# JET PRUNING: Looking for New (BSM) Physics at the LHC with Jets

Steve Ellis – INTRO
Jon Walsh – Validation
Chris Vermilion - Engineering

0903.5081

## Big Picture:

For the next decade the focus of particle physics phenomenology will be on the LHC. The LHC will be both very exciting and very challenging -

- addressing a wealth of essential scientific questions
- with new (not understood) detectors
- operating at high energy *and* high luminosity
- most of the data will be about hadrons (jets)

Theory and Experiment must work together to make the most of the data.

# Outline & Issues

- Brief review of (QCD) jets, including masses
- Search for BSM physics in SINGLE jets – bumps in mass distributions
  $\Rightarrow$ Smooth but Large QCD background
- Consider Recombination (kT) jets $\rightarrow$ natural substructure but also
  algorithm systematics
  and contributions from ISR, FSR, UE and Pile-up
- Improve by PRUNING (removing) large angle, soft branchings
- Validate with studies of surrogate new heavy particle – top q - Jon
- Pruning engineering - Chris

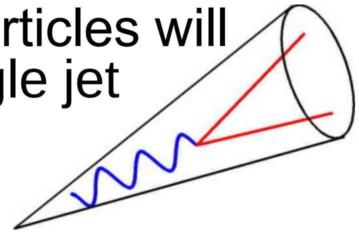
## Why JETS?

Essentially all LHC events involve an important hadronic component, only $Z' \to \mu^+ \mu^-$ avoids this constraint

The primary tool for hadronic analysis is the study of jets, to map long distance degrees of freedom (i.e., detected) onto short distance dof (in the Lagrangian)

Jets used at the Tevatron to test the SM, will be used at the LHC to test for *non-SM-ness*

Most SM particles (top quarks, *W*'s, *Z*'s) and some BSM particles will often be produced with a large enough boost to be in a single jet



**SEARCH** for new particles by focusing on jet masses (bumps in the distribution) and jet substructure - bumps in masses of sub-jets, and …

# Jet Physics: The Basis of QCD Collider Phenomenology

Long distance physics = complicated (all orders showering of colored objects, nonperturbative hadronization = organization into color singlets)

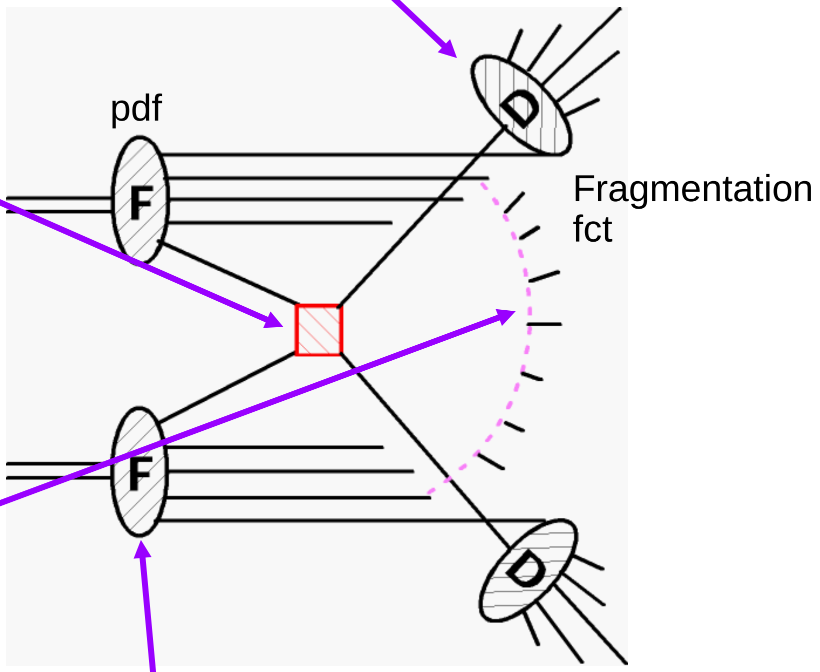
Measure this in the detector

Short distance physics = simple (perturbative)

Want to study this

Correlated by Underlying Event (UE) color correlations + Pile Up

Stuck with this, small?




More long distance physics, but measured in pdfs

# Jets – a brief history at Hadron Colliders

- JETS I – Cone jets applied to data at the ISR, SpbarpS, and Run I at the Tevatron to map final state hadrons onto LO (or NLO) hard scattering, essentially 1 jet ⇔1 parton (test QCD)

Little attention paid to masses of jets or the internal structure, except for energy distribution within a jet

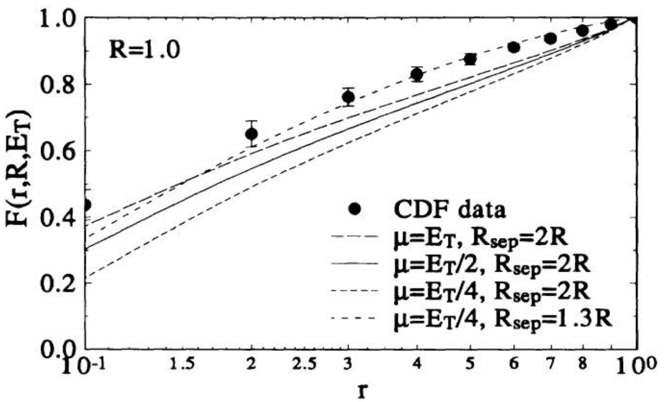



FIG. 2. $F(r,R,E_T)$ vs $r$ for $R=1.0$, $\sqrt{s}=1800$ GeV, $E_T=100$ GeV, and $0.1<|\eta|<0.7$ with $\mu=E_T/4$, $E_T/2$, $E_T$ compared to data from CDF [7]; the dot-dashed curve is explained in the text.

- JETS II – Run II & LHC, starting to look at structure of jets: masses and internal structure – a jet renaissance

# Defining Jets

•Map the observed (hadronic) final states onto the (short-distance) partons by summing up all the approximately collinear stuff, ideally on an event-by-event basis.

• Need rules for summing $\Rightarrow$ jet algorithm

Start with list of particles/towers

End with list of jets (and stuff not in jets)

*E.g.*,

- <u>Cone Algorithms</u>, based on fixed geometry – focus on core of jet

  Simple, "well" suited to hadron colliders with Underlying Events (UE)

- <u>Recombination</u> (or kT) <u>Algorithm,</u> based on pairwise merging to undo shower

  Tends to "vacuum up" soft particles, "well" suited to e+e- colliders

## The good news about jet algorithms:

- 👍 Render PertThy IR & Collinear Safe, potential singularities cancel
- 👍 Simple, in principle, to apply to data and to theory
- 👍 Relatively insensitive to perturbative showering and hadronization

## The bad news about jet algorithms:

- 👎 The mapping of color singlet hadrons on to colored partons can *never* be 1 to 1, event-by-event!
- 👎 There is no unique, perfect algorithm; all have systematic issues
- 👎 Different experiments use different algorithms (and seeds)
- 👎 The detailed result depends on the algorithm

# Recent progress in understanding/using jets

- Improved tools and understanding of algorithms – eg. G. Salam
- Improved analytic descriptions – eg. G. Sterman and collaborators,
  SCET community (C. Lee, I. Fleming, S. Stewart, et al.)
- Jet selection schemes to isolate W/Z, top quarks or Higgs as single jets –
  J. Butterworth and collaborators
  UCB Group (J. Thaler, et al.)
  Johns Hopkins Group (D. Kaplan, et al.)
  Stony Brook Group (G. Sterman, et al.)
- Perturbative results for masses – UW
- Generic search/pruning techniques for BSM searches with jets – focus on masses for now - UW

# Recombination – focus on undoing the shower pairwise (local)

Merge list of partons, particles or towers pairwise based on "closeness" defined by minimum value of

$$k_{T,(ij)}^2 \equiv \mathrm{Min}\left[\left(p_{T,i}^2\right)^{\alpha},\left(p_{T,j}^2\right)^{\alpha}\right]\frac{\left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2}{D^2},$$

$$k_{T,i}^2 = \left(p_{T,i}^2\right)^{\alpha}$$

If $k_{T,(ij)}^2$ is the minimum, merge pair and redo list;
If $k_{T,i}^2$ is the minimum → *i* is a jet!
(no more merging for *i*), 1 parameter *D* *(NLO, equals cone for D = R, $R_{sep}$ = 1)*

$\alpha = 1$, ordinary $k_T$, recombine soft stuff first
$\alpha = 0$, *Cambridge/Aachen* (CA), controlled by angles only
$\alpha = -1$, *Anti-$k_T$*, just recombine stuff around hard guys – cone-like

- 👍 Jet identification is unique – no merge/split stage (Cone issue)
- 👍 Everything in a jet, no Dark Towers (Cone issue)
- 👎 Resulting jets are more amorphous, energy calibration difficult (subtraction for UE?), Impact of UE and pile-up not so well understood, especially at LHC
- 👍 FASTJet version (Gavin Salam) goes like N ln N (only recalculate nearest neighbors), plus has scheme for doing UE correction

# Cone Algorithm – focus on the core of jet (non-local)

- Jet = “stable cone” $\Rightarrow$ 4-vector of cone contents || cone direction
- Well studied – but several issues

- **Cone Algorithm – particles, calorimeter towers, partons in cone of size R, defined in angular space, *e.g.*, (*y*,*φ*),**
- **CONE center -** $\left(y^C, \varphi^C\right)$
- **CONE i ∊ *C iff*** $\Delta R^i \equiv \sqrt{\left(y^i - y^C\right)^2 + \left(\varphi^i - \varphi^C\right)^2} \le R$
- **Cone Contents ⇒ 4-vector** $P_\mu^C = \sum_{i \subset C} p_\mu^i$
- **4-vector direction** $\bar{y}^C = 0.5 \ln\left[\frac{P_0^C + P_z^C}{P_0^C - P_z^C}\right]$ ; $\bar{\varphi}^C = \arctan\left[\frac{P_y^C}{P_x^C}\right]$
- **Jet = stable cone** $\left(\bar{y}^C, \bar{\varphi}^C\right) = \left(y^C, \varphi^C\right)$

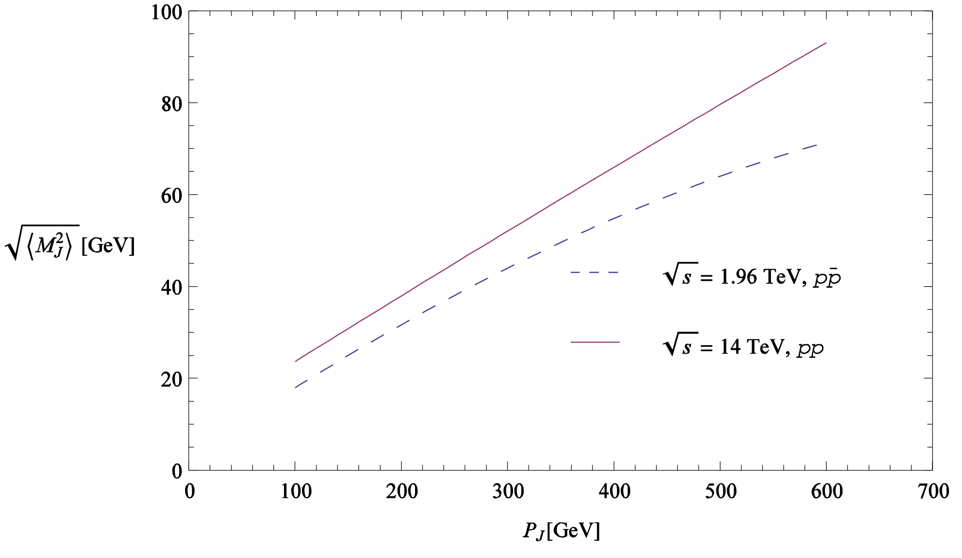
**Find by iteration, *i.e.*, put next trial cone at** $\left(\bar{y}^C, \bar{\varphi}^C\right)$

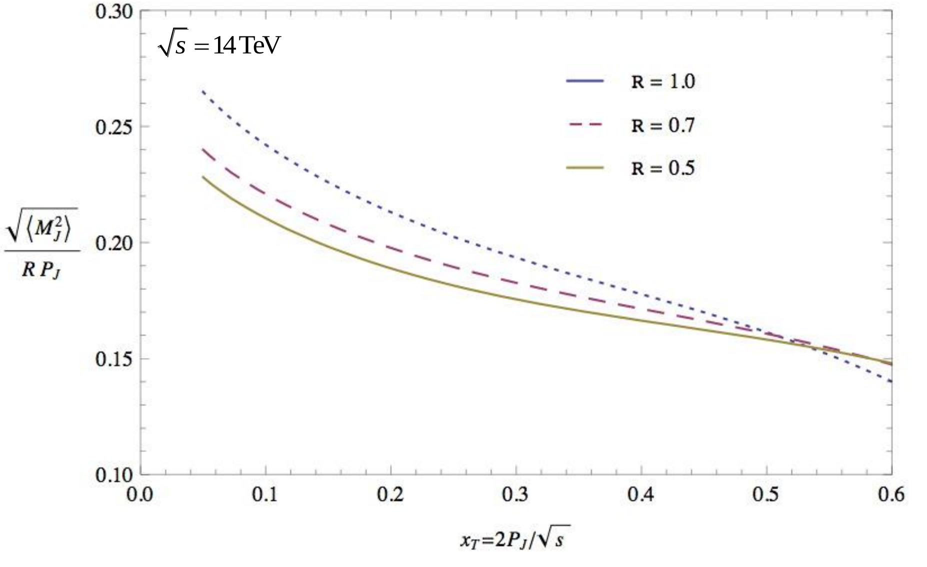
# Jet Masses in QCD: To compare to non-QCD

Dimensions

- In NLO PertThy $\sqrt{p_{J,\mu}p_J^{\mu}} \Rightarrow \sqrt{\langle M^2 \rangle_{NLO}} = f\left(\frac{p_J}{\sqrt{s}}\right)\sqrt{\alpha_s(p_J)}\,p_J R$

Phase space from dpfs, $f \sim 1$ Jet Size, $R \sim \Delta\theta$, determined by jet algorithm

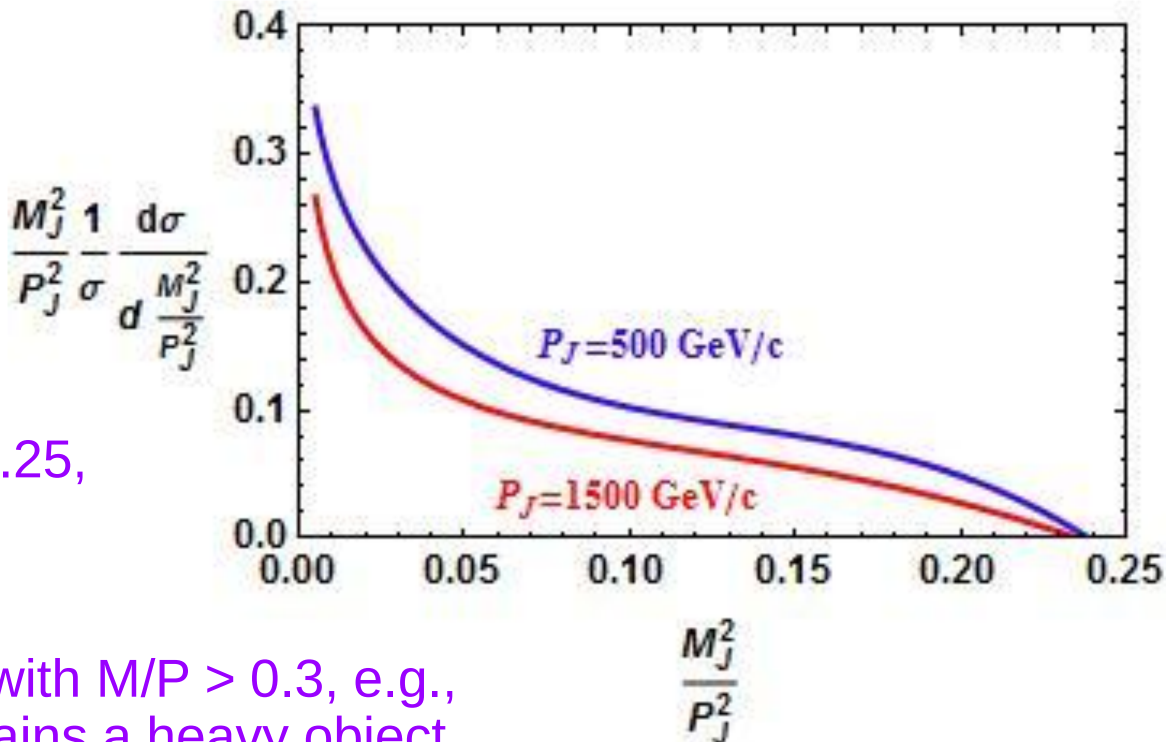



Useful QCD "Rule-of-Thumb" $\Rightarrow \sqrt{\langle M^2 \rangle_{NLO}} \sim 0.2\,p_J R$

# Mass for fixed $P_J$ at NLO

For Cone, R = 0.7
or kT, D = 0.7



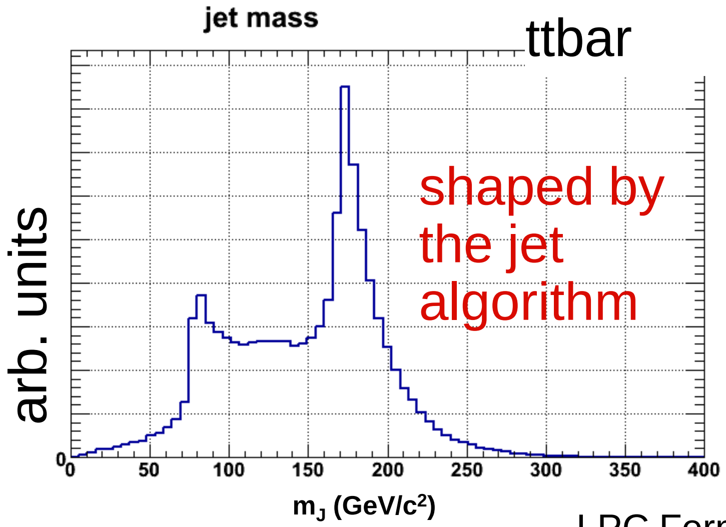

Peaked at low mass,
cuts off for $(M/P)^2 > 0.25$,
$M/P > 0.5$

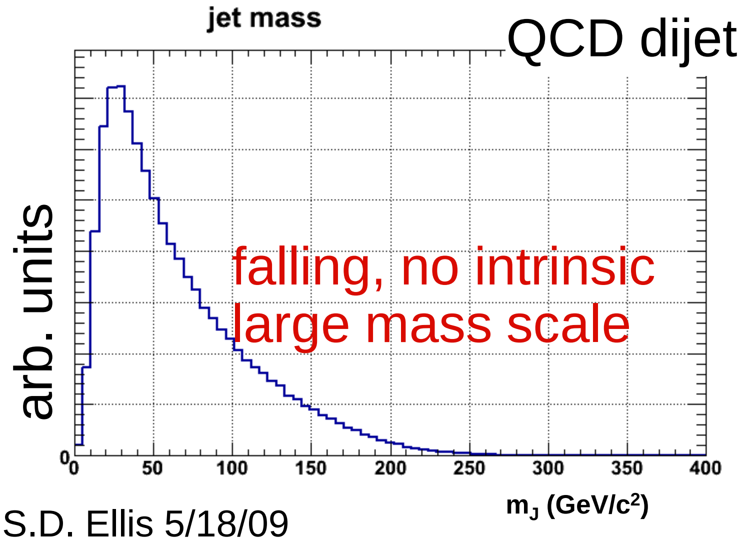
$\Rightarrow$ Selecting on jets with $M/P > 0.3$, e.g., because the jet contains a heavy object, already suppresses the QCD background;

Want heavy particle boosted enough to be in a jet (use large-ish R/D ~1), but not so much to be QCD like ($\sim 2 < \gamma < 5$)

# Finding Heavy Particles with Jets - Issues

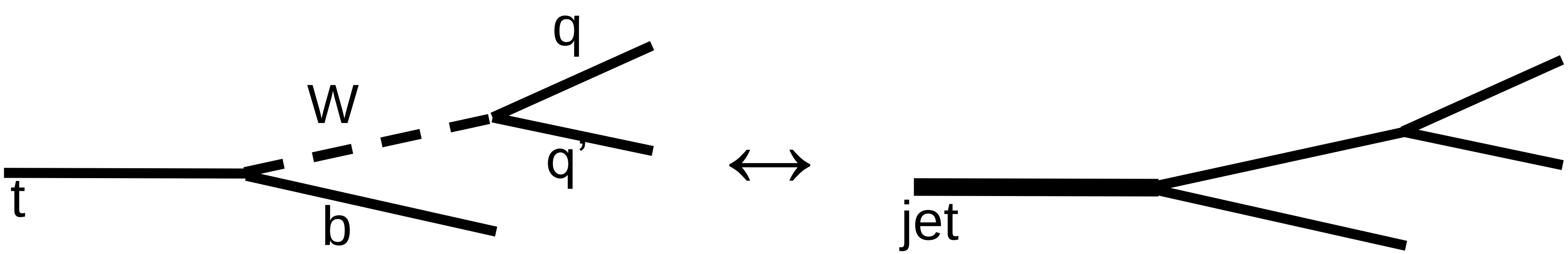
- QCD multijet production rate >> production rate for heavy particles
- In the jet mass spectrum, production of non-QCD jets may appear as local excesses (bumps!) but must be enhanced using analyses
- Use jet substructure as defined by recombination algorithms to refine jets
- Algorithm will systematically shape distributions
- Use top quark as surrogate new particle. $\sigma_{ttbar} \approx 10^{-3} \sigma_{jj}$



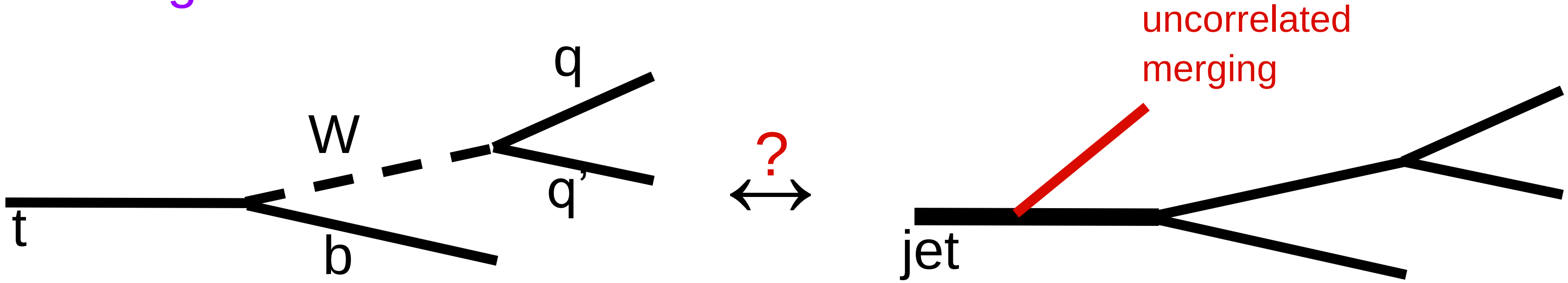
# Using Jet Substructure to separate QCD jets from jets reconstructing heavy particle decays

- Map the kinematics at the vertices onto a decay.
- Masses (jet and subjet) are key variables - strong discriminators between QCD and non-QCD jets.
- How does the choice of algorithm affect the substructure we will observe?



# Reconstruction in Jet Substructure

- Want to identify a heavy particle reconstructed in a single jet.
  - Need correct ordering in the substructure and accurate reconstruction.
  - Must understand how decays and QCD differ in their expected substructure.
  - Makes reconstruction sensitive to systematics of the jet algorithm.
- Jet substructure affected by the systematics of the algorithm.

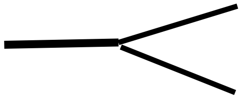


# Systematics of the Jet Algorithm

- Consider generic recombination step: *i,j* ➜ *p*
- Useful variables: (Lab frame) $$z = \frac{\min(p_{T_i}, p_{T_j})}{p_{T_p}} \qquad \theta = \Delta R_{ij}$$
- Merging metrics: $$\rho_{\mathrm{kT}} = p_{T_p} z\theta/D \qquad \rho_{\mathrm{CA}} = \theta/D$$
- In terms of *z*, $\theta$, the algorithms will give different kinematic distributions:
  - CA orders only in $\theta$ : *z* is unconstrained
  - kT orders in $z{\cdot}\theta$ : z and $\theta$ are both regulated
- The metrics of kT and CA will shape the jet substructure.

# Systematics of the Jet Algorithm II

- Subjet masses, mass of jet = $M_J$

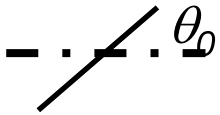


$$a_1 = m_{D,>} / M_J$$

$$a_2 = m_{D,<} / M_J$$

- In jet rest frame (think top decay)
  (note : there is one)

$$\cos\theta_0 = \hat{p}_{D,m>} \cdot \hat{P}_{J,Lab}$$




- Plus an azimuthal angle

- Again angular distributions are strongly shaped by the algorithm, choosing the algorithm is important!

# Studying Systematics: QCD vs ttbar Jets

- We compare the substructure of the kT and CA algorithm by looking at jets in QCD dijet & ttbar events; generated with MadGraph/PYTHIA (DWT tune).
- High pT jets: 300-500 GeV - these jets will be part of a background sample used in later studies on top reconstruction.
- Use a large D jet algorithm: D = 1.0
- Look at LAST recombinations in the jet - these are the parts of the substructure that will be tested to determine whether the jet is likely to come from a heavy particle decay.
- Labeling for the last recombination: 1,2 ➜ J

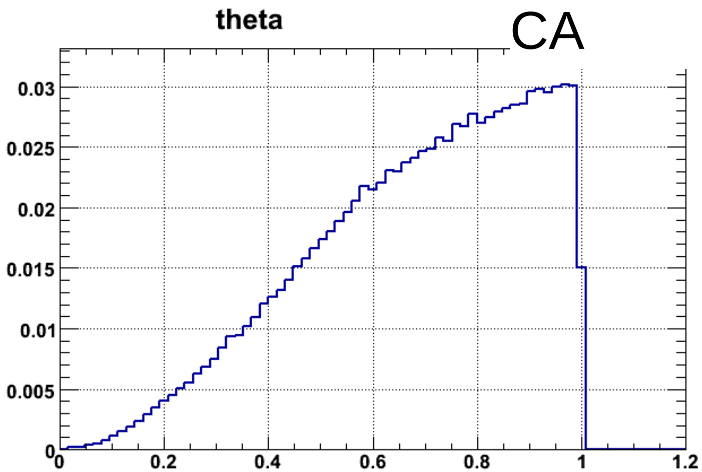
# Systematics of Algorithm: $\theta$

- Consider $\theta$ on LAST recombination for CA and kT (same events, different algorithm)

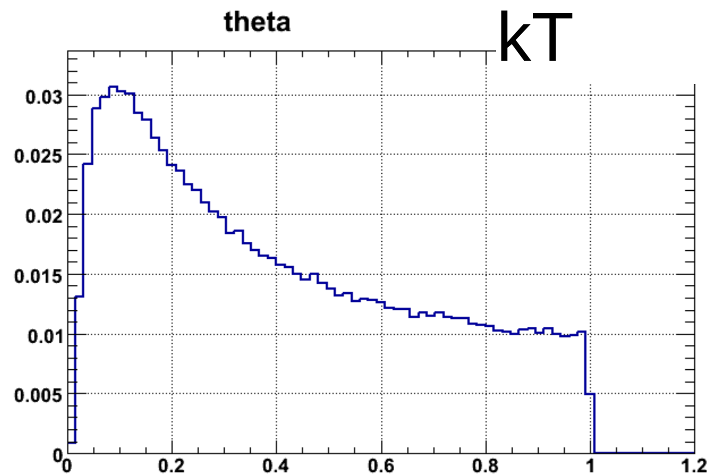
- CA orders only in $\theta$ - means $\theta$ tends to be large (often close to D) at the last merging

- kT orders in $z \cdot \theta$, meaning $\theta$ can be small
  - Get a distribution in $\theta$ that is more weighted towards small $\theta$ than CA

normalized distributions




D

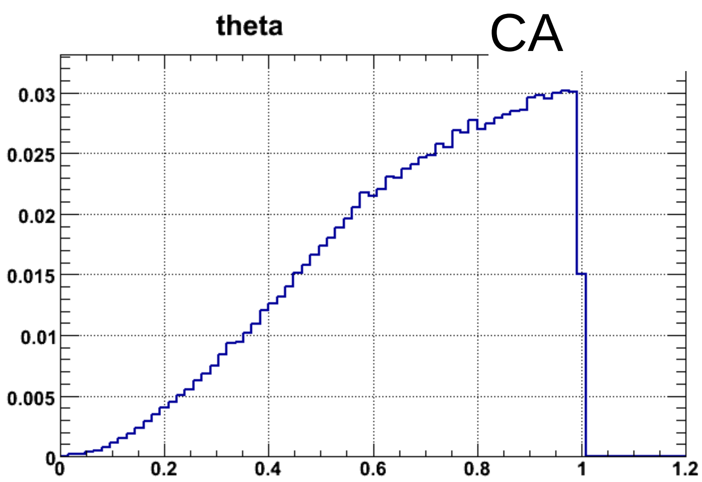


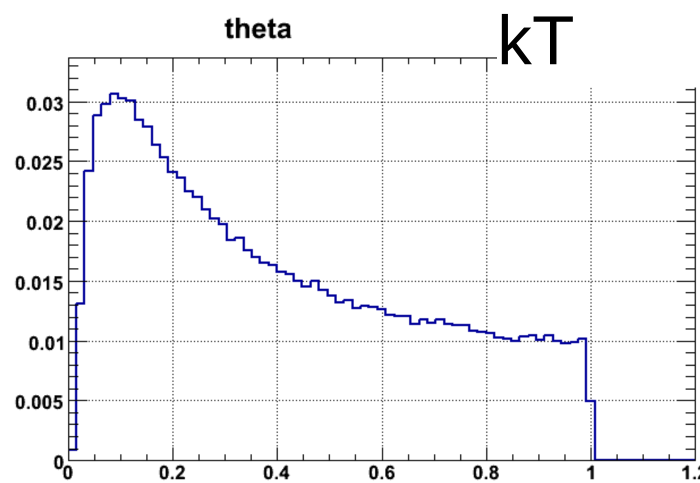

D

QCD

# Systematics of Algorithm: $\theta$ COMPARE

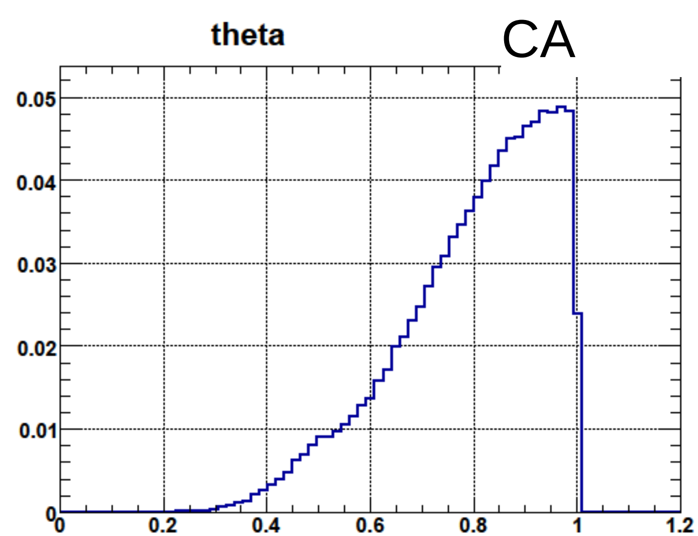
normalized distributions

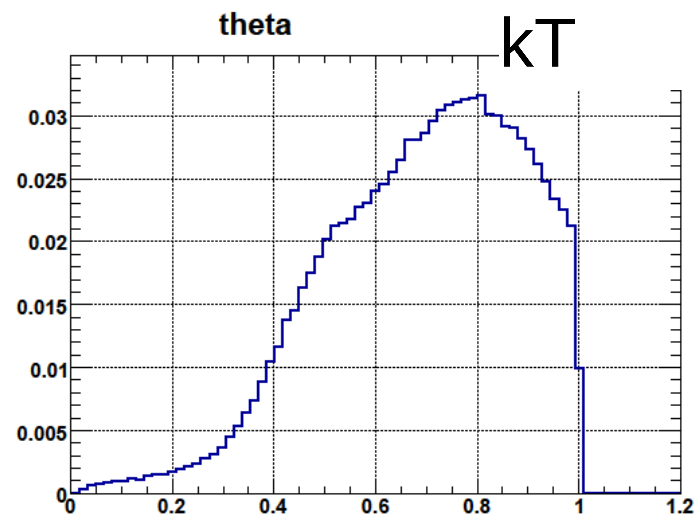



QCD

normalized distributions




ttbar

D

D

# Systematics of Algorithm: z

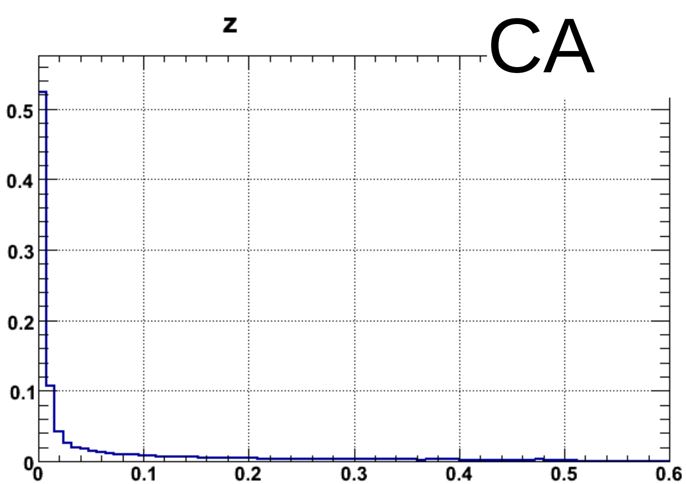
- Consider *z* on LAST recombination for CA and kT.

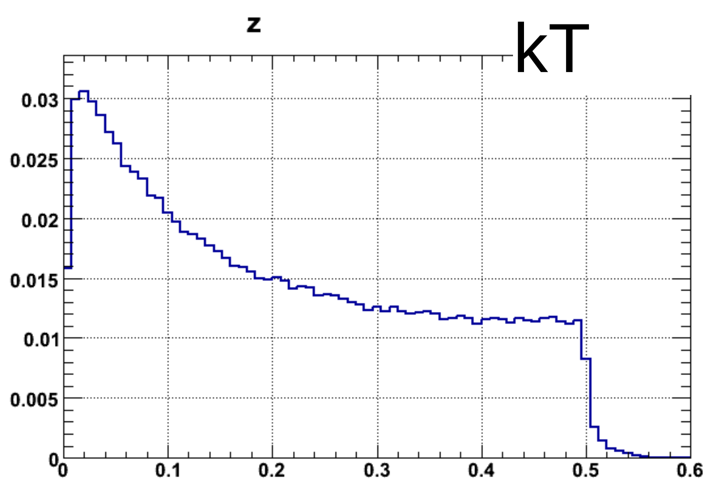
- Metric for CA is independent of z - distribution of *z* comes from the ordering in *θ*

- Periphery of jet is dominated by soft protojets - these are merged early by kT, but can be merged late by CA

- CA has many more low *z*, large *θ* recombinations than kT

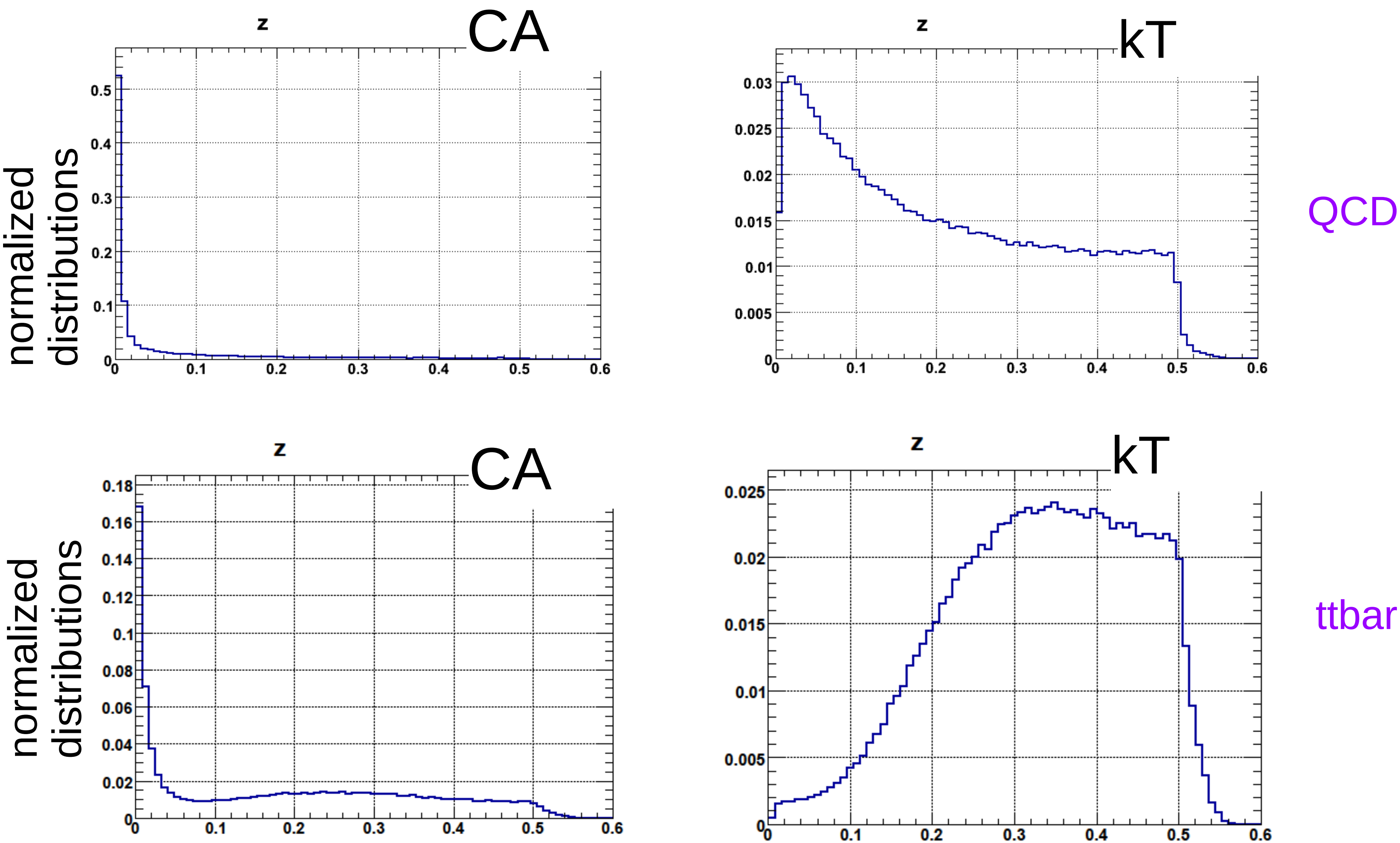
normalized distributions




QCD

# Systematics of Algorithm: z COMPARE



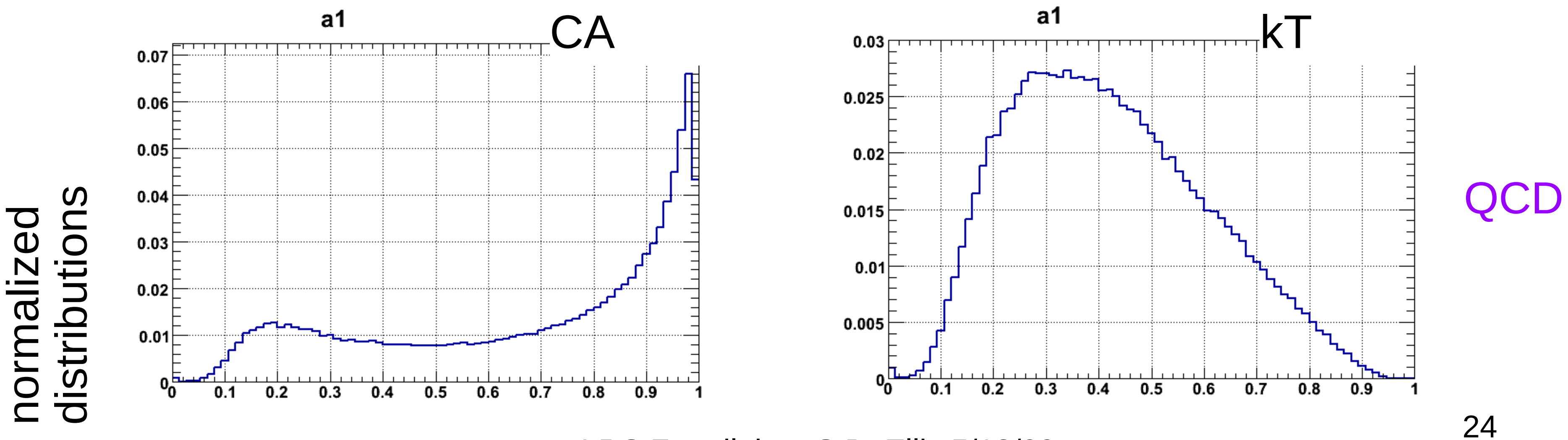
# Systematics of Algorithm: Subjet Masses

- Consider heavier subjet mass at LAST recombination, scaled by the jet mass

$$a_1 = \max(m_1, m_2)/m_j$$

- Last recombinations in CA dominated by small z and large $\theta$
  - Subjet mass for CA is close to the jet mass - $a_1$ near 1

- Last recombinations in kT seldom very soft
  - Subjet mass for kT suppressed for $a_1$ near 1

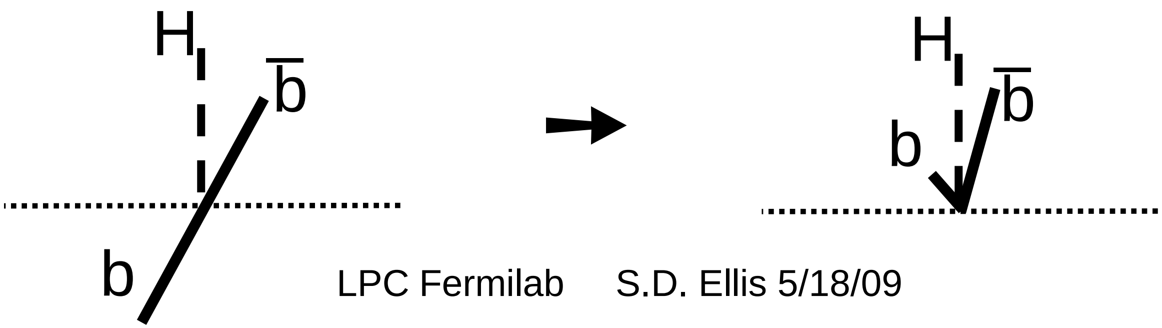


# Systematics in Heavy Particle Reconstruction

- Some kinematic regimes of heavy particle decay have a poor reconstruction rate.

- Example: Higgs decay $H \rightarrow b\bar{b}$ with a very backwards-going b in the Higgs rest frame.
  - The backwards-going b will be soft in the lab frame - difficult to accurately reconstruct.
  - When the Higgs is reconstructed in the jet, the mass distribution is broadened by the likely poor mass resolution.

H rest frame                                        lab frame

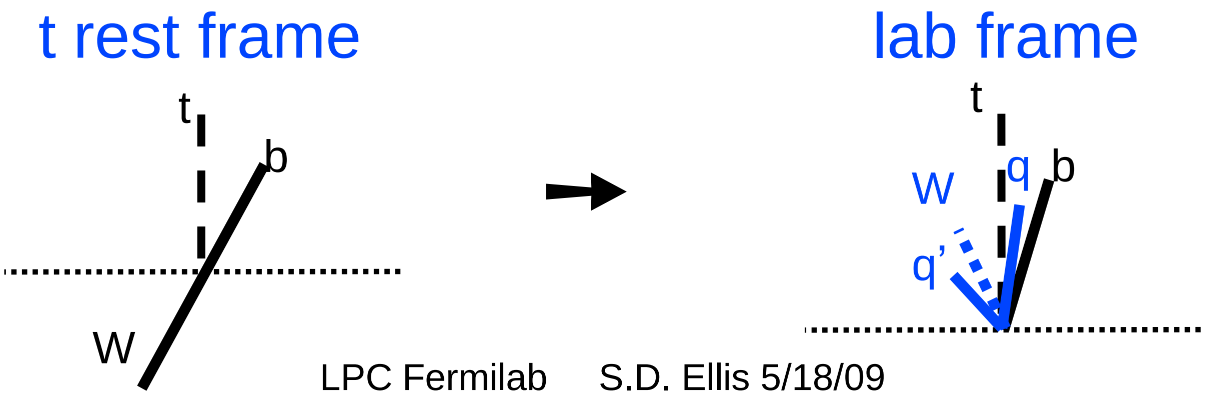


# Systematics in Heavy Particle Reconstruction

- In multi-step decays, kinematic constraints are more severe.

- Example: hadronic top decay with a backwards going W in the top rest frame
  - In the lab frame, the decay angle of the W will typically be larger than the top quark.
  - This geometry makes it difficult to reconstruct the W as a subjet - even at the parton level!

    - One of the quarks from the W will be soft - can mispair the one of the quarks from the W with the b, giving inaccurate substructure



# Summary: Reconstructed Heavy Particles

- Decays resulting in soft (in Lab) partons are less likely to be accurately reconstructed
  - Soft partons are poorly measured → broader jet, subjet mass distributions
  - Soft partons are often recombined in wrong order → inaccurate substructure
- Small z recombinations often arise from
  - Uncorrelated ISR, FSR
  - Underlying event or pile-up contributions
  - → Not indicative of a correctly reconstructed heavy particle –

⇒ Can the jet substructure be modified to reduce the effect of soft recombinations?

# Pruning the Jet Substructure

- Soft, large angle recombinations
  - Tend to degrade the signal (real decays)
  - Tend to enhance the background (larger QCD jet masses)
  - Tend to arise from uncorrelated physics
- This is a generic problem for searches - try to come up with a generic solution

$\Rightarrow$ PRUNE these recombinations and focus on masses

others have tried similar ideas - Salam/Butterworth (Higgs), Kaplan (tops), Thaler/Wang (tops)

# Pruning reveals for hidden truth -

# Pruning :

Procedure:

- Start with the objects (e.g. towers) forming a jet found with a recombination algorithm

- Rerun the algorithm, but at each recombination test whether:
  - $z < z_{cut}$ and $\Delta R_{ij} > D_{cut}$ ($\theta_J$ is angle at final recombination in original found jet)

CA: $z_{cut} = 0.1$ and $D_{cut} = \theta_J/2$

kT: $z_{cut} = 0.15$ and $D_{cut} = \theta_J/2$

- If true (a soft, large angle recombination), prune the softer branch by NOT doing the recombination and discarding the softer branch

- Proceed with the algorithm

$\Rightarrow$ The resulting jet is the pruned jet

# You will see that

- Pruning narrows peaks in jet and subjet mass distributions of reconstructed top quarks

- Pruning improves both signal purity (*R*) and signal-to-noise (*S*) in top quark reconstruction using a QCD multijet background

- The *D* dependence of the jet algorithm is reduced by pruning - the improvements in *R* and *S* using an optimized *D* exhibit only small improvement over using a constant *D* = 1.0 with pruning

- A simple pruning procedure based on *D* = 1.0 CA (or kT) jets can
  - Enhance likelihood of success of heavy particle searches
  - Reduce systematic effects of the jet algorithm, the UE and PU
  - Cannot be THE answer, but part of the answer, e.g., use with b-tagging, require correlations with other jets/leptons (pair production)

# You will conclude that:

- Generic techniques like pruning can help:
  - Reduce widths of bumps in mass distributions of reconstructed heavy particles
  - Improve the power of single-jet heavy particle searches
  - Reduce systematics from the algorithm and sensitivity to soft, uncorrelated physics

- Systematics of the jet algorithm are important in studying jet substructure
  - The jet substructure we expect from the kT and CA algorithms are very different
  - Shaping can make it difficult to determine the physics of a jet

- Should certify *pruning* by finding tops, *W*'s and *Z*'s in single jets in early LHC running (or with Tevatron data)

# For the Future (this week?)

- Many questions remain about using jet substructure:
  - How does the detector affect jet substructure and the systematics of the algorithm? How does it affect techniques like pruning? What are actual mass uncertainties?
  - Which kinematic variables best discriminate between QCD and non-QCD jets? How powerful are these variables?
  - How can jet substructure fit into an analysis? How orthogonal is the information provided by jet substructure to other data from the event?
  - How can theory calculations link up with experimental observations about jet substructure?
  - Will matched MC data sets change our conclusions?

- Jet substructure shows promise to learn about the physics behind jets, but there is still much to discover

# Extra Detail Slides

# Test Pruning:

- Study of top reconstruction:
  - Hadronic top decay as a surrogate for a massive particle produced at the LHC
  - Use a QCD multijet background - separate (unmatched) samples from 2, 3, and 4 hard parton MEs
  - ME from MadGraph, showered and hadronized in Pythia (DWT tune), jets found with homemade code

- Look at several quantities before/after pruning:

$\Rightarrow$ Mass resolution of reconstructed tops (width of bump), small width means smaller background contribution

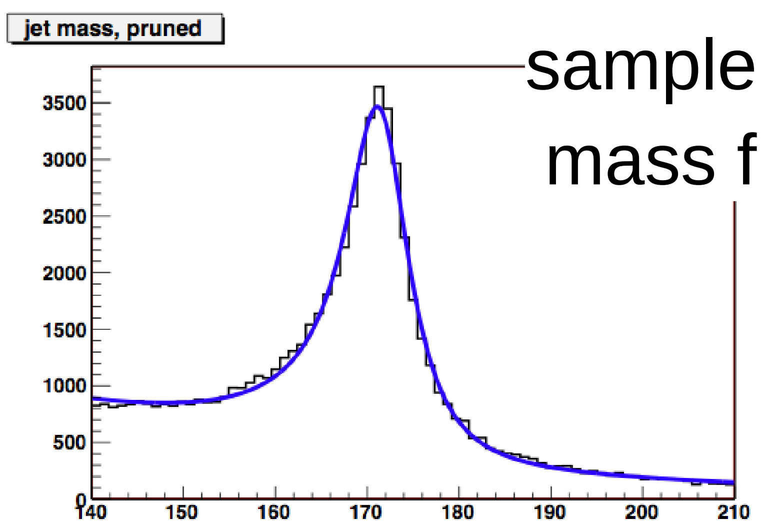
  - pT dependence of pruning effect

  - Dependence on choice of jet algorithm and angular parameter D

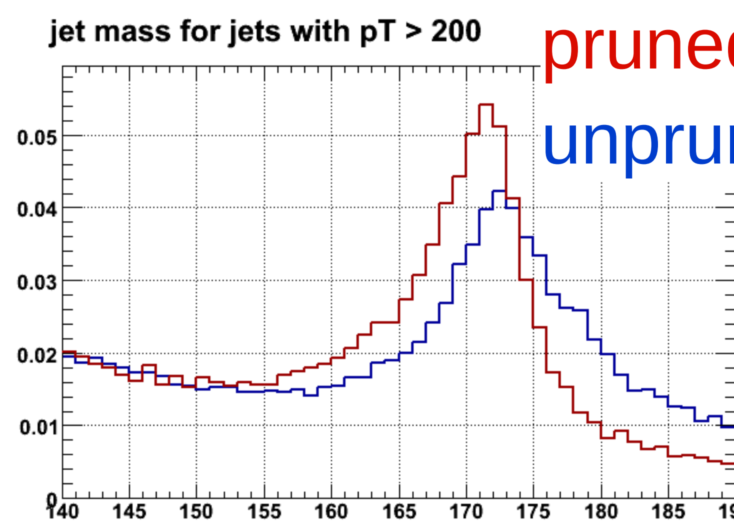
# Defining Reconstructed Tops – Search Mode

- A jet reconstructing a top will have a mass within the top mass window, and a primary subjet mass within the W mass window - call these jets top jets

- Defining the top, W mass windows:
  - Fit the jet mass and subjet mass distributions with (asymmetric) Breit-Wigner plus continuum → widths of the peaks
  - The top and W windows are defined separately for pruned and not pruned - test whether pruning is narrowing the mass distribution



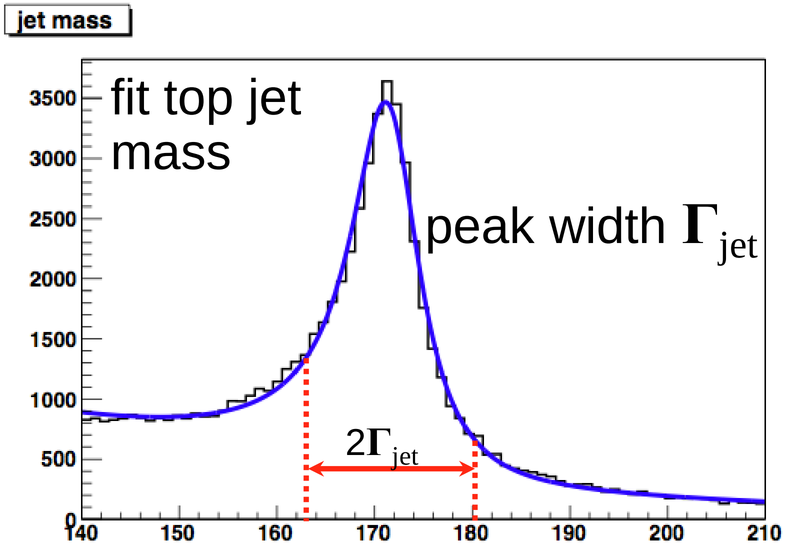

sample mass fit




pruned
unpruned

# Defining Reconstructed Tops

fit mass windows to identify
a reconstructed top quark




peak function: skewed Breit-Wigner

$$M^2\Gamma^2 \frac{[a+b(m-M)]}{(m^2-M^2)^2+M^2\Gamma^2}$$

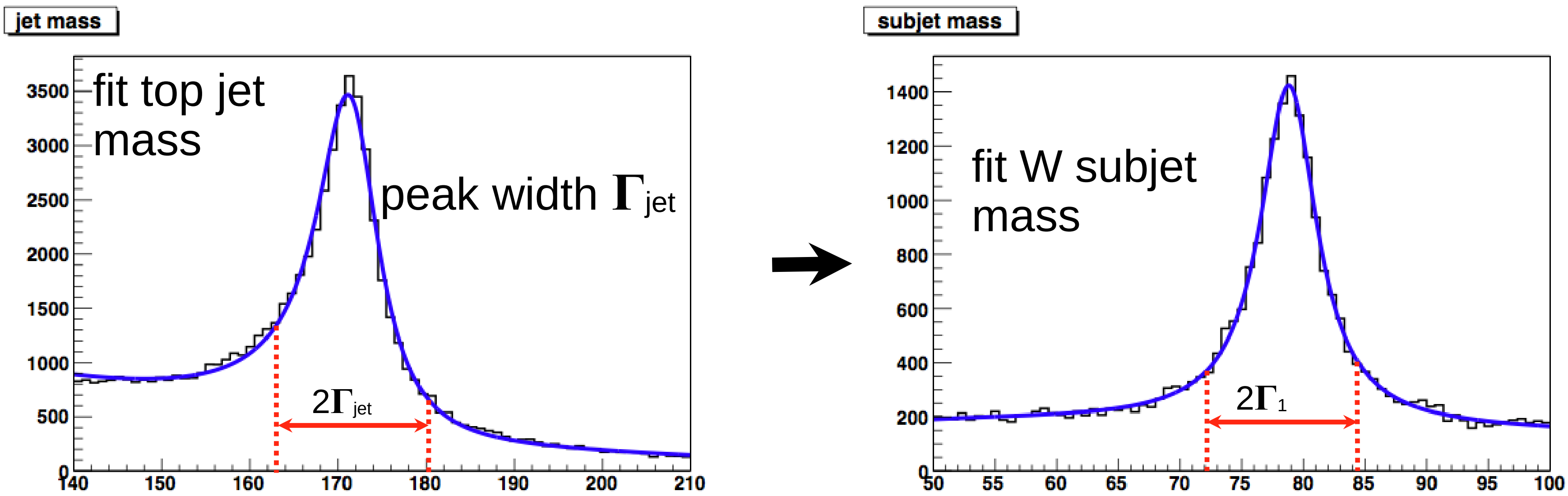
plus continuum background distribution

$$\frac{c}{m}+\frac{d}{m^2}$$

# Defining Reconstructed Tops

fit mass windows to identify a reconstructed top quark

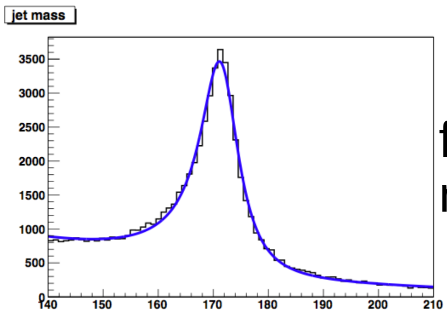
cut on masses of jet (top mass) and subjet (W mass)



# Defining Reconstructed Tops

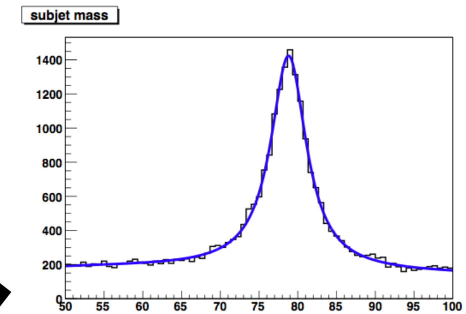
fit mass windows to identify a reconstructed top quark

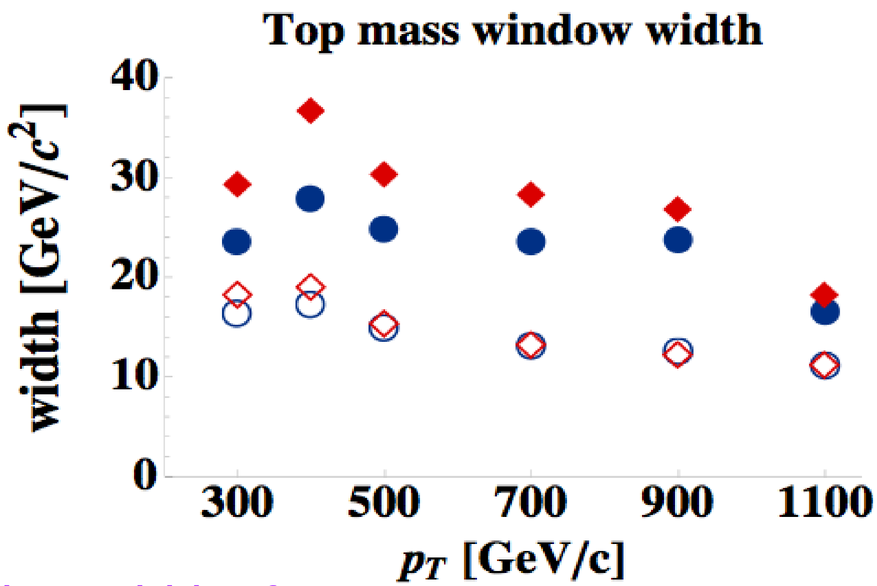
cut on masses of jet (top mass) and subjet (W mass)

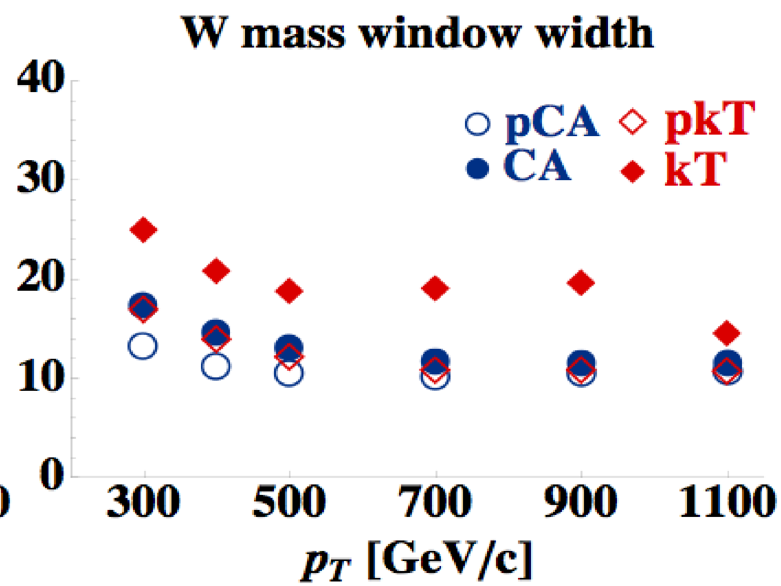



fit top jet mass

fit W subjet mass



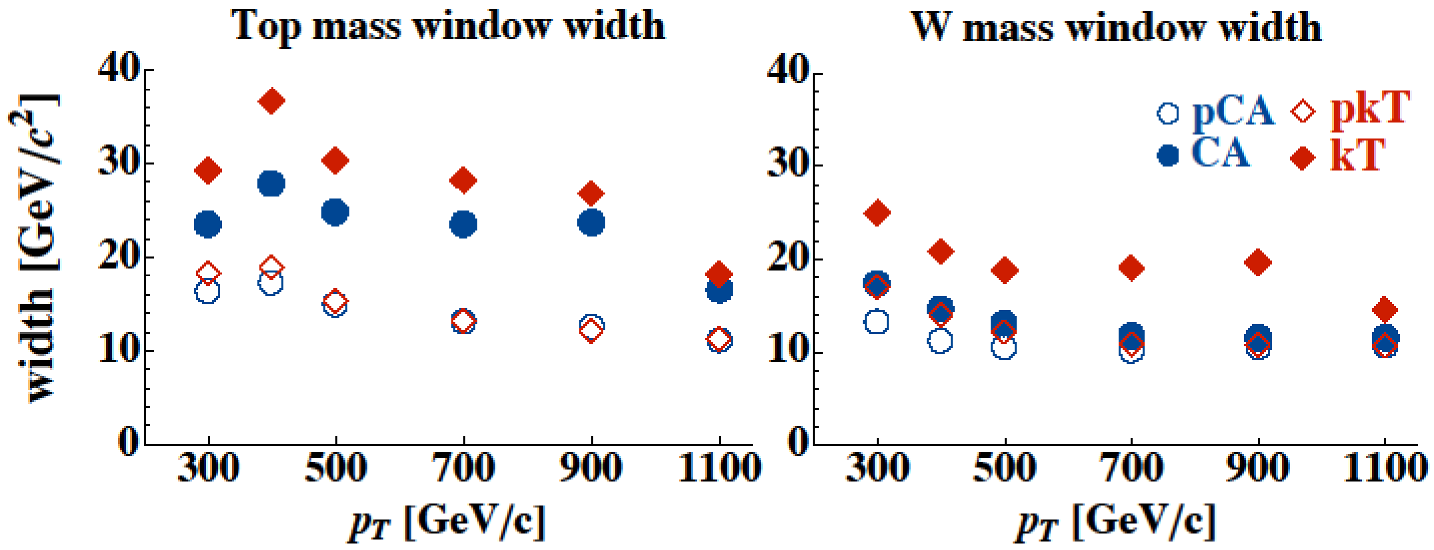

window widths for pruned (pX) and unpruned jets

# Mass Windows and Pruning - Summary

- Fit the top and W mass peaks, look at window widths for unpruned and pruned (pX) cases in (100 - 200 GeV wide) pT bins

⇒ Pruned windows narrower, meaning better mass bump resolution - better heavy particle ID

⇒ Pruned window widths fairly consistent between algorithms (not true of unpruned), over the full range in pT



# Statistical Measures:

- Count top jets in signal and background samples
  - $N_S$ : number of top jets in signal sample
  - $N_B$ : number of top jets in background sample
  - $A$ : unpruned algorithm $\quad pA$ : pruned algorithm

# Statistical Measures:

- Count top jets in signal and background samples
  - $N_S$ : number of top jets in signal sample
  - $N_B$ : number of top jets in background sample
  - $A$ : unpruned algorithm    $pA$ : pruned algorithm

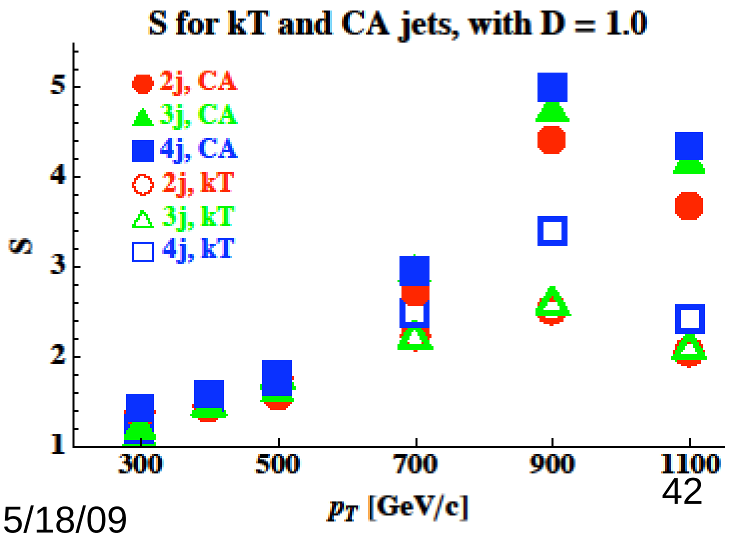
- Have compared pruned and unpruned samples with 3 measures:
  - *ε*, *R*, *S* - efficiency, Sig/Bkg, and Sig/Bkg$^{1/2}$

$$\epsilon = \frac{N_S(pA)}{N_S(A)} \qquad R = \frac{N_S(pA)/N_B(pA)}{N_S(A)/N_B(A)} \qquad S = \frac{N_S(pA)/\sqrt{N_B(pA)}}{N_S(A)/\sqrt{N_B(A)}}$$

Here focus on *S*

*S* > 1 (improved likelihood to see bump if prune), all pT, all bkgs, both algorithms

Turns over at large pT where top decay becomes very narrow

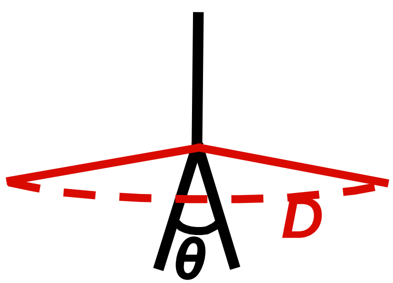


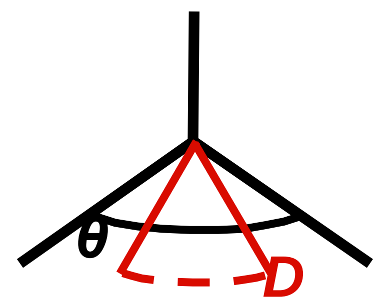
# Heavy Particle Decays and *D* *(Variable D - see Thaler)*

- Heavy particle ID with the unpruned algorithm is improved when *D* is matched to the expected average decay angle

- Rule of thumb: $\theta$ = 2m/pT

- Two cases:




*D* > $\theta$

- lets in extra radiation
- QCD jet masses larger

*D* < $\theta$

- particle will not be reconstructed

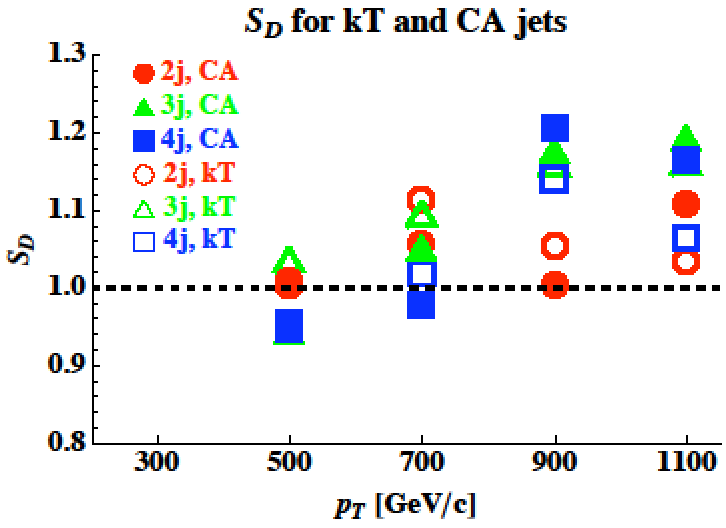
# Improvements in Pruning

- Optimize *D* for each pT bin: *D* = min(2m/pT, 1.0) ⇒ (1.0,1.0,0.8,0.6,0.5,0.4) for our pT bins
- Pruning still shows improvements
- How does pruning compare between fixed *D* = 1.0 and *D* optimized for each pT bin ⇒ $S_D = S_{D\ opt}/S_{D=1}$?

⇒ Little further improvement obtained by varying *D*

⇒ $S_D$ = 1 in first 2 bins

⇒ Pruning with Fixed *D* does most of the work

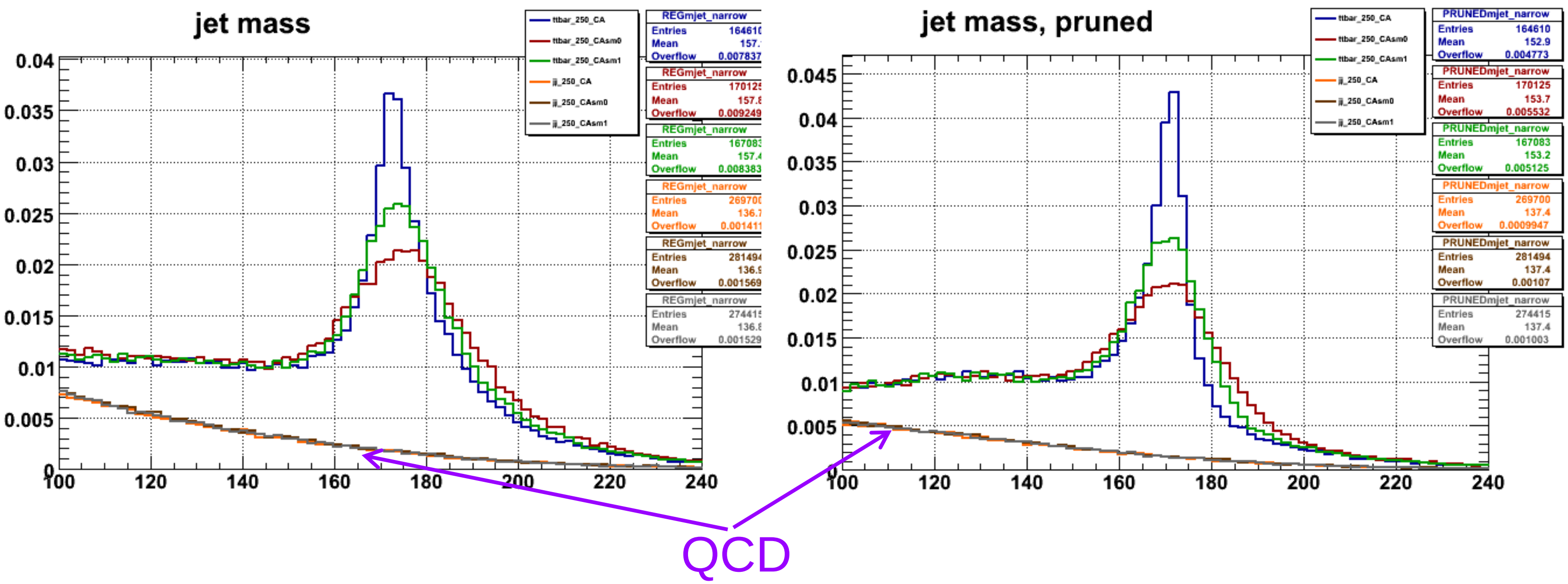


# Consider impact of (Gaussian[1]) smearing

Smear energies in “calorimeter cells” with Gaussian width (300 GeV/c < pT < 500 GeV/c)

$$\sigma_{E,0} = \sqrt{E + 0.01E^2} \ \text{(worst, red curve)} \ [\text{blue curve } \sigma_E = 0]$$

$$\sigma_{E,1} = \sqrt{(0.65)^2 E + (0.05)^2 E^2} \ \text{(realistic, green curve)}$$




⇒ Pruning still helps (pruned peaks are more narrow), but impact is degraded by detector smearing

[1] From P. Loch

# Statistical Measures:

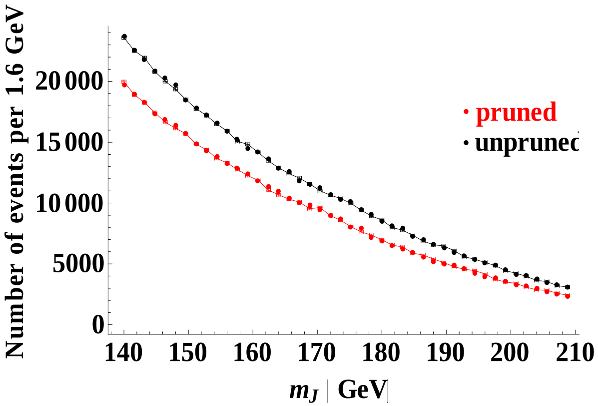
| | | $\varepsilon$ | *R* | *S* |
|---|---|---|---|---|
| No Smearing | pCA/CA | 0.90 | 2.25 | 1.42 |
| | pkT/kT | 0.68 | 3.01 | 1.44 |
| Reasonable Smearing | pCA/CA | 0.98 | 1.75 | 1.31 |
| | pkT/kT | 0.72 | 2.20 | 1.26 |
| Worst Smearing | pCA/CA | 1.00 | 1.59 | 1.26 |
| | pkT/kt | 0.74 | 2.00 | 1.22 |

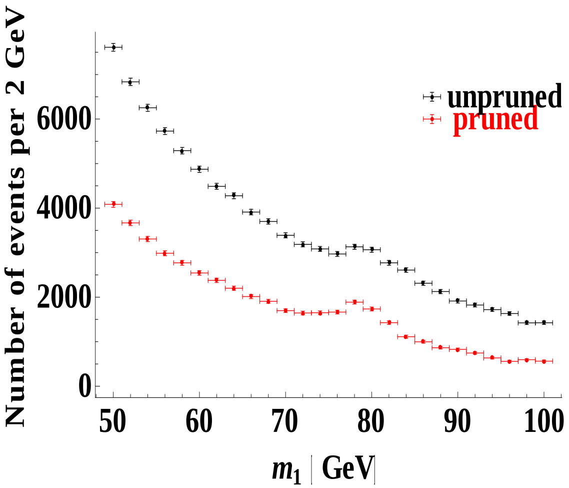
$$\epsilon = \frac{N_S(pA)}{N_S(A)} \qquad R = \frac{N_S(pA)/N_B(pA)}{N_S(A)/N_B(A)} \qquad S = \frac{N_S(pA)/\sqrt{N_B(pA)}}{N_S(A)/\sqrt{N_B(A)}}$$

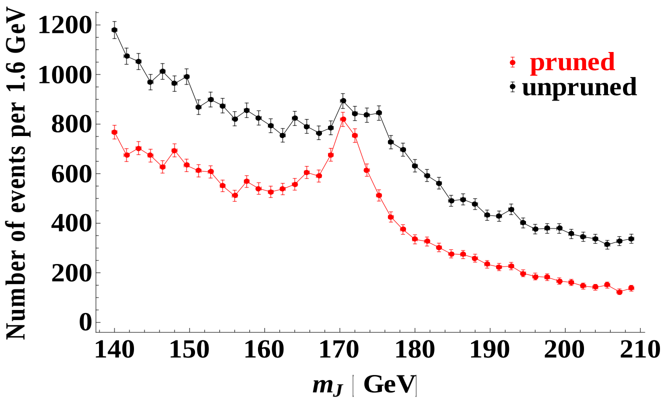
$\Rightarrow$ Smearing degrades but does not eliminate the value of pruning

# "Simulated" data plots (Peskin plots)

- Include signal (tops) and bkg (QCD) with correct ratio and "simulated" statistical uncertainties and fluctuations, corresponding to 1 $fb^{-1}$ (300 GeV/c < pT < 500 GeV/c)



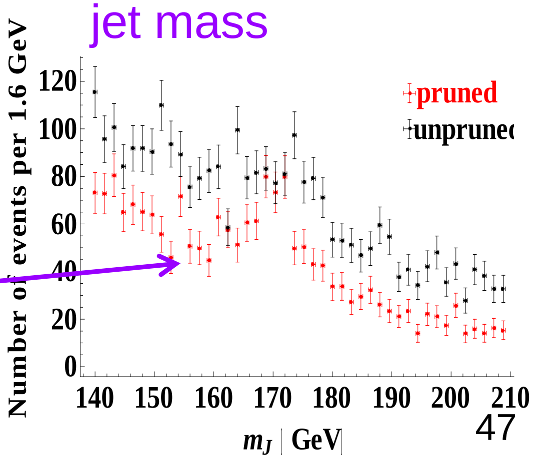

Find (small) mass bump and cut on it

Find daughter mass bump and cut on it

Now a clear signal in jet mass

Pruning enhances the signal, but its still tough in a real search

For known top quark, pruning + 100 $pb^{-1}$ may be enough (especially with b tags)



# Compare to other "Jet Grooming" – CA jets

- PSJ (Kaplan, et al., for tops) – find primary subjets and build "groomed" jet from these (3 or 4 of them)

1. Define $\delta_p = \frac{\min\left[p_{T1}, p_{T2}\right]}{p_{T,J}}$ , $\delta_{p,\mathrm{MIN}} = 0.1\left(p_T < 800 \text{ GeV/c}\right), 0.05\left(p_T > 800 \text{ GeV/c}\right)$

   $\delta_R = \left|\Delta\eta_{12}\right| + \left|\Delta\phi_{12}\right|$ , $\delta_{R,\mathrm{MIN}} = 0.19$
2. Start of top of branch (the jet) and follow hardest daughter at each branching (discarding softer daughters) until reach first branching where $\delta_p > \delta_{p,\mathrm{MIN}}, \delta_R > \delta_{R,\mathrm{MIN}}$ . If does not exist, discard jet.
3. If such a branching exists, start again with each daughter of this branching as top branch as in 2. Again follow along the hardest daughter (discarding softer daughters) until a branching where $\delta_p > \delta_{p,\mathrm{MIN}}, \delta_R > \delta_{R,\mathrm{MIN}}$ . If present, the daughters of this ($2^{nd}$) hard branching are primary subjets. If not present, the original daughter is primary subjet. This can yield 2, 3 or 4 primary subjets.
4. Keep only 3 and 4 subjet cases and recombine the subjets with CA algorithm.

# Compare to other "Jet Grooming" – CA jets

- MDF (Butterworth, et al., for Higgs) – find primary subjets and build "groomed" jet from these (2 or 3 of them)

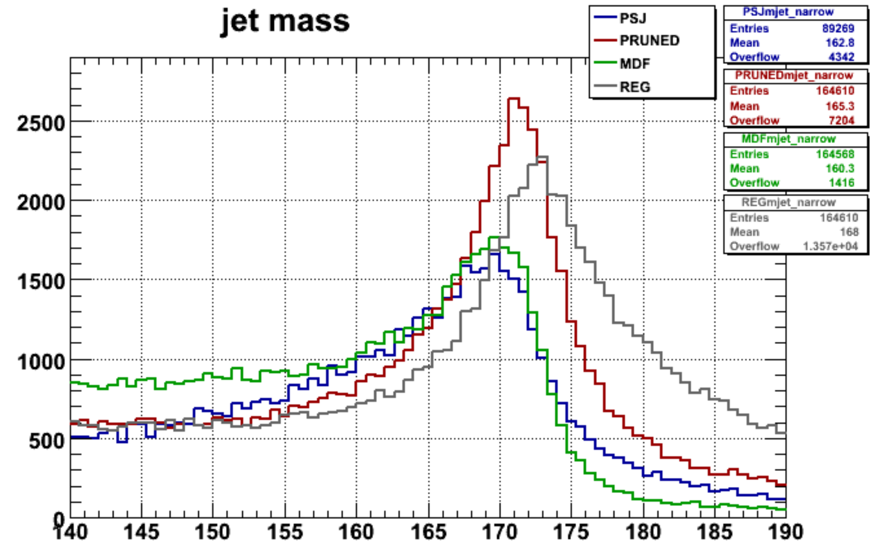
1. For each p → 1,2 branching define $a_1 = \frac{\max[m_1, m_2]}{m_p}$, $\mu = 0.67$

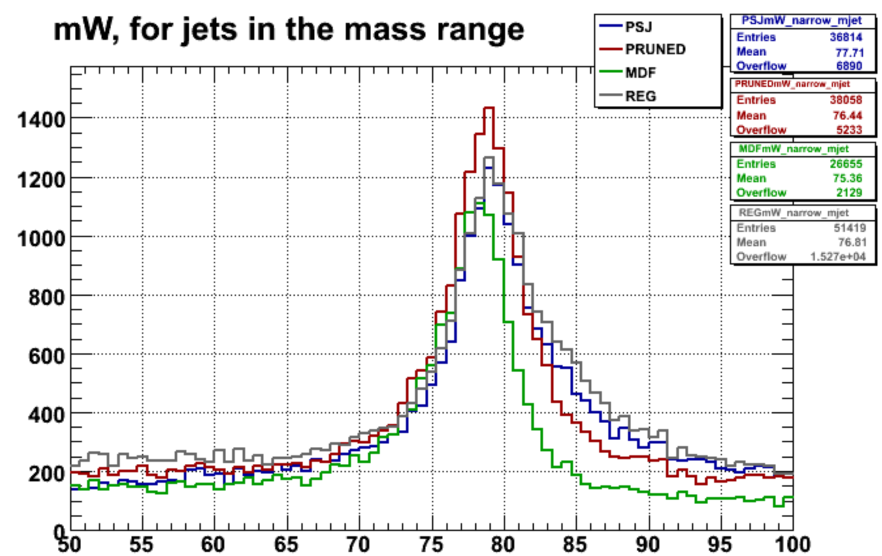
   $y = \frac{\min\left[p_{T,1}^2, p_{T,2}^2\right]}{m_J^2} \Delta R_{12}^2$, $y_{\text{cut}} = 0.09$

2. Start of top of branch (the jet) and follow hardest daughter at each branching (discarding softer daughters) until reach first branching where $a_1 < \mu, y > y_{\text{cut}}$. If does not exist, discard jet.
3. If such a branching exists, define $\Delta R_{bb} = \Delta R_{12}, D_{\text{filt}} = \min\left[0.3, \Delta R_{bb}/2\right]$ and start again with each daughter of this branching as top branch as in 2. Again follow along the hardest daughter (discarding softer daughters) until a branching where $\Delta R < D_{\text{filt}}$, (but $\Delta R > D_{\text{filt}}$ for early branchings). If present, the daughters of this (2nd) hard branching are primary subjets. If not present, the original daughter is primary subjet. This can yield 2, 3 or 4 primary subjets.
4. Keep the 3 hardest subjets (discard 1 subjet case but keep if only 2). Recombine the (2 or) 3 subjets with CA algorithm.

# Plots – first look

- Pruning yields comparable or narrower “bumps” in mass distributions
- Pruning yields comparable or better numbers for $\varepsilon$, *R* and *S*
- Suggests pruning is as effective and generally simpler than other methods



# Statistical Measures:

| | | $\varepsilon$ | $R$ | $S$ |
|---|---|---|---|---|
| 300 GeV/c < pT < 500 GeV/c | pCA/CA | 0.90 | 2.25 | 1.42 |
| | PSJCA/CA | 0.87 | 1.49 | 1.14 |
| | MDFCA/CA | 0.65 | 2.64 | 1.31 |
| 800 GeV/c < pT < 1000 GeV/c | pCA/CA | 2.40 | 8.11 | 4.41 |
| | PSJCA/CA | 2.24 | 8.72 | 4.42 |
| | MDFCA/CA | 2.91 | 3.63 | 3.25 |

$$\epsilon = \frac{N_S(pA)}{N_S(A)} \qquad R = \frac{N_S(pA)/N_B(pA)}{N_S(A)/N_B(A)} \qquad S = \frac{N_S(pA)/\sqrt{N_B(pA)}}{N_S(A)/\sqrt{N_B(A)}}$$

$\Rightarrow$ Pruning is comparable or slightly better than other grooming techniques

# Aside: Rest Frame variables

- Pruning removes branchings ("decays") with
  - $\cos\theta_0 > 0.8$, (heavier daughter forward) most subjet masses
  - $\cos\theta_0 < -0.8$, (heavier daughter backward) small daughter masses only (both daughters $a_2 < a_1 < 0.3$)