Jet Substructure & N-subjettiness

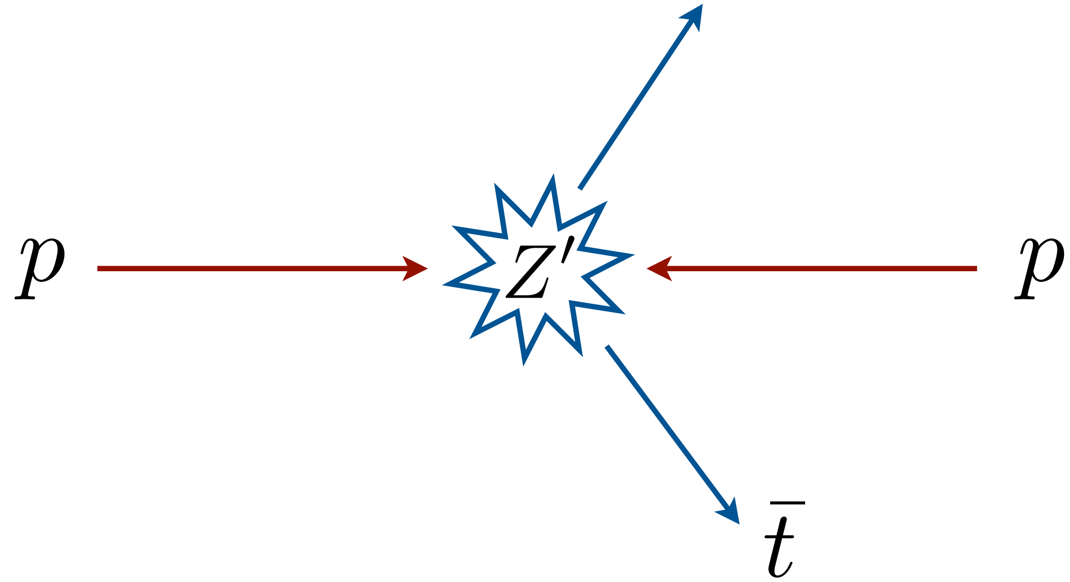
Jesse Thaler

Thanks to Sal Rappoccio and Chris Lee for slides!
David Miller for up-to-the-minute results!

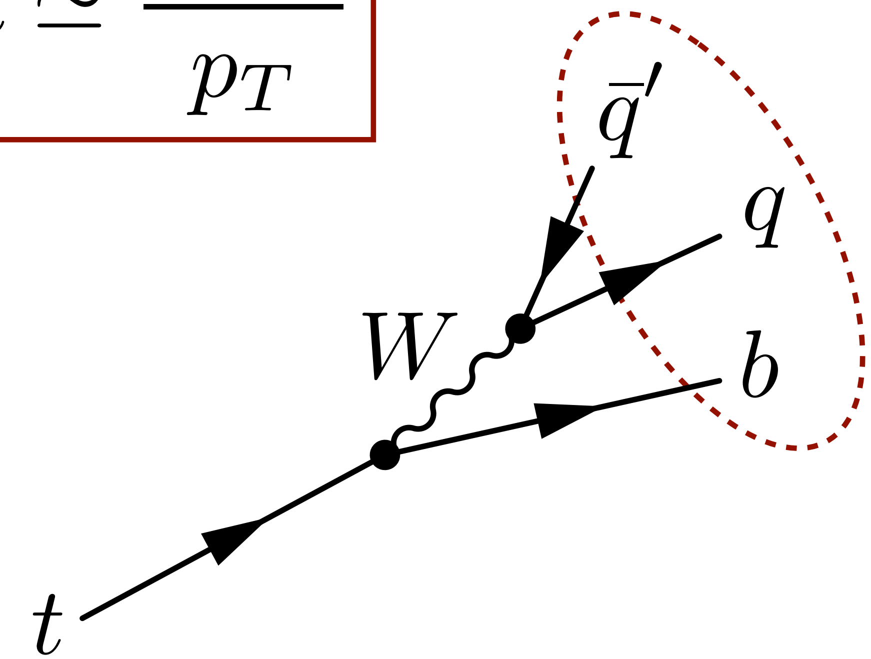
Based on work with Ken Van Tilburg: 1011.2268 & 1108.2701
Preliminary work with Ilya Fiege, Matthew Schwartz, and Iain Stewart

Extreme Kinematics at the LHC

$$\sqrt{\hat{s}} \gg m_{\text{top}}$$



$$\Delta R \simeq \frac{m_{\text{top}}}{p_T}$$

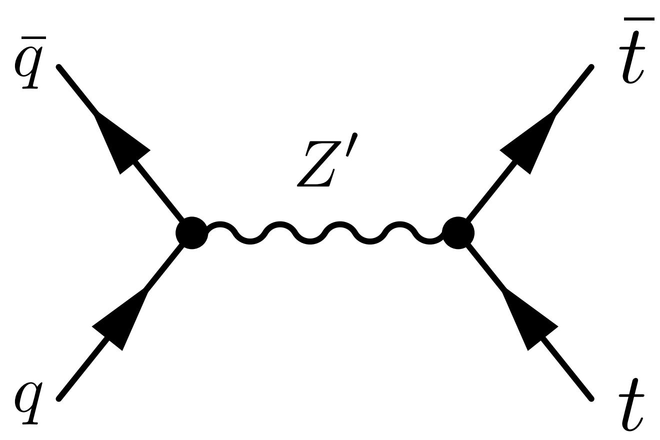


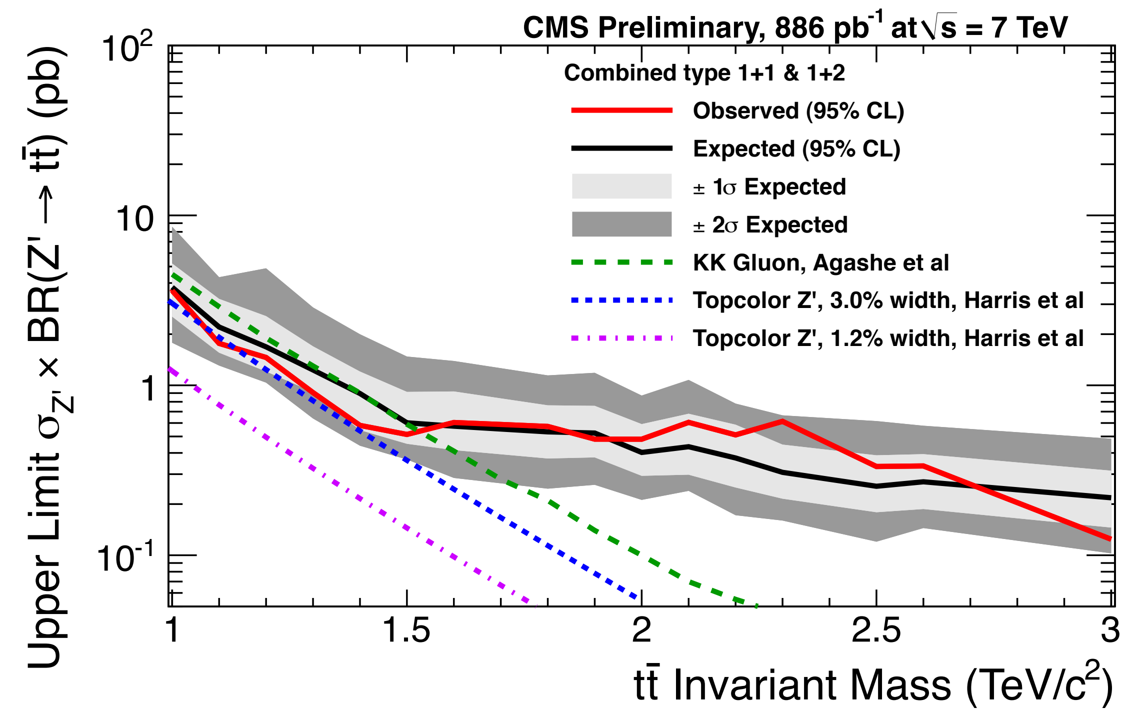
Heavy resonance to boosted tops...
...looks just like QCD dijets.

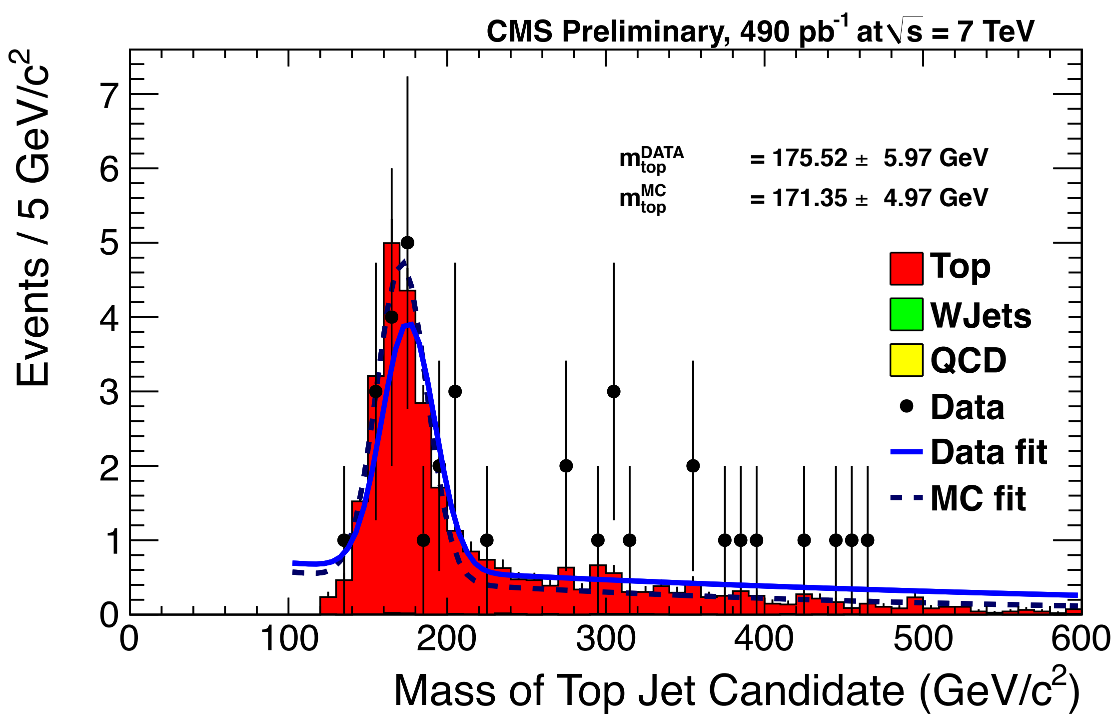
Must reduce QCD by 10^{-3} in boosted regime!
(rejection factor of few x 10^{-2} per jet)

Experimentally... it works!

CMS-PAS-EXO-11-006



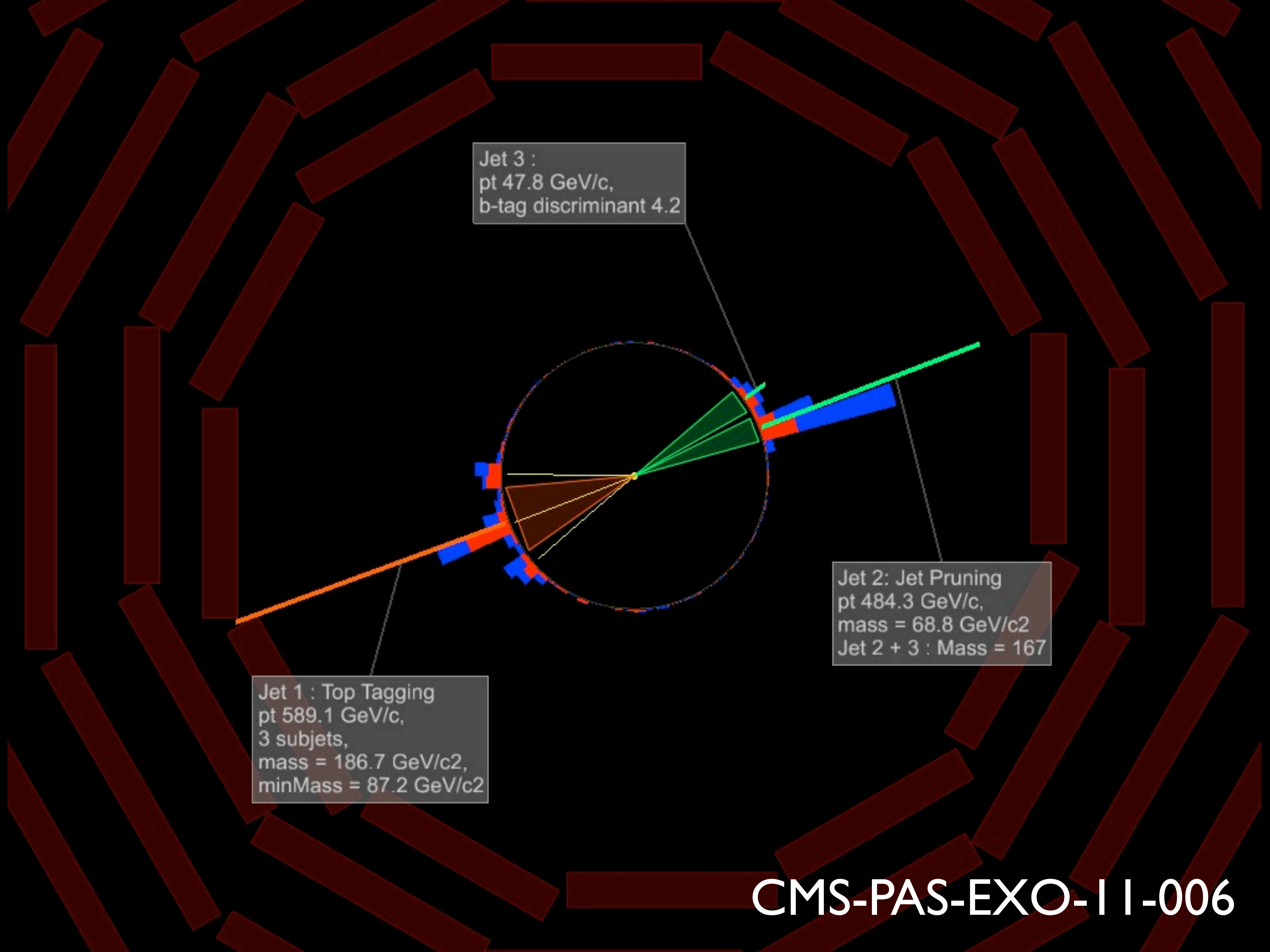




Jet Merging in Boosted Regime
→ “Fat” Jet Identification

Type 1: Top tag
Type 2: W tag + b tag

Competitive with “resolved” searches

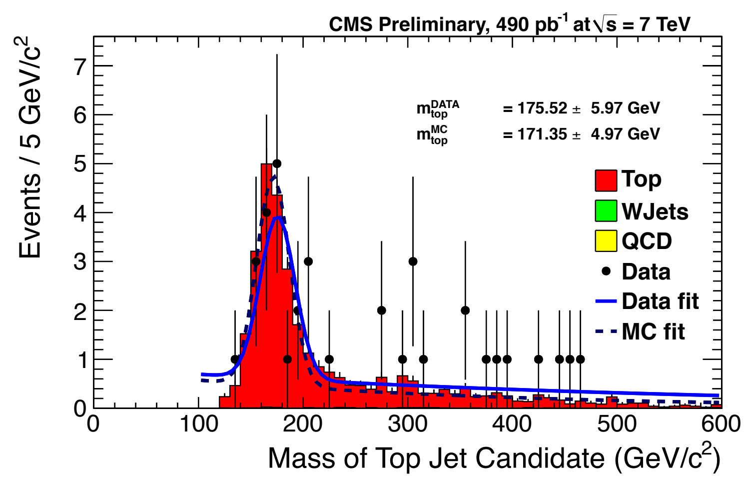


CMS-PAS-EXO-11-006

Jet Substructure & N-subjettiness

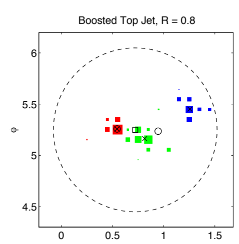
Overview of Jet Substructure

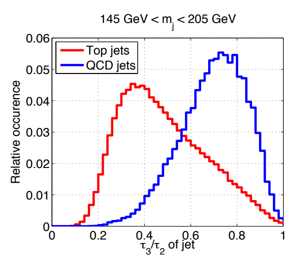
Boosted objects at the LHC



Introducing N-subjettiness

Top tagging with τ_3/τ_2, W/Z/H tagging with τ_2/τ_1



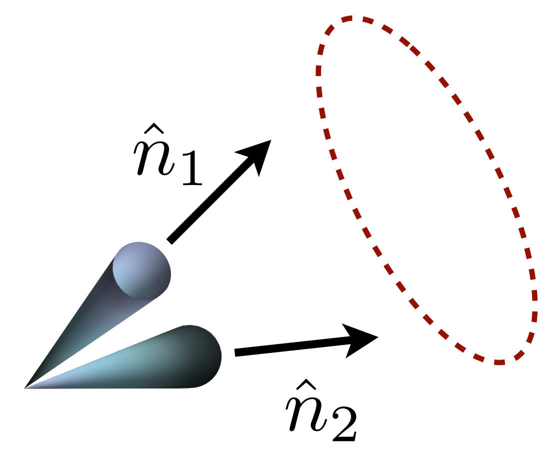


$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min_A \{\Delta R_{A,k}\}$$

[JDT, Van Tilburg: 1011.2268 & 1108.2701]

Ongoing Theoretical Studies

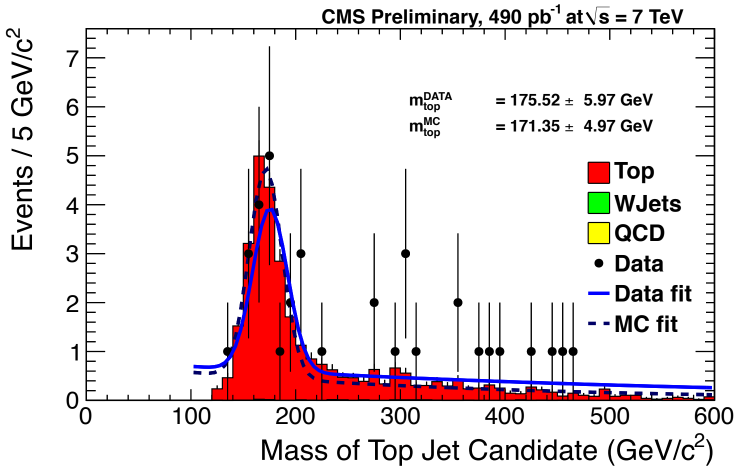
Expanding about the infinite boost limit



[Fiege, Schwartz, Stewart, JDT: 12xx.xxxx]

Overview of Jet Substructure

Boosted objects at the LHC

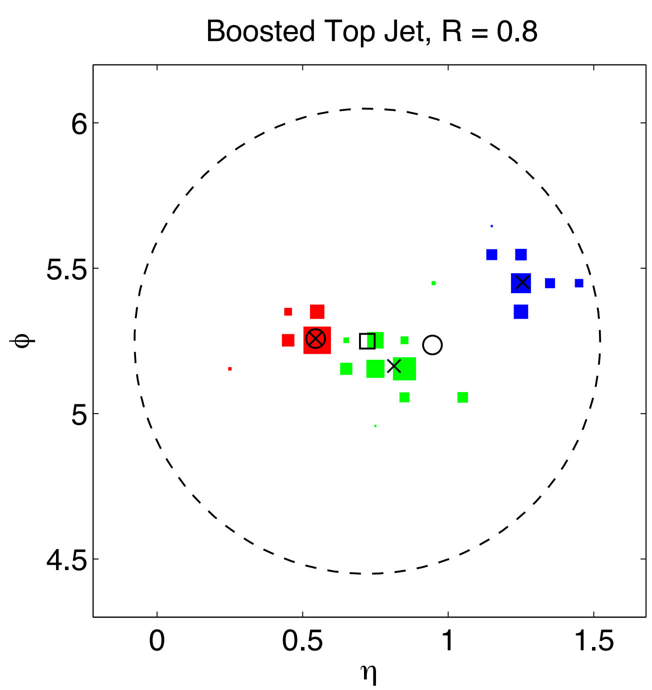


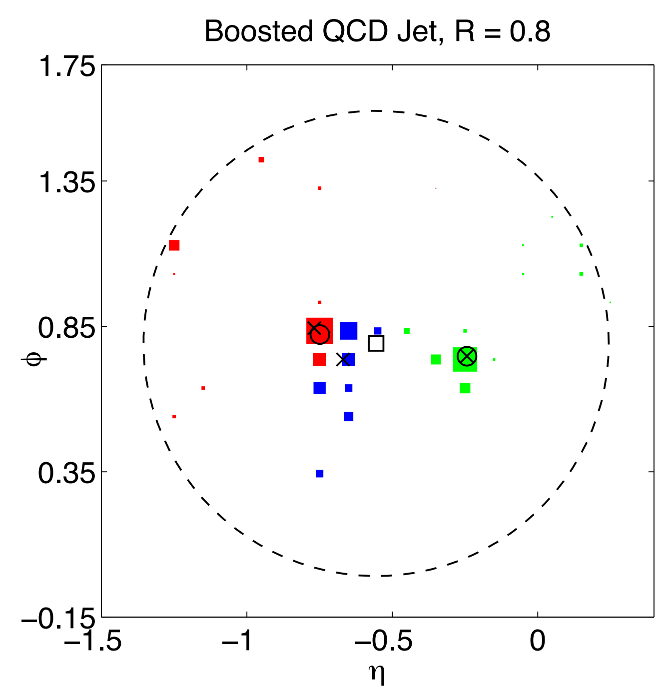
Thanks to Sal Rappoccio and Chris Lee for slides!

(I am assuming you know how to reconstruct jets, R = eta-phi distance, etc.)

Jet Substructure by Eye

Two jets with $m_{\rm jet} = m_{\rm top}$. Coloring by exclusive k_T.





Basic Tagging Strategies:

Algorithmic: $\{p_i\} \to p_{\rm jet}, \text{yes/no}$

Currently used by CMS: JHTT, Pruning

Jet Shape Cut: $f(\{p_i\}) < f_{\rm cut}$

Theoretically appealing, typically less effective

Algorithmic: Johns Hopkins Tagger

Inspired by BDRS Boosted Higgs Method

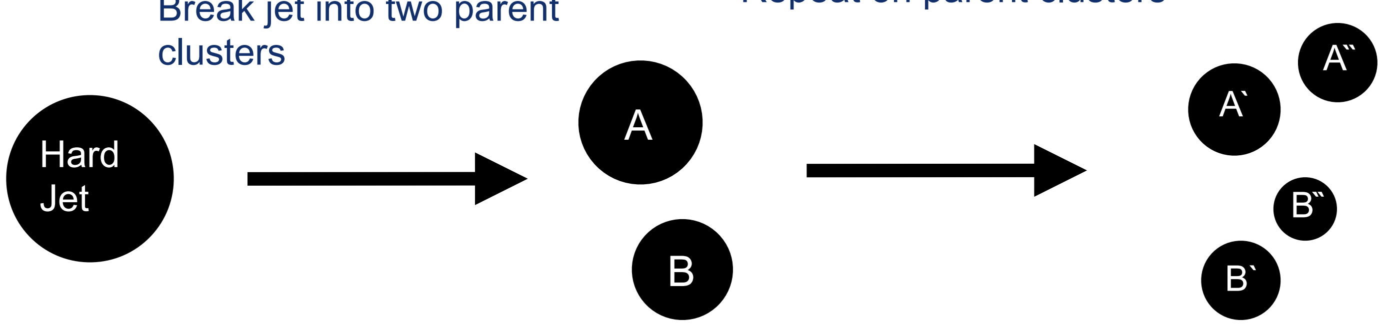
[Kaplan, Rehermann, Schwartz, Tweedie]

Top Tagging Details

- Based on Kaplan et al. (arXiv:0806.0848)
- Cluster particle flow candidates using Cambridge Aachen
- Reverse the clustering sequence in order to find substructure
- Subjets must satisfy two requirements
  - Momentum fraction criterion: pTsubjet > 0.05×pThard jet ← Removes soft clusters
  - Adjacency criterion: ΔR(A, B) > 0.4 - 0.0004×pT ← Removes wide angle clusters
- Iterative process - throw out objects that fail momentum fraction cut and try to decluster again
- Then use :
  - Jet Mass ~ Top mass
  - Minimum mass pairing of subjets ~ W mass

Primary Decomposition:
Break jet into two parent clusters

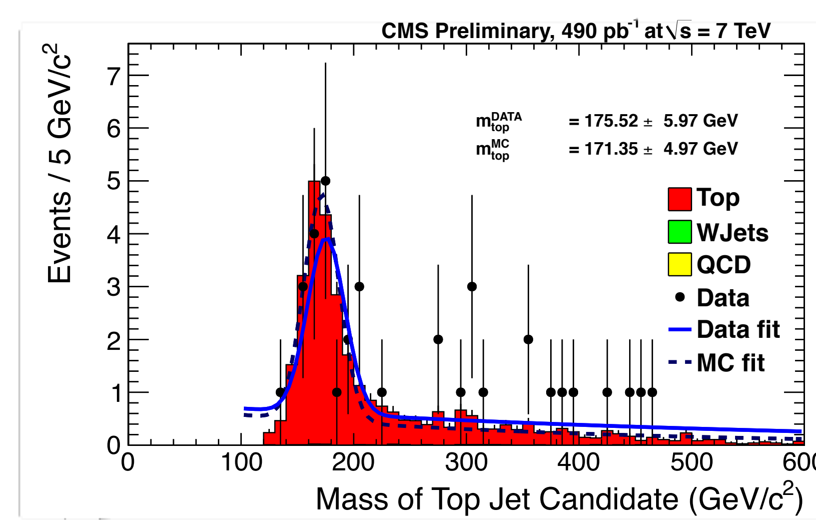
Secondary Decomposition
Repeat on parent clusters



Require >= 3

26 July 2011

Monday, July 25, 2011



Sal Rappoccio — Panic 2011

Algorithmic: Jet Grooming

See also Jet Filtering & Jet Trimming

[Ellis, Vermilion, Walsh]

Jet Pruning Details

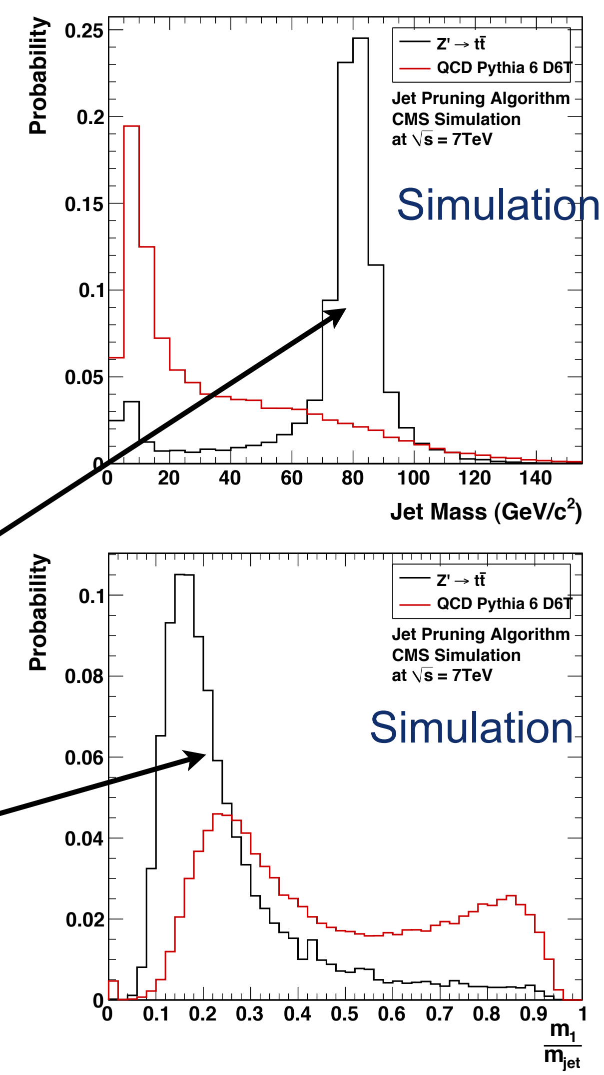
- Ellis et al. (arXiv:0903.5081)
- Improves mass resolution by removing soft, large angle particles from the jet
- Recluster each jet, requiring that each recombination satisfy the following:

$$\frac{\min(p_{T1}, p_{T2})}{p_{Tp}} > 0.1$$

$$\Delta R_{12} < 0.5 \times \frac{m_{\rm jet}}{p_T}$$

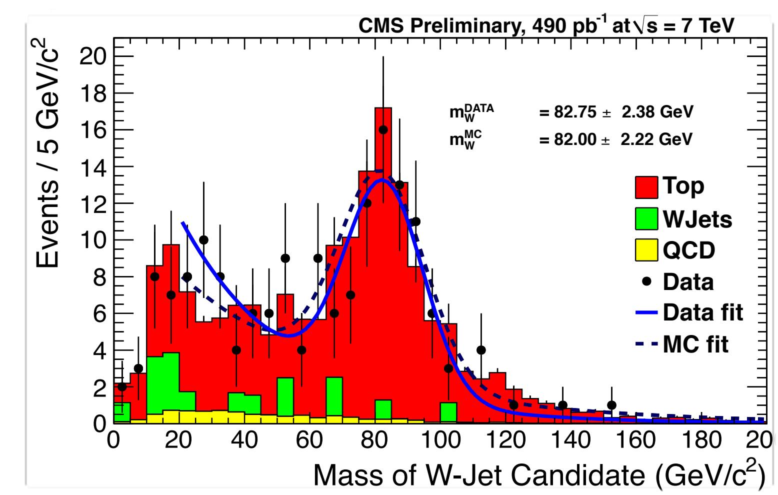
- For W tagging, require:
  - Jet mass in 60-100 GeV/c2
  - Mass drop (mu) < 0.4

$$\mu = \frac{m_{j1}}{m_j}$$

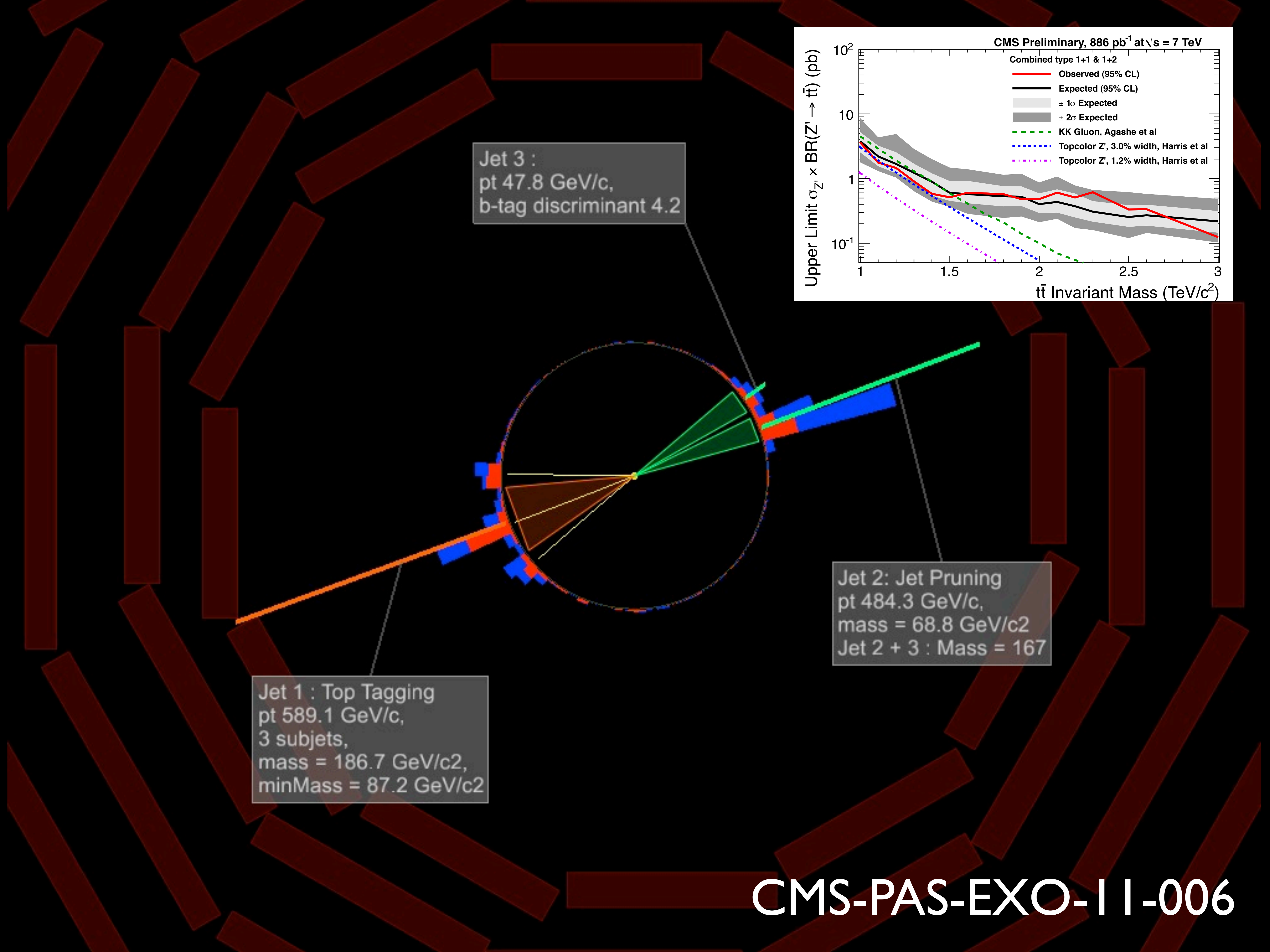


26 July 2011

Monday, July 25, 2011



Sal Rappoccio — Panic 2011



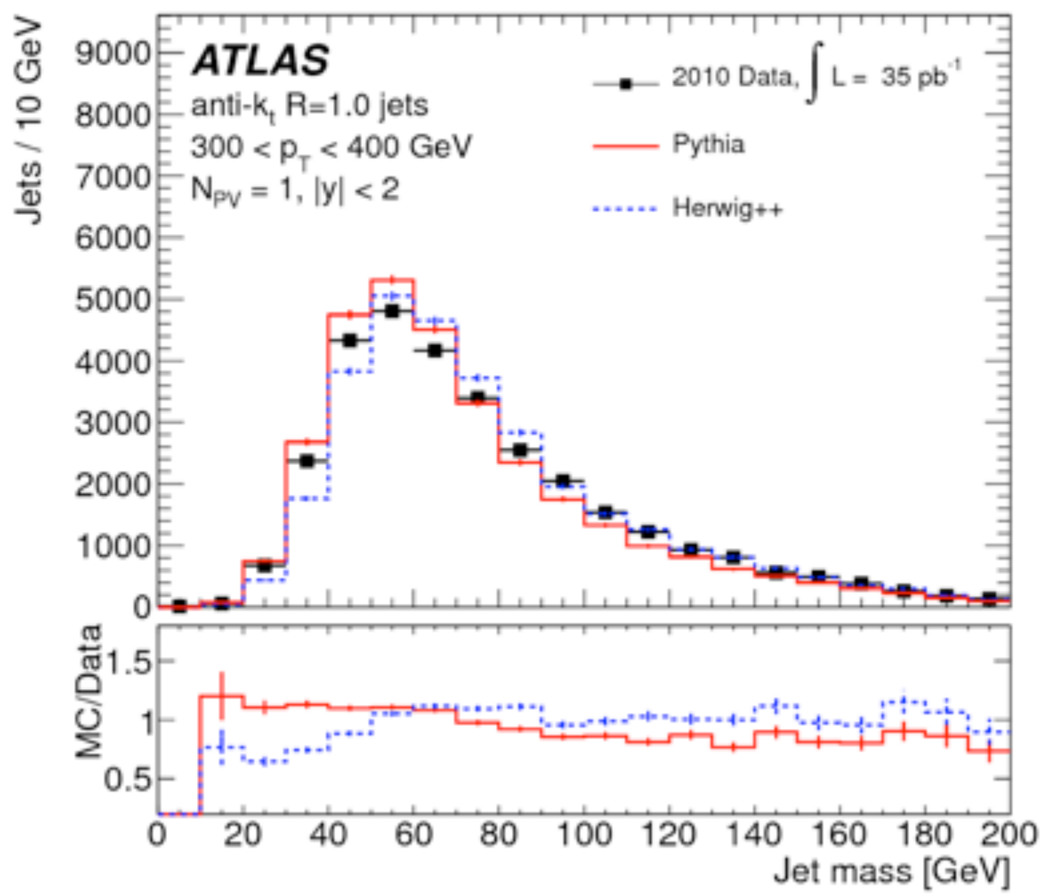
CMS-PAS-EXO-11-006

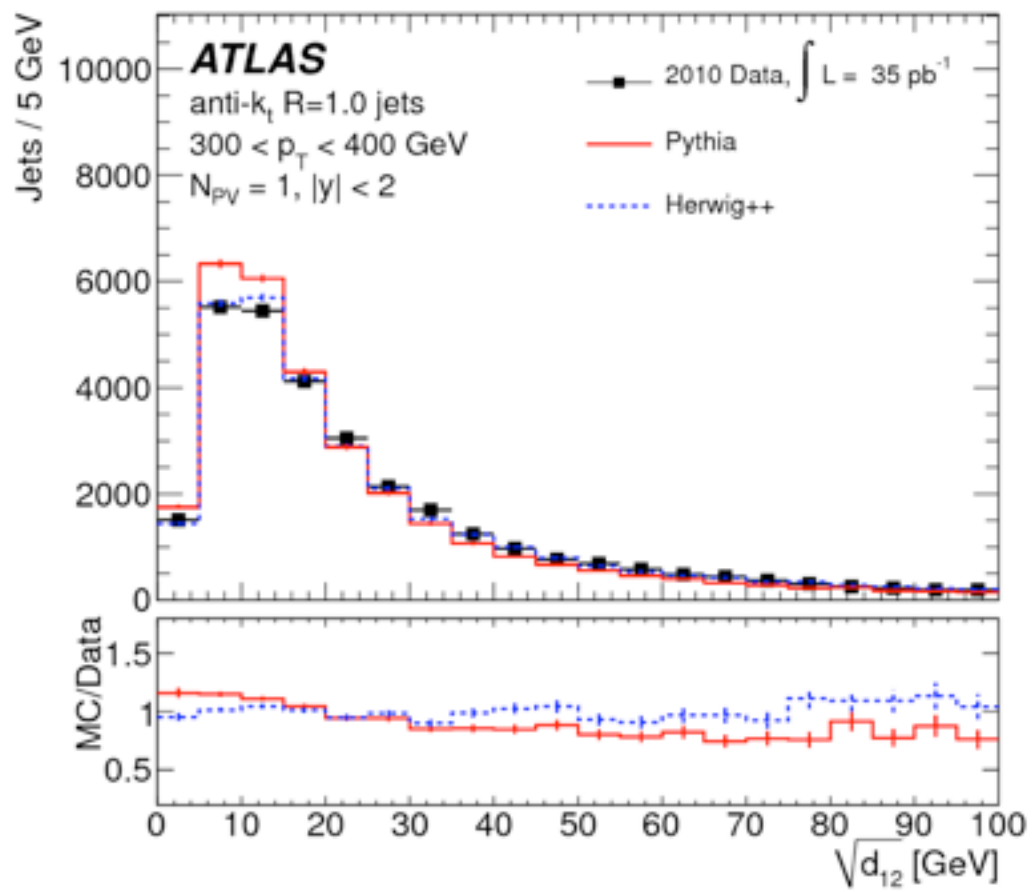
Other Notable Measurements

ATL-STDM-2011-19

Jet Mass

$$\sqrt{d_{12}} = \min(p_{Ta}, p_{Tb}) \times \delta R_{a,b},$$





QCD-like ⟷ Signal-like

QCD-like ⟷ Signal-like

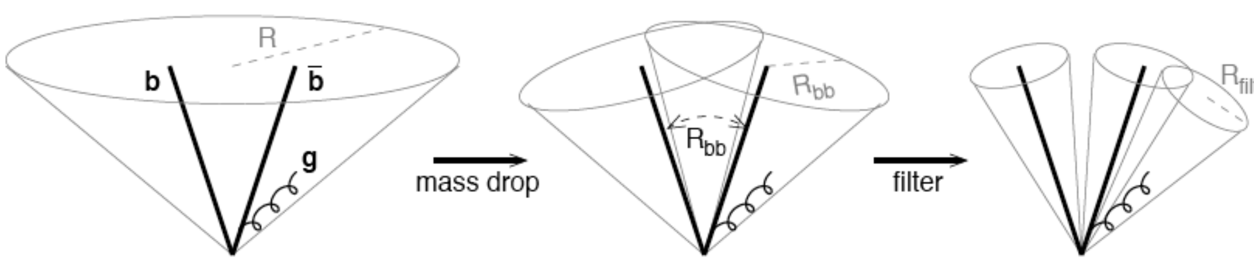
ATLAS Measurement of Jet Substructure in Inclusive QCD Jets

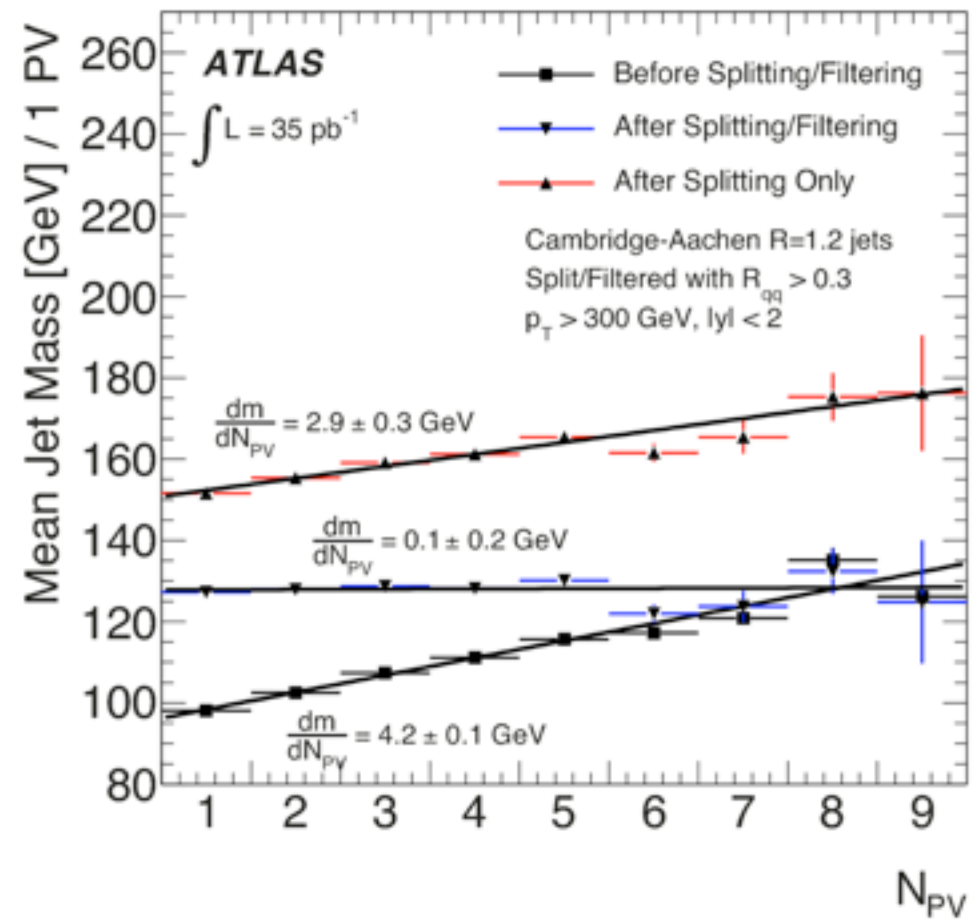
Other Notable Measurements

ATL-STDM-2011-19

BDRS Method

[Butterworth, Davison, Rubin, Salam]





ATLAS Test of Grooming Procedures and Pileup

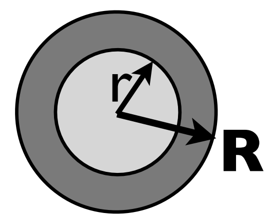
Jet Shapes: Angularities

Close Relatives of N-subjettiness

Calculable Measures: Jet Shapes

- Additional discriminating power by measuring more shapes:

"The" Jet Shape: $\Psi(r/R)$ = energy fraction inside subcone

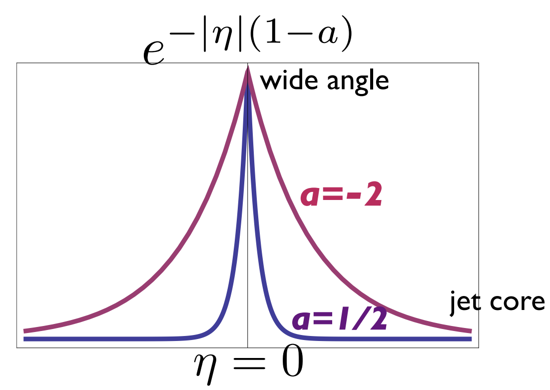


Ellis, Kunzst, Soper

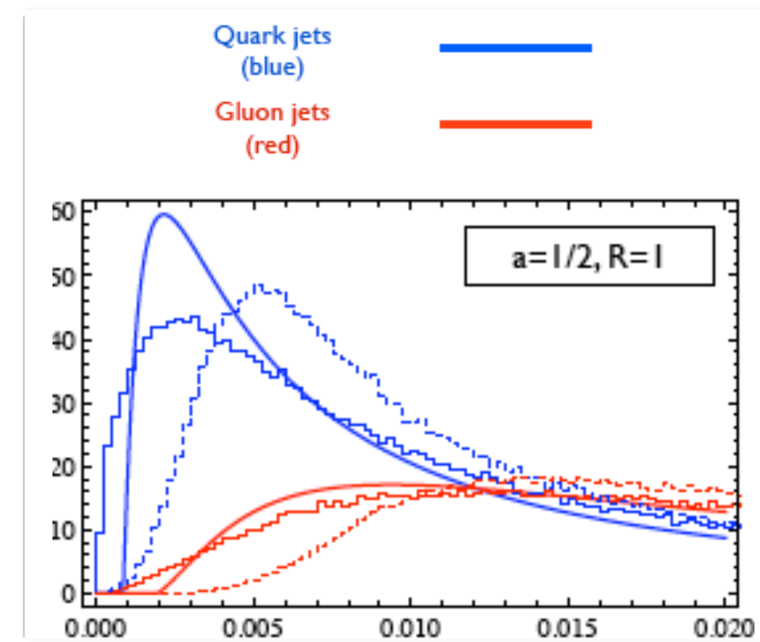
"Angularity" Jet Shapes:

$$\tau_a = \frac{1}{E_J}\sum_{i\in\text{jet}} |\mathbf{p}_T^i| e^{-|\eta_i|(1-a)}$$

Ellis, Hornig, CL, Vermilion, Walsh (2009);
cf. Almeida, S. Lee, Perez, Sterman, Sung, Virzi (2008)
based on event shape
by Berger, Kucs, Sterman (2003)



Quark/Gluon Separation

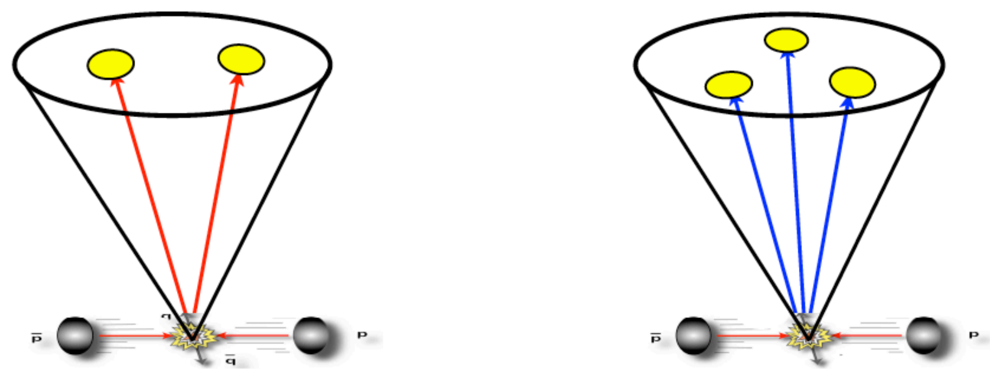


Christopher Lee — Panic 2011

Jet Shapes: Planar Flow

Top Tagging with Jet Shapes

Calculable Measures: Jet Shapes



QCD massive jet   top jet

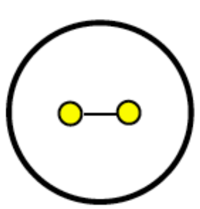
- Planar flow, Pf, measures the energy ratio between two primary axes of cone surface:

(i) "moment of inertia": $$I_E^{kl} = \frac{1}{m_J} \sum_{i \in R} E_i \frac{p_{i,k}}{E_i} \frac{p_{i,l}}{E_i} \, ,$$

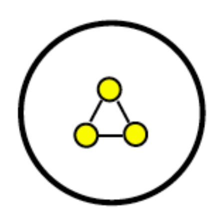
(ii) Planar flow: $$Pf = 4 \frac{\det(\mathrm{I_E})}{\mathrm{tr}(\mathrm{I_E})^2} = \frac{4\lambda_1 \lambda_2}{(\lambda_1 + \lambda_2)^2}$$

Almeida, S. Lee, Perez, Sterman, Sung, Virzi (2008)

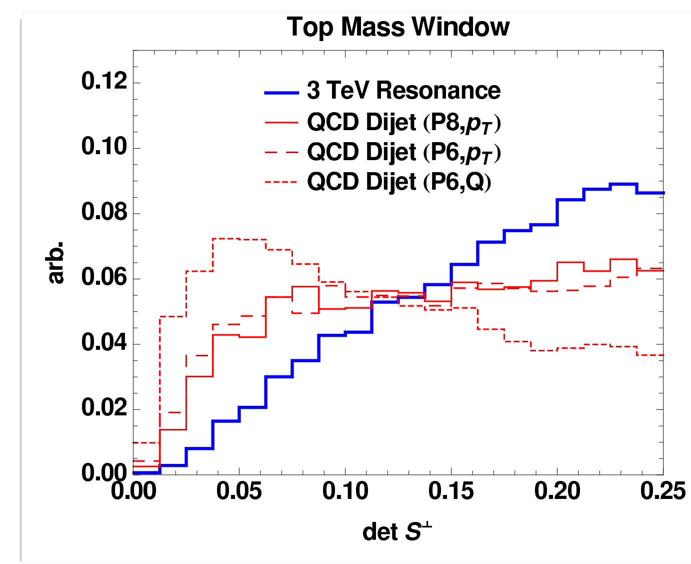
Thaler, Wang (2008)



leading order QCD, Pf=0



top jet, Pf=1



Ok Top/QCD Separation

Christopher Lee — Panic 2011

Additional Observables

Dipolarity

$$\mathcal{D} = \frac{1}{R^2_{j_1 j_2}} \sum_{\alpha \in \mathrm{jet}} \frac{p_{T,\alpha}}{p_{T,\mathrm{jet}}} R^2_\alpha$$

[Hook, Jankowiak, Wacker]

Pull

$$\vec{t} = \sum_{\alpha \in \mathrm{jet}} \frac{p_{T,\alpha}}{p_{T,\mathrm{jet}}} \, |\vec{r}_\alpha| \, \vec{r}_\alpha \, .$$

[Gallicchio, Schwartz]

Substructure without Trees

$$\Delta\mathcal{G}(R) = R \, \frac{\sum_{j_1 \neq j_2} d^{(\mathrm{JADE})}_{j_1 j_2} \, K(R - \Delta R_{j_1 j_2})}{\sum_{j_1 \neq j_2} d^{(\mathrm{JADE})}_{j_1 j_2} \, \Theta(R - \Delta R_{j_1 j_2})}$$

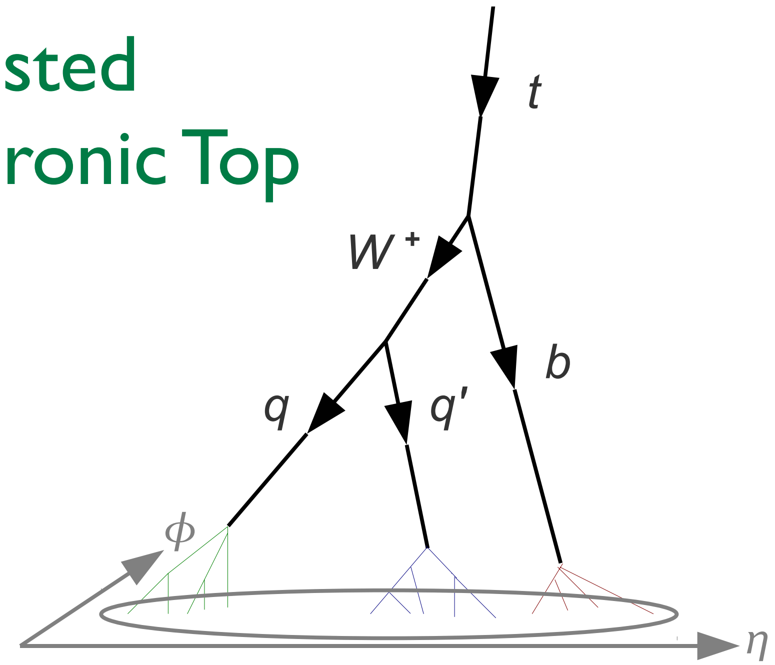
[Jankowiak, Larkoski]

See Seung Lee's talk on Template Overlap Method

Other Algorithmic Methods: HEPTopTagger, YSplitter, Filtripruning

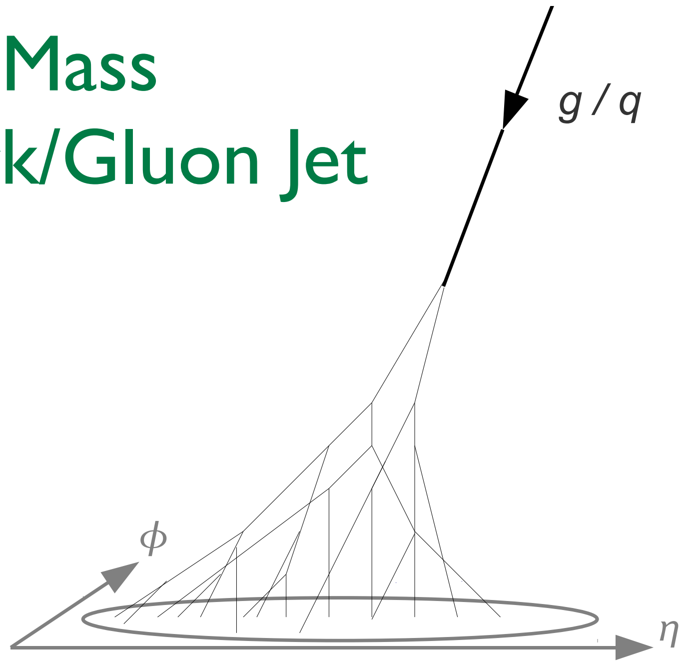
Top Tagging c. 2010

Boosted Hadronic Top

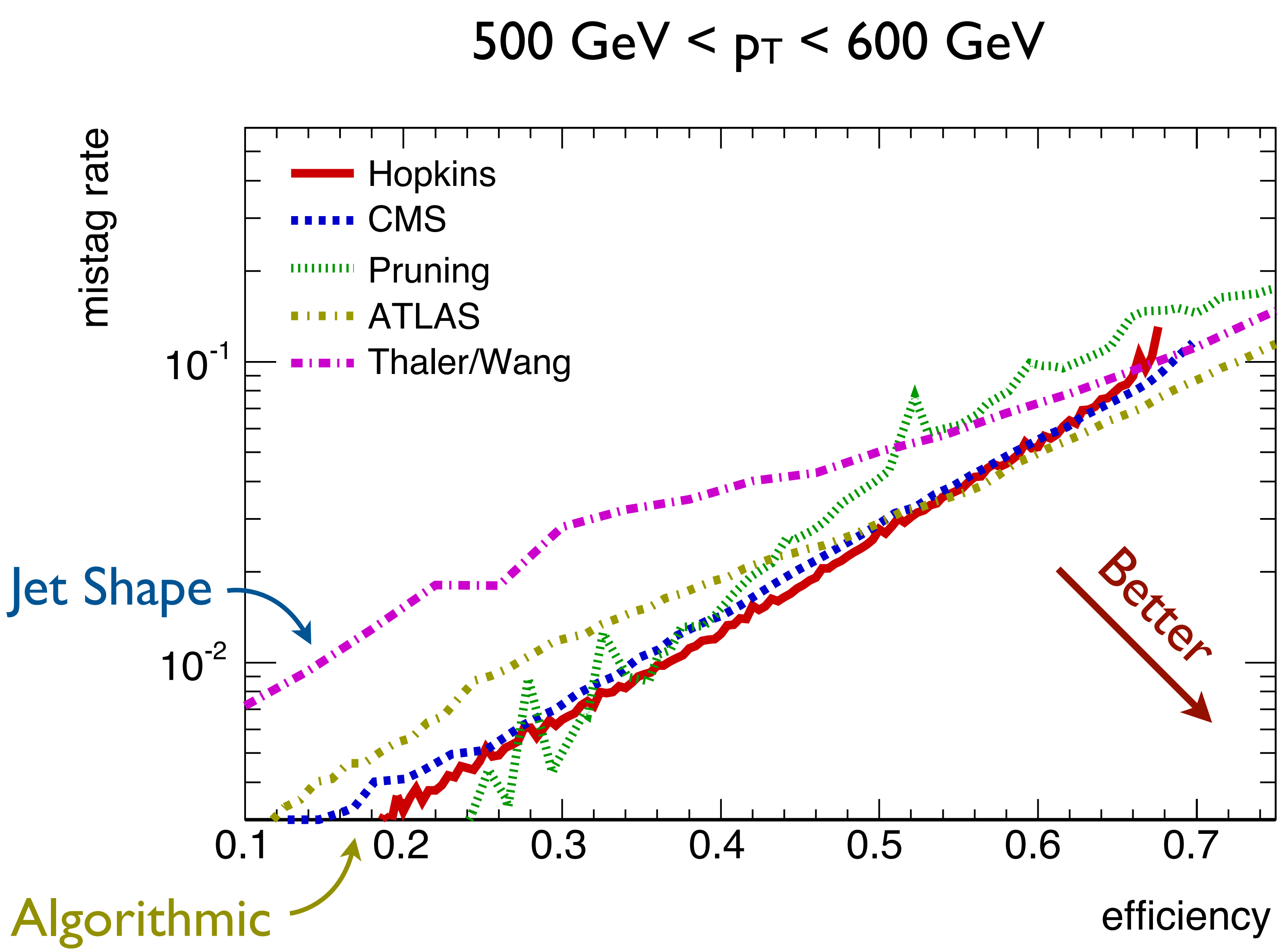


vs.

High Mass Quark/Gluon Jet

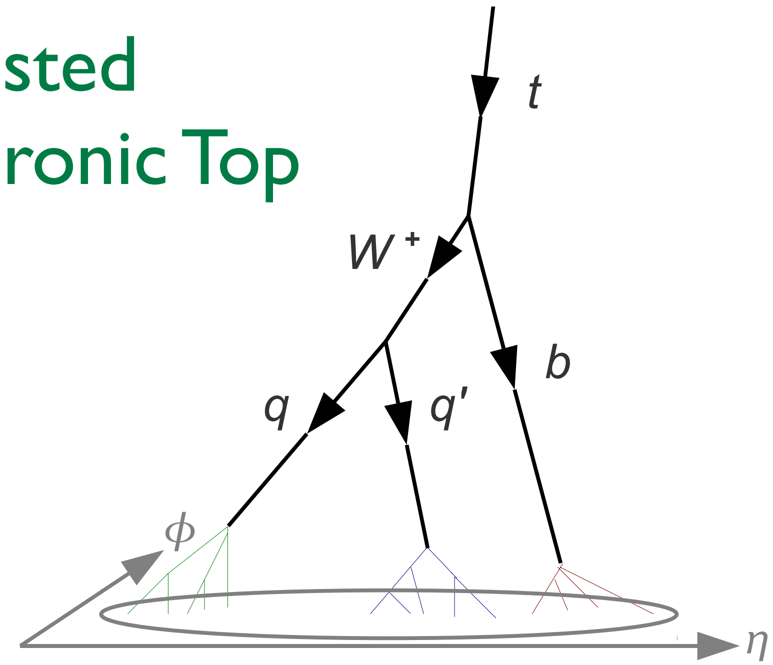


Boost2010 Top Benchmark



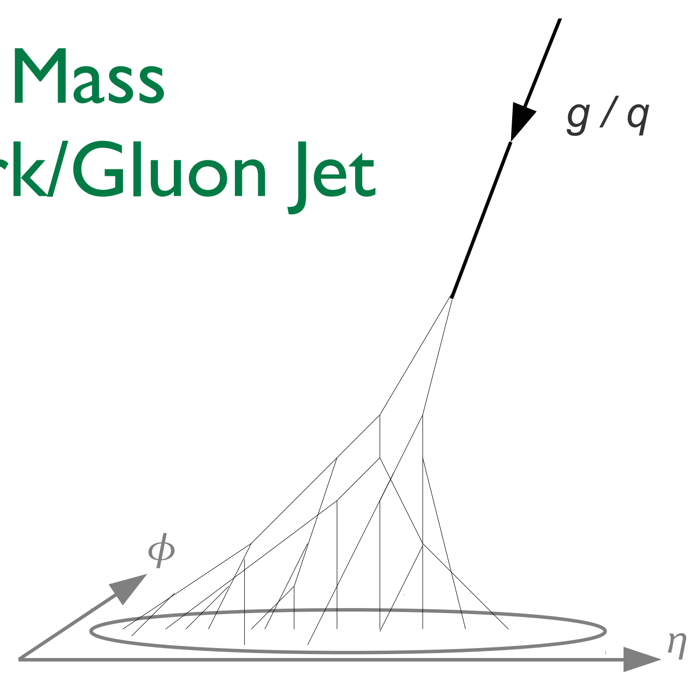
Top Tagging c. 2011

Boosted
Hadronic Top



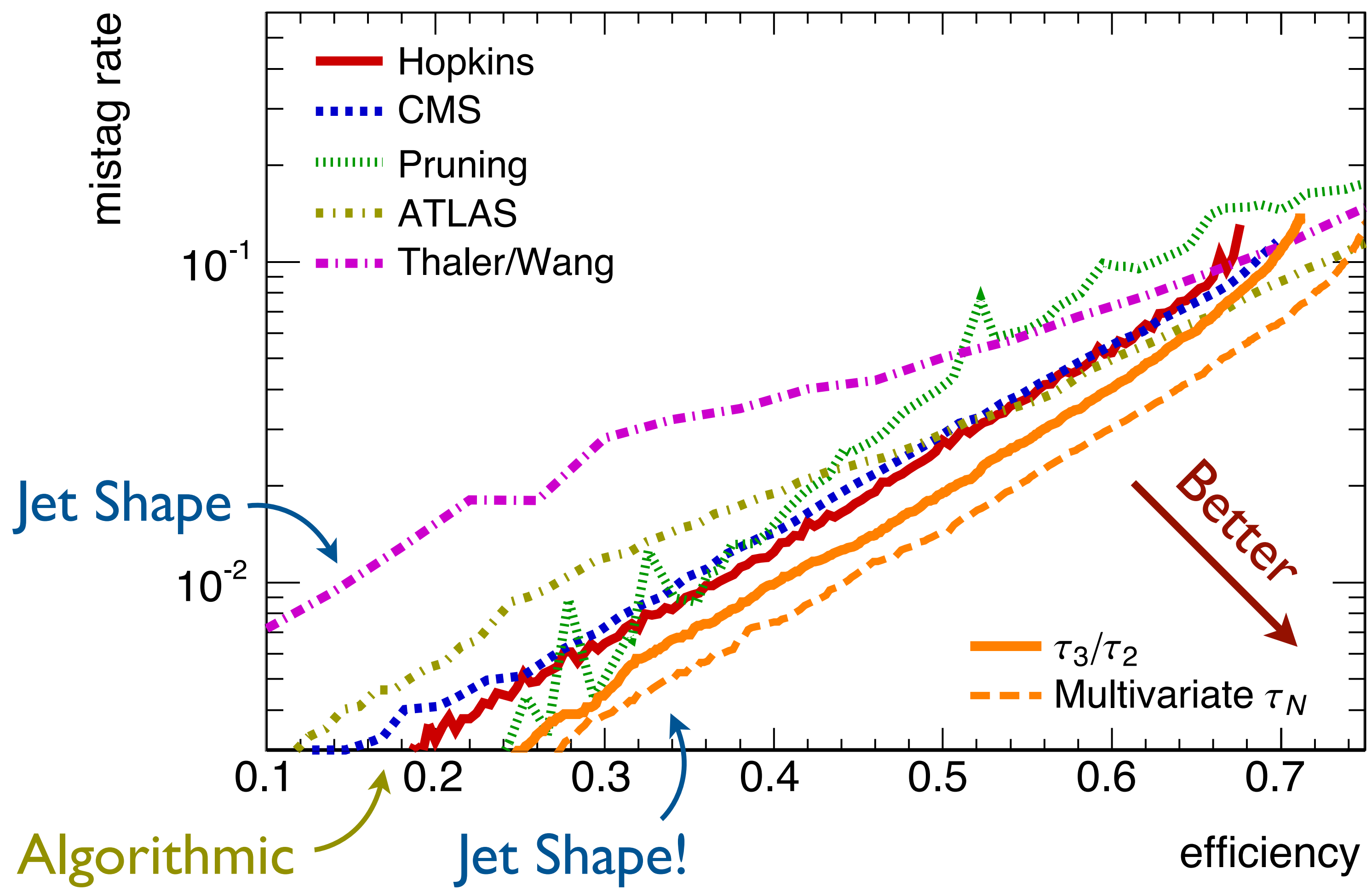
vs.

High Mass
Quark/Gluon Jet



Boost2010 Top Benchmark

500 GeV < p_T < 600 GeV

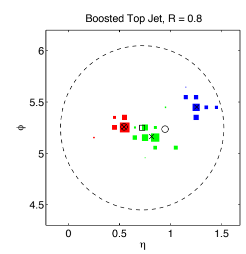


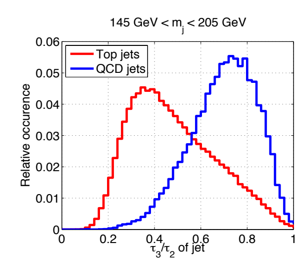
"Thaler/Van Tilburg"

fixed 160 GeV < m_{jet} < 240 GeV cut
one-dimensional cut on τ_3/τ_2

Introducing N-subjettiness

Top tagging with τ_3/τ_2, W/Z/H tagging with τ_2/τ_1



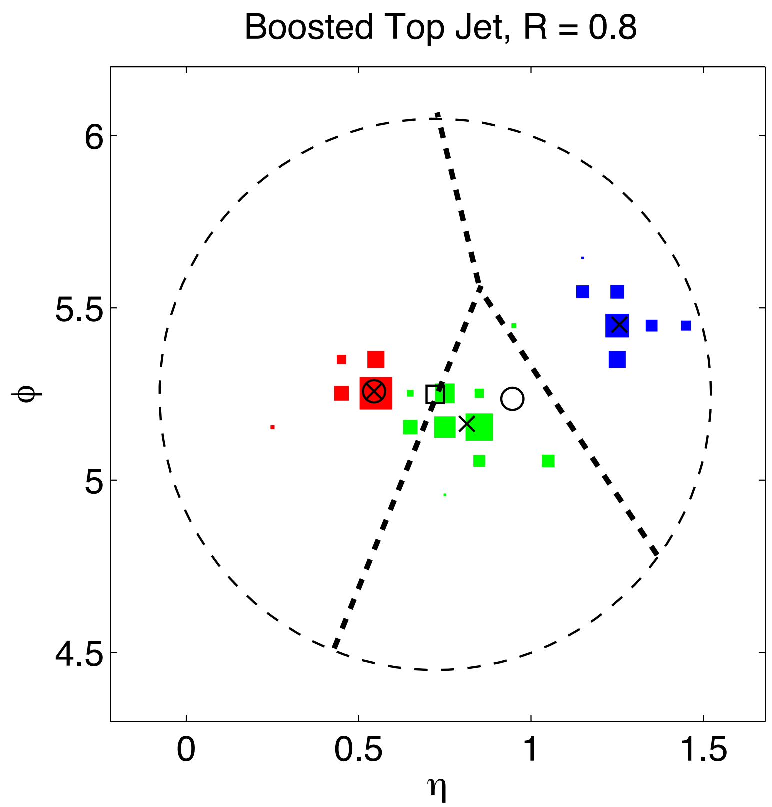


$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min_A \{\Delta R_{A,k}\}$$

[JDT, Van Tilburg: 1011.2268 & 1108.2701]

Introducing N-subjettiness

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{k,1}, \Delta R_{k,2}, \ldots, \Delta R_{k,N}\}$$



Generalization of thrust to multiple (sub)jets!

(strictly speaking, generalization of jet broadening)

Jet shape "counts" number of subjets!

subjets: ≤ N — > N

τ_N: 0 — 1

Adapted from "N-jettiness", used to define exclusive jet bins
[Stewart, Tackmann, Waalewijn: 1004.2489]

Introducing N-subjettiness

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\left\{\Delta R_{k,1}, \Delta R_{k,2}, \ldots, \Delta R_{k,N}\right\}^{\beta}$$

Choice of subjet axes?

Axes that minimize τ_N!

$$\tau_N = \min_{\{p_{\text{axes}}\}} \tilde{\tau}_N$$

Analogous to thrust
Fast implementation by generalized k-means clustering

Choice of angular weighting?

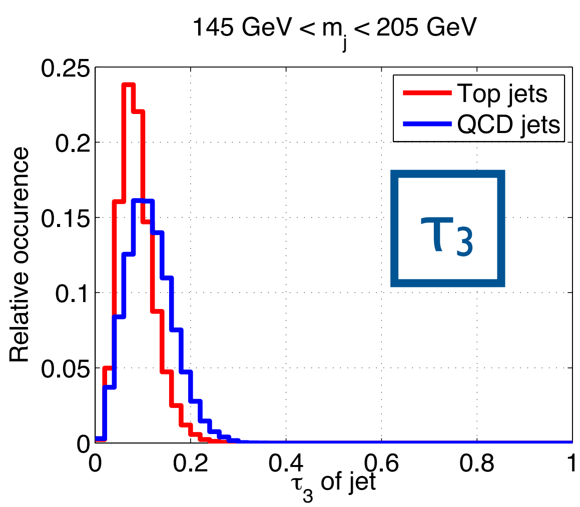
$\beta = 1$: $\approx$ Jet Broadening
$\beta = 2$: $\approx$ Thrust
$\beta = 2\text{-}a$: $\approx$ Angularities

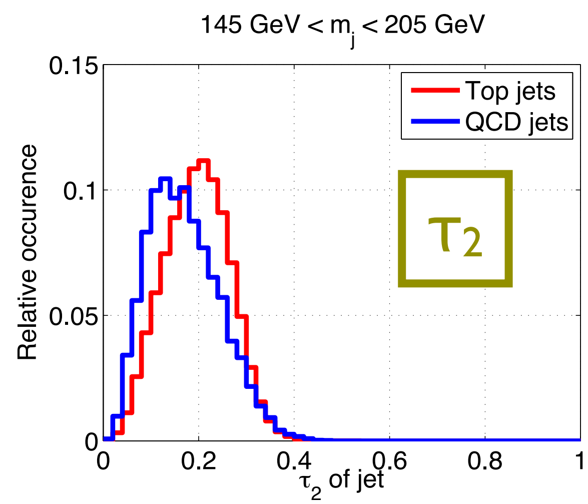
$\beta = 1$ is preferred for boosted object hunting
(Open theoretical question: why?)

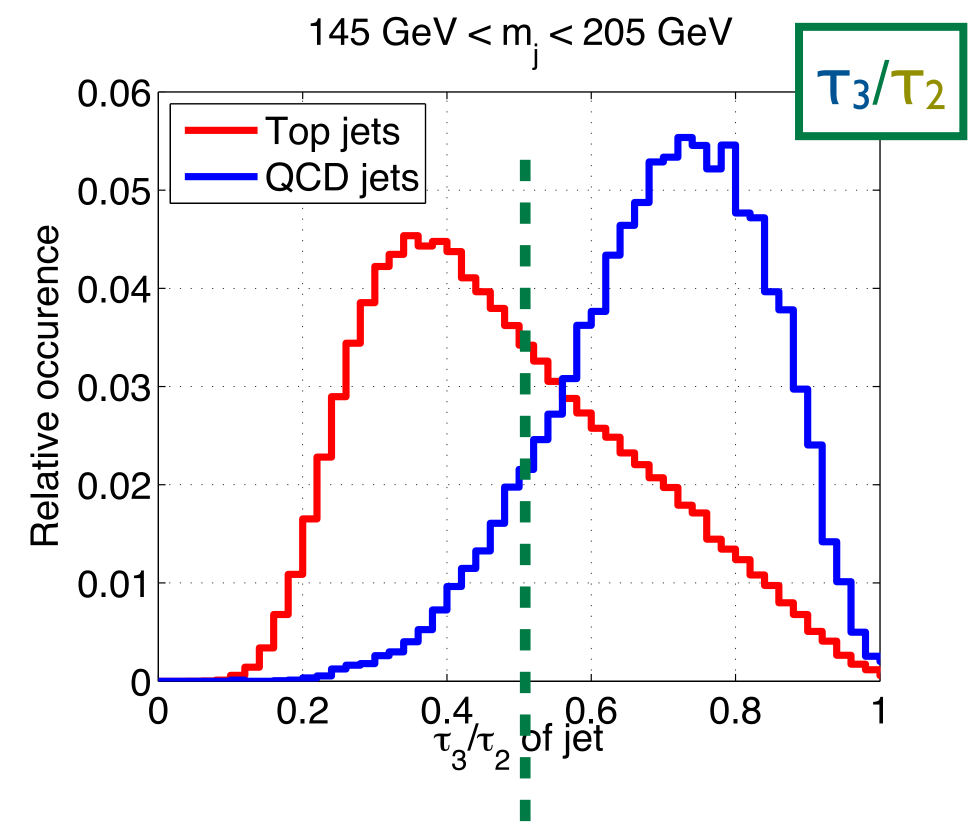
No algorithmic dependence (apart from initial jet finding)

Tagging with τ_N/τ_{N-1}

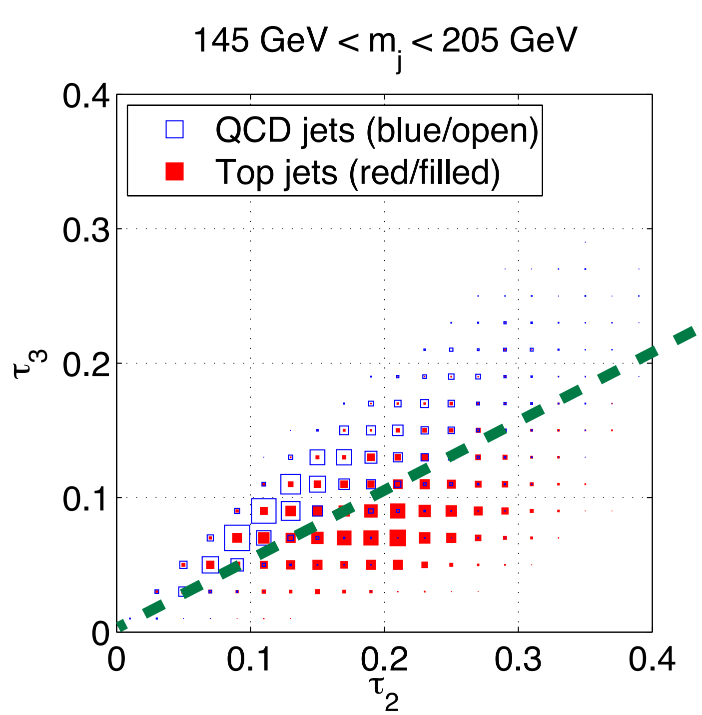
Some raw distinguishing power...







Flexible cut to adjust signal acceptance vs. background rejection



...but suggests multivariate cut!

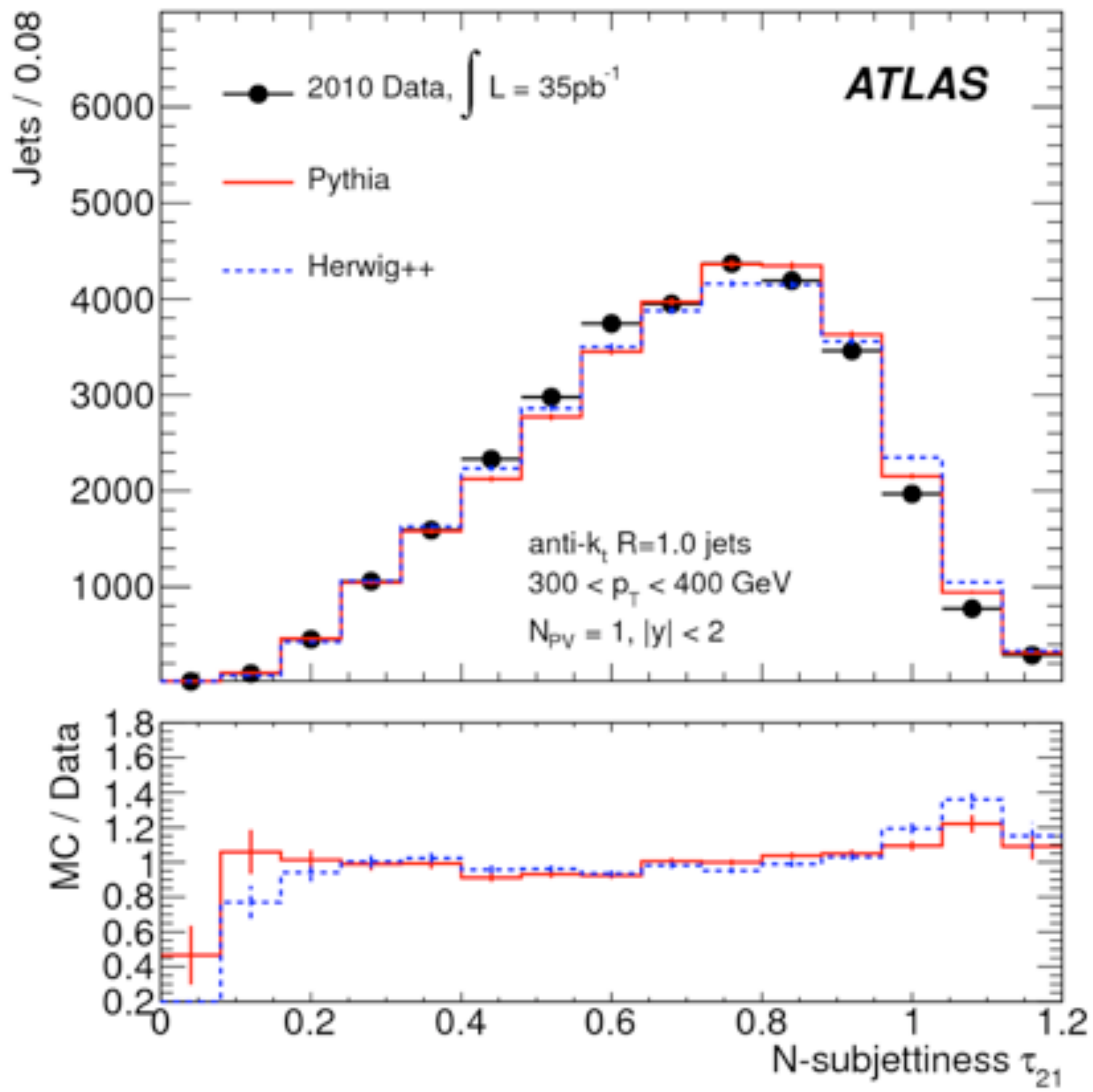
τ_3/τ_2: Boosted Tops

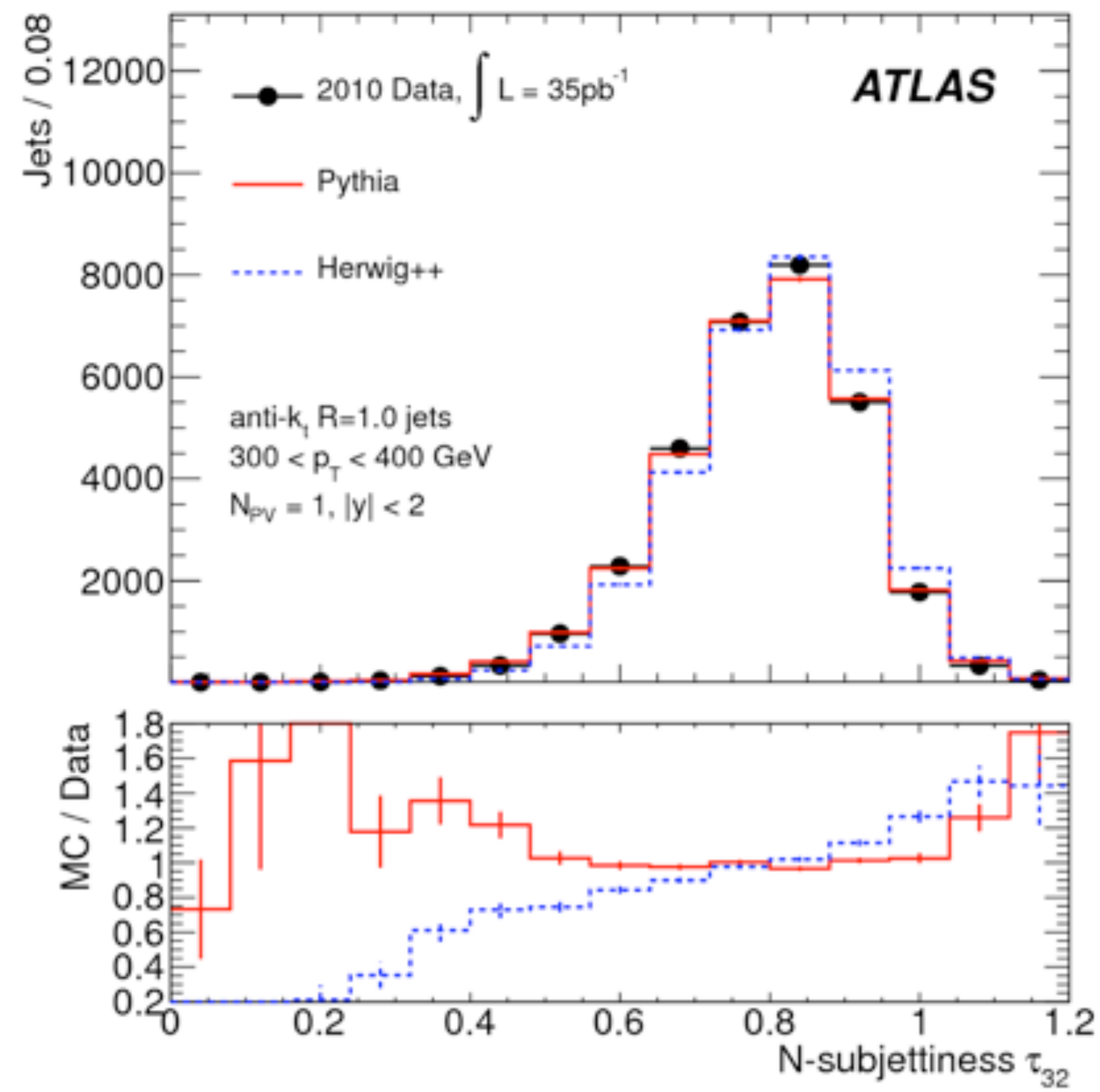
τ_2/τ_1: Boosted W/Z/H

Ratio is quasi-boost invariant

Hot Off the Press from ATLAS

Calibration on Background. Thanks to David Miller!

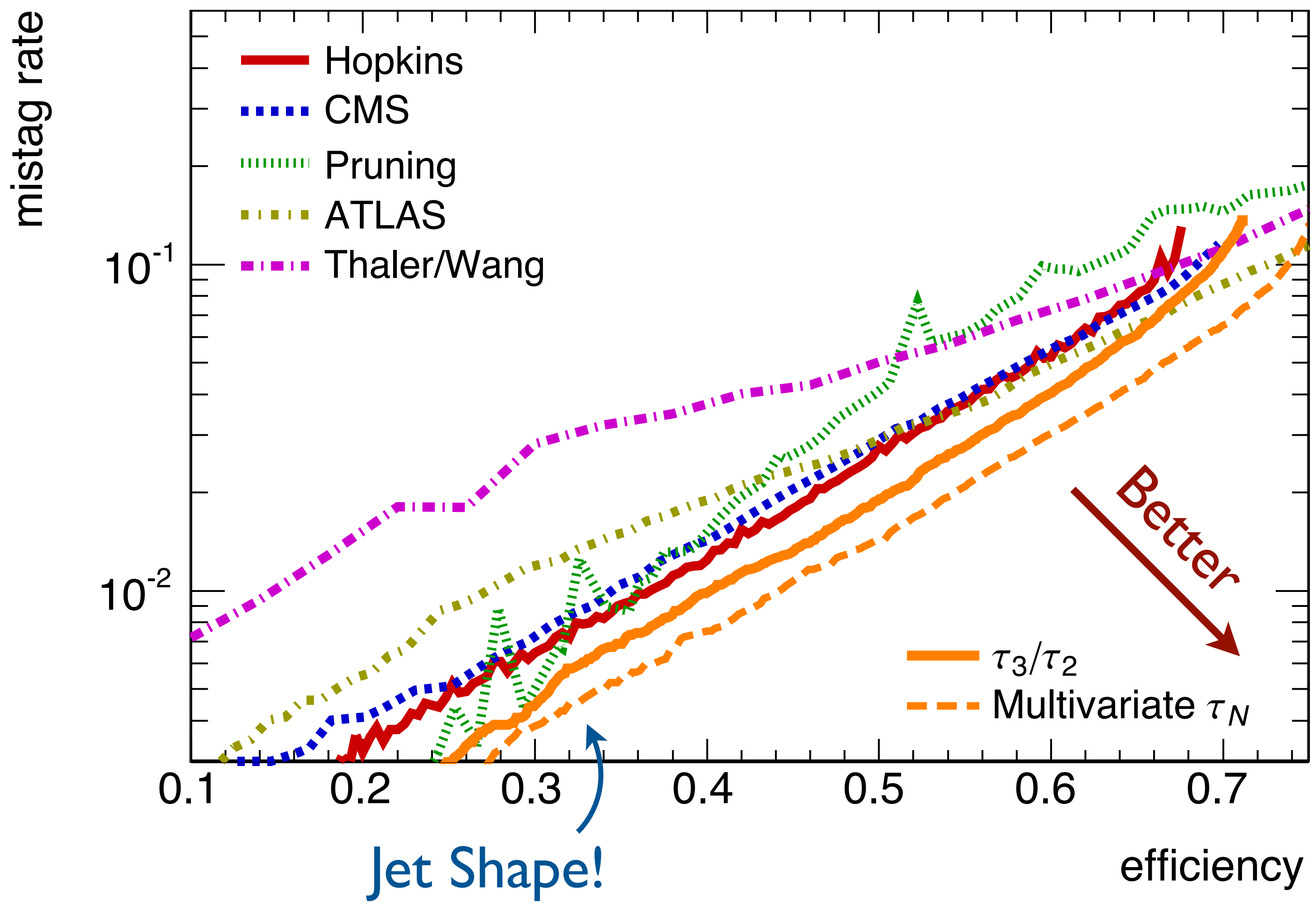




Signal-like ↔ QCD-like

Signal-like ↔ QCD-like

Top Tagging c. 2011



500 GeV < p_T < 600 GeV

fixed jet mass cut:
160 GeV to 240 GeV

one-dimensional cut
on τ_3/τ_2 (from minimum)

dashed = Fisher Discriminant

N-subjettiness:

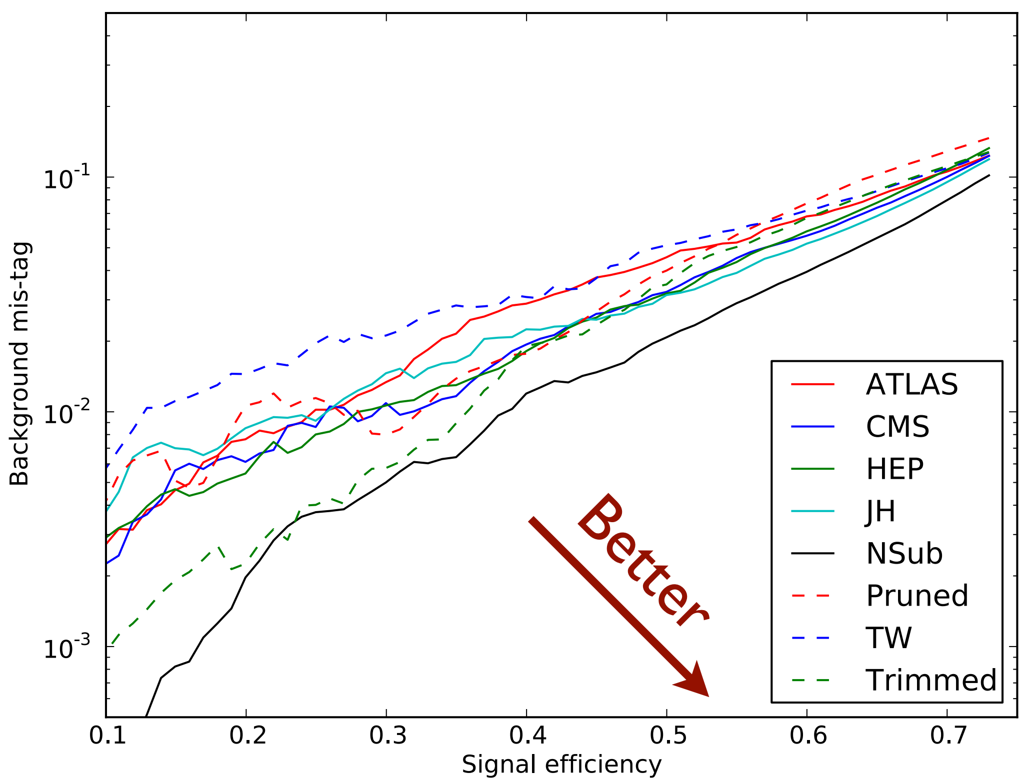
Intuitive measure of "number of subjets"
Based on true jet shape (can use as hybrid, too)
Plays well with multivariate methods
Excellent tagging performance out of the box
Ratio τ_N/τ_{N-1} has reduced JES sensitivity
Plays well with grooming (e.g. pruning)
Calculable (& resummable!) in perturbation theory

Top Tagging c. 2012

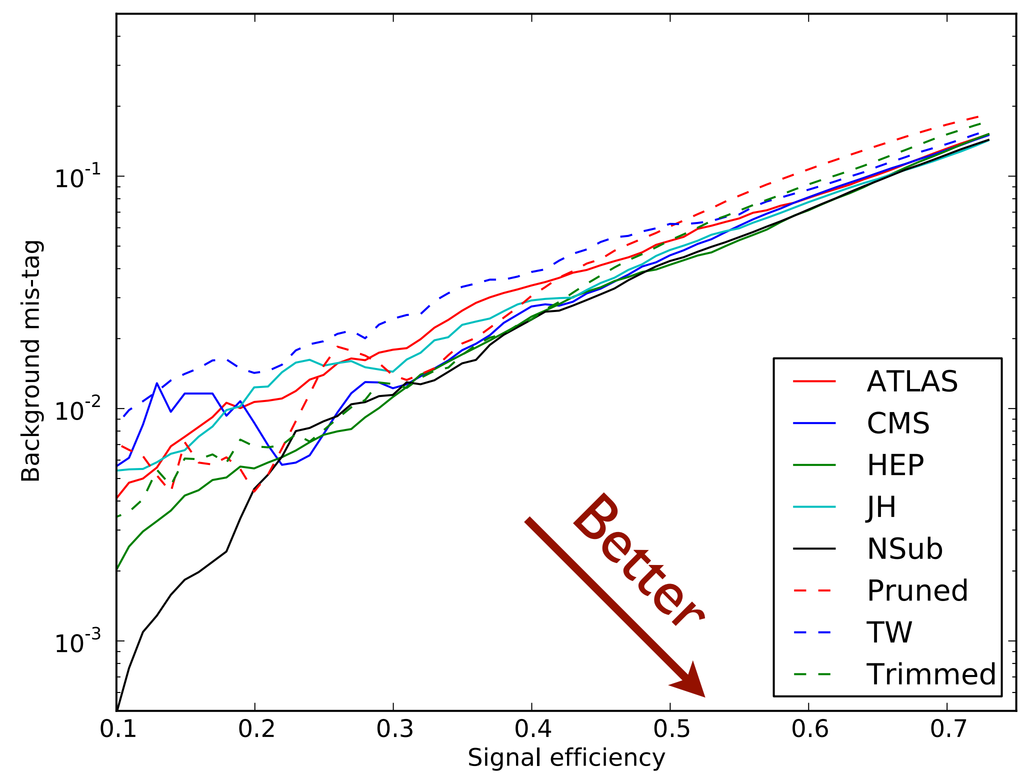
Boost2011 Top Benchmark

500 GeV < p_T < 600 GeV

Herwig 6.5



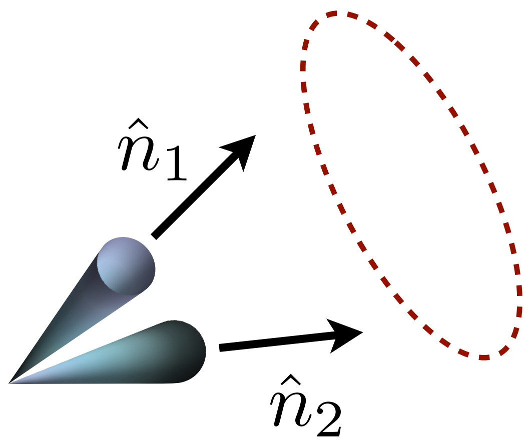
Herwig++



Pythia 6.4 more optimistic, Sherpa more pessimistic,
Detector effects also important.

Ongoing Theoretical Studies

Expanding about the infinite boost limit



[Fiege, Schwartz, Stewart, JDT: 12xx.xxxx]

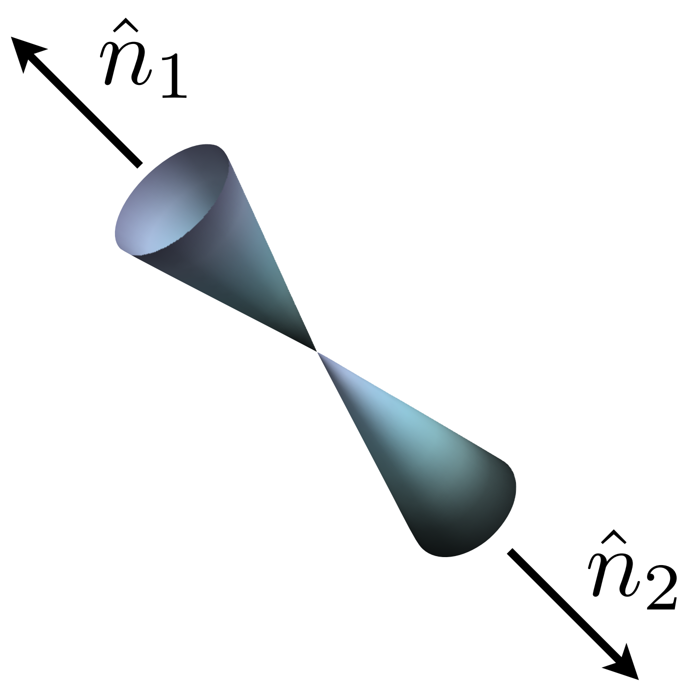
Calculate Idealized 2-subjettiness

In SCET, easier to use "geometric" definition:

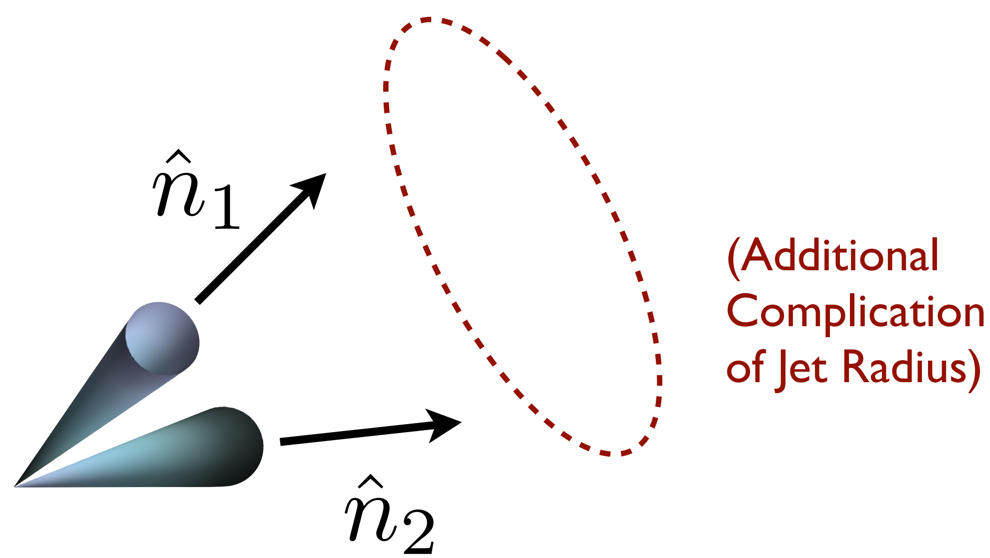
$$\tau_2 = \sum_k \min\{\hat{n}_1 \cdot p_k, \hat{n}_2 \cdot p_k\}$$

Essentially 2-subjettiness with β = 2, using light-like vectors

$$\hat{n}_i \cdot \hat{n}_i = 0 \qquad |\vec{\hat{n}}_i| = 1$$

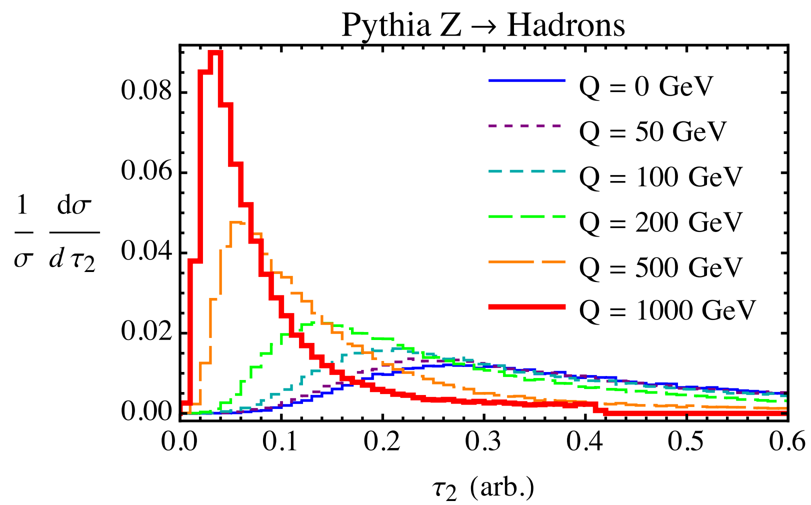


Rest Frame = Classic Thrust

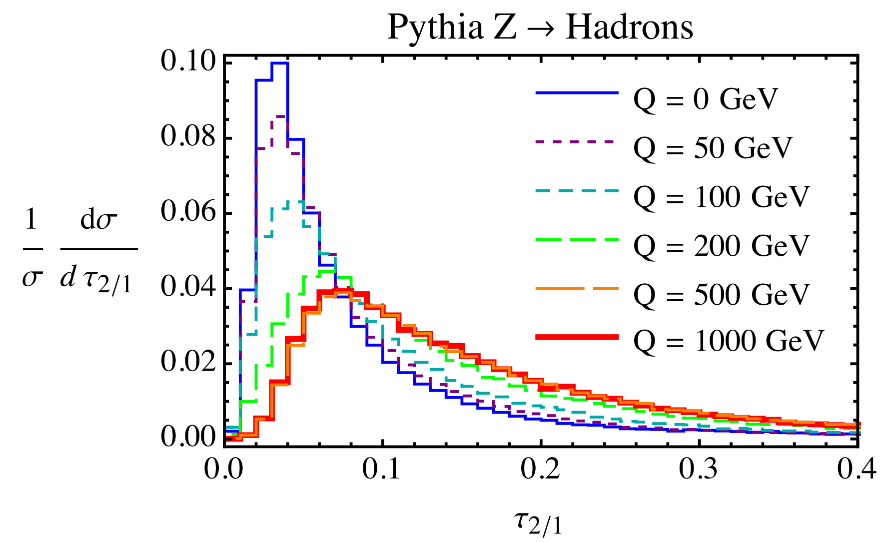


Boosted Frame = 2-subjettiness

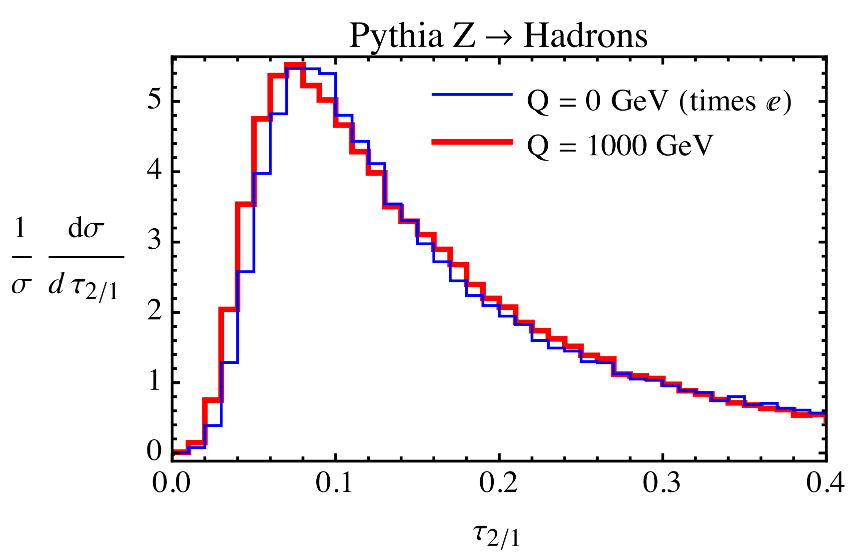
2-subjettiness in Pythia



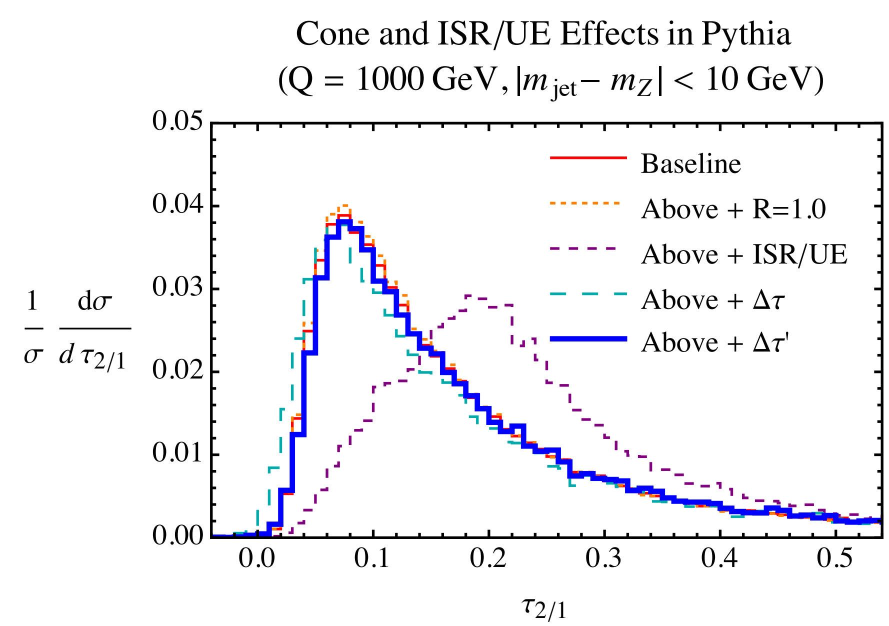
τ_2 at different Q



τ_2/τ_1 saturates as $Q \to \infty$ (!)



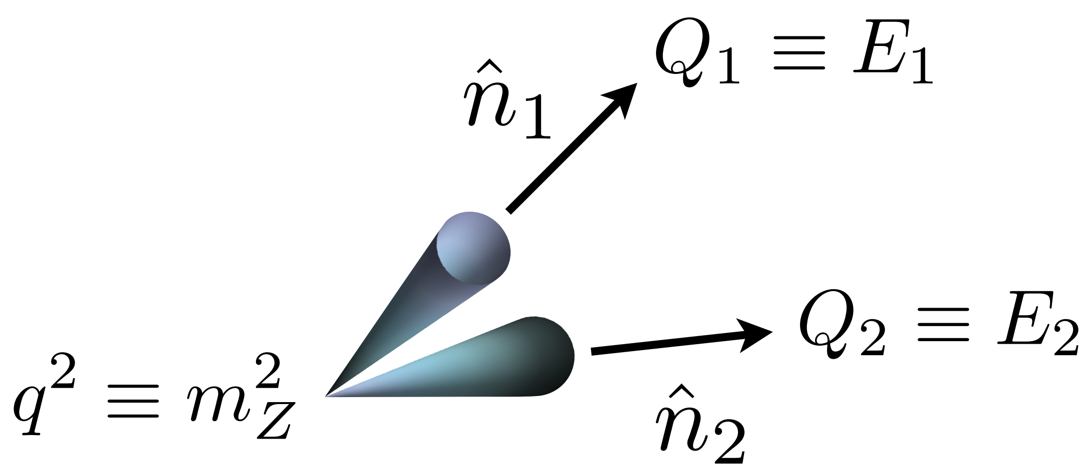
τ_2/τ_1 ($Q = \infty$) ≈ e x thrust (!!)



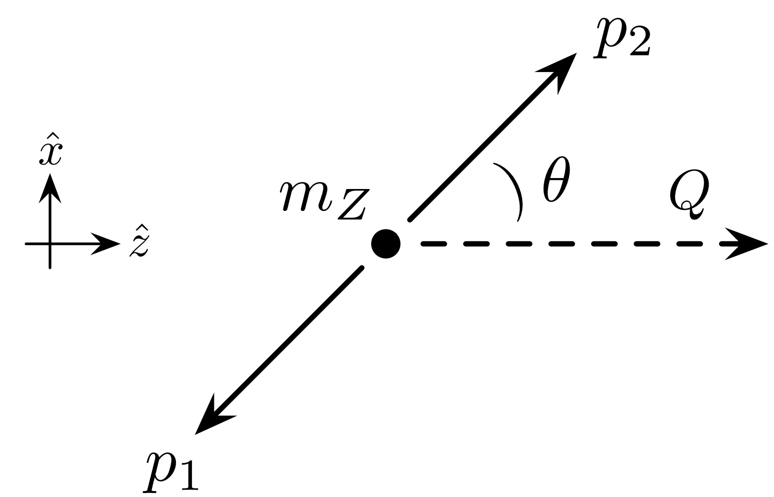
Control over ISR/UE/Pileup (!!!)

Allows systematic jet substructure calculations in 1/Q

2-subjettiness in SCET



$$\sqrt{\hat{s}_{12}} \equiv \sqrt{2\hat{n}_1 \cdot \hat{n}_2}$$



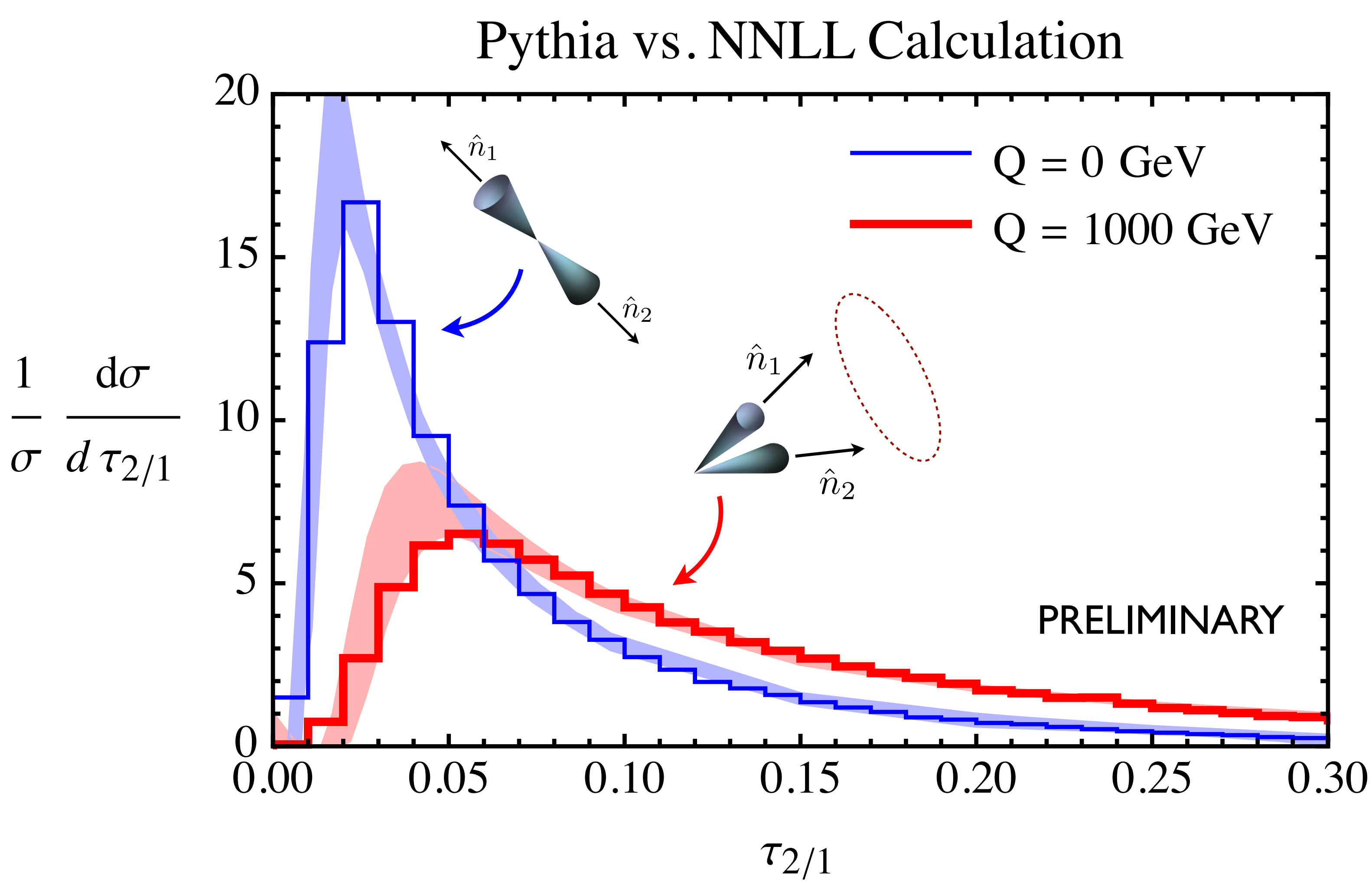
θ = angle between rest frame thrust axis and boost axis

Recycle NNLL thrust by boosting! (and θ integral...)

$$\frac{1}{\sigma_0}\frac{d\sigma}{d\tau_{2/1}} = \widehat{\mathcal{T}}_1^2 \int \frac{d\cos\theta}{2} H(m_Z,\mu_H) U_H(m_Z,\mu_H,\mu_J)$$
$$\times \int dk\, d\ell\, J_{\tau_{2/1}}\Big[\widehat{\mathcal{T}}_1\big(\widehat{\mathcal{T}}_1 \tau_{2/1} - k\big), \Big\{\frac{2Q_i}{\widehat{\mathcal{T}}_1}\Big\}, \mu_J\Big]$$
$$\times U_S^\tau\Big(k-\ell, \frac{\mu_J}{\beta_\theta}, \frac{\mu_S}{\beta_\theta}\Big)\, S_\tau\Big(\ell - \frac{2\overline{\Omega}_1}{\beta_\theta}, \frac{\mu_S}{\beta_\theta}\Big),$$

Boosted Z Signal Predictions

Surprisingly good agreement between Pythia and NNLL



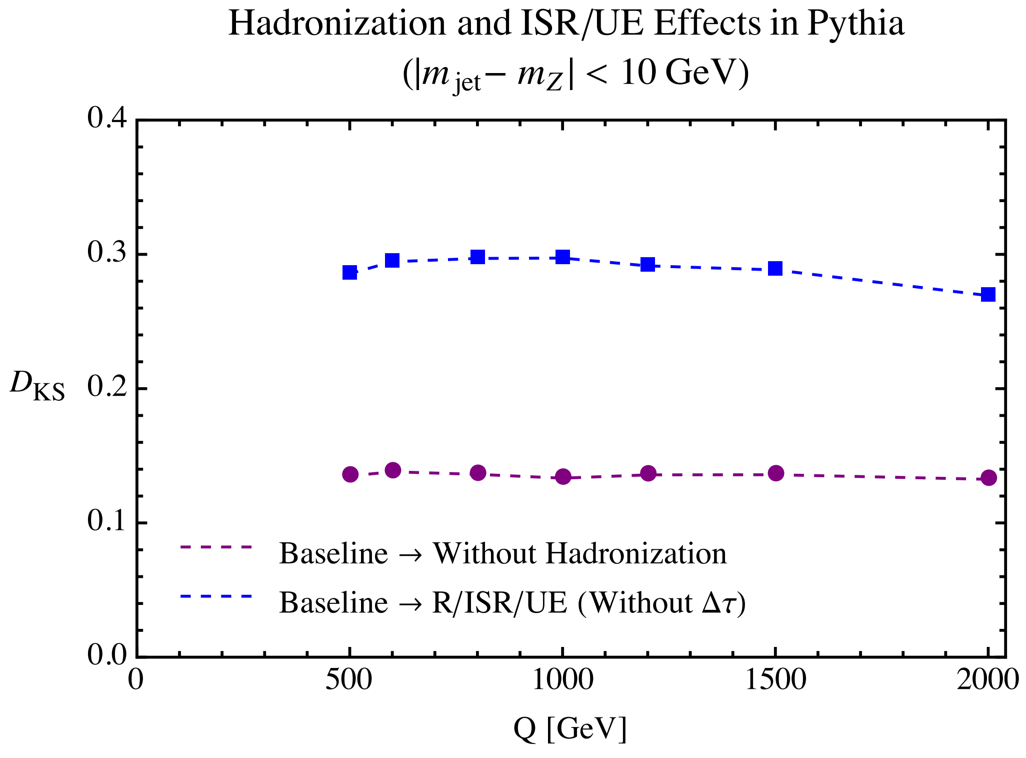
Extra SCET machinery needed for precise peak shape ↔ Tails match beautifully

Next targets: fat jet background, boosted top signal

Correction Factors

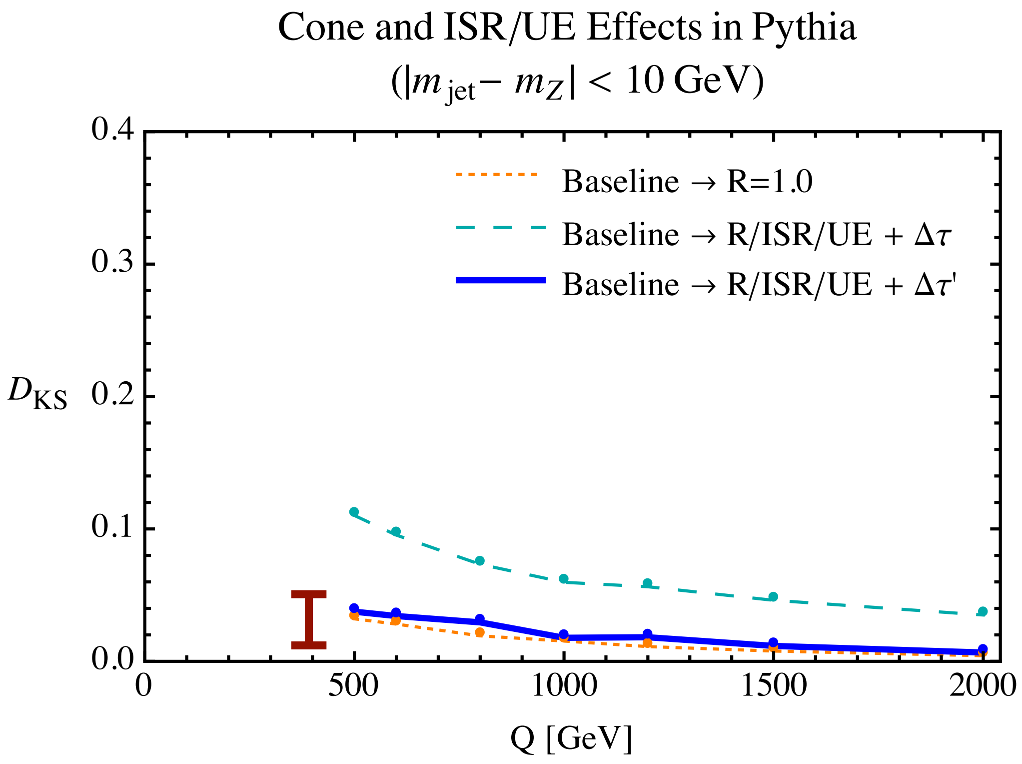
Estimated sizes using Pythia

Effects included in SCET calculation



Fully analytic corrections, hadronization can be derived from LEP thrust

Effects neglected (scale as 1/Q)



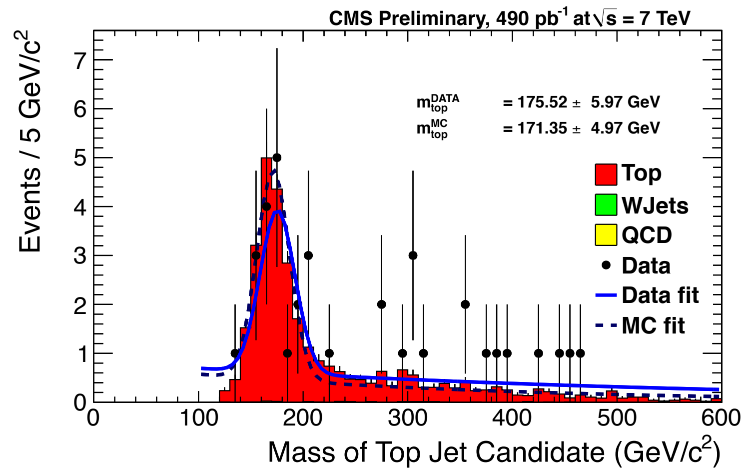
Remaining effects less than 5% (also Γ/m_Z calculable effects)

Excitement at QCD/BSM Boundary

Jet Substructure & N-subjettiness

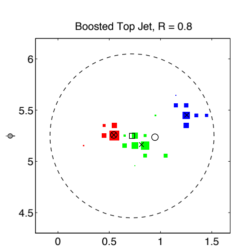
Overview of Jet Substructure

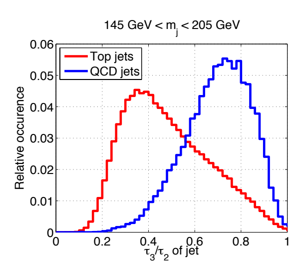
Boosted objects at the LHC



Introducing N-subjettiness

Top tagging with τ_3/τ_2, W/Z/H tagging with τ_2/τ_1



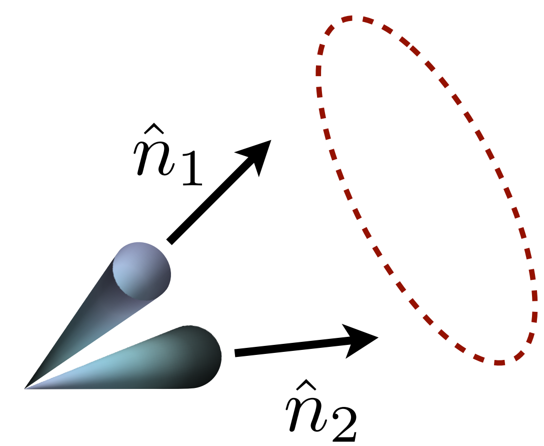


$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min_A \{\Delta R_{A,k}\}$$

[JDT, Van Tilburg: 1011.2268 & 1108.2701]

Ongoing Theoretical Studies

Expanding about the infinite boost limit

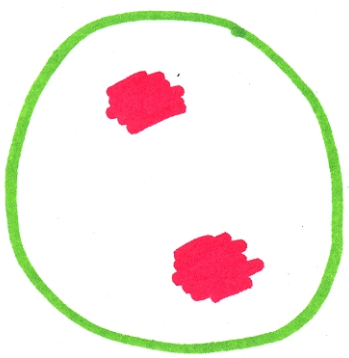


[Fiege, Schwartz, Stewart, JDT: 12xx.xxxx]

Backup Slides

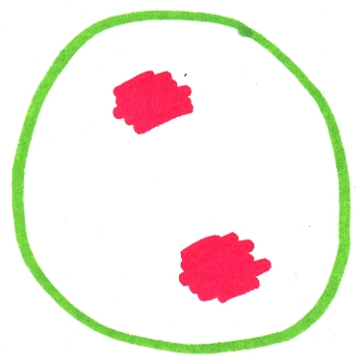
Credit to Original Literature

W



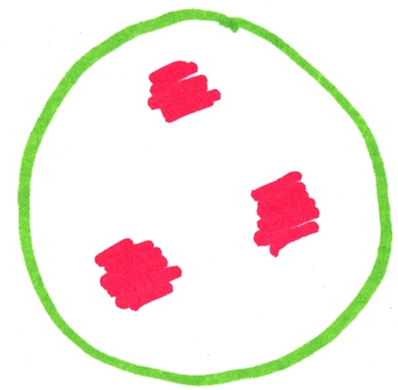
[Seymour; Butterworth, Cox, Forshaw; ...]

Higgs



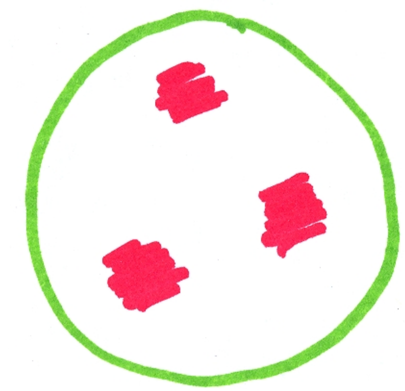
[Butterworth, Davison, Rubin, Salam; Plehn, Salam, Spannowsky; see also Kribs, Martin, Roy, Spannowsky; ...]

Top



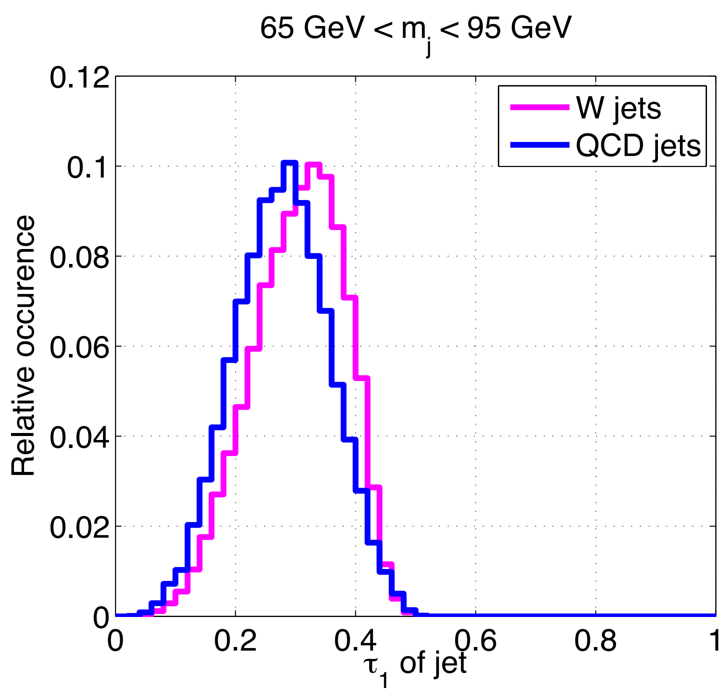
[Brooijmans; Kaplan, Rehermann, Schwartz, Tweedie; Ellis, Vermilion, Walsh; see also JDT, Wang; Almeida, Lee, Perez, Sterman, Sung, Virzi; ...]

SUSY

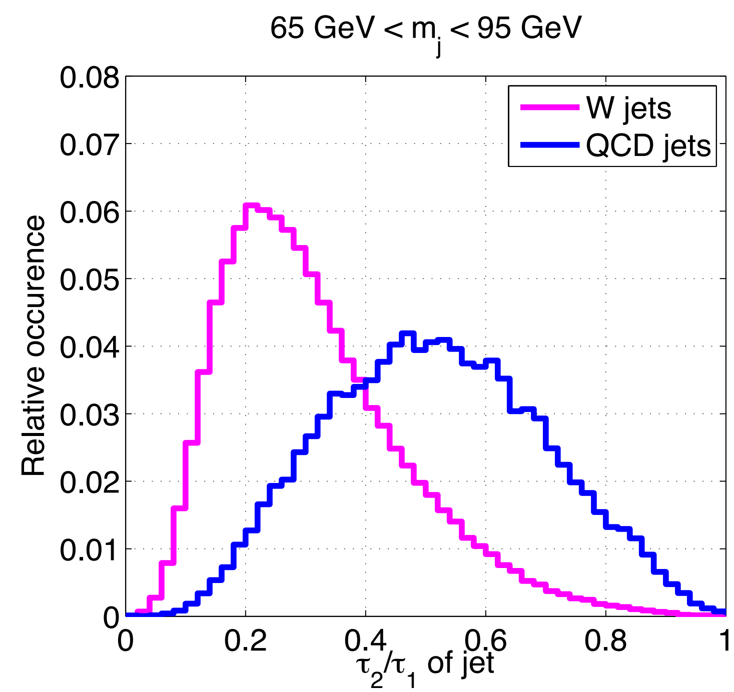


[Butterworth, Ellis, Raklev, Salam; ...]

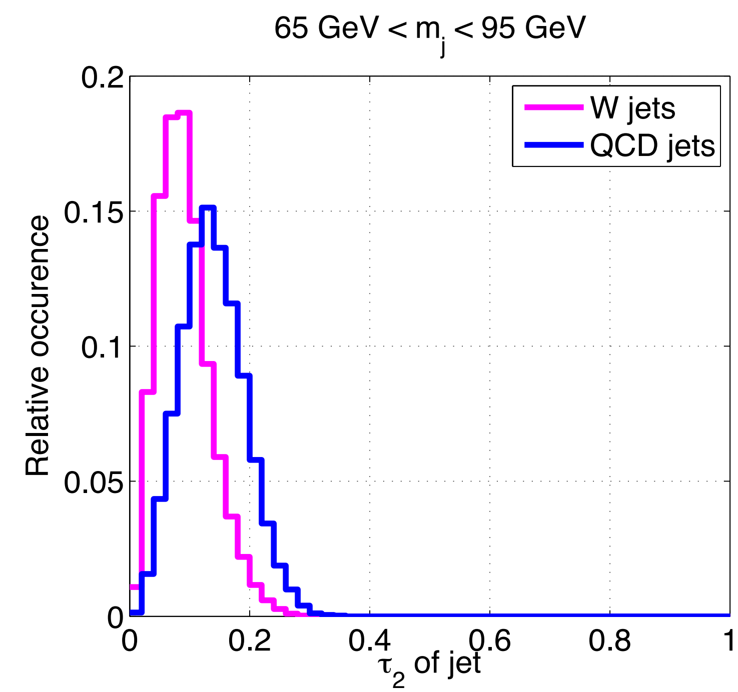
W Tagging



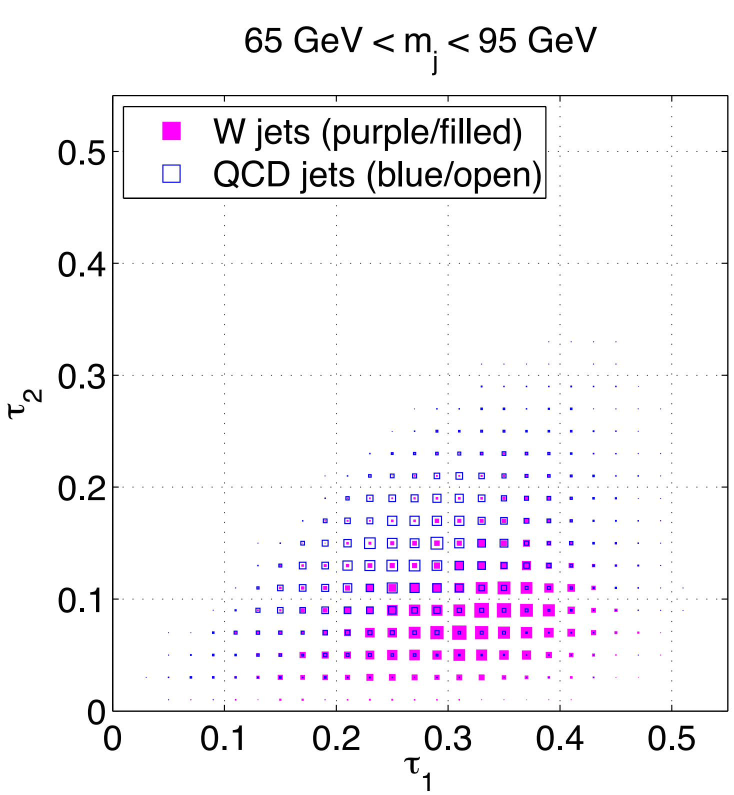
65 GeV < m_j < 95 GeV
W jets
QCD jets
Relative occurence
τ_1 of jet



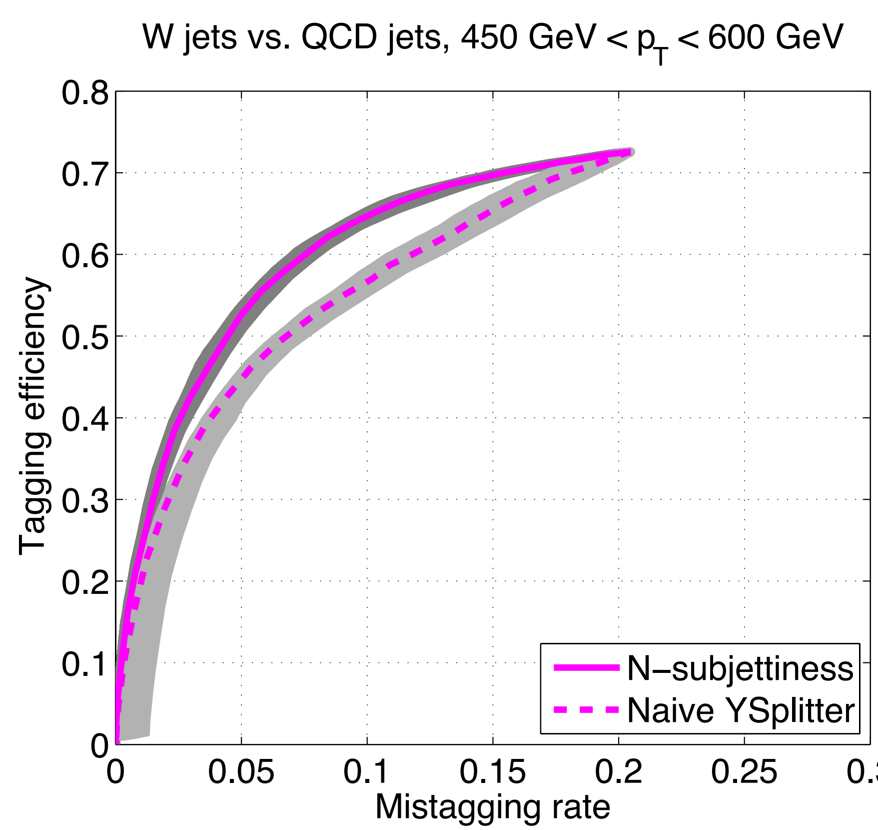
65 GeV < m_j < 95 GeV
W jets
QCD jets
Relative occurence
τ_2/τ_1 of jet



65 GeV < m_j < 95 GeV
W jets
QCD jets
Relative occurence
τ_2 of jet



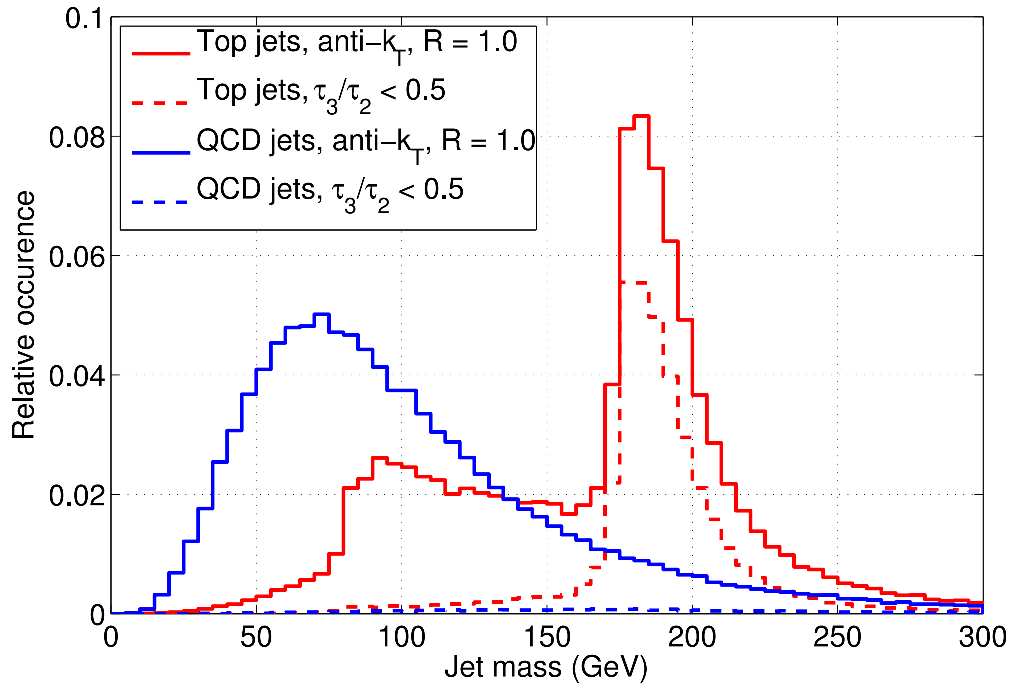
65 GeV < m_j < 95 GeV
W jets (purple/filled)
QCD jets (blue/open)
τ_2
τ_1

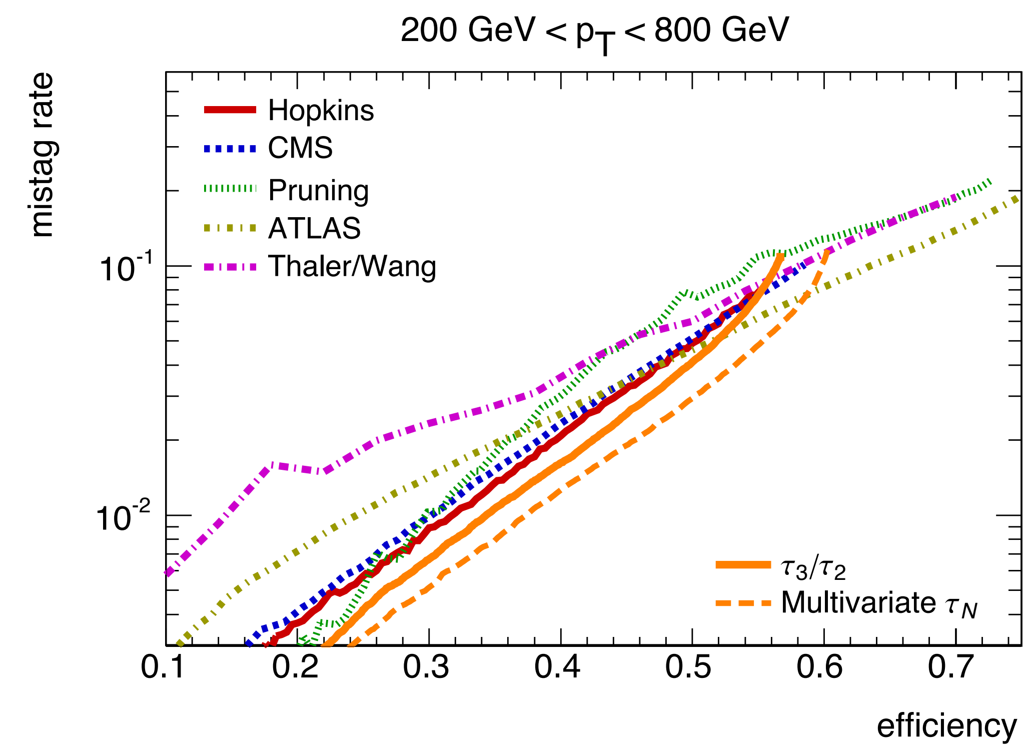


W jets vs. QCD jets, 450 GeV < p_T < 600 GeV
N−subjettiness
Naive YSplitter
Tagging efficiency
Mistagging rate

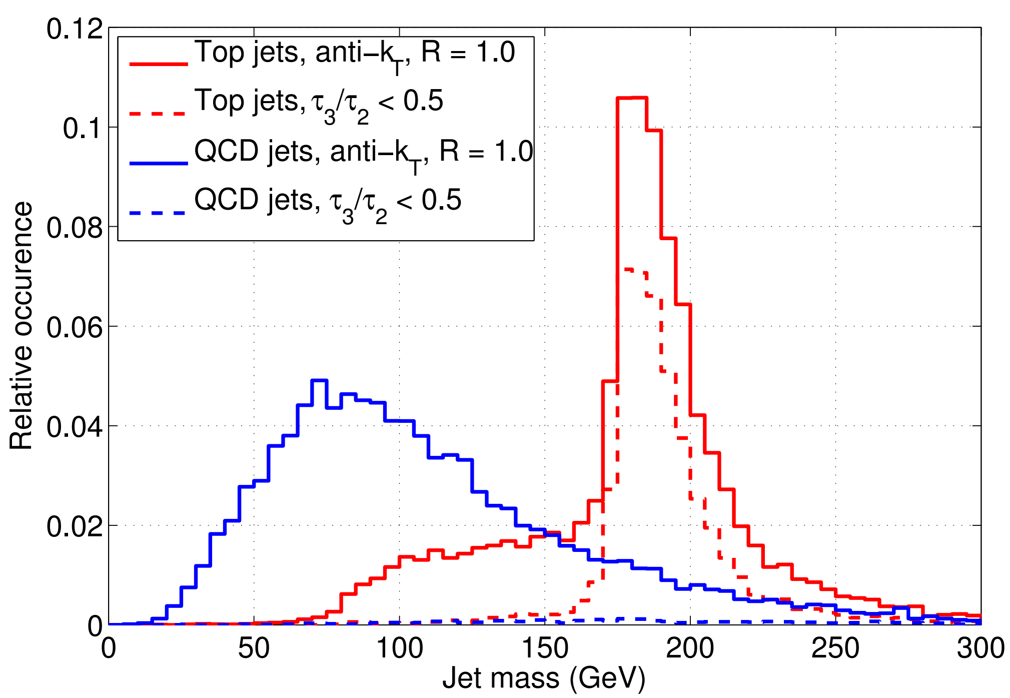
More Plots

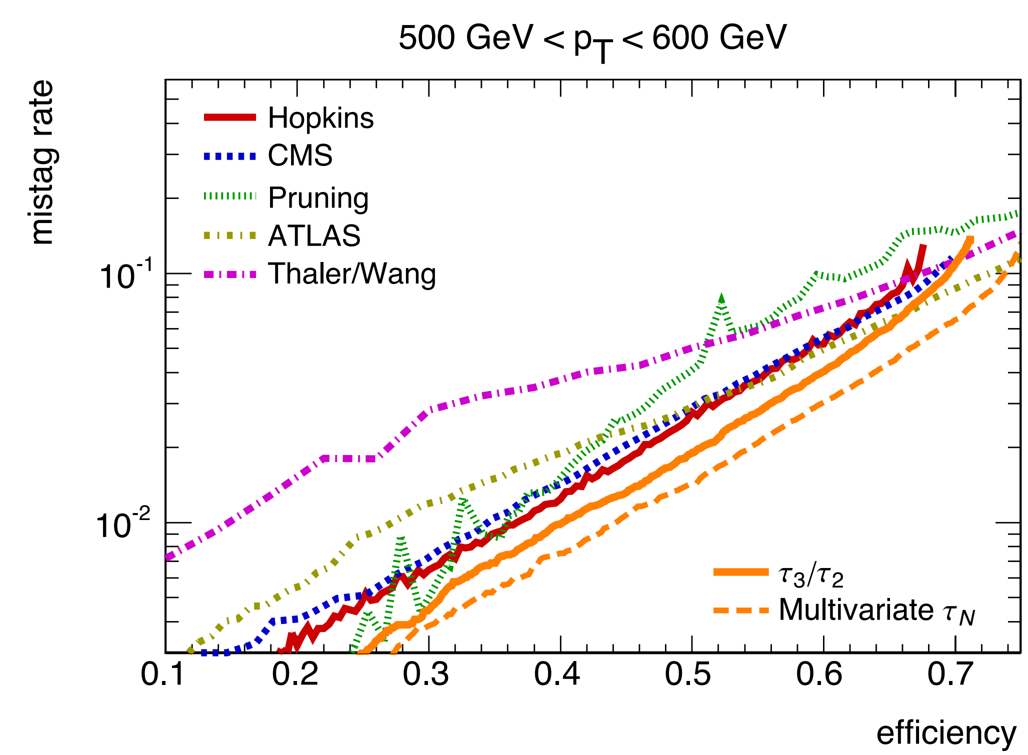
200 GeV < p_T < 800 GeV



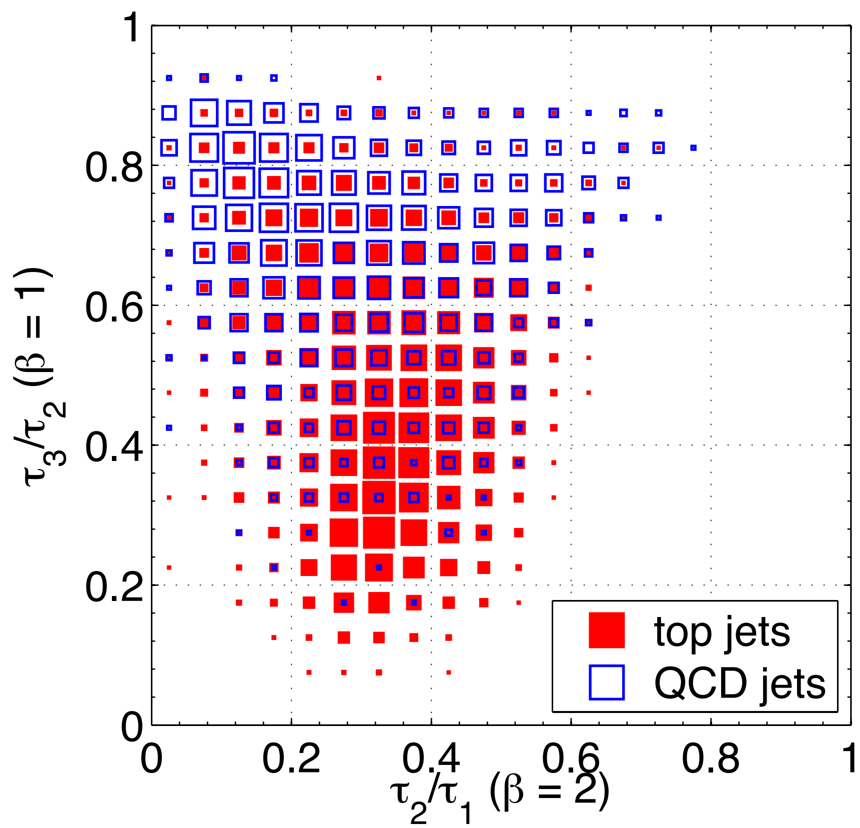


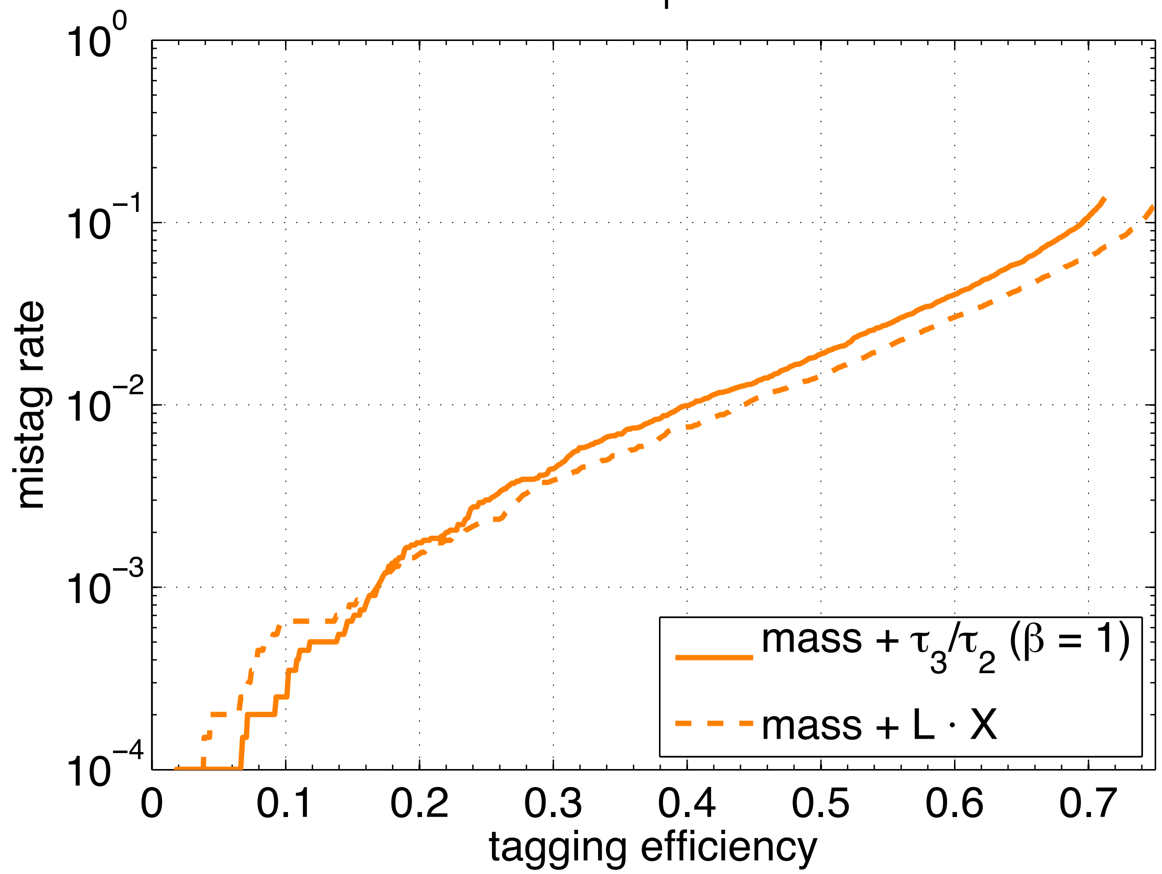
500 GeV < p_T < 600 GeV





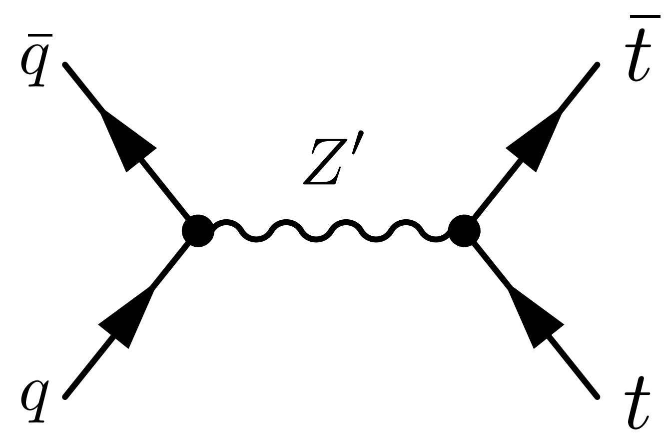
Linear Fisher Discriminant

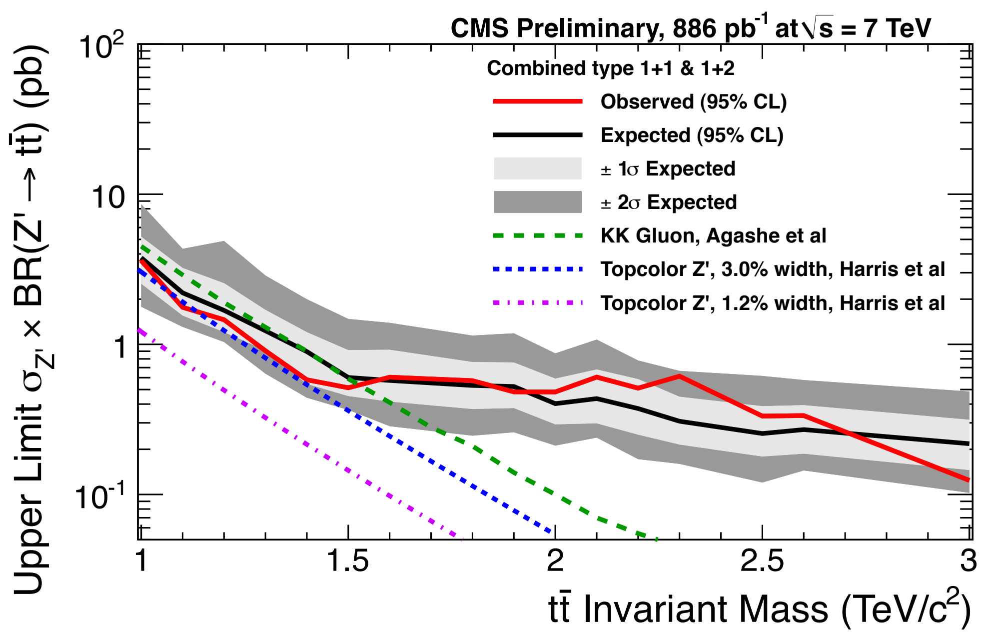




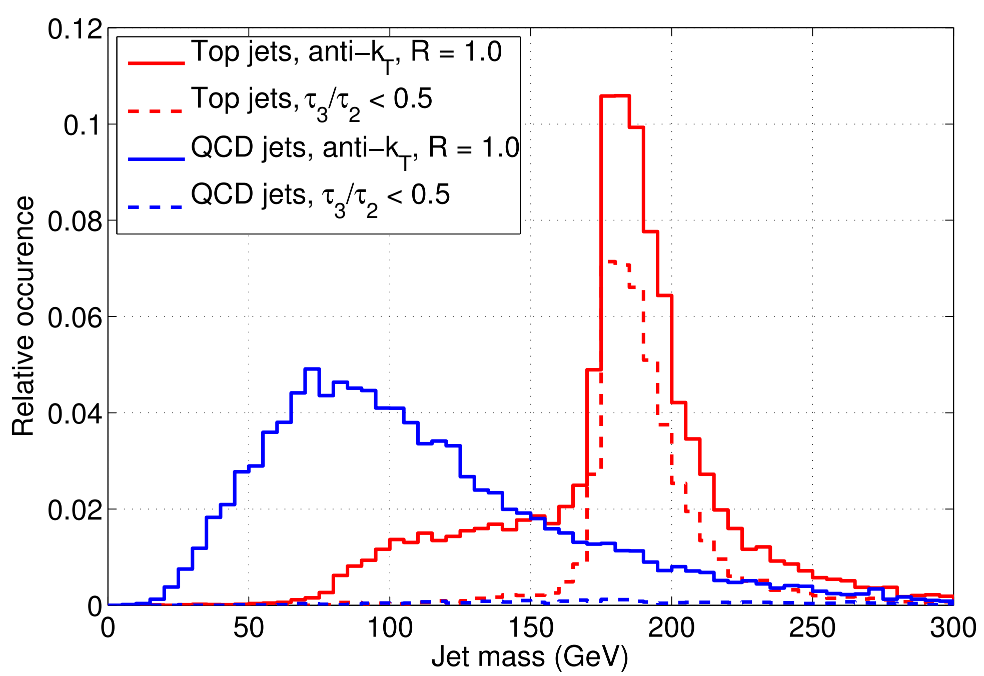
$$160\ \text{GeV} < m_{\text{jet}} < 280\ \text{GeV}$$
$$\vec{L} \cdot \vec{X} < c \text{ with}$$
$$\vec{L} = [2.30,\ \text{-}5.85,\ \text{-}1.89,\ 6.21,\ 7.25,\ \text{-}5.35,\ \text{-}0.86,\ 1.61,\ \text{-}14.07]$$
$$\vec{X} = \left[\tau_1^{(1)}, \tau_2^{(1)}, \tau_3^{(1)}, \frac{\tau_2^{(1)}}{\tau_1^{(1)}}, \frac{\tau_3^{(1)}}{\tau_2^{(1)}}, \frac{\tau_2^{(2)}}{\tau_1^{(2)}}, \frac{\tau_3^{(2)}}{\tau_2^{(2)}}, \left(\frac{\Delta m}{m}\right)_+, \left(\frac{\Delta m}{m}\right)_-\right]$$

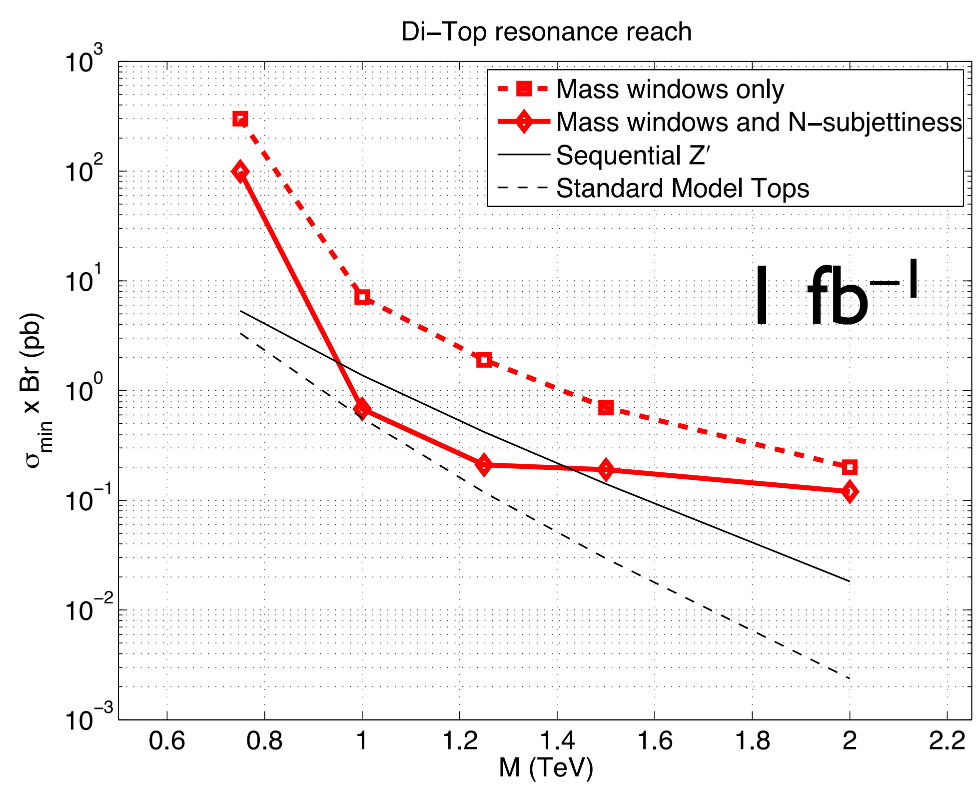
Improving the Reach





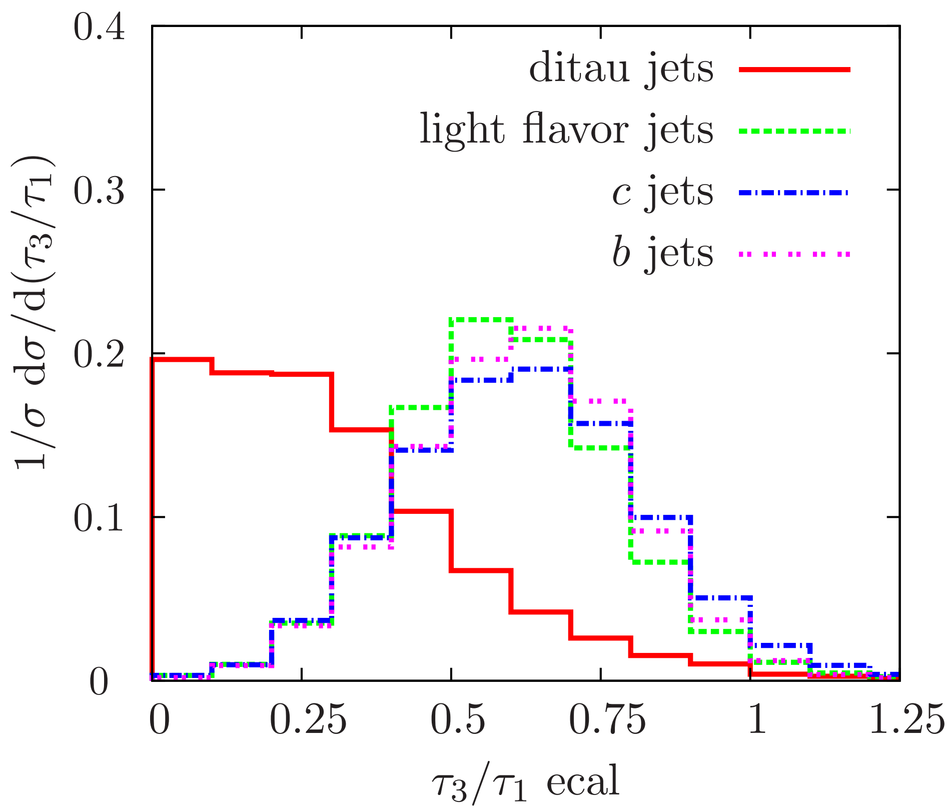
Fight QCD even before applying top mass cut





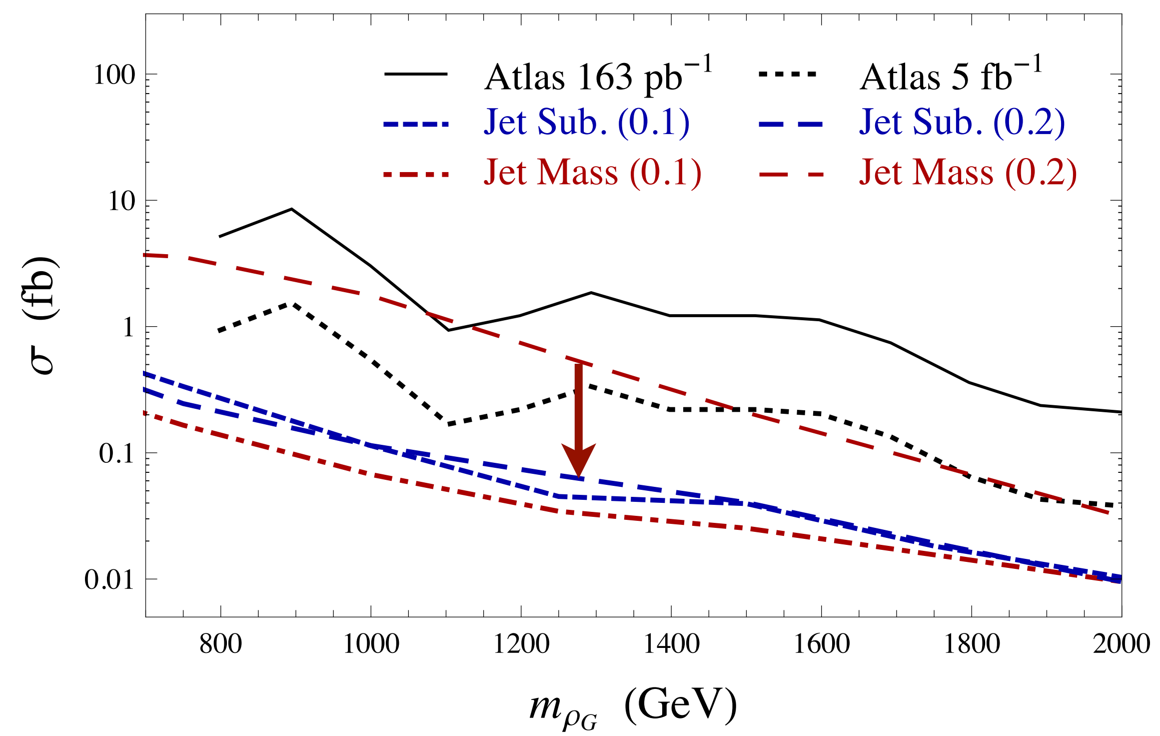
Recent Applications

$h \to aa \to (\tau^+\tau^-)(\tau^+\tau^-)$



[Englert, Roy, Spannowsky]

$\rho_G \to \pi_G \pi_G \to (gg)(gg)$



[Bai, Shelton]

N-subjettiness: Powerful discriminating variable, by itself or as part of a multivariate study

Minimization

Classic problem in computer science!

$$\tau_N = \min_{\{p_{\text{axes}}\}} \tilde{\tau}_N \qquad \tilde{\tau}_N = \frac{1}{d_0} \sum_k p_{T,k} \min_A \{\Delta R_{A,k}^{\beta}\}$$

↑ True Jet Shape

↑ Hybrid Jet Shape

β = 2 (≈ thrust): "k-means clustering"

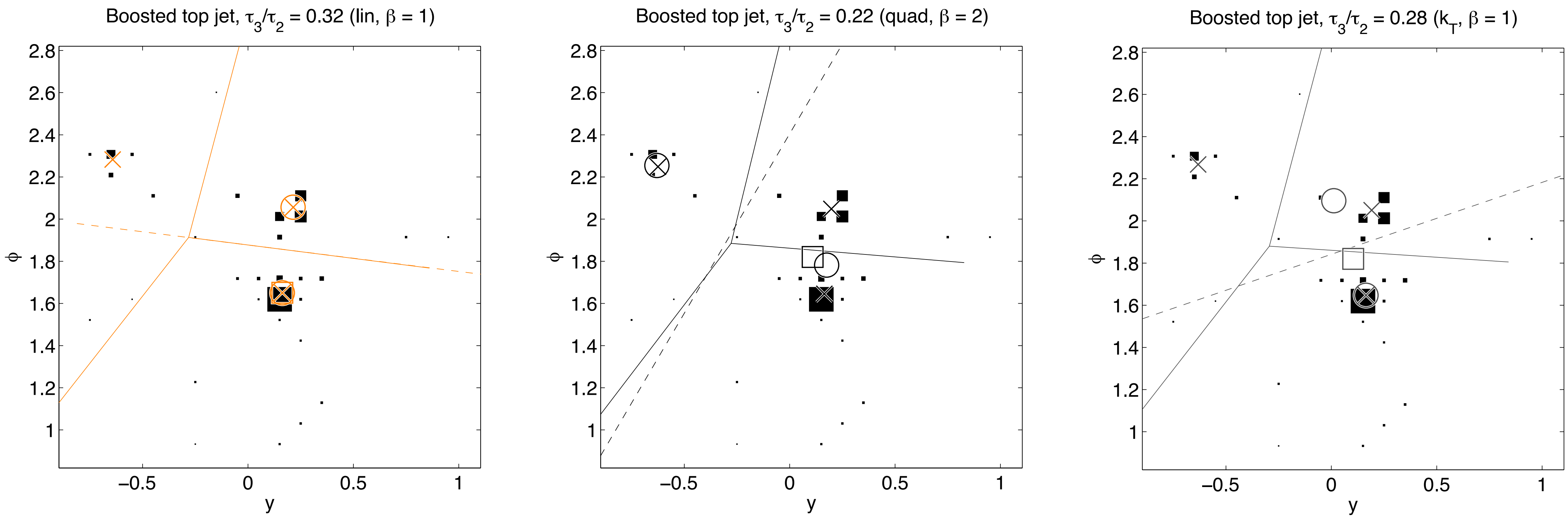
≈ iterative jet finding, fixed number of jets, voronoi boundaries

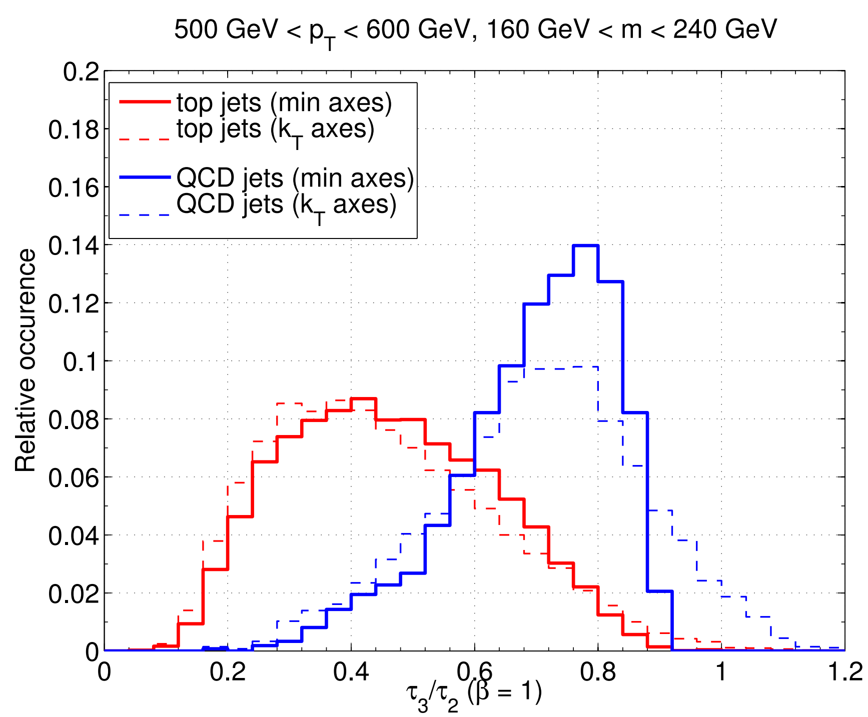
$$\frac{\partial \tilde{\tau}_1}{\partial \vec{p}_{\text{axis}}} = 0 \qquad \Rightarrow \qquad \vec{p}_{\text{axis}} = \sum_k \vec{p}_k$$

i.e. jet axis = jet momentum

β = 1 (≈ jet broadening): We developed new minimization algorithm, jet axis ≠ jet momentum, kind of like a "median" jet

Effect of Minimization



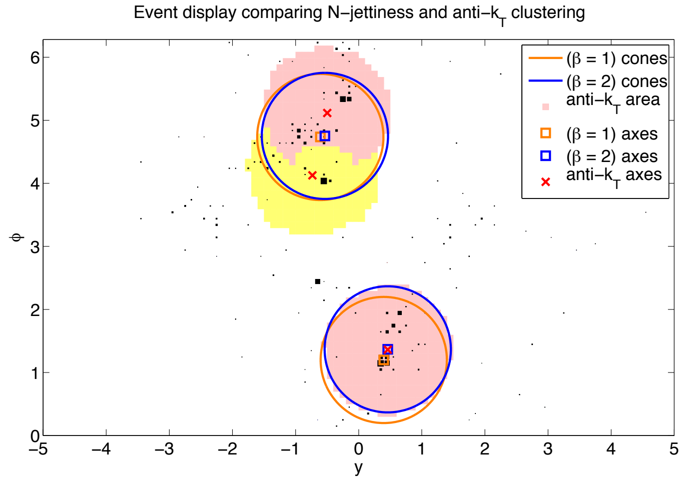


With minimization,
$\tau_N/\tau_{N-1} < 1$

Better Top/QCD Separation

A New/Old Jet Algorithm

Exclusive Cones for Higgs Searches



[JDT, Vermilion: Ongoing]

Thoughts on Jet Algorithms

Hybrid Jet Shape: $\tau_N = f(\{p_k\}, \{p_{\text{axes}}\})$

Minimization of any hybrid jet shape defines a jet algorithm!

Not a new idea... Stable cone finding:

Minimize over axis A: $$\tau_1(R_0) = \sum_k p_{Tk} \min(\Delta R_{A,k}, R_0)^2$$ ← Key!

Solution: $$p_{\text{jet}} = \sum_k p_k$$ in cone

[Ellis, Huston, Tonnesmann]

In CS optimization: cluster finding = minimization

N-jettiness

Original Purpose: Define exclusive jet cross sections
Especially important for Higgs + N jet searches

[Stewart, Tackmann, Waalewijn]

"Exactly N jets"?

Current Method

Run jet algorithm,
require exactly N jets
above jet p_T threshold

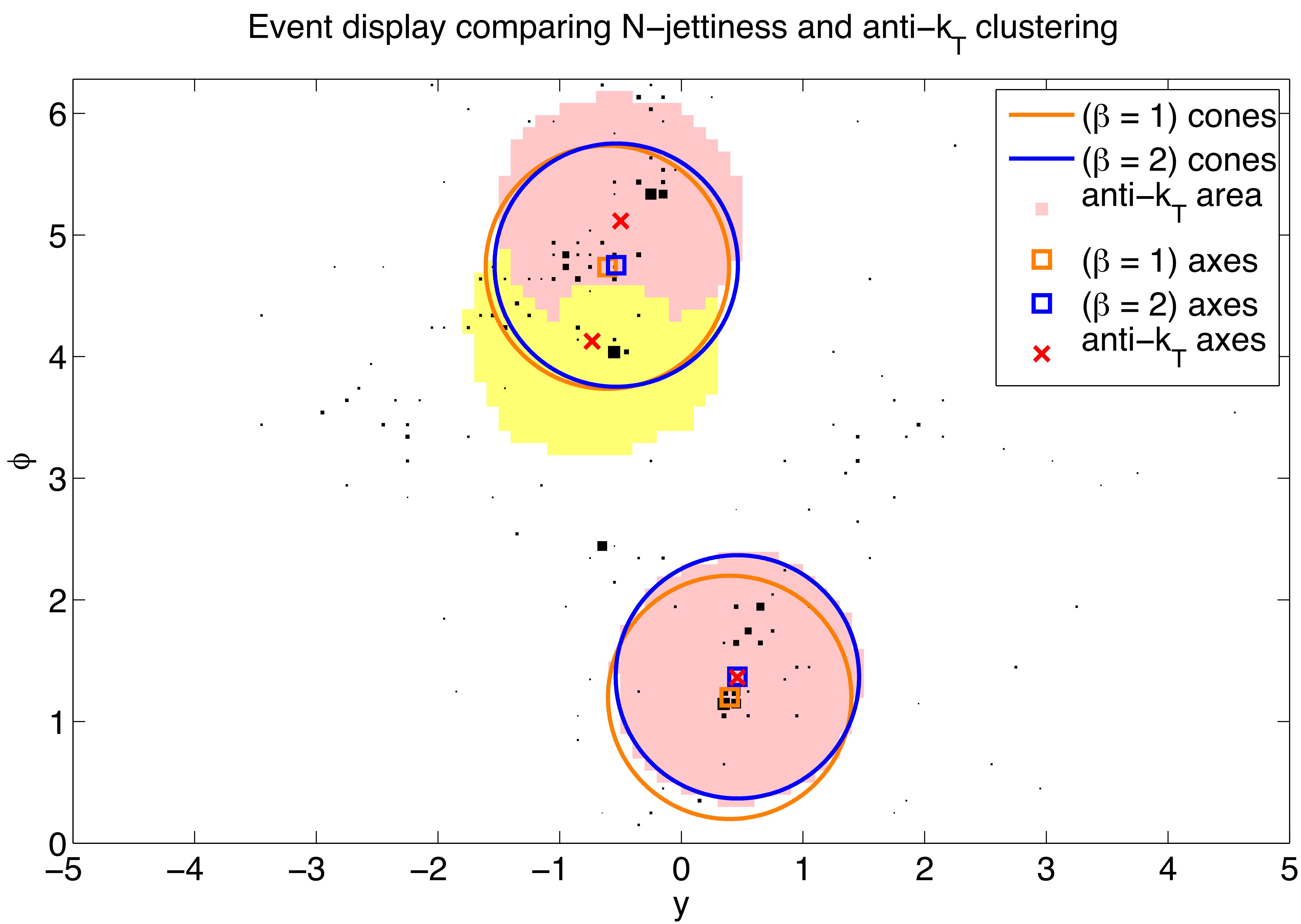
Alternative Method

Find exactly N jets,
restrict event to be below
N-jettiness threshold

With minimization, can use to define N jets!

$$\tau_N^{(\beta)}(R_0) = \sum_i p_{T,i} \min\left\{(\Delta R_{1,i})^\beta, \ldots, (\Delta R_{N,i})^\beta, (R_0)^\beta\right\}$$

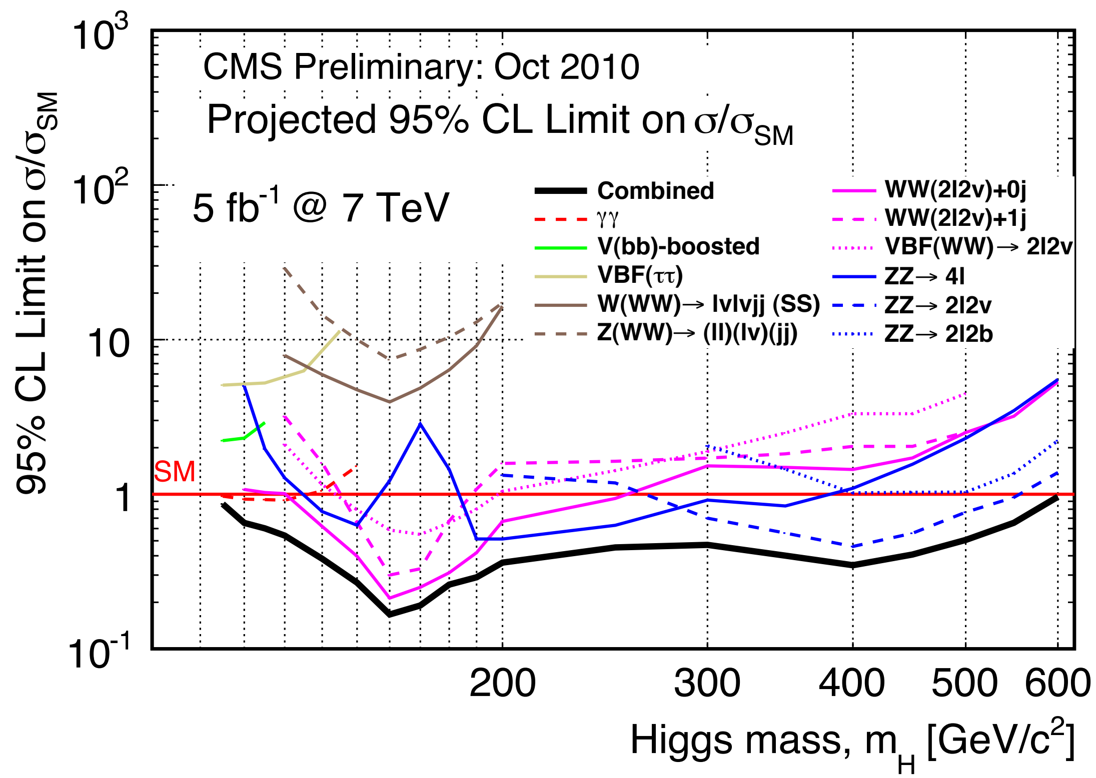
N-jettiness Jets



A Killer App?

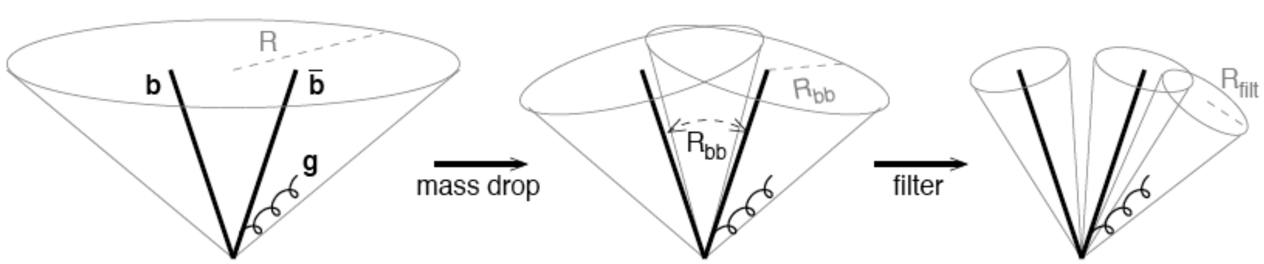
The Boosted Higgs Search

$$q\bar{q} \to Z^* \to ZH,\quad H \to b\bar{b},\quad Z \to \ell^+\ell^-$$



BDRS Method

[Butterworth, Davison, Rubin, Salam]



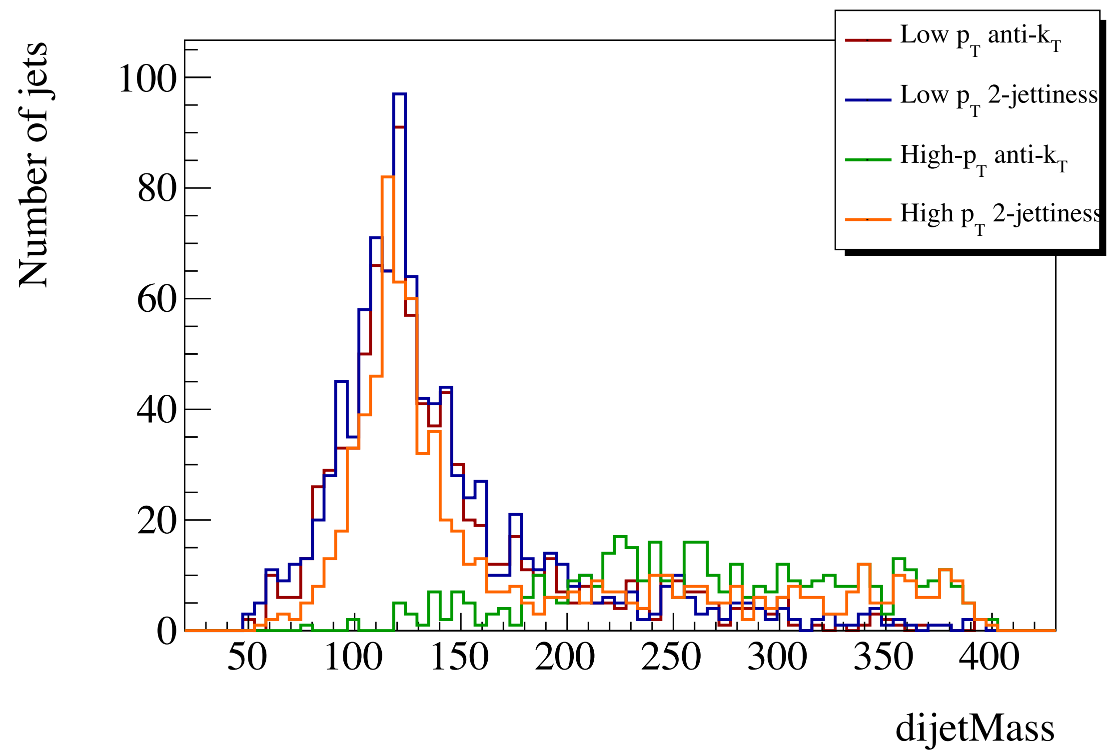
An Alternative?

Signal has exactly 2 jets!
No need for jet substructure,
just minimize 2-jettiness

(Or if you have to use substructure,
use τ_2/τ_1 on fat jets)

Interpolating the Higgs Search

2-jettiness vs. anti-kT: interpolating between regimes

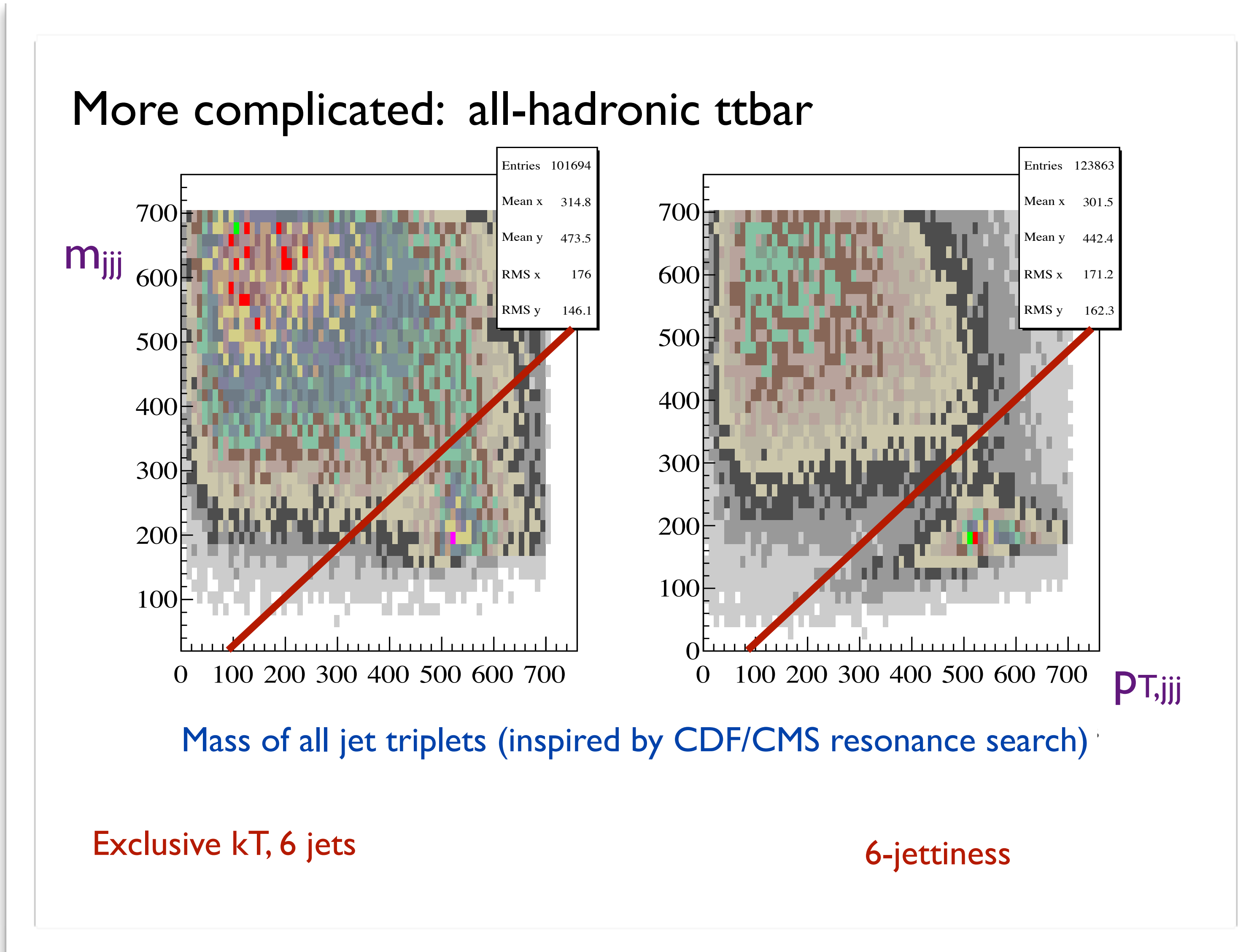


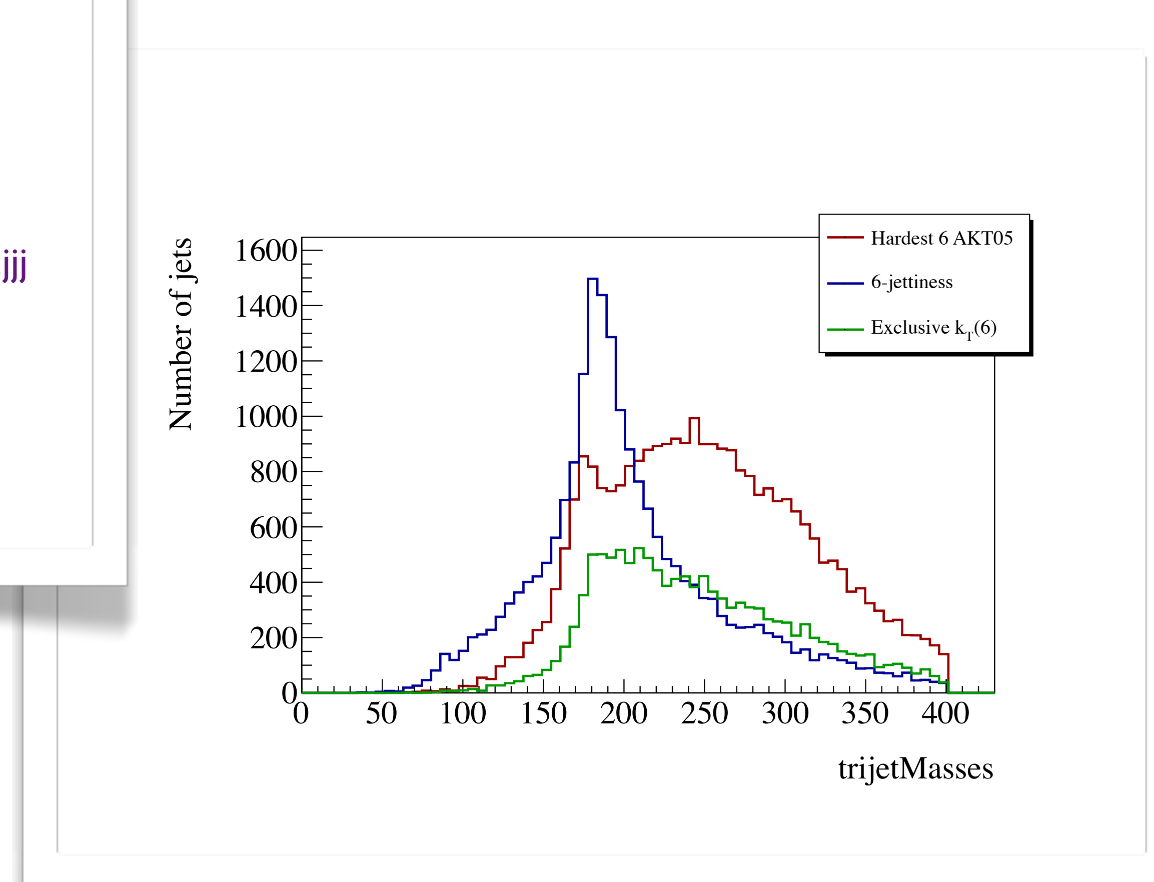
Signal: H(bb)Z(μμ)
low p_T = no cut
high p_T = > 300 GeV

Same analysis works over all(?) p_T!

[Chris Vermilion, very preliminary]

Interpolating for Top?





[Chris Vermilion, very preliminary]