Jet Substructure Without Trees

Andrew Larkoski
SLAC, Stanford University
with Martin Jankowiak
arXiv:1104.1646

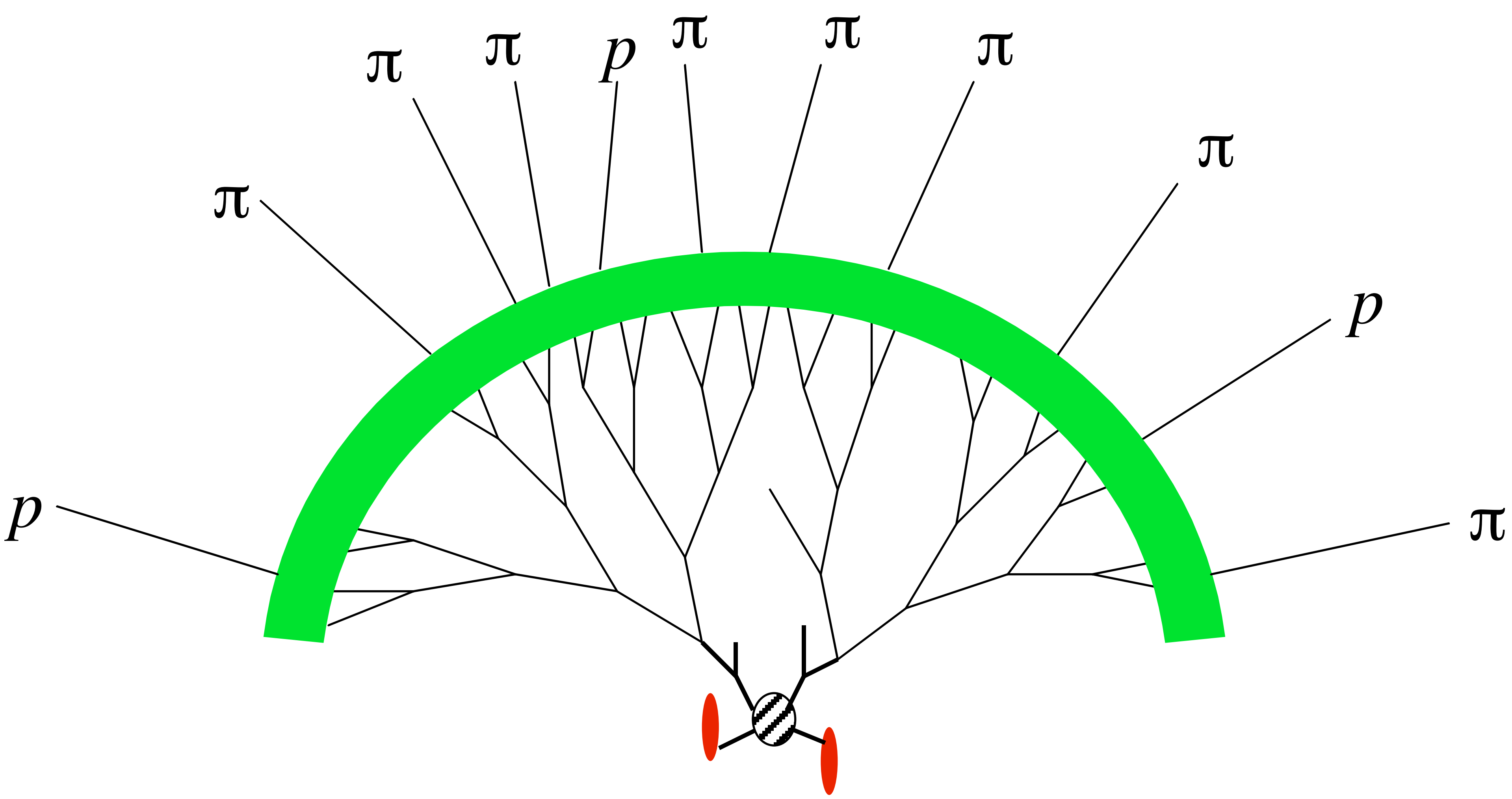
WC ATLAS Forum, May 4, 2011

Introduction

- We are officially in the LHC Era!
- A lot of theoretical effort devoted to LHC Inverse Problem
- One aspect of this is Jet Substructure Program
  - Can we find efficient methods for distinguishing QCD jets from jets which come from heavy particle decays?

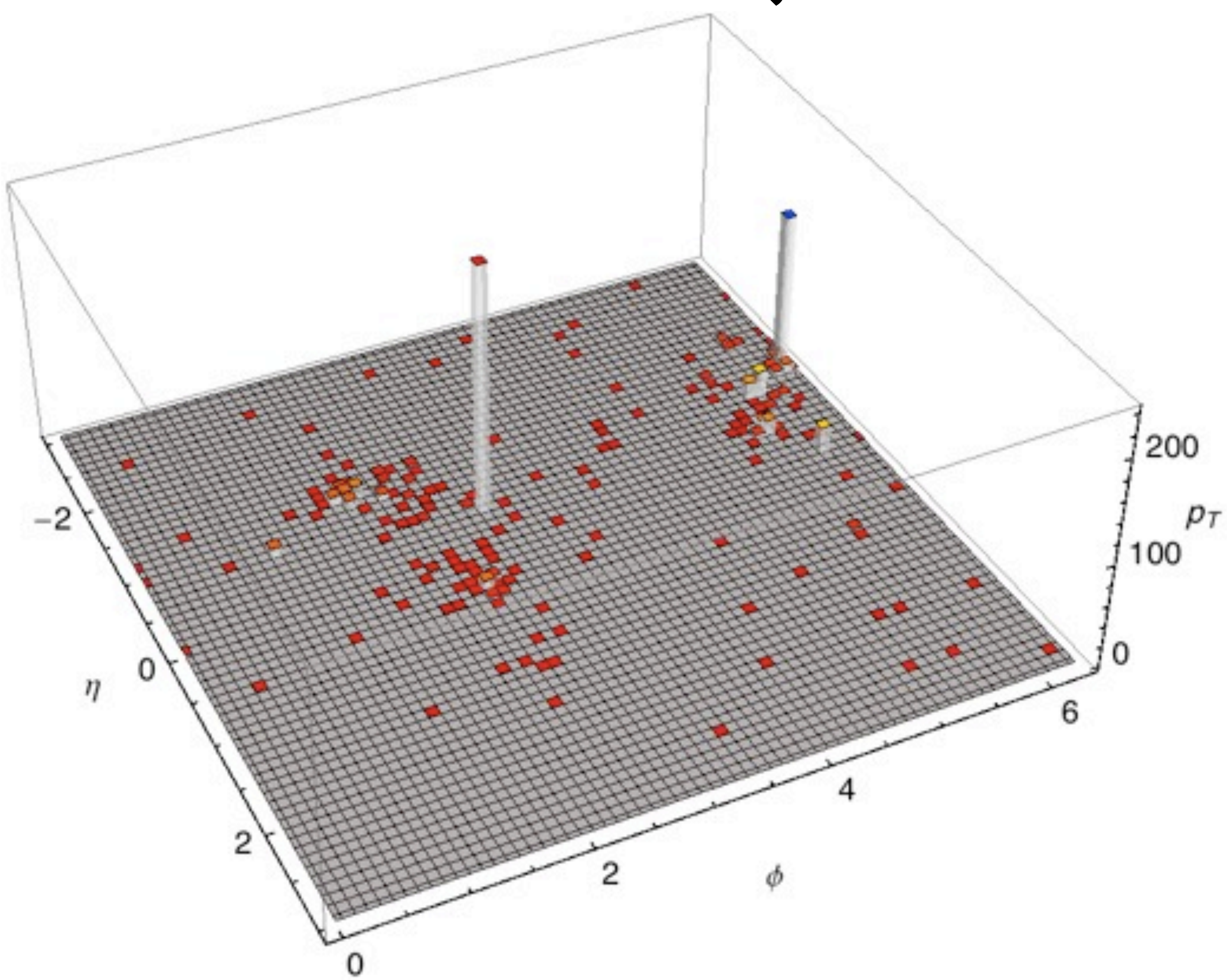
Introduction

- Jets: Our view into perturbative QCD
- Partons shower and hadronize into collimated jets of energy



Introduction

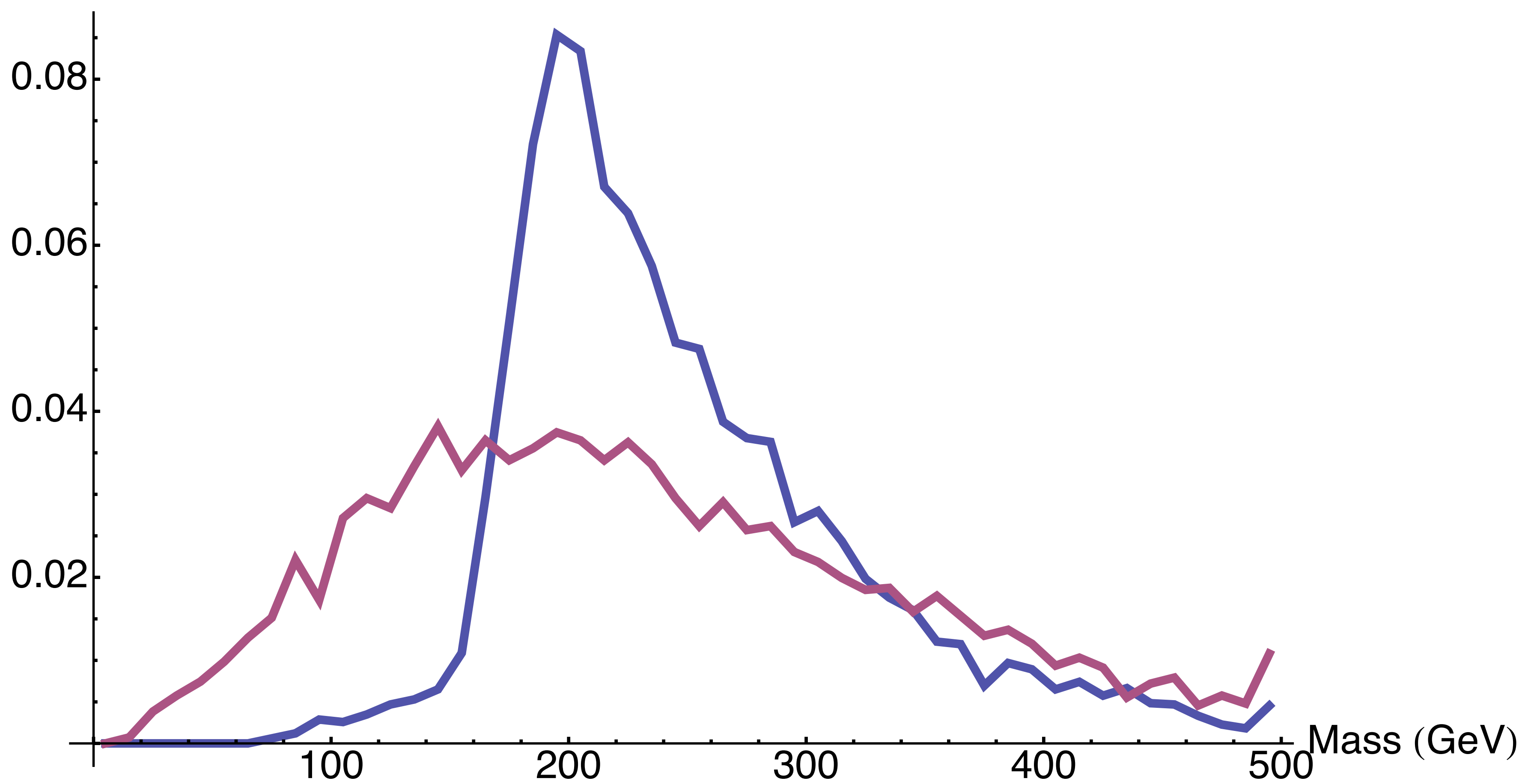
- Jets: Our view into perturbative QCD
- Partons shower and hadronize into collimated jets of energy



- How do we identify the two hard cells as jets?
- Use an algorithm to "undo" parton shower

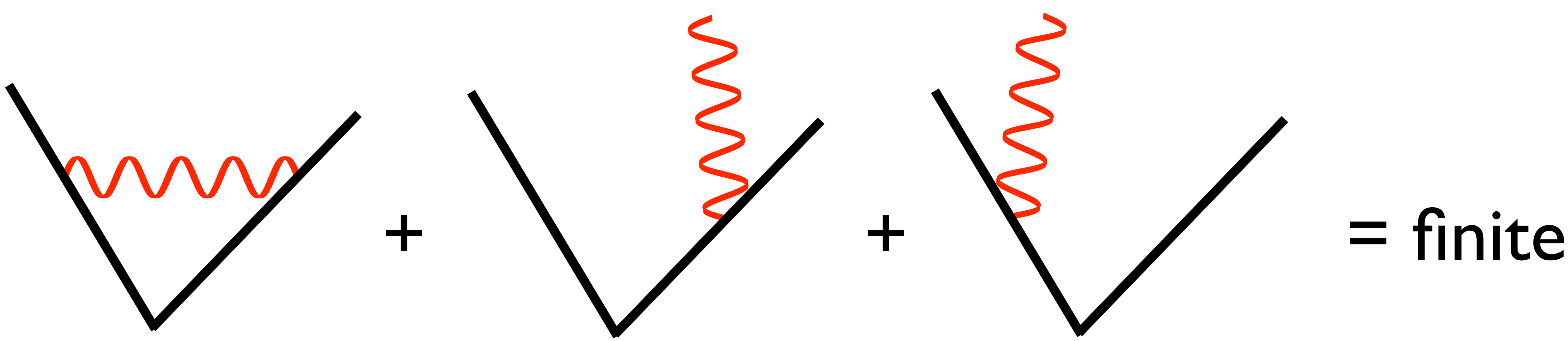
Introduction

- How do we study jets?
  - Jet Observables
  - Simplest Examples: jet mass, jet transverse momenta



Introduction

- How do we study jets?
  - Good observables can be calculated in perturbation theory
    - "IRC safety": Robust against soft and collinear splittings



Introduction

- Unclustering
  - Define a branching tree with a sequential jet algorithm

Introduction

- Unclustering
  - Define a branching tree with a sequential jet algorithm
- *kT*-type sequential jet algorithm

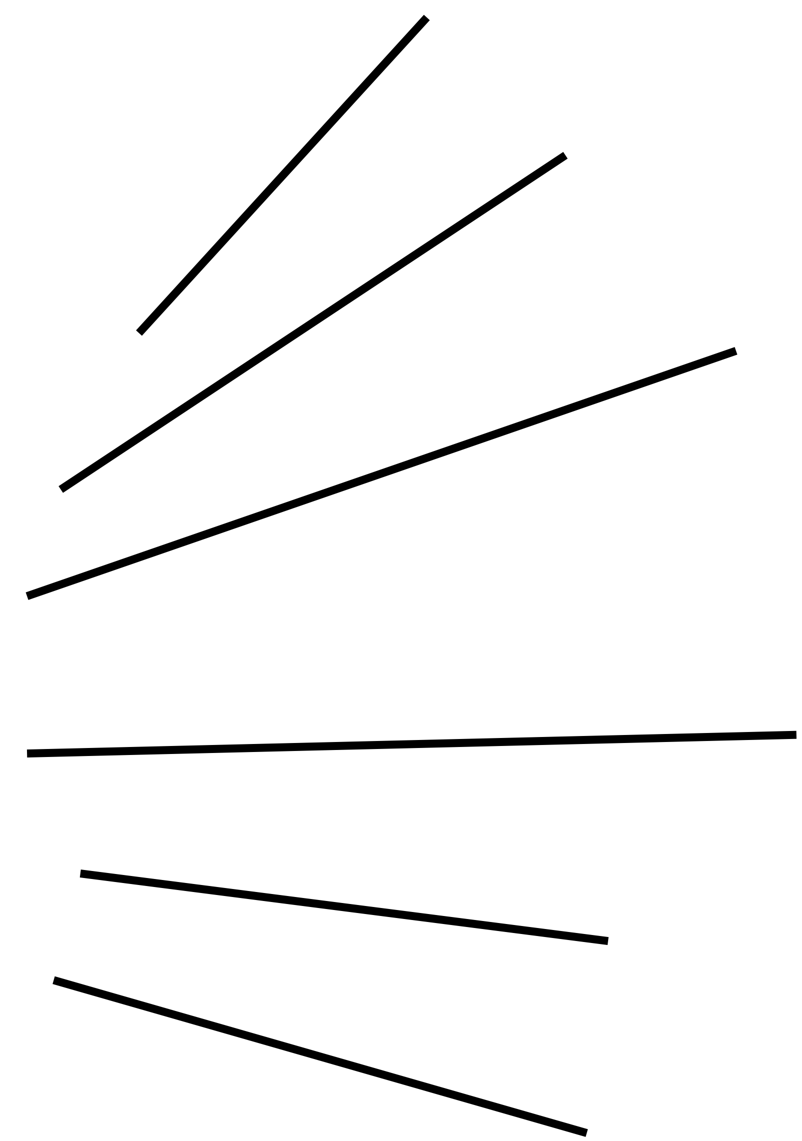
Ellis, Soper
Catani, *et al.*

  - 1) Compute $d_{ij} = \min[p_{T,i}^{2n}, p_{T,j}^{2n}] \frac{\Delta R_{ij}^2}{R^2}$

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$$

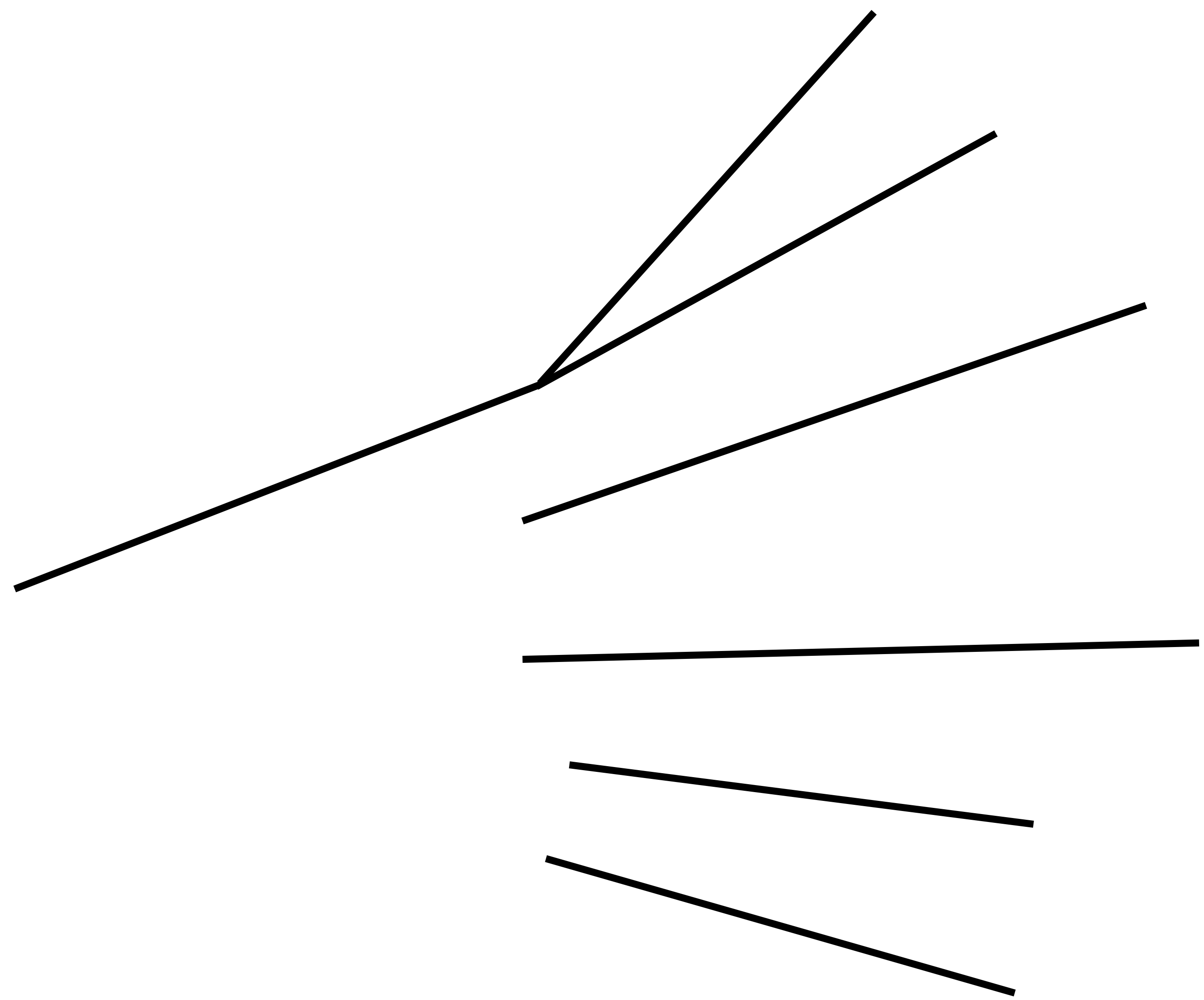
Introduction

- *kT*-type sequential jet algorithm
  - 2) Merge closest pair of particles



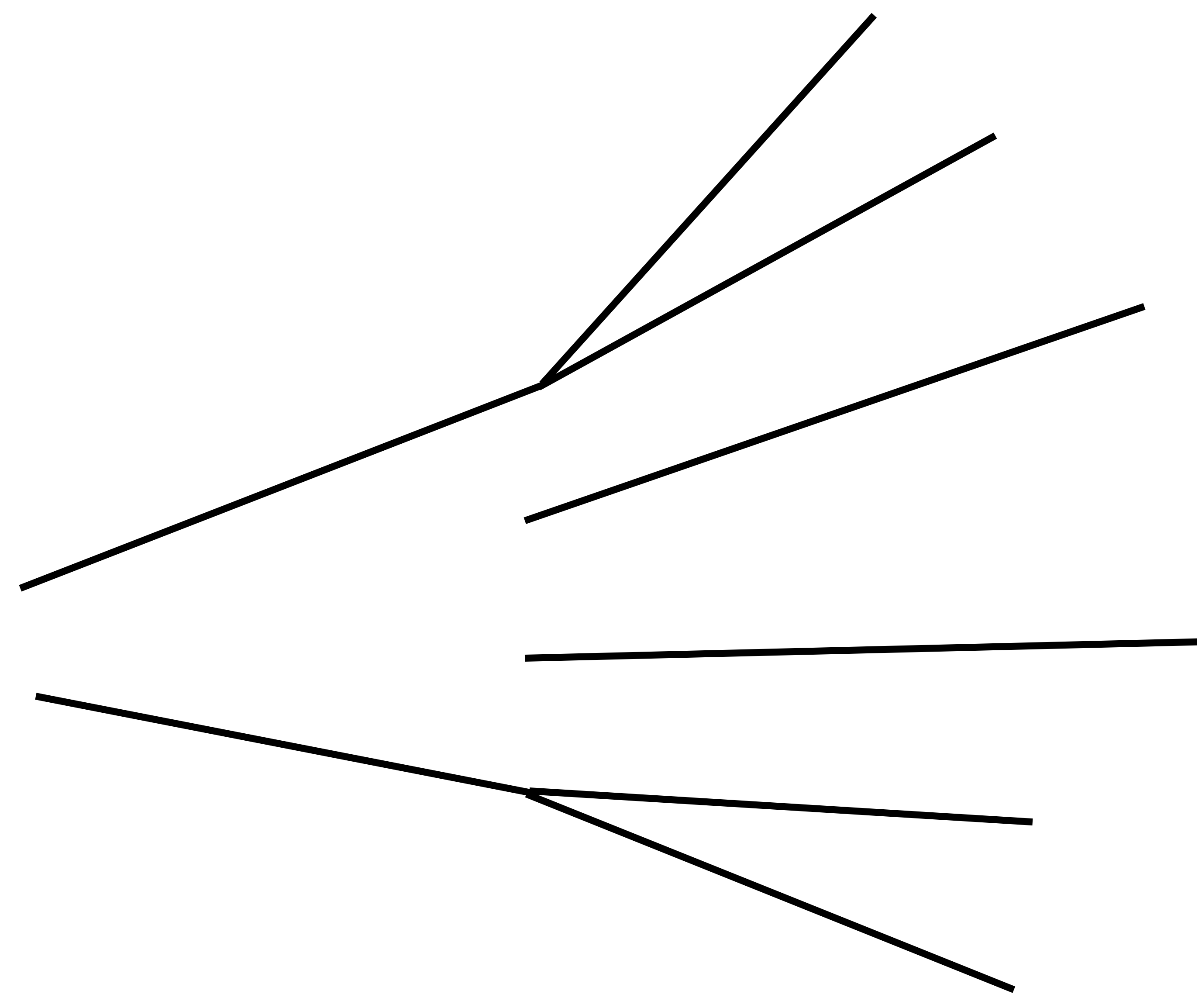
Introduction

- *kT*-type sequential jet algorithm
  - 2) Merge closest pair of particles



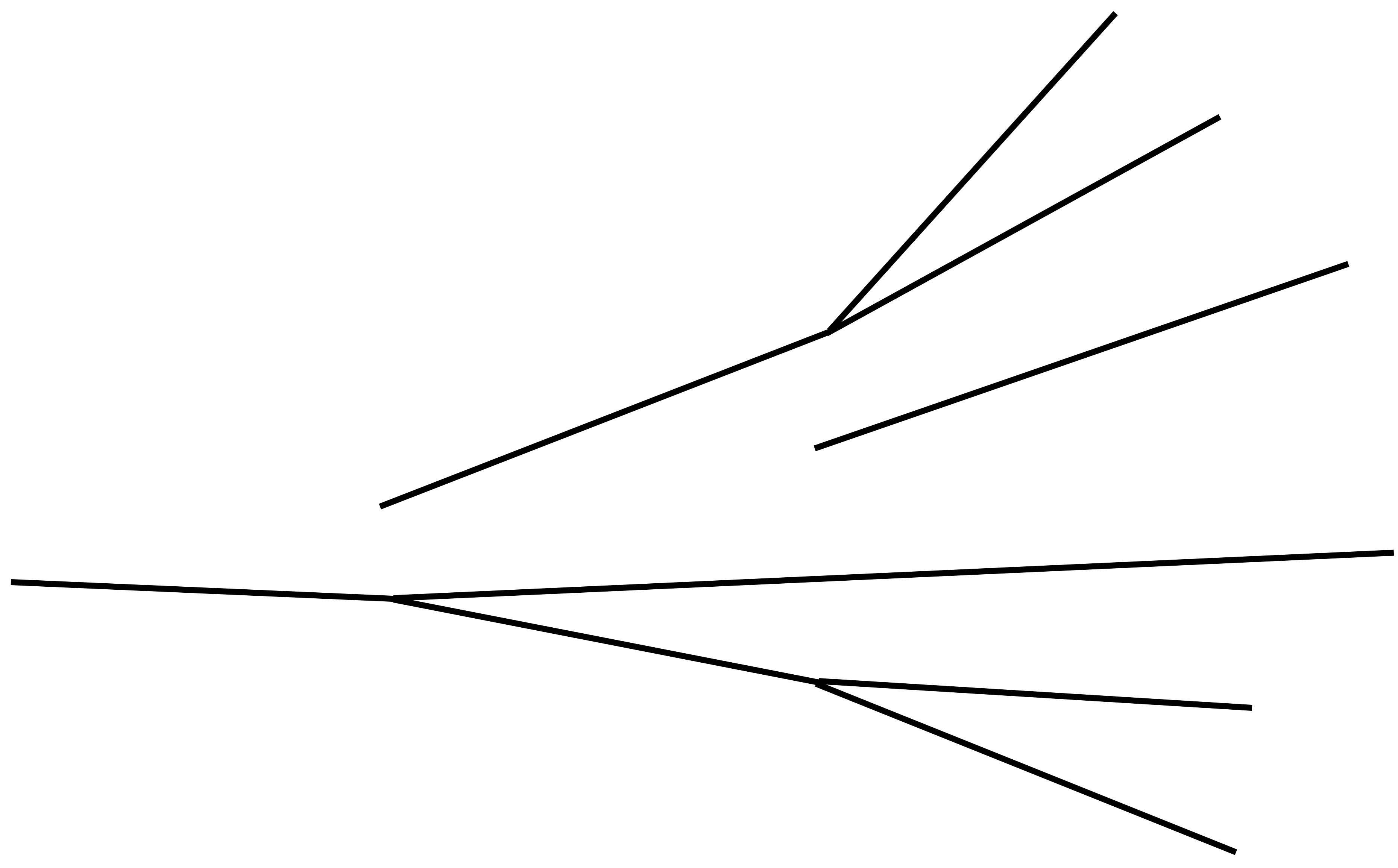
Introduction

- *kT*-type sequential jet algorithm
  - 2) Merge closest pair of particles



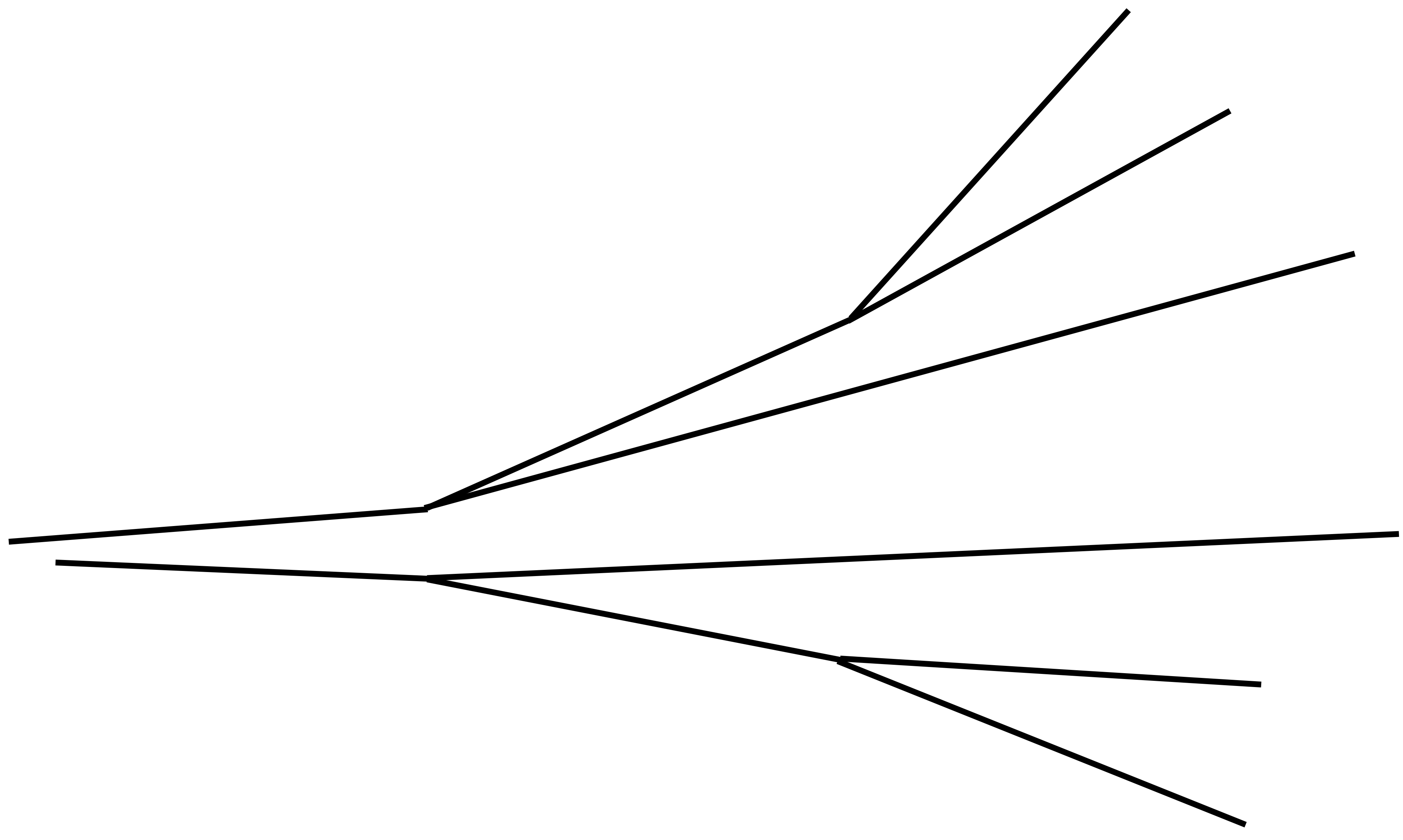
Introduction

- *kT*-type sequential jet algorithm
  - 2) Merge closest pair of particles



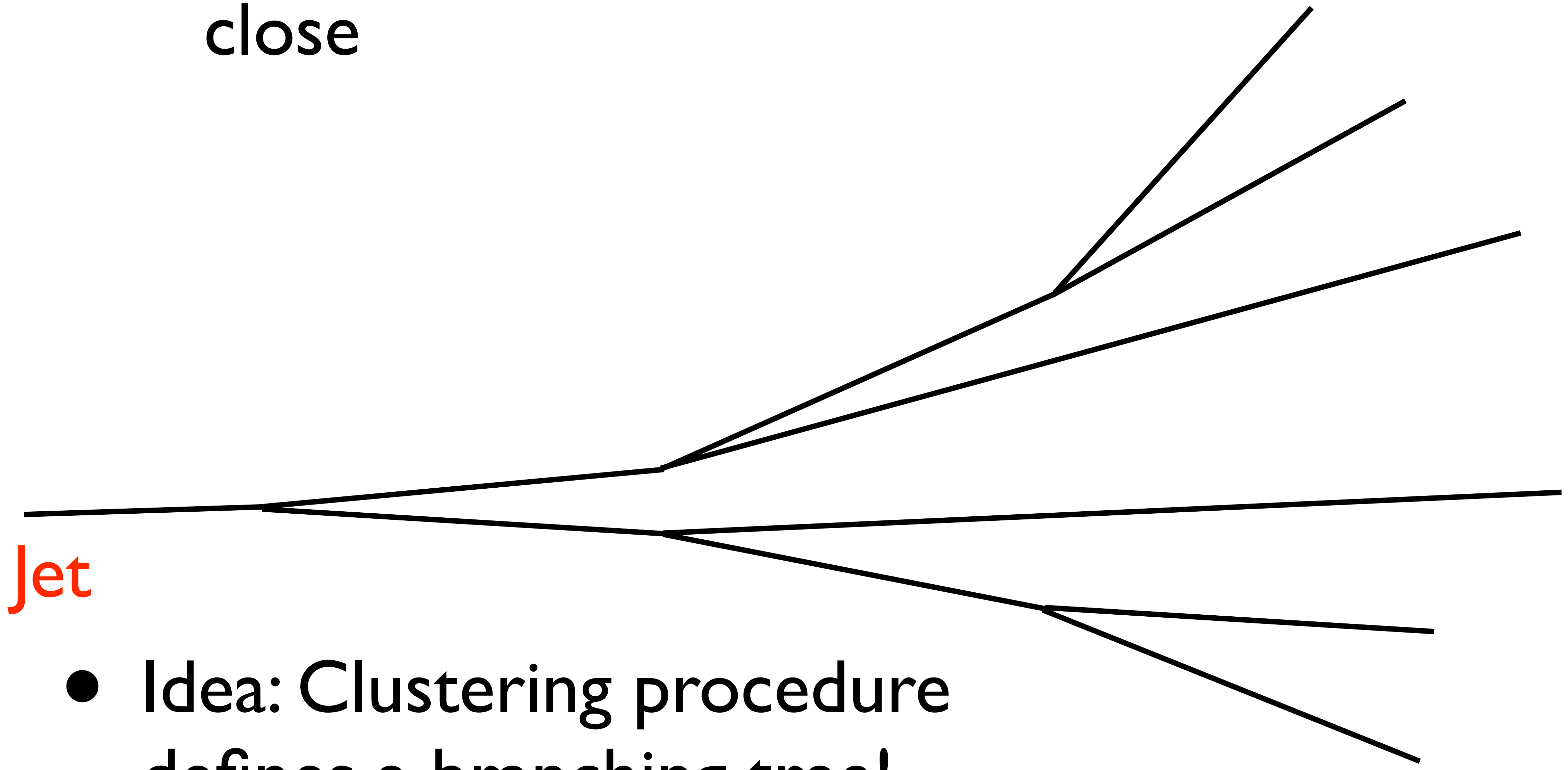
Introduction

- *kT*-type sequential jet algorithm
  - 2) Merge closest pair of particles



Introduction

- *kT*-type sequential jet algorithm
  - 3) Continue until no pair of particles is close



- Idea: Clustering procedure defines a branching tree!

Outline

- Describe a new method for studying jets
  - How to find angular and mass scales within a jet
- Prominence: A trick for reducing noise
- Jet Observables in this framework
- An application: Identifying top quark jets
- Discussion

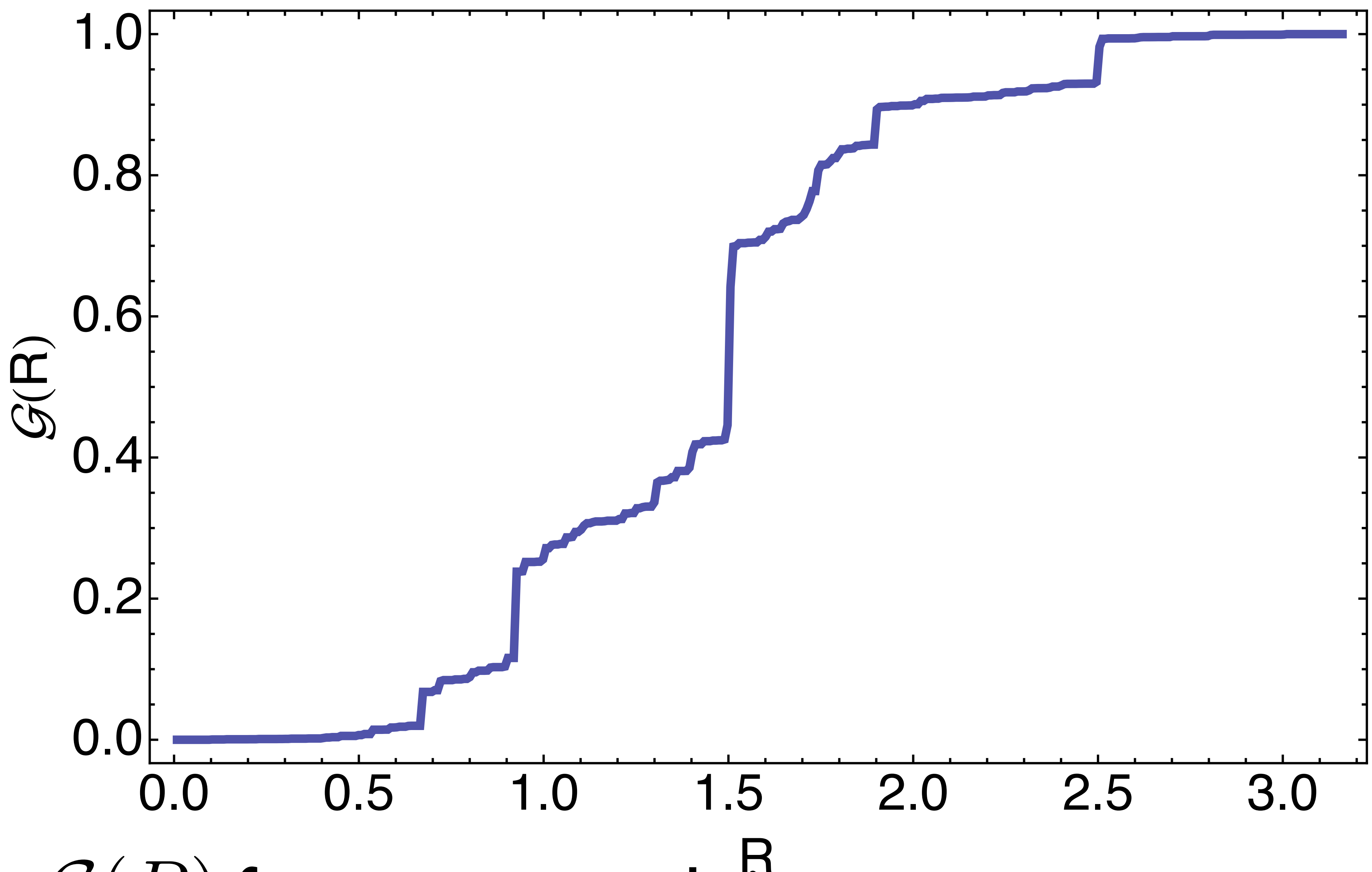
Angular Correlation Function

$$\mathcal{G}(R) \equiv \frac{\sum_{i \neq j} p_{Ti} p_{Tj} \Delta R_{ij}^2 \Theta(R - \Delta R_{ij})}{\sum_{i \neq j} p_{Ti} p_{Tj} \Delta R_{ij}^2} \approx \frac{\sum_{i \neq j} p_i \cdot p_j \Theta(R - \Delta R_{ij})}{\sum_{i \neq j} p_i \cdot p_j}$$

- IRC safe = computable in perturbation theory
- *R* is **not** measured wrt jet center
  - Distinct from angular profile
  - Quantifies jet scaling in an IRC safe way

Angular Correlation Function

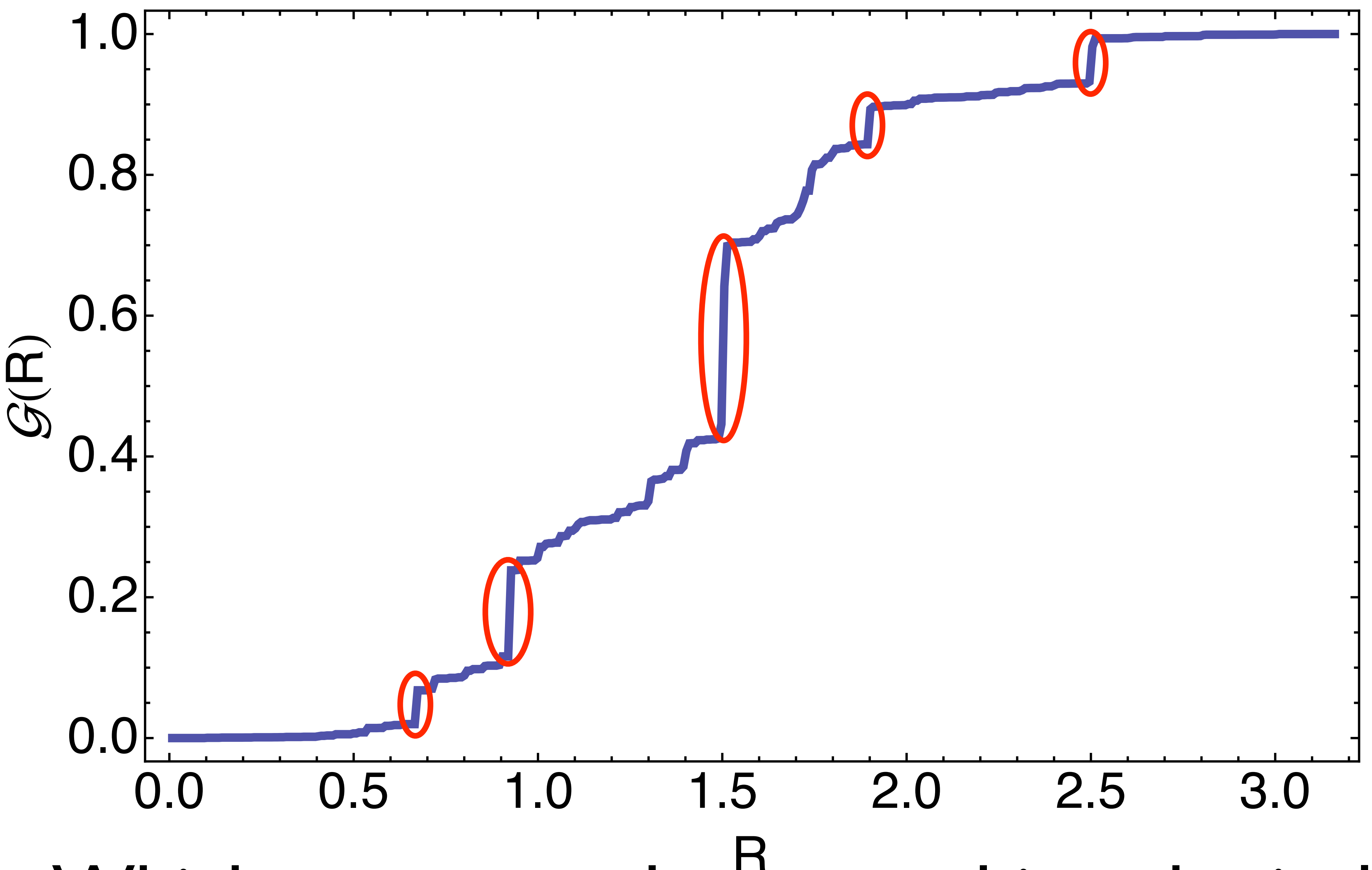
- Ledges in $\mathcal{G}(R)$ = separation of hard subjets



- $\mathcal{G}(R)$ for a top quark jet

Angular Correlation Function

- Ledges in $\mathcal{G}(R)$ = separation of hard subjets



- Which correspond to something physical?

Angular Correlation Function

- How to find ledges
  - Find peaks in the derivative!
  - Problem: really want ratio of masses
  - Take derivative of $\log \mathcal{G}(R)$

Angular Correlation Function

- How to find ledges
  - Find peaks in the derivative!
  - Problem: really want ratio of masses
  - Take derivative of $\log \mathcal{G}(R)$
- QCD is scale invariant
  - Take derivative wrt log *R*
  - Reduces noise at small *R*

Angular Correlation Function

- Look for peaks in $d \log \mathcal{G}(R)/d \log R$

Angular Correlation Function

- Look for peaks in $d \log \mathcal{G}(R) / d \log R$
- Angular structure function:

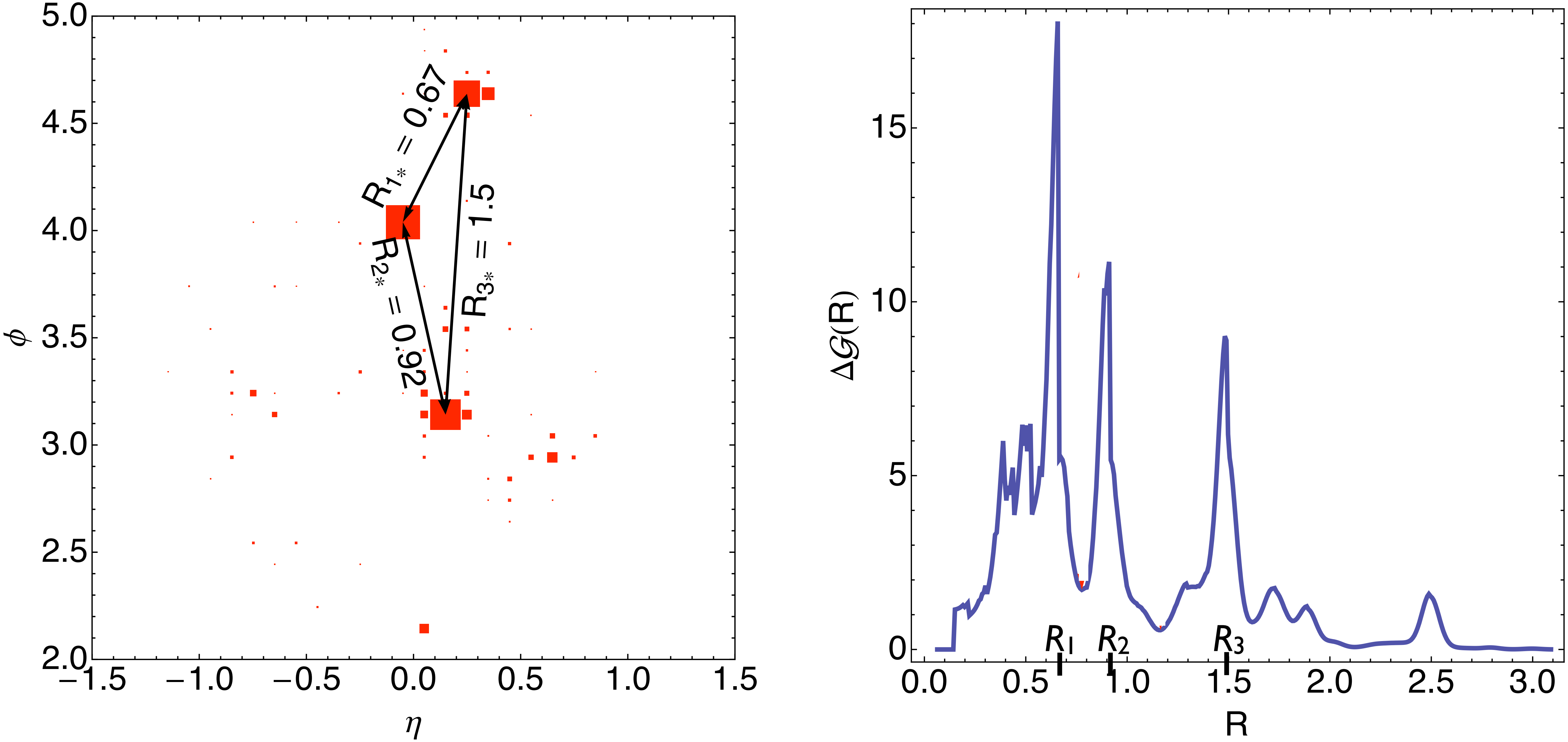
$$\Delta\mathcal{G}(R) \equiv R \, \frac{\sum\limits_{i \neq j} p_{Ti} p_{Tj} \Delta R_{ij}^2 K(R - \Delta R_{ij})}{\sum\limits_{i \neq j} p_{Ti} p_{Tj} \Delta R_{ij}^2 \Theta(R - \Delta R_{ij})}$$

- *K* is taken to be a smooth gaussian kernel:

$$\delta(R - \Delta R_{ij}) \simeq \frac{e^{-\frac{(R - \Delta R_{ij})^2}{dR^2}}}{dR\sqrt{\pi}}$$

Angular Correlation Function

Question: Does $\Delta\mathcal{G}(R)$ determine interesting ledges?



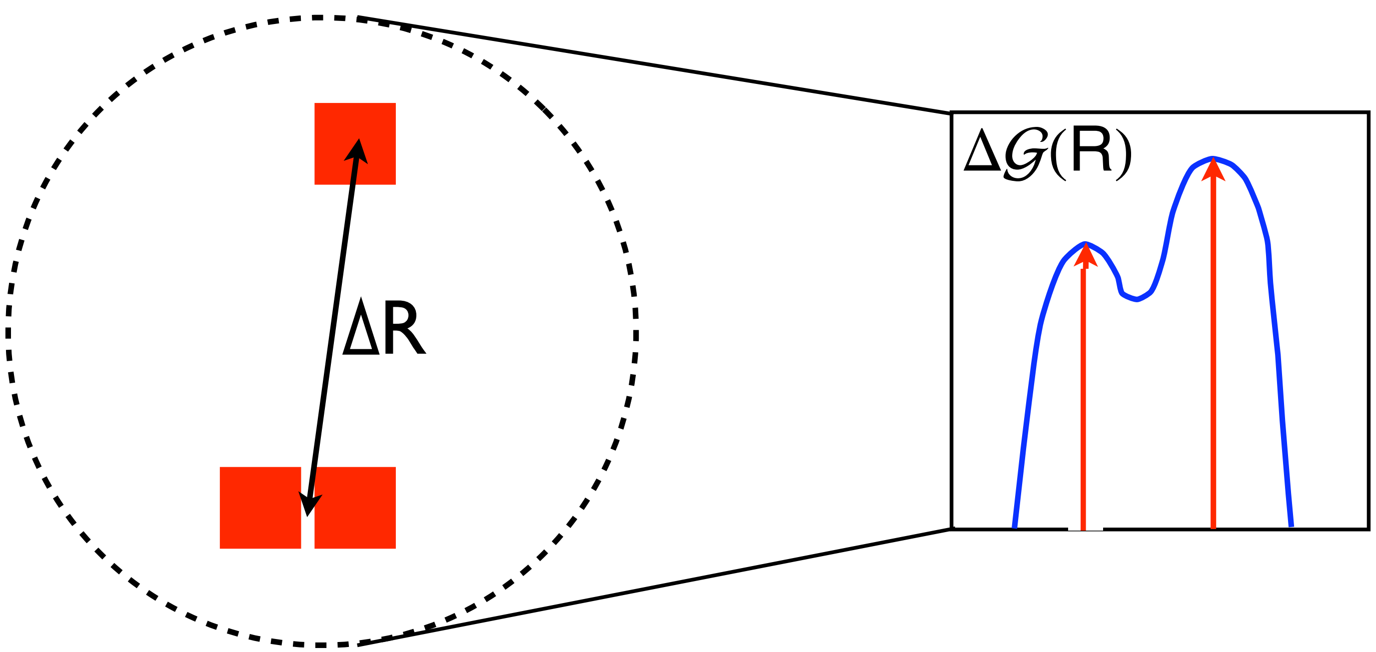
Answer: Yes!

Prominence

- $\Delta\mathcal{G}(R)$ picks out physical peaks beautifully!
  - Still some noise: how can we reduce it?
- How do we define interesting peaks?
  - By height?

Prominence

- $\Delta\mathcal{G}(R)$ picks out physical peaks beautifully!
  - Still some noise: how can we reduce it?
- How do we define interesting peaks?
  - By height?



Is this little bump interesting?

Prominence

- Quiz: What is the highest mountain in the contiguous US?

Prominence

- Quiz: What is the highest mountain in the contiguous US?
  - Mt. Whitney, CA
- What is the most *prominent* mountain in the contiguous US?

Prominence

- Quiz: What is the highest mountain in the contiguous US?
  - Mt. Whitney, CA
- What is the most *prominent* mountain in the contiguous US?
  - Mt. Rainier, WA

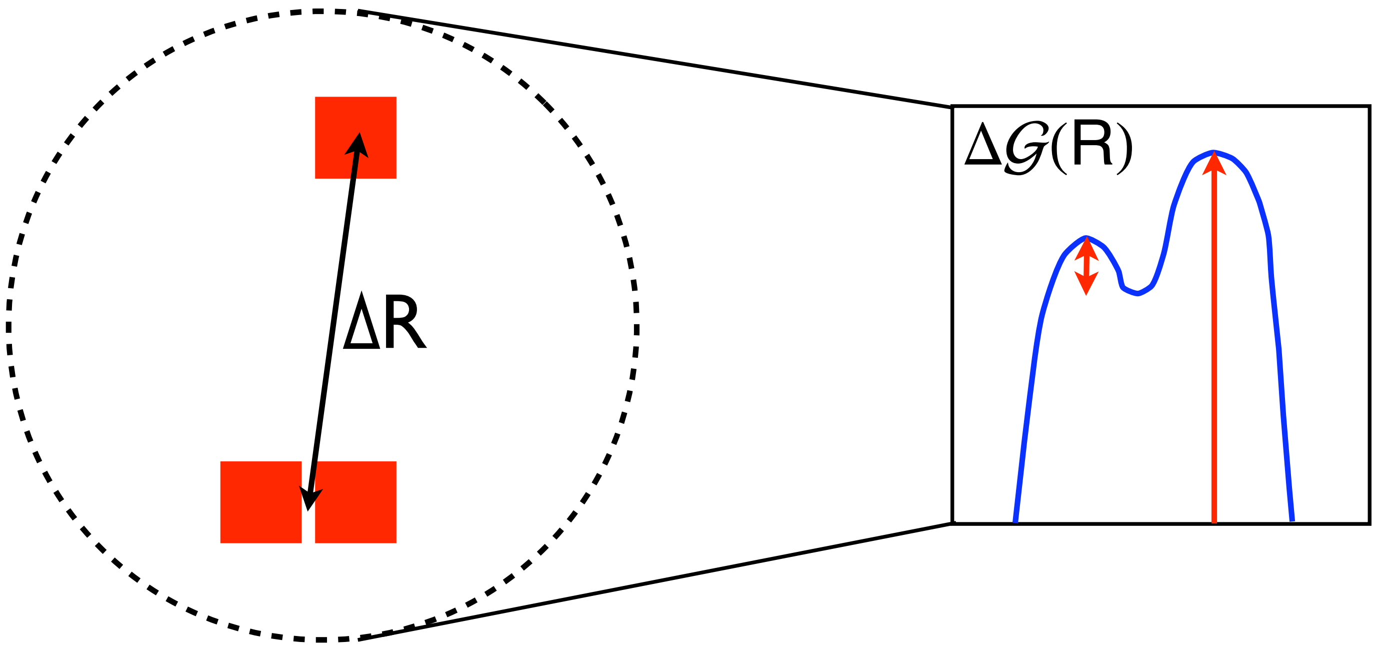
Prominence

- Why is Mt. Rainier the most prominent mountain?
  - Prominence ~ amount mountain sticks out above ambient landscape
- How does this help us?

Prominence

- Why is Mt. Rainier the most prominent mountain?
  - Prominence ~ amount mountain sticks out above ambient landscape
- How does this help us?

Prominence of little bump is tiny!

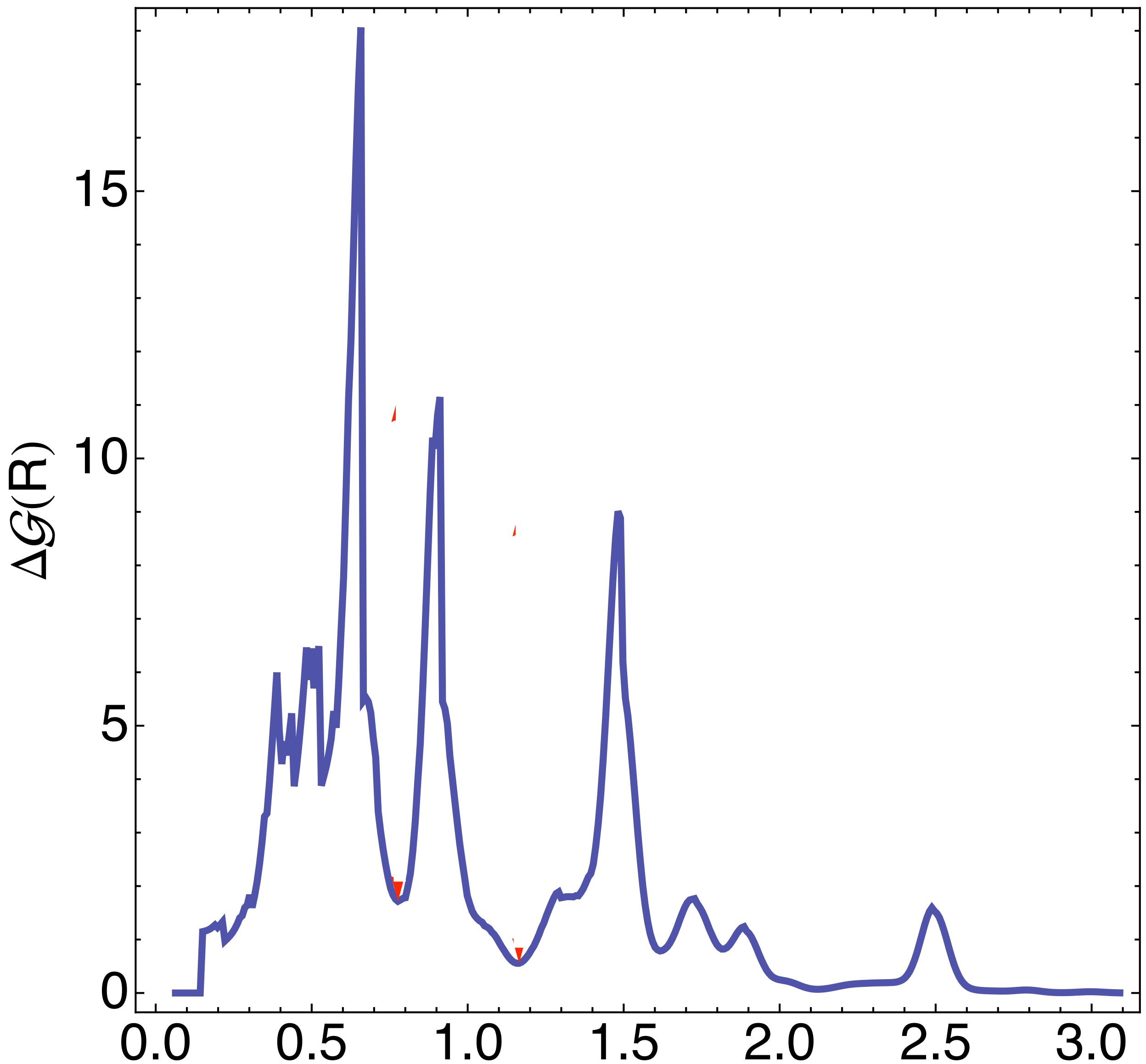


Review

- Our program for studying jets:
  - Define a jet using your favorite algorithm
  - From its constituents, compute $\mathcal{G}(R)$
  - To find ledges in $\mathcal{G}(R)$, find peaks in $\Delta\mathcal{G}(R)$
  - Interesting peaks have a prominence greater than some value
- Two parameters: delta-function smoothing *dR*, minimum prominence

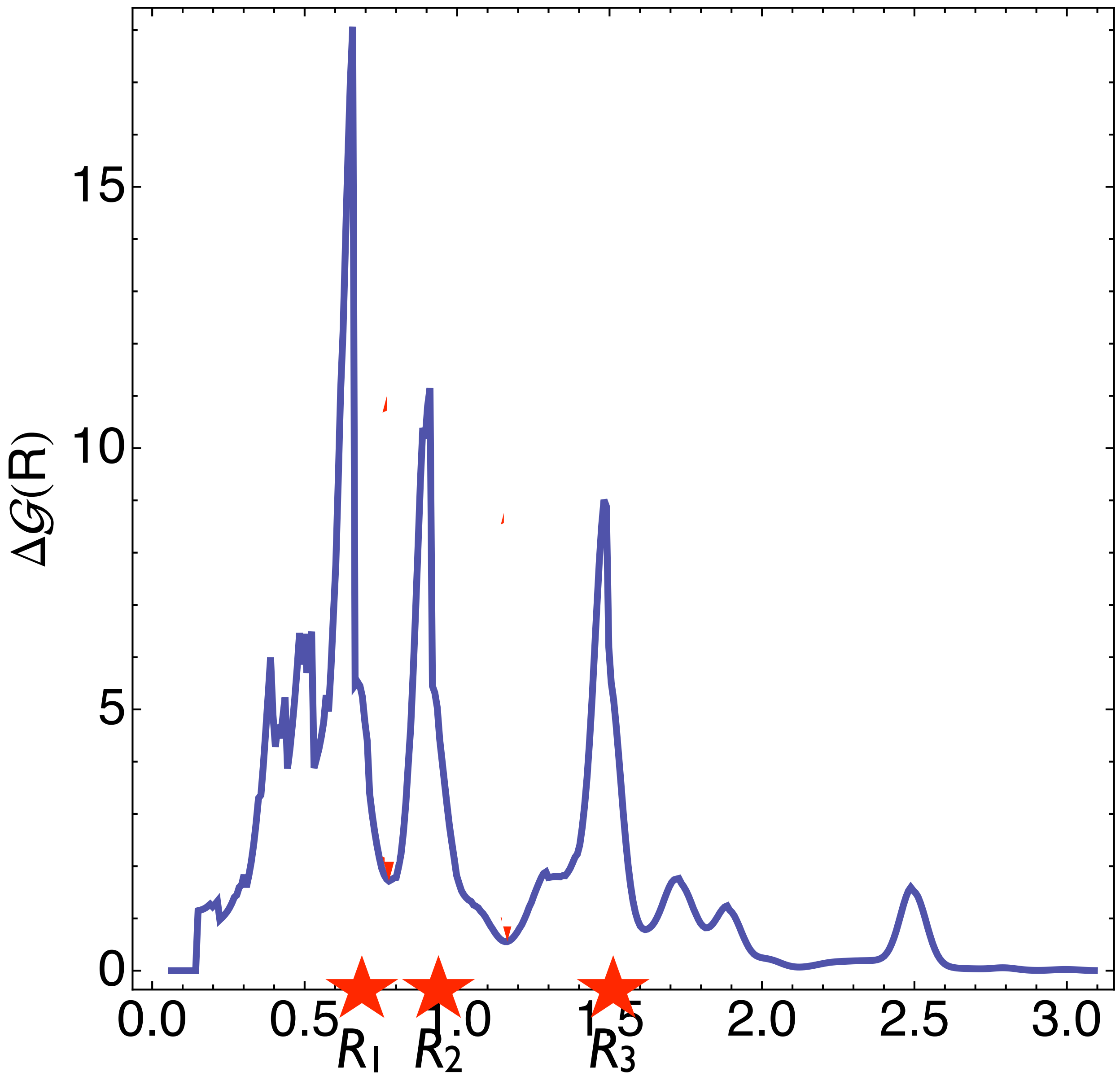
Jet Substructure Observables

- IRC safe observables from $\Delta\mathcal{G}(R)$:



Jet Substructure Observables

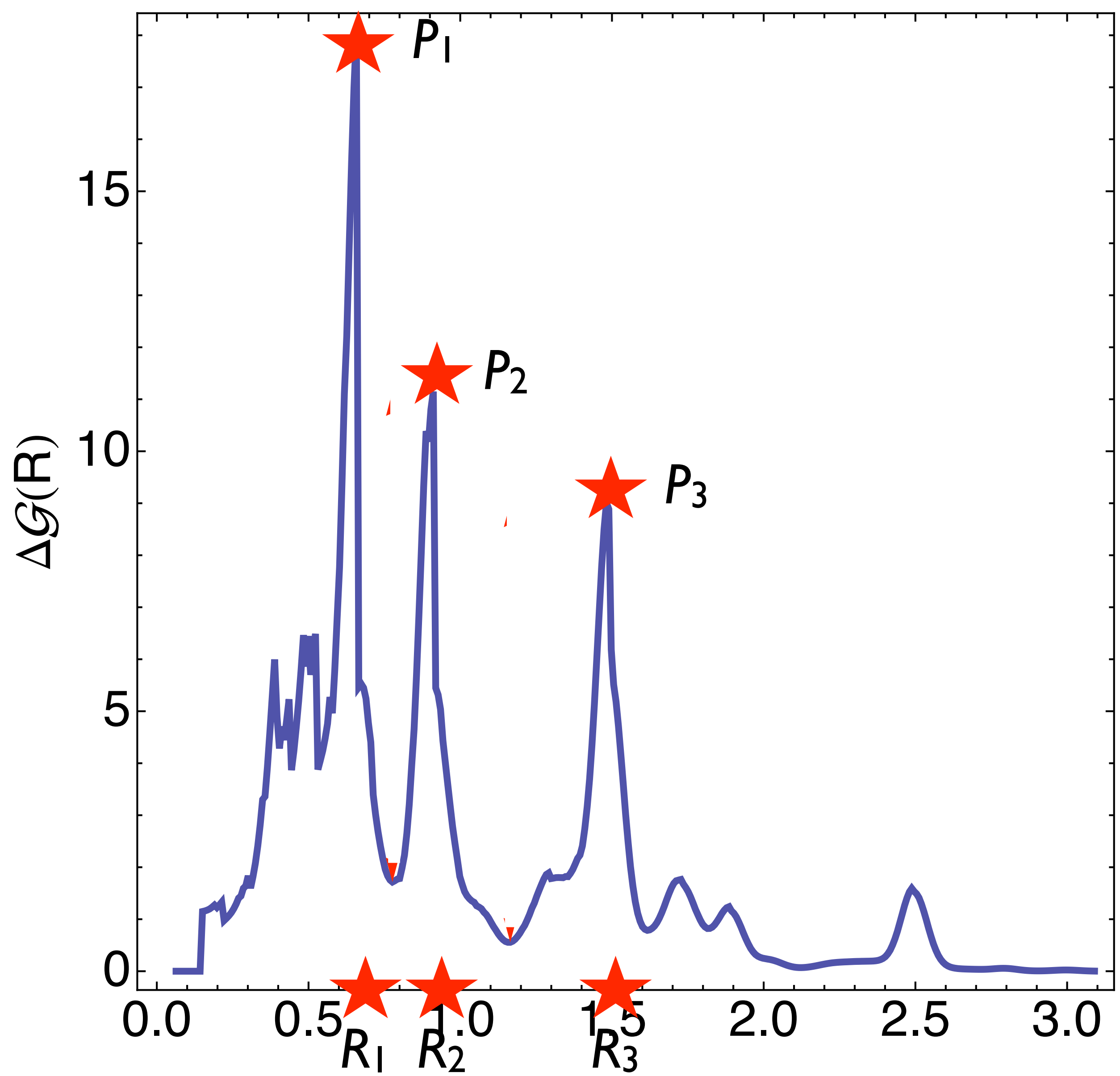
- IRC safe observables from $\Delta\mathcal{G}(R)$:



- Entire curve is IRC safe
- Location of peaks in *R*

Jet Substructure Observables

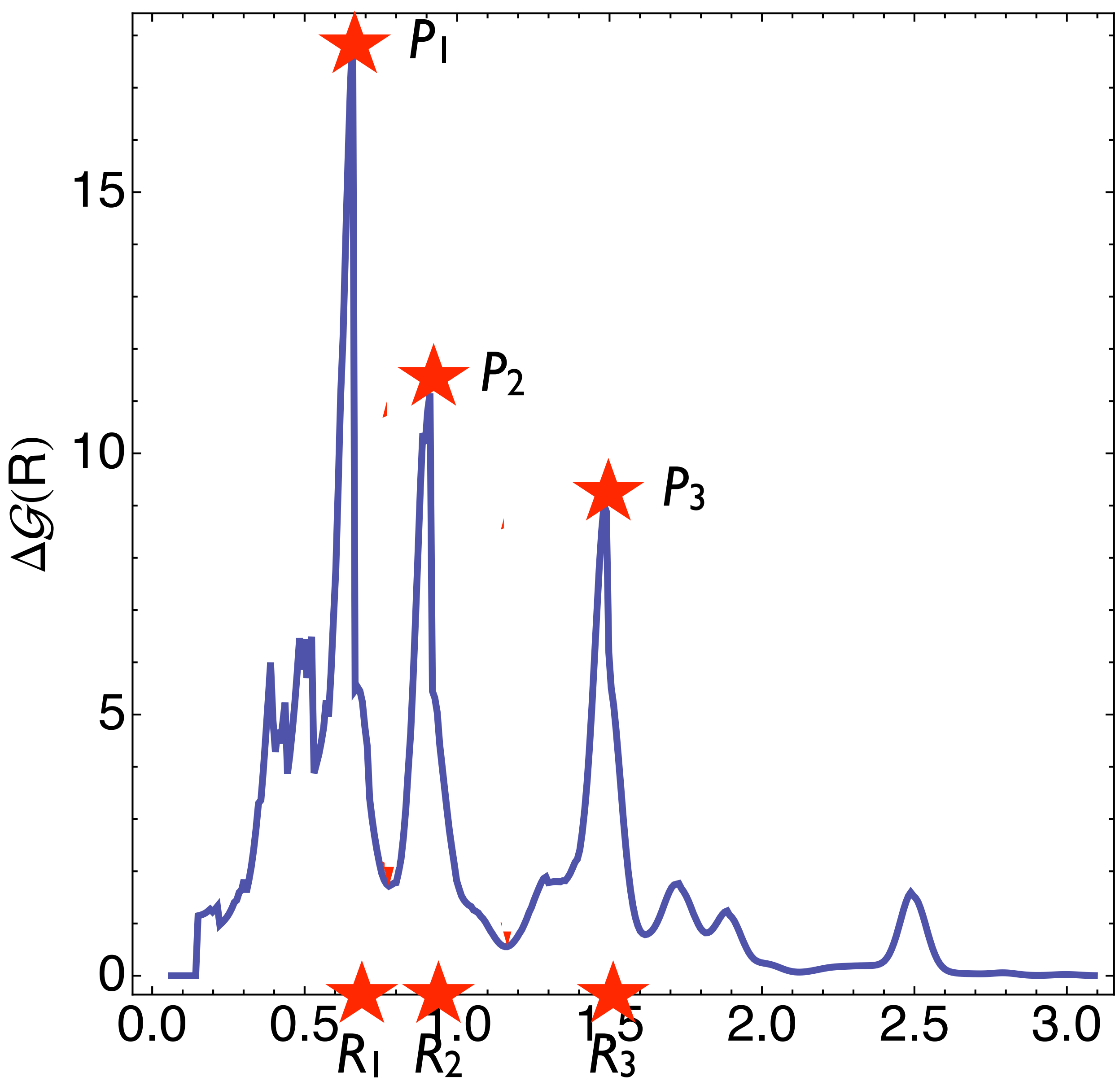
- IRC safe observables from $\Delta\mathcal{G}(R)$:



- Location of peaks in R
- Height of peaks

Jet Substructure Observables

- IRC safe observables from $\Delta\mathcal{G}(R)$:



- Location of peaks in R
- Height of peaks
- Number of peaks

Jet Substructure Observables

- Our approach:
  - Bin jets by the number of peaks
    - Related to the number of subjets

Jet Substructure Observables

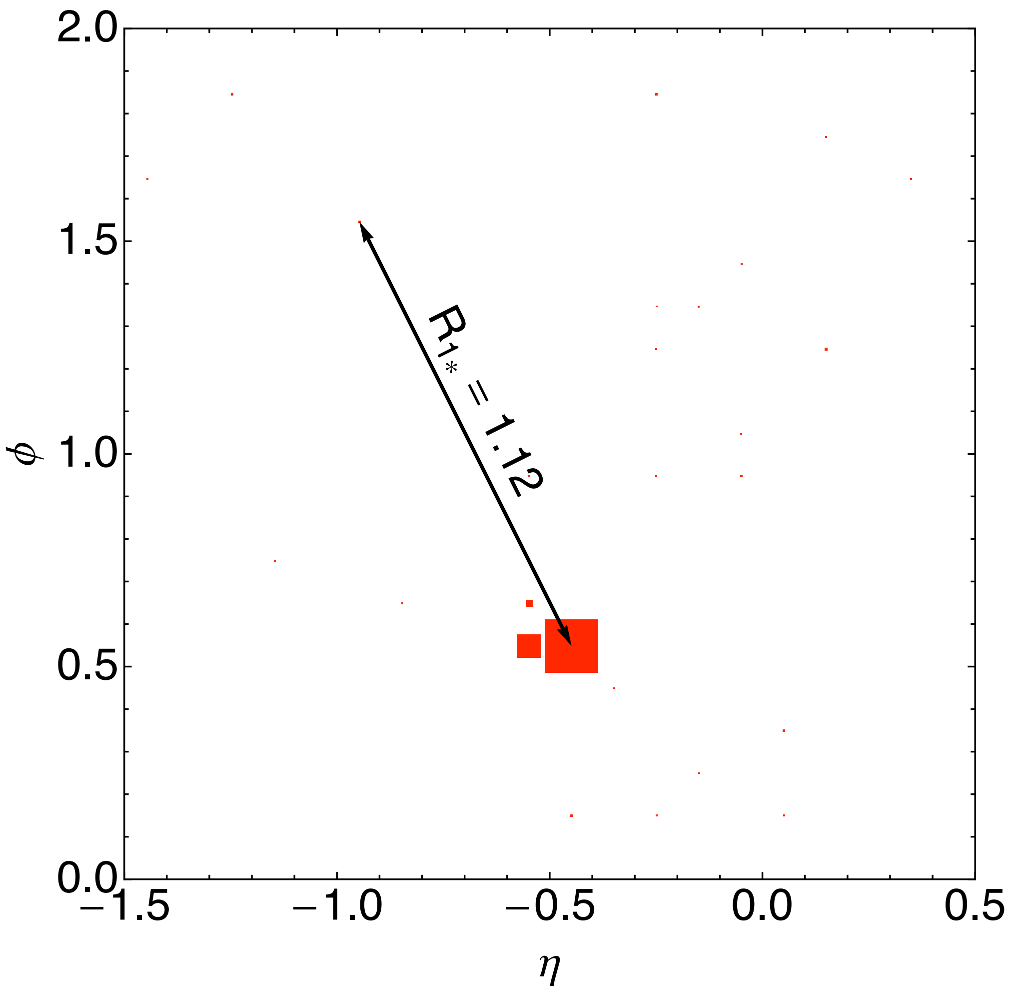
- Our approach:
  - Bin jets by the number of peaks
    - Related to the number of subjets
  - In each peak number bin:
    - Location in R of all peaks
    - Invariant mass of pairs of subjets from peak height
  - An explosion of observables!

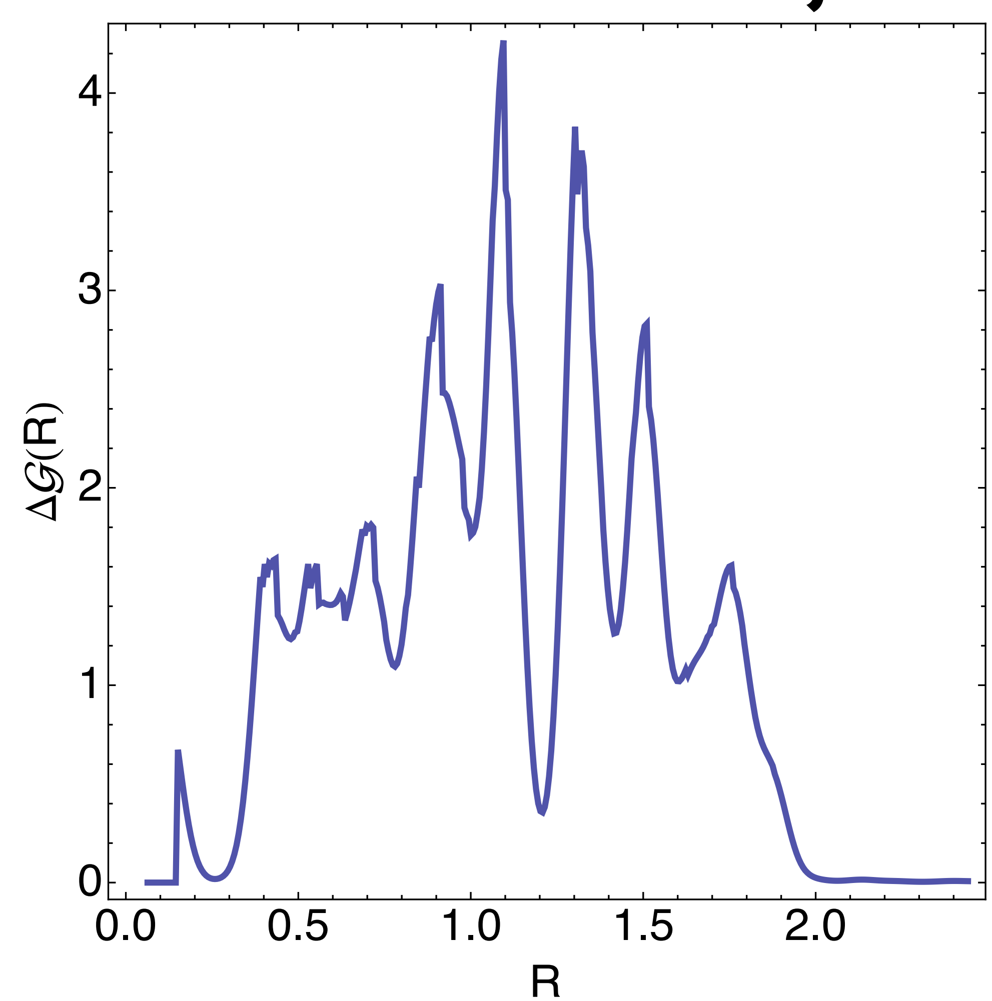
Jet Substructure Observables

- Our expectations:
  - $n_{\text{peaks}} \simeq \binom{n_{\text{subjets}}}{2}$

Jet Substructure Observables

- Our expectations:
  - $n_{\mathrm{peaks}} \simeq \binom{n_{\mathrm{subjets}}}{2}$
  - QCD: Flat distributions of peak locations; small invariant mass of subjets





Jet Substructure Observables

- Our expectations:
  - $n_{\text{peaks}} \simeq \binom{n_{\text{subjets}}}{2}$
  - QCD: Flat distributions of peak locations; small invariant mass of subjets
  - Heavy Particle: Peaks in distributions correlated with masses and momenta

Jet Substructure Observables

- Our expectations:
  - $n_{\text{peaks}} \simeq \binom{n_{\text{subjets}}}{2}$
  - QCD: Flat distributions of peak locations; small invariant mass of subjets
  - Heavy Particle: Peaks in distributions correlated with masses and momenta
  - Should be substantial discrimination power!

Application: Top Tagger

- Jets at the LHC
  - Very energetic, especially once at 14 TeV
  - QCD Jets will have large mass
- Radius of jet ~ $\frac{2m}{p_T}$
  - At Tevatron, tops have small p_T
  - At LHC, tops will be boosted
    - Can have top quark jets!

Application: Top Tagger

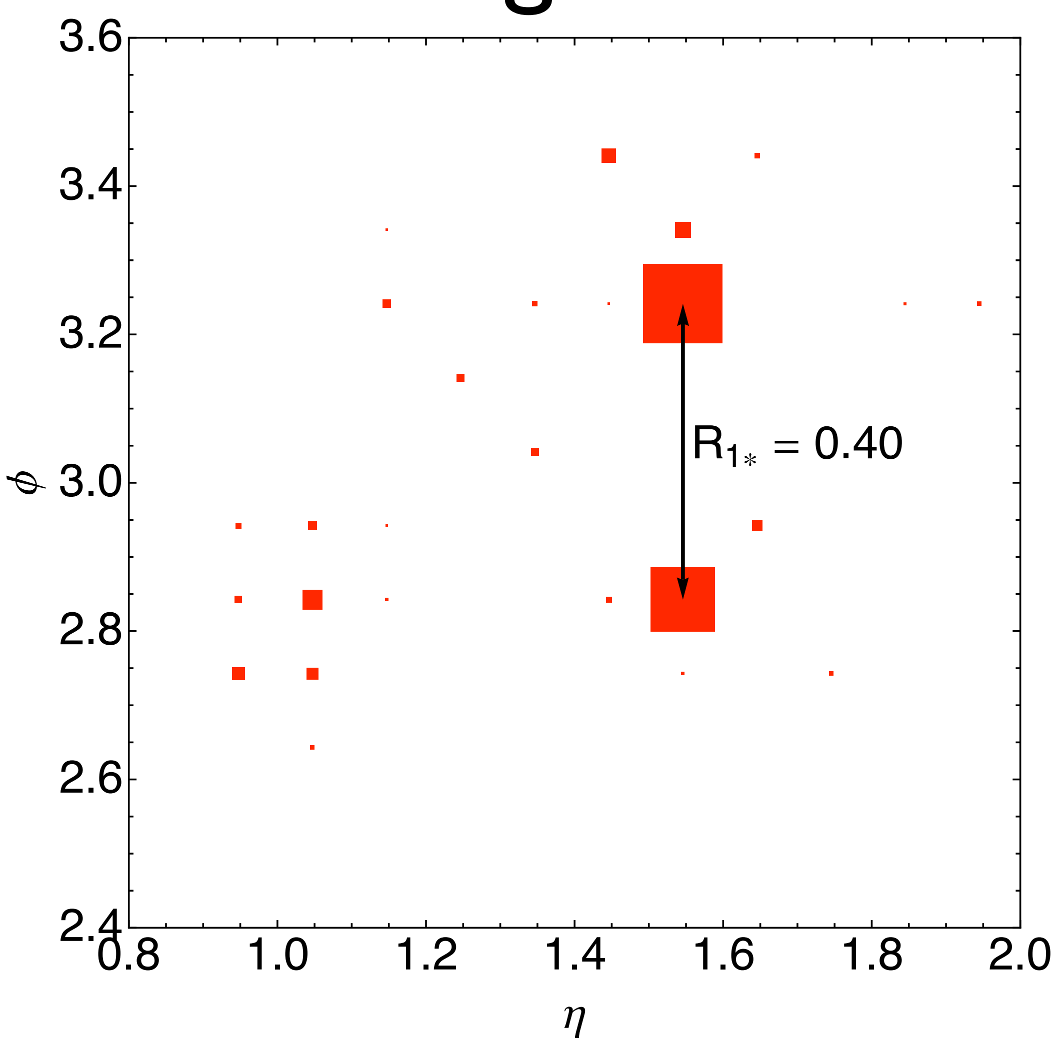
- Our Top Tagger:
  - Use observables from $\Delta\mathcal{G}(R)$
  - Expectations:
    - Top quark jets will have 3 subjets
    - Separation of subjets will be strongly correlated with m_W, m_t
    - Separation of subjets in QCD jet will be uncorrelated with any mass scale

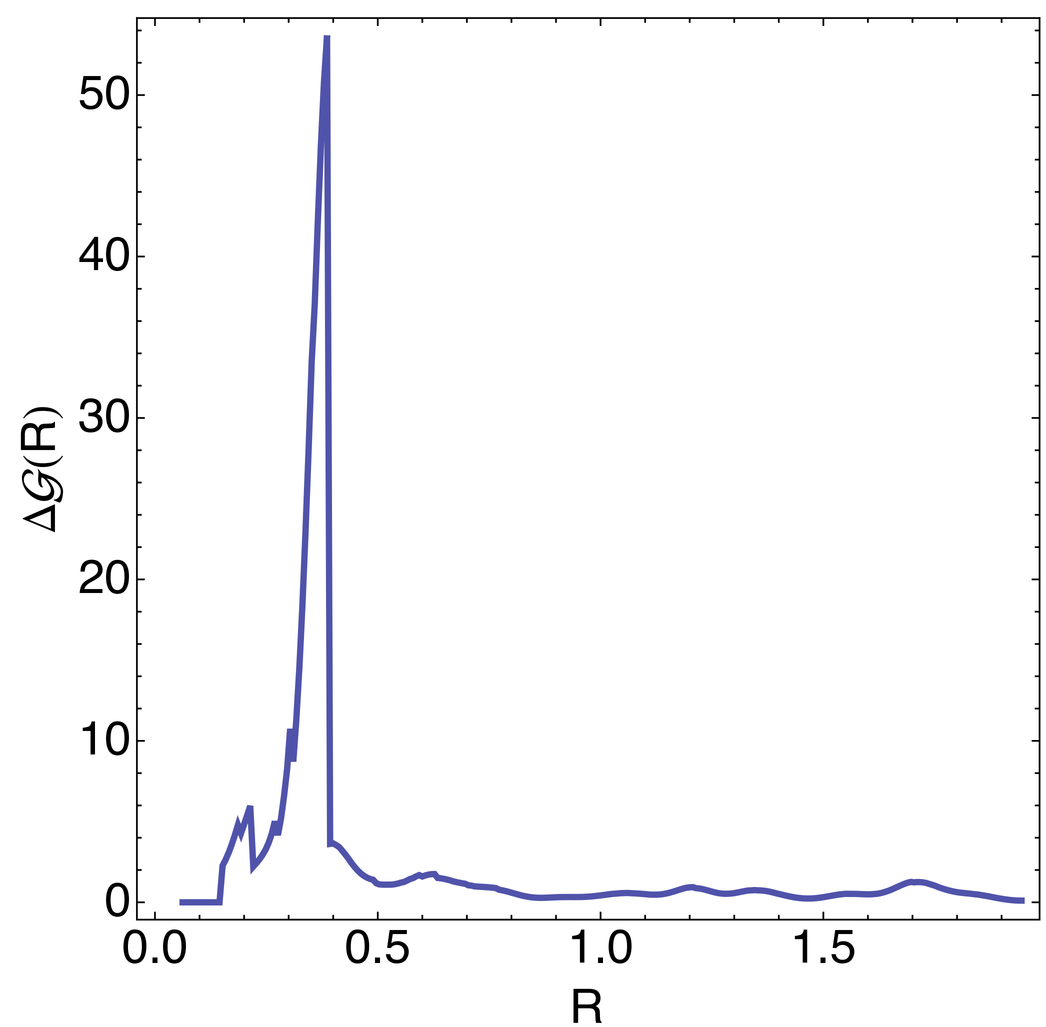
Application: Top Tagger

- Our Top Tagger:
  - Within a p_T bin:
    - Keep jets with 1, 2 or 3 peaks in $\Delta\mathcal{G}(R)$
    - Parton shower can smear subjets
    - Record each location in R and peak height for peaks with minimum prominence
    - Also include mass of jet as discriminant

Application: Top Tagger

- More on the number of peaks:
  - npeak = 1 bin ➔ W decay products are unresolved or subjets form an equilateral triangle





Application: Top Tagger

- More on the number of peaks:
  - npeak = 1 bin ➔ W decay products are unresolved or subjets form an equilateral triangle
  - npeak = 2 bin ➔ W decay products have large mass with b but not with each other or subjets form an isoceles triangle

Application: Top Tagger

- More on the number of peaks:
  - npeak = 1 bin ➔ W decay products are unresolved or subjets form an equilateral triangle
  - npeak = 2 bin ➔ W decay products have large mass with b but not with each other or subjets form an isoceles triangle
  - npeak = 3 bin ➔ All subjets of top are resolved

Application: Top Tagger

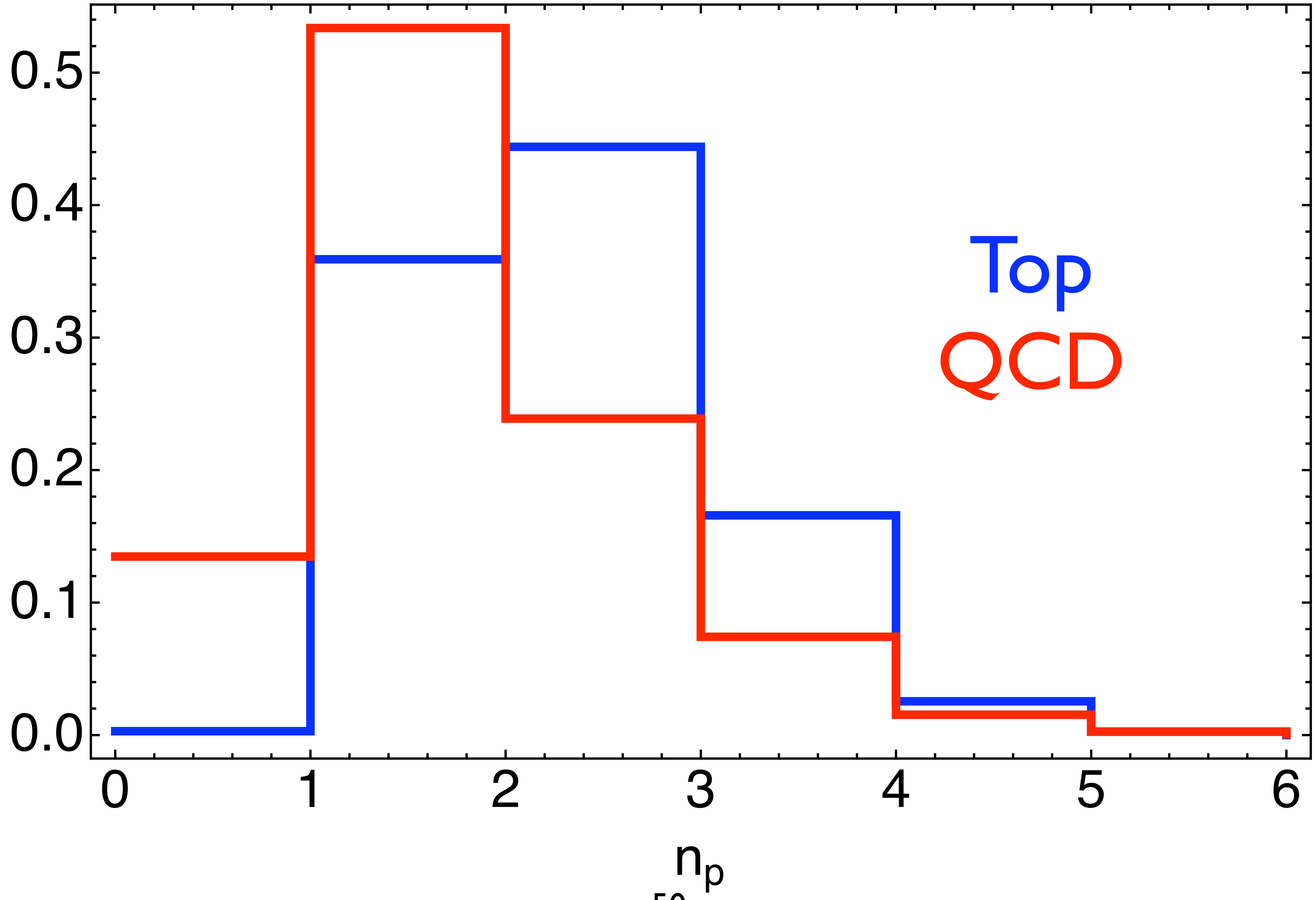
- More on the number of peaks:
  - npeak = 1 bin ➔ W decay products are unresolved or subjets form an equilateral triangle
  - npeak = 2 bin ➔ W decay products have large mass with b but not with each other or subjets form an isoceles triangle
  - npeak = 3 bin ➔ All subjets of top are resolved
  - Higher bins: hard radiation

Application: Top Tagger

- Use data generated for BOOST 2010
  Karagoz, Spannowsky, Vos
- Simulate detector by .1 x .1 binning in η, ϕ
- Find jets in events with FastJet 2.4.2 Cacciari, Salam implementation of CA with R = 1.5
- Study QCD jets and top quark jets in pt bins ranging from 200-800 GeV
- Set minimum prominence to 4.0
- Set delta function smoothing width to 0.06

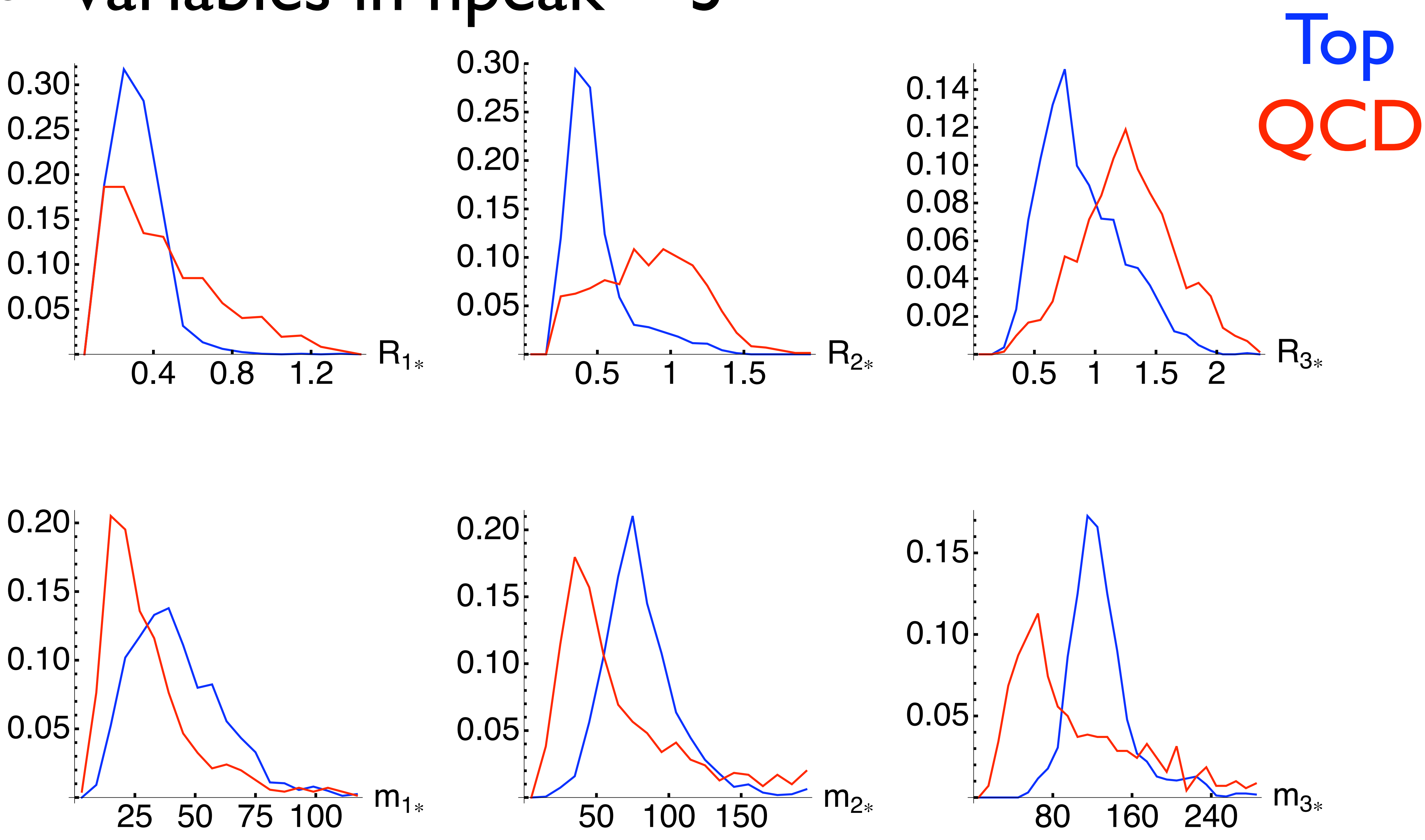
Application: Top Tagger

- Example: study 500-600 GeV pt bin
- Peak number distribution with minprom = 4.0

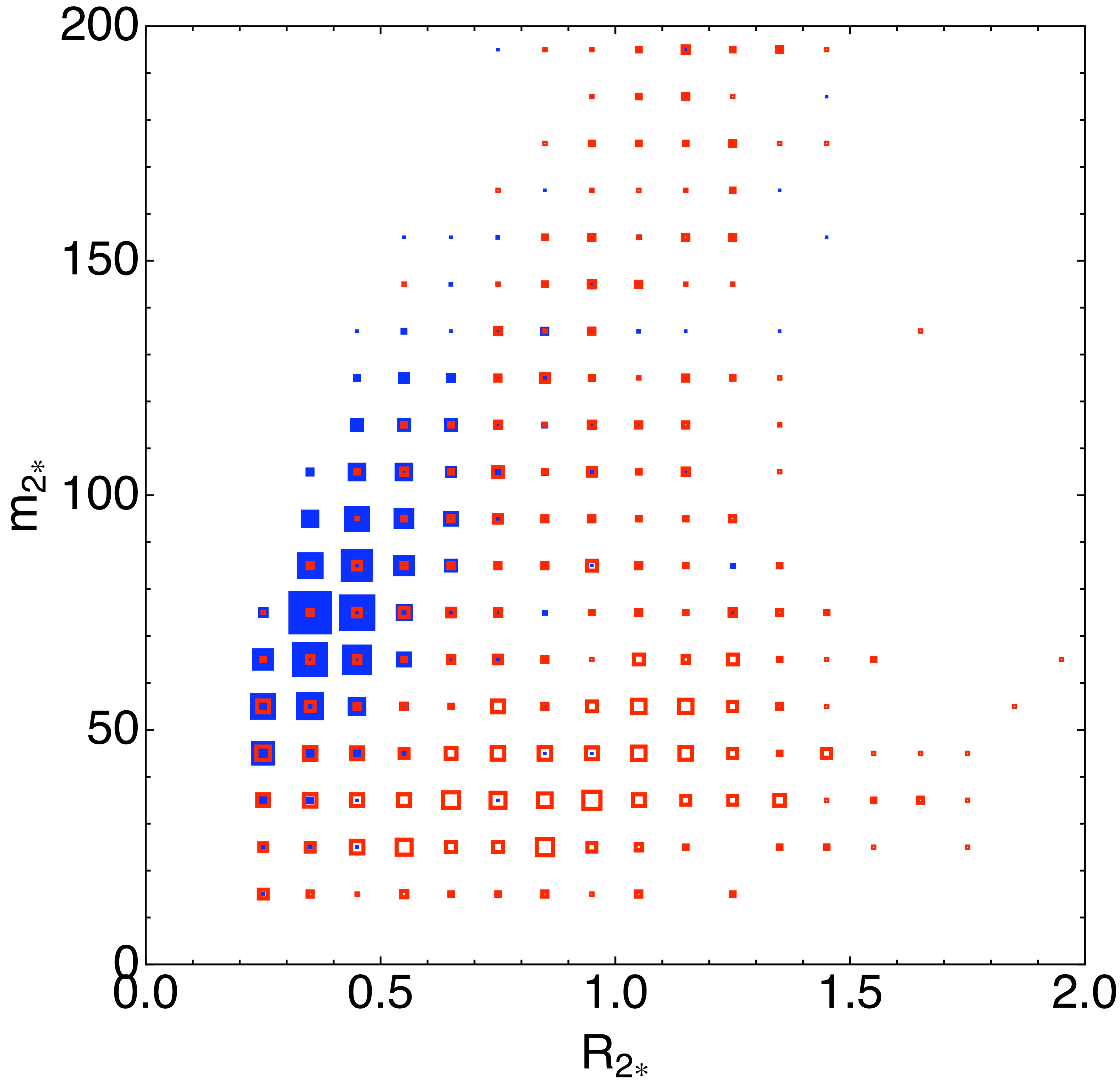


Application: Top Tagger

- Example: study 500-600 GeV pt bin
- Variables in npeak = 3



Application: Top Tagger



Top

QCD

- Correlation of separation of subjets and their invariant mass
- Top: m ~ R
- QCD: m, R uncorrelated

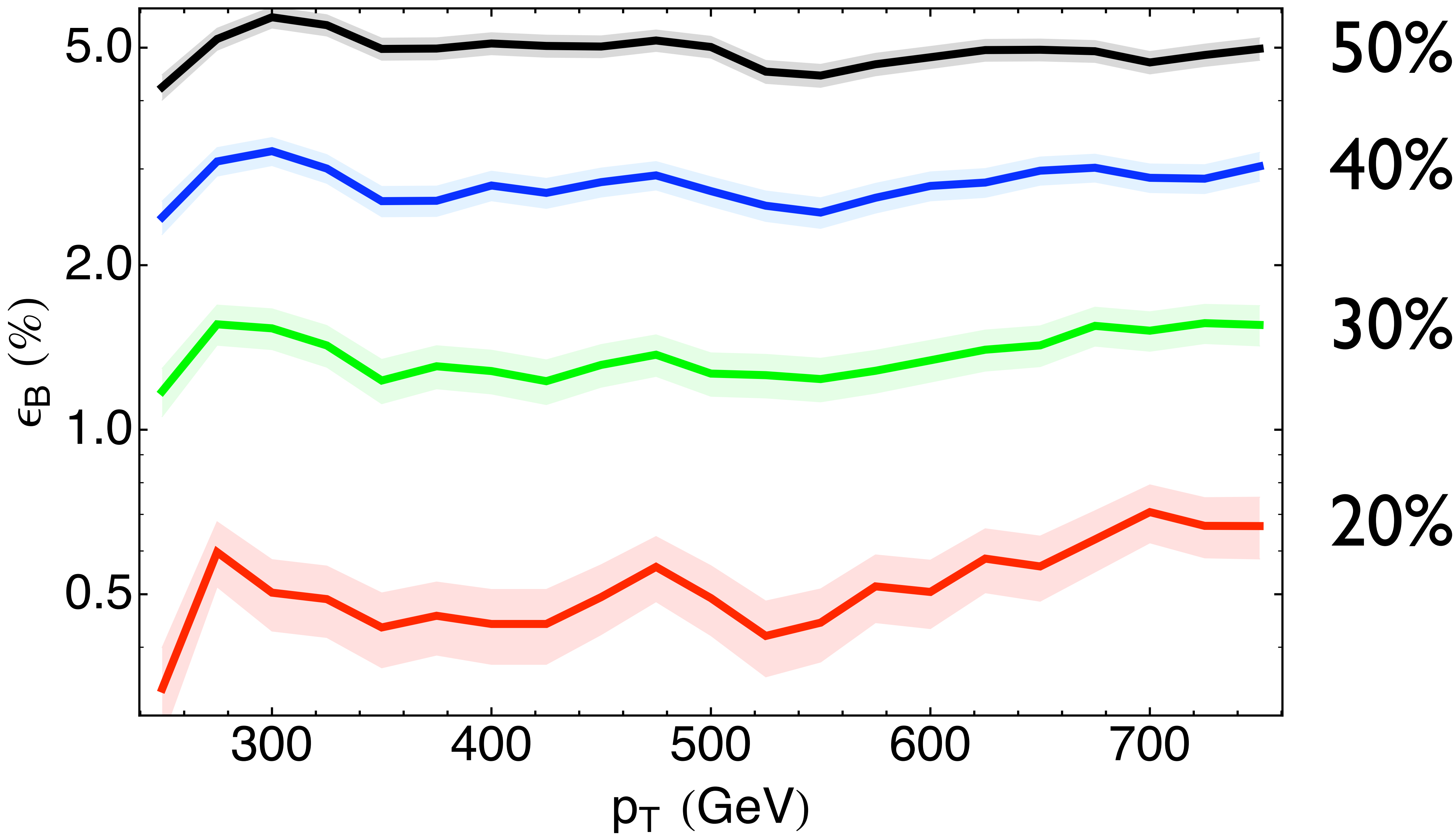
Application: Top Tagger

- General procedure:
  - In each pt and peak bin, use all observables for discrimination
  - Use Monte Carlo to sample cut locations and then compute efficiencies
  - Recombine results from npeak = 1, 2, 3 bins
  - Compute overall efficiencies for signal and background jets

Application: Top Tagger

- Results:

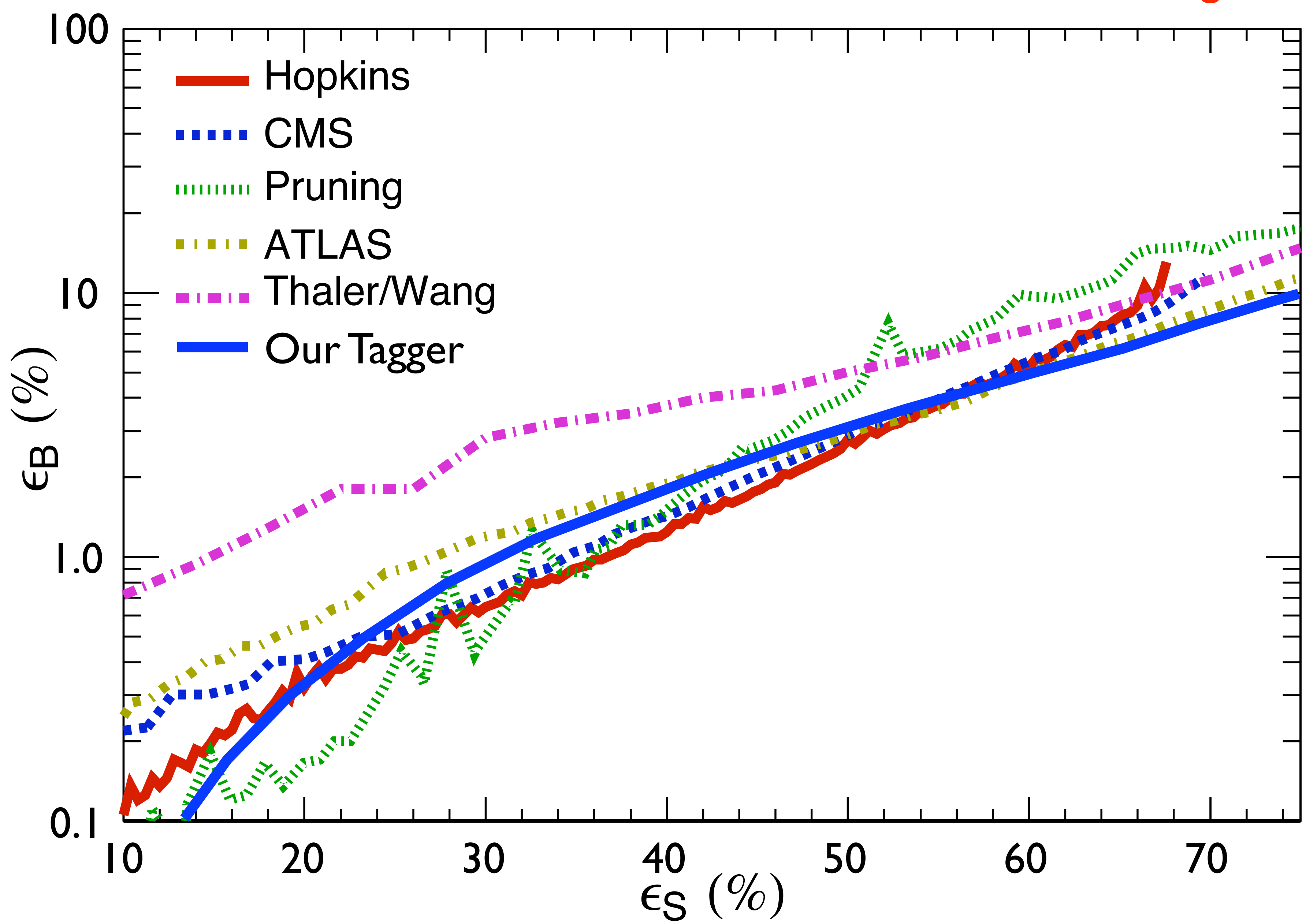
Sig Eff:



Application: Top Tagger

- Comparing to other top taggers:

Karagoz, Spannowsky, Vos



anti-kT, R=1.0, 500 GeV < pT < 600 GeV

Application: Top Tagger

- Top tagger competitive with other methods in the literature
- Important: Still substantial optimization that can be done
  - Better choice of variables?
  - Prominence as a function of pt?
  - Use more information about structure of top jet?

Future/Current Directions

- Probing Perturbative QCD and UE
  - What is ensemble average of $\Delta\mathcal{G}(R)$?
  - What are the quark and gluon contributions?
  - Can we distinguish UE models?
- Adaptive Grooming
  - IRC safe scales are defined
  - Can these scales be used to improve mass resolution?

Summary

- Introduced a method for finding scales within a jet without using branching tree

- First Application: Top Tagger competitive with other methods

- Many other possible applications!