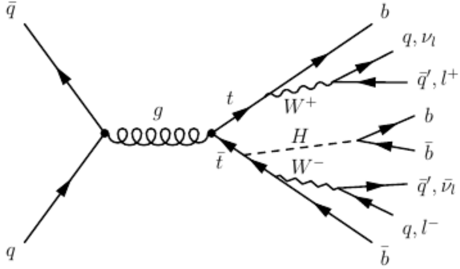
Top Tagging and Matrix Element Method for ttH/tt+jets separation at 13 TeV

Maren Meinhard

CHIPP Winter School - January 23, 2015

Aim of the $t\bar{t}H, H \rightarrow b\bar{b}$ analysis

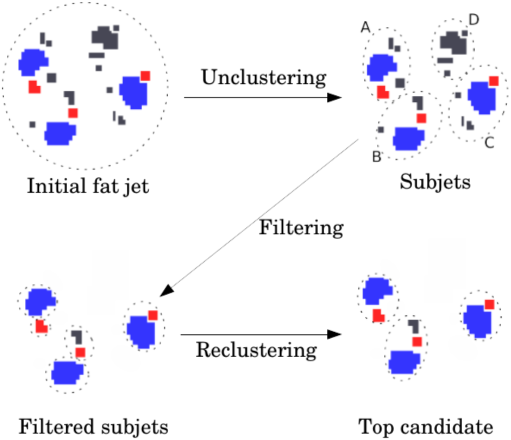
$\Rightarrow$ Measurement of Yukawa coupling between Higgs and top
$\Rightarrow$ Study the $t\bar{t}H$ final state in the semi-leptonic channel



Increase separation power between $t\bar{t}H$ and $t\bar{t} + Jets$ by optimization of event categories and inclusion of jet substructure in the input of the Matrix Element analysis

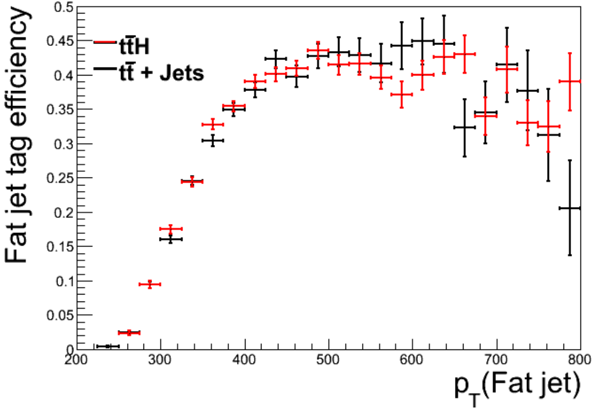
MultiR HEPTopTagger (MultiR HTT)

Input : fat jet constructed with Cambridge-Aachen (CA) algorithm with $R = 1.5$



[G. Kasieczka]

MultiR HTT tagging efficiency



- Low efficiency for small p_T fat jets
- Efficiency as a function of fat jet p_T agrees for ttH and tt+jets (relevant for top tagging in data)

The Matrix Element Method (MEM)

Probability density for an event $\mathbf{x}$ and observables $\boldsymbol{\alpha}$:

$$P(\mathbf{x}|\boldsymbol{\alpha}) = \frac{1}{\sigma_\alpha} \int d\Phi(\mathbf{y})\ |M_\alpha|^2(\mathbf{y})\ W(\mathbf{x}, \mathbf{y})$$

where

- σ_α is the total cross section
- $d\Phi(\mathbf{y})$ is the phase-space measure
- $|M_\alpha|^2(\mathbf{y})$ is the LO matrix element
- $W(\mathbf{x}, \mathbf{y})$ is the transfer function (probability to obtain a detector response $\mathbf{y}$ for an event $\mathbf{x}$)

Define ratio between $P(\mathbf{signal}|\boldsymbol{\alpha})$ and $P(\mathbf{bkg}|\boldsymbol{\alpha})$ as a one dimensional discriminating variable

Definition of some MEM decision variables

$$P_{s/b} = \frac{P(\mathbf{s}|\boldsymbol{\alpha})}{P(\mathbf{s}|\boldsymbol{\alpha}) + P(\mathbf{b}|\boldsymbol{\alpha})}$$

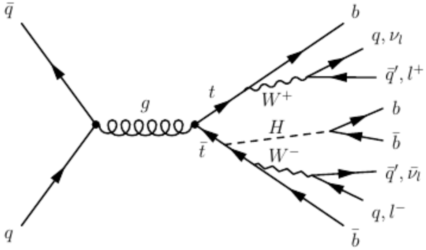
$$\frac{S}{B} = \frac{s_1 \cdot \#\ \textit{signal events with}\ P_{s/b} > 0.65}{s_2 \cdot \#\ \textit{bkg. events with}\ P_{s/b} > 0.65}$$

$$\frac{S}{\sqrt{B}} = \frac{s_1 \cdot \#\ \textit{signal events with}\ P_{s/b} > 0.65}{\sqrt{s_2 \cdot \#\ \textit{bkg. events with}\ P_{s/b} > 0.65}}$$

where -s_i are factors that correct for differences in the number of generated events in both samples and
-$P(\mathbf{s}/\mathbf{b}|\boldsymbol{\alpha})$ is the probability to have an outcome $\boldsymbol{\alpha}$ for a signal/background event

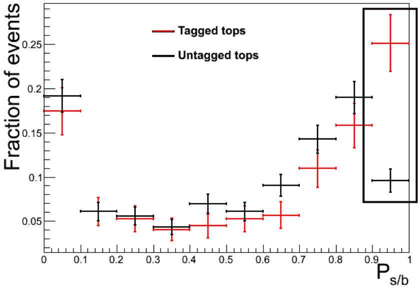
Event selection

- One single lepton with $p_T > 30\,\text{GeV}$
- $N_{jet} \geq 6$ with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.5$
- "W tag" : mass of jets produced by decay in
  - $[60, 100]$ GeV if $N_{jet} = 6$ or
  - $[72, 94]$ GeV if $N_{jet} > 6$

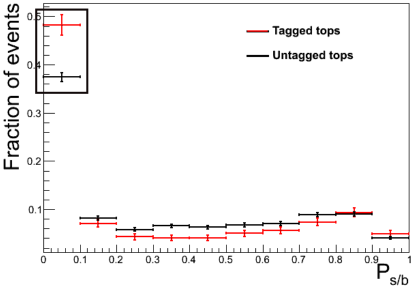


$\Rightarrow$ W decay fully reconstructed

$P_{s/b}$ distribution for tagged and untagged events



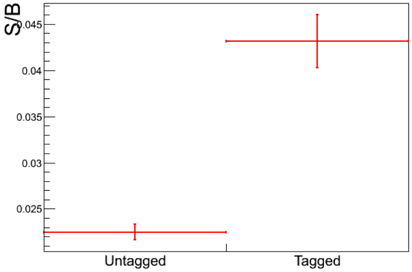
$t\bar{t}H$



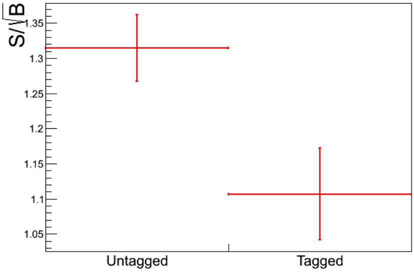
$t\bar{t} + Jets$

Top tagging increases the separation between signal and background events

Correlation MEM - MultiR HTT: S/B and $S/\sqrt{B}$ at $\mathcal{L} = 19.04\,\mathrm{fb}^{-1}$



S/B



$S/\sqrt{B}$

$\frac{S}{B}$ is increased by a factor 2, $\frac{S}{\sqrt{B}}$ decreases when using top tagging

Conclusions and next steps

Summary

- Tagging top quarks in $t\bar{t}H$ and $t\bar{t} + Jets$ events leads to better separation between signal and background events
- This is due to a better jet reconstruction for "tagged" events
- Improves S/B ratio by almost a factor two

Next steps

- Optimize event categories to increase the separation power between signal and background events
- Include events that are not fully reconstructed
- Implement Higgs tagger to further improve the separation

Backup

Cambridge-Aachen (CA) jet algorithm

Define distance between input objects i and j:

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2}$$

with

- $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$, the angular distance between two input objects
- η, the pseudo-rapidity
- ϕ, the azimuthal angle

Recombine closest jets until $\Delta R_{ij} > R$ for all input objects, where R is a parameter of the algorithm

MultiR HTT WP and Data Samples

MultiR HTT Working Point (WP)

- CA, R=1.5 fat jets
- $100\,\text{GeV} < m_t(R_{min}) < 225\,\text{GeV}$
- $f_W(R_{min})$ <0.19
- ΔR_{min} <0.5
- p_T >200 GeV

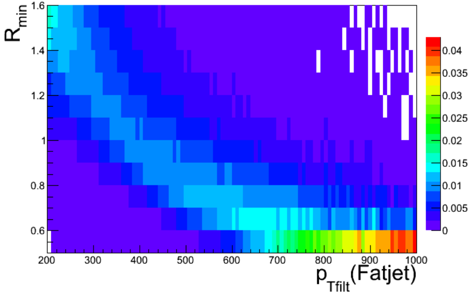
Available samples (CSA 14)

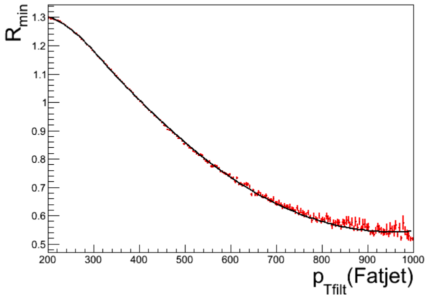
Sample	Event generator	N. generated events	PU
$t\bar{t}H$	Pythia 6	97520	PU20bx25
$t\bar{t} + Jets$	Madgraph	25474122	PU20bx25

MultiR HTT - ΔR_{min}

R_{min} = smallest cone size where mass drop is less than 20%

At higher p_T, top decay products are very collimated
$\Rightarrow R_{min}$ decreases with p_T



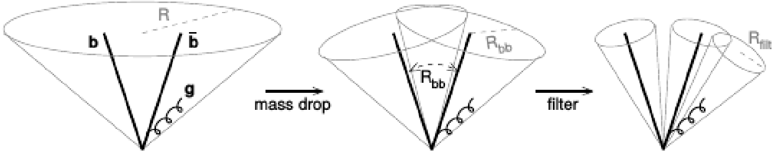


Define

$$\Delta R_{min} = R_{min} - R_{min,fit}$$

Higgs tagging

Similar to top tagging, starts with fat jet j constructed with CA Algorithm



- Undo last algorithm clustering step so that $m_{j1} > m_{j2}$
- If $m_{j1} < \mu m_{j2}$ (significant mass drop) and the splitting is relatively symmetric, keep j_1, j_2
- j is a Higgs candidate if both j_1 and j_2 are b-tagged

[J. Butterworth *et al.*]