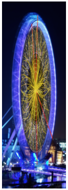
Fuzzy Jets

A new class of jet algorithms that use soft assignments during clustering

Benjamin Nachman

SLAC, Stanford University

Paper in Preparation



BOOST 2014

In collaboration with

Lester Mackey

Ariel Schwartzman

Conrad Stansbury

Jet Clustering as an Unsupervised Learning Task

Fundamental question: *which particles belong together?*

The state-of-the-art in organizing hadronic final states at the LHC is **hierarchical agglomerative** clustering:

hierarchical: hereditary structure to the final state classification
agglomerative: every object starts as a cluster and is sequentially merged

These schemes require metrics on momenta $d_{ij} : \mathbb{R}^4 \times \mathbb{R}^4 \to \mathbb{R}^+, d_{iB} : \mathbb{R}^4 \to \mathbb{R}^+$ and proceed as follows:

1. Assign each particle as a proto-jet.
2. Repeat until there are no proto-jets left: Let $d_{kX} = \text{argmin}_{i,j}\{d_{ij}, d_{iB}\}$. If $X = l$, combine proto-jets k and l into a new proto-jet with $\vec{X}_{\text{new}} = \vec{X}_l + \vec{X}_k$ (E-scheme). Else, declare proto-jet k a jet and remove it from the list.

$\vec{X}$ = 4-vector, with components p_x, p_y, p_z and E.

Membership Functions

Definition For each particle i, its membership function f_i is a map $f_i : \{\text{jets}\} \to [0,1]$ such that $\sum_{j=1}^{k} f_i = 1$.

For the state-of-the-art clustering schemes, every clustered object belongs to exactly one jet (with probability 1) and thus

$$f_i(j) = \begin{cases} 1 & i \text{ was clustered into jet } j \\ 0 & \text{else} \end{cases}$$

New developments[†] have shown that introducing probabilistic memberships can lead to enhanced understanding/performance.

Our goal is to incorporate **fuzziness** during clustering in order to **learn** the features of jetty events and (sub)jetty structure.

$\rightarrow$ Today: a Preliminary Look at Fuzzy Jet Clustering

[†] Examples:

Qjets [1201.1914] Re-cluster with randomness. For particle i and jet j, $f_i(j)$ is the fraction of clusterings in which i is[†] in j.
N.B. Not exactly Qjets, which is sensitive to joint membership functions.

PUPPI [1407.6013] $f_i(j)$ based on local p_T density – high $f_i(j)$ indicates higher likelihood for hard scatter origin.

Probabilistic Clustering

Many standard methods of unsupervised learning incorporate probabilistic membership **as part of the clustering procedure**.

One of the most basic such algorithms is the **mixture model** for k jets,

$$p(\vec{X}) = \sum_{j=1}^{k} \pi_j \Phi(\vec{X}|\vec{\Theta}_j)$$

$\vec{\Theta}_j$ are the jets
π_j are (learned) priors

Which is equivalent to a *generative model* phrased as follows:

$\xi_i \in \{1, ..., k\} \equiv$ jet to which particle i belongs

$\xi_i \overset{IID}{\sim}$ Multinomial$(\vec{\pi}, 1)$, for $\sum_{i=1}^{|\pi|} \pi_i = 1$.

$\vec{X}_i|\xi_i \overset{IID}{\sim} \Phi(\vec{X}_i|\vec{\Theta}_{\xi_i})$, for a probability density Φ depending on parameters $\vec{\Theta}$

The probabilistic part of the clustering is in the random variables ξ_i.

Gaussian Mixture Model (GMM) Jets

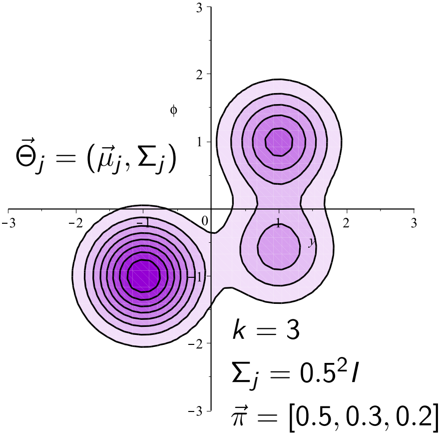
Consider a special case in which Φ is a Gaussian in $y - \phi$ space:

$$\Phi(\vec{\rho}|\vec{\mu}_j, \Sigma_j) = \frac{\exp\left(-\frac{1}{2}(\vec{\rho} - \vec{\mu}_j)^\intercal \Sigma_j^{-1}(\vec{\rho} - \vec{\mu}_j)\right)}{\sqrt{|2\pi\Sigma_j|}}$$

$\vec{\rho}$ is the $y - \phi$ (sub)coordinates of $\vec{X}$.
Σ is a 2×2 invertible matrix.

For a fixed k, the goal is to **learn**

1. $\vec{\mu}$ (jet positions)
2. Σ (jet size and shape)



EM Algorithm

There is a well-known procedure for iteratively determining the distribution of the ξ_i known as the Expectation-Maximization (EM) algorithm:

1. Initialize $\vec{\pi}$, $\vec{\mu}_j$ and Σ_j (more on this later)
2. Alternate until convergence

   E Step Compute membership functions $f_i(j) = \Pr(\xi_i = j | \vec{\mu}_j, \Sigma_j, \vec{\pi})$

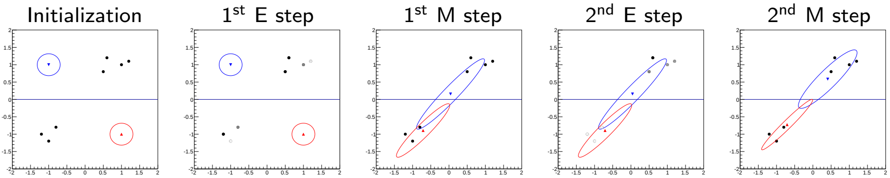
   M Step Given the $f_i(j)$, compute the new cluster centers and shapes:

$$\pi_j = \sum_{i=1}^{n} \frac{f_i(j)}{n} \qquad \vec{\mu}_j = \frac{\sum_i f_i(j)\vec{\rho}_i}{\sum_i f_i(j)} \qquad \Sigma_j = \frac{\sum_i f_i(j)(\vec{\rho}_i - \vec{\mu}_j)(\vec{\rho}_i - \vec{\mu}_j)^\mathsf{T}}{\sum_i f_i(j)}$$

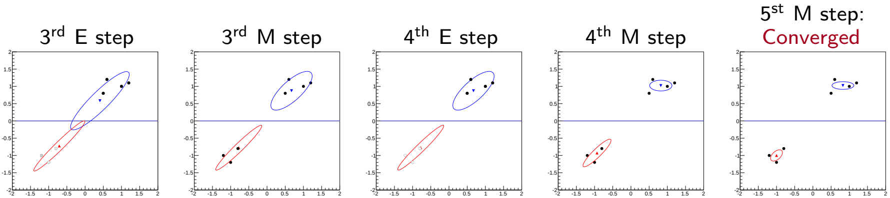
This is not sequential recombination!

N.B. This is Lloyd's algorithm for solving k-means in the limit that $\Sigma = \sigma^2 I \to 0$ and $\pi_i = \frac{1}{k}$.

Illustration: $k = 2$



In the E-step, darker colors correspond to higher value of f_i(blue jet).



The algorithm will converge, but not guaranteed to reach global optimum

Modified Mixture Model (mMM)

Out-of-the-box, Gaussian Mixture Model jets are not IRC safe.

However, there is a simple modification that makes a broad class of *modified* mixture model jets IR(C) safe, given by the likelihood

$$\mathcal{L}(\vec{\pi}, \vec{\Theta}) = \sum_{i=1}^{N} p_{Ti}^{\alpha} \left(\sum_{j=1}^{k} \pi_j \Phi(\vec{\rho}_i | \vec{\Theta}_j) \right)$$

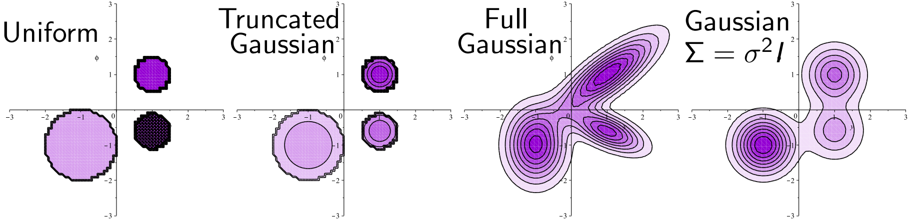
- For $\alpha \to 0$, we recover the standard mixture model.
- For $\alpha = 1$, the jets are IRC safe for any Φ that depends only on $\vec{\rho}$.
- For $\alpha > 0$, the jets are generally IR safe.

Physically: α gives a larger weight to the hard structure of a jet.

Modification to the EM Algorithm

The EM algorithm still applies with $\sim$one minor change: $f_i(j) \mapsto p_{Ti}^{\alpha} f_i(j)$.

For any kernel Φ, one needs to re-derive the M-step, but this is straightforward for simple shapes:



Preliminary results shown in the following slides constructed with PYTHIA 8.170 and FASTJET 3.0.3.

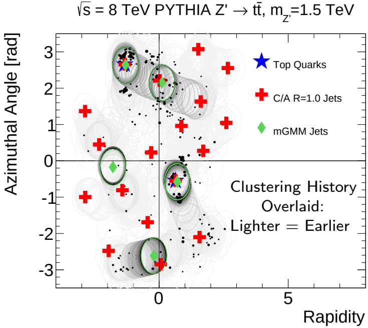
Choosing the Initialization

For sequential recombination, R is specified ahead of time and k is learned.

For mMM jets, k is specified ahead of time and (more than) R is learned.

There are many possibilities for initializing; for example,

1. Run a sequential recombination scheme for k and the $\vec{\mu}_{i=1..k}^{\text{initial}}$
2. Choose k using some standard procedure (e.g. *gap statistic*) and then initialize uniformly.
3. Let every particle be a cluster and let clusters merge.



$\pi \propto 1, \Sigma = 0.5^2 I, \alpha = 1$

Cluster Merging Scheme

More plots in backup

Learning σ for mGMM Fuzzy Jets

We have studied several kernels Φ and many configurations.

For illustration, the next slides show mGMM jets with $\Sigma \propto \sigma^2 I$.

`Initialization`: anti-k_t jets with $R = 1.0$ and $p_T > 5$ GeV.

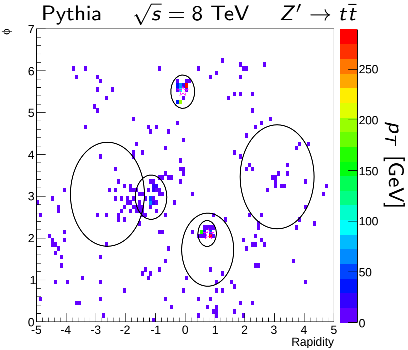
Background	QCD dijets, $\hat{p}_T > 200$ GeV
3-prong	$Z' \to t\bar{t}$ (fully hadronic)
2-prong	$W' \to W(\to qq')Z(\to e\bar{e})$

N.B. Signal events are weighted to match the background p_T spectrum.

Define the kinematics of fuzzy jet j by

$$\vec{X}_j^{\text{jet}} = \sum_{i=1}^{N} \vec{X}_i^{\text{particle}} \times \delta(\max_q f_i(q), j)$$

More information on jet kinematics ▸ in the backup

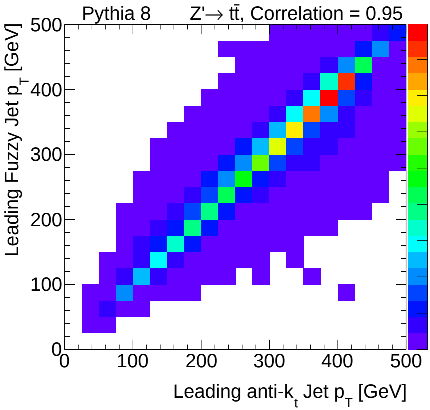


Ellipses are the 1σ contours
Pink stars are the top quarks

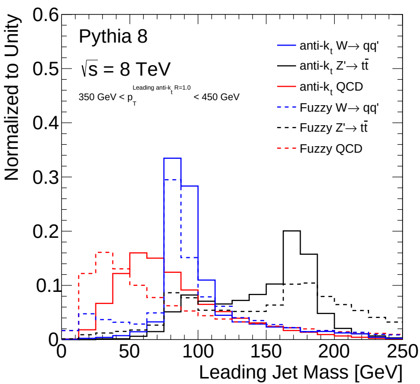
Fuzzy jets can (and do) overlap!

Comparing Jet Kinematics: Fuzzy Jets vs anti-k_t

Fuzzy mGMM jets with $\Sigma \propto \sigma^2 I$ are very similar to anti-k_t R=1.0 jets.



p_T is determined mostly by the hard core: fuzzy jets learn the **same** core

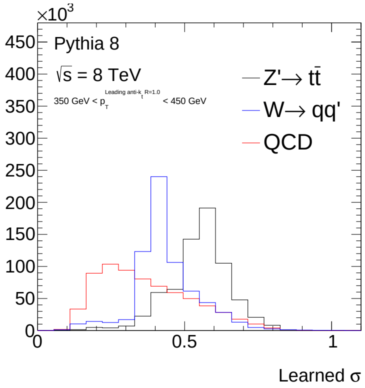


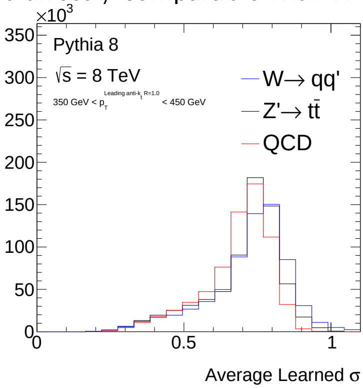
Lower p_T bin ▼ in the backup

Mass depends on the shape of the jet: fuzzy jets learn **different** shapes

The size σ of fuzzy jets

Unlike usual sequential recombination[†], jet size is **learned** by fuzzy jets. One measure is the Gaussian σ, though not directly comparable with R.





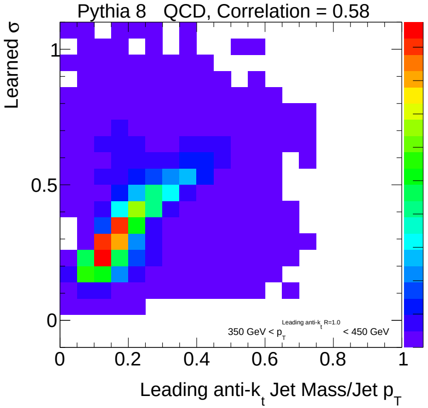
For fixed p_T, fuzzy jets is learning the sizes for these processes!

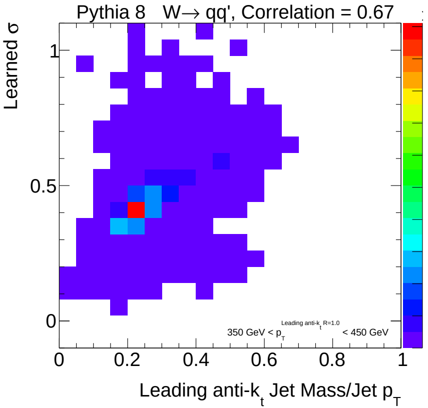
The subleading jets tend to be larger on average than the leading jets.

† With modification, size can vary in sequential recombination - see *Jets with Variable R*

Correlations between σ and anti-k_t jet mass

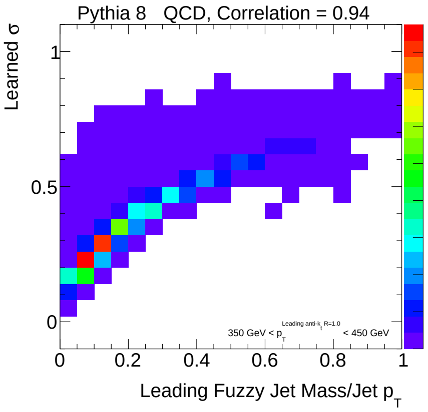
As expected, for a fixed p_T, the higher mass jets have a larger size.

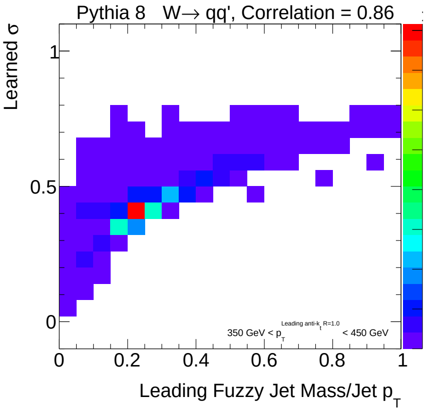




However, the correlation is not 100% - the fuzzy jet σ is not just the characteristic anti-k_t jet size.

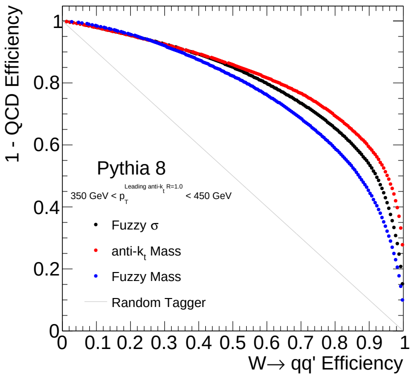
Even though σ has a component uncorrelated with the characteristic anti-k_t jet size, it is highly correlated with the characteristic fuzzy jet size.

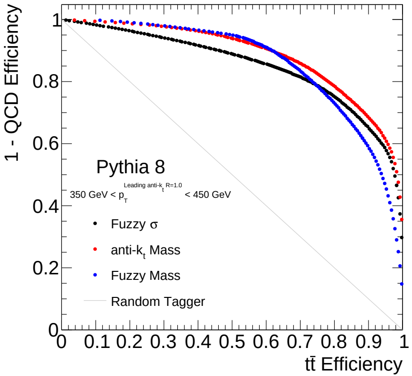




Tagging Hadronic W boson and top quark decays

In some regimes, the uncorrelated component of σ (and the mass) may be useful in discriminating heavy particle decays.





Still need to understand correlations with other (substructure) variables.

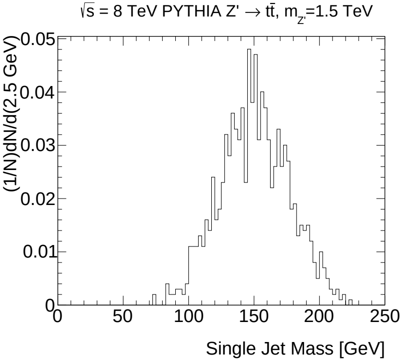
Extensions for Fuzziness

Today's focus was on learning. In order to fully take advantage of the fuzzy part of mMM jets, one needs to incorporate correlations.

The generative model for mMM jets assume independence – fuzzy jets 'volatility' (à la Qjets) could be more powerful with correlations.

There are unsupervised learning algorithms that do this, but scale with complexity.

Incorporating correlations may necessary to learn the tree structure of QCD radiation.

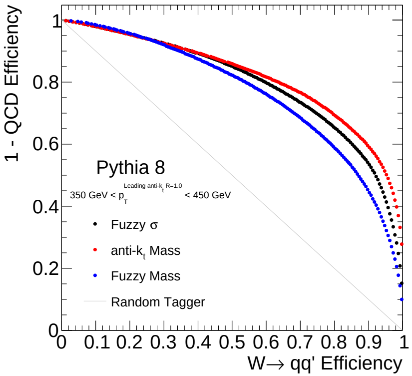


With $f_i(j)$, can compute a *distribution* for any jet observable. Details ▸ in the backup

Conclusions

Fuzzy jets are an alternative clustering scheme that naturally incorporate probabilistic membership to learn features of hadronic final states.

- Given any probability density Φ that only depends on y and ϕ, there is a corresponding modified mixture model jet.
- Unlike sequential recombination schemes, these algorithms learn the shape of jets.
- In some regimes, this learning may help improve tagging performance.



Further exploration and comparisons are required to fully understand and exploit the potential of **Fuzzy Jets**!

BACKUP

Table of Contents

Cluster Merging for mGMM Jets

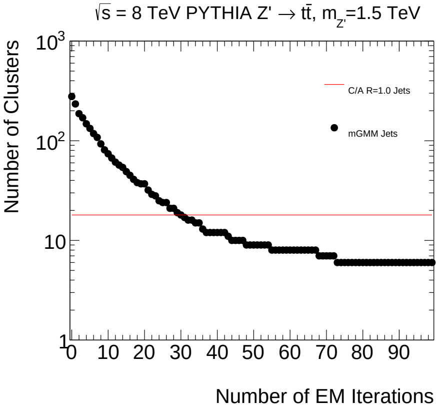
Back to main slides

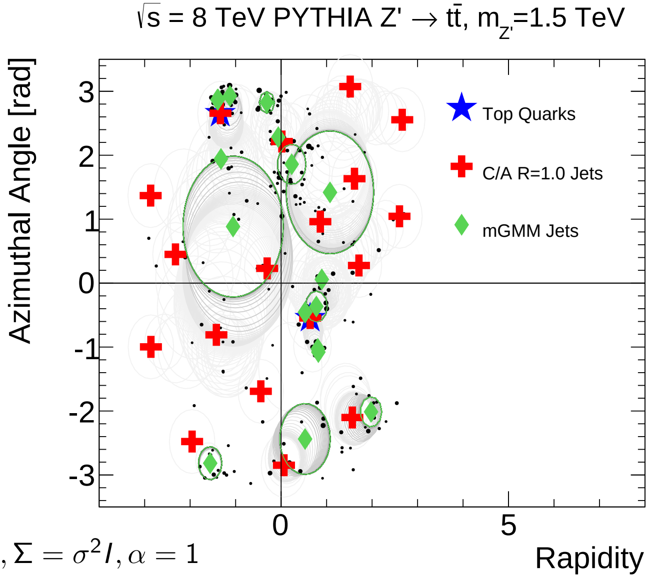
Need to fix a parameter r_{merge}. We find that if small enough, the actual value does not matter.

(chosen here to be $r = 0.01$)

With the appropriate choices of σ and α, one recovers a set of jets very similar to C/A (only depend on $\vec{\rho}$).

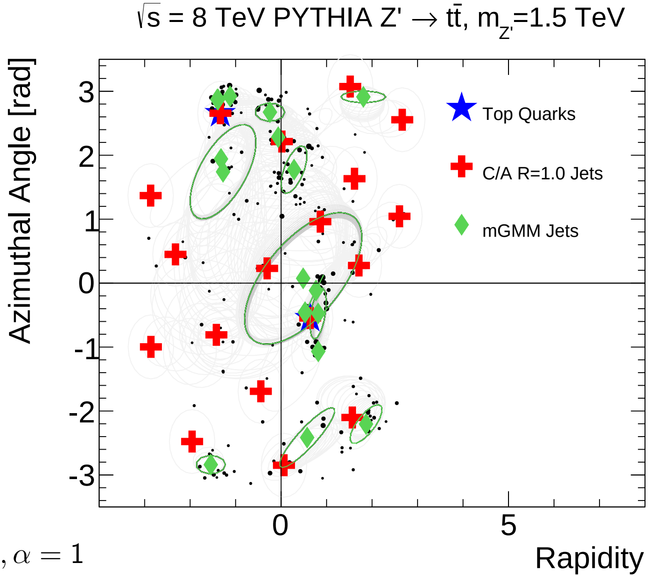
The next two slides show some examples of event displays with the clustering history overlaid, with learning $\Sigma = \sigma^2 I$ and a full covariance matrix Σ.





$\pi \propto 1, \Sigma = \sigma^2 I, \alpha = 1$

Cluster Merging Scheme



$\pi \propto 1, \alpha = 1$

Cluster Merging Scheme

Jet Observables with Membership Functions

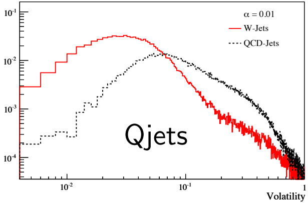
Membership functions can generate *distributions* for each jet observable.

$$\Pr(\mathcal{O}(J) \leq c) = \sum_{\mathcal{S}} \Pr(\mathcal{O}(J) \leq c | \mathcal{S} \in J) \Pr(\mathcal{S} \in J)$$

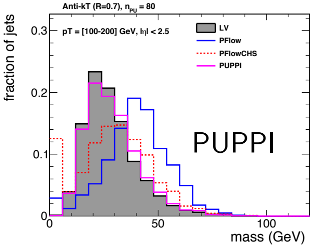
J is a jet
$\mathcal{O}$ is a jet observable (like mass)
$\mathcal{S}$ is a set of constituents
$\Pr(\mathcal{S} \in J)$ is determined by the $f_i(J)$

From this distribution, one can use various moments as *new jet observables*.

e.g. coefficient of variation for the mass (Qjets volatility)



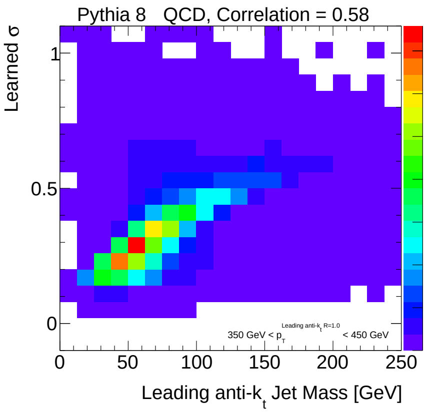
arXiv:1201.1914

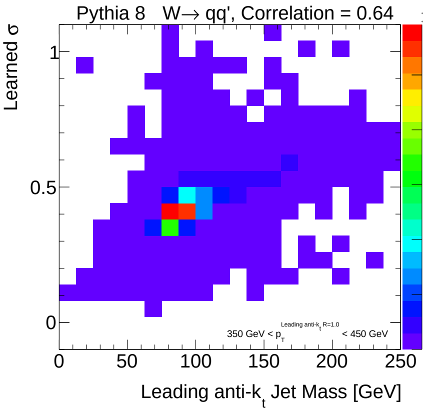


arXiv:1407.6013

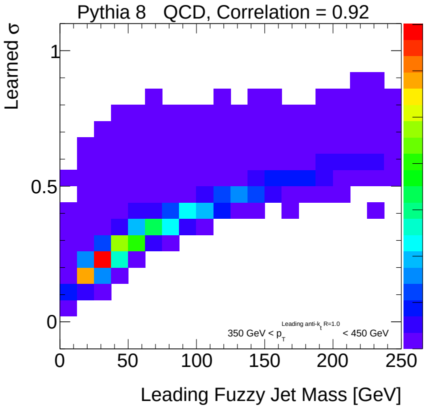
Correlations between σ and anti-k_t jet mass

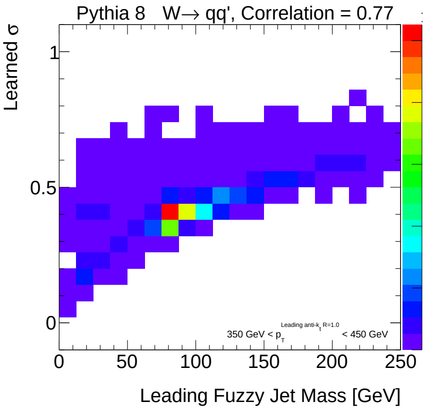
Back to TOC





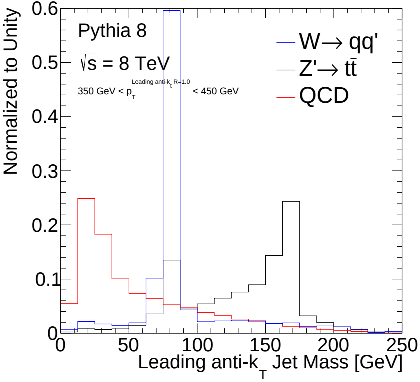
Correlations between σ and fuzzy jet mass

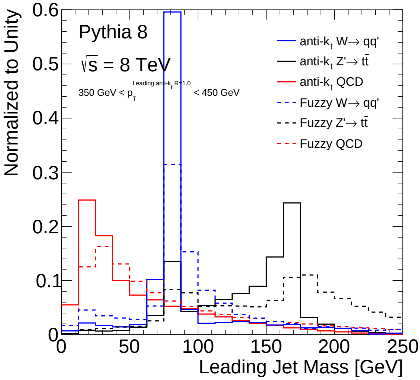




Comparisons with Trimmed Jet Mass

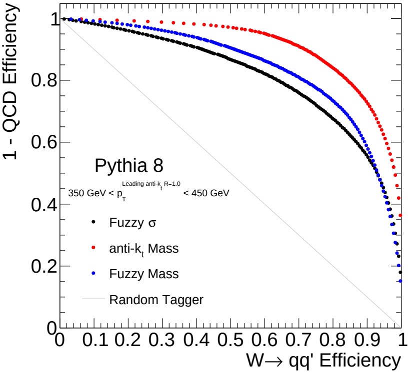
Back to TOC

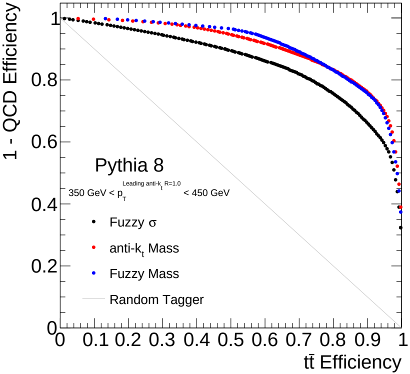




Comparisons with Trimmed Jet Mass II

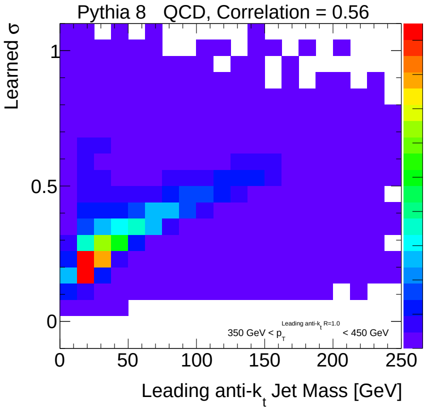
Back to TOC

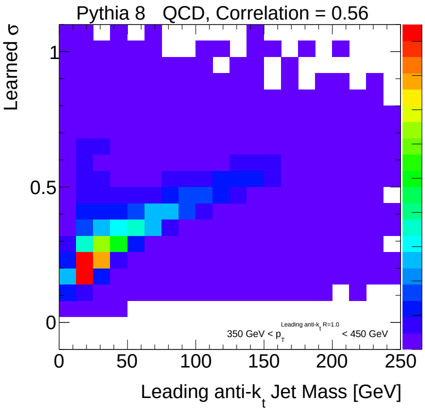


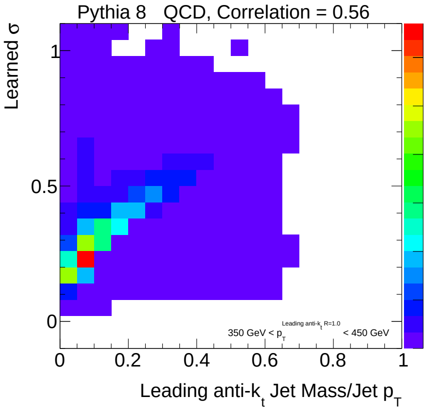


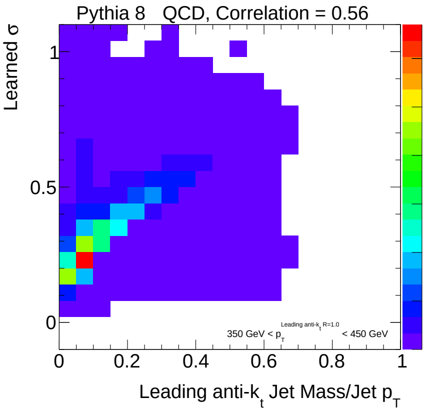
Comparisons with Trimmed Jet Mass III

Back to TOC



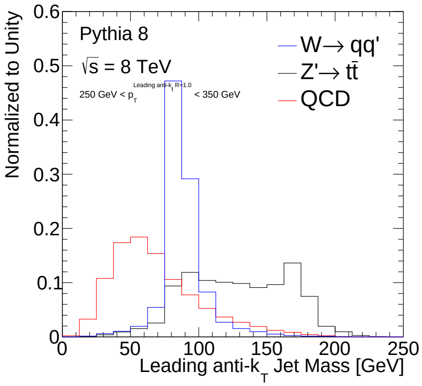



Pythia 8 QCD, Correlation = 0.56
Learned σ
1
0.5
0
0 0.2 0.4 0.6 0.8 1
Leading anti-k_t Jet Mass/Jet p_T
350 GeV < $p_T^{\text{Leading anti-}k_t\ R=1.0}$ < 450 GeV


Pythia 8 QCD, Correlation = 0.56
Learned σ
1
0.5
0
0 0.2 0.4 0.6 0.8 1
Leading anti-k_t Jet Mass/Jet p_T
350 GeV < $p_T^{\text{Leading anti-}k_t\ R=1.0}$ < 450 GeV

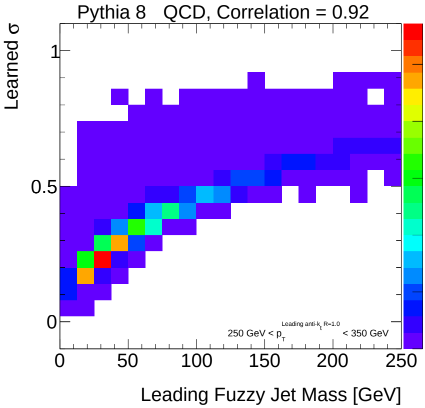
Lower p_T bin: Jet Mass

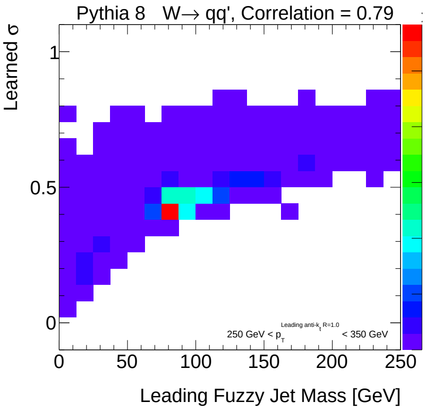
Back to main slides



Lower p_T bin: Fuzzy Mass and σ

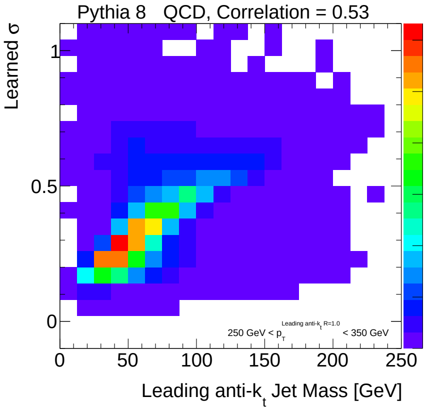
Back to main slides

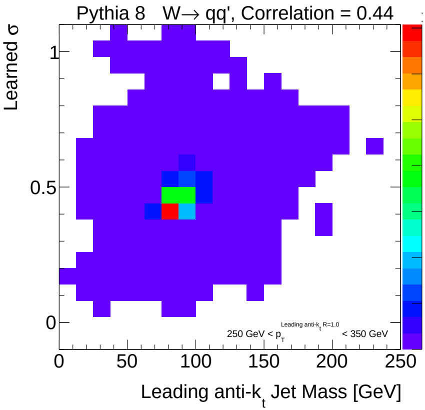




Lower p_T bin: Jet Mass and σ

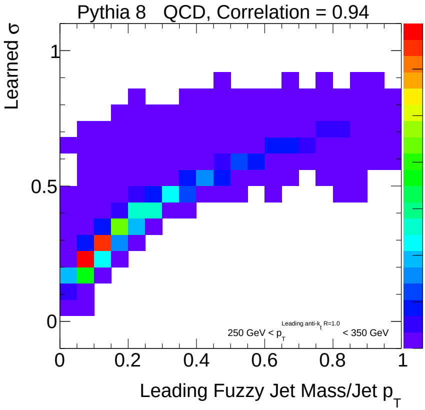
Back to main slides

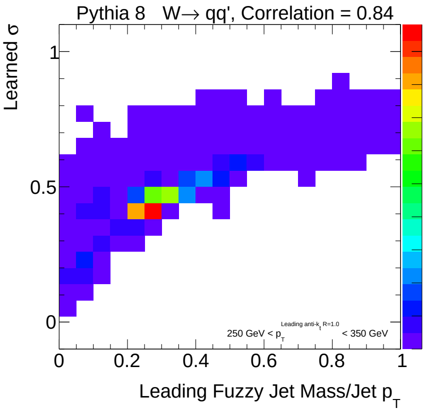




Lower p_T bin: Fuzzy Mass/p_T and σ

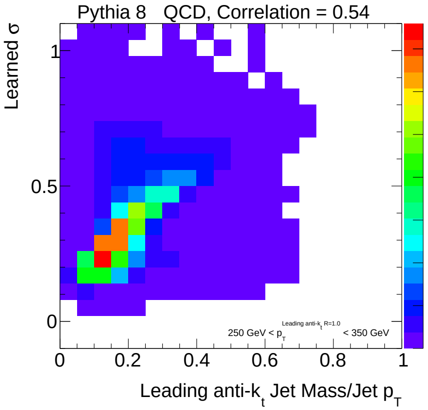
Back to main slides

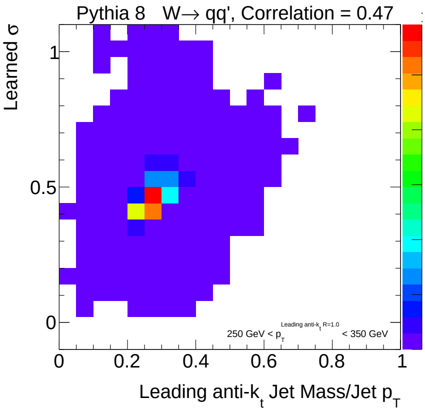




Lower p_T bin: Jet Mass/p_T and σ

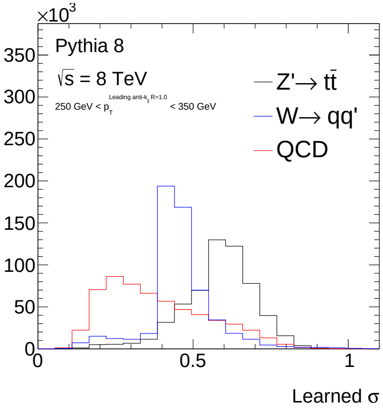
▸ Back to main slides

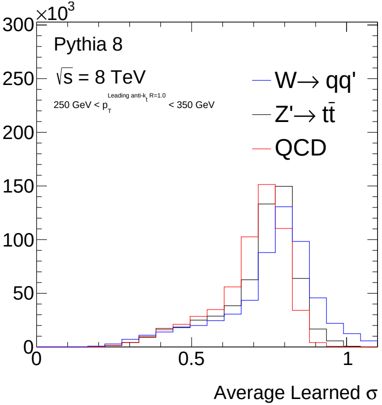




Lower p_T bin: Learned σ

▸ Back to main slides

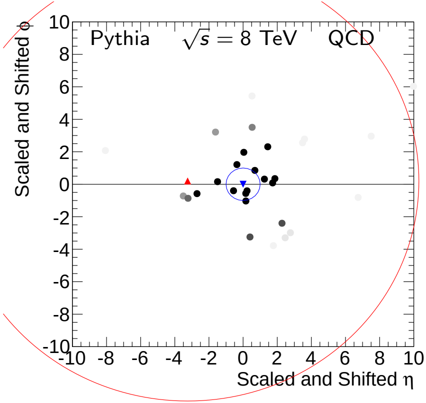




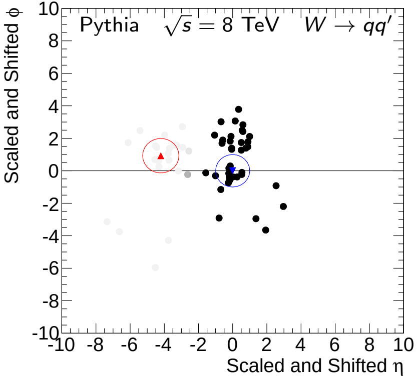
Learning Substructure

One can apply the same techniques to learn the shapes of jet substructure. For example, $k = 2$ fuzzy jets with anti-k_t $R = 1.0$ constituents as input:

Coordinates rotated and scaled - leading subjet is at $(0, 0)$ and has $\sigma = 1$.



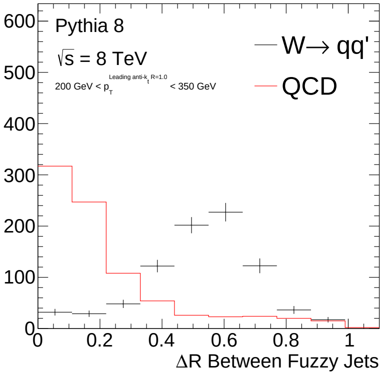
One small and one broad subjet

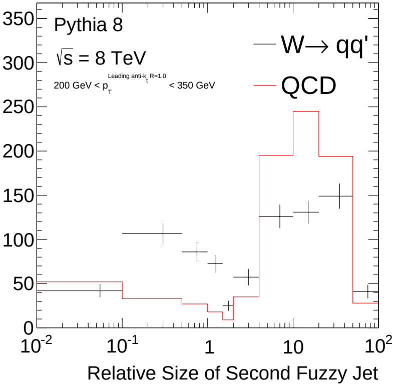


Relatively symmetric subjets

Learning Substructure II

Two observables based on the representative event displays: ΔR between learned fuzzy jets and the relative size of the second fuzzy subjet (σ_2/σ_1).



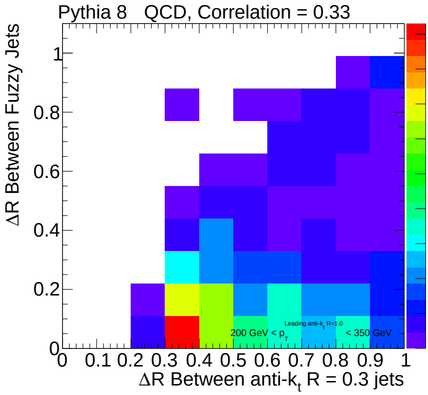


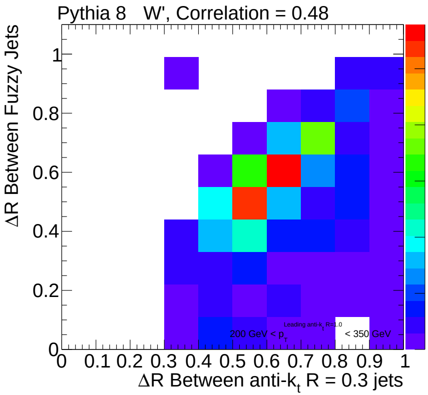
W Further apart (correlated with ΔR of seeded R=0.3 anti-k_t jets)

W subjets are more symmetric in size than for QCD.

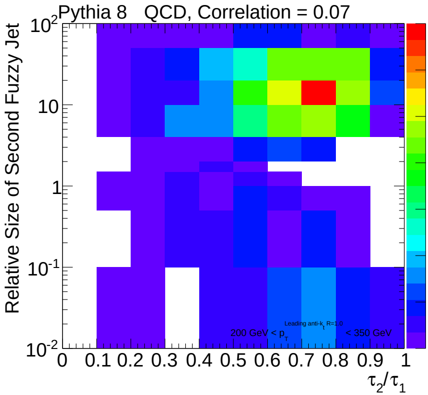
More in the backup

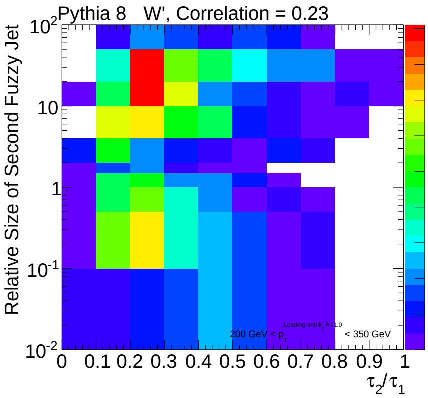
Back to main slides





SLAC


Pythia 8 QCD, Correlation = 0.07
Relative Size of Second Fuzzy Jet
10^{2}
10
1
10^{-1}
10^{-2}
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
τ_2/τ_1
Leading anti-k_t R=1.0
200 GeV < p_T < 350 GeV


Pythia 8 W', Correlation = 0.23
Relative Size of Second Fuzzy Jet
10^{2}
10
1
10^{-1}
10^{-2}
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
τ_2/τ_1
Leading anti-k_t R=1.0
200 GeV < p_T < 350 GeV