Boosted WW cross – section in the semileptonic channel

PPSMC

12th of December 2014

Fabian Spettel

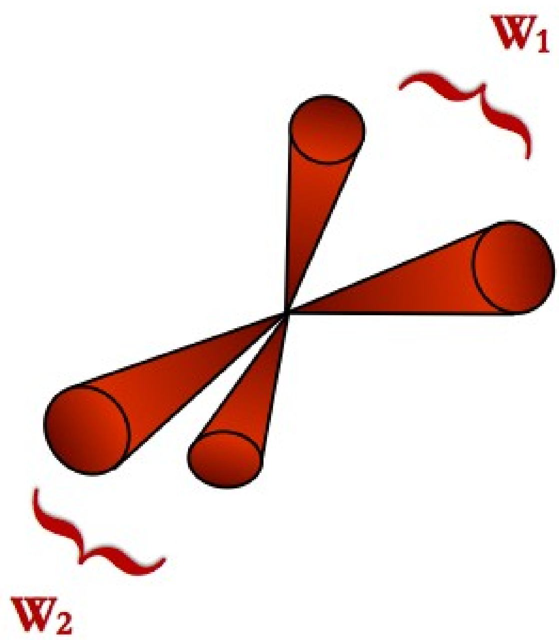
Max-Planck-Institut für Physik
(Werner-Heisenberg-Institut)

ATLAS EXPERIMENT
http://atlas.ch

אוניברסיטת תל-אביב
TEL AVIV UNIVERSITY

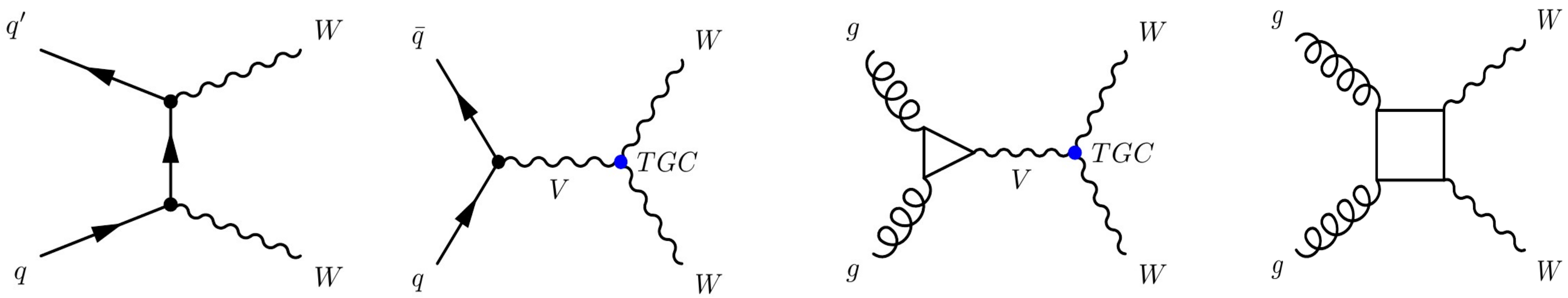
Outline

1. Why measure “boring“ SM cross-section?
2. Boosted objects and jet - substructure
3. Multivariate Methods



Motivation

- WW - production sensitive to anomalous triple gauge boson couplings (aTGCs)



- Important test for SM
- Deviations from SM become more enhanced with increasing CM energy
- Mild excess in σ found by ATLAS and CMS
- Dominant background for $H \rightarrow W W^*$
- With higher CM energy, boosted topologies become much more important
- Lately a large number of jet substructure methods on the market
- Can be nicely used to discriminate S/B

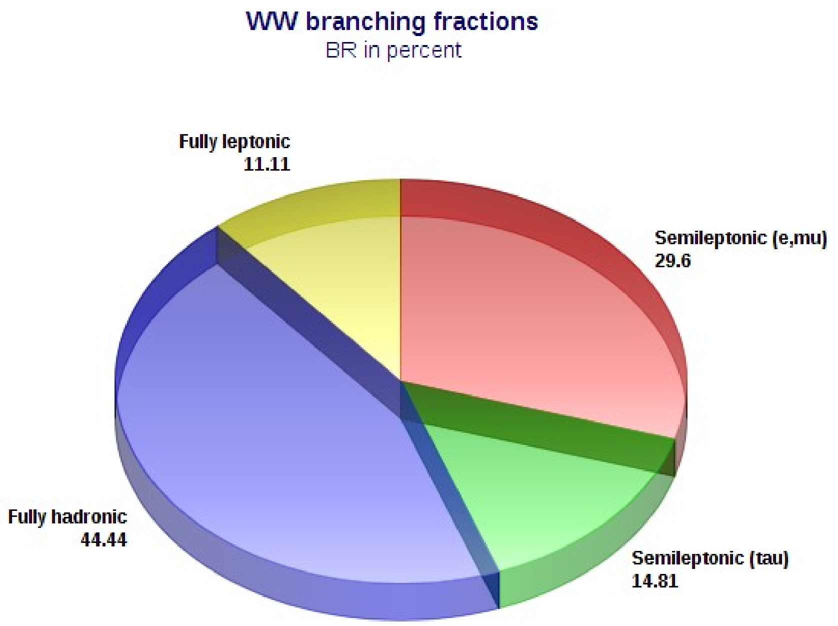
Motivation

- Semileptonic channel has a high branching ratio

$$BR\,(WW \rightarrow l\nu l\nu) \approx 5\%, \quad (l = e, \mu)$$

$$BR\,(WW \rightarrow jj l\nu) \approx 29\%, \quad (l = e, \mu)$$

→ Gain factor 6 in statistics



- Full hadronic channel would allow full control over WW system
- Already attempted → Impossible!
- Need hard lepton to trigger the event and suppress QCD - multijet events
- **<u>Downside:</u>**

  Detector resolution too low to distinguish jets from W and Z

  - Have to measure combined WW+WZ cross - section
  - In turn gives better sensitivity to aTGCs (arXiv:1410.7238v1)

ATLAS @ LHC

2012 Data: $\sqrt{s}=8$ TeV, $\int \mathscr{L}\, dt = 21\,\mathrm{fb}^{-1}$

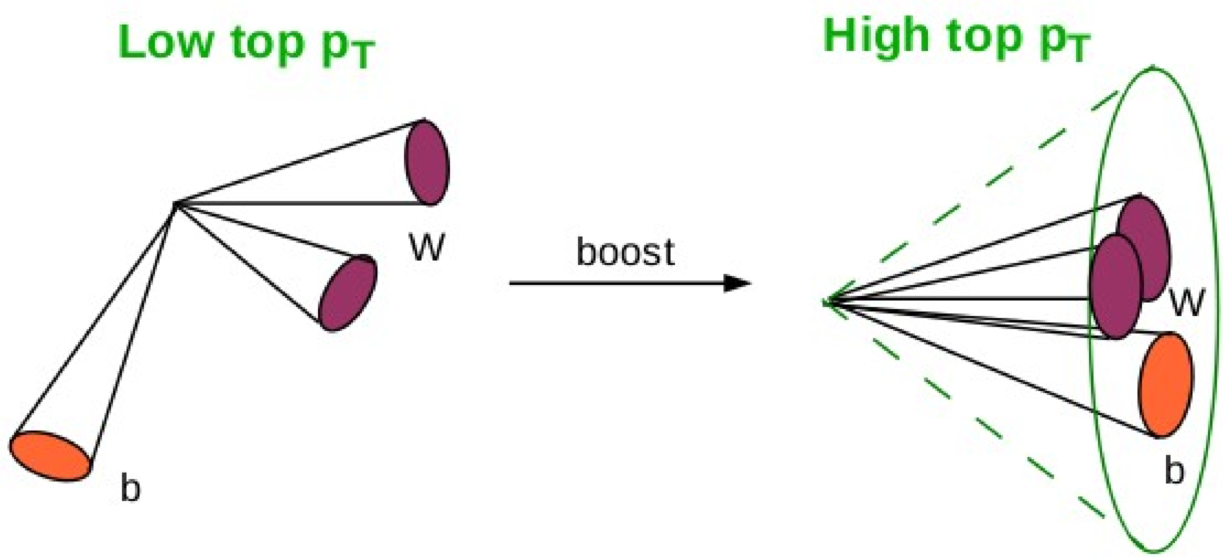
- In the Muonstream alone, there are 725M recorded events, 46k of which are WW
- After event selection: Only 500 - 1000 signal events left

2. Boosted topologies and jet - substructure

Why jet substructure?

- **High center of mass energy at the LHC:**
  - Large amount of heavy particles is produced boosted and decaying in a collimated (single jet like) final state
  - Decay products are clustered into one jet with size $R \approx \frac{2m}{p_T}$
  - Final state not resolvable with standard (narrow jet) techniques anymore

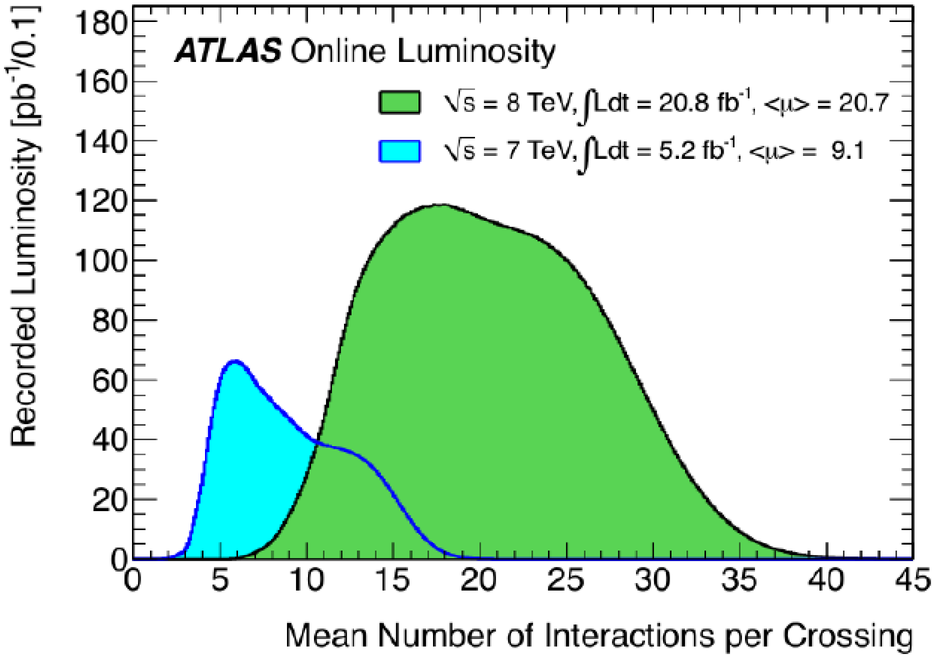
  $\rightarrow$ Go to "fat jets"



  - Fat jet mass is an important variable to identify decayed particles

Why jet substructure

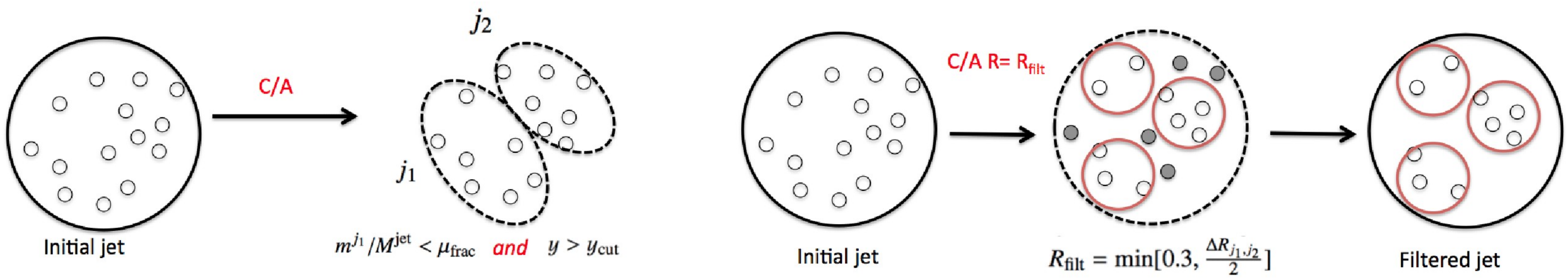
- **High luminosity:**
  - Additional pp - collissions per bunch crossing (pile-up) deteriorate jet mass and shape
  - Need technique to separate internal energy flow structures from diffuse pile-up contributions for mass reconstruction



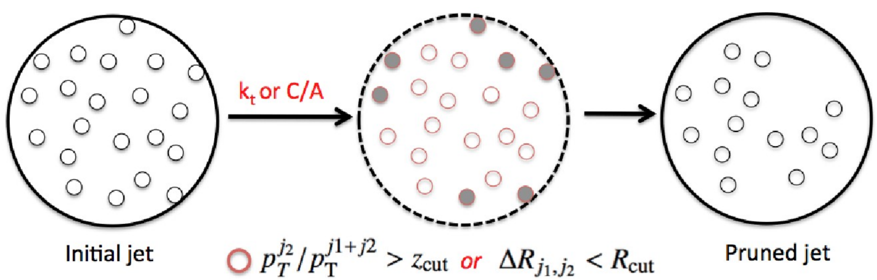
- **Jet grooming:**
  1. Filtering
  2. Pruning
  3. Trimming
- **Jet substructure:**
  - Different techniques/variables to distinguish gluon - jets from from heavy particle - jets

Jet grooming

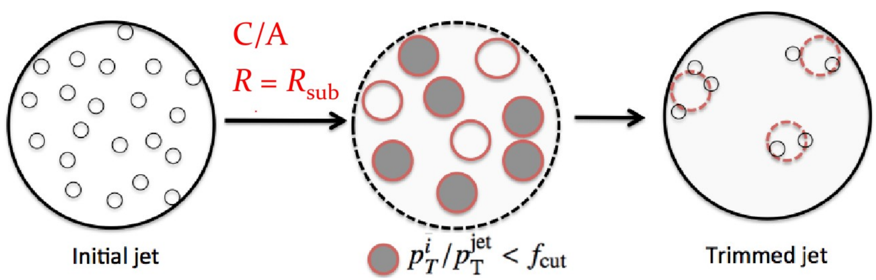
Mass drop tagging plus filtering:

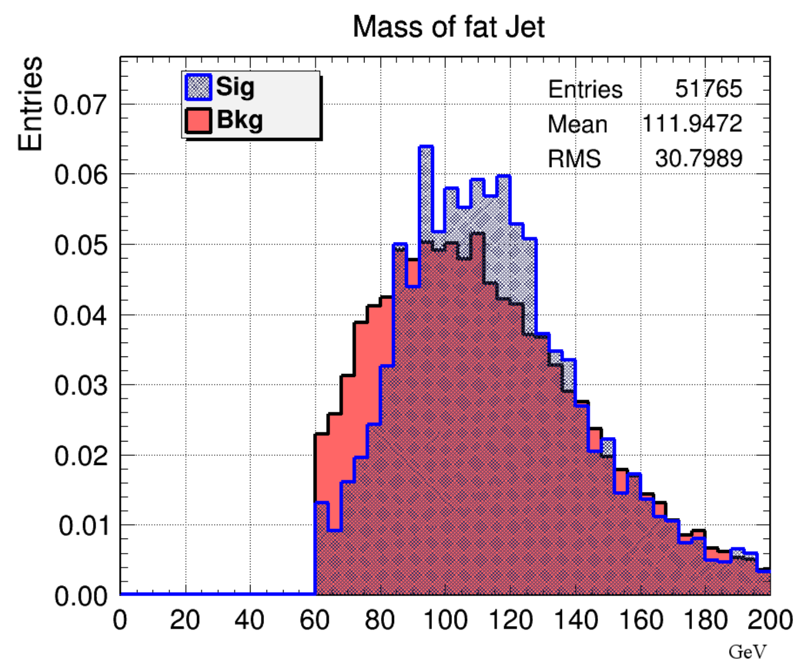


Pruning:



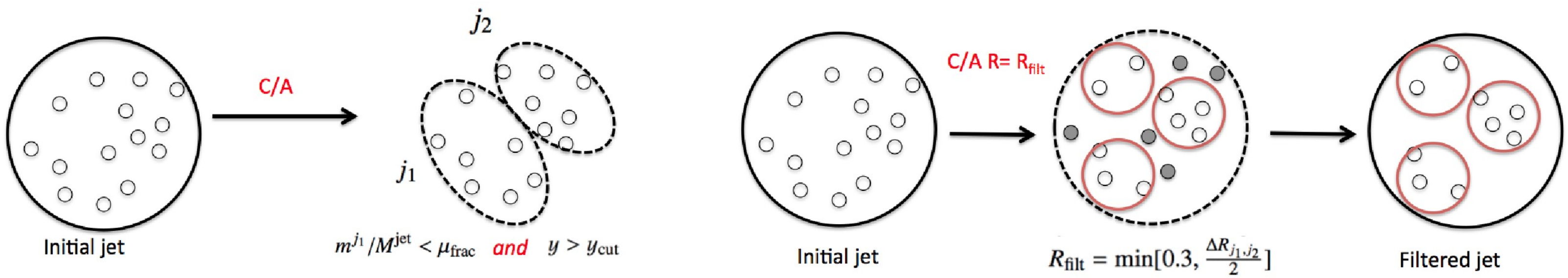
Trimming:



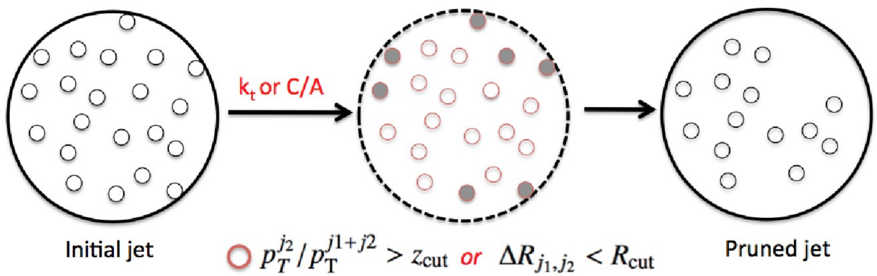


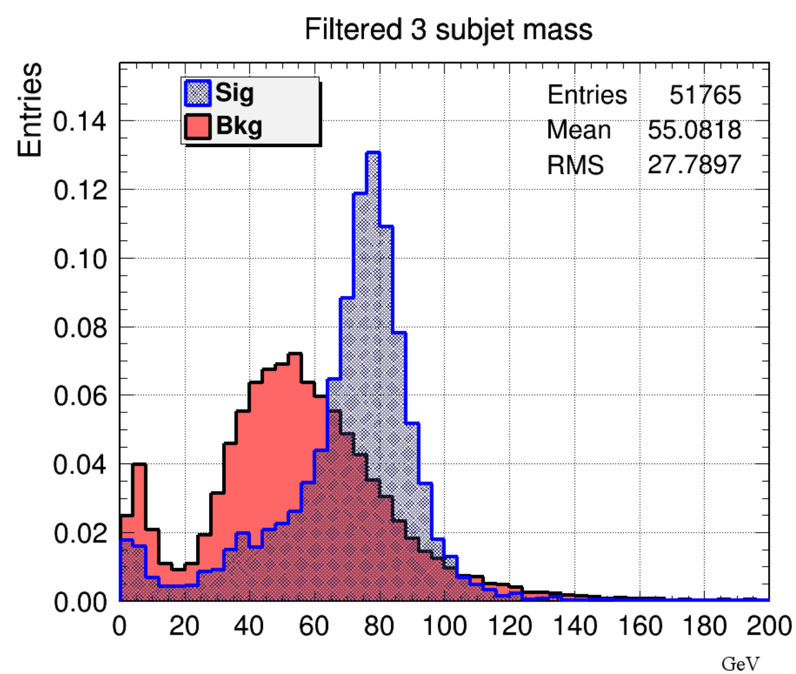
Jet grooming

Mass drop tagging plus filtering:

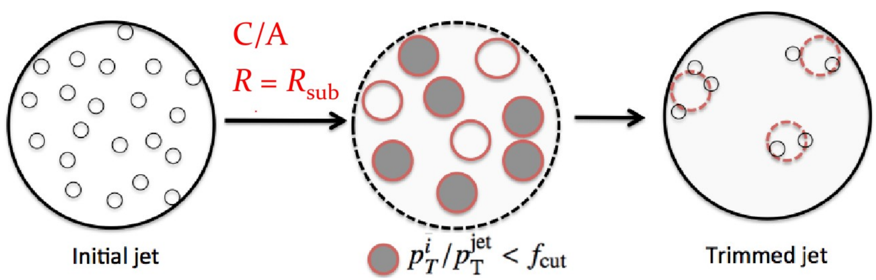


Pruning:



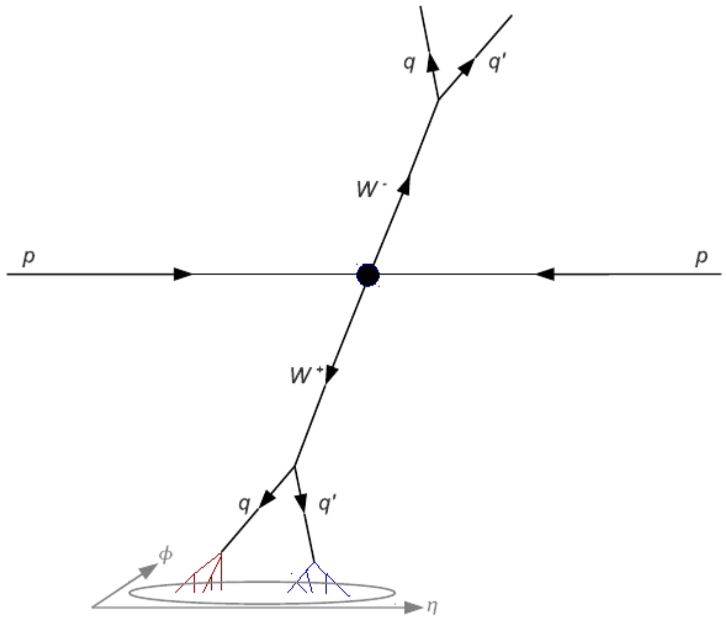


Trimming:

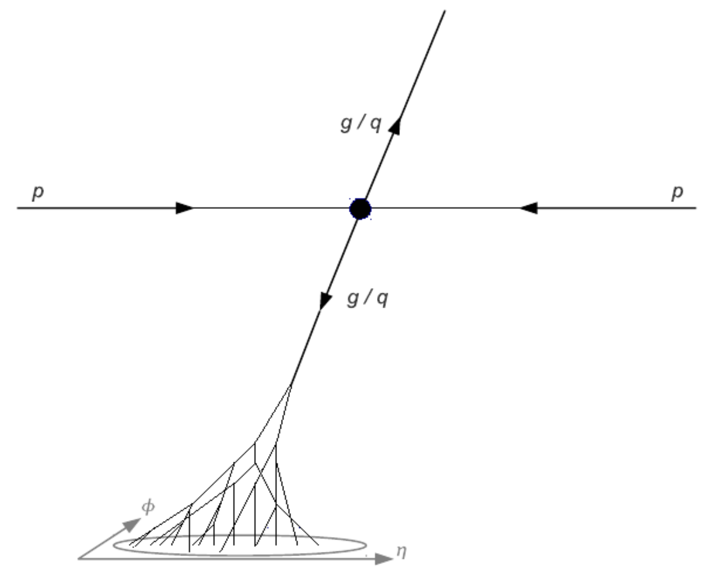


Example: N - subjettiness

- **N - subjettiness:** N can go from 1 to infinity
- **What it means:**
  - τ_N means to what degree a particular jet can be regarded as a jet composed of N subjets
  - $\tau_N \approx 0$: All radiation aligned with candidate subjets $\rightarrow$ N or fewer subjets
  - $\tau_N \gg 0$: Significant energy distributed away from subjet directions $\rightarrow$ At least N+1 subjets



Typical W - jet



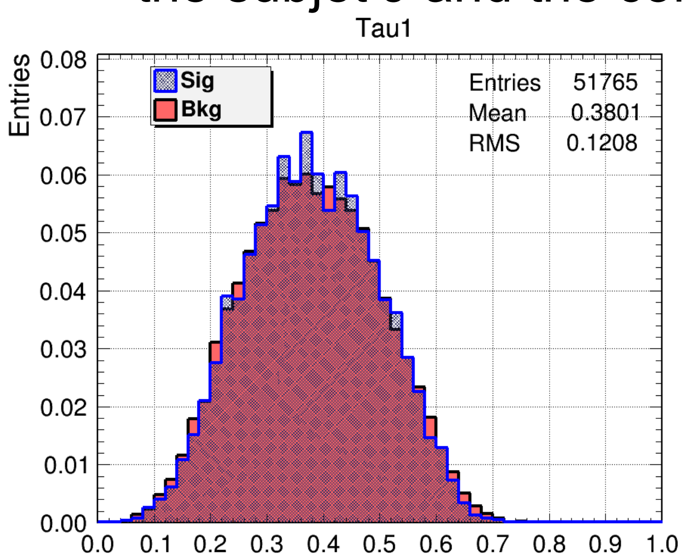
Typical gluon - jet

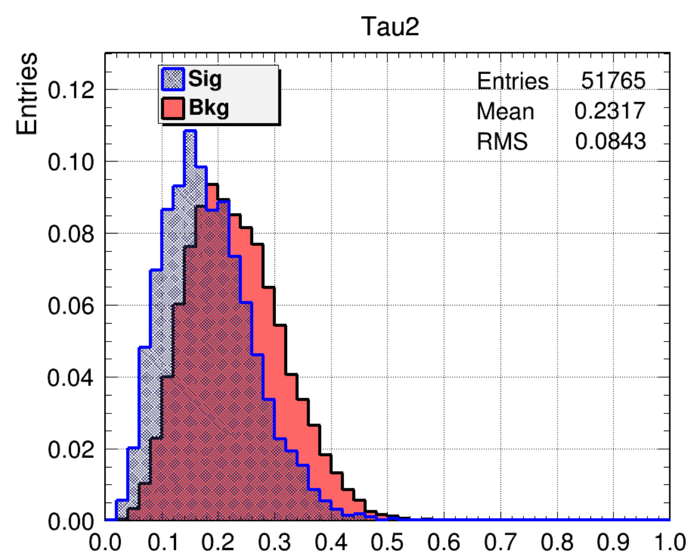
Example: N - subjettiness

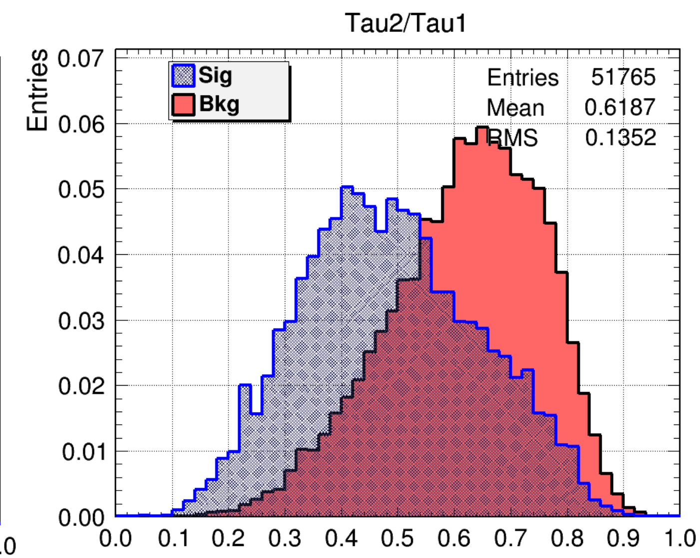
- How is it calculated? (arXiv:1011.2268)
- We look at one fat - jet
- Then we identify N candidate subjets (force k_T - algorithm to return exactly N jets)

$$\tau_N = a \cdot \sum_k p_{T,k} \, min(\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k})$$

- *k* runs over the jet constituents and $\Delta R_{J,k}$ is the distance between the subjet *J* and the constituent *k*







- Best discrimination by using the ratio $\frac{\tau_{N+1}}{\tau_N}$

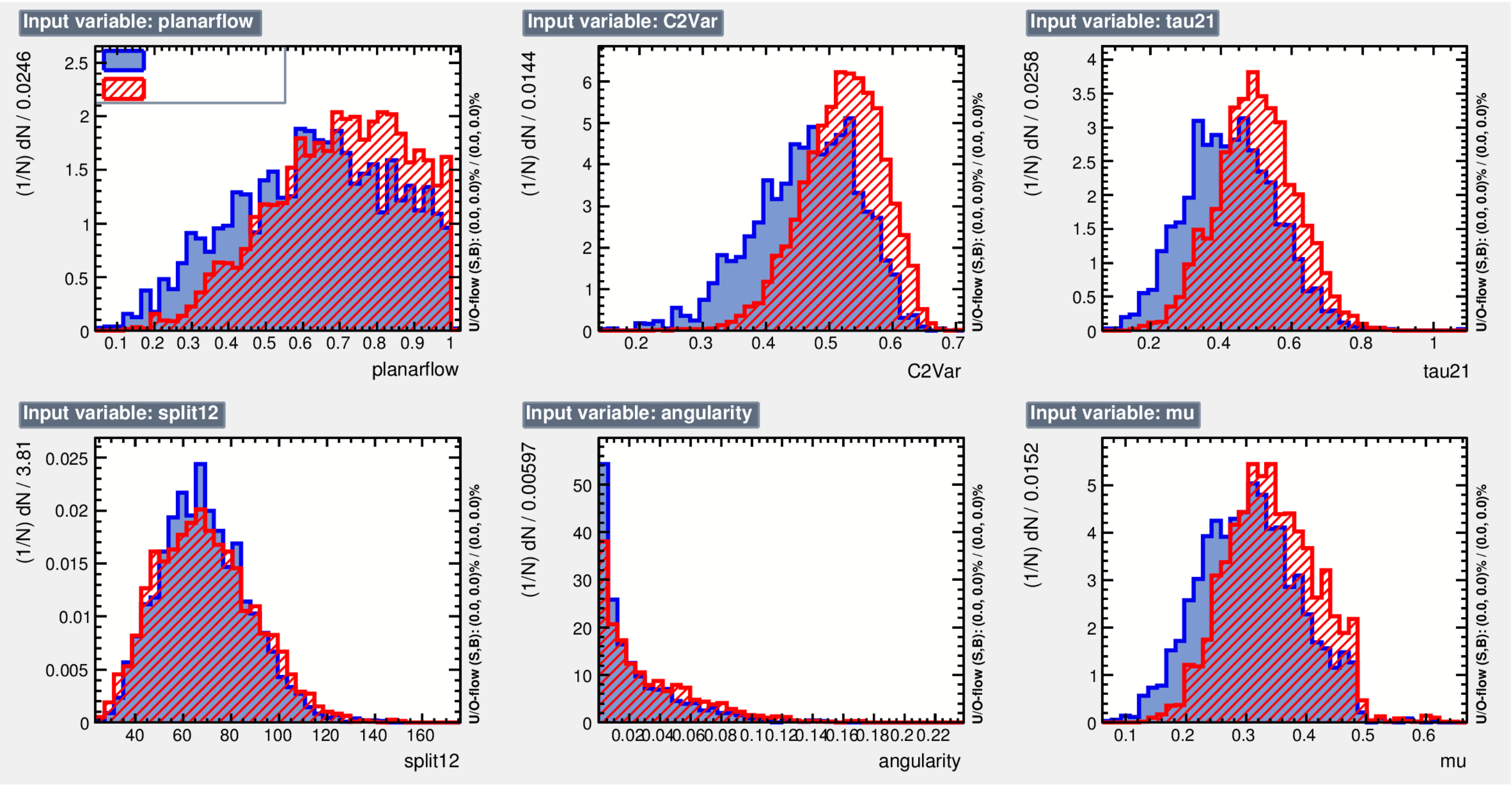
Event selection

- Won't go into details, only important cuts are mentioned
  1. Require exactly one hard, well reconstructed lepton (electron or muon)
  2. Missing transverse energy from the neutrino of > 30 GeV
  3. At least one Cambridge-Aachen (R = 1.2) jet with:
     - Mass > 60 GeV (W - mass)
     - p_T > 150 GeV
  4. Jet has to pass mass drop tagging and undergo filtering, filtered jet has to fall into W - mass window (60 GeV < m_J < 100 GeV)
- Still low S/B fraction S/(S+B) ~ 3%:
  - → Make use of substructure variables and combine them in a multivariate analysis
    1. Boosted Decision Tree
    2. Neural Network
    3. Likelihood Function

3.Multivariate Methods

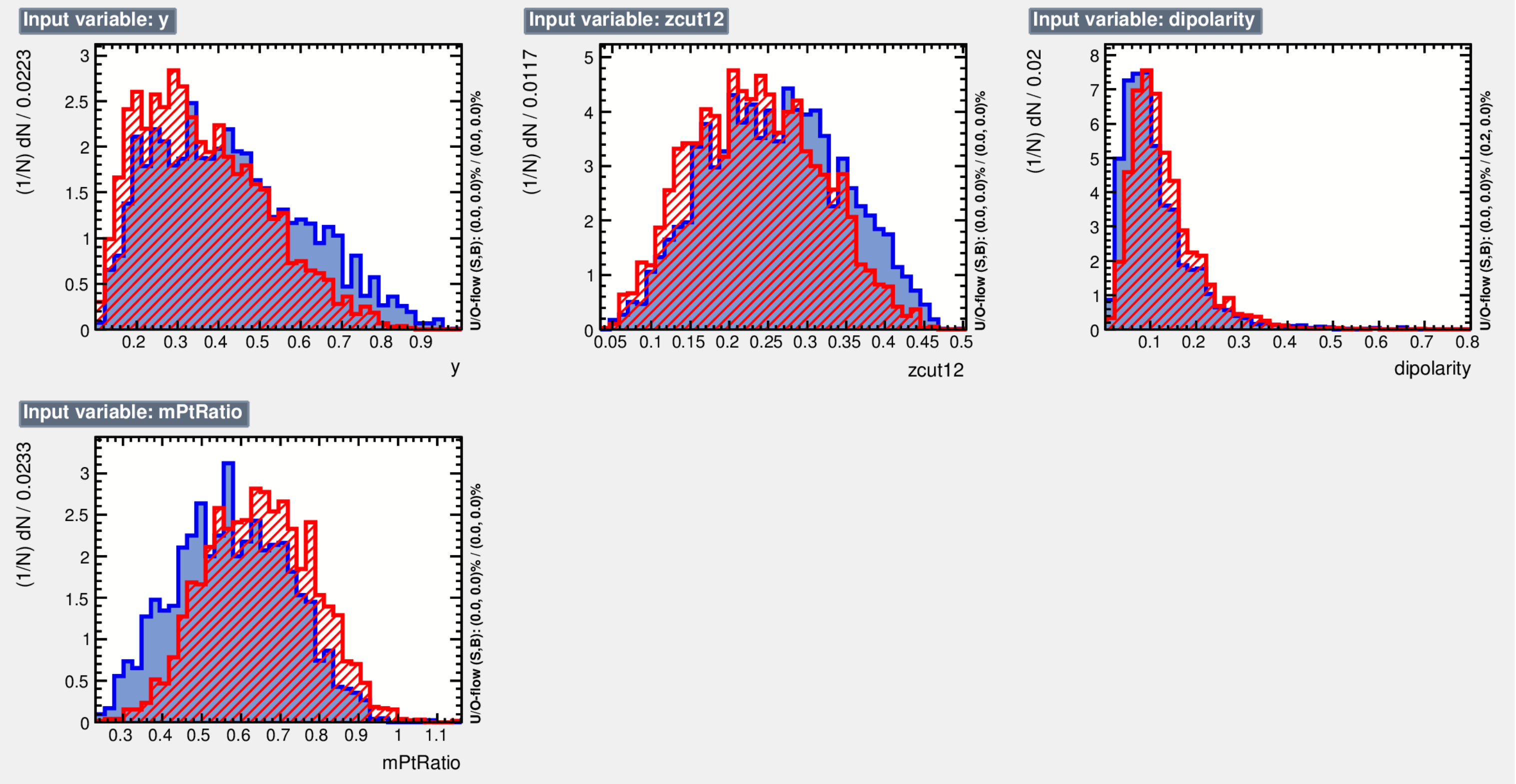
Multivariate Methods

- Idea of MV methods is to combine separating power of several variables into one
- Plain cuts would reject too many events
- MV methods take into account correlations between variables
- Have to be trained with Monte Carlo events
- Generally a very robust way to classify signal and background events as such



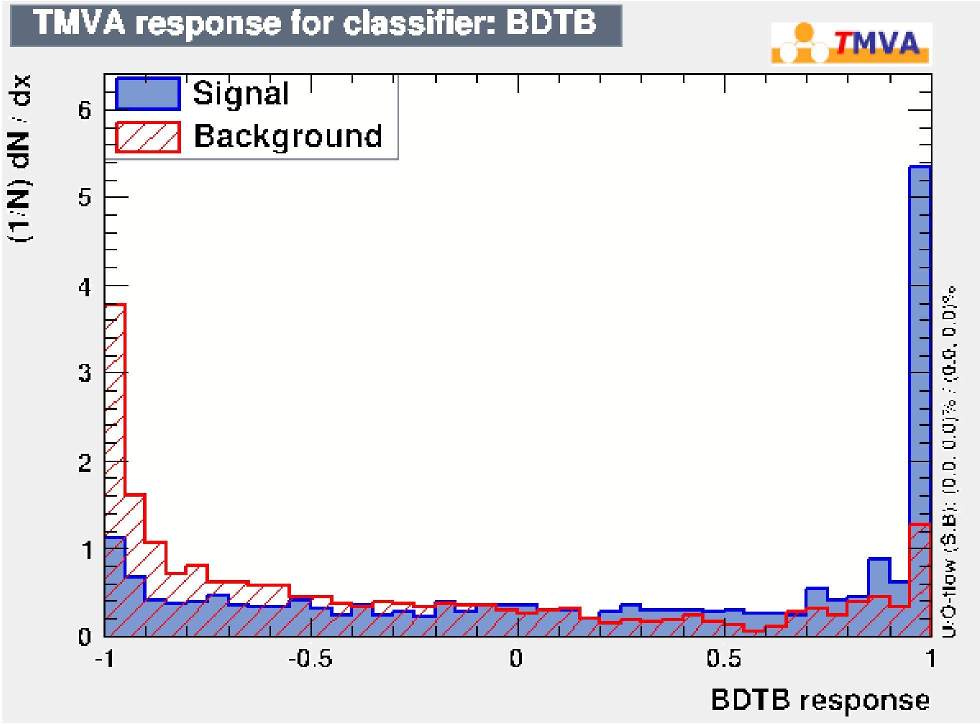
Multivariate Methods

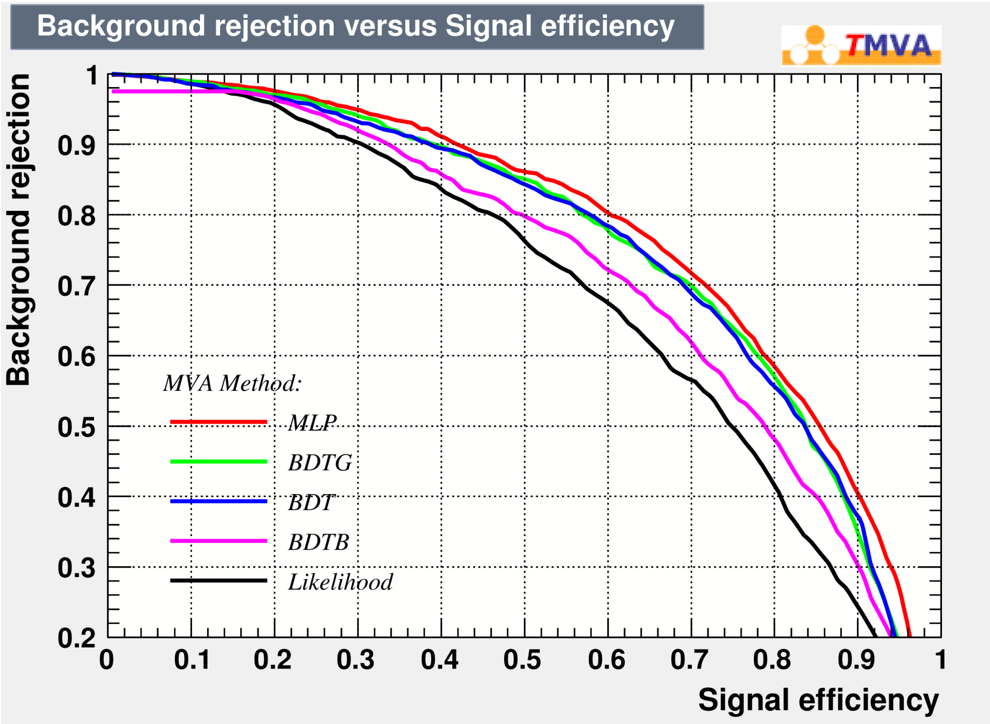
- Idea of MV methods is to combine separating power of several variables into one
- Plain cuts would reject too many events
- MV methods take into account correlations between variables
- Have to be trained with Monte Carlo events
- Generally a very robust way to classify signal and background events as such



Multivariate Methods

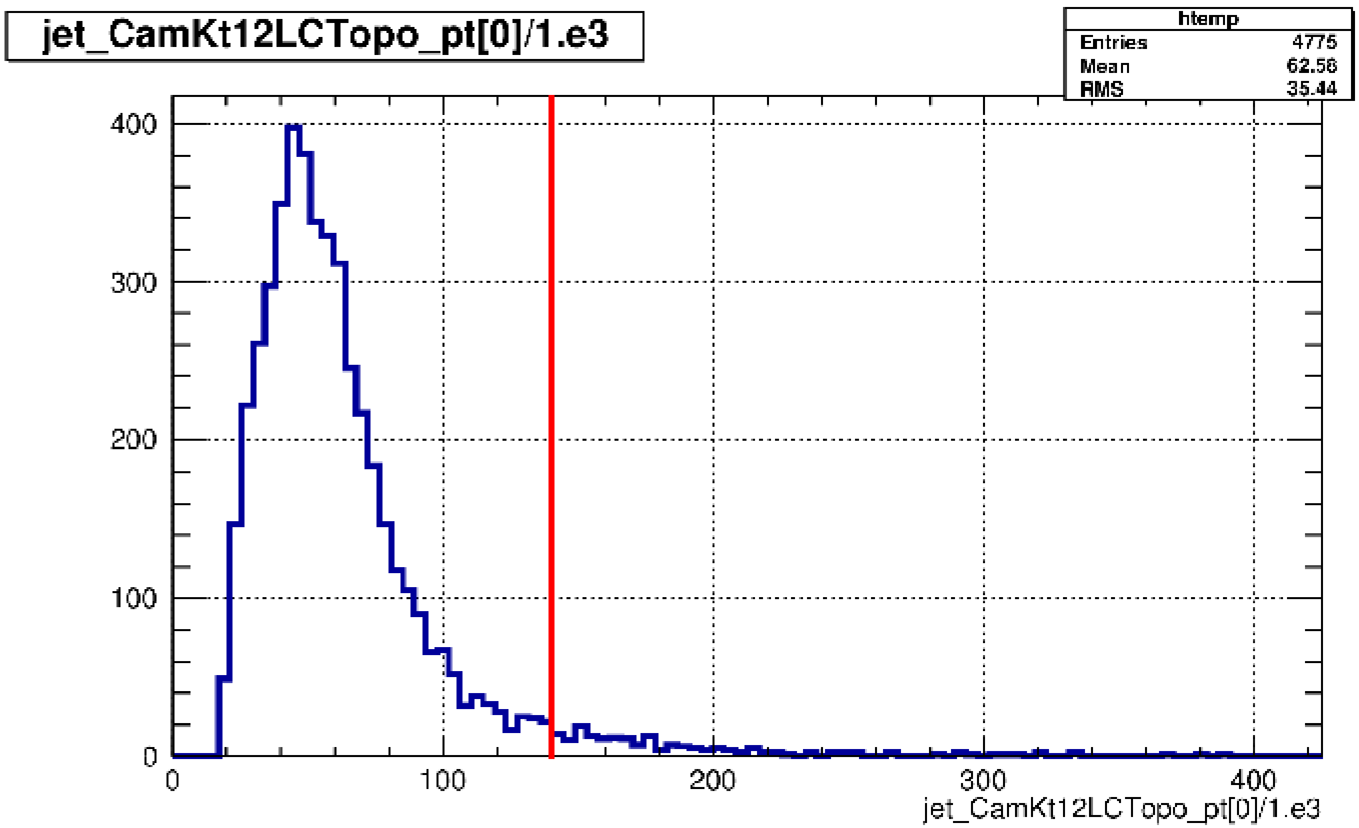
- Idea of MV methods is to combine separating power of several variables into one
- Plain cuts would reject too many events
- MV methods take into account correlations between variables
- Have to be trained with Monte Carlo events
- Generally a very robust way to classify signal and background events as such





Conclusion

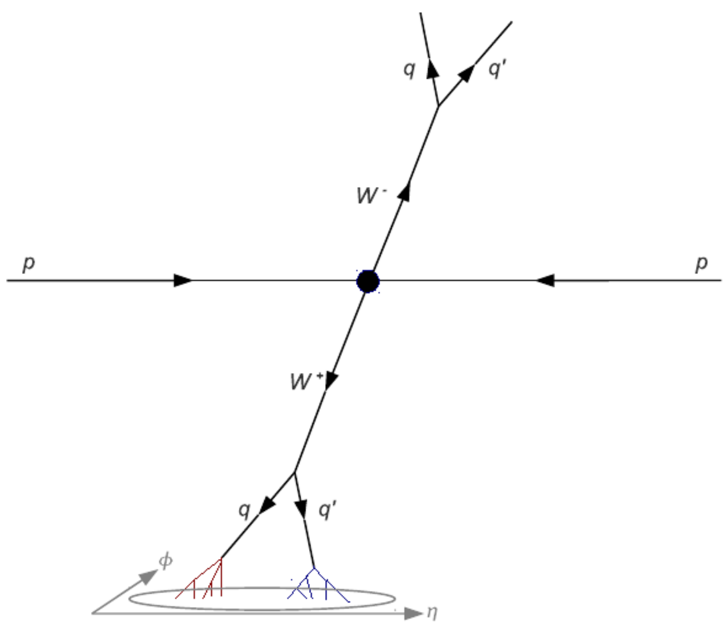
1. Boosted topologies are a wide an interesting field which becomes increasingly important
2. Not the easiest channel, boost - requirement already throws away around ~96% of all signal events
3. Substructure is very powerful to disentangle S from B
4. Multivariate tools provide a good way to combine variables

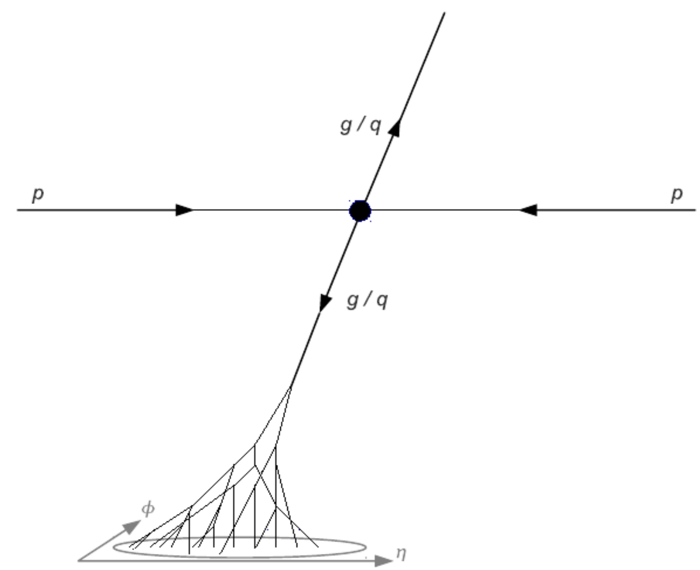


Backup Slides

Example: N-subjettiness

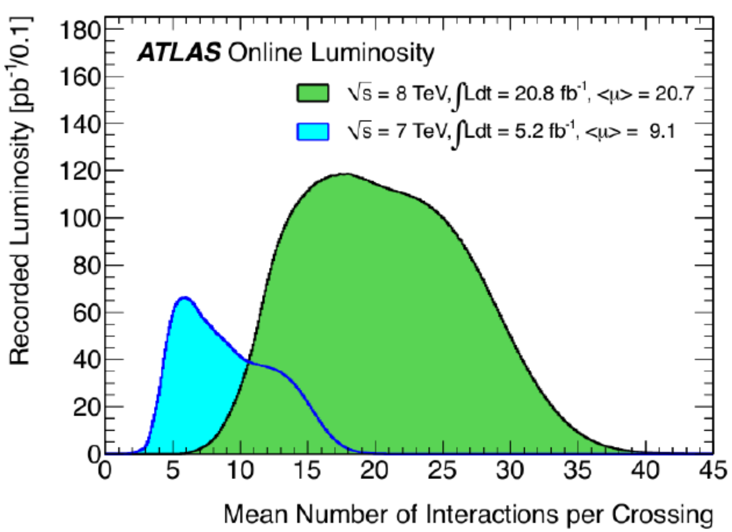
- One typical variable: N-subjettiness (arXiv:1011.2268)
- We look at one W-jet (jet has W mass)
- Then we identify N candidate subjets (hardest p_T reclustered jets)
- $\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \, min(\Delta R_{1,k}, \Delta R_{2,k}, \ldots, \Delta R_{N,k})$
- k runs over the jet constituents and $\Delta R_{J,k}$ is the distance between the subjet J and the constituent k
- $d_0 = \sum_k p_{T,k} R_0$ is the original jet's radius.

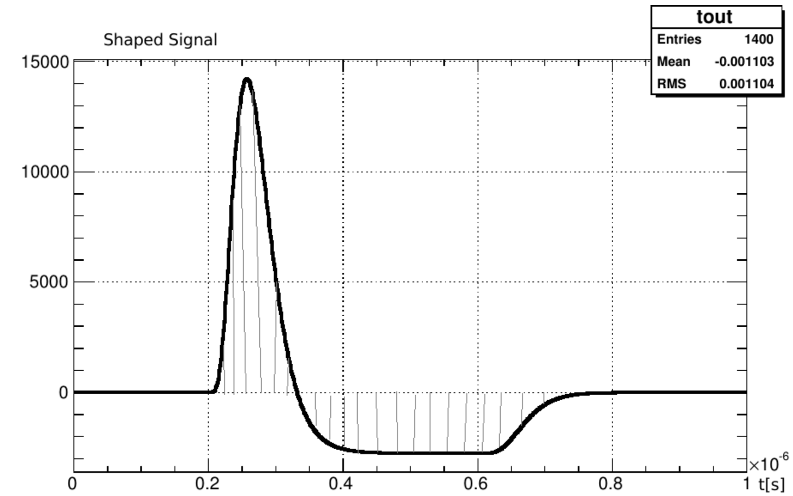




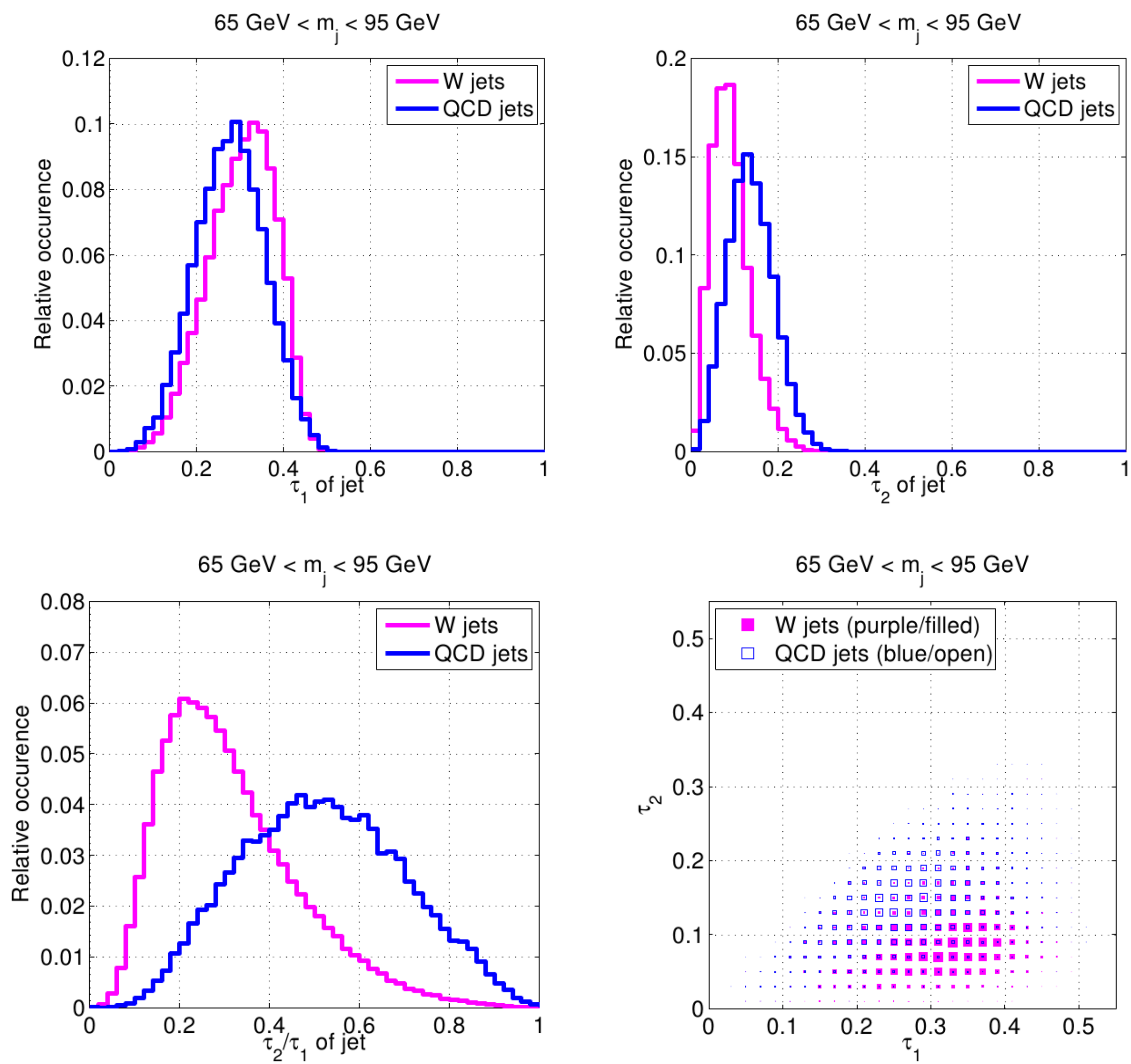
Pile-up

- Two types of pileup
  1. In time pileup
  2. Out of time pileup
- **In time pileup:**
  - Activity in the event from *pp* collisions in the same bunch crossing
  - Can be characterized by N_{PV} (number of primary vertizes)
- **Out of time pileup:**
  - Remaining signal in calorimeters from previous bunch crossings, due to long integration times → leads to negative cells/clusters



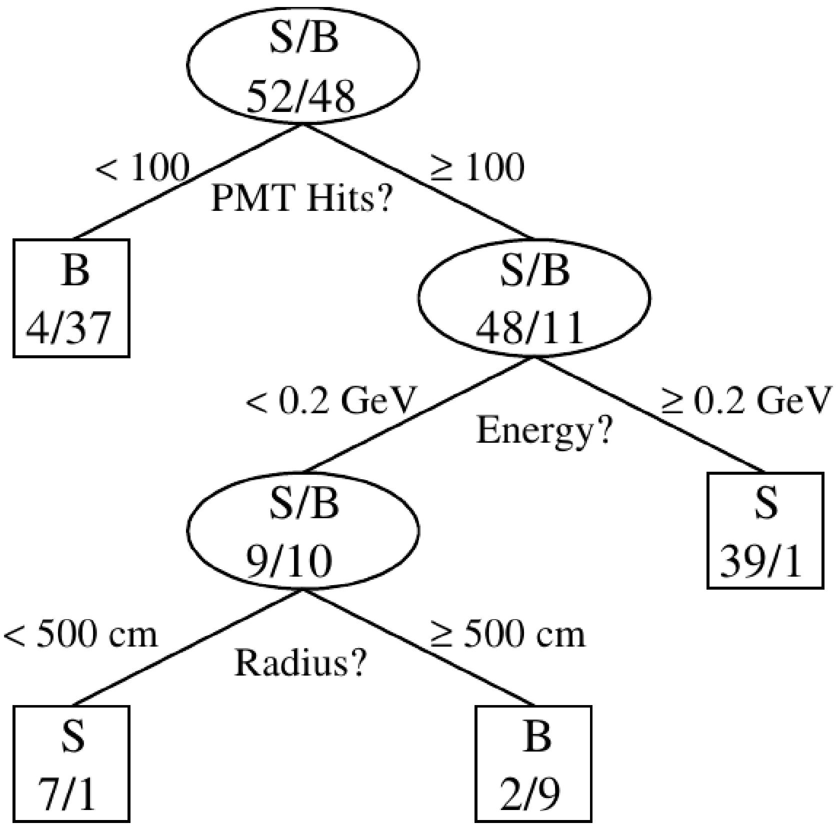


N-subjettiness



Boosted decision tree

- Train tree with S/B events
- In each node, split tree according to best separation
- Until only leaves with purity above a certain threshold are left
- Increase the weight of events that fell on a wrong leaf
- Make a new tree
- Iterate this procedure N times
- As signal classified event gets output value of 1, other wise 0
- Sum over all trees and compute average output value
- This is the final output variable

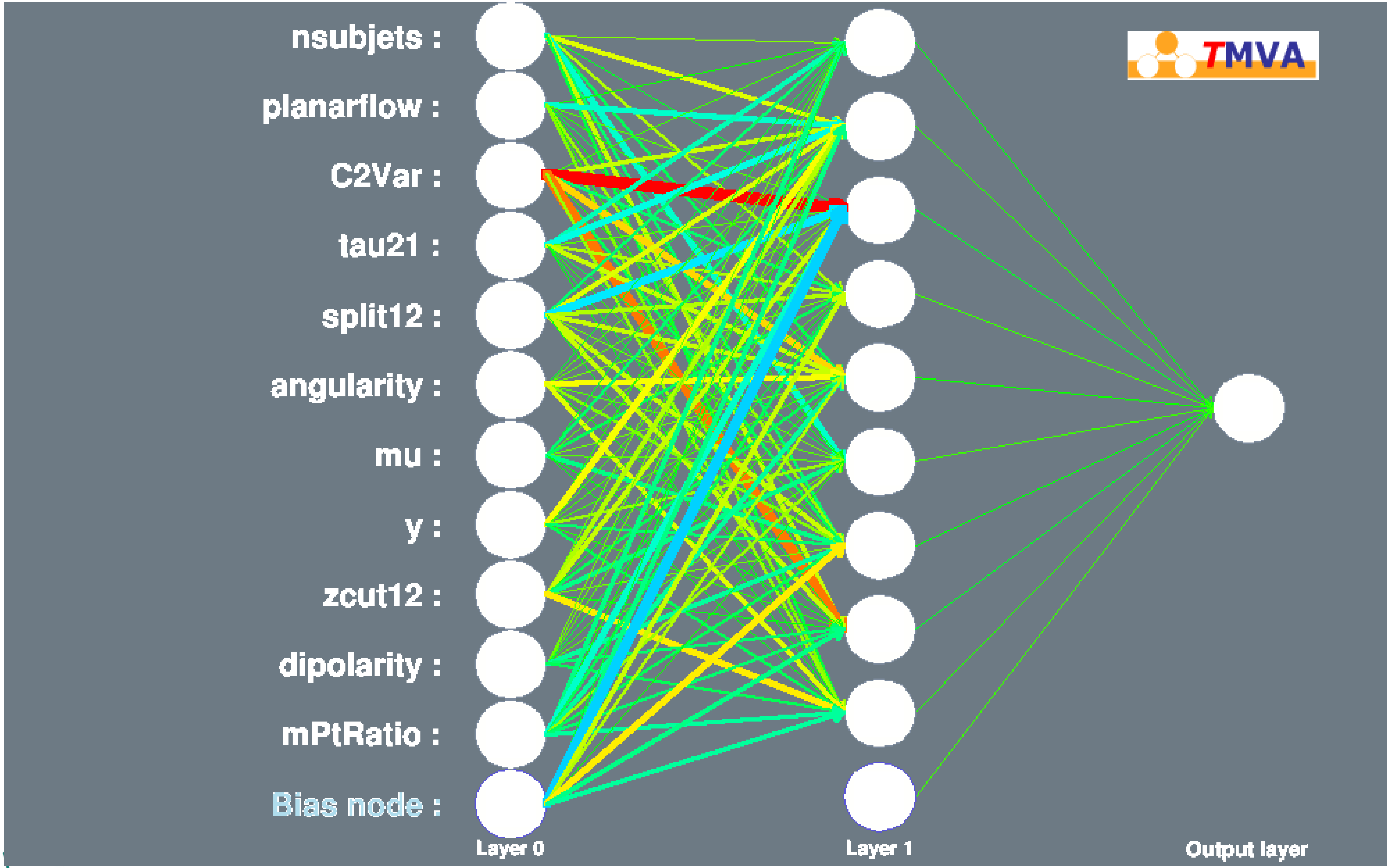


Separating signal and background

- Idea is to combine separating power of many little separating variables
- Solution: **Multivariate method** → Artificial Neural Network (ANN)
- Very powerful at recognizing patterns → Classification
- Has to be trained with many signal and background events (~50k each)

How does it work?

Separating signal and background



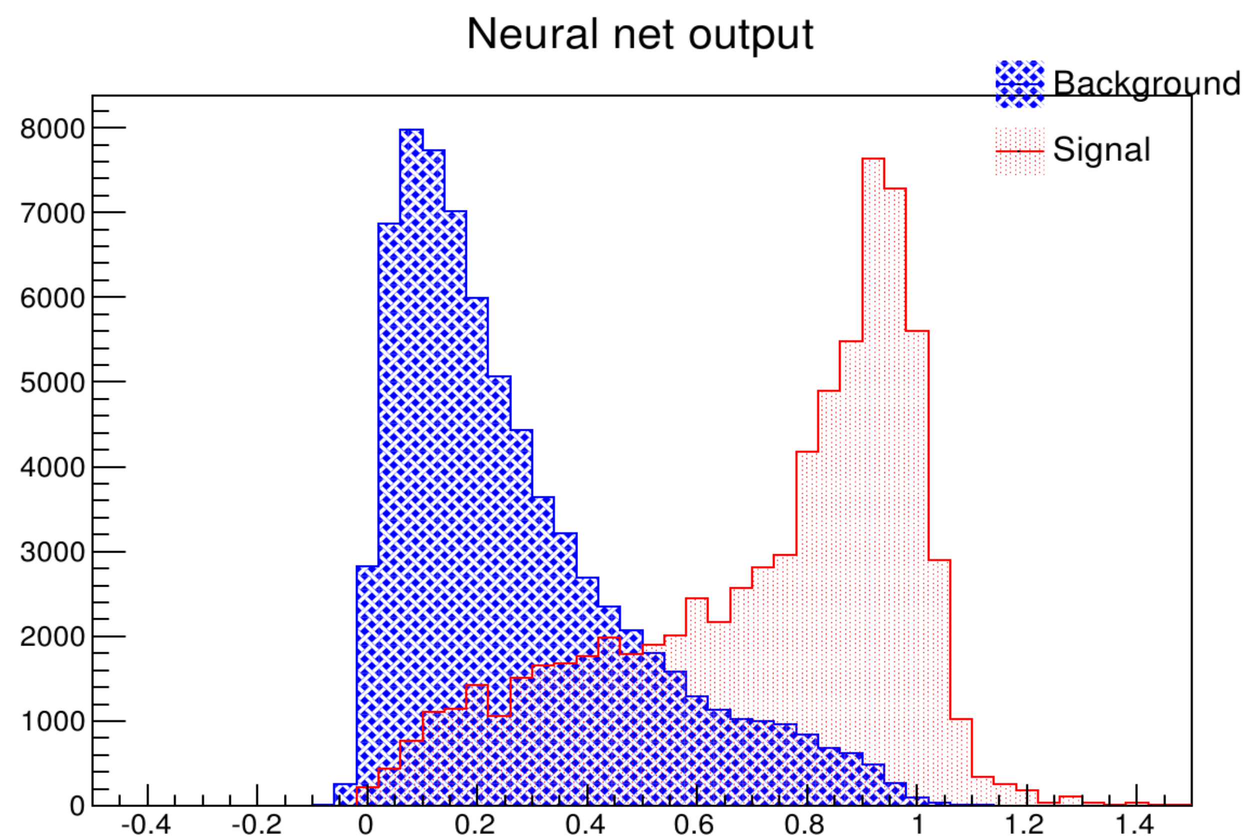
Separating signal and background

How does it work?

- Not programmed, but *trained*
- Supervised learning
- Feed the NN with all signal and bkg. events, where each is flagged as such
- After that (1 epoch) adjust weights of every synapse and node
- In signal case, answer at output layer shall be 1, other wise 0
- Train with some thousand epochs
- Apply trained network on data

Separating signal and background

- For training, use data instead of MC as background
- Assumption data is bkg only very well justified: $\frac{S}{B} \approx 2 \cdot 10^{-5}$



WW - Cross section

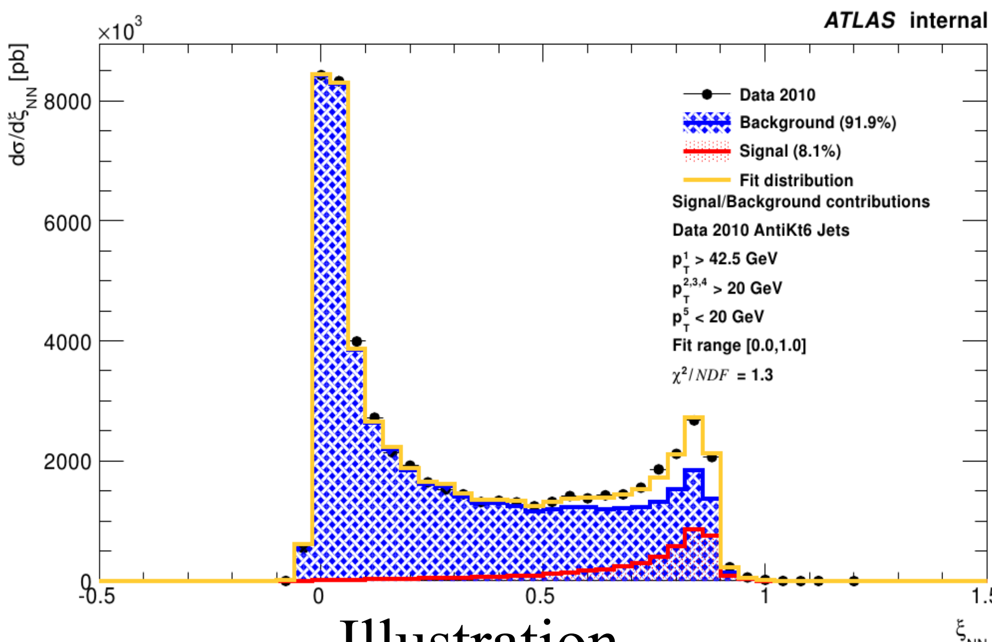
Separating signal and background

- Absolute amount of signal events still very small
- Cutting on the NN output would further reduce signal rate
- Better possibility:

  Run NN on data and then perform a fit of sig. and bg. output distributions to determine fraction of signal in data

- Access to the cross section

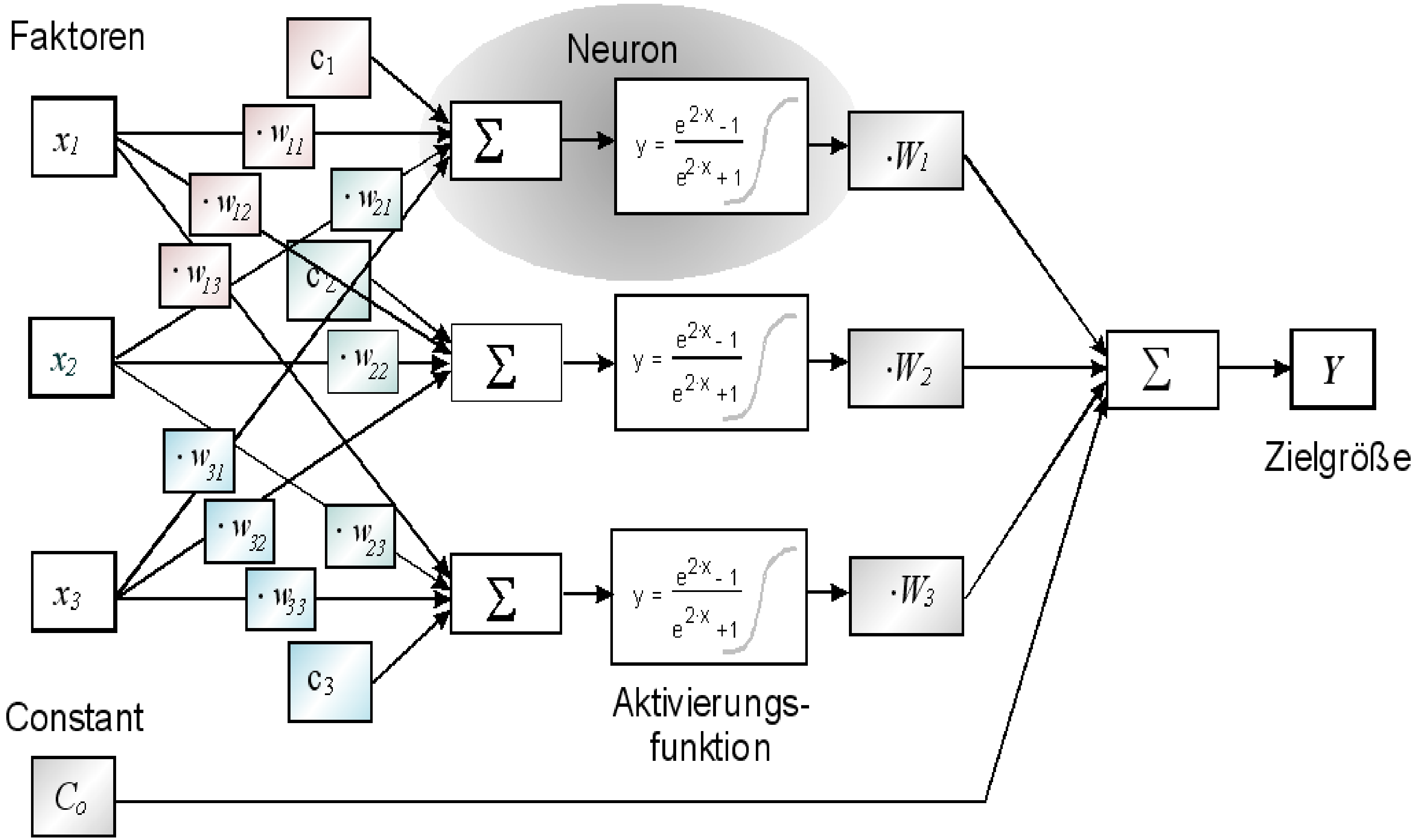
$$F(out) = f_{Sig} \cdot NN_{Sig} + (1 - f_{Sig}) \cdot NN_{Bkg}$$



Illustration

Orel Gueta,
Tel Aviv University

Neural Net



Faktoren
Neuron
c_1
x_1
$\cdot w_{11}$
$\cdot w_{12}$
$\cdot w_{21}$
$\cdot w_{13}$
c_2
x_2
$\cdot w_{22}$
$\cdot w_{31}$
$\cdot w_{32}$
$\cdot w_{23}$
x_3
$\cdot w_{33}$
c_3
Constant
C_0
Σ
$y = \frac{e^{2 \cdot x} - 1}{e^{2 \cdot x} + 1}$
$\cdot W_1$
$\cdot W_2$
$\cdot W_3$
Y
Zielgröße
Aktivierungs-
funktion