



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

Steve Ellis, Jon Walsh and Chris Vermilion

0903.5081

0907.XXXX

- go to [tinyurl.com/jetpruning](http://tinyurl.com/jetpruning)

## Big Picture:

The LHC will be both very exciting and very challenging –

- most of the data will be about hadrons (jets)
- many interesting objects ( $W$ 's,  $Z$ 's, tops, SUSY particles) will be **boosted** enough to appear in single jet
- must be able to ID/reconstruct these jets to find the BSM physics



Department of Physics  
University of Washington

Giving New Physics a Boost - 2009  
SLAC 7/09/09



# Outline & Issues

- Brief review of (QCD) jets, including masses
- *Search* for BSM physics in SINGLE jets, want generic techniques
  - bumps in jet mass distributions
- ⇒ Large but *Smooth* QCD background
- Consider Recombination (kT) jets → natural substructure but also
  - algorithm systematics (shaping of distributions)
  - 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



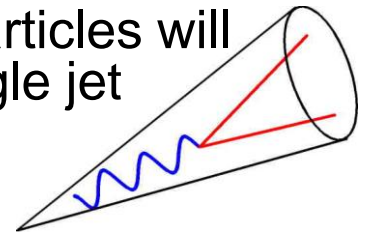
## Why JETS?

Essentially all LHC events involve an important hadronic component, only  $Z' \rightarrow \mu^+ \mu^-$  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 ...



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

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

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

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

If  $k_{T,(ij)}^2$  is the minimum, merge pair and redo list;

If  $k_{T,i}^2$  is the minimum  $\rightarrow i$  is a jet!

(no more merging for  $i$ ), 1 parameter  $D$  (NLO, equals cone for  $D = R$ ,  $R_{\text{sep}} = 1$ )

$\alpha = 1$ , 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 (Cacciari, Salam & Soyez) goes like N In N (only recalculate nearest neighbors), plus has scheme for doing UE correction



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

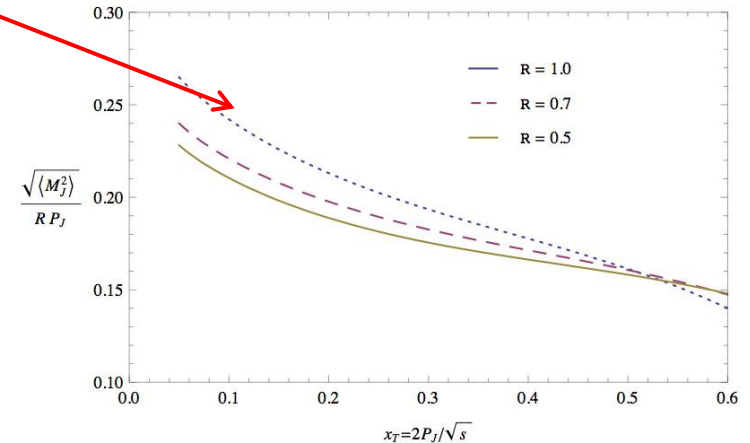
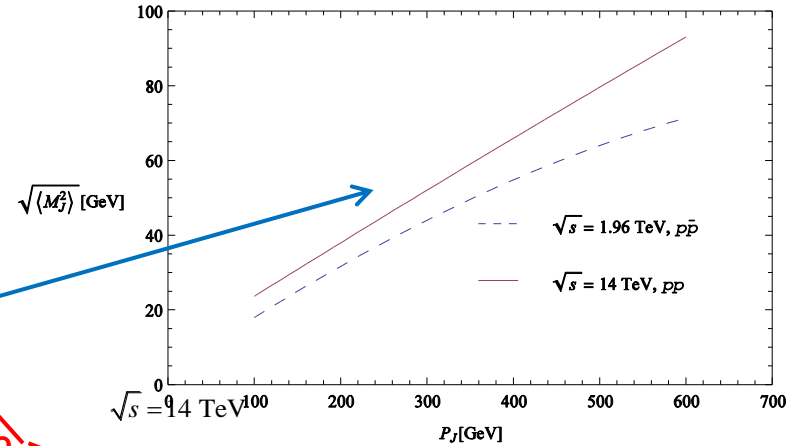
- In NLO PertThy

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

Phase space from pdfs,  
 $f \sim 1$  & const

Dimensions

Jet Size,  $R = D \sim \Delta\theta$ , determined by jet algorithm



Useful QCD “Rule-of-Thumb”  $\Rightarrow \sqrt{\langle M^2 \rangle_{NLO}} \sim 0.2 p_J R (1 \pm 0.25)$



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

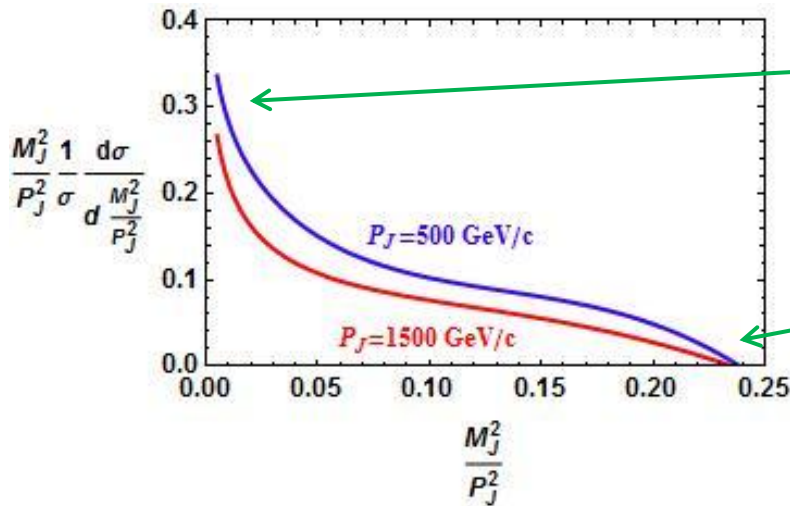
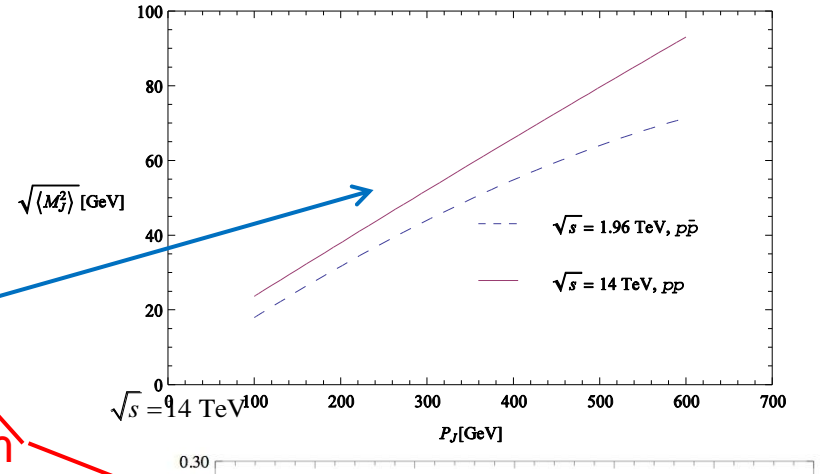
- In NLO PertThy

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

Phase space from pdfs,  $f \sim 1$  & const

Dimensions

Jet Size,  $R = D \sim \Delta\theta$ , determined by jet algorithm



Peaked at low mass  
(log(m)/m behavior),

cuts off for  $(M/P)^2 > 0.25 \sim R^2/4$   
( $M/P > 0.5$ ) large mass can't fit in  
fixed size jet, QCD suppressed for  
 $M/P > 0.3$

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

Useful QCD "Rule-of-Thumb"  $\Rightarrow \sqrt{\langle M^2 \rangle}_{NLO} \sim 0.2 p_J R (1 \pm 0.25)$



# Finding Heavy Particles with Jets - Issues



QCD multijet production rate  $\gg$  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



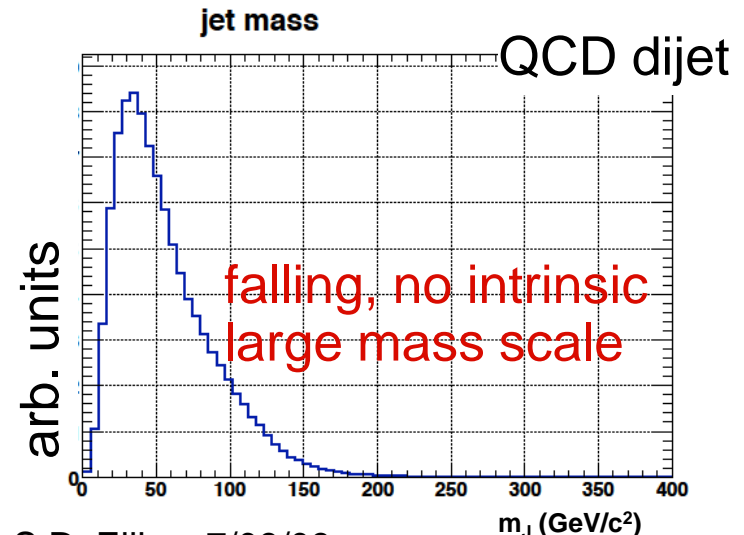
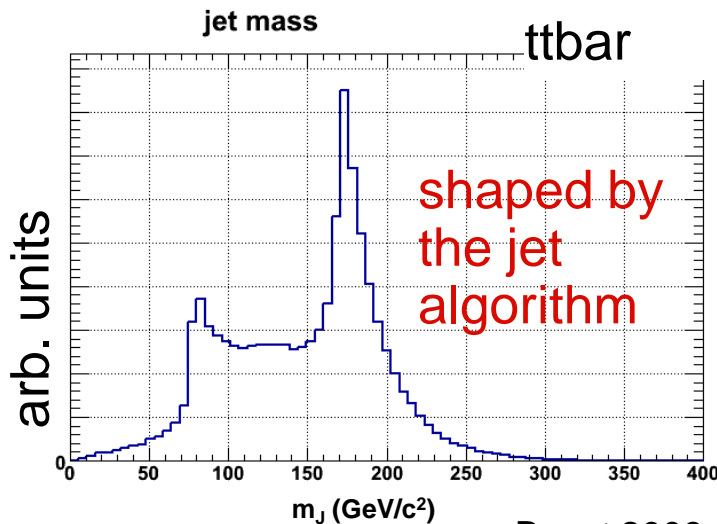
Use jet substructure as defined by recombination algorithms to refine jets



Algorithm will systematically shape distributions

- Use top quark as surrogate new particle.

$$\sigma_{t\bar{t}} \approx 10^{-3} \sigma_{jj}$$

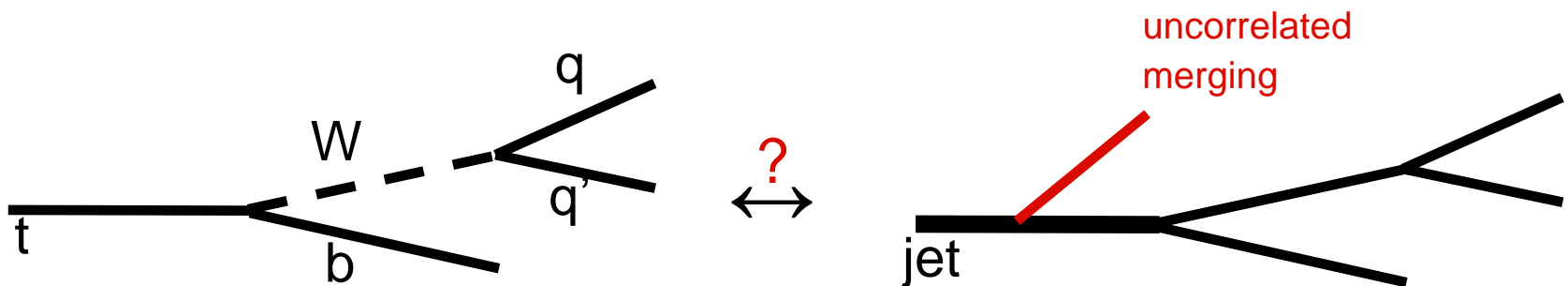




# Reconstruction of Jet Substructure – QCD vs Heavy Particle

- Want to identify a heavy particle reconstructed in a single jet.
  - Need correct ordering in the substructure and accurate reconstruction (to obtain masses accurately)
  - Need to understand how decays and QCD differ in their expected substructure, e.g., distributions at branchings.

⇒ But jet substructure affected by the systematics of the algorithm, and by kinematics when jet masses/subjet masses are fixed.





# Systematics of the Jet Algorithm

- Consider generic recombination step:  $i, j \rightarrow p$
- Useful variables:  $z = \frac{\min(p_{T_i}, p_{T_j})}{p_{T_p}}$   $\theta = \Delta R_{ij}$   
(Lab frame)
- Merging metrics:  $\rho_{\text{kT}} = p_{T_p} z \theta / D$   $\rho_{\text{CA}} = \theta / D$
- 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.

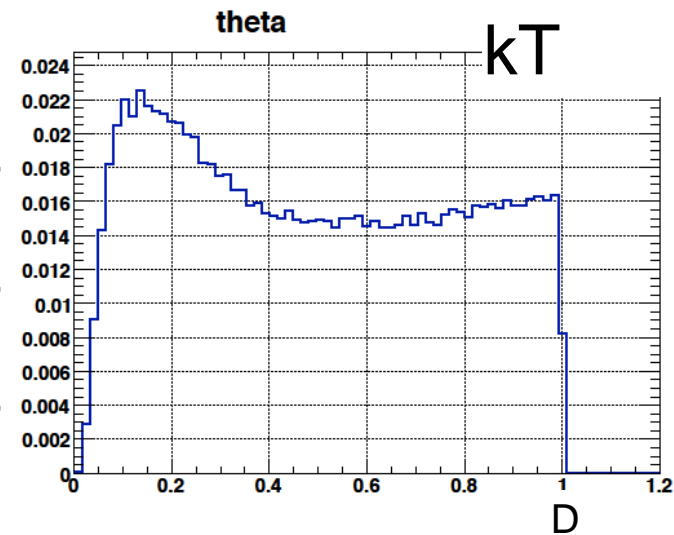
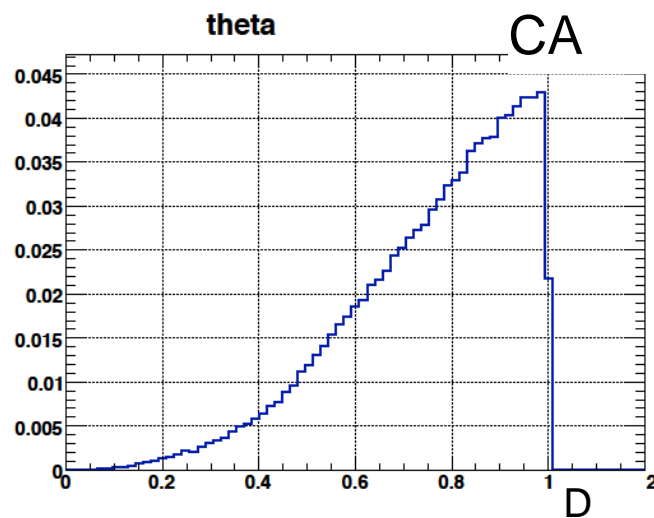


# Systematics of Algorithm: $\theta$

MadGraph/PYTHIA (DWT tune) data

- Consider  $\theta$  of **LAST** recombination for CA and kT (same events, different algorithm) for QCD dijets ( $200 < p_T < 500$ ),  $D = 1$
- 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

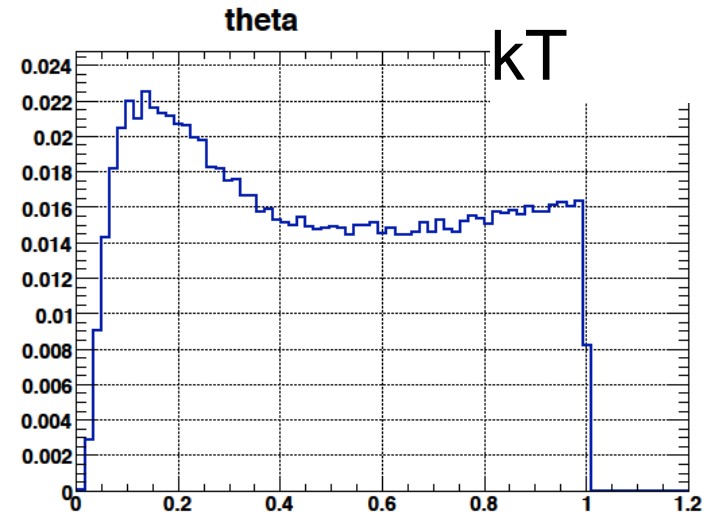
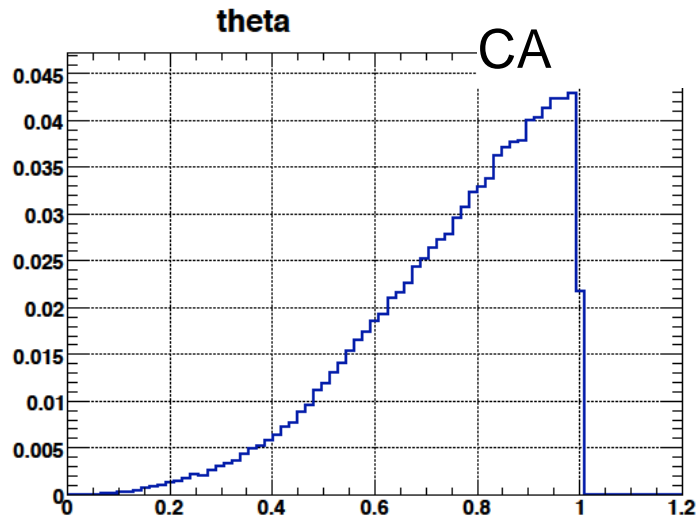


QCD



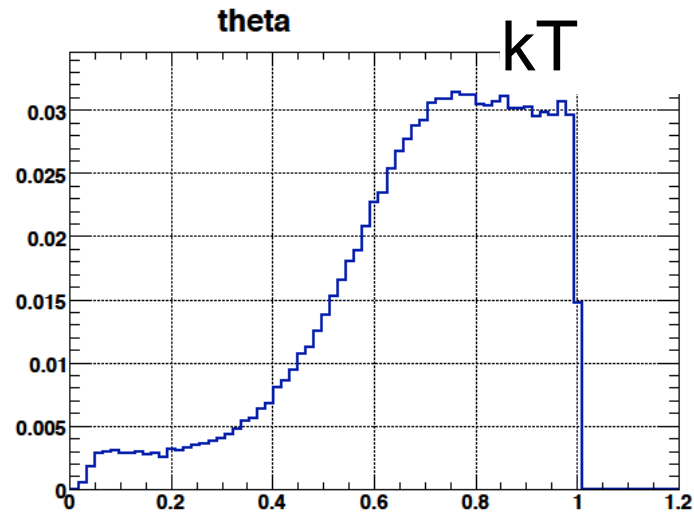
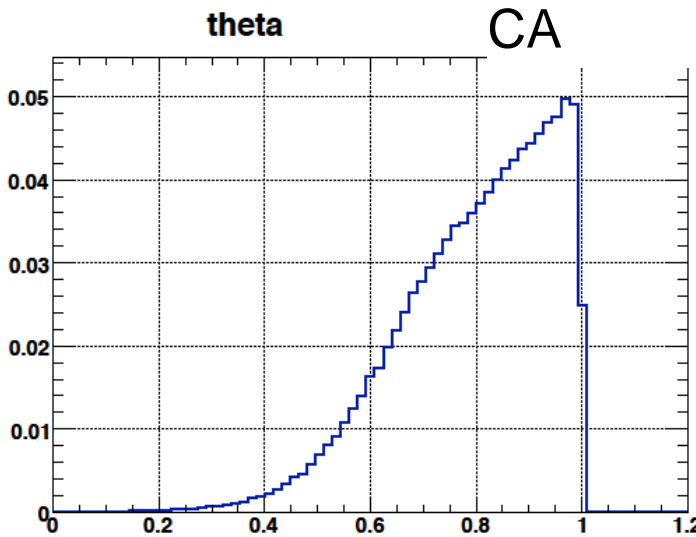
# Systematics of Algorithm: $\theta$ COMPARE

normalized  
distributions



QCD

normalized  
distributions



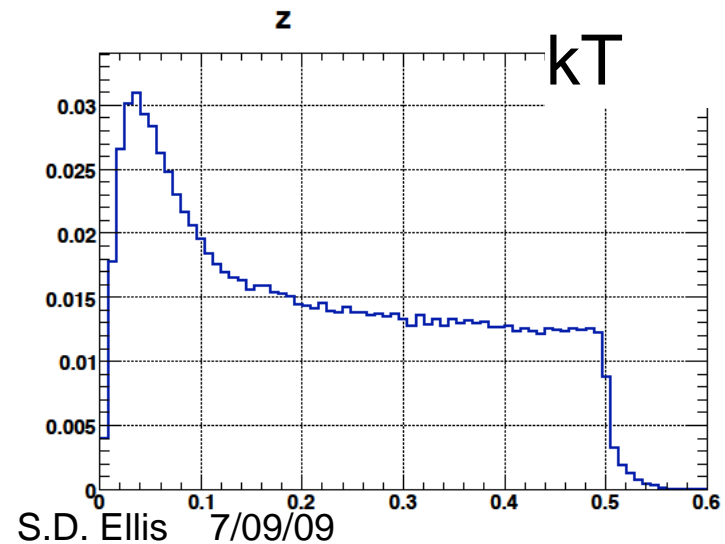
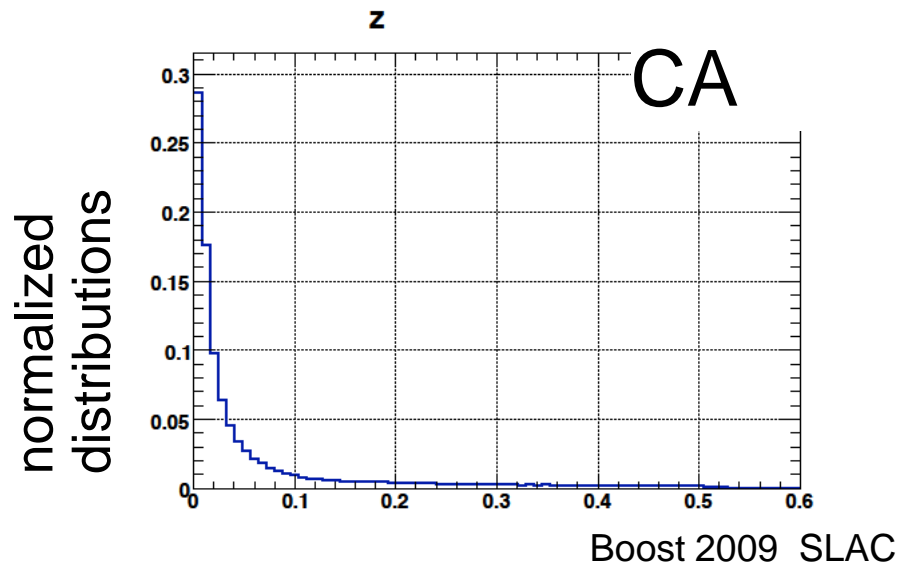
ttbar

All  $200 \text{ GeV} < p_T < 500 \text{ GeV}$



# Systematics of Algorithm: z

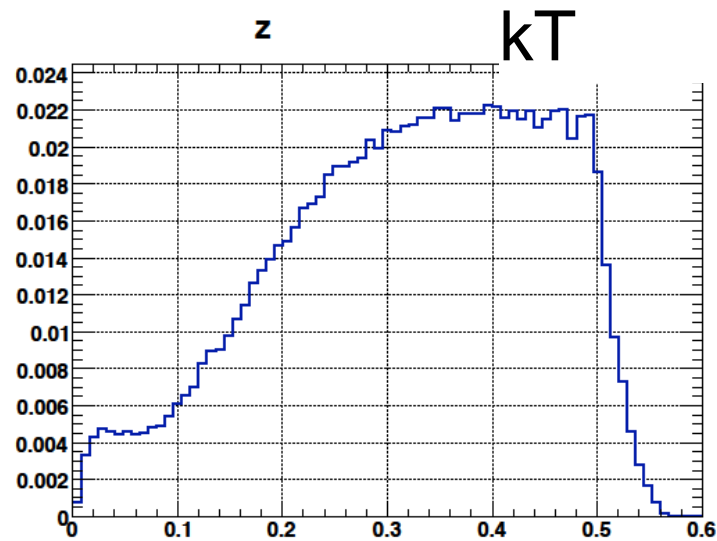
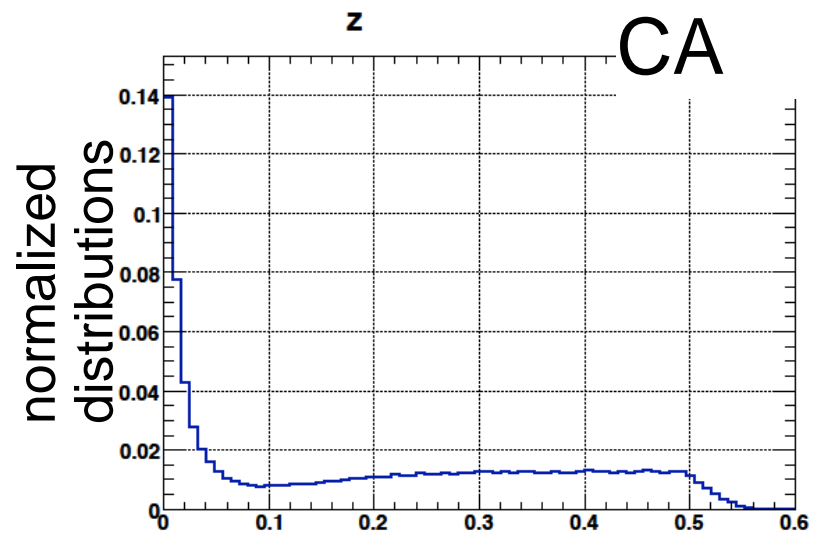
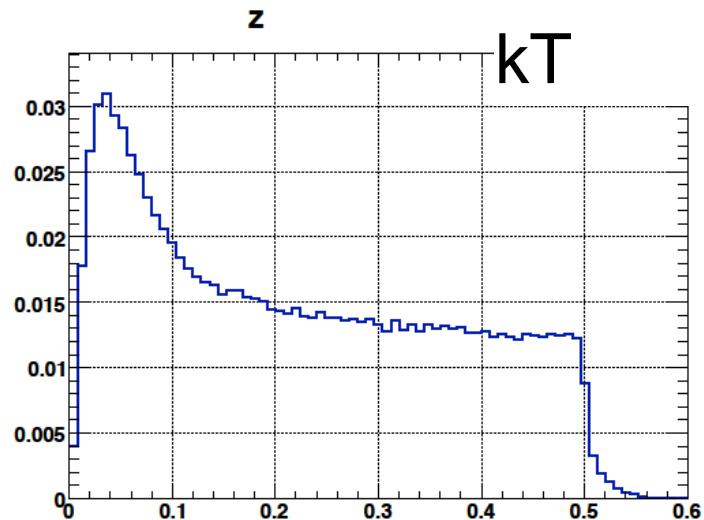
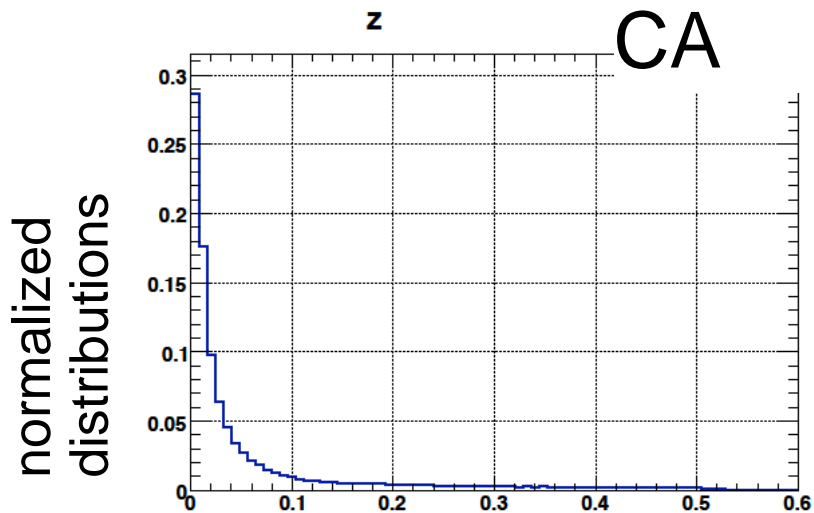
- Consider z on **LAST** recombination for CA and kT.
- Metric for CA is independent of z - distribution of z comes from the ordering in  $\theta$
- Periphery of jet is dominated by soft protojets - these are merged early by kT, but can be merged late by CA
- CA has many more low z, large  $\theta$  recombinations than kT



QCD



# Systematics of Algorithm: $z$ COMPARE



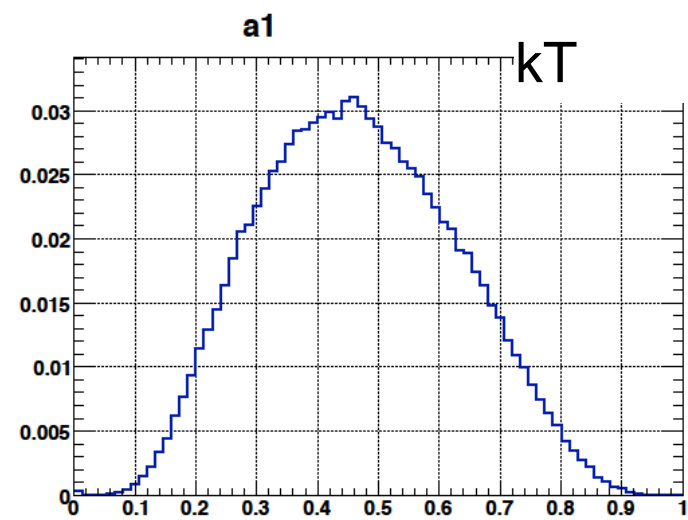
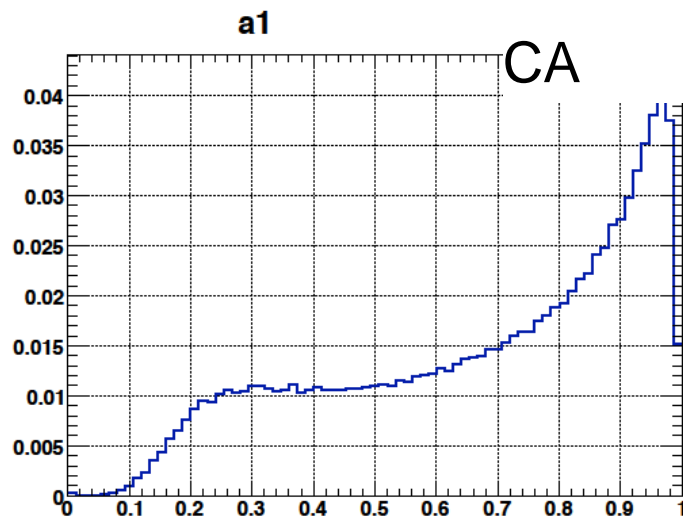


# Systematics of Algorithm: Subjet Masses

- 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

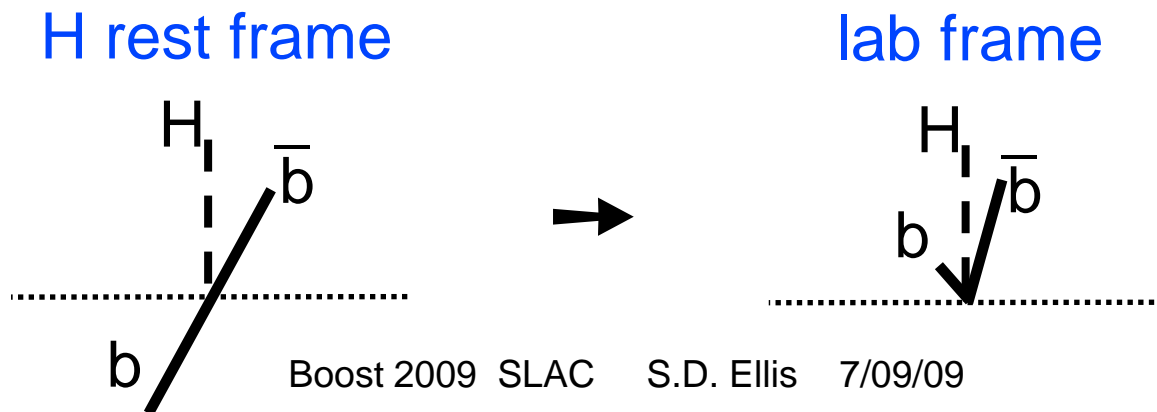


QCD



# 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.





# Summary: Reconstructed Heavy Particles

- “Real” 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 also 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
- ⇒ PRUNE these recombinations and focus on masses



Others have tried similar ideas (and earlier), e.g. – Butterworth, Davison, Rubin & Salam, (Higgs)

Kaplan, Rehermann, Schwartz & Tweedie (tops)

Thaler & Wang (tops)

Almeida, Lee, Perez, Sung and Virzi (tops)



# 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:

- $z < z_{\text{cut}}$  and  $\Delta R_{ij} > D_{\text{cut}}$

CA:  $z_{\text{cut}} = 0.1$  and  $D_{\text{cut}} = m_J/P_{T,J}$

- $m_J/P_{T,J}$  is IR safe measure of opening angle of found jet

kT:  $z_{\text{cut}} = 0.15$  and  $D_{\text{cut}} = m_J/P_{T,J}$

- 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

⇒ The resulting jet is the pruned jet



# Test Pruning:

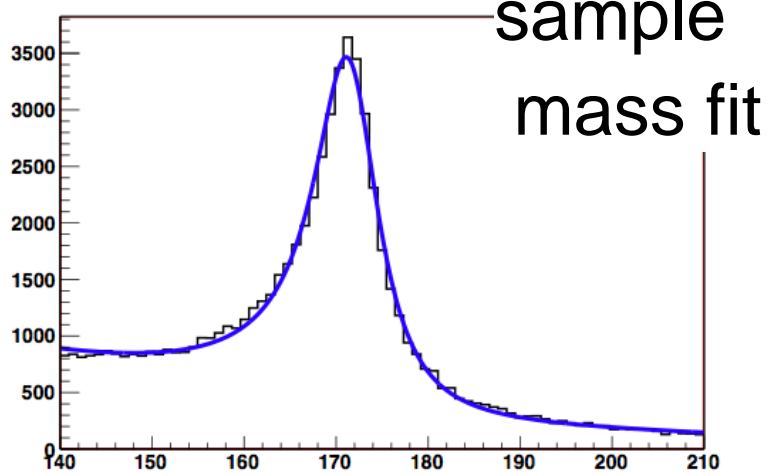
- Study of top reconstruction:
  - Hadronic top decay as a surrogate for a massive particle produced at the LHC
  - Use a QCD multijet background based on matched samples from 2, 3, and 4 hard parton MEs
  - ME from MadGraph, showered and hadronized in Pythia (DWT tune), jets found with FastJet
- Look at several quantities before/after pruning:
  - ⇒ Mass resolution of reconstructed tops (width of bump), small width means smaller background contribution
  - $p_T$  dependence of pruning effect
  - Dependence on choice of jet algorithm and angular parameter  $D$



# Defining Reconstructed Tops – Search Mode

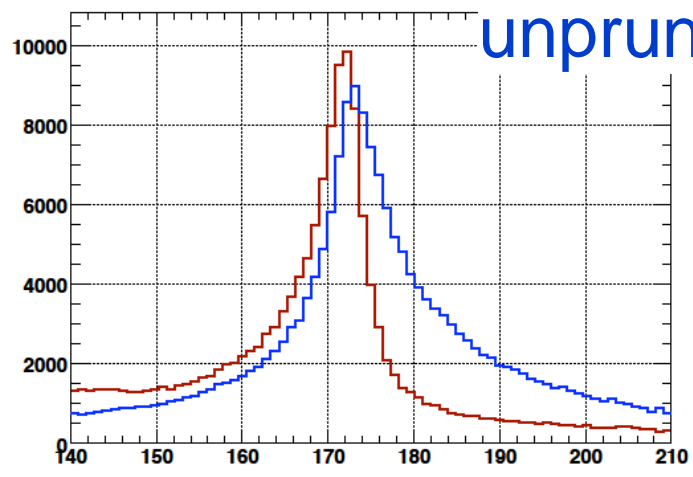
- A jet reconstructing a top will have a mass within the top mass window, and a primary subjet mass within the W mass window - call these jets **top jets**
- Defining the top, W mass windows:
  - **Fit** the observed jet mass and subjet mass distributions with (asymmetric) Breit-Wigner plus continuum → 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

jet mass, pruned



pruned  
unpruned

jet mass

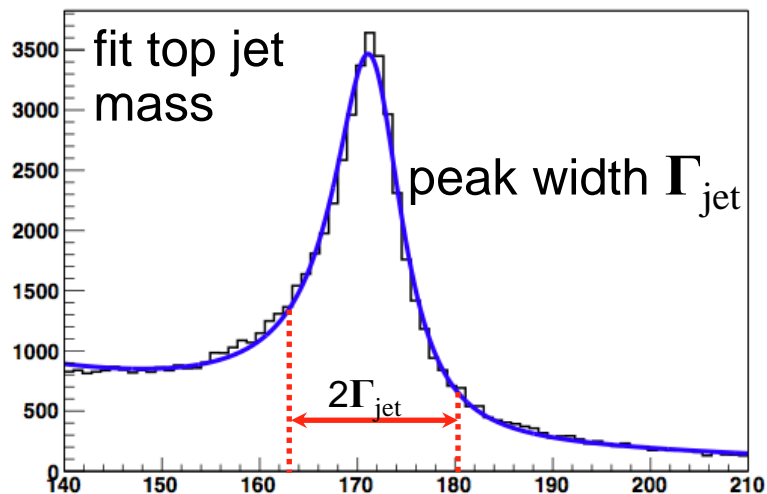




# Defining Reconstructed Tops

fit mass windows to identify  
a reconstructed top quark

jet mass



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

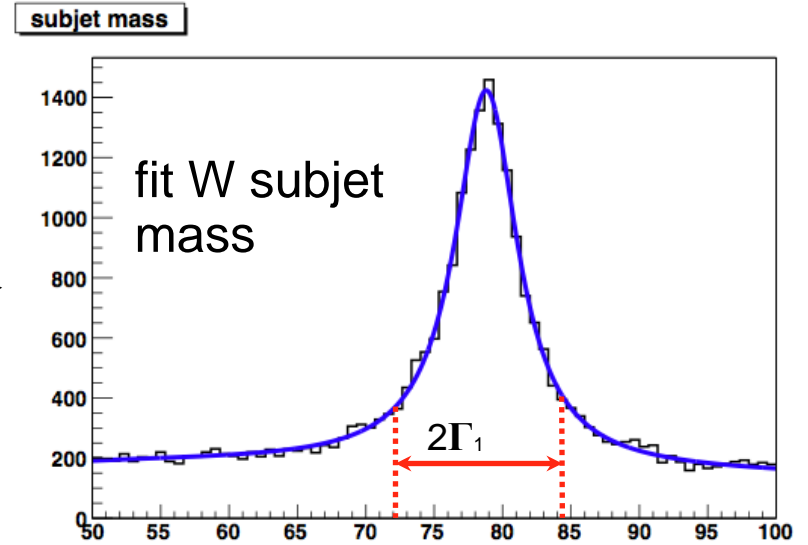
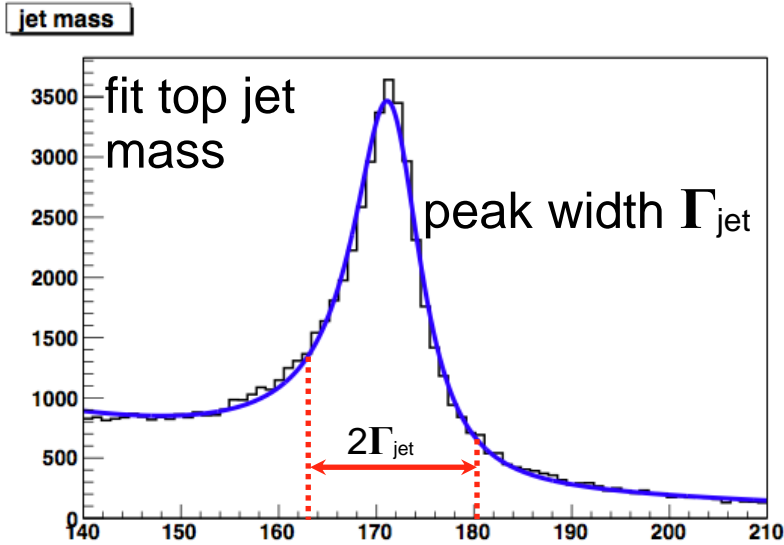
$$\frac{c}{m} + \frac{d}{m^2}$$



# Defining Reconstructed Tops

fit mass windows to identify a reconstructed top quark

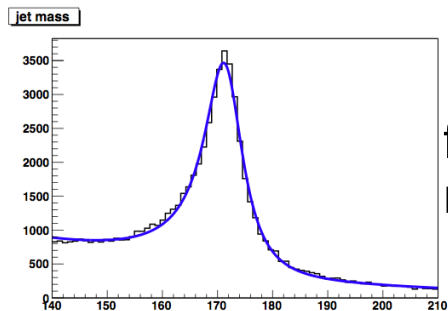
cut on masses of jet (top mass) and subjet (W mass)





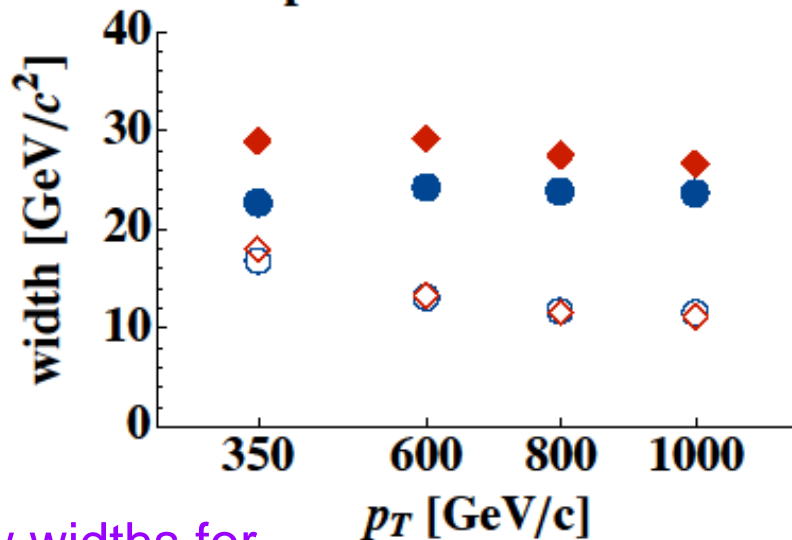
# Defining Reconstructed Tops

fit mass windows to identify a reconstructed top quark

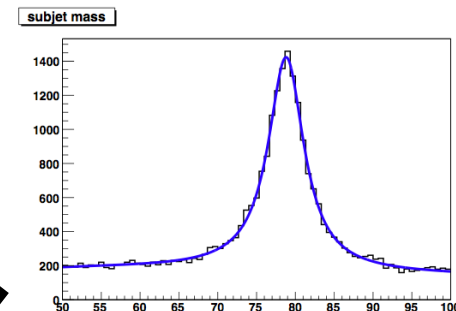


fit top jet mass

Top mass window width

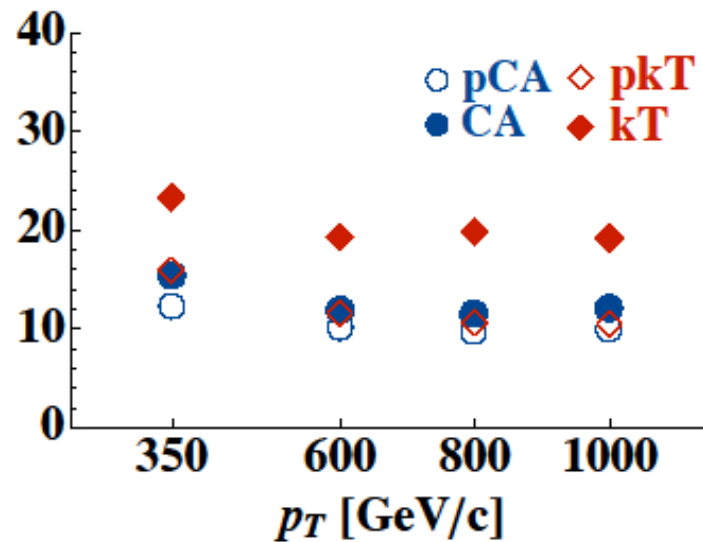


cut on masses of jet (top mass) and subjet (W mass)



fit W subjet mass

W mass window width

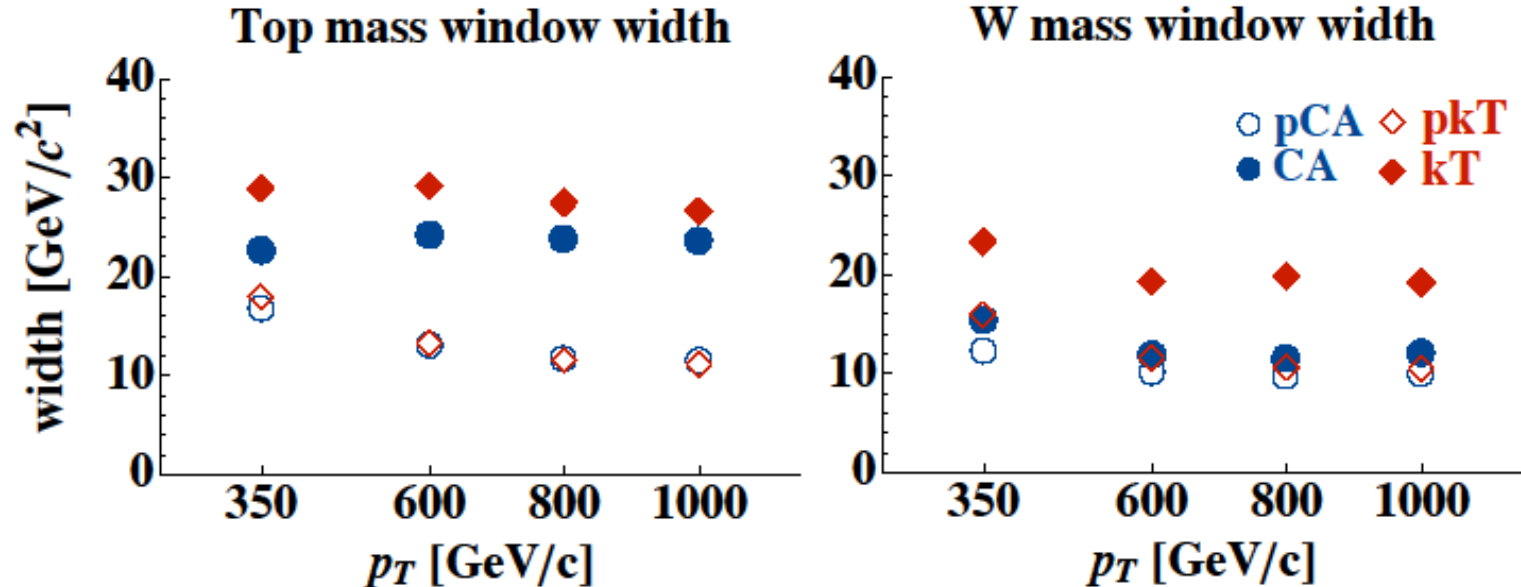


window widths for pruned (pX) and unpruned jets



# Mass Windows and Pruning - Summary

- Fit the top and W mass peaks, look at window widths for unpruned and pruned (pX) cases in (200 - 300 GeV wide)  $p_T$  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  $p_T$





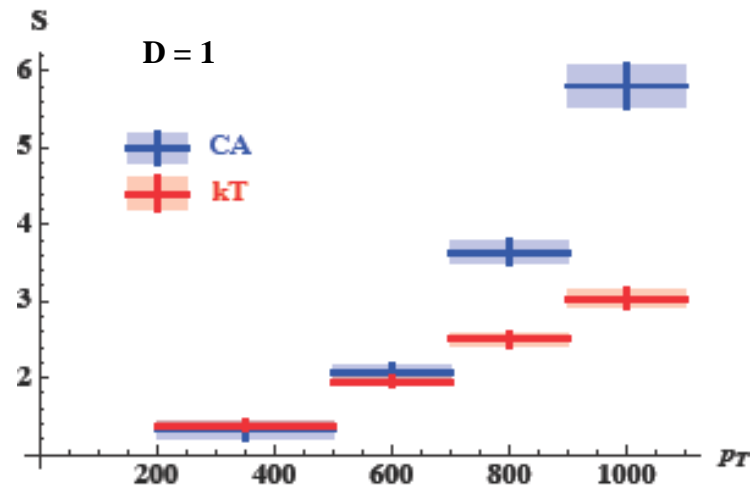
# Statistical Measures:

- 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:
  - $\epsilon$ ,  $R$ ,  $S$  - efficiency, Sig/Bkg, and Sig/Bkg<sup>1/2</sup>

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

Here focus on  $S$

$S > 1$  (improved likelihood to see bump if prune), all  $p_T$ , all bkg, both algorithms

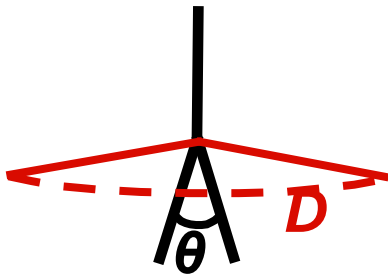




# Heavy Particle Decays and $D$

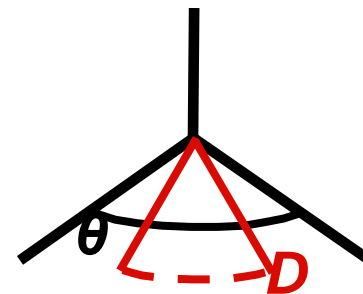
(See also - Variable R... – Krohn, Thaler & Wang )

- Heavy particle ID with the unpruned algorithm is improved when  $D$  is matched to the expected average decay angle
- Rule of thumb (as above):  $\theta = 2m/pT$
- Two cases:



$D > \theta$

- lets in extra radiation
- QCD jet masses larger



$D < \theta$

- particle will not be reconstructed



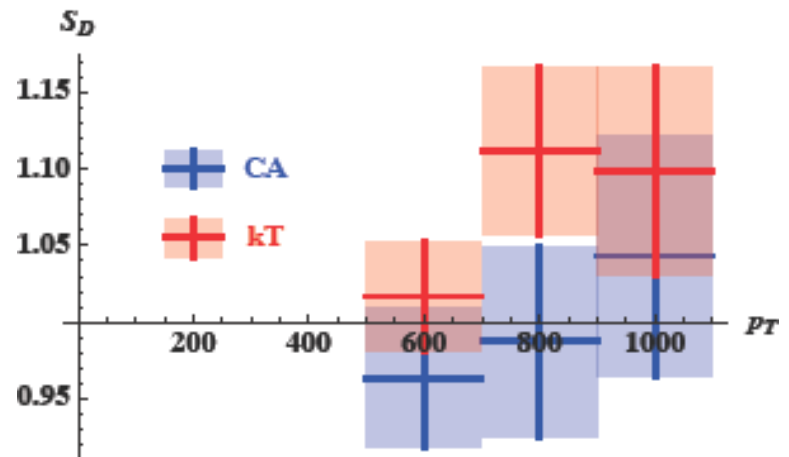
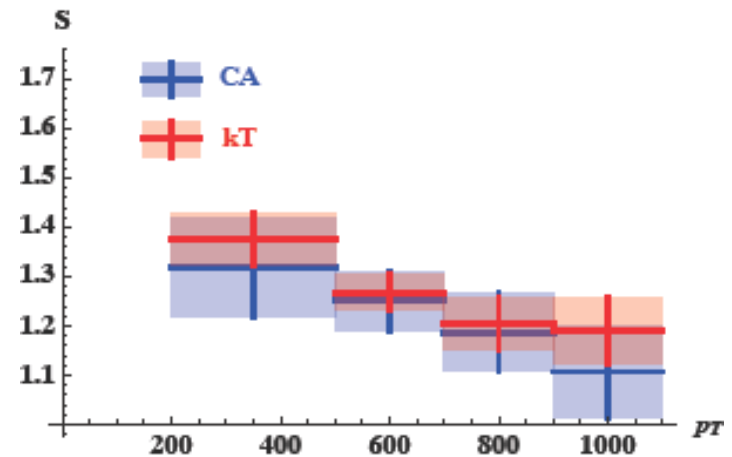
# Improvements in Pruning

- Optimize  $D$  for each pT bin:  $D = \min(2m/pT_{\min}, 1.0) \Rightarrow (1.0, 0.7, 0.5, 0.4)$  for our pT bins
- Pruning still shows improvements
- How does pruning compare between fixed  $D = 1.0$  and  $D$  optimized for each pT bin  $\Rightarrow S_D = S_{D \text{ opt}}/S_{D=1}$ ?

$\Rightarrow$  Little further improvement obtained by varying  $D$

$\Rightarrow S_D = 1$  in first bin

$\Rightarrow$  Pruning with Fixed  $D$  does most of the work





# Learned 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 *generic* 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)
- software at [tinyurl.com/jetpruning](http://tinyurl.com/jetpruning)



# And:

- 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)
- Much left to understand about jet substructure (here?), *e.g.*,
  - How does the detector affect jet substructure and the systematics of the algorithm? How does it affect techniques like pruning? What are experimental jet mass uncertainties?
  - How can jet substructure fit into an overall analysis? How orthogonal is the information provided by jet substructure to other data from the event?