USE OF MULTIVARIATE TECHNIQUES AND MULTIPLE TAGGING ALGORITHMS

US ATLAS Hadronic Final State Forum

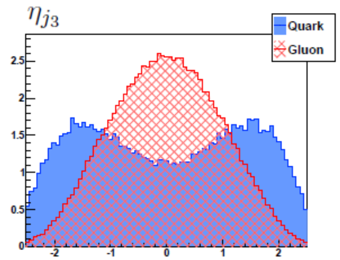
December 3, 2012

Matthew Schwartz
Harvard University

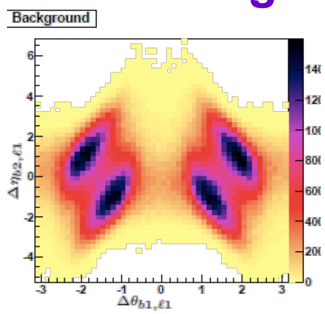
Outline

- Introduction to multivariate methods
- Applications to
  - Finding pure samples of quark and gluon jets
  - Distinguishing quark and gluon jets
- Jets and parton showers
  - Is there a optimal jet algorithm?
  - Qjets
- Conclusions
- Jet charge (time permitting)

Why use a multivariate approach?



- We can think about and visualize **single variables**



- Two variables are harder

- Multidimensional distributions are not well-suited for visualization.

- There are things that **computers are just better** at.

- Multivariate approach lets you figure out how well you could **possibly do**

EFFICIENCY
Save you the trouble or looking for good variables (project killer)

FRAMING
See if simple variables can do as well (establishes the goal)

POWER
Sometimes they are really necessary (e.g. ZH)

Multivariate (MVA) basics

Lots of methods (all in the TMVA package for root)

- Boosted Decision Trees
- Artificial Neural Networks
- Fischer Discriminants
- Rectangular cut optimization
- Projective Likelihood Estimator
- H-matrix discriminant
- Predictive learning/Rule ensemble
- Support Vector Machines
- K-nearest neighbor
- …

Useful in many areas of science, such as artificial intelligence

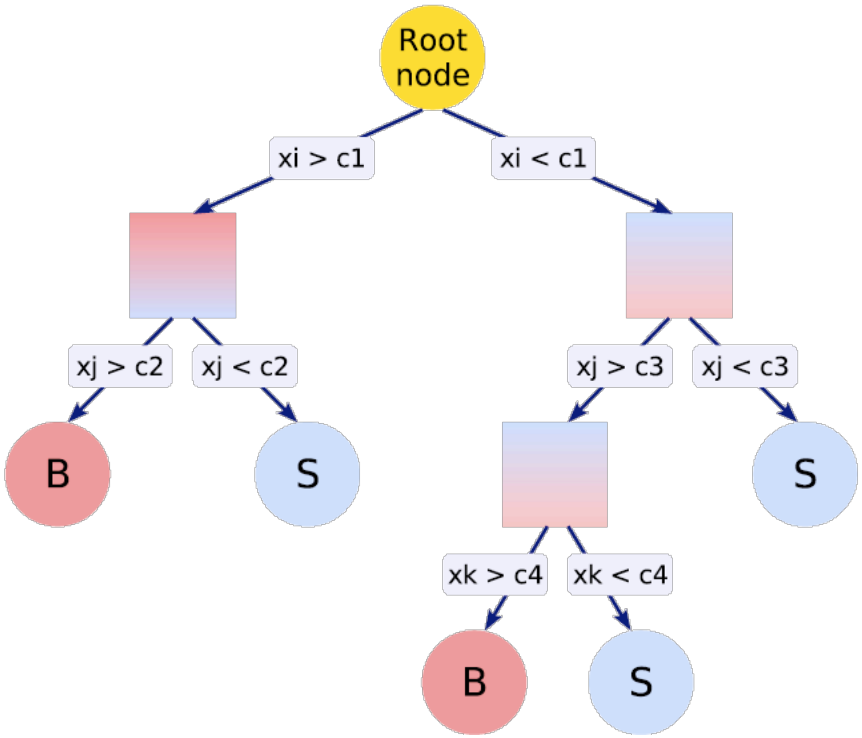
For particle physics, **Boosted Decision Trees** work best

Easy to understand

Train fast

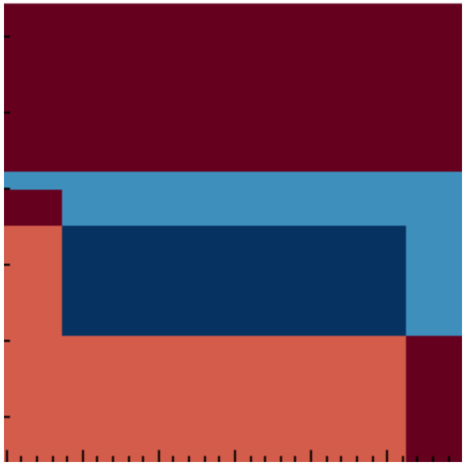
Nearly optimal efficiencies

Boosted Decision trees



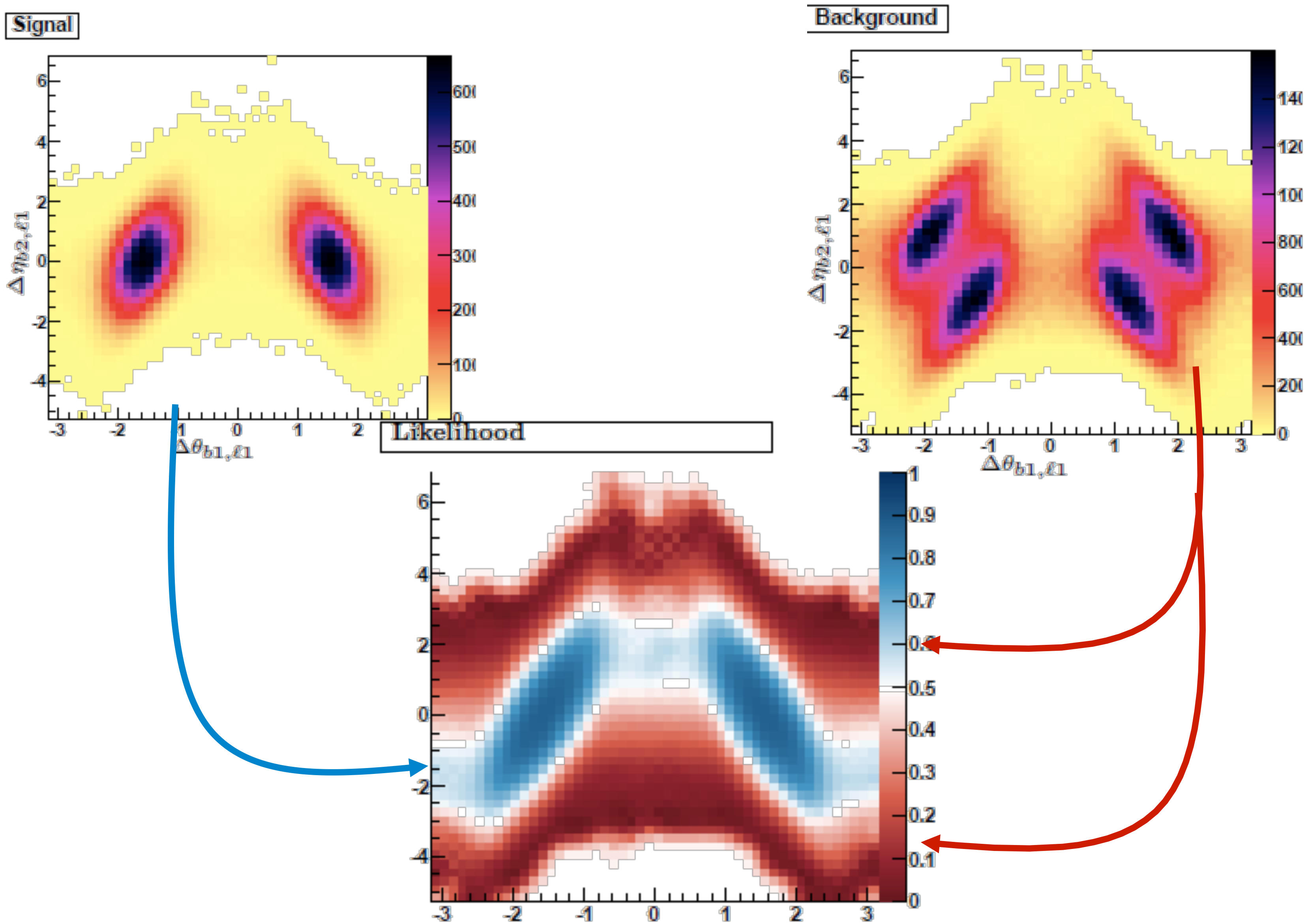
One decision tree

Two decision trees

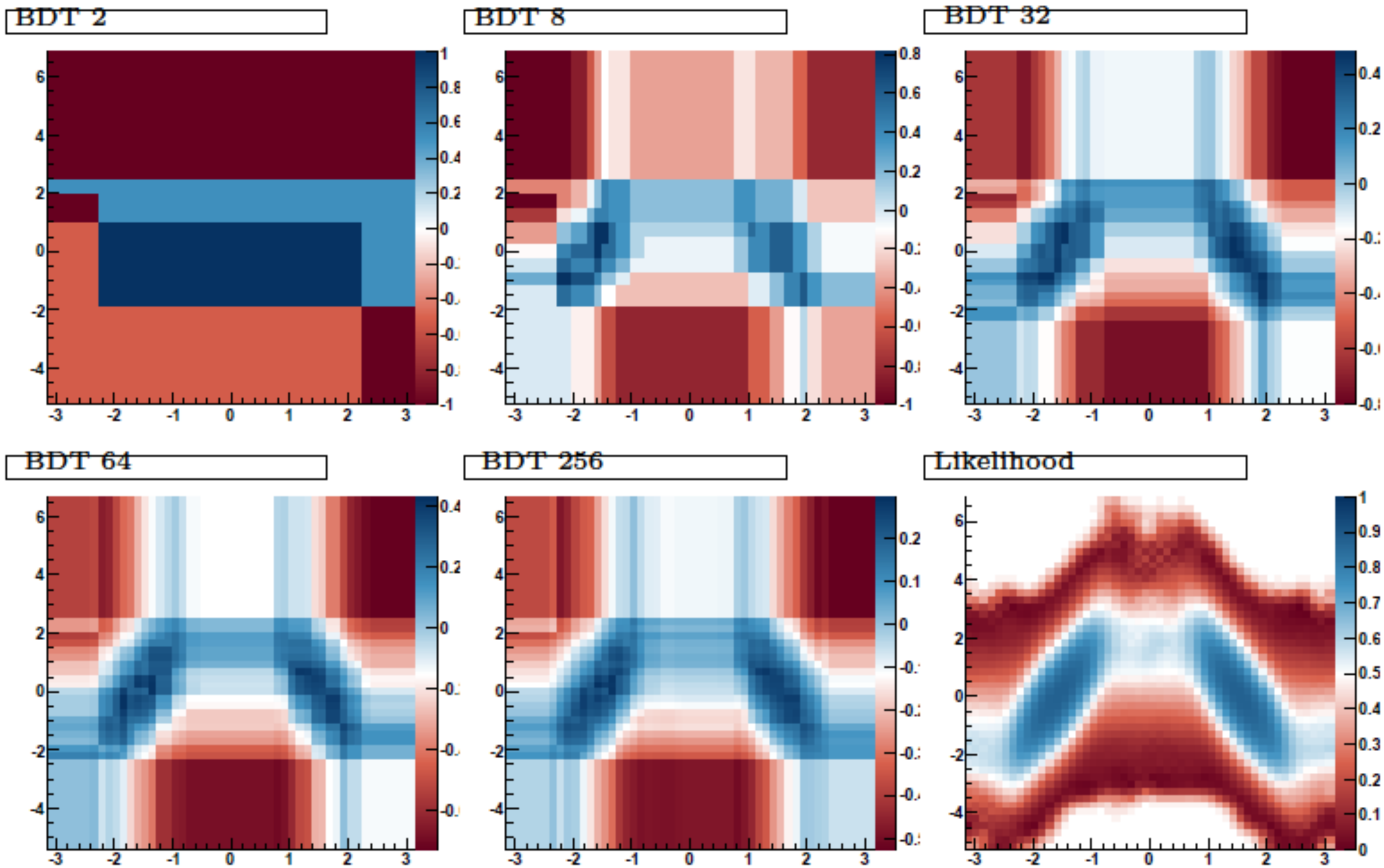


•**Boosting** : train successive trees on misclassified events by enhancing their importance

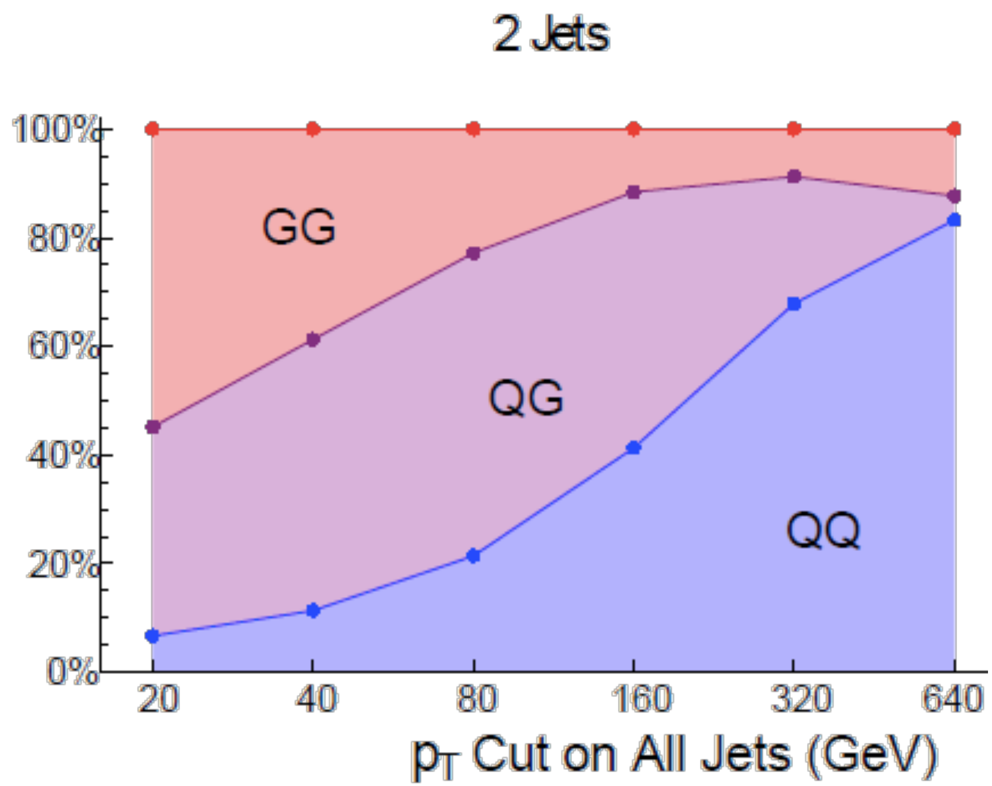
Exact solution: likelihood



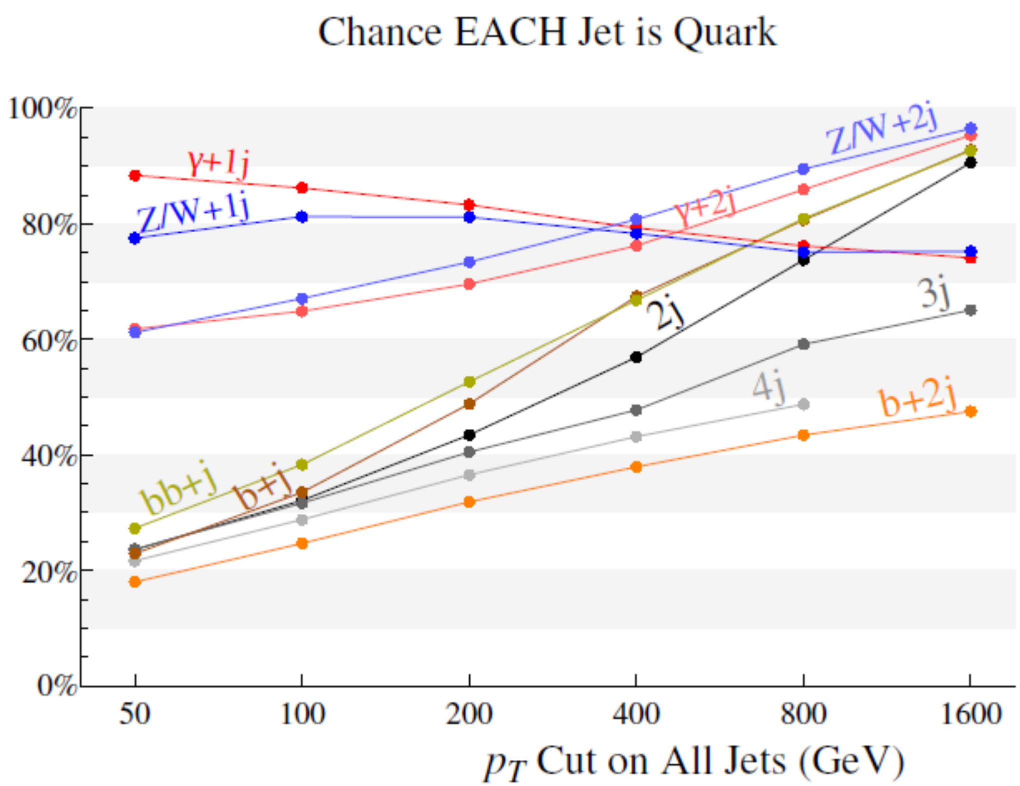
Multidimensions: approximate



Example: where are the quark jets?



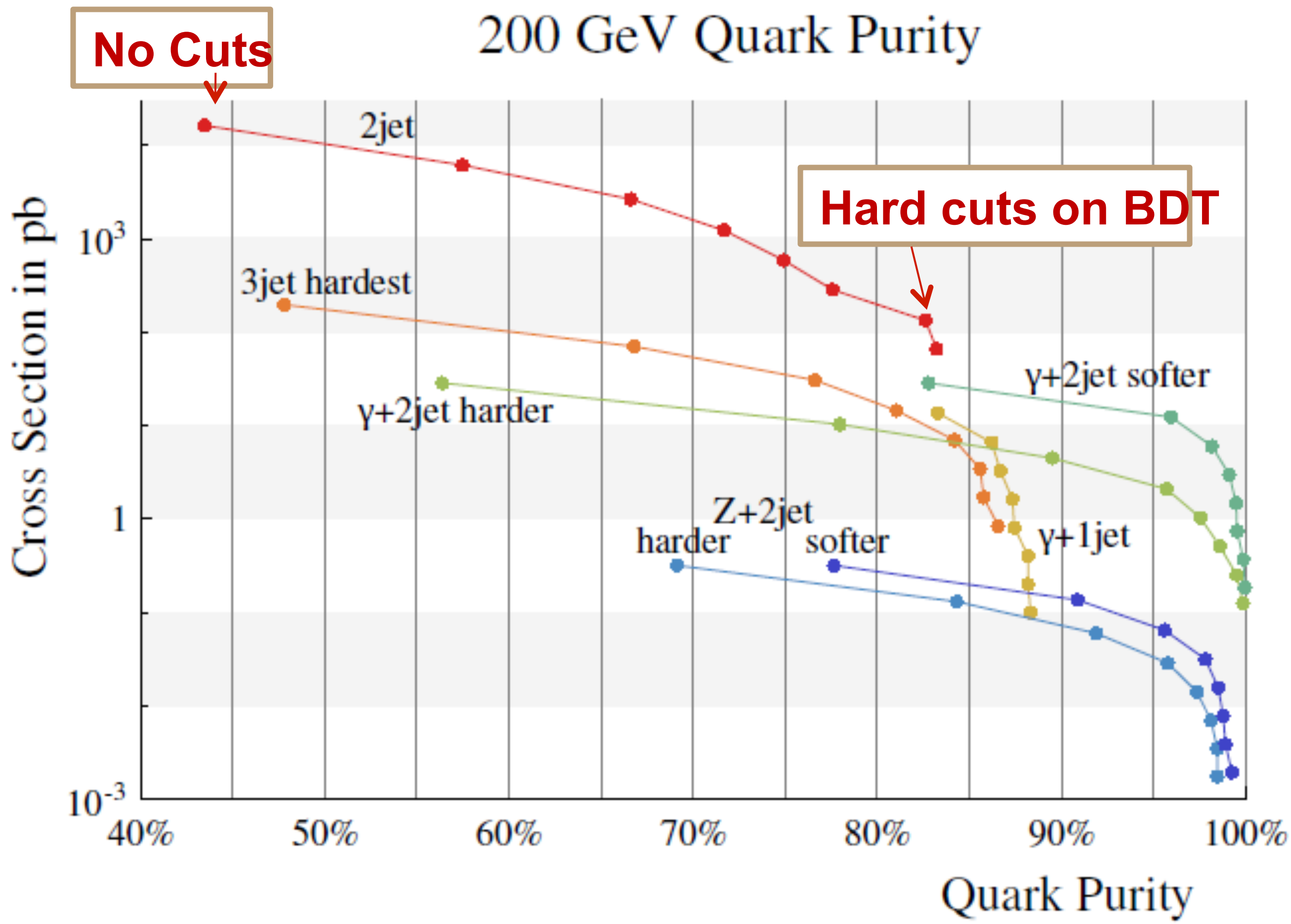
Look at all samples



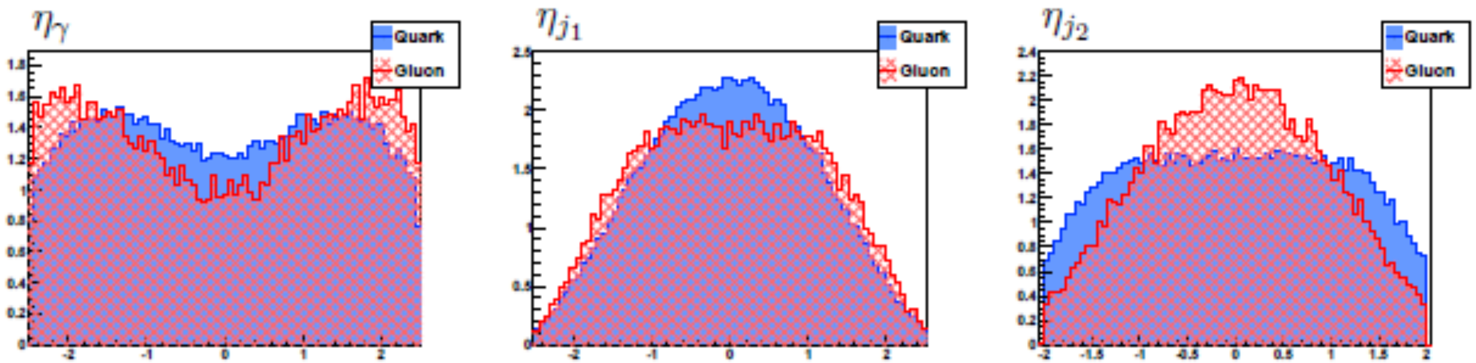
•Can cuts purify the samples?

Throw them into the BDT

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



Now look at the $\gamma + 2$ jets sample

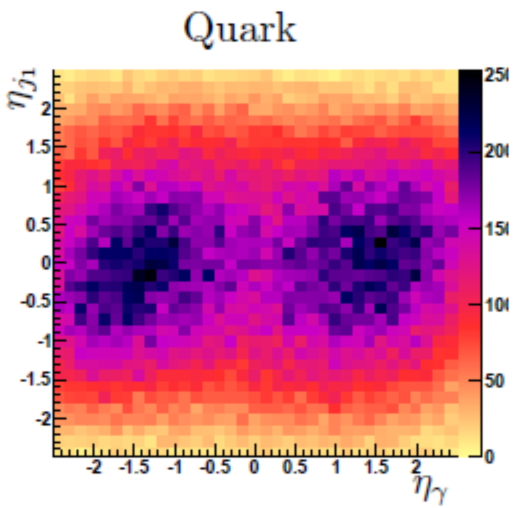


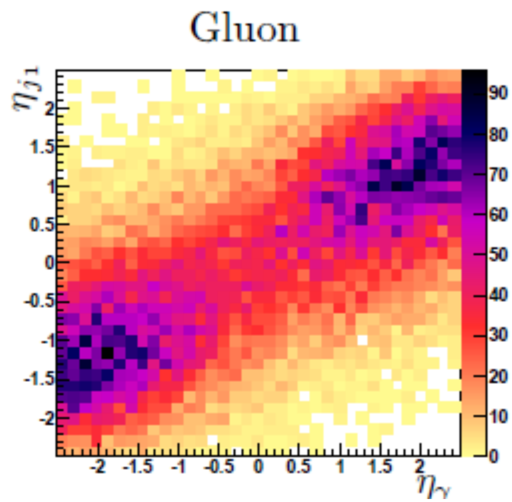
- Look at the best discriminants, ranked by cuts

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

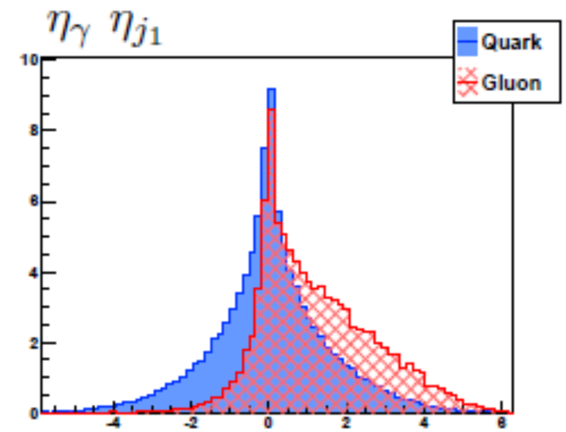
- But cutting on just η_γ or just η_{J2} does not help much

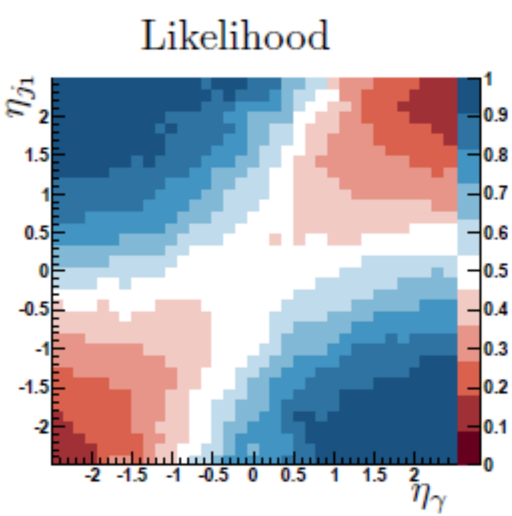
Look at correlations



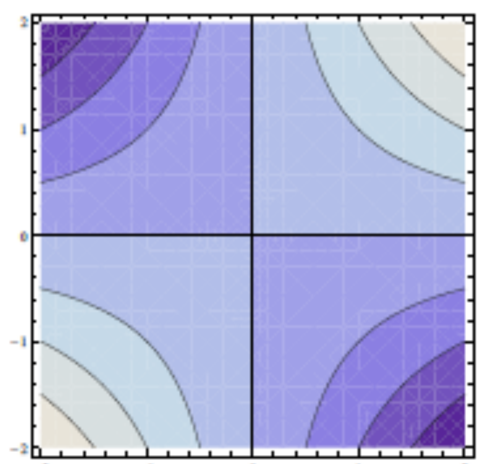


Distribution of $\eta_\gamma \eta_{j1}$

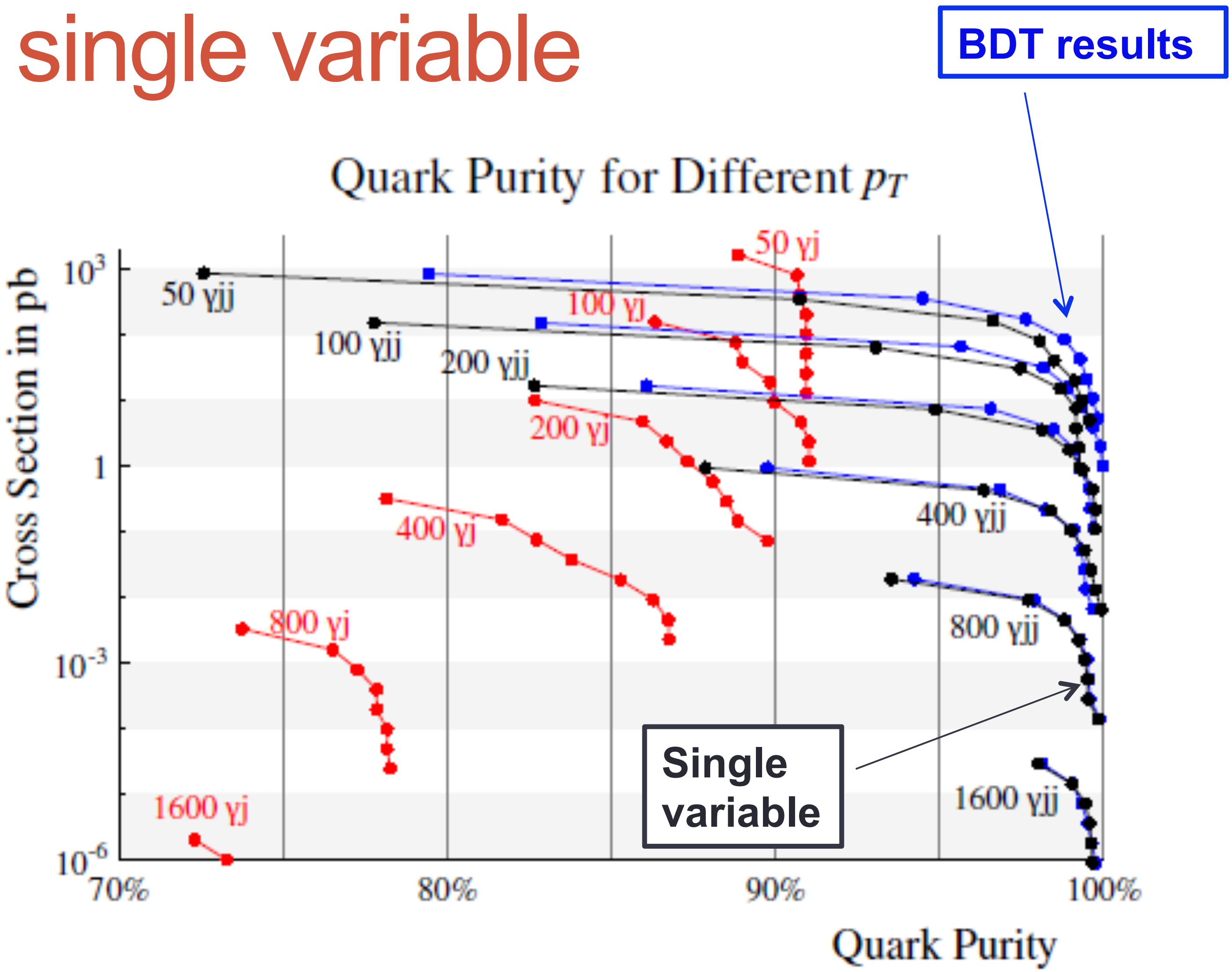




Contours of $\eta_\gamma \eta_{j1}$



Best single variable



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

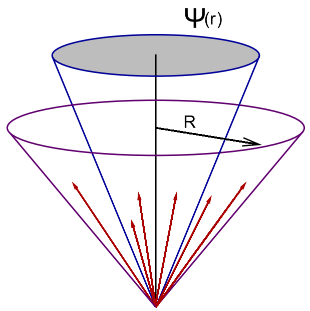
Pure quark and gluon samples

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

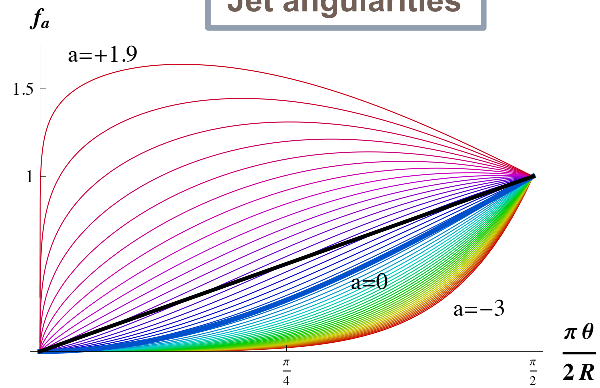
Efficiencies and other possibilities discussed in Gallicchio and MDS 1104.1175 (JHEP)

Quark and gluon jet substructure

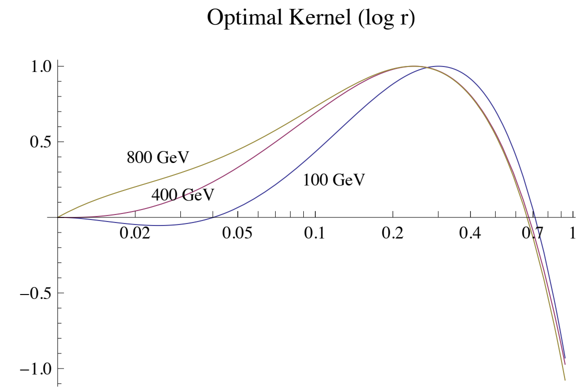
Integrated/differential "Jet Shape"



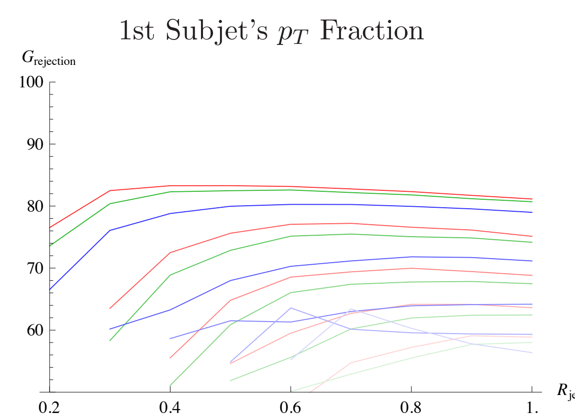
Jet angularities



Iteratively optimized radial profile

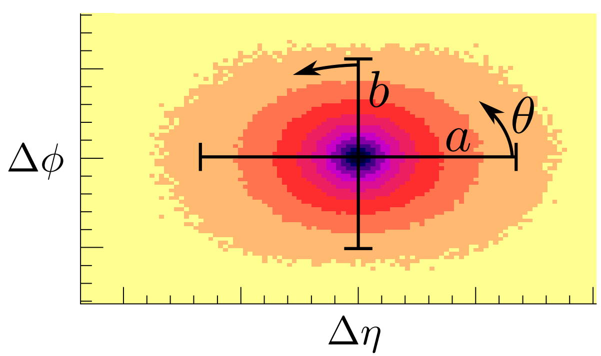


Subjet counts and properties



Properties of Covariance tensor

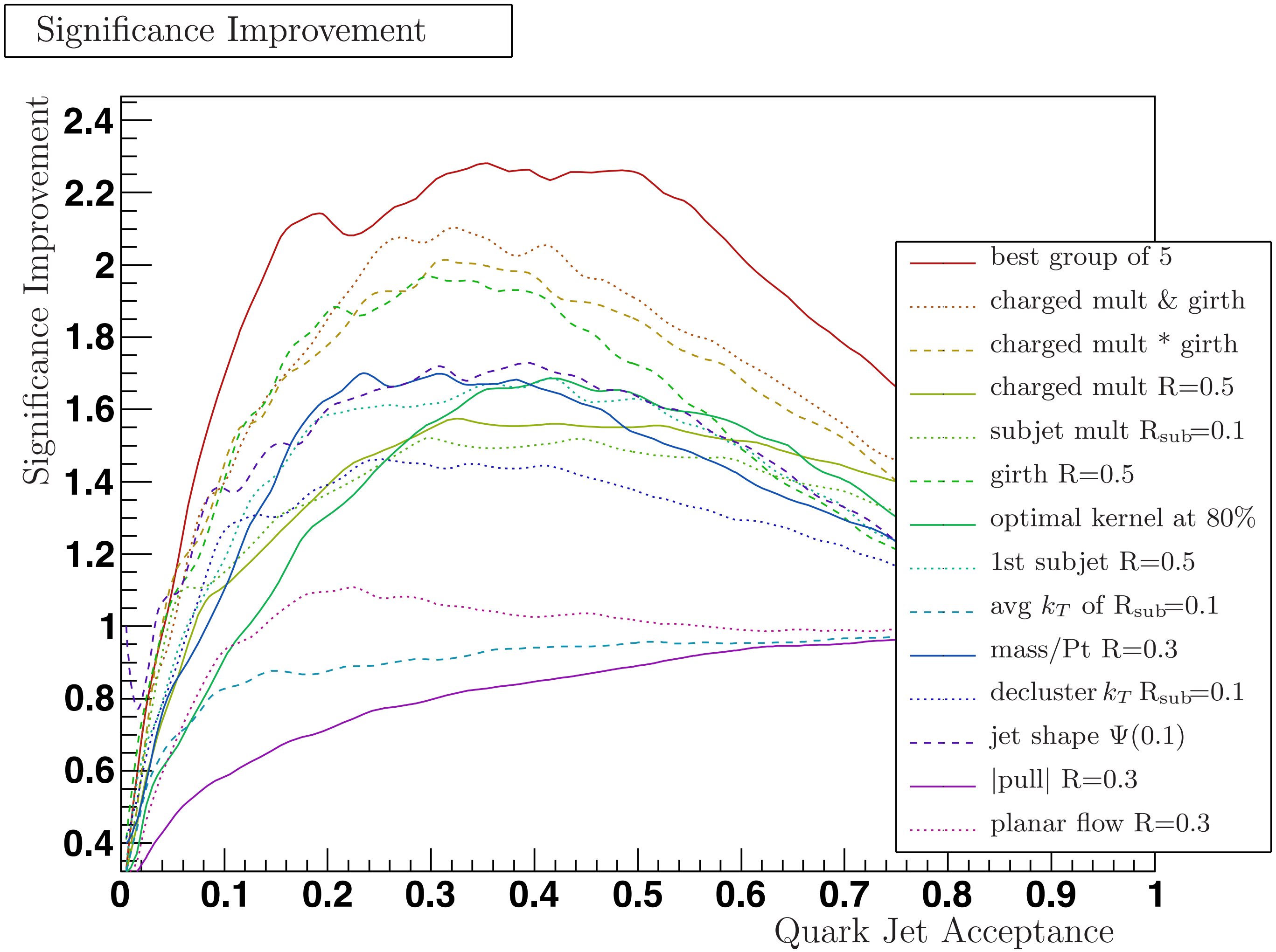
$$\mathbf{C} = \sum_{i\in\text{jet}} \frac{p_T^i}{p_T^{jet}} \begin{pmatrix} \Delta\eta_i\Delta\eta_i & \Delta\eta_i\Delta\phi_i \\ \Delta\phi_i\Delta\eta_i & \Delta\phi_i\Delta\phi_i \end{pmatrix}$$



Combination of Eigenvalues

Eigenvalues:	$a > b$
Quadratic Moment:	$g = \sqrt{a^2 + b^2}$
Determinant:	$det = a \cdot b$
Ratio:	$\rho = b/a$
Eccentricity:	$\epsilon = \sqrt{a^2 - b^2}$
Planar Flow:	$pf = \frac{4ab}{(a+b)^2}$
Orientation:	θ

Quark and gluon jet substructure



Quark and gluon tagging: results

	Gluon Efficiency % at 50% Quark Acceptance	**50 GeV** Particles P8	**50 GeV** Particles H++	**50 GeV** Tracks P8	**50 GeV** Tracks H++	**200 GeV** Particles P8	**200 GeV** Particles H++	**200 GeV** Tracks P8	**200 GeV** Tracks H++
Single variables	2-Point Moment β=1/5	8.7*	17.8*	13.7*	22.8*	8.3	15.9	13.2	19.6
	1-Subjettiness β=1/2	9.3	18.5	14.2	22.9	7.6	16.2	12.3	19.4*
	2-Subjettiness β=1/2	9.2	18.6	13.9	23.6	6.8	15.7*	9.8	18.7
	3-Subjettiness β=1	9.1	19.3	14.6	24.4	5.9*	16.7	8.6*	19.5
	Radial Moment β=1 (Girth)	10.3	20.5	16.1	24.9	11.2	18.9	15.3	21.9
	Angularity $a = +1$	10.3	20.0	15.8	24.5	12.0	19.3	14.0	21.6
	Det of Covariance Matrix	11.2	21.2	18.1	27.0	9.4	20.9	13.5	24.6
	Track Spread: $\sqrt{< p_T^2 >}/p_T^{\text{jet}}$	16.5	25.3	16.5	25.3	9.3	20.1	9.3	20.1
	Track Count	17.7	26.4	17.7	26.4	8.9	21.0	8.9	21.0
	Decluster with k_T, ΔR	15.8	24.5	20.1	28.4	13.9	20.1	16.9	23.4
	Jet m/p_T for R=0.3 subjet	13.1	25.9	16.3	27.7	11.9	24.2	14.8	26.2
	Planar Flow	28.7	34.4	28.7	34.4	39.6	42.9	39.6	42.9
	Pull Magnitude	37.0	39.0	32.9	35.6	30.6	30.2	29.6	30.6
Pairs of variables	Track Count & Girth	9.9	20.1	13.4	23.2	7.1	17.3	7.7*	18.7
	R=0.3 m/p_T & R=0.7 2-Point β=1/5	7.9*	17.7	12.2*	22.1	5.7	14.4*	8.5	17.9
	1-Subj β=1/2 & R=0.7 2-Point β=1/5	8.5	17.3*	12.9	22.1	6.0	14.6	8.6	17.7*
	Girth & R=0.7 2-Point β=1/10	12.6	21.9	12.6	21.9*	9.2	18.0	9.2	18.0
	1-Subj β=1/2 & 3-Subj β=1	8.9	18.0	14.0	23.2	5.6*	15.0	8.4	18.4
3,4,5 variables	Best Group of 3	7.5	17.0	11.0	20.9	4.7	14.0	6.9	16.6
	Best Group of 4	7.1	16.7	10.6	20.5	4.5	13.7	6.2	16.3
	Best Group of 5	6.9	16.4	10.4	20.0	4.3	13.3	6.1	15.9

Gallicchio and MDS arXiv:1211.7038

Quark and Gluon tagging

- Discrimination easier at **higher p_T**
- Using **all particles** works better than just **charged tracks**
- **Pythia** gives bigger Q/G difference than **Herwig**
- 80-90% gluon rejection at 50% quark acceptance is realistic
- **Data** on single variable and multivariate distributions would be **very useful**

Jet grooming

Basic idea: remove soft radiation which is not collinear

Filtering (Butterwort et al 2008)

- Recluster fat jet into R=0.3 subjets
- Keep 3 hardest subjets

- ❑ Boosted Higgs
- ❑ Boosted top

Trimming (Krohn et al 2008)

- Recluster fat jet into R=0.3 subjets
- Keep subjets which have energy > 5% jet energy

- ❑ Parton momentum reconstruction
- ❑ Pileup removal

Pruning (Ellis et al 2008)

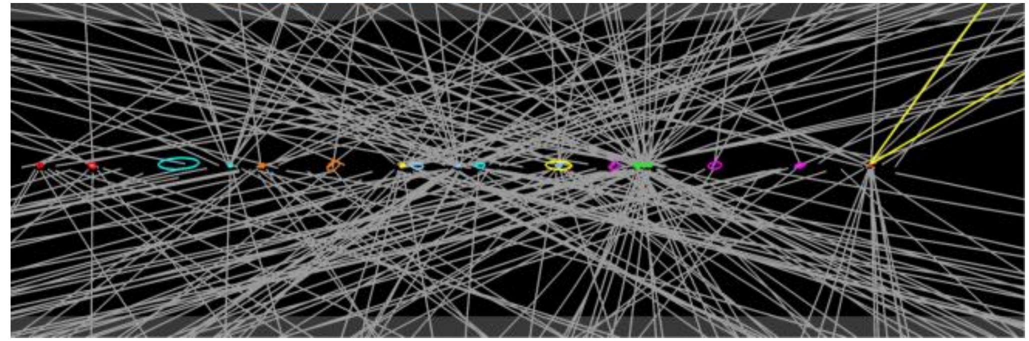
- Undo clustering steps
- Cluster 1 with 2 if
  - E_1 , E_2 > 0.1 (E_1+E_2)
  - or $R_{12} < 0.2$
  - otherwise, drop softer of 1,2

- ❑ Jet mass searches
- ❑ Qjets

- ❑ All help with jet substructure
- ❑ All help with pileup removal

Trimming

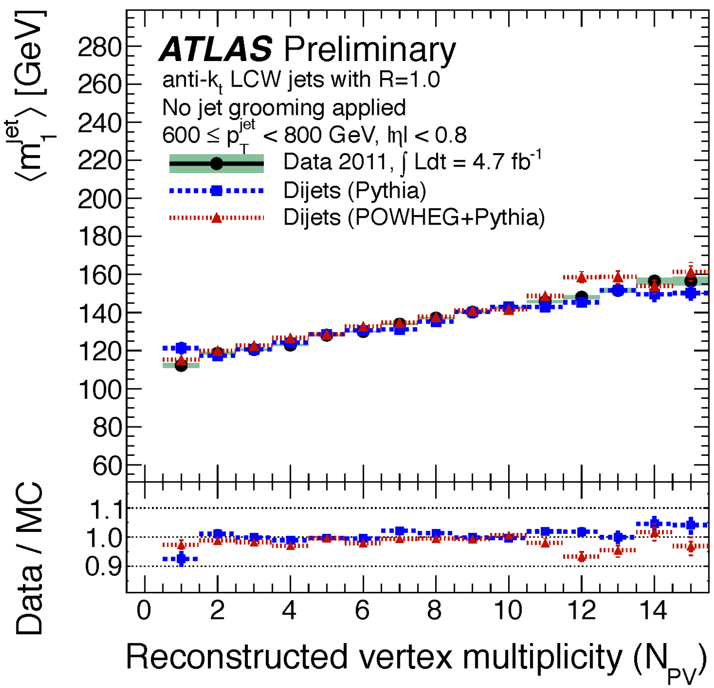
Helps experimentally
with pileup subtraction



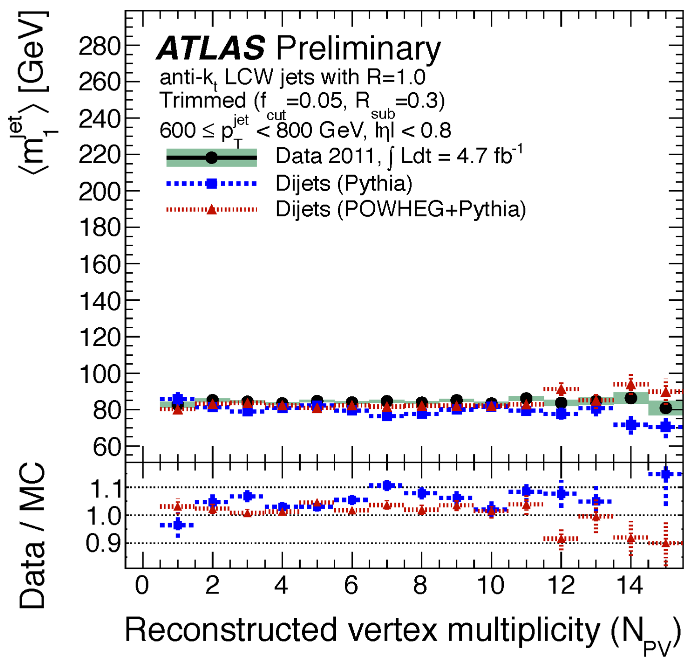
$N_{PV} = 30$

Jet mass dependence
On $N_{\text{Pileup-Vertices}}$

Before trimming

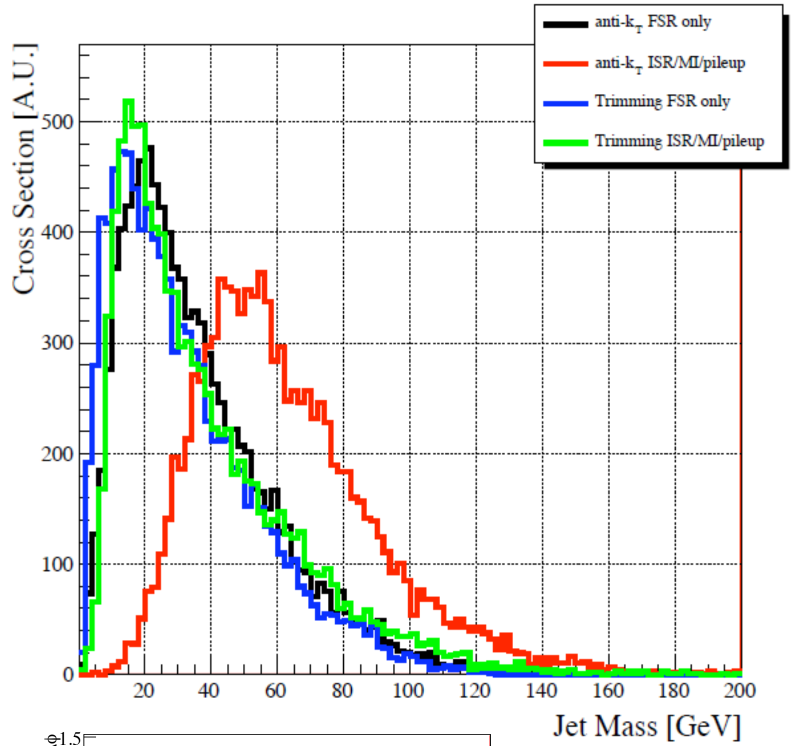


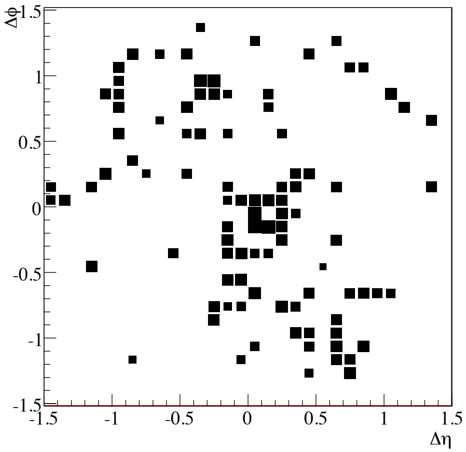
After trimming



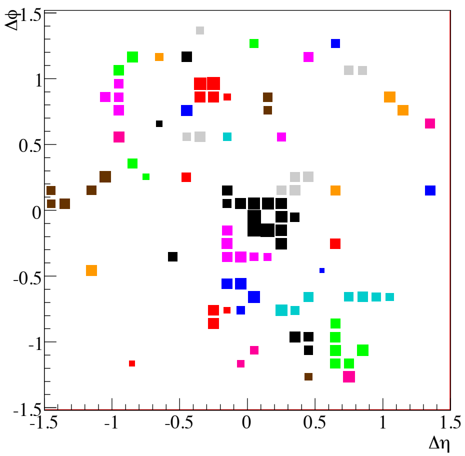
Trimming

1) Make seed jet with anti-k_T (R_0 large)
2) Recluster into subjets with k_T (R_0 small)
3) Remove subjets if $p_T < f_{cut}\,\Lambda_{hard}$
4) Kept subjets give trimmed jet

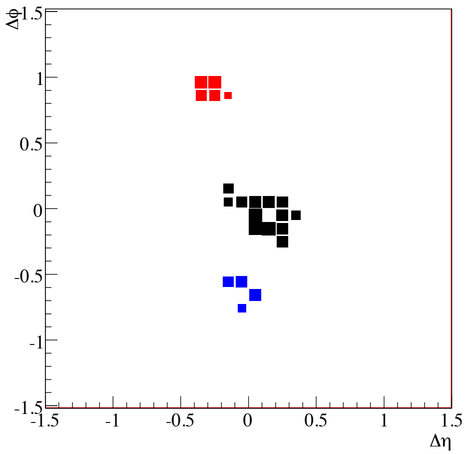




After anti-k_T
$R_0 \sim 1.5$



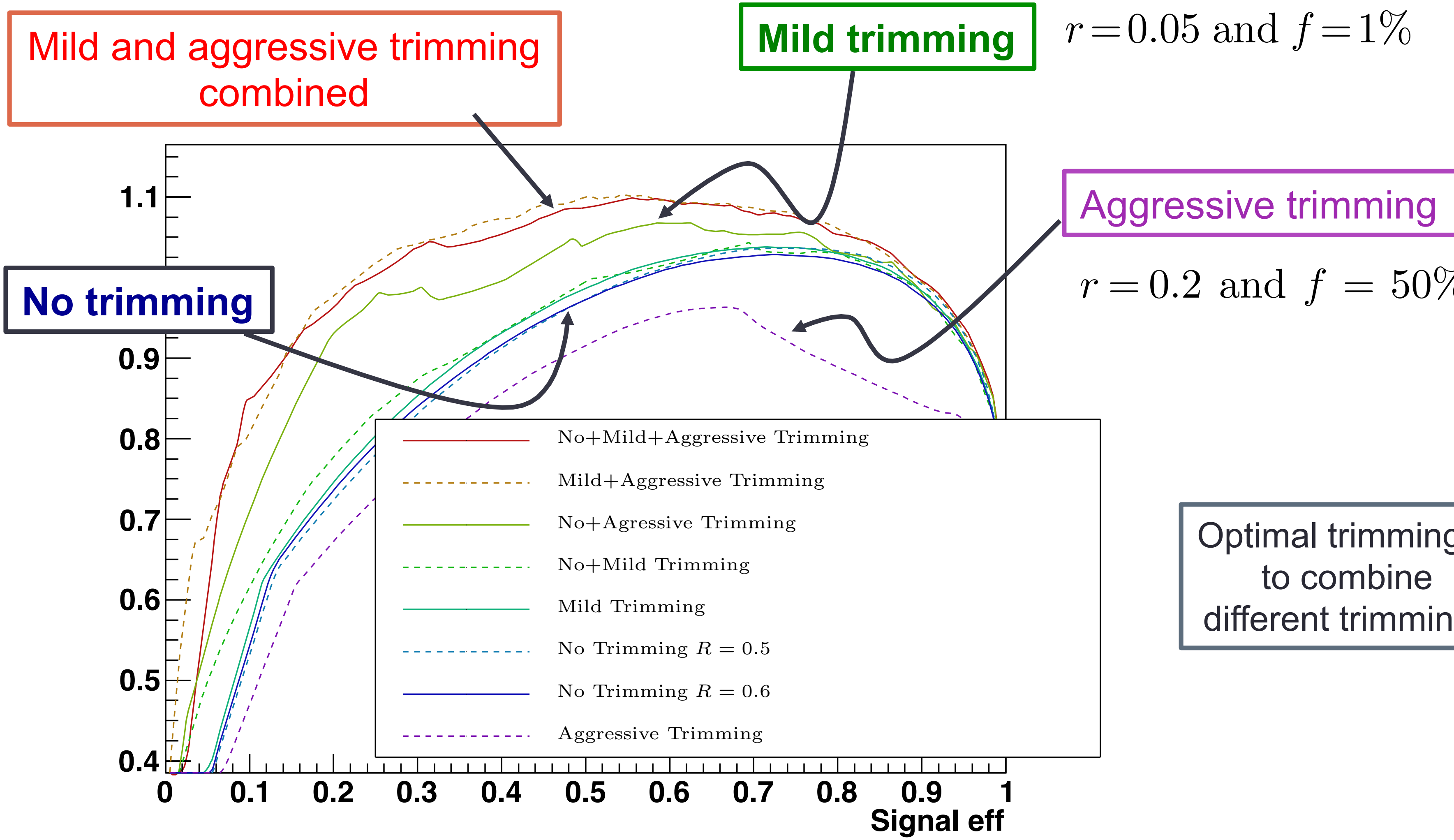
After k_T
$R_0 \sim 0.2$



No subjets below
1% of total event p_T

Interesting MVA result

Black, Gallicchio, Huth, Kagan, Schwartz, Tweedie arXiv:1010.3698 (JHEP)

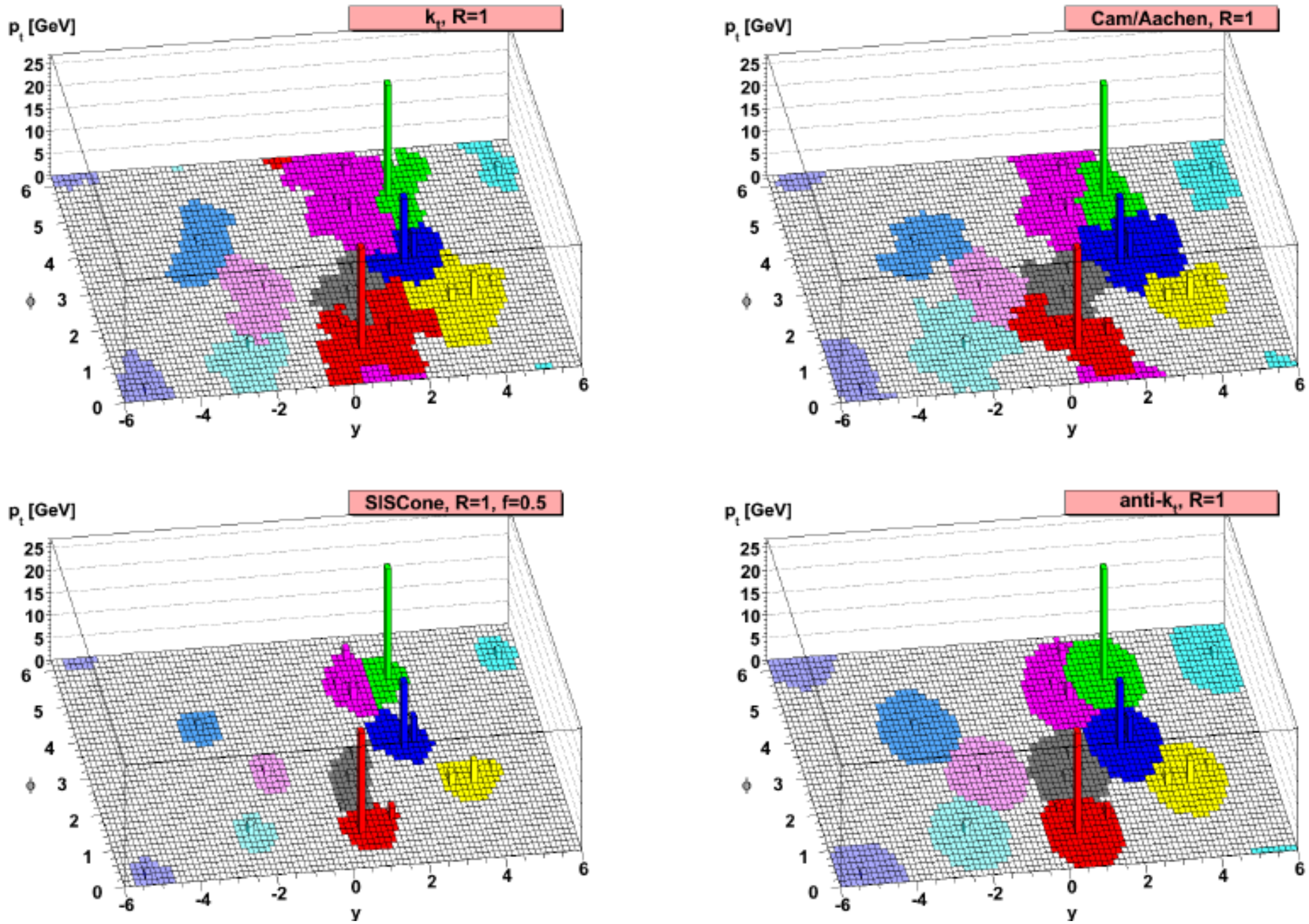


Optimal trimming is to combine different trimmings

Optimal trimming is not “pure” trimming:

$$|\mathrm{trim}_1\rangle + |\mathrm{trim}_2\rangle$$ works better than $|\mathrm{trim}_1\rangle$ or $|\mathrm{trim}_2\rangle$

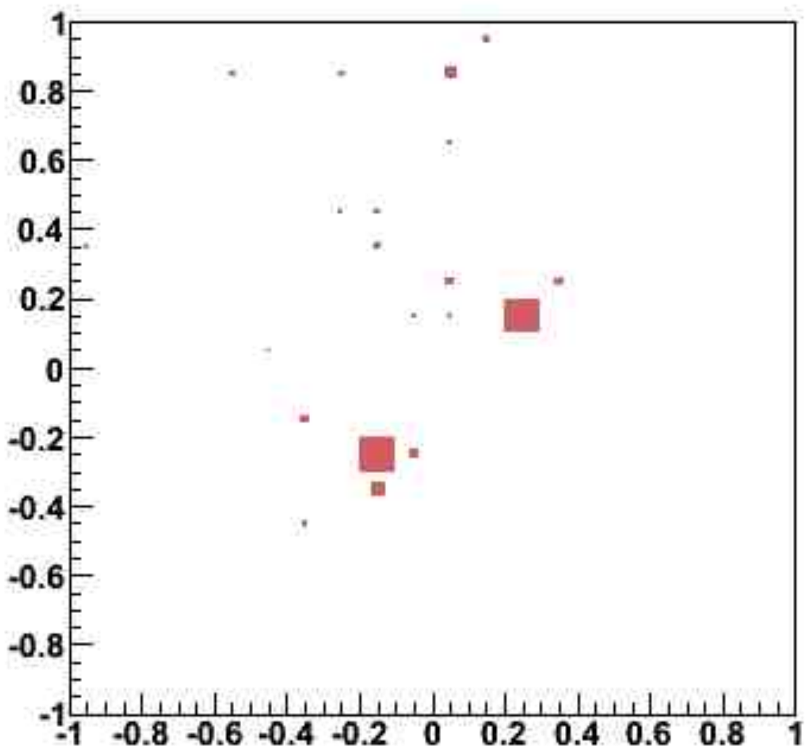
Different algorithms, different results

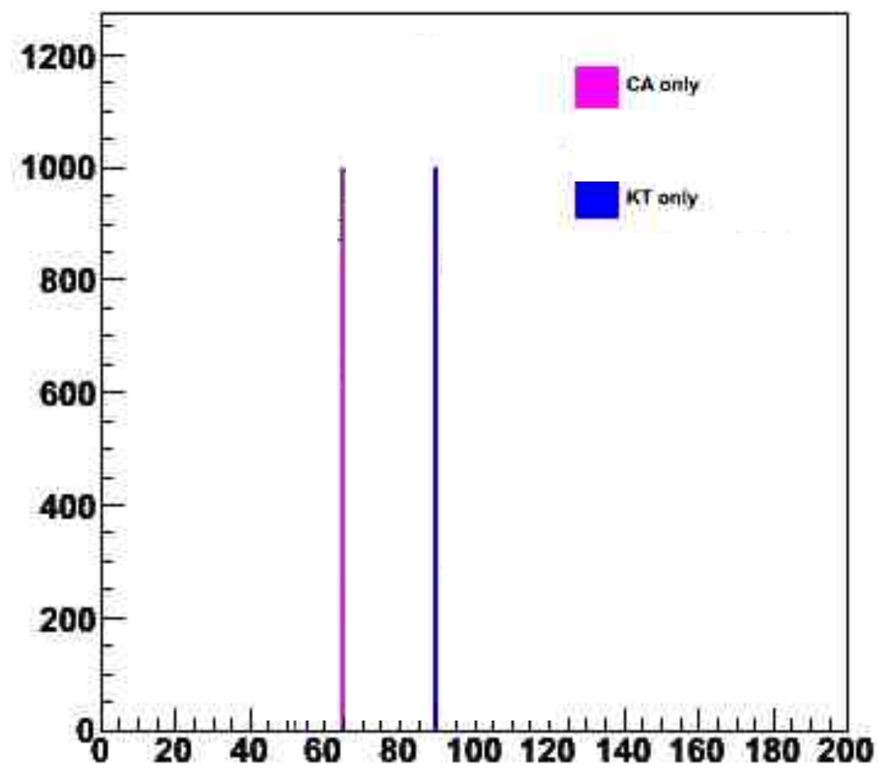


Cacciari, Salam, Soyez JHEP 0804:063 (2008)

e.g. reconstruct W invariant mass

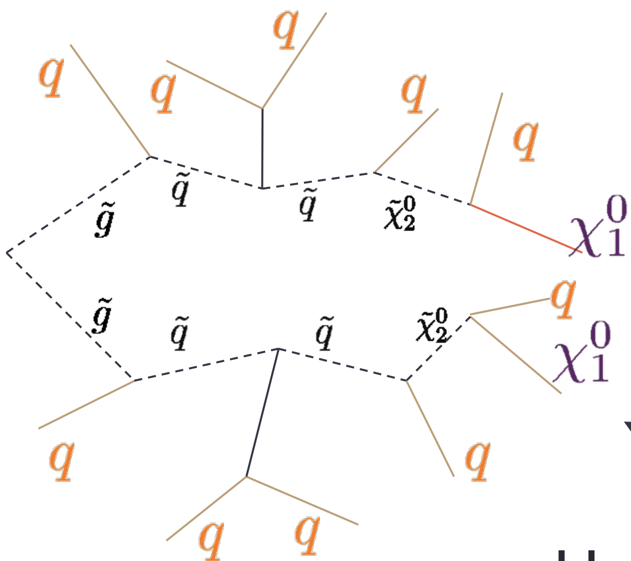
$$W \to \bar{q}q$$





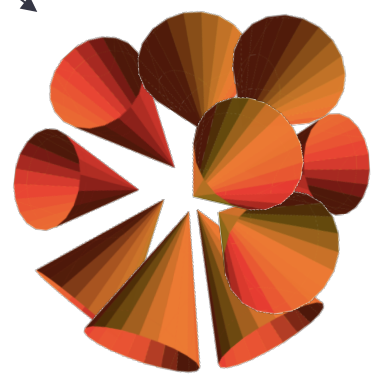
Jet-to-parton map

We want to see quarks and gluons:



Parton shower

We observe jets:



+ missing energy

How can we **invert** ?

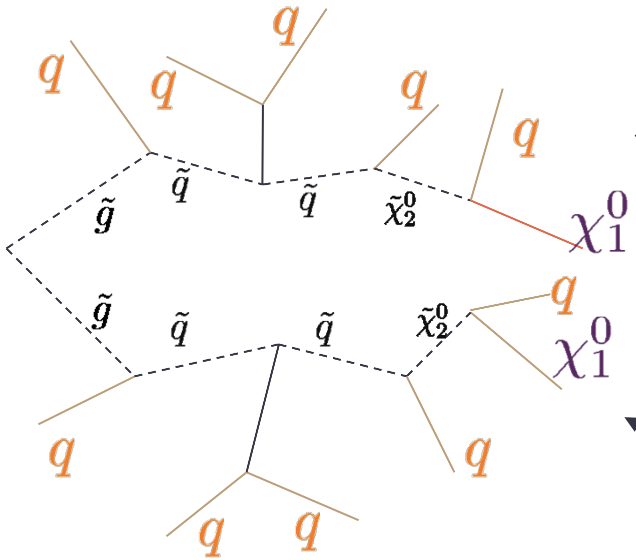
$\Downarrow$

- Find jet momenta
- Set quark momenta = jet momenta

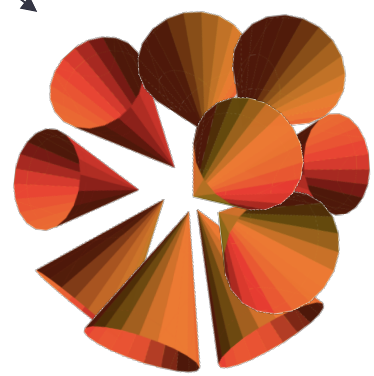
Jet-to-parton map

We want to see quarks and gluons:

We observe jets:



Parton shower



+ missing energy

Jet algorithms

Reality: **this** exists

Assumption: **this** exists

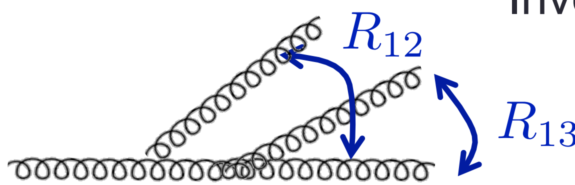
Parton-shower is ***not* invertible**

Different distance measures

Cambride/Aachen algorithm

$$d_{ij} = \left(\frac{R_{ij}}{R_0}\right)^2$$

- clusters closest radiation first

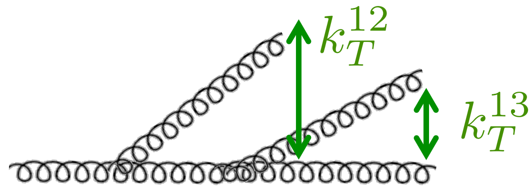


Inversion of **Herwig shower**

k_T algorithm

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

- clusters hard collinear radiation first



Inversion of **Pythia shower**

anti k_T algorithm

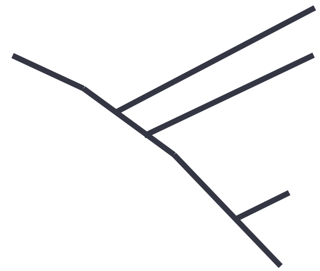
$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2})\left(\frac{R_{ij}}{R_0}\right)^2$$

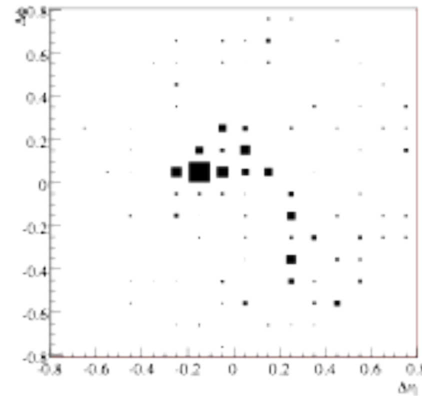
- Clusters farthest first
- No inverse parton-shower interpretation

- Produces round jets
- Experiment friendly

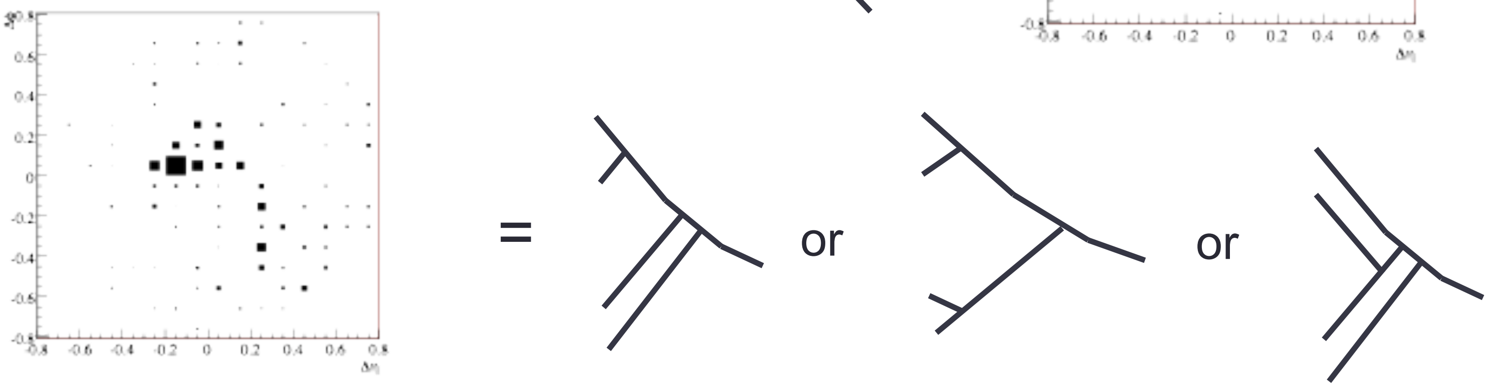
Parton shower is not invertible

Parton shower gives an event





What is the **inverse**?



- Is there a way to have "fuzzier" jets
  which account for non-unique inverse?

One possibility: Qjets

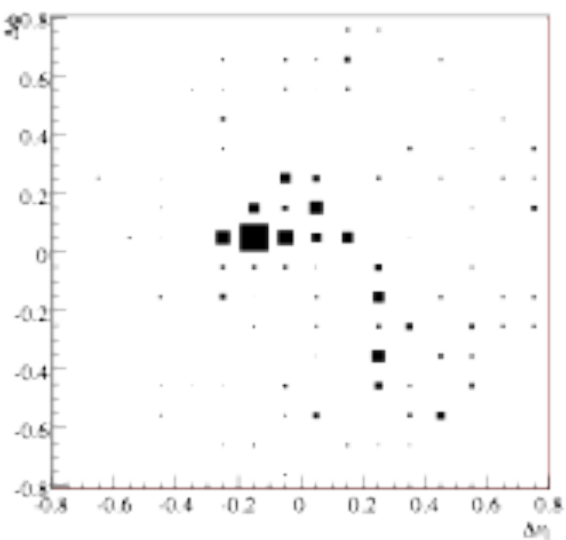
Ellis,Hornig, Krohn, Roy, MDS
arXiv:1201.1914 (PRL)

Add randomness into the jet algorithm

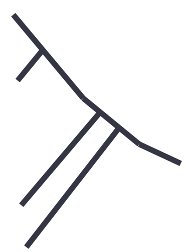
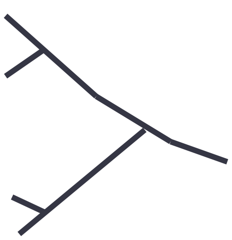
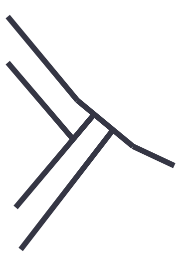
Instead of choosing smallest d_{ij}, choose pair with a probability

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

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



$$|\psi\rangle = \sum_j a_j |\mathrm{tree}_j\rangle$$

=  and  and 

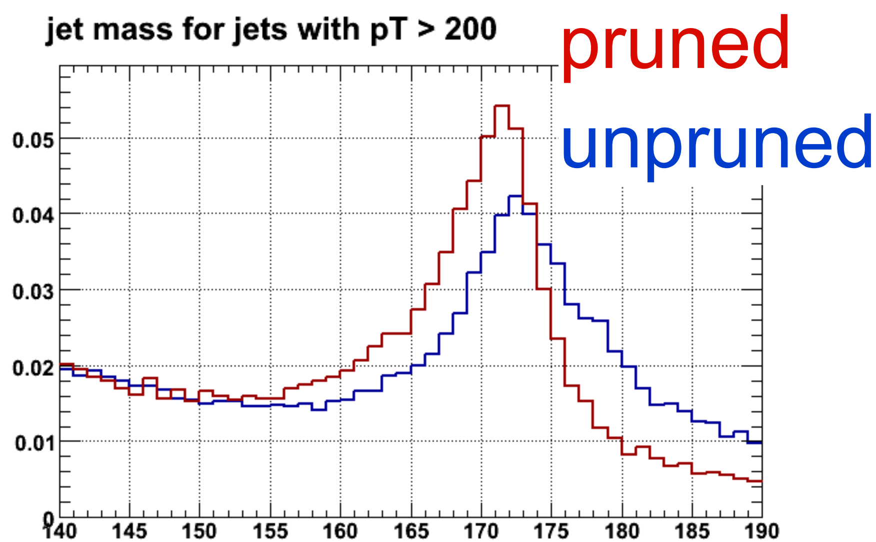
What do we do with the Qjets?

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

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

$$z_{ij} \equiv \frac{\min\left(p_{T_i}, p_{T_j}\right)}{|\vec{p_{T_i}} + \vec{p_{T_j}}|} < z_{\text{cut}} \qquad \Delta R_{ij} > D_{\text{cut}}$$

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



Pruned Qjets

Event with a boosted W boson

This is ***one*** event

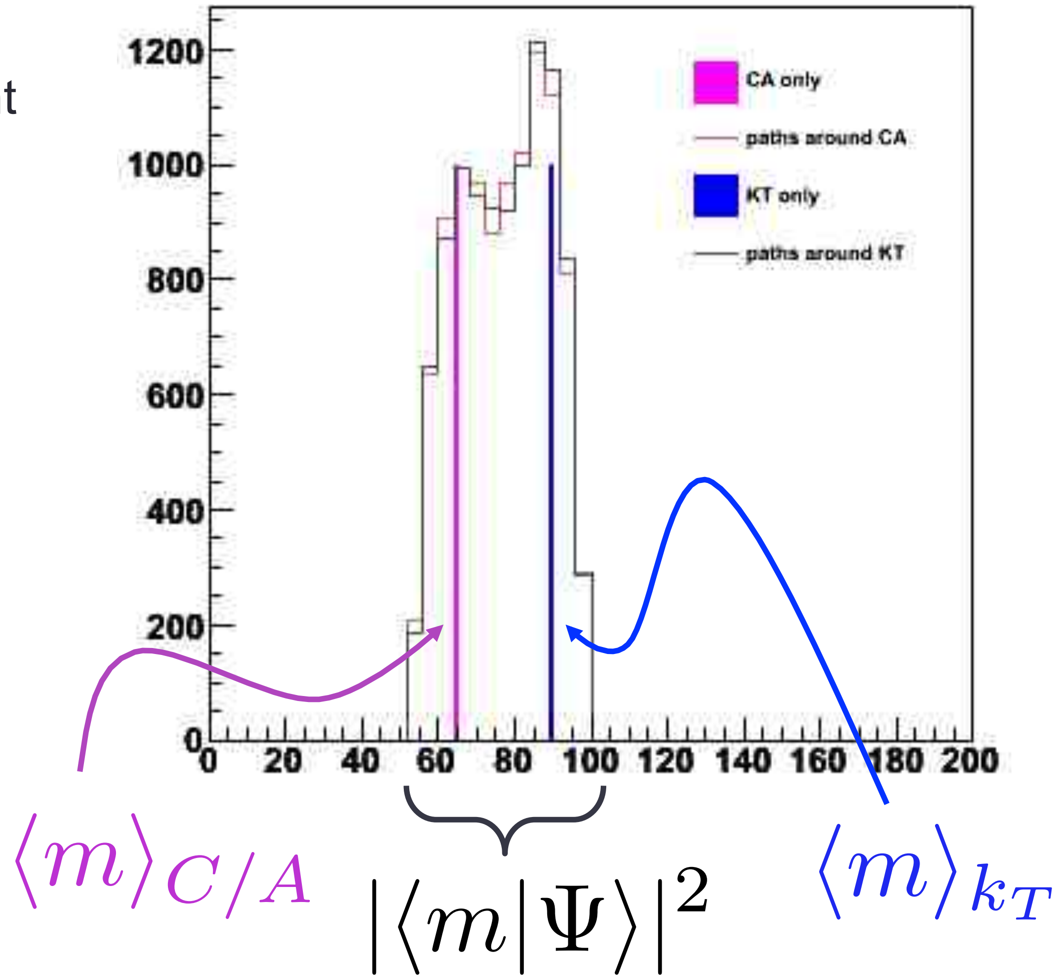
- Construct 100 trees
  from each jet in each event

- Apply pruning to each tree

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

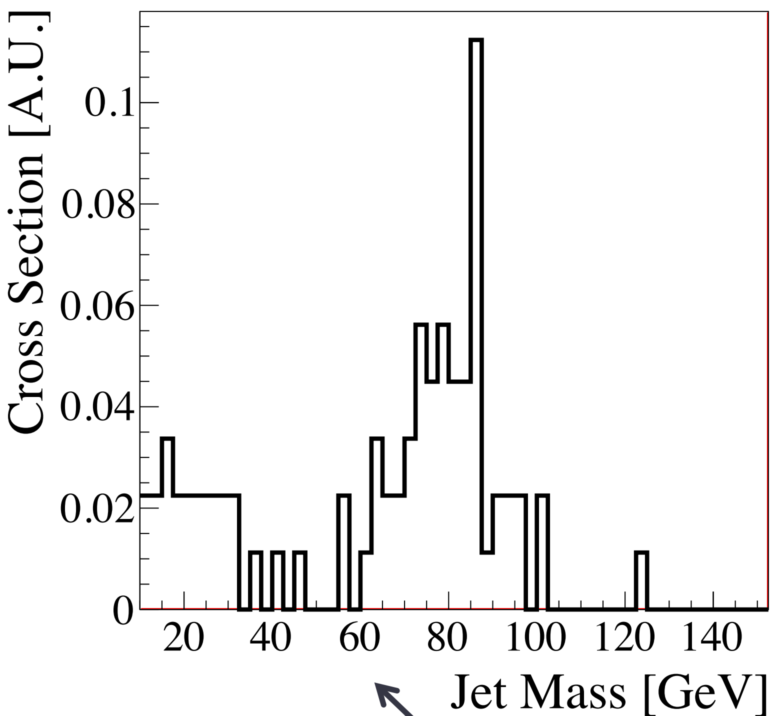
$$\Delta R_{ij} > D_{\text{cut}}$$

- Histogram resulting masses

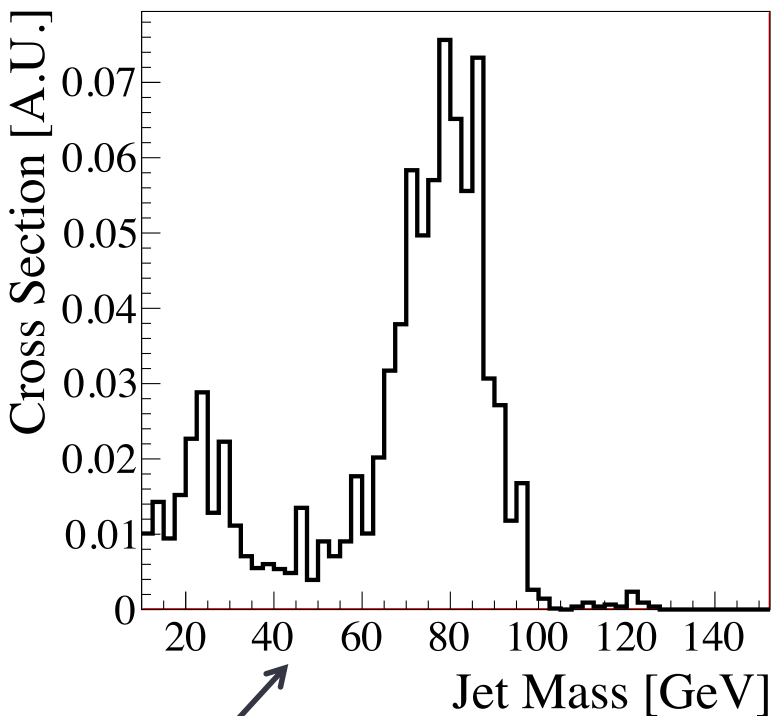


Distributions become much smoother

Classical anti-k_T



Pruned Qjets anti-k_T



The same 100 events

Need fewer events for same precision

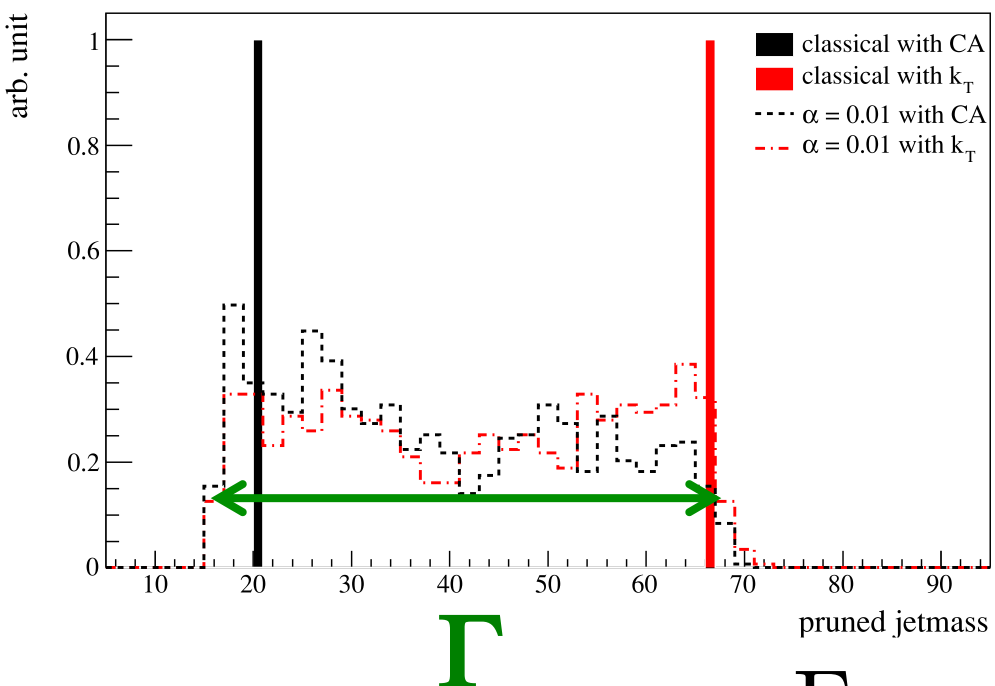
For example,
- Take 10 boosted W events (p_T>500)
- Construct jet mass
- Look at **variance** of the the **mean** W-jet mass over many pseudo-experiments

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	$\boxed{0.45}$

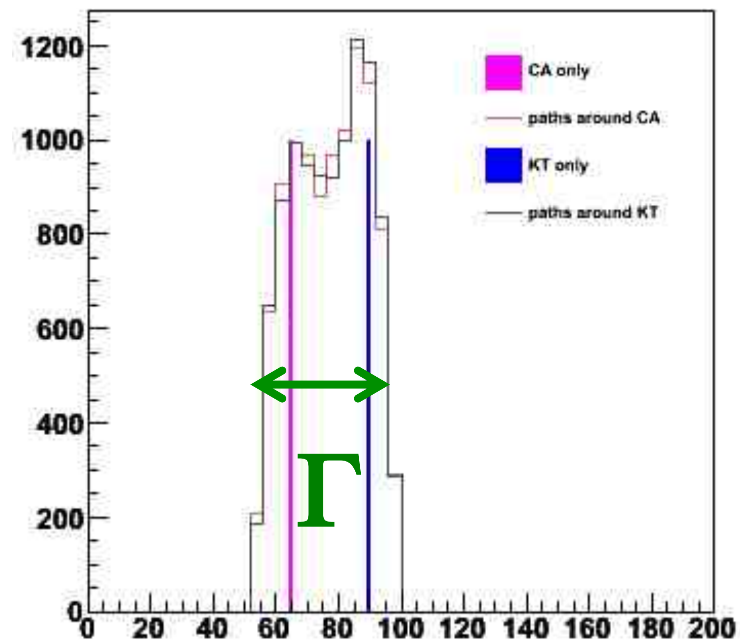
Qjets needs **half as much luminosity** as conventional jet algorithms

Signal vs background

QCD jets (one event)



W jets (one event)

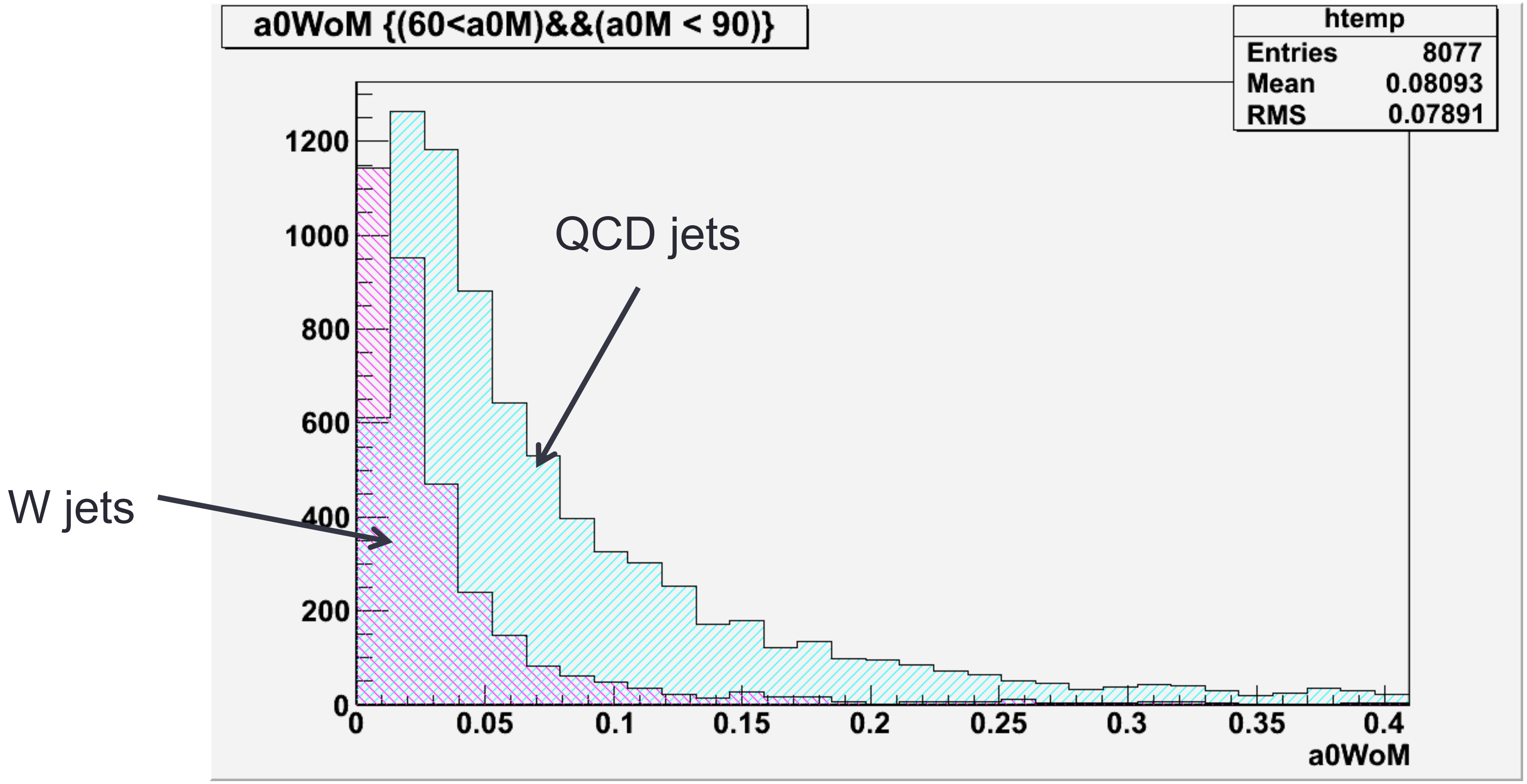


Volatility $\mathcal{V} = \dfrac{\Gamma}{\langle m \rangle}$ is a purely Q-observable

Volatility

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

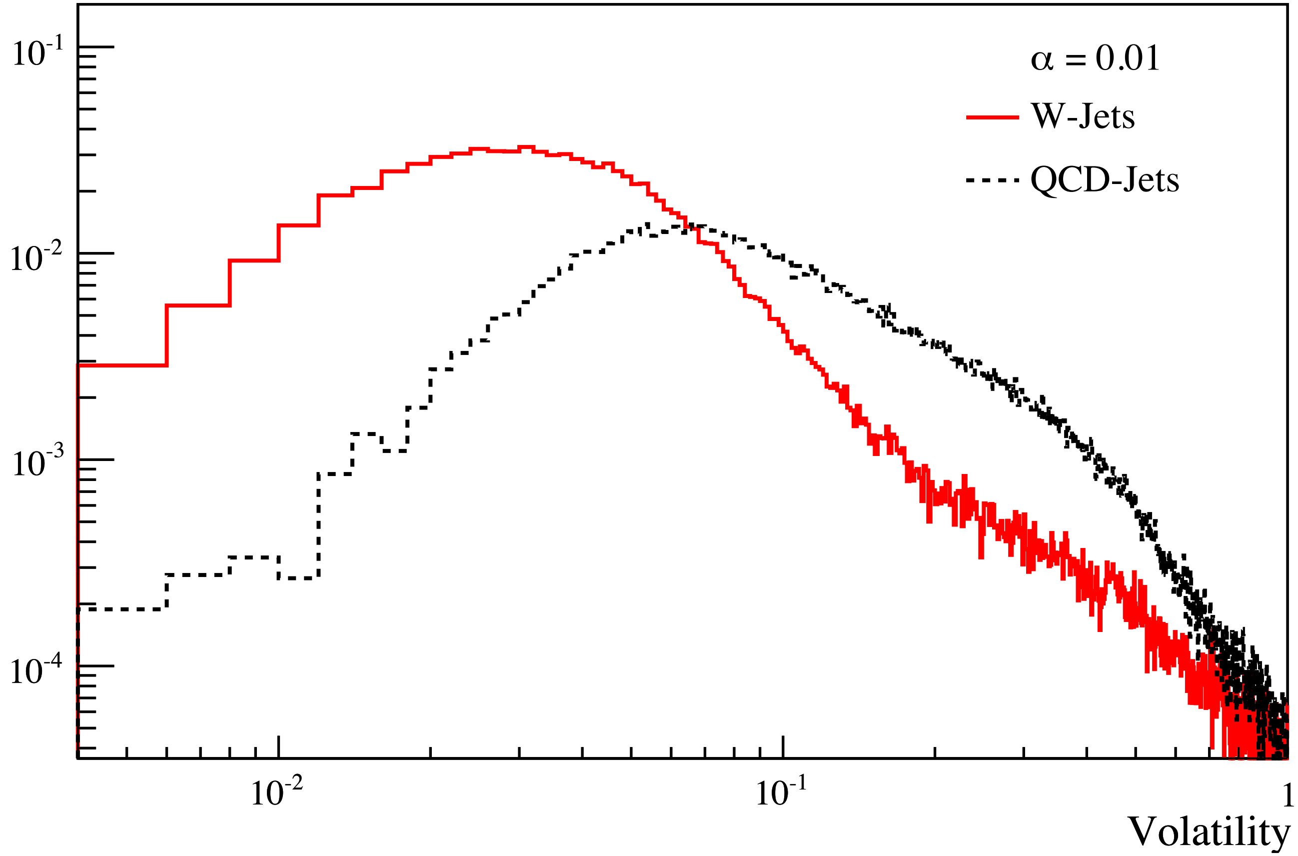
QCD jets are broader than boosted W jets



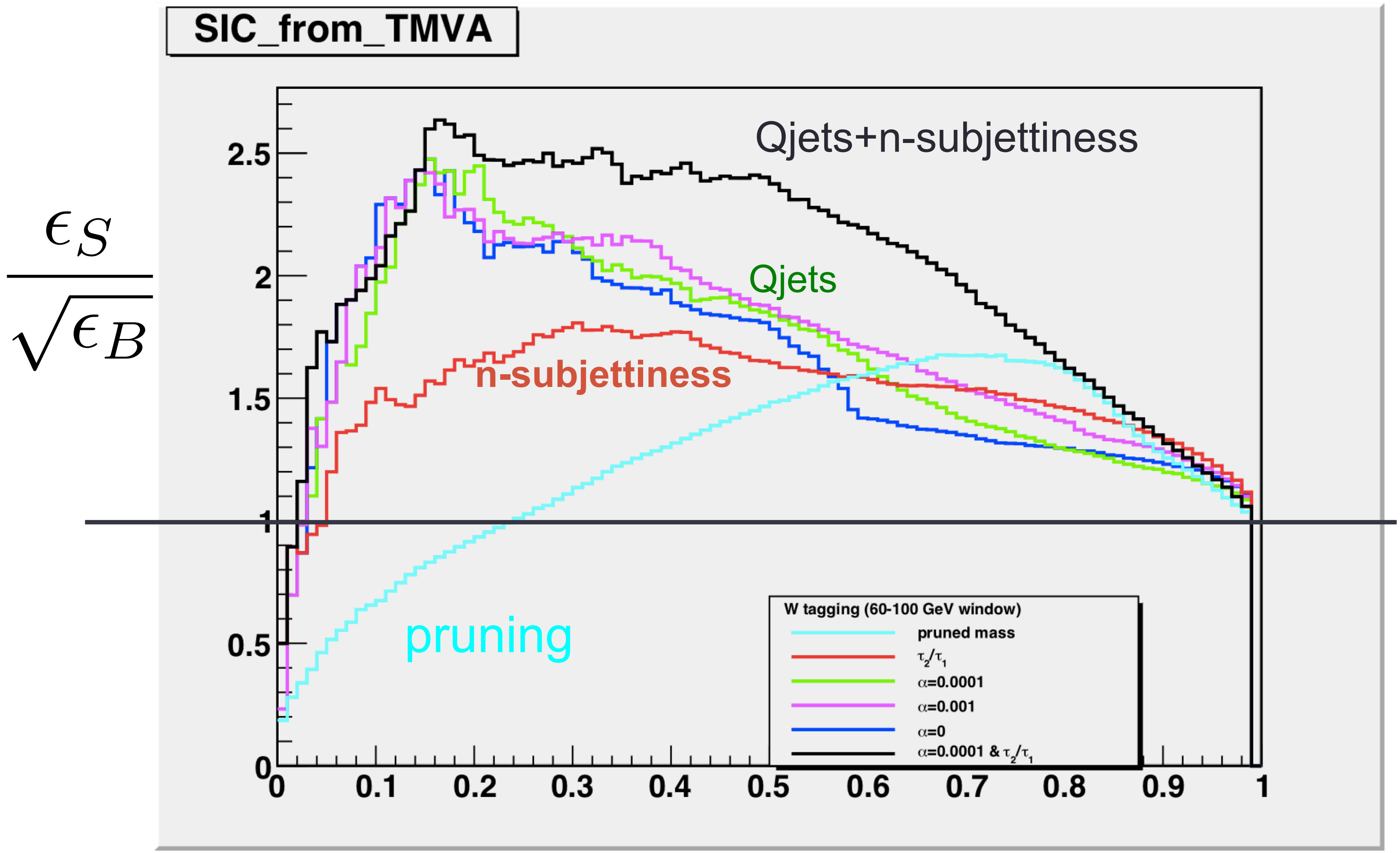
Volatility

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

QCD jets are broader than boosted W jets



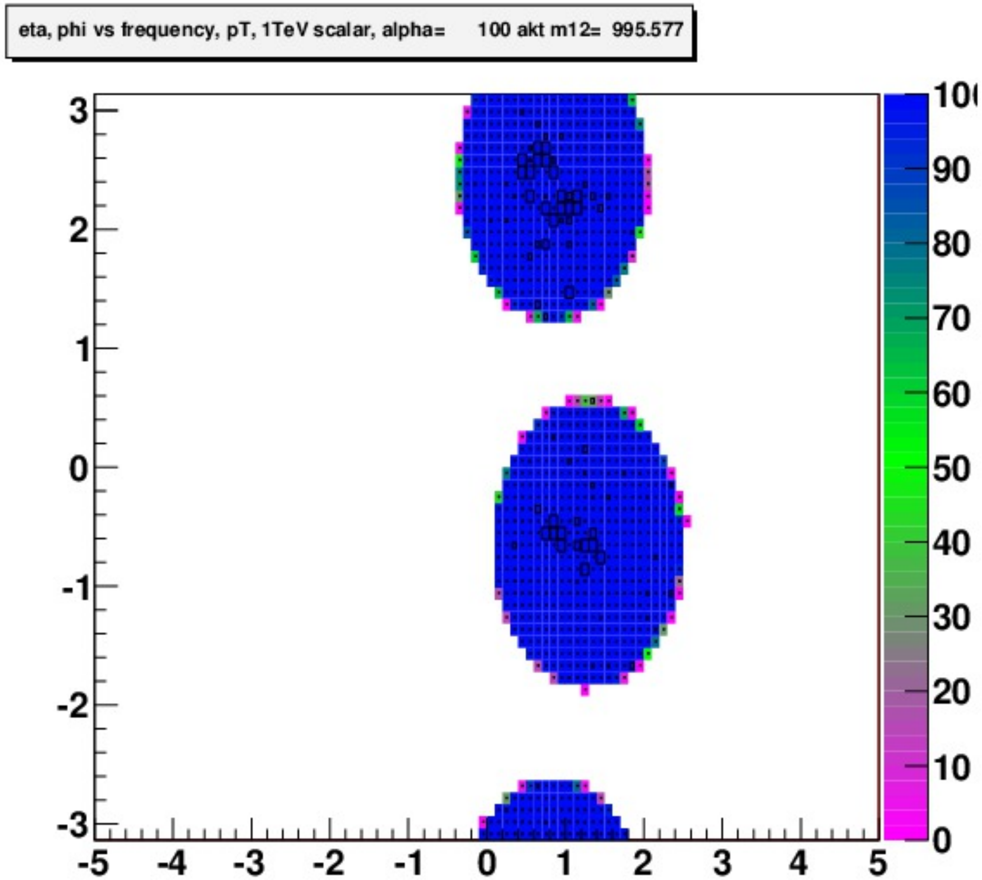
W-tagging: cut on volatility



Qjets in dijet events (no pruning)

$\alpha = 100$

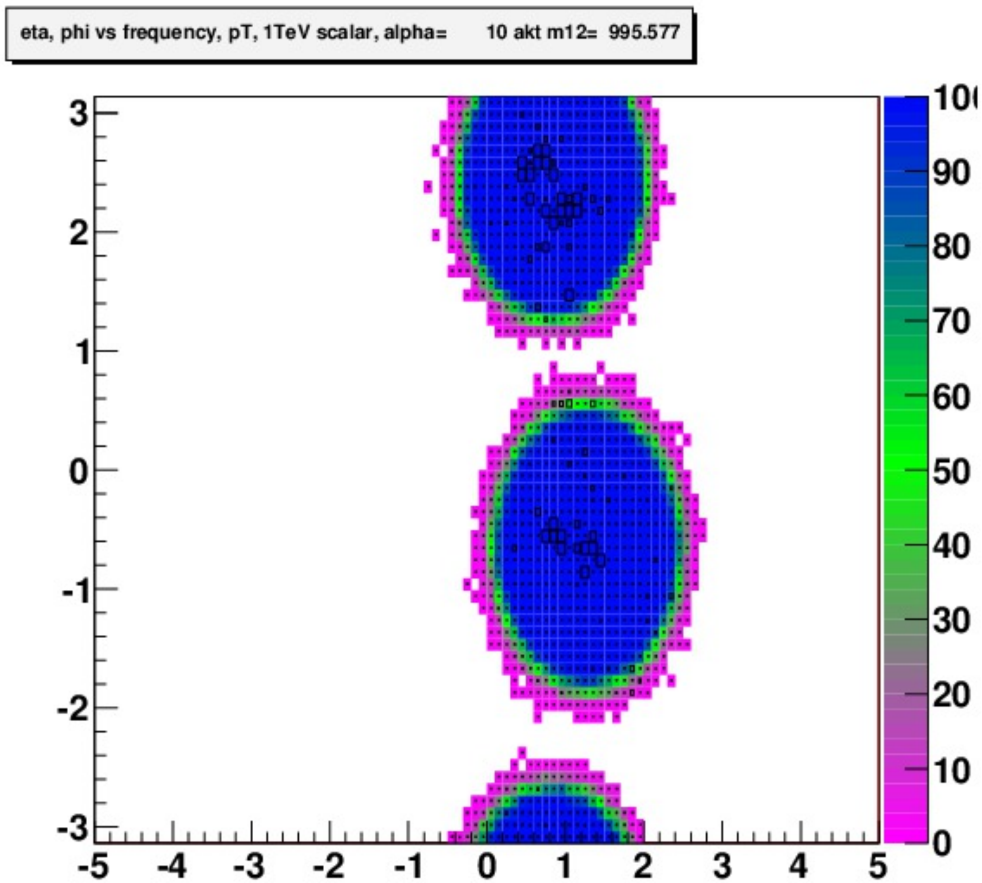
(classical anti-kT)



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

Qjets in dijet events (no pruning)

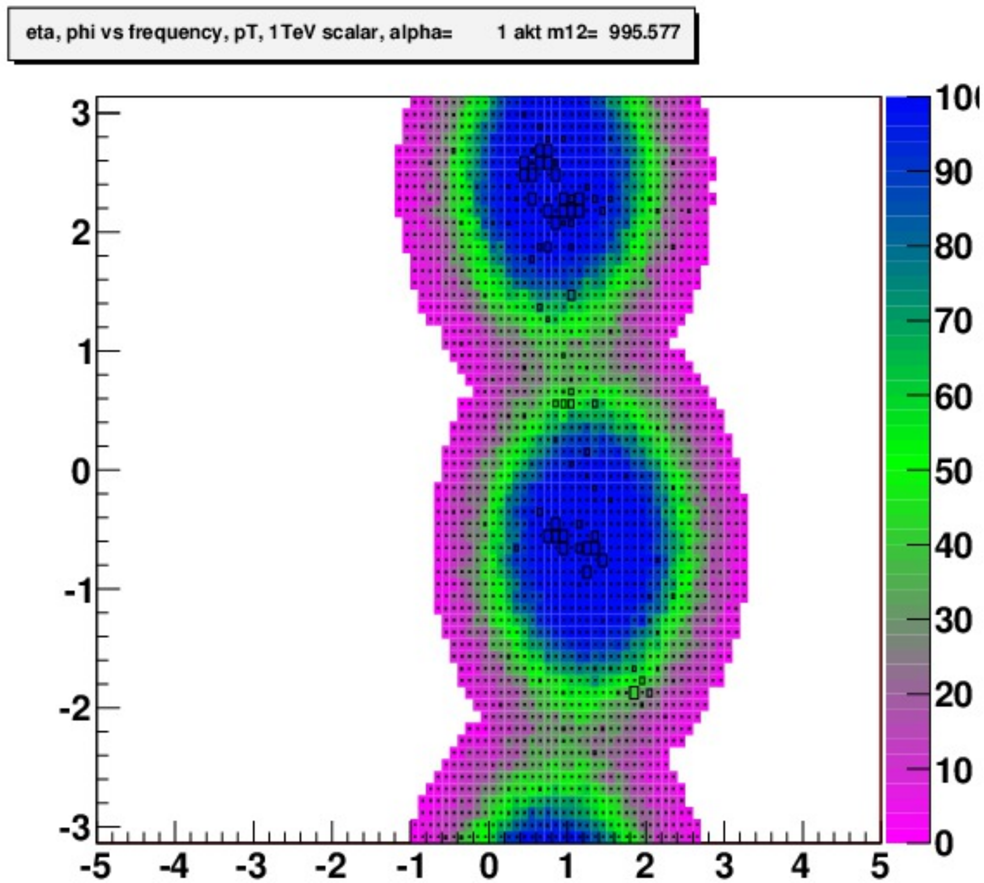
$\alpha = 10$



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

Qjets in dijet events (no pruning)

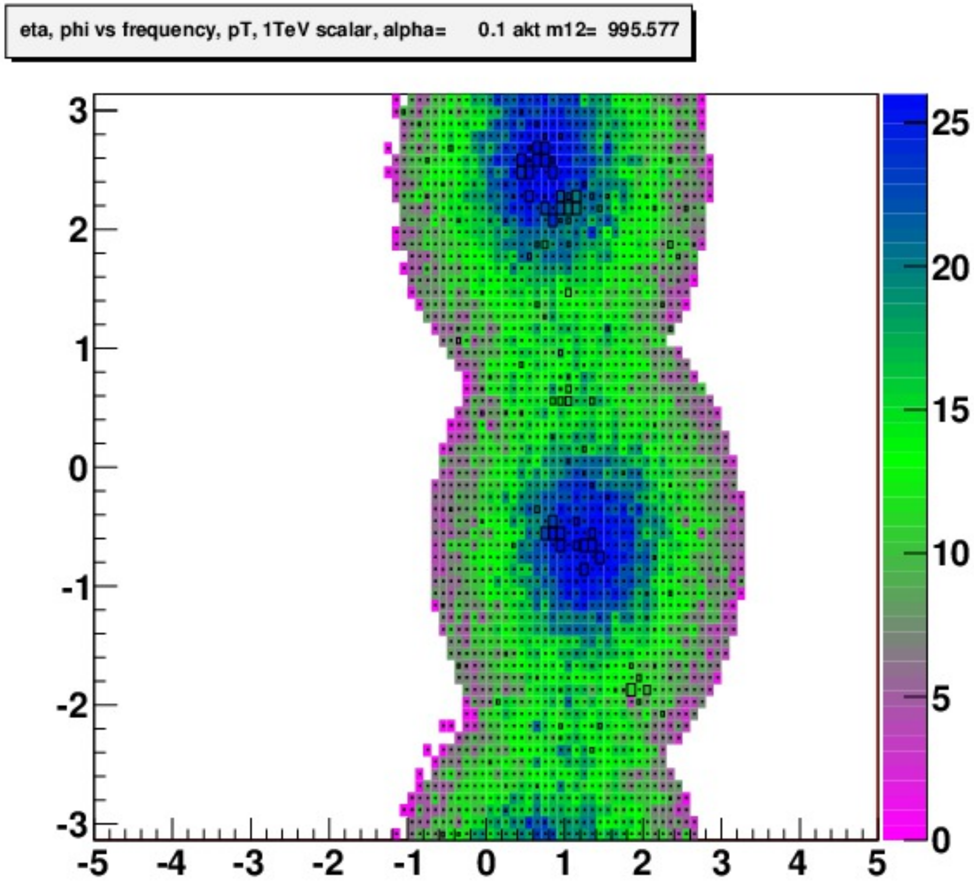
$\alpha = 1$



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

Qjets in dijet events (no pruning)

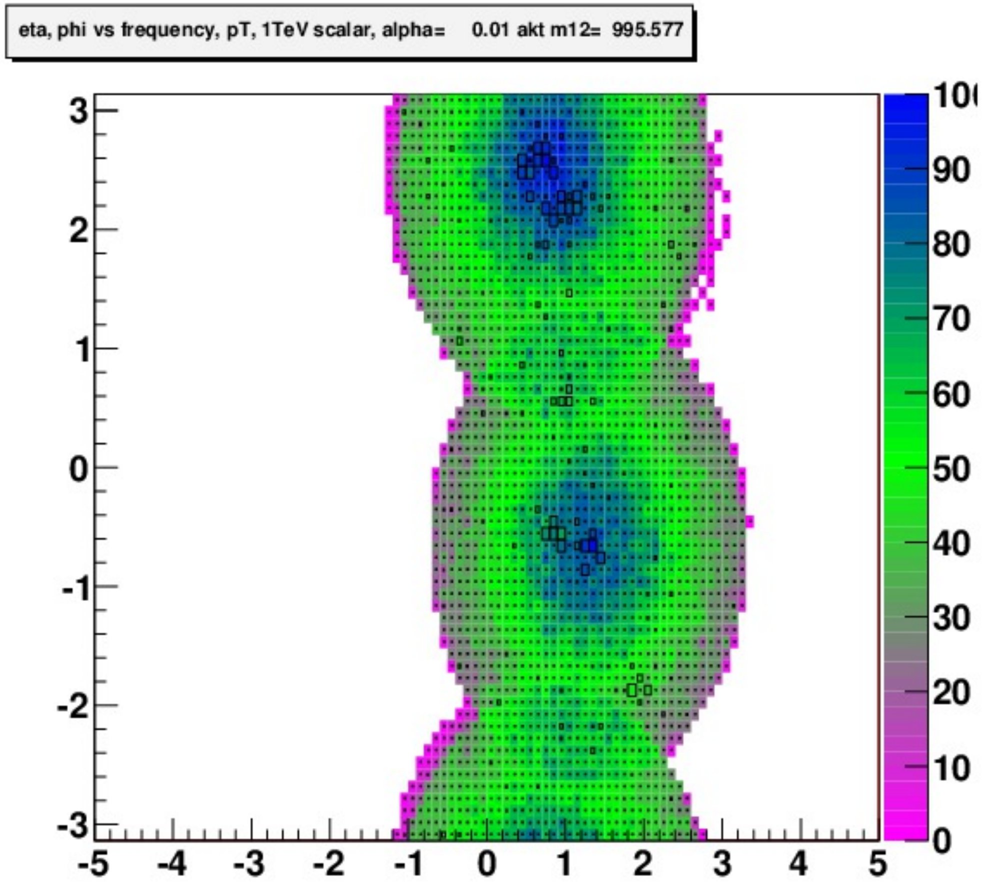
$\alpha = 0.1$



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

Qjets in dijet events (no pruning)

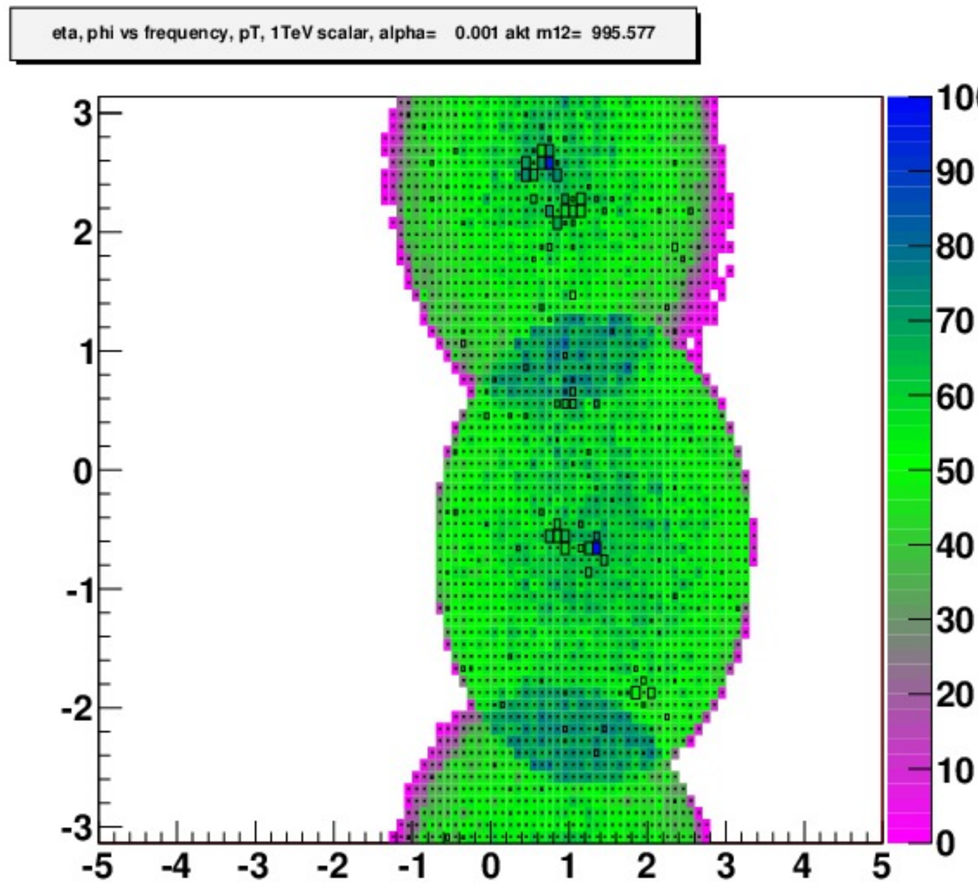
$\alpha = 0.01$



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

Qjets in dijet events (no pruning)

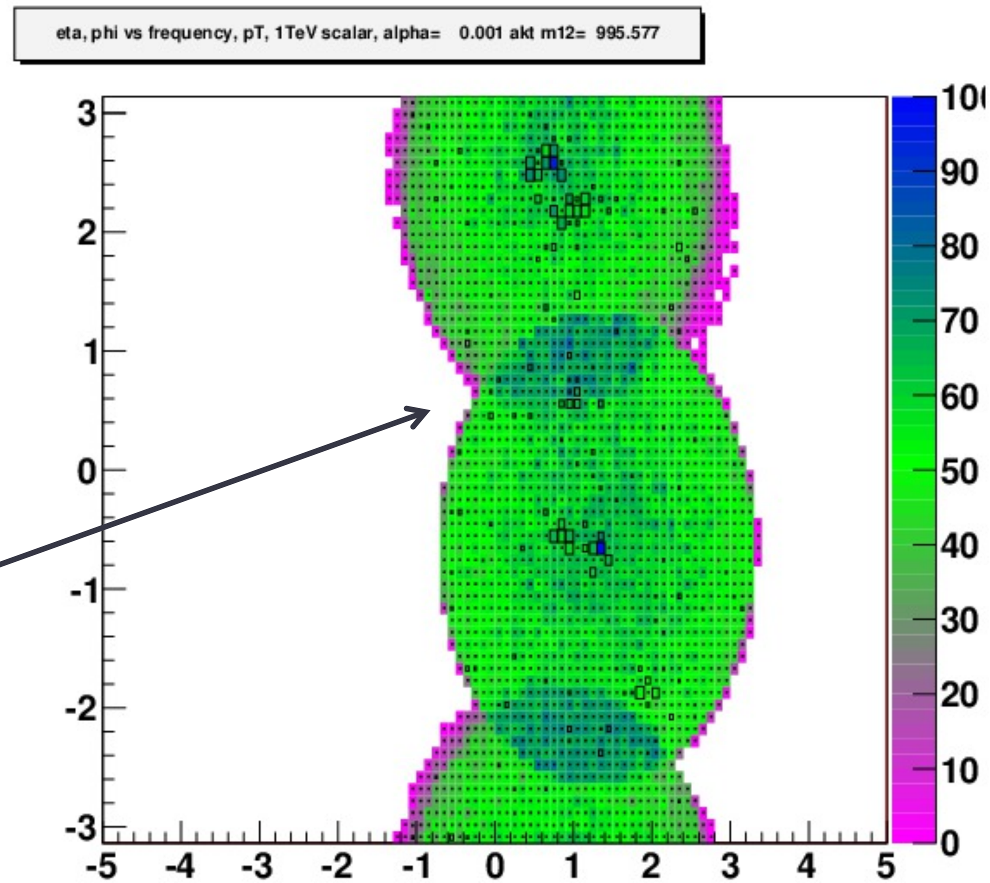
$\alpha = 0.001$



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

Qjets in dijet events (no pruning)

$\alpha = 0.001$



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

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

Conclusions

Multivariate analysis quickly tells you how well you could **possibly do**

FRAMING

See if simple variable can do as well

POWER

Sometimes MVA is really necessary

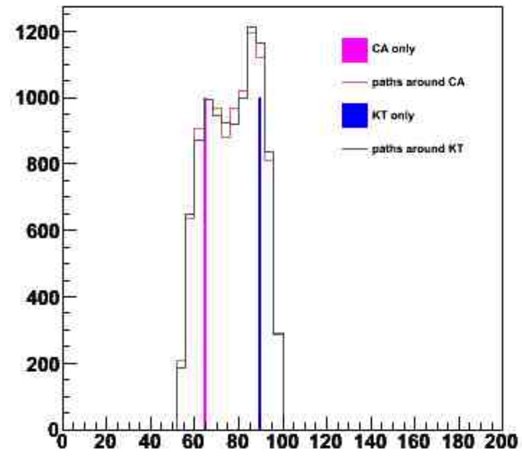
EFFICIENCY

Saves you the trouble or looking for good variables

- Combing 2-3 variables for quark-gluon discrimination seems to help
- Need to measure **correlations in real data**
  - With 1D distributions data, we can determine which generator is better
  - With 2D distribution, generators and models can be tested and improved

- Best description of a jet (or an event) may be multiple descriptions

- Qjets: non-deterministic jet algorithm
- Randomness gives ensemble of interpretations of each jet



JET CHARGE

Jet charge should be measured!

p_T weighted jet charge:

$$\mathcal{Q}^i_\kappa = \frac{1}{(p_T^{\text{jet}})^\kappa} \sum_{j \in \text{jet}} Q_j (p_T^j)^\kappa$$

- Used in DIS
- Used at LEP
- Used at Tevatron
- Not used at LHC

Can be measured

Is useful for BSM

Can be calculated

Jet charge is a July 5 analysis

July 4: find the Higgs.
July 5: what are its properties?

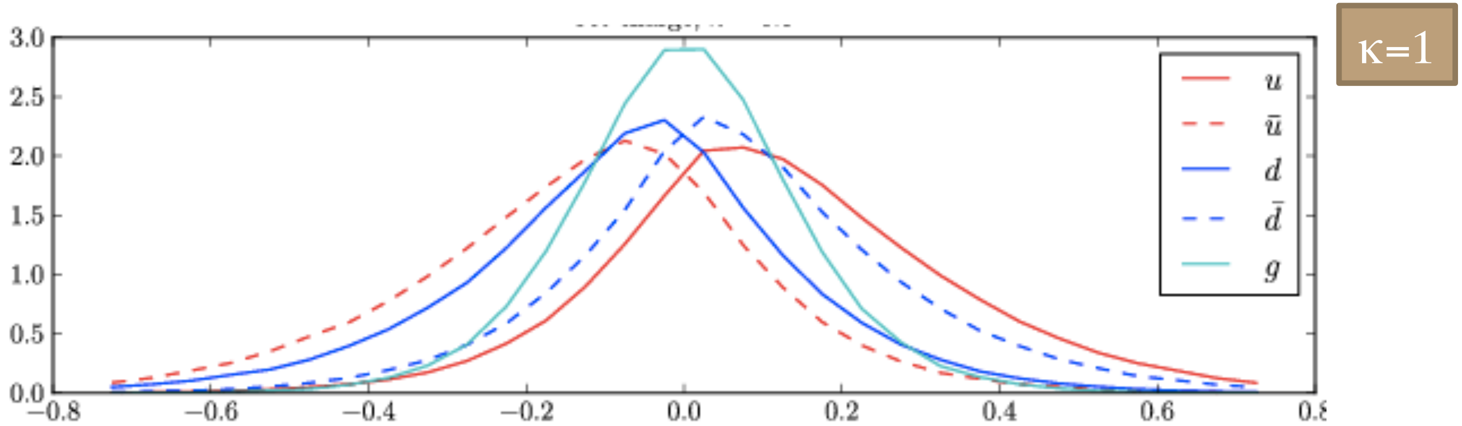
mm/dd/yyyy: Find BSM
Mm/dd+1/yyyy: What are its properties?

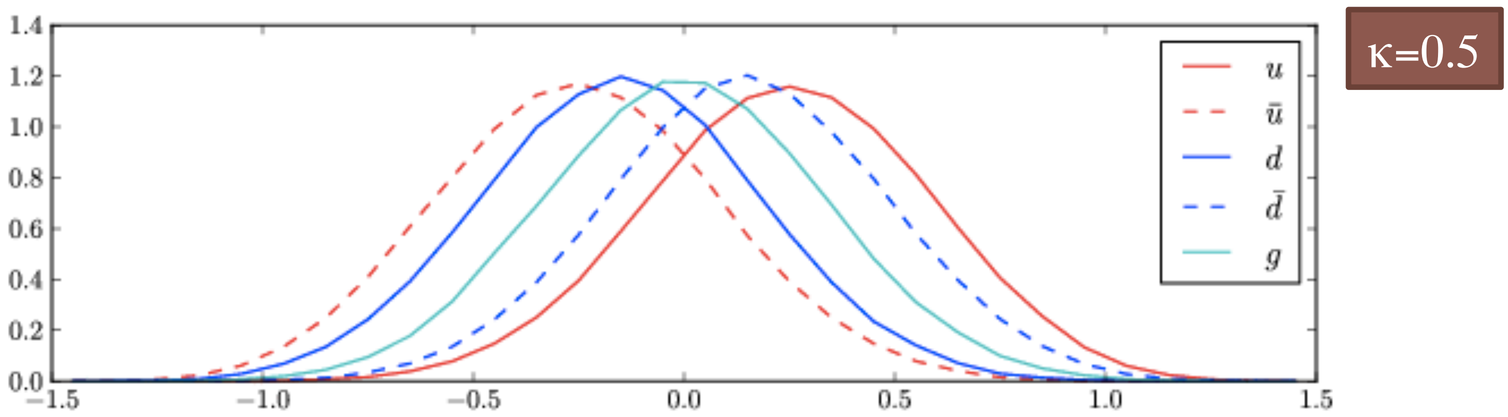
Distinguishings charge $-\frac{2}{3}, -\frac{1}{3}, 0, \frac{1}{3}, \frac{2}{3}$

Measured the p_T-weighted **jet charge**:

Krohn, Lin, MDS, Waalewijn
arXiv:1209.2421

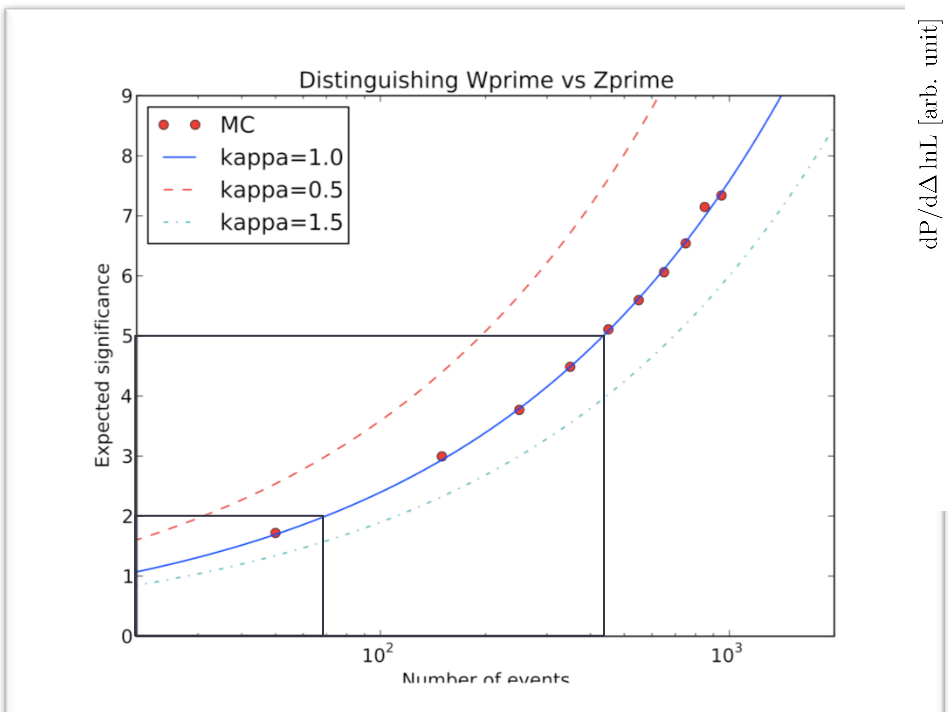
$$Q_\kappa^i = \frac{1}{(p_T^{\text{jet}})^\kappa} \sum_{j \in \text{jet}} Q_j (p_T^j)^\kappa$$

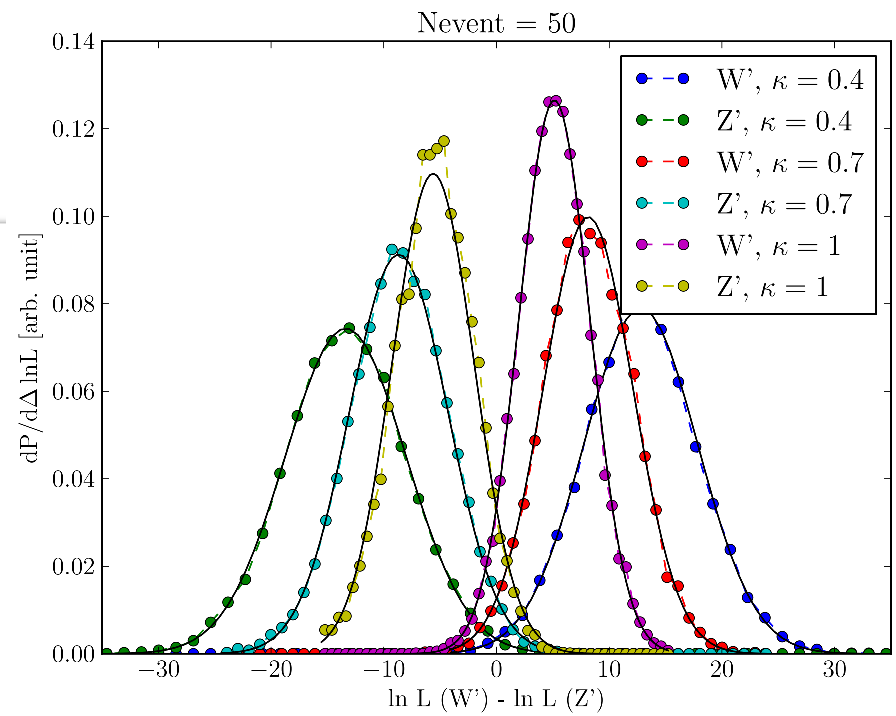




Distinguishes W ' from Z '

Log-likelihood distribution for 1 TeV resonance, various κ

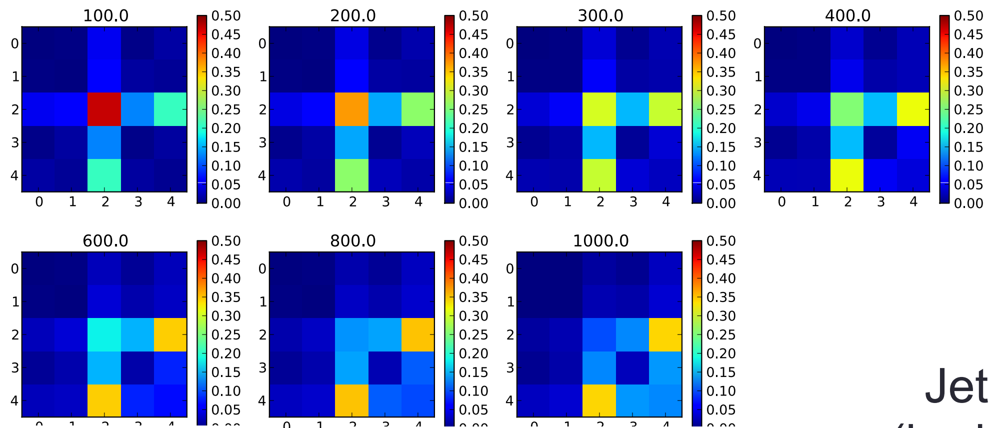




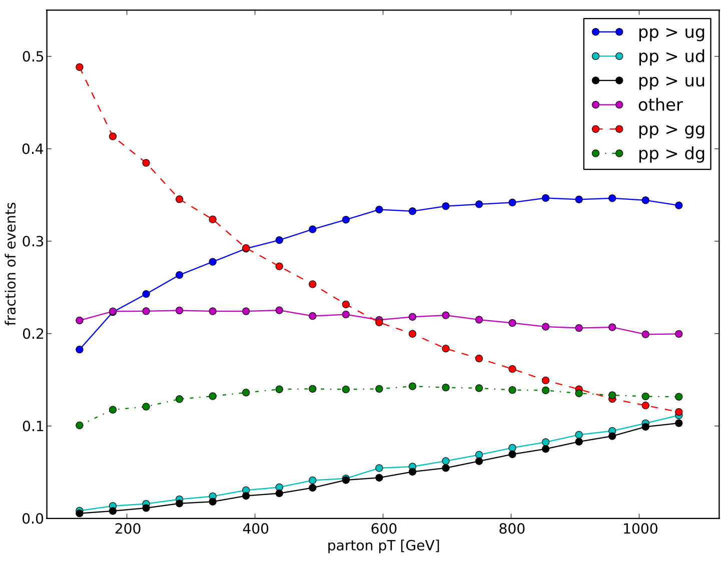
2σ with 30 events
5σ with 200 events

Calibrate on standard model

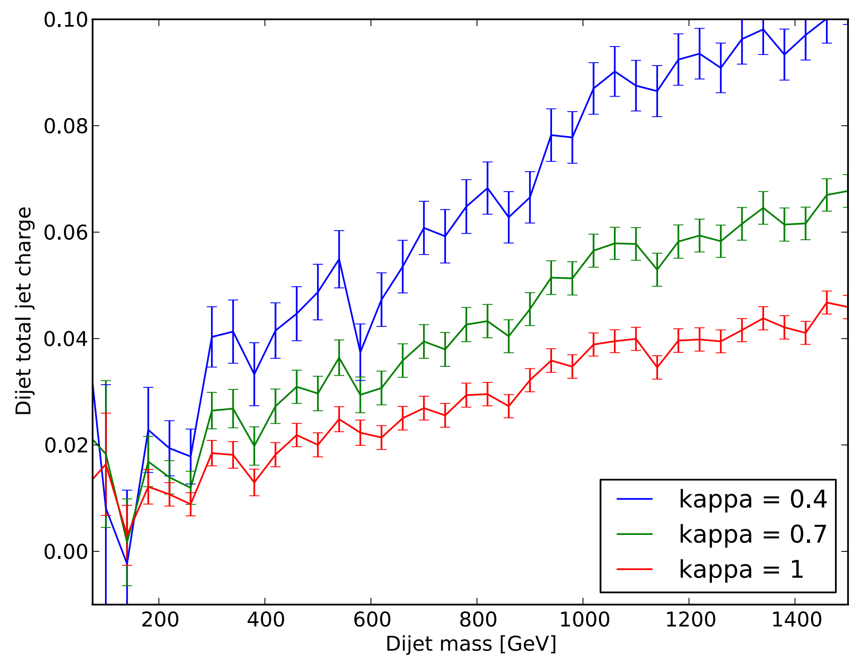
2D charges (parton level)
for different pT



Fractions
(parton level)

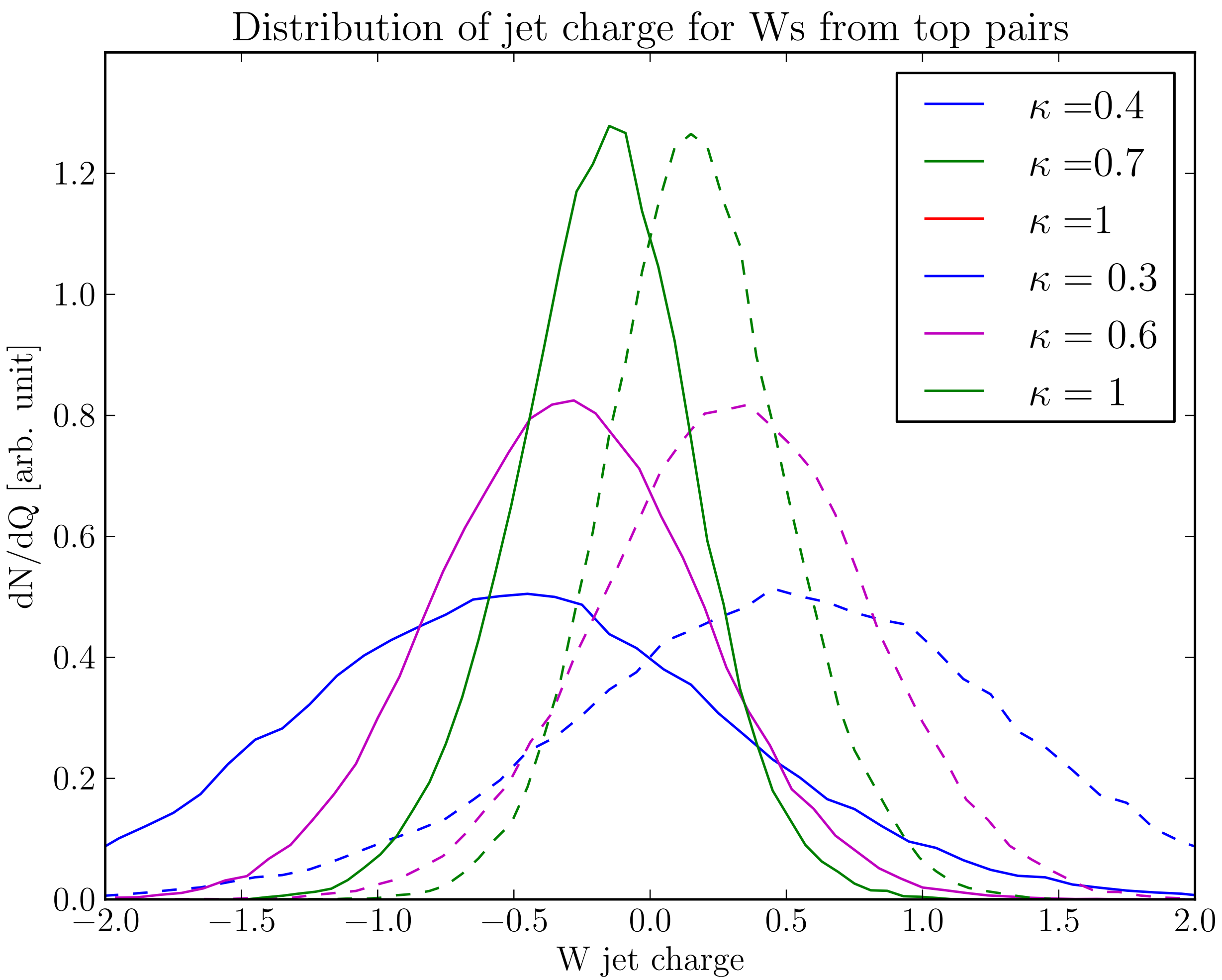


Jet charge
(hadron level)



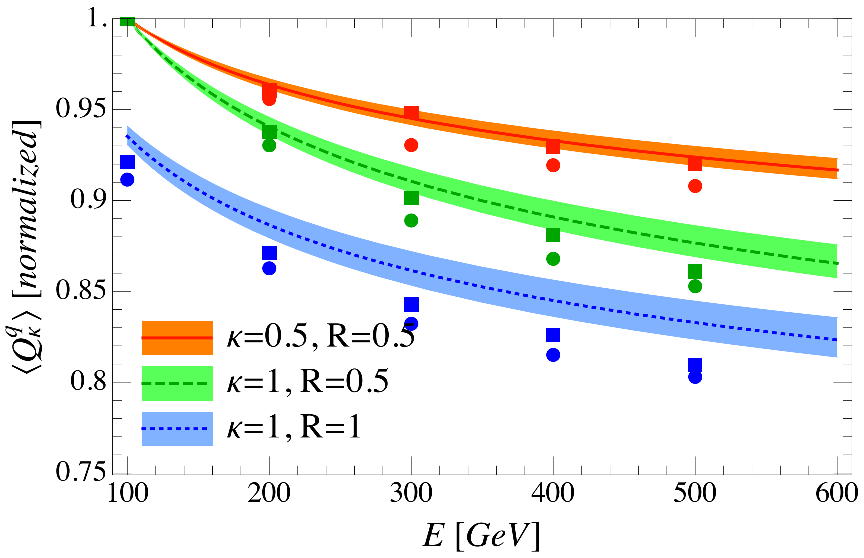
Hadronic W decays in tops

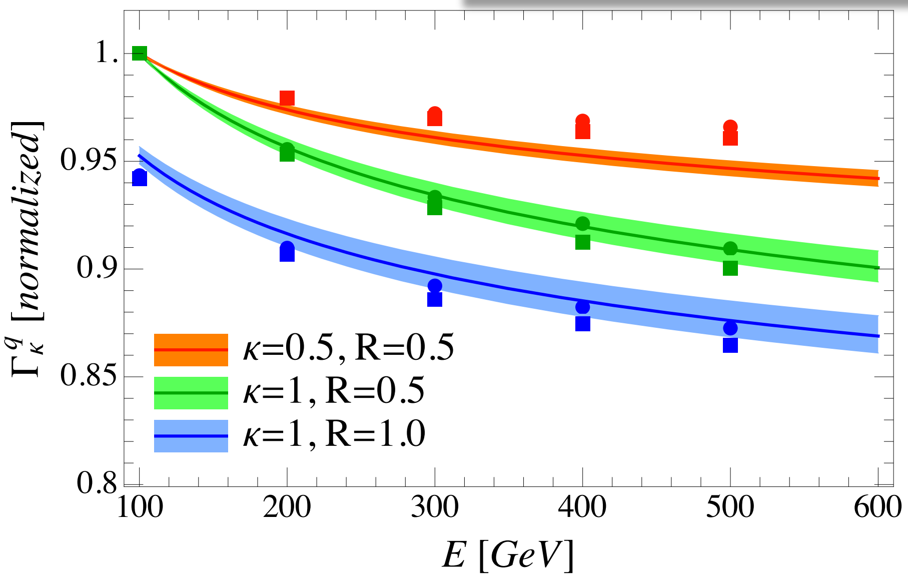
$$\mathcal{Q}_\kappa^i = \frac{1}{(p_T^{\text{jet}})^\kappa} \sum_{j\in\text{jet}} Q_j (p_T^j)^\kappa$$



Mean and width are calculable

Krohn, Lin, MDS, Waalewijn
arXiv:1209.2421



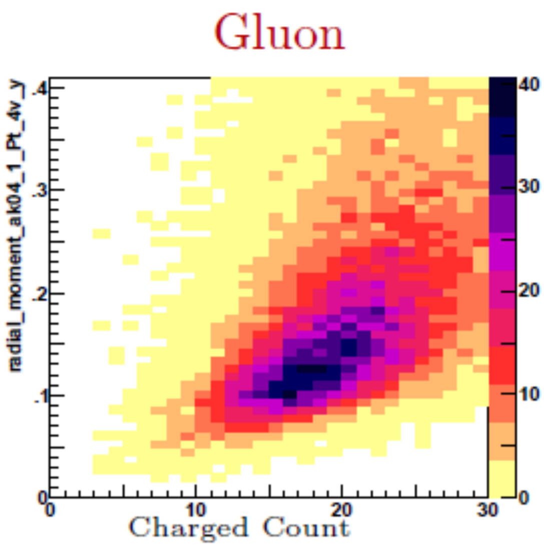


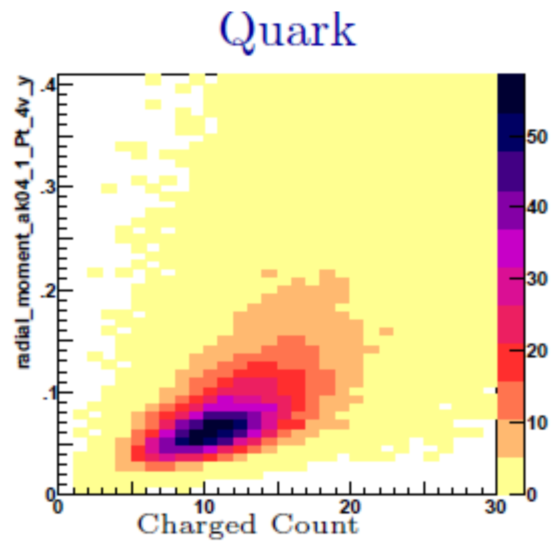
- Moments of charge distribution
  calculable from moments of fragmentation functions
- Evolution of these moments can be done precisely in QCD

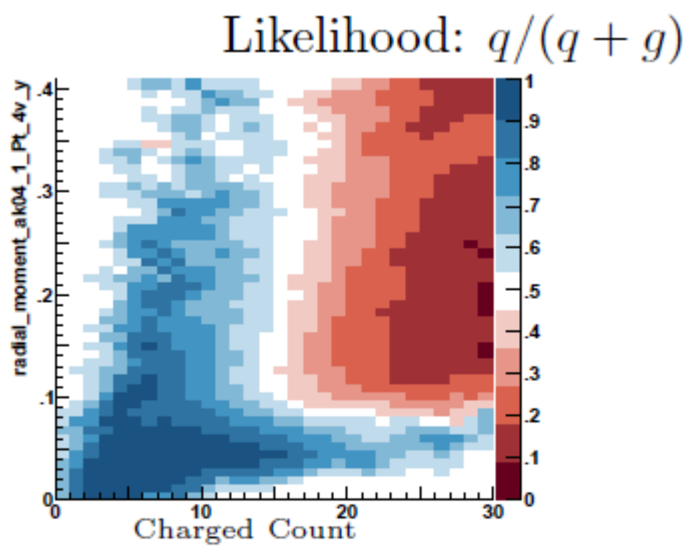
Please measure these moments!

BACKUP SLIDES

Best pair works pretty good

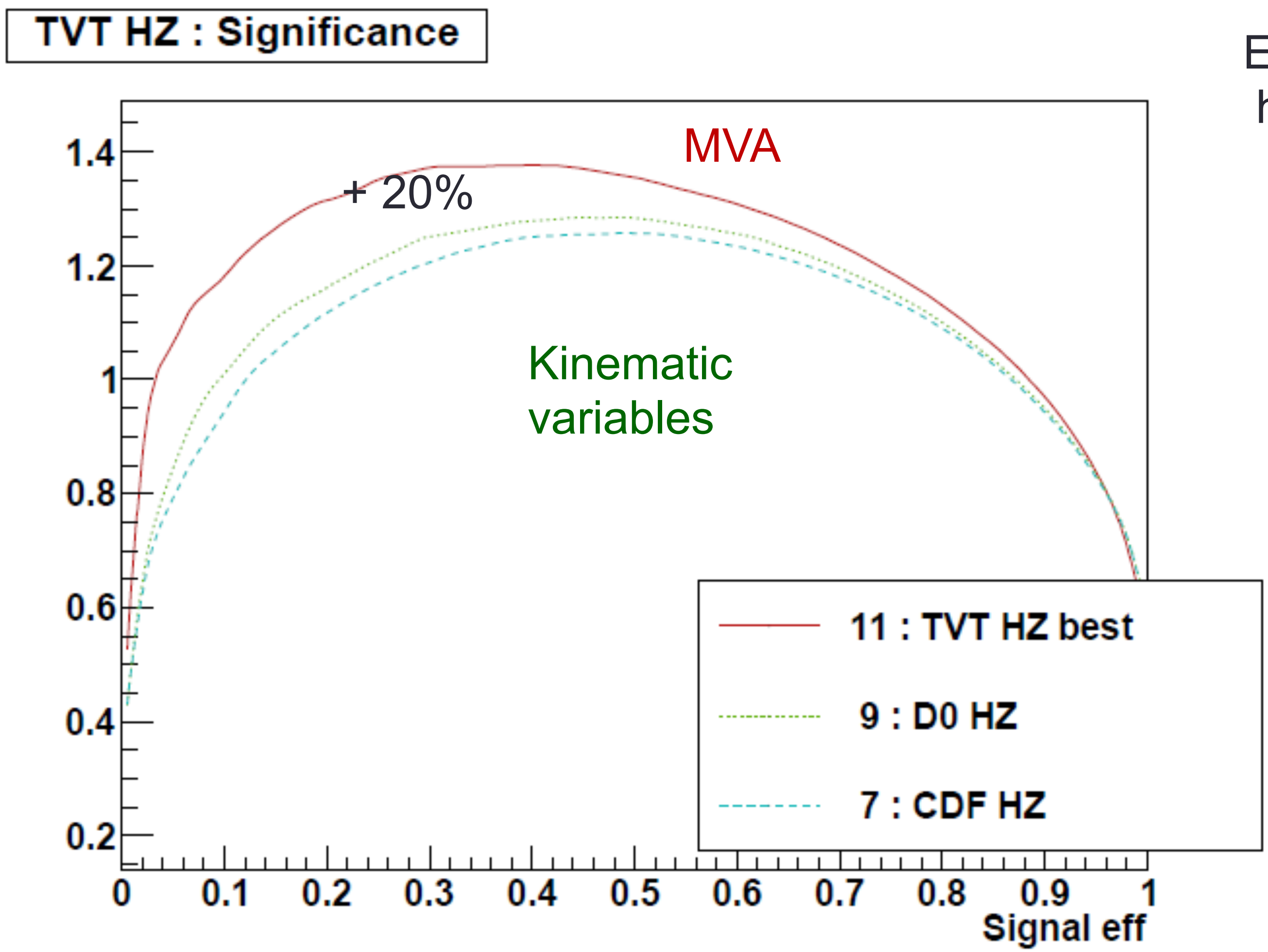






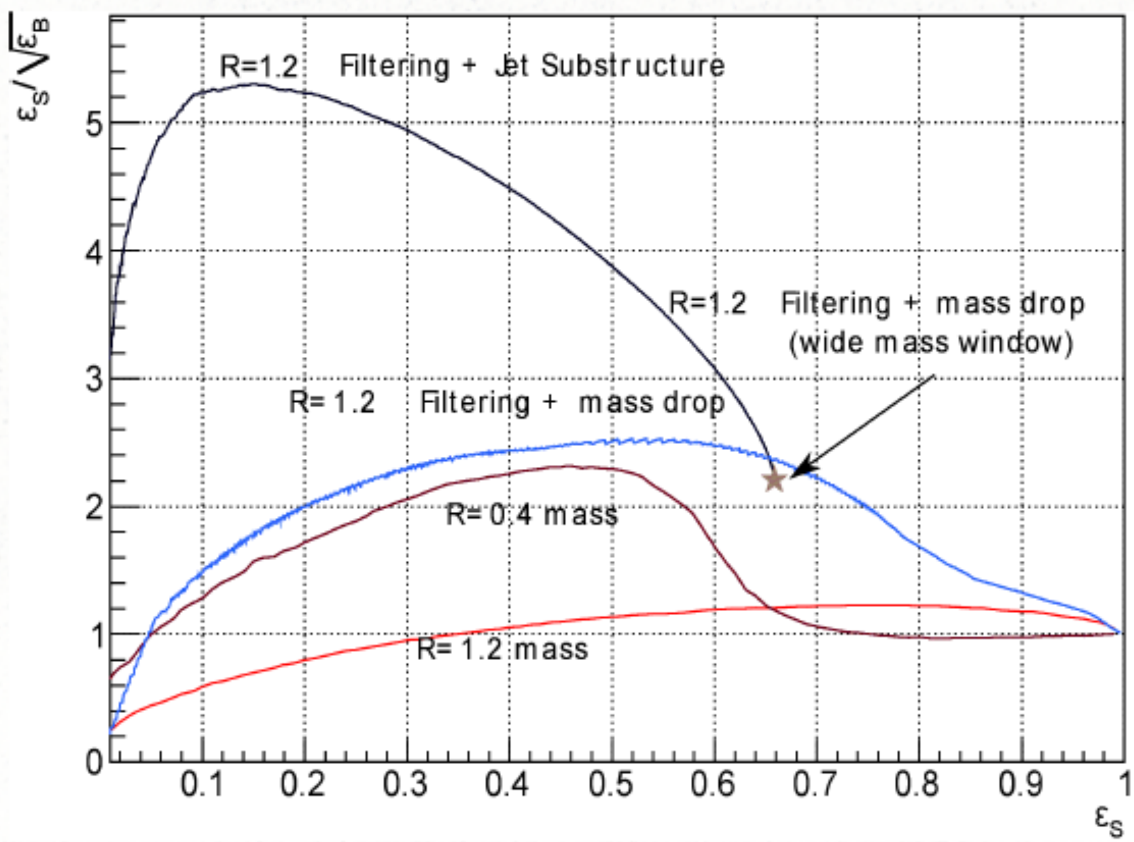
- Can cut on single variable
  Girth x Multiplicity

MVA POWERFUL FOR HIGGS SEARCH



Every little bit helps

Optimize W tagging



<u>7 MC insensitive variables</u>

1. Jet mass with R=0.5
2. Jet mass with R=0.4
3. Filtered Jet mass
4. Mass of hardest subjet, from filtering
5. Mass of second hardest subjet, from filtering
6. Ratio of the pT's of the two hardest subjets
7. Planar Flow

This is the **ultimate goal** of W-tagging

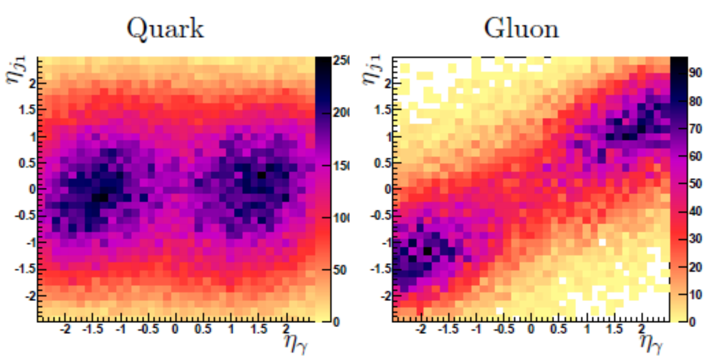
Challenge: can fewer variables do as well?

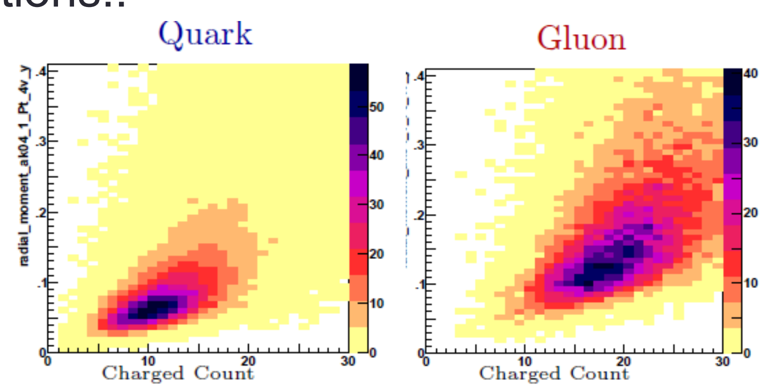
General observations

- For matrix-element level (fixed # dof)
  - BDT much **easier to implement** than matrix element method
  - MVA kinematic discriminants can be **similar to cutting on one** smart variable (e.g. in finding pure samples of Quark and Gluon jets)

- For showered samples (W-tagging, HZ)
  - **Subtle correlations** make **thinking difficult**
  - Some discriminants (like pull) **don't work on their own** but work well as the 5th or 6th variable added to a BDT
  - Sometimes **few variables are enough** (e.g. Quark vs Glue)
  - Sometimes **many variables are needed** (e.g. W-tagging)

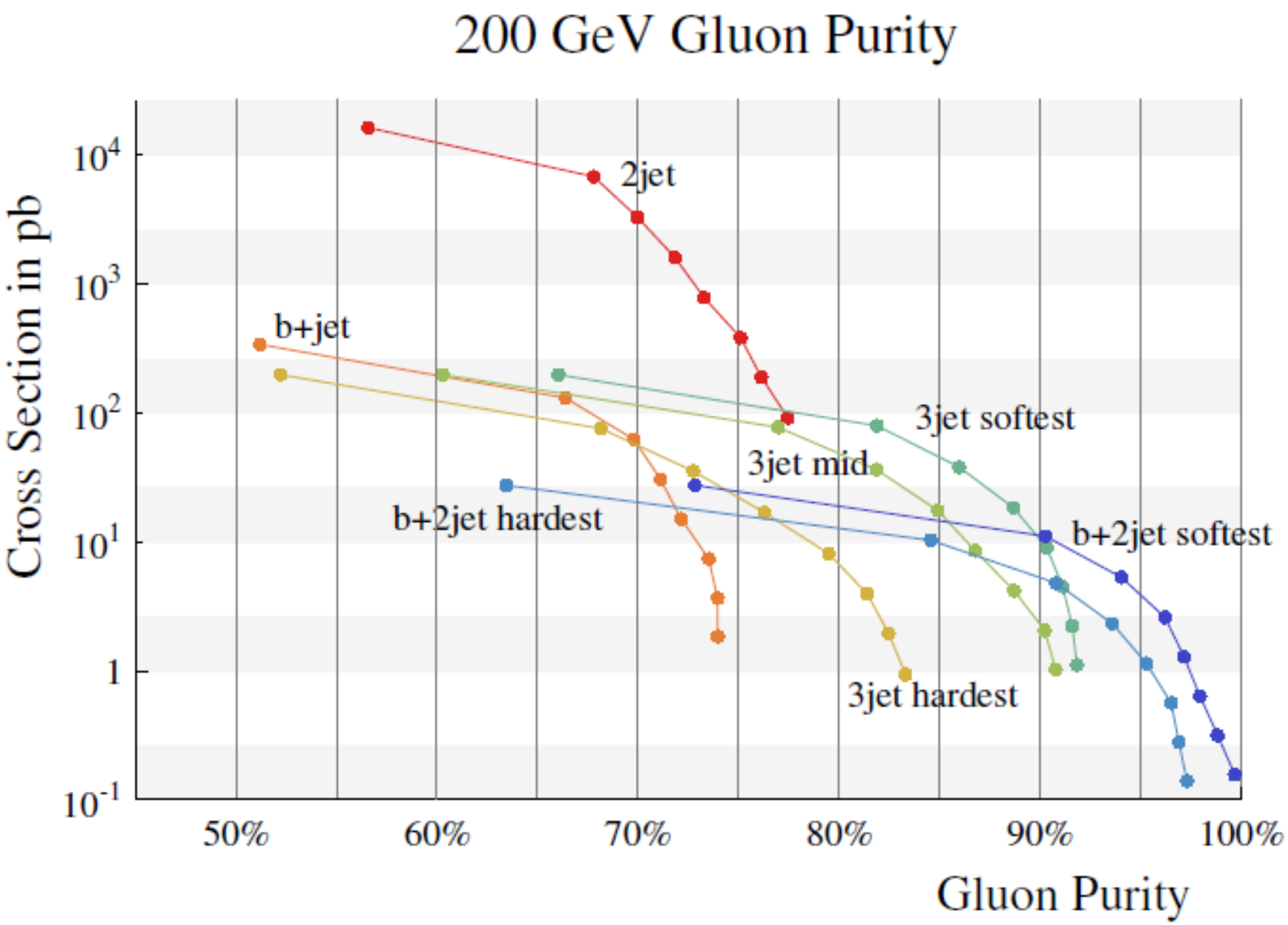
- We need data on 2d correlations!!





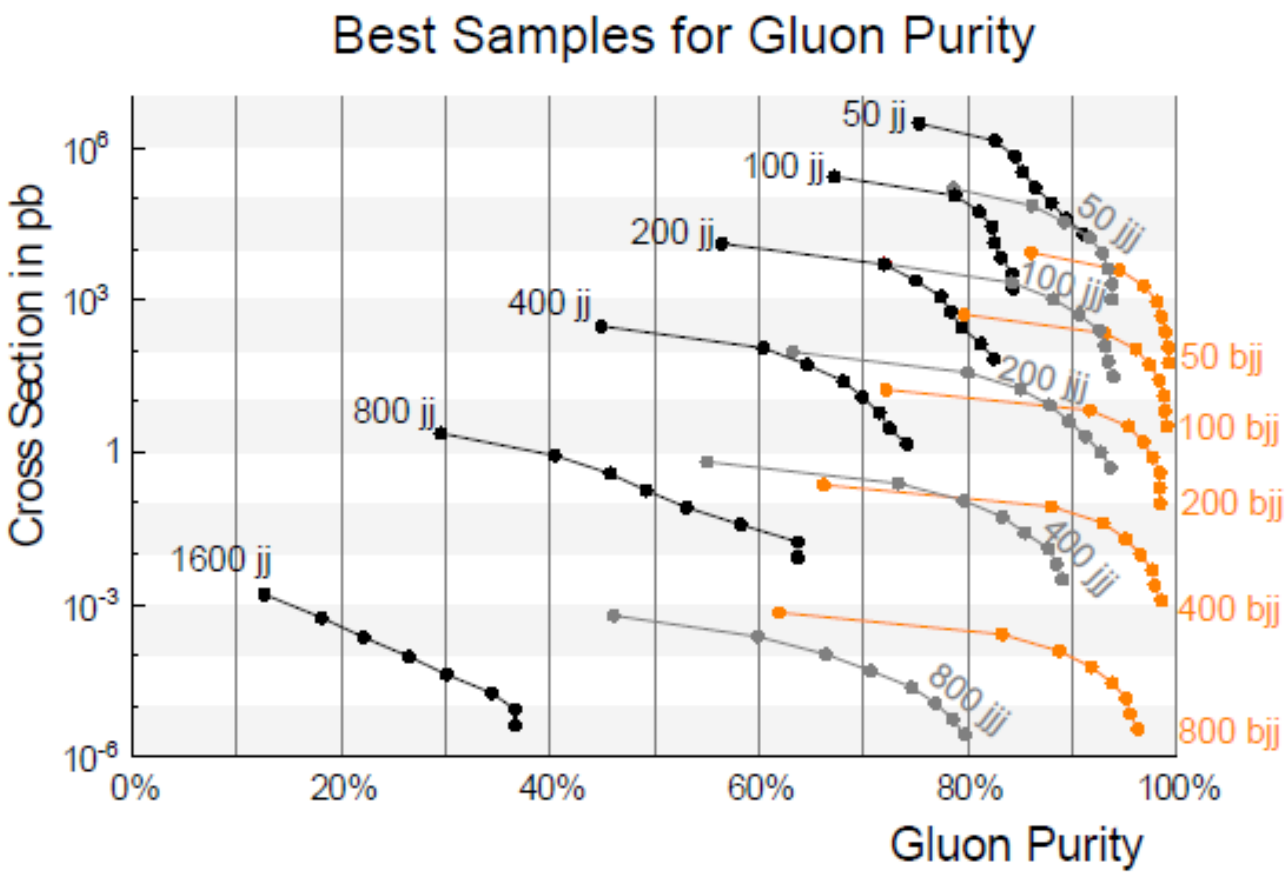
Most useful with **pure sampes** (quark, glue, hadronic-W, hadronic-top)

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