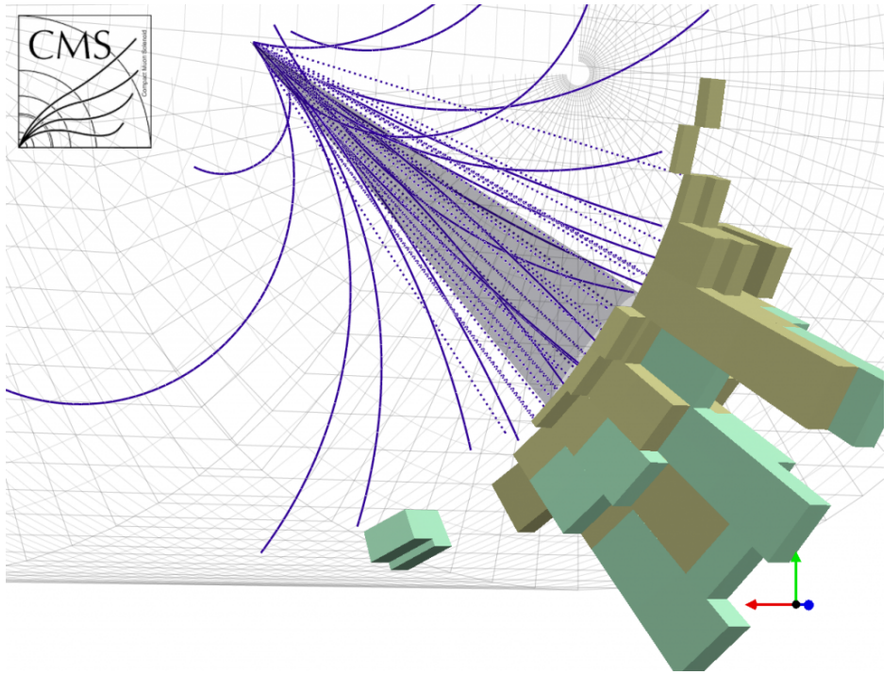
Substructure and tagging in CMS

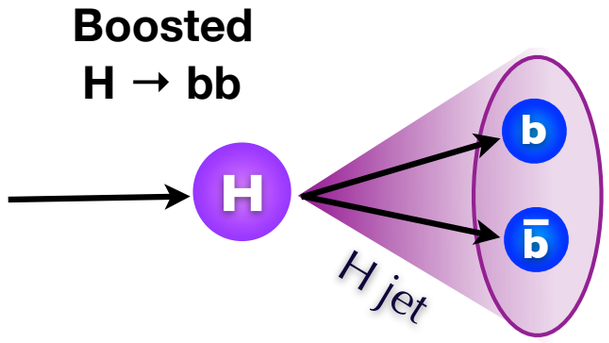
James Dolen
Purdue University Northwest

Large Hadron Collider Physics (LHCP) conference - Bologna Italy - June 6th, 2018

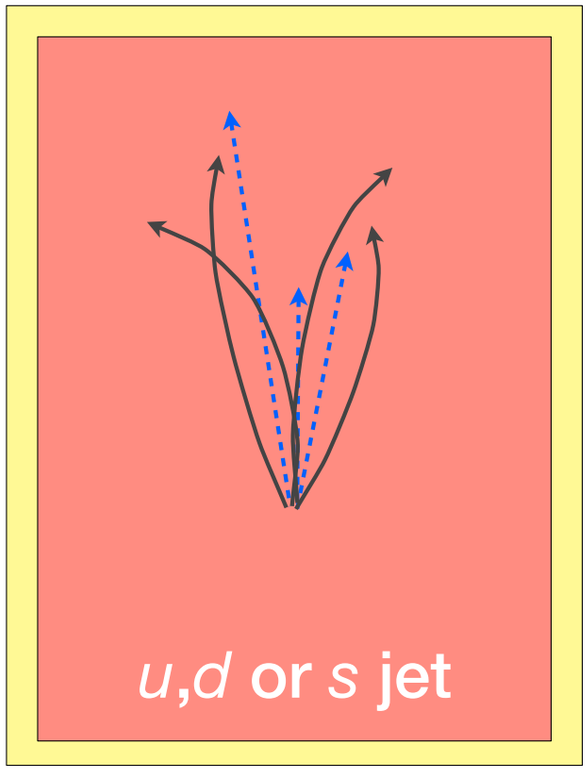
Jet Tagging at CMS

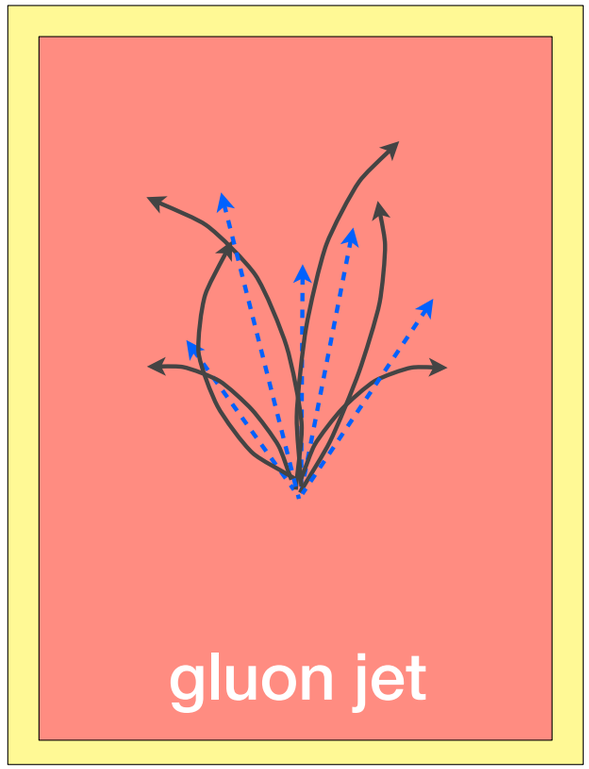
- Jets are ubiquitous at hadron colliders.
  - Collimated spray of hadrons resulting from an initial state quark or gluon
- LHC jet tagging revolution
  - Enabled by new reconstruction techniques (ex. particle flow, sequential recombination jet algorithms)
  - Discriminate quark and gluon jets, noise and pileup jets
  - Boosted heavy object jet tagging (ex. top, W, Z, Higgs)
    - all decay products of reconstructed within one jet
  - Advanced b-tagging methods
    - Subjet b-tagging, double b-tagging
- Critical tools for discovery!
- Critical tools for measurements!

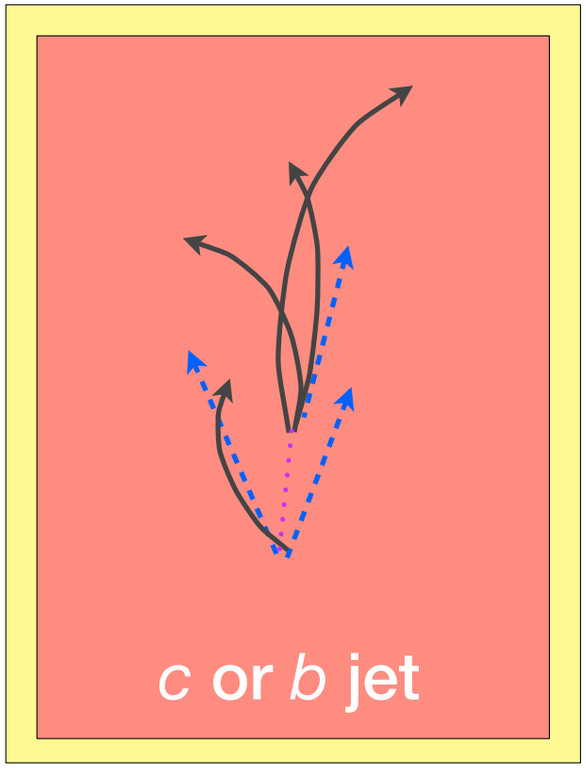


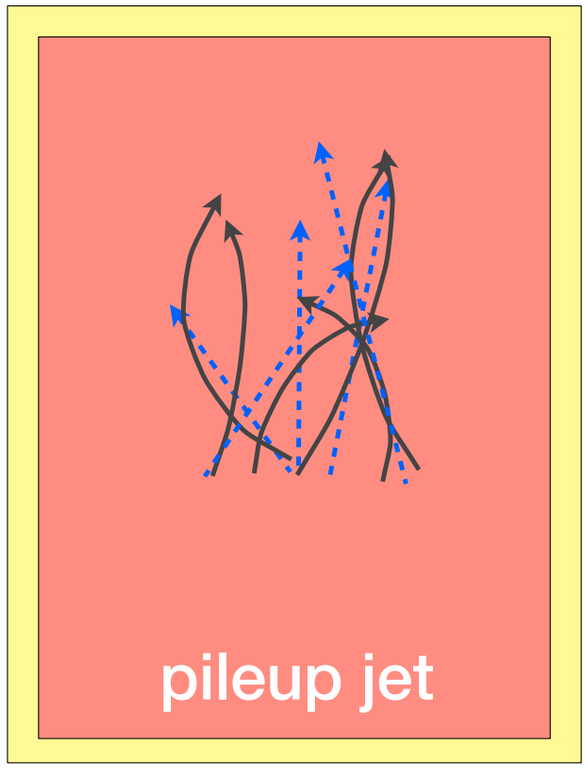


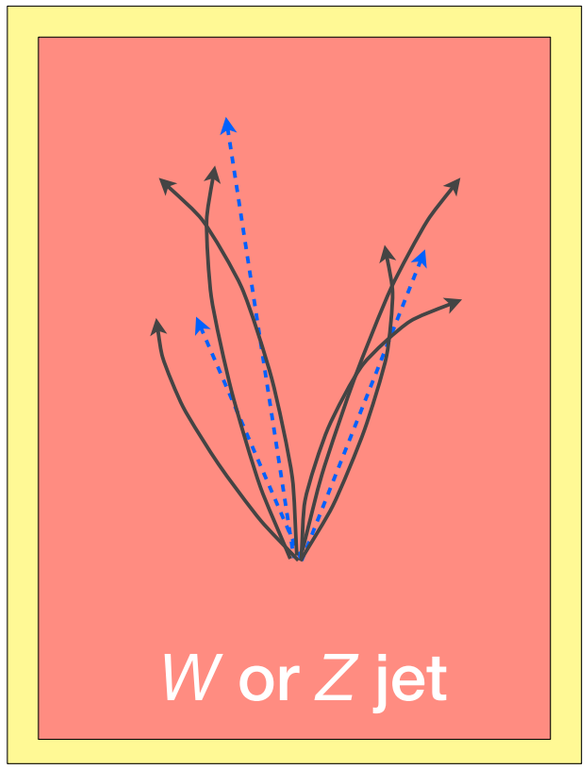
Jet tagging

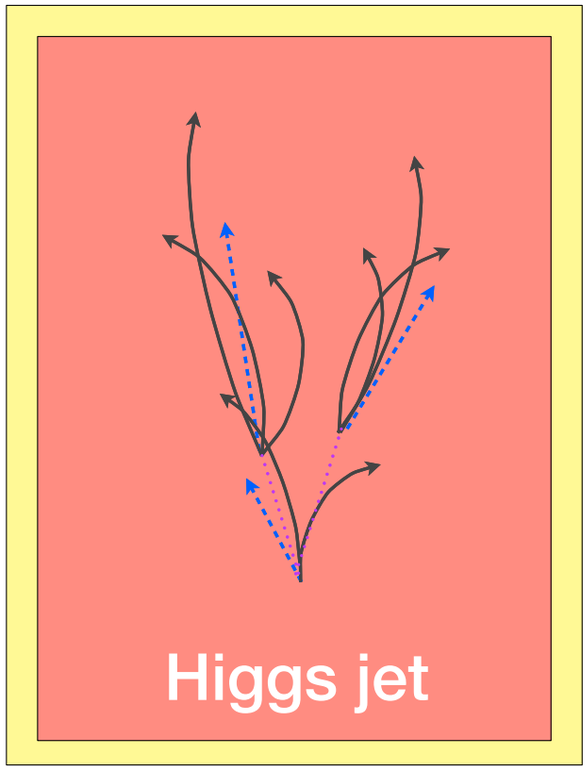












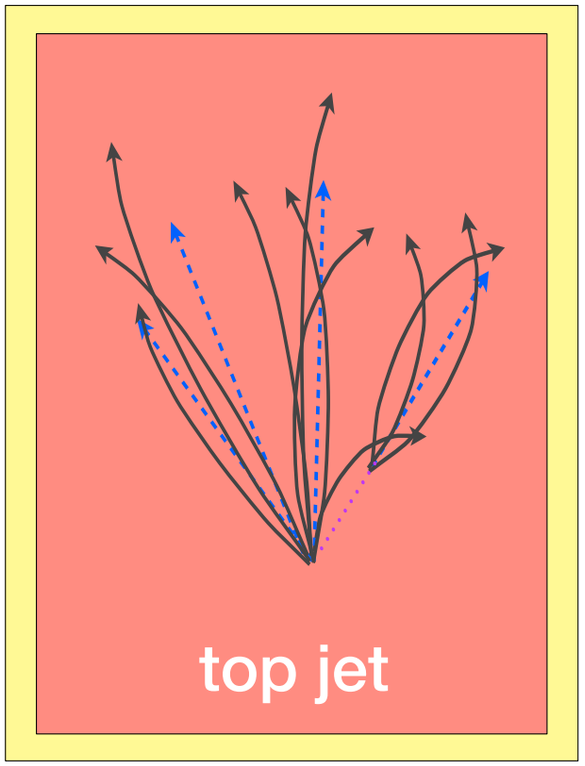
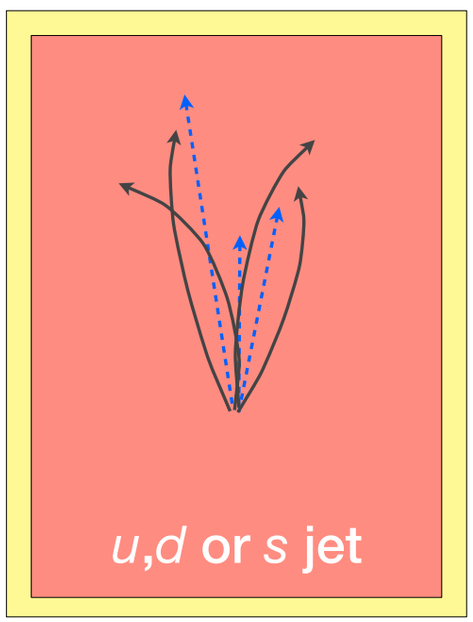


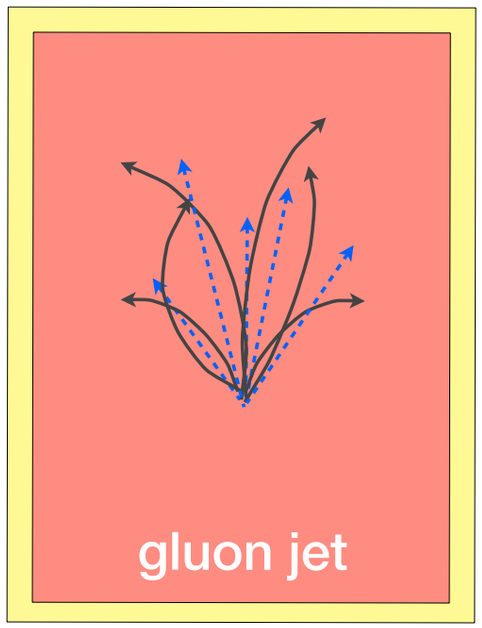
Diagram by
Nhan Tran

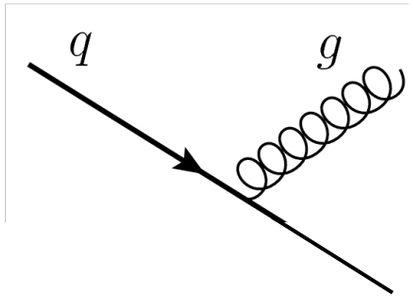
?

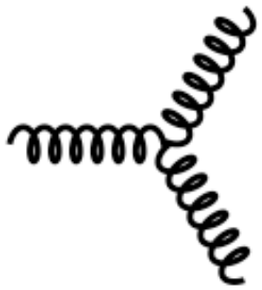
Identifying quark and gluon jets

- Gluon more likely to radiate a gluon (C_A =3 vs C_F = 4/3)
- Gluon fragmentation function softer
- Gluon jets
  - Wider than quark jets (η-Φ plane)
  - Larger multiplicities
  - Fewer hard particles
- Quark jets
  - Narrow
  - Smaller multiplicities
  - Asymmetrical energy shared between constituents



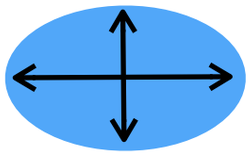






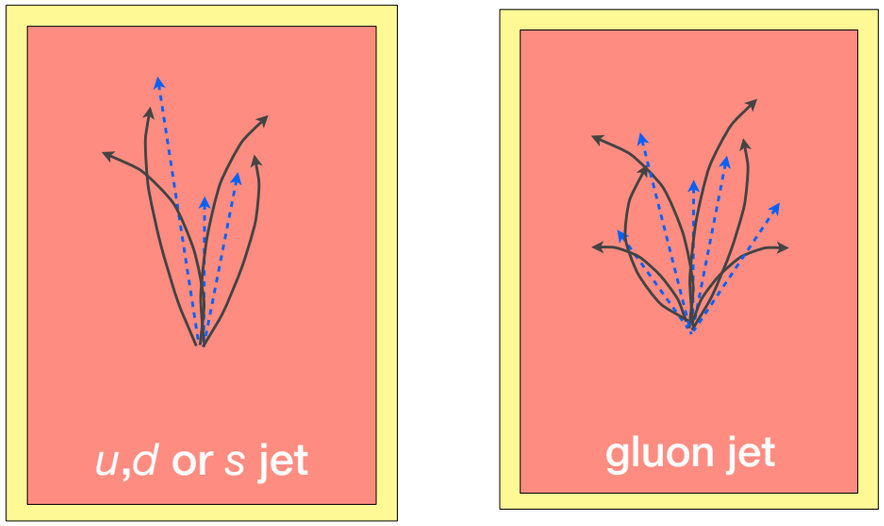
Quark/gluon discrimination at CMS

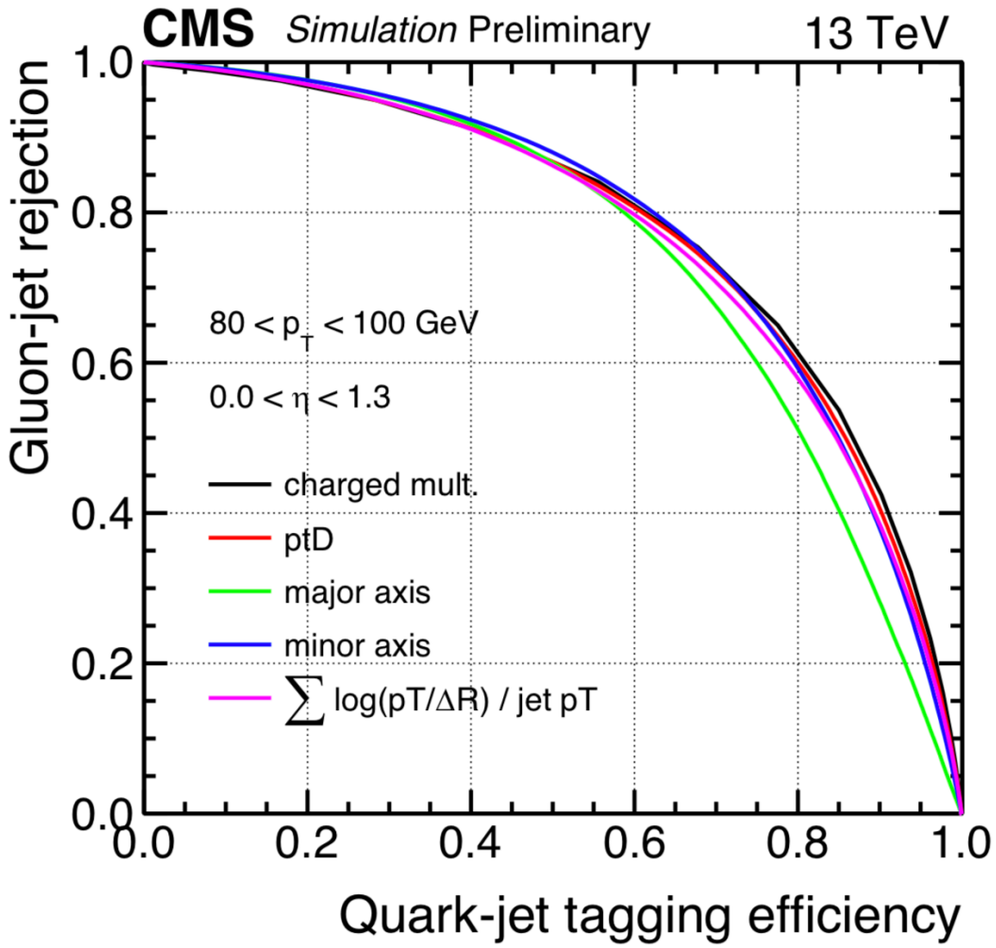
- Three tools:
  - Likelihood discriminator (3 variable)
  - BDT discriminator (5 variable)
  - Deep neural net (jet constituents)
- Quark/gluon discriminator variables:
  - Jet particle multiplicity
    - Total or charged
  - Jet shape
    - major axis width
    - minor axis width
    - ΔR weighted p_T $\sum_i \frac{\log(\frac{p_{T,i}}{\Delta R})}{\text{jet } p_T}$
  - Energy sharing (p_TD)



$$p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}}$$

→ 1 if all momentum carried by one particle
→ 0 if jet has infinite number of particles

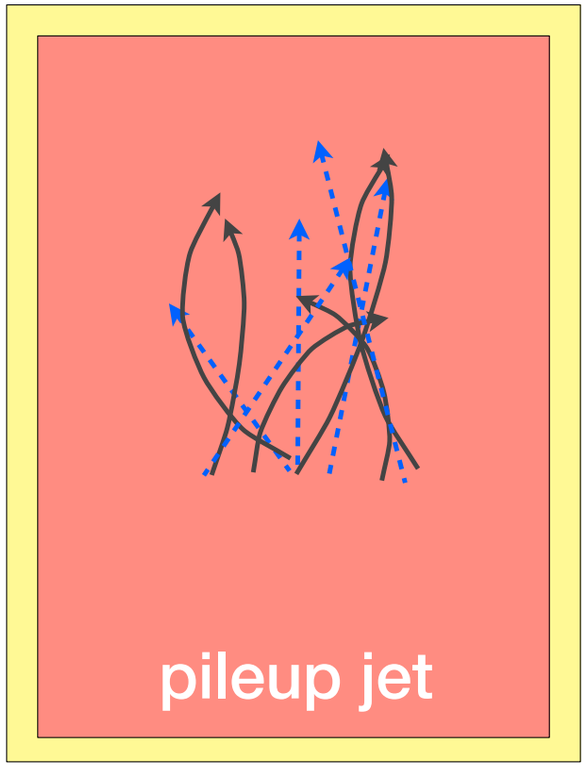


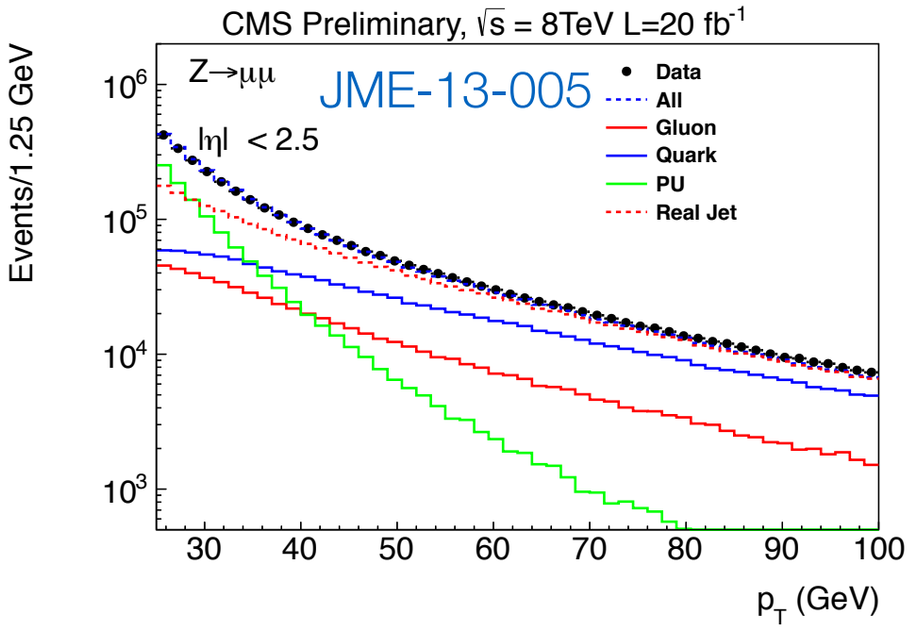


CMS DP Note 2017/027

Pileup jet tagging

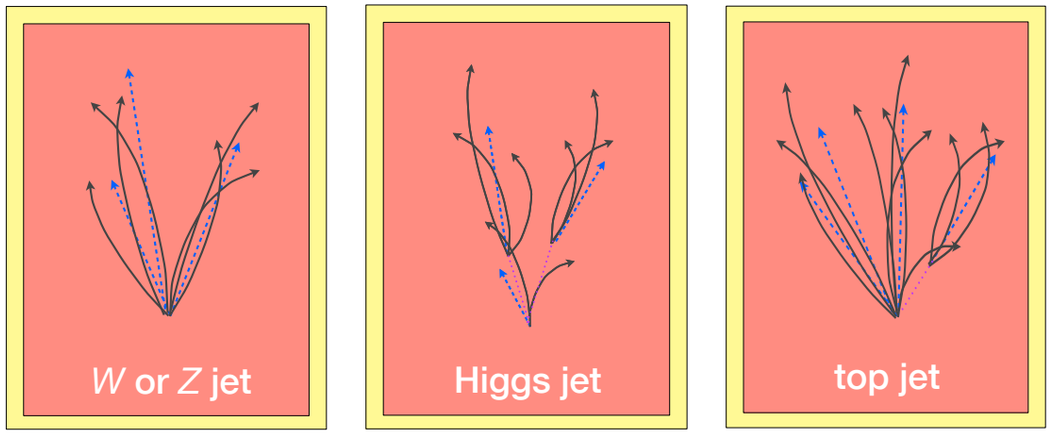
- Pileup jets are formed from overlapping low p_T pileup particles
  - Each pileup vertex contributes ~0.7 GeV of energy per unit area (η,Φ) of the detector
- Pileup jet tagging - tool used to identify and reject jets originating from pileup
  - Utilize vertex, track, and jet shape information
  - Charged particles inside pileup jets are not associated with the primary vertex
  - Pileup jets are more defuse (overlapping soft particles from multiple vertices)



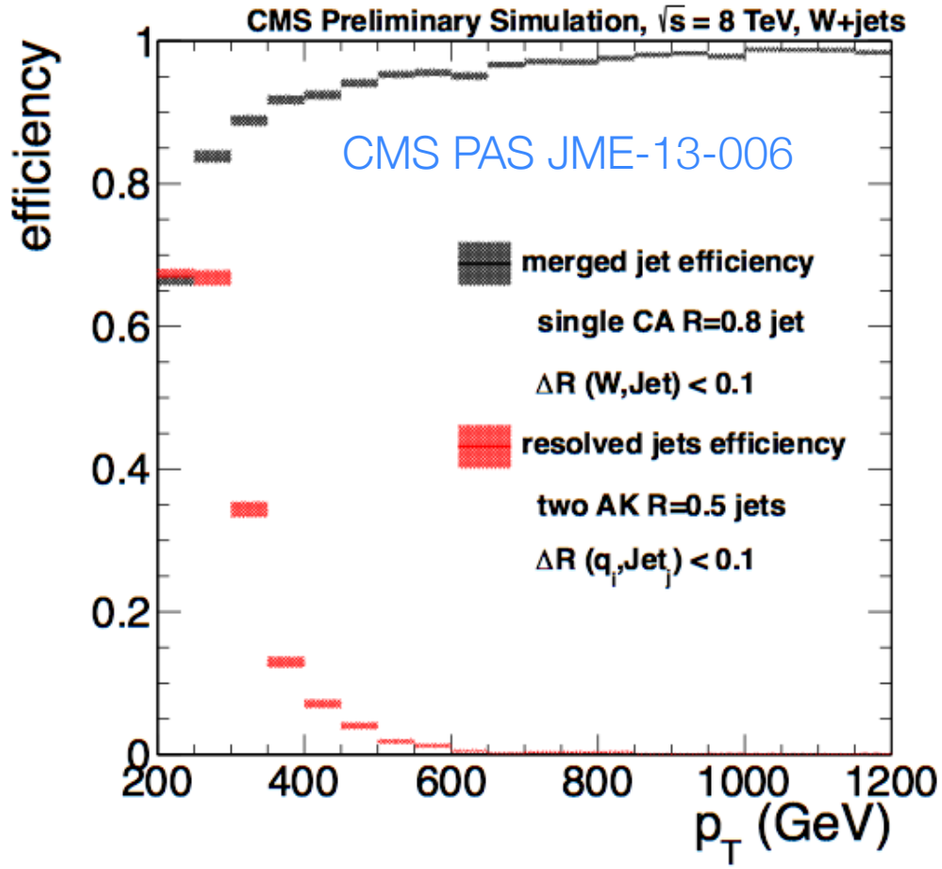


Boosted heavy object tagging

- Example: 2 body decay
  - decay product angular separation $\Delta R \sim 2M/p_T$

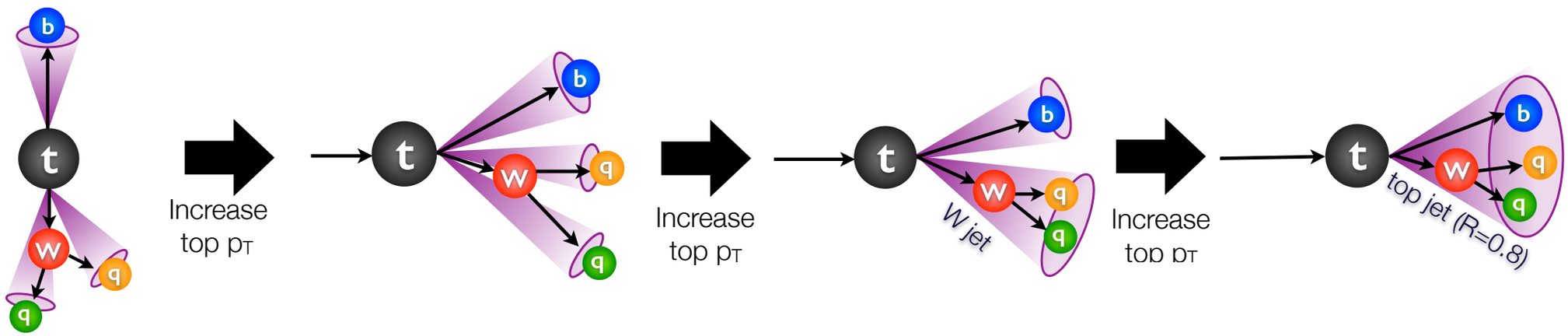


- Boosted heavy object jets are identifiable based on their mass and via jet substructure
  - The jet, it's constituents, and it's clustering history, contain useful information which can be used to identify these objects
    - examples: jet mass, "subjettiness", W mass within a top jet

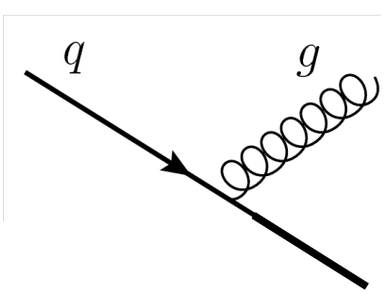


Jet mass as a tagging variable

- Merged top/W/Z/H jet → LO jet mass is the heavy object mass



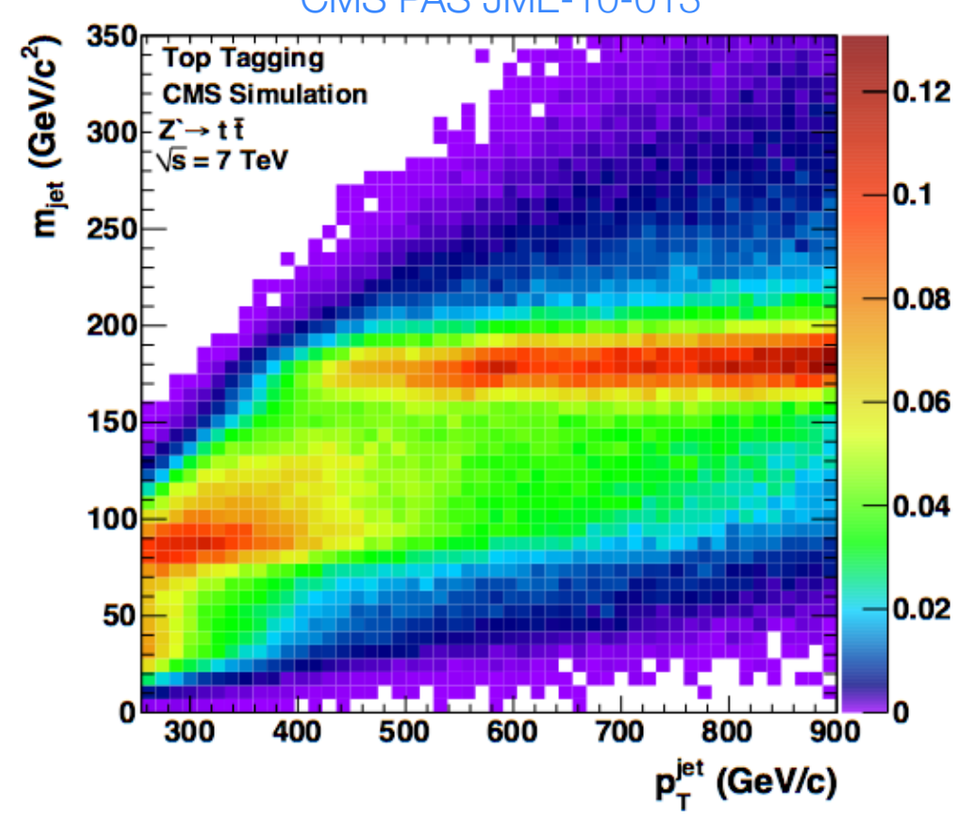
- For background jets (quark/gluon) the LO jet mass is ~ 0, but perturbative effects lead to measured mass



$$\langle M^2 \rangle \simeq C \cdot \frac{\alpha_s}{\pi} p_t^2 R^2$$

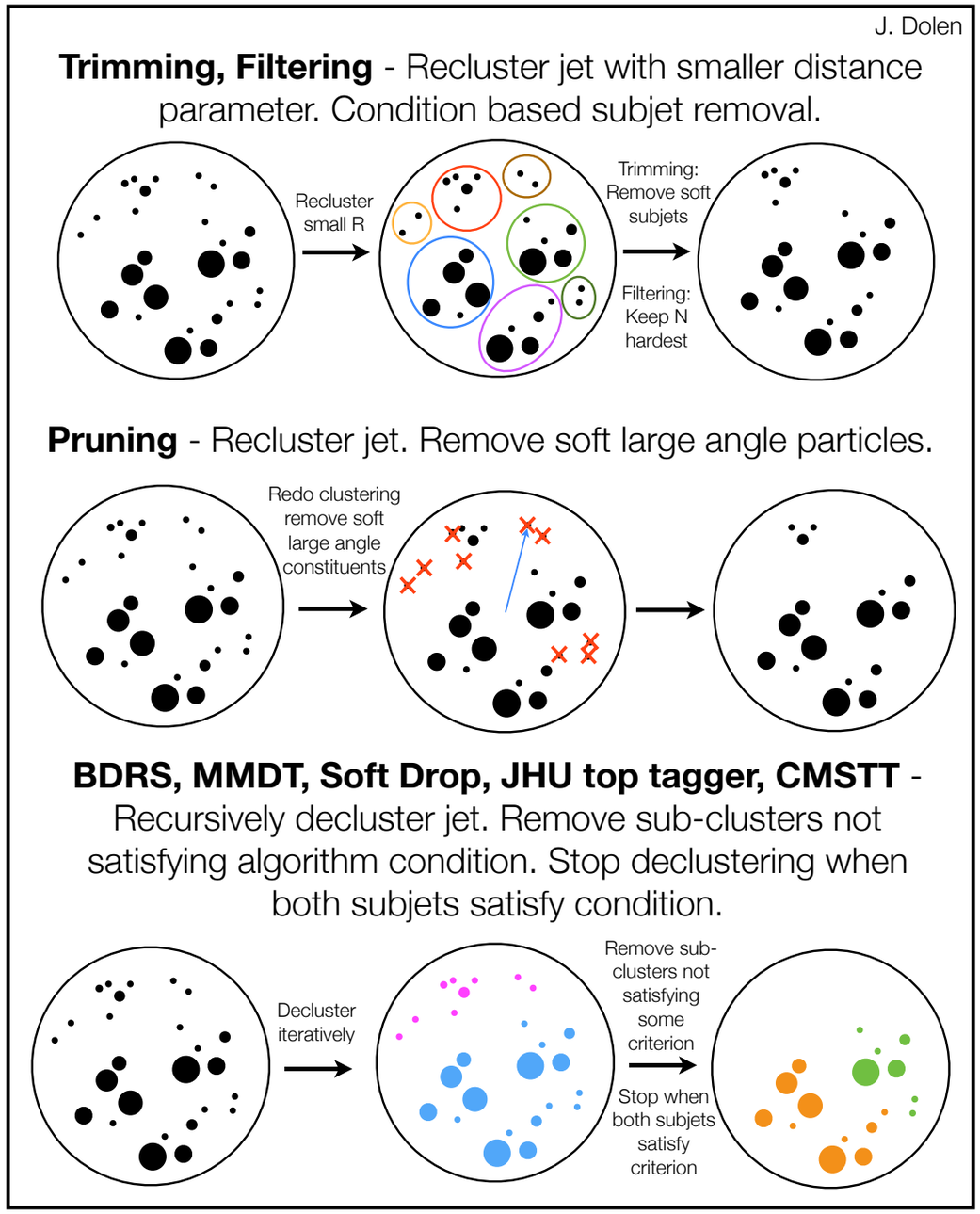
Ellis et al. Prog. Part. Nucl. Phys. 60 (2008)

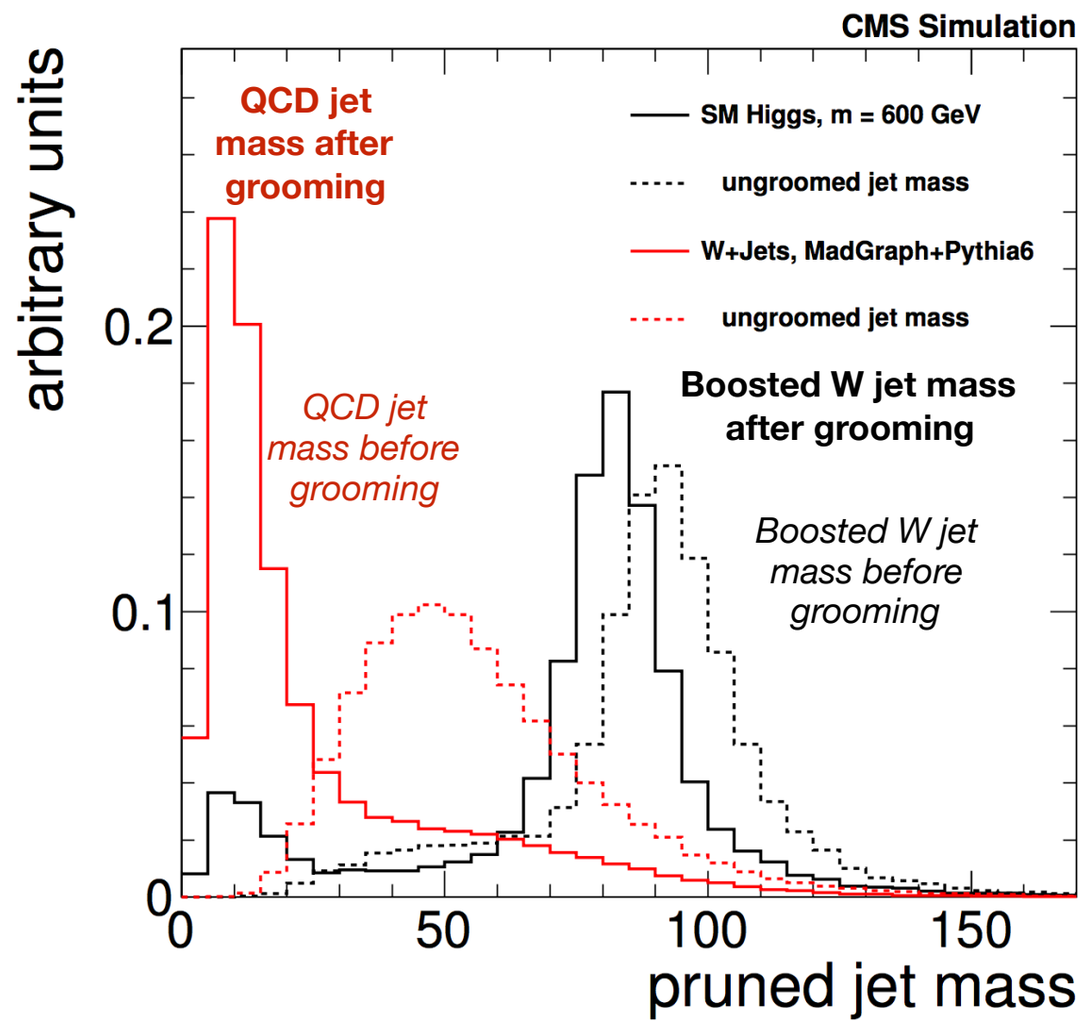
CMS PAS JME-10-013



Jet grooming

Algorithmic jet substructure techniques designed to remove isolated soft radiation in jets (contamination from ISR, UE, pileup)





CMS HIG-13-008
H → WW → lνqq

Reduces jet mass dependence on pileup

Reduces measured QCD jet mass (improves discrimination)

Improved jet mass resolution for boosted heavy object

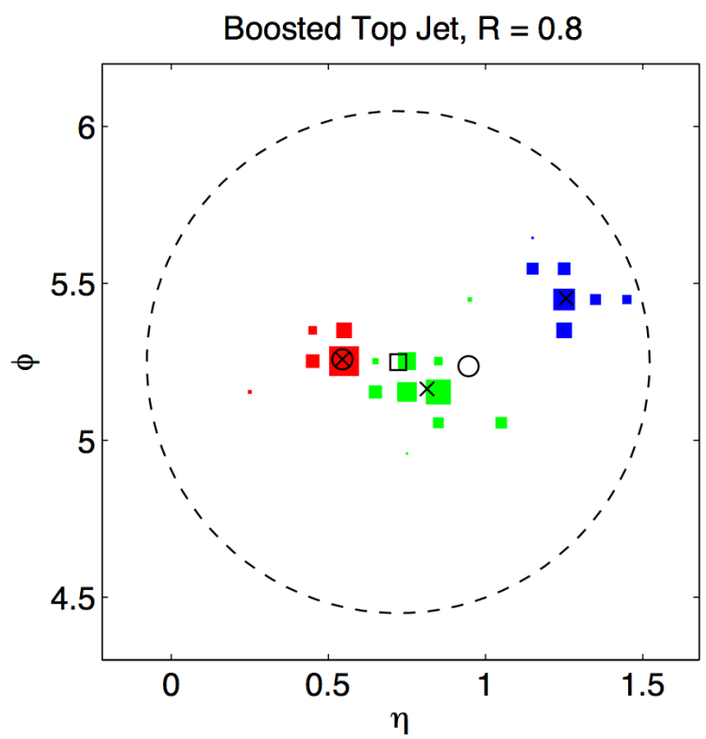
Jet shape tagging variables

Particle energy pattern within a jet used to identify "multi-prong" jets

N-subjettiness

$$\tau_N = \frac{\sum_{i=1}^{n_{\text{constituents}}} p_{\mathrm{T},i} \min\{\Delta R_{1,i}, \Delta R_{2,i}, ..., \Delta R_{N,i}\}}{\sum_{i=1}^{n_{\text{constituents}}} p_{\mathrm{T},i} R}$$

Determines how consistent a jet is with having N or fewer subjets



$\tau_N \rightarrow 0 \Longrightarrow$ energy spread is close to the subjet axes

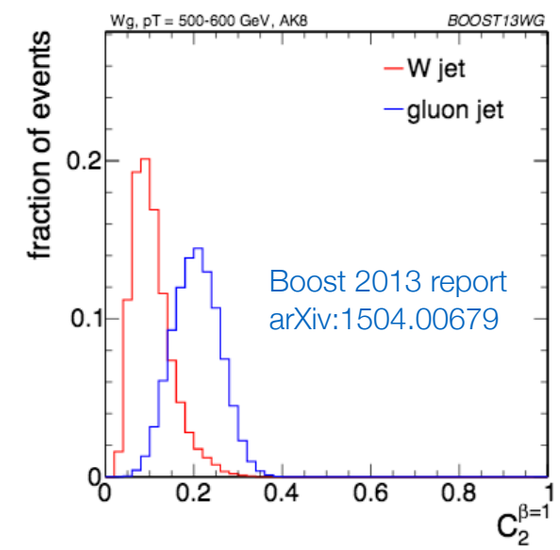
Better discrimination by using ratios (ex. τ_3/τ_2)

J. Thaler, K. Van Tilburg, JHEP 2011:15

Energy Correlation Functions (ECF)

$$\mathrm{ECF}(N,\beta) = \sum_{i_1<i_2<...<i_N \in J} \left(\prod_{a=1}^{N} p_{T i_a}\right) \left(\prod_{b=1}^{N-1} \prod_{c=b+1}^{N} R_{i_b i_c}\right)^{\beta}$$

Jet constituent based observables sensitive to N subjet substructure



$$C_N^{(\beta)} = \frac{\mathrm{ECF}(N+1,\beta)\,\mathrm{ECF}(N-1,\beta)}{\mathrm{ECF}(N,\beta)^2} \qquad D_2(\beta) = \frac{\mathrm{ECF}(3,\beta)\mathrm{ECF}(1,\beta)^3}{\mathrm{ECF}(2,\beta)^3}$$

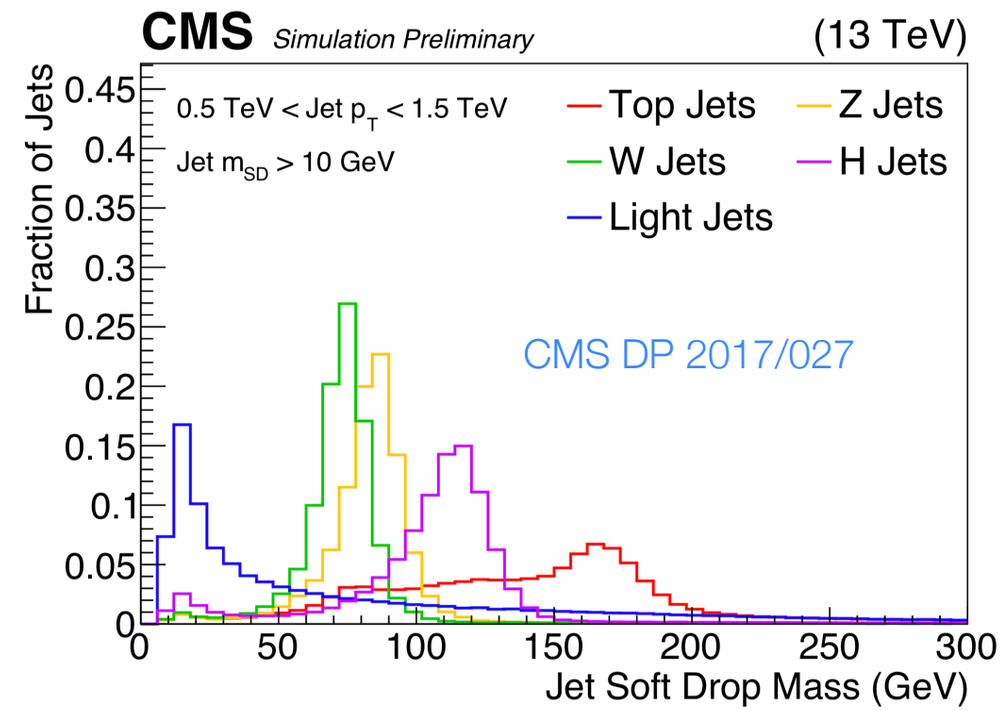
ECF ratios (C2,D2, etc.) for boosted object discrimination (Small C2 $\Longrightarrow$ 2 subjets)

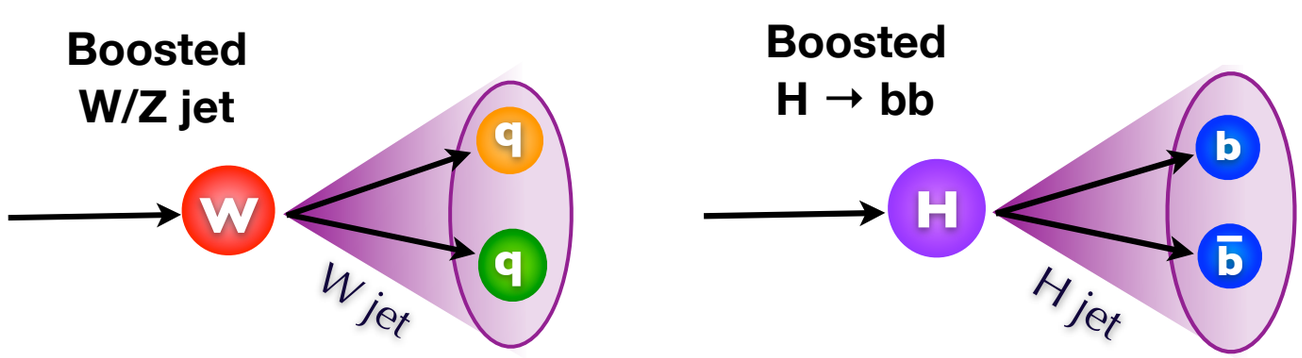
β parameter allows access to different angular scales

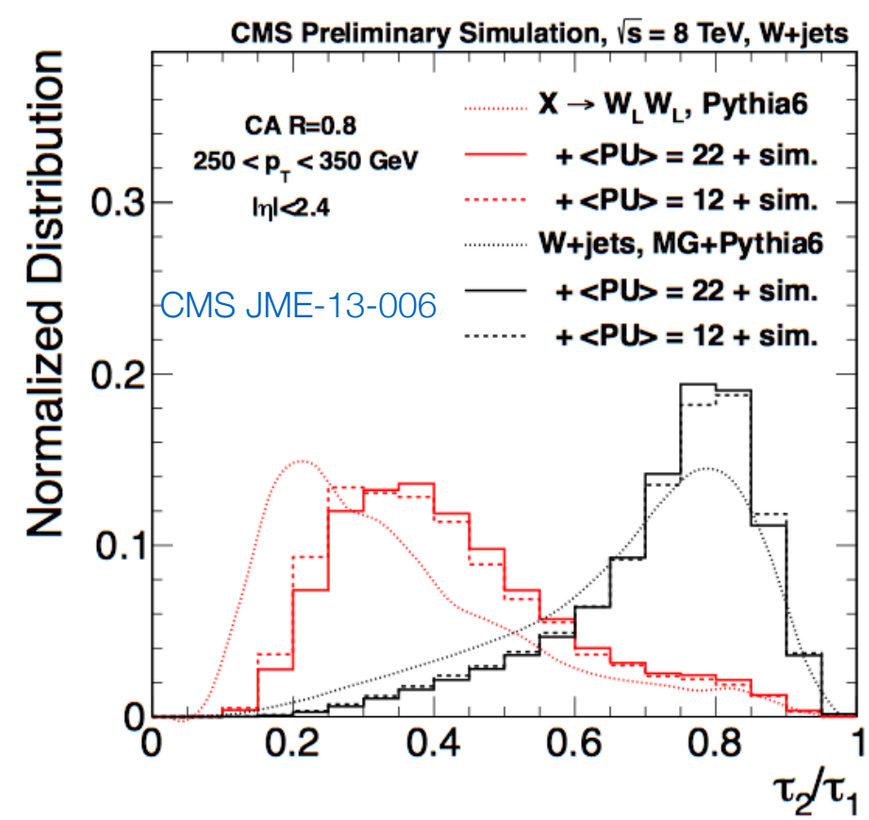
A. Banfi, G. P. Salam, and G. Zanderighi, JHEP 0503 (2005) 073
M. Jankowiak and A. J. Larkoski JHEP 1106 (2011) 057
A. Larkoski, G. Salam, J. Thaler JHEP 2013:108

2-prong tagging "V-tagging"

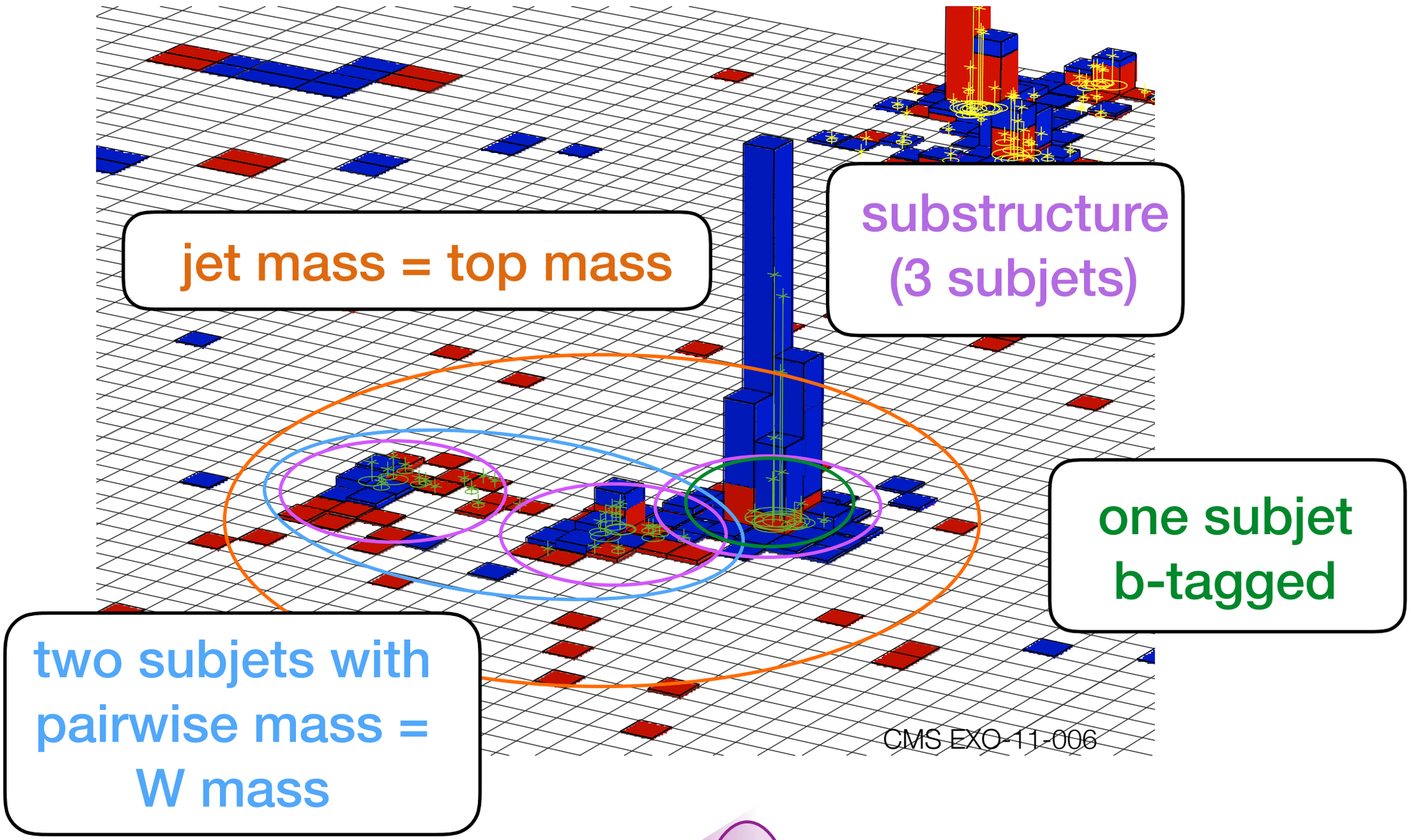
- Boosted W, Z, H tagging
- General technique: select V jets based on groomed jet mass and jet shape variable
- CMS default:
  - Anti-kt R=0.8 PUPPI jet
  - MMDT/Soft drop jet mass
  - N-subjettiness ratio τ_2/τ_1

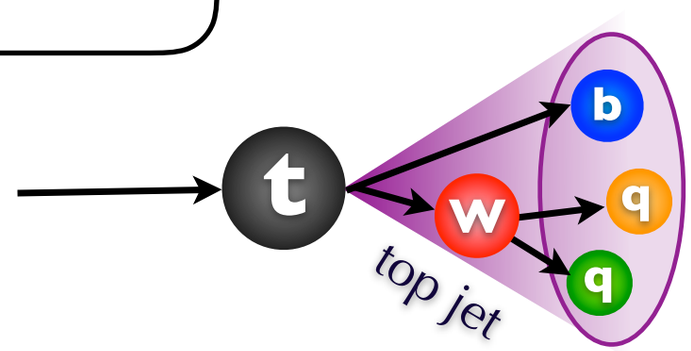




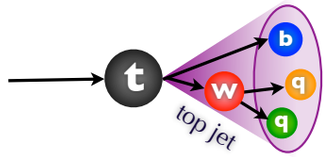


Top jet tagging



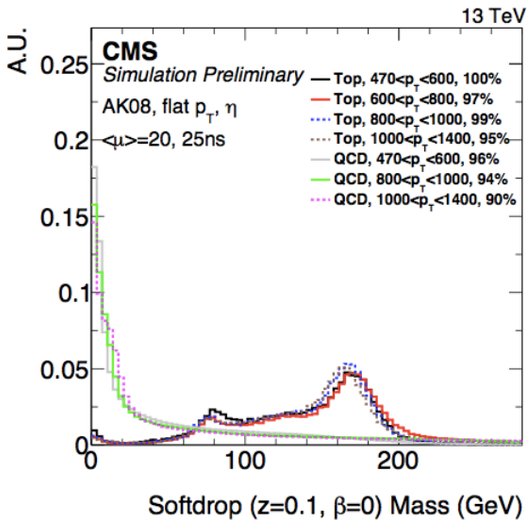


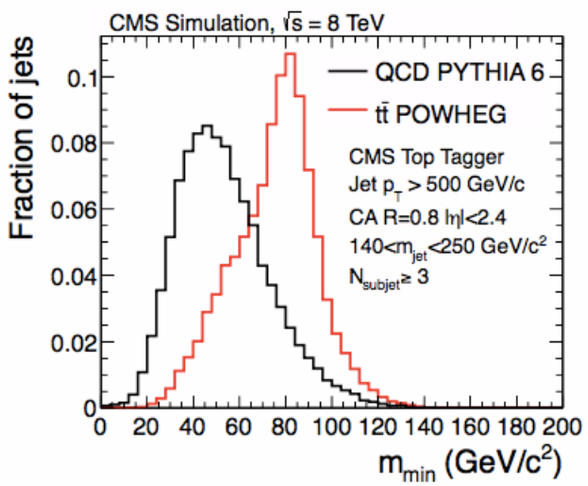
Top tagging algorithms

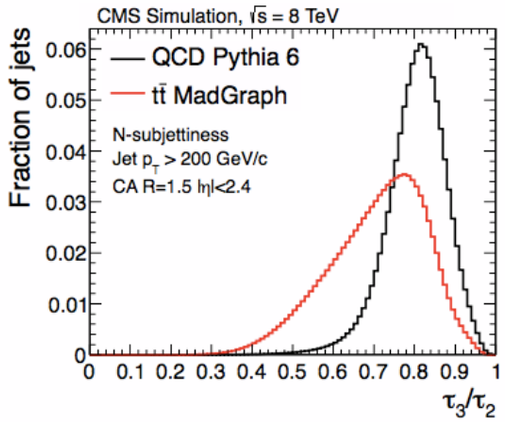


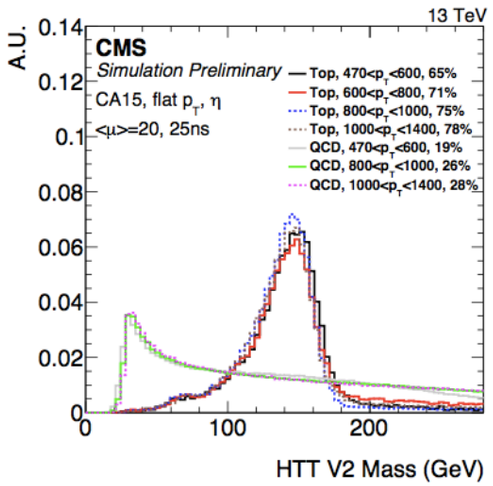
- Taggers utilizing general algorithms
  - Groomed mass + N-subjettiness (τ_3/τ_2)
  - Shower deconstruction [1]
  - Splitting scale
  - Subjet b-tagging
- Dedicated algorithms
  - CMS Top Tagger (JHU Top Tagger) [2]
    - Decluster jet twice to find 1-4 subjets
    - Select tops based on top mass, W mass, Nsubjets
  - HEP Top Tagger v1 [3]
    - Very large jets (R=1.2-1.5)
    - Multistep decluster + filter procedure
    - Select tops based on top mass, W mass
  - HEP Top Tagger v2 [4]
    - Multiple algorithm improvements + multi R approach
    - Select tops based on top mass, W mass, optimal jet size

CMS JME-15-002



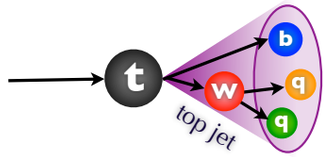




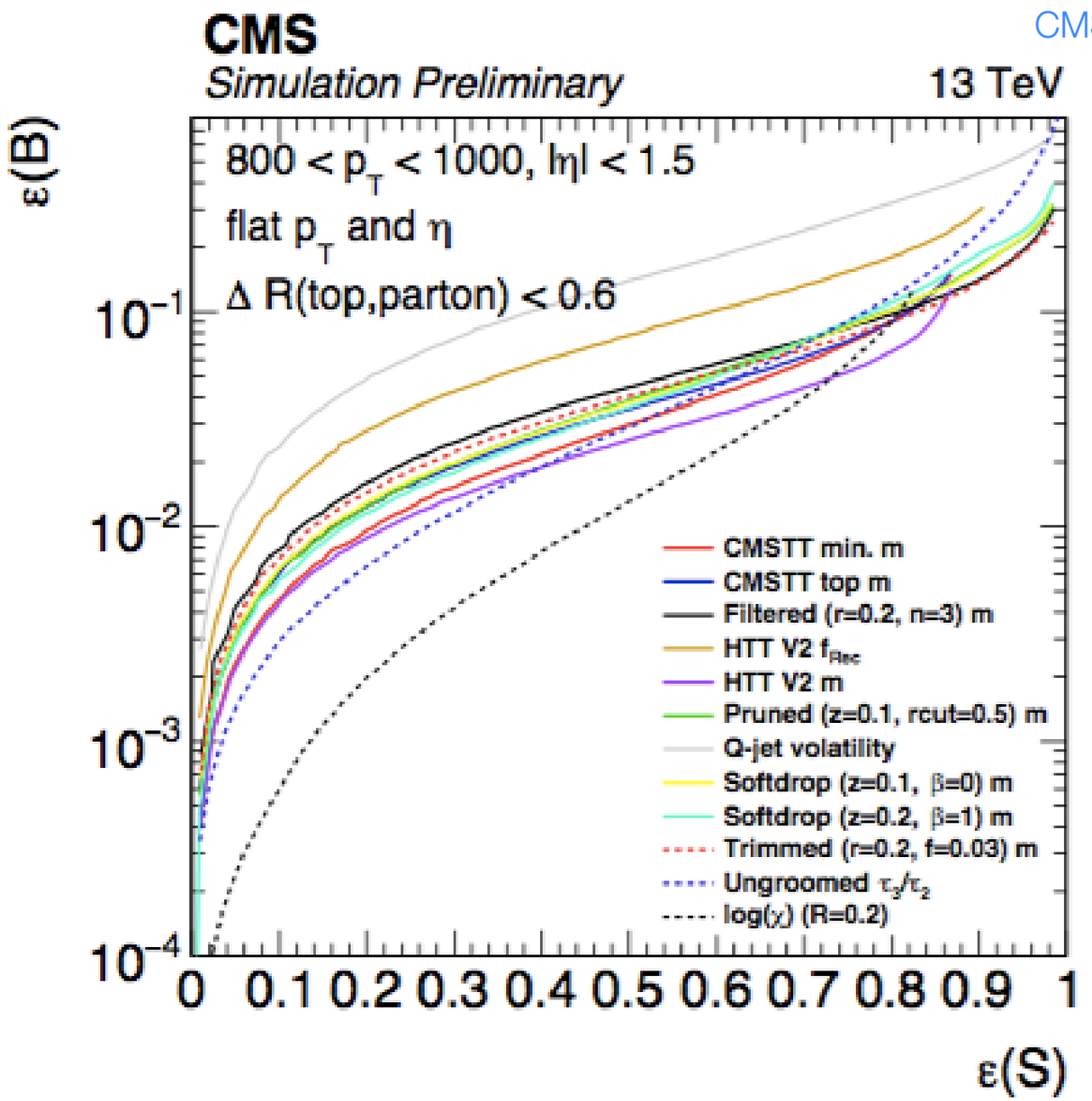


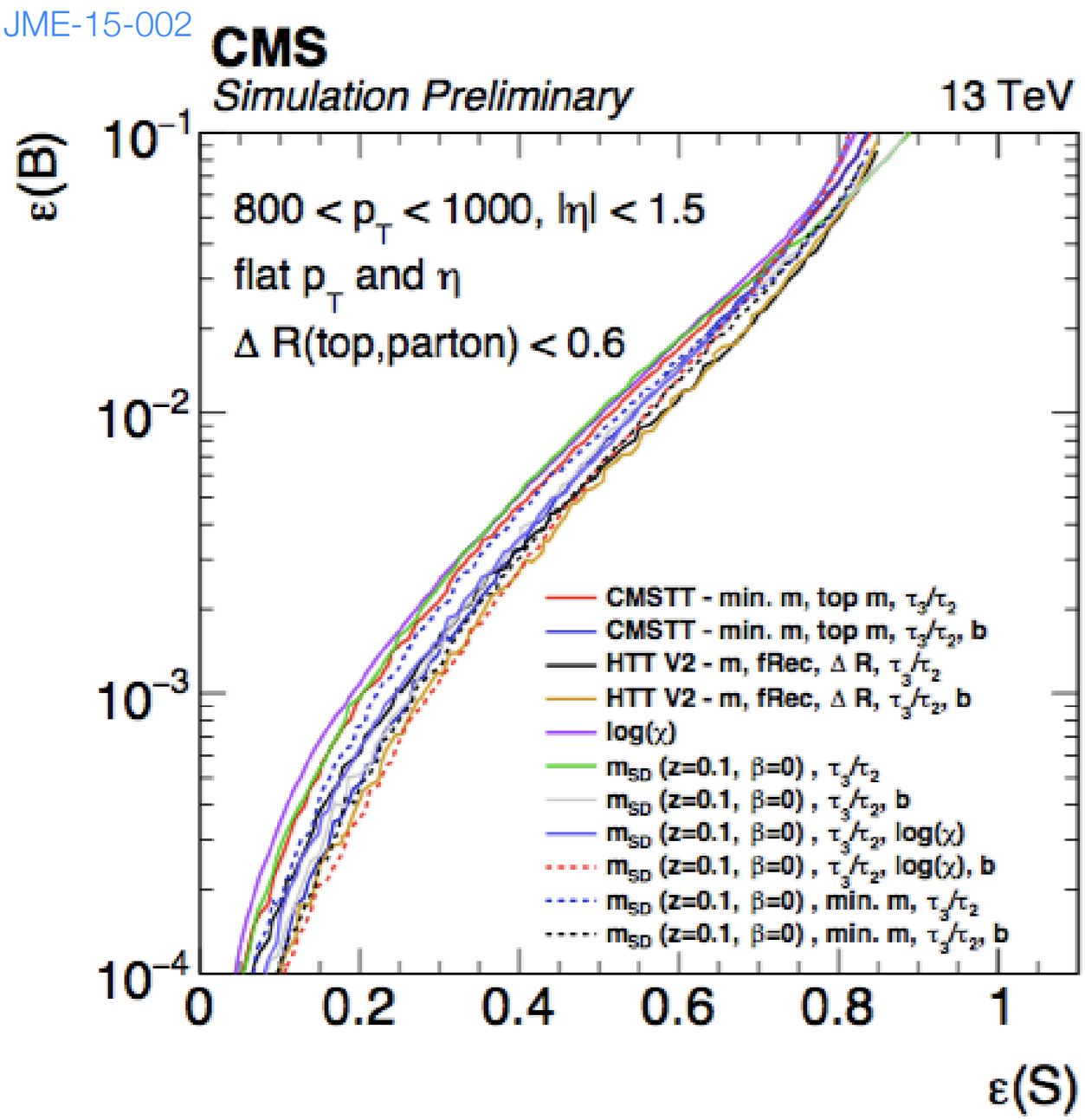
[1] D. Soper, M. Spannowsky arXiv:1211.3140
[2] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, arXiv:0806.0848
[3] T. Plehn, M. Spannowsky, M. Takeuchi, D. Zerwas, arXiv:1006.2833
[4] G. Kasieczka, T. Plehn, T. Schell, T. Strebler, G. Salam arXiv:1503.05921

Top tagging performance



CMS JME-15-002



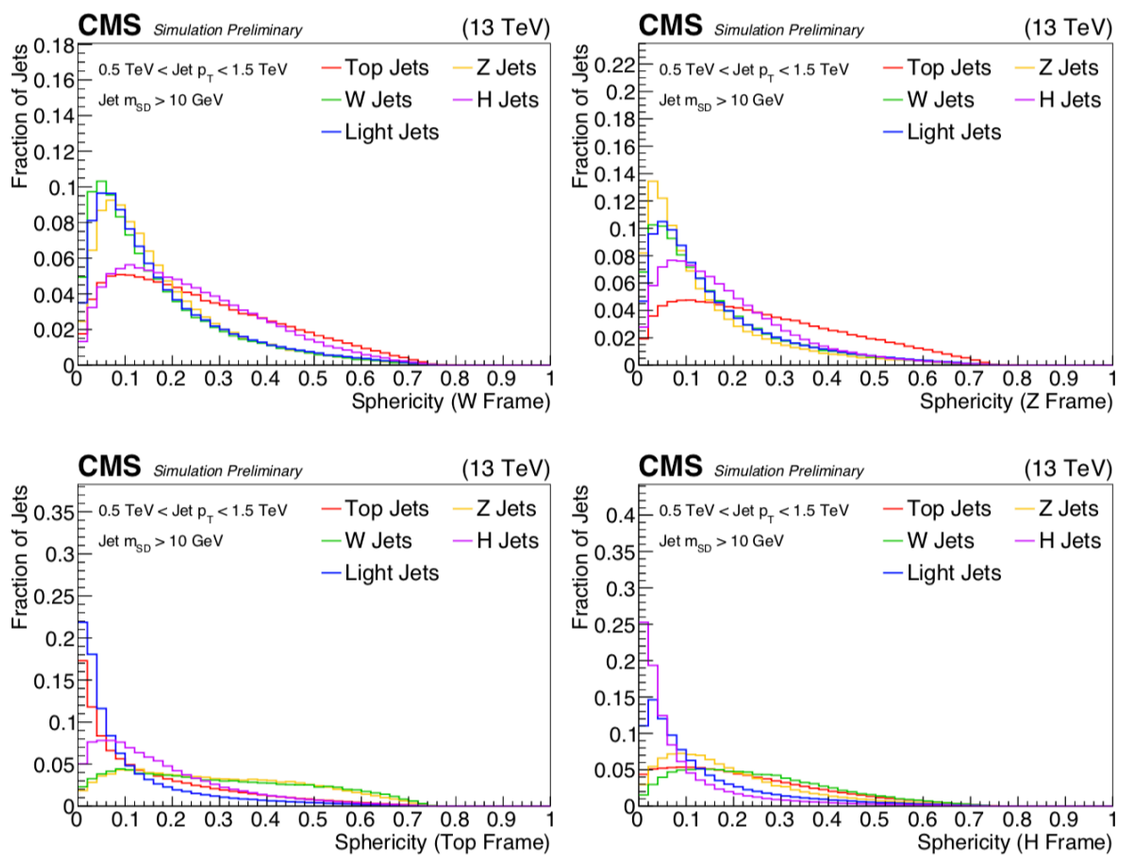


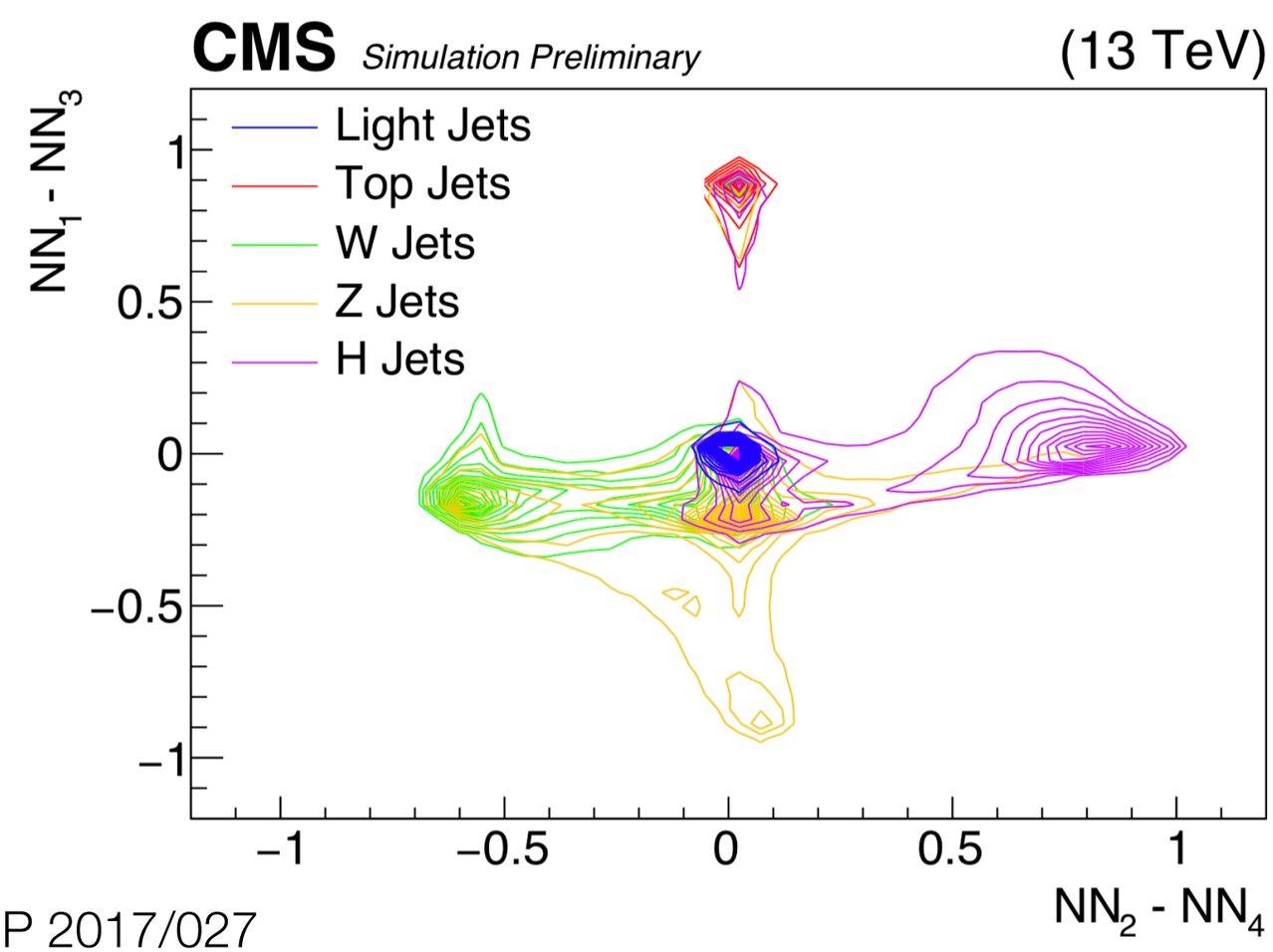
CMS default:

- Anti-kt R=0.8 PUPPI jet
- MMDT/Soft drop jet mass
- N-subjettiness ratio τ_3/τ_2

Multi-category taggers

- BEST (Boosted Event Shape Tagger)
- Deep AK8

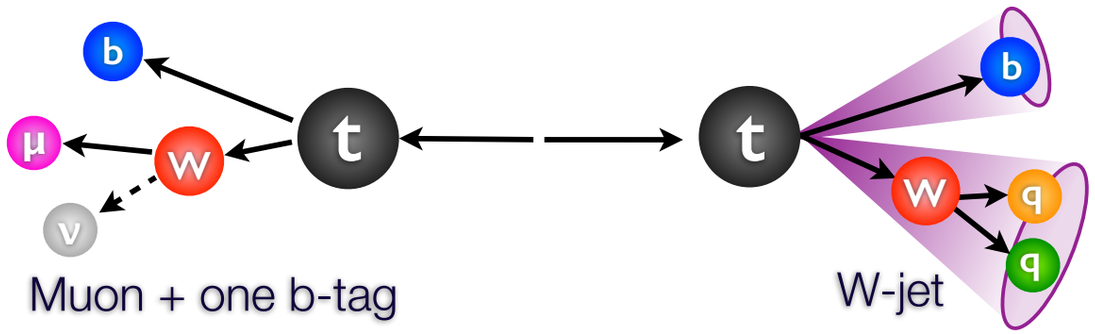


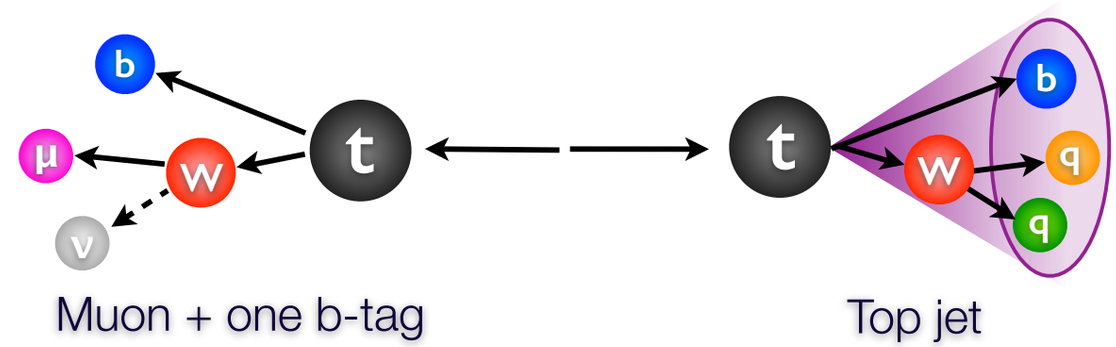


CMS DP 2017/027

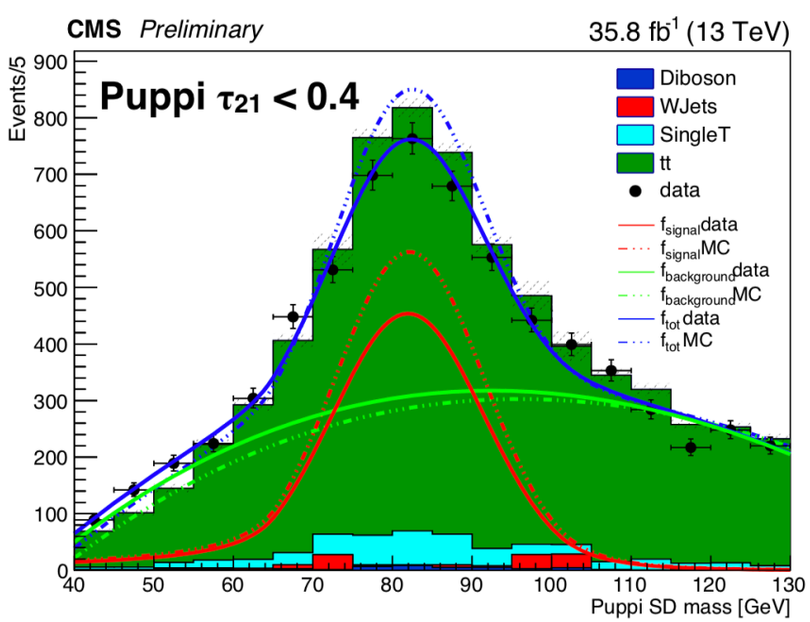
Top and W jet validation in data

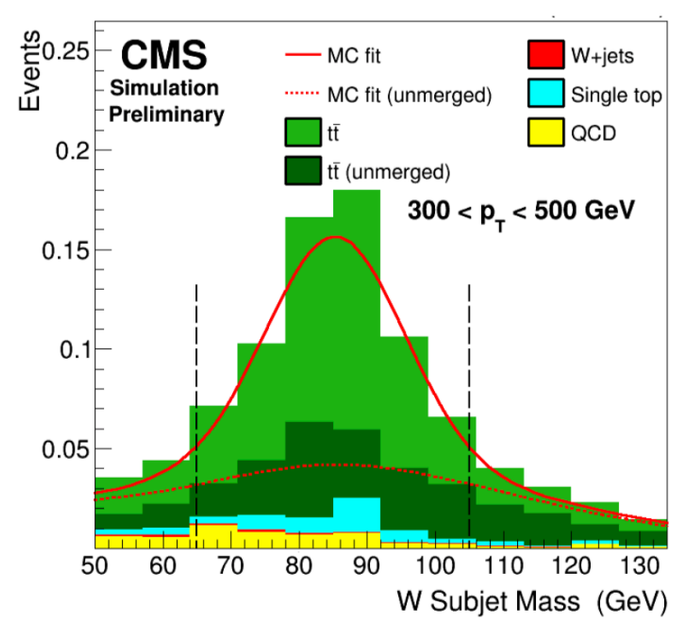
- Semileptonic ttbar selection → very pure sample of boosted Ws

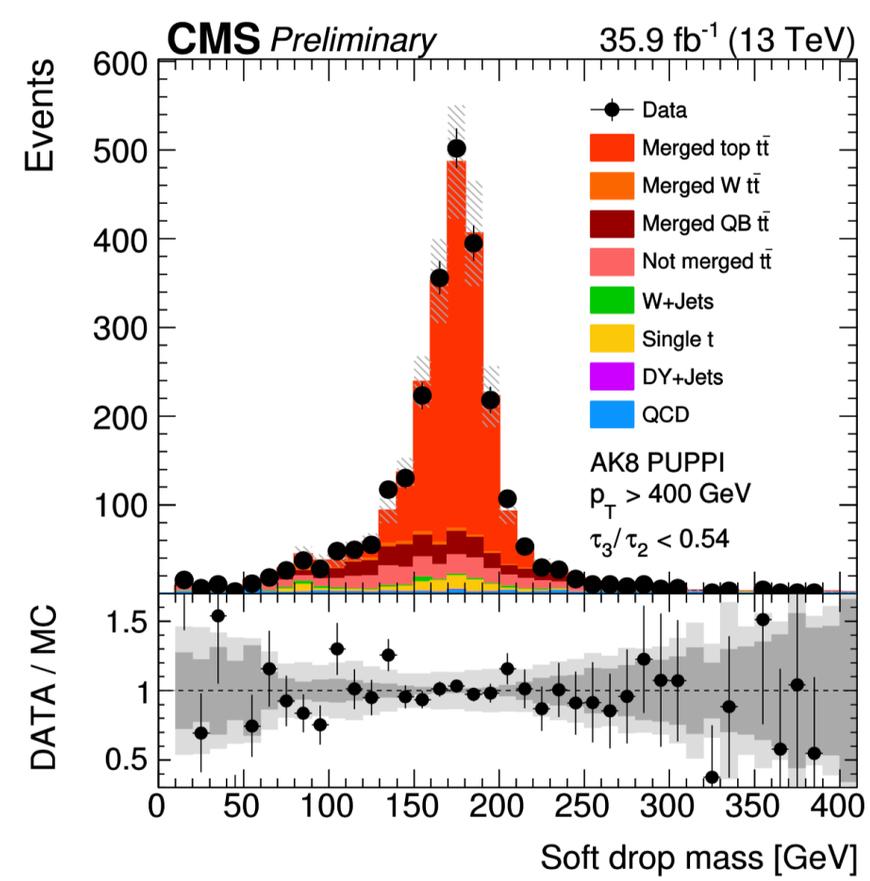




- Data-MC scale factors measured



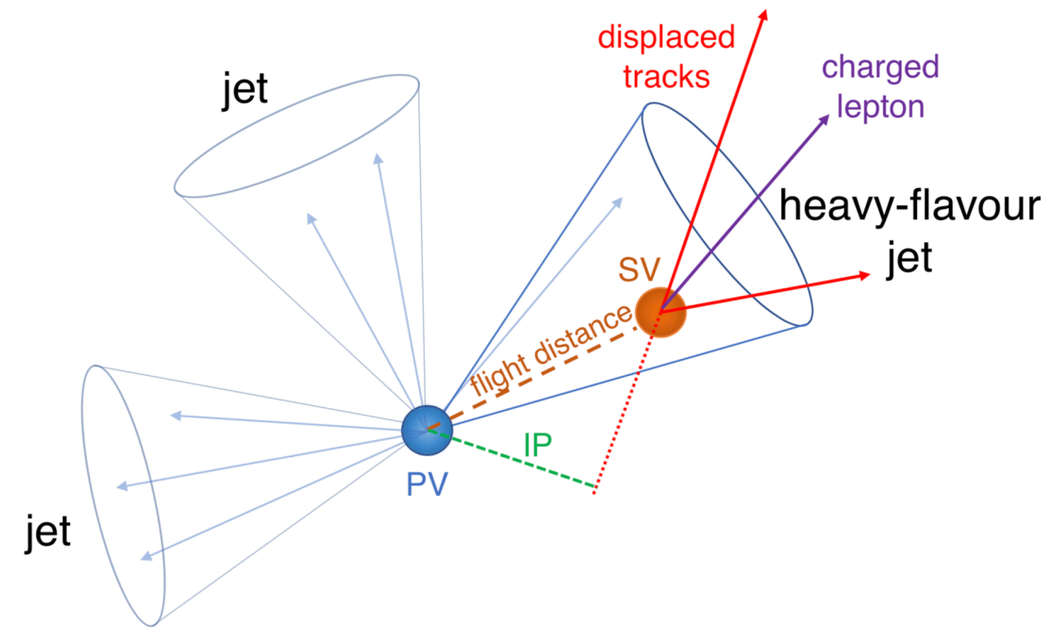


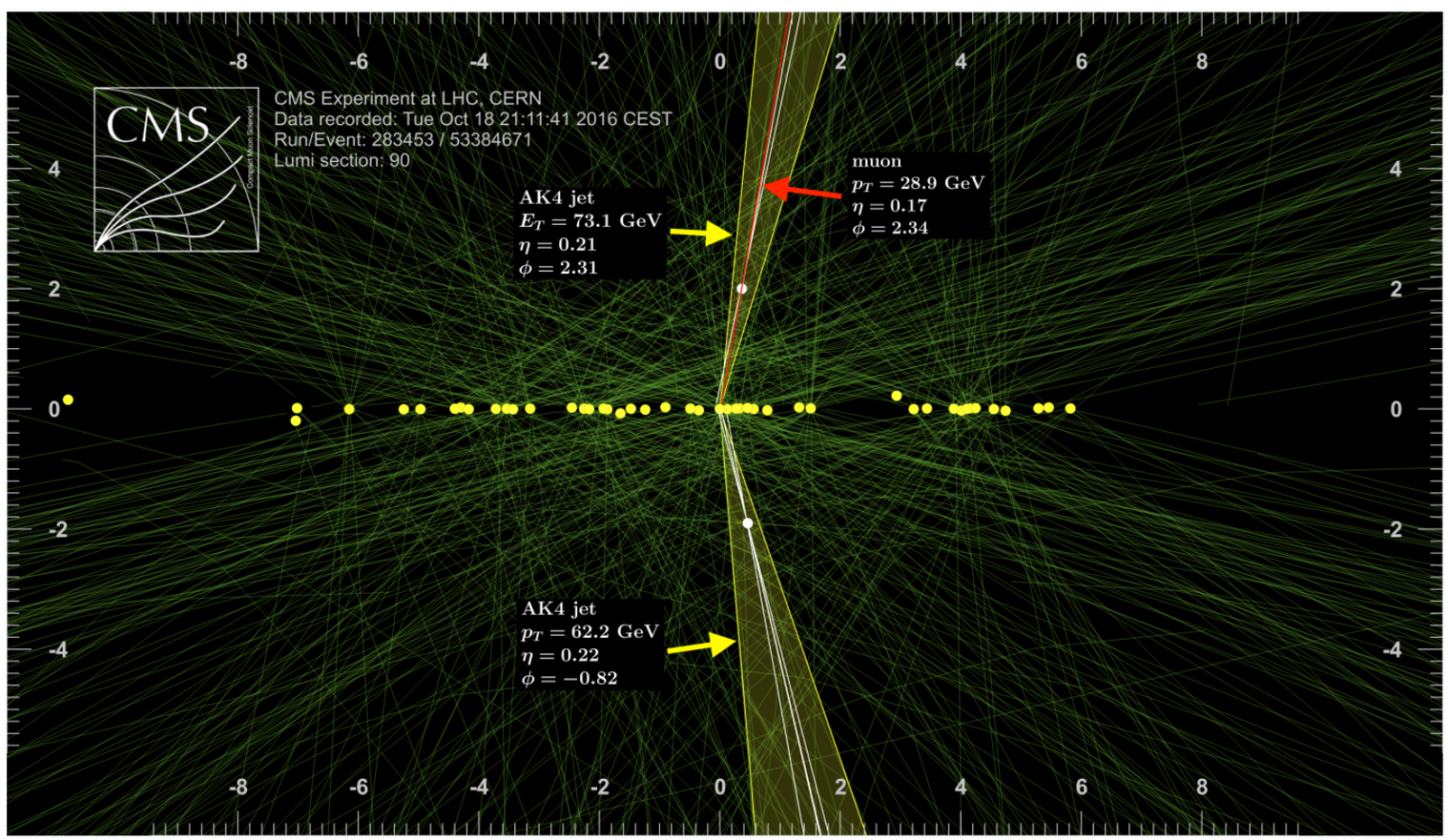


CMS DP 17/026

Heavy flavour tagging at CMS

- Identify jets originating from a b or c quark
- Tagging algorithms rely on track and secondary vertex information

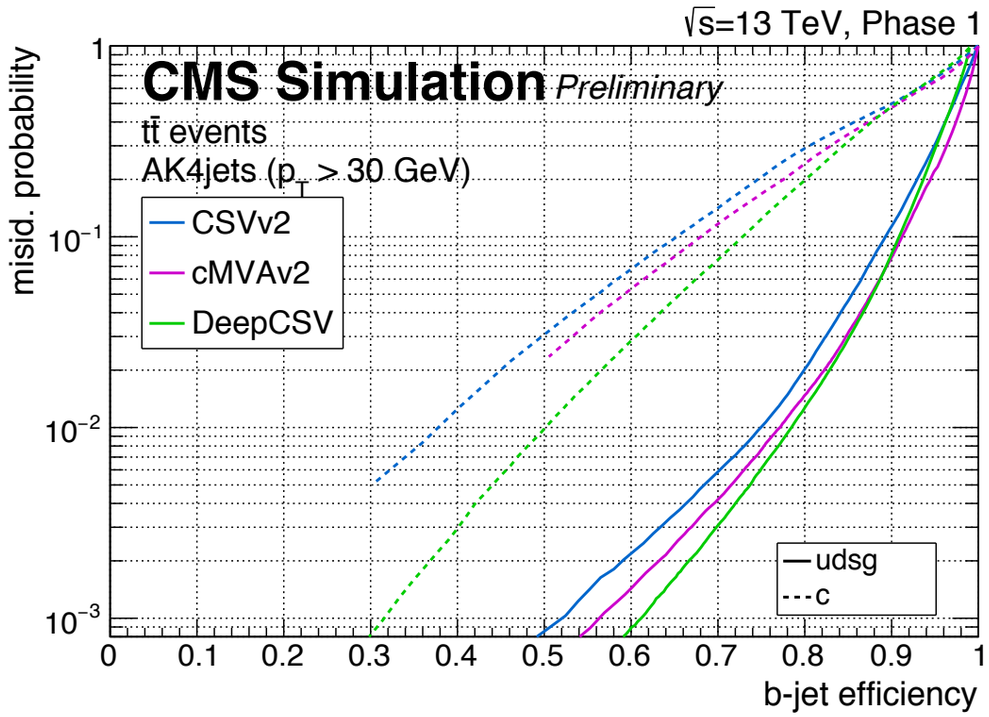


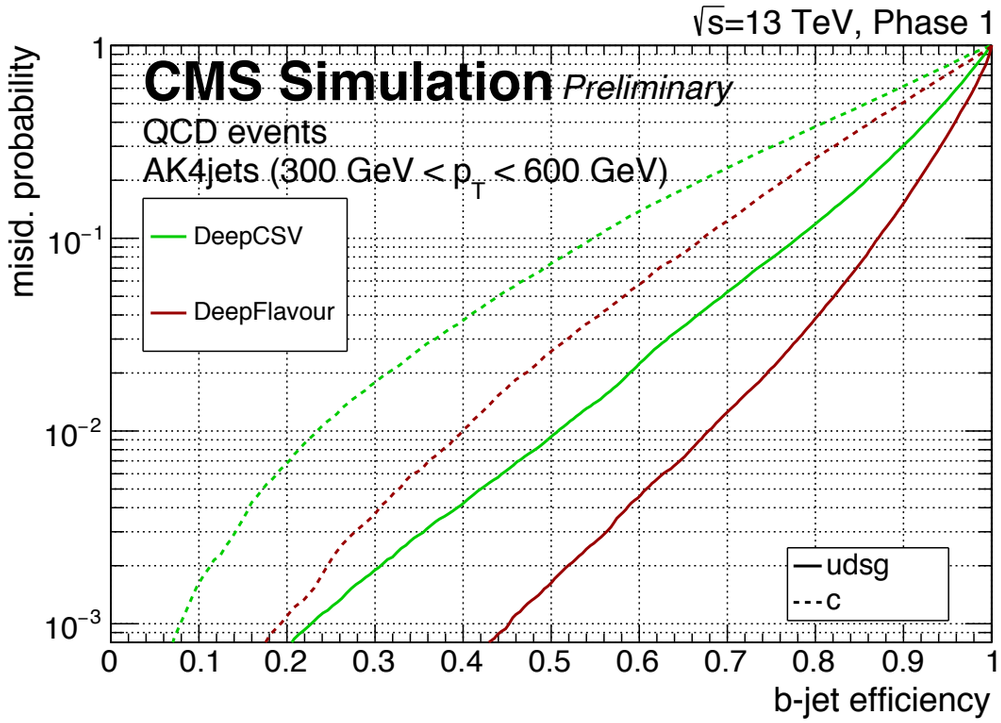


CMS DP 2017/032

Heavy flavour taggers

- Optimized algorithms for Run II
- CSVv2
  - 19 variable shallow neural network (1 hidden layer)
- cMVAv2
  - BDT combines CSVv2 with taggers based on soft leptons and track probability information
- DeepCSV
  - Similar to CSVv2 but with a deep neural network (4 hidden layers with 100 notes each) and with more tracks considered
- Deep Flavour (*in development*)
  - Deep neural network using particle flow particles and secondary vertices as input

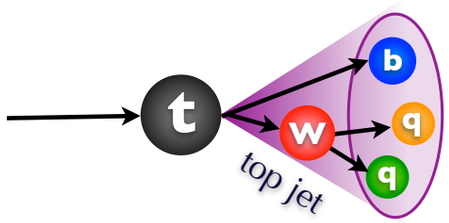




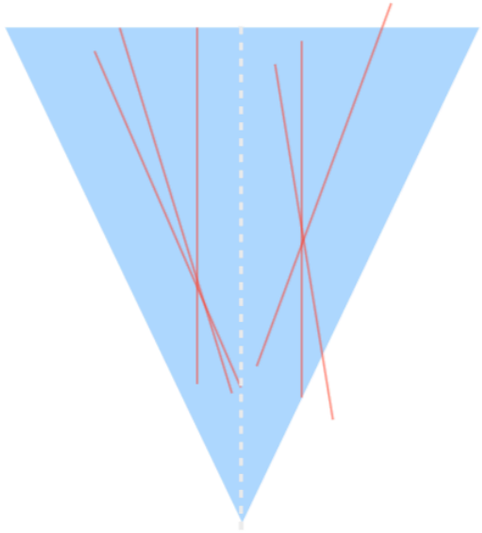
CMS DP 2017/013

b-tagging in boosted topologies

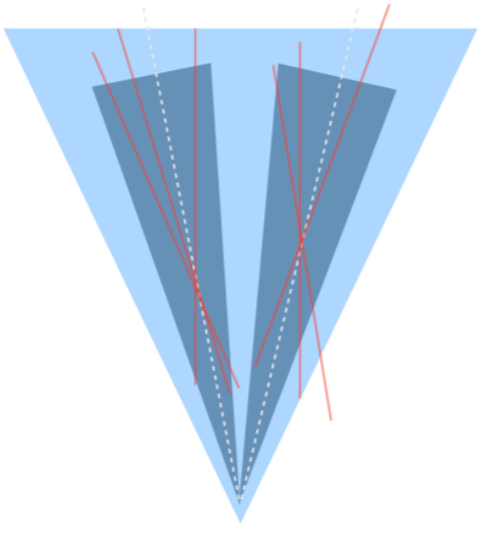
- Boosted top
  - **Fat jet method** - require the large AK8 jet be b-tagged
  - **Subjet method** - Use grooming algorithms to identify subjets and require 1 b-tagged subjet



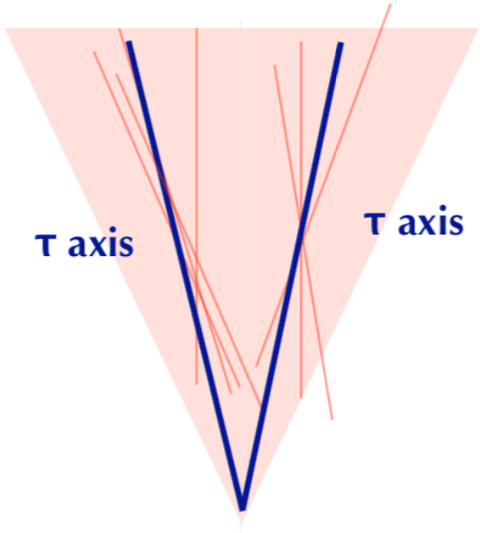
- Boosted Higgs → bb, boosted Z → bb, etc.
  - **Subjet method** - identify two subjets which can be b-tagged
  - **Dedicated double b-tagger** - trained specifically for this topology



AK8 jet

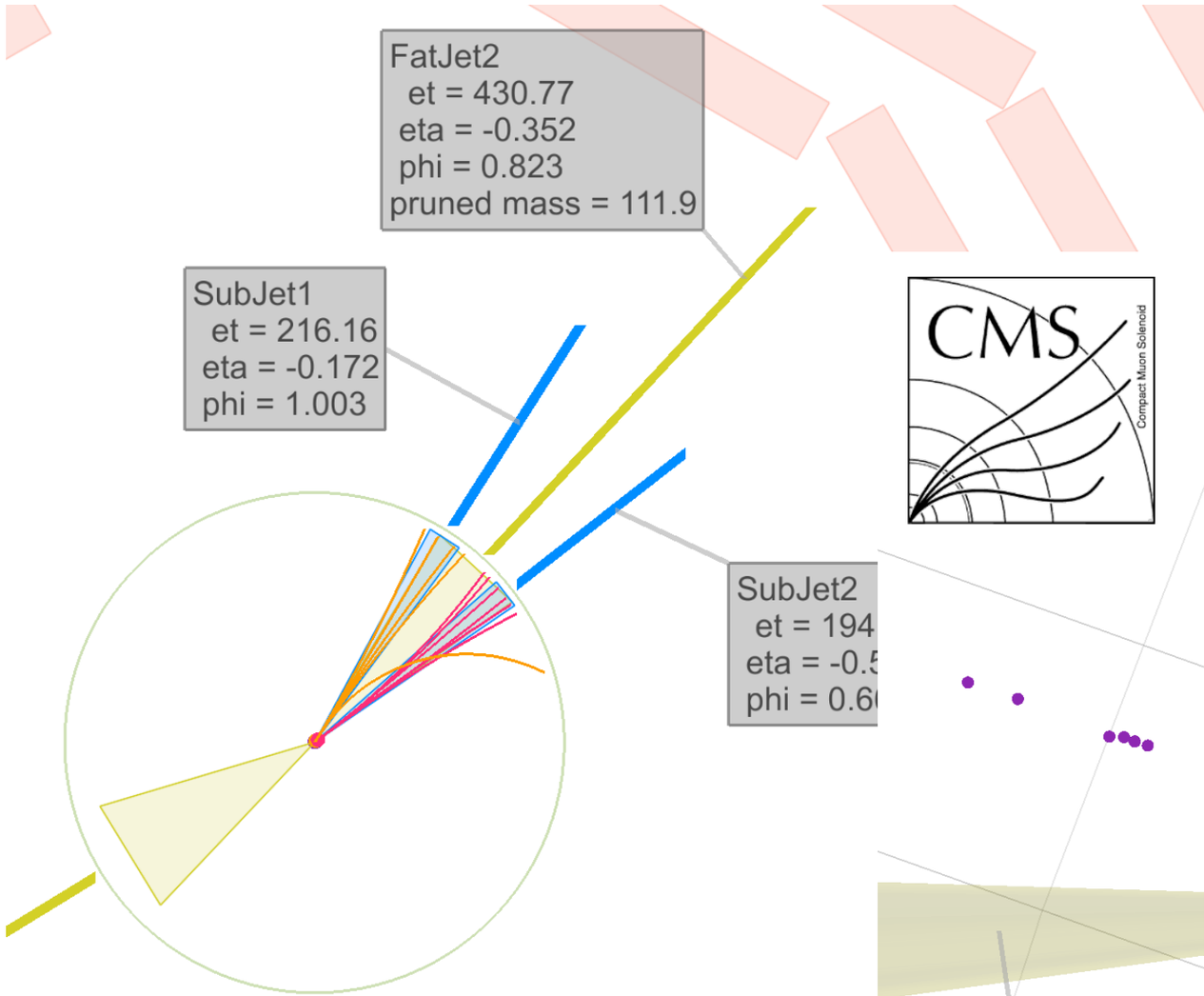


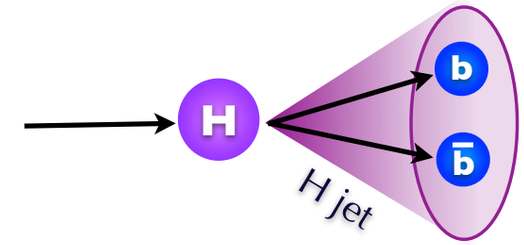
subjets

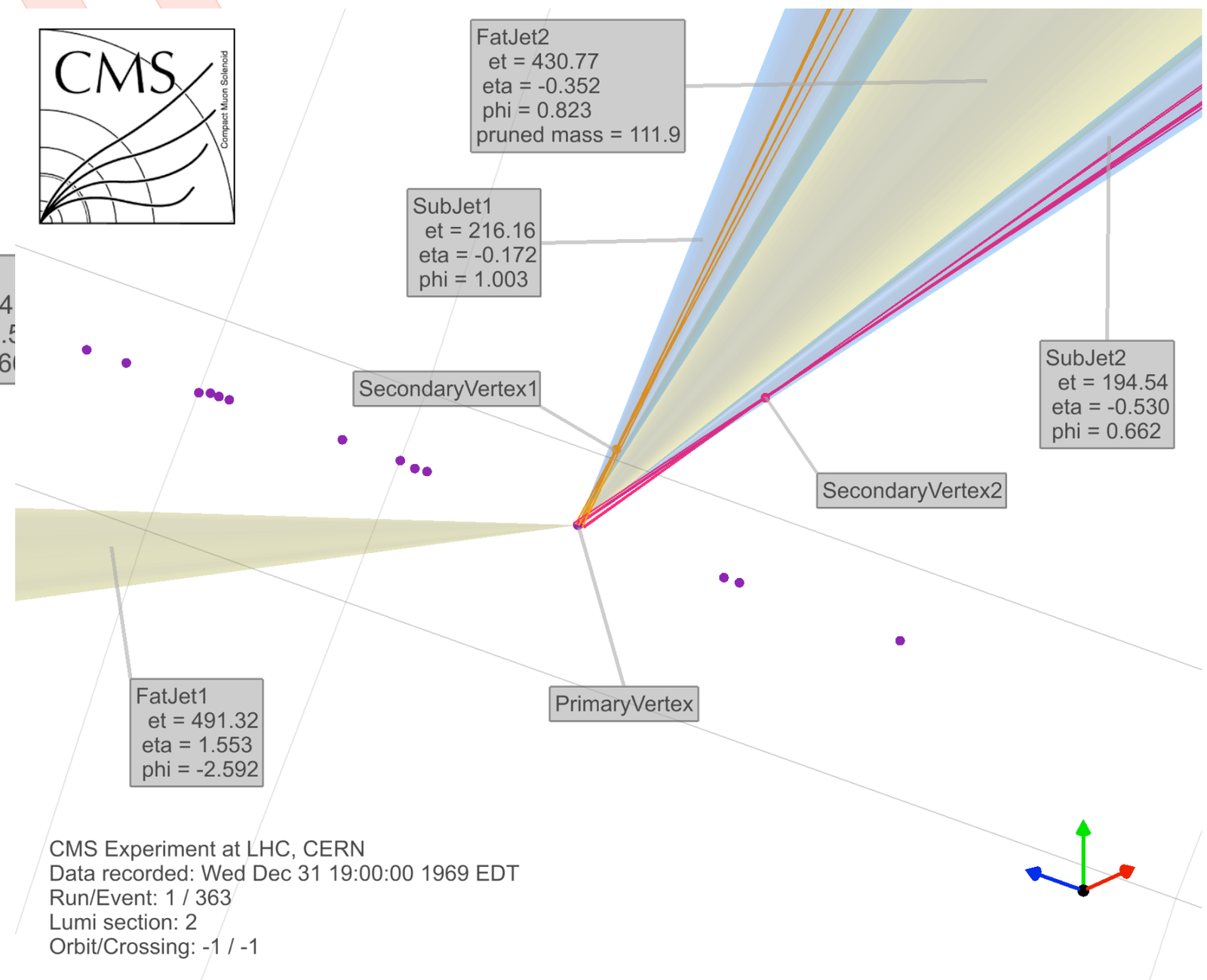


double-b

Double b-tagged H-jet





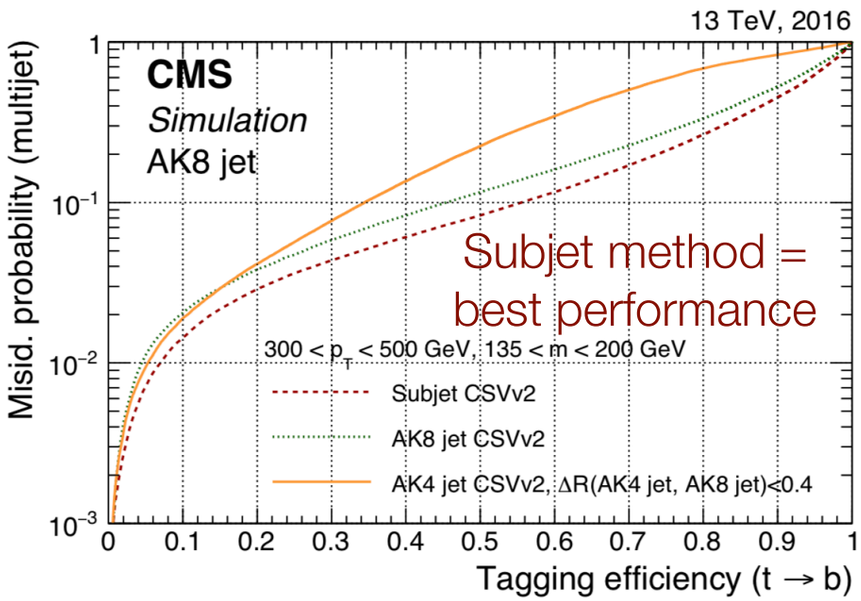


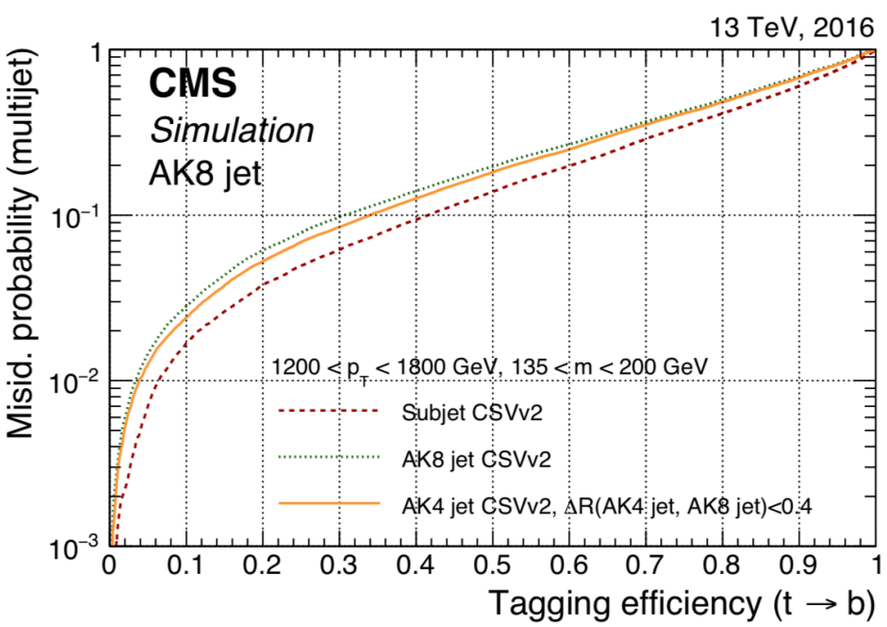
CMS DP 2017/032

b-tagging in boosted topologies

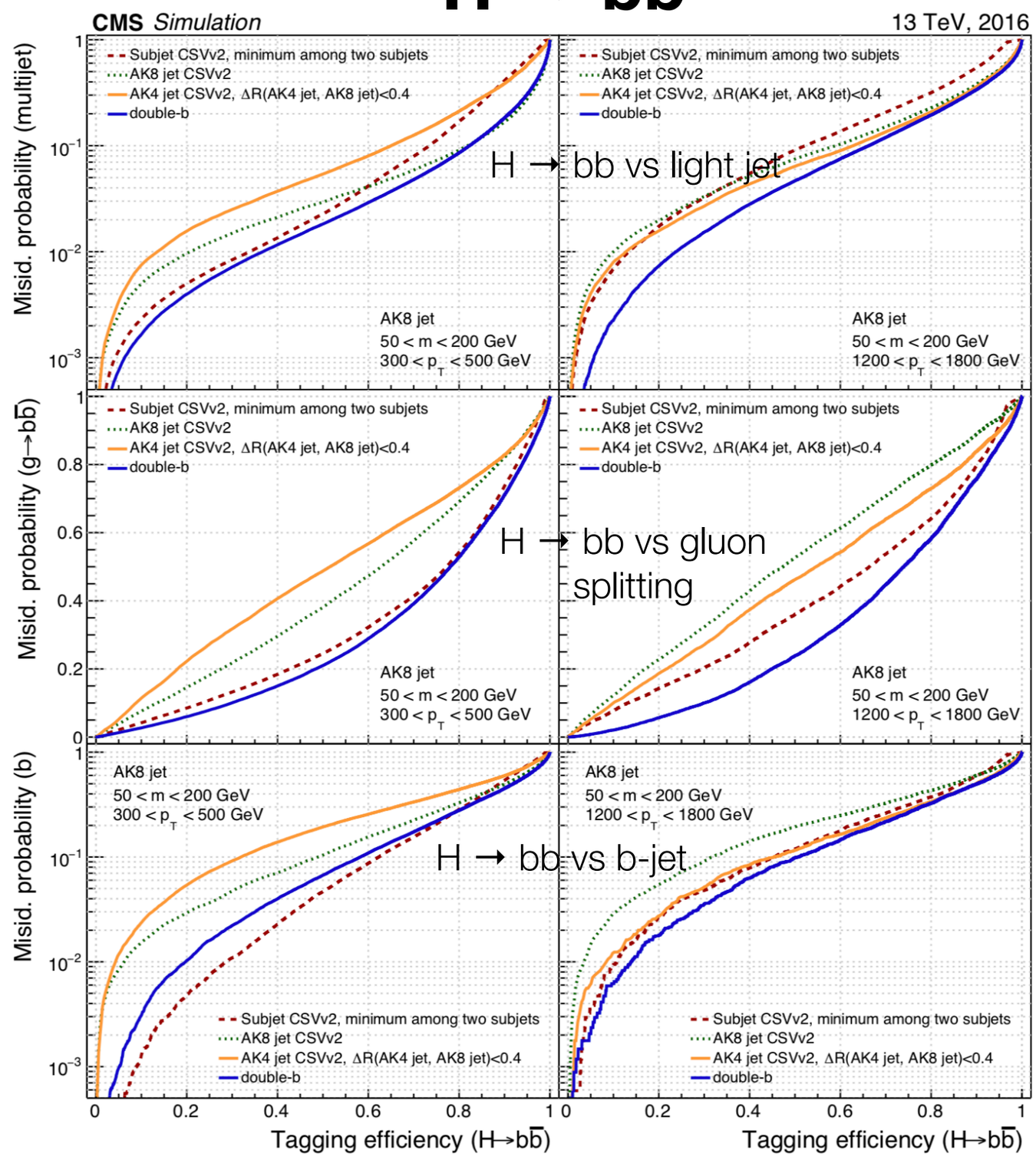
CMS PAS BTV-16-002 ARXIV 1712.07158

t → bW





H → bb



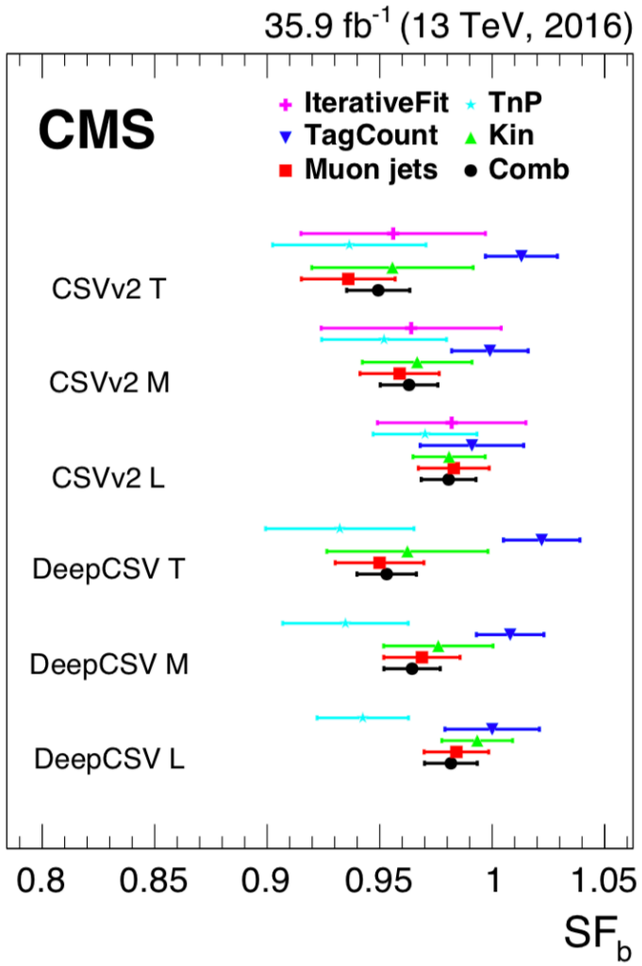
Conclusions

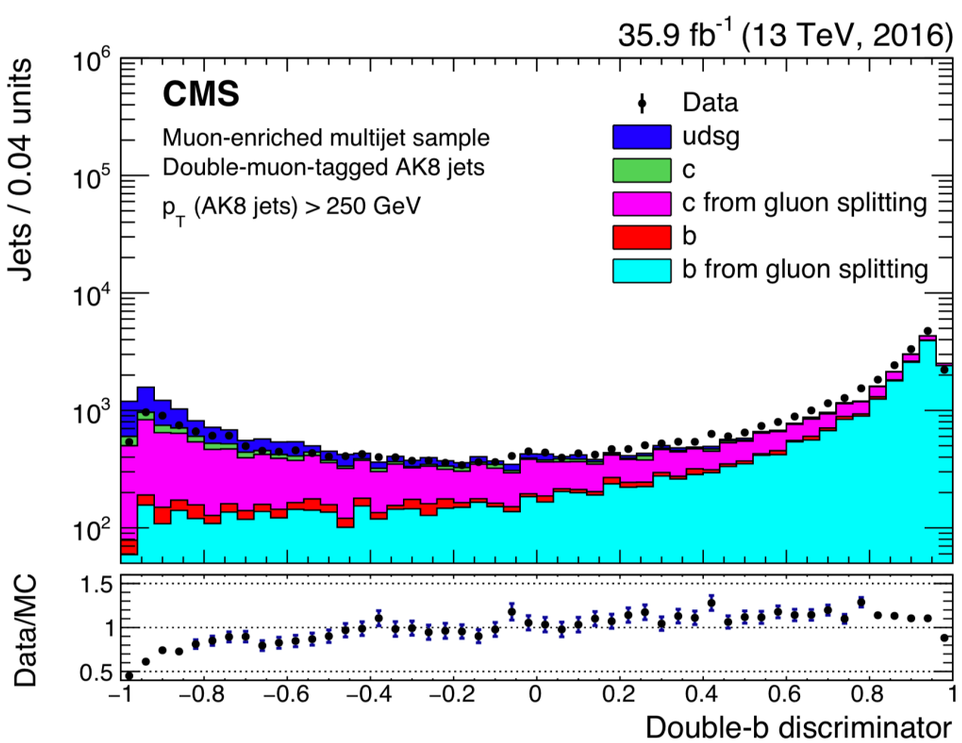
- New jet tagging techniques have proven to be indispensable to LHC analyses
  - Impressive progress in the last 7 years developing and commissioning jet tools
  - Boosted objects are now mainstream
    - Thoroughly calibrated and commonly used
    - example: groomed jet mass is now in the CMS trigger
    - Strong community of theorists and experimenters
      - Short turnaround time between new ideas and results
  - Modern techniques have led to advances in heavy flavour tagging

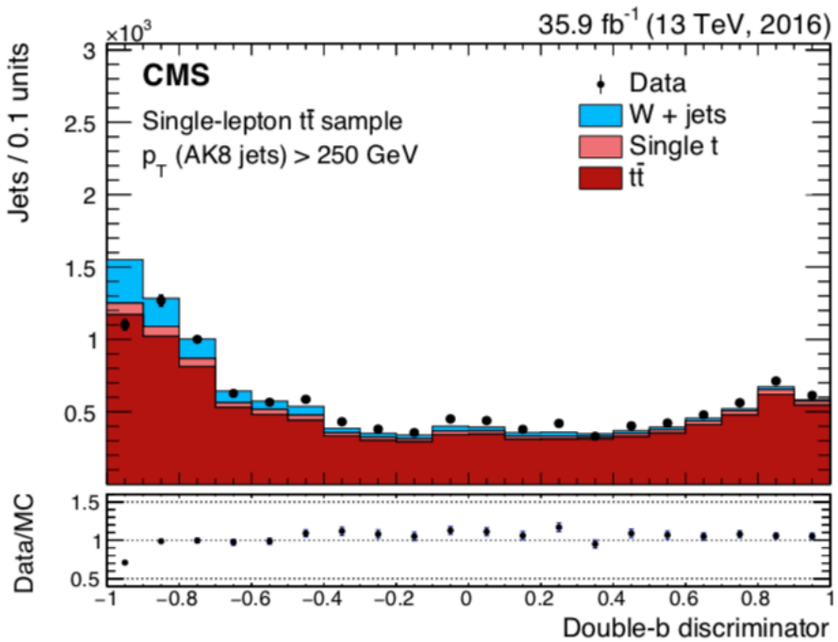

BACKUP

b-tagging validation in data

- b-jet rich events
  - Events with muons from b-decays
  - tt events
- QCD, DY enriched + c-jet enriched topologies
- Double b-tagger
  - g → bb jet with two tagged muons

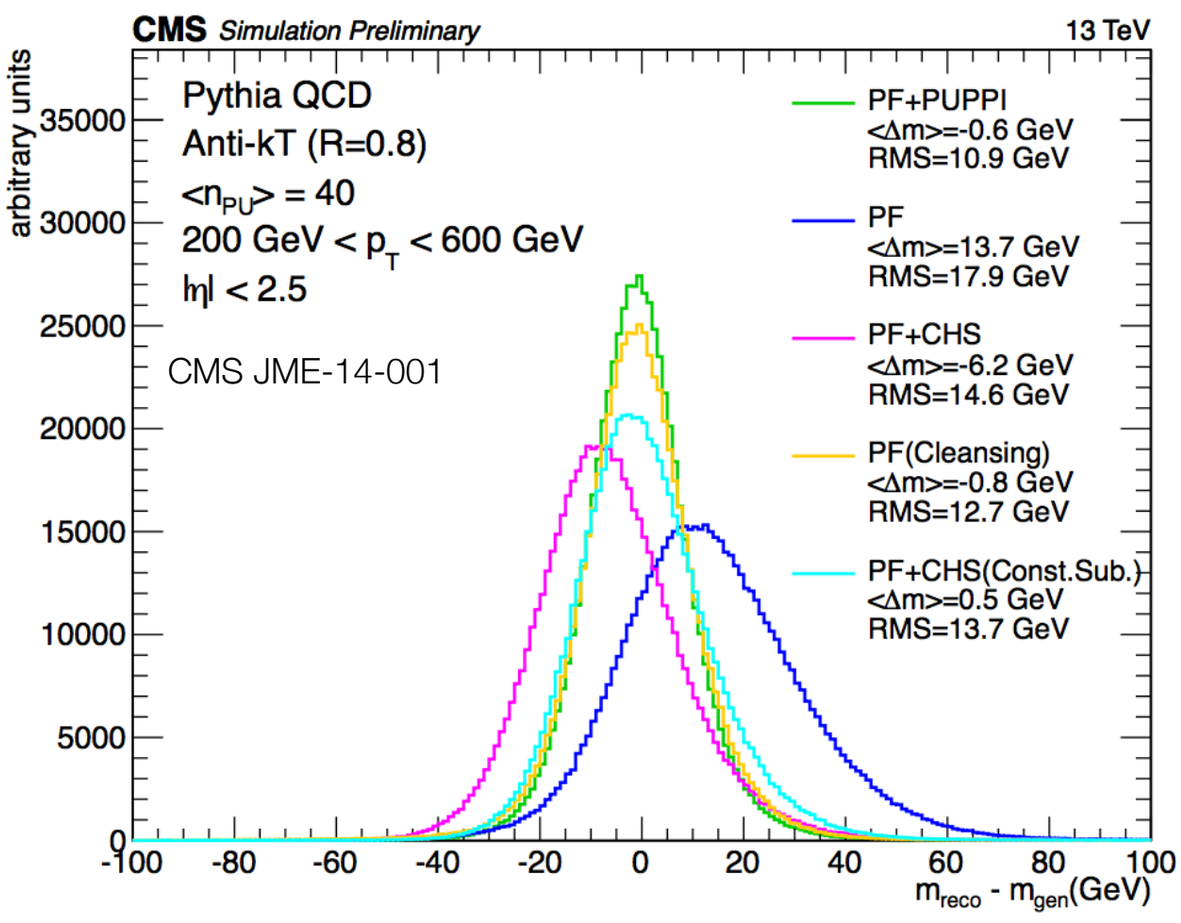






Pileup removal before jet clustering

- Charged Hadron Subtraction (CHS)
  - enabled by particle flow
  - remove charged particles originating from pileup vertices before clustering jets
  - does not remove neutral pileup



- New tools which utilize additional information for pileup removal
  - ***Constituent subtraction [1]*** - per particle area subtraction
  - ***Jet cleansing [2]*** - charged particle vertex information used to correct jets at the subjet level
  - ***Soft killer [3]*** - progressively remove soft particles until the average pileup density in the event is 0
  - ***Pileup Per Particle Identification (PUPPI) [4]*** - jet shape and charged particle vertex information used to suppress pileup

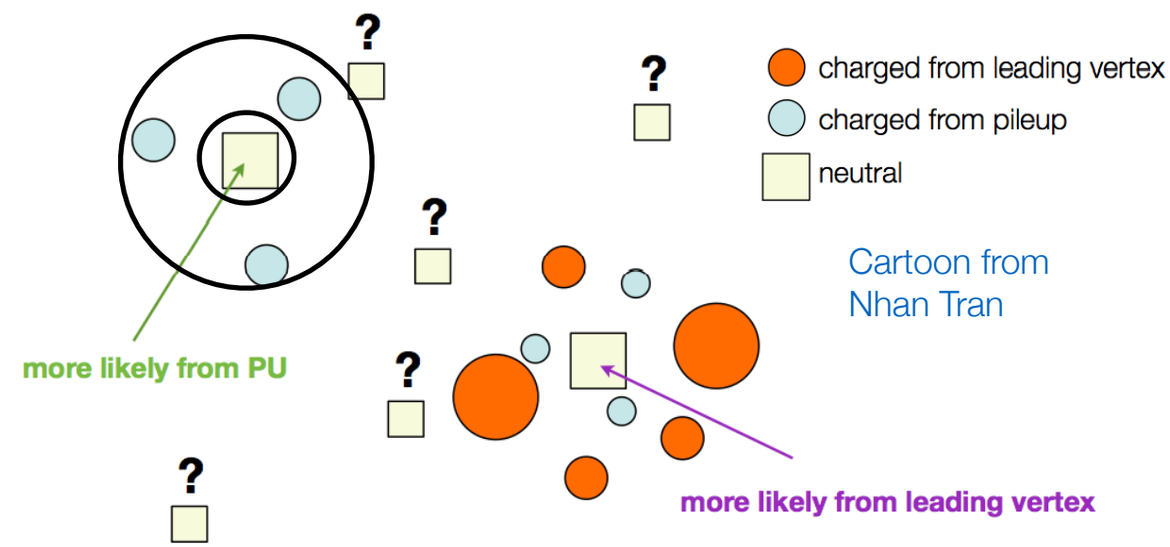
Jet mass resolution improved by Charged Hadron Subtraction (CHS)

Jet mass resolution is further improved by new techniques which also remove neutral pileup

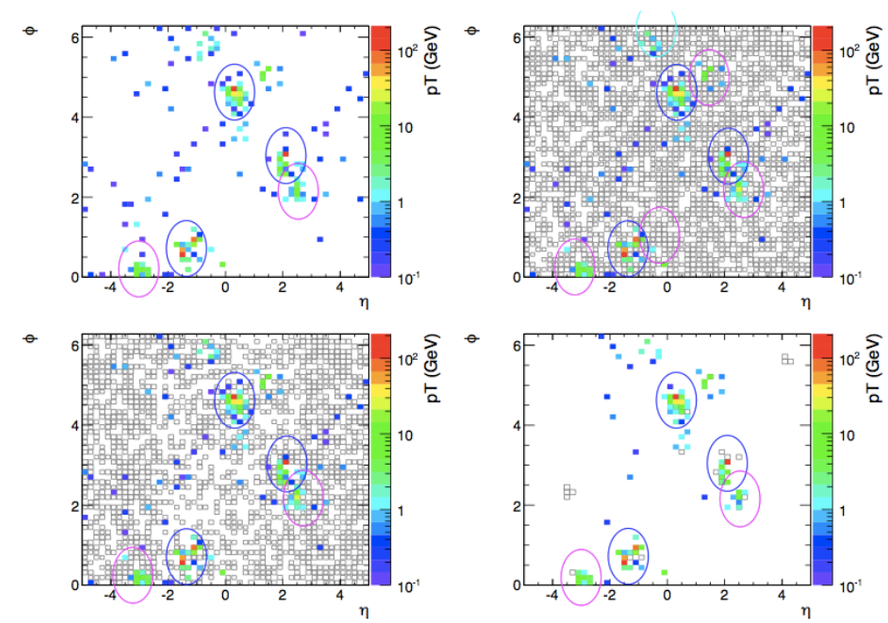
[1] P. Berta, M. Spousta, D. Miller, R. Leitner arXiv:1403.3108
[2] D. Krohn, M. Low, M. Schwartz, L. Wang, arXiv:1309.4777
[3] M. Cacciari, G. Salam, G. Soyez, arXiv:1407.0408
[4] D. Bertolini, P. Harris, M. Low, N. Tran, arXiv:1407.6013

PileUp Per Particle Identification (PUPPI)

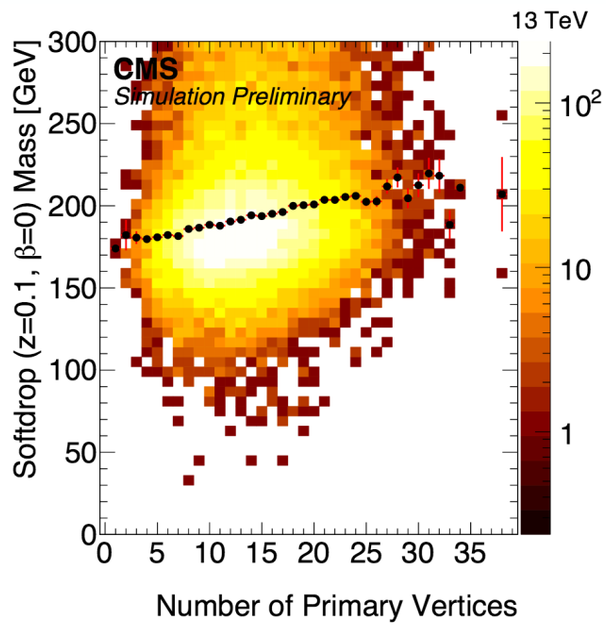
- Pileup handles:
  - Tracking/vertexing→ we know which charged particles come from pileup and which ones come from the hard scatter
  - Pileup is randomly distributed, while collinear radiation from a particle from the hard scatter is preferentially radiated at small angles
  - p_T spectrum of pileup falls quickly

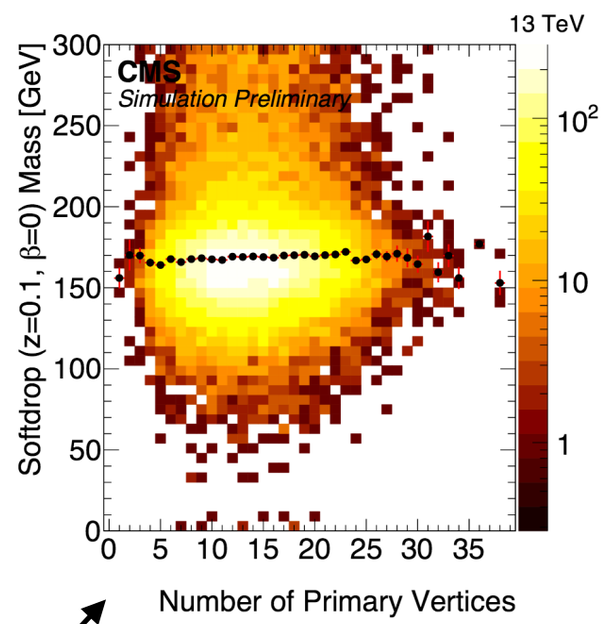


- PUPPI assigns each particle a weight based on the likelihood it originated from pileup
  - Handle: Calculate $\alpha=\sum p_{Tj}/\Delta R_{ij}$ for each particle j within an annulus around particle i
    - α small for particles from pileup (nearby particles are soft and at large angle)
    - α large for particles from the LV (radiation around the particle harder and at small angle)
  -

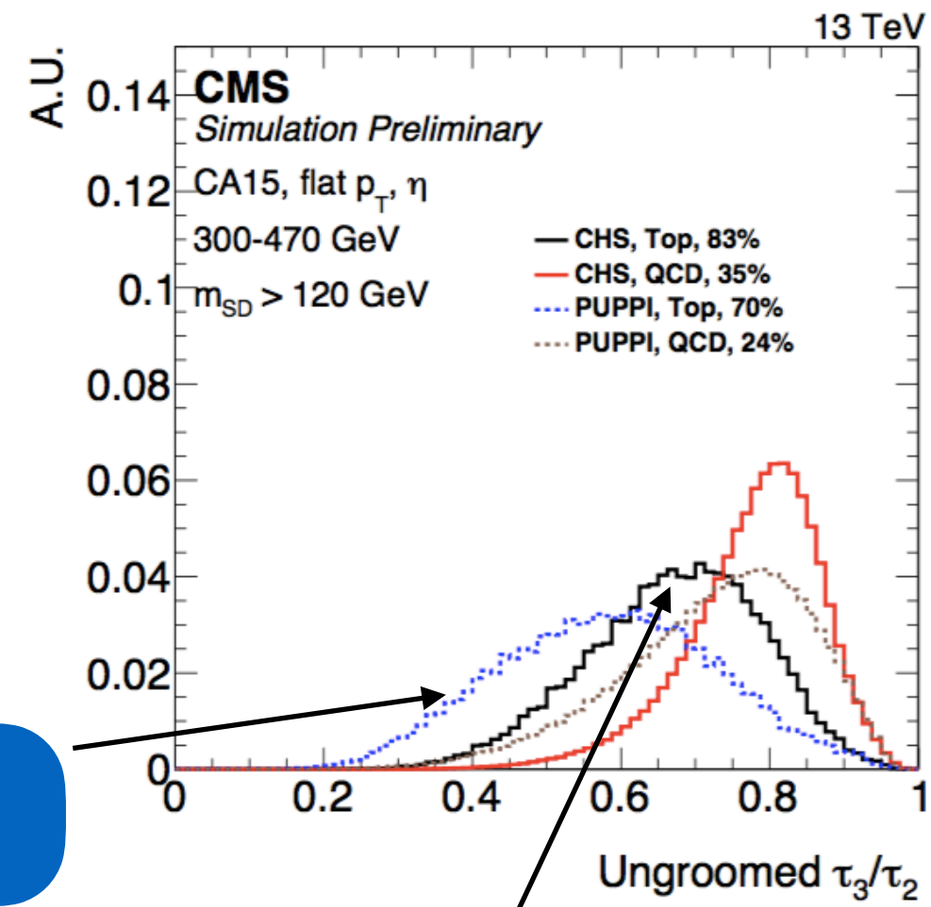


PUPPI performance





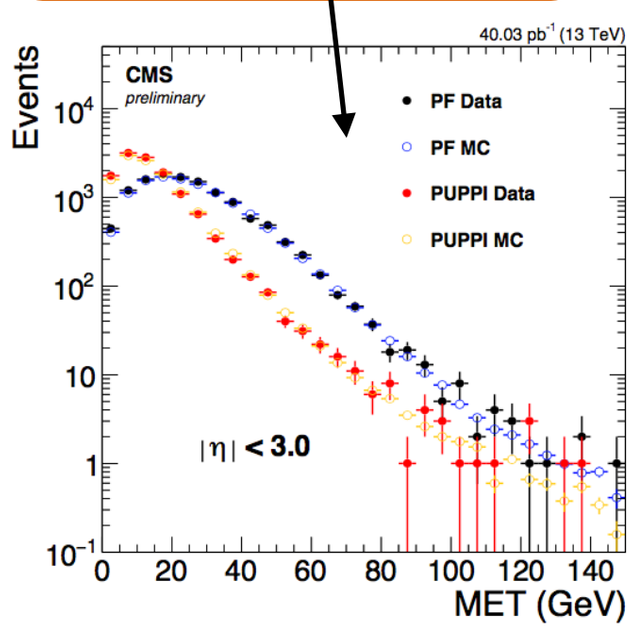
PUPPI jet mass stable with pileup



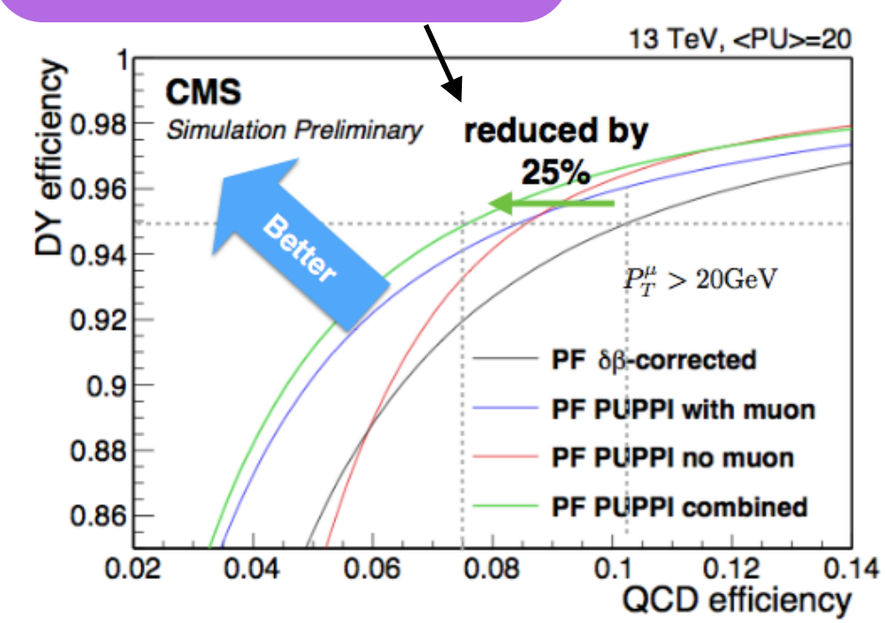
PUPPI corrects jet shapes for pileup → improves tagging discrimination

With pileup the top jet N-subjettiness value is shifted to be more QCD like

Improved MET resolution

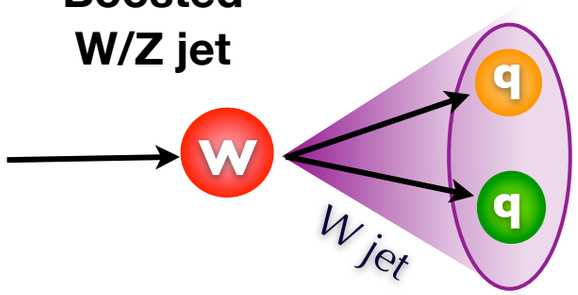


Improved lepton isolation

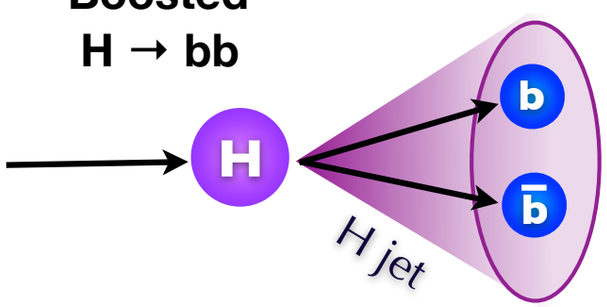


Boosted jet menu

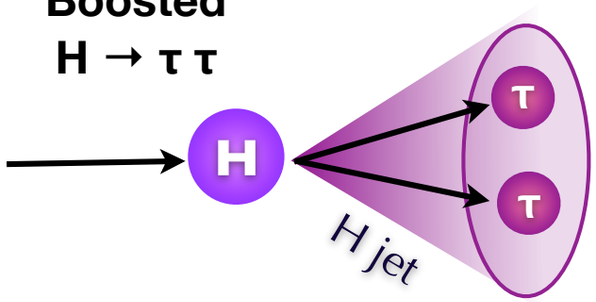
Boosted W/Z jet



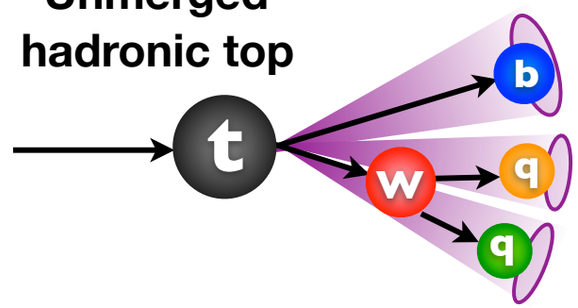
Boosted H → bb



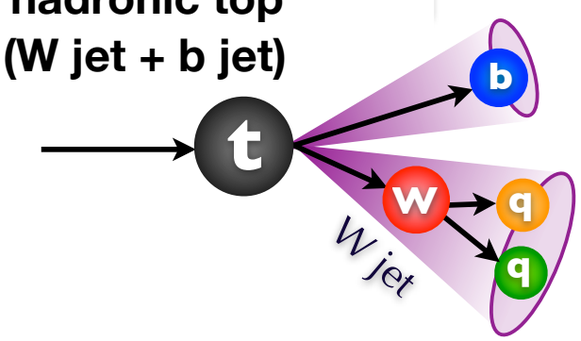
Boosted H → τ τ



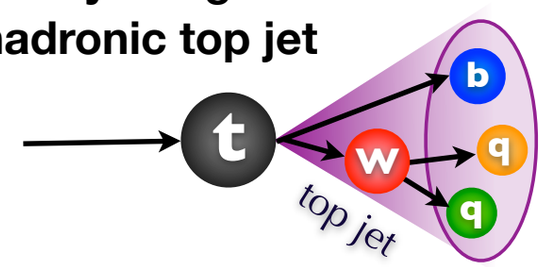
Unmerged hadronic top



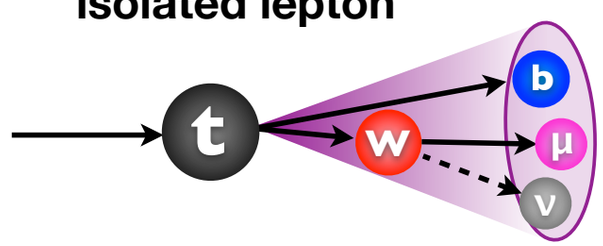
Partially merged hadronic top (W jet + b jet)



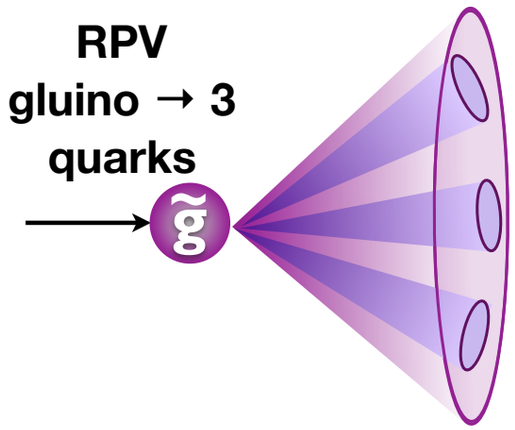
Fully merged hadronic top jet



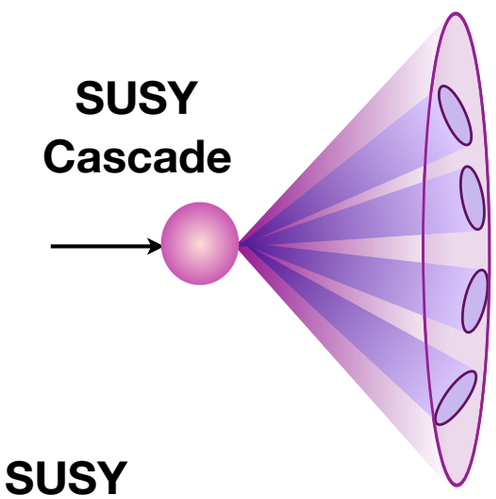
Leptonic top with non-isolated lepton



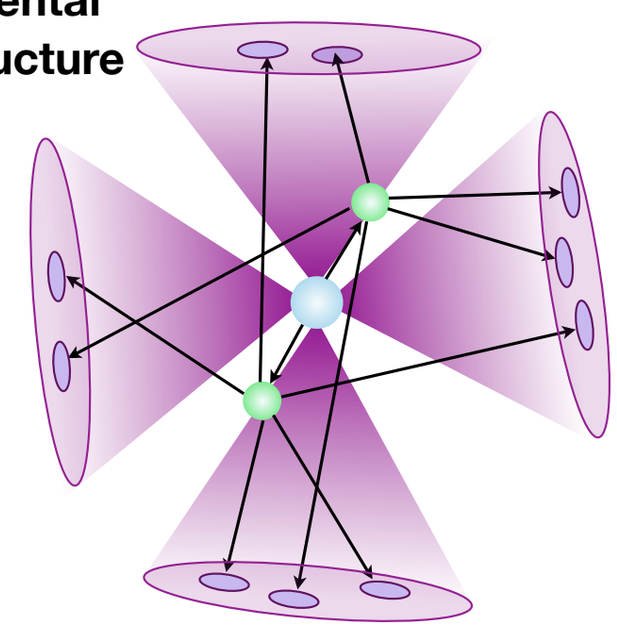
RPV gluino → 3 quarks



SUSY Cascade



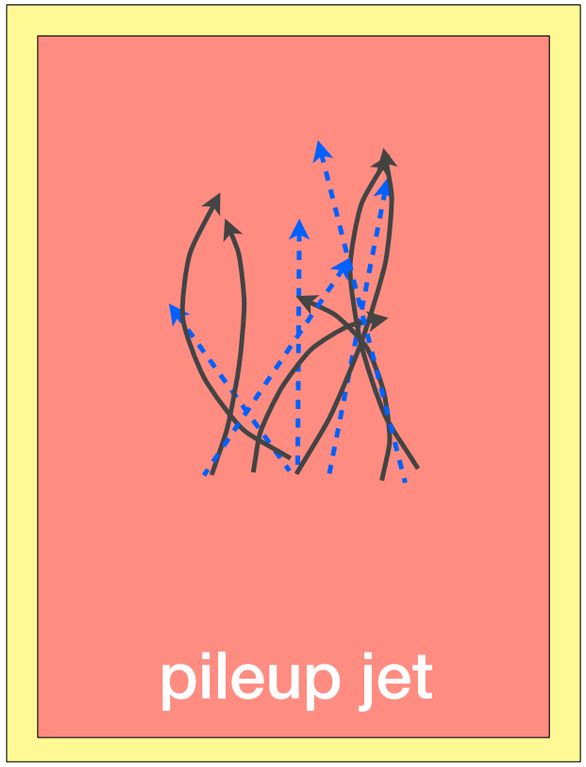
SUSY Accidental Substructure

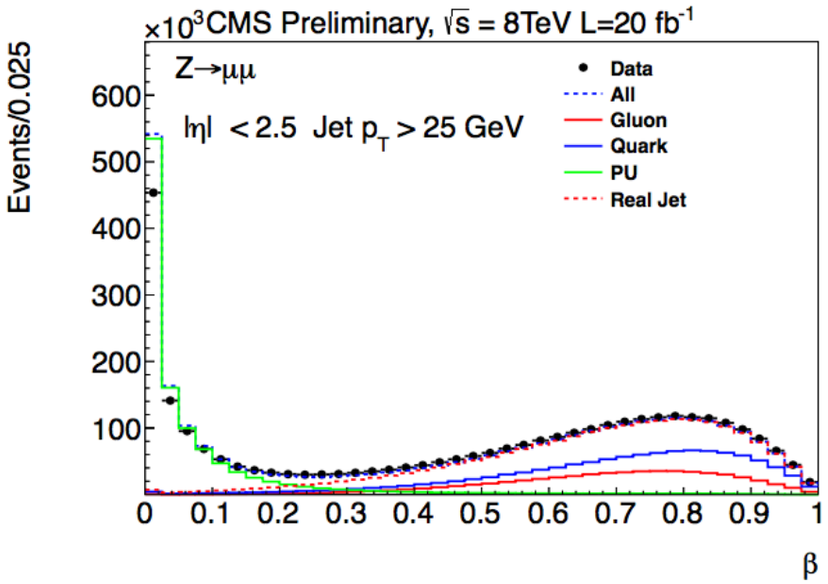


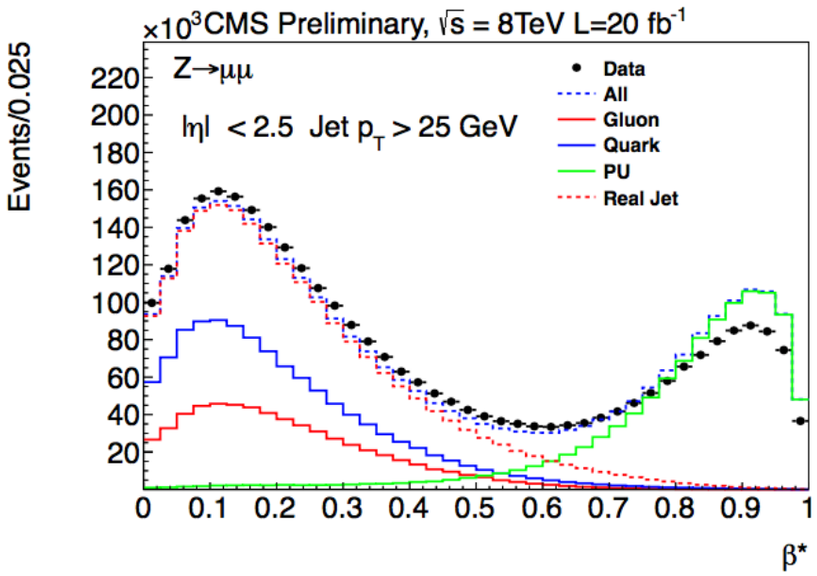
Pileup jets

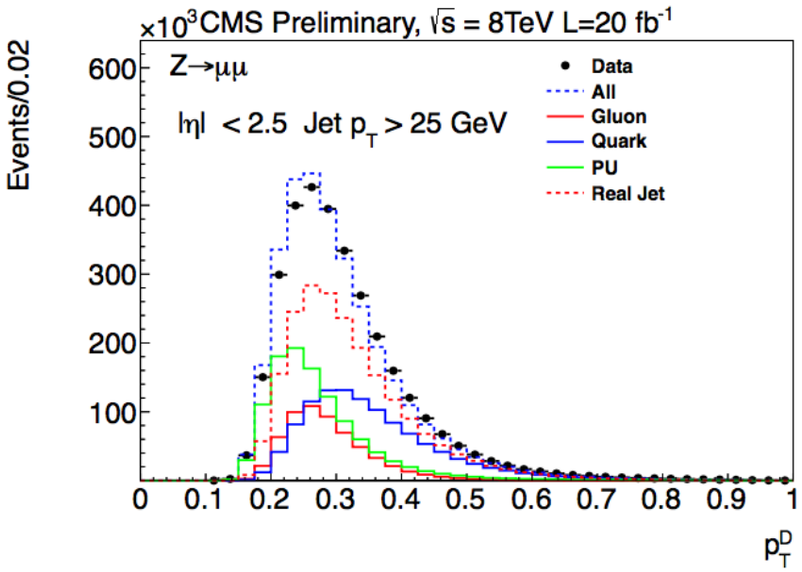
- Many overlapping soft particles from PU can add jet of substantial pT
- Identifying pileup jets relies on charged PU vertex jet shape

$$\beta = \frac{\sum_{i \in PV} p_{Ti}}{\sum_i p_{Ti}} \qquad \beta^* = \frac{\sum_{i \in otherPV} p_{Ti}}{\sum_i p_{Ti}}$$





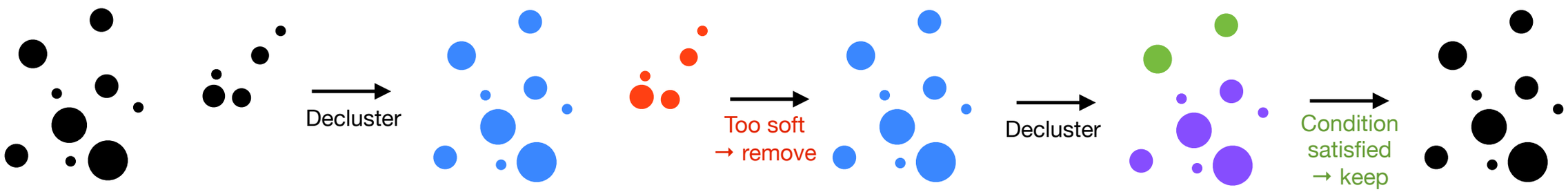




Soft drop

- Recursively decluste

Soft Drop Condition: $$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\rm cut} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$ soft drop condition is



J. Dolen

- Soft wide angle radiation fails the condition
  - As z_{cut} ⇑ ⇒ more aggressive grooming
  - As β ⇓ ⇒ more aggressive grooming
- Example (z_{cut} = 0.1) :
  - If β = 0, remove softer subjet if p_T fraction < 0.1 (~equivalent to MMDT)
  - If β > 0, remove softer subjet if p_T fraction < x, where x increases with ΔR and has maximum value 0.1
  - β → ∞ no grooming
  - β <0 soft drop becomes a tagger instead of a groomer (finds jets with hard, large angle subjets)