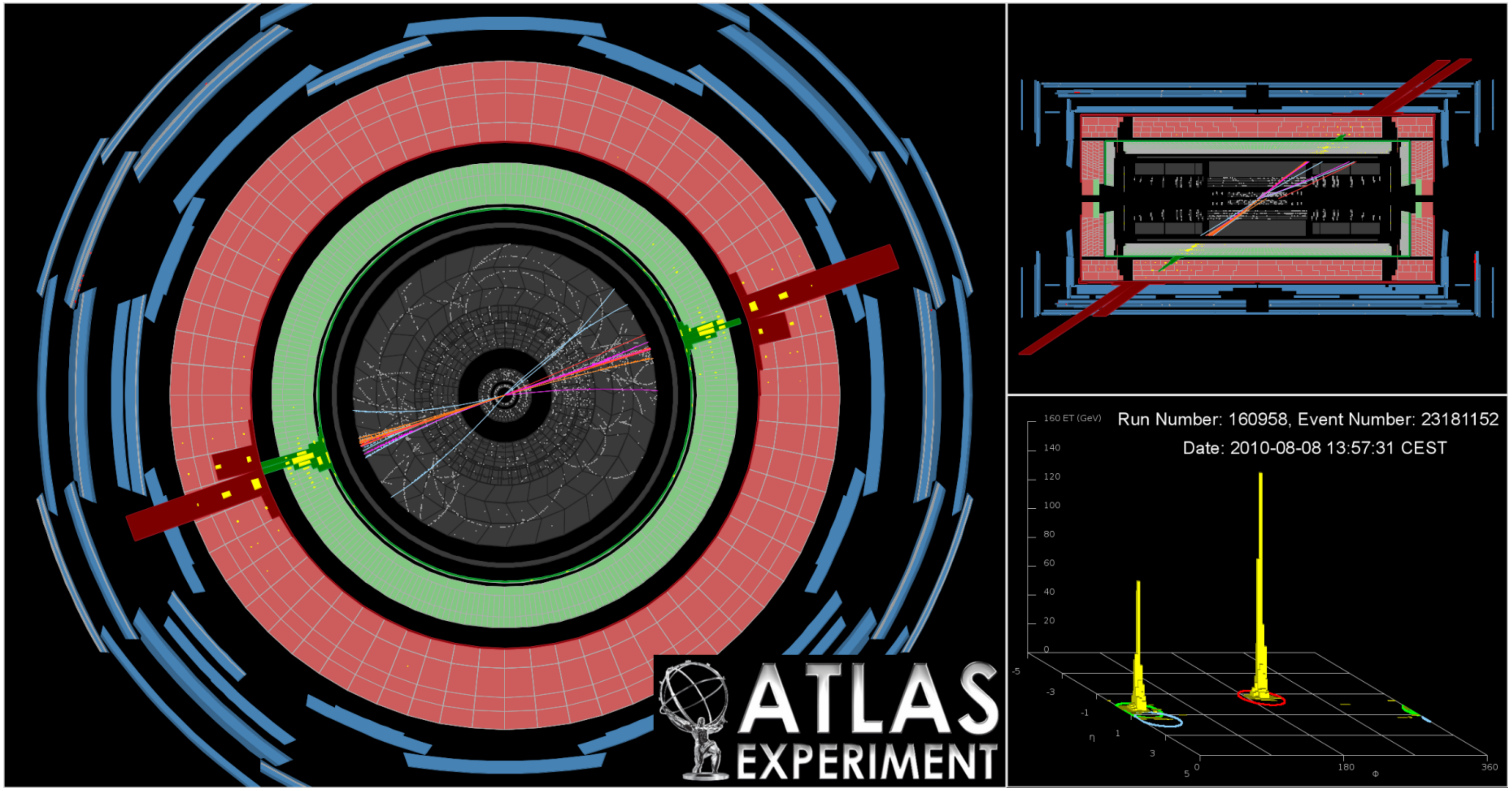
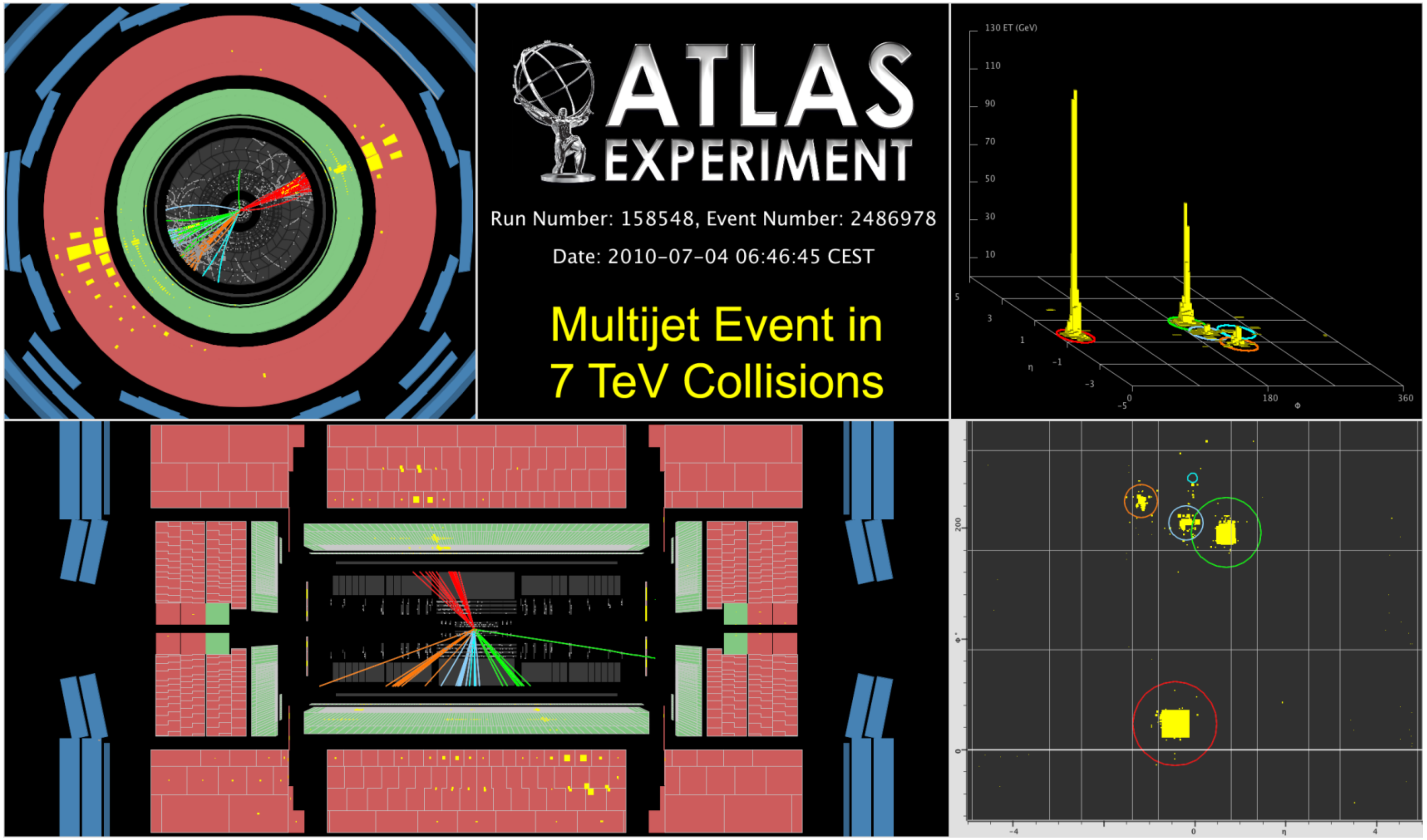
THE INSIDE OF A JET

Matthew Schwartz
Harvard University

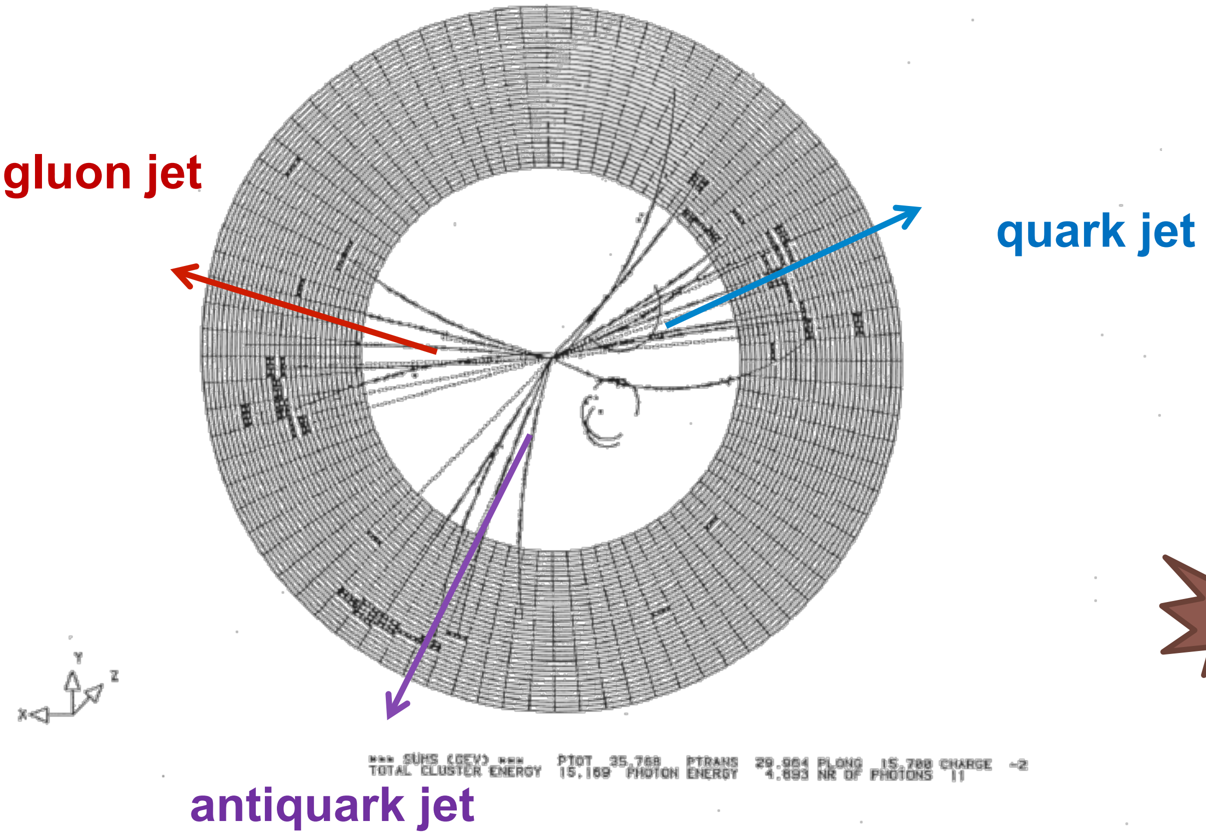
A dijet event from atlas



A multijet event

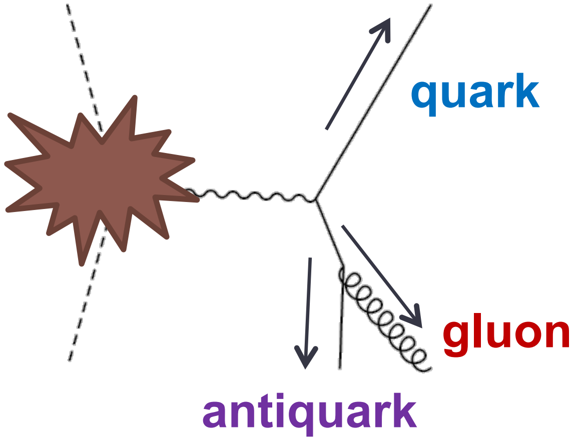


What is a jet?

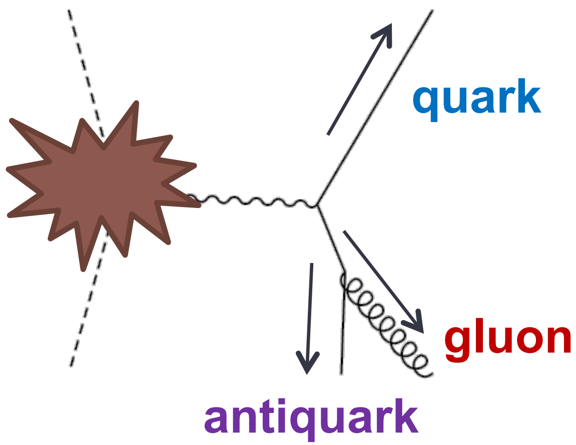


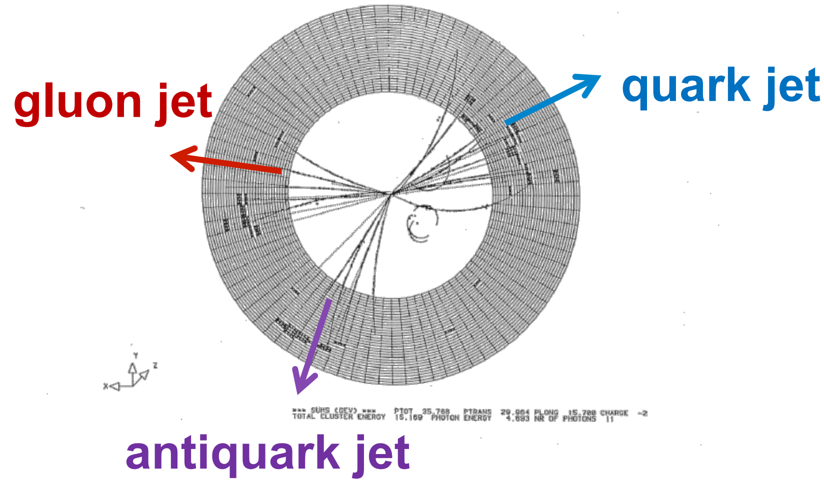
PETRA (DESY) 1979

First 3 jet event



What is a jet?





- Quarks and gluons (partons) are produced at **short distance**, where QCD is weak
- Partons radiate and fragment into stable hadrons at long distance
- No interference between **short** and long distances

<u>Factorization</u>

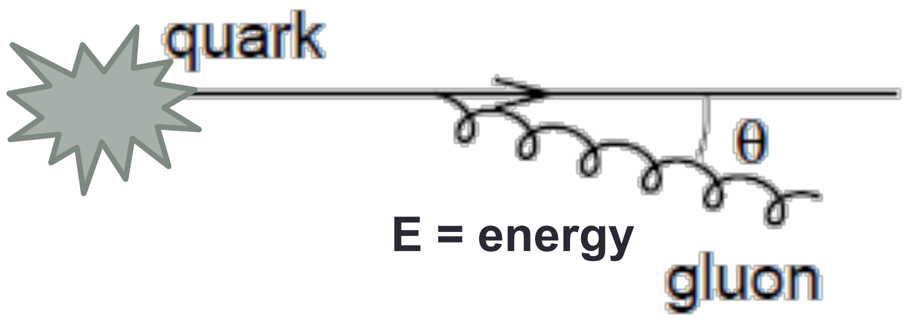
$d\sigma$ = [production] x [hadronization]

short distance ↖ long distance ↗

Can calculate separately

Short distance physics (e.g. supersymmetry) imprinted on **jets**!

What is a jet?



Cross section for producing an a gluon is dominated by small angles

$$d\sigma \approx \alpha_s \frac{dE}{E} \frac{d\theta}{\theta}$$

Small E
→ Soft divergence

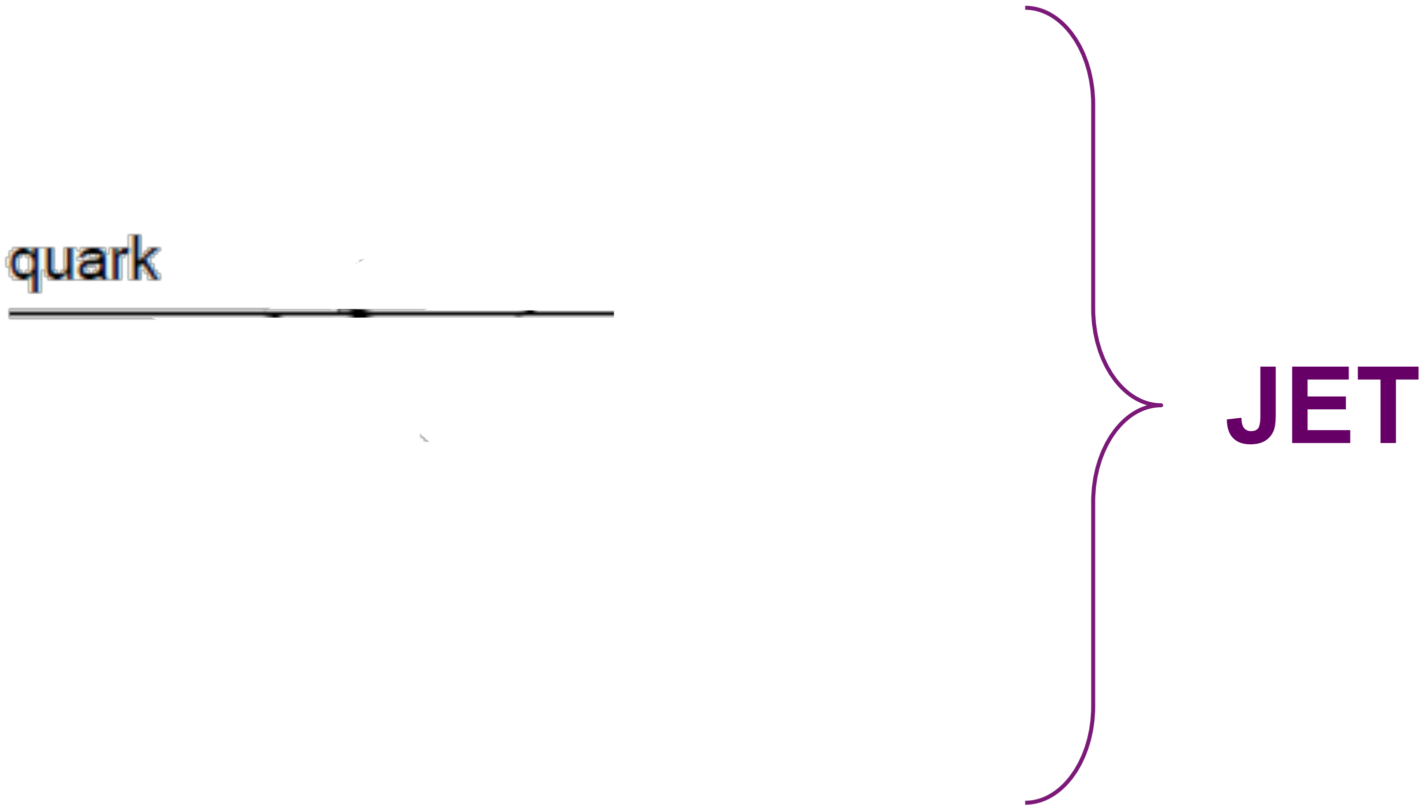
Small angle
→ collinear divergence

Interference is subleading
in the collinear limit
↓
(semi-classical) **probability**
for emission

Parton shower

Jet production is well-described by **semi-classical parton shower** picture

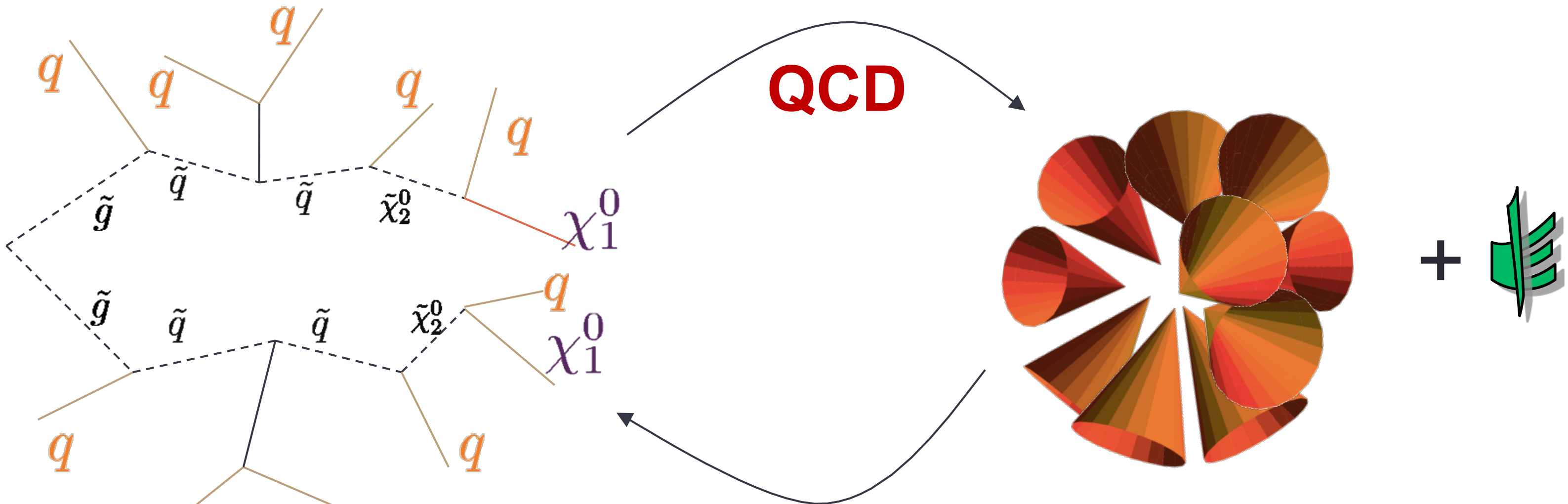
- Quark starts out off-shell at short distances ~ 10^{-3} fm
- As it moves out, has **probability for emission**
- When it gets ~1 fm away, shower stops and hadronization occurs



Jet to Parton map

Short distance process produces **quarks**

We observe **jets**:



Can we **invert** ?

How do we study jets

Jets are not well defined. Many different jet algorithms

- Cone algorithms
- Cluster algorithms
- Global algorithms (event shapes)
- …

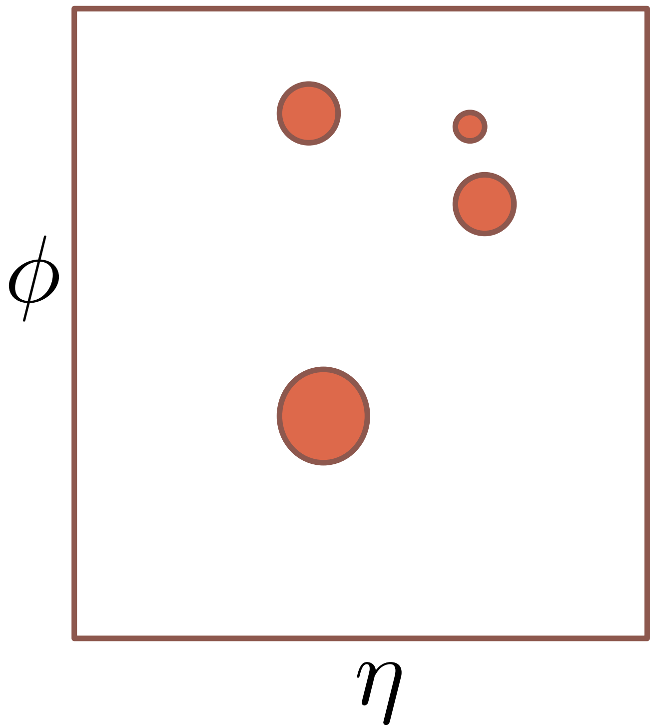
Radiation is assigned to jets differently

Iterative jet algorithms

- Start with input 4-vectors
  - e.g. stable particles, topoclusters, calorimiter cells, etc.

- **Calculate** the pairwise distance

$$R_{ij} = \sqrt{(\theta_i - \theta_j)^2 + (\eta_i - \eta_j)^2}$$



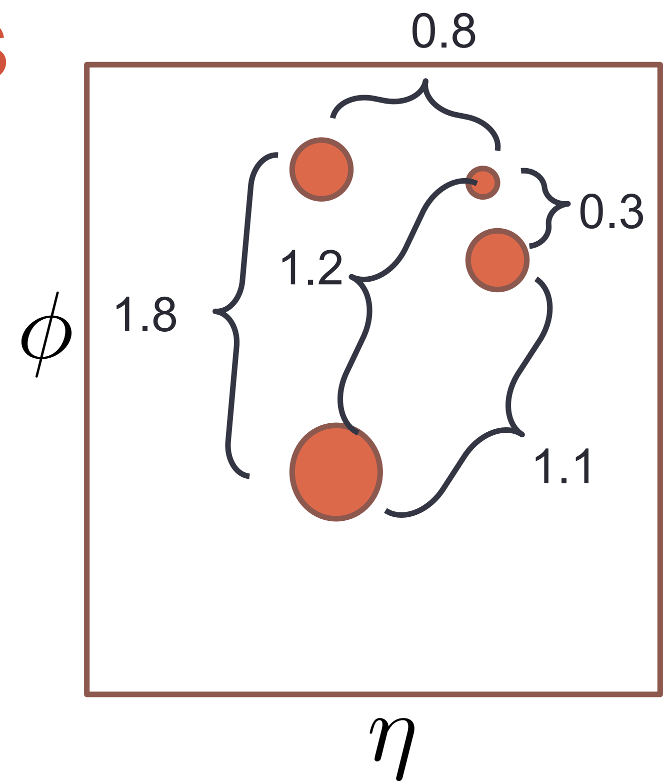
Iterative jet algorithms

- Start with input 4-vectors
  - e.g. stable particles, topoclusters, calorimiter cells, etc.

- **Calculate** the pairwise distances

$$R_{ij} = \sqrt{(\theta_i - \theta_j)^2 + (\eta_i - \eta_j)^2}$$

- **Merge** the two closest particles



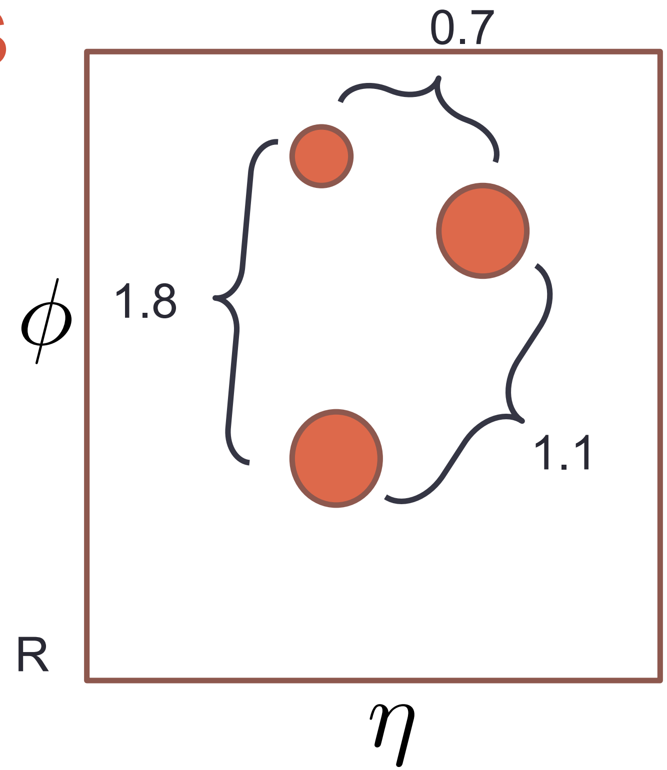
Iterative jet algorithms

- Start with input 4-vectors
    - e.g. stable particles, topoclusters, calorimiter cells, etc.

- **Calculate** the pairwise distance

$$R_{ij} = \sqrt{(\theta_i - \theta_j)^2 + (\eta_i - \eta_j)^2}$$

- **Merge** the two closest particles
- **Repeat** until no two particles are closer than R



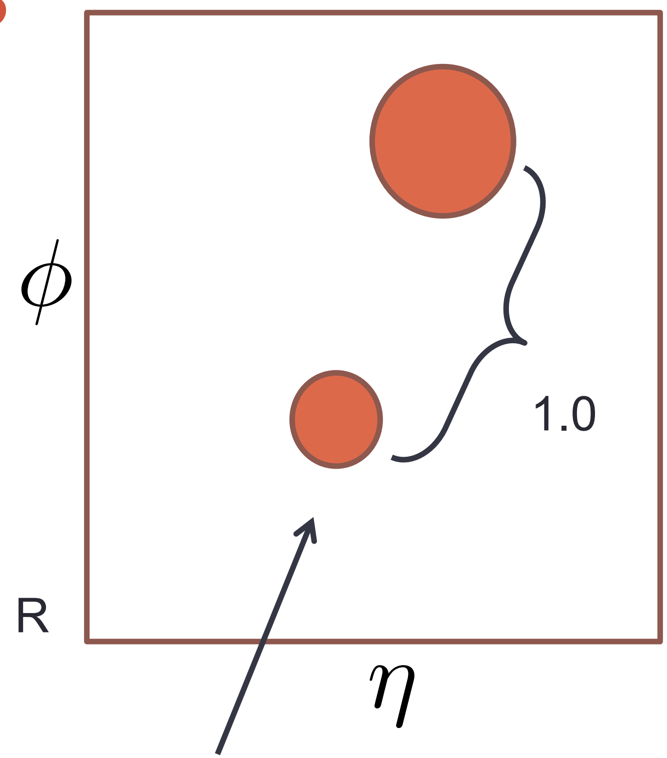
Iterative jet algorithms

- Start with input 4-vectors
    - e.g. stable particles, topoclusters, calorimiter cells, etc.

- **Calculate** the pairwise distance

$$R_{ij} = \sqrt{(\theta_i - \theta_j)^2 + (\eta_i - \eta_j)^2}$$

- **Merge** the two closest particles
- **Repeat** until no two particles are closer than R



Two R=1.0 Jets

Different distance measures

- k_T algorithm

$$d_{ij} = \min(p_{Ti}^2, p_{Tj}^2)\left(\frac{\Delta R}{R_0}\right)^2,\ d_{iB} = p_{Ti}^2$$

- C/A algorithm

$$d_{ij} = \left(\frac{\Delta R}{R_0}\right)^2,\ d_{iB} = 1$$

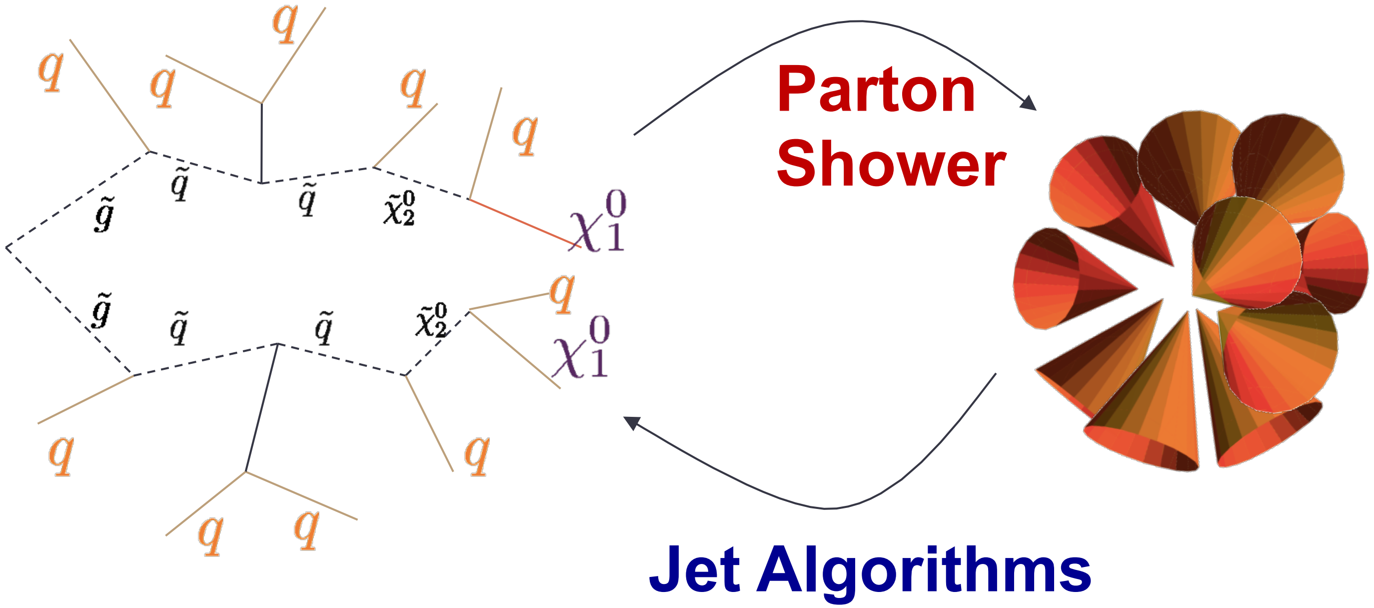
- anti-k_T algorithm

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2})\left(\frac{\Delta R}{R_0}\right)^2,\ d_{iB} = p_{Ti}^{-2}$$

Can be thought of as **inversions** of different **parton showers**

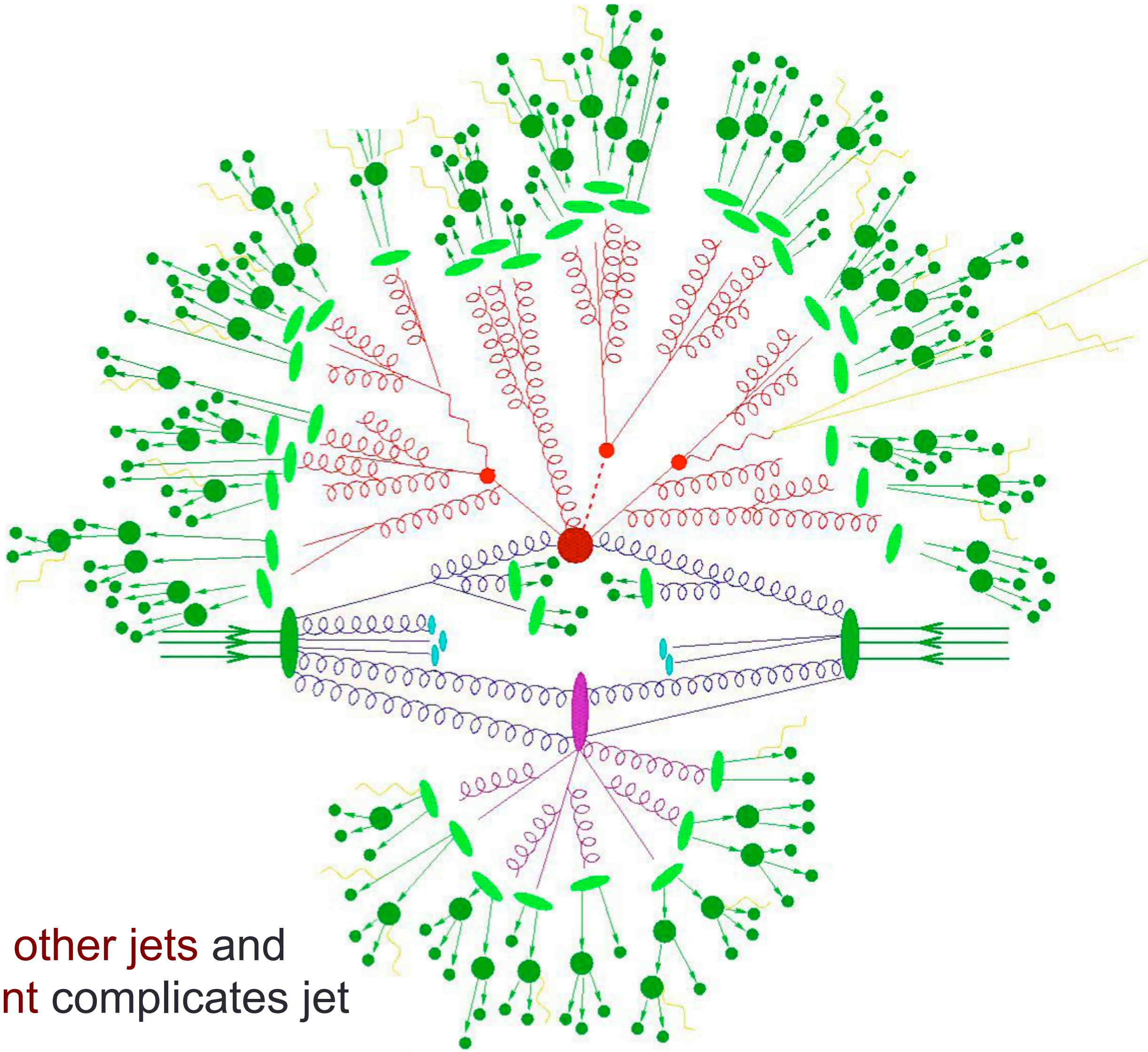
Summary so far

- Existence of **jets** comes from **collinear singularities** in QCD
- In collinear limit, interference is unimportant
  and **semi-classical picture** applies
- **Parton showers** use a semi-classical Markov process to **simulate QCD**
- **Jet algorithms** attempt to invert the **parton-to-jet mapping**



- This is a great first approximation,
  but **reality** is **much more interesting**

Real events are complicated!!



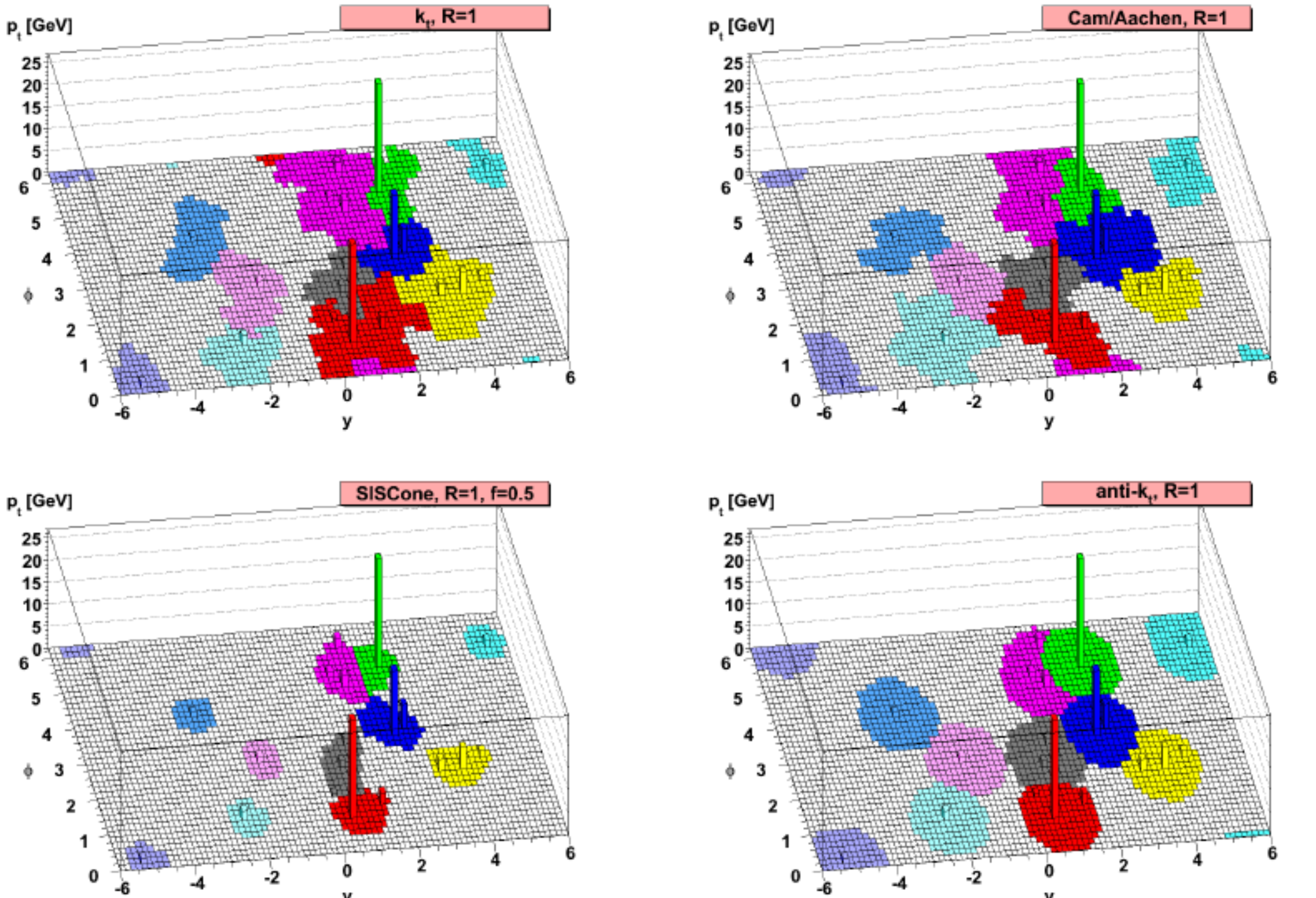
Radiation from other jets and underlying event complicates jet reconstruction

Image taken from R. Teuscher

Beyond the first approximation

- Contamination from underlying event/other jets
  - **Parton-shower** is ***not* invertible**
- Jet = parton worked great at LEP and Tevatron
  - At LHC, detectors are so good, we can **look *inside* jets**
- **Interference** is sometimes important
  - Critical for measuring **Color correlations** of jets
- Last few years have seen many qualitatively new ways of thinking about jets

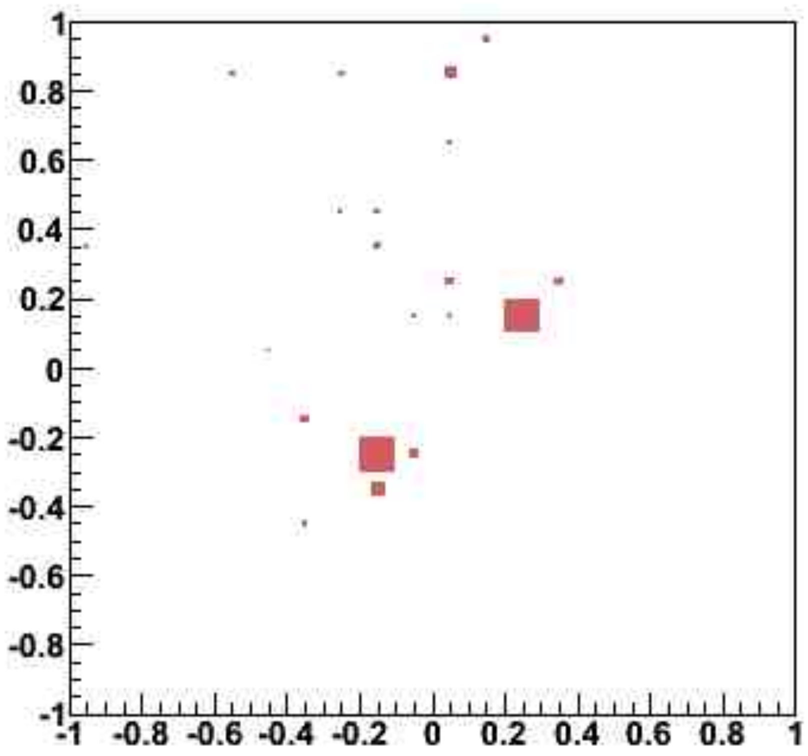
Different algorithms, different results

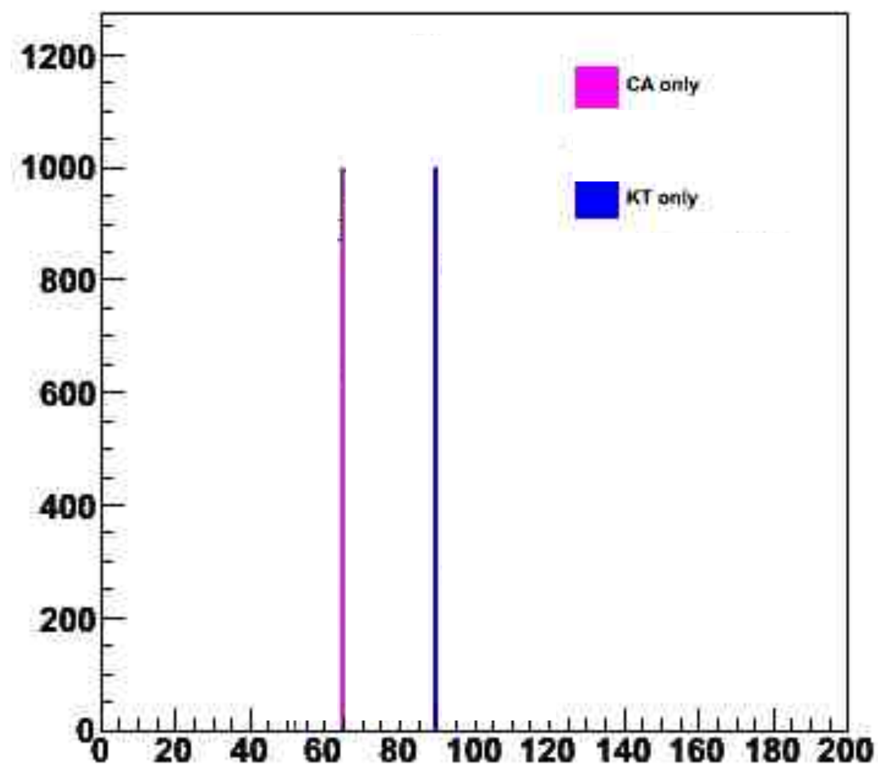


Cacciari, Salam, Soyez JHEP 0804:063 (2008)

e.g. reconstruct W invariant mass

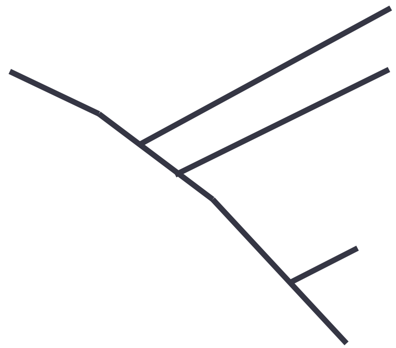
$$W \to \bar{q}q$$

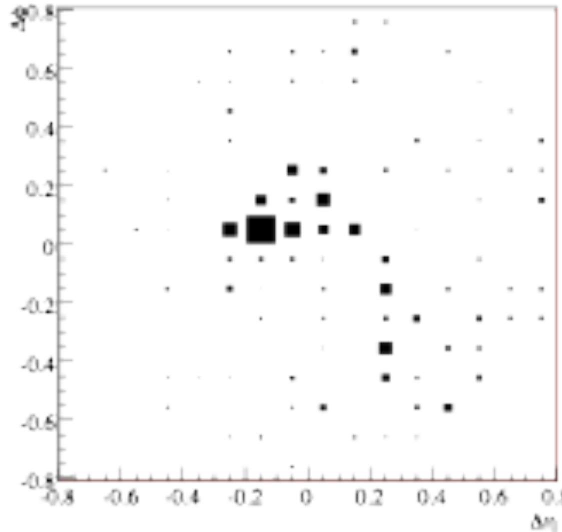




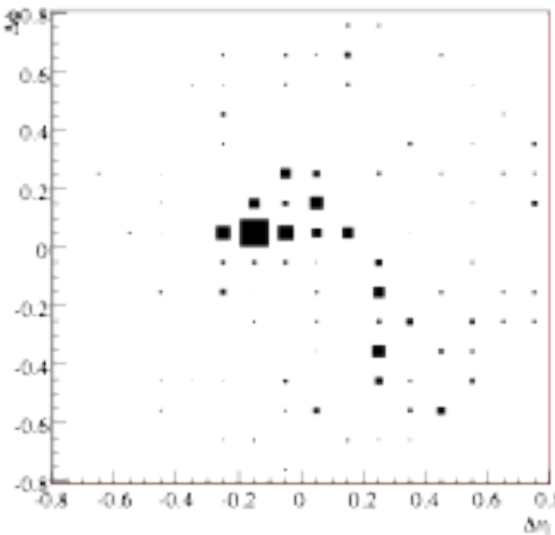
Parton shower is not invertible

Parton shower gives an event

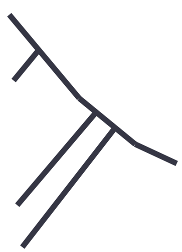




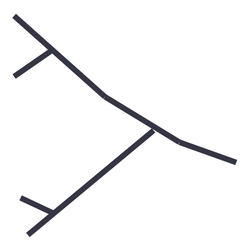
What is the **inverse**?



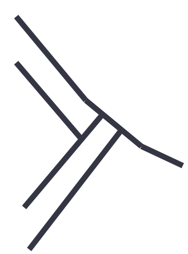
=



or



or



?

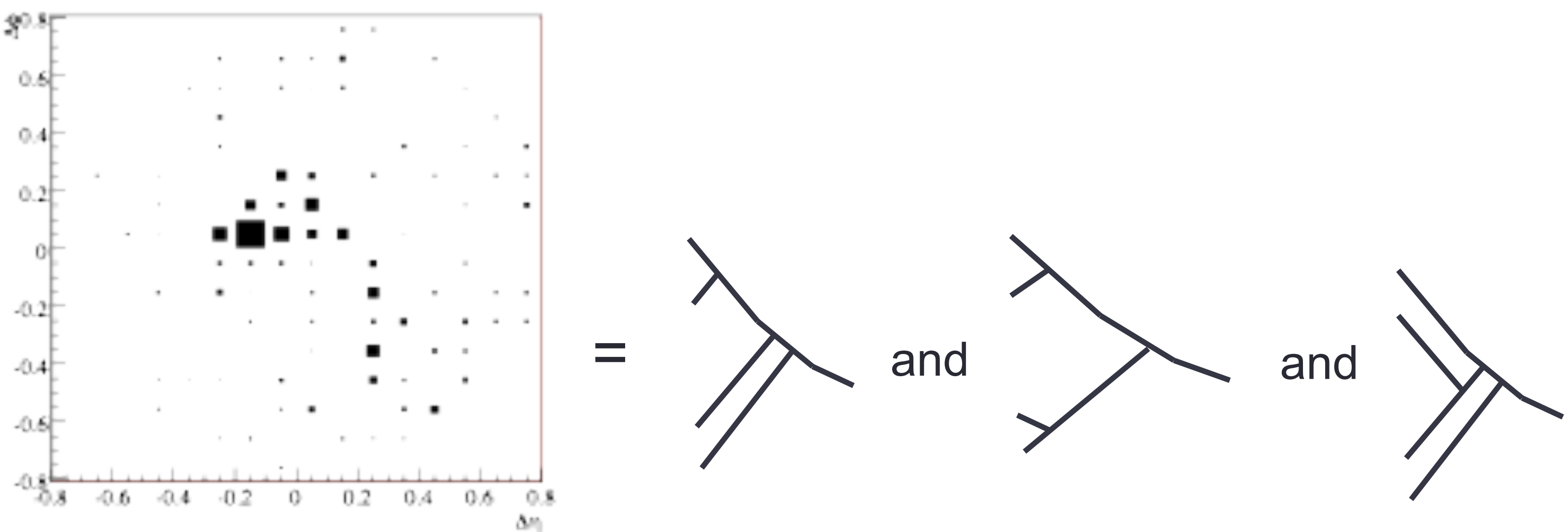
All of them! -- Qjets

Ellis, Horning, Krohn, Roy and MDS
PRL (2012) to appear

- **Add randomness** into the jet algorithm
  - Instead of choosing smallest d_{ij}, choose pair with a probability

$$P \propto \exp(-\alpha d_{ij})$$

- Generates **ensemble of trees** for **each event**



What did we do with the Qjets?

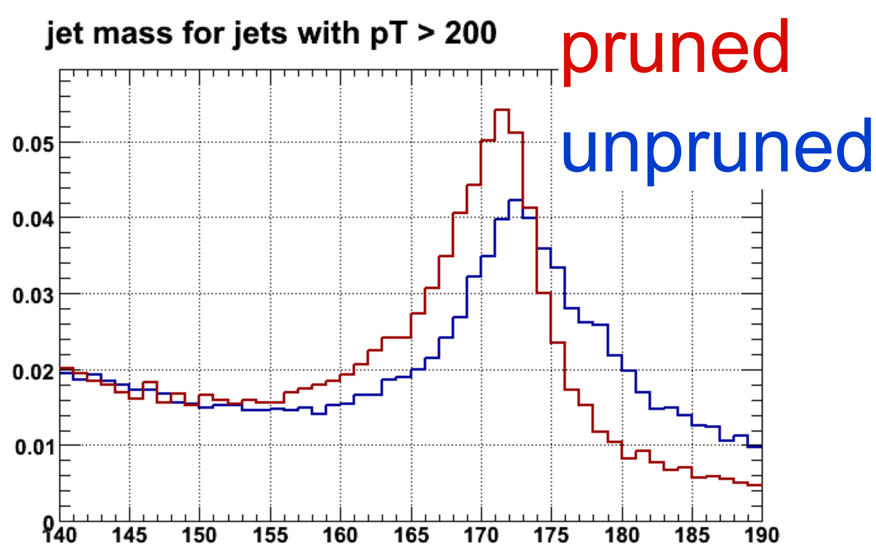
As an example, we can **prune** them

Ellis, Vermilion, Walsh Phys.Rev. D80 (2009)

- Pruning **discards** radiation in clustering that is **soft but not collinear**

$$z_{ij} \equiv \frac{\min\left(p_{T_i}, p_{T_j}\right)}{\left|\vec{p_{T_i}} + \vec{p_{T_j}}\right|} < z_{\rm cut}$$

$$\Delta R_{ij} > D_{\rm cut}$$

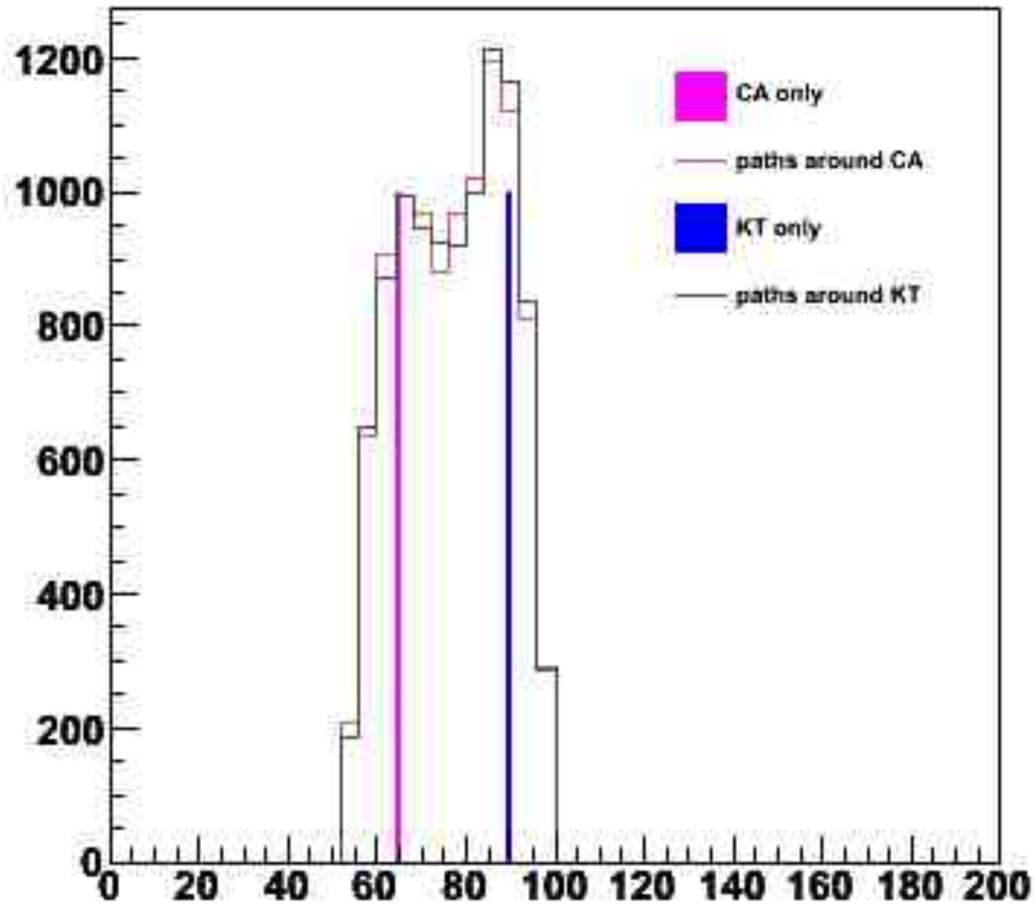


Other variants **filtering** or **trimming** work similarly

Butterworth, Cox, Forshaw Phys.Rev. D65 (2002)

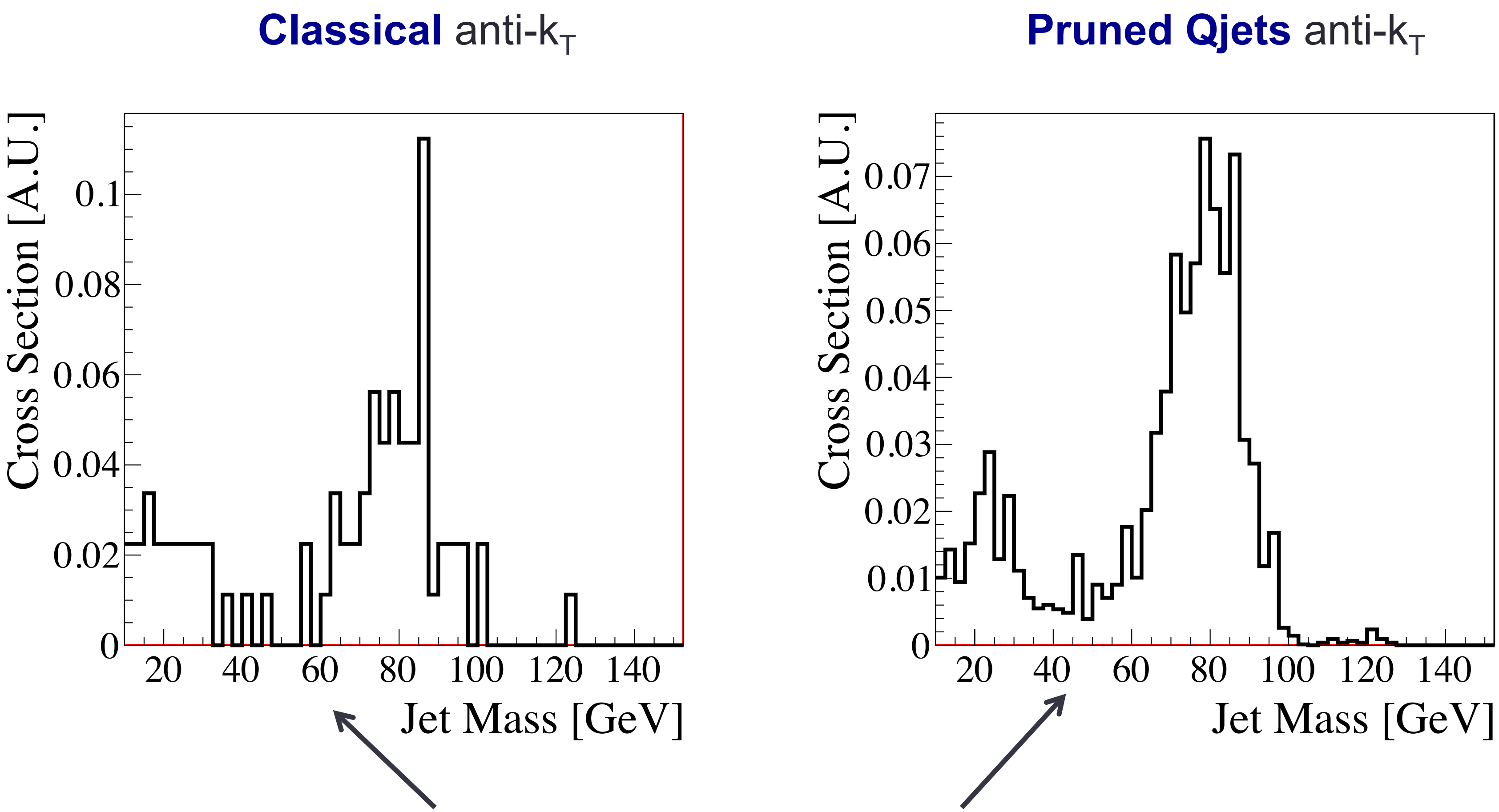
Krohn, Thaler, Wang JHEP 1002 (2010)

W jet mass with pruned Qjets



This is ***one*** event

Distributions become much smoother



The same 100 events

Need fewer events for same precision

For example,

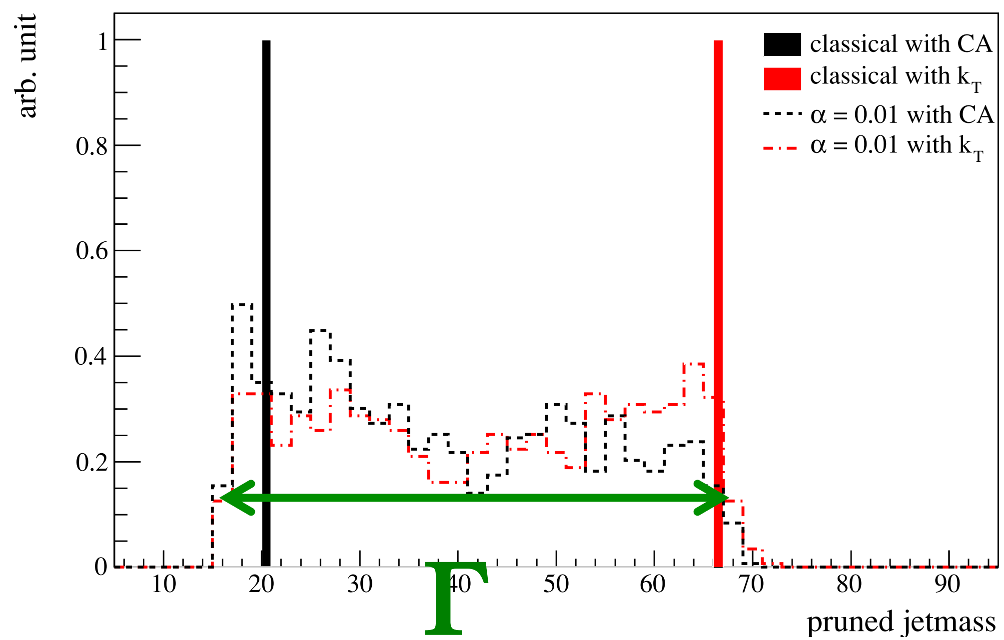
- Take 10 boosted W events (p_T>500)
- Construct jet mass
- Look at **variance** of the the **mean** W-jet mass over many pseudo-experiments

Algorithm	Mass uncertainty $\delta\langle m\rangle$	Relative Luminosity required
k_T	3.15 GeV	1.00
Qjets α=0	2.20 GeV	0.50
Qjets α=0.001	2.04 GeV	0.45

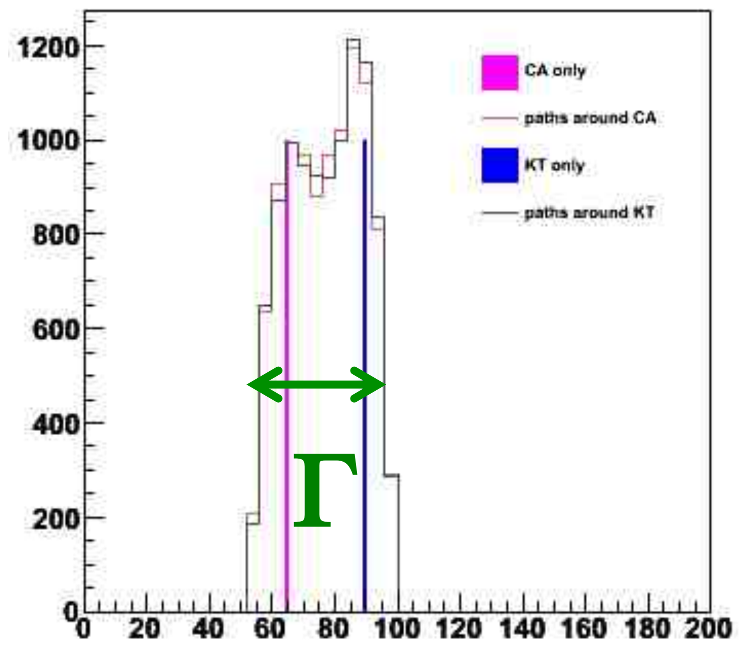
Qjets needs **half as much luminosity** as conventional jet algorithms

Signal vs background

QCD jets (one event)



W jets (one event)

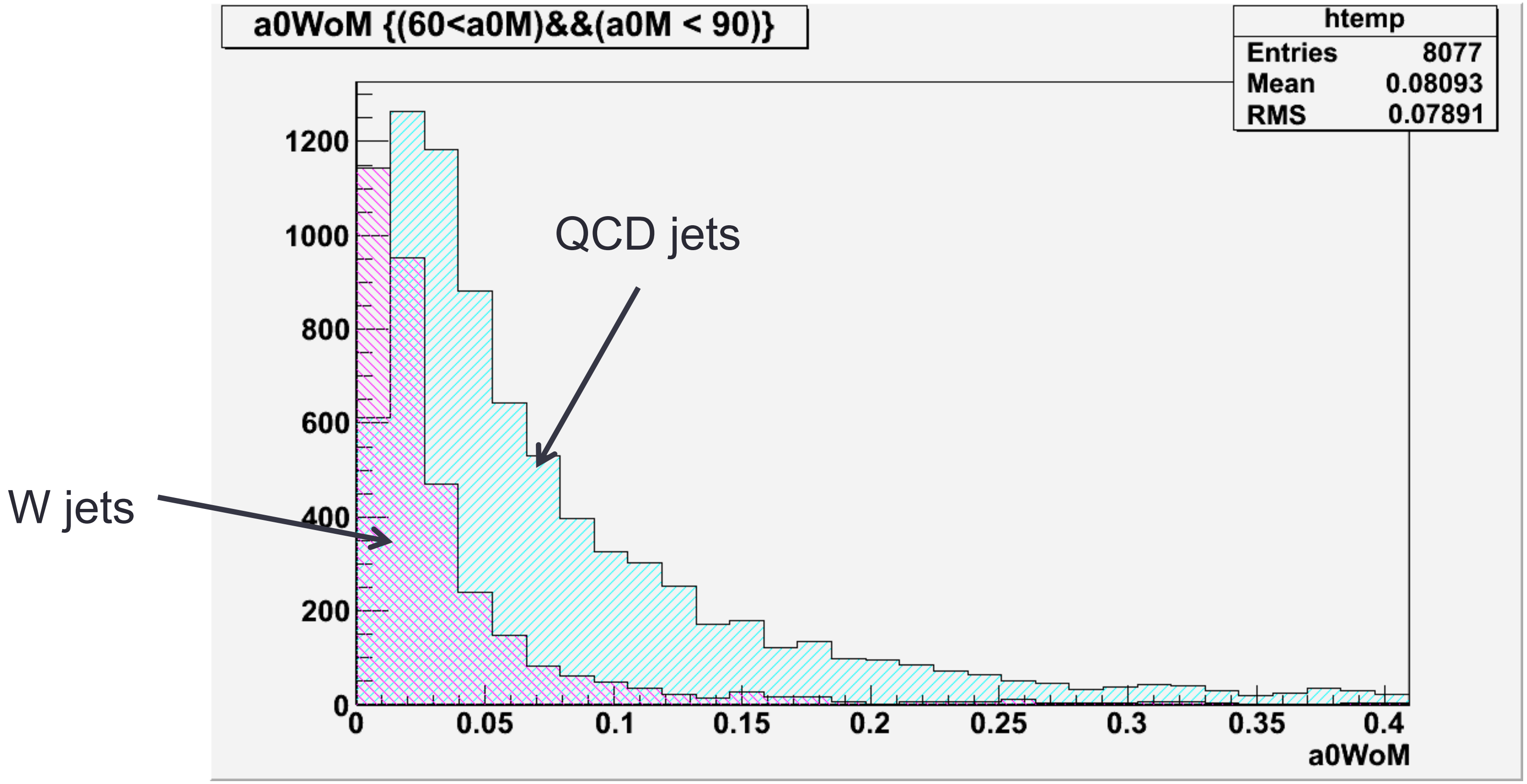


Volatility $\mathcal{V} = \frac{\Gamma}{\langle m \rangle}$ is a purely Q-observable

Volatility

$$\mathcal{V} = \frac{\Gamma}{\langle m \rangle}$$

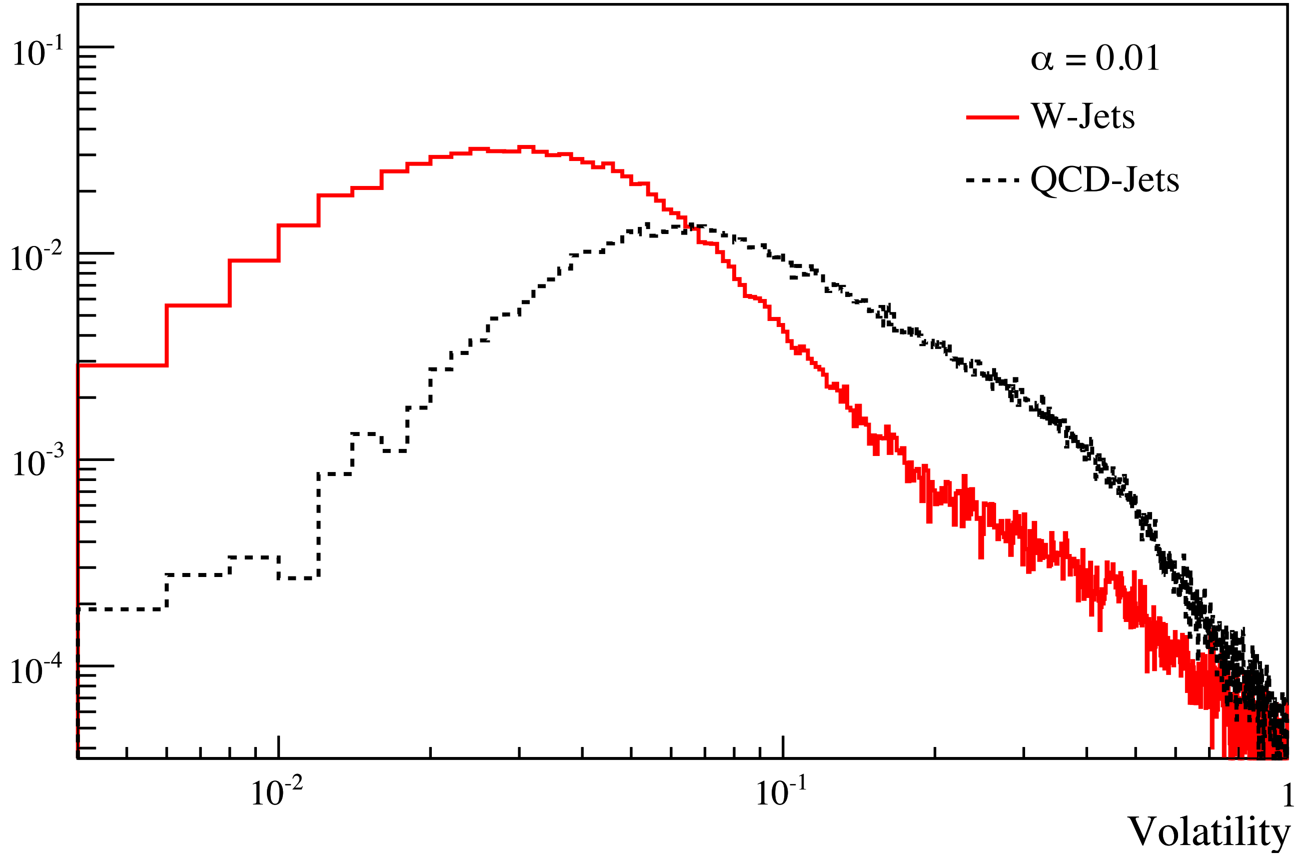
QCD jets are broader than boosted W jets



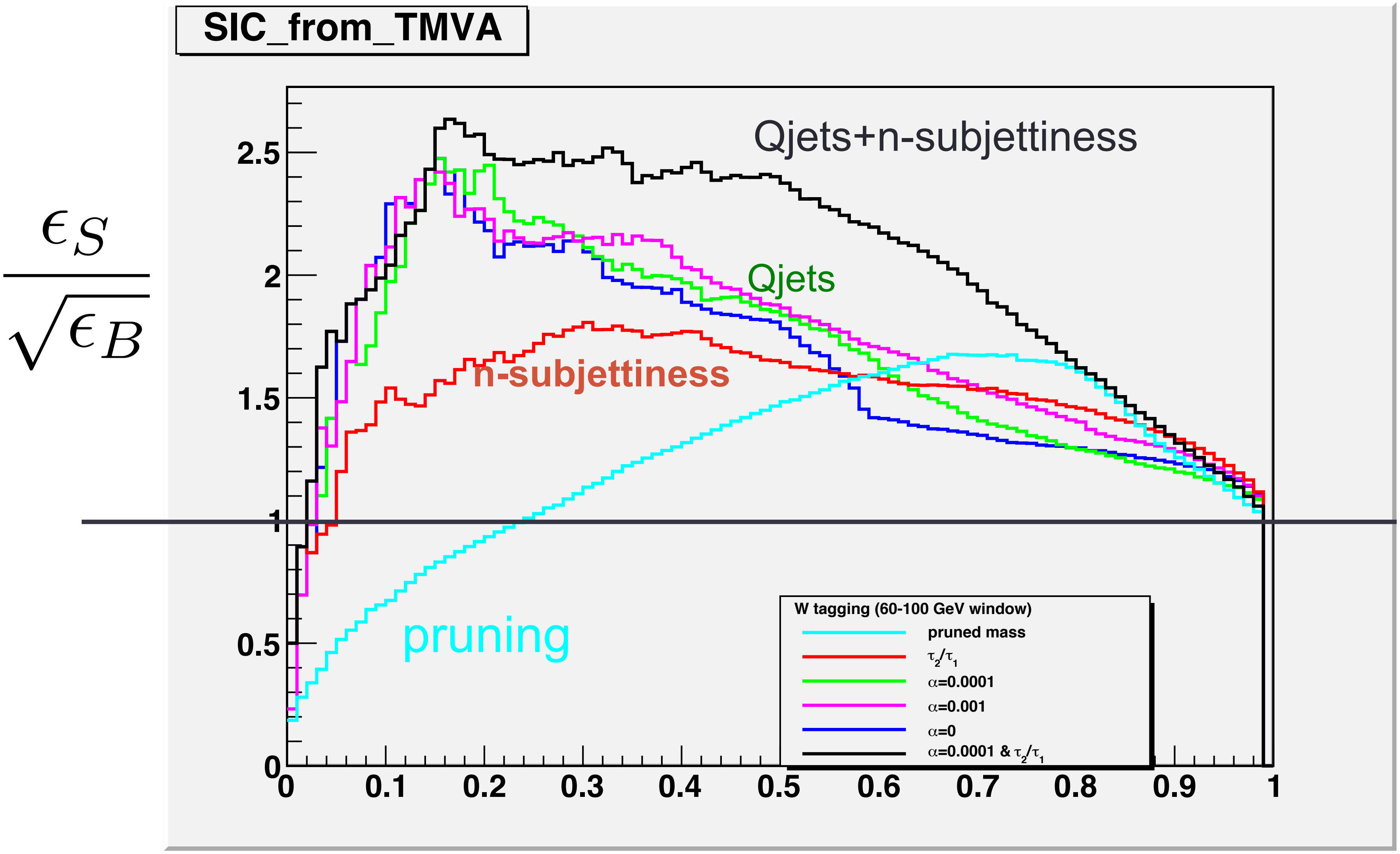
Volatility

$$\mathcal{V} = \frac{\Gamma}{\langle m \rangle}$$

QCD jets are broader than boosted W jets



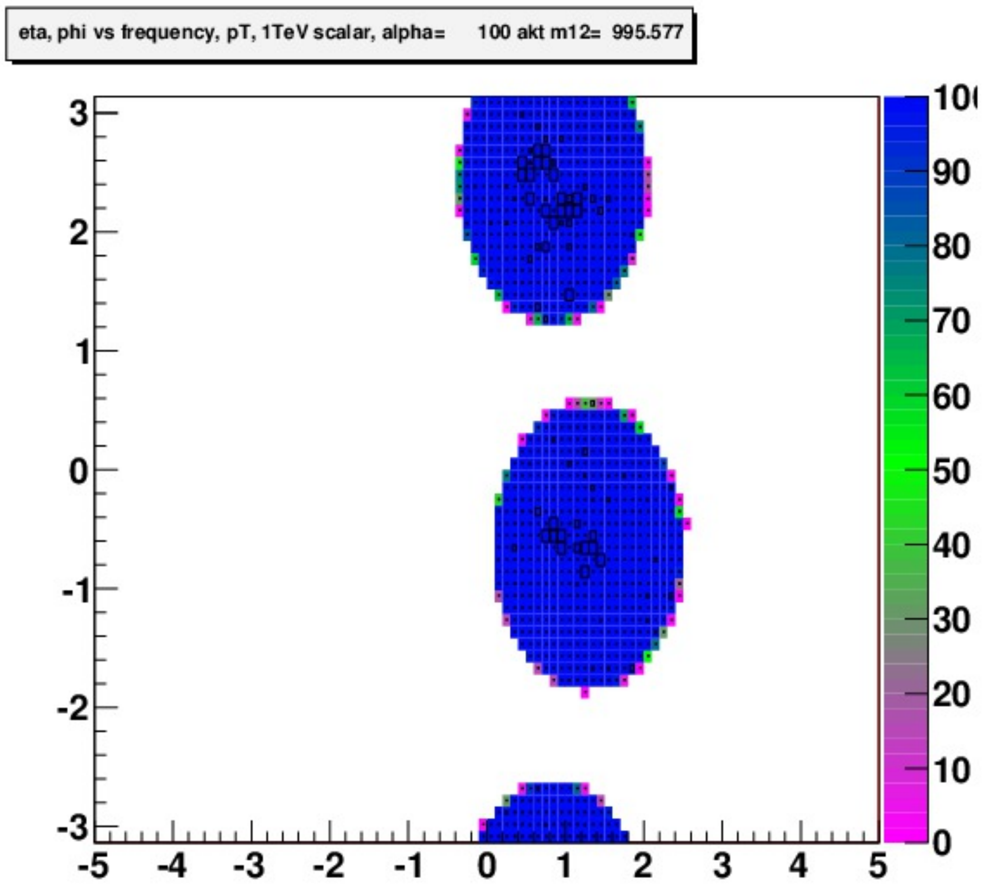
W-tagging: cut on volatility



Qjets on dijet events (no pruning)

α = 100

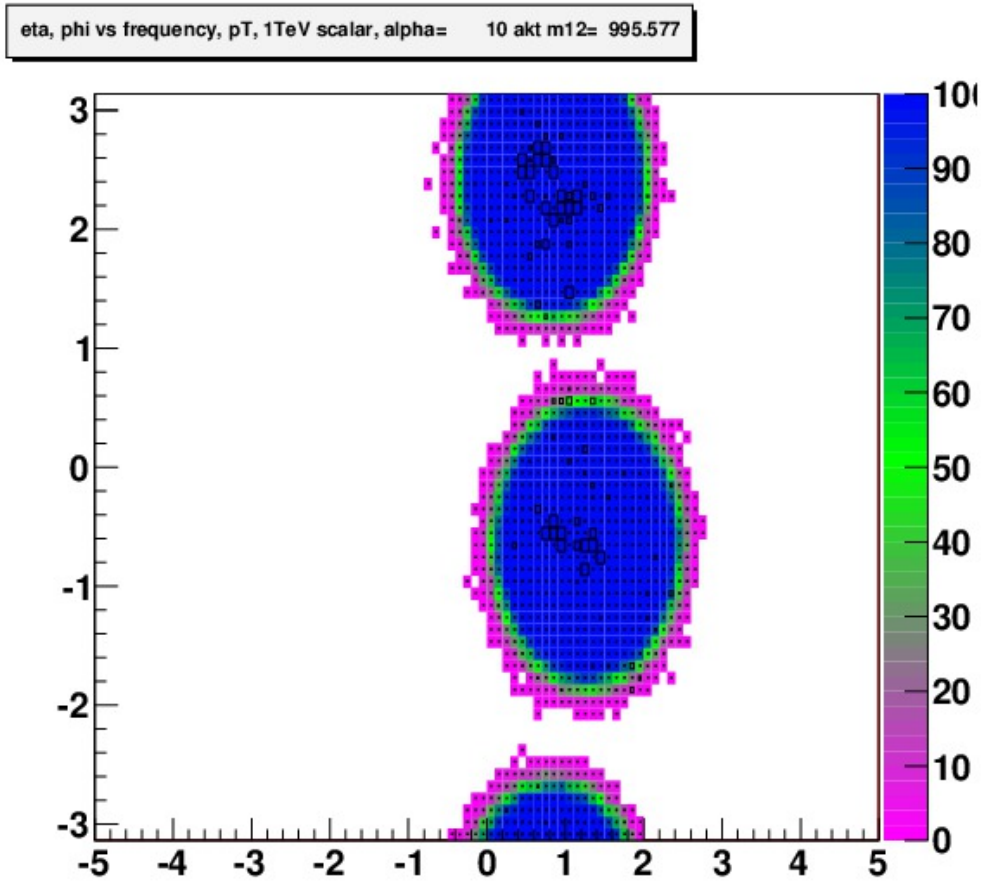
(classical anti-kT)



Work in progress, with D. Krohn and D. Kahawala

Qjets on dijet events (no pruning)

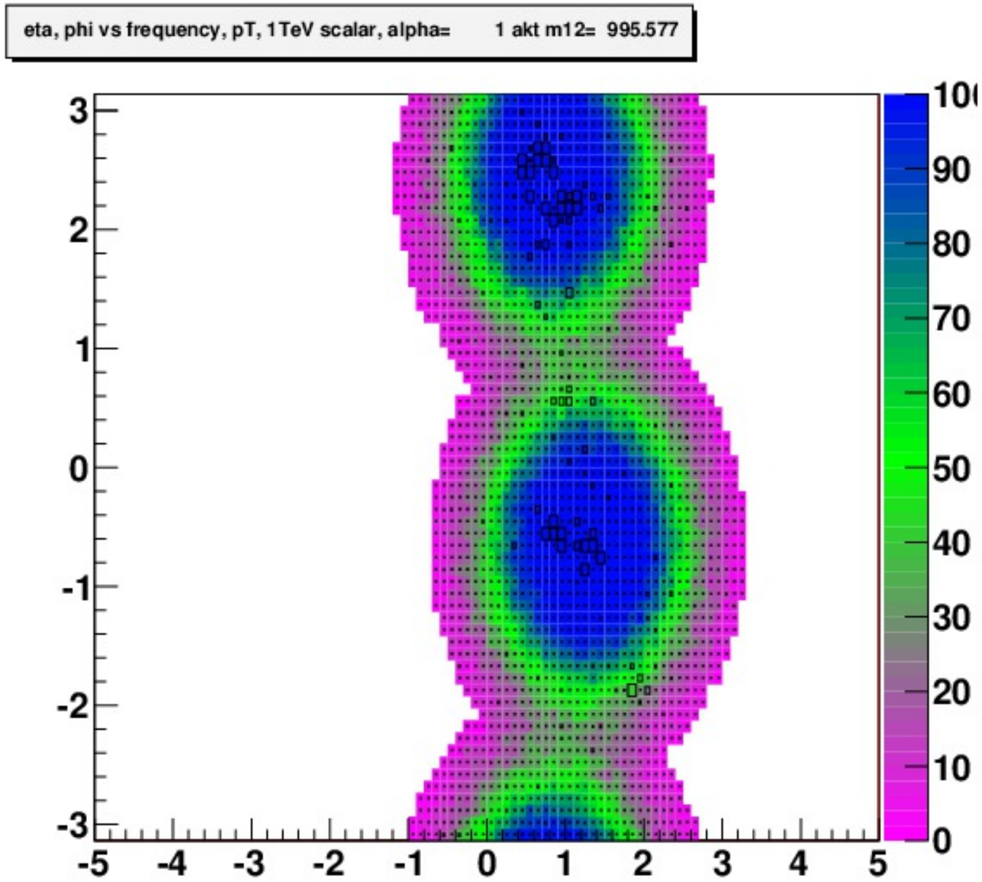
$\alpha = 10$



Work in progress, with D. Krohn and D. Kahawala

Qjets on dijet events (no pruning)

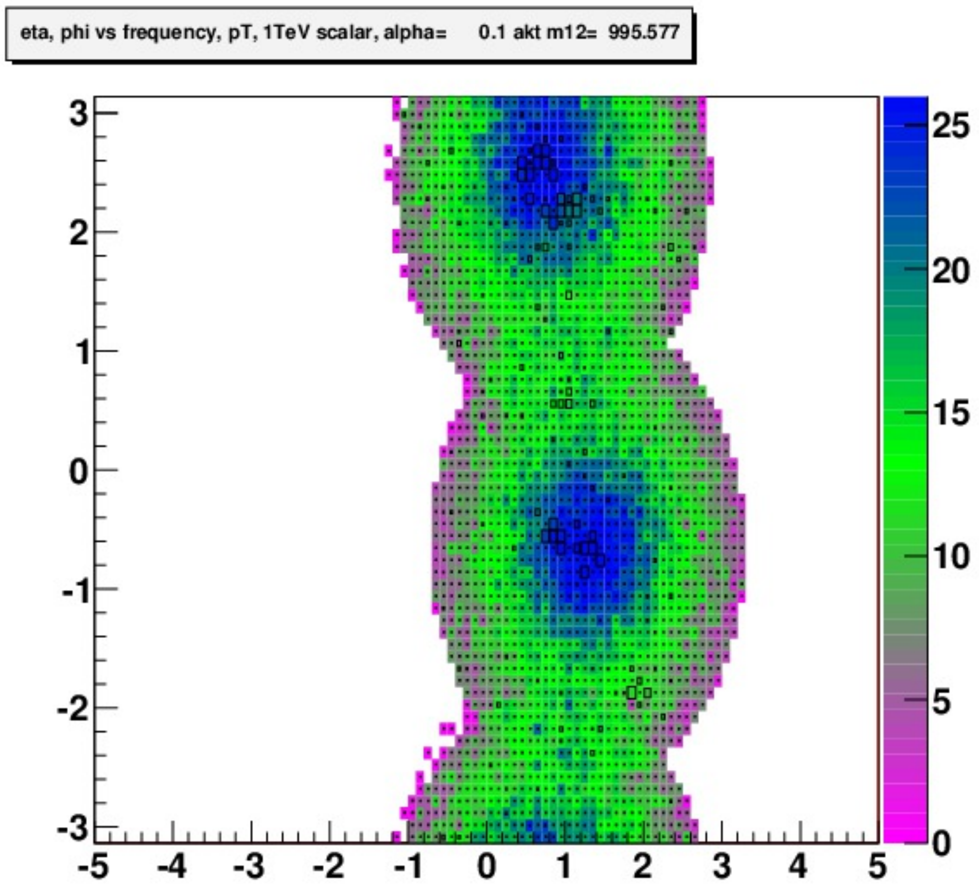
$\alpha = 1$



Work in progress, with D. Krohn and D. Kahawala

Qjets on dijet events (no pruning)

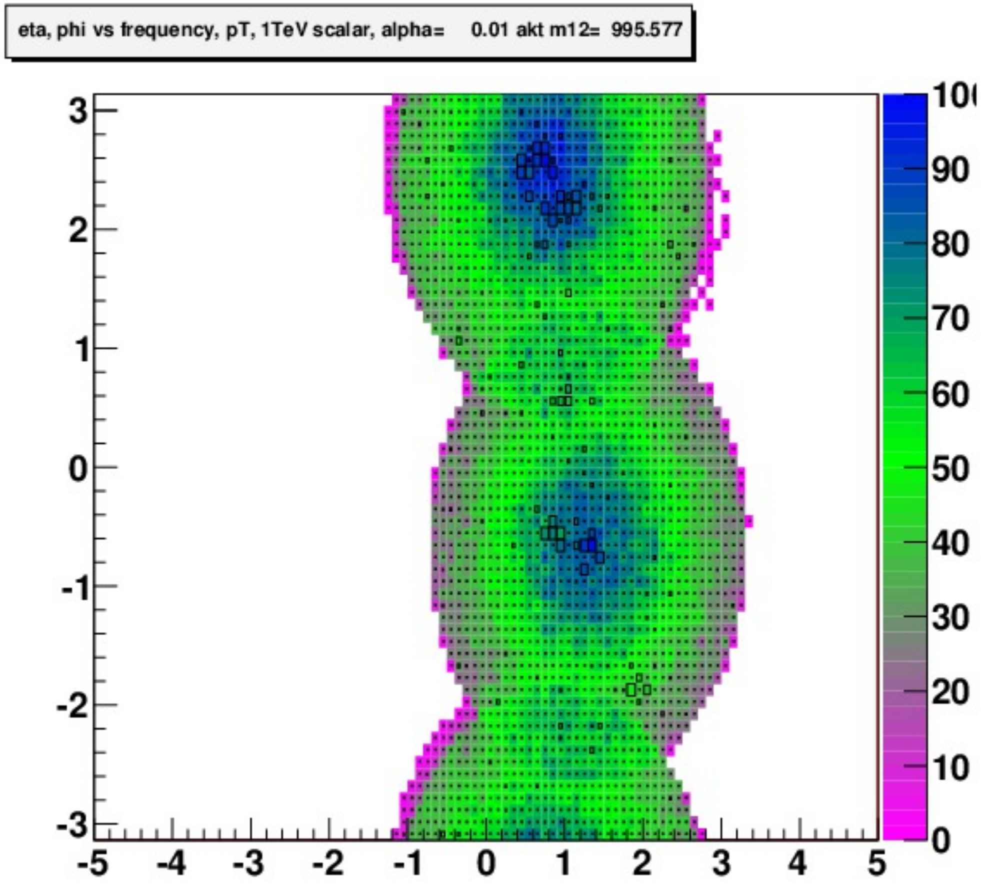
$\alpha = 0.1$



Work in progress, with D. Krohn and D. Kahawala

Qjets on dijet events (no pruning)

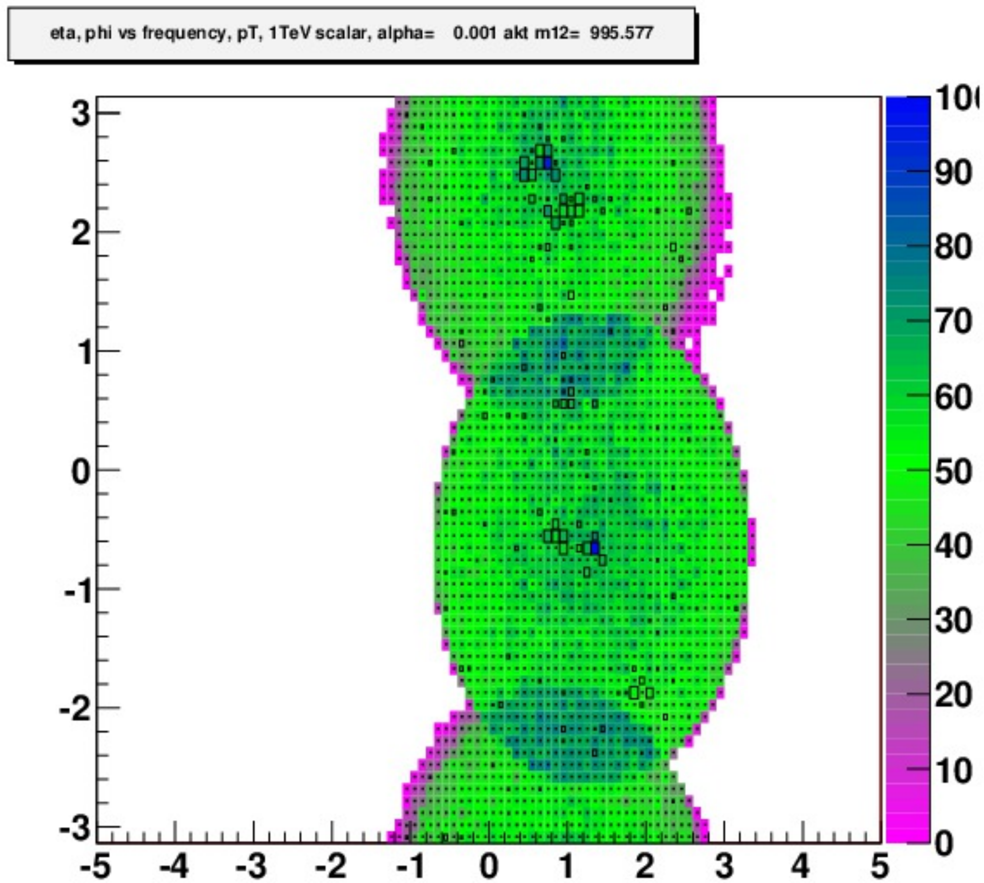
$\alpha = 0.01$



Work in progress, with D. Krohn and D. Kahawala

Qjets on dijet events (no pruning)

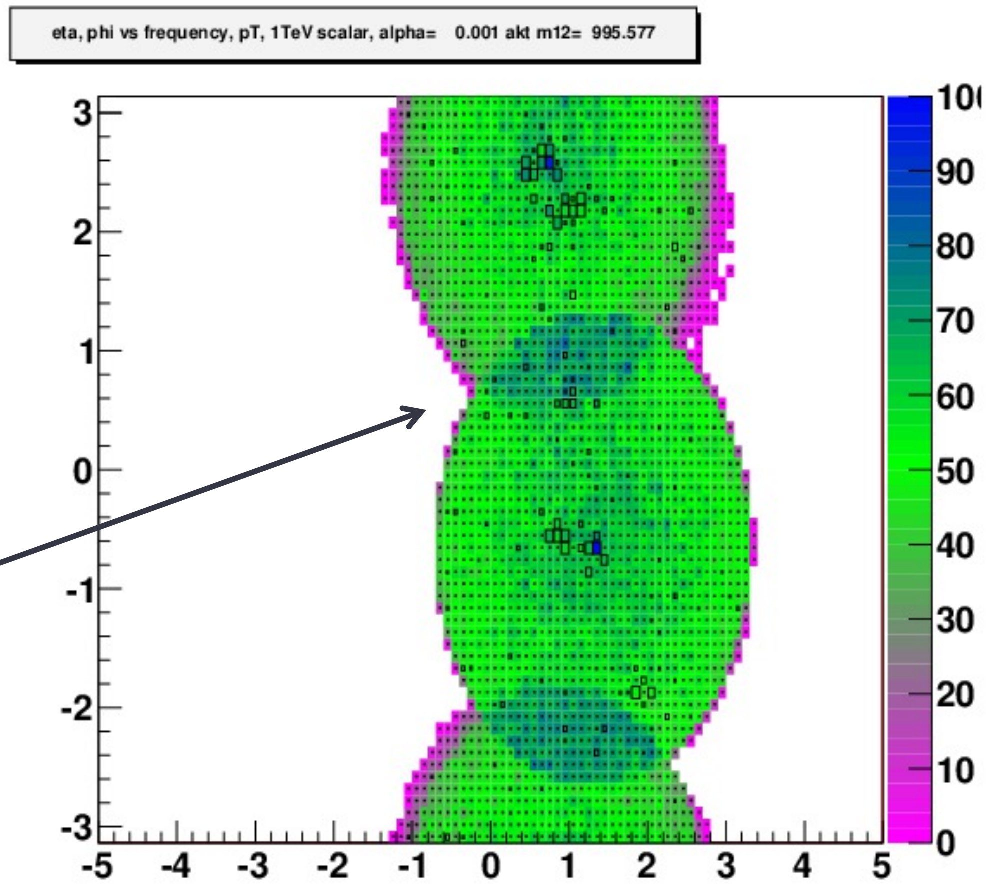
$\alpha = 0.001$



Work in progress, with D. Krohn and D. Kahawala

Qjets on dijet events (no pruning)

$\alpha = 0.001$



May help **resolve ambiguities** with **overlapping jets**

Work in progress, with D. Krohn and D. Kahawala

Summary of Qjets

- **Parton shower** is **not invertible**: Jet-to-parton map is not unique
  - Why always pick the most-probable shower history?
  - Use **all** possible **shower histories**!
- **Robust** to choice of jet algorithm
  - Don't need algorithm at all (at least with pruning)

$$\omega_{ij}^{(\alpha)} \equiv \exp\left\{-\alpha \frac{(d_{ij} - d^{\min})}{d^{\min}}\right\}$$

α=0 works great!

- W-mass measurement: 0.45 times luminosity required as with classical jets
- Boosted W's versus QCD jets background
  - **Significance improvement of 2.5** over simple mass window cut
  - Significance improvement of 1.7 over n-subjettiness
- **Lots of potential applications** – we're just starting to think about them

What else is wrong with the jet-to-parton map?

It treats jets as 4-vectors

- Jets have **color** , and color connections
  - Used by D0 (published) and ATLAS (Moriond, hopefully)

Gallichio and MDS **Phys.Rev.Lett.** 105 (2010) 022001

- Quark and gluon jets may be different
  - **New physics** is **quark heavy**, backgrounds are **gluon heavy**
  - Although difficult, quark and gluon discrimination could be extremely useful

Gallichio and MDS **Phys.Rev.Lett.** 107 (2011) 172001

- Jets have **charge**

Krohn, Lin, MDS, work in progress

- Jets from boosted objects have **substructure**
  - E.g. top-tagging from boosted top jets – used by CMS!

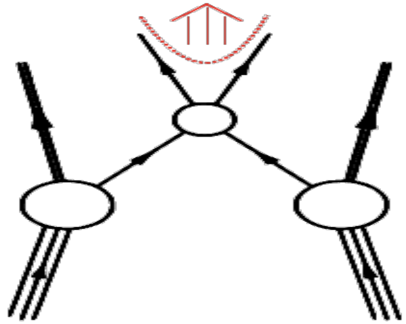
Kaplan, Rehermann, MDS, Tweedie **Phys.Rev.Lett.** 101 (2008) 142001

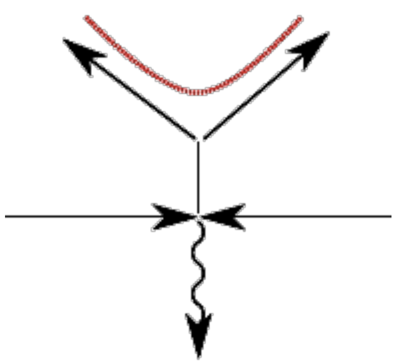
  - Boosted Higgs searces

Butterworth, Davison Rubin, Salam Phys.Rev.Lett. 100 (2008)

measuring Color flows in jets

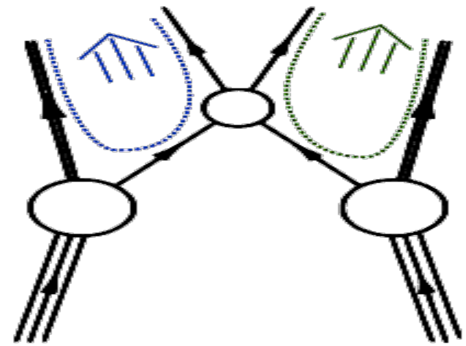
Signal

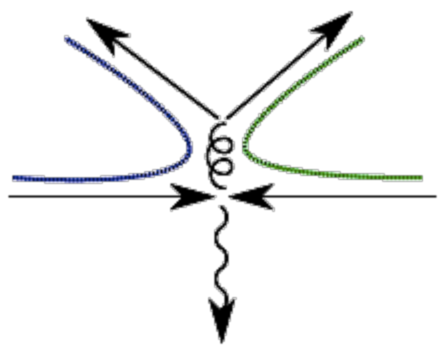




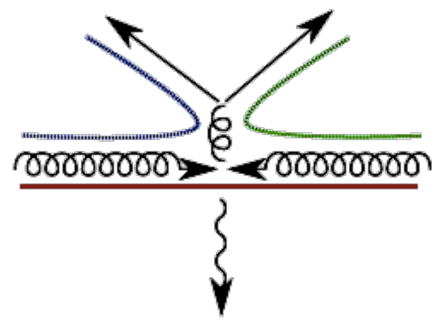
$H \to b\bar{b}$

Background





$q\bar{q} \to Zb\bar{b}$

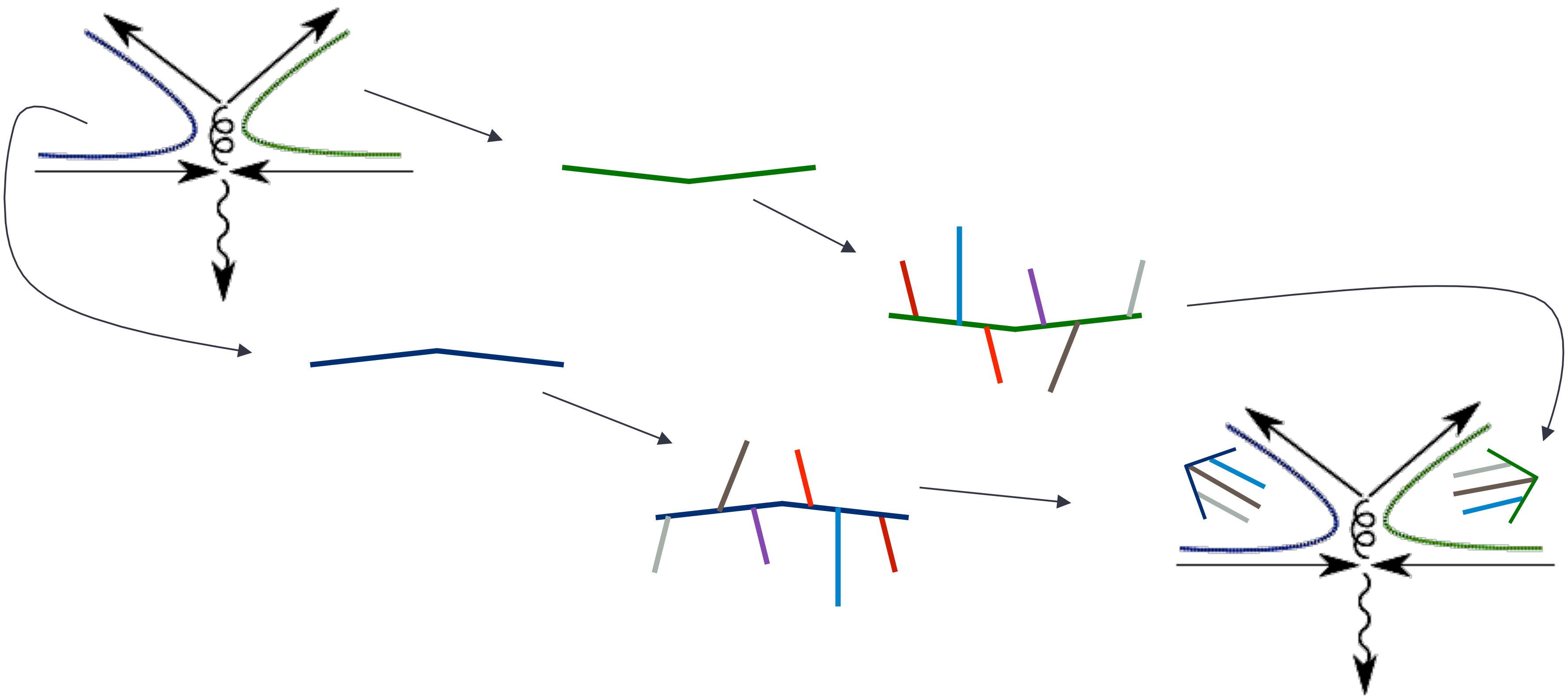


$gg \to Zb\bar{b}$

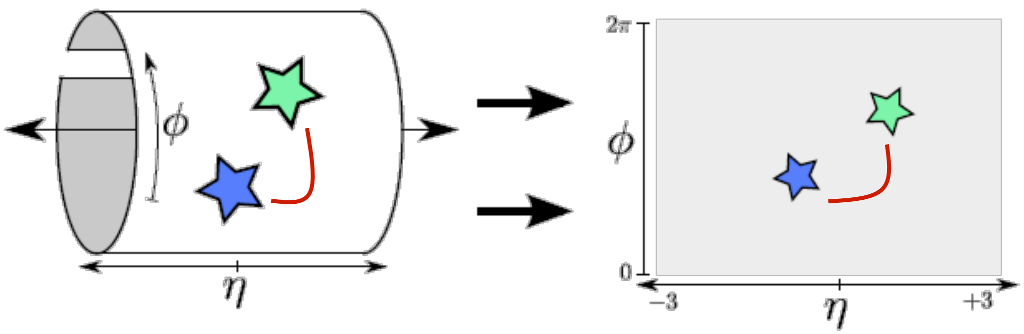
How do they show up?

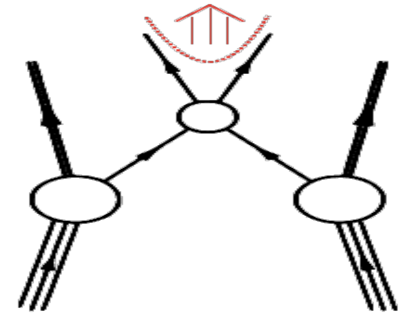
Monte Carlo simulation

- **Color coherence** (angular ordering, e.g. Herwig)
- Color string showers in its rest frame (pt ordering, e.g. Pythia)
  - Boost → **string showers** in **string-momentum** direction



How do they show up?



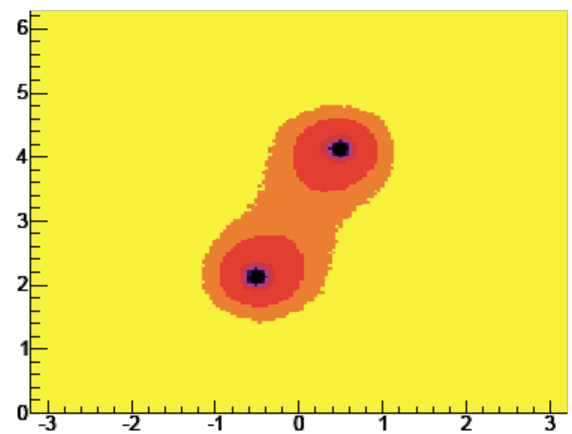


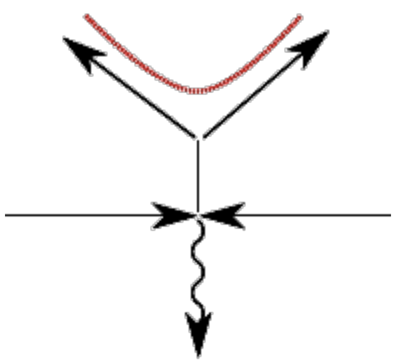
Shower same event
millions of times

Higgs:

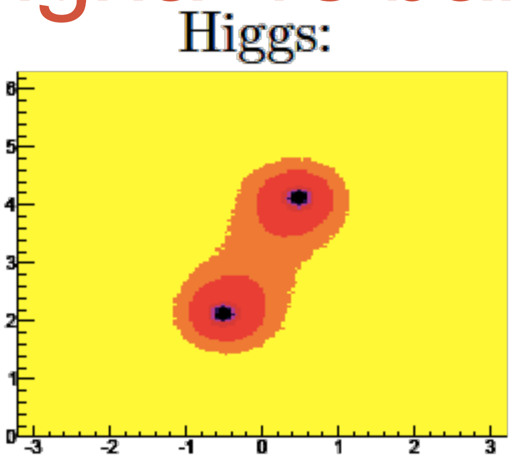
$\Delta\eta_{b\bar{b}} = 1$
$\Delta\phi_{b\bar{b}} = 2$

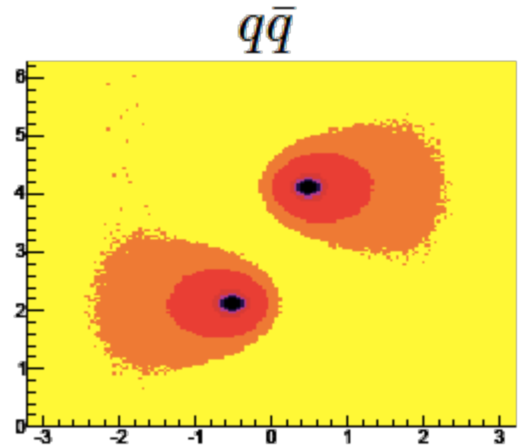
Add up E_T in each cell:

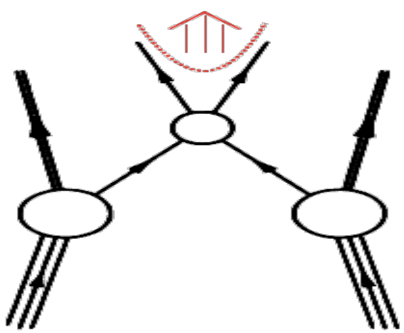


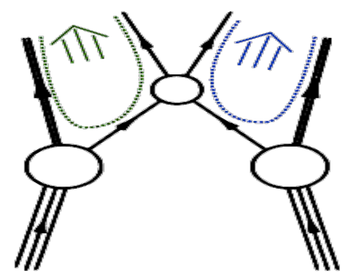


Signal vs bakcground





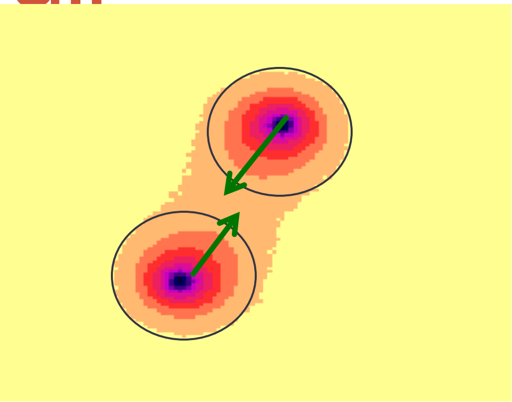


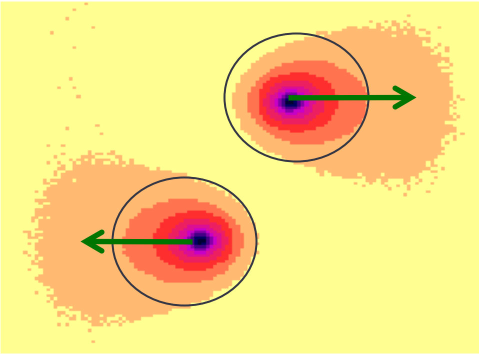


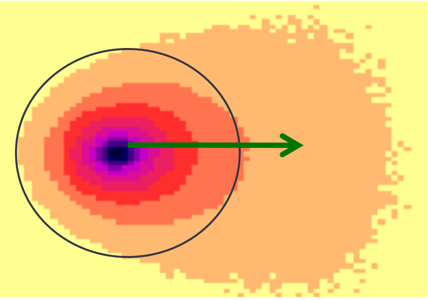
Signal (Higgs)
Color singlet

Background (QCD)
Color connected to beam

Pull

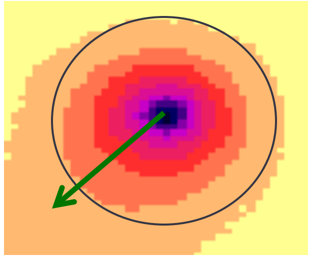






- Find **jets** (e.g. anti-k_T)
- Construct **pull vector** (~ dipole moment) on radiation in **jet**

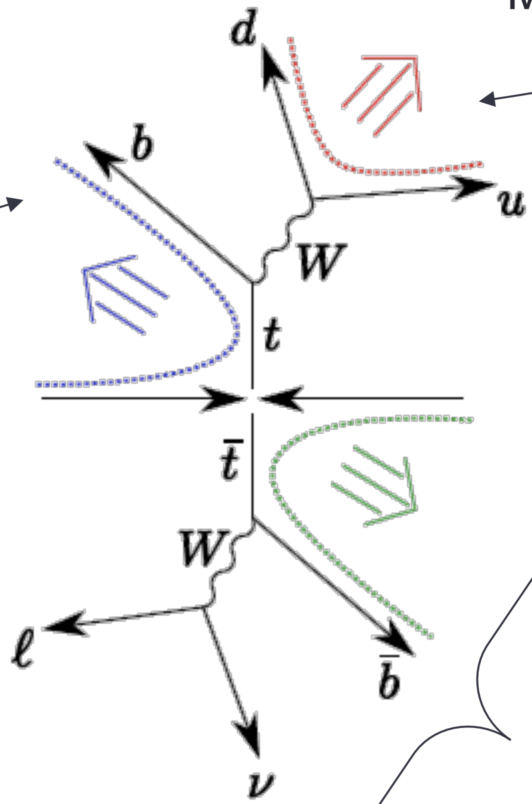
$$\vec{p} = \sum_i \frac{E_T^i \, |r_i|}{E_T^{jet}} \, \vec{r}_i$$



Can we validate? Yes! on ttbar

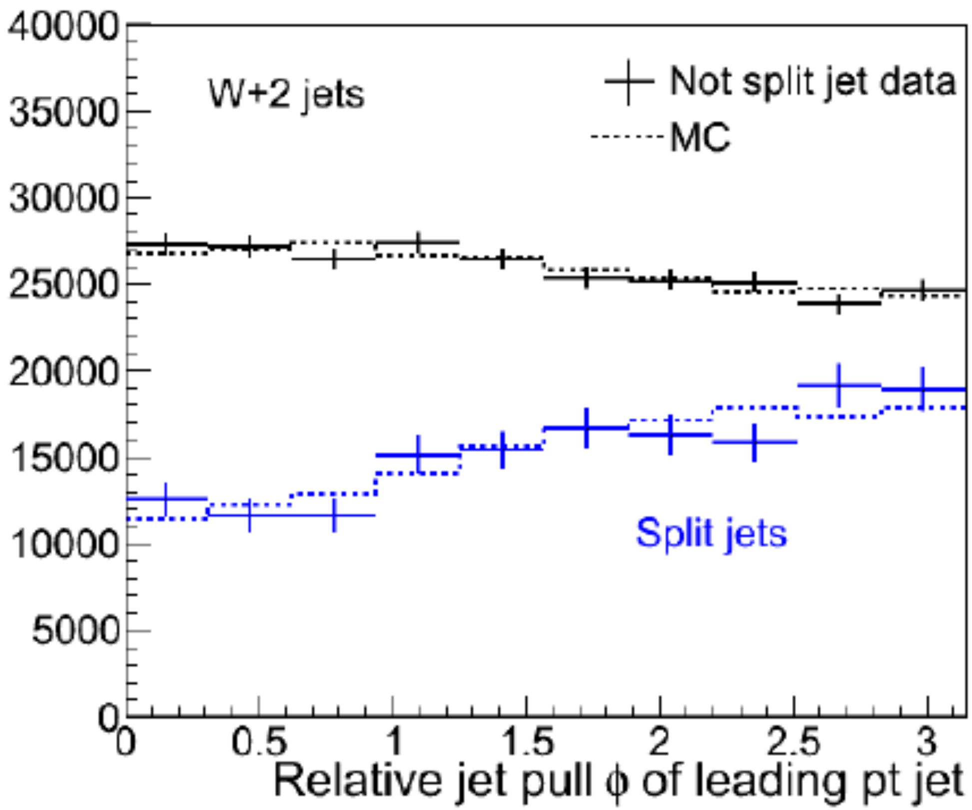
b-tag

Measure pull

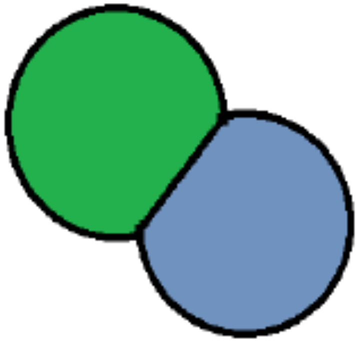


Clean top tag on leptonic side

Measured by D0



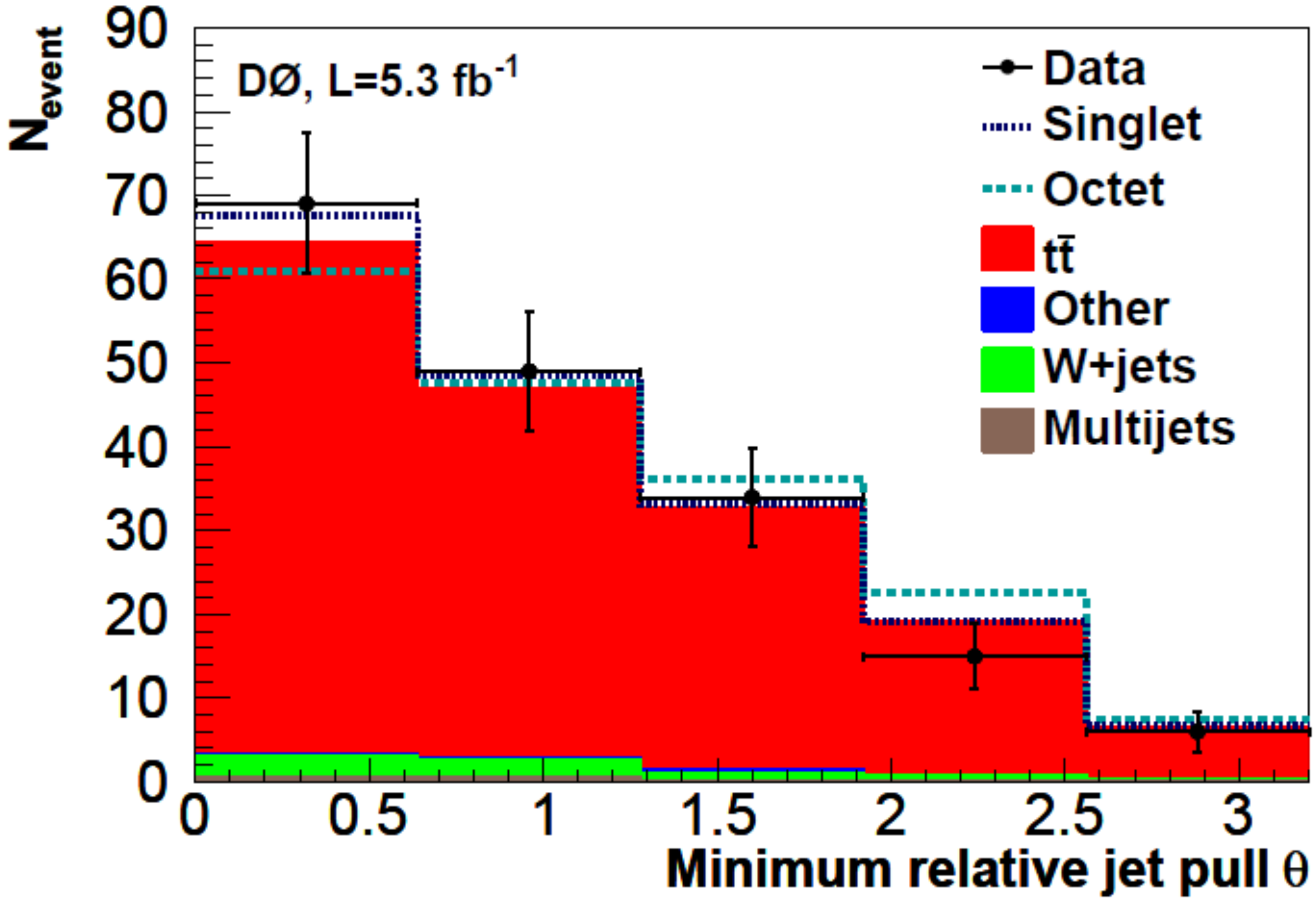
Noise/pileup area smaller towards other jet!



Cells are assigned to the *nearest jet*

D0 ruled out color octet W

Andy Haas and Yvonne Peters, hep-ex:1101.0648



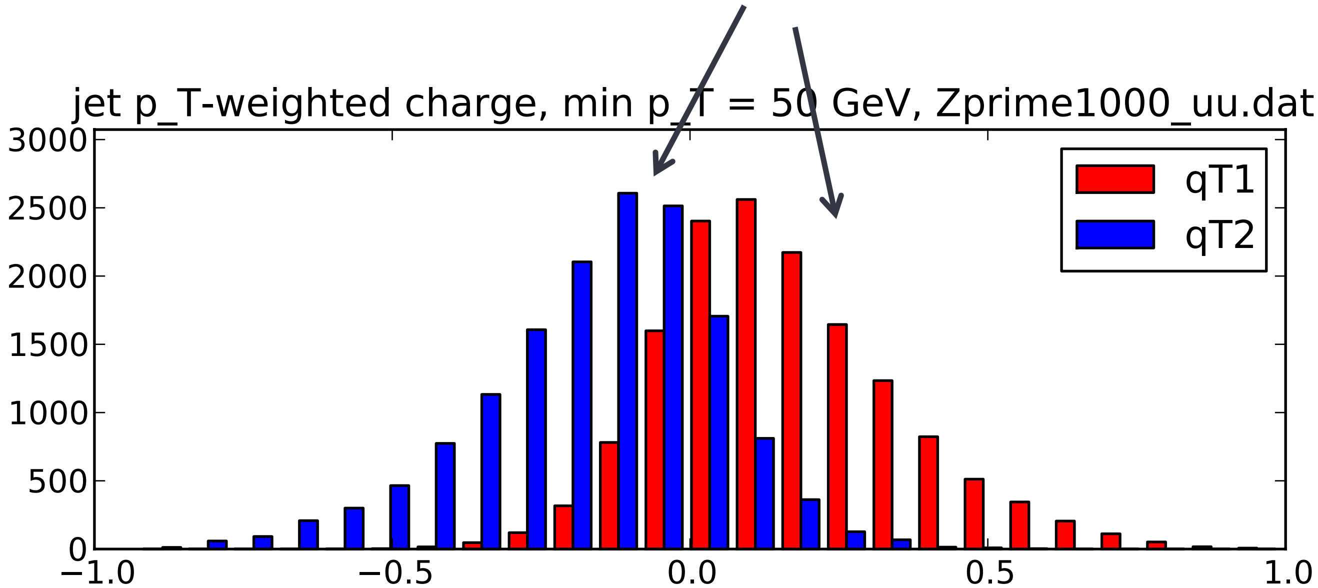
Jet Charge

Work in progress with
David Krohn and Tongyan Lin

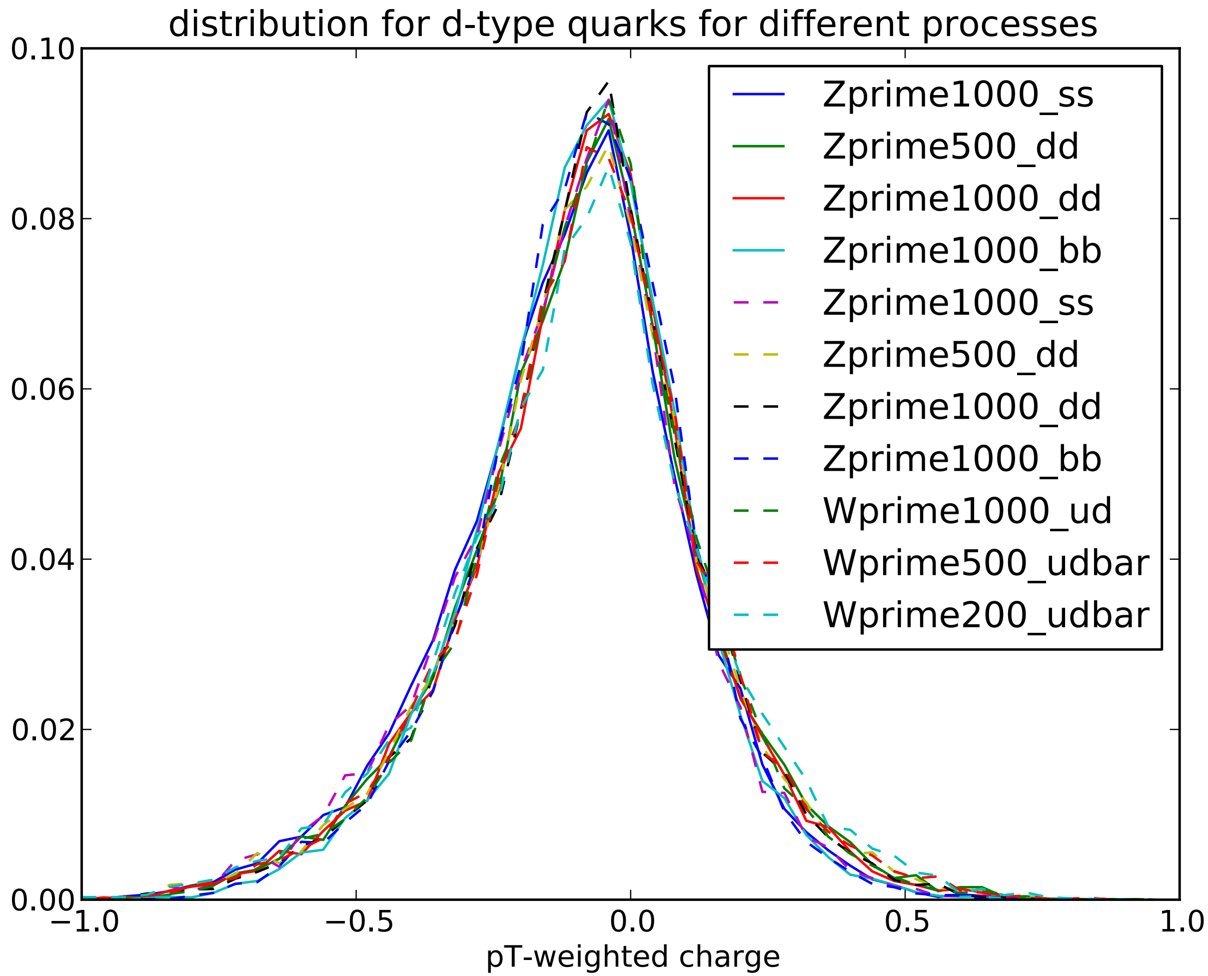
- **Measured** at LEP for light-quark forward backward asymmetries

$$\langle Q \rangle = \sum_i p_T^i Q^i$$

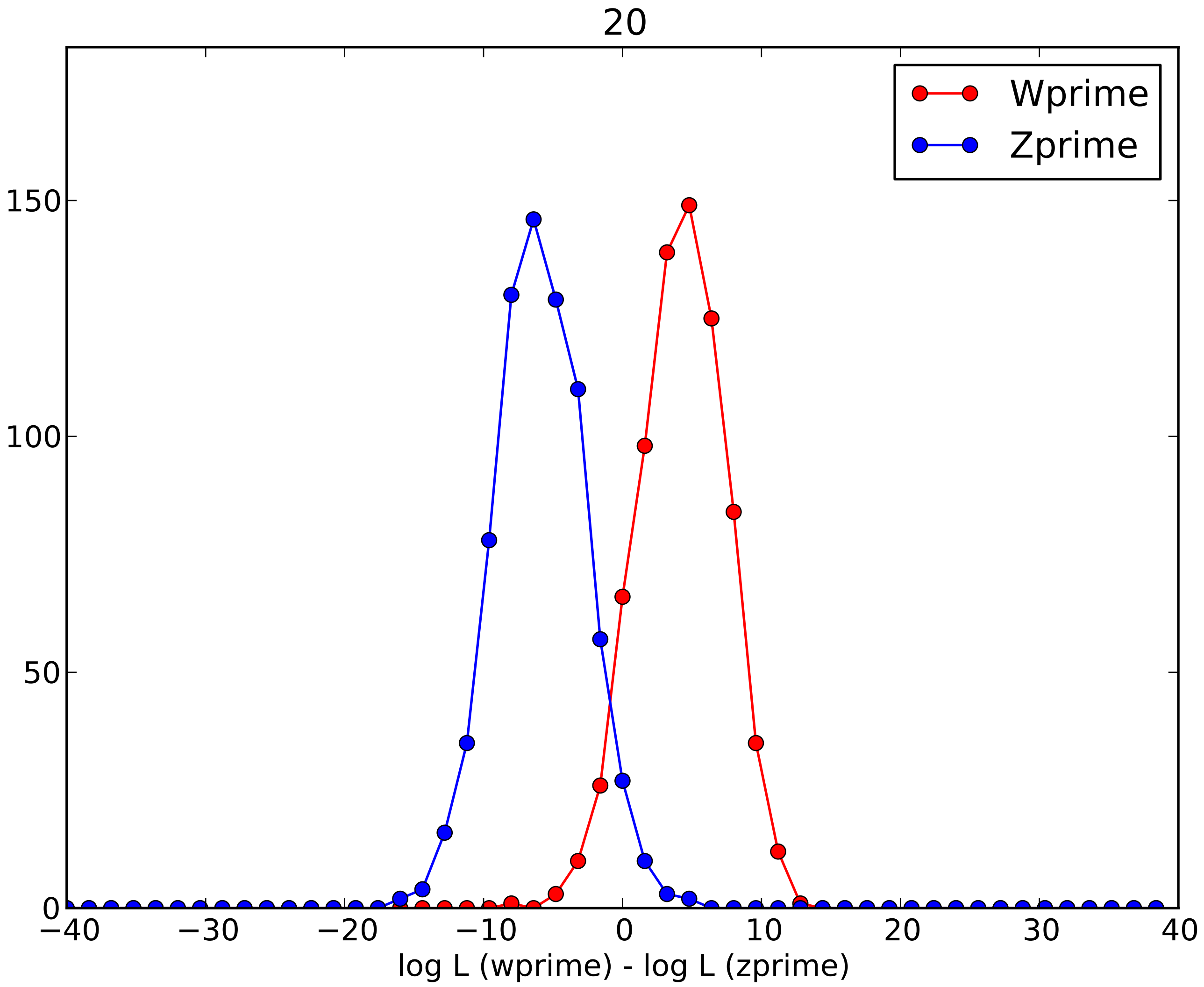
$$Z' \to \bar{u} u$$



Consistent among flavors



W' vs Z' log likelihood



Quark versus Gluon jets

Subtle subject

- Monte Carlo event generators **may not be trustworthy**
- Some data from LEP, but not at the precision that ATLAS and CMS can measure

Two parts

1. Assuming Pythia is correct, **how can we distinguish Q from G**?

   Gallichio and MDS **Phys.Rev.Lett.** 107 (2011) 172001

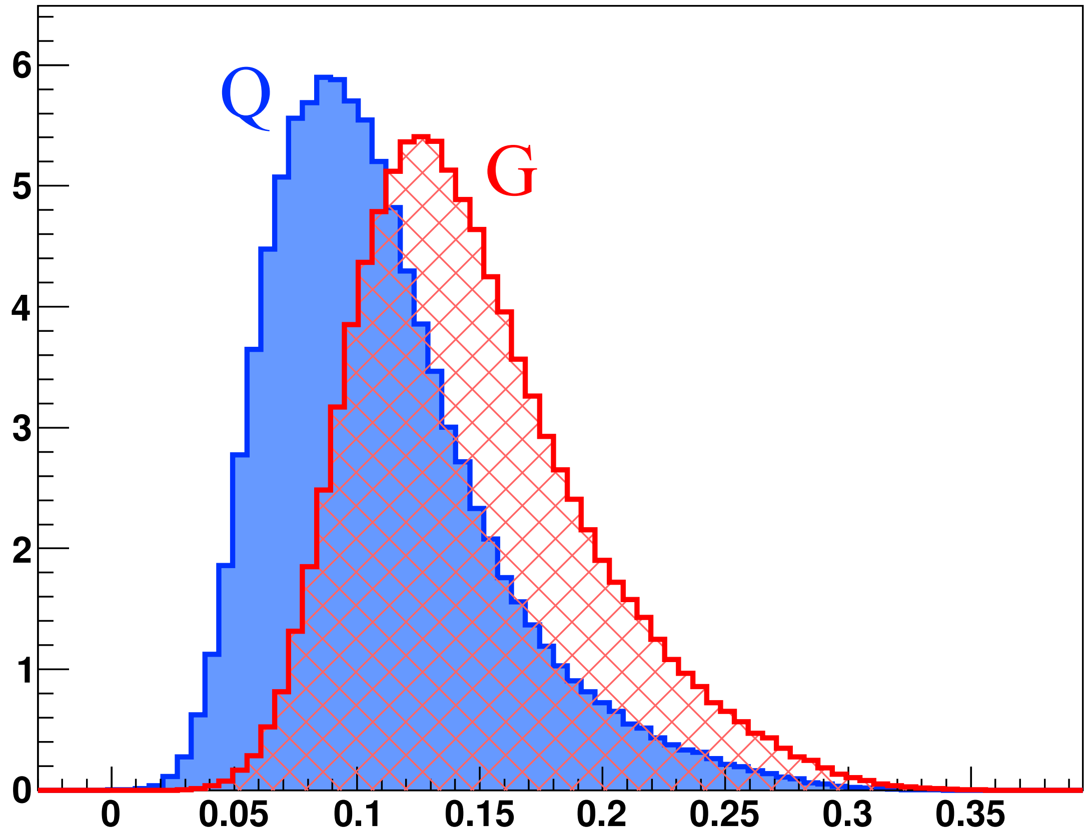
2. How can we **validate on data**?
   - Where do we find pure samples of quark and gluon jets?

   Gallichio and MDS **JHEP** 1110 (2011) 103

How to compare variables?

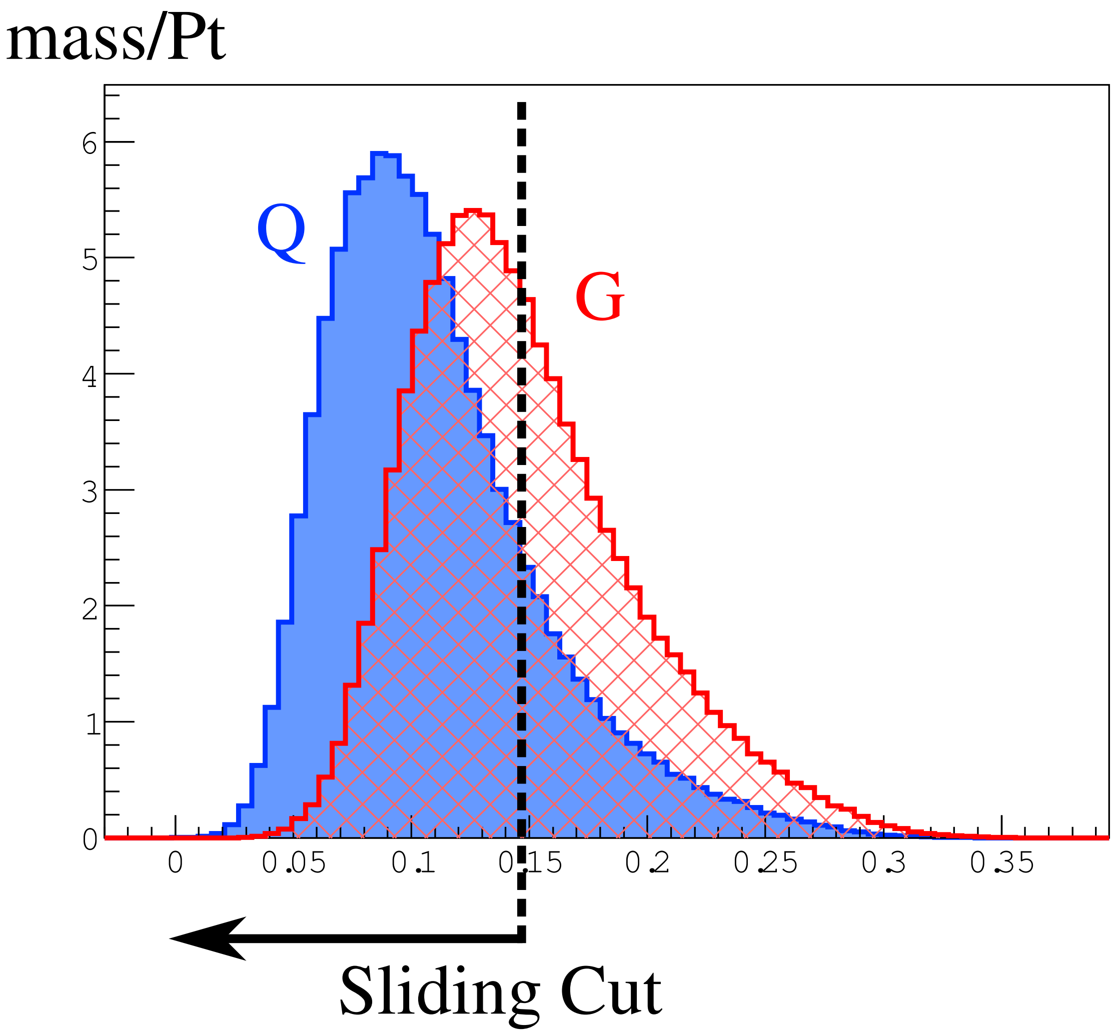
- Look at distributions of each variable, normalized to equal area

mass/Pt



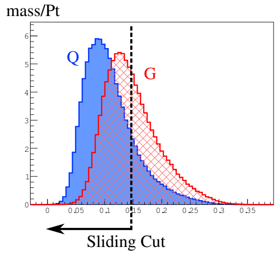
How to compare variables?

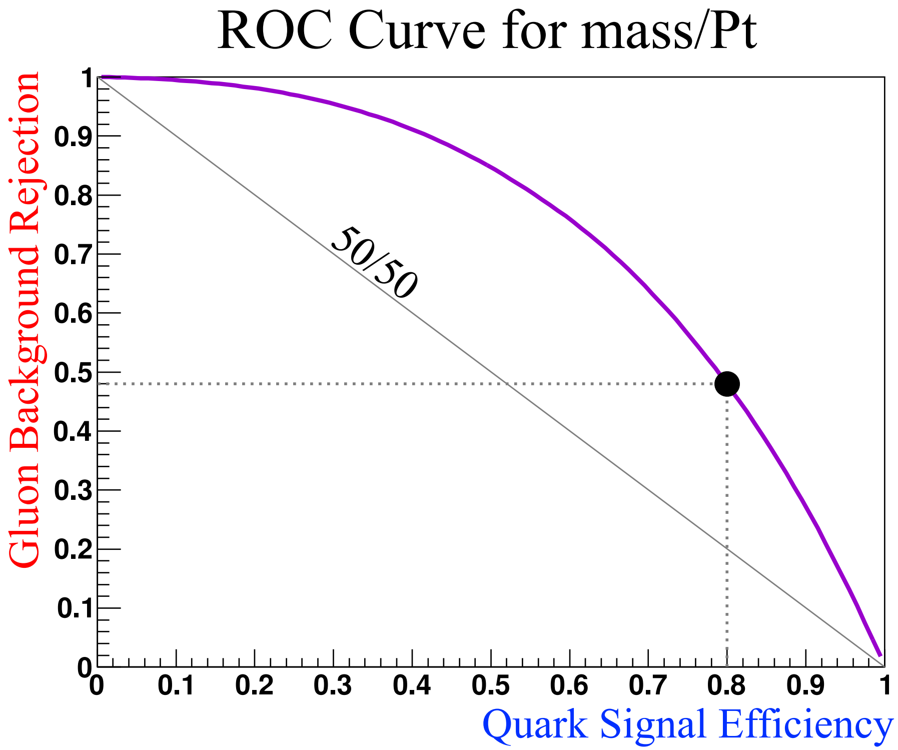
- Look at distributions of each variable, normalized to equal area
- Look at efficiencies as a function of sliding cut



How to compare variables?

This generates the "Receiver Operator Characteristic" (ROC)





We looked at 10,000 variables

The menu, including varying jet size

- Distinguishable particles/tracks/subjets
  - multiplicity, $\langle p_T \rangle$, σ_{p_T}, $\langle k_T \rangle$,
  - charge-weighted p_T sum
- Moments
  - mass, girth, jet broadening
  - angularities
  - optimal kernel
  - 2D: pull, planar flow
- Subjet properties
  - Multiplicity for different algorithms and R_{sub}
  - First subjet's p_T, 2nd's p_T, etc.
  - Ratios of subjet p_T's.
  - k_T splitting scale

Show http://jets.physics.harvard.edu/qvg

We looked at 10,000 variables

Best 2 were

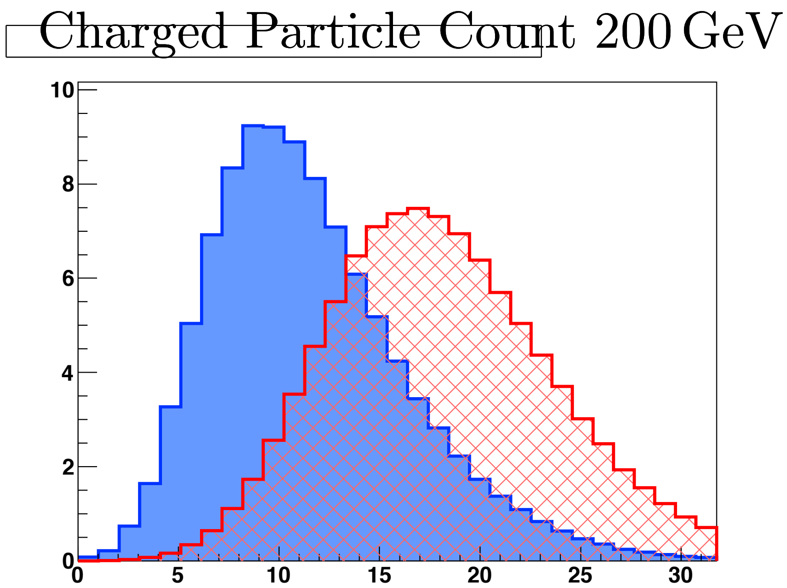
1. Charged particle count
    - Better spatial and energy resolution works better
        - e.g. particles > topoclusters > calorimeter cells > subjets

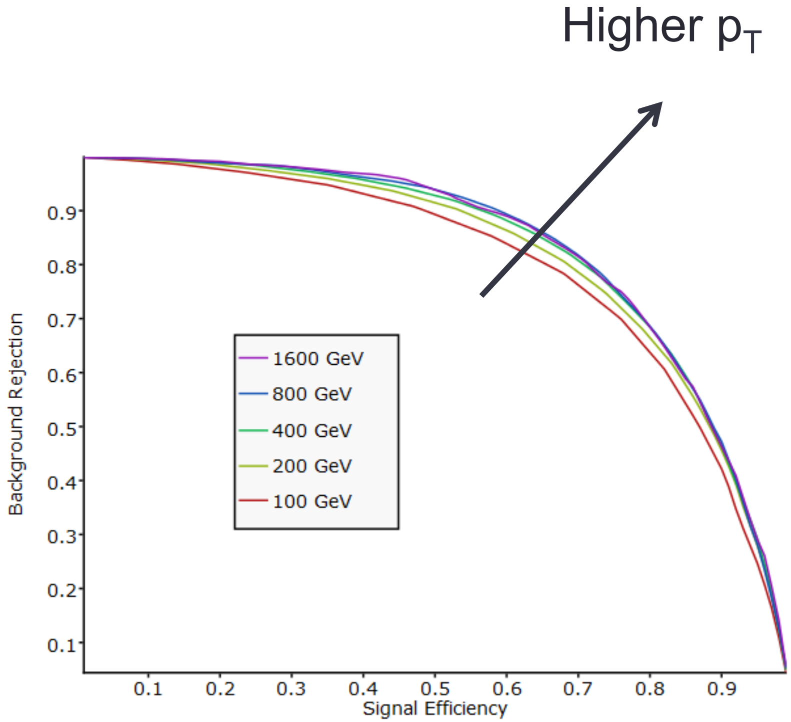
and

2. Linear radial moment (girth)
    - Similar to jet broadening

Show http://jets.physics.harvard.edu/qvg

Charged Particle Count



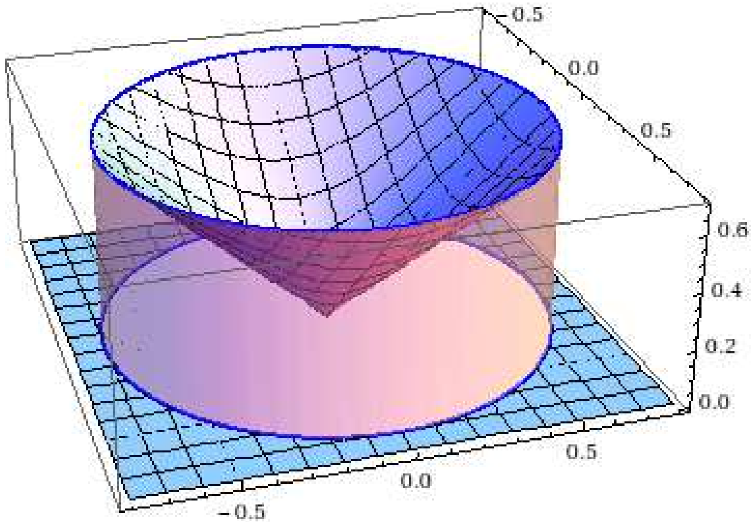


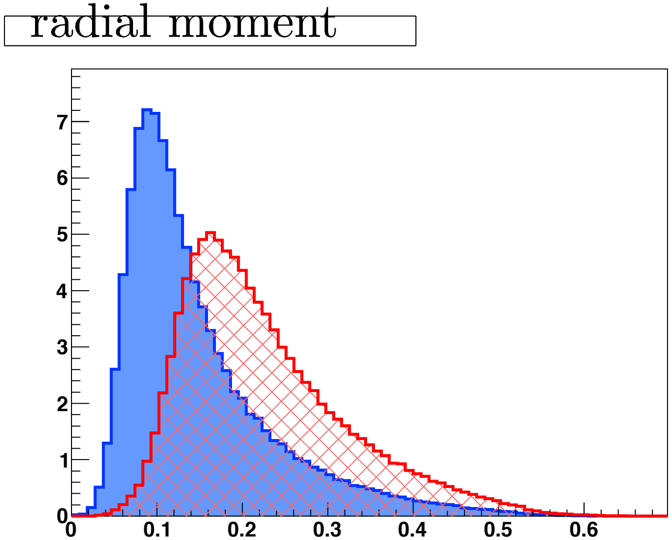
Higher p_T means more tracks and more 'time' to establish C_A/C_F.

Girth

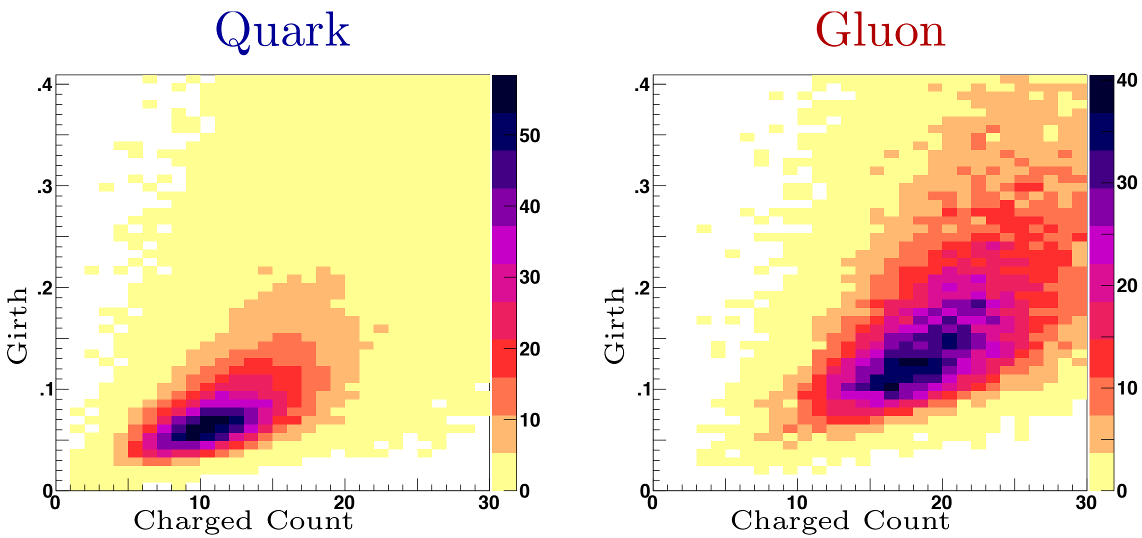
Weight p_T deposits by distance from jet center

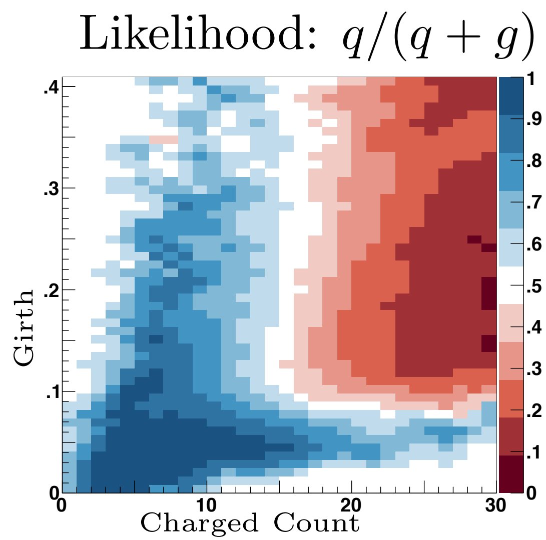
Radial Moment, or Girth :
$$g = \frac{1}{p_T^{jet}} \sum_{i \in \text{jet}} p_T^i |r_i|$$

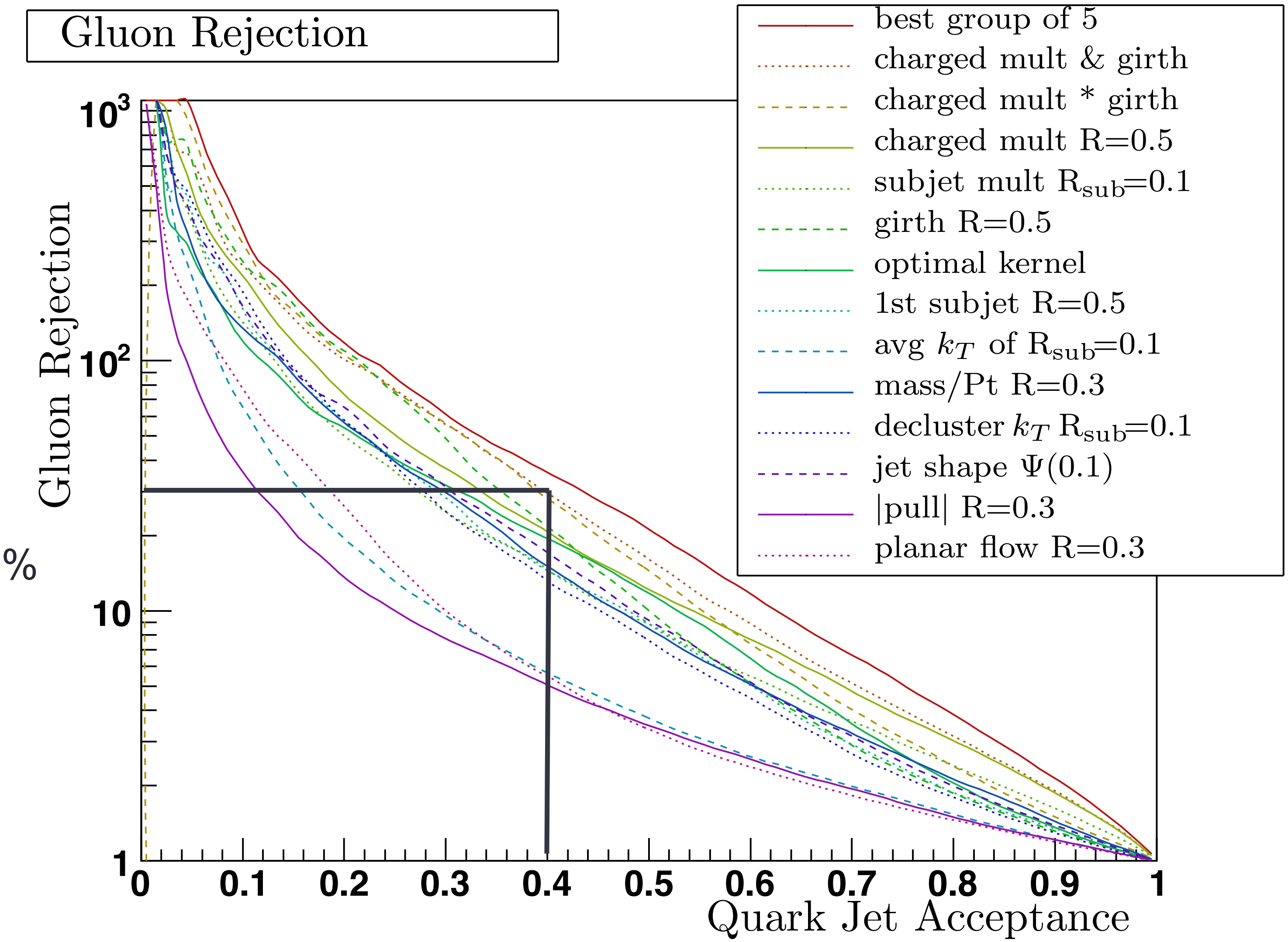




2D distributions show that they are fairly uncorrelated



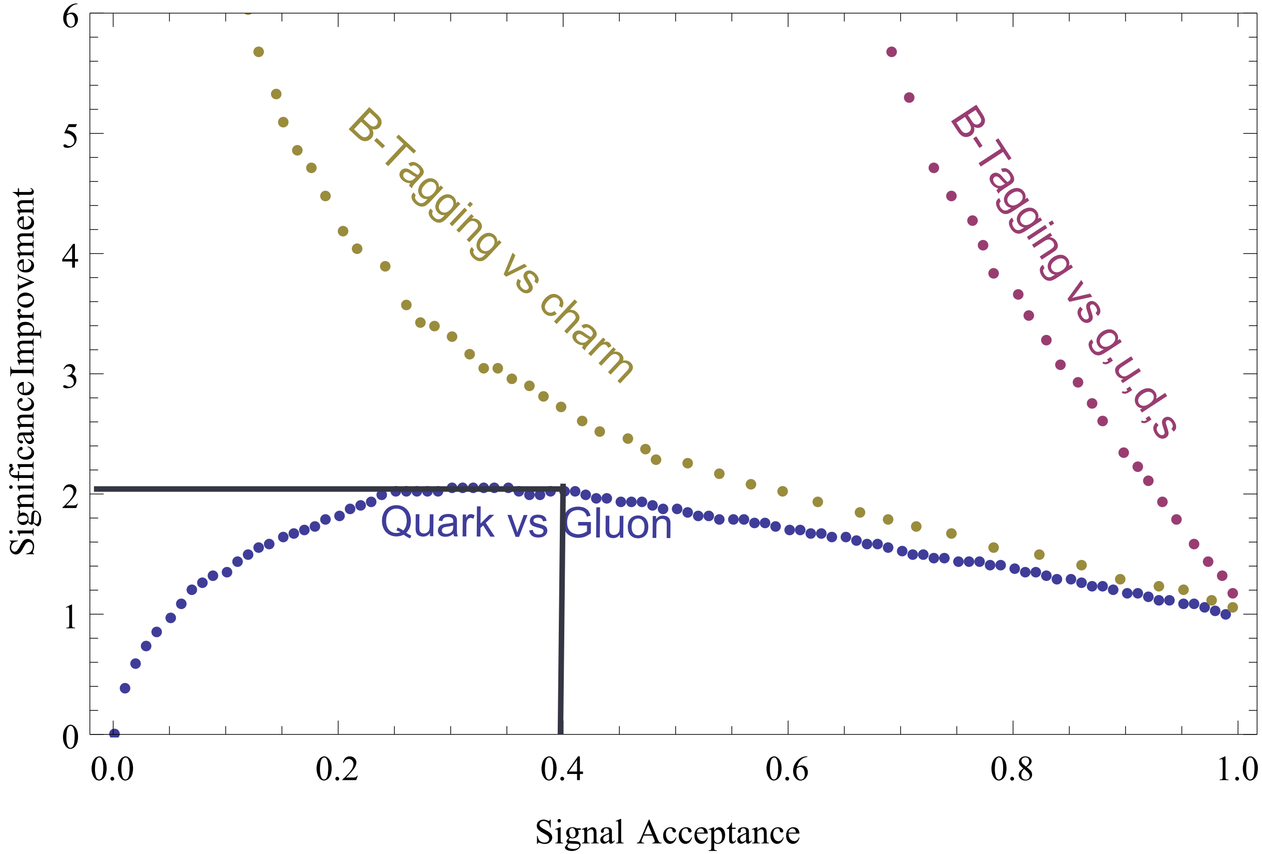




1/30 = 3% gluon

40% quarks

Result



Significance Improvement of $$\frac{0.4}{\sqrt{1/30}} = 2.19$$

Conclusions

"These are not your daddy's jets" -- Steve Ellis

The **LHC is so great** that we can go **well-beyond the jet-to-parton map**

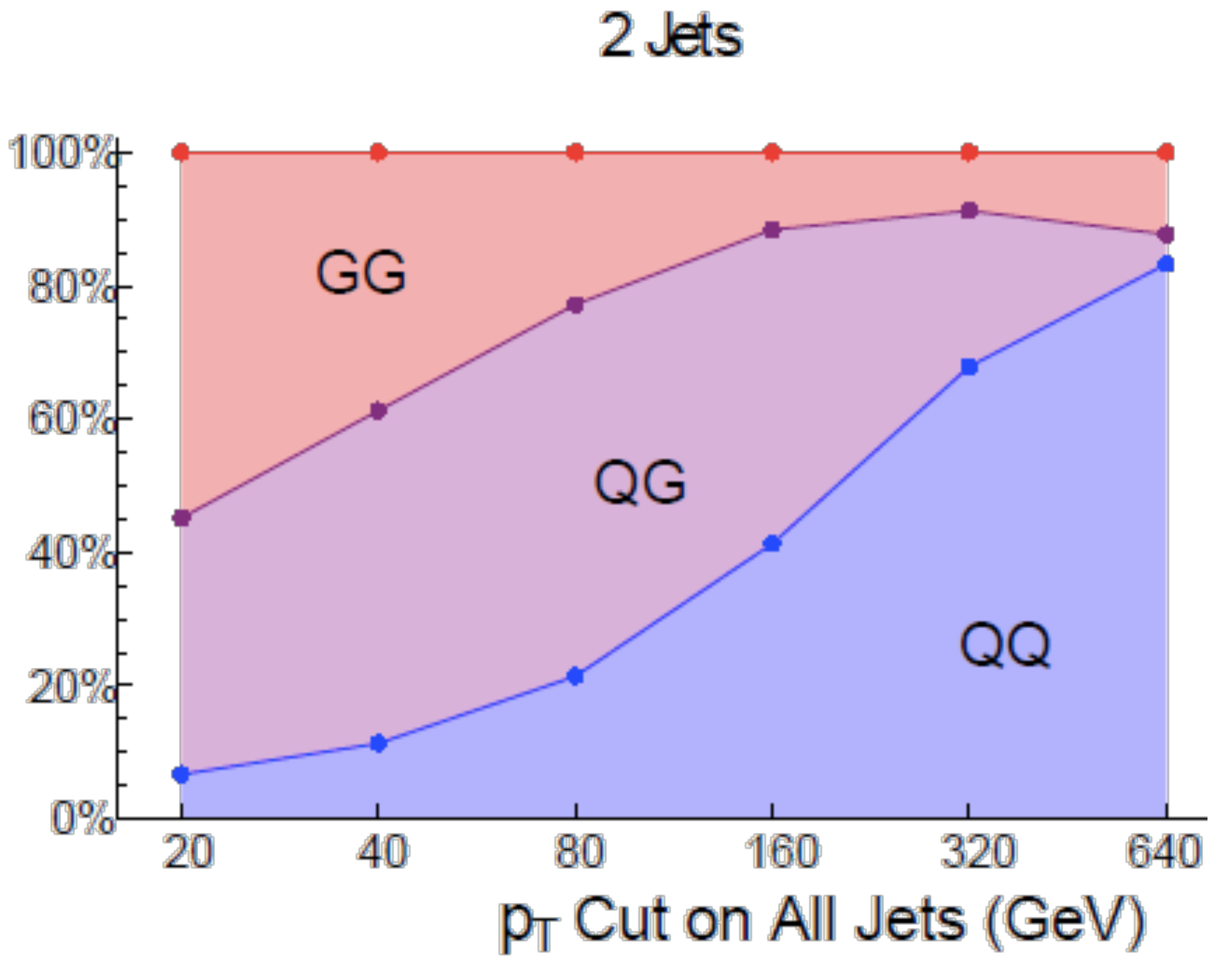
- Detectors can measure jet **substructure**
- Need to look at substructure to find new physics in huge backgrounds

Beyond the jet-to-parton map

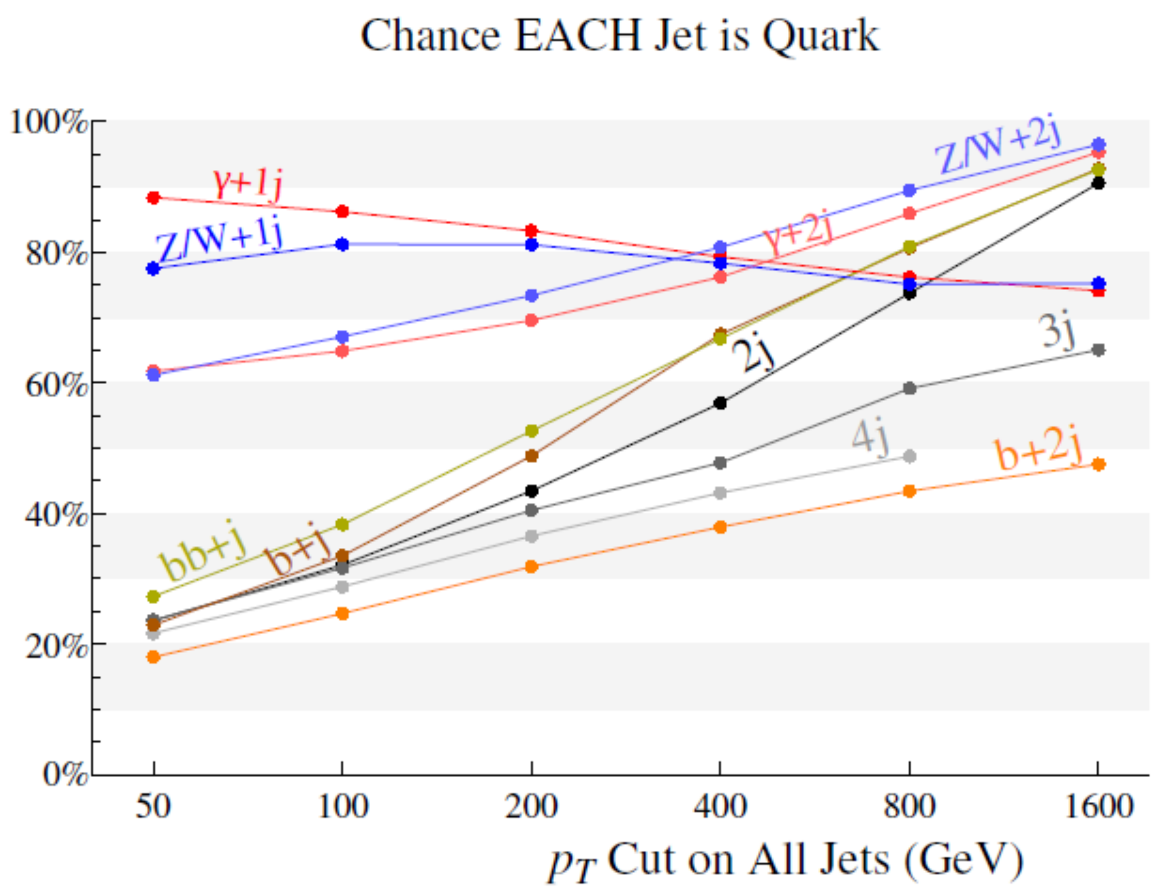
- **Qjets**
  - Not mostly likely shower history,
    but **weighted distribution of *all* shower histories**
  - Can **improve mass resolution** and **discovery potential** in boosted objects
  - Volatility does better for W-tagging than any other *single* variable
- Jets have **color**
  - Color represetations can be measured: adjoint vs singlet
  - **Color connections**
- **Quark** jets and **gluon** jets distinguishable: 40% Q vs 3% G
  - Charge particle count and linear radial moment work best
- Jet substructure
  - **many** worked out **applications** over the last few years

Backup Slides

Where are the quark jets?



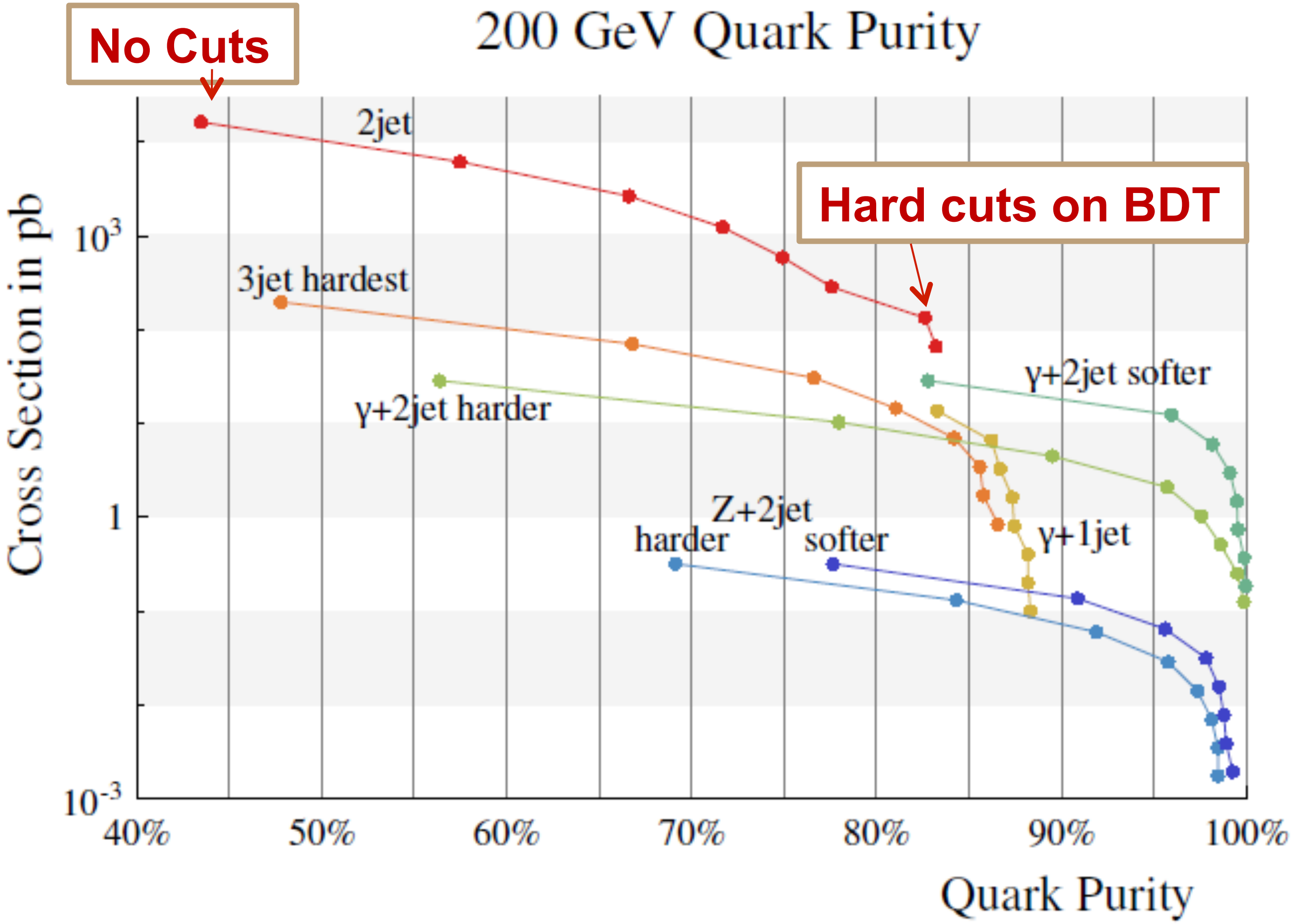
Look at all samples



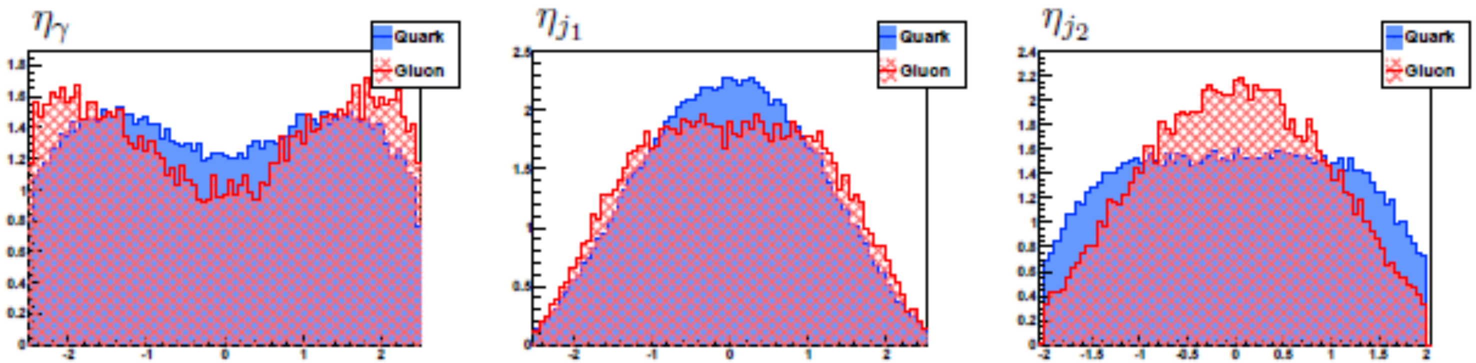
- What about **cross sections**?
- Can **cuts purify** the **samples**?

Throw them into a Boosted Decision Tree

Optimize efficiency using BDT classifier with parton momenta as inputs (6 or 9 inputs)



Now look at the γ + 2 jets sample

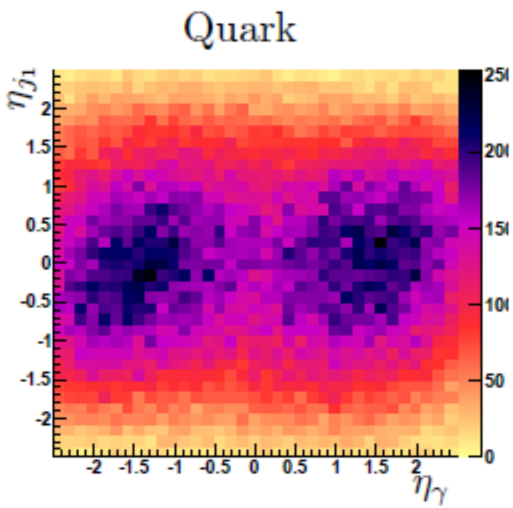


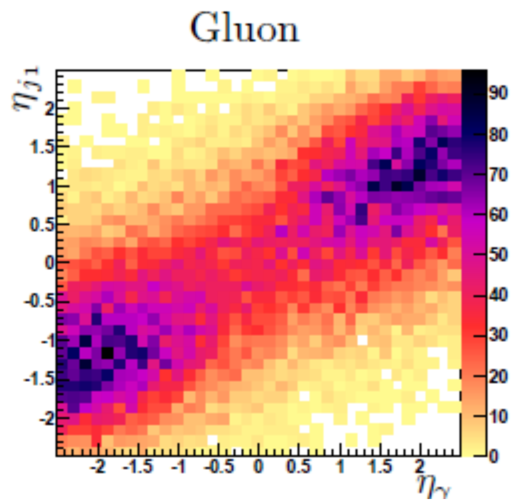
- Look at the best discriminants, ranked by cuts

- The rapidity of the photon and
  the rapidity of the second hardest jet look good

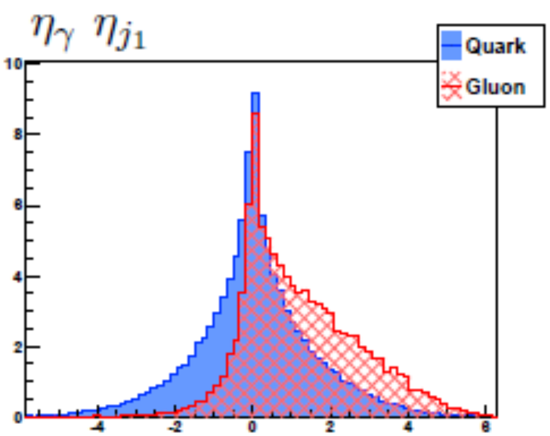
- But cutting on just η_{γ} or just η_{J2} does not help much

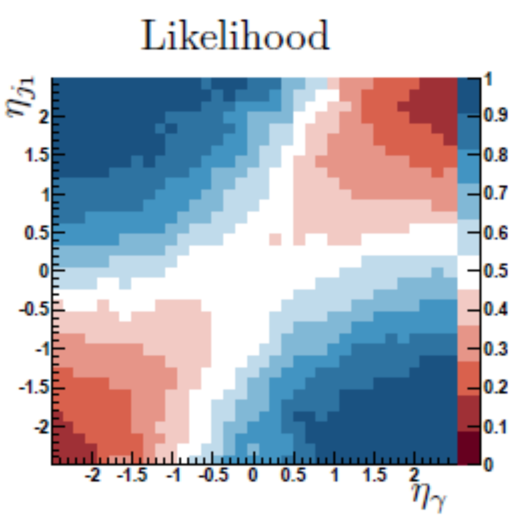
Look at correlations



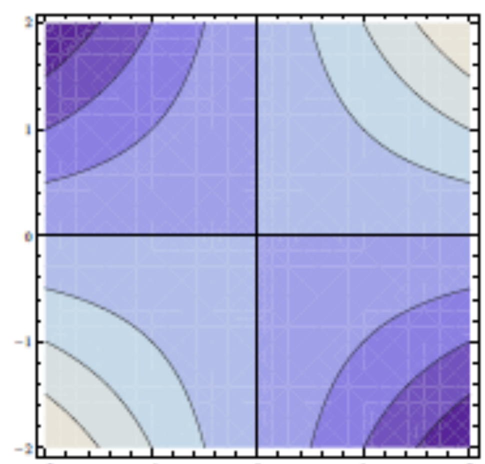


Distribution of $\eta_{\gamma}\eta_{j1}$

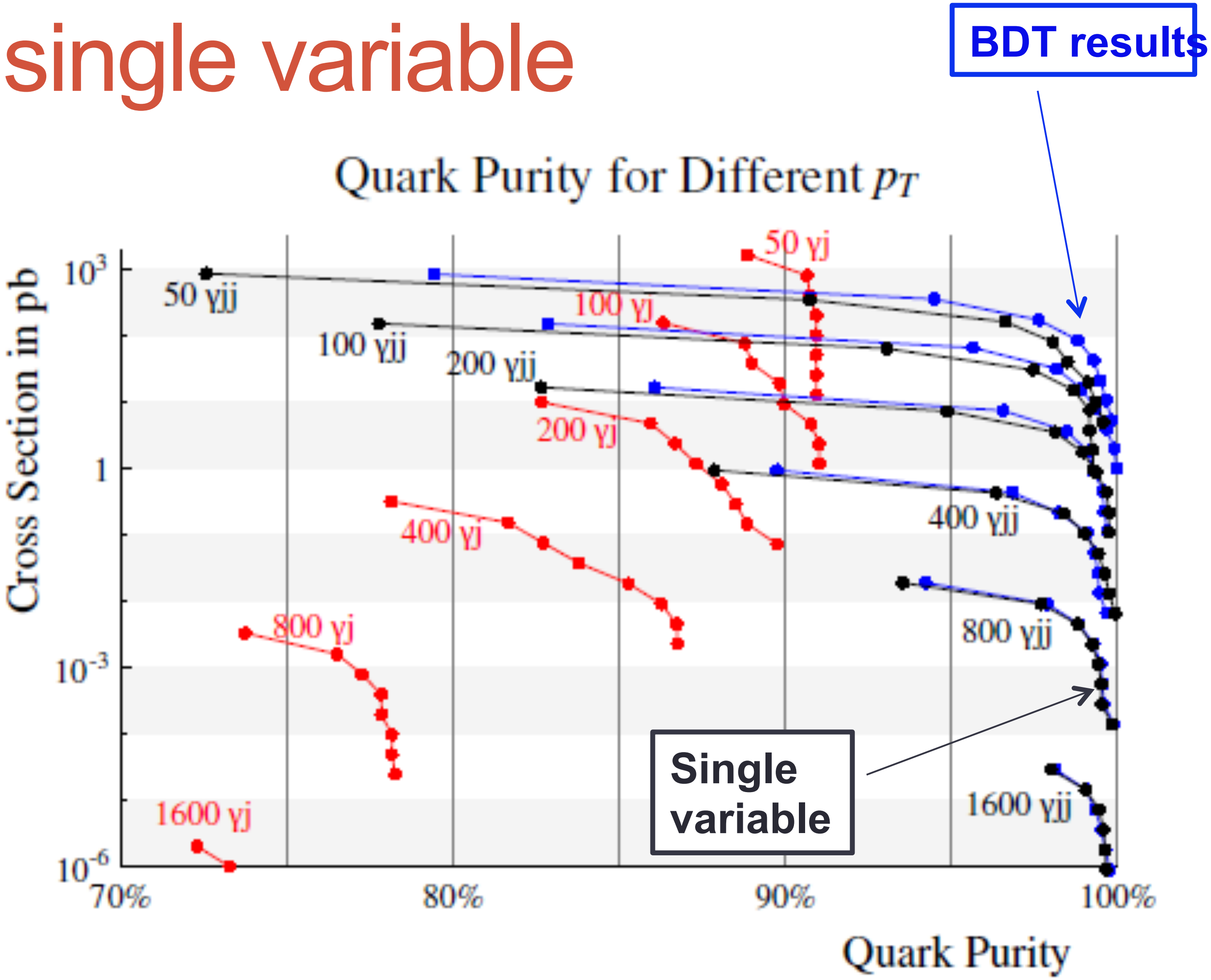




Contours of $\eta_{\gamma}\eta_{j1}$

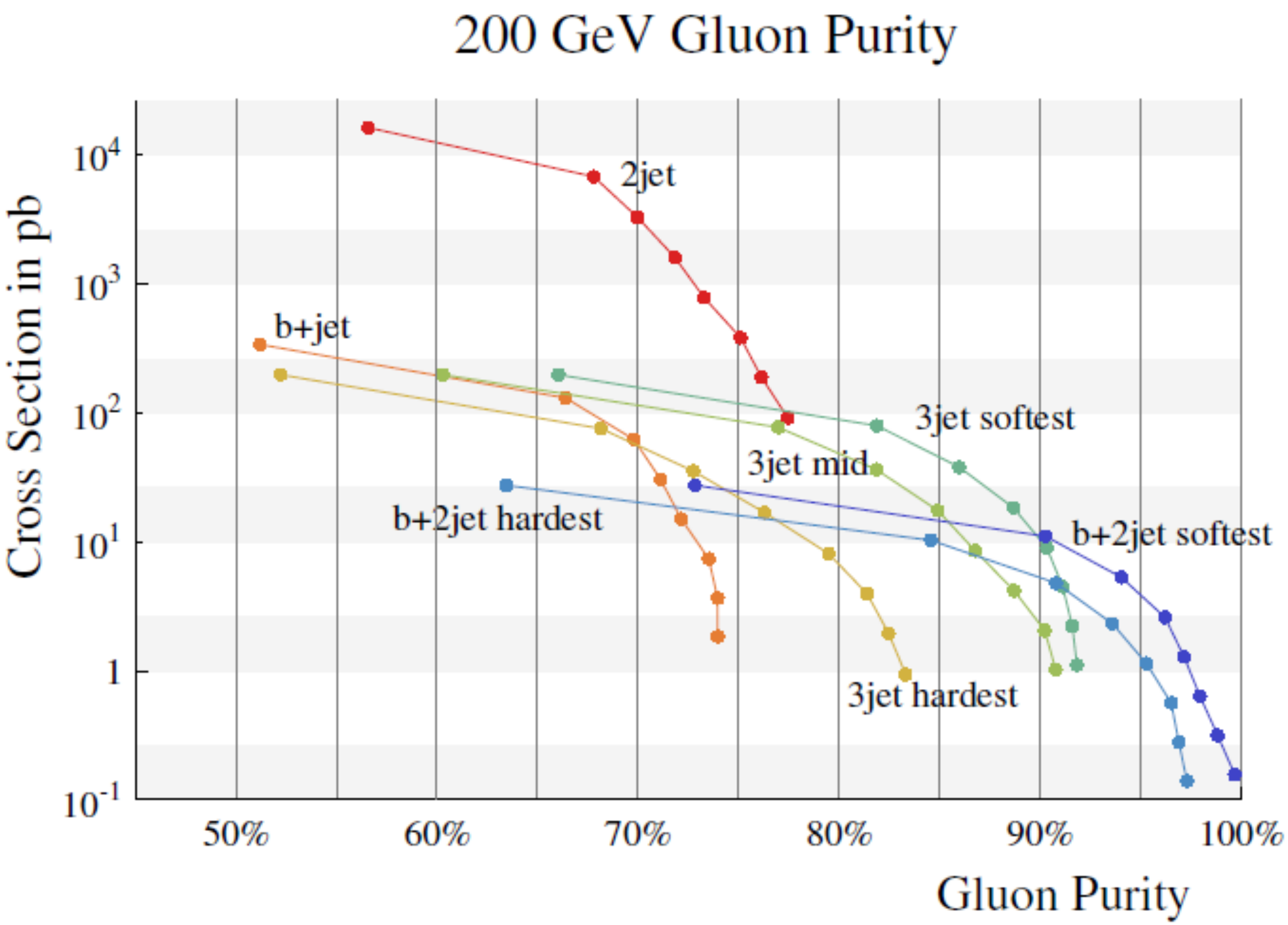


Best single variable



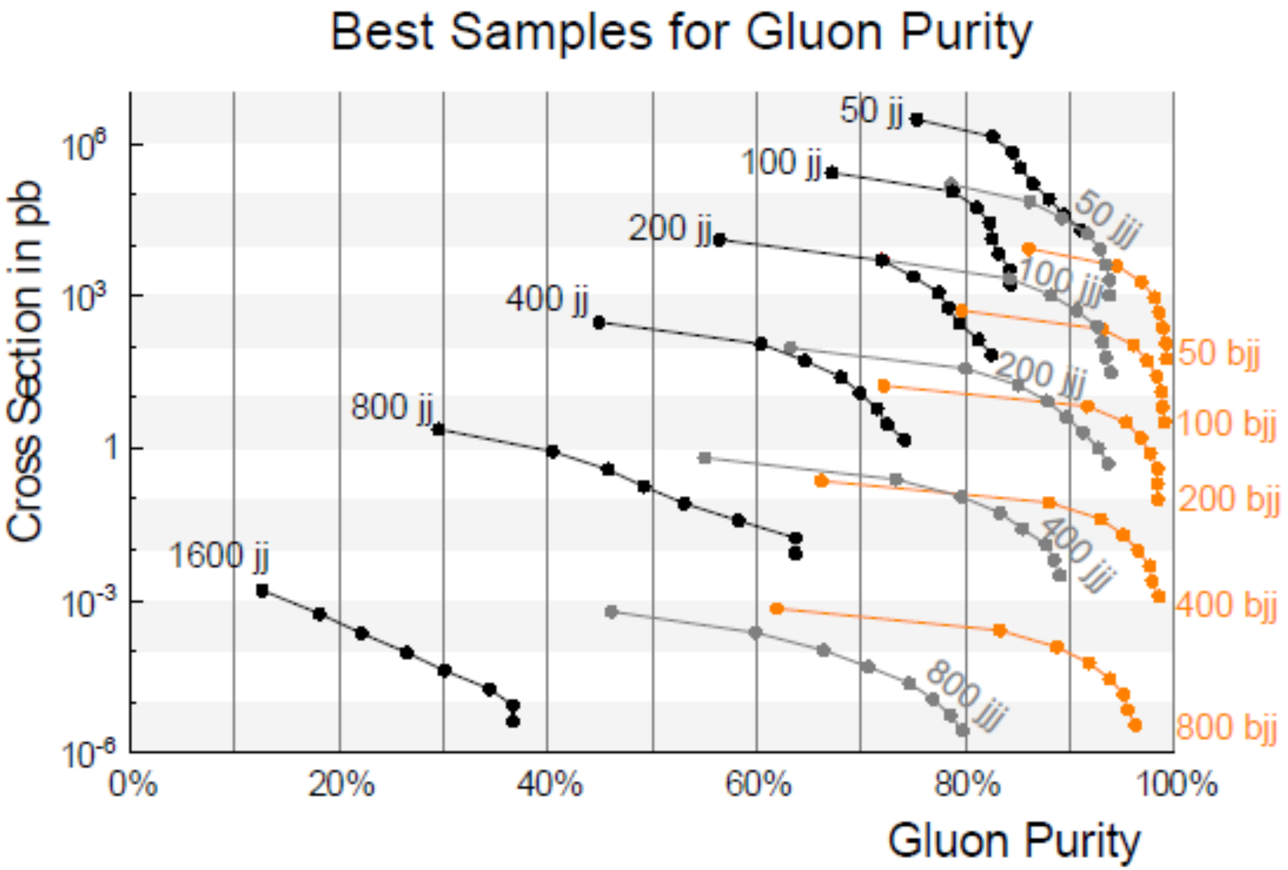
BDTs led us to the variable,
but with the variable we **don't need BDTs**

What about pure gluons?



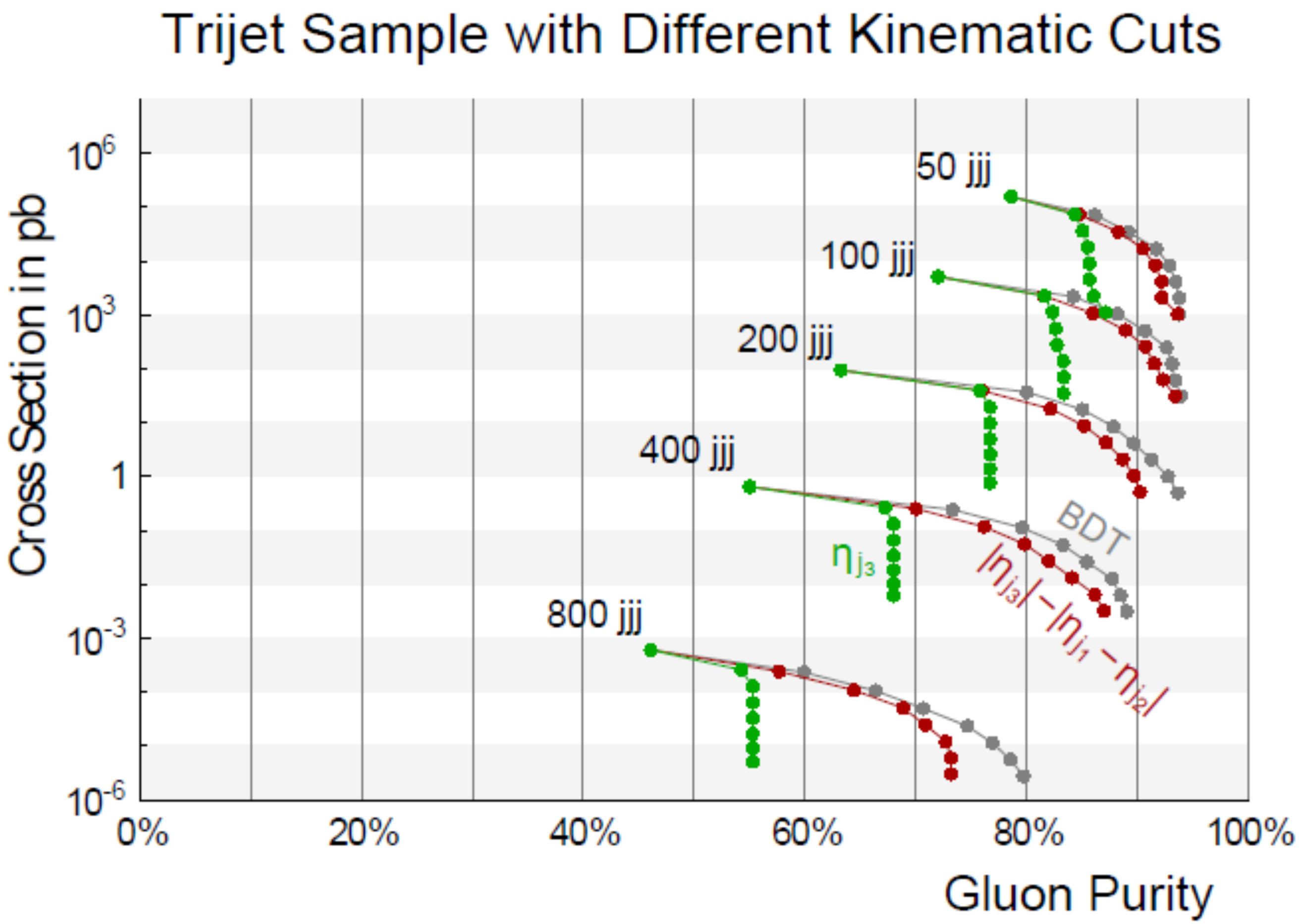
b+2 jets or trijets look promising

Throw it at the BDT



- Now try to find a single variable that works as well…

Finding Pure Gluon jets



Summary of finding quarks/gluons

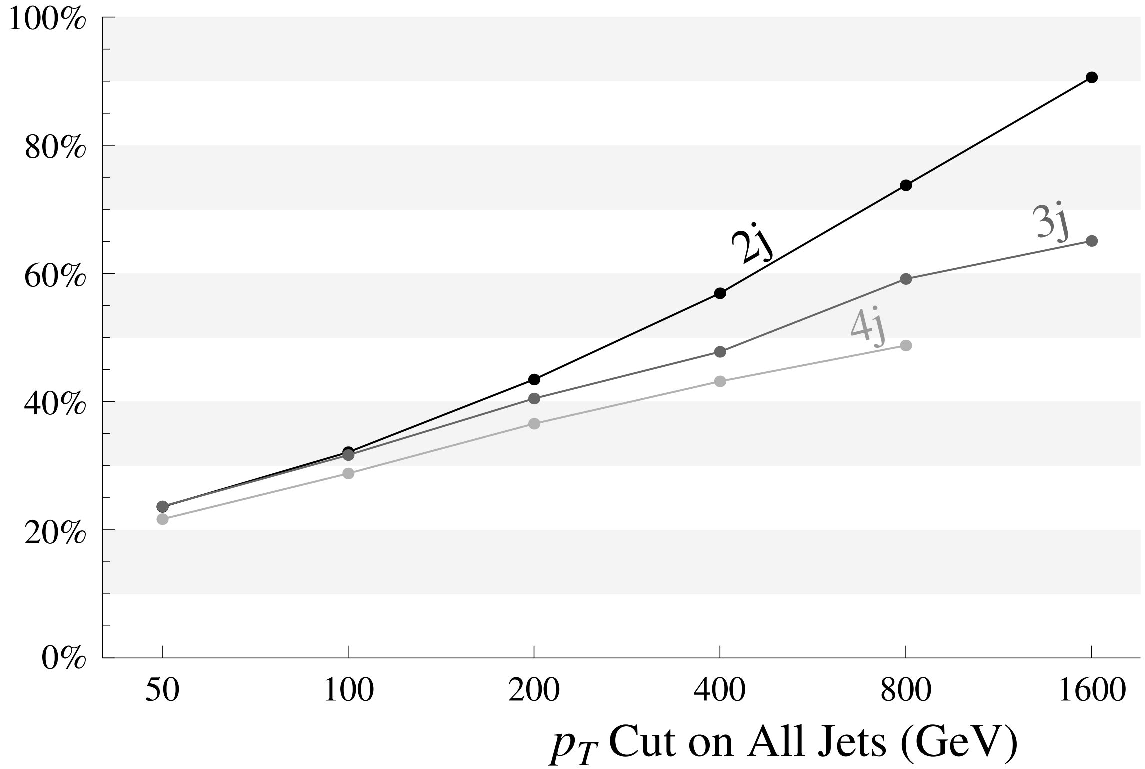
- For quarks, look at gamma + jet
  - cut on

$$\eta_{\gamma}\eta_{j1} + \Delta R_{\gamma j_2}$$

- For gluons
  - Look at b+2 jets
  - look at trijets
    - Cut on

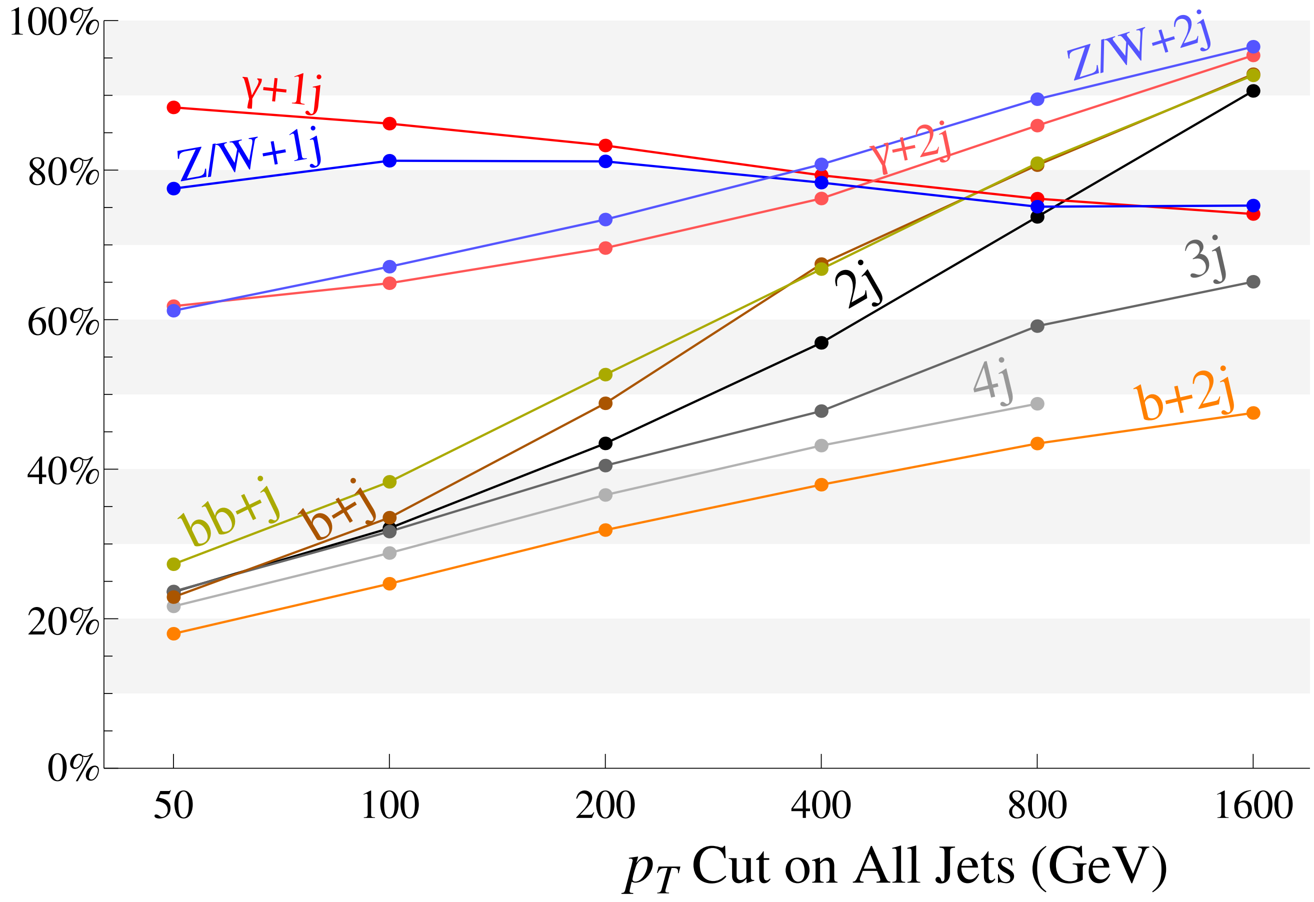
$$|\eta_{j_3}| - |\eta_{j_1} - \eta_{j_2}|$$

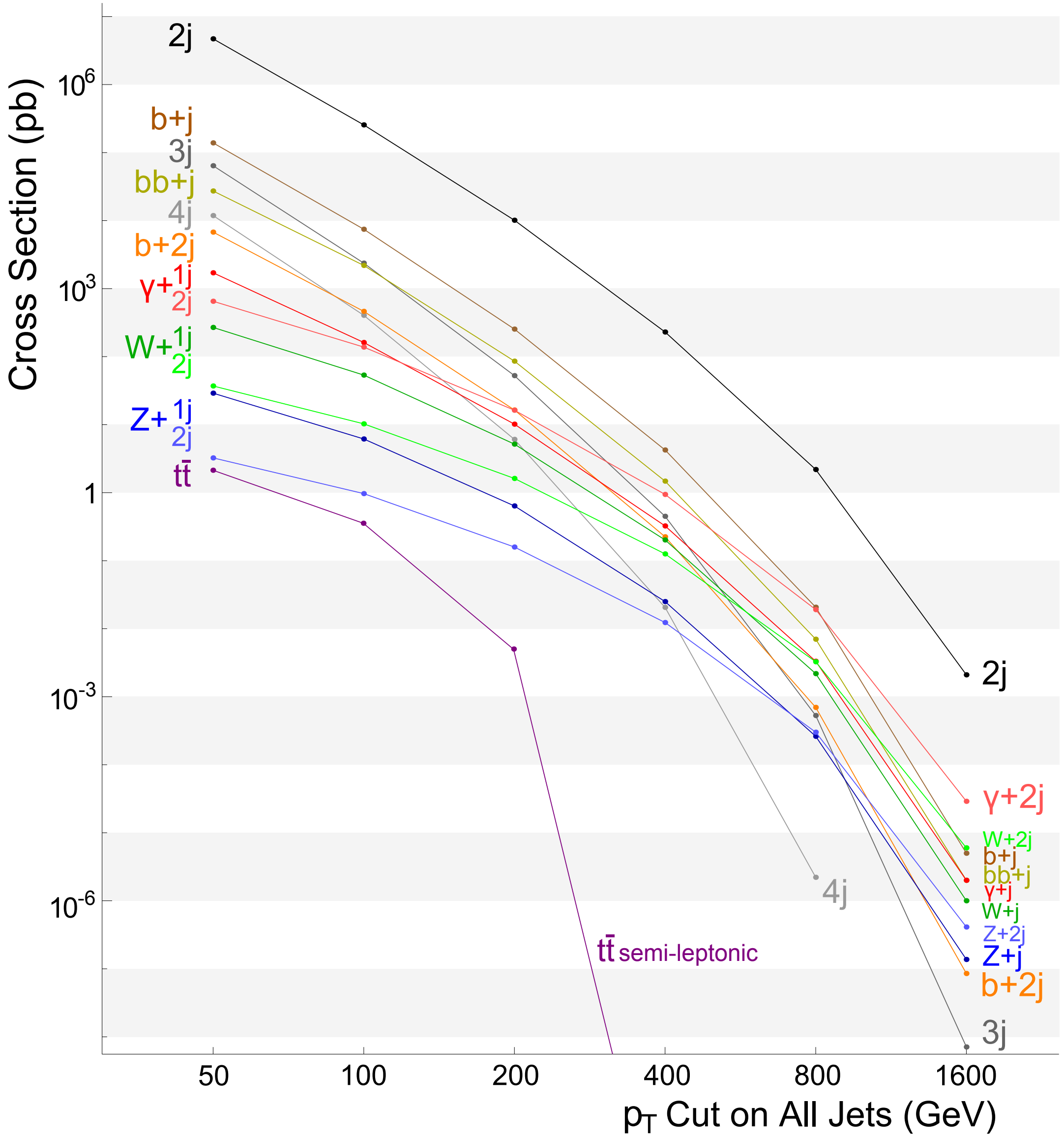
Chance EACH Jet is Quark



So chance that all 4 jets $\gtrsim 100\,GeV$ are quark $\approx (30\%)^4 \approx 1/125$

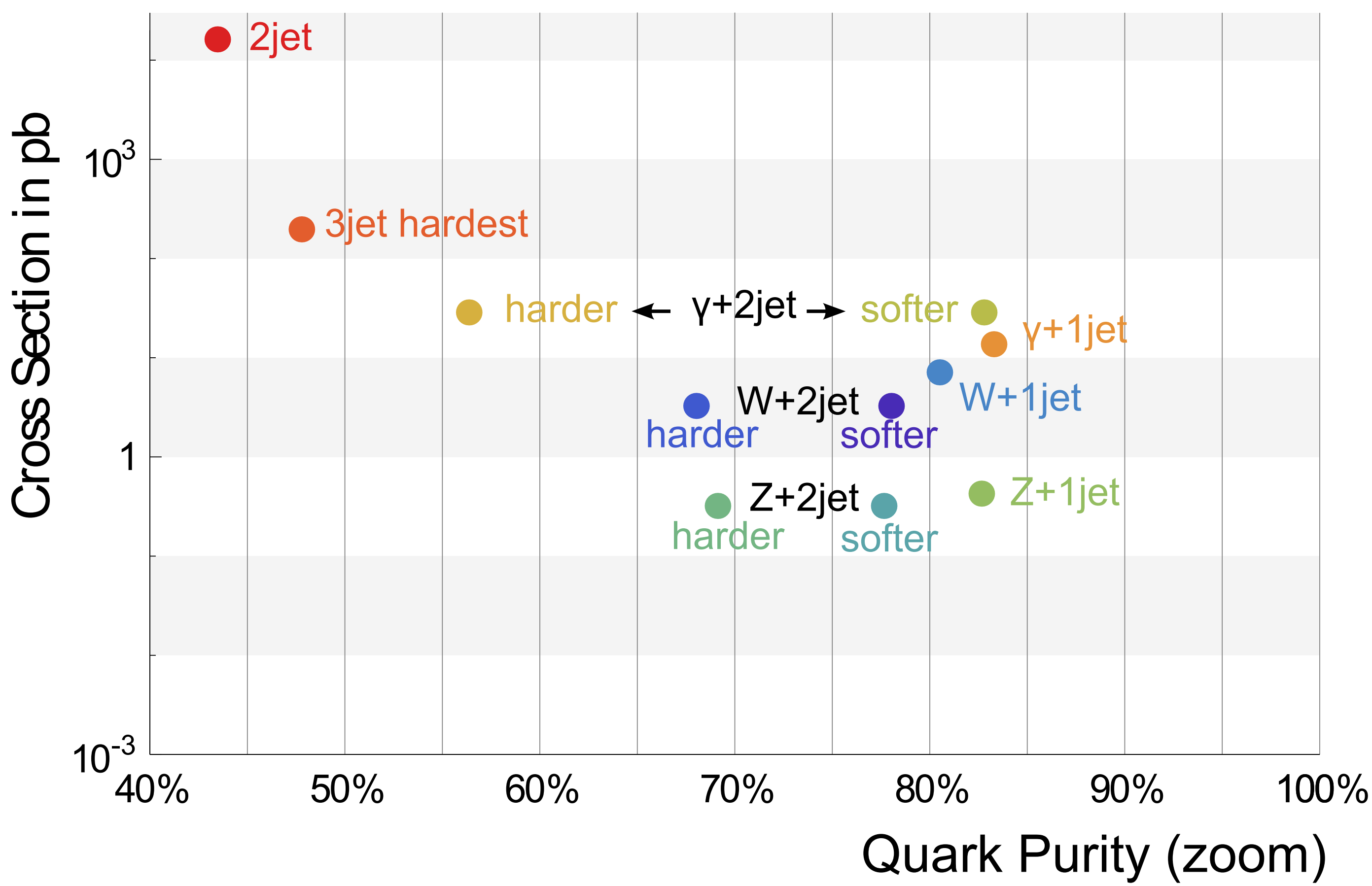
Chance EACH Jet is Quark





Cross Section (pb)
10^6
10^3
1
10^-3
10^-6
2j
b+j
3j
bb+j
4j
b+2j
γ+1j 2j
W+1j 2j
Z+1j 2j
tt̄
tt̄ semi-leptonic
2j
γ+2j
W+2j
b+j
bb+j
γ+j
W+j
Z+2j
Z+j
b+2j
3j
4j
50
100
200
400
800
1600
pT Cut on All Jets (GeV)

200 GeV Quark Purity



200 GeV Quark Purity