

# Produzione associata del bosone di Higgs con una coppia di quark top ad alto impulso trasverso

---

**Silvia Biondi** (Università degli studi di Bologna & INFN)

101° Congresso della Società Italiana di Fisica  
“Sapienza”, Roma, 21 - 25 Settembre 2015



# Outline

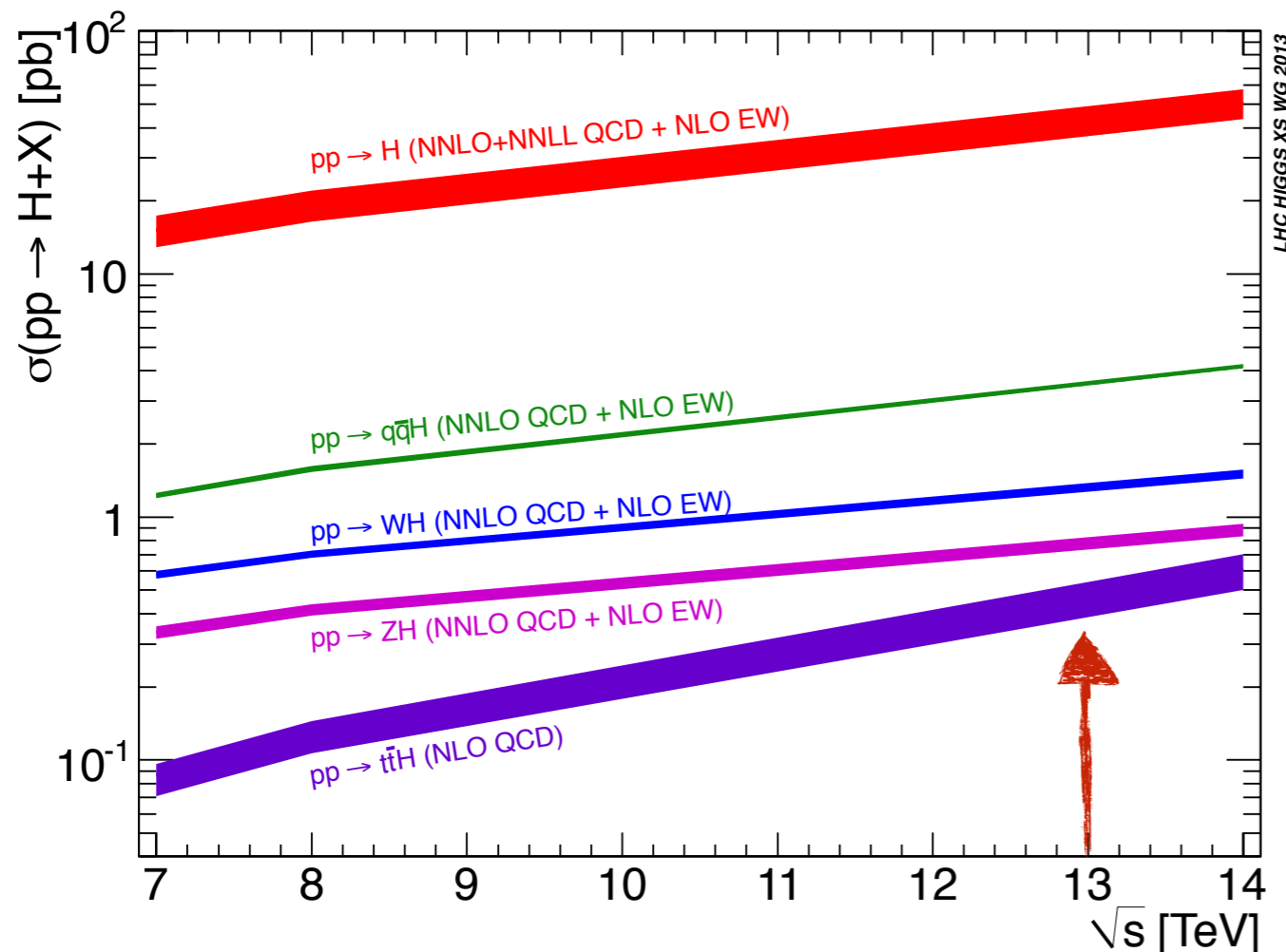
- Descrizione della topologia  $t\bar{t}H$
- Motivazioni e vantaggi dell'analisi in regioni ad alto impulso trasverso
- Studi preliminari con campioni MonteCarlo
- Studio delle prestazioni dell'algoritmo di top-tagging con MC a 13TeV
- Conclusioni e piani futuri



# Bosone di Higgs e quark Top

- Bosone di Higgs: osservato nel 2012 dalle collaborazioni ATLAS e CMS  
 ( $M_{\text{Higgs}} = 126.0 \pm 0.4 \text{ (stat)} \pm 0.4 \text{ (syst)} \text{ GeV}$ )
- Ricerche attuali: studio delle proprietà (SM Higgs?, spin, CP) e degli accoppiamenti e fisica oltre il Modello Standard
- Top: quark più massivo  $\rightarrow M_{\text{top}} = 173.34 \pm 0.27 \text{ (stat)} \pm 0.71 \text{ (syst)} \text{ GeV}$ 
  - accoppiamento più alto con il bosone di Higgs ( $\sim 1$ )

@13TeV!



cross section (pb)			
channel	7 TeV	8 TeV	13 TeV
ggH	15	19	43
VBF	1.2	1.6	3.7
WH	0.5	0.7	1.4
ZH	0.3	0.4	0.9
ttH	0.09	0.13	0.50

la produzione associata  $t\bar{t}H$  presenta l'incremento più grande di sezione d'urto tra tutti i processi!

<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections>

# Topologia resolved $t\bar{t}H$

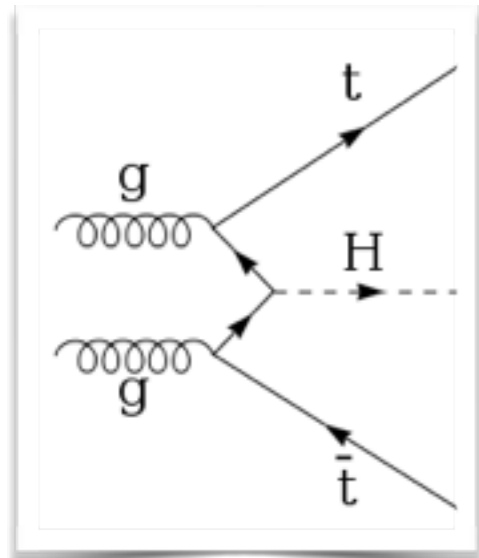


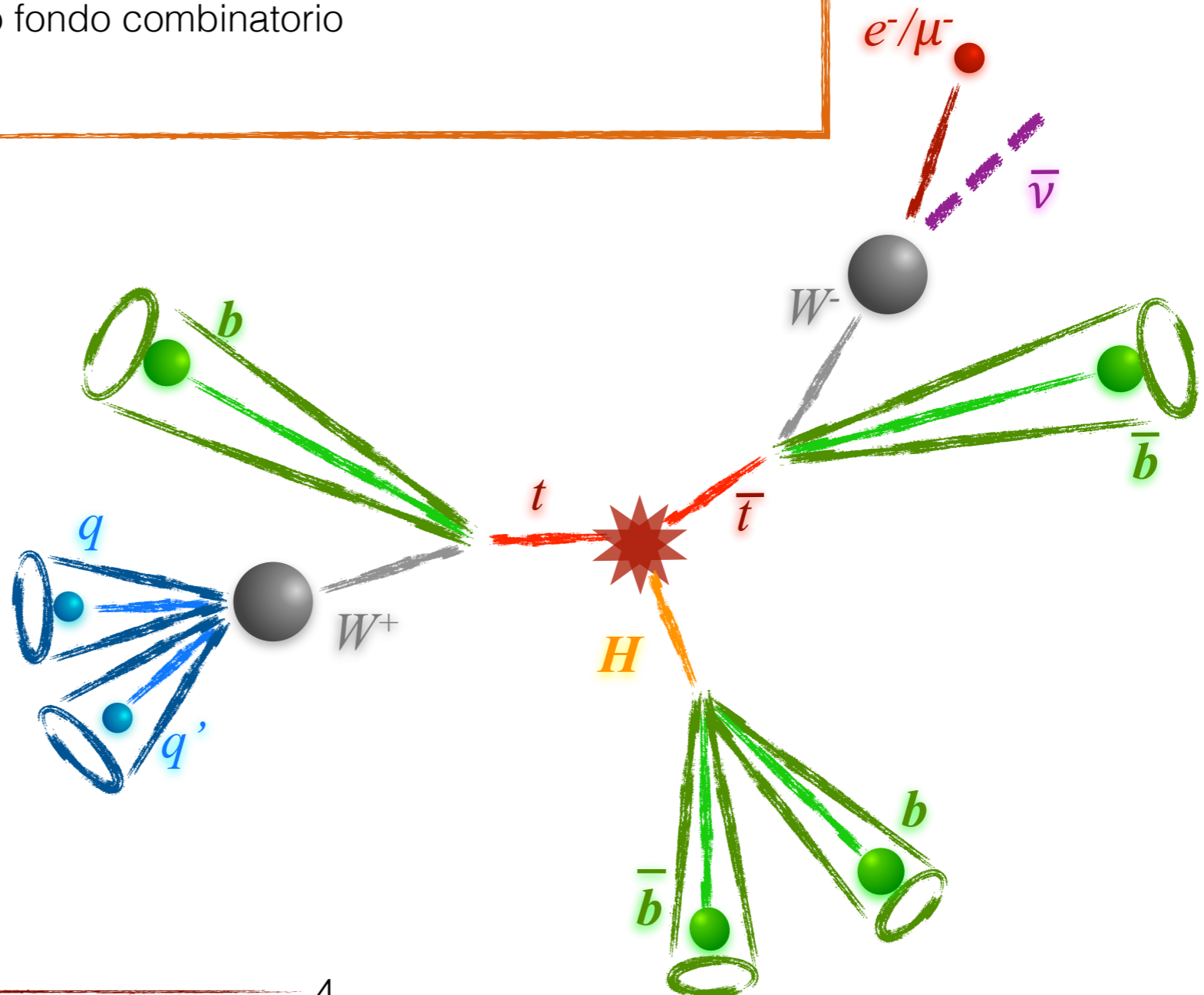
Diagramma della produzione associata  $t\bar{t}H$

- 6 jet distinti (“resolved”)
  - algoritmi standard di ricostruzione dei jet

$$\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$$

- alto fondo combinatorio

basso impulso trasverso ( $p_T$ )



# Topologia boosted $t\bar{t}H$

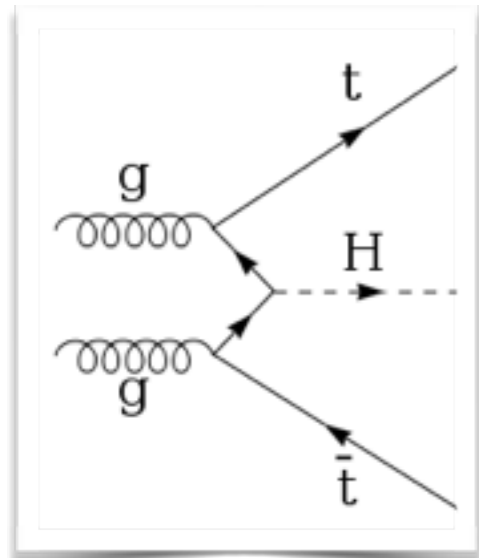
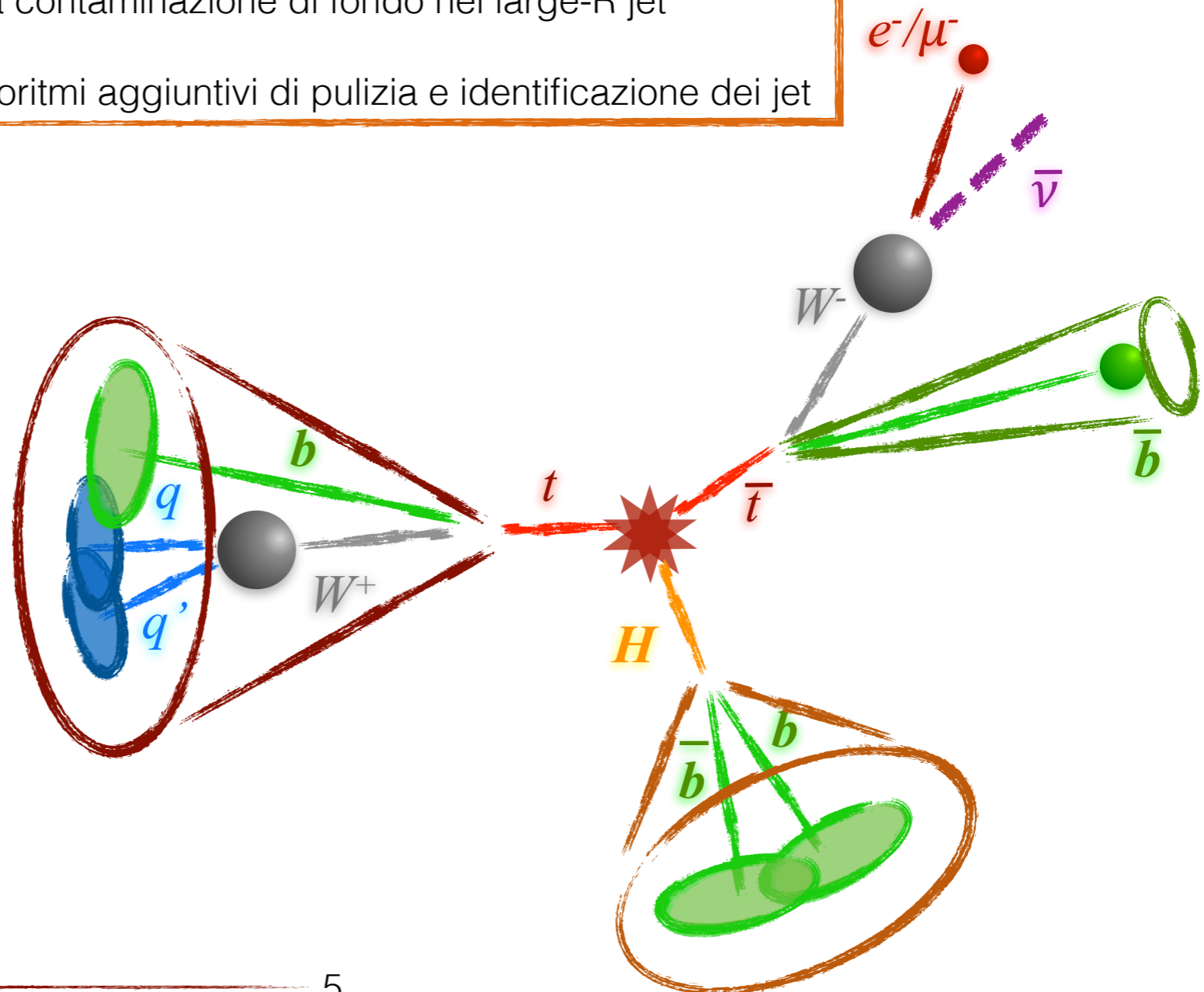


Diagramma della produzione associata  $t\bar{t}H$

- 2 large-R jet (“**boosted**”)
    - algoritmi standard di ricostruzione dei jet
- $$\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \geq 1.0$$
- alta contaminazione di fondo nei large-R jet
  - algoritmi aggiuntivi di pulizia e identificazione dei jet

alto impulso trasverso ( $p_T$ )



# Algoritmi di ricostruzione e “pulizia” dei large-R jet

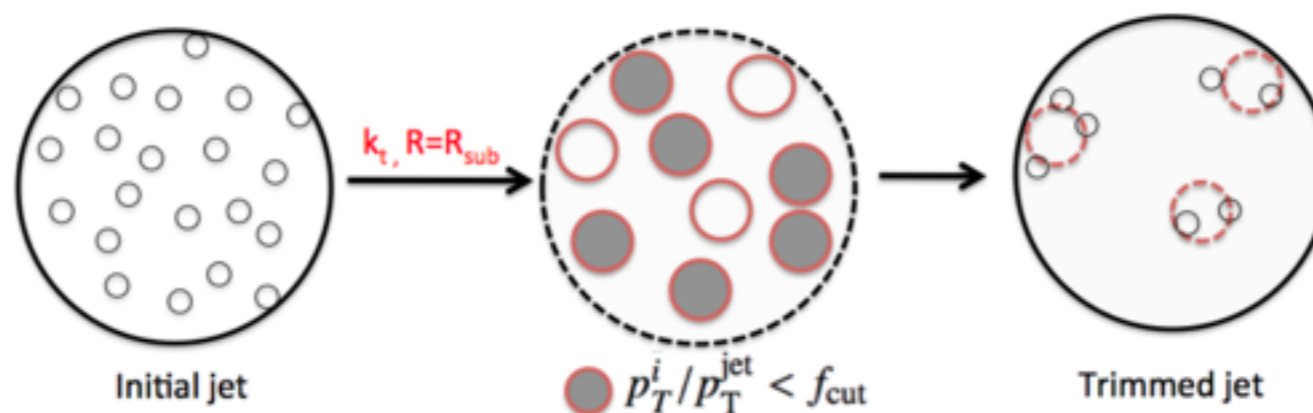
## Algoritmo di ricostruzione

I jet sono ricostruiti con un algoritmo iterativo che combina i depositi calorimetrici entro un determinato raggio  $R = 1.0$

i jet vengono poi “puliti” dalla contaminazione dovuta all’alta concentrazione di particelle

## Algoritmo di trimming

I costituenti di un jet vengono ricostruiti nuovamente in jet di raggio inferiore  $R_{\text{sub}}$  (subjet).  
I subjet con un  $p_T$  inferiore di una frazione  $f_{\text{cut}}$  del  $p_T$  del jet iniziale vengono scartati.  
Il jet finale viene ricostruito utilizzando solamente i subjet rimanenti.



$$R_{\text{sub}} = 0.2$$
$$f_{\text{cut}} = 0.05$$

# Studi preliminari Monte Carlo

## ◆ dettagli del processo

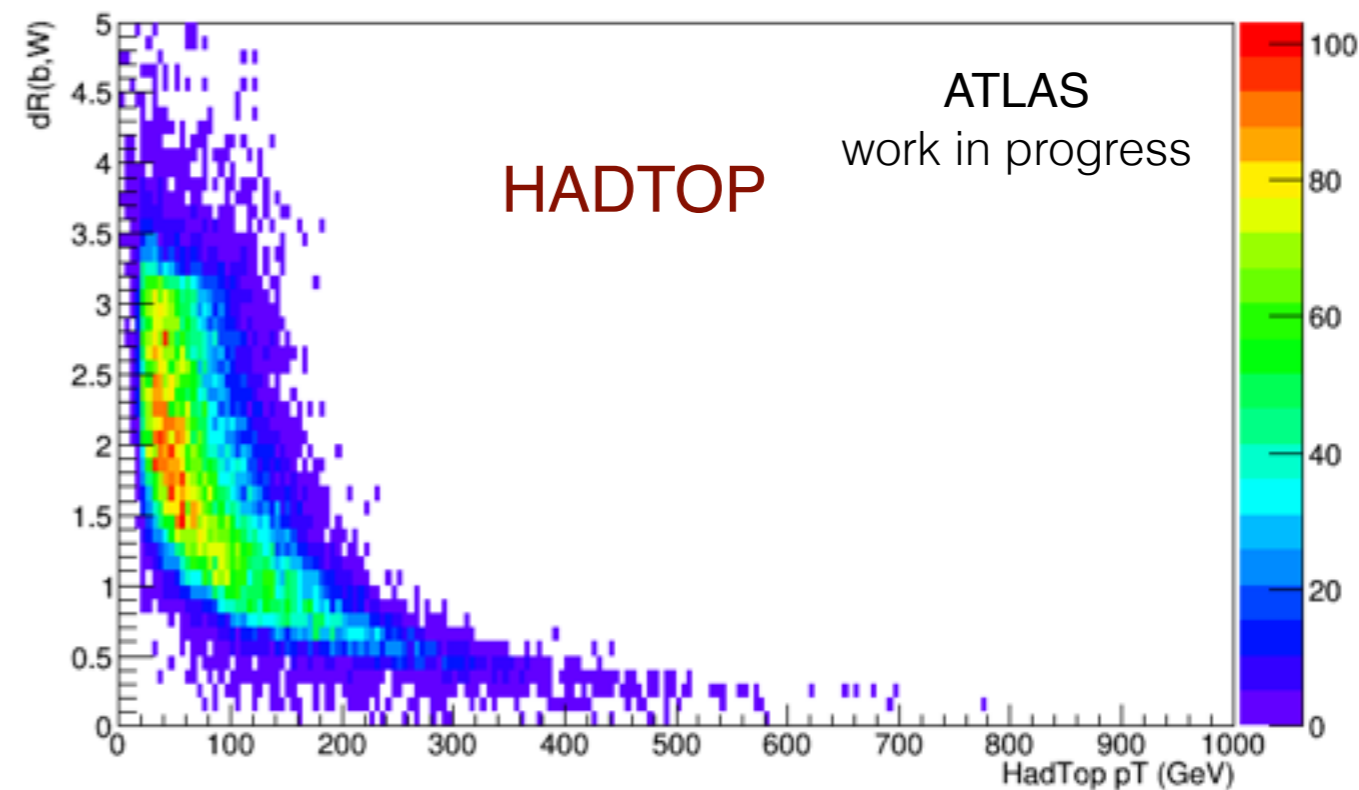
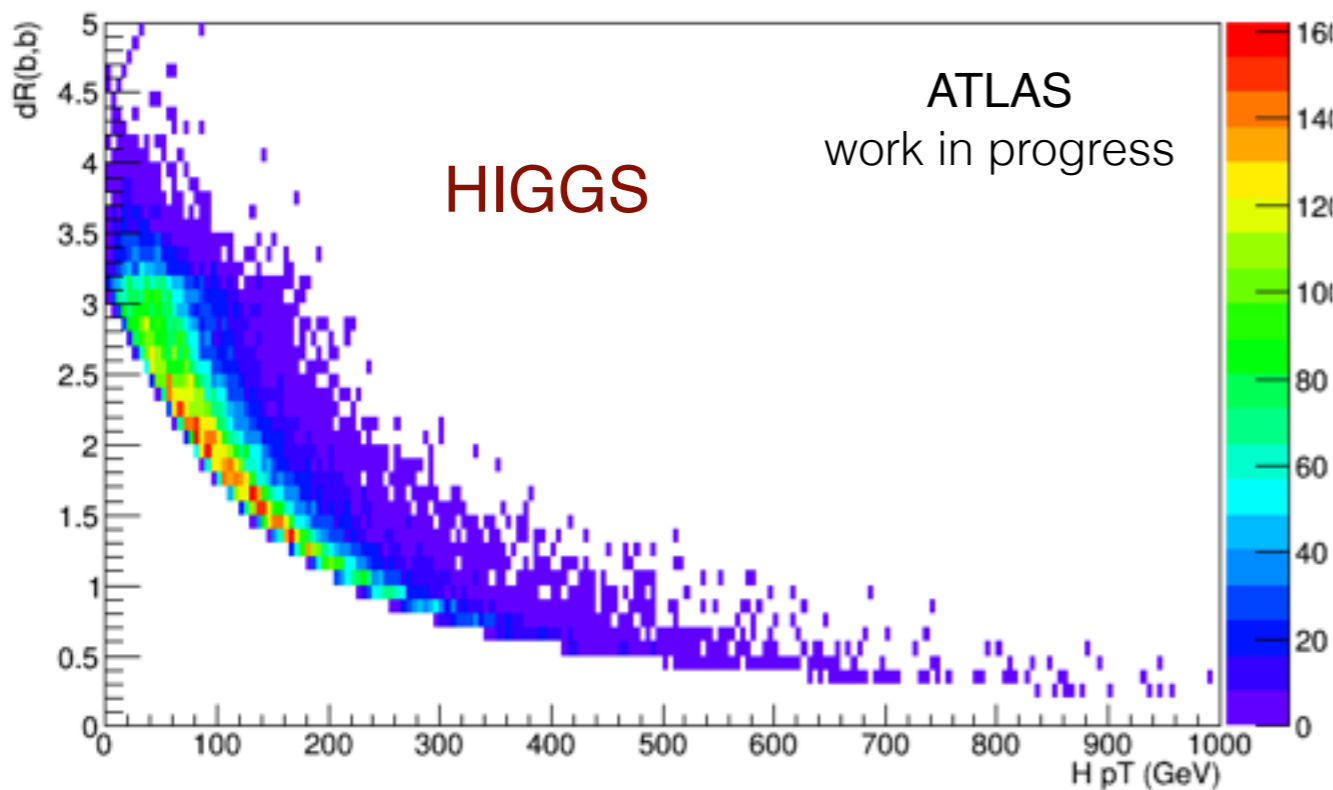
- ▶ processo generato:  $pp \rightarrow t\bar{t}H \rightarrow l+jets+b\bar{b}$  ( $l = e, \mu$ )
- ▶ energia del fascio: 13 TeV
- ▶ numero di eventi: 30000
- ▶ sezione d'urto:  $0.04016 \pm 8.86e-5$  pb

MadGraph\_aMC@NLO  
(v 2.2.1)

La topologia studiata nel caso dell'analisi boosted ttH richiede uno stato finale in cui ci sia un quark top ( $p_T > 300$  GeV) e/o un bosone di Higgs ( $p_T > 250$  GeV) che decadono adronicamente.

dR(b,b) distribution vs Higgs pT

dR(b,W) distribution vs HadTop pT



# Studi preliminari Monte Carlo

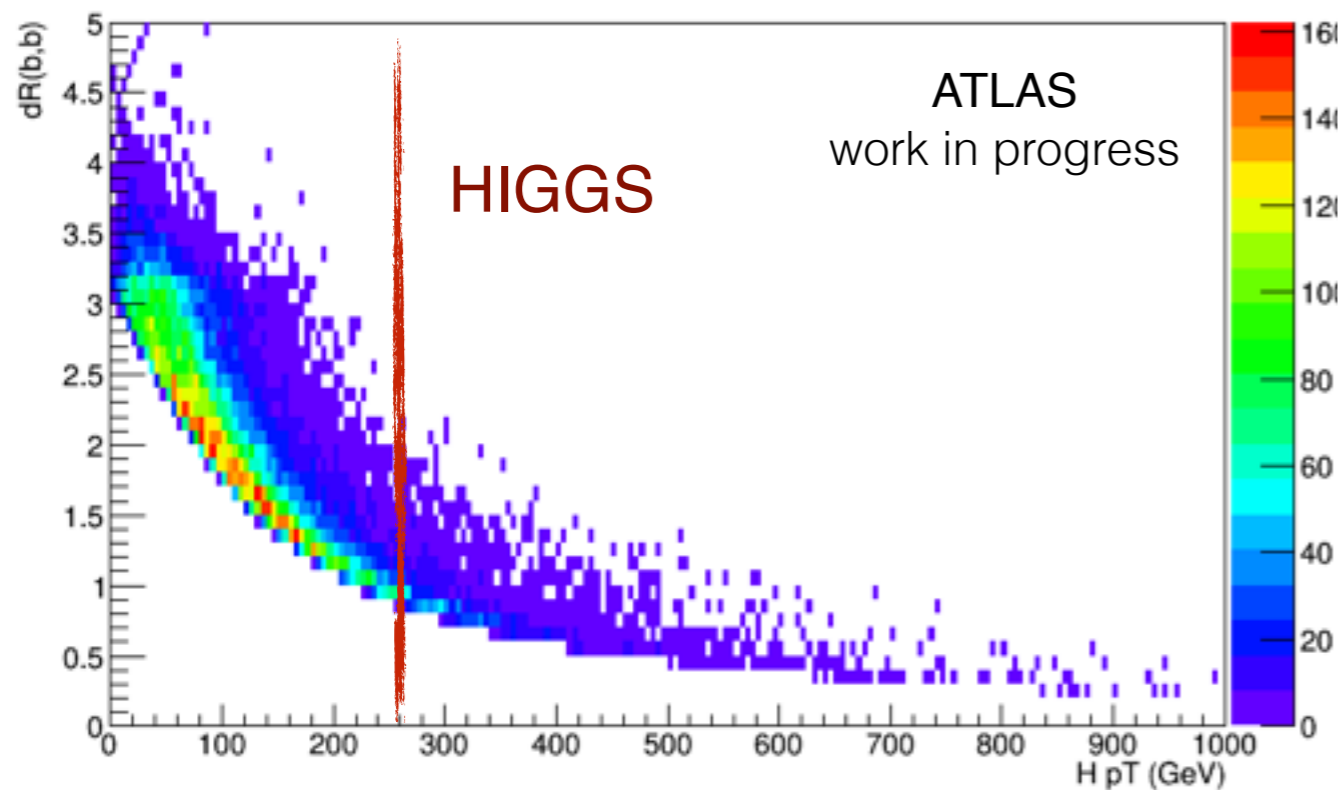
## ◆ dettagli del processo

- ▶ processo generato:  $pp \rightarrow t\bar{t}H \rightarrow l+jets+b\bar{b}$  ( $l = e, \mu$ )
- ▶ energia del fascio: 13 TeV
- ▶ numero di eventi: 30000
- ▶ sezione d'urto:  $0.04016 \pm 8.86e-5$  pb

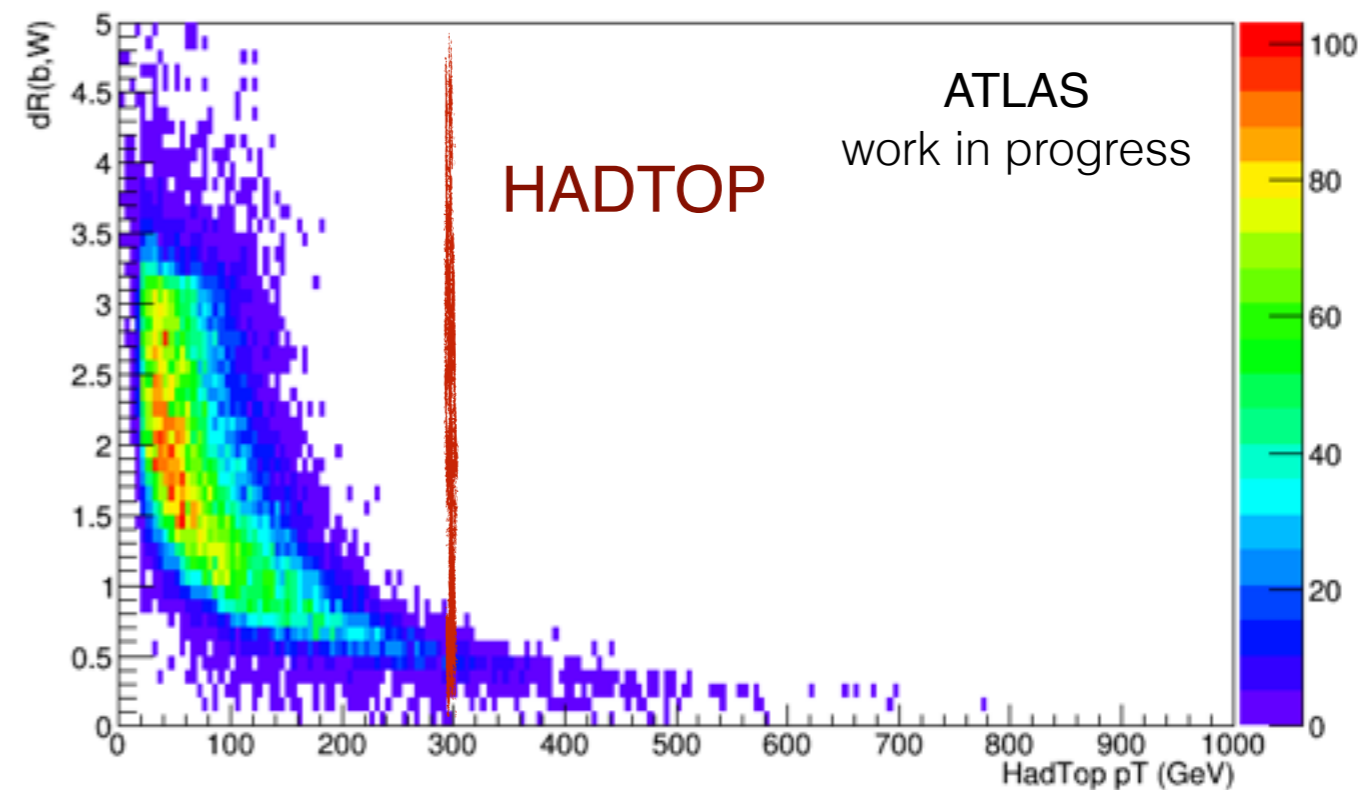
MadGraph\_aMC@NLO  
(v 2.2.1)

La topologia studiata nel caso dell'analisi boosted ttH richiede uno stato finale in cui ci sia un quark top ( $p_T > 300$  GeV) e/o un bosone di Higgs ( $p_T > 250$  GeV) che decadono adronicamente.

dR(b,b) distribution vs Higgs pT



dR(b,W) distribution vs HadTop pT





# Studi preliminari Monte Carlo

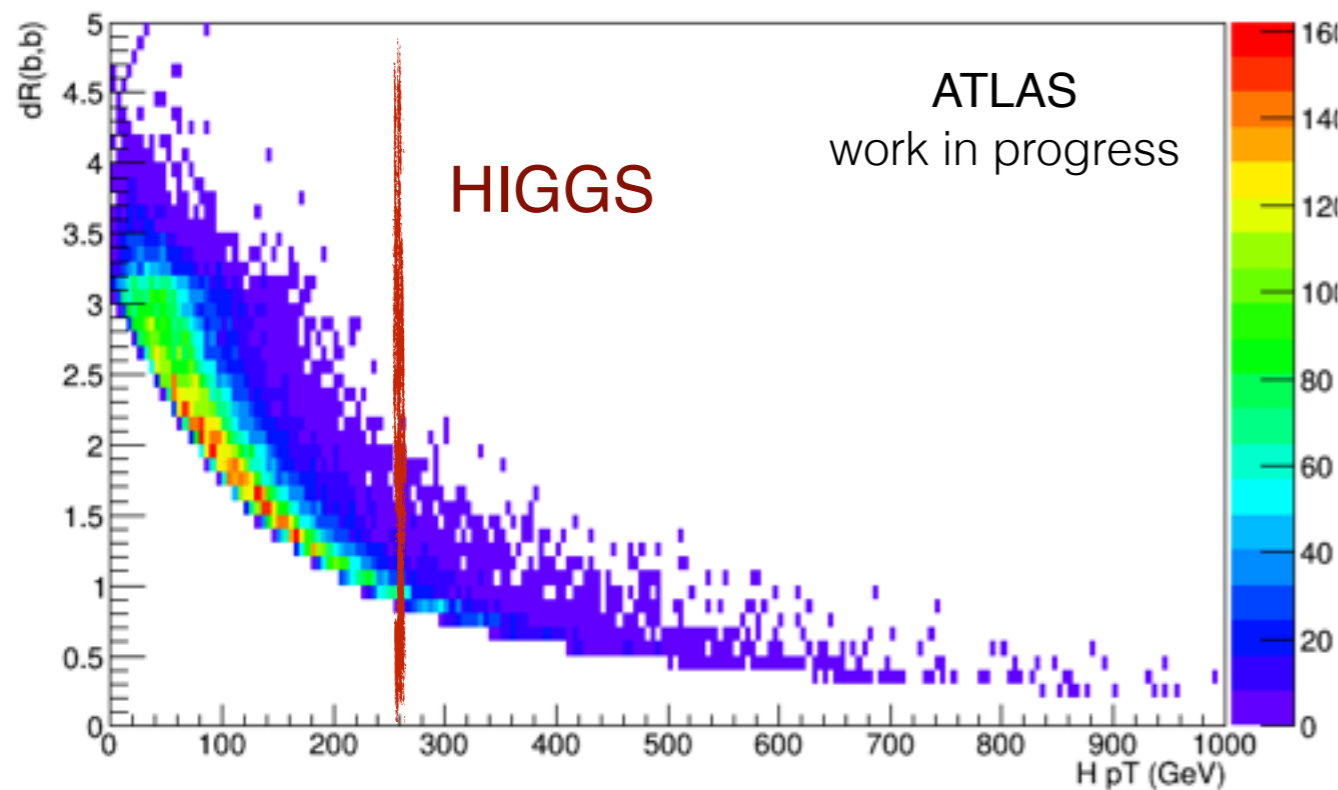
## ◆ dettagli del processo

- ▶ processo generato:  $pp \rightarrow t\bar{t}H \rightarrow l+jets+b\bar{b}$  ( $l = e, \mu$ )
- ▶ energia del fascio: 13 TeV
- ▶ numero di eventi: 30000
- ▶ sezione d'urto:  $0.04016 \pm 8.86e-5$  pb

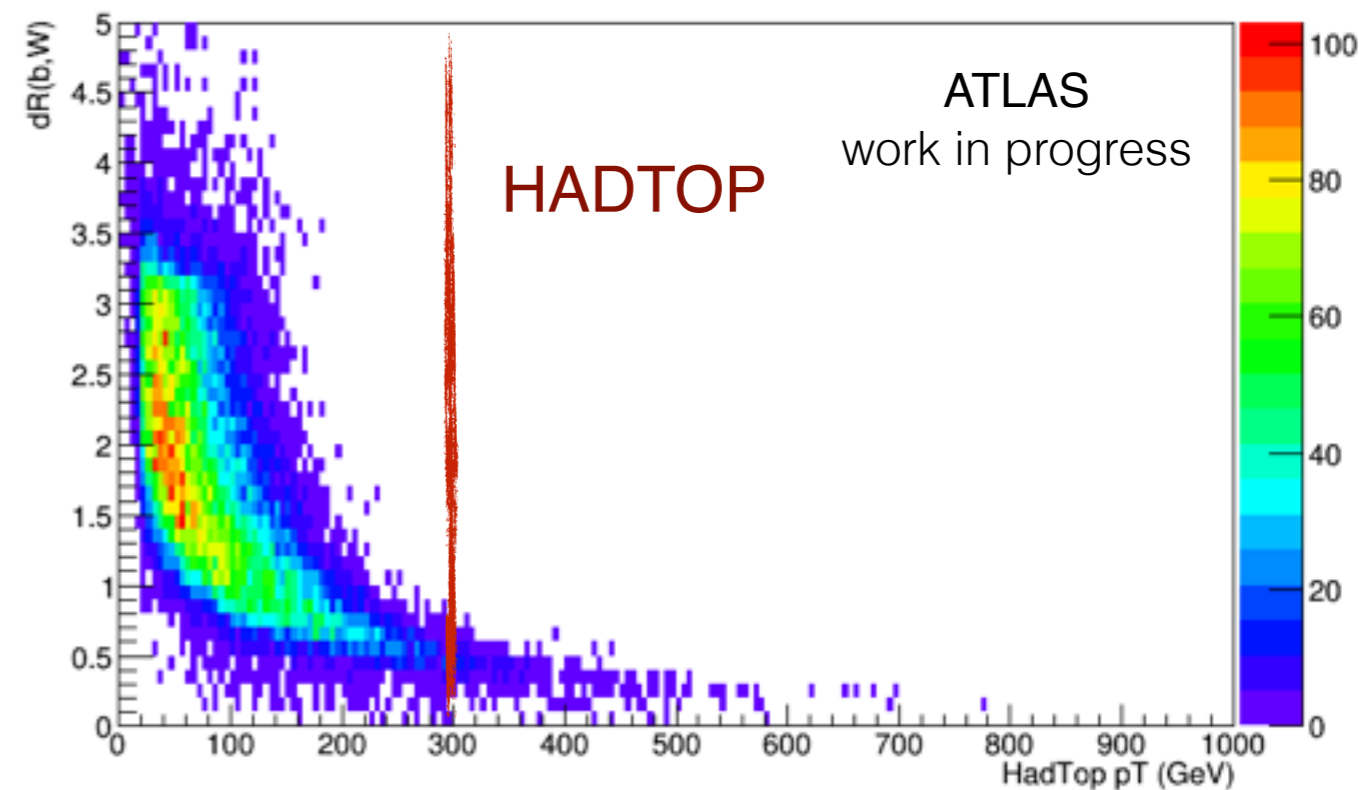
MadGraph\_aMC@NLO  
(v 2.2.1)

La topologia studiata nel caso dell'analisi boosted ttH richiede uno stato finale in cui ci sia un quark top ( $p_T > 300$  GeV) e/o un bosone di Higgs ( $p_T > 250$  GeV) che decadono adronicamente.

dR(b,b) distribution vs Higgs pT



dR(b,W) distribution vs HadTop pT



☹️ riduzione della statistica: ~ 20% rimanente

😊 studio di una zona estremamente interessante nel modo più efficiente e preciso possibile

# Tecniche di top-tagging

- Esistono diverse tecniche di top-tagging
- Studio preliminare a 13TeV è concentrato inizialmente su un semplice algoritmo che prevede dei tagli su due variabili del jet:

## La massa calibrata del jet

$$(m_{\text{jet}}^{\text{uncalib}})^2 = \left( \sum_i E_i \right)^2 - \left( \sum_i \vec{p}_i \right)^2$$

calibrazione necessaria a causa dei contributi a basso  $p_T$  e grande angolo

## Il rapporto $\tau_{32}$ tra le variabili N-subjettiness

$$\tau_N = \frac{1}{d_0} \sum_k p_{T_k} \times \min(\delta R_{1k}, \delta R_{2k}, \dots, \delta R_{Nk})$$

$\downarrow$  numero di subjet ricostruiti       $\rightarrow$  distanza tra subjet i e costituenti k

$$d_0 = \sum_k p_{T_k} \times R.$$

$$\rightarrow \tau_{32} = \tau_3 / \tau_2.$$

- La bontà di un algoritmo di top-tagging è data da:

## Efficienza di segnale

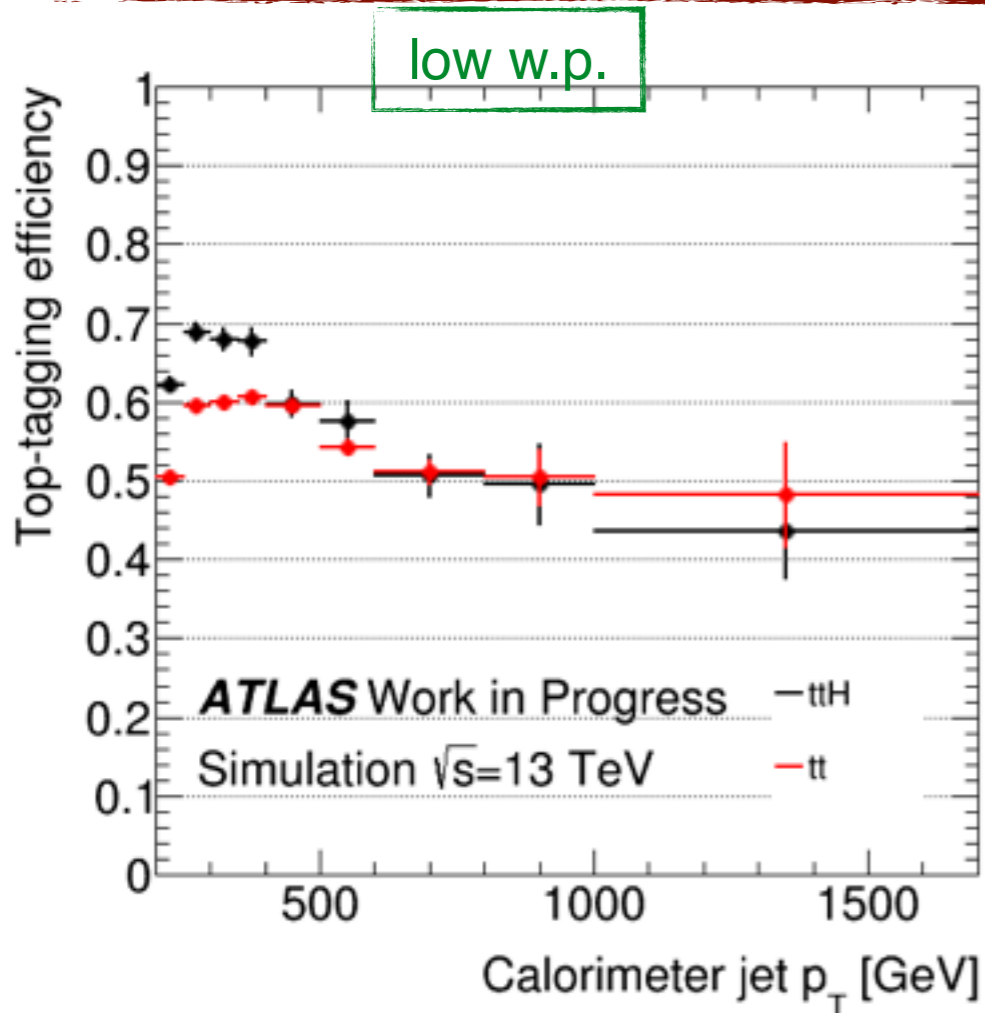
$$\epsilon = \left( \frac{N_{\text{tagged}}}{N_{\text{total}}} \right)_{\text{signal}}$$

## Reiezione del fondo

$$r = \left( \frac{N_{\text{total}}}{N_{\text{tagged}}} \right)_{\text{background}}$$

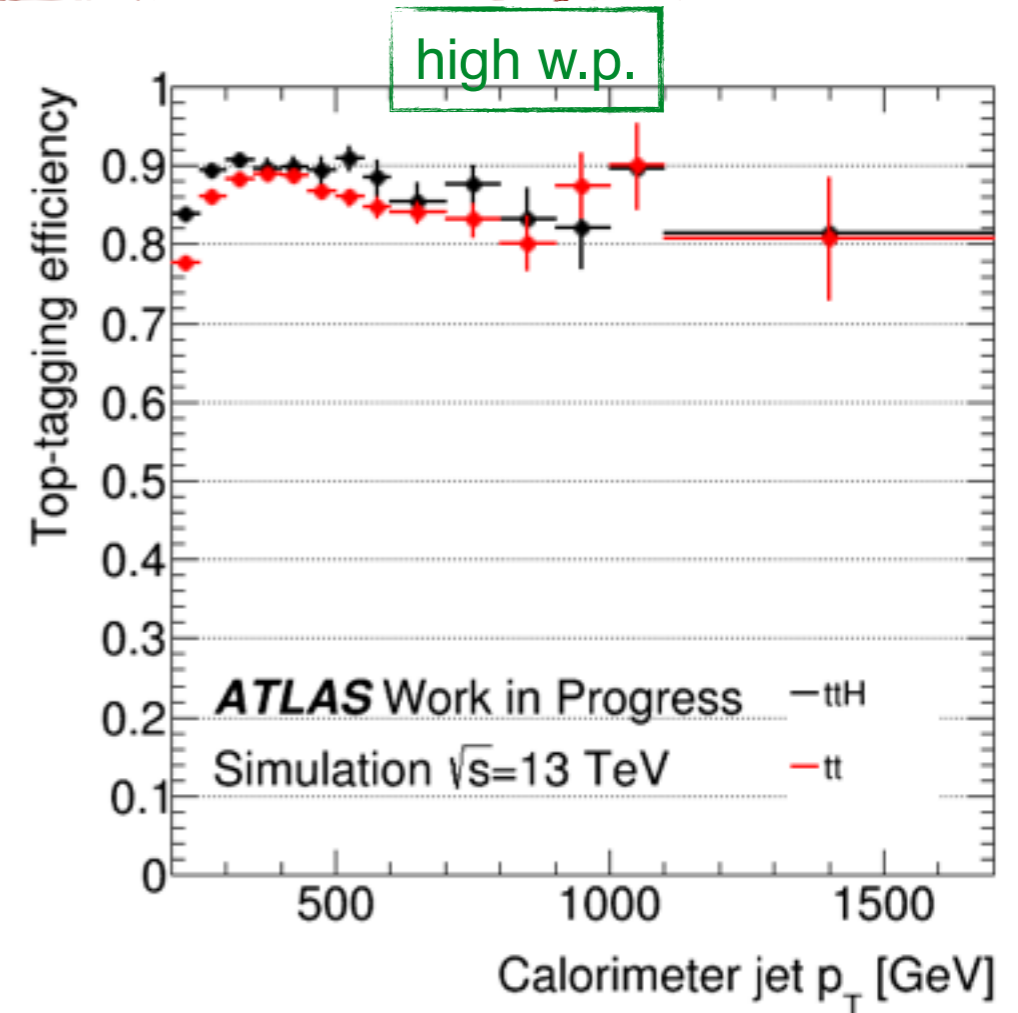
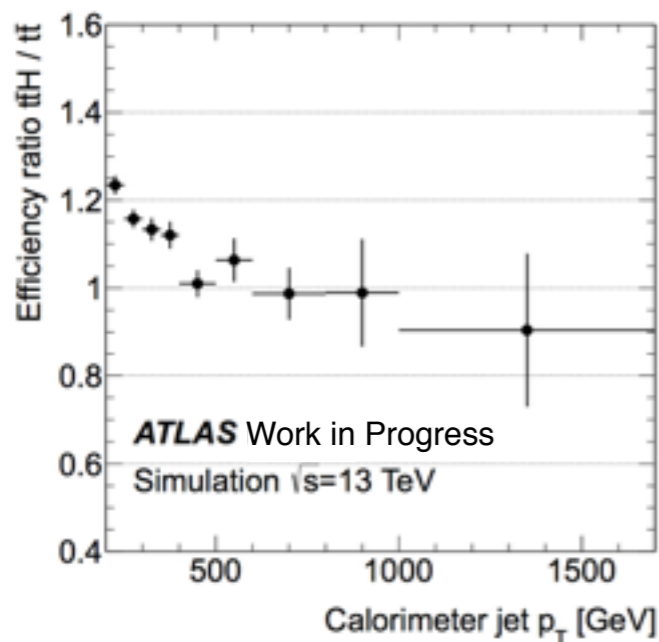
- Le tecniche di top-tagging sono state principalmente utilizzate per analisi di processi  $t\bar{t}$ .
- Confronto tra efficienze in  $t\bar{t}$  e  $t\bar{t}H$ .

# Prestazioni dell'algoritmo di top-tagging



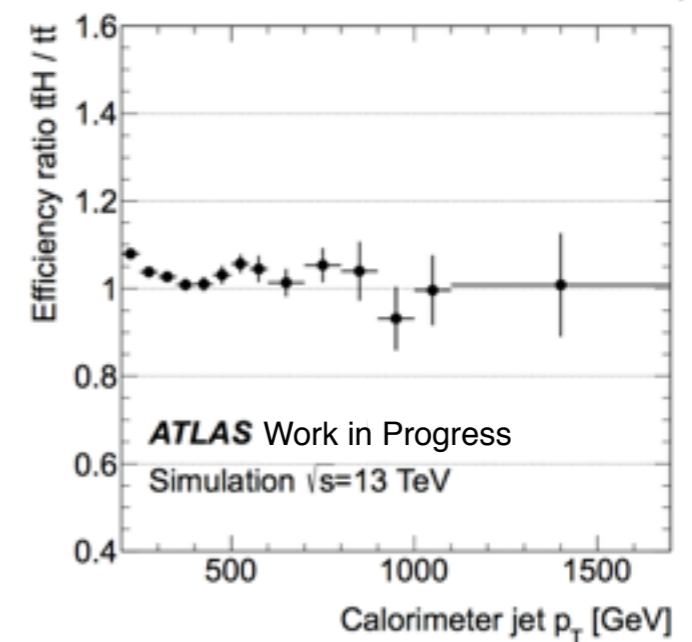
Due diversi punti di lavoro: a bassa e alta efficienza di segnale (valori indicativi calcolati ad alto  $p_T$ ).

Entrambi i punti di lavoro presentano un'efficienza di top-tagging più alta a  $p_T$  relativamente bassi per la configurazione  $t\bar{t}H$  considerata in questo studio.



Necessità di studi sulla capacità di reiezione del fondo

alta efficienza e alta reiezione



# Conclusioni e Piani futuri

## ★ “The importance of being boosted ttH” (~ Oscar Wilde)

- la sezione d'urto presenta un aumento maggiore rispetto a quello degli altri canali
  - 0.13 pb → 0.50 pb
- le **tecniche “boosted”** aiutano a risolvere il problema di:
  - elevato fondo combinatorio
  - perdita di efficienza degli algoritmi standard di ricostruzione

## ★ Studi fatti finora

- studi MC con una **generazione MC** personale
  - comprensione della topologia del canale
- prestazioni dell'**algoritmo di top-tagging nello stato finale ttH**
  - confronto con il fondo dominante (tt)

## ★ Piani futuri

- studiare i **primi dati a 13TeV**
- riprodurre i risultati ottenuti per l'**algoritmo di top-tagging**
- introdurre algoritmi di **top-tagging alternativi e boson-tagging**
- finalizzare una **strategia di analisi completa**





Grazie per  
l'attenzione!

# Backup

---

# Boosted $t\bar{t}H$ - Motivazioni

2009

## Fat Jets for a Light Higgs

Tilman Plehn,<sup>1</sup> Gavin P. Salam,<sup>2</sup> and Michael Spannowsky<sup>3</sup>

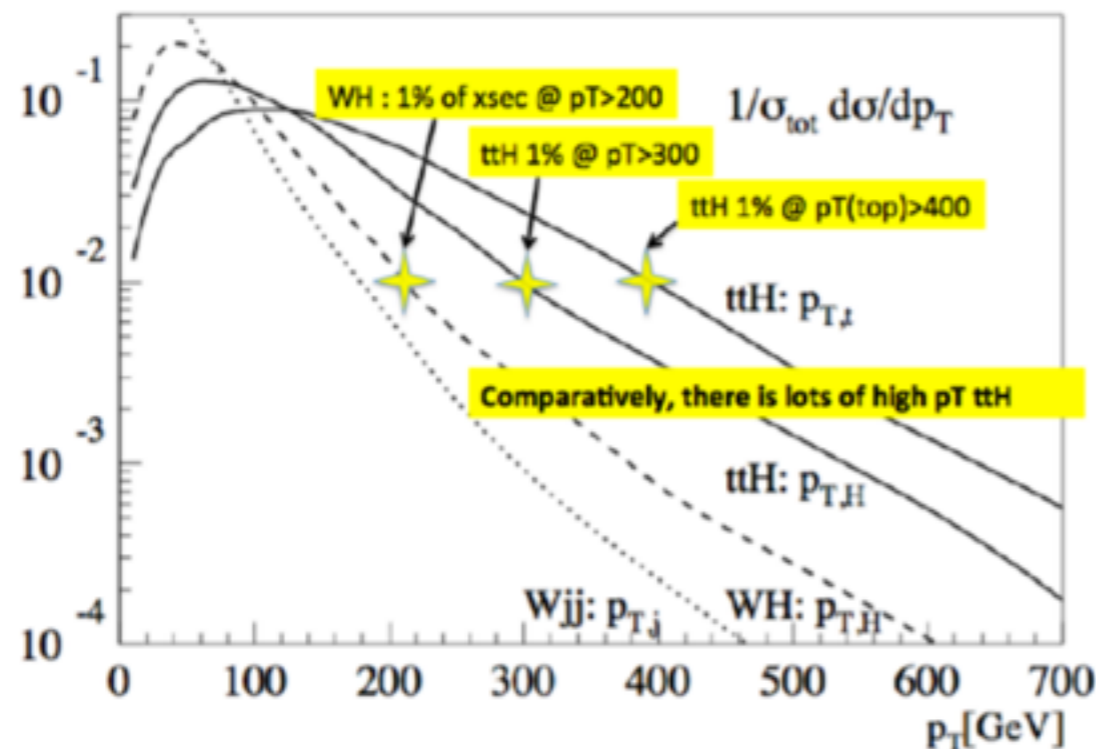
<sup>1</sup>*Institut für Theoretische Physik, Universität Heidelberg, Germany*

<sup>2</sup>*LPTHE, UPMC Univ Paris 6, CNRS UMR 7598, Paris, France*

<sup>3</sup>*Institut für Theoretische Physik, Universität Karlsruhe, KIT, Germany and  
Department of Physics and Institute of Theoretical Science, University of Oregon, Eugene, USA*

arXiv:0910.5472v2

At the LHC associated top quark and Higgs boson production with a Higgs decay to bottom quarks has long been a heavily disputed search channel. Recently, it has been found to not be viable. We show how it can be observed by tagging massive Higgs and top jets. For this purpose we construct boosted top and Higgs taggers for Standard Model processes in a complex QCD environment.

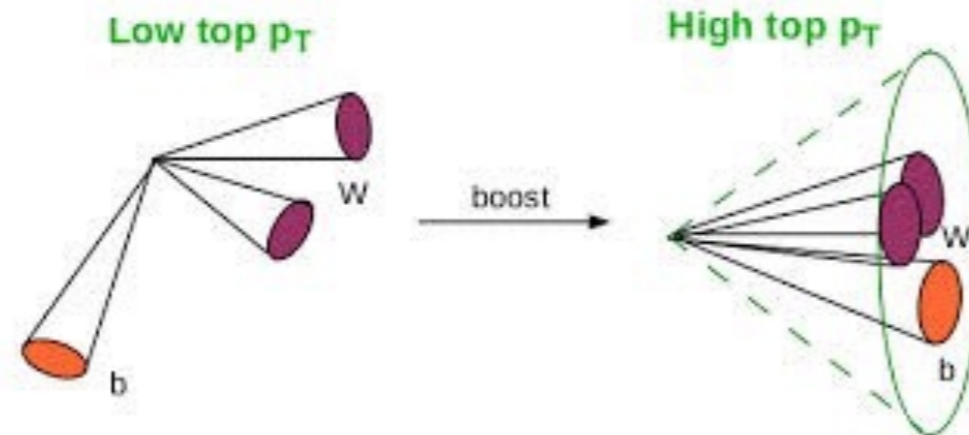


grande importanza di nuovi algoritmi di tagging sia per il top che per l'Higgs boosted che siano efficienti anche in eventi di LHC con un'alta concentrazione di particelle nello stato finale

FIG. 1: Normalized top and Higgs transverse momentum spectra in  $t\bar{t}H$  production (solid). We also show  $p_{T,H}$  in  $W^-H$  production (dashed) and the  $p_T$  of the harder jet in  $W^-jj$  production with  $p_{T,j} > 20$  GeV (dotted).

# Boosted $t\bar{t}H$ - Motivazioni

- **Boosted regime:** contaminazione dei jet  
algoritmi standard perdono efficienza



separazione dei prodotti di decadimento

$$\Delta R \approx \frac{2m}{p_T}$$

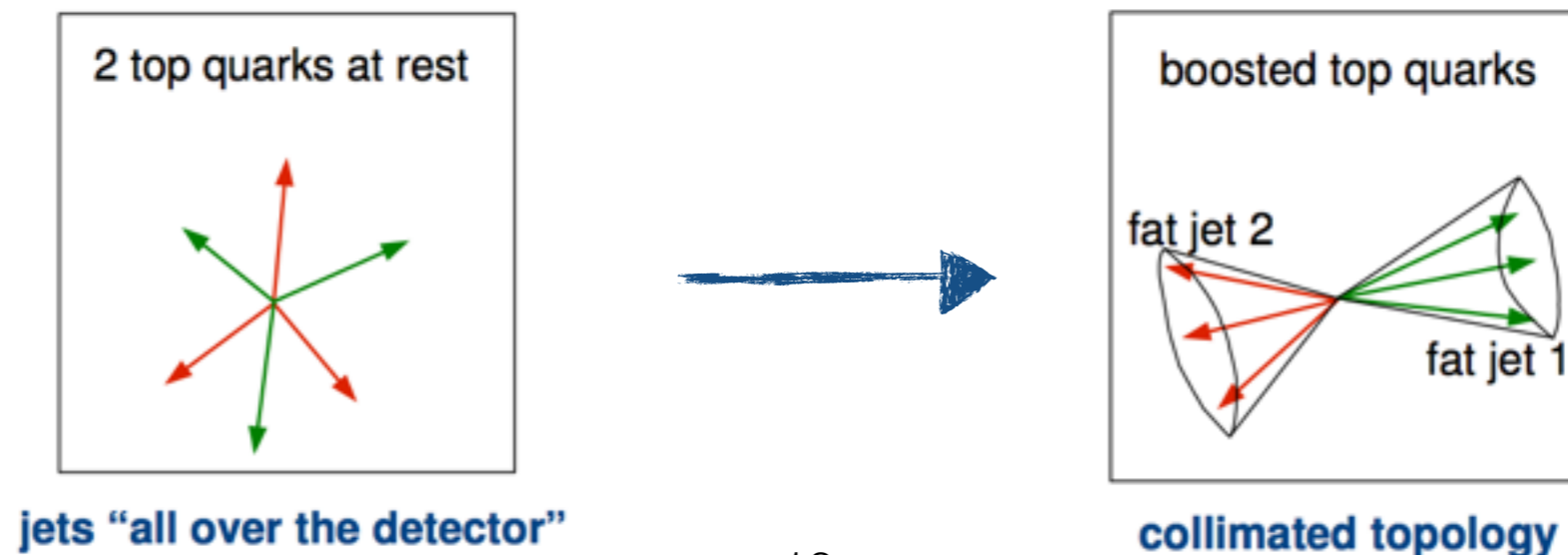
$$\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$$

→ soluzione: considerare un jet unico ("fat jet" o "large-R jet")

- **Sistema  $t\bar{t}H$ :** alta concentrazione di particelle di decadimento nello stato finale  
fondo combinatorio molto grande



→ soluzione: boosted regime  
considerare un jet unico





# Generazione MC: informazioni preliminari

generator: MadGraph\_aMC@NLO  
(v 2.2.1)

## ◆ dettagli del processo

- ▶ processo generato:  $pp \rightarrow t\bar{t}H \rightarrow l+jets+b\bar{b}$  ( $l = e, \mu$ )
- ▶ energia del fascio: 13 TeV
- ▶ numero di eventi: 30000
- ▶ sezione d'urto:  $0.04016 \pm 8.86e-5$  pb
  - ▶ somma dei 4 processi:

$p \rightarrow gg, t \rightarrow l\nu b + t \rightarrow qq b + H \rightarrow bb$   
 $p \rightarrow gg, t \rightarrow qq b + t \rightarrow l\nu b + H \rightarrow bb$   
 $p \rightarrow qq, t \rightarrow l\nu b + t \rightarrow qq b + H \rightarrow bb$   
 $p \rightarrow qq, t \rightarrow qq b + t \rightarrow l\nu b + H \rightarrow bb$

from MadGraph output

Graph	Cross-Section ↓	Error
P0 sum	0.04015953483	
<a href="#">P0_gg_ttxh_t_wpb_wp_lvl_tx_wmbx_wm_qq_h_bbx</a>	0.01456	5.15e-05
<a href="#">P0_gg_ttxh_t_wpb_wp_qq_tx_wmbx_wm_lvl_h_bbx</a>	0.01456	5.41e-05
<a href="#">P0_qq_ttxh_t_wpb_wp_qq_tx_wmbx_wm_lvl_h_bbx</a>	0.005526	3.39e-05
<a href="#">P0_qq_ttxh_t_wpb_wp_lvl_tx_wmbx_wm_qq_h_bbx</a>	0.005515	3.36e-05

## ▶ tagli a livello di generazione:

- $P_{Tjets} > 20$  GeV
- $P_{Tb} > 20$  GeV
- $P_{Tlep} > 20$  GeV
- $|\eta| < 2.5$

# Jets inputs

---

Three types of inputs are given to jet reconstruction:

- **3D topological clusters** (topo-clusters) are used to build fully reconstructed **calorimeter jets**. Calorimeter cells are first clustered together using a three-dimensional topological clustering algorithm that includes noise suppression. The resulting topo-clusters are considered as massless four-momenta. They are classified as either electromagnetic or hadronic based on their shape, depth and energy density. In the calibration procedure, corrections are applied to the energy in order to calibrate the clusters to the hadronic scale.
- **Stable particles** with a lifetime of at least 10 ps (excluding muons and neutrinos) are used in simulated samples to build **generator-level jets**. These particles retain their respective masses when input to the jet reconstruction algorithm.
- **Charged particle tracks** are used to build **track jets**. Tracks are assigned the pion mass when used as input to the jet algorithm. Quality selection criteria are applied in order to ensure that good quality tracks originating from the reconstructed hard scattering vertex are used to build track-jets. The requirements are that the track should have transverse momentum greater than 0.5 GeV, transverse impact parameter less than 1.0 mm, longitudinal impact parameter less than 1.0 mm, and at least one associated hit in the pixel detector and six associated hits in the silicon strip detector. The hard scattering vertex is selected as the primary vertex that has the largest  $\Sigma(p_T)^2$  in the event and that contains at least two tracks. This type of jet provides information independent of the calorimeter systematic uncertainties.

# Anti- $k_t$ and $k_t$ algorithms

- The iterative recombination procedure works by first cleaning a list of all objects (either hadrons, topo-clusters or tracks) in an event.
- The ordering of the list is irrelevant and proto-jets are built from these objects.
- Two distance measures in  $y$ - $\phi$ -space are associated to each member of the list, between the proto-jet and its closest neighbor:

$$\rho_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{(\Delta R_{ij})^2}{R^2}$$

and between the proto-jet and the beam:

$$\rho_{iB} = p_{Ti}^{2p}$$

measure of the opening angle  
between the two constituents

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$$

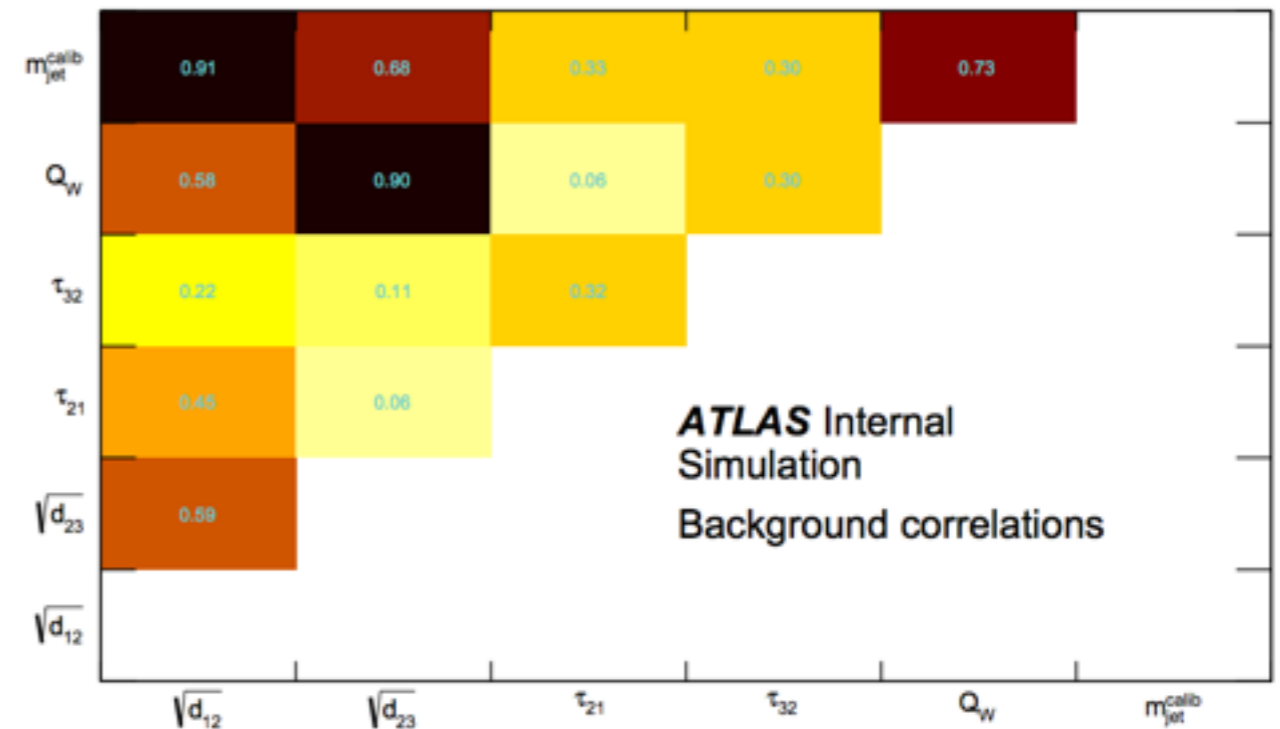
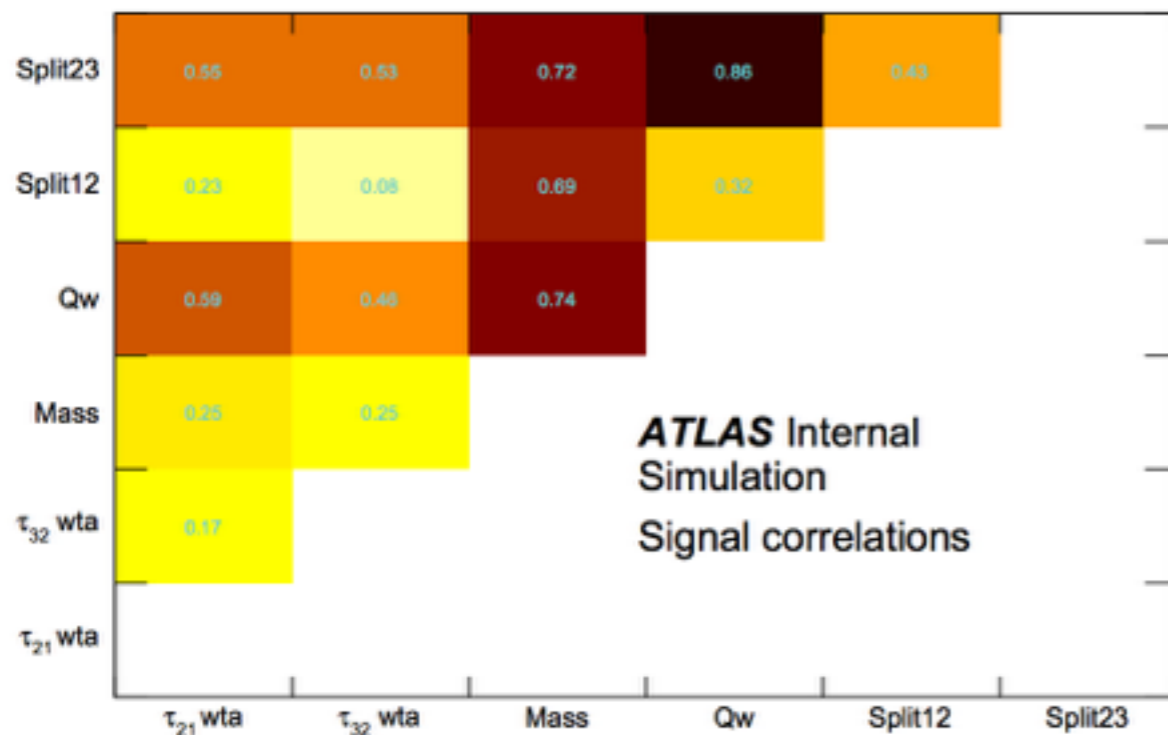
- ➔ If  $\rho_{iB} < \rho_{ij}$ : the proto-jet is closer to the beam than it is to any other proto-jet in the event, so it is defined as a jet and removed from the list.
- ➔ If  $\rho_{iB} > \rho_{ij}$ : the two proto-jets  $i$  and  $j$  are combined into one, thereby forming a new proto-jet. This procedure continues through all proto-jets in the event.
- ◆ If  $p = +1$   $\longrightarrow$   **$k_t$  algorithm**: proto-jets with the smallest  $p_T$  tend to be clustered first, so that the highest  $p_T$  proto-jets are clustered last.
- ◆ If  $p = -1$   $\longrightarrow$  **anti- $k_t$  algorithm**: proto-jets with the largest  $p_T$  are clustered first. A consequence of this is that isolated anti- $k_t$  jets tend to be very close to circular in  $\eta$ - $\phi$  space, because the axis of the jet is relatively fixed after the first few steps of recombination. This stability makes anti- $k_t$  jets more robust than  $k_t$  jets in high multiplicity environments.

# Choosing the variables

The two variables  $m_{\text{jet}}^{\text{calib}}$  and  $\tau_{32}$  were chosen from a set of substructure variables, including other N-subjettiness ratio ( $\tau_{21}$ ), splitting scale variables ( $\sqrt{d_{12}}, \sqrt{d_{23}}$ ) and the minimum dijet mass from the three subjets ( $Q_w$ ).

- The two chosen variables show a **good background rejection** at 50% and 80% signal efficiency.
- The  $\tau_{ij}$  variables are **uncorrelated** with respect to the mass and energy scale variables.

This combination of strong performance and lack of correlated behavior motivates the choice of tagging variables.



Correlation matrices for the different studied variables, for signal (left) and background (right).