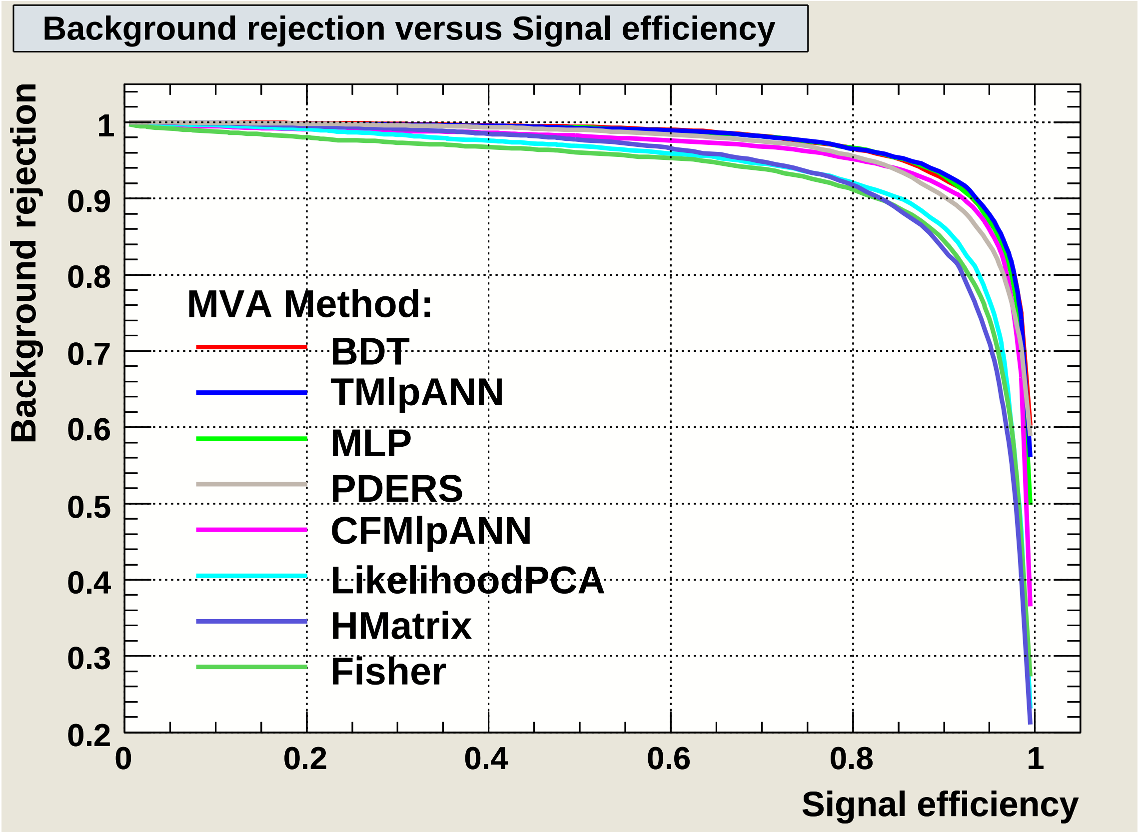


One hidden layer: N:N$_h$:1

- for 14 variables: 14:15:14:1 $\sim$ 14:30:1 $\surd$ & 14:32:1
- similar plots for any number of variables: not yet



TMVA classifiers



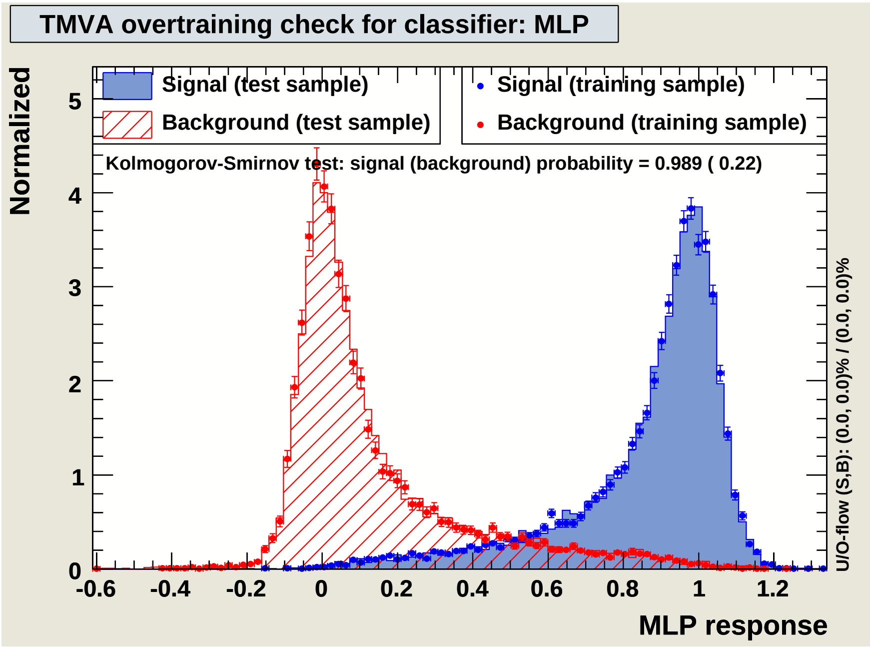
- **Parameters of classifiers: taken from example code of TMVA package**
- **Boosted Decision Trees (BDT) and Artificial Neural Networks (TMlpANN and MLP) have similar results.**

TMVA classifiers

```
--- Factory      : Evaluation results ranked by best signal efficiency and purity (area)
--- Factory      : -----------------------------------------------------------------------------
--- Factory      : MVA              Signal efficiency at bkg eff. (error):  |  Sepa-    Signifi-
--- Factory      : Methods:         @B=0.01    @B=0.10    @B=0.30    Area   |  ration:  cance:
--- Factory      : -----------------------------------------------------------------------------
--- Factory      : BDT            : 0.605(03)  0.932(02)  0.988(00)  0.973  |  0.753    1.874
--- Factory      : TMlpANN        : 0.578(04)  0.936(01)  0.987(00)  0.973  |  0.758    1.951
--- Factory      : MLP            : 0.579(04)  0.932(02)  0.986(00)  0.971  |  0.750    1.772
--- Factory      : PDERS          : 0.491(04)  0.903(02)  0.985(00)  0.965  |  0.695    1.654
--- Factory      : CFMlpANN       : 0.252(03)  0.921(02)  0.981(01)  0.959  |  0.716    1.748
--- Factory      : LikelihoodPCA  : 0.210(03)  0.852(02)  0.964(01)  0.936  |  0.615    1.330
--- Factory      : HMatrix        : 0.325(03)  0.831(03)  0.952(01)  0.934  |  0.612    1.194
--- Factory      : Fisher         : 0.069(02)  0.826(03)  0.959(01)  0.927  |  0.614    1.088
--- Factory      : -----------------------------------------------------------------------------
```

- **Boosted Decision Trees (BDT): slowest**
- **TMlpANN and MLP: a clear speed advantage for the MLP**
- **MLP:** $\surd$

Overtraining of neural network

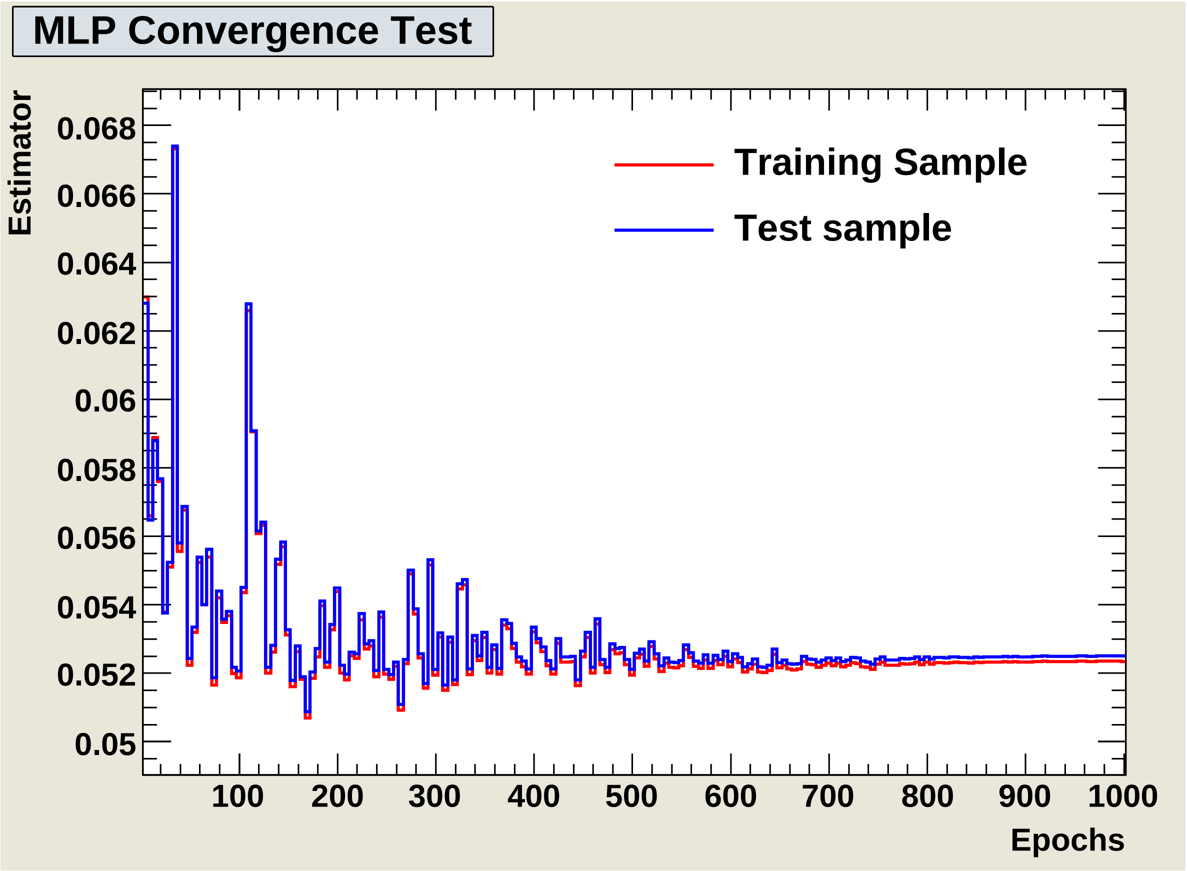



```
--- Factory      : Testing efficiency compared to training efficiency (overtraining check)
--- Factory      : -----------------------------------------------------------------------------
--- Factory      : MVA          Signal efficiency: from test sample (from traing sample)
--- Factory      : Methods:        @B=0.01              @B=0.10             @B=0.30
--- Factory      : -----------------------------------------------------------------------------
--- Factory      : MLP           : 0.559 (0.554)       0.929 (0.931)      0.986 (0.987)
--- Factory      : -----------------------------------------------------------------------------
```

- **NN's response for test sample and training sample are similar.**

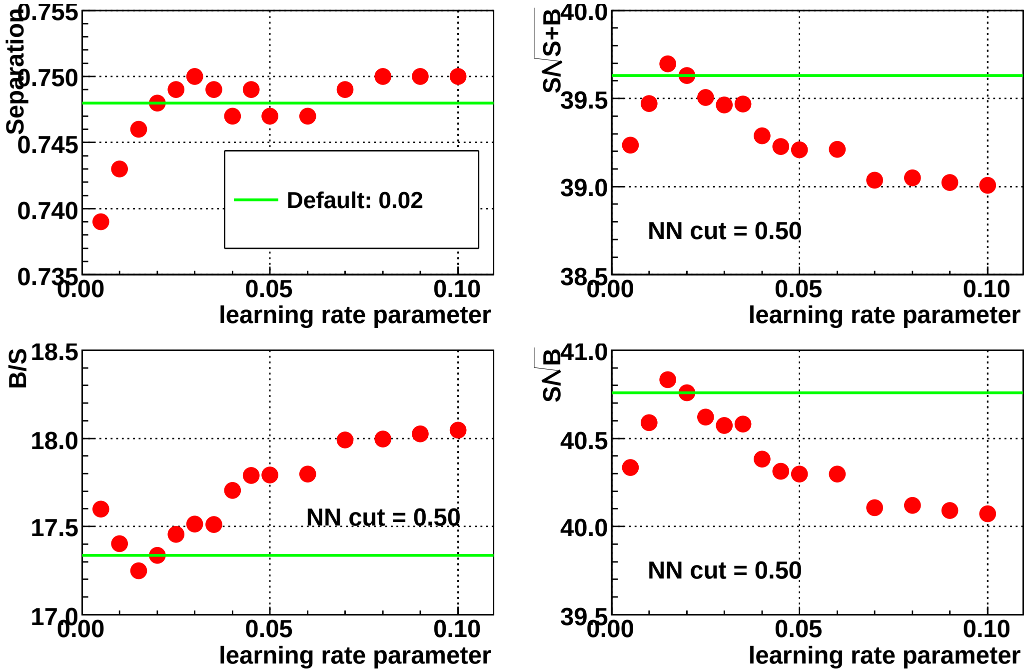
Neural network learning parameters

- **an epoch is one complete iteration through all available training samples.**
- **number of training cycles: 700**



Neural network parameters: learning rate η

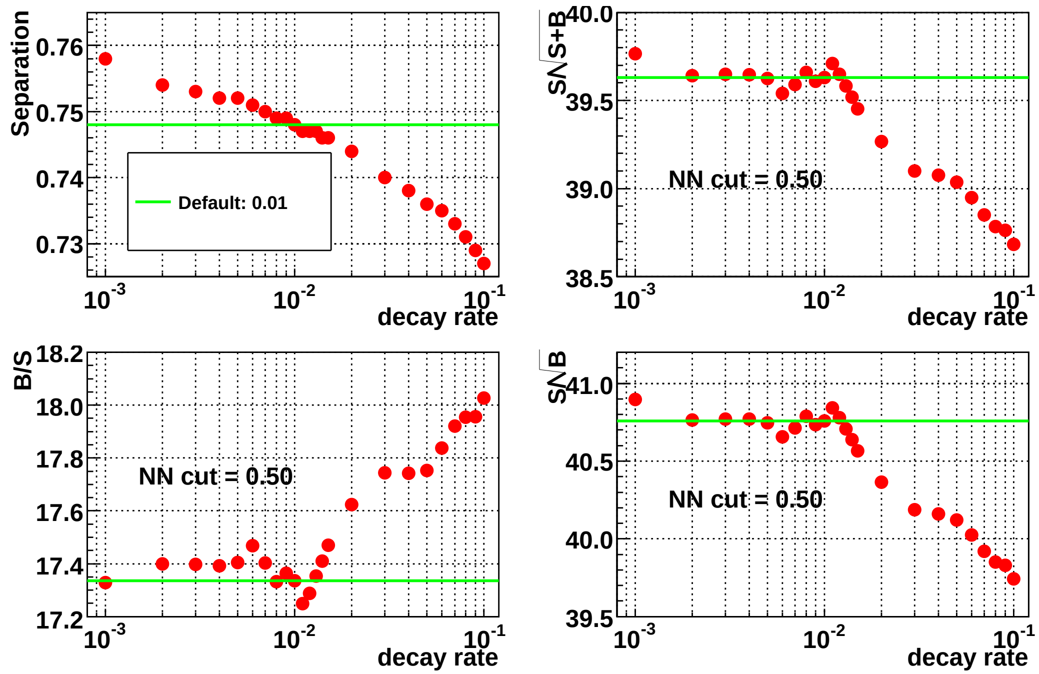
- **the positive learning rate η is a factor in updating weights**



η	ZH	WW	ZZ	QQ	B/S	
0.015	4350	131	3432	14288	4.1034	
0.020	4351	157	3533	13707	3.9984	$\surd$

Neural network parameters: decay rate

- **the decay rate is a factor for learning parameter; smaller decay rate, larger number of training cycles**



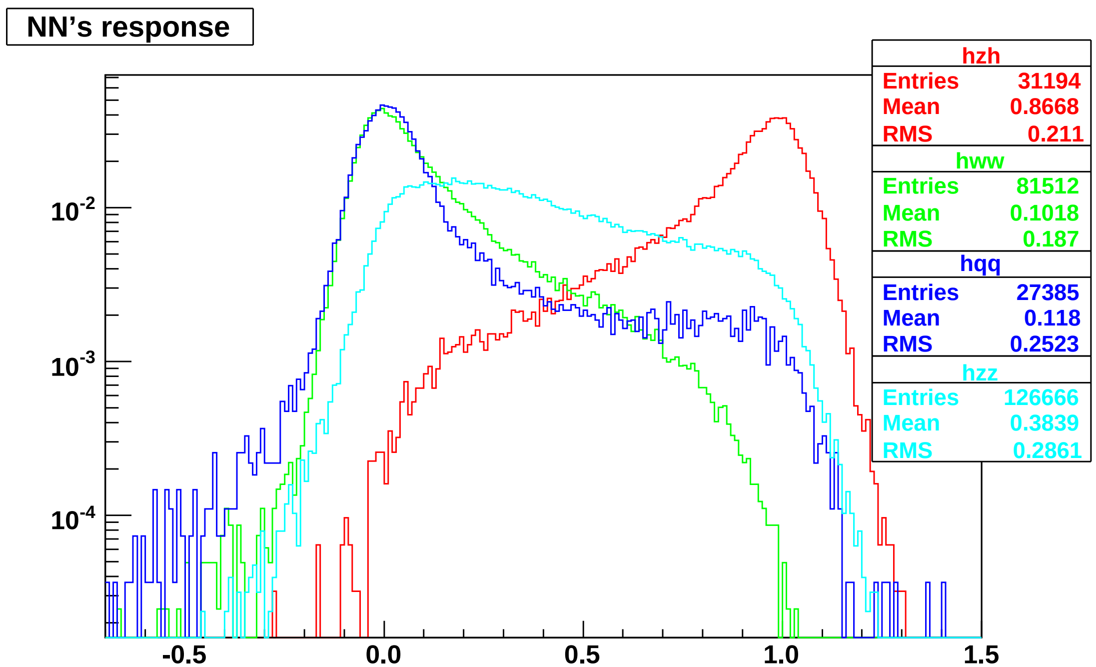
DecayRate	ZH	WW	ZZ	QQ	B/S	
0.010	4351	157	3533	13707	3.9984	$\surd$
0.011	4350	157	3505	13939	4.0462	

Neural network architecture

- **neural network architecture 14:30:1**
  - **input layers: 14 variables**
  - **one hidden layer: 30 nodes**
  - **output layers: one**
- **learning parameters**
  - **number of training cycles: 700**
  - **learning rate: 0.02 (default)**
  - **decay rate: 0.01 (default)**
- **neuron activation function: sigmoid (default)**
- **synapsis function: sum (default)**
- **learning mode: sequential (default)**

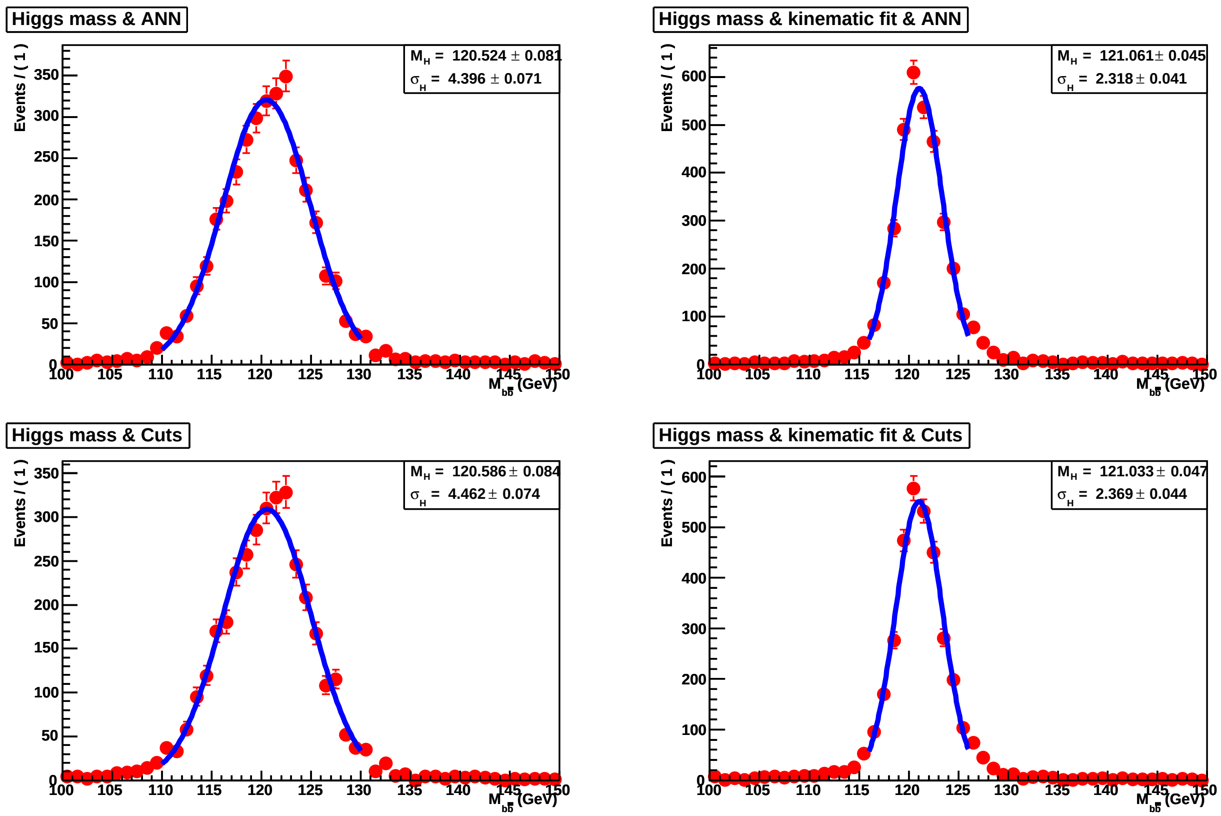
Neural network's response

- **the signal peaks around 1.0; the backgrounds peak around 0.0**
- **plots are normalizes to one**
- **NN response $\in (0, 1)$ by a sigmoid function $1/(1 + e^{1.25-4.5x})$: not yet**



Higgs mass: neural network vs. cuts-based method

	ZH	WW	ZZ	QQ	B/S	$S/\sqrt{B}$
cuts	4350	704	11623	33571	10.5514	20.3044
neural network	4351	157	3533	13707	3.9984	32.9877



Summary

- A analysis code is ready for $e^+e^- \longrightarrow ZH \longrightarrow q\bar{q}b\bar{b}$ study
  - cuts-based method; neural network: MLP @ TMVA package
  - neural network has a better signal/background ratio, slightly improves higgs mass resolution.

	ZH	WW	ZZ	QQ	B/S	$S/\sqrt{B}$
cuts	4350	704	11623	33571	10.5514	20.3044
neural network	4351	157	3533	13707	3.9984	32.9877

- wait for simulated samples for detector models @ 250 GeV
  - Try to use neural network @ 350 GeV, NOT final answer

- b-tag effect: we do not show results
  - use training sample in b-tag package
  - is it OK for LDC00SC at Mokka 6.2 ???