Jet Substructure Without Trees

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

BOOST, May 22, 2011

Goal: Find Hard Scales in Jets

- Define an angular correlation function
- Define an angular transform to find important scales
- Construct a robust method for finding structure in the transform
- Define IRC safe observables that can be used to characterize the structure of a jet
- Test by constructing a top tagger

Angular Correlations

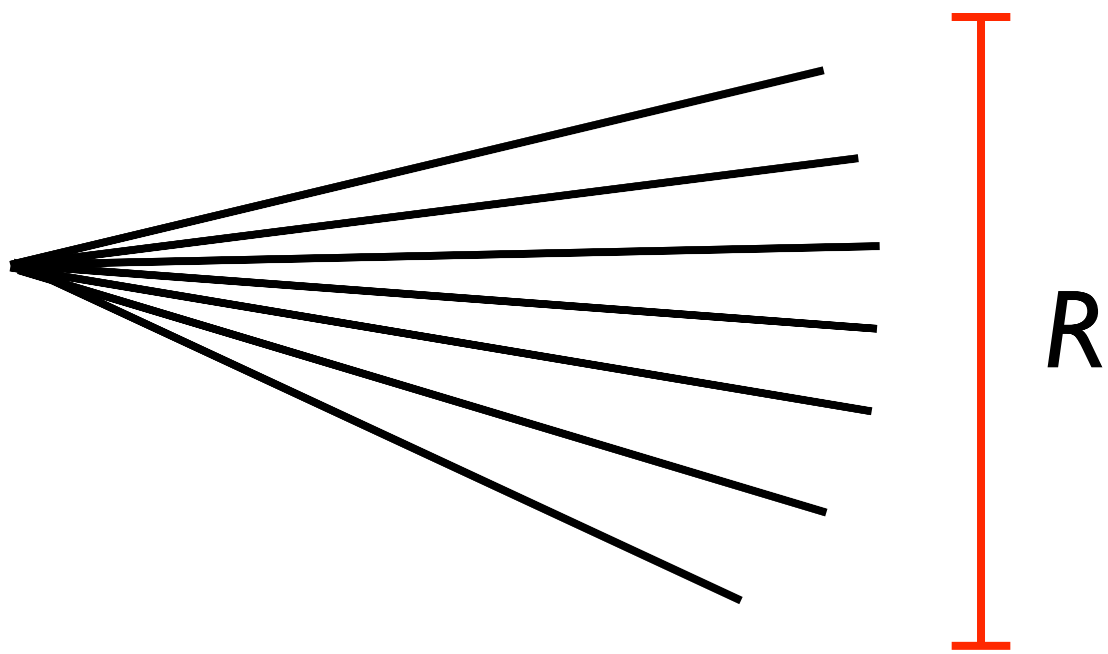
- For any IRC safe set of particles {*i*}:

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

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

Angular Correlations

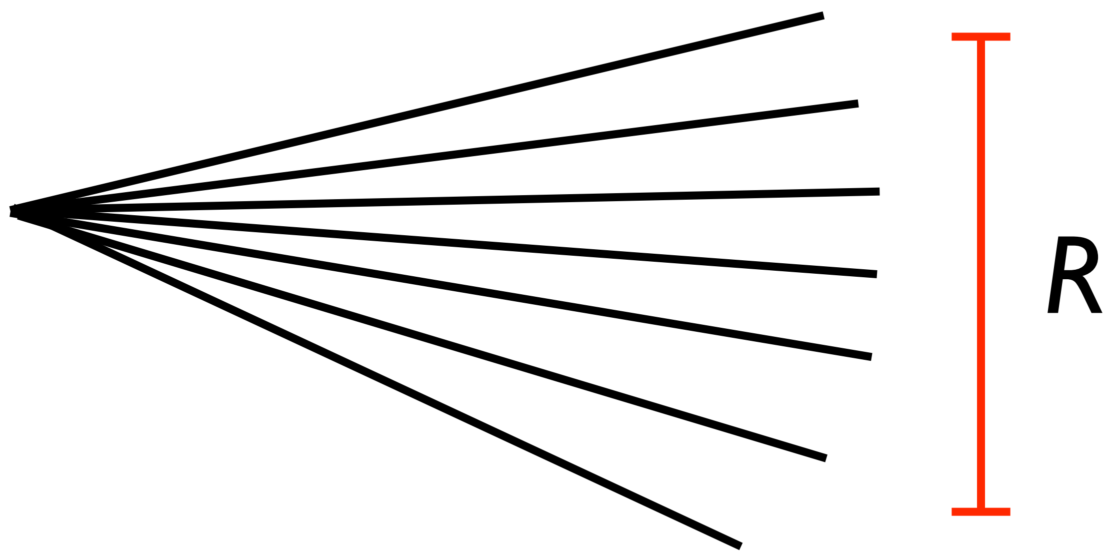
- What is the physical picture?
  - Jet with ~no substructure



  - Angular correlation function is smooth
  - In QCD, $\mathcal{G}(R)$ ~ R^2

Angular Correlations

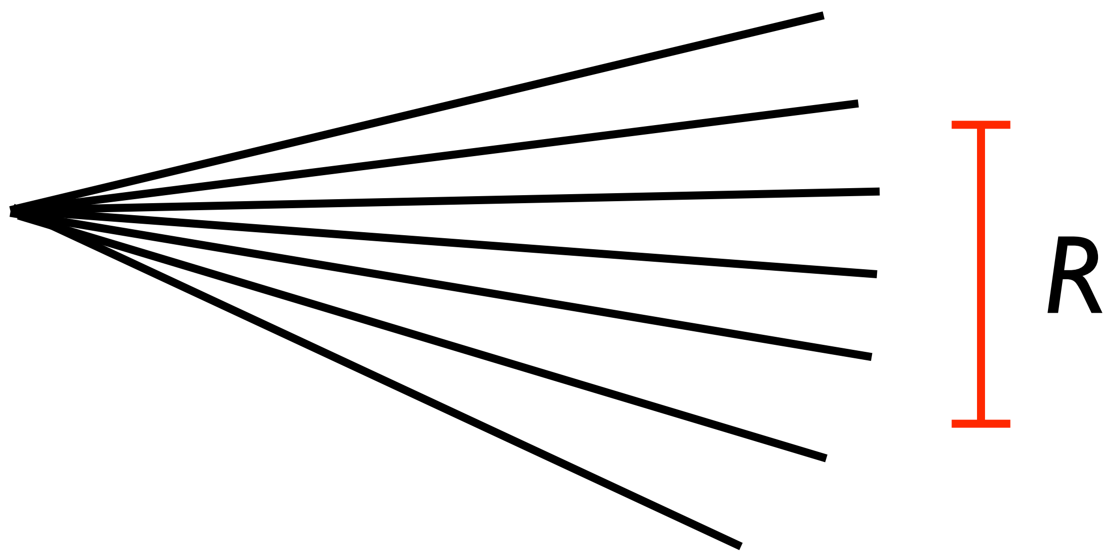
- What is the physical picture?
  - Jet with ~no substructure



  - Angular correlation function is smooth
  - In QCD, $\mathcal{G}(R)$ ~ R^2

Angular Correlations

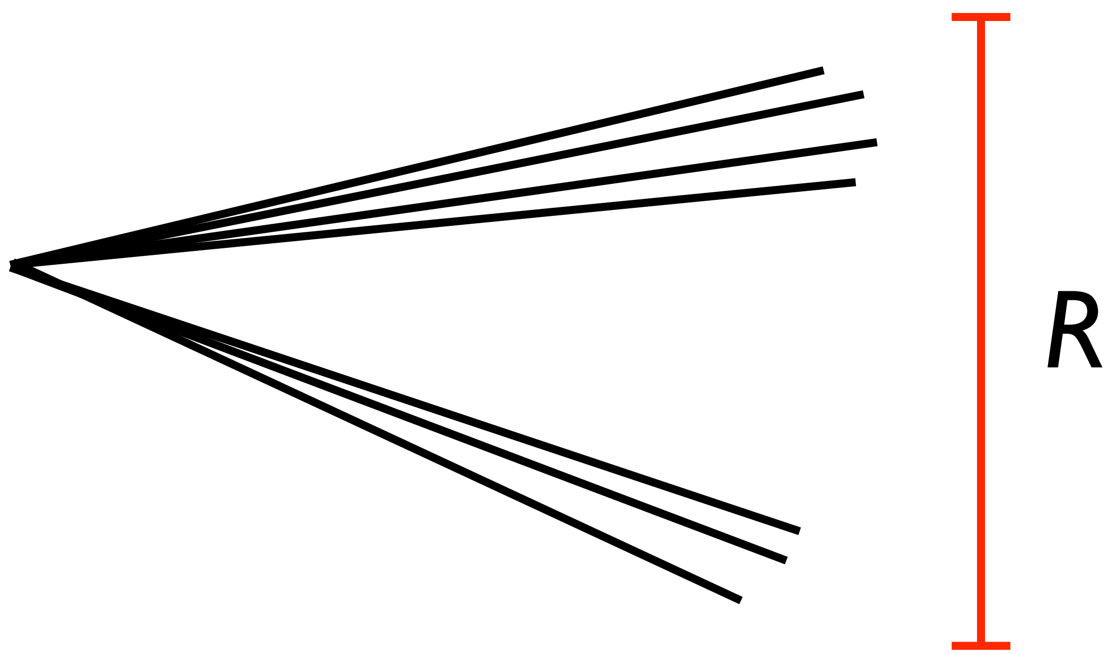
- What is the physical picture?
  - Jet with ~no substructure



  - Angular correlation function is smooth
  - In QCD, $\mathcal{G}(R)$ ~ R^2

Angular Correlations

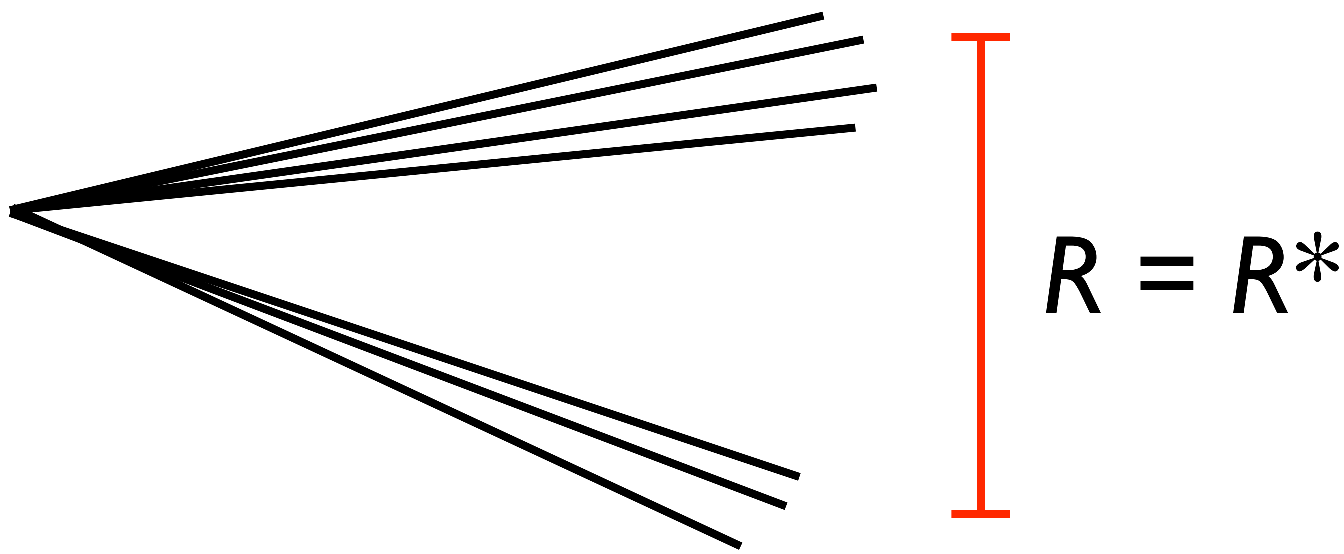
- What is the physical picture?
  - Jet with substructure



  - $\mathcal{G}(R)$ ~ 1

Angular Correlations

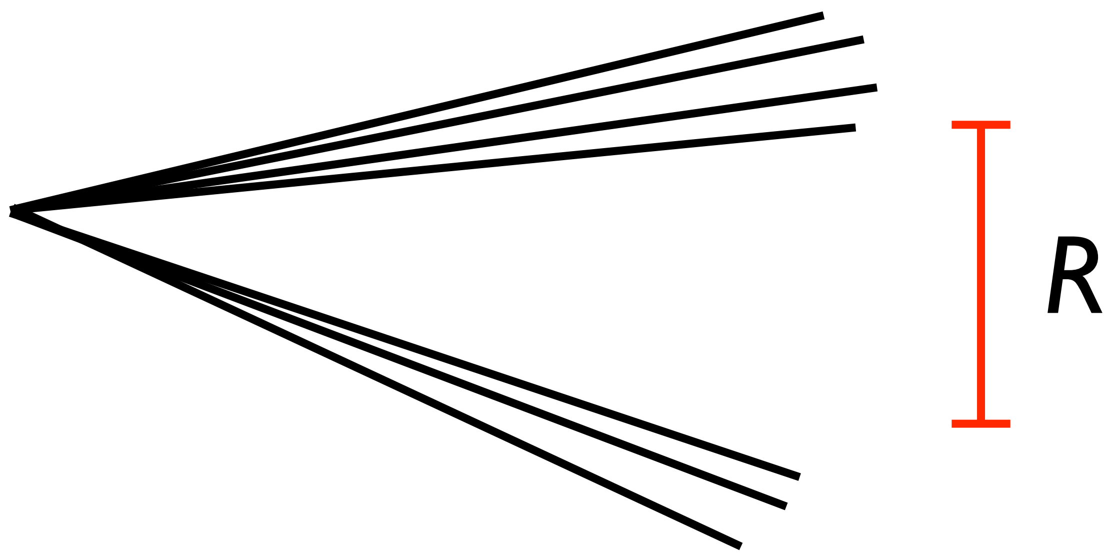
- What is the physical picture?
  - Jet with substructure



  - $\mathcal{G}(R) \sim 1$

Angular Correlations

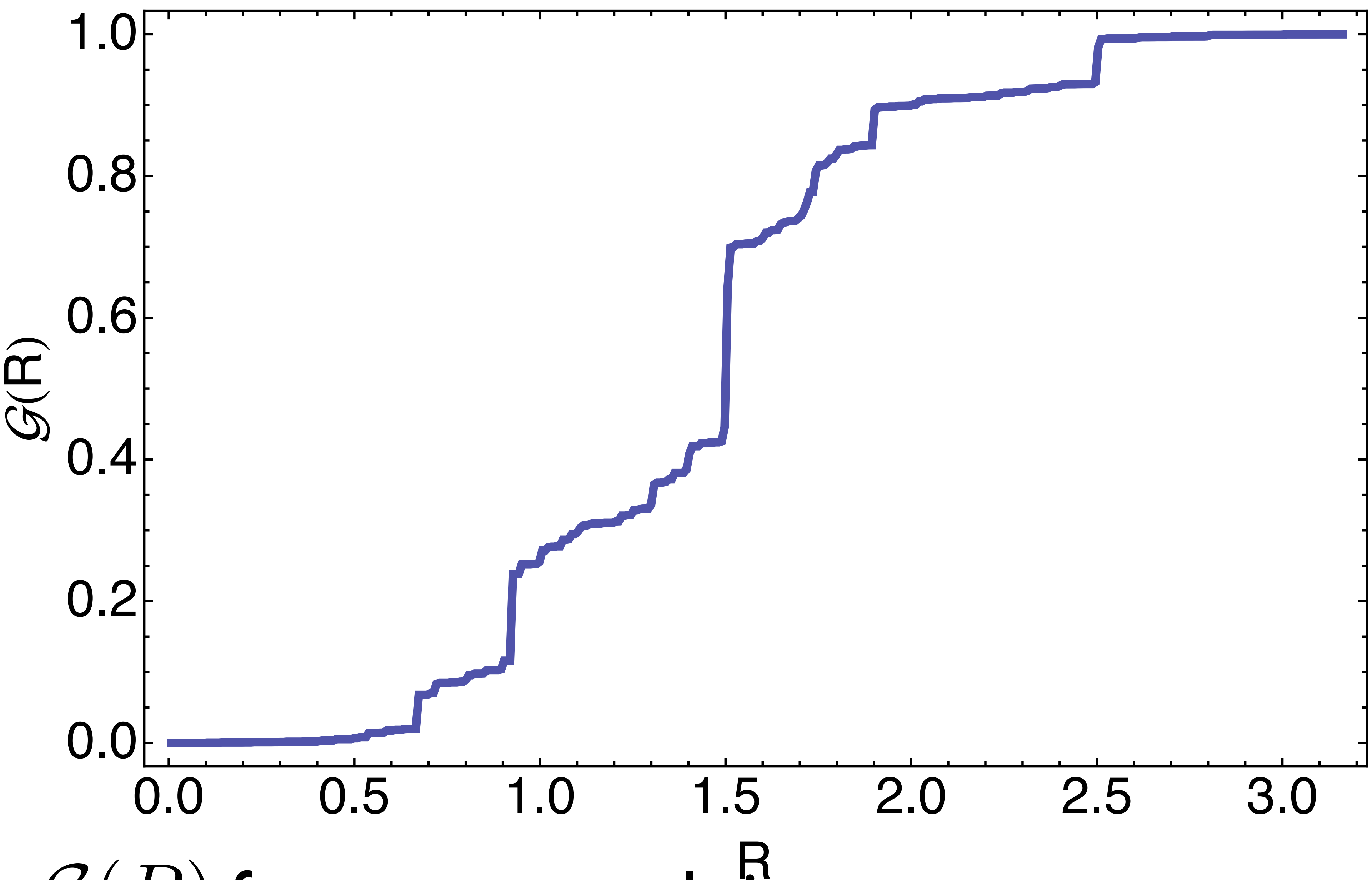
- What is the physical picture?
  - Jet with substructure



  - $\mathcal{G}(R)$ ~ 0
  - Angular correlation is discontinuous at *R**

Angular Correlations

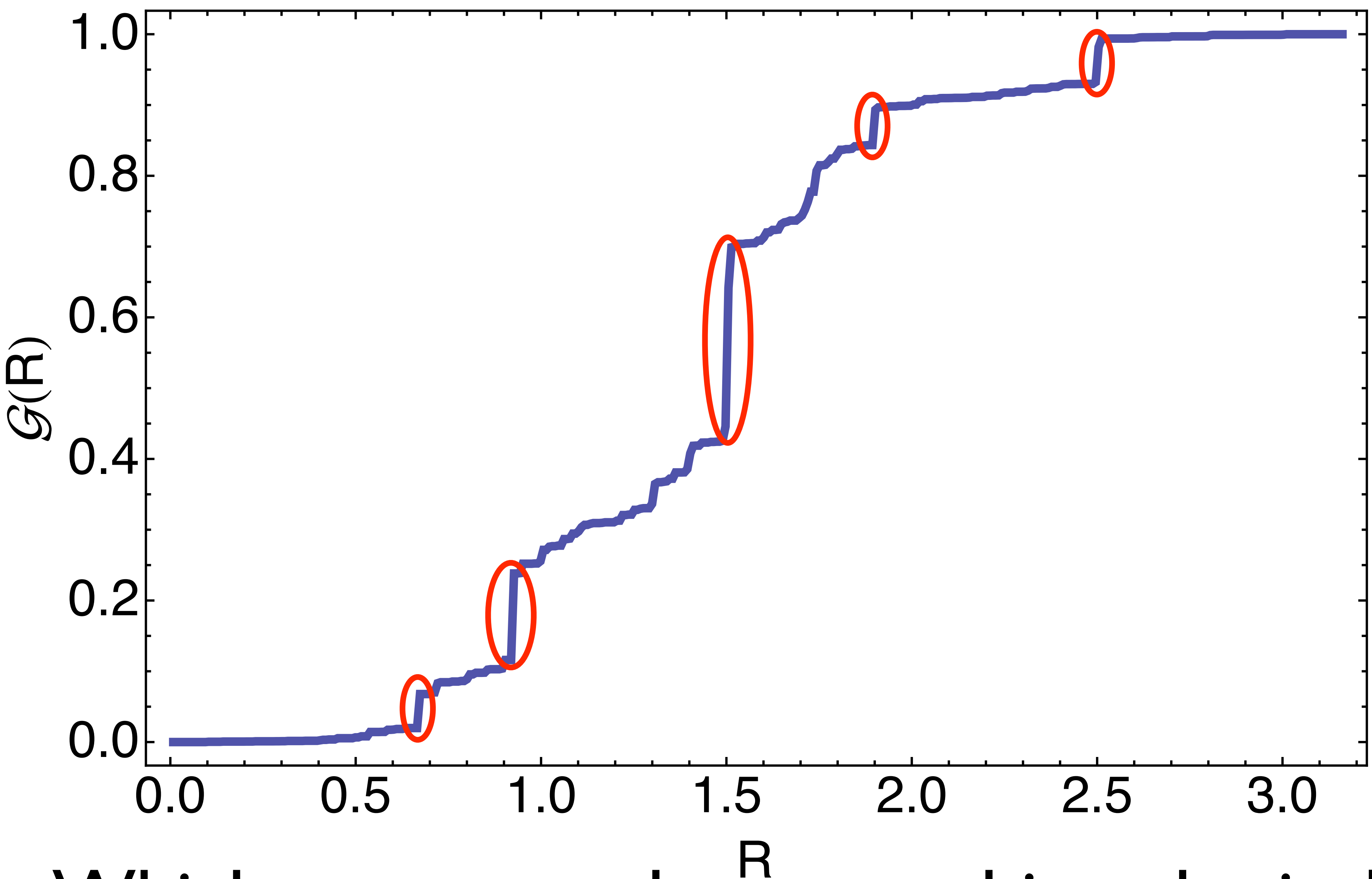
- Ledges in $\mathcal{G}(R)$ = separation of hard subjets



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

Angular Correlations

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



- Which correspond to something physical?

Angular Structure

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

Angular Structure

- 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 Structure

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

Angular Structure

- 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:

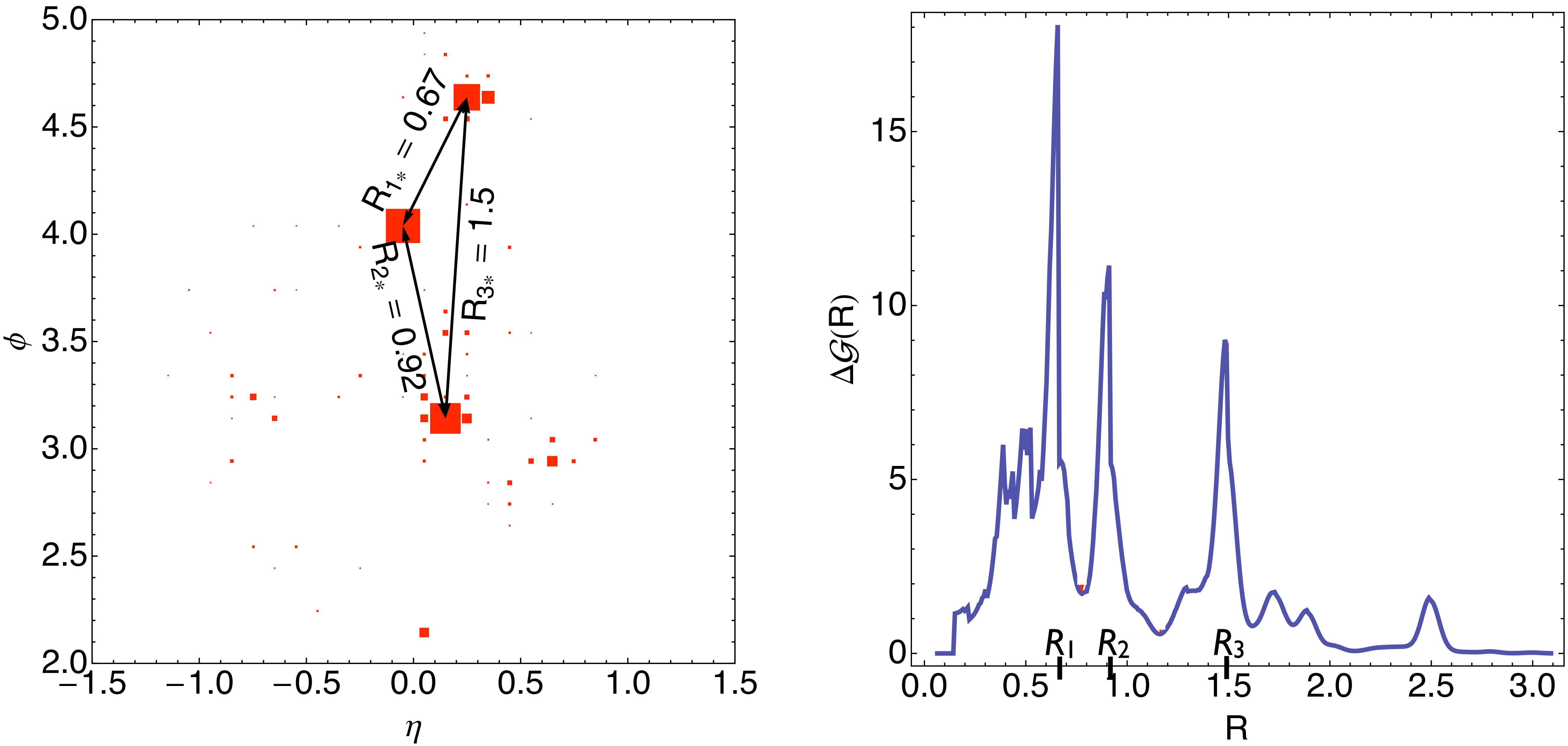
$$\delta(R - \Delta R_{ij}) \simeq \frac{e^{-\frac{(R - \Delta R_{ij})^2}{2 dR^2}}}{dR \sqrt{2\pi}}$$

Angular Structure

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

Angular Structure

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

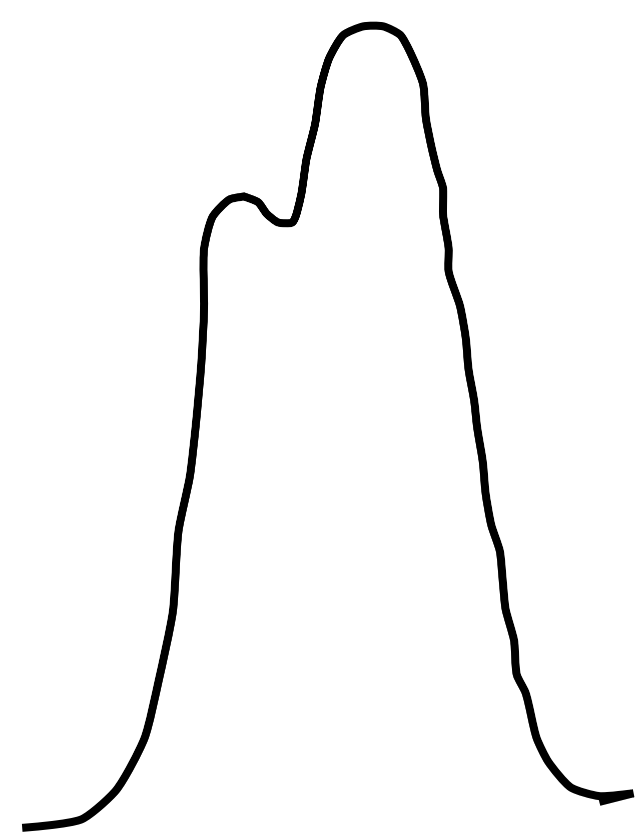


Answer: Yes!

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? Why?

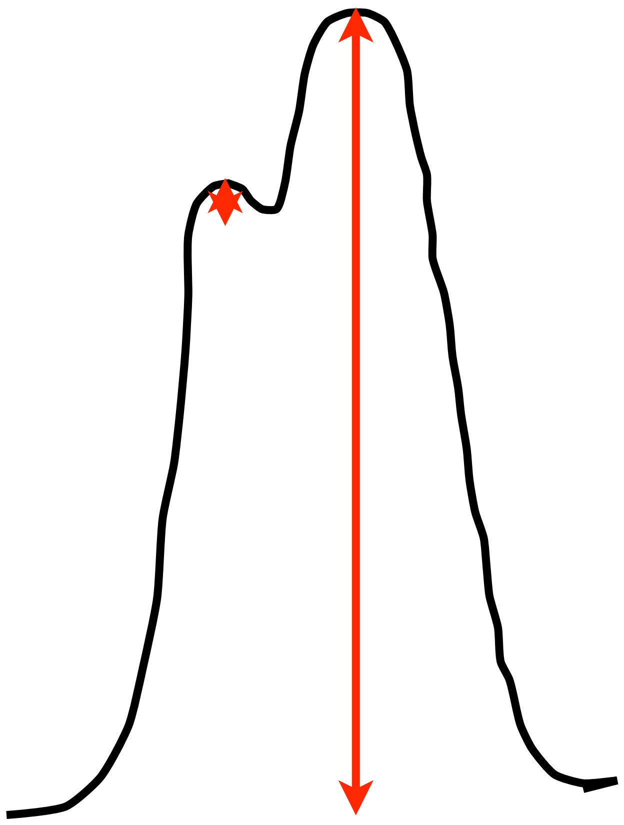
Is the little bump interesting?



Prominence

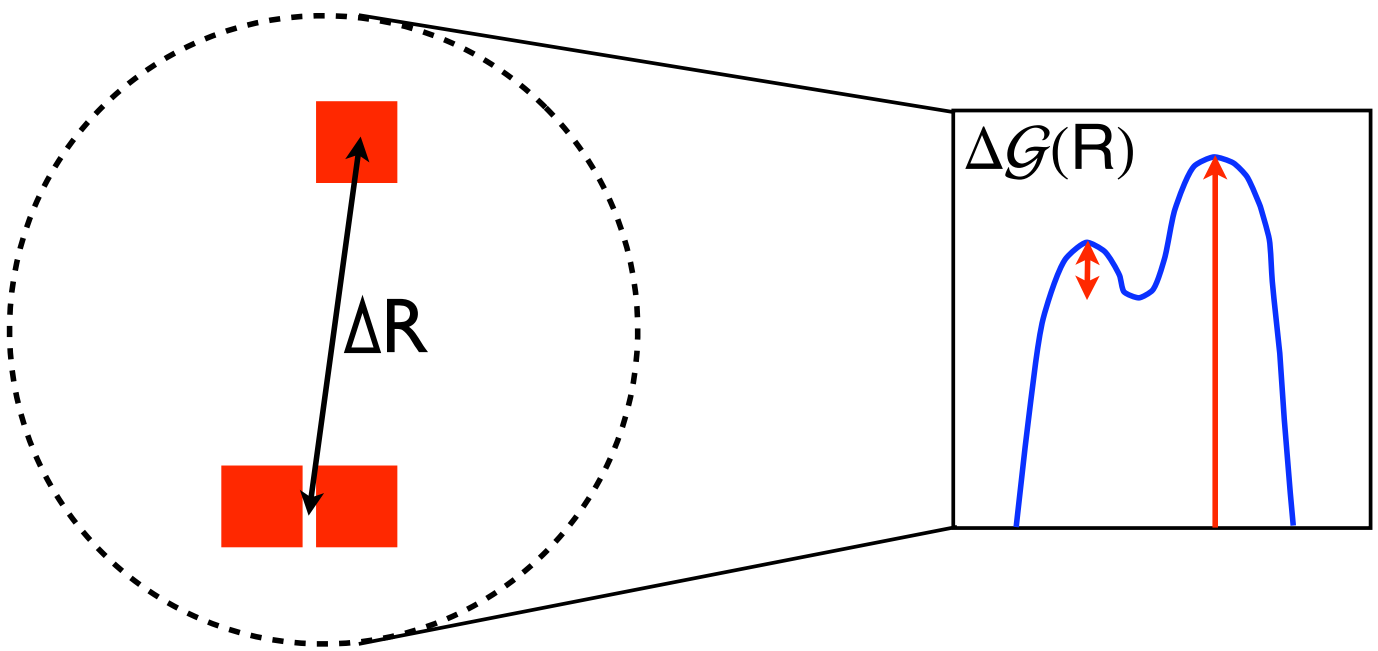
- Proposition: Define peaks by their *prominence*
- Prominence = amount peak sticks out above ambient background

Prominence
of little bump
is tiny!



Prominence

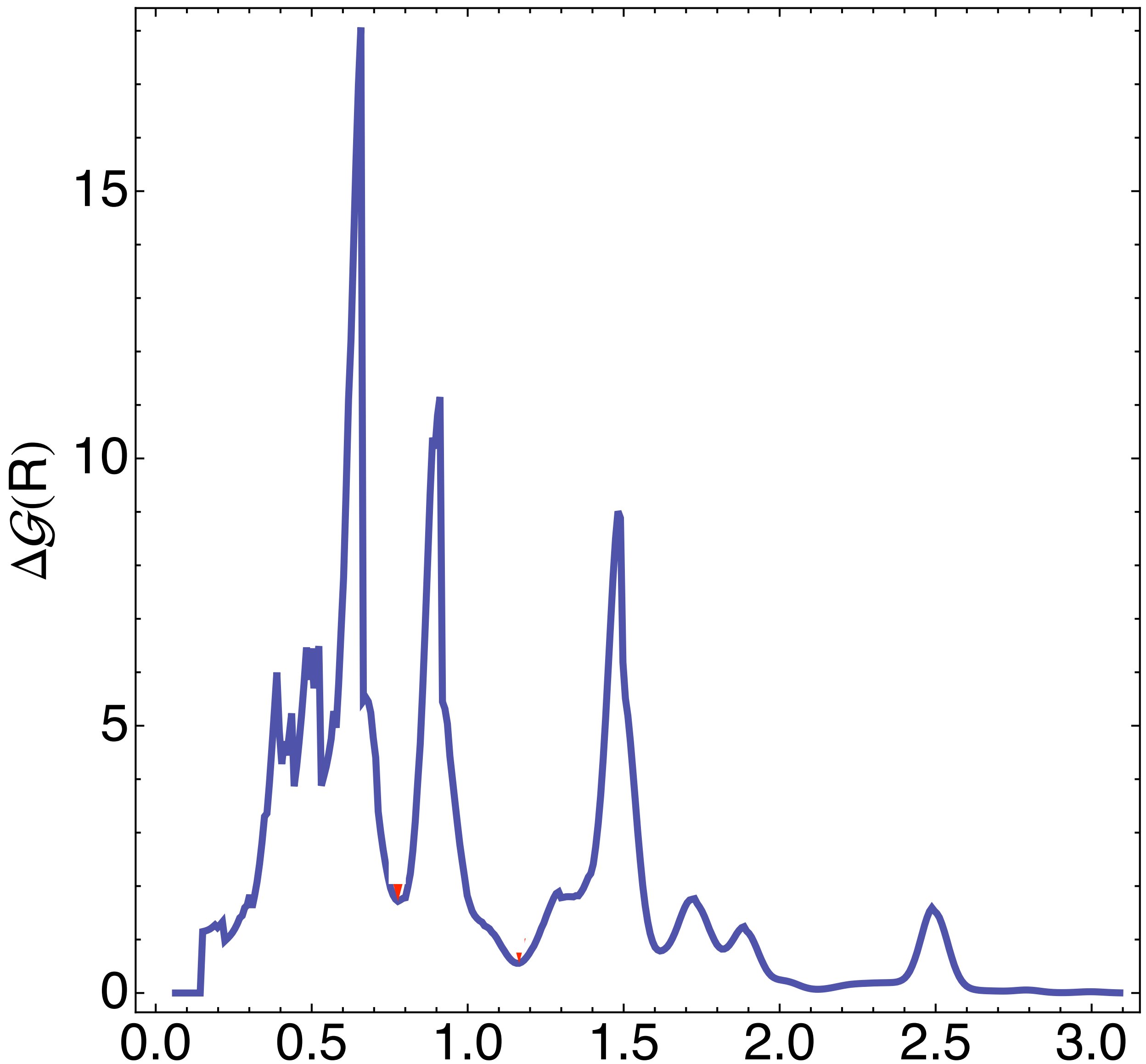
- Possible double counting of angular scales



- Defining interesting peaks by prominence removes double counting ambiguity

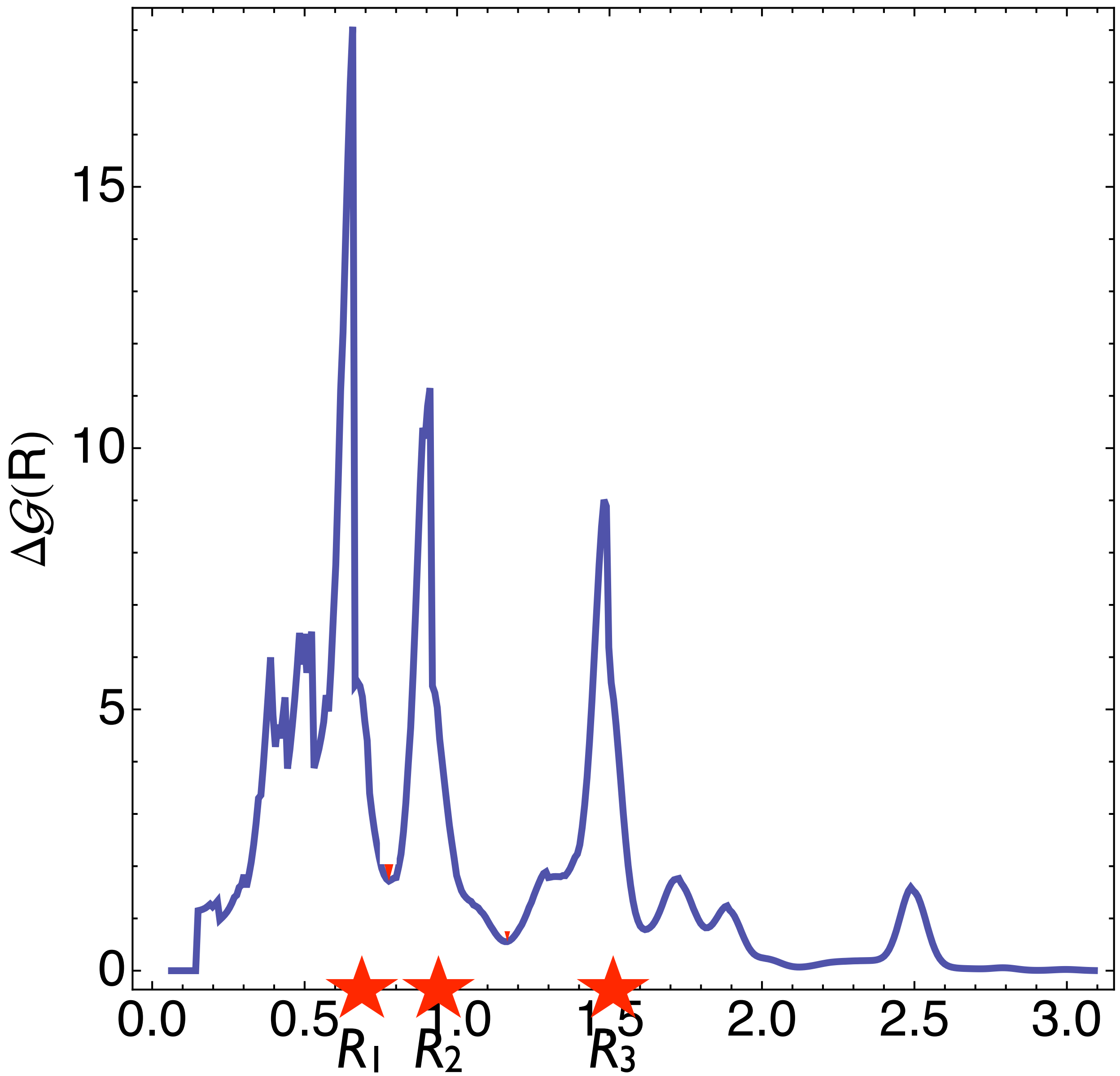
Defining Observables

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



Defining Observables

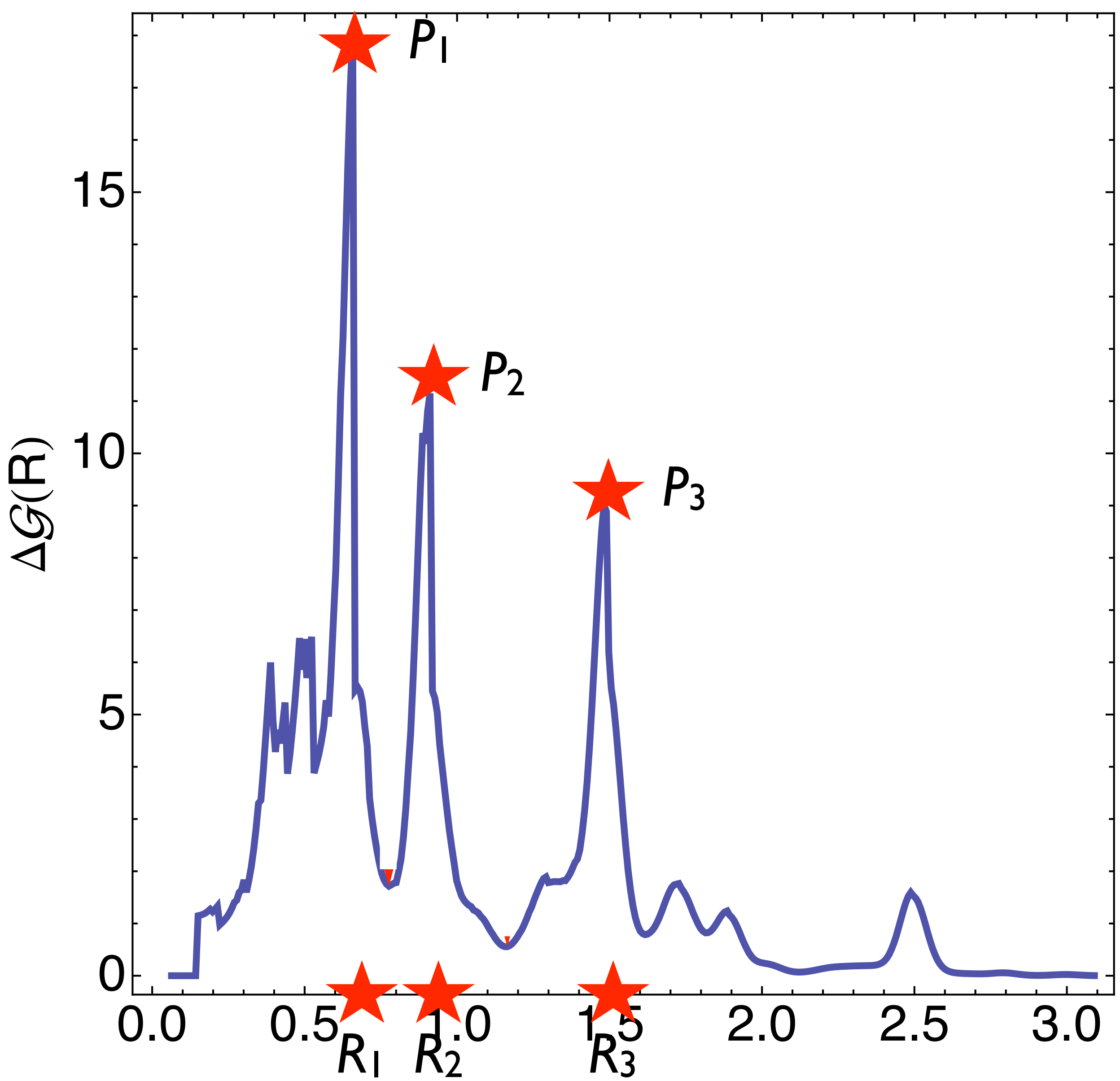
- IRC safe observables from $\Delta\mathcal{G}(R)$:



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

Defining Observables

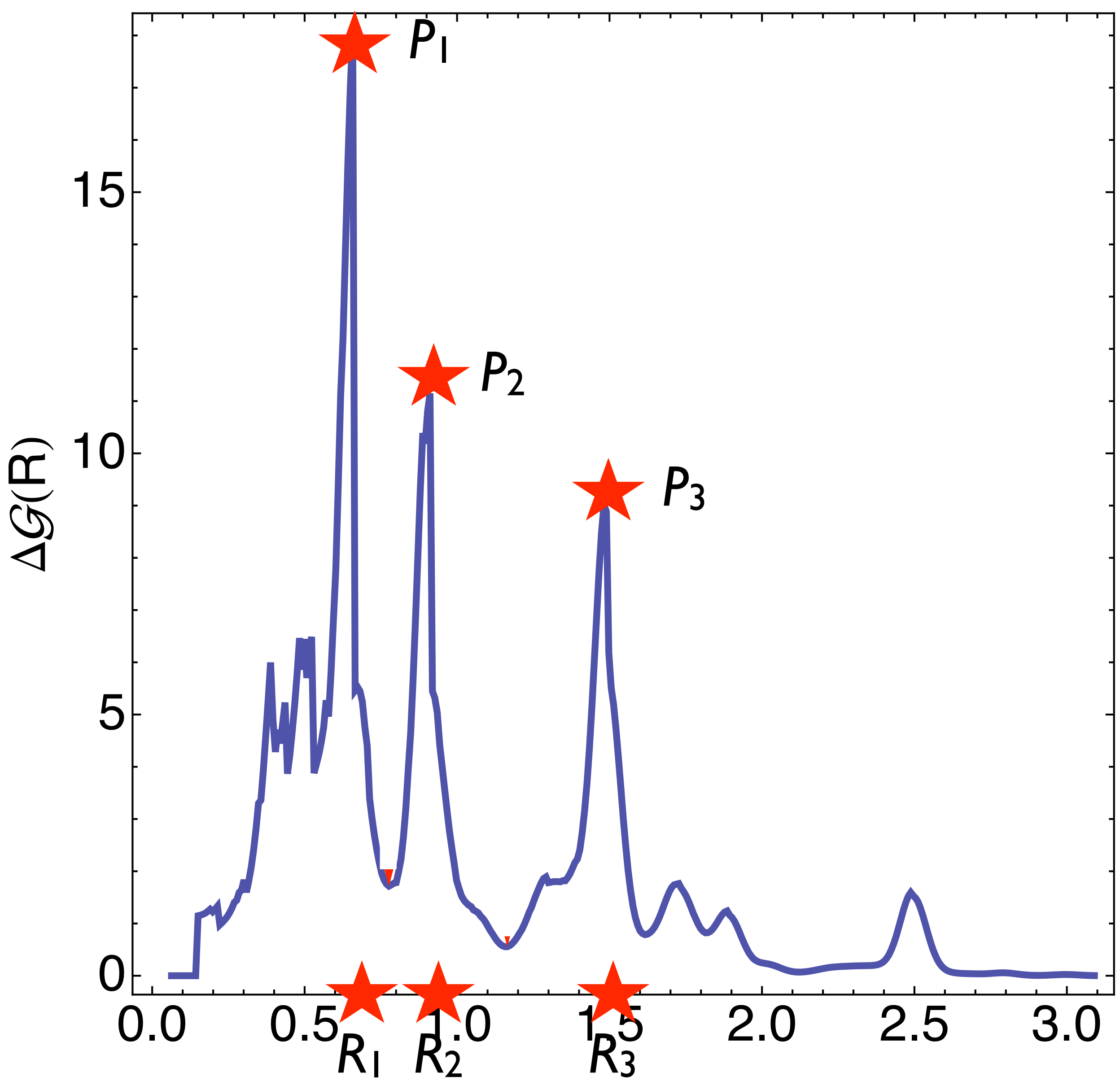
- IRC safe observables from $\Delta\mathcal{G}(R)$:



- Location of peaks in R
- Height of peaks

Defining Observables

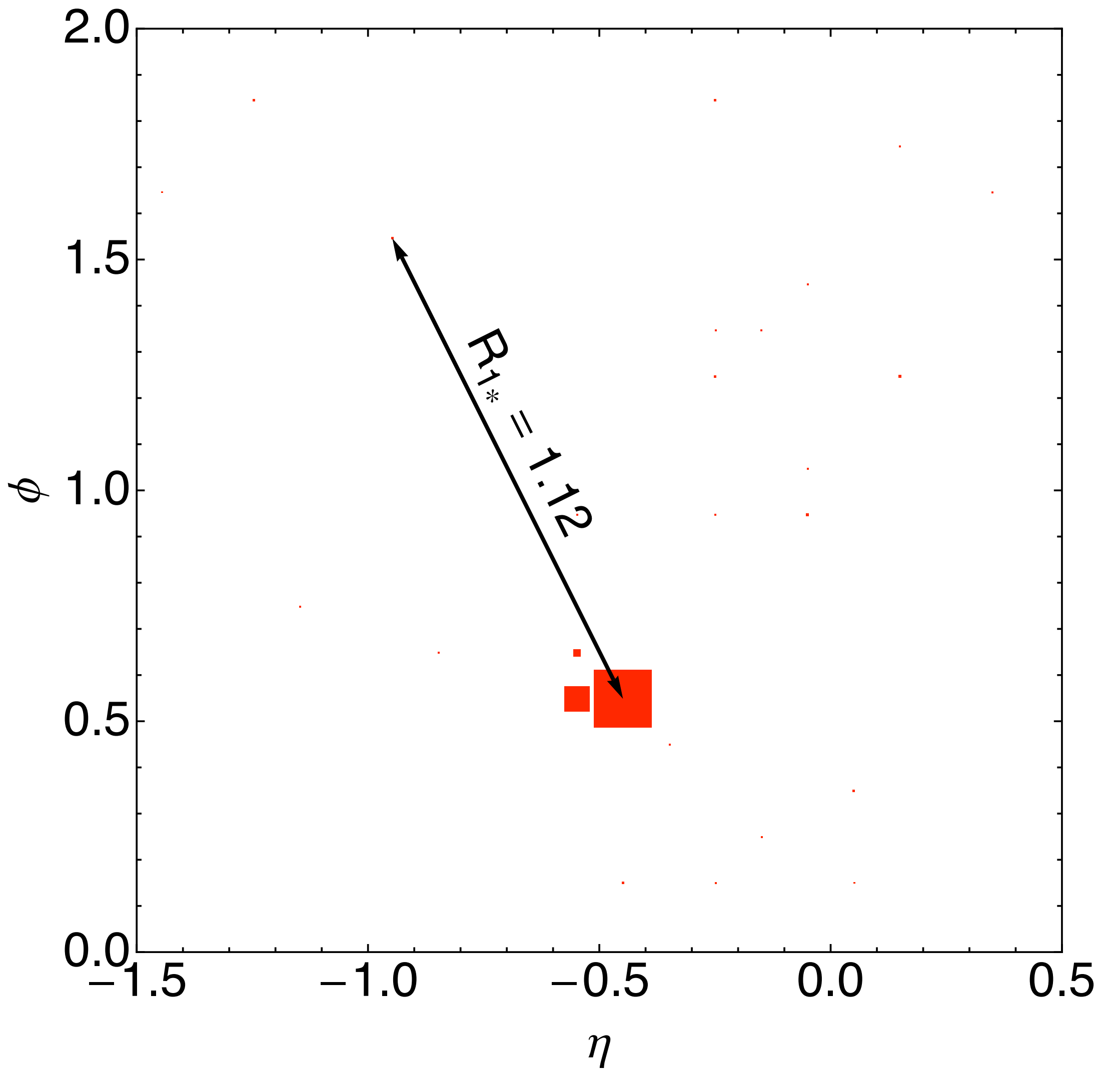
- IRC safe observables from $\Delta\mathcal{G}(R)$:

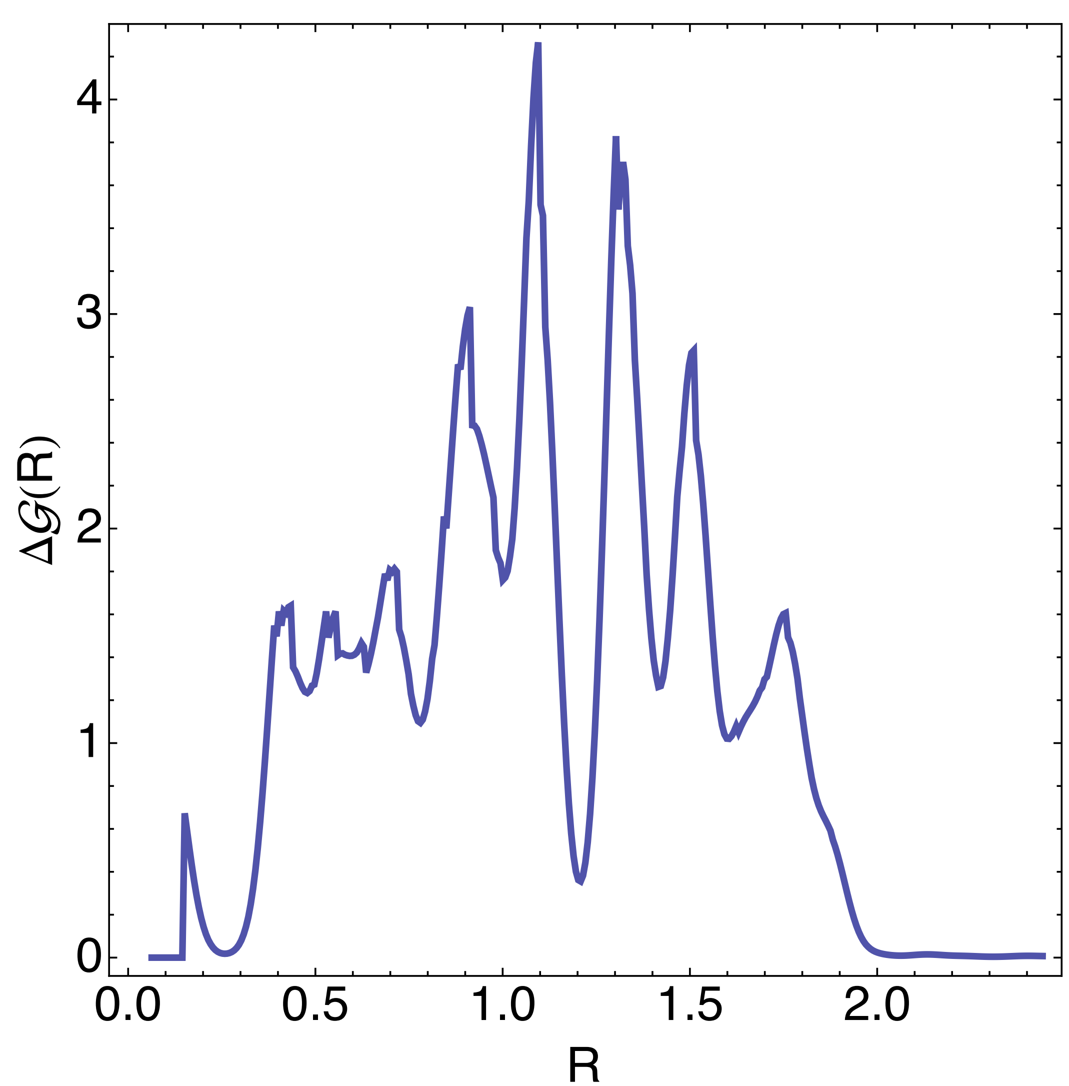


- Location of peaks in R
- Height of peaks
- Number of peaks

Current/Future Directions

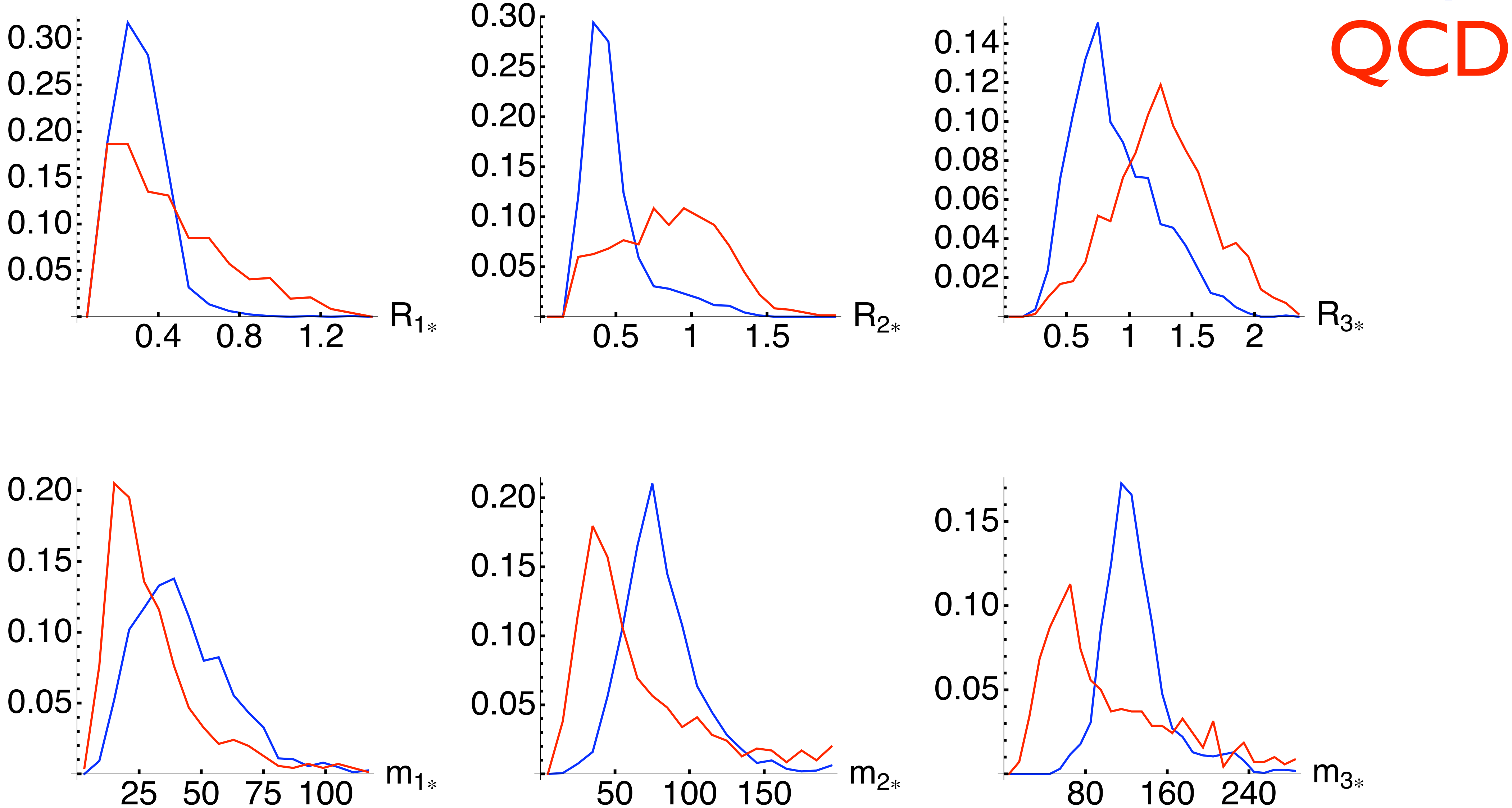
- Constructed a top tagging algorithm competitive with others in the literature (discussed in paper)





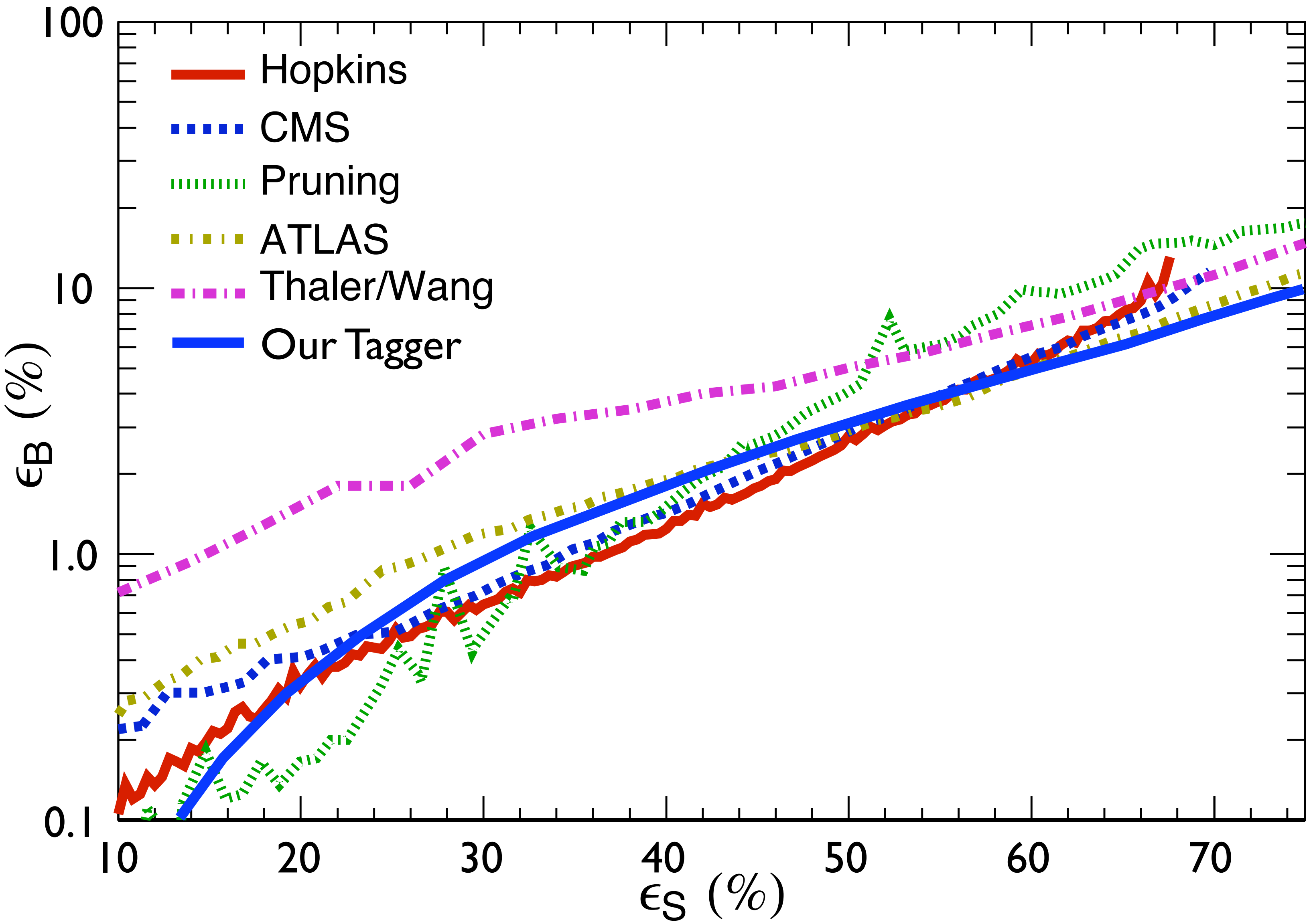
Current/Future Directions

- Constructed a top tagging algorithm competitive with others in the literature (discussed in paper)



Current/Future Directions

- Constructed a top tagging algorithm competitive with others in the literature



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

Back-up Slides

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

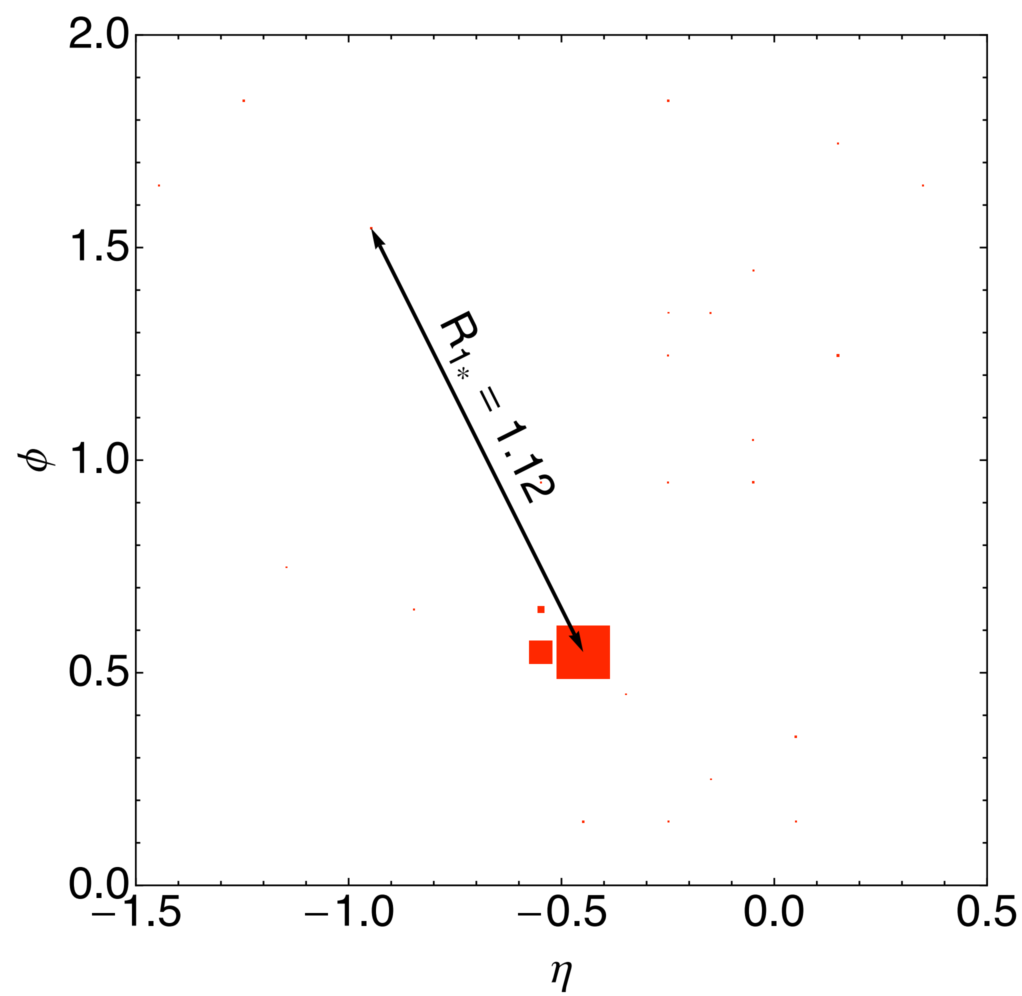
- Quiz: What is the highest mountain in the contiguous US?
  - Mt. Whitney, CA
- What is the most *prominent* mountain in the contiguous US?
  - Mt. Rainier, WA

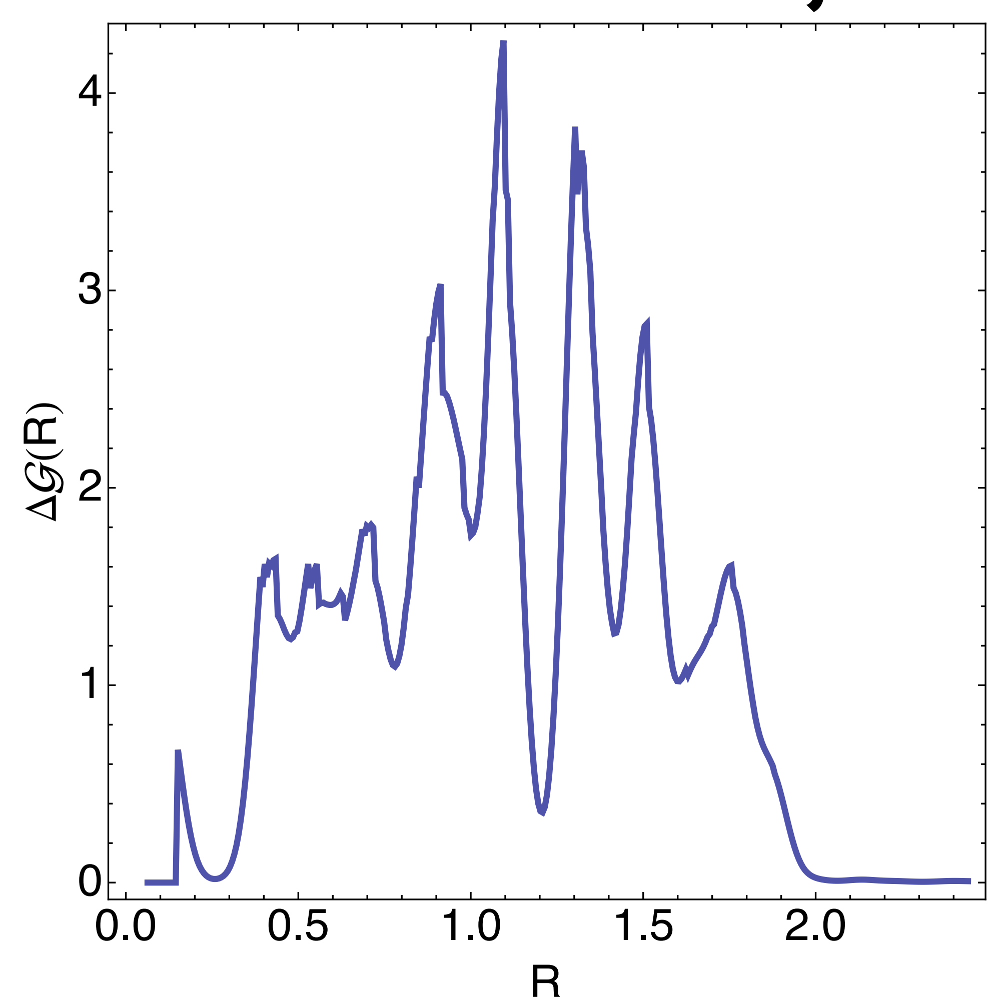
Jet Substructure Observables

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

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





Jet Substructure Observables

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

Application: Top Tagger

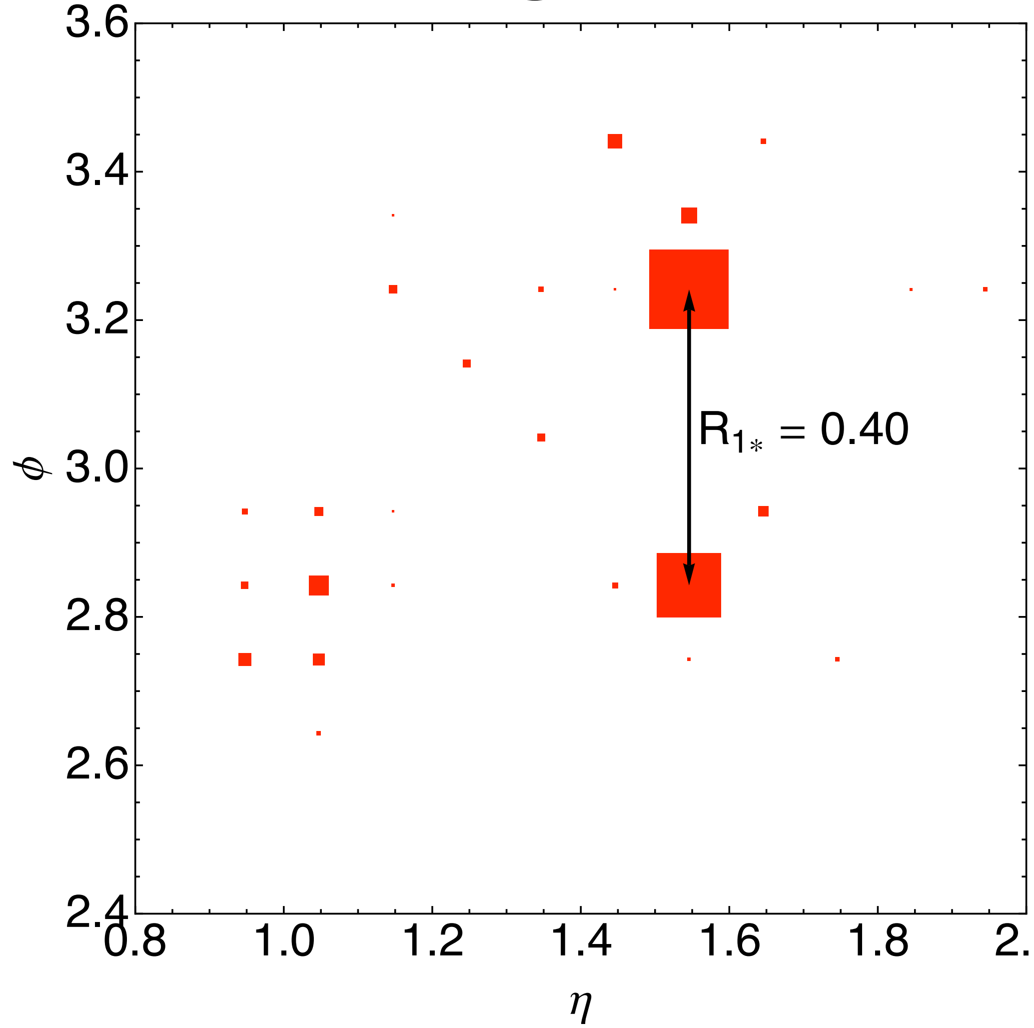
- Our Top Tagger:
  - Use observables from $\Delta\mathcal{G}(R)$
  - Expectations:
    - Top quark jets will have 3 subjets
    - Separation of subjets will be strongly correlated with m_W, m_t
    - Separation of subjets in QCD jet will be uncorrelated with any mass scale

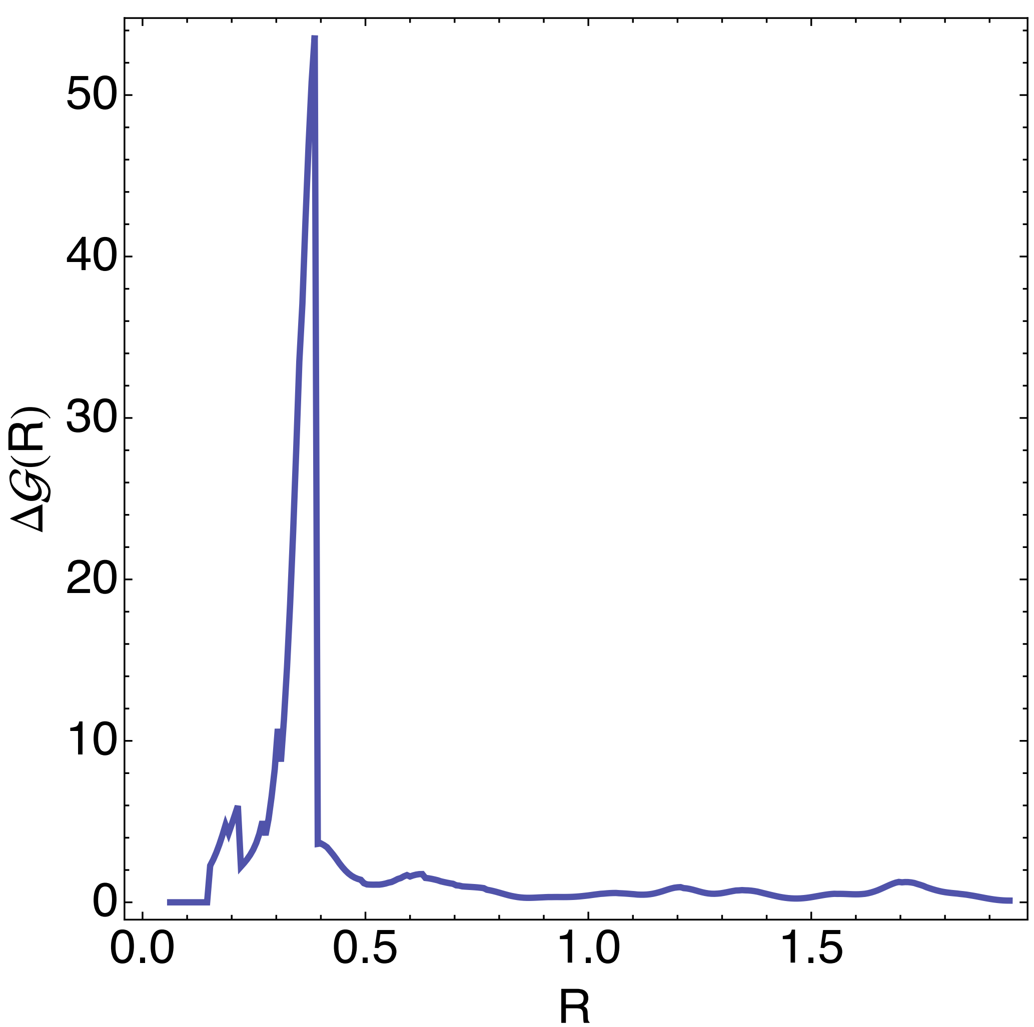
Application: Top Tagger

- 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

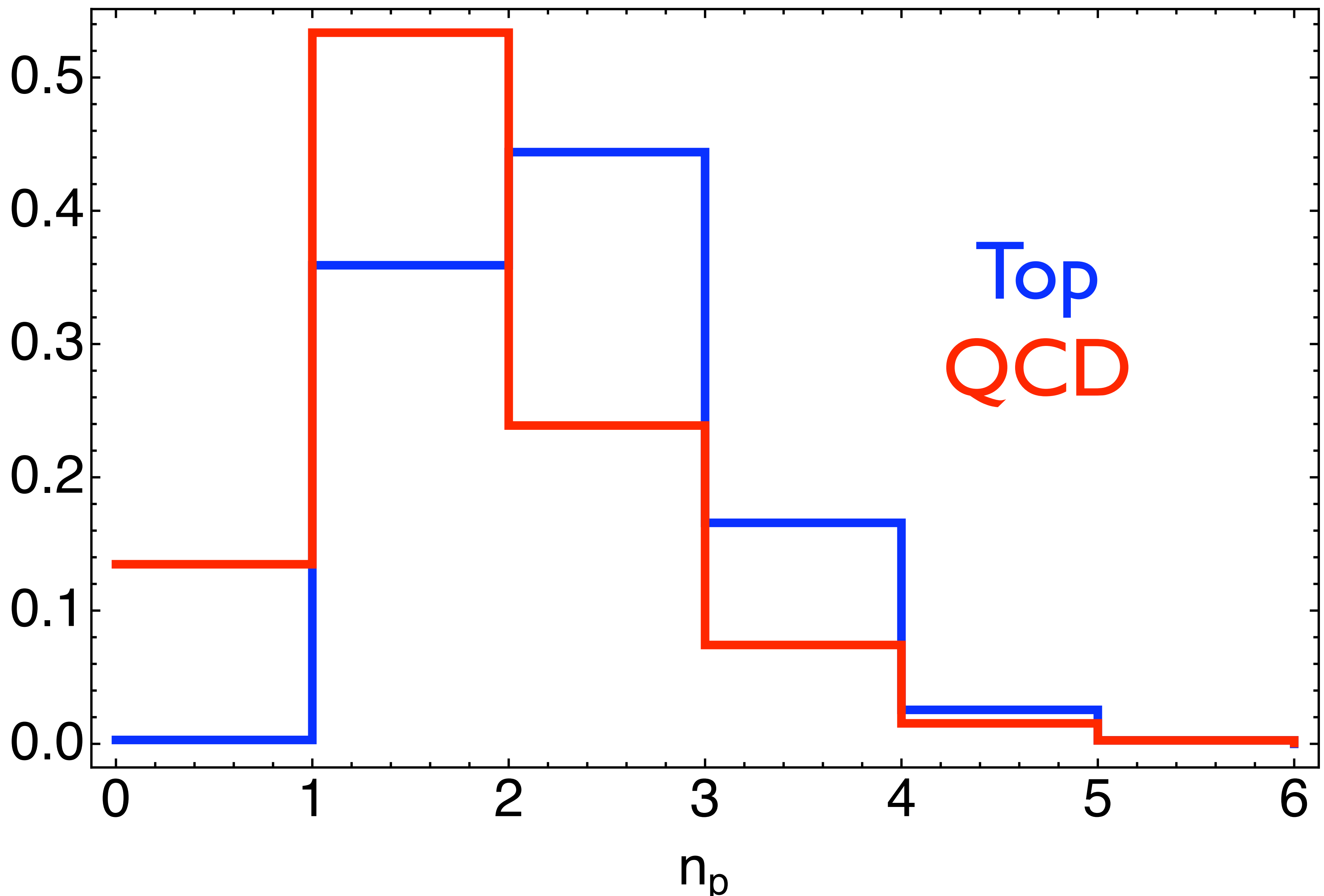
- More on the number of peaks:
  - npeak = 1 bin ➔ W decay products are unresolved or subjets form an equilateral triangle
  - npeak = 2 bin ➔ W decay products have large mass with b but not with each other or subjets form an isoceles triangle
  - npeak = 3 bin ➔ All subjets of top are resolved
  - Higher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

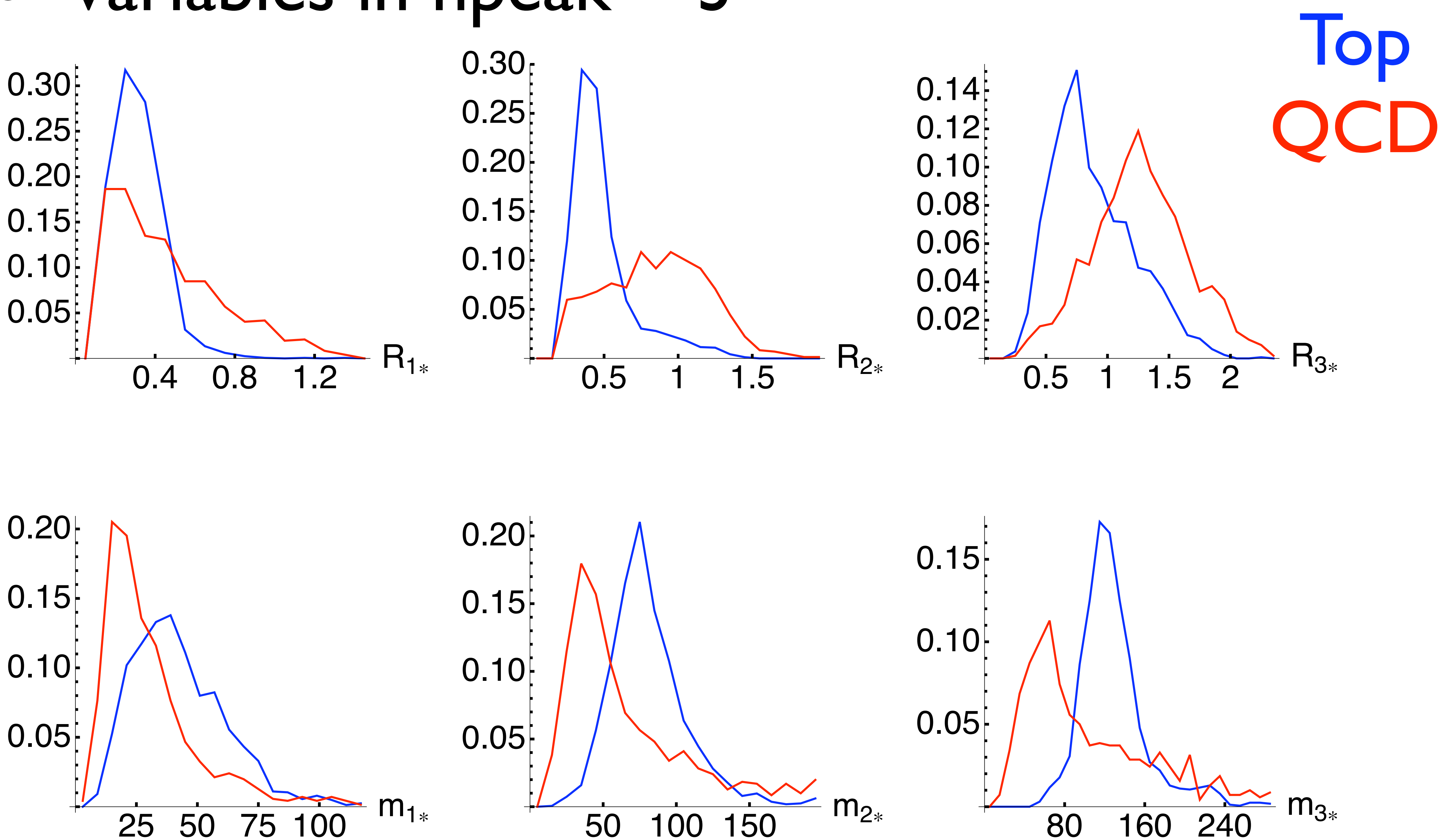
Application: Top Tagger

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

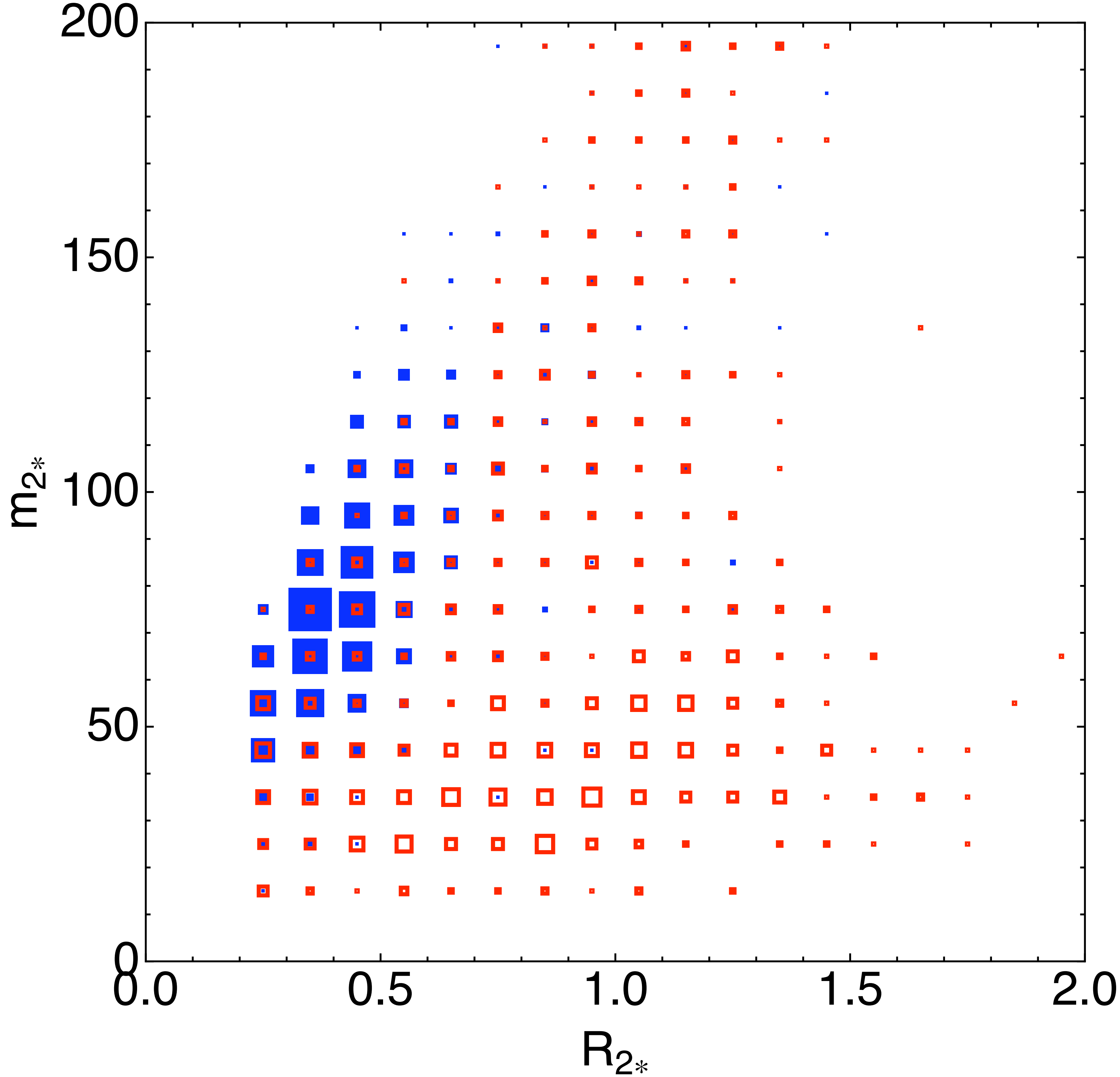


Application: Top Tagger

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



Application: Top Tagger



Top

QCD

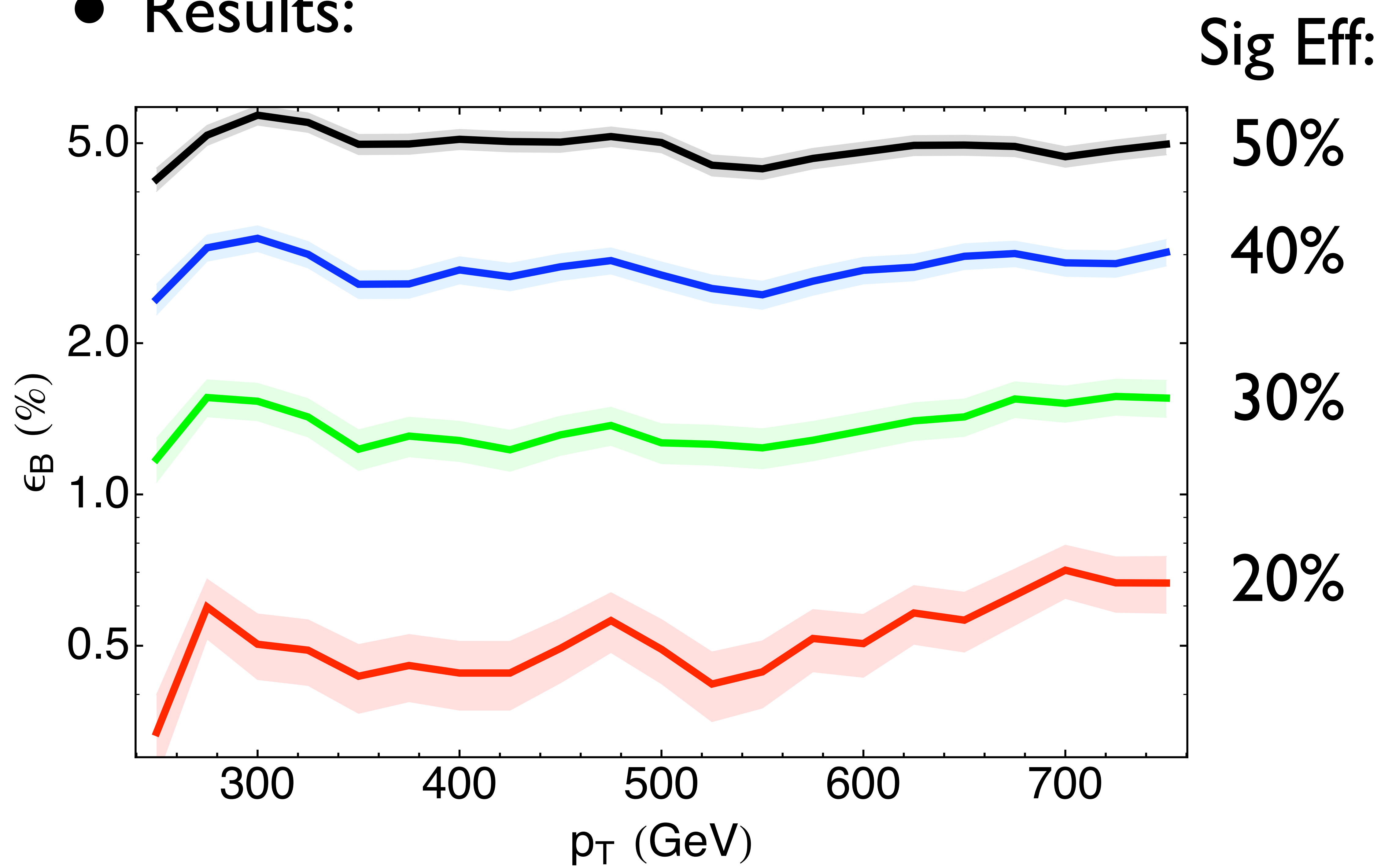
- Correlation of separation of subjets and their invariant mass
- Top: m ~ R
- QCD: m, R uncorrelated

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

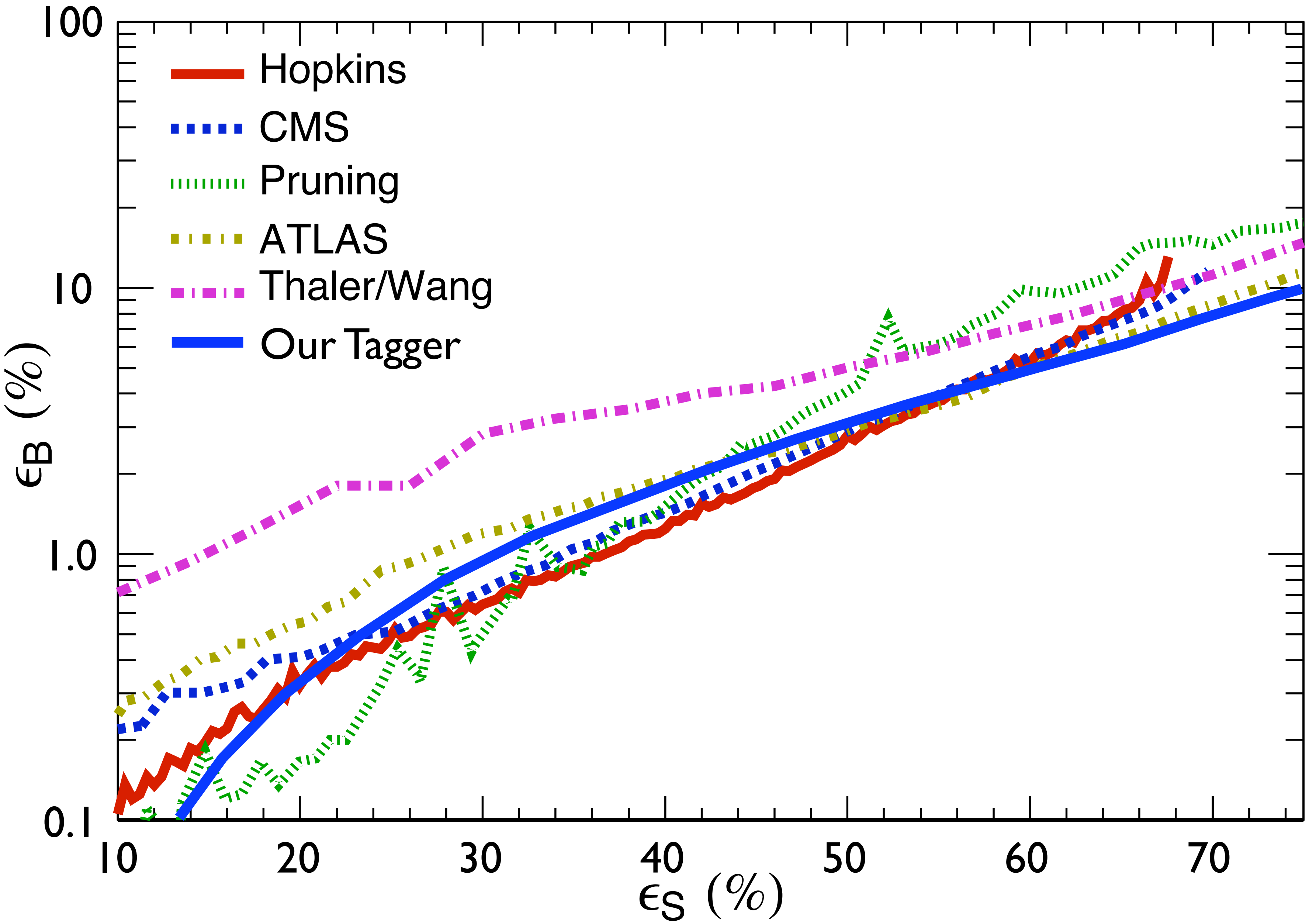
- Results:



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?