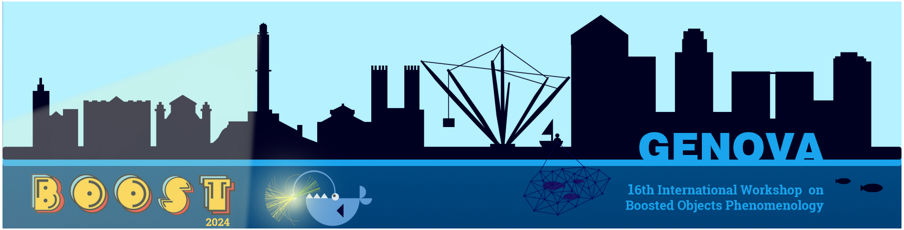
BOOST Camp (theory)

Giovanni Stagnitto

Milano Bicocca University & INFN



from the BOOST 2024 website:

BOOST 2024 is the 16th conference of a series of successful joint theory/experiment workshops that bring together the world's leading experts from theory and LHC/RHIC experiments to discuss the latest progress and develop new approaches on the reconstruction of and use of jet substructure to study Quantum Chromodynamics (QCD) and look for physics beyond the Standard Model.

from the BOOST 2024 website:

BOOST 2024 is the 16th conference of a series of successful joint theory/experiment workshops that bring together the world's leading experts from theory and LHC/RHIC experiments to discuss the latest progress and develop new approaches on the reconstruction of and use of jet substructure to study Quantum Chromodynamics (QCD) and look for physics beyond the Standard Model.

Outline

Discalaimer

The referencing is minimal. Apologies for any relevant omission.

I have taken inspiration (and stolen material) from:

- lectures by Matteo Cacciari and Gavin Salam
- previous BOOST Camps
- the two books in the Conclusion slides

I will not follow an historical approach, rather I will focus on concepts.

I will mostly present textbook knowledge, rather than topics still in developments.

Outline

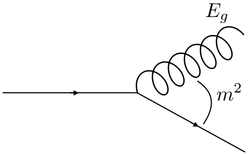
All the QCD we will need in this talk

Larkoski (1709.06195)

Assumption: at high energies,

1. the coupling of QCD, α_s, is small, so we can use perturbation theory.
2. QCD has no intrinsic scale, it is a scale-invariant quantum field theory.

Probability for a quark to emit a gluon:



Two degrees of freedom, E_g and m^2, with

$$m^2 = 2p_q \cdot p_g = 2E_q E_g(1 - \cos\theta)$$

Scale invariance means that:

$$P(\lambda E_g, \lambda^2 m^2)\mathrm{d}(\lambda E_g)\mathrm{d}(\lambda^2 m^2) = P(E_g, m^2)\mathrm{d}E_g \mathrm{d}m^2$$

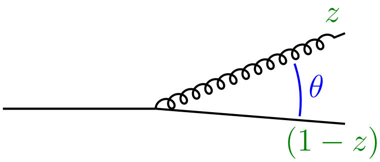
All the QCD we will need in this talk

Larkoski (1709.06195)

The simplest form turns out to be the correct one:

$$P(E_g, m^2)\mathrm{d}E_g\mathrm{d}m^2 = \frac{\alpha_s C_i}{\pi}\frac{\mathrm{d}E_g}{E_g}\frac{\mathrm{d}m^2}{m^2}$$

C_i is a colour factor: $C_i = C_F = 4/3$ for $q \to qg$ and $C_i = C_A = 3$ for $g \to gg$.



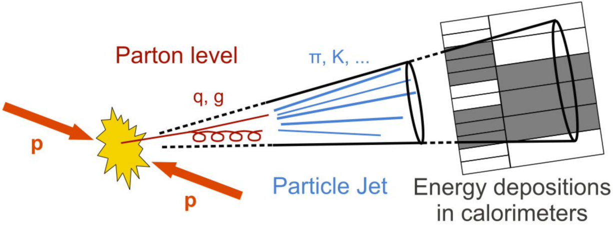
In term of dimensionless quantities, θ and $z = E_g/(E_q + E_g)$, and by taking the small angle limit, $\theta \ll 1$

$$P(z, \theta^2)\mathrm{d}z\mathrm{d}\theta^2 = \frac{\alpha_s C_F}{\pi}\frac{\mathrm{d}z}{z}\frac{\mathrm{d}\theta^2}{\theta^2}$$

QCD dynamics favours emission of soft ($z \to 0$) and/or collinear ($\theta \to 0$) particles.

What is a jet?

A jet is the macroscopic manifestation of QCD dynamics at high energies
i.e. most of the particles are soft or tend to be emitted at small angles



Naive definition: collimated bunch of hadrons flying roughly in the same direction
Proper definition: a collection of hadrons defined by means of a jet algorithm
At the LHC we usually adopt *sequential recombination clustering algorithms*
that can applied to objects at parton, particle or detector level.

Example: the gen-k_t family of clustering algorithms

Given distances d_{ij} and beam distances d_{iB} defined as:

$$d_{ij} = \min\left(p_{ti}^{2p}, p_{tj}^{2p}\right) \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = p_{ti}^{2p}$$

with transverse momentum p_t and angular distance $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with y rapidity y and azimuthal angle ϕ, apply the following algorithm:

1. identify all initial objects as *pseudo-jets*
2. find the minimum distance:
   - d_{ij}: recombine the pseudo-jet (i, j) into a new pseudo-jet k by summing the 4-momenta and update the distances
   - d_{iB}: declare the pseudo-jet i as a final jet and remove all distances involving i.
3. iterate until there are no pseudo-jets left.

The parameter R is called *jet radius* (usually taken between 0.4 and 1)

Example: the gen-k_t family of clustering algorithms

The value of the parameter p defines the algorithm:

$$d_{ij} = \min\left(p_{ti}^{2p}, p_{tj}^{2p}\right) \frac{\Delta R_{ij}^2}{R^2}\,, \quad d_{iB} = p_{ti}^{2p}$$

- $p = 1$: k_t algorithm
- $p = 0$: Cambridge/Aachen (C/A) algorithm
- $p = -1$: anti-k_t algorithm

Example: the gen-k_t family of clustering algorithms

The value of the parameter p defines the algorithm:

- $p = 1$: k_t algorithm $\rightarrow$ mass/virtuality ordering

$$d_{ij} = \min\left(p_{ti}^2, p_{tj}^2\right) \frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = p_{ti}^2$$

  Distance measure reflects the splitting probability $P_{k \rightarrow ij}$:

$$P_{k \rightarrow ij} \sim \frac{\alpha_s}{\min\left(p_{ti}^2, p_{tj}^2\right)\ \Delta R_{ij}^2} \sim \frac{1}{d_{ij}}$$

  (at the LHC we use variables invariant under longitudinal boosts, such as p_t and Δ_R; energies and angles are not invariant)

- $p = 0$: Cambridge/Aachen (C/A) algorithm
- $p = -1$: anti-k_t algorithm

Example: the gen-k_t family of clustering algorithms

The value of the parameter p defines the algorithm:

- $p = 1$: k_t algorithm $\rightarrow$ mass/virtuality ordering
- $p = 0$: Cambridge/Aachen (C/A) algorithm $\rightarrow$ pure angular ordering

$$d_{ij} = \frac{\Delta R_{ij}^2}{R^2}\,, \quad d_{iB} = 1$$

- $p = -1$: anti-k_t algorithm

Example: the gen-k_t family of clustering algorithms

The value of the parameter p defines the algorithm:

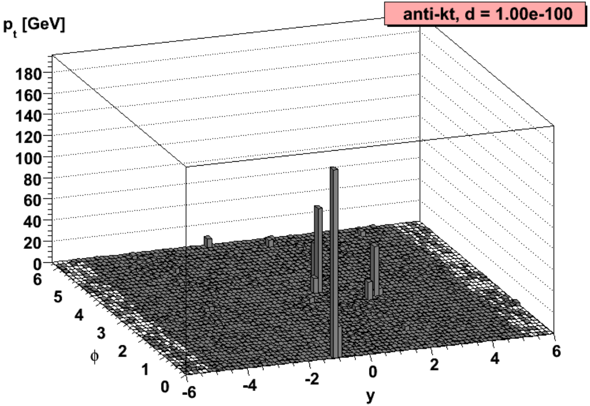
- $p = 1$: k_t algorithm $\rightarrow$ mass/virtuality ordering
- $p = 0$: Cambridge/Aachen (C/A) algorithm $\rightarrow$ pure angular ordering
- $p = -1$: anti-k_t algorithm $\rightarrow$ unphysical?

$$d_{ij} = \min\left(\frac{1}{p_{ti}^2}, \frac{1}{p_{tj}^2}\right)\frac{\Delta R_{ij}^2}{R^2}, \quad d_{iB} = \frac{1}{p_{ti}^2}$$

It tends to favour clustering involving hard particles rather than first recombining soft particles or particles close in angle.

Anti-k_t in action

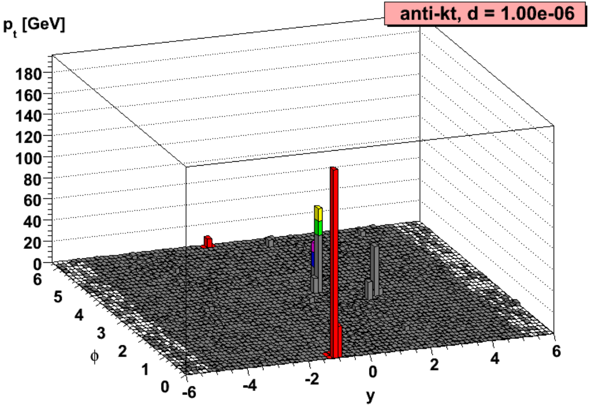
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

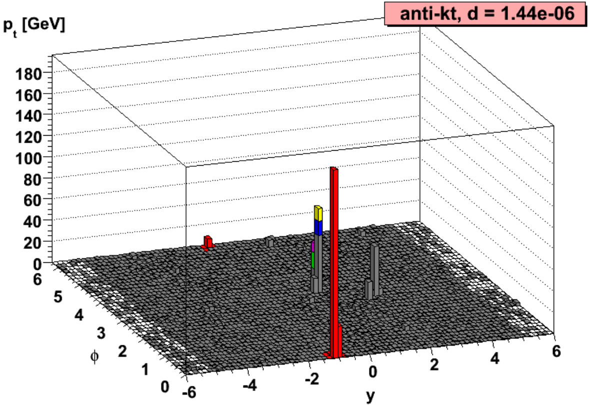
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

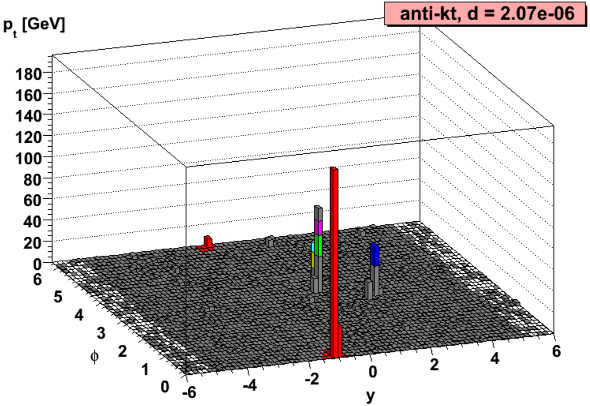
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

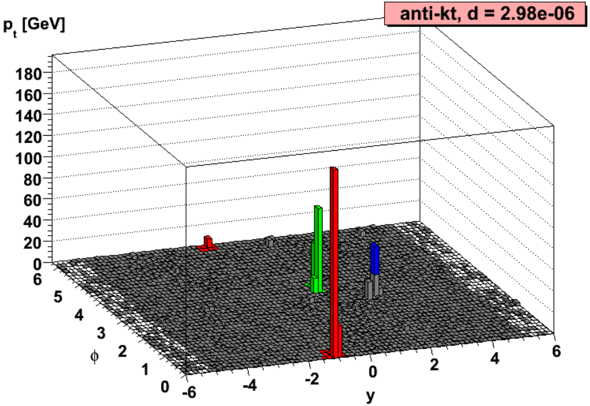
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

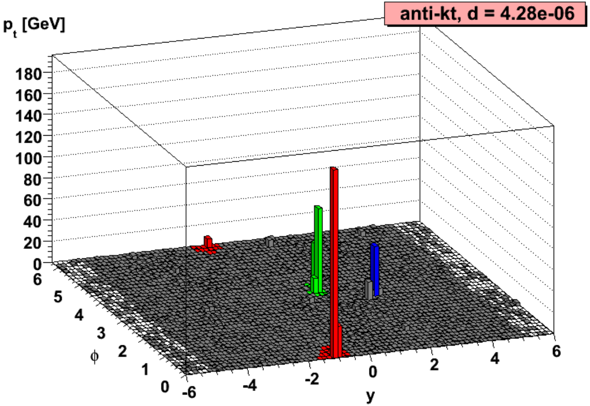
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

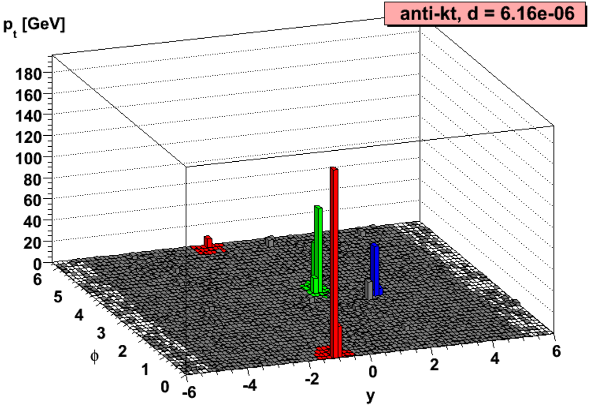
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

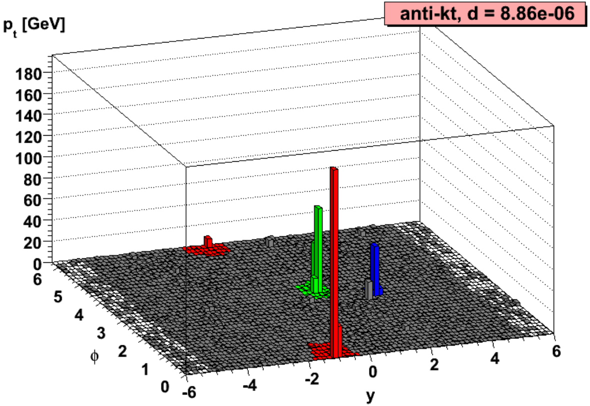
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

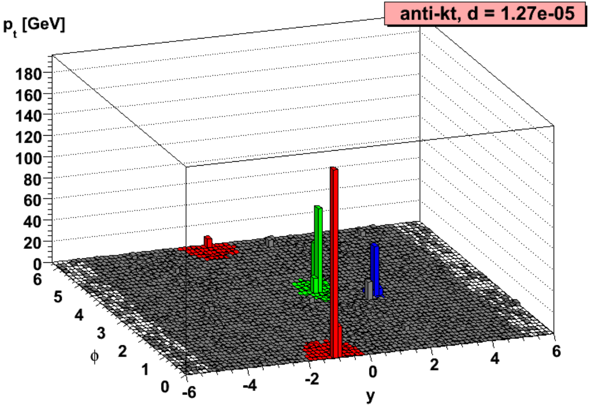
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

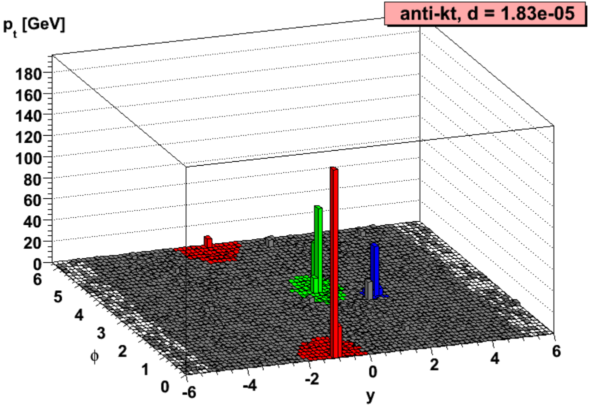
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

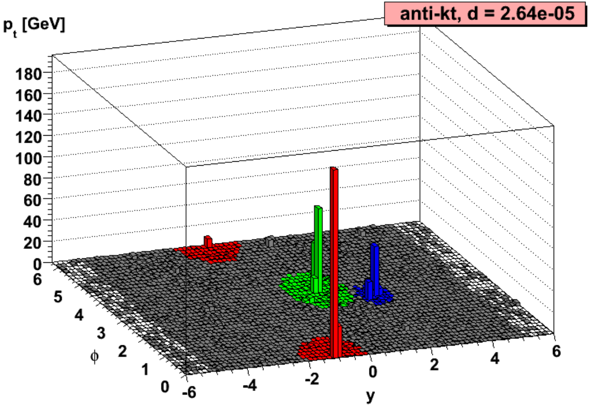
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

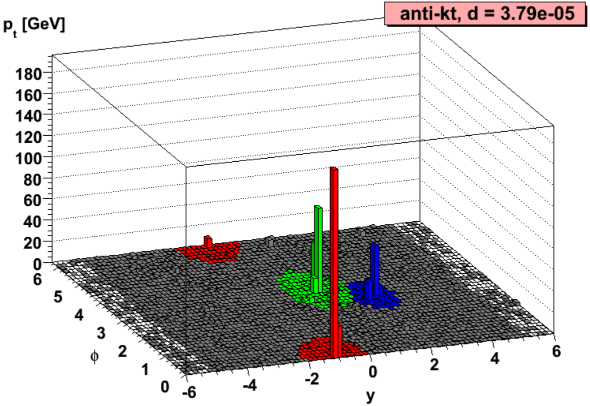
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

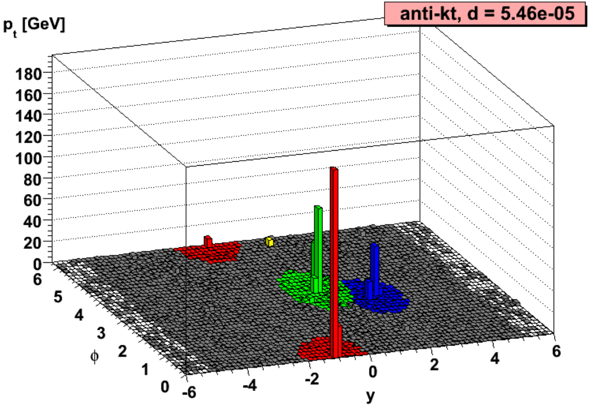
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

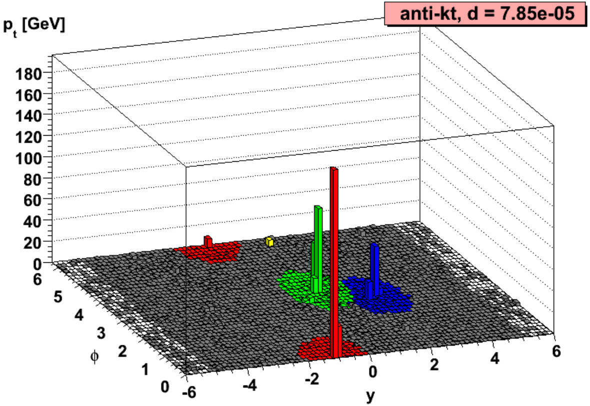
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

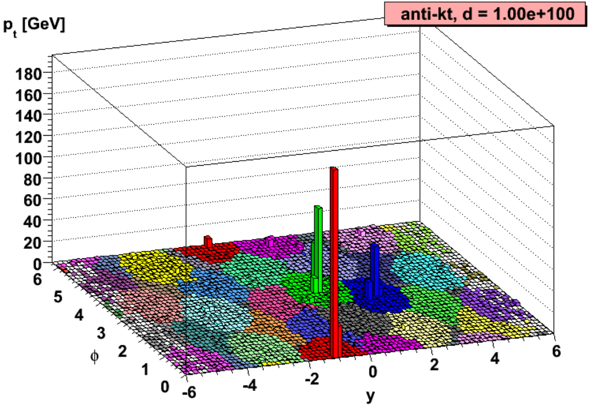
Animations by M. Cacciari



Clustering growing around hard cores

Anti-k_t in action

Animations by M. Cacciari



Circular jets in a theory-friendly way

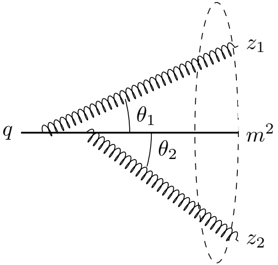
Warmup calculation: jet mass

Defined as the squared sum of the 4-momenta of all particles in a jet. At leading order (a single quark or gluon), it is zero. Jets acquire mass through showering!

$$m^2 = \left(\sum_{i \in jet} k_i \right)^2 .$$

In the collinear limit, it can be written as:

$$m^2 = R^2 E_J^2 \sum_{i \in jet} z_i \, \theta_i^2$$



with z_i the momentum fraction of the emission (rescaled by the jet energy E_J) and θ_i the angular distance from the jet axis (rescaled by the jet radius R). We define

$$\rho = \sum_{i \in jet} z_i \, \theta_i^2 = \sum_{i \in jet} \rho_i$$

Warmup calculation: jet mass

Cumulative distribution $\Sigma(\rho)$ ($=$ probability of measuring a mass $< \rho$)

$$\Sigma(\rho) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\prod_{i=1}^{n} \int \mathrm{d}z_i \mathrm{d}\theta_i^2 P(z_i, \theta_i^2) \left[\Theta(\theta_i < 1) \Theta\left(\sum_{j=1}^{n} \rho_j < \rho \right) + \Theta(\theta_i > 1) - 1 \right] \right)$$

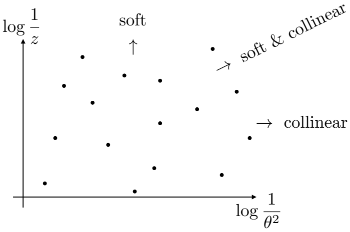
- We sum over all possible number of emissions, with each emission coming with a probability $P(z_i, \theta_i^2)$
- Real emissions inside jet $(\theta_i < 1)$ are allowed only if they leads to a value of the jet mass less than ρ $(\sum_{j=1}^{n} \rho_j < \rho)$
- Real emissions outside jet $(\theta_i > 1)$ as well as virtual emissions ("-1") are always allowed, as they don't affect the value of jet mass

Warmup calculation: jet mass

Assuming gluon emissions off a quark:

$$P(z,\theta^2)\mathrm{d}z\mathrm{d}\theta^2 = \frac{\alpha_s C_F}{\pi}\frac{\mathrm{d}z}{z}\frac{\mathrm{d}\theta^2}{\theta^2} = \frac{\alpha_s C_F}{\pi}\mathrm{d}\left(\log\frac{1}{z}\right)\mathrm{d}\left(\log\frac{1}{\theta^2}\right)$$

Crucial observation: emissions are uniform in the $(\log 1/\theta^2, \log 1/z)$ plane
$\rightarrow$ they are exponentially apart in (θ^2, z) $\rightarrow$ a single emission dominates the jet mass



"Lund" plane

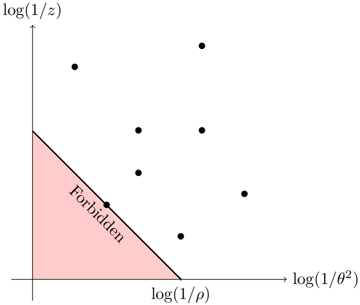
$$\Theta\left(\sum_{i=1}^{n}\rho_i<\rho\right) \simeq \Theta\left(\max_i \rho_i<\rho\right) = \prod_{i=1}^{n}\Theta(\rho_i<\rho)$$

i.e. factorisation of contributions

Warmup calculation: jet mass

By exploiting our observation, we obtain *exponentiation*:

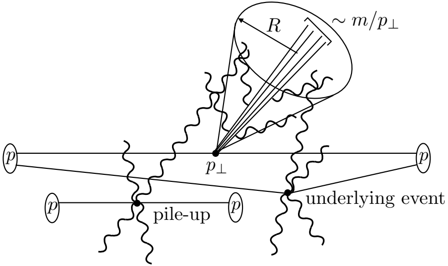
$$\Sigma(\rho) = \sum_{n=0}^{\infty} \frac{1}{n!}\left(-\int \mathrm{d}z_i \mathrm{d}\theta_i^2 P(z_i,\theta_i^2)\,\Theta(\theta_i<1)\Theta(\rho_i>\rho)\right)^n \equiv \exp\left[-V(\rho)\right]$$



Graphical interpretation: we are vetoing all real contributions that would lead to a value of the mass larger than $\rho \to$ only virtual contributions survive in the red forbidden region.
The no-emissions probability $V(\rho)$ is proportional to the area of the forbidden region:

$$V(\rho) = \frac{\alpha_s C_F}{\pi}\int_0^1 \frac{\mathrm{d}z_i}{z_i}\int_0^1 \frac{\mathrm{d}\theta_i^2}{\theta_i^2}\,\Theta(z_i\theta_i^2>\rho) = \frac{\alpha_s C_F}{\pi}\frac{1}{2}\log^2\frac{1}{\rho}$$

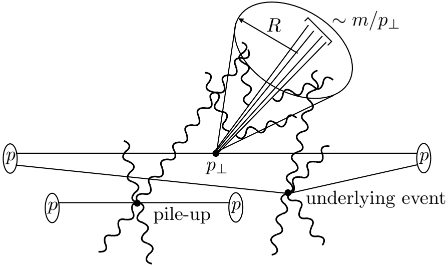
Jet mass at the LHC

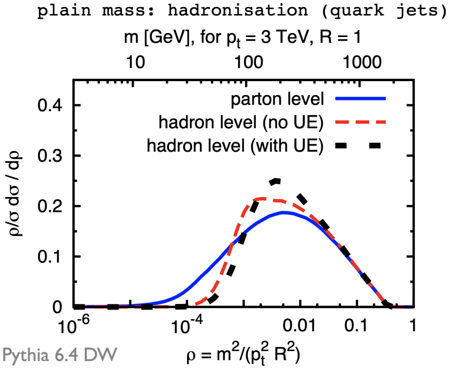


Comparison to data is complicated by non-perturbative physics:

- hadronisation
- underlying event (UE)
- pile-up

Jet mass at the LHC

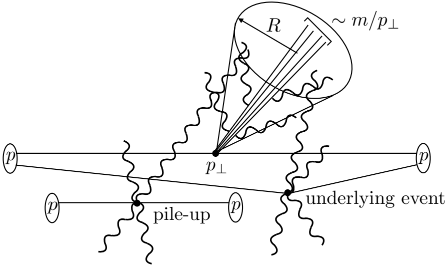




Comparison to data is complicated by non-perturbative physics:

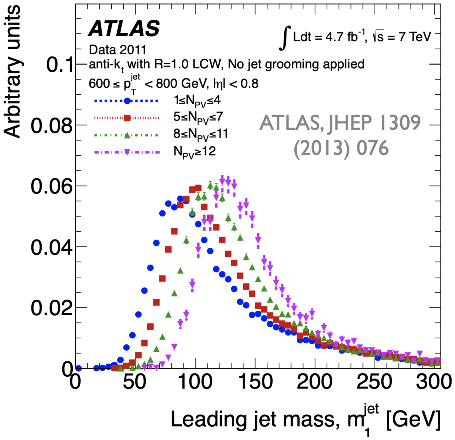
- hadronisation
- underlying event (UE)
- pile-up

Jet mass at the LHC

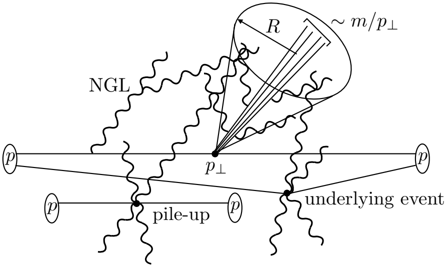


Comparison to data is complicated by non-perturbative physics:

- hadronisation
- underlying event (UE)
- pile-up



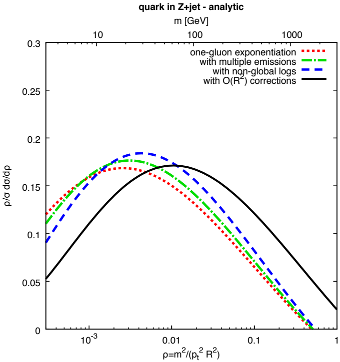
Jet mass at the LHC



Comparison to data is complicated by non-perturbative physics:

- hadronisation
- underlying event (UE)
- pile-up

Also effects of perturbative origin, such as non-global logarithms (NGLs), jet radius corrections, multiple emissions render calculations more complicated



Outline

QCD and jets

Jet substructure
- Boosted objects
- Reclustering
- Tagger and groomers
- Analytic understanding: Soft Drop

Selected topics

Conclusions

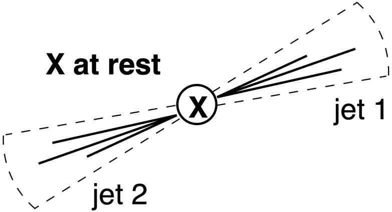
What are boosted objects?

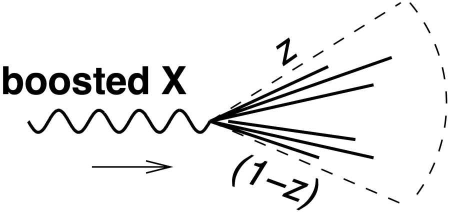
Let's consider the 2-prong decay of a heaavy particle X.

Remember that $m^2 \simeq z(1-z)p_t^2\theta^2 \to \theta^2 \propto \frac{m^2}{p_t^2}$

Standard analysis: $p_t \lesssim m$
Decay products are reconstructed as two separate jets.

Boosted analysis: $p_t \gg m$
Decay products are collimated and they end up in a single jet.

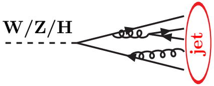




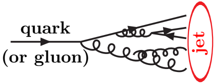
The boosted scenario is common at the LHC
e.g. electroweak particle with mass $m \sim 100$ GeV produced with $p_t \sim 1$ TeV

Jet mass as discriminant?

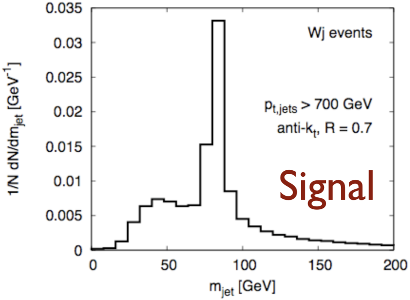
How to discriminate

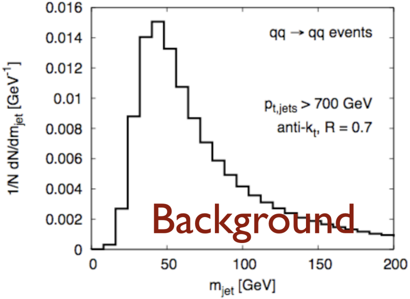


from



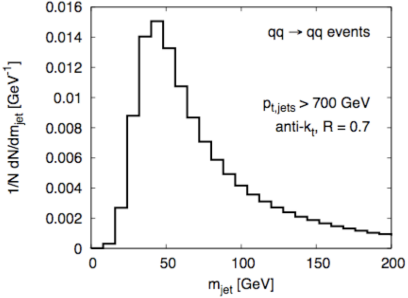
?

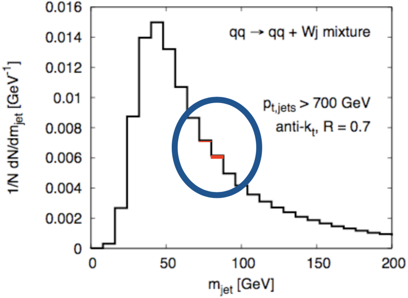




The normalised jet mass distribution peaks around the W-boson mass, whereas the jet mass of the QCD background is peaked towards smaller values of the mass.

Jet mass as discriminant?



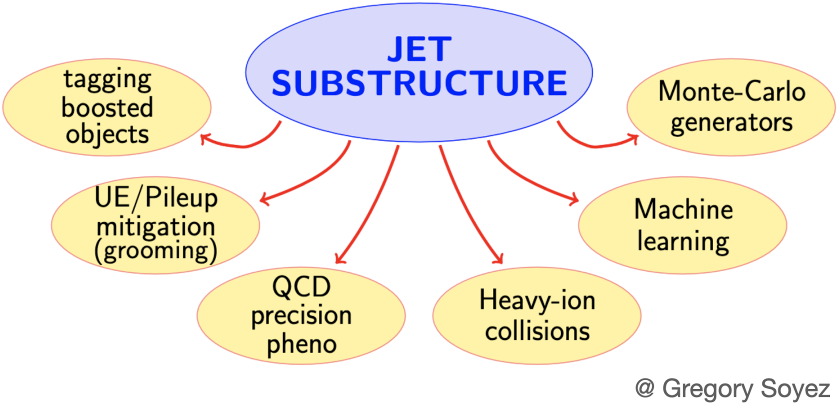


However, when we add signal on top of background with appropriate cross section, the distributions are practically identical!

$\rightarrow$ we need to go beyond the monolithic picture of a jet by studying its substructure i.e. identify hard prongs and quantify amount of radiation around them.

The world of jet substructure

First application was tagging of boosted objects, but nowadays is a very broad topic.



Re-cluster the jet

- The starting point of many jet substructure technique: given a jet, *recluster* it i.e. get the clustering sequence by using an algorithm different from the one adopted from jet reconstruction.

Re-cluster the jet

- The starting point of many jet substructure technique: given a jet, *recluster* it i.e. get the clustering sequence by using an algorithm different from the one adopted from jet reconstruction.
- Usually, jets are first found with anti-k_t with a large radius $R \sim 1$, in order to collect all decay products. But anti-k_t is not suited for jet substructure!

Re-cluster the jet

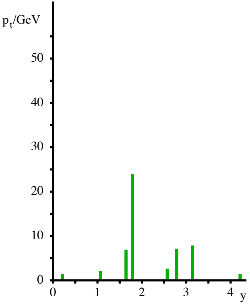
- The starting point of many jet substructure technique: given a jet, *recluster* it i.e. get the clustering sequence by using an algorithm different from the one adopted from jet reconstruction.
- Usually, jets are first found with anti-k_t with a large radius $R \sim 1$, in order to collect all decay products. But anti-k_t is not suited for jet substructure!
- Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event e.g. W decay

Which algorithm is most suited to jet substructure?

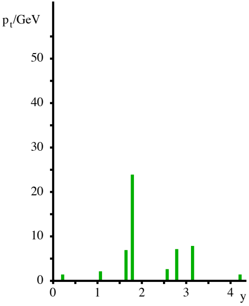
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.

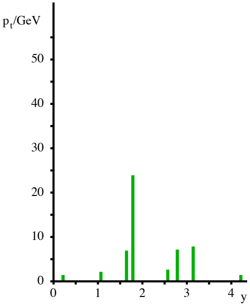
anti-k_t algorithm



k_t algorithm



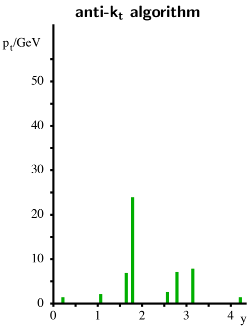
Cambridge/Aachen algorithm

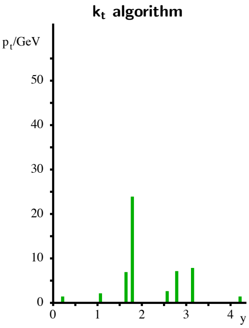


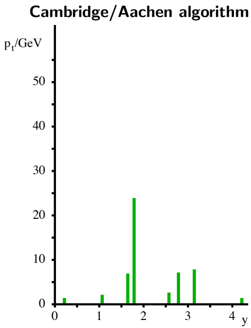
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.





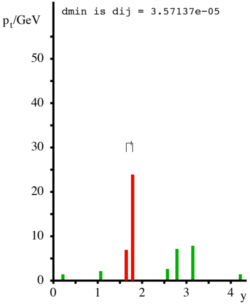


Which algorithm is most suited to jet substructure?

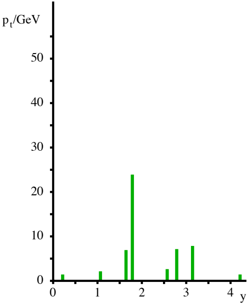
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.

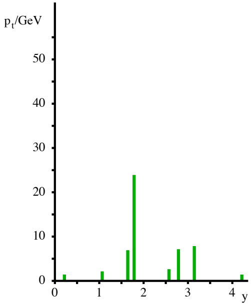
anti-k_t algorithm



k_t algorithm



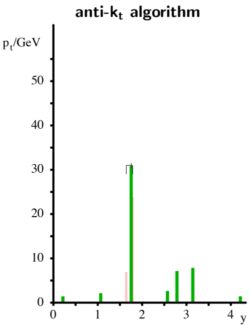
Cambridge/Aachen algorithm

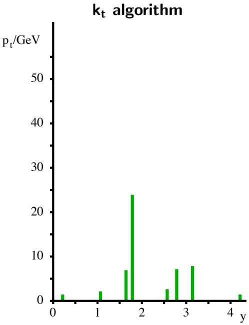


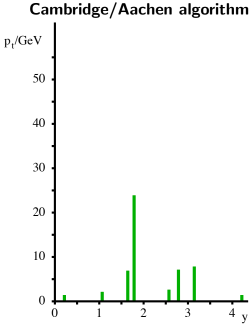
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



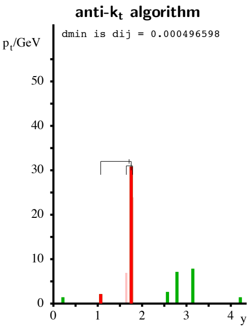


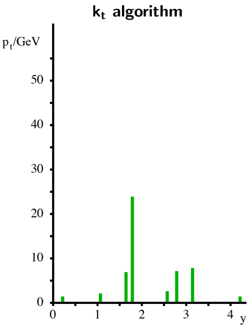


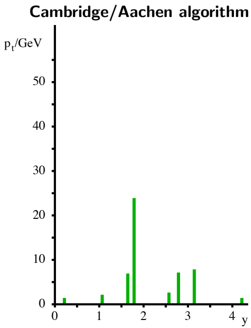
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



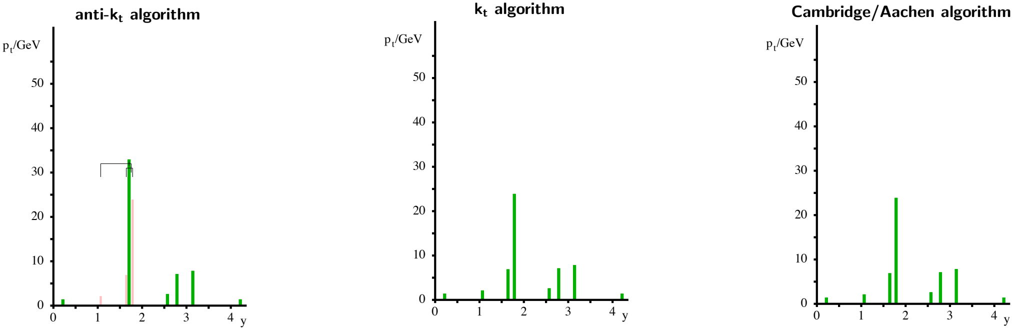




Which algorithm is most suited to jet substructure?

Animations by G. Salam

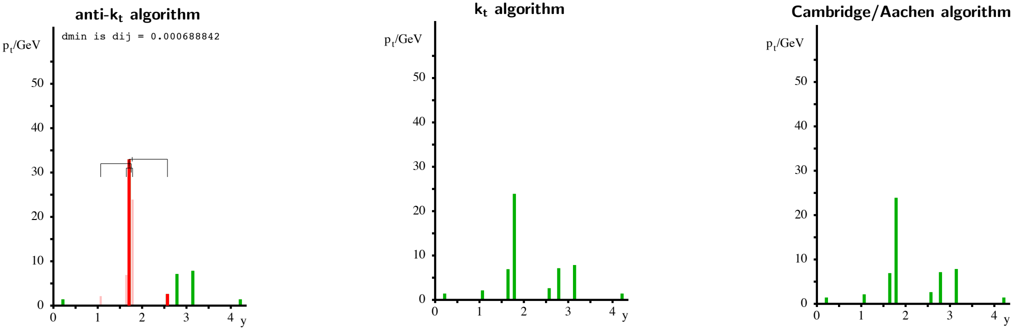
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Which algorithm is most suited to jet substructure?

Animations by G. Salam

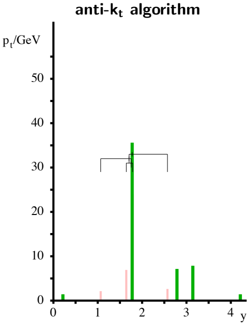
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.

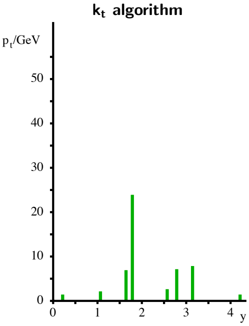


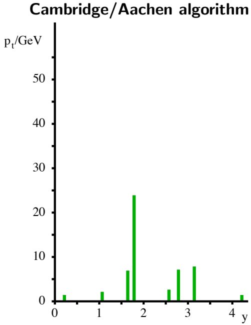
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



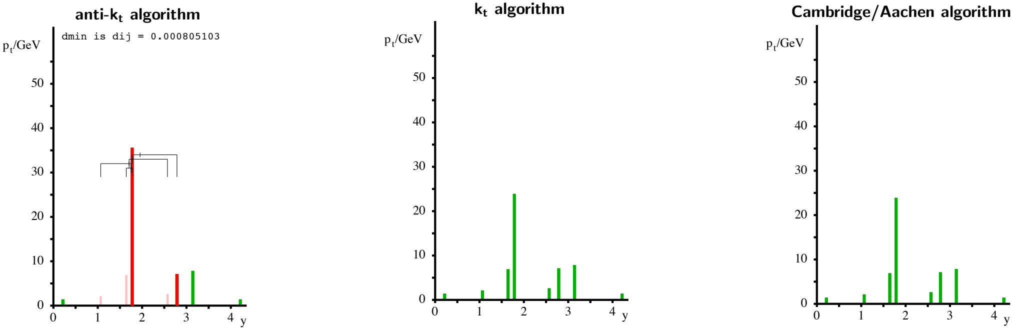




Which algorithm is most suited to jet substructure?

Animations by G. Salam

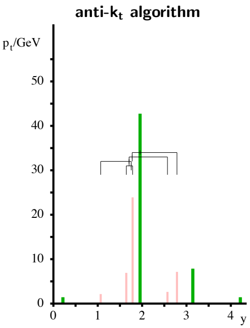
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.

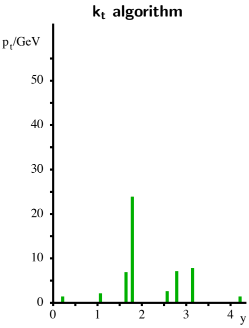


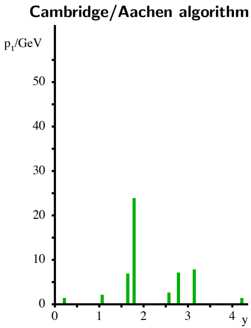
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



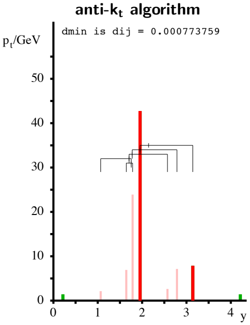


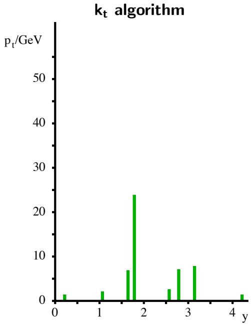


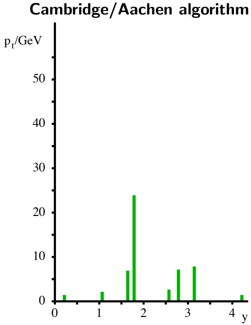
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.





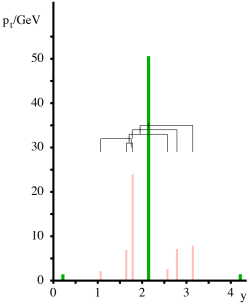


Which algorithm is most suited to jet substructure?

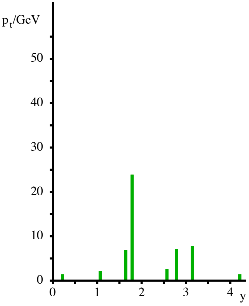
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.

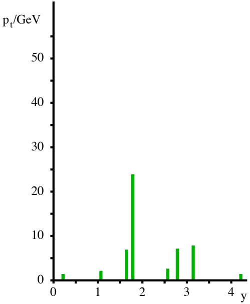
anti-k_t algorithm



k_t algorithm



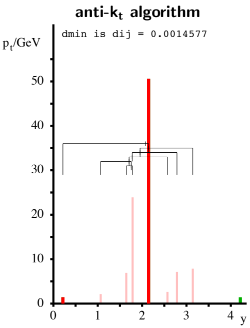
Cambridge/Aachen algorithm

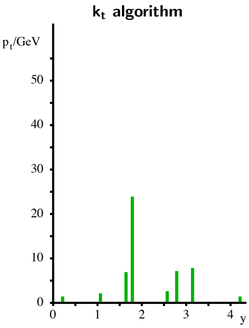


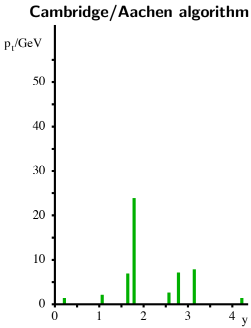
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



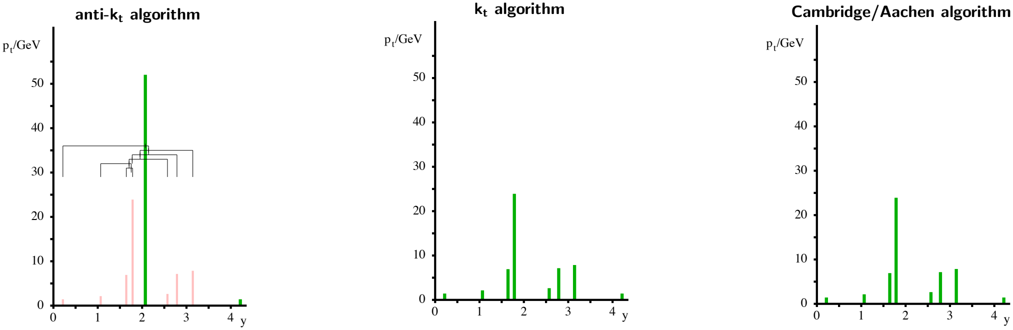




Which algorithm is most suited to jet substructure?

Animations by G. Salam

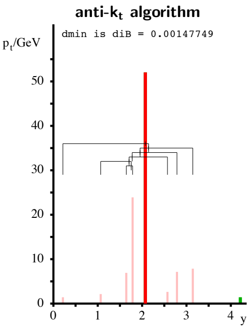
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.

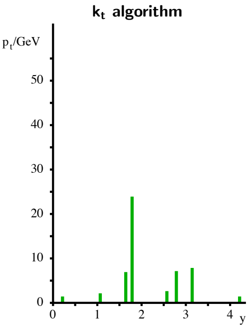


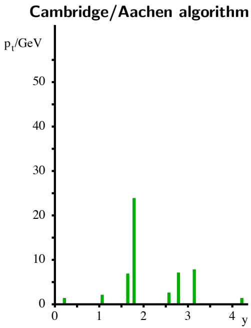
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



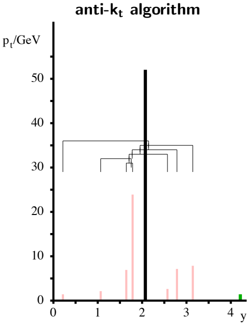


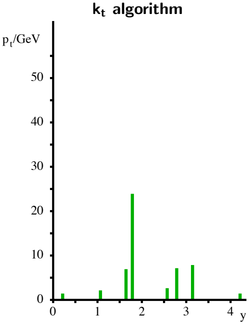


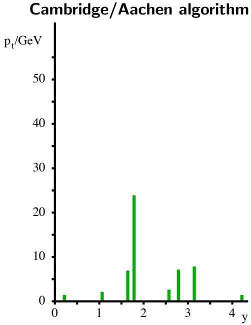
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



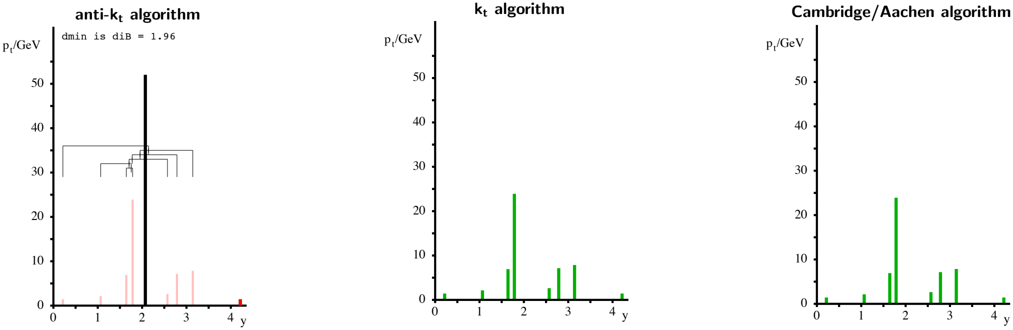




Which algorithm is most suited to jet substructure?

Animations by G. Salam

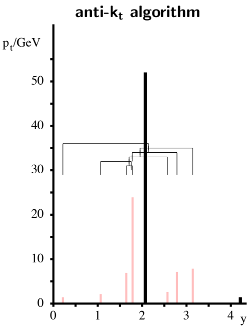
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.

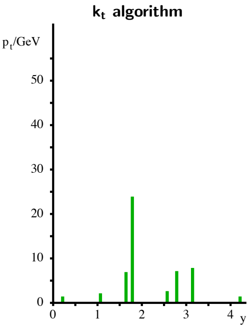


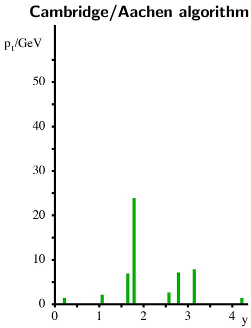
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



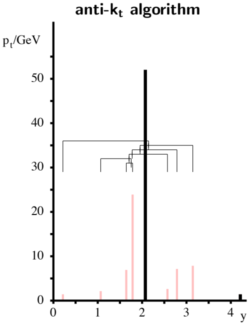




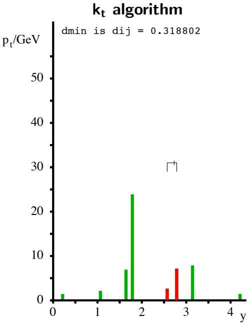
Which algorithm is most suited to jet substructure?

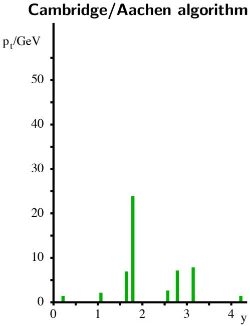
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.

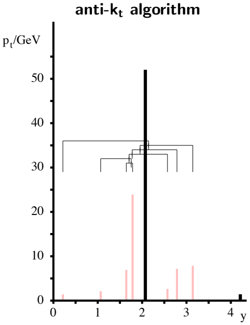




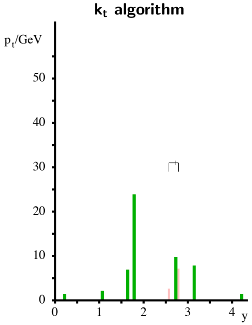
Which algorithm is most suited to jet substructure?

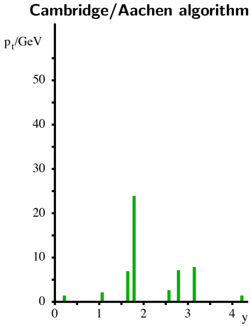
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs lost early in the history.

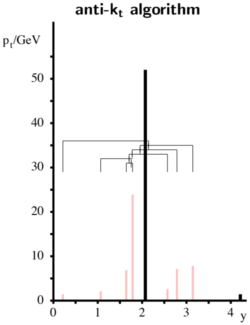




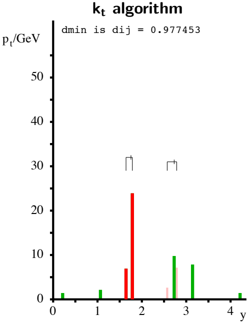
Which algorithm is most suited to jet substructure?

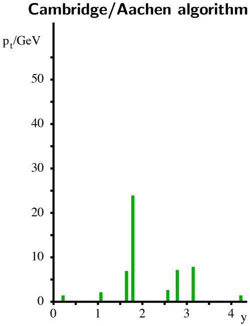
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs
lost early in the history.

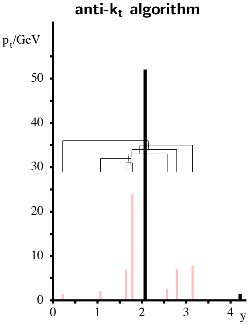




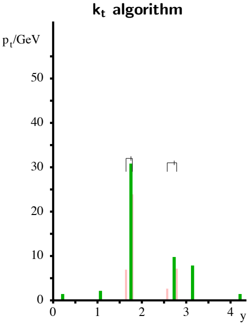
Which algorithm is most suited to jet substructure?

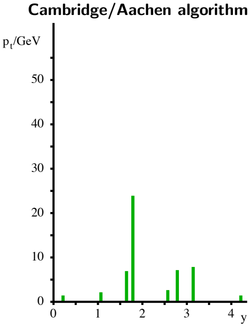
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs lost early in the history.

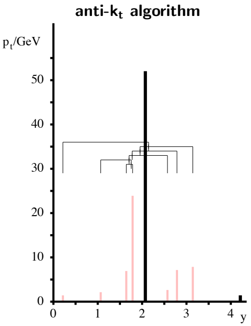




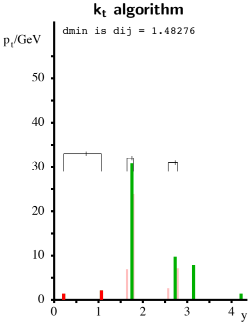
Which algorithm is most suited to jet substructure?

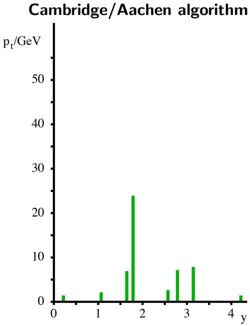
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs
lost early in the history.

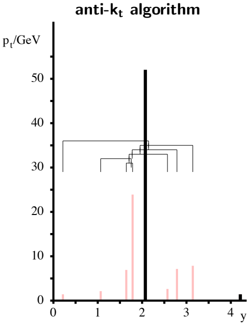


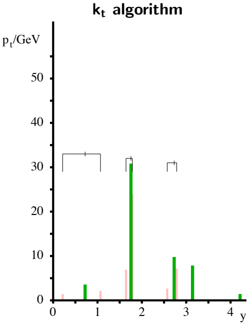


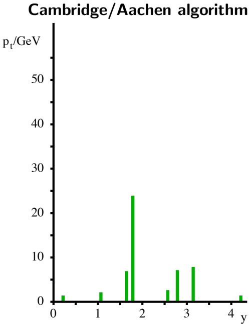
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.





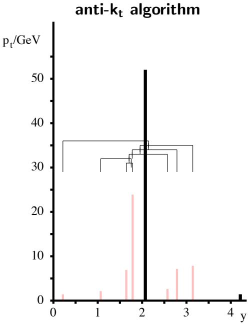


Information about prongs
lost early in the history.

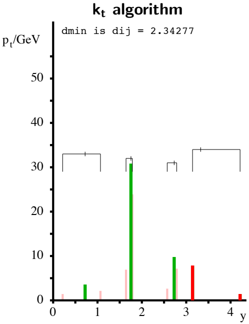
Which algorithm is most suited to jet substructure?

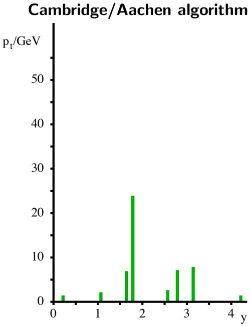
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs
lost early in the history.

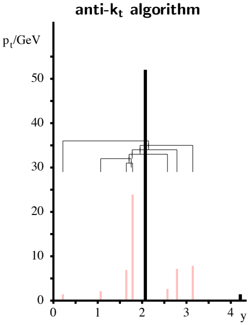




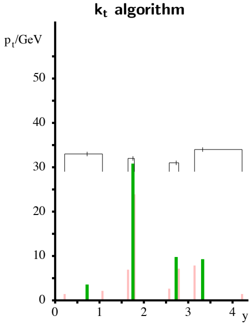
Which algorithm is most suited to jet substructure?

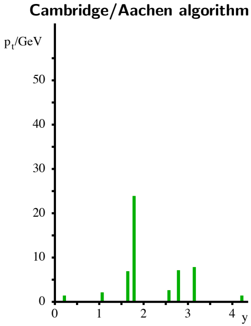
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.

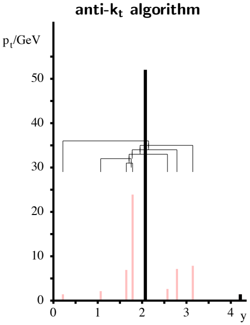




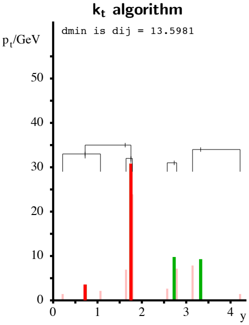
Which algorithm is most suited to jet substructure?

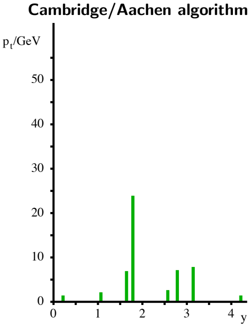
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs
lost early in the history.

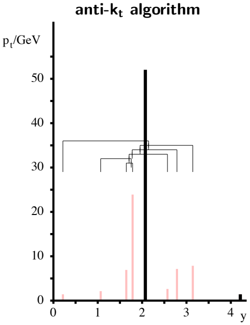




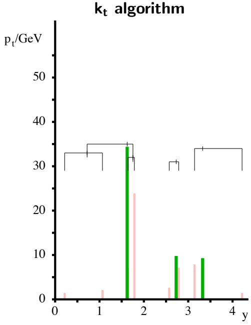
Which algorithm is most suited to jet substructure?

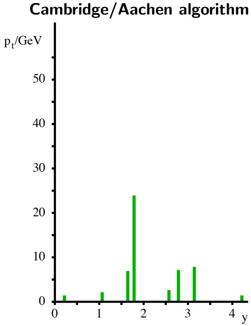
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.



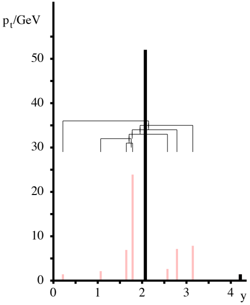


Which algorithm is most suited to jet substructure?

Animations by G. Salam

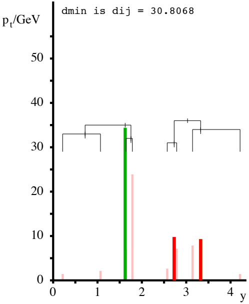
Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.

anti-k_t algorithm

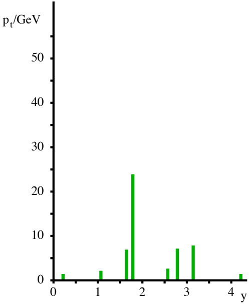


Information about prongs lost early in the history.

k_t algorithm



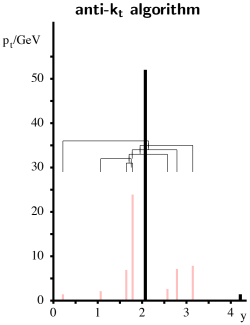
Cambridge/Aachen algorithm



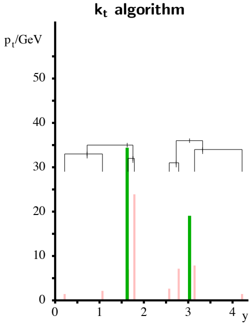
Which algorithm is most suited to jet substructure?

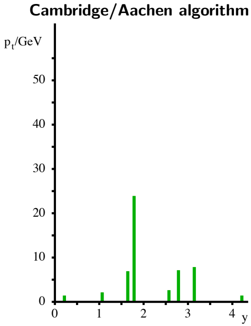
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.

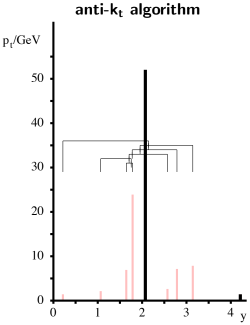




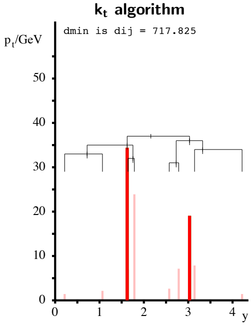
Which algorithm is most suited to jet substructure?

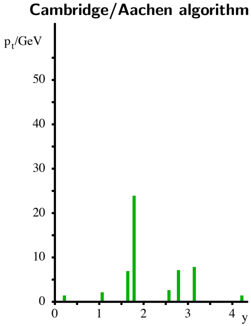
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.

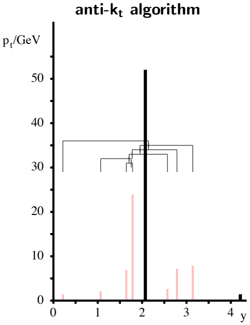




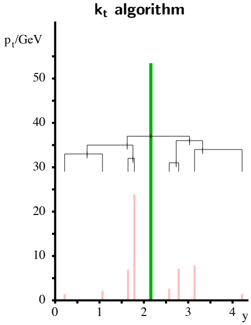
Which algorithm is most suited to jet substructure?

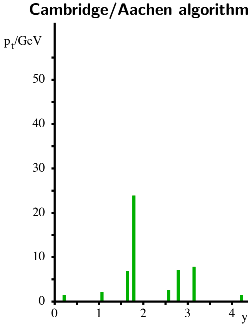
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs
lost early in the history.

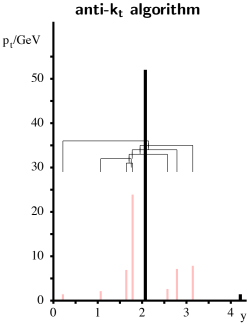


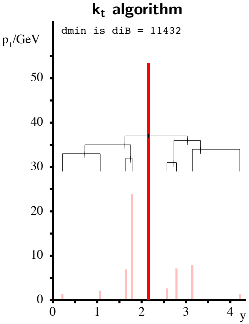


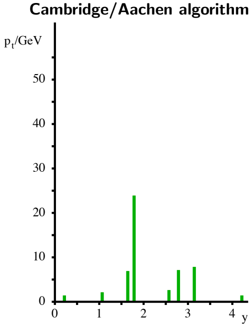
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.





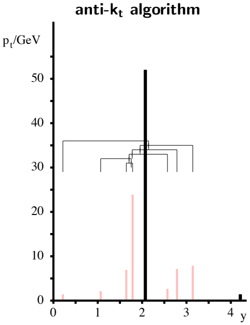


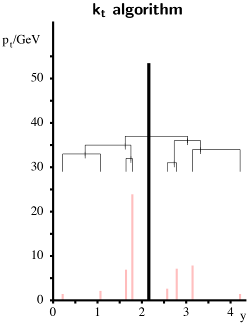
Information about prongs lost early in the history.

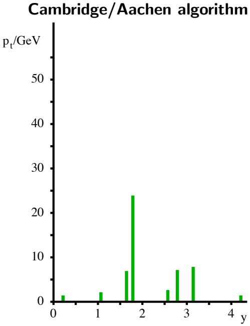
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.





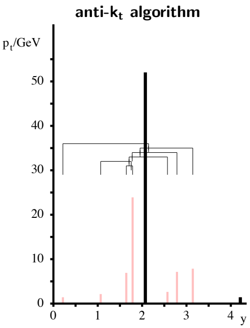


Information about prongs lost early in the history.

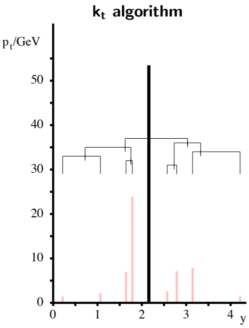
Which algorithm is most suited to jet substructure?

Animations by G. Salam

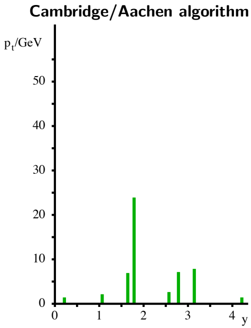
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs lost early in the history.



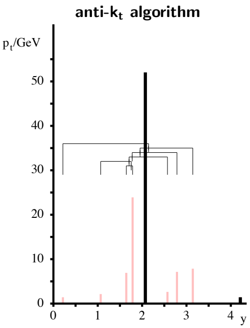
Prongs are recombined in the last step of clustering.



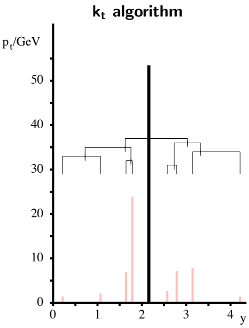
Which algorithm is most suited to jet substructure?

Animations by G. Salam

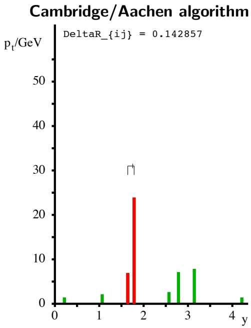
Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.



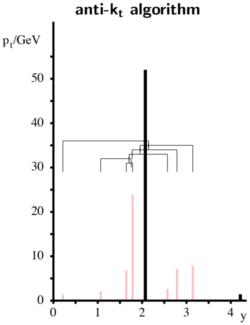
Prongs are recombined in the last step of clustering.



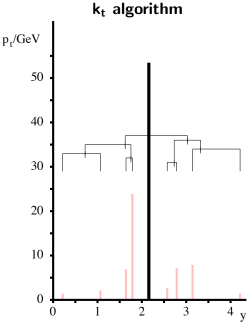
Which algorithm is most suited to jet substructure?

Animations by G. Salam

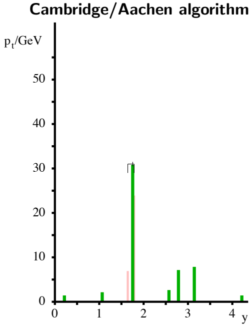
Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.



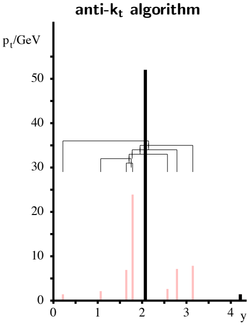
Prongs are recombined in the last step of clustering.



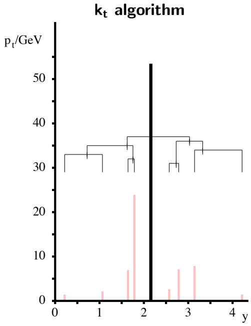
Which algorithm is most suited to jet substructure?

Animations by G. Salam

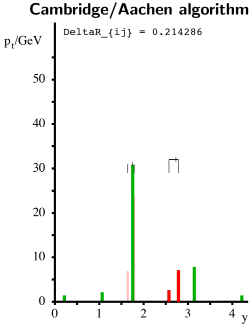
Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.



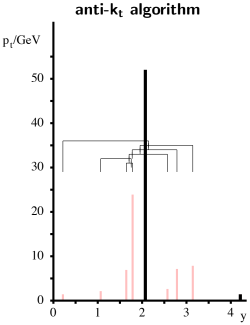
Prongs are recombined in the last step of clustering.



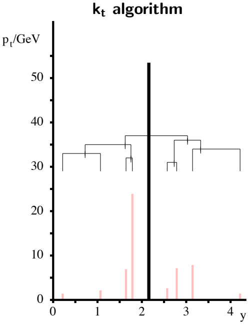
Which algorithm is most suited to jet substructure?

Animations by G. Salam

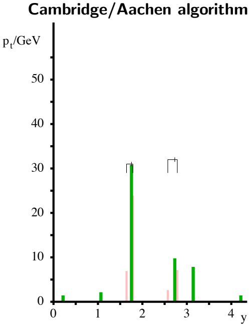
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs lost early in the history.



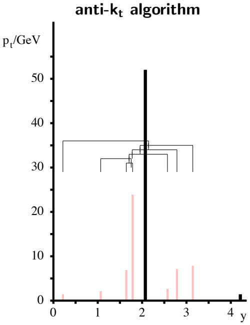
Prongs are recombined in the last step of clustering.



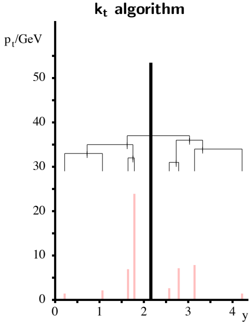
Which algorithm is most suited to jet substructure?

Animations by G. Salam

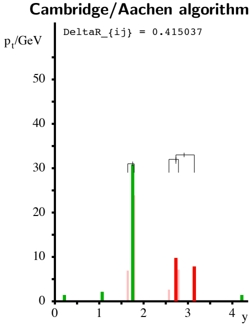
Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.



Prongs are recombined in the last step of clustering.

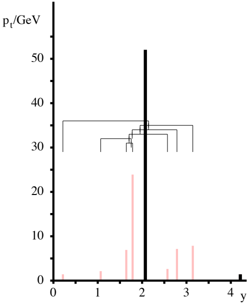


Which algorithm is most suited to jet substructure?

Animations by G. Salam

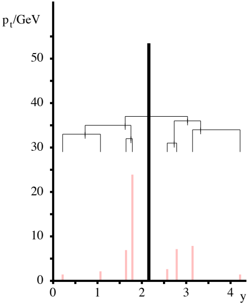
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.

anti-k_t algorithm



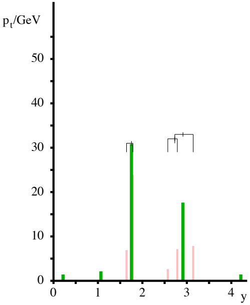
Information about prongs lost early in the history.

k_t algorithm



Prongs are recombined in the last step of clustering.

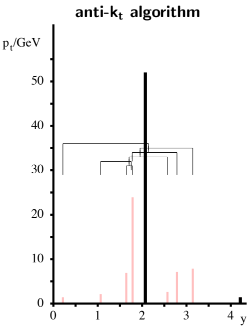
Cambridge/Aachen algorithm



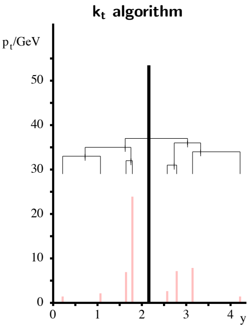
Which algorithm is most suited to jet substructure?

Animations by G. Salam

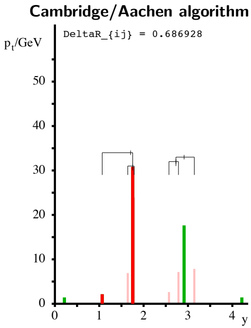
Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.



Prongs are recombined in the last step of clustering.

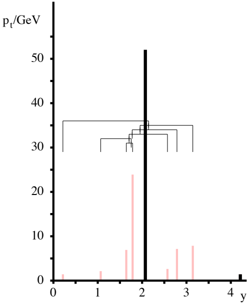


Which algorithm is most suited to jet substructure?

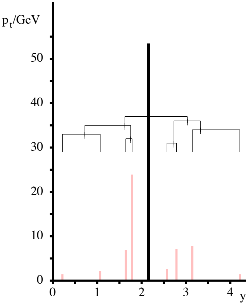
Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.

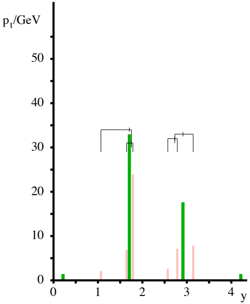
anti-k_t algorithm



k_t algorithm



Cambridge/Aachen algorithm



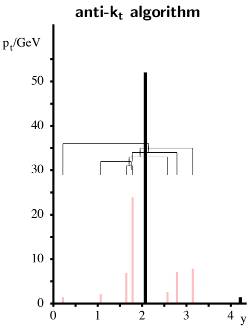
Information about prongs lost early in the history.

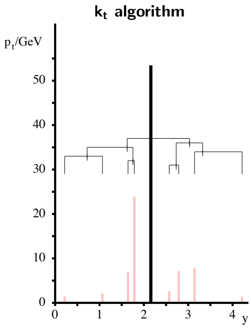
Prongs are recombined in the last step of clustering.

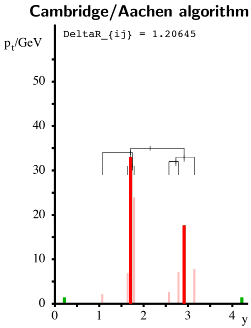
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.







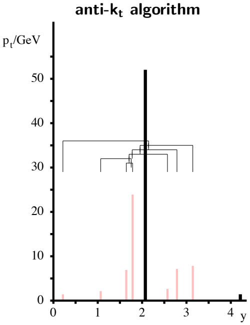
Information about prongs lost early in the history.

Prongs are recombined in the last step of clustering.

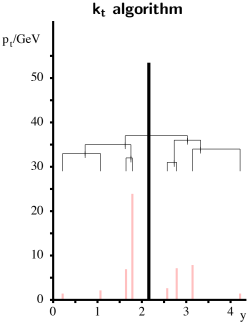
Which algorithm is most suited to jet substructure?

Animations by G. Salam

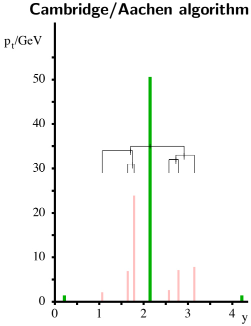
Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.



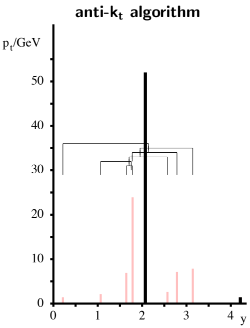
Prongs are recombined in the last step of clustering.



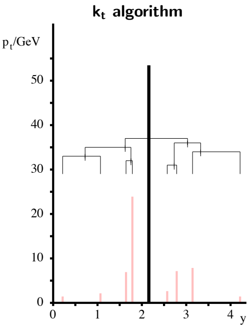
Which algorithm is most suited to jet substructure?

Animations by G. Salam

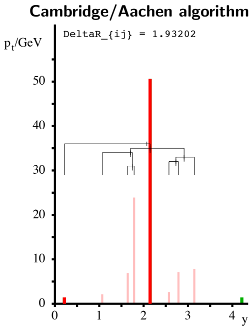
Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.



Information about prongs lost early in the history.



Prongs are recombined in the last step of clustering.

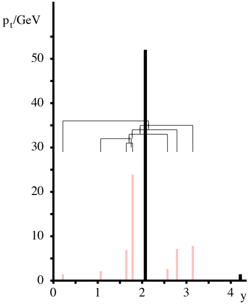


Which algorithm is most suited to jet substructure?

Animations by G. Salam

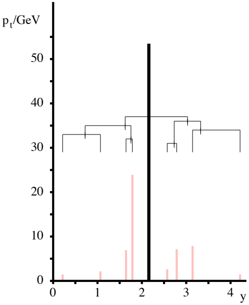
Let's investigate the clustering sequences of the gen-k_t family when applied to a 2-prong-like event.

anti-k_t algorithm



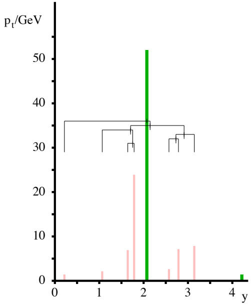
Information about prongs lost early in the history.

k_t algorithm



Prongs are recombined in the last step of clustering.

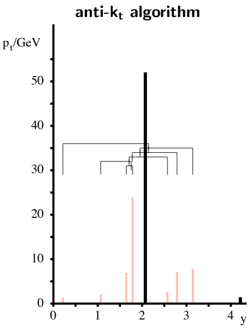
Cambridge/Aachen algorithm

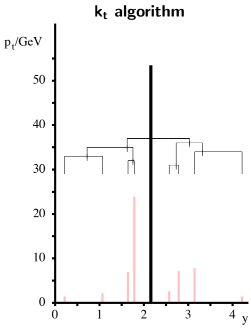


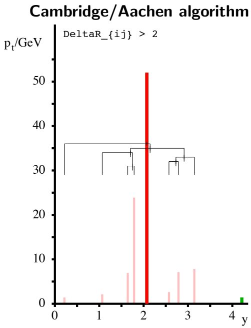
Which algorithm is most suited to jet substructure?

Animations by G. Salam

Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.







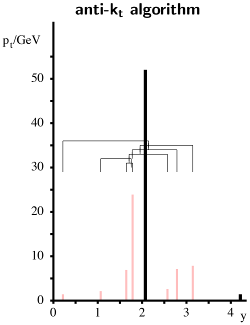
Information about prongs lost early in the history.

Prongs are recombined in the last step of clustering.

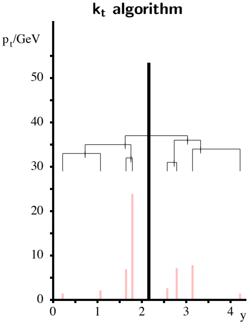
Which algorithm is most suited to jet substructure?

Animations by G. Salam

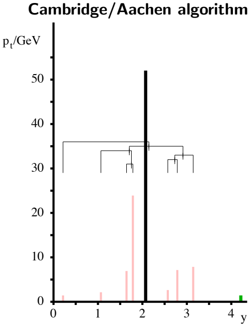
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.



Information about prongs lost early in the history.



Prongs are recombined in the last step of clustering.

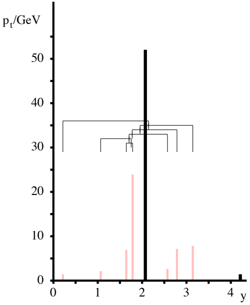


Which algorithm is most suited to jet substructure?

Animations by G. Salam

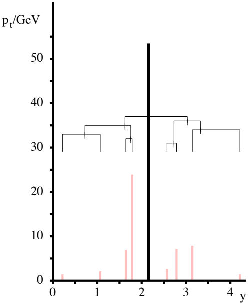
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.

anti-k_t algorithm



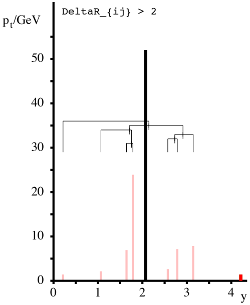
Information about prongs lost early in the history.

k_t algorithm



Prongs are recombined in the last step of clustering.

Cambridge/Aachen algorithm

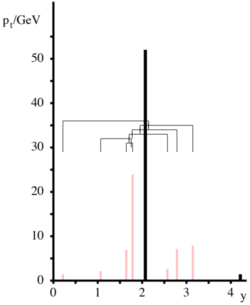


Which algorithm is most suited to jet substructure?

Animations by G. Salam

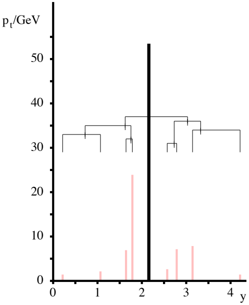
Let's investigate the clustering sequences of the gen-k_t family
when applied to a 2-prong-like event.

anti-k_t algorithm



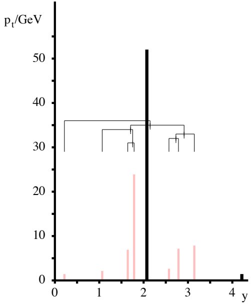
Information about prongs lost early in the history.

k_t algorithm



Prongs are recombined in the last step of clustering.

Cambridge/Aachen algorithm



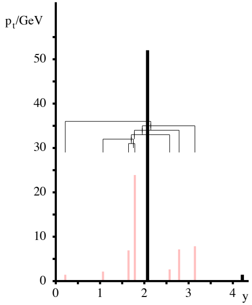
Angular ordering with some soft contamination.

Which algorithm is most suited to jet substructure?

Animations by G. Salam

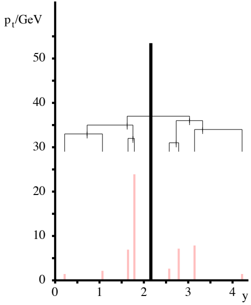
Anti-k_t bad (unphysical), k_t good (physical),
C/A very good together with some "cleaning" of soft junk

anti-k_t algorithm



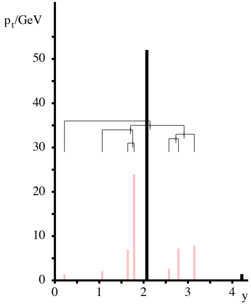
Information about prongs lost early in the history.

k_t algorithm



Prongs are recombined in the last step of clustering.

Cambridge/Aachen algorithm

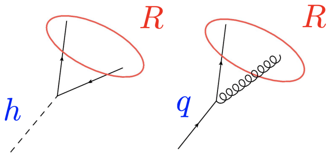


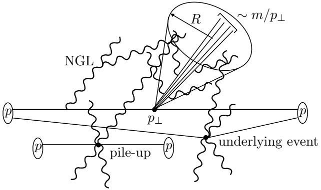
Angular ordering with some soft contamination.

Tagging & grooming

Once the jet has been reclustered, we can *tag* and/or *groom* it.

Tagging: find a particular structure inside the jet e.g. 2-prong decay of Higgs boson or 3-prong decay of top quark.
Key observation: energy sharing in signal is mostly simmetrical ($P_h(z) \sim 1$), whereas in background is mostly asymmetrical ($P_q(z) \sim 1/z$).

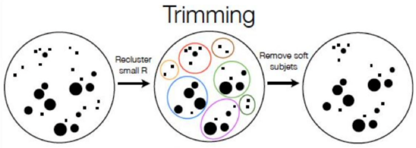




Grooming: remove background contamination in jet, while keeping the bulk of perturbative radiation.
Key observation: NGLs, UE, pile-up mostly appear as soft wide-angle radiation inside the jet.

Example of grooming: trimming and pruning

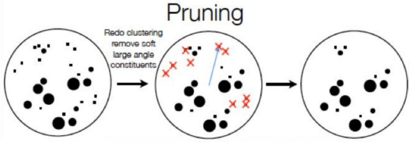
Trimming (top-down approach)



1. Recluster with C/A with a smaller radius $R_{\rm trim}$
2. Keep all subjects with

$$p_t > f_{\rm trim}\, p_{t,\rm jet}$$

Pruning (bottom-up approach)



1. Define (dynamical) pruning radius $R_{\rm prune} = 2\, f_{\rm prune}\, m_{jet}/p_{t,\rm jet}$
2. For every step $i + j \to k$ of C/A check:
   - Large angle? : $\Delta R_{ij} < R_{\rm prune}$
   - Soft? : $\min(p_{t,i}, p_{t,j}) \geq z_{\rm prune}\, p_{t,k}$
3. If neither criteria are met, eliminate softer subjet and keep the harder one

Example of tagging: N-subjettiness

Thaler, Van Tilburg (1011.2268)

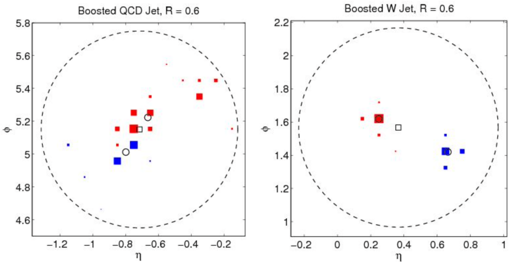
N-subjettiness τ_N quantifies the amount of radiation around N prongs. After having introduced a set of N axes $\{a_1, \ldots, a_N\}$ (e.g. the ones given by gen-k_t algorithm or the ones that minimize τ_N), τ_N is defined as:

$$\tau_N^{(\beta)} = \sum_{i \in jet} p_{ti} \, \min(\Delta R^\beta_{ia_1}, \ldots, \Delta R^\beta_{ia_N}) \,, \quad \text{with e.g. } \beta = 1, 2$$

The N-subjettiness ratio:

$$\tau_{N,N-1}^{(\beta)} = \frac{\tau_N^{(\beta)}}{\tau_{N-1}^{(\beta)}}$$

is a good discriminating variable for N-prong signal jets against QCD background e.g. a cut on $\tau_{21} < \tau_{\rm cut}$ useful to discriminate $W/Z/H$ vs. QCD and $\tau_{32} < \tau_{\rm cut}$ useful for top vs. QCD.



Soft Drop: tagger and groomer

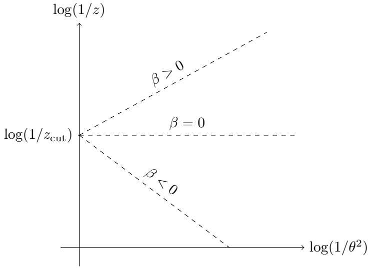
Larkoski, Marzani, Soyez, Thaler (1402.2657)

First, recluster jet constituents with C/A.

1. Undo the last stage of C/A, $(i+j) \to i+j$
2. Check Soft Drop condition:

$$\frac{\min(p_{t,i}, p_{t,j})}{p_{t,i}+p_{t,j}} > z_{\rm cut} \left(\frac{\Delta R_{ij}}{R}\right)^{\beta}$$

   In soft limit, $z > z_{\rm cut}\theta^{\beta}$. If subjets pass it, then declare $(i+j)$ as the soft-drop jet.
3. Otherwise, iterate on the subjet with the largest p_t.



- $\beta > 0$: remove all soft radiation and some soft-collinear
- $\beta = 0$ (mMDT): remove all soft-collinear (just symmetry condition)
- $\beta < 0$: remove also hard-collinear radiation

Soft Drop: tagger and groomer

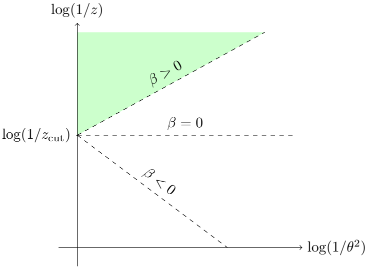
Larkoski, Marzani, Soyez, Thaler (1402.2657)

First, recluster jet constituents with C/A.

1. Undo the last stage of C/A,
   $(i+j) \to i+j$
2. Check Soft Drop condition:

$$\frac{\min(p_{t,i}, p_{t,j})}{p_{t,i}+p_{t,j}} > z_{\text{cut}} \left(\frac{\Delta R_{ij}}{R}\right)^{\beta}$$

   In soft limit, $z > z_{\text{cut}}\theta^{\beta}$.
   If subjets pass it, then declare $(i+j)$ as the soft-drop jet.
3. Otherwise, iterate on the subjet with the largest p_t.



- $\beta > 0$: remove all soft radiation and some soft-collinear
- $\beta = 0$ (mMDT): remove all soft-collinear (just symmetry condition)
- $\beta < 0$: remove also hard-collinear radiation

Soft Drop: tagger and groomer

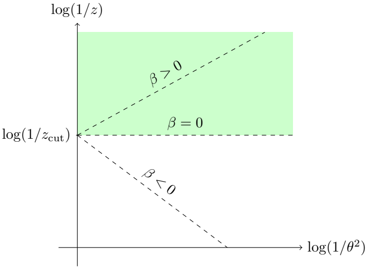
Larkoski, Marzani, Soyez, Thaler (1402.2657)

First, recluster jet constituents with C/A.

1. Undo the last stage of C/A, $(i+j) \to i+j$
2. Check Soft Drop condition:

$$\frac{\min(p_{t,i}, p_{t,j})}{p_{t,i}+p_{t,j}} > z_{\rm cut}\left(\frac{\Delta R_{ij}}{R}\right)^{\beta}$$

   In soft limit, $z > z_{\rm cut}\theta^{\beta}$.
   If subjets pass it, then declare $(i+j)$ as the soft-drop jet.
3. Otherwise, iterate on the subjet with the largest p_t.



- $\beta > 0$: remove all soft radiation and some soft-collinear
- $\beta = 0$ (mMDT): remove all soft-collinear (just symmetry condition)
- $\beta < 0$: remove also hard-collinear radiation

Soft Drop: tagger and groomer

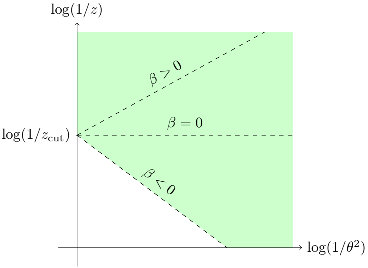
Larkoski, Marzani, Soyez, Thaler (1402.2657)

First, recluster jet constituents with C/A.

1. Undo the last stage of C/A, $(i+j) \to i+j$
2. Check Soft Drop condition:

$$\frac{\min(p_{t,i}, p_{t,j})}{p_{t,i}+p_{t,j}} > z_{\text{cut}} \left(\frac{\Delta R_{ij}}{R}\right)^{\beta}$$

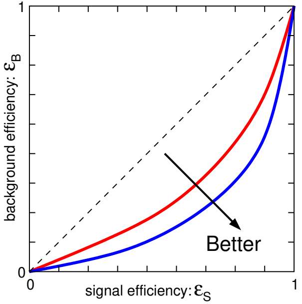
   In soft limit, $z > z_{\text{cut}}\theta^{\beta}$.
   If subjets pass it, then declare $(i+j)$ as the soft-drop jet.
3. Otherwise, iterate on the subjet with the largest p_t.



- $\beta > 0$: remove all soft radiation and some soft-collinear
- $\beta = 0$ (mMDT): remove all soft-collinear (just symmetry condition)
- $\beta < 0$: remove also hard-collinear radiation

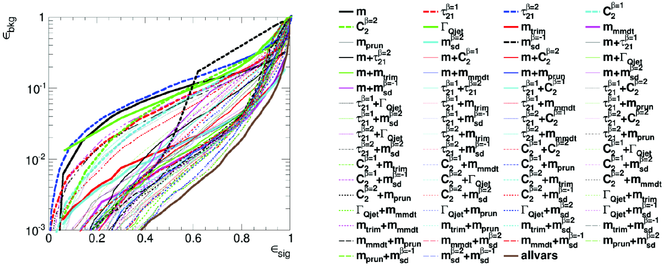
Comparison of different taggers

Evaluate performances with ROC curves:



Dashed line is a random ragger e.g. flipping a coin.

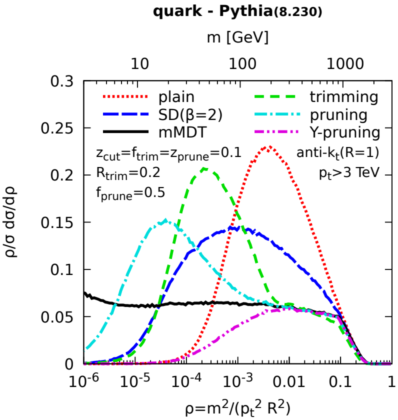
Example of ROCs for combinations of taggers and groomers for W tagging (slide from BOOST 2013):

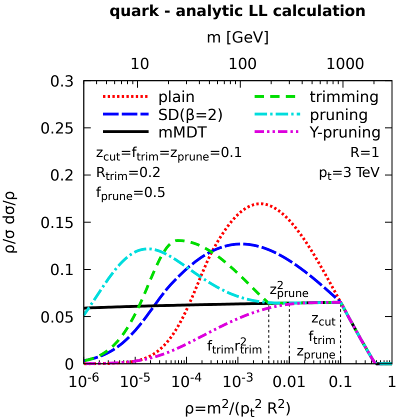


Can we understand these curves from first principles?

Revolution driven by analytical studies.

Jet mass after grooming: Monte Carlo vs. analytics





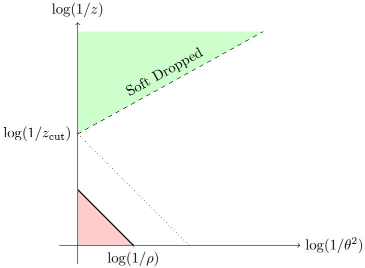
Transition points and shapes can be traced back to analytical calculations.

Example: jet mass after Soft Drop with $\beta \geq 0$

Let's revise the jet mass calculation:

$$\Sigma_{\rm SD}(\rho) = \exp\left[-\Theta(\rho > z_{\rm cut}) V(\rho) - \Theta(\rho < z_{\rm cut}) V_{\rm SD}\right]$$

For $\rho > z_{\rm cut}$, Soft Drop has no effect.



The no-emission probability is the same of the plain jet mass, $V(\rho)$.

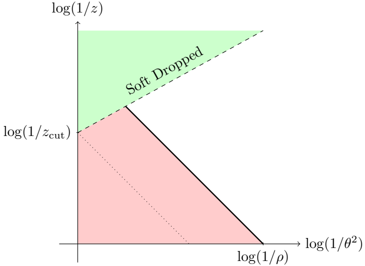
$$V(\rho) = \frac{\alpha_s C_F}{\pi} \int_0^1 \frac{{\rm d}z_i}{z_i} \int_0^1 \frac{{\rm d}\theta_i^2}{\theta_i^2}\, \Theta(z_i \theta_i^2 > \rho) = \frac{\alpha_s C_F}{\pi} \frac{1}{2} \log^2 \frac{1}{\rho}$$

Example: jet mass after Soft Drop with $\beta \geq 0$

Let's revise the jet mass calculation:

$$\Sigma_{\text{SD}}(\rho) = \exp\left[-\Theta(\rho > z_{\text{cut}})V(\rho) - \Theta(\rho < z_{\text{cut}})V_{\text{SD}}\right]$$

For $\rho < z_{\text{cut}}$, Soft Drop removes soft radiation (and with $\beta = 0$ also soft-collinear radiation).



The shared red area is:

$$V_{\text{SD}}(\rho) = \frac{\alpha_s C_F}{\pi}\int_0^1 \frac{dz_i}{z_i}\int_0^1 \frac{d\theta_i^2}{\theta_i^2}\,\Theta(z_i\theta_i^2 > \rho)\,\Theta(z > z_{\text{cut}}\theta^\beta)$$
$$= \frac{\alpha_s C_F}{\pi}\left[\frac{\beta}{2+\beta}\frac{1}{2}\log^2\frac{1}{\rho} + \frac{2}{2+\beta}\left(\frac{1}{2}\log^2\frac{1}{z_{\text{cut}}} + \log\frac{z_{\text{cut}}}{\rho}\log\frac{1}{z_{\text{cut}}}\right)\right]$$

With $\beta = 0$ double logarithms of $\log(1/\rho)$ are absent.

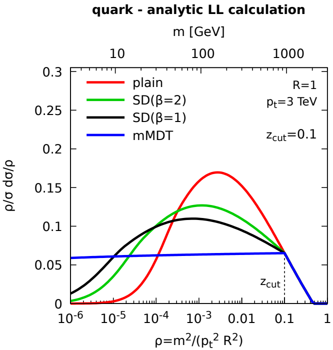
Example: jet mass after Soft Drop with $\beta \geq 0$

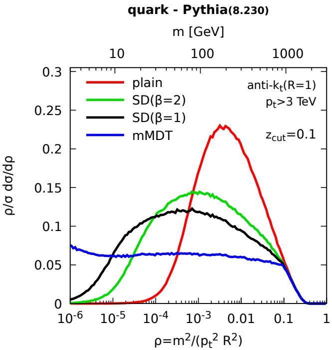
The jet mass spectrum is given by

$$\frac{\rho}{\sigma}\frac{d\sigma}{d\rho} = \frac{d\Sigma(\rho)}{d\log\rho}$$

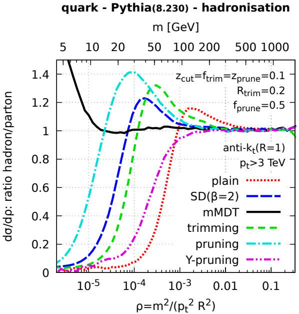
reproducing behaviour of Monte Carlo simulations.

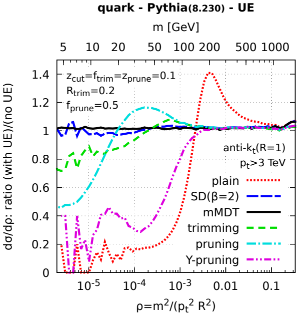
mMDT basically flat (good background).

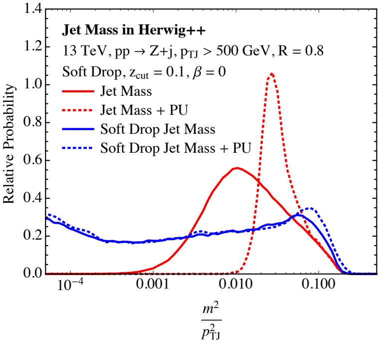




Example: jet mass after Soft Drop with $\beta \geq 0$

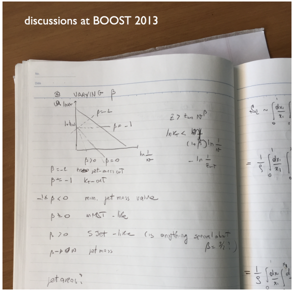


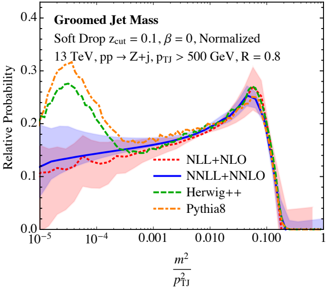


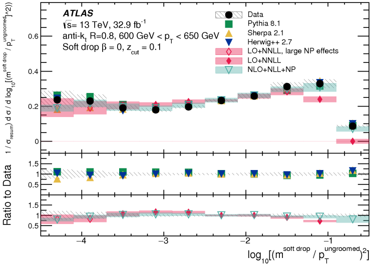


Reduced sensitivity to non-perturbative physics compared to other tools.

Soft Drop: a BOOST success story





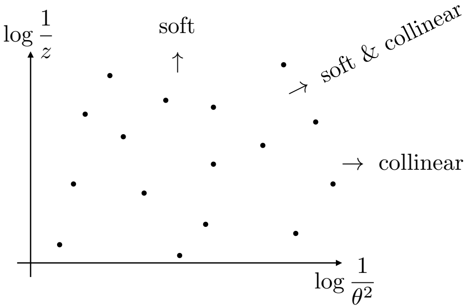


Born after discussions at BOOST 2013.
In the recent years, very precise calculations [e.g. Frye, Larkoski, Schwartz,Yan (2016)] compared to precise measurement [e.g. ATLAS Phys. Rev. Lett. 121 (2018) 092001]

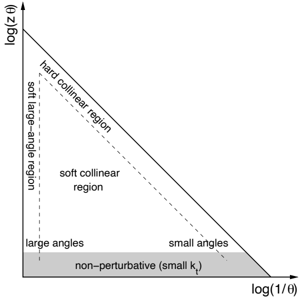
Outline

The Lund plane

Powerful way of depicting the pattern of QCD radiation.
We have already used the (θ, z) version, but $(\theta, k_t = z\theta)$ is also possible.



Soft and collinear in ortogonal directions



The non-pertubative region in the bottom

The Lund plane may constitute an observable itself!

The Lund Jet plane

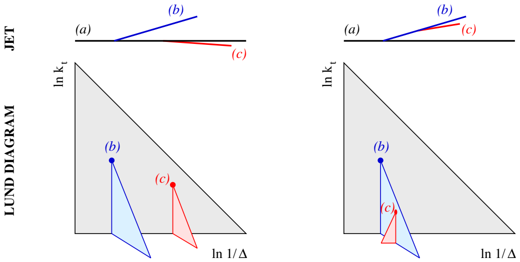
Dreyer, Salam, Soyez (1807.04758)

First decluster the jet with C/A. For each step of the declustering, involving pseudo-jets a and b with $p_{t,a} > p_{t,b}$, we record the variables:

$$\Delta \equiv \Delta_{ab} = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2}, \quad k_t = p_{t,b}\Delta_{ab}, \quad z = \frac{p_{t,b}}{p_{t,a} + p_{t,b}}$$

and we plot them in the Lund $(\ln 1/\Delta, \ln k_t)$ or $(\ln 1/\Delta, \ln 1/z)$ plane.

We have a *primary* Lund plane (if related to an emission off the hardest branch), possibly branching into a *secondary*, *tertiary*, etc. Lund plane.



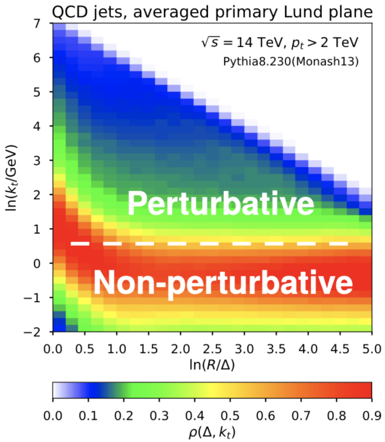
Lund Jet plane for precision QCD

Lund plane density, defined as:

$$\rho(\Delta, k_t) = \frac{1}{N_{\text{jet}}} \frac{\mathrm{d}n_{\text{emissions}}}{\mathrm{d}\ln k_t \,\mathrm{d}\ln 1/\Delta}$$

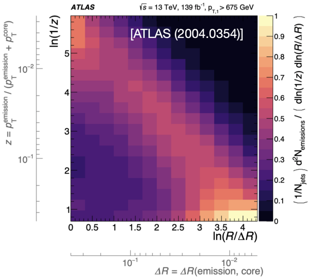
with N_{jet} the total number of jets. At leading order we have:

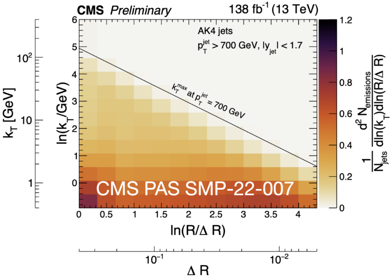
$$\rho_i \simeq \frac{2\alpha_s(k_t)\, C_i}{\pi}; \quad C_q = C_F,\, C_g = C_A$$

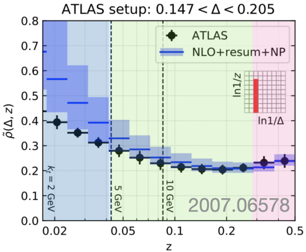


Lund Jet plane for precision QCD

Lund Jet plane density measured at the LHC and precisely calculated!

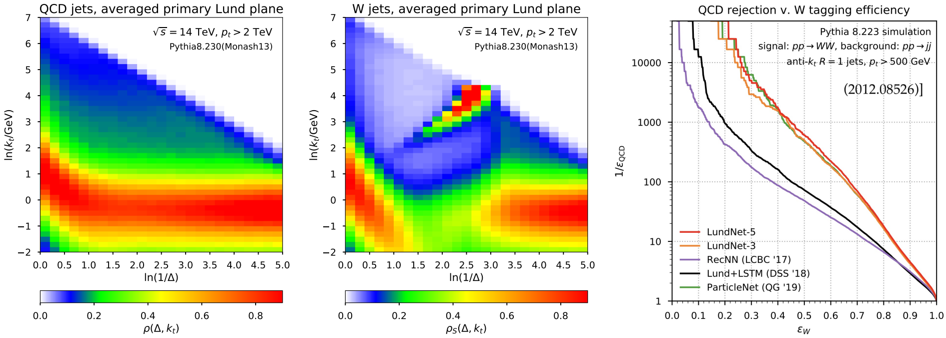






Non-pert. Resummation NLO

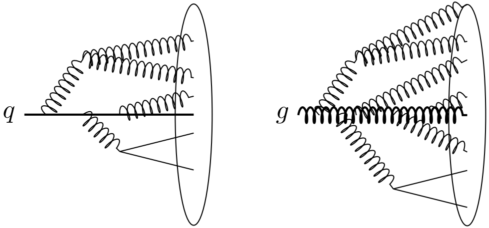
Lund Jet plane for signal/background discrimination



Lund plane images look differently between background and signal jets!
Lund images and trees can be adopted as theory-friendly input to machine learning models, even reaching state-of-the-art performances.

Quark- vs. gluon-jet discrimination

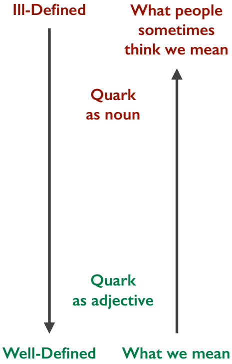
Disentangle jets that can be thought of as originating from the fragmentation of a high-energy quark from the ones originating from a gluon.



Important either as **observable** (precision α_s studies, PDF extraction) or as **tool** (isolation of specific production channels, search of new physics).

What is a quark jet?

The proper definition is not free from ambiguities



A quark parton

A Born-level quark parton

The initiating quark parton in a final state shower

An eikonal line with baryon number 1/3 and carrying triplet color charge

A quark operator appearing in a hard matrix element in the context of a factorization theorem

A parton-level jet object that has been quark-tagged using a soft-safe flavored jet algorithm (automatically collinear safe if you sum constituent flavors)

A phase space region (as defined by an unambiguous hadronic fiducial cross section measurement) that yields an enriched sample of quarks (as interpreted by some suitable, though fundamentally ambiguous, criterion)

(from Les Houches workshop 2015)

The simplest q/g discriminant

The now familiar jet mass offers a possible discriminant:

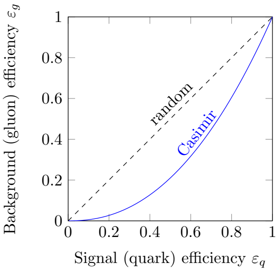
$$\Sigma_q(\rho) = \exp\left[-\frac{\alpha_s C_F}{\pi}\frac{1}{2}\log^2\frac{1}{\rho}\right], \quad \Sigma_g(\rho) = \exp\left[-\frac{\alpha_s C_A}{\pi}\frac{1}{2}\log^2\frac{1}{\rho}\right]$$

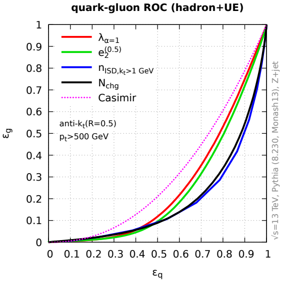
The cumulative distributions for q/g only differ by the colour (Casimir) factor.

The ROC curve is given by:

$$\varepsilon_g = \Sigma_g(\Sigma_q^{-1}(\varepsilon_q)) = (\varepsilon_q)^{C_A/C_F}$$

This feature is called Casimir scaling and provides a benchmark for expectation when using more sophisticated taggers (e.g. angularities)





Outline

QCD and jets

Jet substructure

Selected topics

Conclusions

What I haven't talked about

I hope I provided you with enough background to understand the talks of this week, but notable omissions are:

- Energy-Energy Correlators (EEC)
- Track Functions
- Physics-aware tools to interpret what Machine Learning gives us

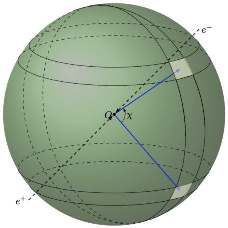
What I haven't talked about

I hope I provided you with enough background to understand the talks of this week, but notable omissions are:

- Energy-Energy Correlators (EEC)

  Long tradition, but renewed interest in the community

$$\frac{\mathrm{d}\sigma}{\mathrm{d}\theta} = \sum_{i,j} \int d\sigma \frac{E_i E_j}{Q^2} \delta\left(\theta - \theta_{ij}\right) \sim \langle \Psi \left| \mathcal{E}\left(\hat{n}_1\right) \mathcal{E}\left(\hat{n}_2\right) \right| \Psi \rangle$$



correlation functions of energy flow operators
(see Kyle Lee's theory overview talk BOOST 2023)

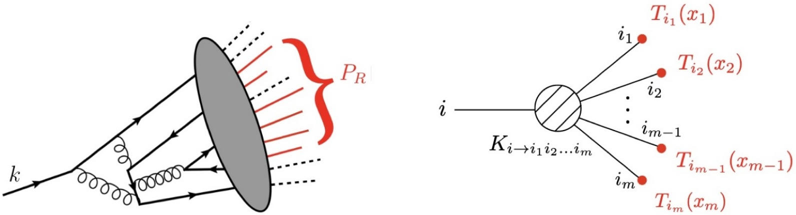
- Track Functions
- Physics-aware tools to interpret what Machine Learning gives us

What I haven't talked about

I hope I provided you with enough background to understand the talks of this week, but notable omissions are:

- Energy-Energy Correlators (EEC)
- Track Functions

Describe fragmentation of a parton into a subset of final state hadrons
They naturally encode multi-hadron fragmentation



(Chang, Procura, Thaler, Waalewijn (2013))

- Physics-aware tools to interpret what Machine Learning gives us

What I haven't talked about

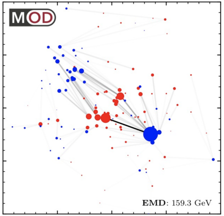
I hope I provided you with enough background to understand the talks of this week, but notable omissions are:

- Energy-Energy Correlators (EEC)
- Track Functions
- Physics-aware tools to interpret what Machine Learning gives us

Energy Flow Polynomials (EFP), Energy Flow Networks (EFN), Energy Mover's Distance (EMD), etc. [Komiske, Metodiev, Thaler, et al.]

$$\text{EFN} = F\left(\sum_{i=1}^{M} z_i \Phi(\hat{p}_i)\right)$$

$$\text{EFP}_G = \sum_{i_1=1}^{M} \cdots \sum_{i_N=1}^{M} z_{i_1} \cdots z_{i_N} \prod_{(k,\ell)\in G} \theta_{i_k i_\ell}$$



What I haven't talked about

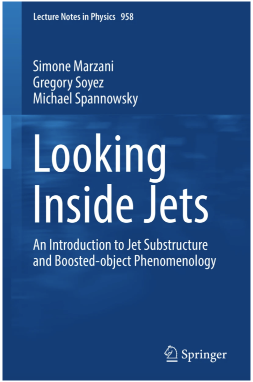
I hope I provided you with enough background to understand the talks of this week, but notable omissions are:

- Energy-Energy Correlators (EEC)
- Track Functions
- Physics-aware tools to interpret what Machine Learning gives us

Moreover, I haven't mentioned:

- Monte Carlo developments (more accurate parton showers, ...)
- Calculations of jet substructure observables in SCET
- Heavy quarks
- Jet substructure in heavy ion environment
- ...

If you want to know more...



arXiv:1901.10342 (188 pages)

QCD Masterclass Lectures on Jet Physics and Machine Learning

Andrew J. Larkoski

Email: larkoa@gmail.com

July 9, 2024

Abstract

These lectures were presented at the 2024 QCD Masterclass in Saint-Jacut-de-la-Mer, France. They introduce and review fundamental theorems and principles of machine learning within the context of collider particle physics, focused on application to jet identification and discrimination. Numerous examples of binary discrimination in jet physics are studied in detail, including $H \to b\bar{b}$ identification in fixed-order perturbation theory, generic one- versus two-prong discrimination with parametric power counting techniques, and up versus down quark jet classification by assuming the central limit theorem, isospin conservation, and a convergent moment expansion of the single particle energy distribution. Quark versus gluon jet discrimination is considered in multiple contexts, from using additive, infrared and collinear safe observables, to using hadronic multiplicity, and to including measurements of the jet charge. While many of the results presented here are well known, some novel results are presented, the most prominent being a parametrized expression for the likelihood ratio of quark versus gluon discrimination for jets on which hadronic multiplicity and jet charge are simultaneously measured. End-of-lecture exercises are also provided.

arXiv:2407.04897 (130 pages)

Enjoy BOOST 2024!

Boccadasse neighbourhood, Genoese for "donkey's mouth" ...

Enjoy BOOST 2024!

... which after the appearance of the Lund plane fish has been called *Boostadasse* neighbourhood, Genoese for "donkey's boost"