Jets at the LHC: Looking Forward and Backward

Steve Ellis

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.

Department of Physics
University of Washington

Outline

- Why jets?
- Old and New lessons for Cone and Recombination (kT) jets
- Understanding jet masses & substructure
- Searching Beyond the Standard Model using single jets

$\Rightarrow$ Pruning to improve searches

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/QCD, and many lessons were learned – QCD is correct but jets have systematic issues

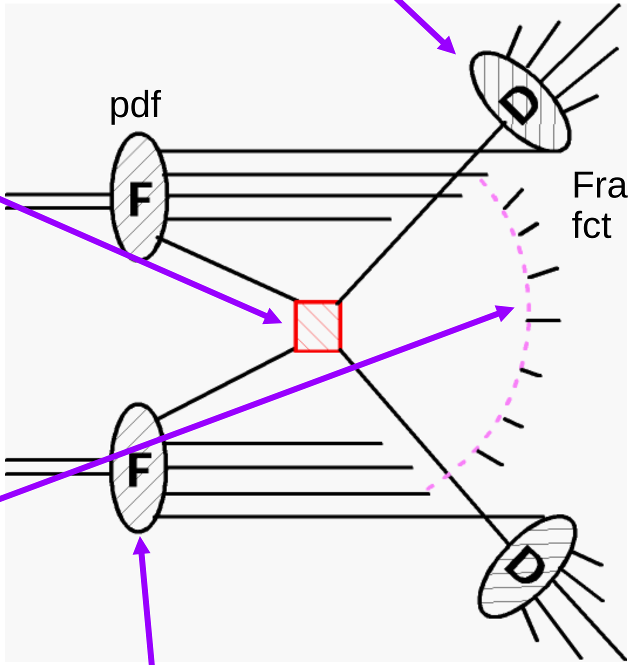
Jets will be used differently at the LHC

new detectors – need to understand them, may be better for jets

look for BSM physics in single jets (*non-SM-ness*), use properties of jets – masses, substructure to tag non-QCD jets

better theoretical understanding (eg., G. Salam, et al.) ⇒ new algorithms but need to understand them in real detectors

Jet Physics: The Basis of QCD Collider Phenomenology – Looking Back



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 talk about this

Correlated by Underlying Event (UE) color correlations + PU

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, initially 1 jet ⇔1 parton (test QCD)

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

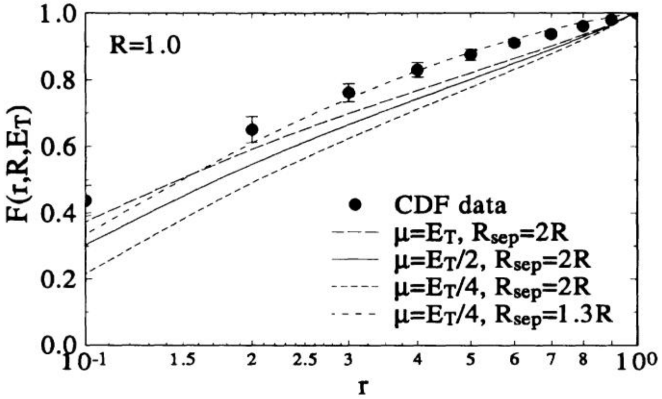


FIG. 2. $F(r,R,E_T)$ vs r for $R=1.0$, $\sqrt{s}=1800$ GeV, $E_T=100$ GeV, and $0.1<|\eta|<0.7$ with $\mu=E_T/4$, $E_T/2$, E_T compared to data from CDF [7]; the dot-dashed curve is explained in the text.

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

Defining Jets – No Unique/Correct Answer

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

- Need rules for summing ⇒ jet algorithm
  - Start with list of particles/towers
  - End with list of jets (and stuff not in jets)

E.g.,

- Cone Algorithms, based on geometry – “non-local” sum over core of shower

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

- Recombination (or kT) Algorithm, based on “local” pair-wise merging of local objects to “undo” shower

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

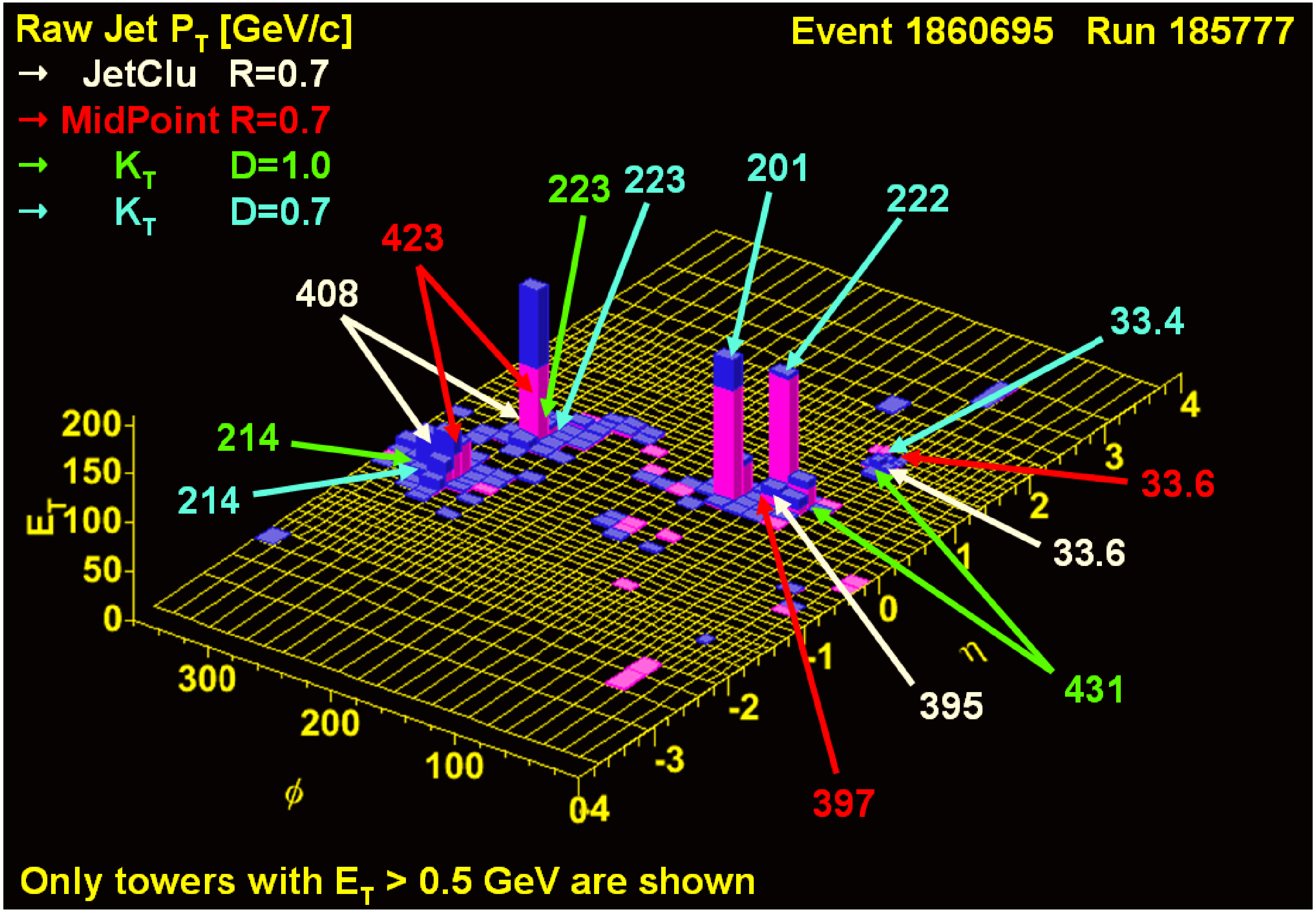
The good news about jet algorithms:

- 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 tend to use different algorithms
- The detailed results (masses, substructure) depend on the algorithm

Different algorithms ⇒ different jets (same CDF event)



EM, **Hadronic**

“Look Back” at Lessons about the Systematics

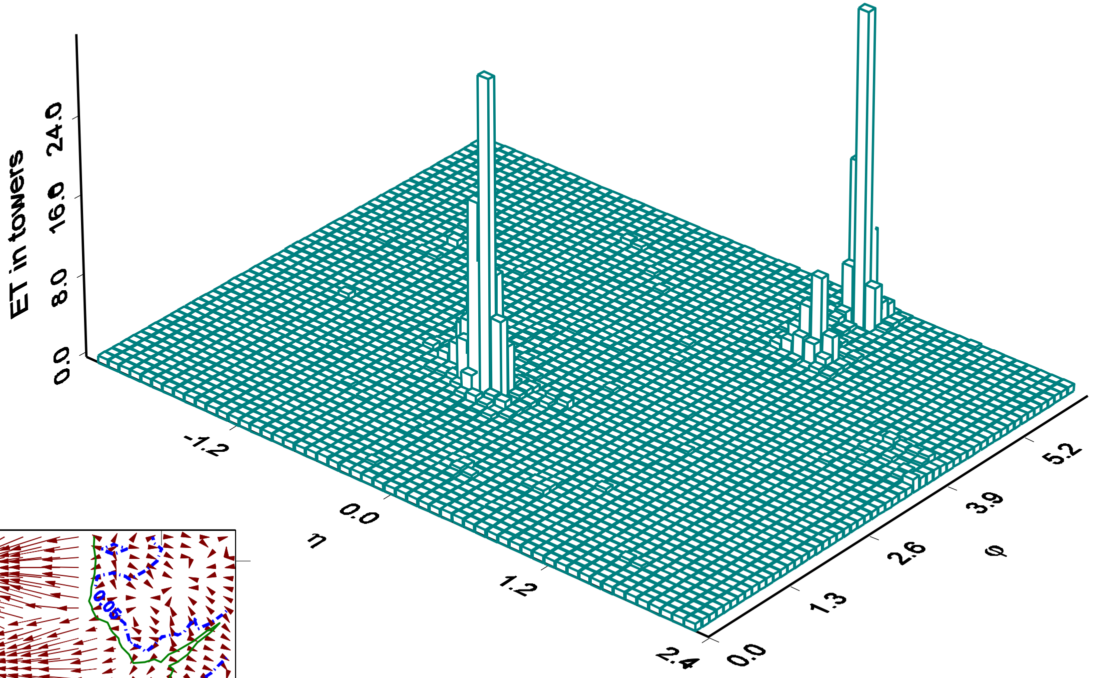
Cone Algorithm – focus on the core of jet (1990 Snowmass)

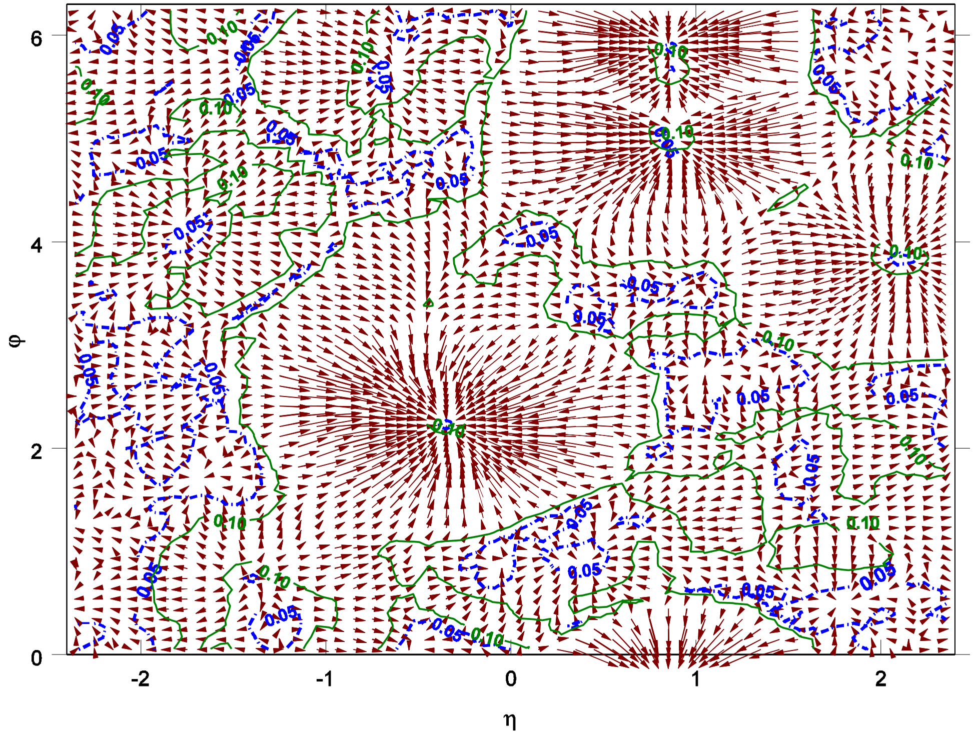
- Jet = “stable cone” ⇒ 4-vector of cone contents || cone direction
- Well studied – several issues

- **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 $\left(\bar{y}^C,\bar{\varphi}^C\right)$

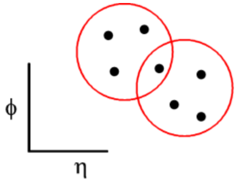
Think of as flow problem to the minima of the Snowmass Potential





$$V_{\text{Snowmass}}\left(y^C,\varphi^C\right) \propto \left[\left(y^C-\bar{y}^C\right)^2+\left(\varphi^C-\bar{\varphi}^C\right)^2\right]$$

Cone Lessons: (The devil we know)



1) Stable Cones can and do *overlap*, need to define rules for merging and splitting (and which cones participate) ⇒ more parameters, merge if shared energy fraction > f_{merge}, else split (but CDF and D0 choose different parameters)

   ⇒ Need $f_{merge} > 0.5$ to avoid too much merging ⇒ huge jets and high sensitivity to UE and PU, e.g., jet area grows with PU

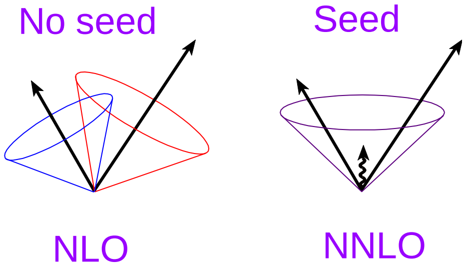
   ⇒ Need $f_{merge} < 0.8$ to avoid too much splitting ⇒ reduced jet sizes and sensitivity to UE and PU, e.g., jet area grows with PU

2) Seeds – experiments only look for jets near very active regions (save computer time, no longer a problem)

   ⇒ Problem for theory, IR sensitive (Unsafe?) at NNLO

   ⇒ Don't find "possible" central jet between two well separated proto-jets (partons)

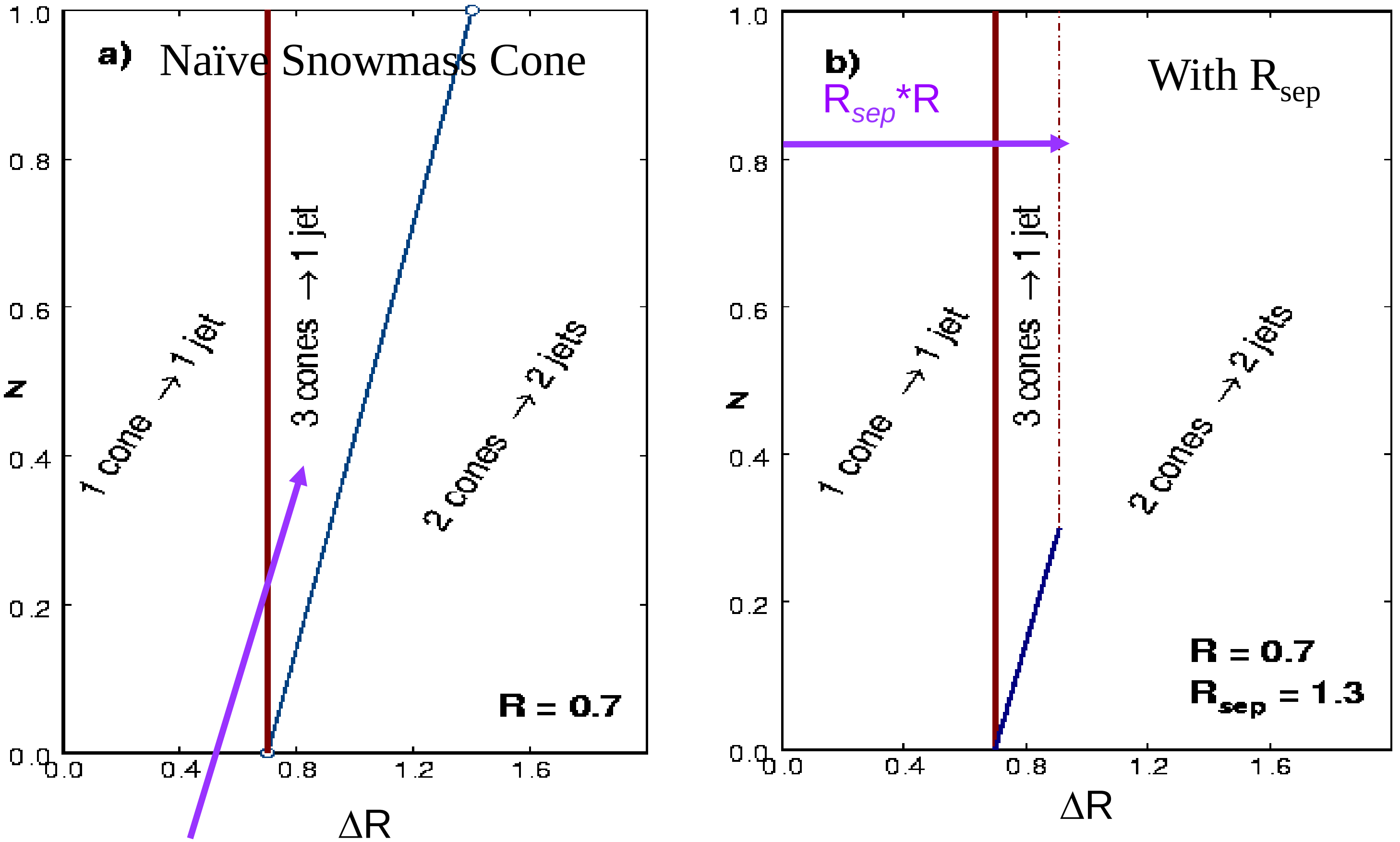
   ⇒ Simulated with R_{sep} (eliminate $\Delta R > R_{sep} * R$)



Seeds:

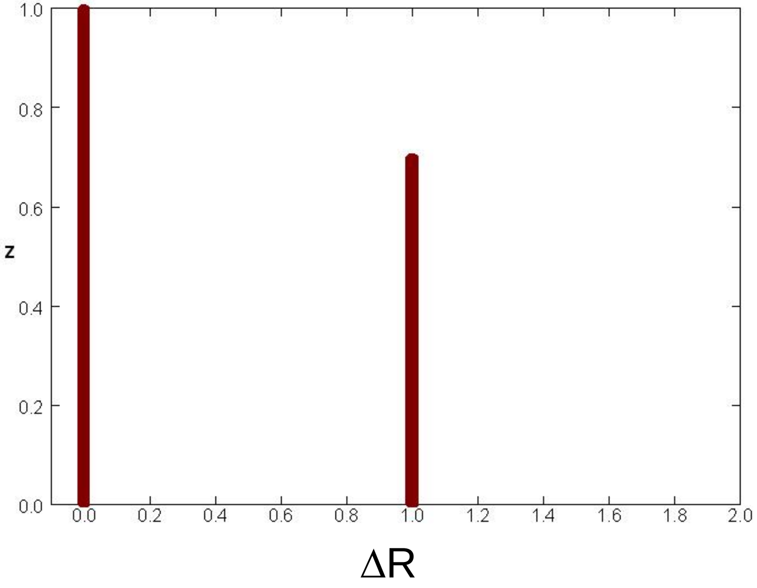
- Seeds can mean missed configurations with 2 partons in 1 Jet, NLO Perturbation Theory – ΔR = parton separation, $z = p_2/p_1$,,

  Simulate the missed middle cones with R_{sep}

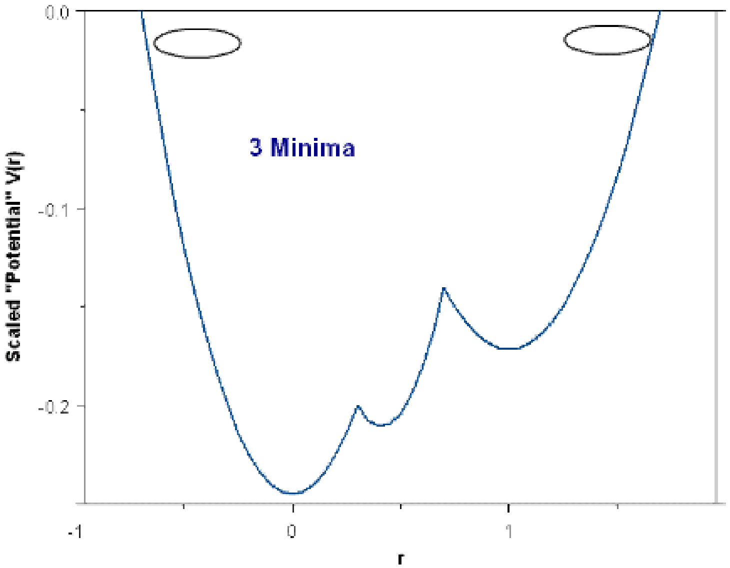


~10% of cross section here

Simple Theory Model - 2 partons (separated by ΔR < 2R): yield potential with 3 minima – trial cones will migrate to minima from seeds near original partons ⇒ miss central minimum



Snowmass Potential

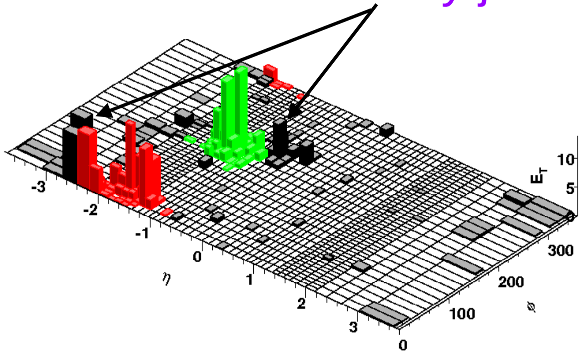


$z = p_{\min} / p_{\max}$, ΔR = separation

Smearing of order R

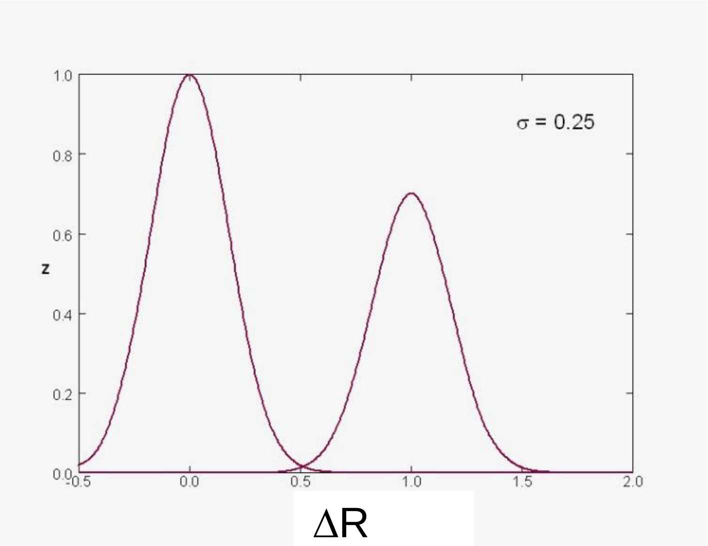
Cone Lessons: (The devil we know)

3) Dark Towers - Energy in secondary showers may not be clustered in any jet

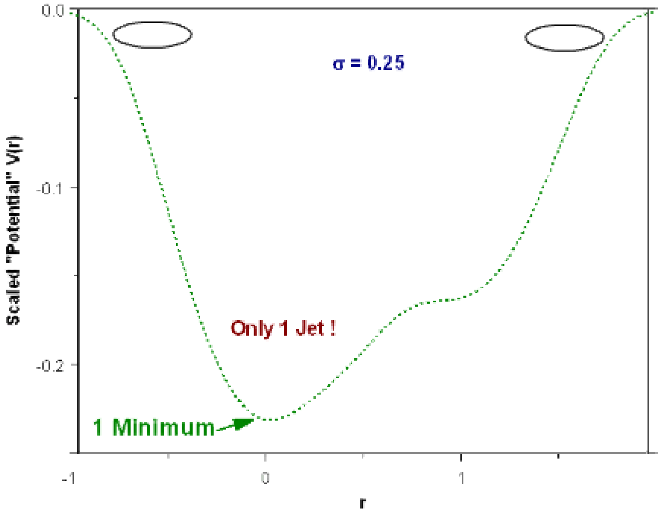


- Expected stable cone *not* stable due to smearing from showering/hadronization (compared to PertThy)
- Under-estimate E_T (~ 5% effect for jet cross section)

Include Gaussian smearing



$\Rightarrow$



Cone Fixes -

1. All experiments use the same split/merge parameters and $0.8 > f_{merge} > 0.5$ to avoid over-merging, over-splitting (jet size stable *vs* jet pT or PU)

   Not true at the Tevatron…

2. NOTE: “progressive-removal” seeded cones - find cone jets one at a time starting with largest pT seed and REMOVE jet constituents from further analysis. This is NOT collinear safe!

3. Use seedless cone algorithm (e.g., SIScone), or correct data for seed effects

   Small effect (1-2 %) in data, big issue in pert Thy

4. No good solution yet to Dark towers except to look for 2nd pass jets after removing the 1st pass jets from the analysis.

Recombination – focus on undoing the shower pairwise

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

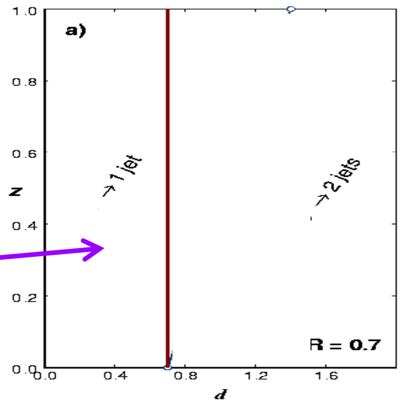
$$k_{T,(ij)}^2 \equiv \mathrm{Min}\left[\left(p_{T,i}^2\right)^\alpha, \left(p_{T,j}^2\right)^\alpha\right] \frac{\left(y_i - y_j\right)^2 + \left(\phi_i - \phi_j\right)^2}{D^2}, k_{T,i}^2 = \left(p_{T,i}^2\right)^\alpha$$

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)



α *= 1,* ordinary *k_T,* recombine soft stuff first (undo kT ordered shower)

α *= 0, Cambridge/Aachen* (CA)*,* controlled by angles only (undo angle ordered shower)

α *= -1, Anti-k_T,* just recombine stuff around hard guys – cone-like with seeds
THE NEW GUY!! (not matched to showers)

Recombination Lessons:

- 👍 Jet identification is unique – no merge/split stage

- 👍 Everything in a jet, no Dark Towers

- 👎 Resulting jets are more amorphous for $\alpha \geq 0$, energy calibration more difficult (subtraction for UE + PU?)

  jet area grows with jet pT, shrinks with PU (size of effect depends on α)*

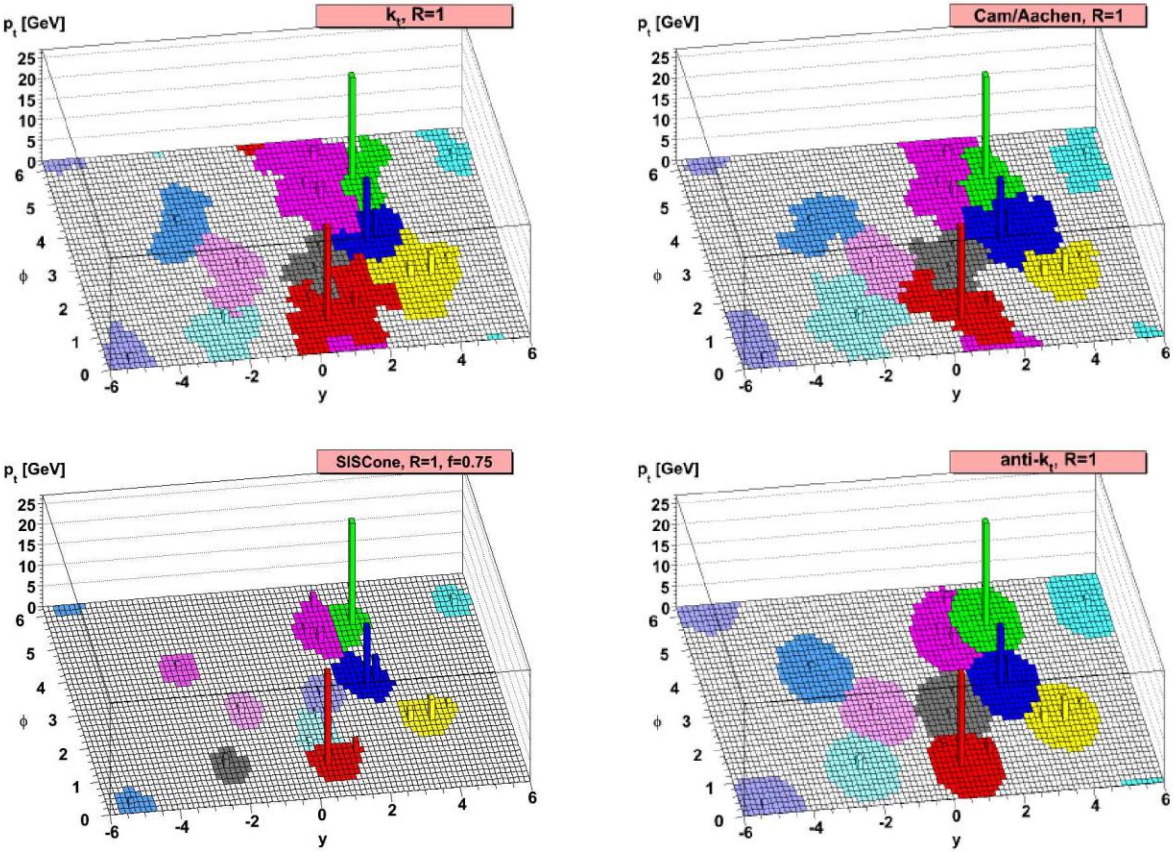
- 👍 But for $\alpha < 0$, Anti-kT (G. Salam et al.), jet area seems stable and geometrically regular * - the “real” cone algorithm

- 👎 Analysis can be very computer intensive (time grows like N^3, recalculate list after each merge)

- 👍 New version FASTJet (G. Salam et al.) goes like N^2 or N ln N ($\alpha \geq 0$), plus scheme for finding areas (and UE correction)

* From J. Huston & B. Martin

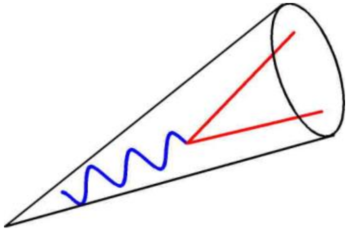
Jet Areas – from G. Salam



Anti-kT very regular leading jets

Goals at LHC Different ⇒ Different Role for Jets!

- Find Physics Beyond the Standard Model (BMS)
- BSM Event structure likely different from QCD, more jets? Different structure within jets? Must be able to reconstruct masses from multi-jets & also from *single jets*
- Want to select events/jets by non-QCD-ness



- Highly boosted SM and non-SM particles – W, Z, top, Higgs, SUSY ⇒ *single* jet instead of 2 or 3 jets, focus on masses and substructure of jets
- Much recent progress, but lots of work still to be done – need real data!!

Looking for hidden truth -

Recent progress in 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.)
- Better understanding of jet masses – jets have a rest frame! (S. Ellis et al.)
- Jet tagging schemes to ID W/Z, top quarks or Higgs (or other BSM particles) as single jets –
  J. Butterworth and collaborators (e.g., G. Salam)
  UCB Group (J. Thaler, et al.)
  Johns Hopkins Group (D. Kaplan, et al.)
  Stony Brook Group (G. Sterman, et al.)
- Generic search/pruning techniques for BSM searches with single jets – focus on masses for now - UW group (with C. Vermilion & J. Walsh)

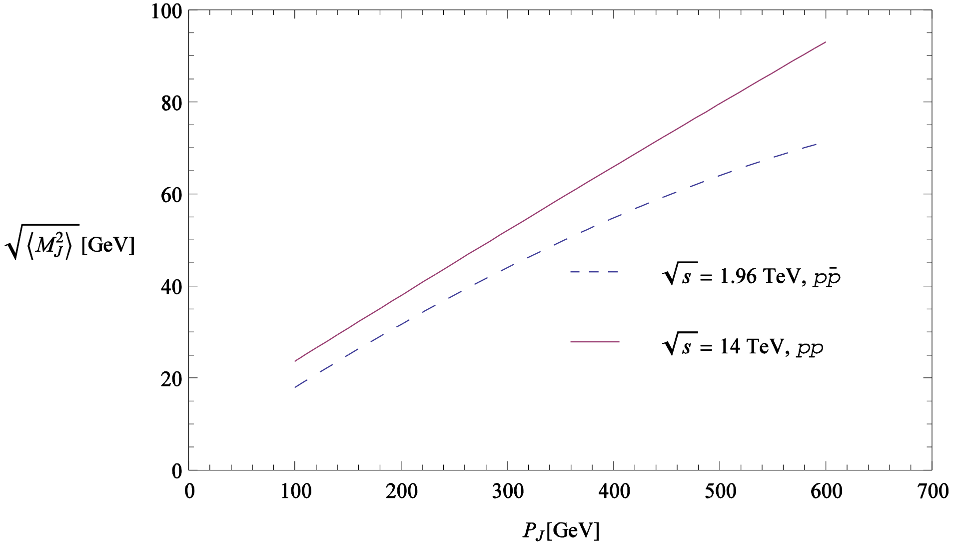
Jet Masses in QCD: To compare to non-QCD

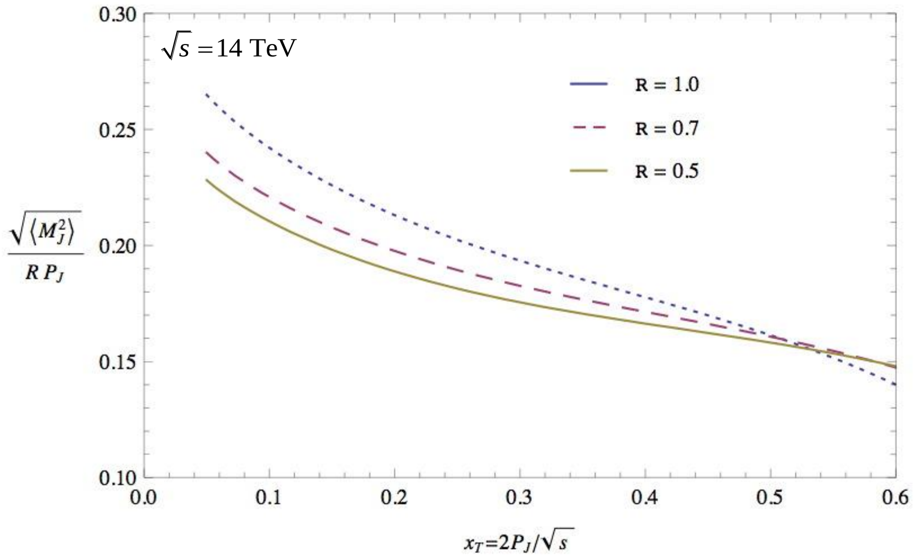
Dimensions

- 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* ~ 1

Jet Size, *R*, *D* ~ $\Delta\theta$, determined by jet algorithm

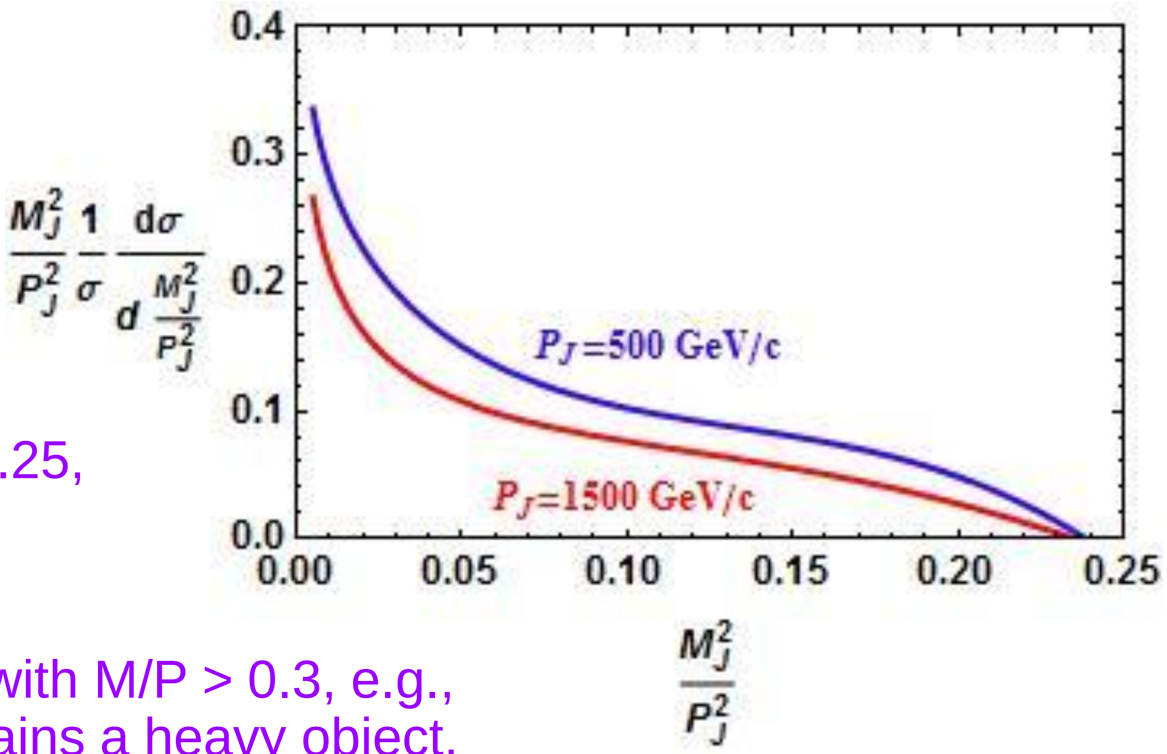




Useful QCD “Rule-of-Thumb” $\Rightarrow \sqrt{\langle M^2\rangle_{NLO}} \sim 0.2\,p_J R$

Mass for fixed P_J at NLO

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



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

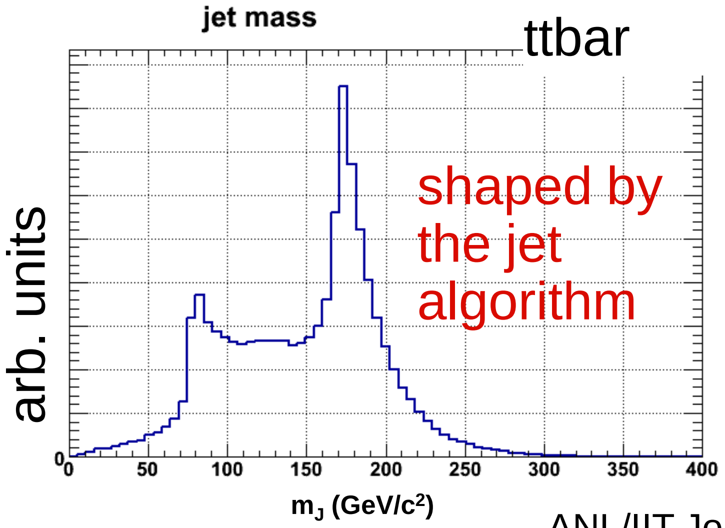
$\Rightarrow$ Selecting on jets with $M/P > 0.3$, e.g.,
because the jet contains a heavy object,
already suppresses the QCD background;

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

Finding Heavy Particles with Jets - Issues

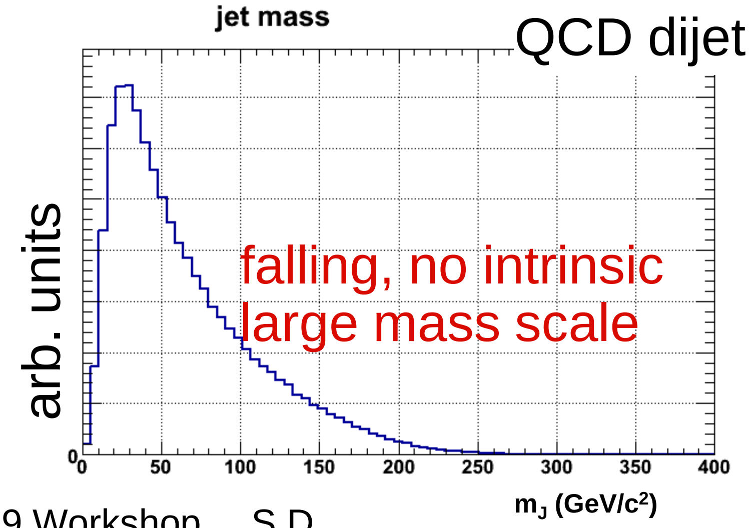
- 👎 QCD multijet production rate >> production rate for heavy particles
- 👍 In the jet mass spectrum, production of non-QCD jets may appear as local excesses (bumps!) but must be enhanced using analyses
- 👍 Use jet substructure as defined by recombination algorithms ($\alpha \geq 0$) to refine jets
- 👎 Algorithm will systematically shape distributions
- Use top quark as surrogate new particle.

$\boldsymbol{\sigma}_{\text{ttbar}} \approx 10^{-3}\boldsymbol{\sigma}_{\text{jj}}$



ttbar

shaped by the jet algorithm

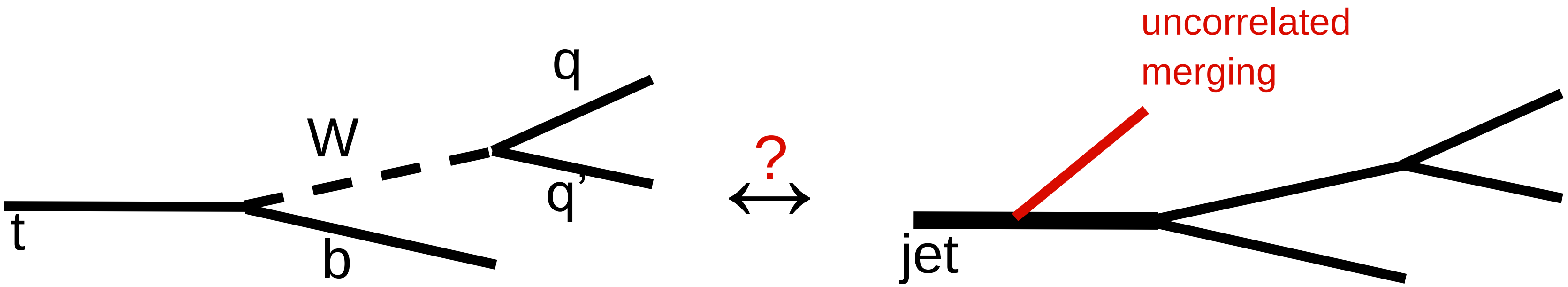


QCD dijet

falling, no intrinsic large mass scale

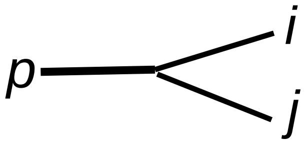
Reconstruction in Jet Substructure – separating jets with heavy mass scale from QCD (scale = Λ_{QCD})

- 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
- Masses (jet and subjet) are robust variables - strong discriminators between QCD and non-QCD jets



Systematics of the Jet Algorithm

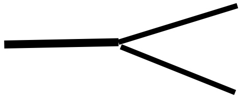
- Consider generic recombination step: *i,j* ➜ *p*



- Useful variables: (Lab frame) $z = \dfrac{\min(p_{T_i}, p_{T_j})}{p_{T_p}}$ $\theta = \Delta R_{ij}$
- Merging metrics: $\rho_{\mathrm{kT}} = p_{T_p} z\theta / D$ $\rho_{\mathrm{CA}} = \theta / D$
- In terms of *z, θ*, the algorithms will give different kinematic distributions:
  - CA orders only in *θ* : *z* is unconstrained
  - kT orders in *z·θ* : z and *θ* 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



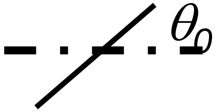
$$a_1 = m_{D,>} / M_J$$

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

- 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

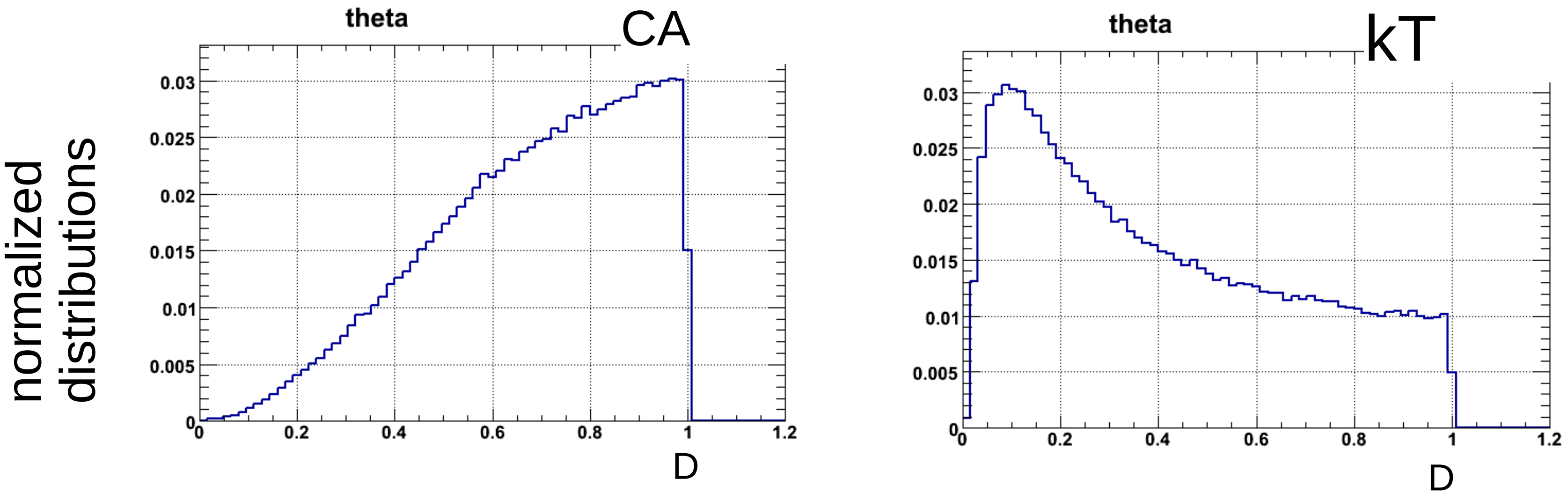
- Again angular distributions are strongly shaped by the algorithm, choosing the algorithm is important!

Studying Systematics: QCD vs ttbar Jets

- 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: θ

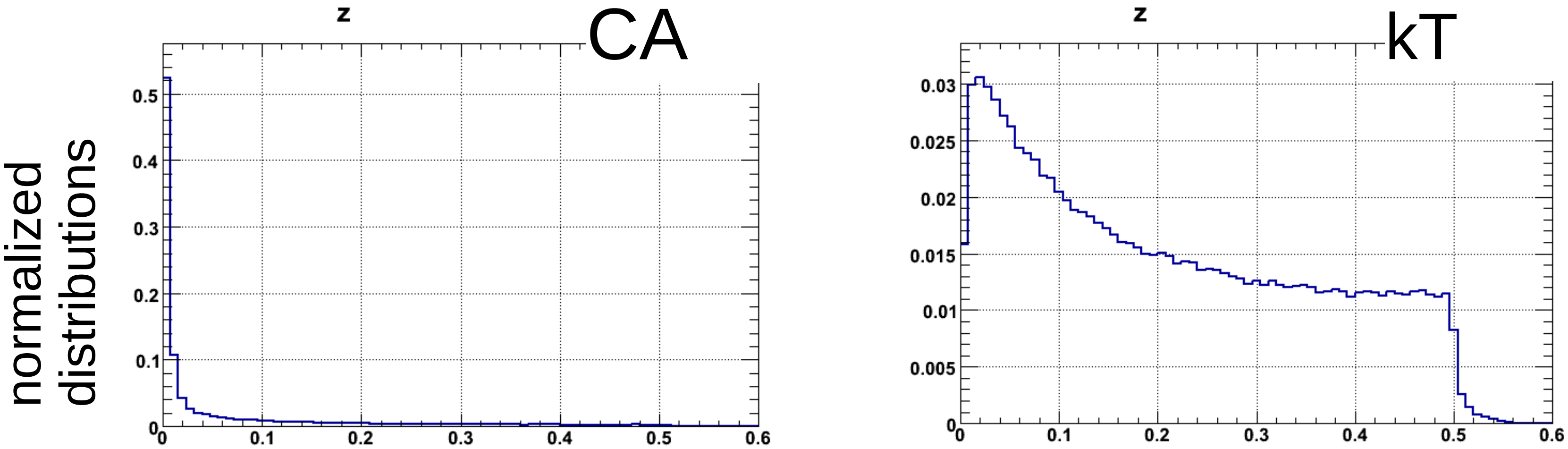
- Consider θ on LAST recombination for CA and kT.
- CA orders only in θ - means θ tends to be large (often close to D) at the last merging.
- kT orders in $z \cdot \theta$, meaning θ can be small
  - Get a distribution in θ that is more weighted towards small θ than CA (even though "jets" are the same)



QCD

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

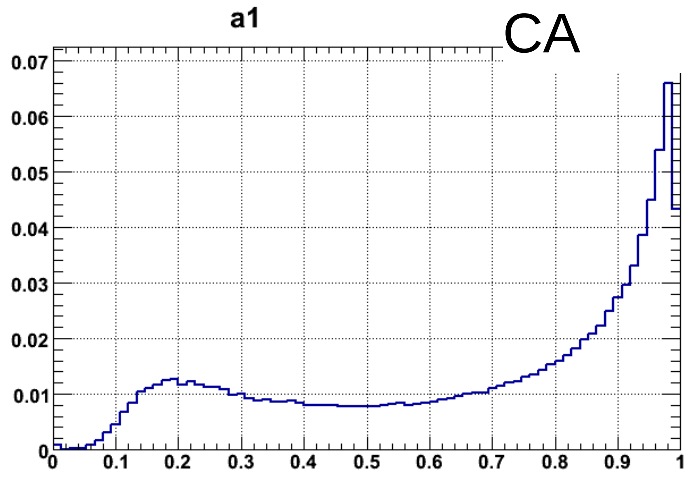


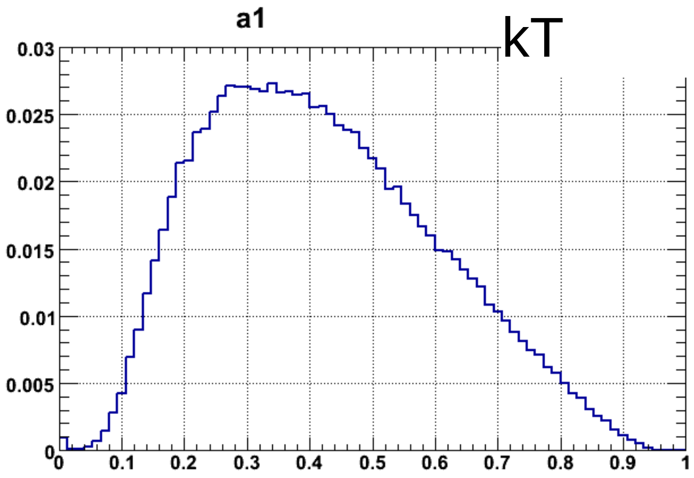
QCD

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 θ
  - 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

normalized distributions





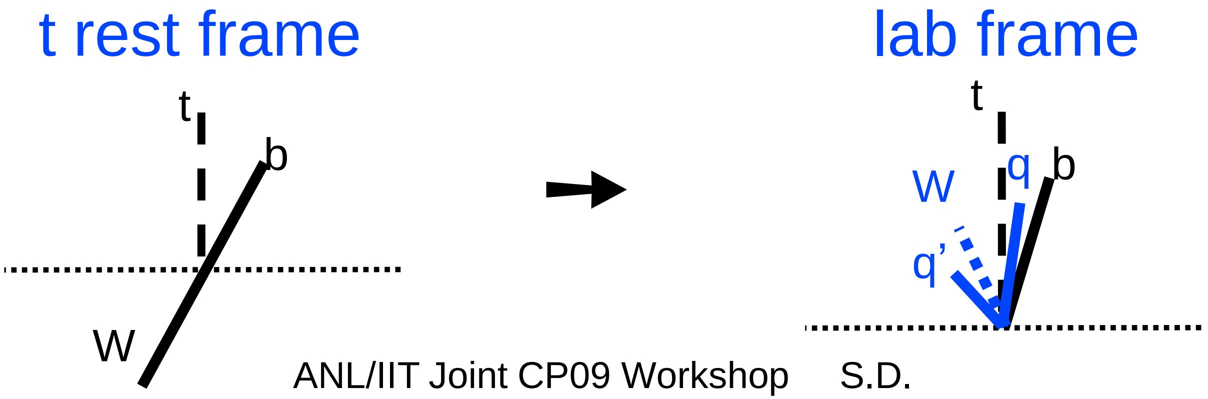
QCD

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

⇒ PRUNE these recombinations and focus on masses

others have tried similar ideas -
Salam/Butterworth (Higgs),
Kaplan (tops),
Thaler/Wang (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 $\Delta R_{ij} > D_{\text{cut}}$ (θ_J is angle at final recombination in original found jet)

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

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

- 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

⇒ 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 - 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:

⇒ Mass resolution of reconstructed tops (width of bump), small width means smaller background contribution

  - pT dependence of pruning effect

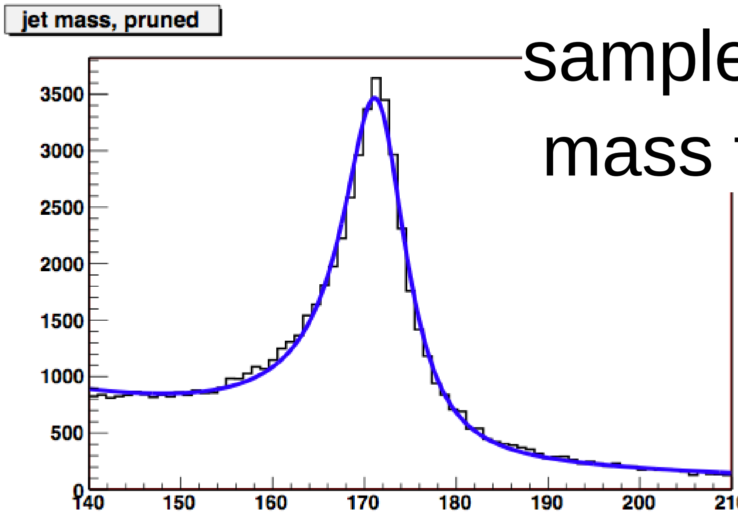
  - Dependence on choice of jet algorithm and angular parameter D

Defining Reconstructed Tops – Search Mode

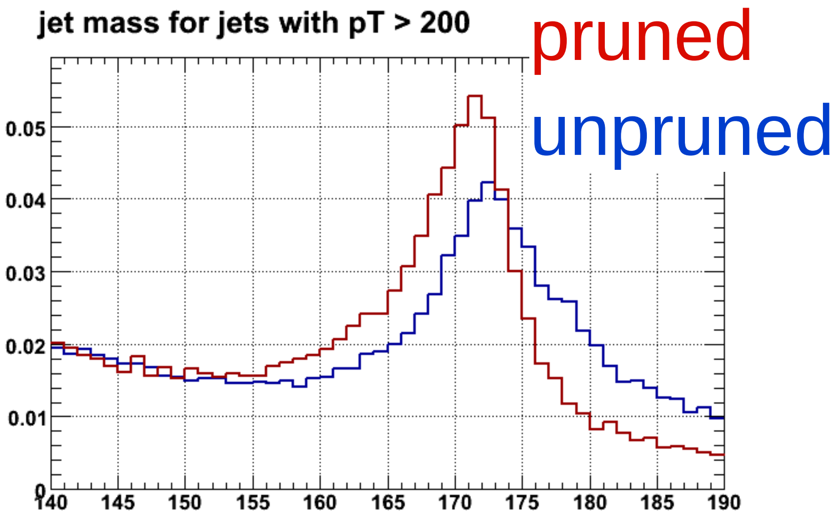
- A jet reconstructing a top will have a mass within the top mass window, and a primary subjet mass within the W mass window - call these jets top jets

- 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

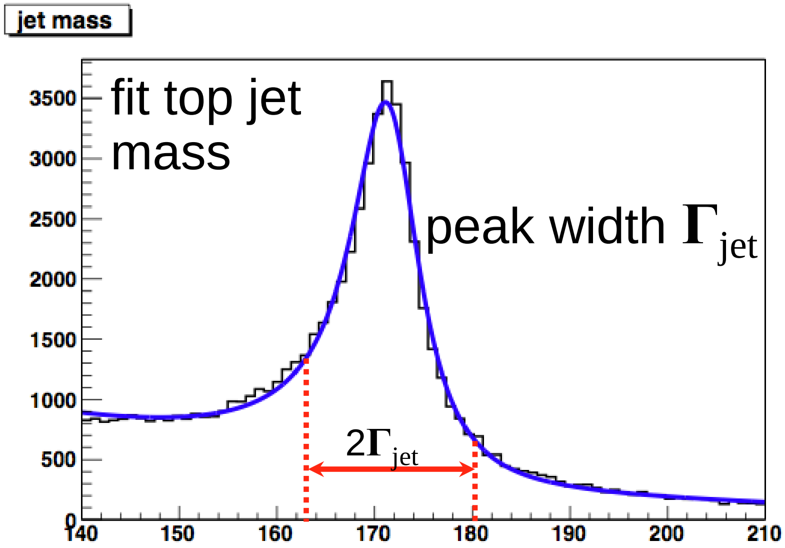


sample mass fit



Defining Reconstructed Tops

fit mass windows to identify
a reconstructed top quark



fit top jet mass

peak width $\mathbf{\Gamma}_{\text{jet}}$

peak function: skewed Breit-Wigner

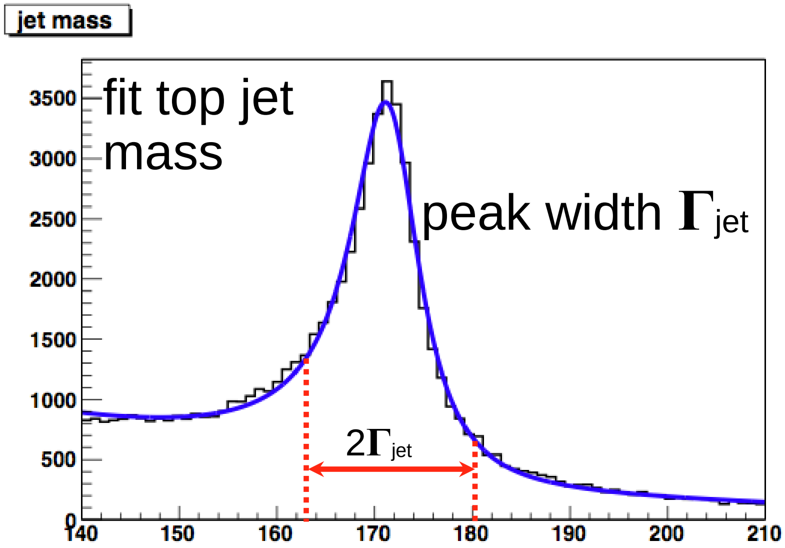
$$M^2\Gamma^2 \frac{[a+b(m-M)]}{(m^2-M^2)^2+M^2\Gamma^2}$$

plus continuum background distribution

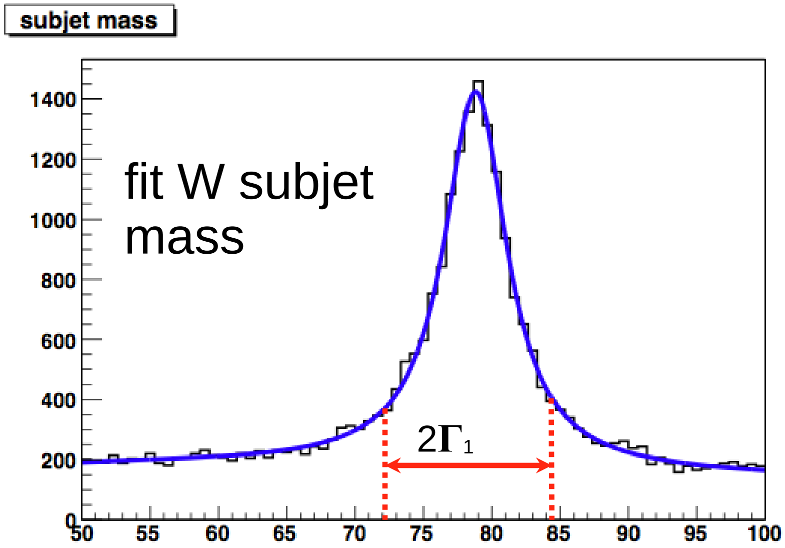
$$\frac{c}{m}+\frac{d}{m^2}$$

Defining Reconstructed Tops

fit mass windows to identify
a reconstructed top quark



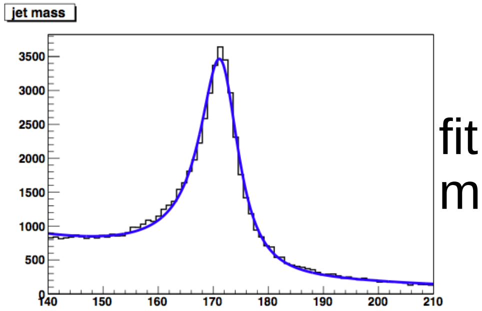
cut on masses of jet (top mass)
and subjet (W mass)



Defining Reconstructed Tops

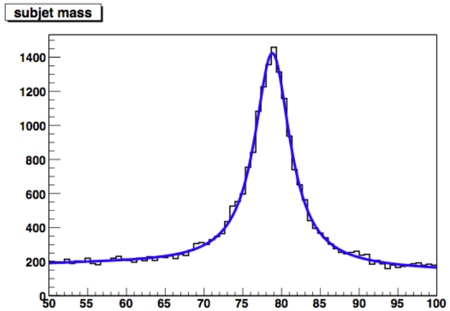
fit mass windows to identify a reconstructed top quark

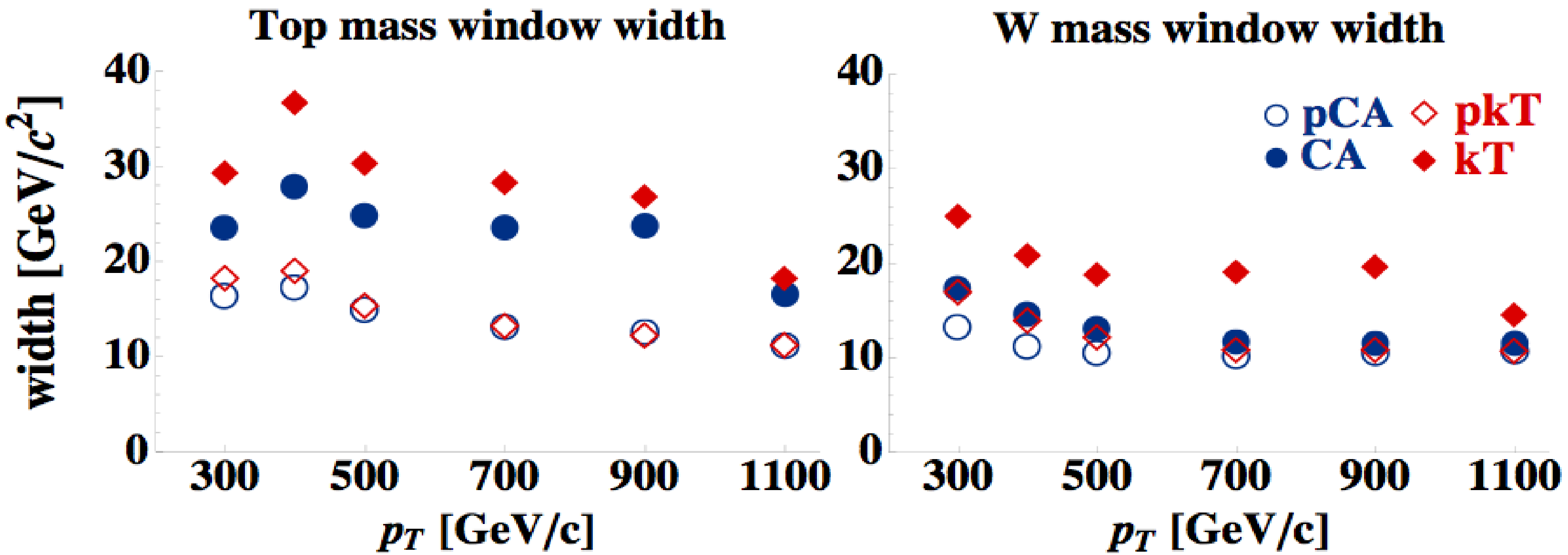
cut on masses of jet (top mass) and subjet (W mass)



fit top jet mass

fit W subjet mass





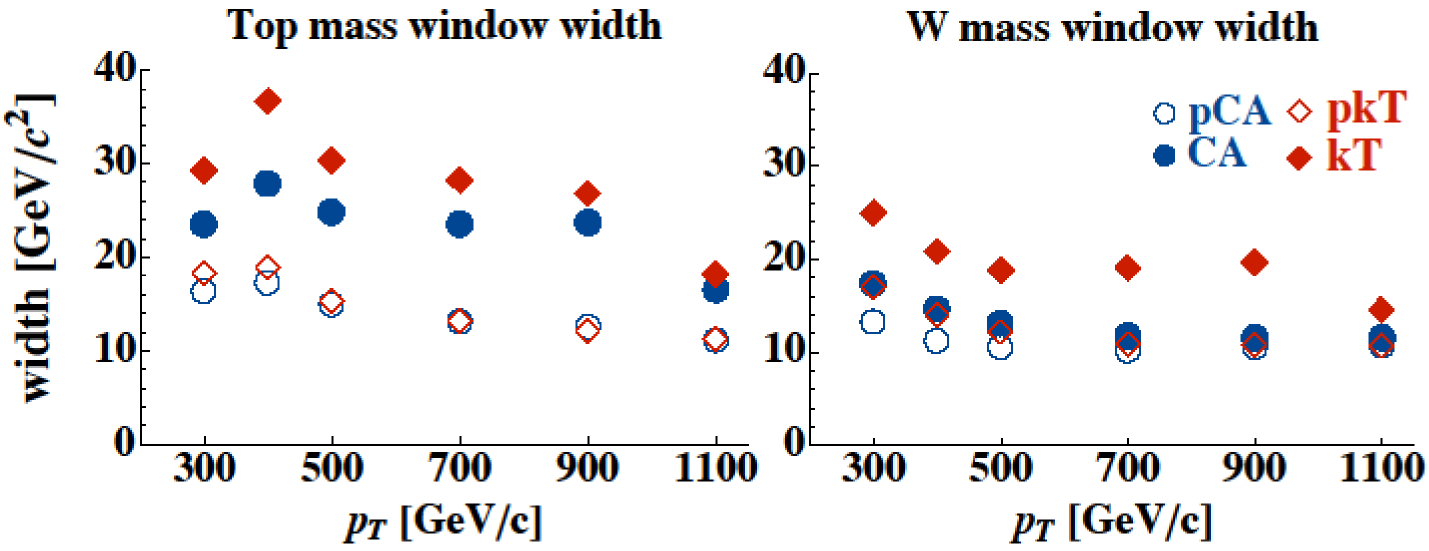
window widths for pruned (pX) and unpruned jets

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:

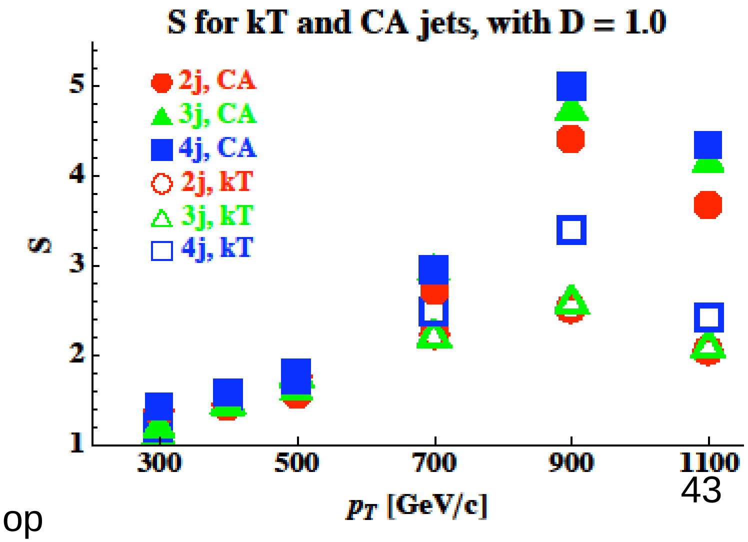
- Count top jets in signal and background samples
  - N_S : number of top jets in signal sample
  - N_B : number of top jets in background sample
  - A : unpruned algorithm pA : pruned algorithm
- Have compared pruned and unpruned samples with 3 measures:
  - *ε*, *R*, *S* - efficiency, Sig/Bkg, and 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



Summary/Conclusions:

- It will take time to understand the SM at the LHC, but we understand jets much better now than we did at the beginning of Run I
- It is essential to test and validate a variety of jet algorithms – the familiar ones (cones) and the less familiar ones (Anti-kT) – they will likely have different uses
- It is essential that the different Collaborations document the algorithms they use – and try to use the same ones some of the time
- It is essential to study and understand the role of the Underlying Event and Pile-Up in jets
- It is essential to study and understand the properties of jets – masses and substructure – validate by IDing top jets, W/Z jets in LHC data

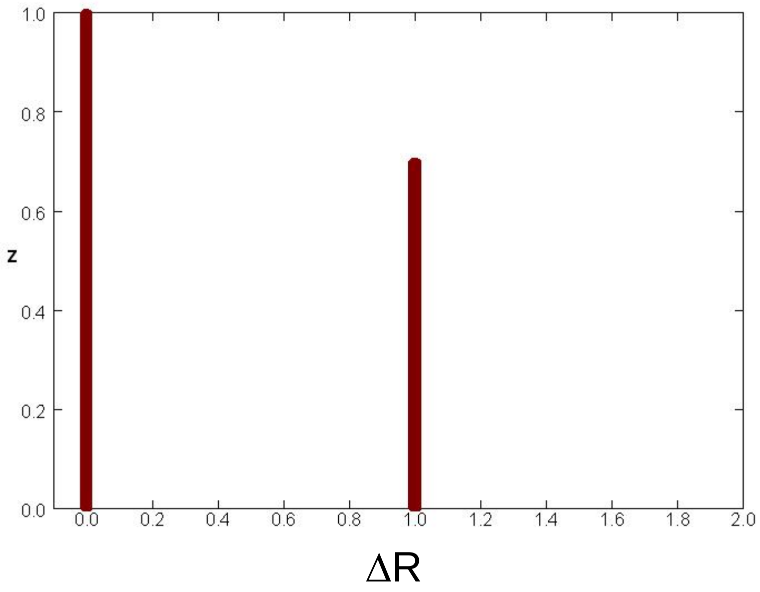
⇒ single jets will likely play a role in the search for BSM physics, along with heavy flavor tags, correlations with other jets (pair production), MET, *etc*.

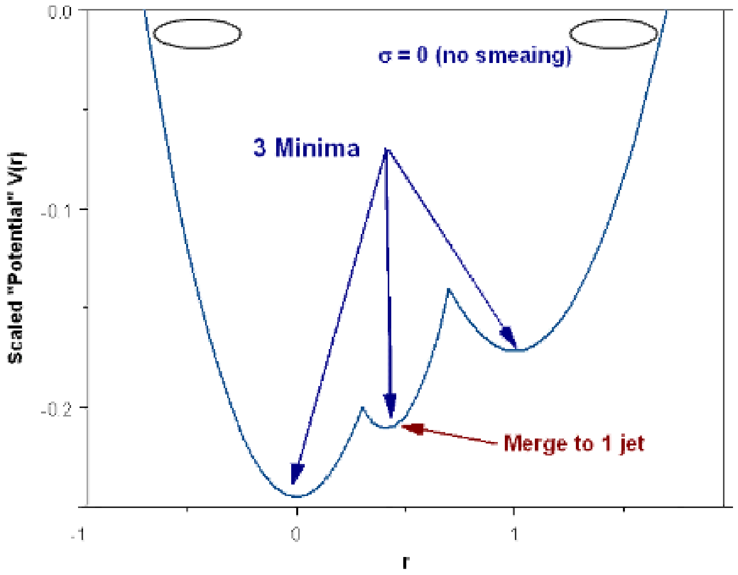
Extra Detail Slides

Simple Theory Model - 2 partons (separated by d < 2R): yield potential with 3 minima – trial cones will migrate to minima from seeds near original partons ⇒ miss central minimum

Add Midpoint cone

Snowmass Potential





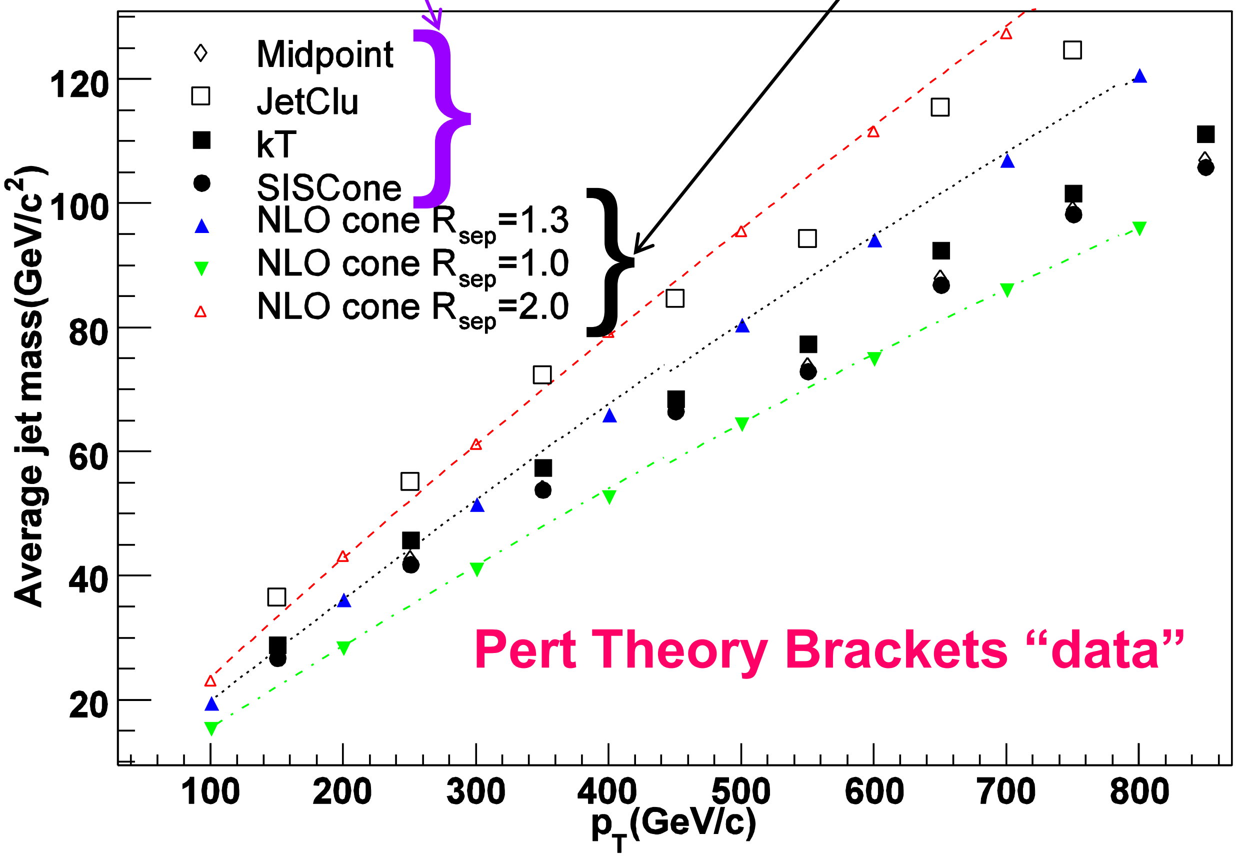
$z = p_{\min} / p_{\max}$, ΔR = separation

Smearing of order R

Compare to (simulated) LHC data: (R_{sep} scales R)

Various algorithms applied to simulated LHC data (diamond, square, circle)

NLO Cone Theory, various R_{sep} values (lines, triangles)



Pert Theory Brackets "data"

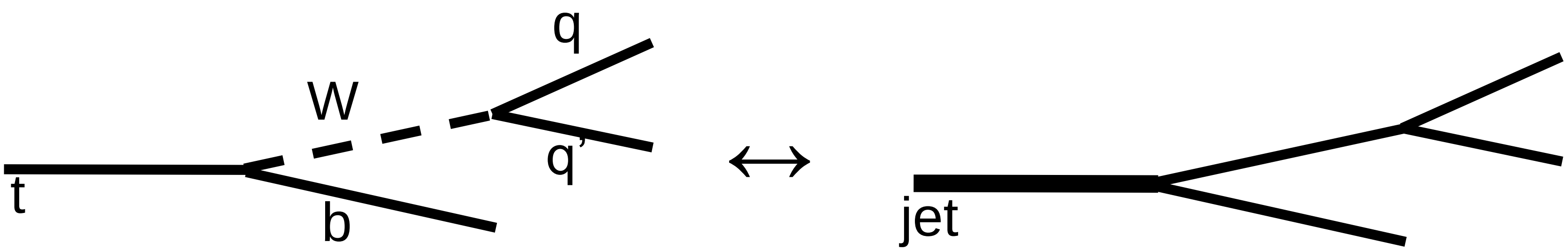
R_{sep} = 2, Snowmass

R_{sep} =1.3 EKS

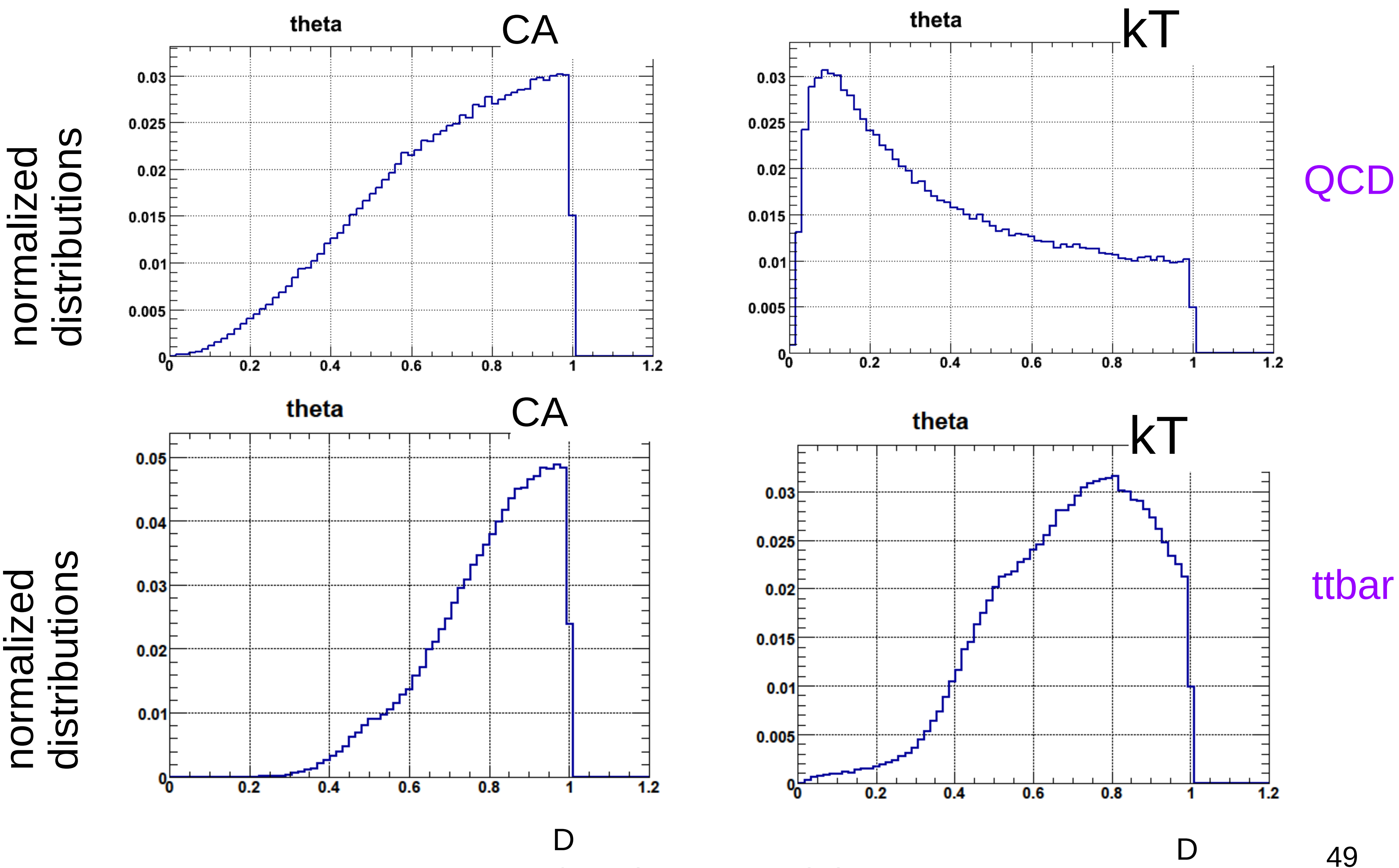
R_{sep} = 2, kT

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?

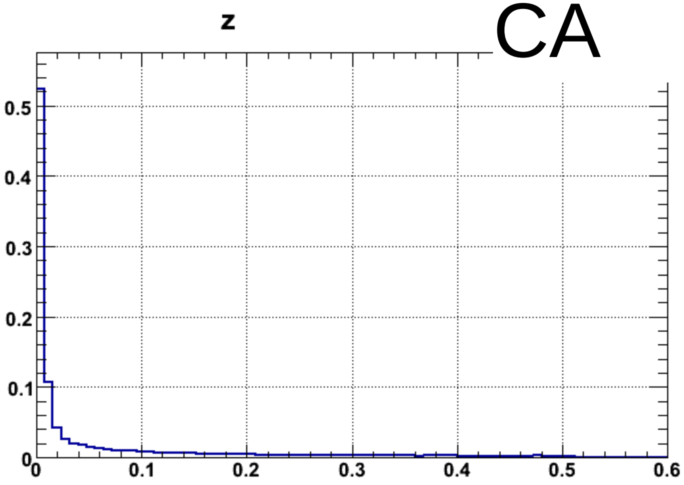


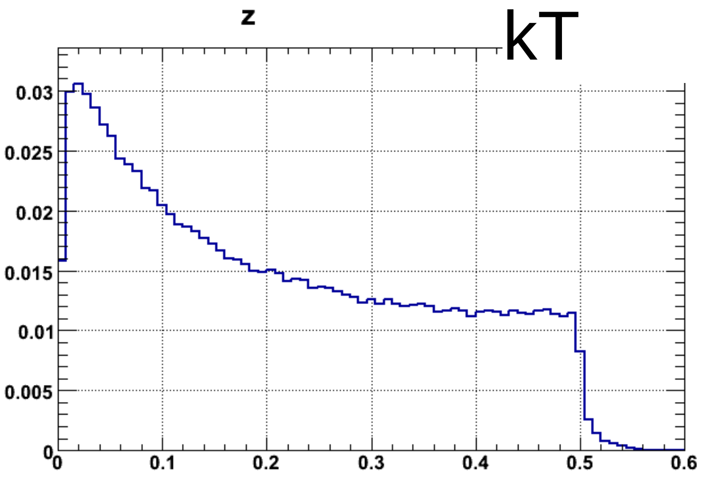
Systematics of Algorithm: θ COMPARE



Systematics of Algorithm: z COMPARE

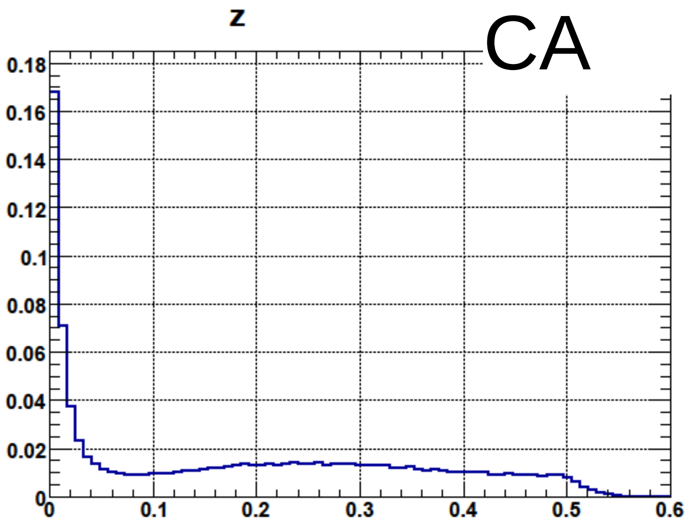
normalized distributions

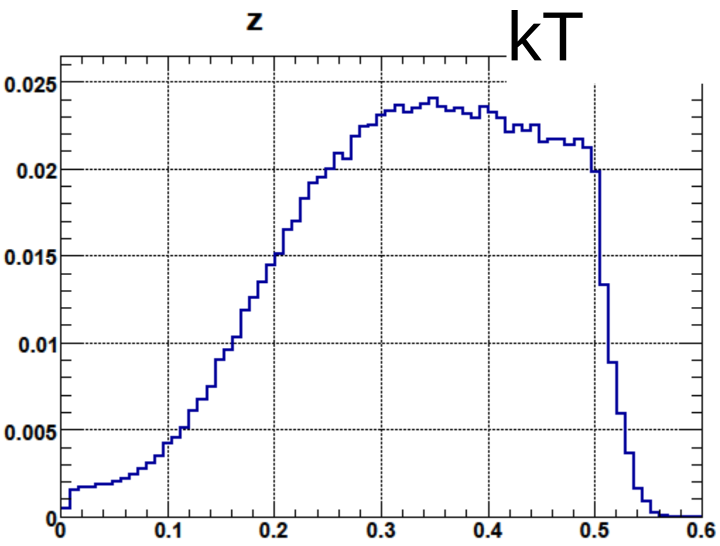




QCD

normalized distributions



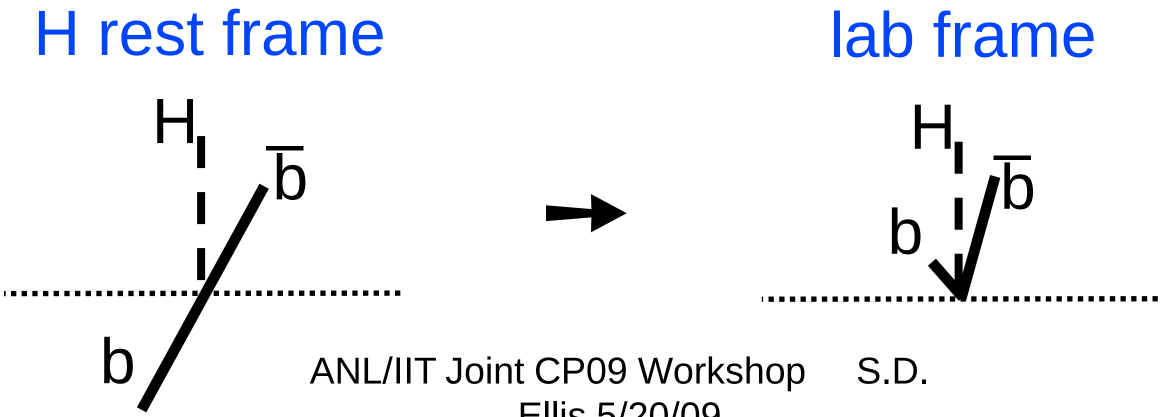


ttbar

Systematics in Heavy Particle Reconstruction

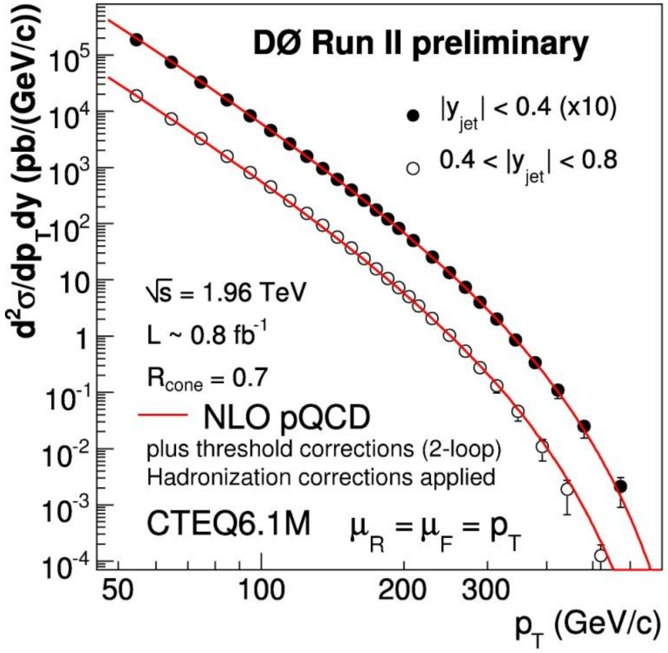
- Some kinematic regimes of heavy particle decay have a poor reconstruction rate.

- Example: Higgs decay H ➜ $b\bar{b}$ 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.

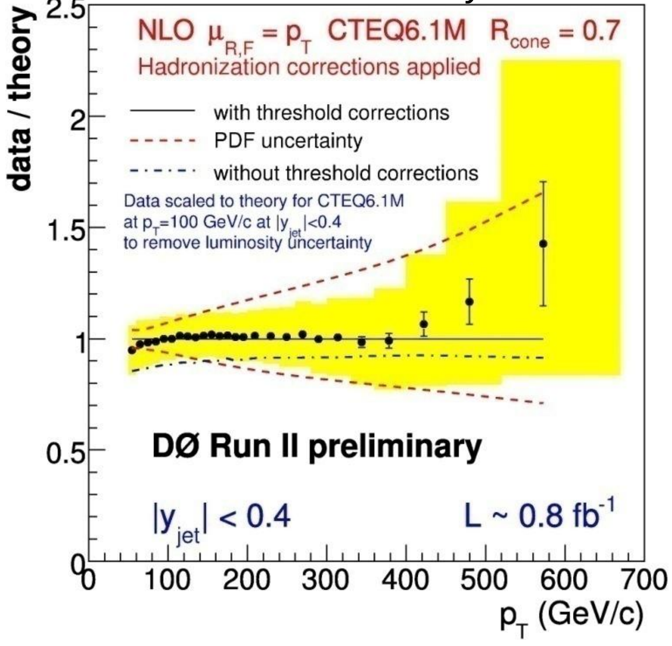


- At the Tevatron jet studies have been driven by “testing” QCD, comparing data and PertThy for inclusive jet cross section – [Cone, DØ]

Inclusive Jet cross section



Ratio data/NLO theory



Range ~ 10^8

Uncertainty ~ 10%
(1 % goal at the LHC)

Similar situation for kT jets [kT, CDF]