

# Understanding boosted top tagging with N-subjettiness and Prong Finding

J. Helliwell  
with: M. Dasgupta

Work in progress



BOOST  
2021

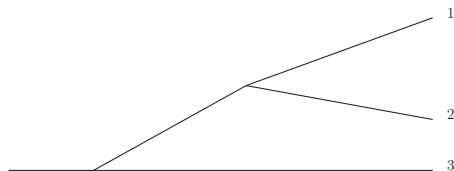
- 1 Introduction and motivation
- 2 Tagger definitions
- 3 Monte Carlo study
- 4 Background calculations
- 5 Signal calculations
- 6 What have we learnt?

- Tagging hadronically decaying top quarks, reconstructed as a single jet.
- Grooming, prong finding and jet shapes used in combination ((ATLAS 2016, CMS 2016))
- Understand what drives this type of tagging procedure
- Focus on  $Y_m$ -Splitter (M Dasgupta, M Guzzi, J Rawling, G Soyez, 2018) with a cut on the N-subjettiness ratio  $\tau_{32}$  (J. Thaler, K, Tilburg 2011), and grooming with mMDT/ Soft drop (M. Dasgupta, A Fregoso, S Marzani, G Salam, 2013) (A. Larkoski, S Marzani, G Soyez, J. Thaler, 2014)



figure adapted from arXiv:1909.12285FERMILAB-PUB-19-492-CMS-E

- Cluster jet using Gen- $k_t(p = 1/2)$  (M. Cacciari, G. Salam, G. Soyez, 2011)  $d_{ij} = \min(z_i, z_j)\theta_{ij}^2$
- Undo last **two** clusterings

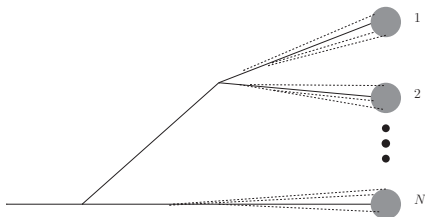


$$\min(m_{12}, m_{13}, m_{23}) > m_{\min}$$

$$m_{\min} \lesssim m_w$$

$$\min(z_1, z_2, z_3) > \zeta$$

Measures the radiation pattern around  $N$  "prongs"



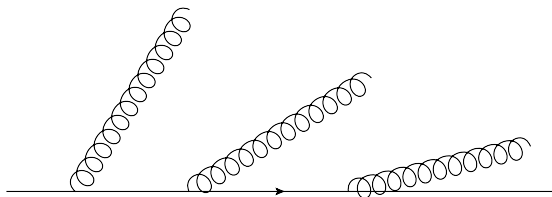
$$\tau_N^{(2)} \simeq \sum_i z_i \min((\theta_{1,i})^2, (\theta_{2,i})^2, \dots, (\theta_{N,i})^2)$$

Use Gen- $k_t$  ( $p = 1/2$ ) axes  
 $(d_{ij} = \min(z_i, z_j)\theta_{ij}^2)$

$$\tau_{32}^{(2)} = \frac{\tau_3^{(2)}}{\tau_2^{(2)}}$$

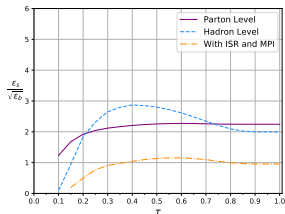
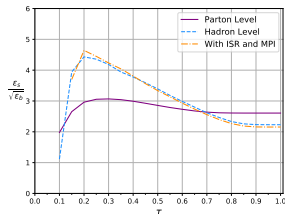
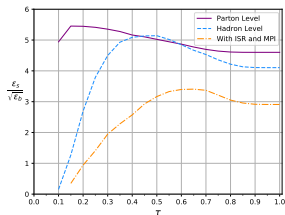
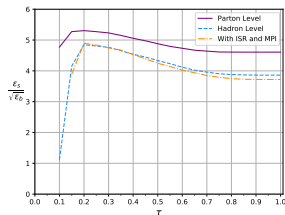
- 1 Cluster jet using Cambridge Aachen algorithm
- 2 Undo last clustering
- 3 Check if lower  $p_t$  sub-jet has energy fraction,  $z > z_{cut}\theta^\beta$ . If not, discard this sub jet and go back to 2.
- 4 If it does, this is the jet.

Use  $z_{cut} = \zeta = 0.05$  throughout this work



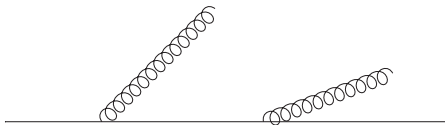
- Performance
- Impact of non-perturbative effects
- Using Pythia 8
- $160 \text{ GeV} < m_{\text{jet}} < 225 \text{ GeV}$

## Signal significance


 Cut on  $\tau_{32}$ 

 Cut on  $\tau_{32}$  after grooming with mMDT.

 Cut on  $\tau_{32}$  after application of  $Y_m$ -Splitter.

 Cut on  $\tau_{32}$  after application of  $Y_m$ -Splitter and grooming with mMDT.



- $Y_m$ -Splitter requires 3 particles (2 emissions)
- $\tau_{32} = 0$  at  $\mathcal{O}(\alpha_s^2)$
- $\rho = \frac{m^2}{p_t^2 R^2}$
- double logarithmic accuracy



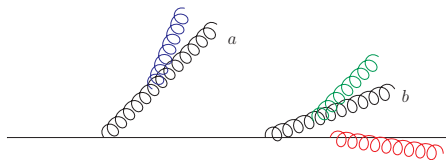
$$\rho \frac{d\Sigma}{d\rho} \stackrel{\zeta > \frac{\rho_{min}}{\rho}}{=} \left( \frac{\alpha_s C_F}{\pi} \right)^2 \ln \left( \frac{\rho}{\rho_{min}} \right) \ln^2(\zeta).$$

$$\stackrel{\zeta < \frac{\rho_{min}}{\rho}}{=} \left( \frac{\alpha_s C_F}{\pi} \right)^2 \ln^2 \left( \frac{\rho_{min}}{\rho} \right) \left( \frac{3}{2} \ln(\zeta) + \ln \left( \frac{\rho_{min}}{\rho} \right) \right)$$

(M Dasgupta, M Guzzi, J Rawling, G Soyez, 2018)

- Leading log accuracy
- Strong ordering in angle and mass
- $\tau_2 = \min(z_1\theta_1^2, z_2\theta_2^2)$ ,  $\tau_3 = \sum_{i=3}^{\infty} z_i\theta_i^2$
- LO result multiplied by Sudakov factor

$$\rho_i = z_i\theta_i^2 \quad \rho_a = \max(z_1\theta_1^2, z_2\theta_2^2) \quad \rho_b = \min(z_1\theta_1^2, z_2\theta_2^2)$$



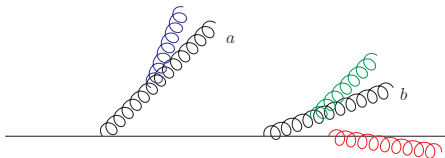
$$S^{(\text{fixed-coupling, soft})} = \exp \left[ -\frac{C_F\alpha_s}{2\pi} \ln^2 \frac{1}{\tau\rho_b} - \frac{C_A\alpha_s}{2\pi} \ln^2 \frac{\rho_a}{\tau\rho_b} - \frac{C_A\alpha_s}{2\pi} \ln^2 \frac{1}{\tau} \right],$$

Accuracy goal of  $\alpha_s^n L_\rho^n f_n(\tau)$ .

$L_\rho$  is a log or a mass scale i.e.  $\ln(\rho)$  or  $\ln(\rho_{\min})$

Strong ordering in angle only

$$\tau_{32}(\{p_i\}) = \frac{\sum_{i=1} \rho_i - \rho_a - \rho_b}{\sum_{i=1} \rho_i - \rho_a}$$



Result valid for  $\tau < \frac{1}{2}$

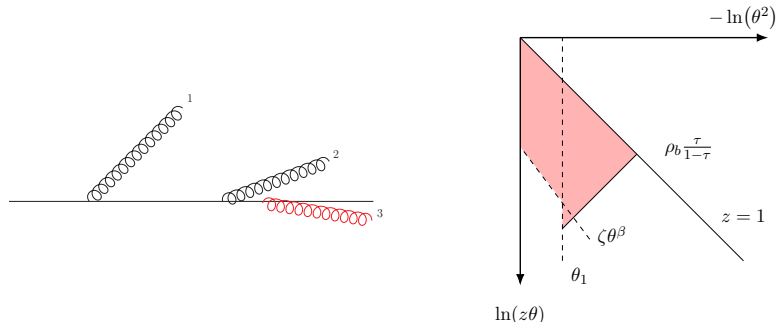
$$\frac{\rho}{\sigma} \frac{d\sigma}{d\rho} \stackrel{\tau < 1/2}{=} \text{LO} \times {}_2F_1(1, R', 1 + R', \tau) \frac{\exp[-R - \gamma_E R']}{\Gamma[1 + R']}$$

$$R^{\text{fixed coupling}} = \frac{\alpha_s C_F}{2\pi} \ln^2 \left( \rho_b \frac{\tau}{1 - \tau} \right)$$

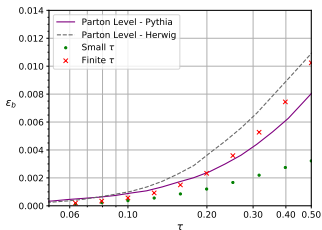
Following method of (D. Napoletano, G. Soyez: 2018)

## Background: Grooming

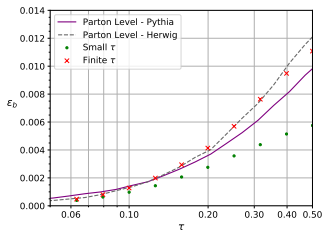
- Sudakov modified as soft wide angle emissions removed
- Complicated by emission 3 setting  $\tau_3$  but emission 1 being able to stop the groomer
- Grooming only affects primary emissions (Angular ordering)
- Removes sensitivity to most non perturbative region of phase space



$$R^{(\text{primary})}\left(\rho_b \frac{\tau}{1-\tau}, \theta_1\right) = R_{\text{SD}}\left(\rho_b \frac{\tau}{1-\tau}\right) + R_{\text{SD}}^{\text{angle}}\left(\theta_1, \rho_b \frac{\tau}{1-\tau}\right).$$



No Grooming.



With  $\beta = 2$  soft drop

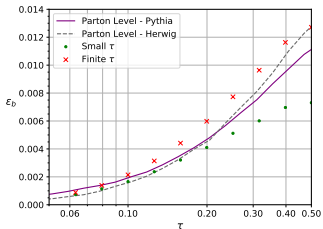
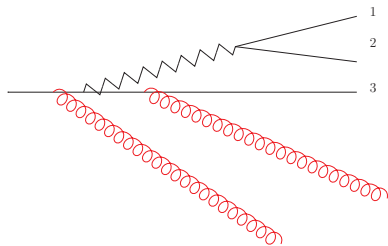


Figure 5: With mMDT

Key difference between signal and background:

- No soft/collinear divergences at LO
- At LO invariant mass is just the top mass



$$\Sigma^{\tau < \frac{1}{2}}(\rho_{\max}, \rho_{\min}, \zeta, \tau) =$$

$$\frac{1}{\sigma_0} \int |M_{t \rightarrow bq\bar{q}}|^2 d\Phi_3 \delta\left(\frac{s_{123}}{R^2 p_t^2} - \rho_t\right) \Theta_{\text{Clust}} \Theta_{Y_m\text{-Splitter}} \frac{e^{-R - \gamma_E R'}}{\Gamma[1 + R']}$$

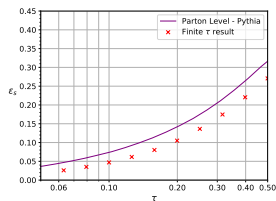
$$R = \frac{\alpha_s C_F}{2\pi} \ln^2 \left( \min \left( \frac{\tau}{1 - \tau} \min(d_{12}, d_{13}, d_{23}), \rho_{\max} - \rho_t \right) \right)$$

Condition for jet mass constraint to be stronger than the cut on  $\tau_{32}$ :

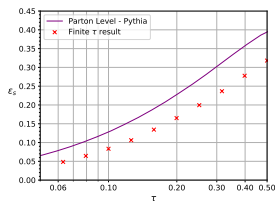
$$M_{\max}^2 < M_T^2 + \frac{\tau}{1 - \tau} \min(d_{12}, d_{13}, d_{23}) p_T^2.$$

For for the cut  $\tau_{32} < 0.3$  the jet mass constraint is only dominant for

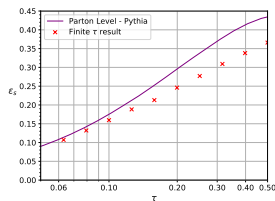
$$M_{\max} \lesssim 181 \text{ GeV}$$



No Grooming.

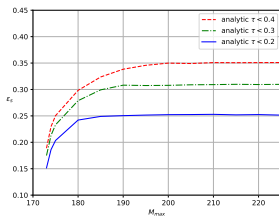


With  $\beta = 2$  soft drop

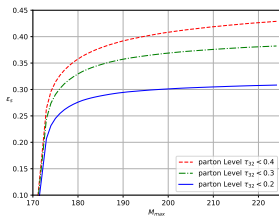


With mMDT

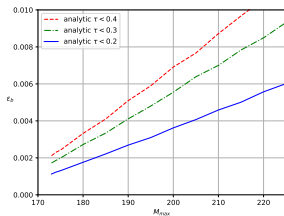
- Reducing  $M_{\max}$  suppresses the background with very little effect on the signal



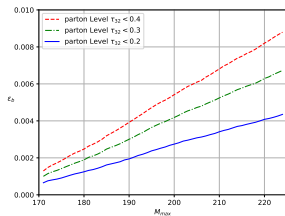
Top jets - Analytic



Top Jets - Pythia



Light quark jets - Analytic



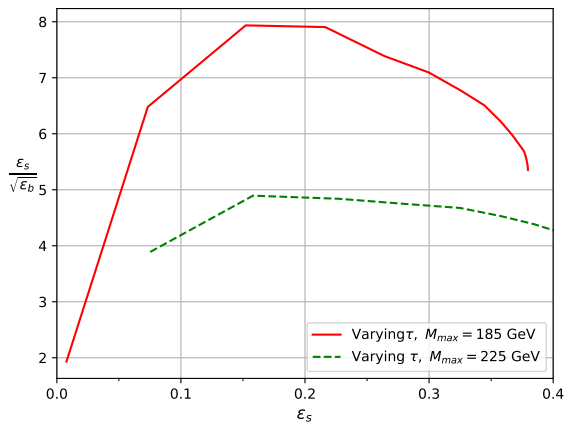
Light quark jets - Pythia

Jets Pre-groomed with mMDT.



## Making the most of the jet mass

- This holds up at hadron level, leading to gains in signal significance.
- Hadronisation corrections  $< 15\%$  for  $M_{\max} = 185$  GeV and  $\tau = 0.4$ .



Hadron level jets with UE from Pythia, pre-groomed with mMDT, tagged with  $Y_m$ -Splitter and a cut on  $\tau_{32}$

- $m\text{MDT} + Y_m\text{-Splitter} + \tau_{32}$  is an effective top-tagger and resilient to non-perturbative effects.
- Performed analytic calculations for both signal and background distributions
- Understood the interplay between mass and  $\tau_{32}$  cuts allowed us to use the mass cut to greatest effect.