MANCHESTER
1824

The University of Manchester

Understanding top taggers

Mrinal Dasgupta

PSR 2019 Vienna

Based on work with Gregory Soyez, Marco Guzzi, Jacob Rawling JHEP 1809 2018

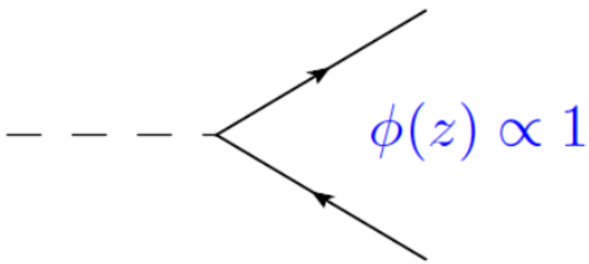
European Research Council
Established by the European Commission

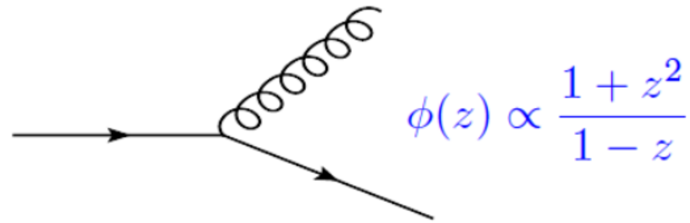
Outline

- Tools and lessons from W/Z/H two-pronged substructure
- CMS top tagger and issues
- Modified top taggers and Y_m splitter
- Analytics for top tagging
- Comparisons to parton showers
- Conclusions and prospects

Aim : Identify the main physics principles that govern performance using resummation and showers

Tool types

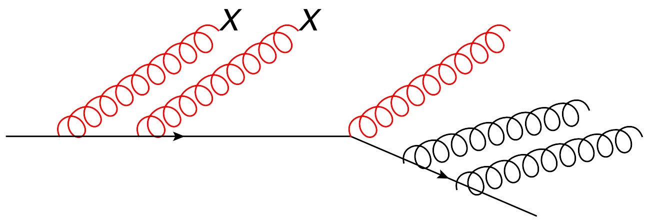


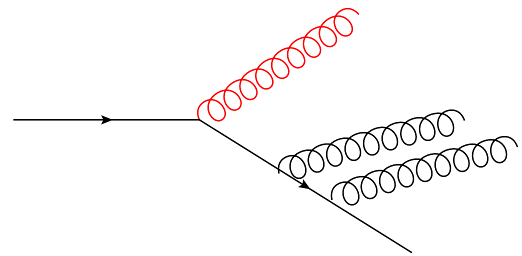


- Prong finders : find N=2,3,… hard prongs. Look to exploit differences in splitting functions. Different ways of defining asymmetry variable z.
- Groomers : remove uncorrelated radiation from jets.
- Radiation constrainers e.g. jet shapes : exploit differences in radiation patterns.

Impact of prong finders

Consider two distinct types :



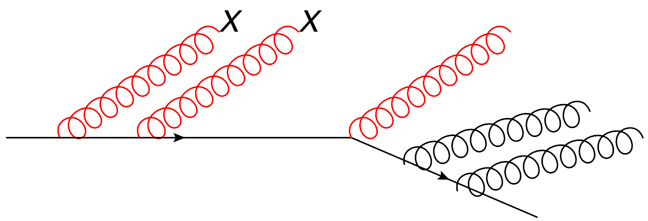


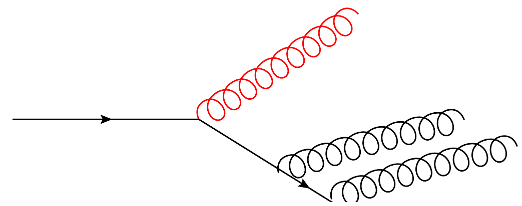
mMDT/SoftDrop
CA declustering and recurses through jet until finds branching with $z > \zeta_{\rm cut}$

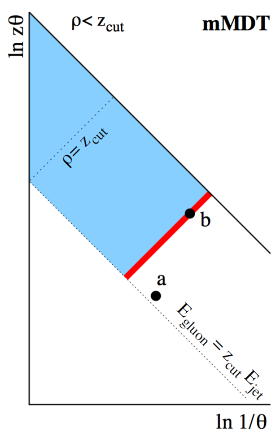
Ym-splitter
gen-kt (p=1/2) declustering and examines 1st emission only.

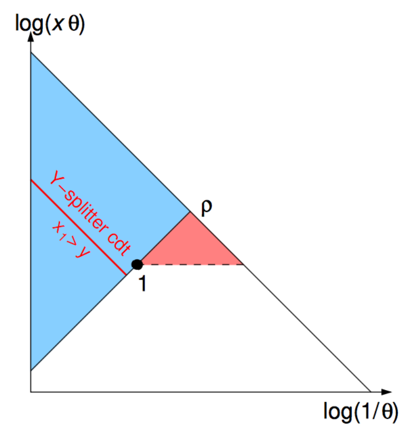
$$\left(\rho \frac{d\sigma}{d\rho}\right)^{\rm LO} = \frac{C_F \alpha_s}{\pi} \ln \frac{1}{\zeta_{\rm cut}} \qquad \rho = \frac{m^2}{p_T^2}$$

mMDT/SoftDrop vs Ym-splitter







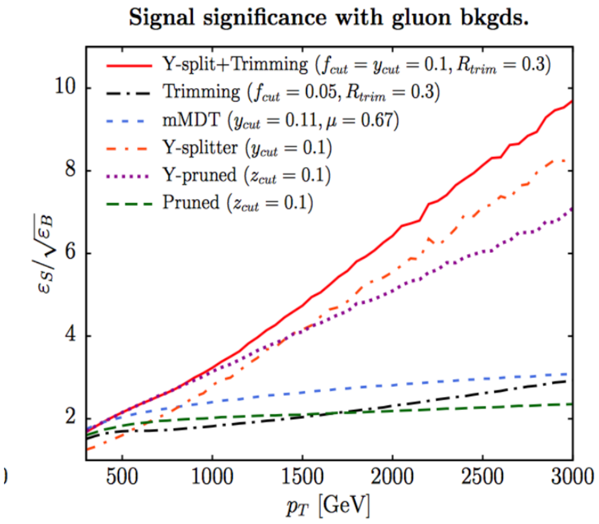


Large Sudakov suppression

$$\rho \frac{d\sigma}{d\rho}^{\mathrm{mMDT}} = \frac{C_F \alpha_s}{\pi} \ln \frac{1}{\zeta_{\mathrm{cut}}} \exp \left[-\frac{C_F \alpha_s}{\pi} \ln \frac{1}{\zeta_{\mathrm{cut}}} \ln \frac{1}{\rho} \right]$$

$$\rho \frac{d\sigma}{d\rho}^{\mathrm{Ym-splitter}} = \frac{C_F \alpha_s}{\pi} \ln \frac{1}{\zeta_{\mathrm{cut}}} \exp \left[-\frac{C_F \alpha_s}{2\pi} \ln^2 \frac{1}{\rho} \right]$$

Performance



MD, Powling, Siodmok 2016

- Ym-splitter needs to be supplemented by grooming to improve signal efficiency.
- Gives important performance gains relative to other methods due to Sudakov.
- Large hadronisation ~ 40% effects **help performance**

The University of Manchester

MANCHESTER 1824

Analytics for top taggers

- Want to identify the main relevant physics effects. Start with the CMS tagger and Y-splitter (used in early ATLAS top tagger).

CMS-PAS-JME-09-001, CMS-PAS-JME-13-007
ATL-COM-PHYS-2008-001

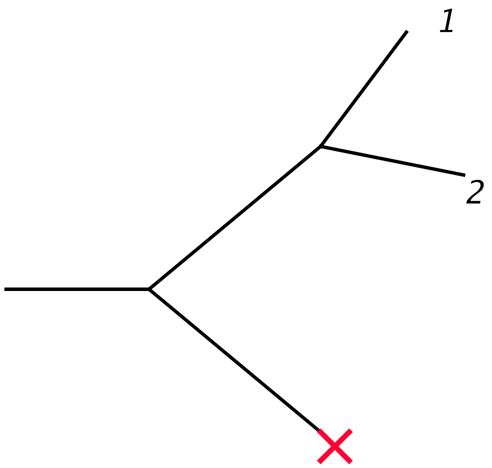
- CMS tagger descends from JH top tagger.

Kaplan, Rehermann, Schwartz and Tweedie 2008

- Both CMS tagger and Y-splitter offer ways of identifying three prongs relevant to top decays.

CMS top tagger

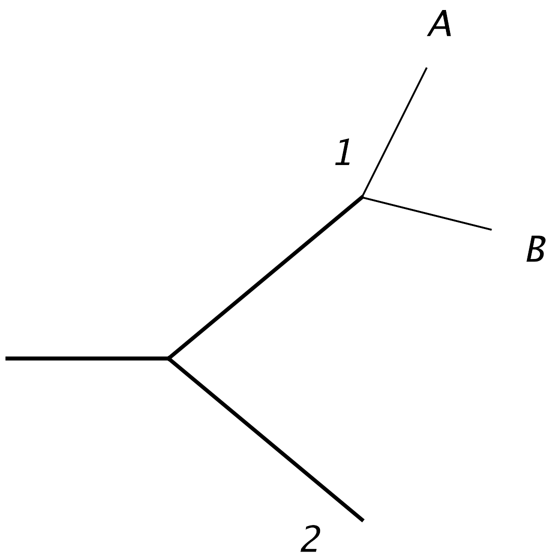
Primary Decomposition

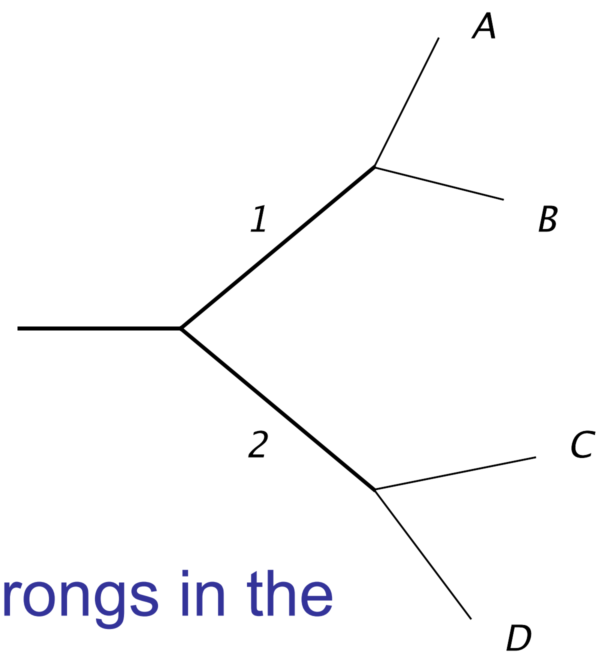


- Perform a C/A de-clustering of the jet and find two prongs.
- Use condition $p_t^{\rm prong} > \zeta_{\rm cut}\, p_t$ where p_t is jet rather than local p_t

CMS top tagger

Secondary decomposition

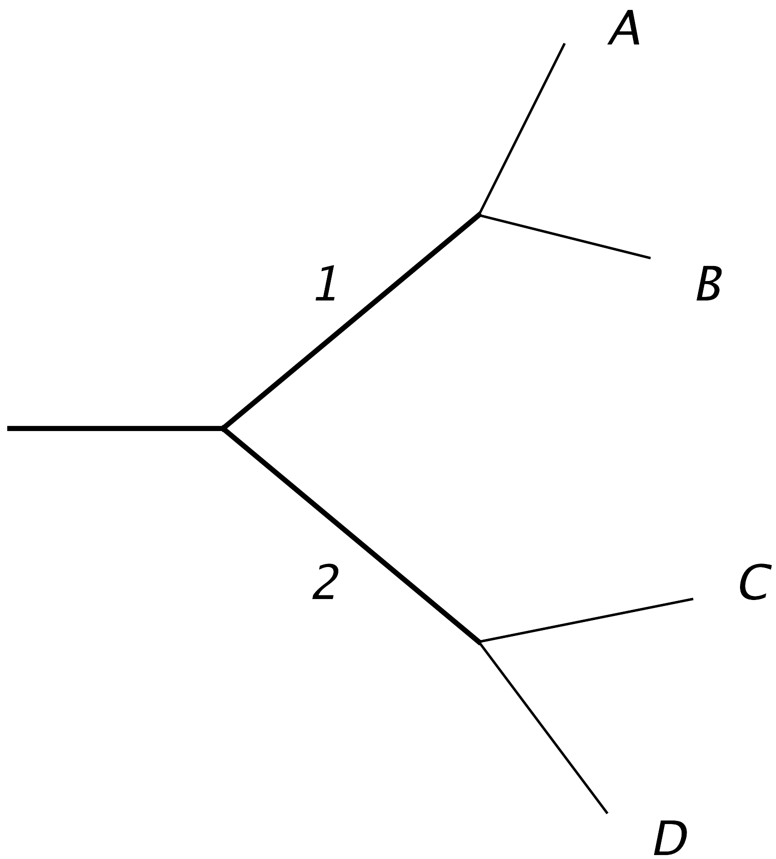




- Decluster both primary prongs in the same way.
- End up with 2, 3, or 4 prongs.
- Select 3 or 4 prong cases as top candidates.

CMS top tagger

Selecting 3 prongs from 4



CMS tagger selects three hardest objects say A,B,C.

Imposes an m_{min} condition

$$\min\left(m_{AB}, m_{BC}, m_{CA}\right) > m_{\min}$$

This method is collinear unsafe!

CMS tagger with angular cut

Original CMS tagger suffers from collinear unsafety CMS-PAS-JME-09-001

A later version introduces an angular cut in addition to the $\zeta_{\rm cut}$

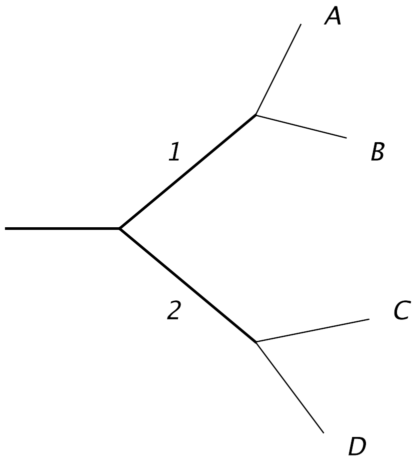
$$\Delta R_{ij} > 0.4 - A p_T$$

CMS-PAS-JME-13-007

with A = 0.0004 GeV^{-1}. Cuts off collinear divergence but vanishes at 1 TeV.

Modified taggers

IRC unsafe tagger may not be reliable so create modified tools



- CMS^{3p,mass} finally selects only the larger invariant mass de-clustering.This restores collinear safety with no ΔR

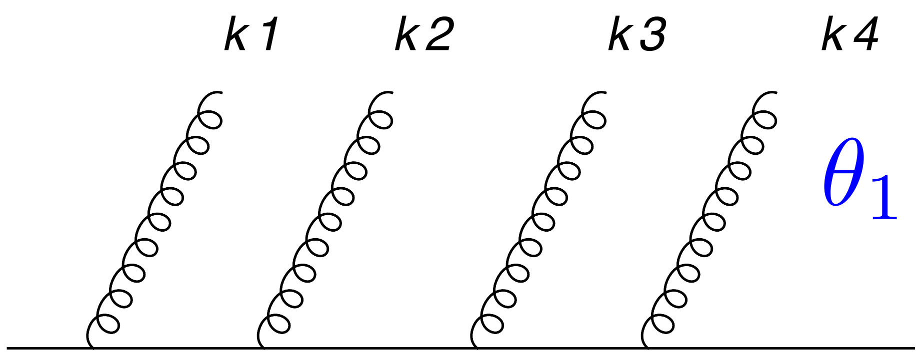
MD, Guzzi, Rawling, Soyez

Modified taggers

Another method : TopSplitter

MD, Guzzi, Rawling, Soyez 2018

Take not largest angle emission but emission that "dominates prong mass" as product of declustering.



$$\theta_1 > \theta_2 > \theta_3 > \theta_4$$

$$z_1 > \zeta_{\rm cut}$$

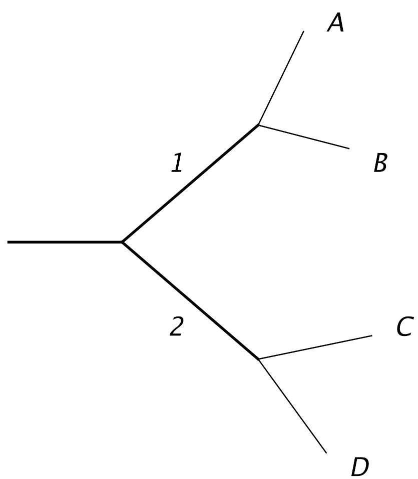
Follow hardest branch and go all the way down C/A tree to find largest $pti\,\theta_i^2$ emission.

Y_m-splitter

- Uses gen-kt (p=1/2) algorithm for de-clustering. Equivalent to mass ordering in soft limit.

MD, Guzzi, Rawling, Soyez 2018

- Not recursive but continue to use ζ_{cut}



Consider prong with larger gen-kt value as declustered if ζ_{cut} passes.

Also needs grooming

MD, Powling, Schunk, Soyez 2016
MD, Powling, Siodmok 2015

Analytics for QCD jets

We calculate jet mass distribution after application of taggers.

Define

$$\rho = \frac{m^2}{p_T^2 R^2}$$

and

$$\rho_{\min} = \frac{m_{\min}^2}{p_T^2 R^2}$$

Compute $\dfrac{d\sigma}{d\rho}$ for fixed $\rho_{\min}$ and related quantities.

MANCHESTER 1824

The University of Manchester

Analytics for QCD jets

With m ~ m_t and m_{min} ~ m_w at high p_t :

$$\rho, \rho_{\min} \ll 1$$

$$L_\rho = \ln \frac{1}{\rho} \gg 1$$

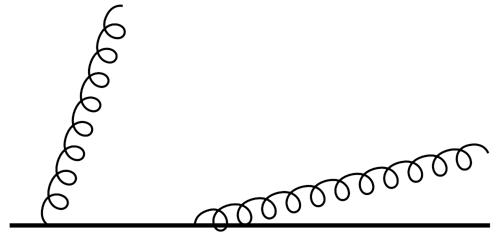
→ **Resum large logs**

Also we have no strong ordering in these masses.

$$L_\rho \sim L_{\rho_{\min}} \gg 1$$

and $\zeta_{\rm cut} \sim 0.05$ $\quad L_{\rho,\rho_{\min}} \gg L_\zeta$

Leading order calculation



Two real emissions to pass the tagger so starts with $\mathcal{O}\left(\alpha_s^2\right)$. For simplicity take limit $\rho \gg \rho_{\min}$

$$\frac{d\sigma}{d\rho} \sim \frac{\alpha_s^2}{\rho} \ln^2 \frac{1}{\zeta_{\rm cut}} \ln \frac{\rho}{\rho_{\min}}$$

soft **strong-ordered**

Compare to QCD jet $\frac{\alpha_s^2}{\rho} \ln^3 \frac{1}{\rho}$

Reduced background after tagging.

MANCHESTER 1824

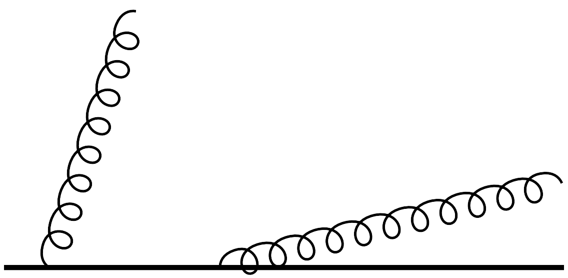
The University of Manchester

Triple collinear limit for QCD jet

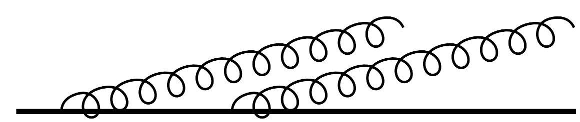
But $\ln \frac{\rho}{\rho_{\min}}, \ln \frac{1}{\zeta_{\rm cut}}$ are not too large.

Need to lift strong ordering and soft approx.

$$\frac{\alpha_s^2}{\rho} \ln^2 \frac{1}{\zeta_{\rm cut}} \ln \frac{\rho}{\rho_{\min}} \quad \Rightarrow \quad \frac{\alpha_s^2}{\rho} \times \mathcal{O}(1)$$

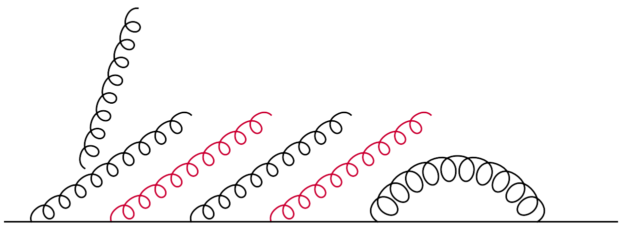


Standard LO DGLAP or PS evolution



Triple collinear splitting functions

All orders



- Beyond leading order : constraints on real emissions arise from ρ and $\rho_{\rm min}$ conditions.

  ⇒ **Sudakov form factors**

- Our resummation accuracy is modified LL. Resums all double logs $\frac{1}{\rho}\alpha_s^n L^{2n-1}$

- Counts $\ln\frac{1}{\rho}, \ln\frac{1}{\rho_{\rm min}}, \ln\frac{1}{\zeta_{\rm cut}}, \ln\frac{\rho}{\rho_{\rm min}}$ all on same footing

- Also includes NLL effects from running coupling and hard collinear emissions.

The University of Manchester

MANCHESTER 1824

Results

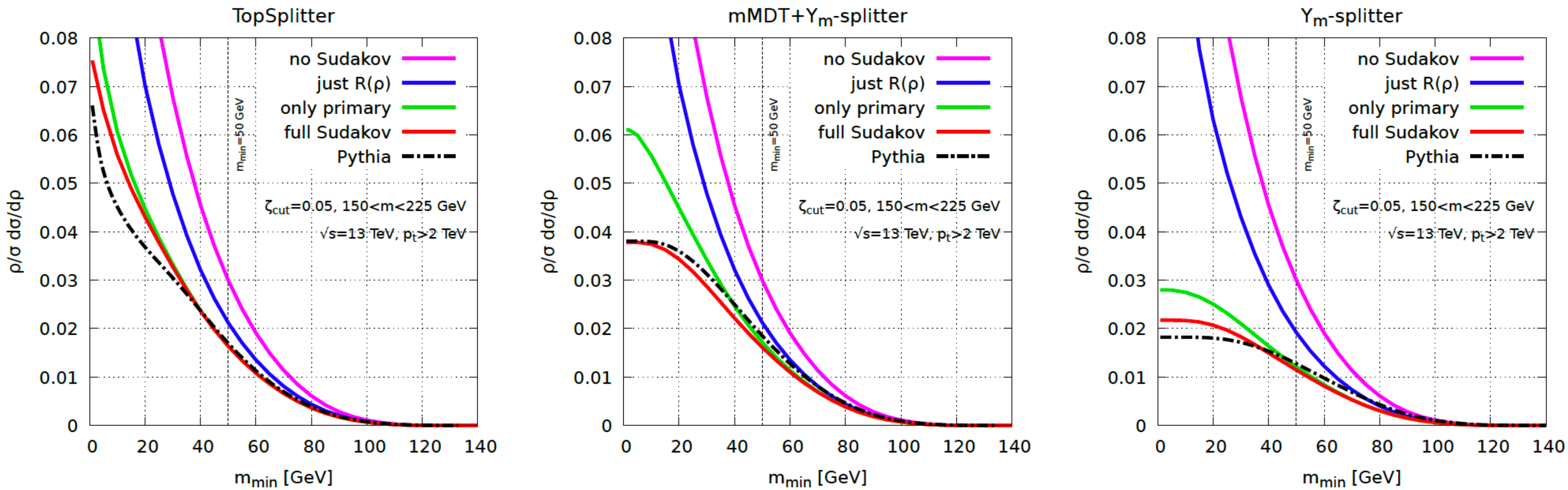
General form:

$$\rho \frac{d\sigma}{d\rho} = \int d\Phi_3 \frac{\hat{P}}{s_{123}^2} \frac{\alpha_s\left(k_{t1}\right)}{2\pi} \frac{\alpha_s\left(k_{t2}\right)}{2\pi} \Theta^{\text{jet}} \Theta^{\text{tagger}} \delta\left(\rho - \frac{s123}{p_t^2 R^2}\right) \times e^{-R}$$

- Prefactor computed using triple-collinear splitting functions and phase space
- Convoluted with a Sudakov form factor accounting for all leading log terms
- Running coupling and hard-collinear effects included
- Matching of Sudakov to triple-collinear phase space.
- Aims to be as accurate as triple-collinear result at LO and reproduce all leading-log terms beyond.

Campbell and Glover 1997, Catani and Grazzini 1998

Results and comparisons to PS

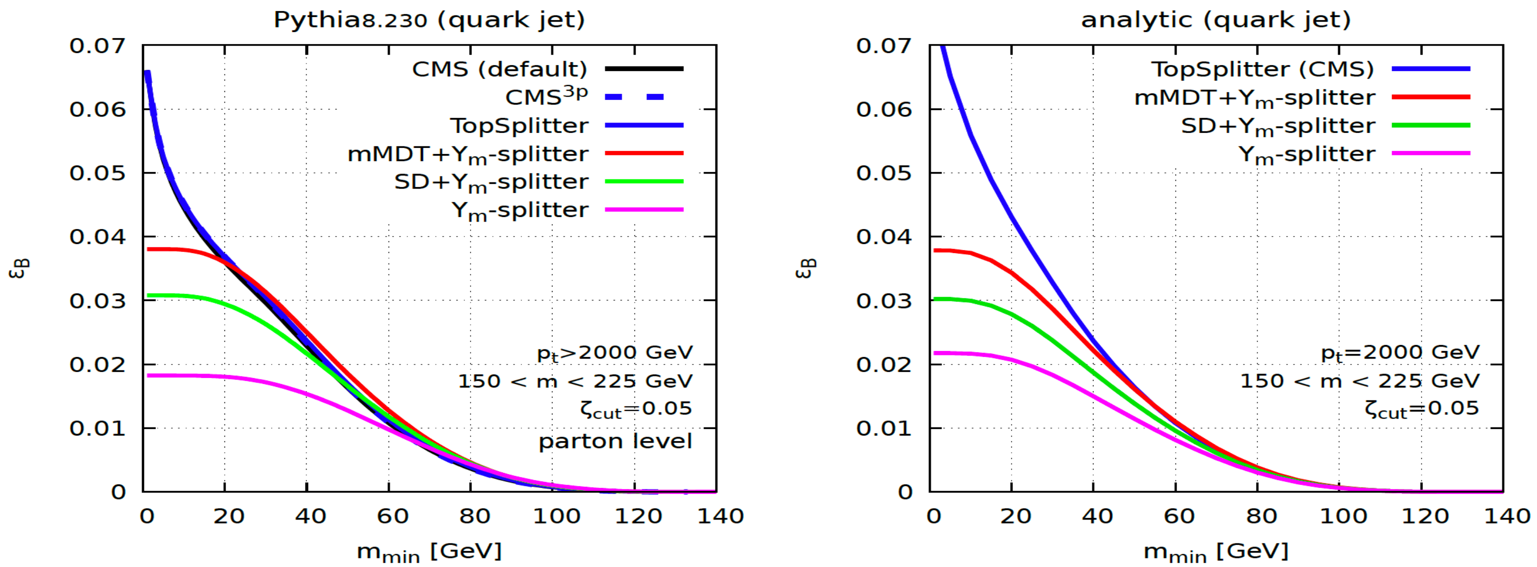


MD, Guzzi, Rawling, Soyez. Preliminary

- Plots reflect that resummation of $\ln \frac{\rho}{\rho_{\min}}$ terms does matter
- Inclusion of secondary emissions important at small m_{min}
- Overall a good agreement with PS.

The University of Manchester

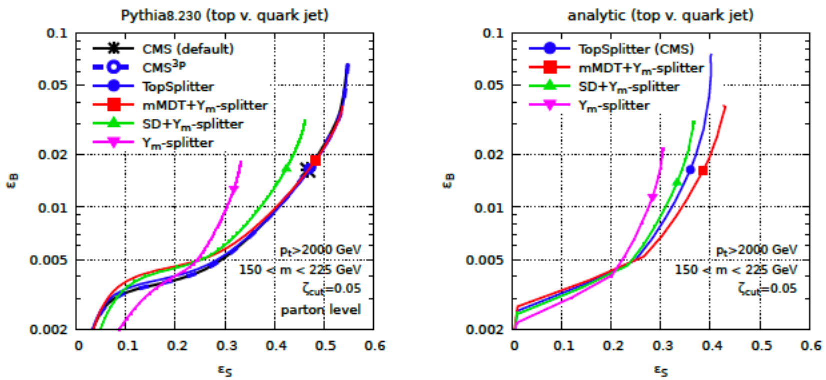
Tagger comparisons for QCD jets



MD, Guzzi, Rawling and Soyez, 2018

- MC and analytics agree on comparative performance
- Y_m splitter best at suppressing QCD jets
- CMS and variants are basically identical for performance
- Groomed Y_m splitter comparable with CMS. Differences largely due to secondary emissions.

Signal jets



MD, Guzzi, Rawling, Soyez. 2018

- For W/Z/H decays impact on background key to final performance. Taggers like Y-splitter are high-performance owing to large Sudakov

- For coloured top this is not the case due to signal Sudakov suppression. Also analysed signal jets with a basic Sudakov. Groomed Y-splitter comparable to CMS and variants.

MANCHESTER 1824

The University of Manchester

Conclusions

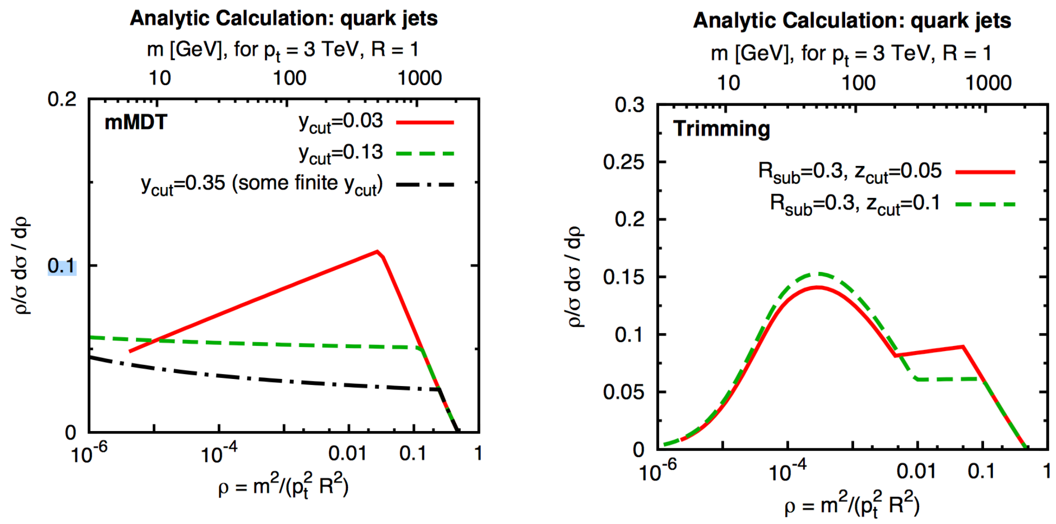
- A first analytic study of aspects of top taggers carried out.
- Shows analytic control over basic features
- Large Sudakov effects not necessarily desirable and hurt signal efficiency.
- CMS tagger become potentially unsafe at high $p_{t..}$ Potentially harmful for precision studies. Easy to design safe variants with no change in performance.
- Plan to nvestigate combinations with jet shape variables like τ_{32} as next step.

BACK UP MATERIAL

MANCHESTER 1824

The University of Manchester

Analytical insight

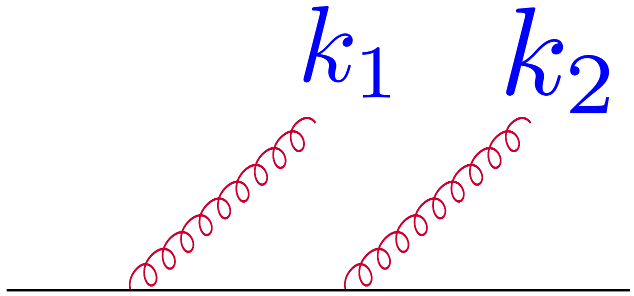


- Traditional approach : Construct taggers on simple intuitive ideas. Leave details to MC studies. Lots of freedom to create many new tools.

- Analytical approach : Worry about details. Get main physics principles. Then construct optimal tools.

MD, Fregoso, Marzani and Salam 2013

Results



$$\otimes \quad \exp[-R]$$

$$\rho_2 = z_2\theta_2^2 < z_1\theta_1^2$$

- The key differences between taggers come from the Sudakov.
- Y_m-splitter has a plain jet mass double log Sudakov in ρ_2
- TopSplitter and safe variants of CMS have an mMDT style single-log Sudakov
- mMDT/SoftDrop grooming + Y_m-splitter inherits grooming Sudakov structure. MD, Fregoso, Marzani and Salam 2013. Larkoski, Marzani, Thaler and Soyez 2014.

Top tagging methods

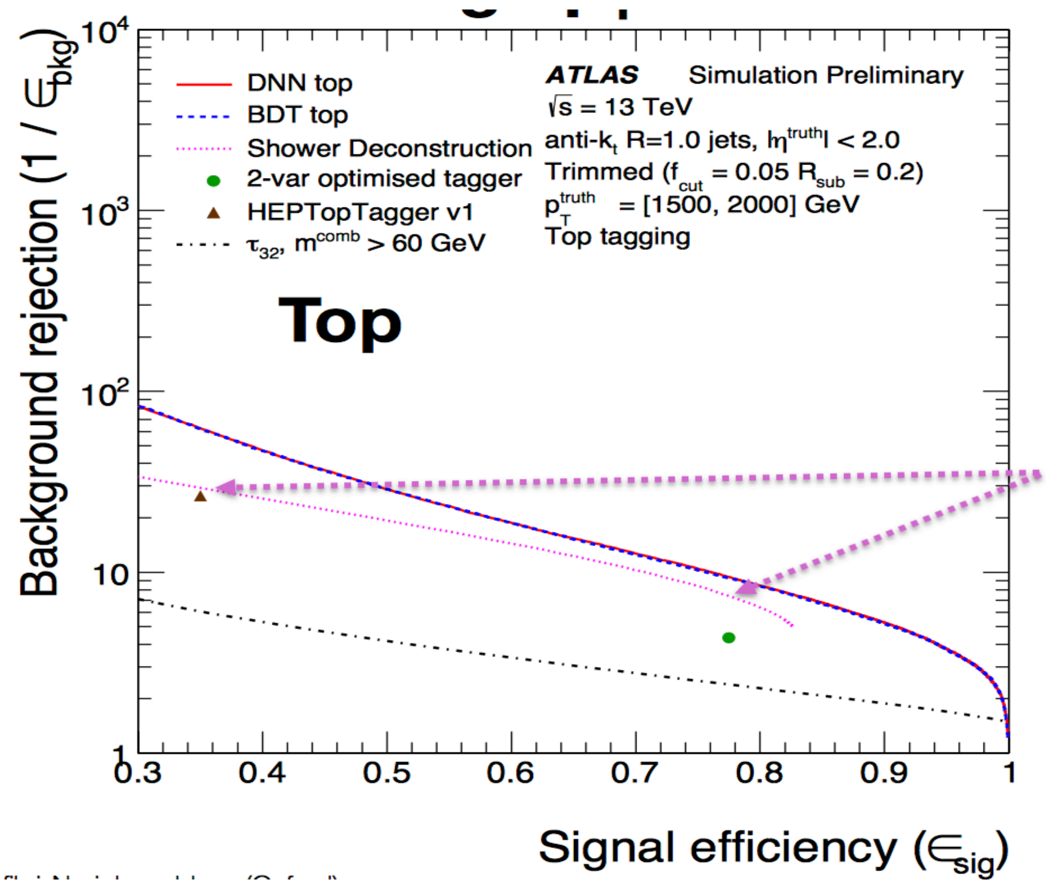


Figure from talk by N.Norjoharuddin on behalf of ATLAS, Boost 2017