Measurement of the $t\bar{t}$ production cross section in the MET+jets channel in 2.2fb-1

CDF Note 9873

Gabriele Compostella,
INFN-CNAF and University of Padova
compostella@pd.infn.it

Silvia Amerio, Donatella Lucchesi
INFN and University of Padova

CDF Italia, Trieste
September 2, 2009

Analysis Motivation

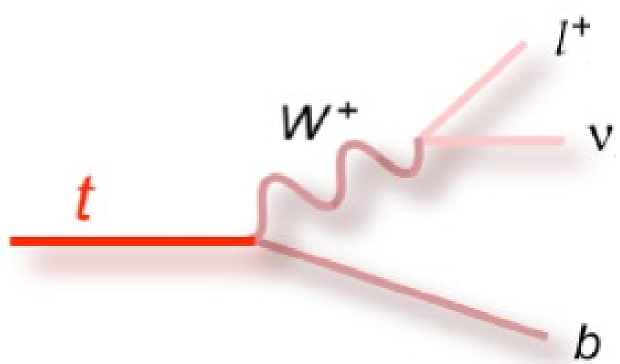
$$p\bar{p} \to t\bar{t} \to W^+ b W^- \bar{b} \to \text{b-jet}$$

$W^+ \to$ 2 jets, $b \to$ b-jet, $W^- \to l^- \bar{\nu}_l$

Signature:

- Large Jet Multiplicity (>=4 jets)
- Large MET from neutrino
- At least one b-tagged jet

Extension of previous measurement on 311 pb^-1 by G.Cortiana and collaborators [Phys. Rev. Lett. **96**, 202002 (2006)] : use 2.2fb^-1 of data and improve signal selection using a Neural Network. ...Preblessing tomorrow!



Analysis focuses on MET from neutrino rather than on lepton identification

- sensitive to leptonic W decays regardless of the lepton type
- large acceptance with respect to $W \to \tau\,\nu$ decays.

- Complementary and independent results wrt
  - Lepton +jets and
  - all-had measurements
- Large impact on the combination!

Analysis Summary

Channel has a large background, mainly from:

- QCD
- EWK+HF

- need an optimized kinematical selection
- need b-jet identification to increase S/N ratio.

Method Summary:

1. Decay channel is selected using a **cut on MET Significance** in order to choose high MET events
2. A single **Neural Network** is used to discriminate Signal vs Background
3. Search for **SecVTX tags** in the event
4. Background estimation performed using a **method 1 approach**: build a b-tag matrix (data driven) parametrized wrt to relevant variables
5. **Cut on NN output** and do a counting experiment to get the xsec measurement

TOP_MULTI_JET dataset up to **p13, 2.2 fb^{-1}** (gsetkd/h/i, gsetmi/j)

L1: at least 1 cal. tower with $E_T \geq 10$ GeV

L2: at least 4 cal. clusters with $E_T \geq 15$ GeV, $\sum E_T \geq 175$ GeV *

L3: at least 4 jets (Cone Radius = 0.4) , $E_T \geq 10$ GeV

MC : Pythia ttbar $M_{top} = 172.5$ GeV/c^2 (ttop25)

*introduced after run 194328 (TMJ-v5), was $\sum E_T \geq 125$ GeV

Clean up cuts:

- Good Run List v26 “em_mu_si_cmxignored”
- Trigger simulation
- Vertex requirements
- Tight leptons (e/μ) veto
- $\not{E}_T^{sig} \geq 3$
- NJets ($E_T > 15$ GeV, $|\eta| \leq 2.0$) ≥ 3

- **No overlap w/ other L+J top analyses**
- **Increased relative contribution from W→ τ ν +jets channel**

- **High, physics induced MET**
- **No overlap with the all-hadronic analysis**

Background Estimate Method

We will require SecVtx tags in the selected sample, need to estimate the background after selection

Basic Idea:

b-jet identification rates are different on ttbar and background processes, this allows to discriminate between the two components.

Method:

- Derive b-tag rates directly from TOP_MULTI_JET data
- Use **3** (E_T > 15 GeV, $|\eta|$ < 2.0) **jet** events
- Take the vars by which the tag-rate mainly depends to build a tag matrix

Signal contamination needs to be as low as possible in the sample used to parametrize the tagging rates in order to avoid biases in the background estimate!

In our case $F_{top} \sim 0.1\%$

Bkg Estimation:

Use the Tagging rate dependencies observed in 3-jet data events to predict the number of tagged jets at higher jet multiplicities and on kinematically selected data samples.

Warning:

Variables used for the tagging rate parametrization need to be able to track possible sample composition changes introduced by a given selection cut.

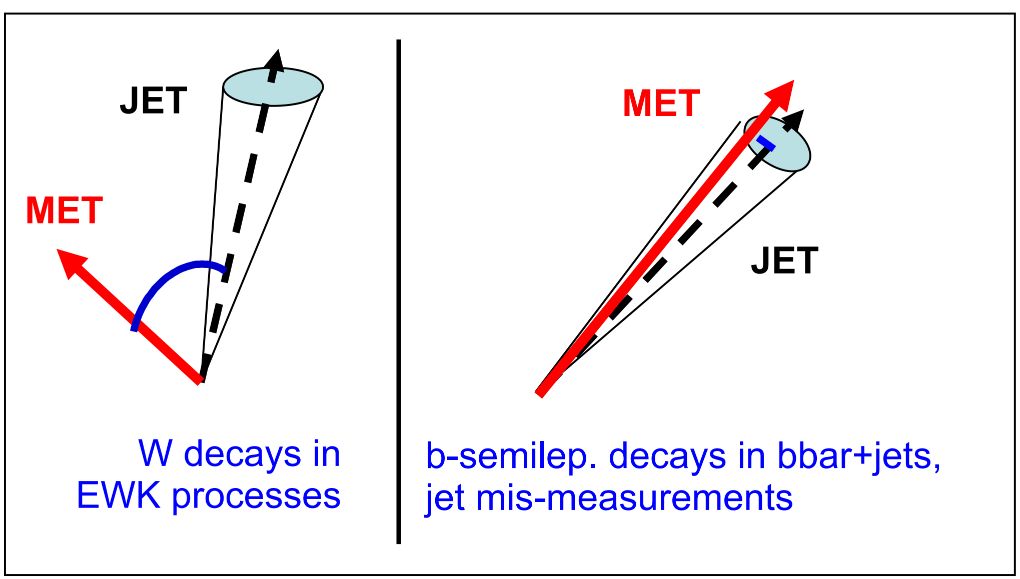
Method assumes that the tag rate does not depend on jet multiplicity, need to verify it!

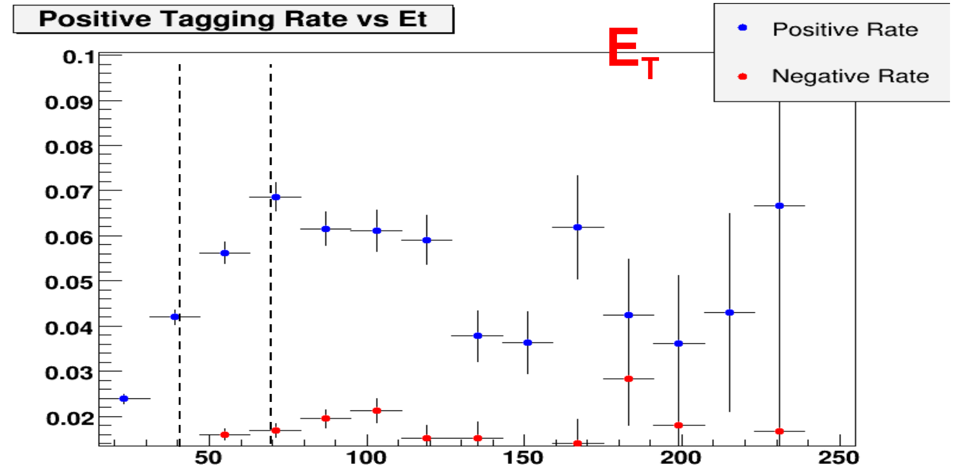
Background b-tagging rates

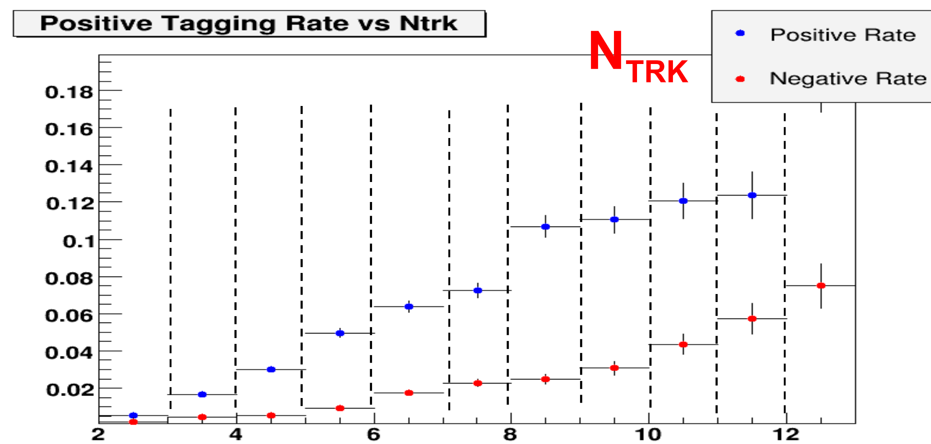
Build a b-tag matrix:

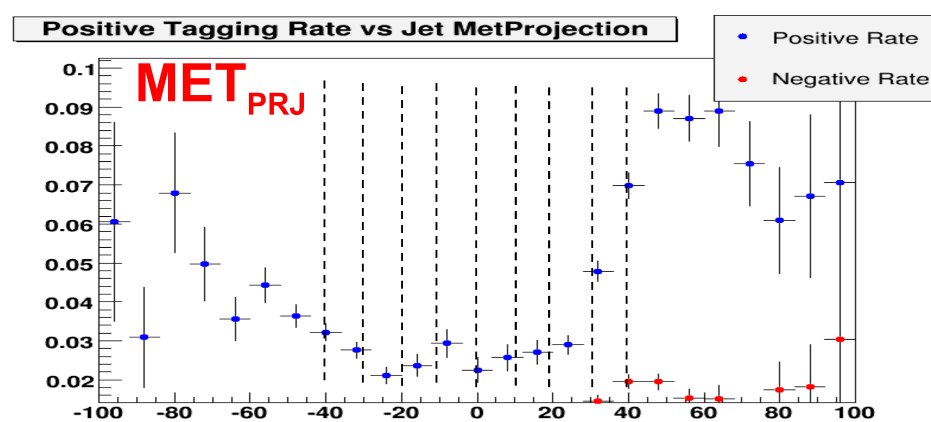
$$P\left(E_T, N_{trk}, MET_{PRJ}\right) = \frac{N_{jets}^{b-tagged}\left(E_T, N_{trk}, MET_{PRJ}\right)}{N_{jets}^{good}\left(E_T, N_{trk}, MET_{PRJ}\right)}$$

MET_{PRJ} has a consistent correlation with the heavy flavor component of the sample and allows to distinguish MET origins in relation to geometrical properties







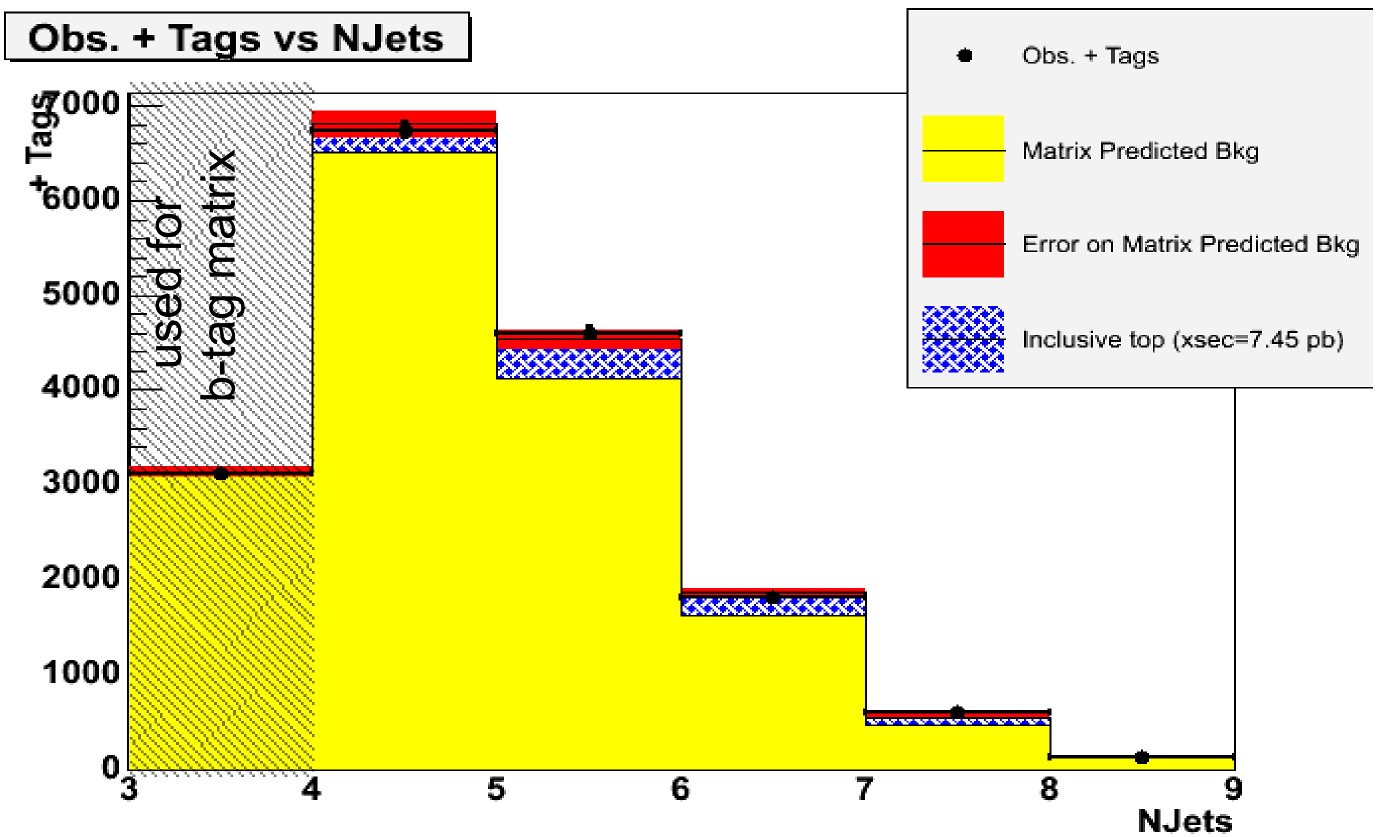


-- matrix binning

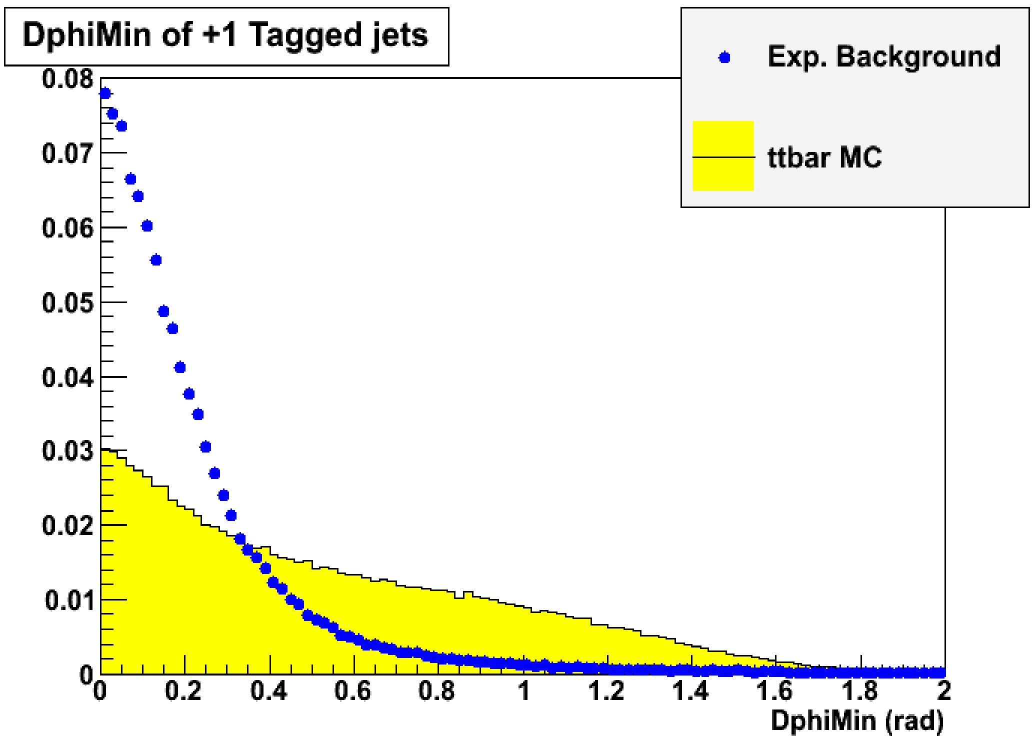
Need to verify that tagging matrix background prediction (parametrized on 3-jet events) works well for events with different jet multiplicity
Expected tags are determined from the tag rate parametrization using:

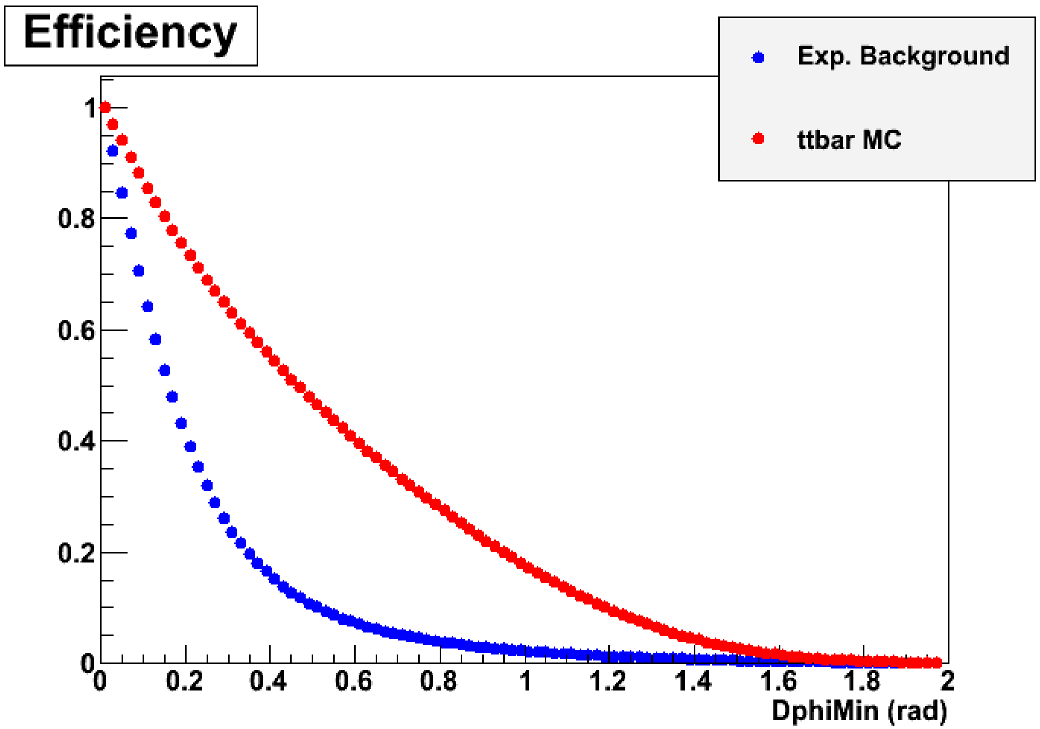
$$N_{exp} = \sum_{i}^{N_{evts}} \sum_{j}^{N_{taggable\ jets}} P(E_T^j, N_{trk}^j, MetProj^j)$$

- Data After preselections only, sample dominated by background
- Iterative correction applied in all bins to take into account top contamination in data
- Stat. Errors only



In order to further clean up our sample, we remove events with high MET and low angle between Jets and MET (which are mainly due to energy mismeasurements and are difficult to model) using an additional cut on DphiMin > 0.4





(Normalized distributions)

Plots show Matrix expected background behaviour versus ttbar MC tagged events.

The chosen DphiMin cut has efficiency **53.5% on MC** and **21.3% on Data** after prerequisites.

Signal Selection

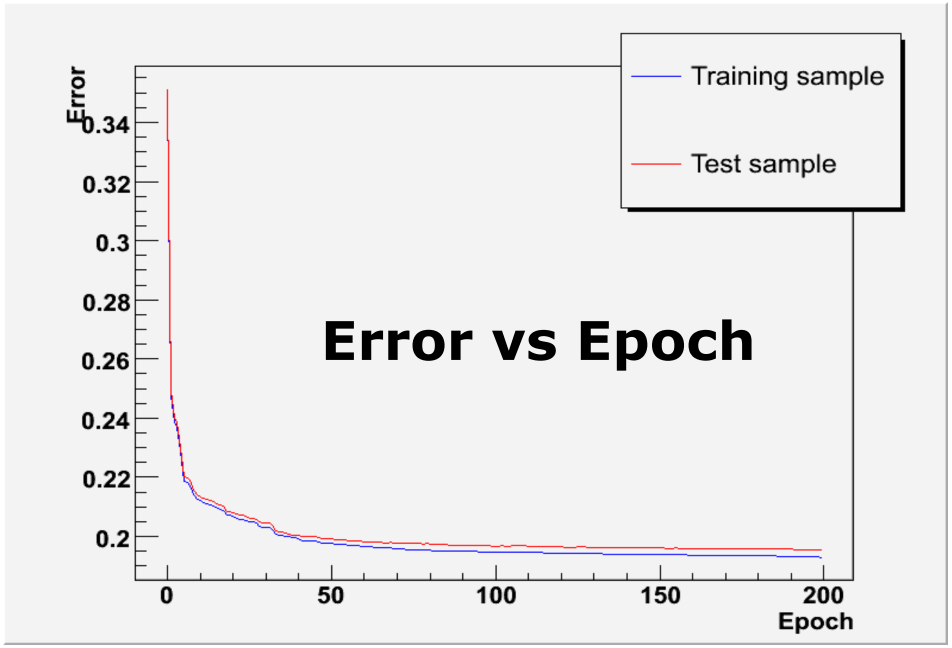
- In order to enhance S/N we use a Neural Network (ROOT TMultiLayerPerceptron) to discriminate signal and background events.
- Training is done on a set containing:
  - All Background (Data) events (~20,000 events, with signal contamination ~3.5%, which we consider negligible for the pourpose of NN training)
  - Equal amount of Signal (MC) events

  ...both passing preselection, DphiMin>0.4 cut and Njets>3

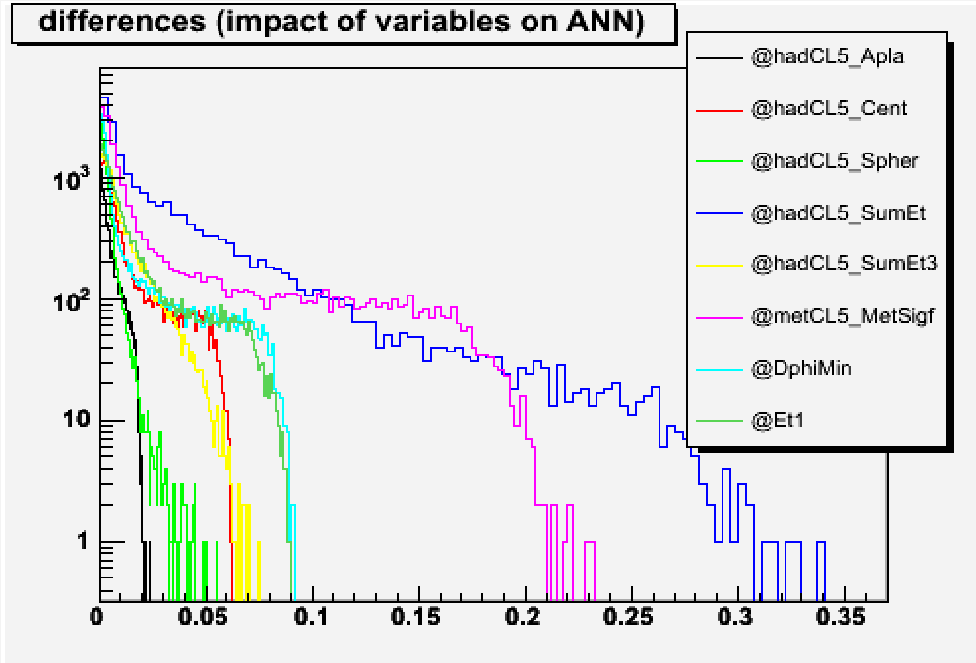
  (half of the set will be used for training, half for test/validation during the training process)
- Training Epochs: ~200, avoid overtraining

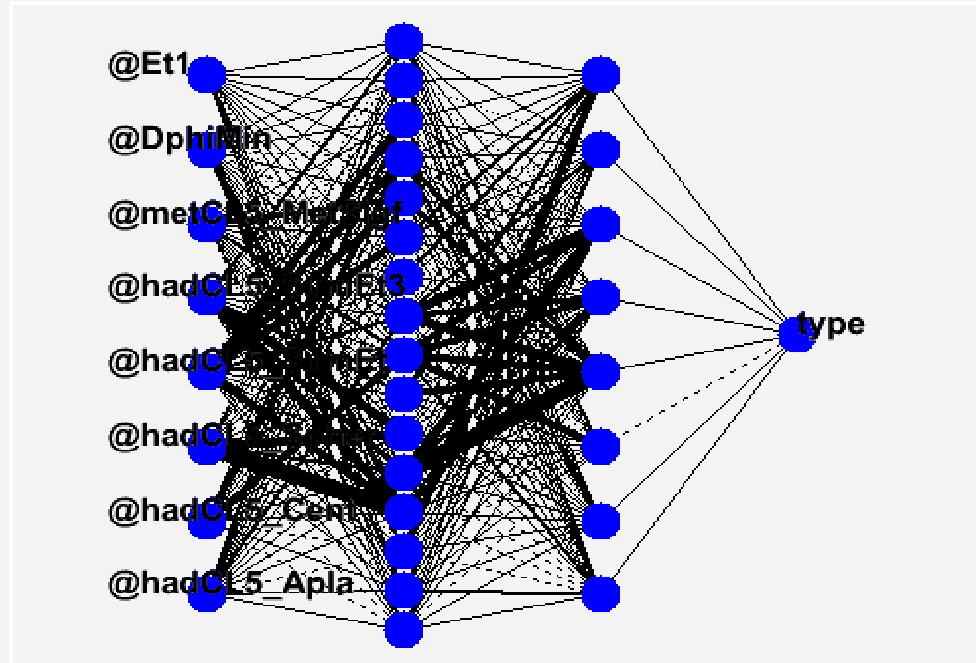
Input Variables

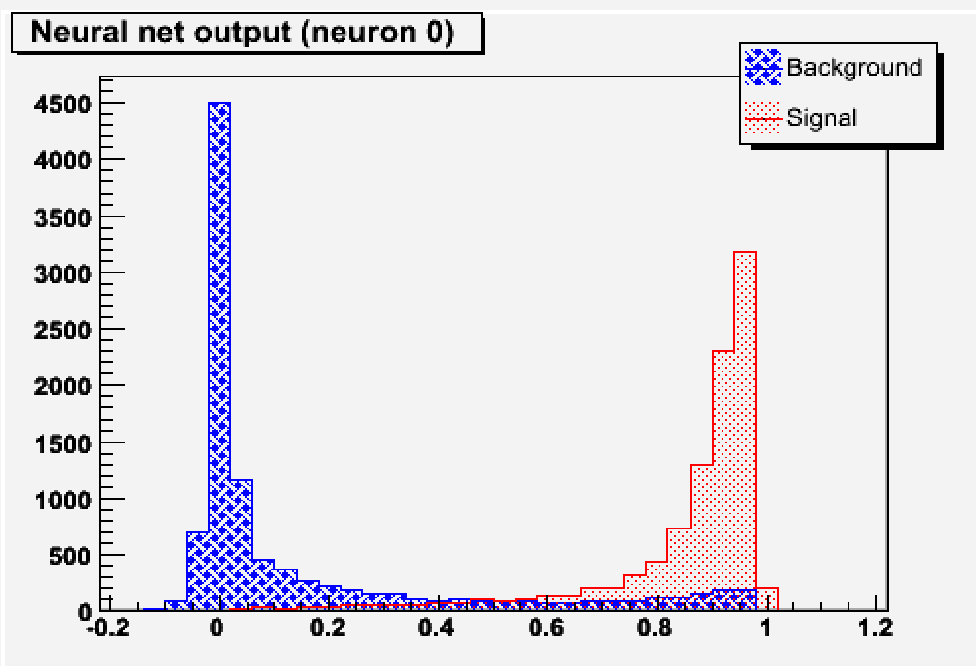
- Et1
- DphiMin
- MET Significance
- SumEt
- SumEt3
- Sphericity
- Centrality
- Aplanarity

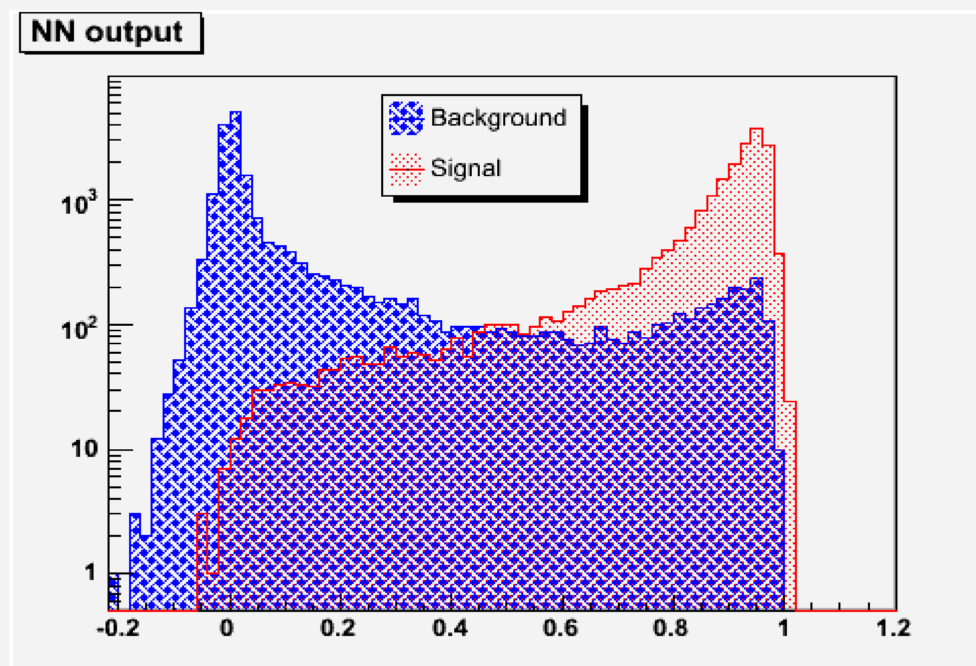


Neural Network Training



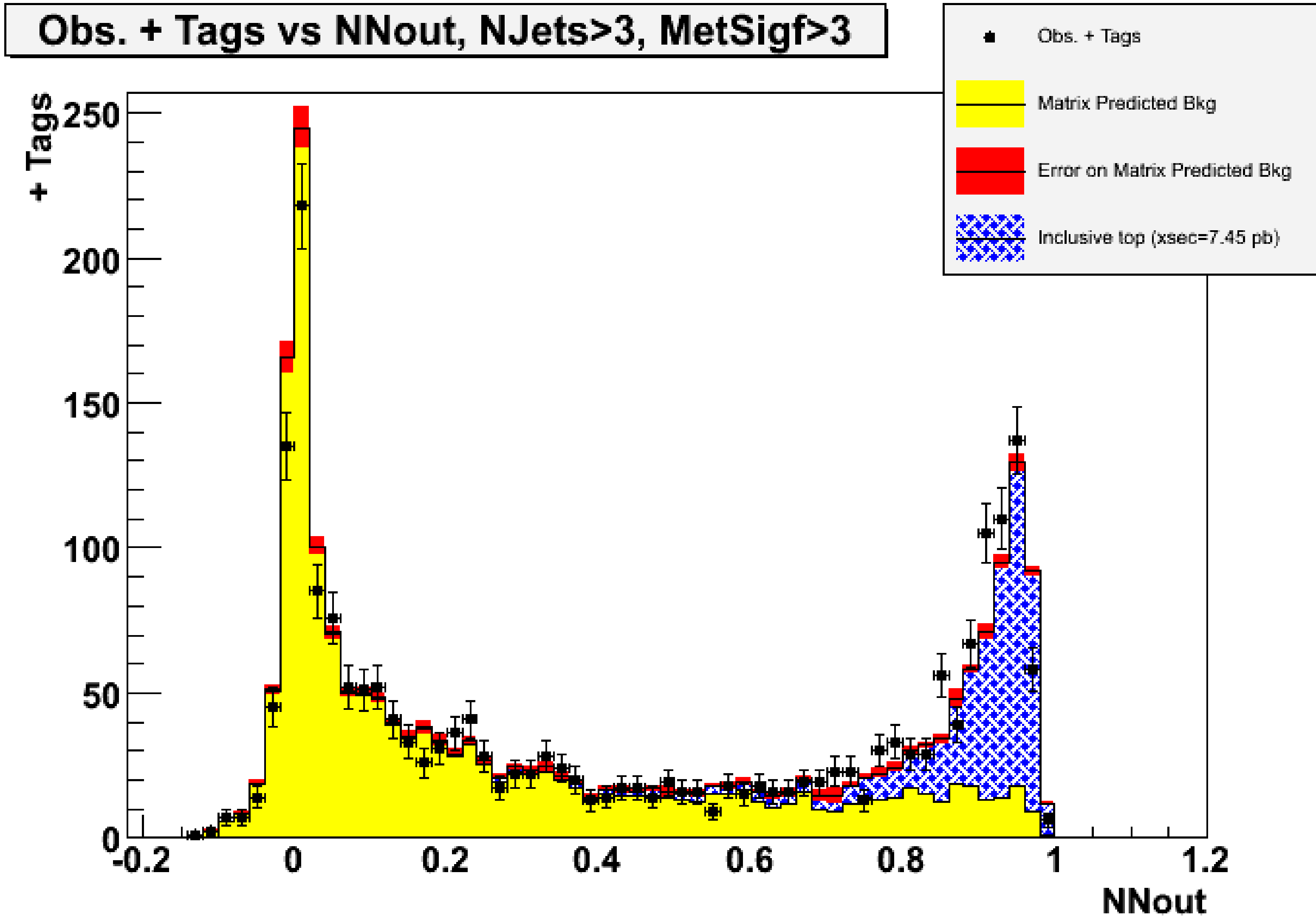






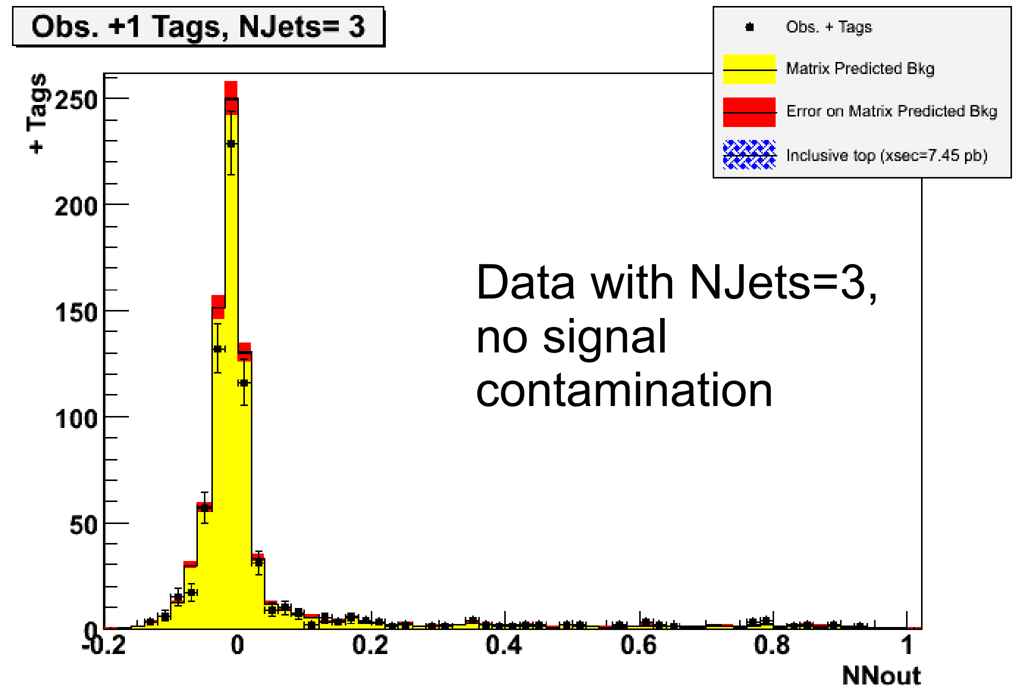
Obs +Tags vs NNout after preselection cuts

- Data After preselections and DphiMin cut, Njets>3
- Iterative correction applied in all bins to take into account top contamination in data
- Stat. Errors only

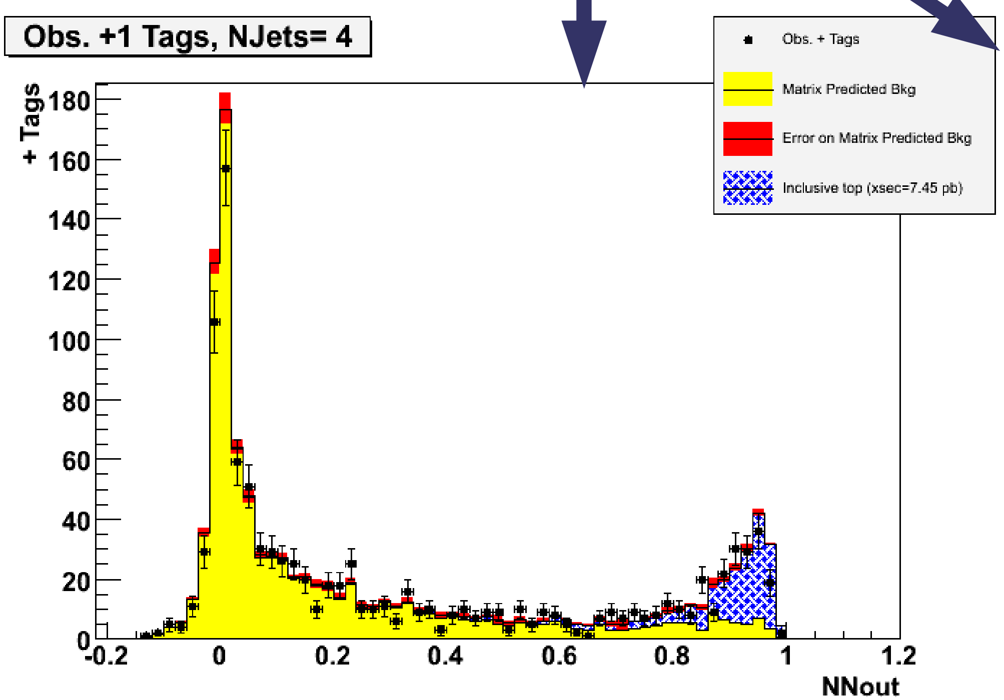


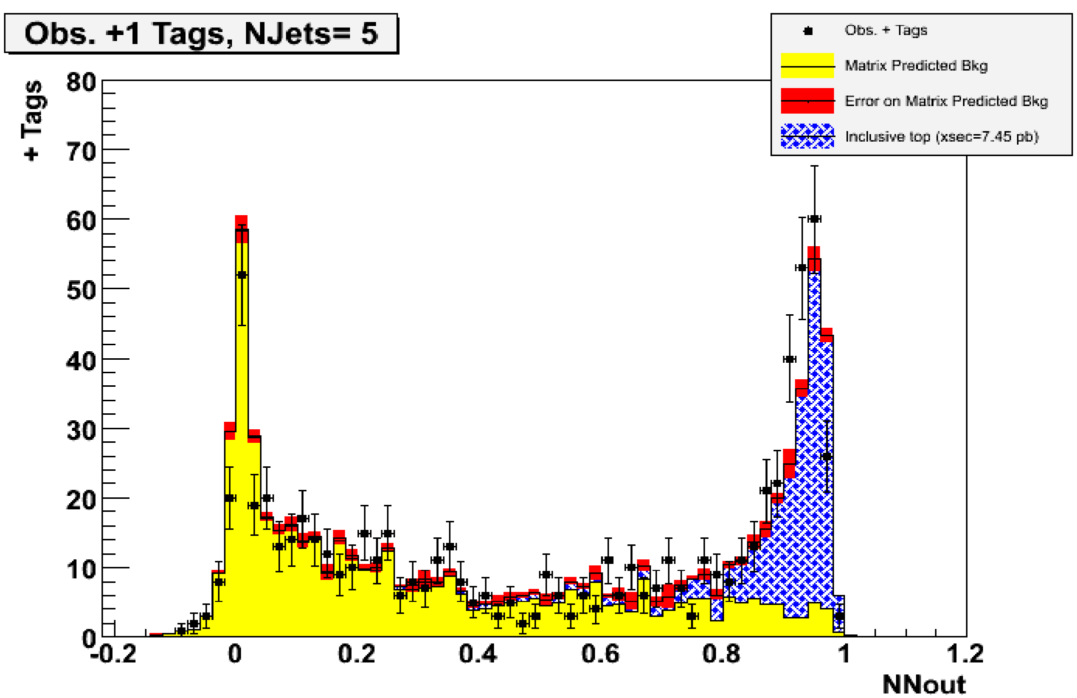
Network output

Matrix prediction reproduces correctly the number of +Tags observed in the 3 Jets sample



Matrix predictions for Njets=4 and 5 are consistent with bkg+signal





We evaluate all sources of systematics as a function of the applied cut on the output of the Neural Network.

Systematics on background determination:
Control sample

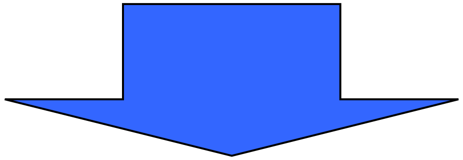
$$\sigma_{ttbar} = \frac{N_{obs}^{tag} - N_{exp}^{tag}}{\varepsilon_{kin} \cdot \varepsilon_{tag}^{ave} \cdot L}$$

Systematics on luminosity determination

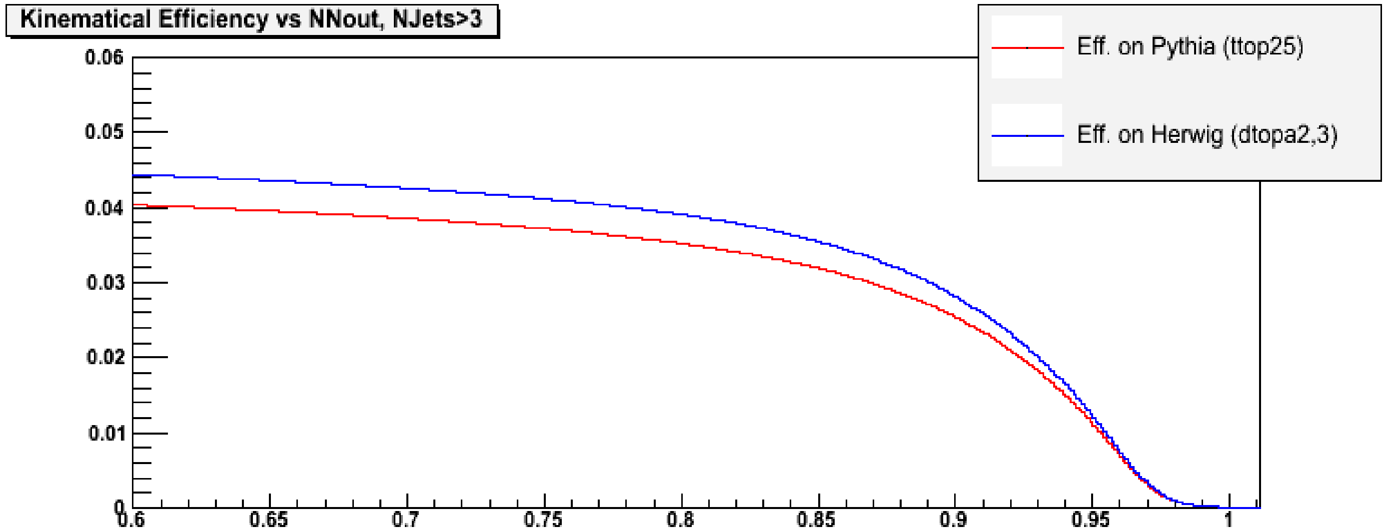
Systematics on kinematical efficiency determination:
MC Gen, PDF, ISR/FSR, trigger, JES, color reconnection

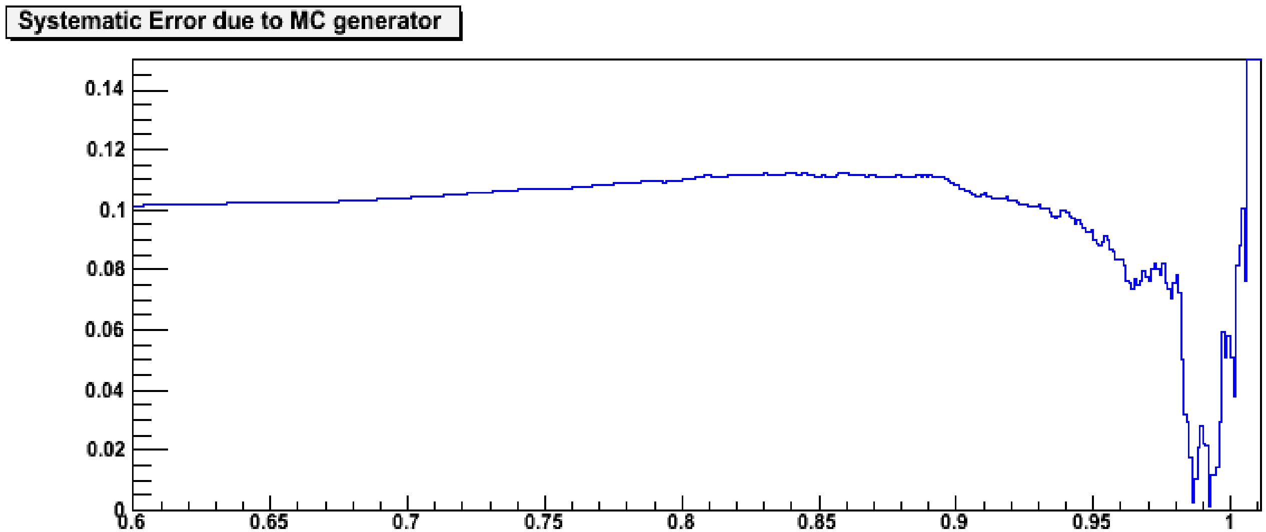
Systematics on average tagging efficiency determination:
SecVtX scale factor

Our main systematics is due to Pythia/Herwig differences, and is almost flat at ~11% for any choice of cut in the range 0.7-0.9.



...we choose to optimize the cut to minimize the statistical error only





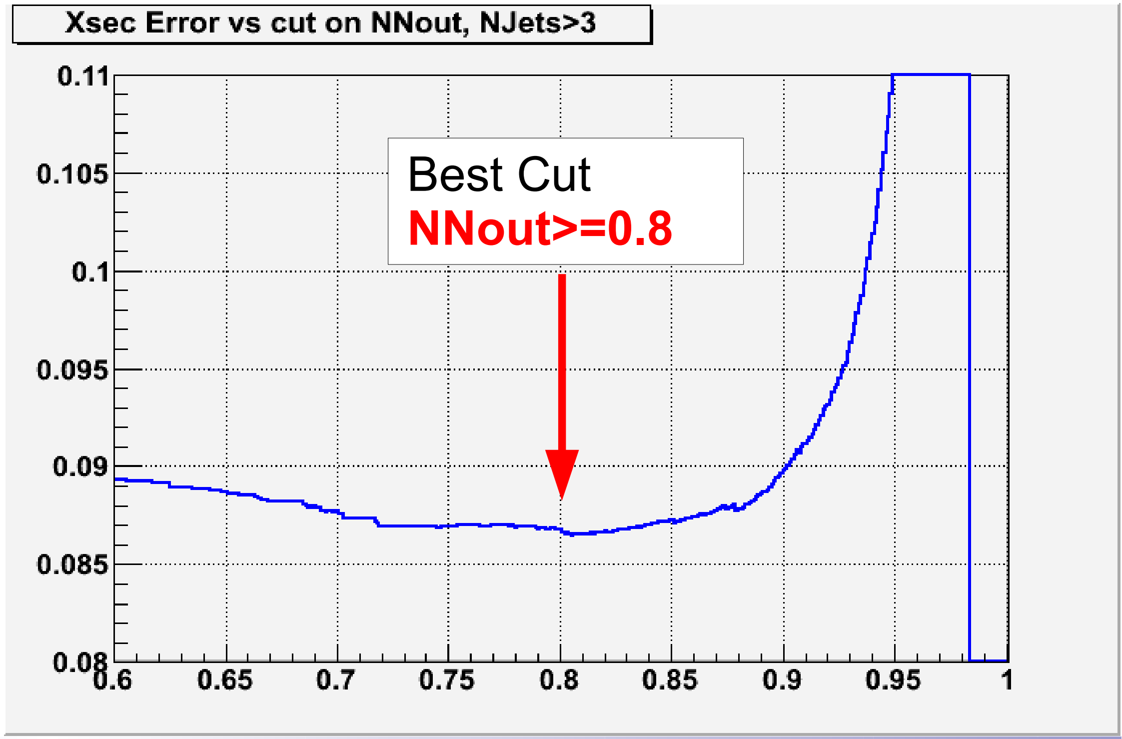
NN Cut optimization

We minimize the relative statistical error on xsec using both the expected amount of tags for signal (from MC) and background (from matrix)

- Start by selecting >3 jets (matrix is computed with = 3jet events)
- Scan NNout cuts
- Calculate the amount of expected bkg tags for a given cut
- Instead of N_{obs}^{tag} use $N^{tag}_{mc} + N^{tag}_{exp}$

$$\sigma_{ttbar} = \frac{N_{obs}^{tag} - N_{exp}^{tag}}{\varepsilon_{kin} \cdot \varepsilon_{tag}^{ave} \cdot L}$$

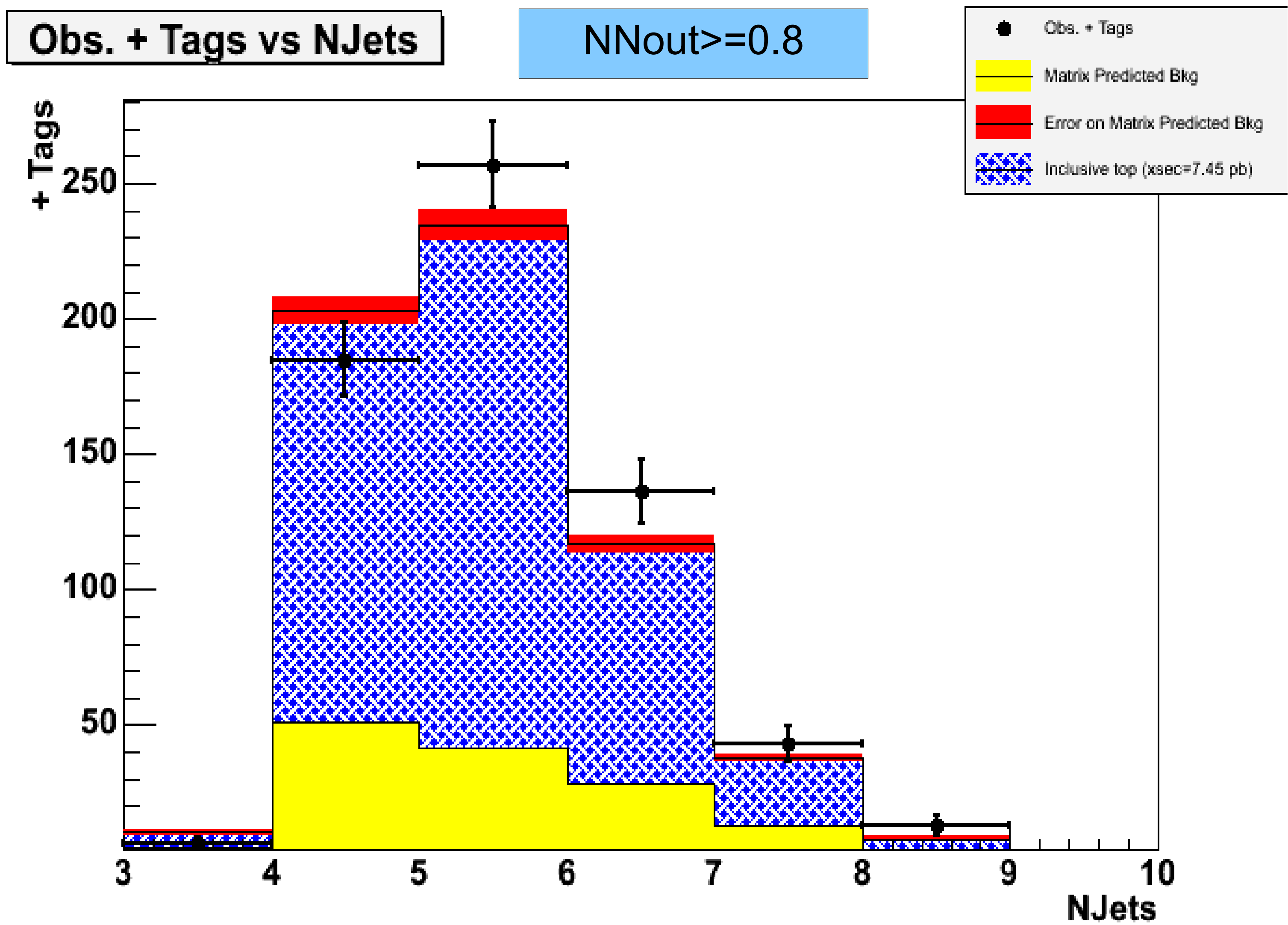
Choose the cut that minimizes the expected (stat. only) relative error on xsec



NN selection: Results

Data After NNout>=0.8
Iterative correction applied in all bins to take into account top contamination in data
Stat. Errors only

- **S/B ~ 3.5**
- **Cut based analysis lumi scaled would give S/B ~ 1.5**



Systematics Summary

Source	Method	Uncertainty
ϵ_{kin} systematics		
Generator dependence	$\frac{\lvert\epsilon_{PYTHIA}-\epsilon_{HERWIG}\rvert}{\epsilon_{PYTHIA}}$	11 %
PDFs	MC reweighting	1.2 %
ISR/FSR	samples comparison	2.7 %
Color Reconnection	samples comparison	2.3 %
Jet Energy Scale	$\frac{\lvert\epsilon_{jetcorr,+1\sigma}-\epsilon_{jetcorr,-1\sigma}\rvert}{2\epsilon_{kin}}$	4.2 %
Trigger simulation	turn-on curves	5 %
Primary Vertex $Z0$		0.2 %
ϵ_{tag} systematics		
SecVtX scale factor	$\frac{\lvert\epsilon_{tag,+1\sigma}-\epsilon_{tag,-1\sigma}\rvert}{2\epsilon_{tag}}$	3.9 %
Tagging matrix systematics		
Data control sample	N_{obs}/N_{exp}	2.5 %
Luminosity systematics		
Luminosity measurement	–	6.0 %

Once all systematics errors have been accounted for, the cross section can then be calculated maximizing a likelihood function whose input parameters are subject to gaussian constraints

$$\mathcal{L} = e^{-\frac{(L-\bar{L})^2}{2\sigma_L^2}} \cdot e^{-\frac{(\epsilon_{kin}-\bar{\epsilon}_{kin})^2}{2\sigma^2_{\epsilon_{kin}}}} \cdot e^{-\frac{(\epsilon^{ave}_{tag}-\bar{\epsilon}^{ave}_{tag})^2}{2\sigma^2_{\epsilon^{ave}_{tag}}}} \cdot e^{-\frac{(N_{exp}-\bar{N}_{exp})^2}{2\sigma^2_{N_{exp}}}} \cdot \frac{(\sigma_{t\bar{t}} \cdot \epsilon_{kin} \cdot \epsilon^{ave}_{tag} \cdot L + N_{exp})^{N_{obs}}}{N_{obs}!} \cdot e^{-(\sigma_{t\bar{t}} \cdot \epsilon_{kin} \cdot \epsilon^{ave}_{tag} \cdot L + N_{exp})}$$

$$\sigma_{ttbar} = \frac{N^{tag}_{obs} - N^{tag}_{exp}}{\varepsilon_{kin} \cdot \varepsilon^{ave}_{tag} \cdot L}$$

Variable	Symbol	Input Value	Output Value
Integrated Luminosity (pb^{-1})	$\mathcal{L}$	2207.5 ± 132.5	2212.1 ± 133.7
Observed Tags	N_{obs}	636	–
Expected Tags	N^{corr}_{exp}	131 ± 9.6	130.6 ± 9.6
Kin. efficiency (%)	ϵ_{kin}	3.53 ± 0.47	3.51 ± 0.52
Ave. b-tagging efficiency	ϵ^{ave}_{tag}	0.811 ± 0.032	0.811 ± 0.032

...giving a cross section measurement of: (assuming M_{top} = 172.5 GeV/c^2)

$$\sigma_{t\bar{t}} = 8.02 \pm 0.42\,(stat) \pm 1.38\,(syst)\; pb$$
$$= 8.02 \pm 1.44\; pb$$

Overall uncertainty of the measurement ~18%, systematics dominated

Conclusions

- In TOP_MULTI_JET data up to 2.2 fb^-1 using a Neural Network we have selected a top sample with a sizeable contribution of tau+jets events, orthogonal to any other used in the cross section analyses at CDF.
- We measured the ttbar production cross section to be:

$$\begin{aligned}\sigma_{t\bar{t}} &= 8.02 \pm 0.42\,(stat) \pm 1.38\,(syst)\ pb \\ &= 8.02 \pm 1.44\,pb\end{aligned}$$

- Documentation on the analysis is available in **CDF Note 9873**
- **Preblessing this analysis tomorrow...**

- Backup -

TOP_MULTI_JET revisions up to p13 can be placed into 3 major groups:

- L1_JET10 and L2_FOUR_JET15_SUMET125 (V3-4) [P0-P1]
- L1_JET10 and L2_FOUR_JET15_SUMET175 (V5-V8) [P2-P7]
- L1_JET20 and L2_FOUR_JET15_SUMET175 (V9) [P8-P12]

On **DATA** taken before TMJ-v5 we simulate the new L2 requirements
(4 L2 Clusters with Et>15GeV and SumEt@L2 >175 GeV)

On **MC** events:
Full simulation of the trigger path is performed:
Previous studies have shown that if L2 is fired, then L1 and L3 are 99% efficient.

L2: we simulate the trigger requirements using Scale Factors developed by A.Mitra to correct the simulation of L2 Cluster Energies in the MC.

L1: Additionally, to cope with the L1_JET20 requirement for p8 and later, we derive a data driven **L1&&L2/L2** turnon rate from Tower10 to reweigh the corresponding MC events.

L3: We perform the simulation of the L3 requirements.

Pre-tag iterative top subtraction

We need to correct the tagging matrix prediction in order to account for the ttbar presence in the pre-tagging sample by using an iterative method:

$$N'_{\exp} = N^{fix}_{\exp} \frac{N_{evt} - N^{ttbar}_{evt}}{N_{evt}} = N^{fix}_{\exp} \frac{N_{evt} - \dfrac{N_{obs} - N_{\exp}}{\varepsilon^{ave}_{tag}}}{N_{evt}}$$

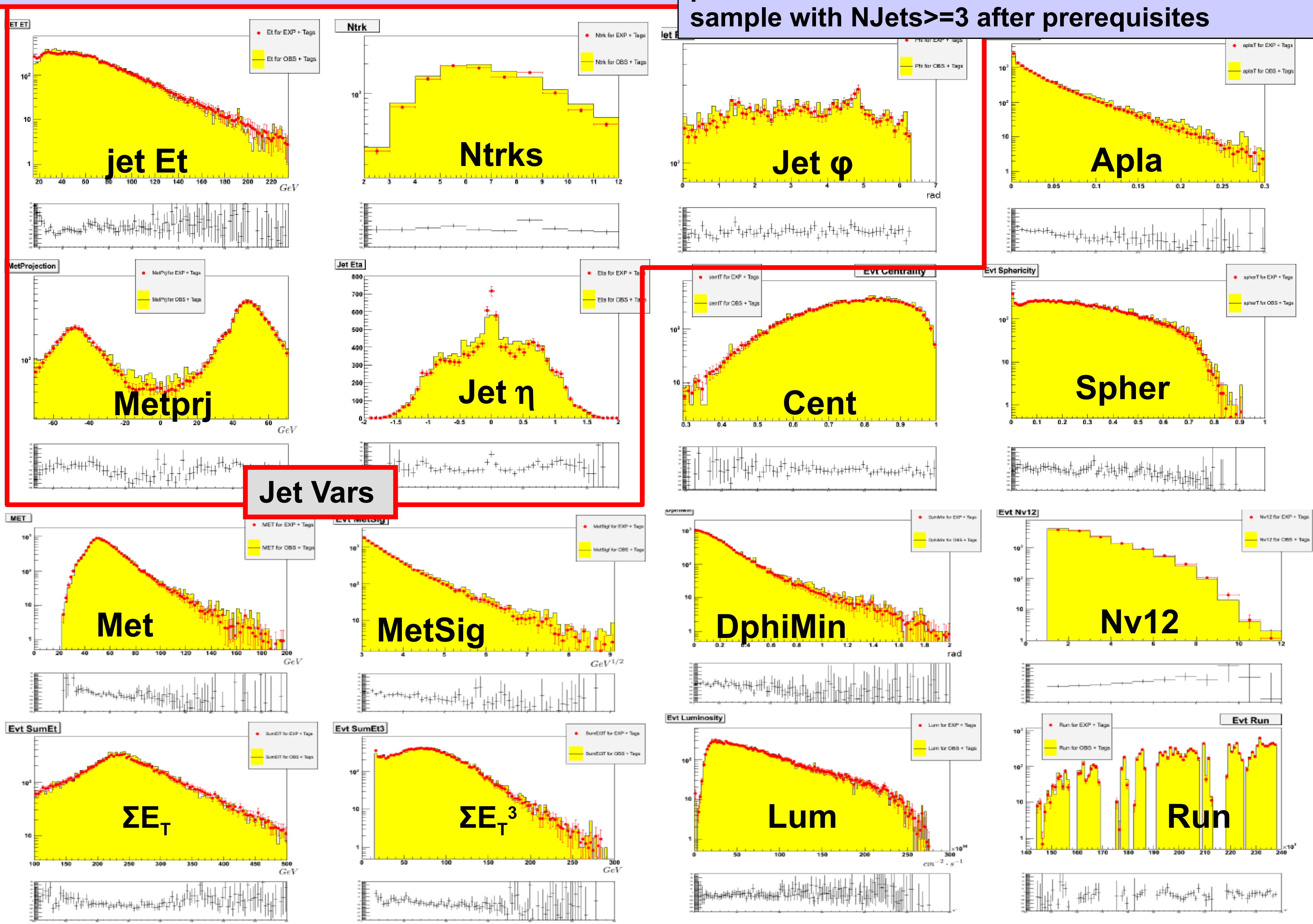
The procedure stops when $|N_{exp}' - N_{exp}| < 1\%$

"top ad BKG" correction (apply matrix to MC, subtract contribution to exp tags predicted on the whole sample):

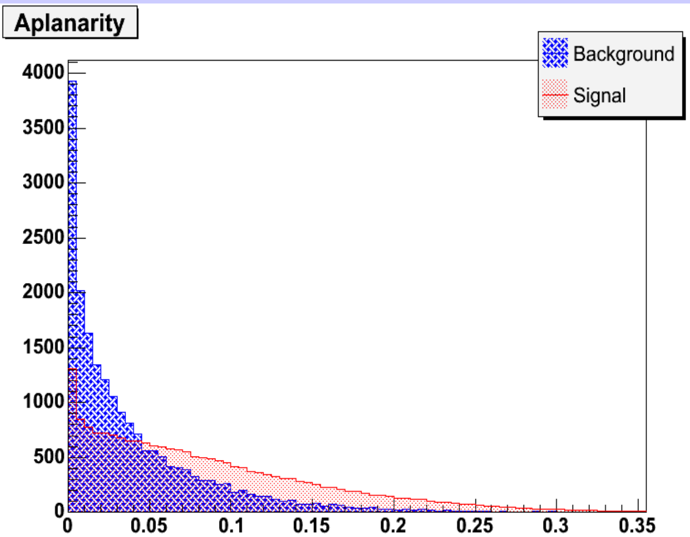
$$N'_{exp} = N_{exp} - \sigma_{t\bar{t}} \epsilon_k t_b L$$

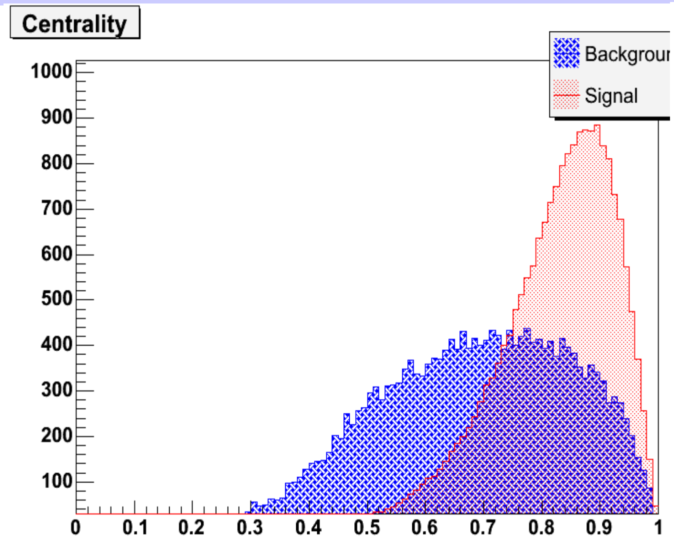
Matrix Checks

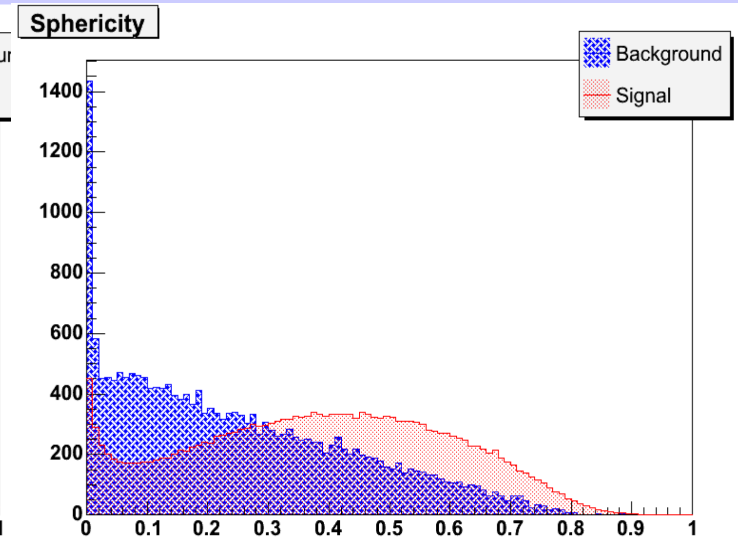
The tag rate parameterization allows to correctly predict kinematical distributions in the data sample with NJets>=3 after prerequisites

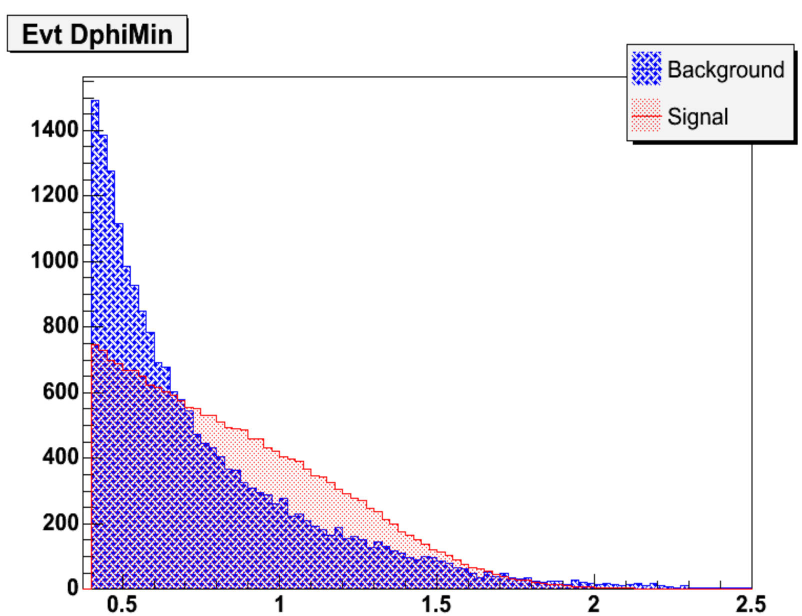


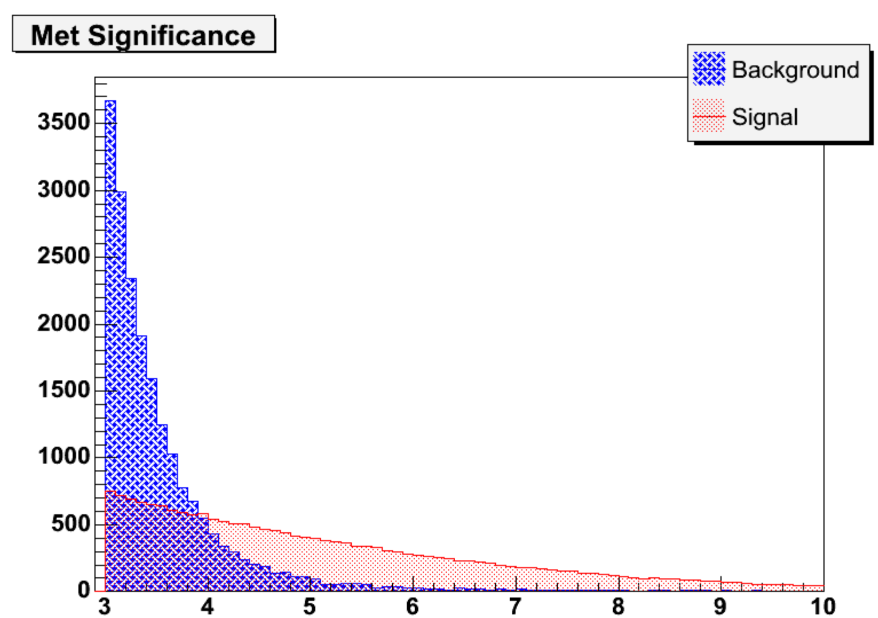
Neural Network Inputs 1/2



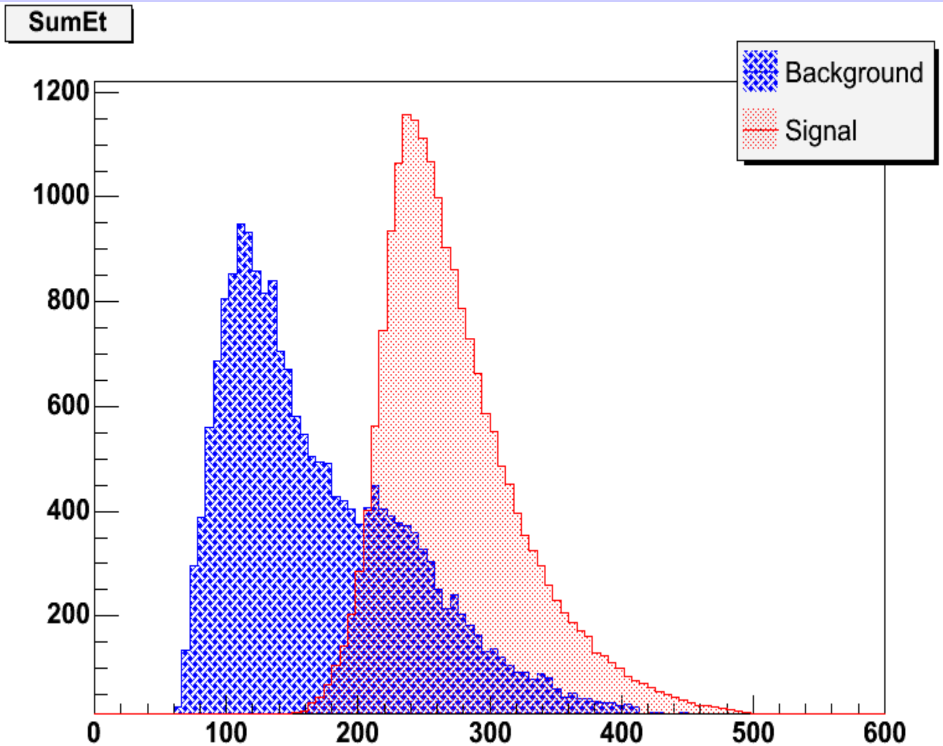


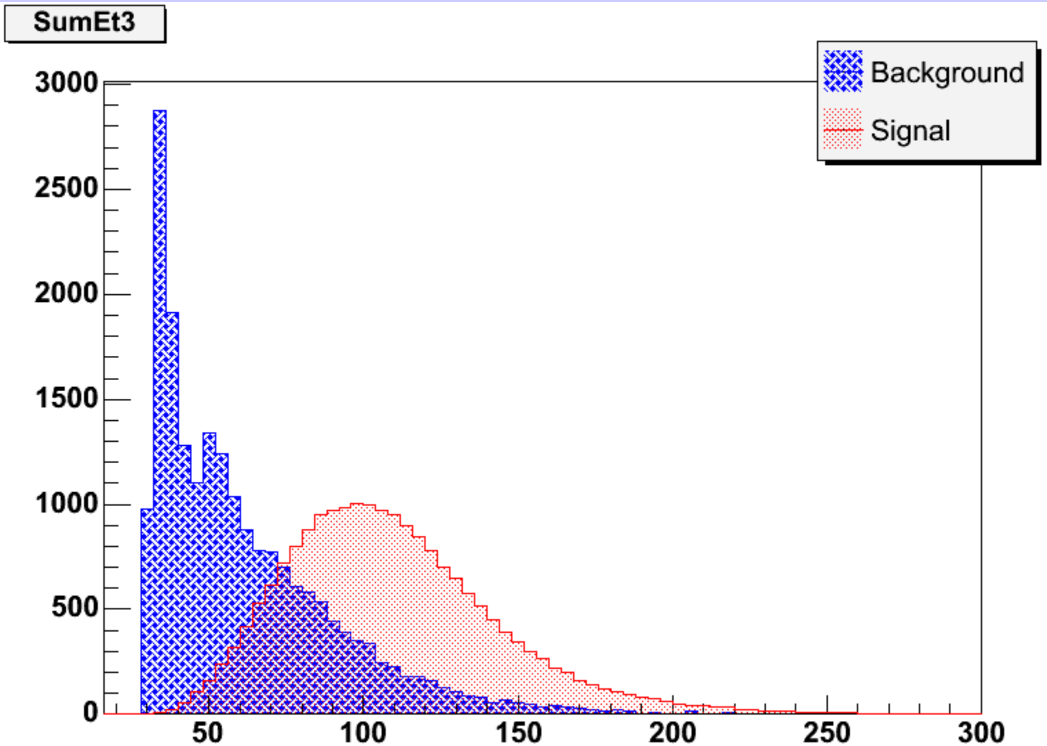


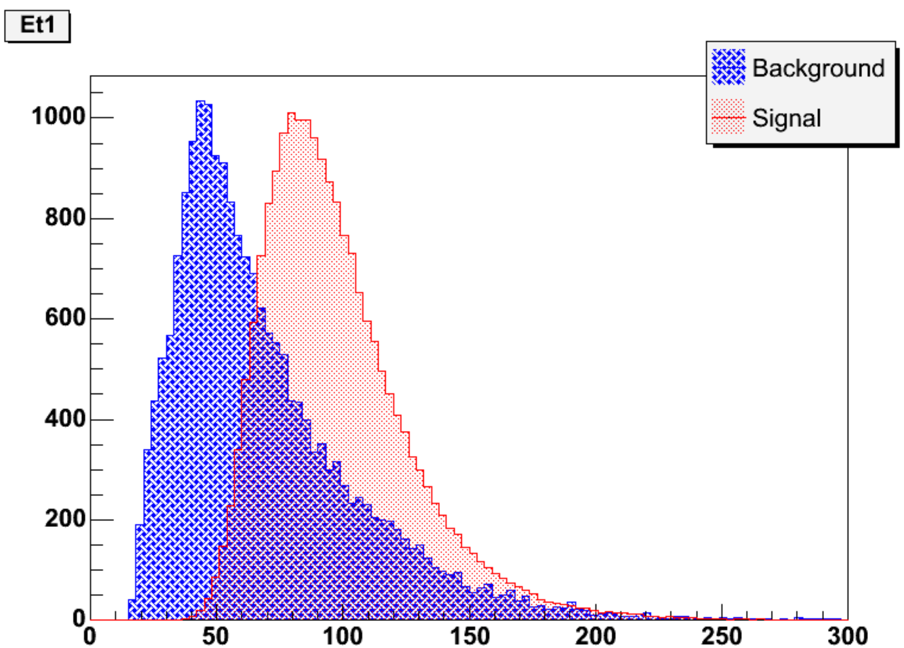


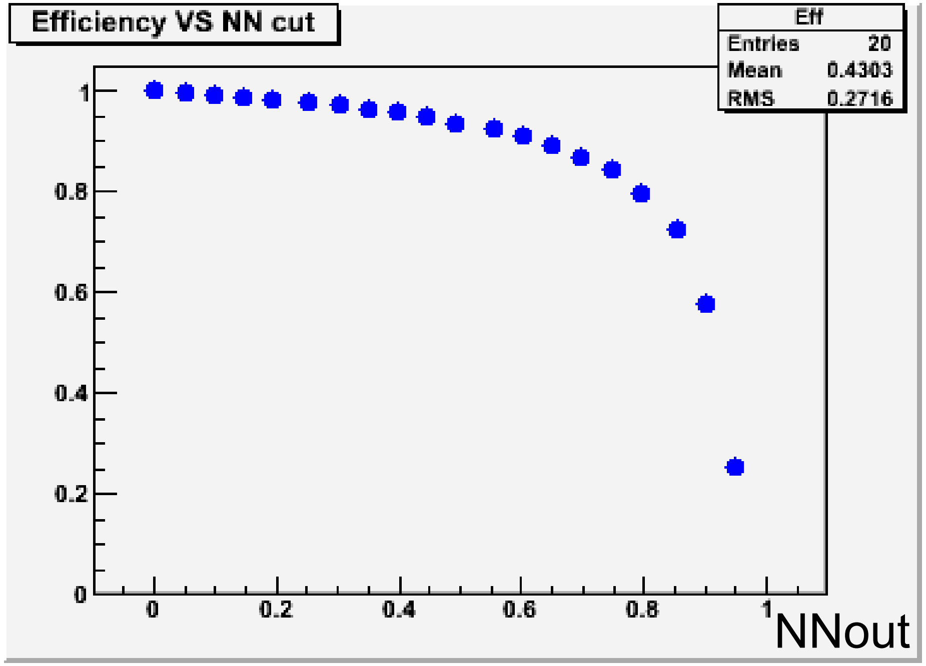


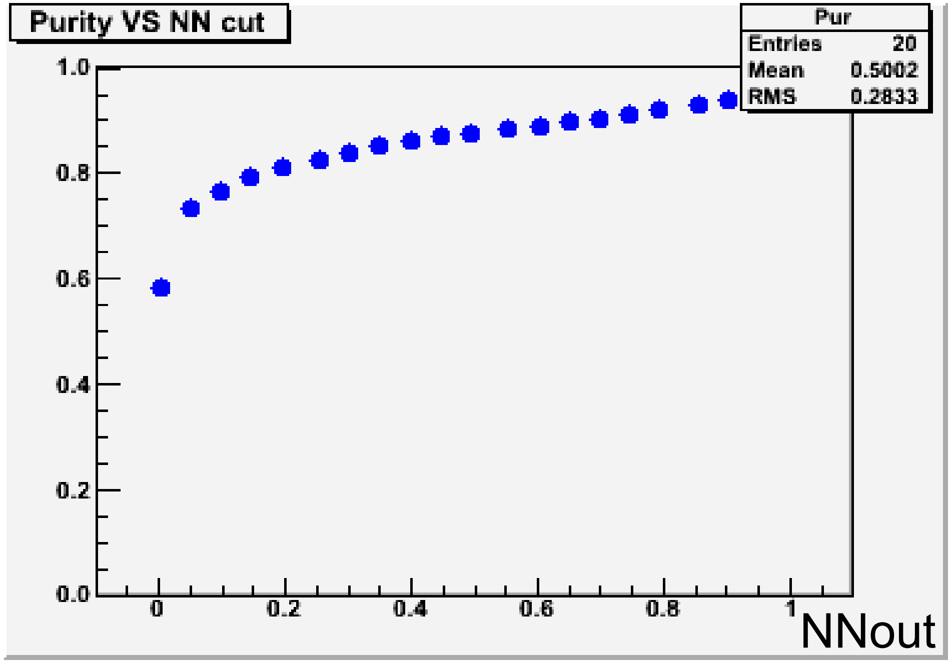
Neural Network Inputs 2/2



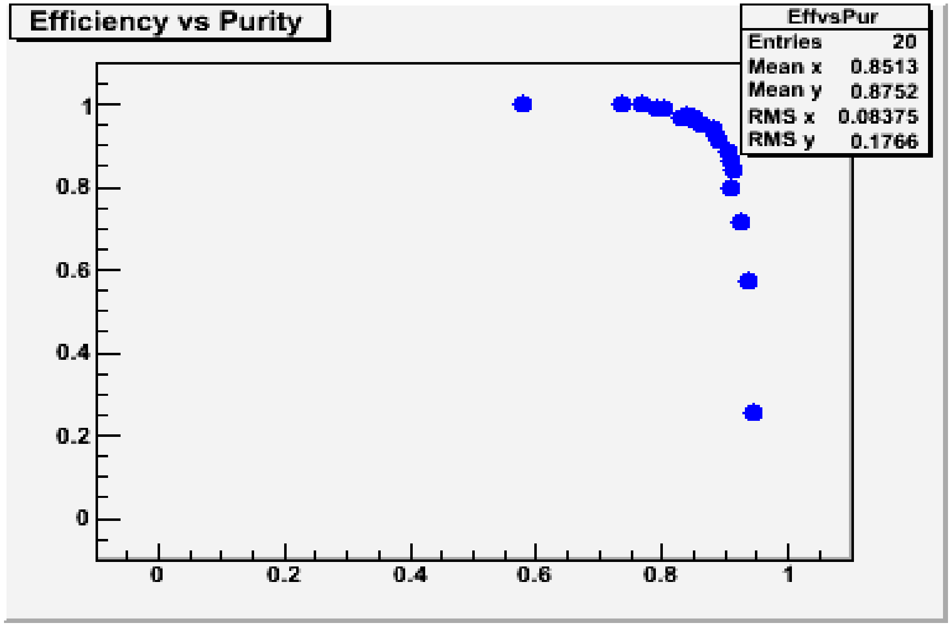








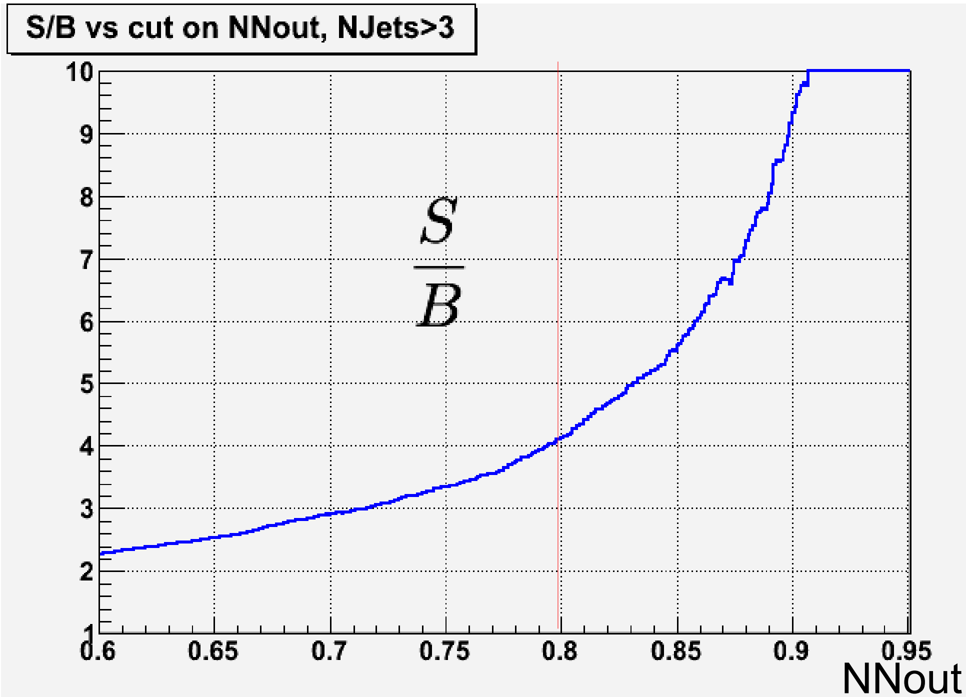
(Test Sample has same number of signal and background events)

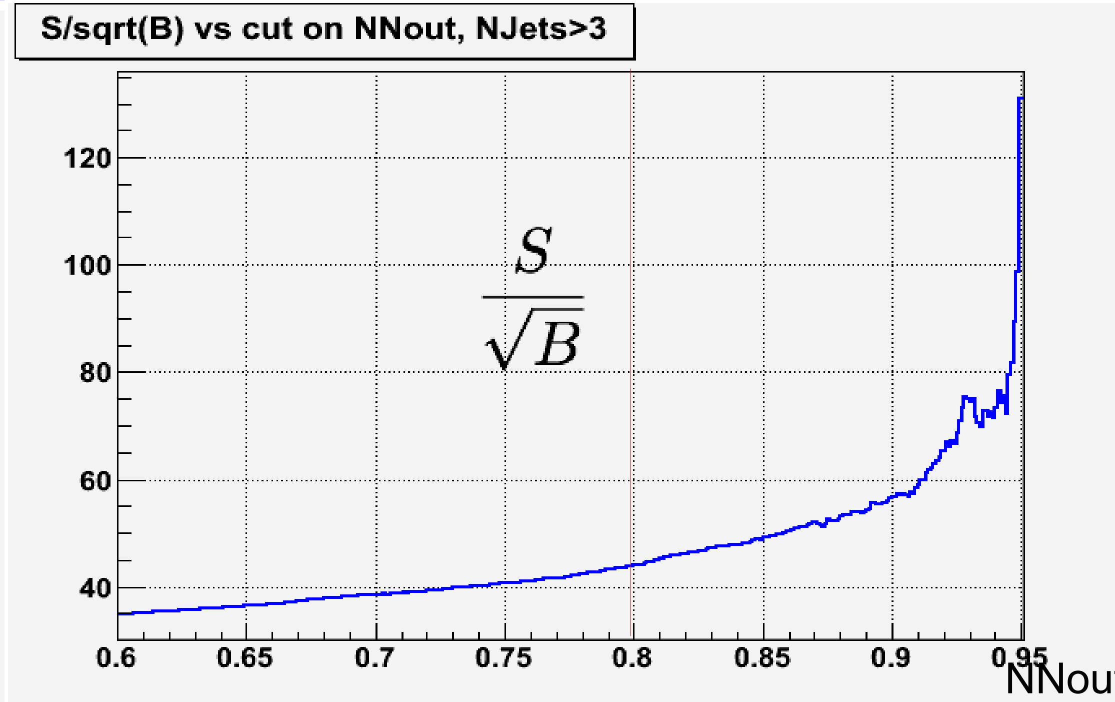


Eff (cut) =
Signal Events passing cut /
Total Signal Events

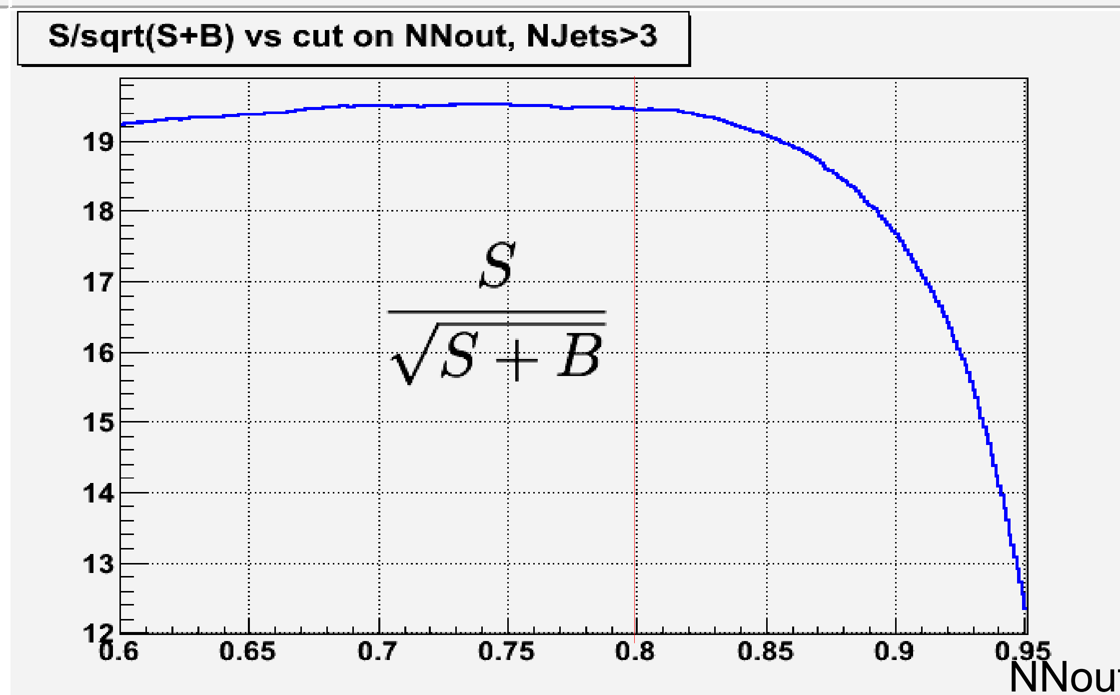
Pur (cut) =
Signal Events passing cut /
Total Events passing cut

NN Cut optimization



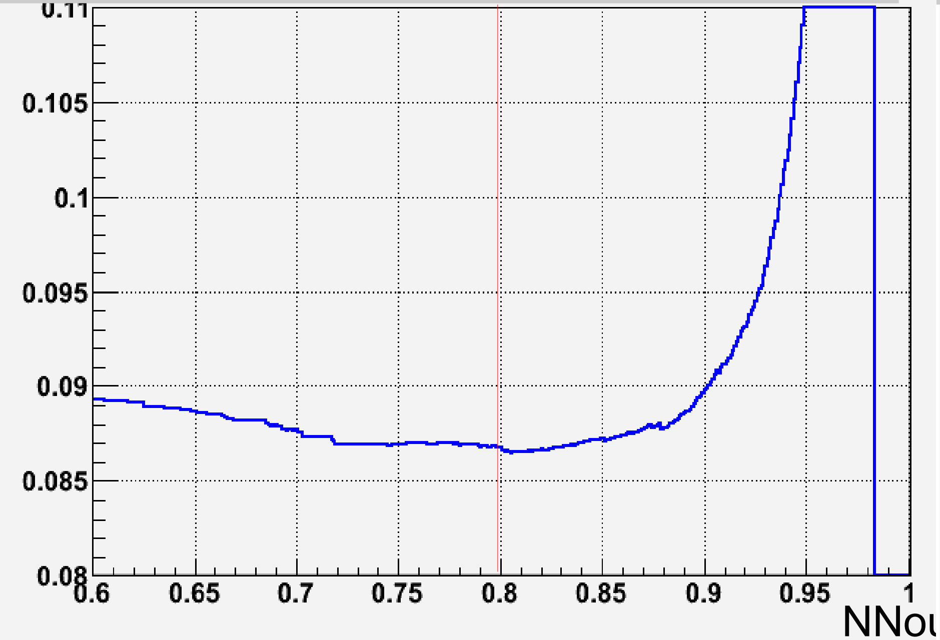


Expected values using MC and tag Matrix to derive the number of Observed tags in data

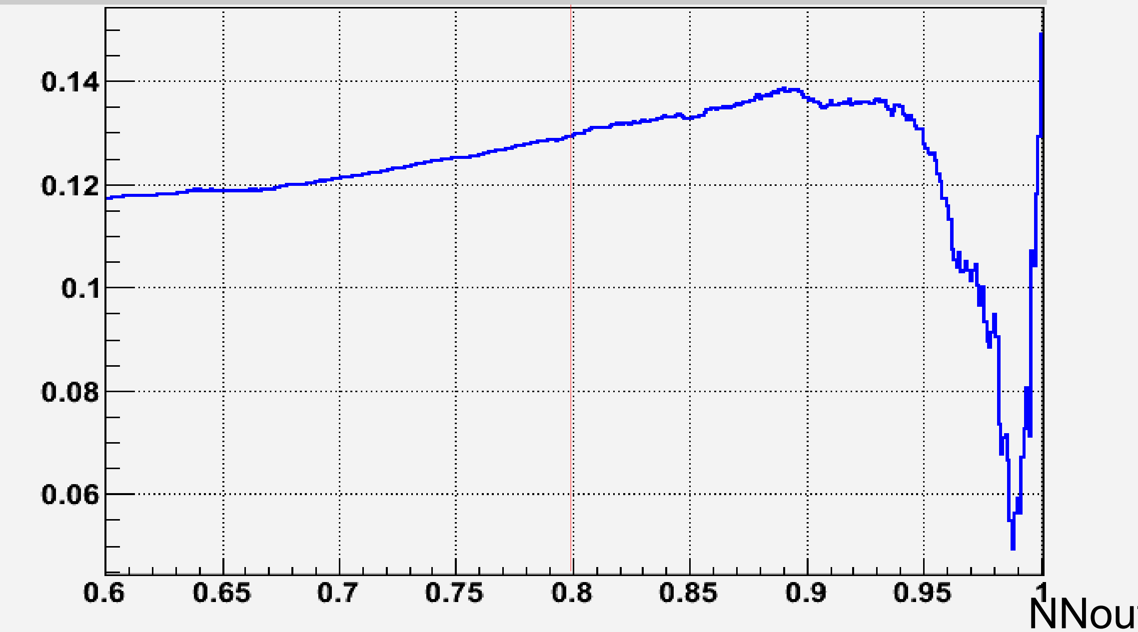


Expected Error

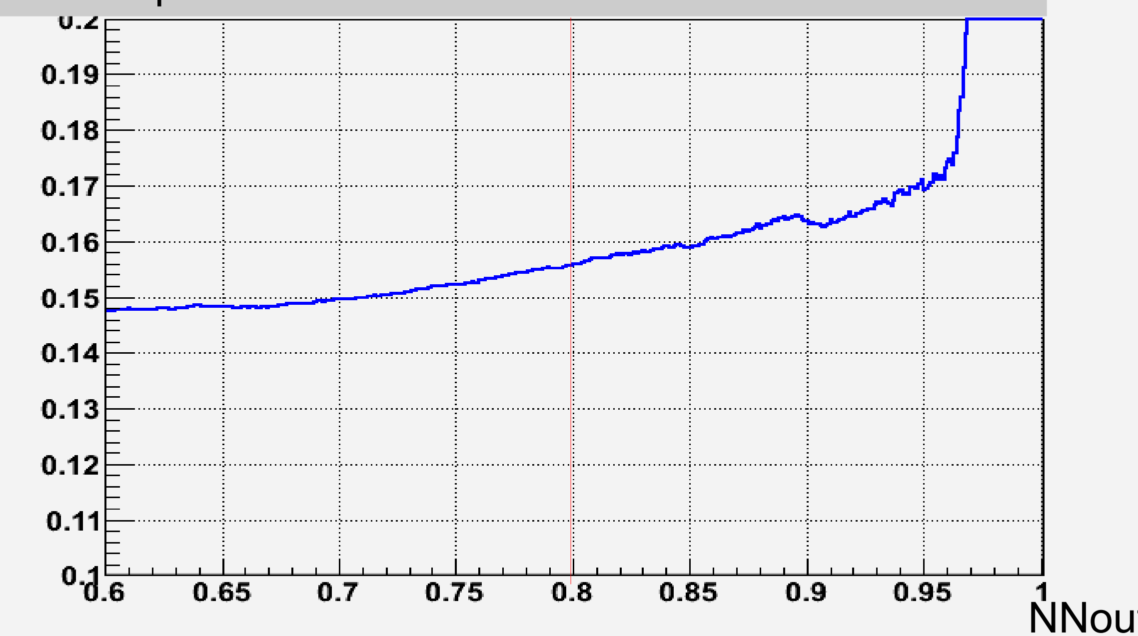
Exp. Stat. Error



Total Systematic Error, excluding tag matrix syst



Total expected Error on cross section



Expected Stat. Error is evaluated using MC and tag Matrix to derive the number of Observed tags in data.