Jet Recombination, Substructure and Pruning or Looking for New (BSM) Physics at the LHC with Single Jets

Steve Ellis, Jon Walsh and Chris Vermilion

0903.5081
0912.0033

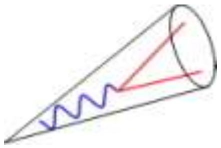
- go to tinyurl.com/jetpruning

Big Picture:

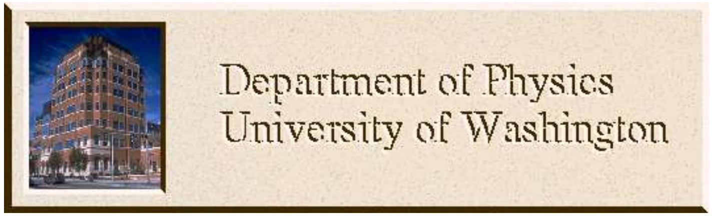
The LHC will be both very exciting and very challenging –

• most of the data will be about hadrons (jets), which have substructure

• 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



SCET - 2010
Ringberg Castle 08.04.10

Outline & Issues

- Brief review of (QCD) jets

  - defined by algorithms (no intrinsic definition)

  - jets have substructure, including masses (not just 1 parton, 1 jet)

  - need precise theoretical description for multiscale problem
  $\rightarrow$ SCET

- Focus on Recombination (kT) jets

  👍 natural substructure, but also

  👎 algorithm systematics (shaping of distributions)

  👎 contributions from (uncorrelated) ISR, FSR, UE and Pile-up

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 – except at leading order or with MC

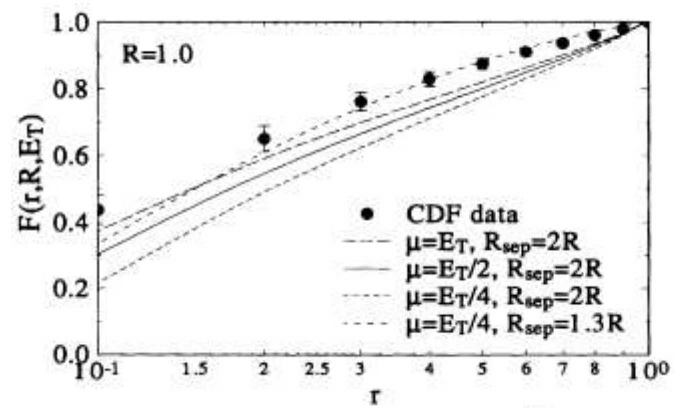


FIG. 2. $F(r,R,E_T)$ vs r for $R=1.0$, $\sqrt{s}=1800$ GeV, $E_T=100$ GeV, and $0.1<|\eta|<0.7$ with $\mu=E_T/4$, $E_T/2$, E_T compared to data from CDF [7]; the dot-dashed curve is explained in the text.

- JETS II – Run II & LHC, starting to look at structure of jets: masses and internal structure – a jet renaissance, need SCET for better tools as here

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.,

- Cone Algorithms, based on fixed geometry – focus on core of jet

  Simple, "well" suited to hadron colliders with Underlying Events (UE), but found jets can/do overlap

- Recombination (or kT) Algorithm, based on pairwise merging to undo shower

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

Outline & Issues (cont'd)

- *Search* for BSM physics in SINGLE jets at the LHC, want generic techniques

  – bumps in jet mass distribution

  ⇒ large but *Smooth* QCD background

  ⇒ bumps degraded by algorithm systematics and uncorrelated UE and Pile-Up contributions

- Need to "clean-up" the jets, e.g., PRUNE them

  - remove large angle, soft branchings

- Validate with studies of surrogate new heavy particle – top q

k_T Algorithm – focus on undoing the shower pairwise, $\Rightarrow$ Natural definition of substructure

Merge partons, particles or towers pairwise based on "closeness" defined by minimum value of k_T, *i.*e. make list of metric values (rapidity y and azimuth ϕ, p_T transverse to beam)

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

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

If $k_{T,(ij)}$ is the minimum, merge pair (add 4-vectors), replace pair with sum in list and redo list;

If $k_{T,i}$ is the minimum $\rightarrow$ *i* is a jet! (no more merging for *i*, it is isolated by *D*),

1 angular size parameter *D* *(NLO, equals Cone for D = R, R_{sep} = 1),* plus

α = *1,* ordinary k_T, recombine soft stuff first

α = *0, Cambridge/Aachen* (CA), controlled by angles only

α = *-1, Anti-k_T,* just recombine stuff around hard guys – cone-like (with seeds)

k_T Algorithm – the good and bad news

- Jet identification is unique – no merge/split stage as in Cone
- Everything in a jet, no Dark Towers as in Cone
- Resulting jets are more amorphous, energy calibration difficult (need area for subtraction for UE?), Impact of UE and pile-up not so well understood, especially at LHC
- Analysis can be very computer intensive (time grows like N^3, recalculate list after each merge)
- New version (Cacciari, Salam & Soyez) goes like N ln N (only recalculate nearest neighbors) , plus scheme for doing UE correction
- They have been used and understood at the Tevatron
- Using Anti-k_T at LHC, which is not so well understood, nor does it provide useful substructure, but could find jets with Anti-k_T and substructure with CA/k_T

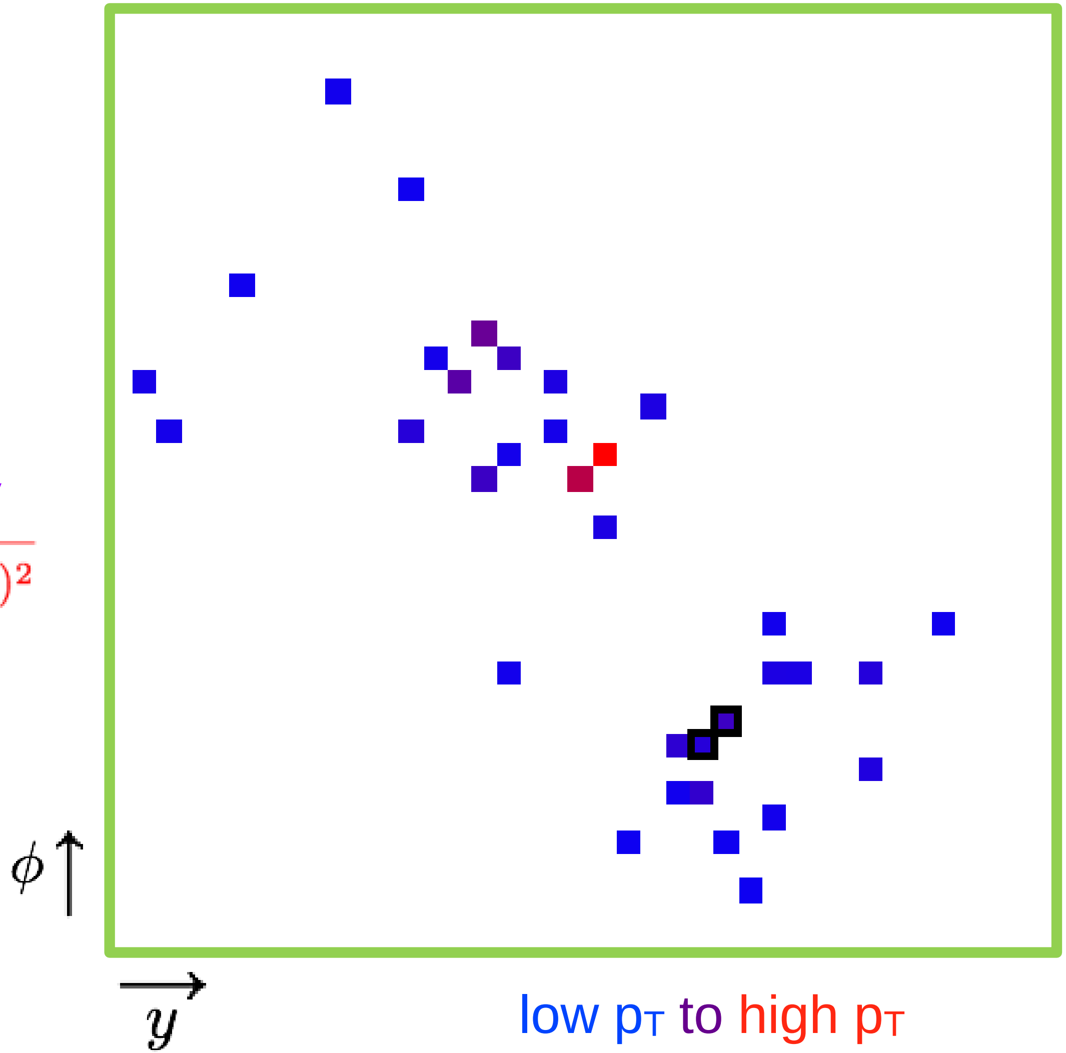
k_T Algorithm – in action, here CA algorithm on QCD jet

Think of starting with calorimeter cells, recombine "closest" pair at each step leading to larger p_T

For CA close in quantity

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}$$

(0.05 x 0.05) Cells with E > 1 GeV

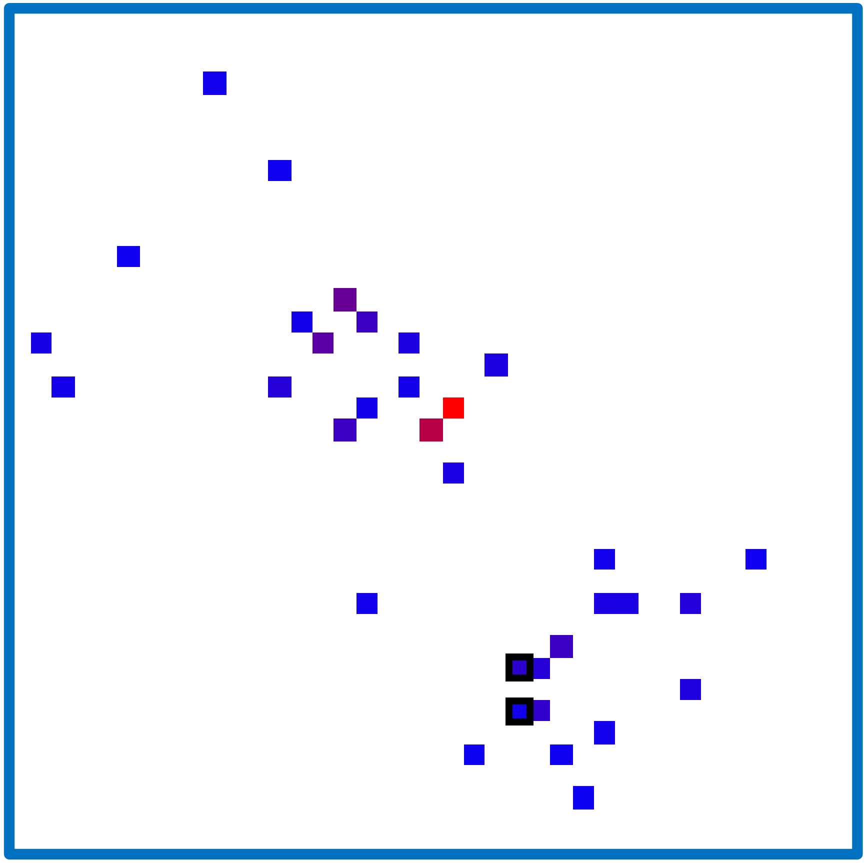


low p_T to high p_T

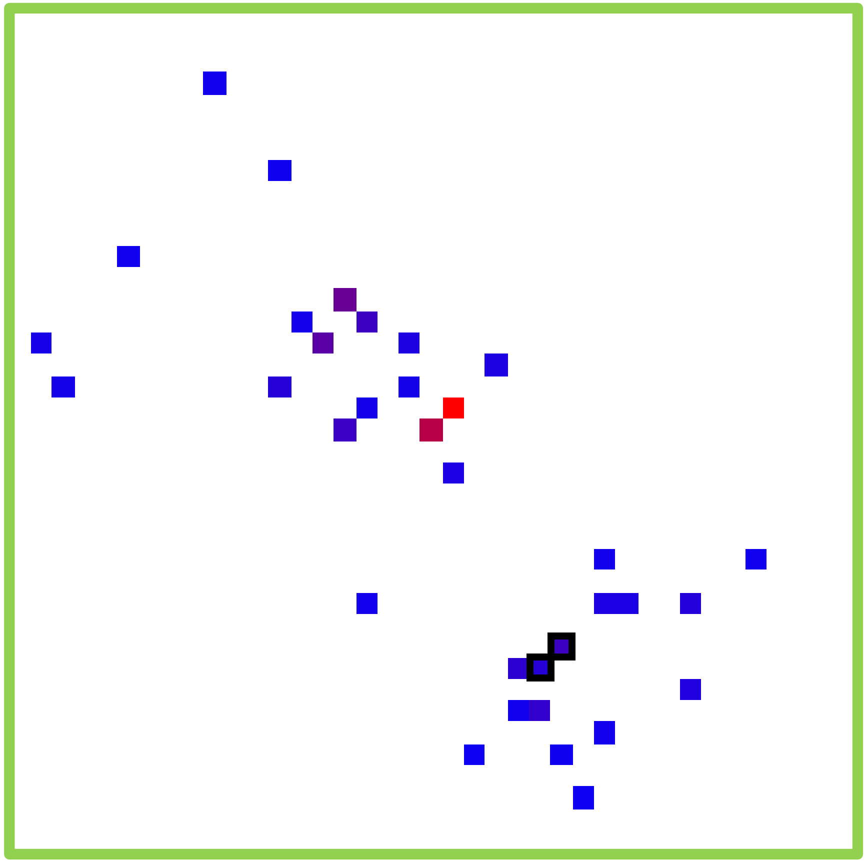
Animations from the studios of J. Walsh

Note: the details of the substructure (at each step) depend on the algorithm

kT (pT and angle)



CA (just angle)



Finding Heavy Particles with Jets - Issues

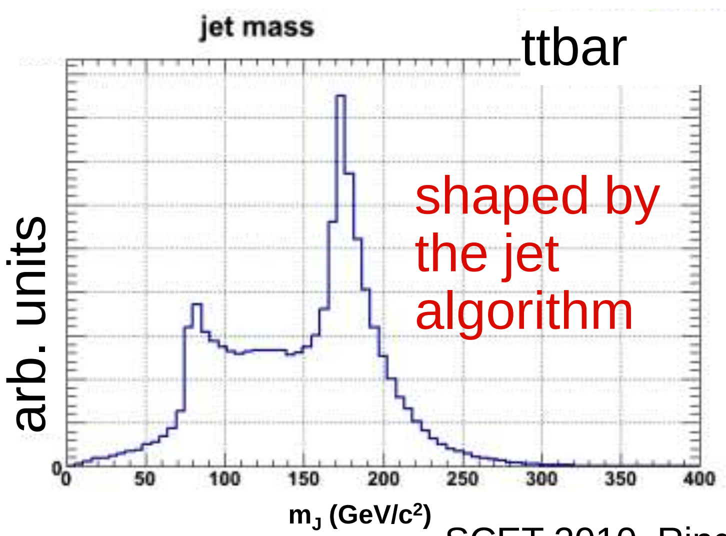
- QCD multijet production rate >> production rate for heavy particles

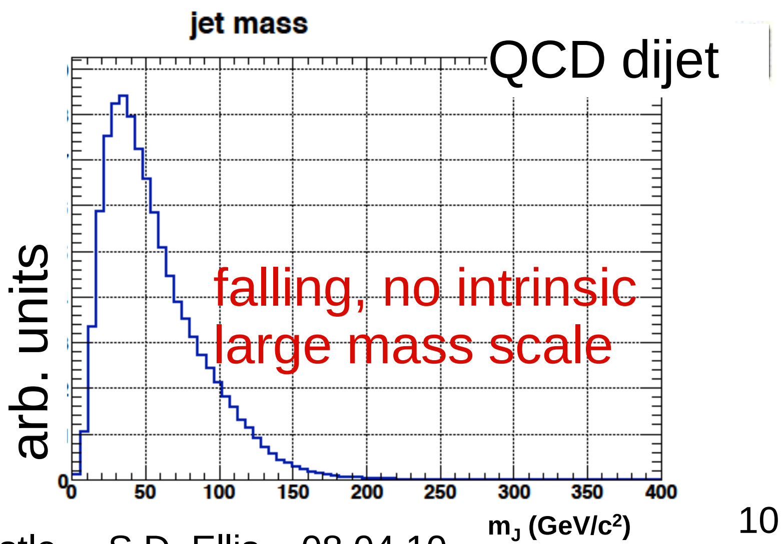
- In the jet mass spectrum, production of non-QCD jets may appear as local excesses (bumps!) but must be enhanced using analyses

- Use jet substructure as defined by recombination algorithms to refine jets

- Algorithm will systematically shape distributions

- Use top quark as surrogate new particle. $\sigma_{ttbar} \approx 10^{-3} \sigma_{jj}$

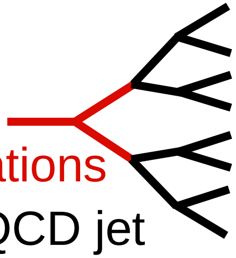




Reconstruction of Jet Substructure – QCD vs Heavy Particle

- 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.



interpret last recombinations as a heavy particle or QCD jet

The algorithm metric affects the substructure - introduces bias

Systematics of the Jet Algorithm

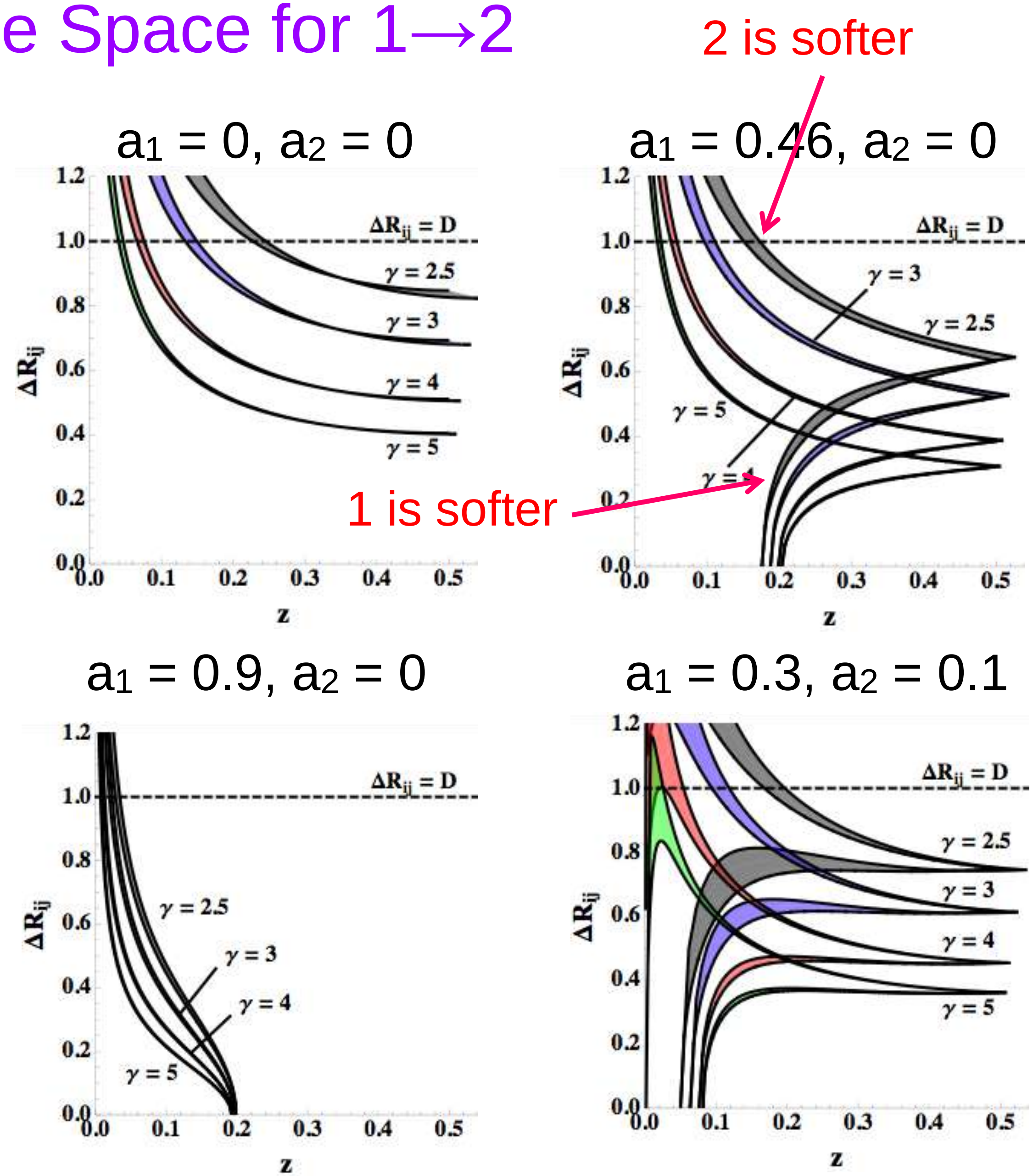
- Consider generic recombination step: *i,j* ➜ *p*
- Useful variables: $z = \frac{\min(p_{T_i}, p_{T_j})}{p_{T_p}}$ ΔR_{ij}
  (Lab frame)
- Merging metrics: $\rho_{kT} = p_{Tp} z \, \Delta R / D$ $\rho_{CA} = \Delta R / D$
- Daughter masses (scaled by jet mass) : $a_{i,j} \equiv \frac{m_{i,j}}{m_J}$
- 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.

Phase Space for 1→2

- The allowed phase space in ΔR, z for a fixed γ (m_J and p_T) is nearly one-dimensional

$$x_J \equiv \frac{m_J^2}{p_{T,J}^2} = \frac{1}{\gamma^2 - 1} \approx z(1-z)\Delta R^2$$

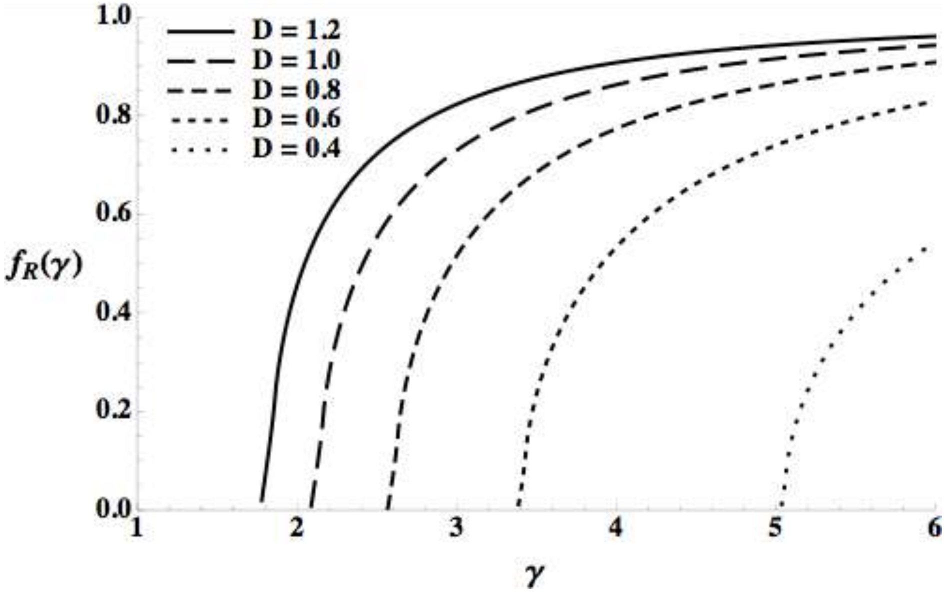
- QCD and decays will weight the phase space differently
- Cutoffs on variables set by the kinematics, not the dynamics
- Sample of phase space slices for different subjet masses,



1→2 Decay in a Jet

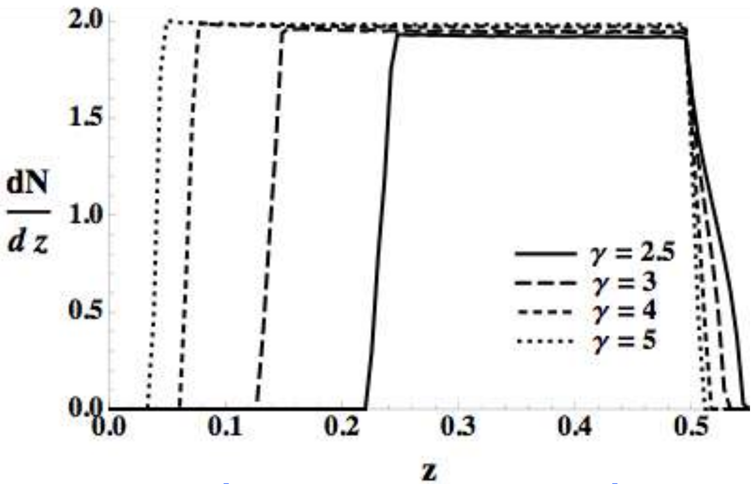
- Goal is to identify jets reconstructing a heavy particle and separate them from QCD jets
- Consider a 1→2 decay (J→ 1,2) reconstructed in a jet, massless daughters ($a_1 = a_2 = 0$)
- Requirement to be in a jet: $\Delta R_{12} < D$ - algorithm independent
- Look at the decay in terms of the algorithm variables

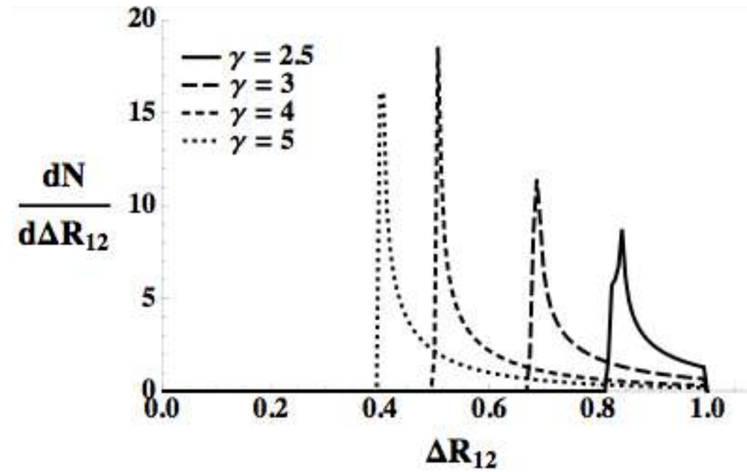
Large *D* is needed to reconstruct jets with a lower boost - use $D = 1.0$, sweet spot $\gamma \sim 3$, $p_T \sim 3\, m_J$, $x_J \sim 0.1$



Recall for QCD jet $p_T \sim 5 \langle m_J \rangle$

1→2 Decay in a Jet (unpolarized)



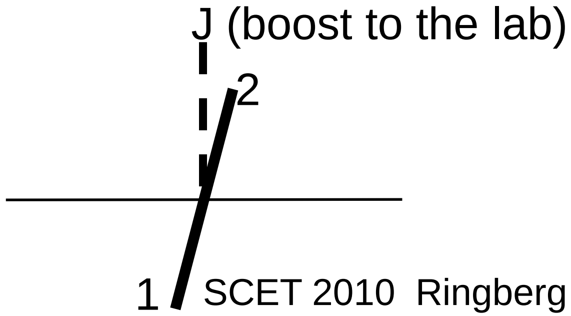


No enhancement at the lower limit in z - unlike QCD

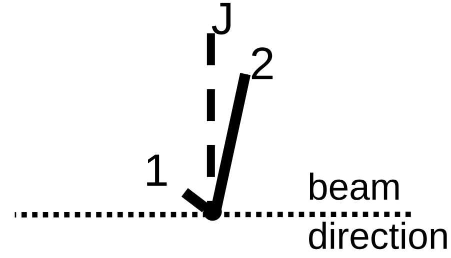
Enhancement at the lower limit for ΔR_{12} - like QCD

Decays not reconstructed: small z, large ΔR_{12}

J rest frame

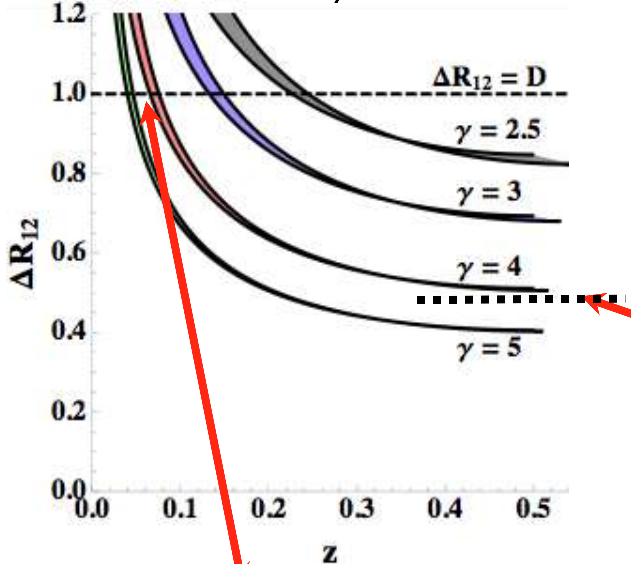


lab frame

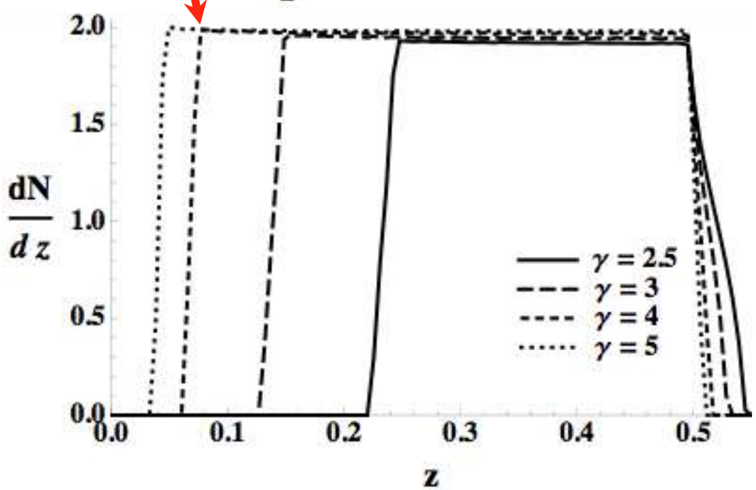


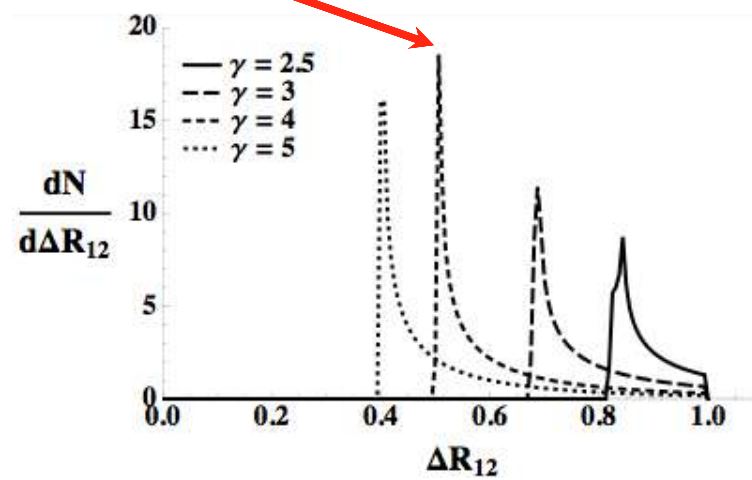
1→2 Decay in a Jet

$a_1 = 0$, $a_2 = 0$



Cutoffs are set by the kinematics - same between QCD and decay with fixed γ





No enhancement at the lower limit in z - unlike QCD

Enhancement at the lower limit for ΔR_{12} - like QCD

QCD Splittings

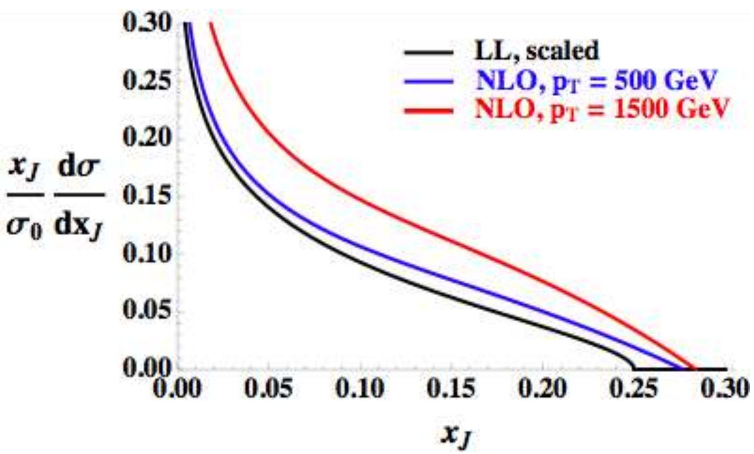
Take a leading-log approximation of QCD: $\frac{d\sigma}{\sigma_0} \propto \frac{dz}{z}\frac{d\Delta R_{12}}{\Delta R_{12}}$

For small angles - good approximation for a splitting in a jet:

$$x_J \equiv \frac{m_J^2}{p_{T,J}^2} = \frac{1}{\gamma^2 - 1} \approx z(1-z)\Delta R^2$$

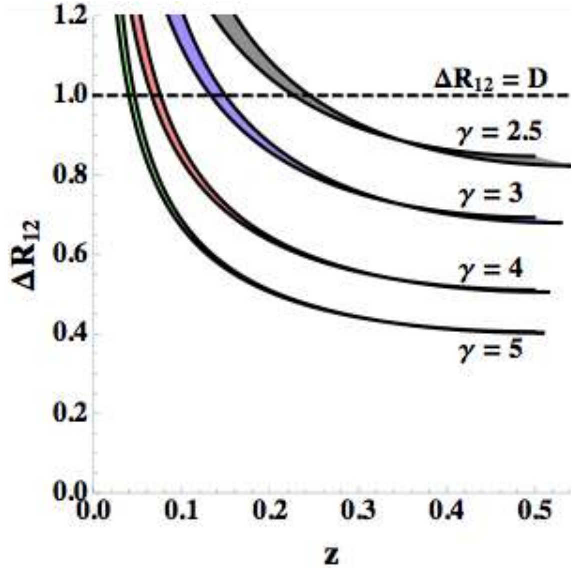
This lets us fix x_J (or γ). Distribution in x_J :

$$\frac{d\sigma}{dx_J} \propto \int \frac{dz\,d\Delta R_{12}}{z\Delta R_{12}}\delta\left(x_J - z(1-z)\Delta R_{12}^2\right) = -\frac{\ln\left(1-\sqrt{1-\frac{4x_J}{D^2}}\right)}{2x_J}\Theta(D^2 - 4x_J)$$



QCD Splittings: ΔR_{12} and z

$a_1 = 0$, $a_2 = 0$



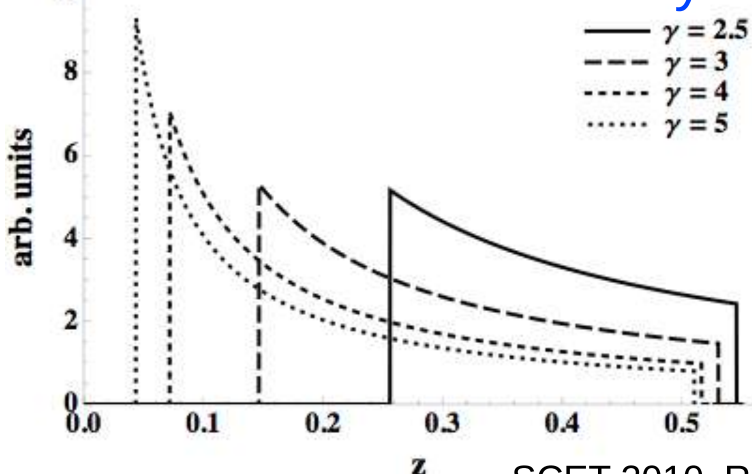
Fix $\gamma(x_J)$, find distributions in ΔR_{12} and z

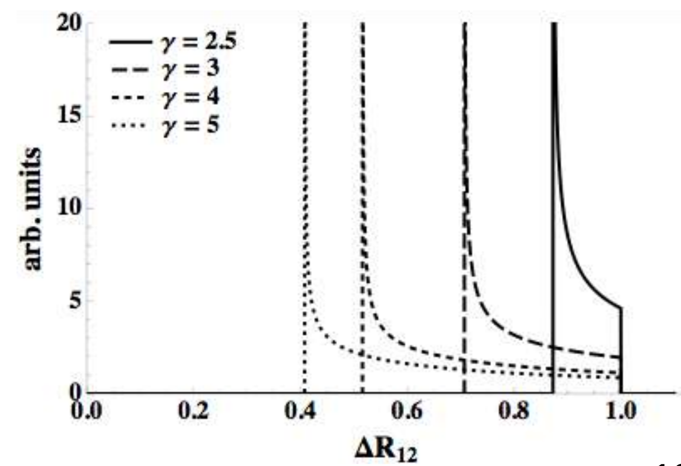
Limits set by the kinematics

QCD will have many more soft (small z) splittings than decays do - QCD splittings are small z, small x_J enhanced

Enhancement at the lower limit in ΔR_{12} - like decays

Enhancement at the lower limit in z - unlike decays





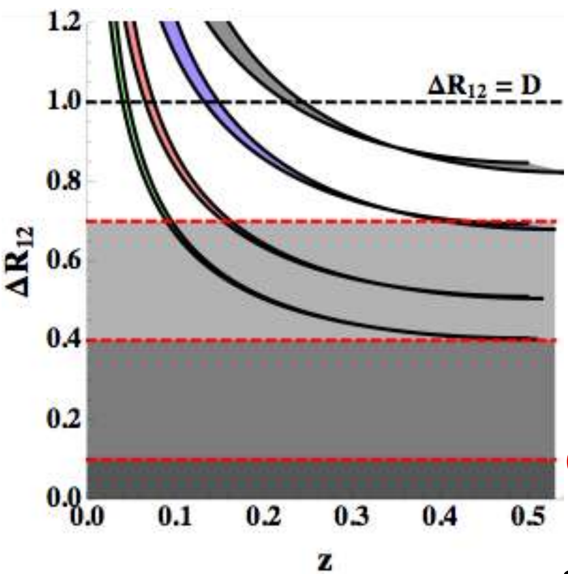
Summary of Dynamics of QCD Vs Decays:

- Distributions in ΔR very similar (for fixed boost)
- QCD enhanced at small z, x_J
- Will these be represented in the *last recombinations* of a jet?

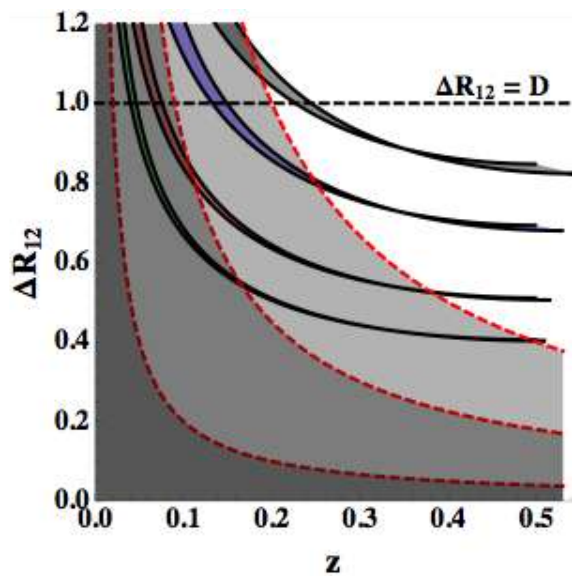
Effects of the Jet Algorithm – Algorithm Bias

- Recombination metrics:
$$\rho_{kT}(i,j) = p_{Tp} z_{ij} \, \Delta R_{ij} / D$$
$$\rho_{CA}(i,j) = \Delta R_{ij} / D$$
- Recombinations are almost always monotonic in the metric
  - The algorithm cuts out phase space in (z, ΔR_{ij}) as it proceeds
- Certain decays will be reconstructed earlier in the algorithm, or not at all

CA



k_T

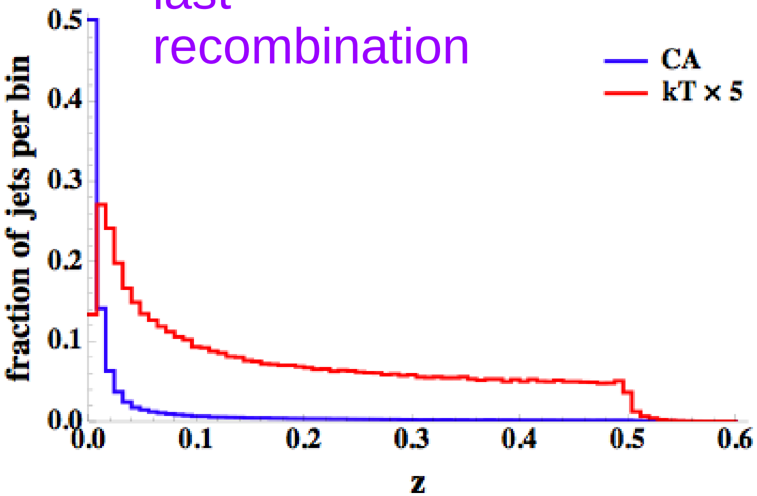


Typical Recombinations

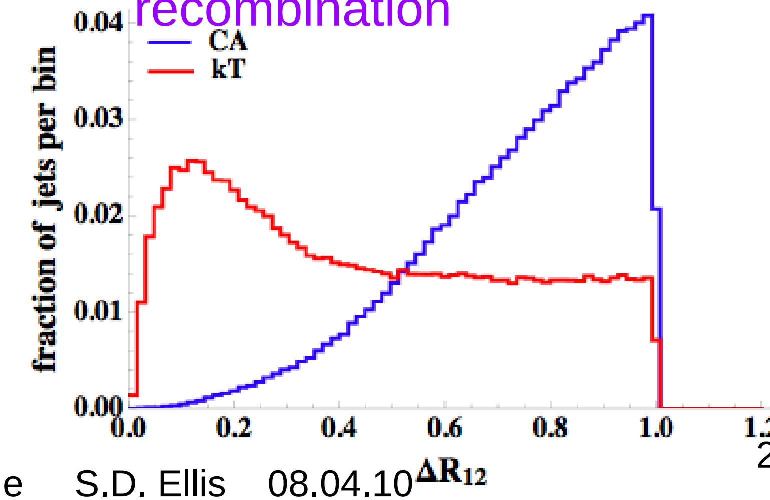
- *Late* recombinations are set by the available phase space
  - For CA, ΔR must be near D, and the phase space tends to create small z recombinations
  - For kT, $z\ \Delta R$ will be larger, with a p_T dependent cut
- The soft (small z) radiation is recombined *earlier* in k_T, meaning it is harder to identify - leads to poorer mass resolution

Matched QCD sample (2, 3, 4 partons) from MadGraph/Pythia, jet p_T between 500-700 GeV

last recombination

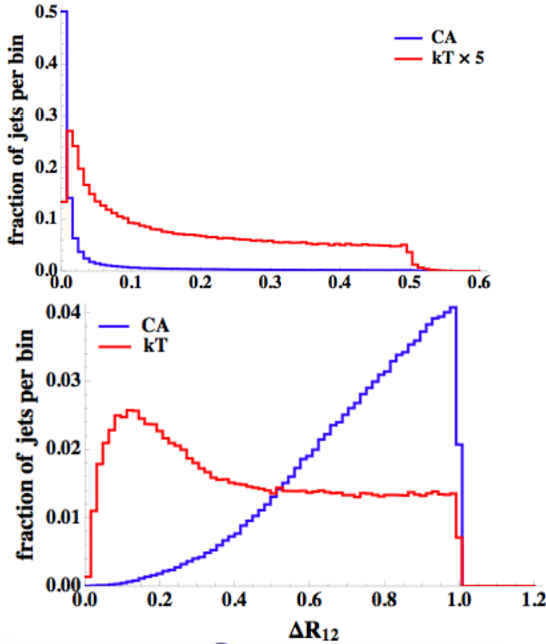


last recombination

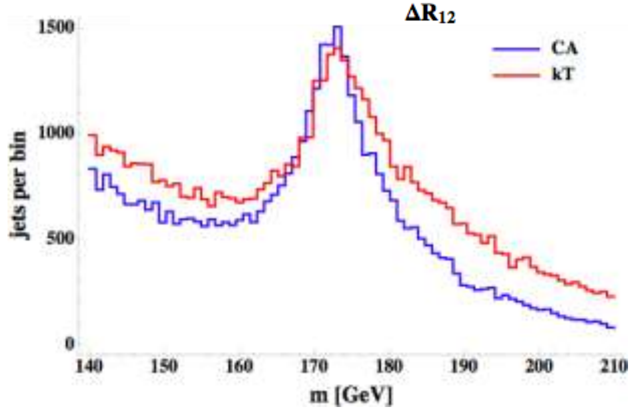


Comparing CA and kT:

Matched QCD, jet p_T between 500-700 GeV



- Final recombinations for CA not QCD-like
  - No enhancement at small ΔR
- Final recombinations for kT more QCD-like
  - Enhanced at small *z and* ΔR
- k_T has poorer mass resolution
  - Soft objects recombined early in algorithm - more merged



tt sample from MadGraph/Pythia jet p_T between 500-700 GeV

Summary: Identifying Reconstructed Decays in Jets

- Reconstruction of a decay can be hidden in the substructure
- Small z recombination unlikely to accurately give decay
- Small z recombinations also arise from UE and pile-up
- The jet algorithm significantly shapes the jet substructure – less so for k_T but has poorer mass resolution
- Proposing a method to deal with these issues: modify the jet substructure to reduce algorithm effects and improve mass resolution, background rejection, and heavy particle identification - ***pruning***

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

Pruning :

Procedure:

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

- Rerun with kT or CA algorithm, but at each recombination test whether soft – large angle:
  - $z < z_{\mathrm{cut}}$ and $\Delta R_{ij} > D_{\mathrm{cut}}$

- 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

Other Jet Grooming techniques:

- *MassDrop Filtering* - Butterworth, Davison, Rubin & Salam (0802.2470) - reprocess jets (top down) to find (fixed number) of subjets (2 or 3)

- *Top Tagging* - Kaplan, Rehermann, Schwartz & Tweedie (0806.0848) – look for specific substructure of tops (3 or 4)

  Thaler & Wang (0806.0023)

- *Trimming* – Krohn, Thaler & Wang (0912.1342) – reprocess to find primary subjets (pT > pTcut, any number of subjets)

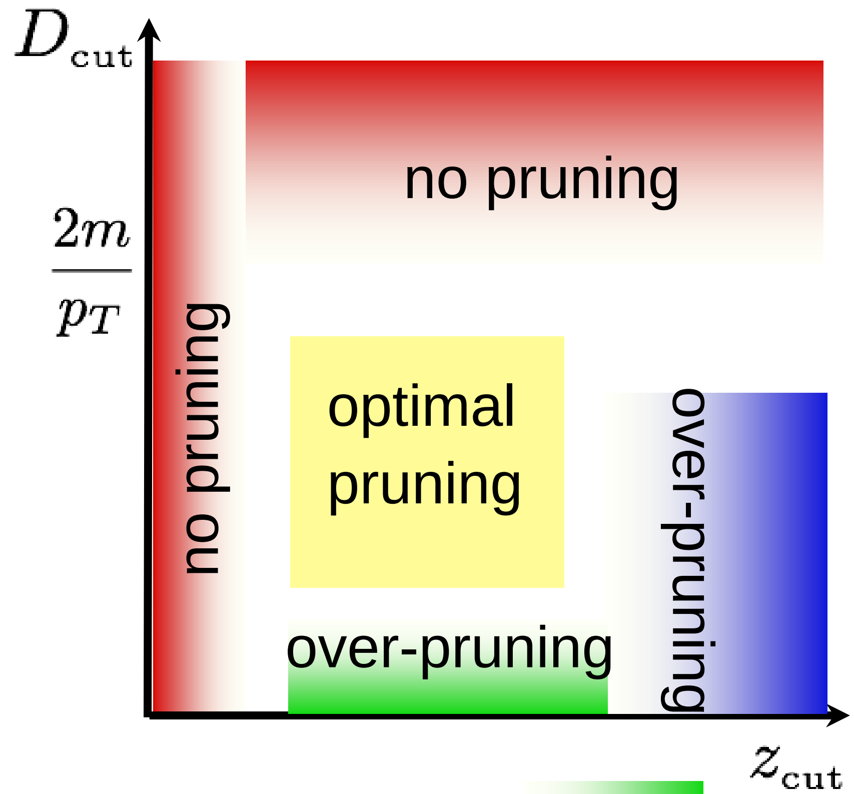
Choices of the pruning parameters

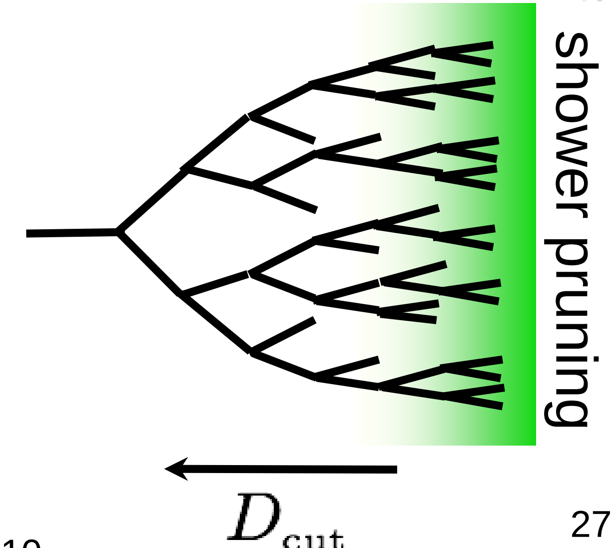
After studies we choose:

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

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

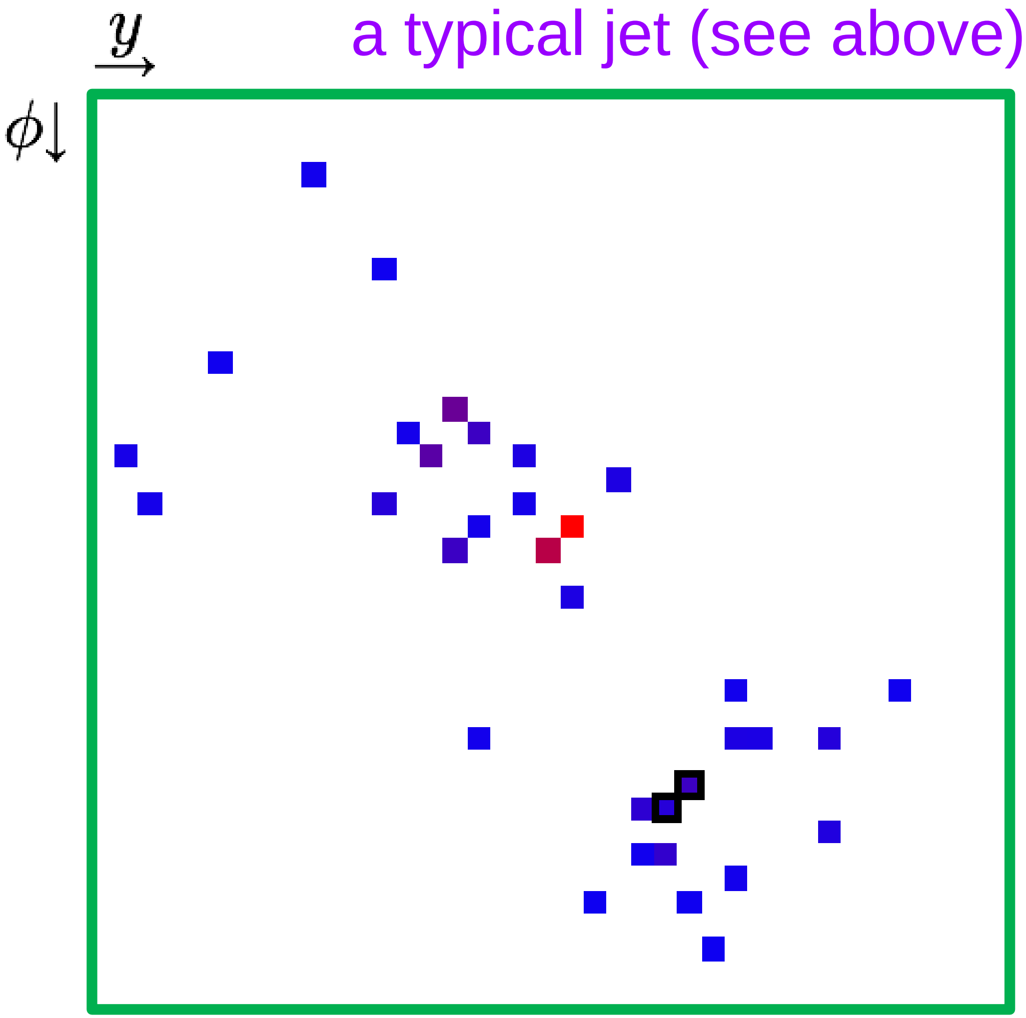
★ $m_J/P_{T,J}$ is *IR safe* measure of opening angle of found jet





Pruning in Action

a typical jet (see above)



p_T: 600 → 590 GeV

mass: 170 → 160 GeV

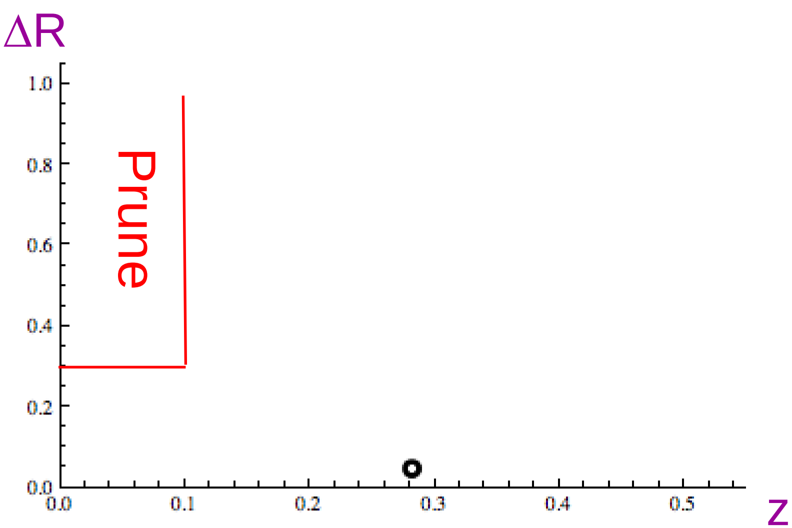
Pruning of a QCD jet near the top mass with the CA algorithm

Red is higher p_T

Blue is lower p_T

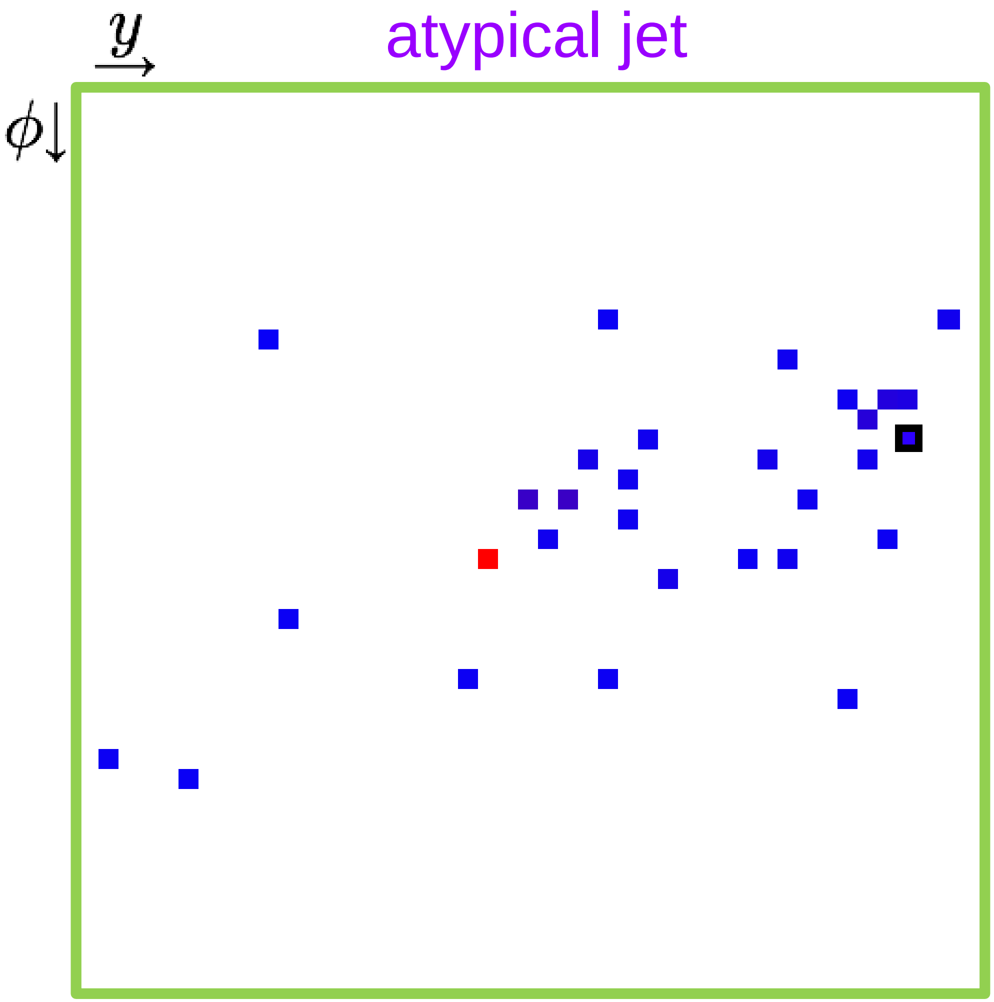
Green X is a pruning

Start with cells with energy > 1 GeV



Pruning in Action

atypical jet



p_T: 600 → 550 GeV

mass: 180 → 30 GeV

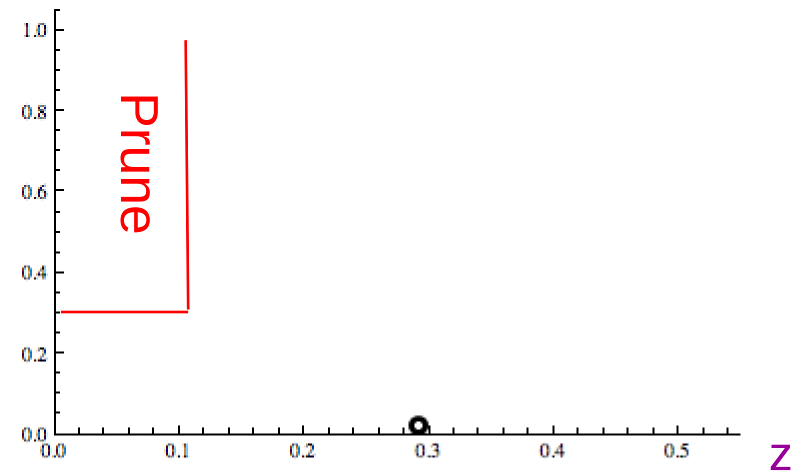
Pruning of a QCD jet near the top mass with the CA algorithm

Red is higher p_T

Blue is lower p_T

Green X is a pruning

Start with cells with energy > 1 GeV



Impact of Pruning – qualitatively just what we want!

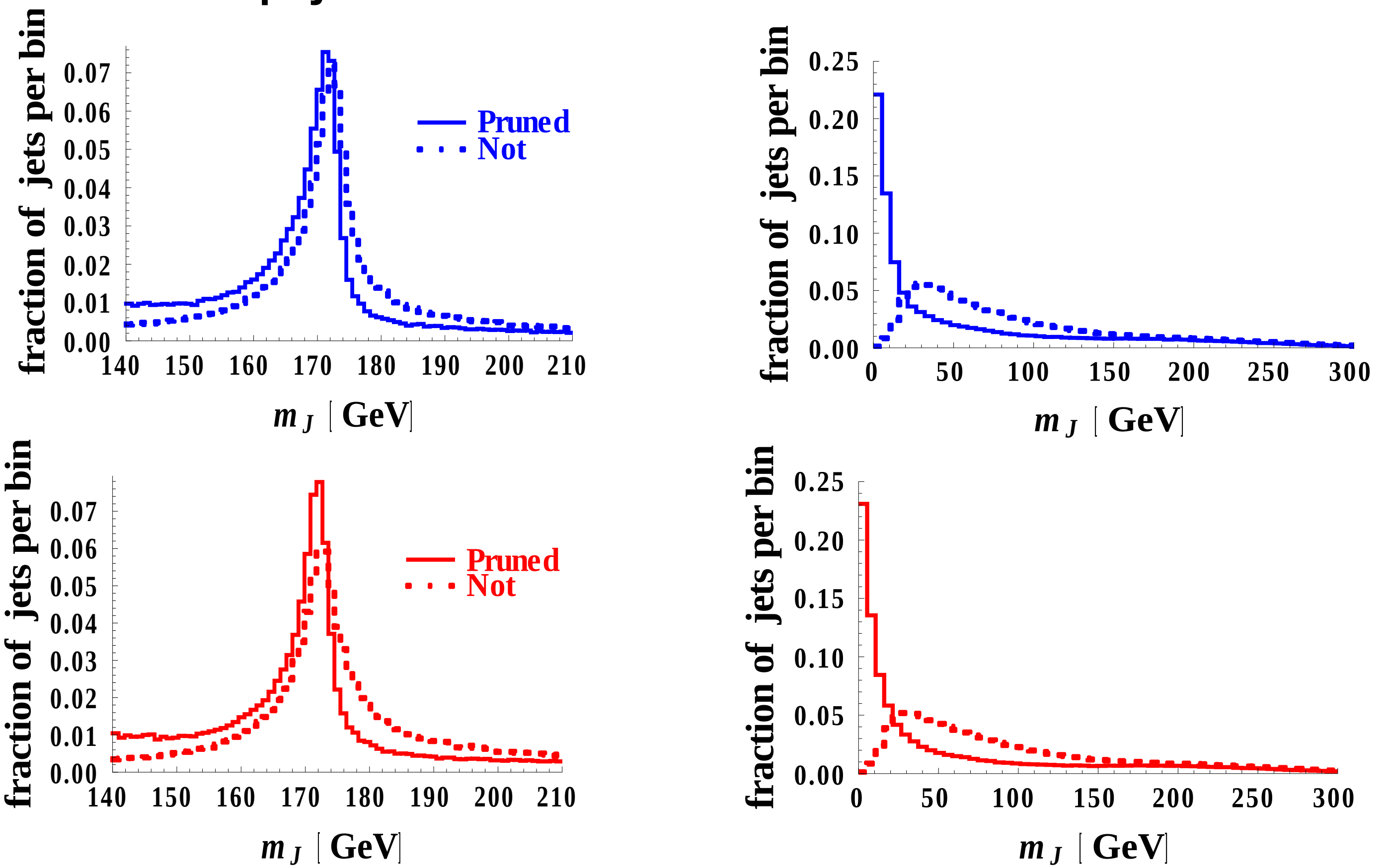
⇒The mass resolution of *pruned* top jets is narrower

⇒ *Pruned* QCD jets have lower mass, sometimes much lower

Top jets | QCD jets

CA

kT



500 < p_T < 700 GeV

Test Pruning in more detail:

- 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, jets found with FastJet

- Look at several quantities before/after pruning:

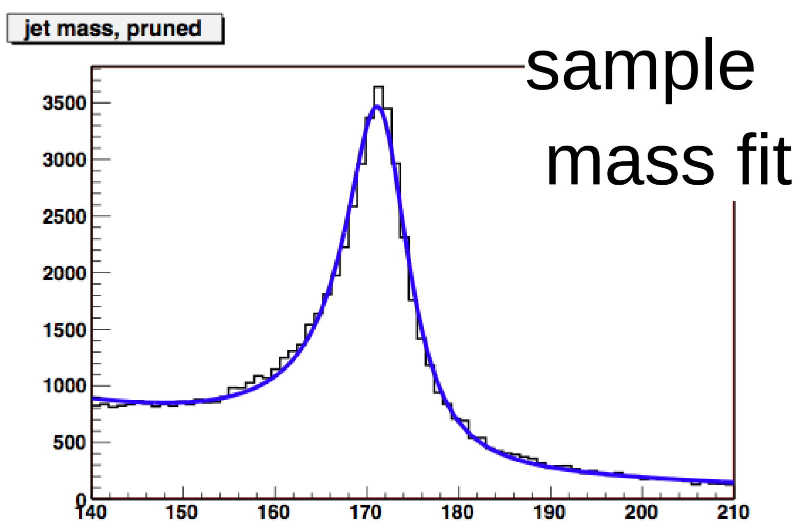
$\Rightarrow$ Mass resolution of reconstructed tops (width of bump), small width means smaller background contribution

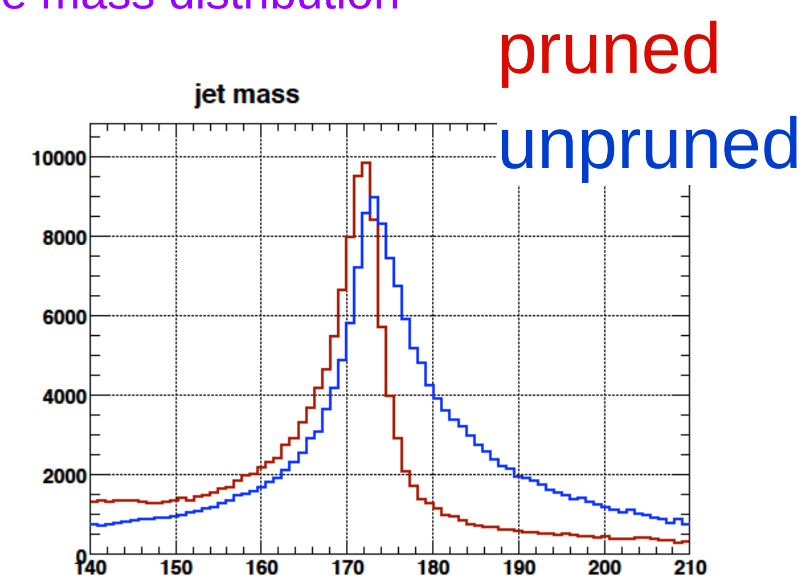
  - pT dependence of pruning effect
  - Dependence on choice of jet algorithm and angular parameter D
  - UE dependence

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

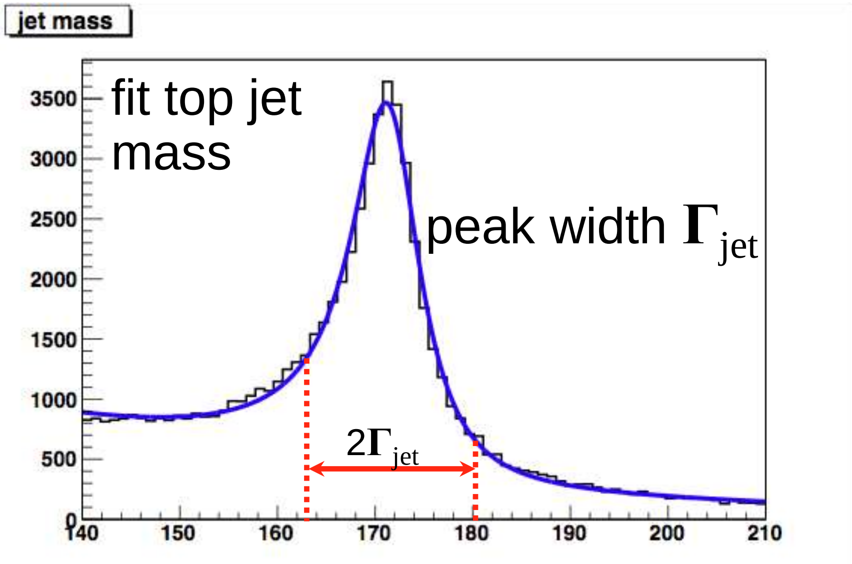
- Defining the top, W mass windows:
  - **Fit** the observed jet mass and subjet mass distributions with (asymmetric) Breit-Wigner plus continuum $\rightarrow$ widths of the peaks
  - The top and W windows are defined separately for pruned and not pruned - test whether pruning is narrowing the mass distribution





Defining Reconstructed Tops

fit mass windows to identify a reconstructed top quark



peak function: skewed Breit-Wigner

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

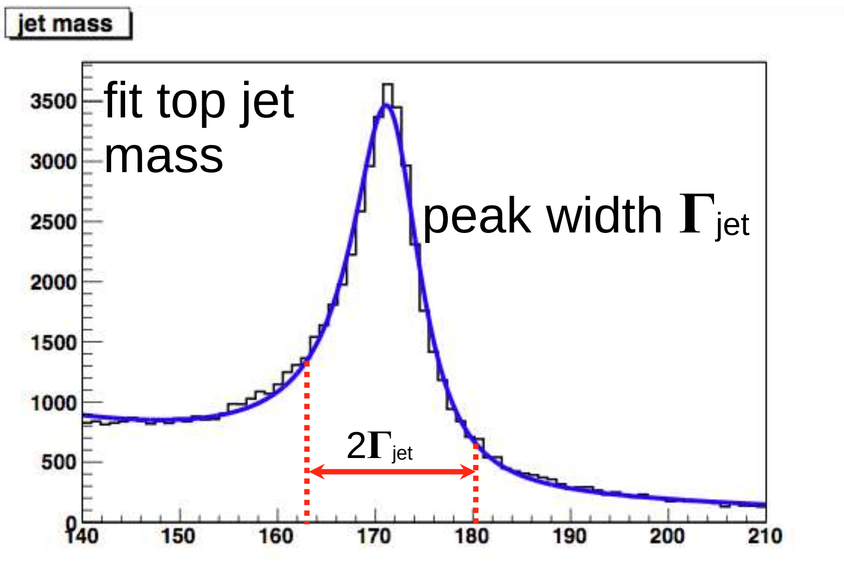
plus continuum background distribution

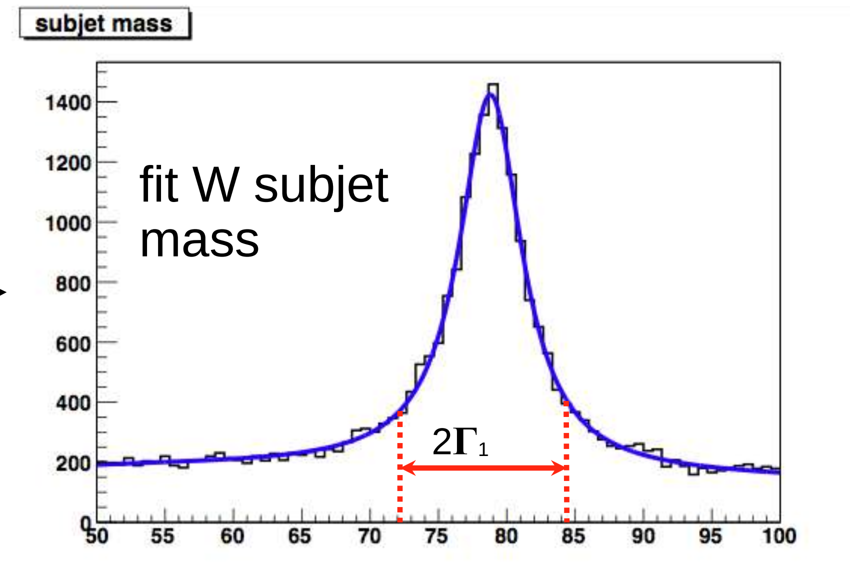
$$\frac{c}{m} + \frac{d}{m^2}$$

Defining Reconstructed Tops

fit mass windows to identify a reconstructed top quark

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

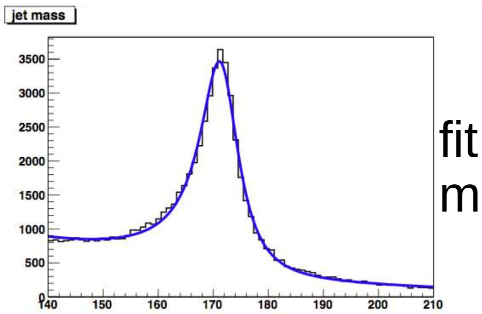




Defining Reconstructed Tops

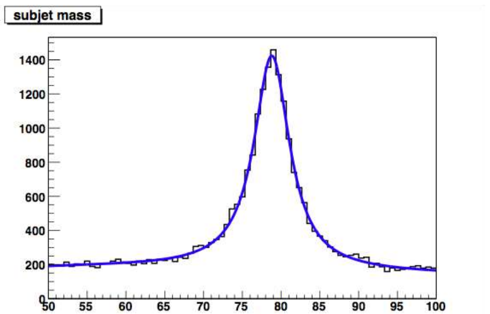
fit mass windows to identify a reconstructed top quark

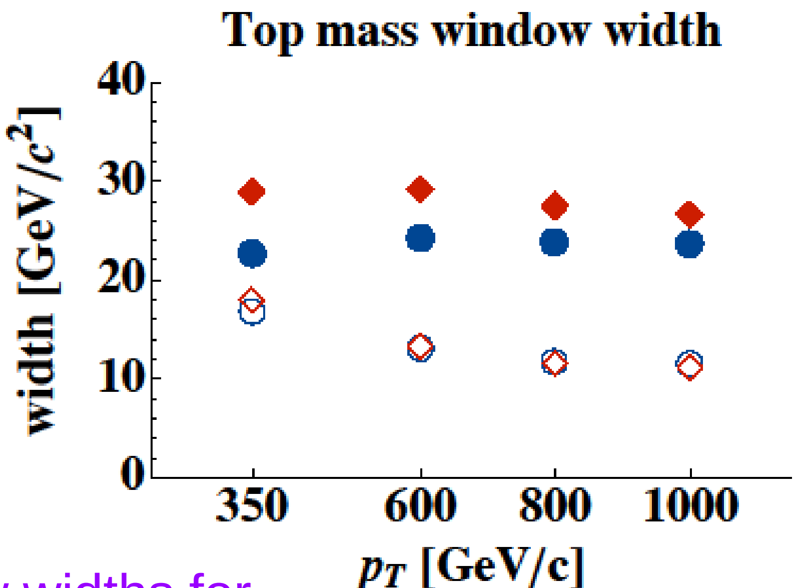
cut on masses of jet (top mass) and subjet (W mass)

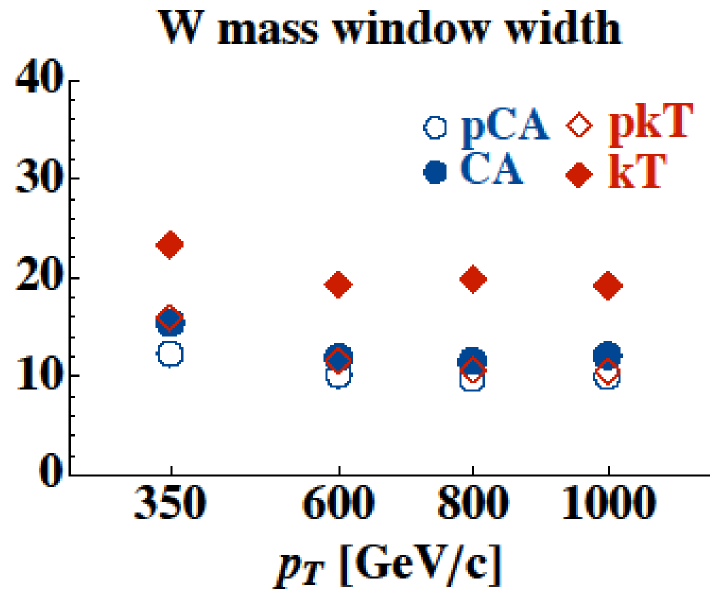


fit top jet mass

fit W subjet mass







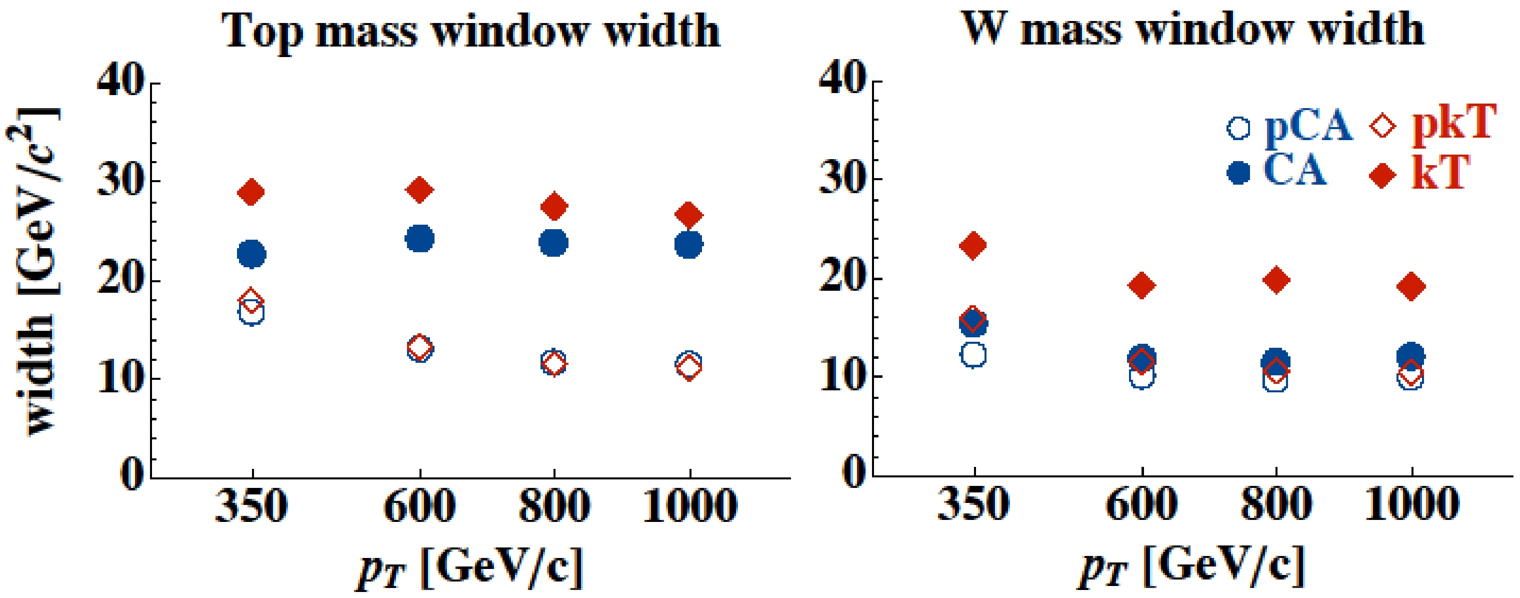
window widths for pruned (pX) and unpruned jets

Mass Windows and Pruning - Summary

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



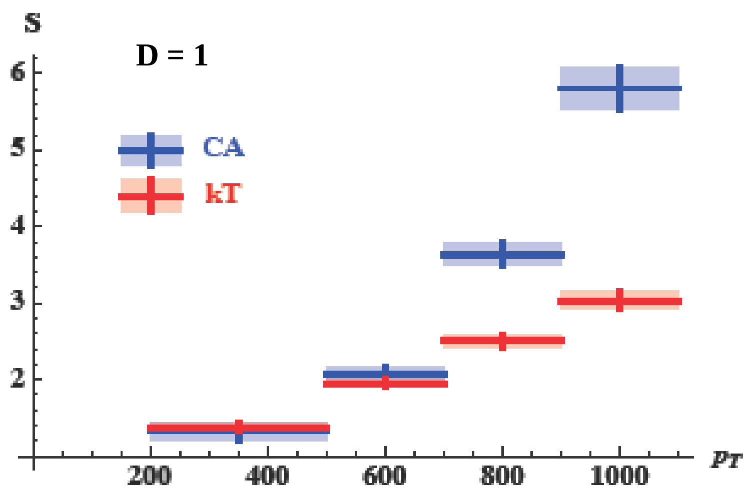
Statistical Measures:

- Count top jets in signal and background samples in fitted bins
  - 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 $\text{Sig/Bkg}^{1/2}$

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

Here focus on *S*

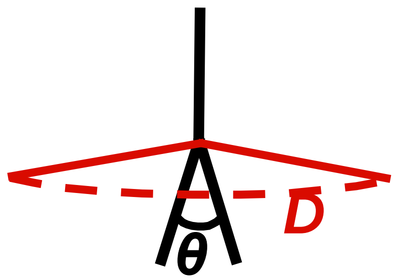
S > 1 (improved likelihood to see bump if prune), all pT, all bkgs, both algorithms



Heavy Particle Decays and *D*

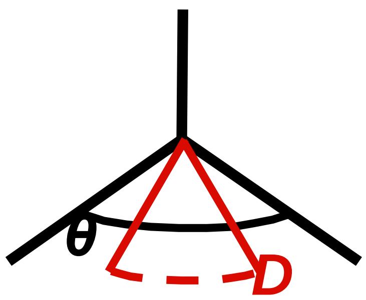
See also Krohn, Thaler & Wang (0903.0392)

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



D > *θ*

- lets in extra radiation
- QCD jet masses larger

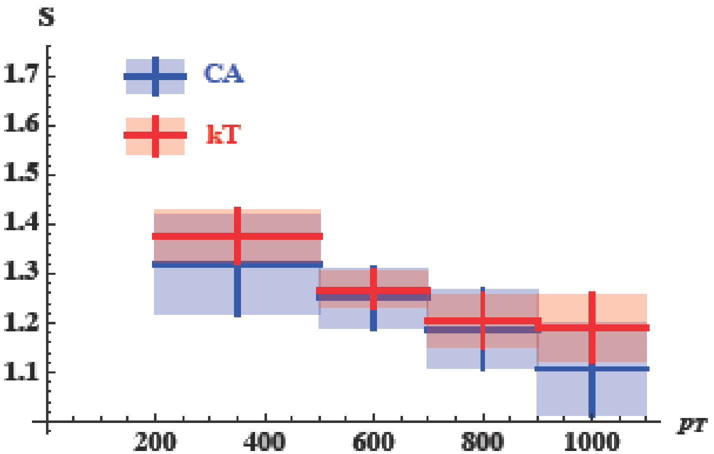


D < *θ*

- particle will not be reconstructed

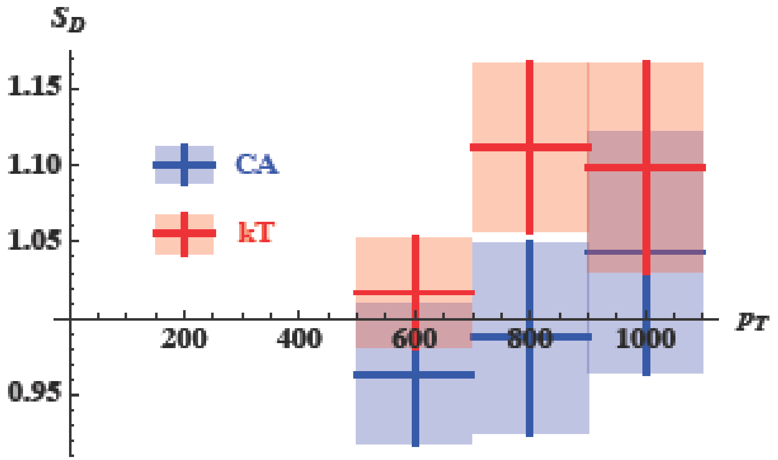
Improvements in Pruning

- 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 $\Rightarrow S_D = S_{D\ opt}/S_{D=1}$?



⇒ Little further improvement obtained by varying *D*

⇒ $S_D = 1$ in first bin

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

Underlying Event Rejection with Pruning

The mass resolution of *pruned* jets is (essentially) unchanged with or without the underlying event

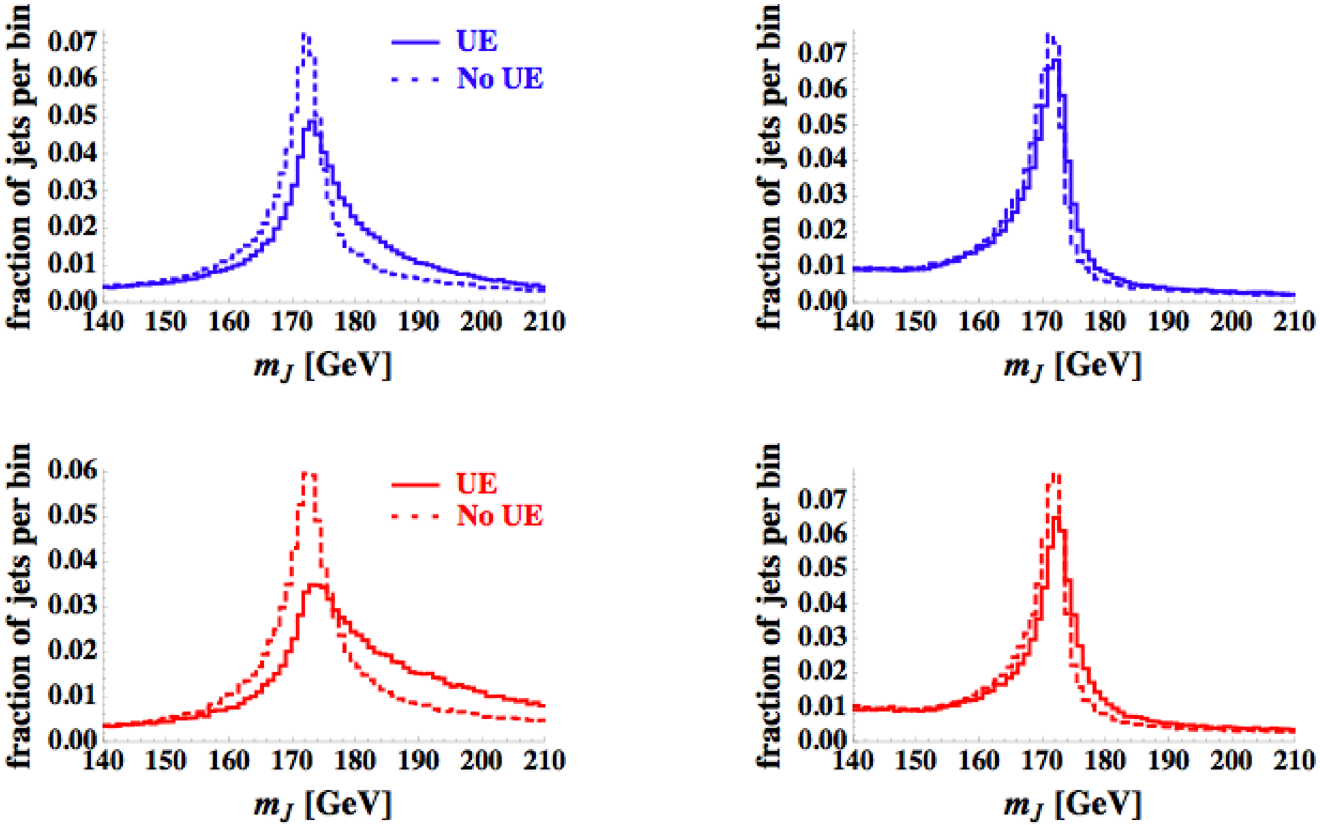
top study

no pruning

pruning

CA

k_T



500 < p_T < 700 GeV

Underlying Event Rejection with Pruning

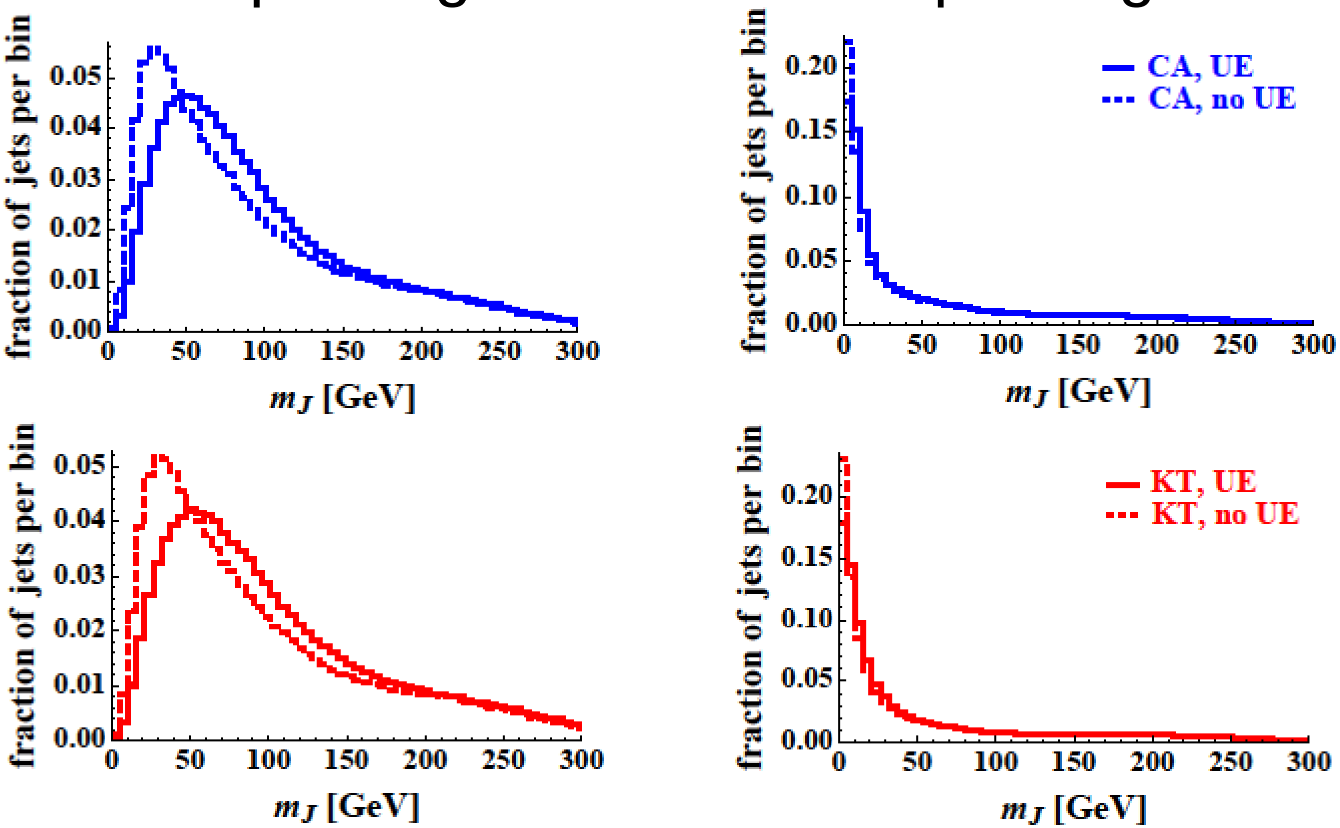
The jet mass distribution for QCD jets is significantly suppressed for *pruned* jets (essentially) independent of the underlying event

QCD study

no pruning | pruning

CA

k_T



500 < p_T < 700 GeV

Summary

- 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)

And:

- Systematics of the jet algorithm are important in studying jet substructure
  - The jet substructure we expect from the kT and CA algorithms are quite 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? Better theory tools – SCET?

More Information:

- software at tinyurl.com/jetpruning

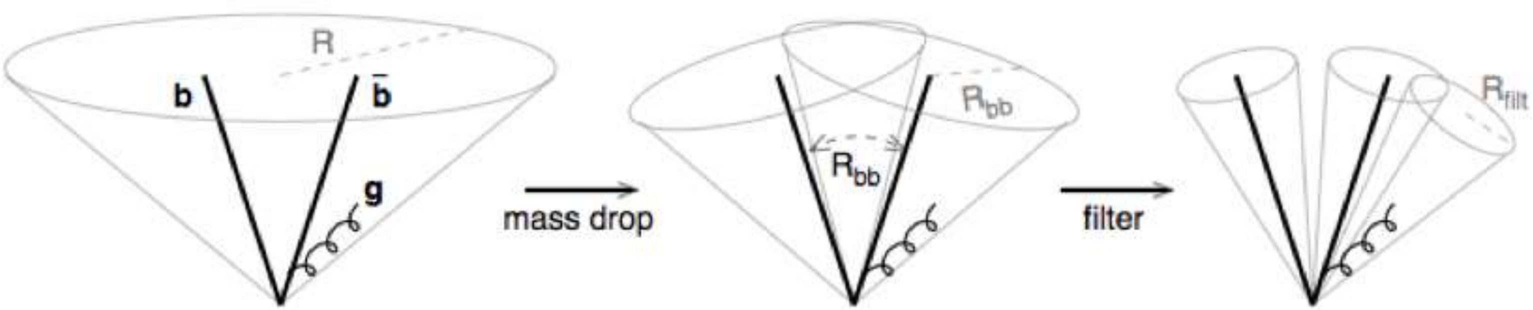
- See comparisons from Jet Substructure Workshop in Seattle in January 2010 (HW for Boost 2010)
  WiKi at
  http://librarian.phys.washington.edu/lhc-jets/index.php/Main_Page

- Jet tools available e.g.,

  http://librarian.phys.washington.edu/lhc-jets/index.php/SpartyJet

Extra Detail Slides

- "Boosted Higgs" (arXiv:0802.2470; Butterworth, Davison, Rubin, Salam)
  1. Starting with found jet, traverse merging history along heavier branch, looking for mass drop and a splitting that is not too asymmetric:
     $m_{daughter}/m_{branch} < \mu_{cut}(= 0.67)$,
     $y \equiv \frac{\min(p_{Ti}^2, p_{Tj}^2)}{m_{branch}^2} \Delta R_{ij}^2 > y_{cut}(= 0.09)$.
     This branching must have two b-tags.
  2. Uncluster below this branching down to
     $\Delta R = R_{cut}(= \min(0.3, R_{b\bar{b}}/2))$.
  3. Take 3 hardest subjets — capturing hardest radiation, but eliminating soft UE.



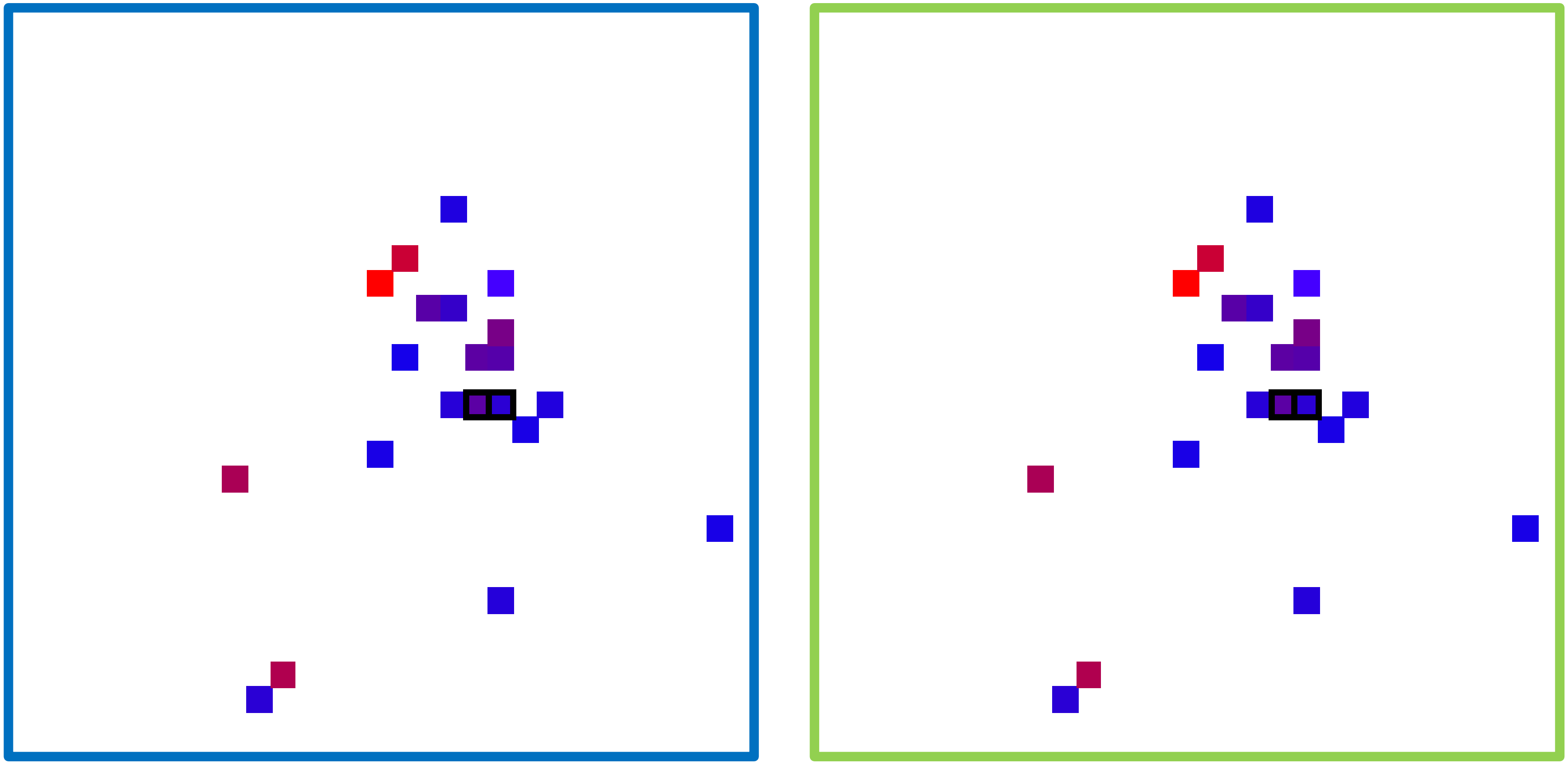
- "Top Tagging" (arXiv:0806.0848; Kaplan, Rehermann, Schwartz, Tweedie)
  1. Starting with found jet, traverse merging history along harder branch, looking for splitting with

     $$z_i \equiv p_T^i / p_T^{jet} > z_{cut},$$
     $$\Delta R > R_{cut}.$$

     This is the top-level splitting.
     - Throw out branches with $z_i < z_{cut}$ and continue. If both z_i fail, this is an irreducible branching.
     - If $\Delta R < R_{cut}$, stop. This is an "irreducible" splitting.
  2. Repeat on the two daughters of the found branch.
  3. Result is 1-4 subjets. Require 3 or 4.
  4. Additional cuts can be made on the subjet kinematics. . .

Note: Top jet with CA

No Pruning **Pruning**



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

- Jet = "stable cone" ⇒ 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 $\left(\bar{y}^C, \bar{\varphi}^C\right)$

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**

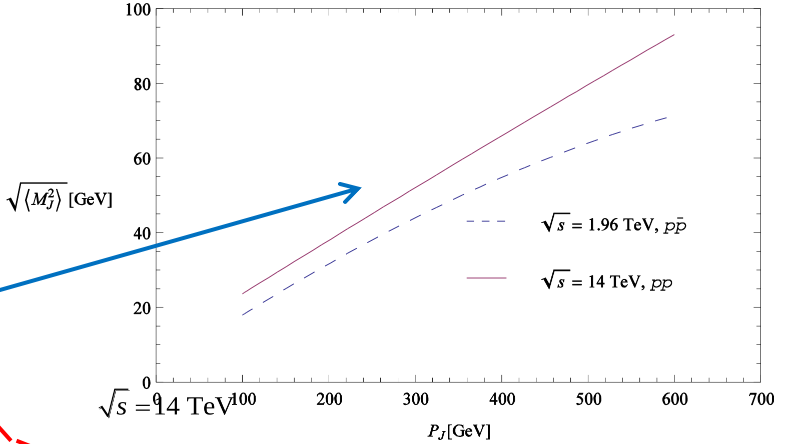
The bad news about jet algorithms:

- 👎 **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**

Jet Masses in QCD: To compare to non-QCD

- In NLO PertThy

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

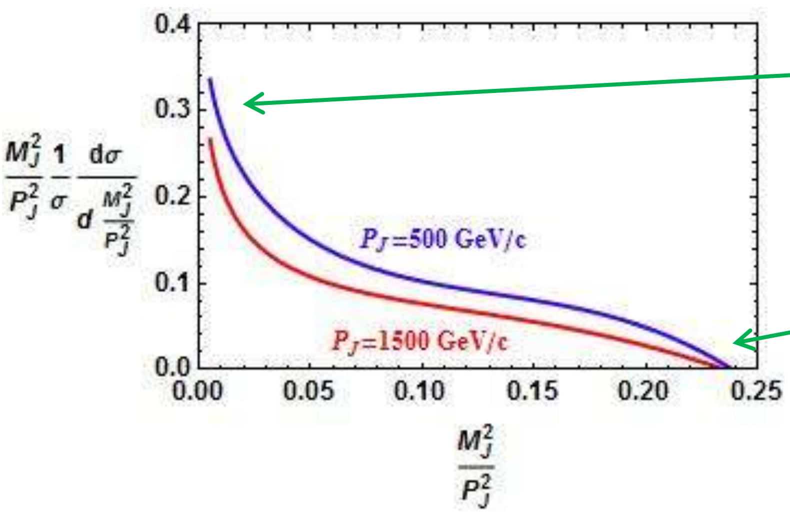


$\sqrt{s}$ =14 TeV

Phase space from pdfs,
f ~ 1 & const

Dimensions

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



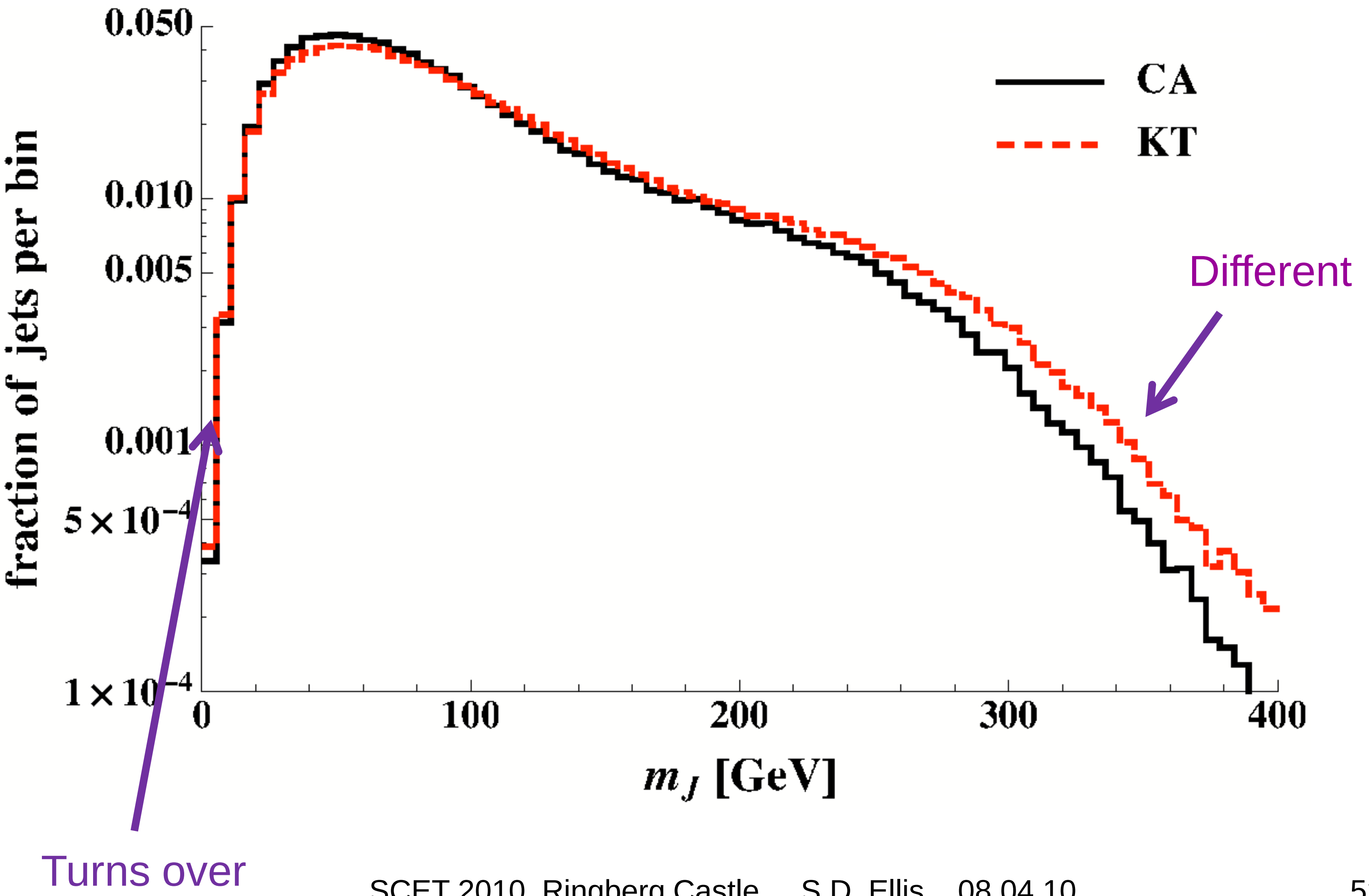
Peaked at low mass (log(m)/m behavior),

cuts off for $(M/P)^2 > 0.25 \sim D^2/4$ (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 D ~1), but not so much to be QCD like (~ 2 < γ < 5)

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

Jet Mass in PYTHIA (matched set)
D = 1, 500 GeV/c < pT < 700 GeV/c

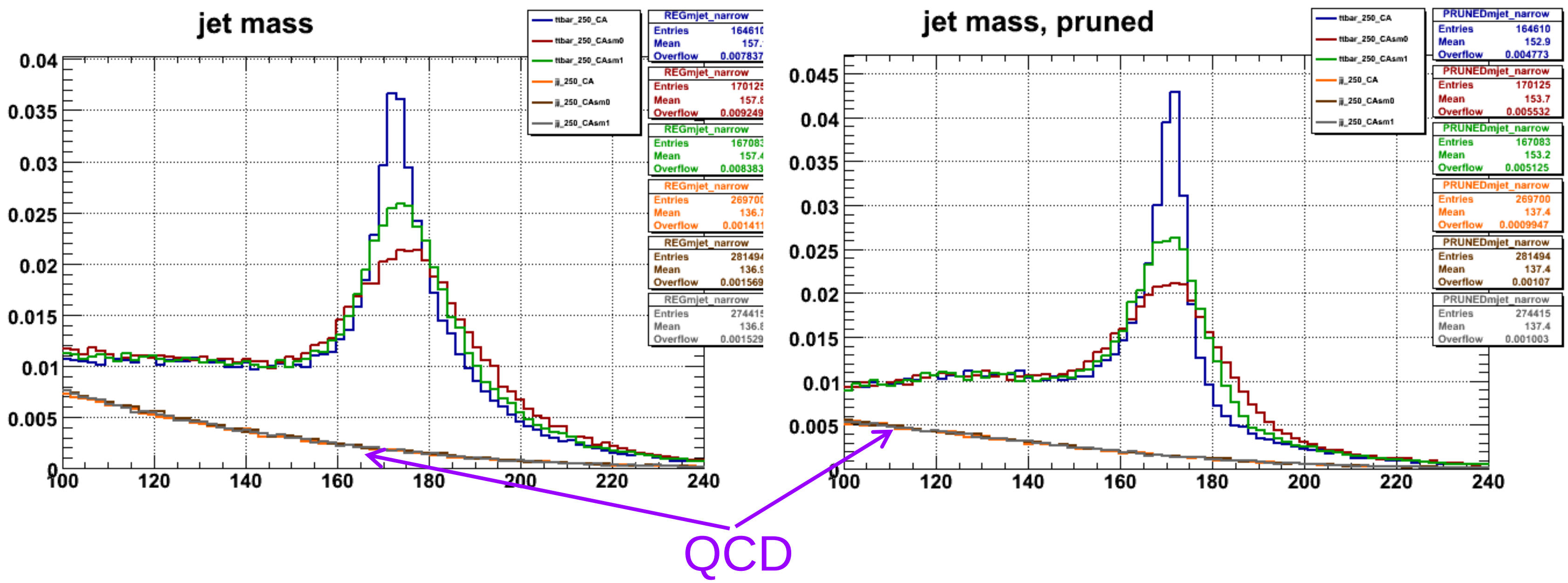


Consider impact of (Gaussian[1]) smearing

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:

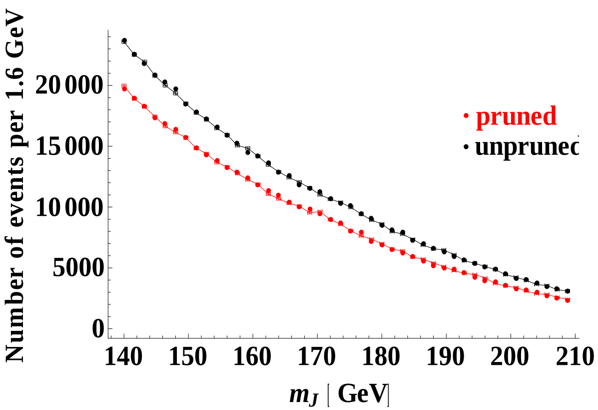
		ε	R	S
No Smearing	pCA/CA	0.90	2.25	1.42
	pkT/kT	0.68	3.01	1.44
Reasonable Smearing	pCA/CA	0.98	1.75	1.31
	pkT/kT	0.72	2.20	1.26
Worst Smearing	pCA/CA	1.00	1.59	1.26
	pkT/kt	0.74	2.00	1.22

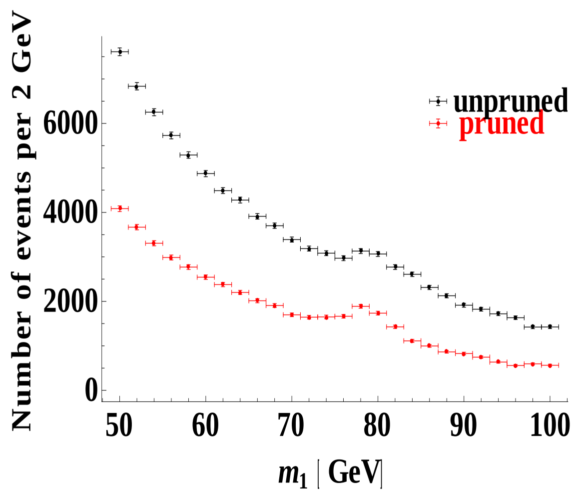
$$\epsilon = \frac{N_S(pA)}{N_S(A)} \qquad R = \frac{N_S(pA)/N_B(pA)}{N_S(A)/N_B(A)} \qquad S = \frac{N_S(pA)/\sqrt{N_B(pA)}}{N_S(A)/\sqrt{N_B(A)}}$$

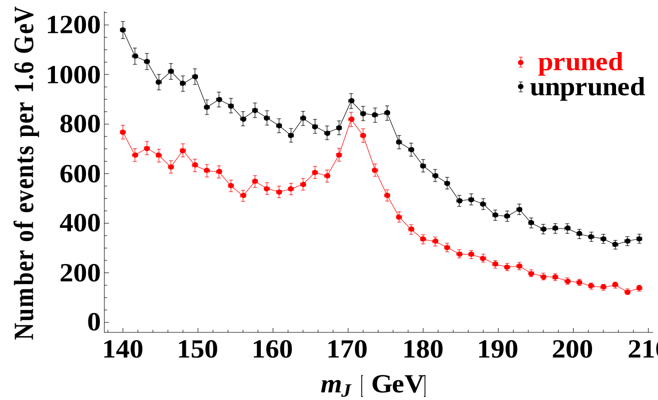
$\Rightarrow$ Smearing degrades but does not eliminate the value of pruning

“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)







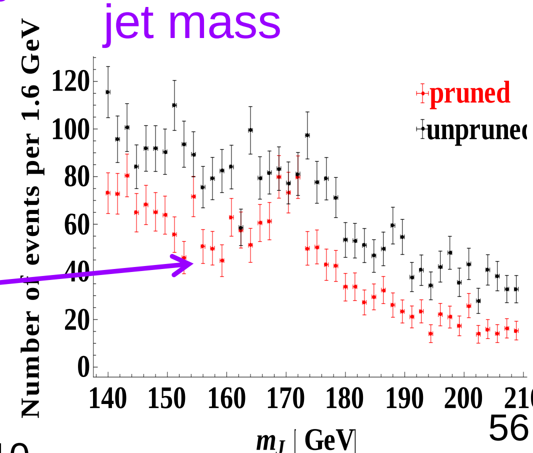
Find (small) mass bump and cut on it

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 (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 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 (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 the 3 hardest subjets (discard 1 subjet case but keep if only 2). Recombine the (2 or) 3 subjets with CA algorithm.