Summer School on Symmetries,
Fundamental Interactions and Cosmology
September 2018, Fraueninsel/Chiemsee

Jets

Matteo Cacciari
LPTHE Paris and Université Paris Diderot

Lecture 1 - Jet algorithms

Lecture 2 - Jet substructure

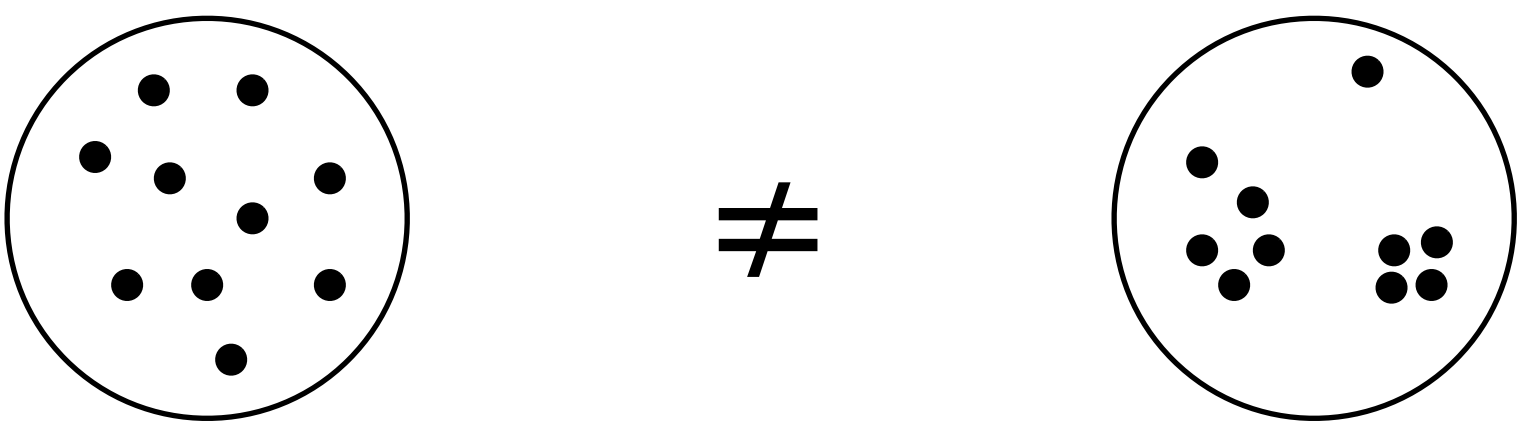
[Includes material from
Gavin Salam and Grégory Soyez]

Jet substructure

At the end of a jet finding (i.e. clustering) procedure, a jet is a **collection of constituents** to which we assign a 4-momentum

(related to the sum of the 4-momenta of the constituents)

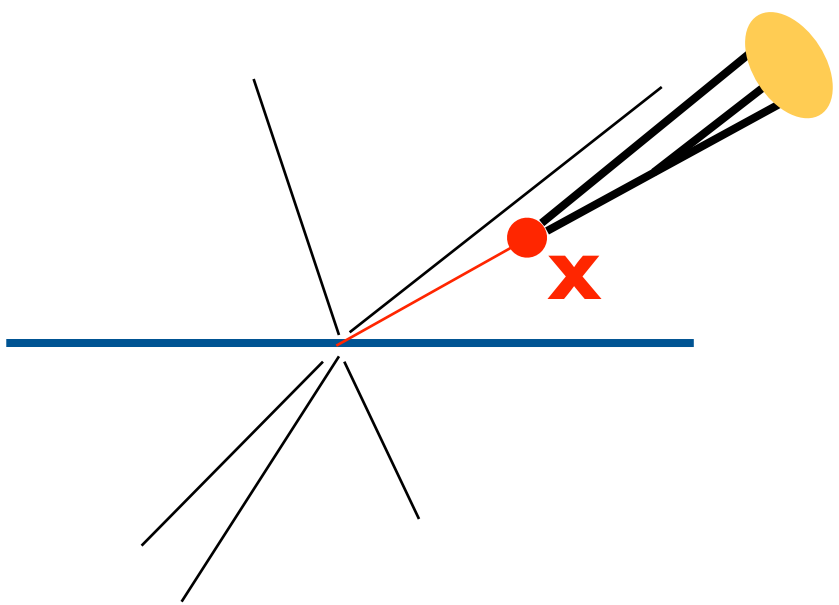
What is the **arrangement** of the constituents inside the jet?



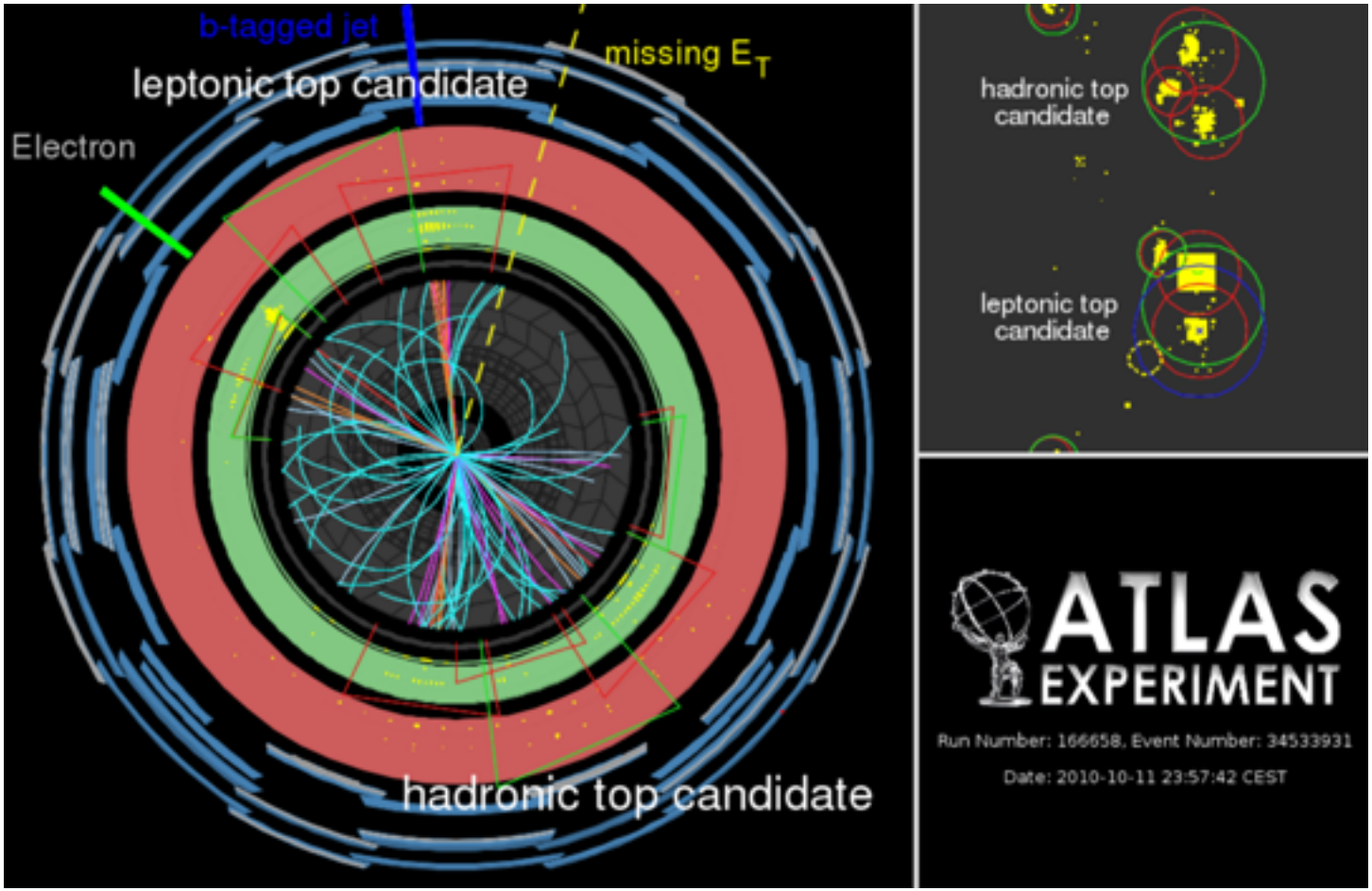
Jet substructure

First studied by Mike Seymour in the early '90s to distinguish W jets from QCD jets

Topic revived about 10 years ago in order to study boosted objects

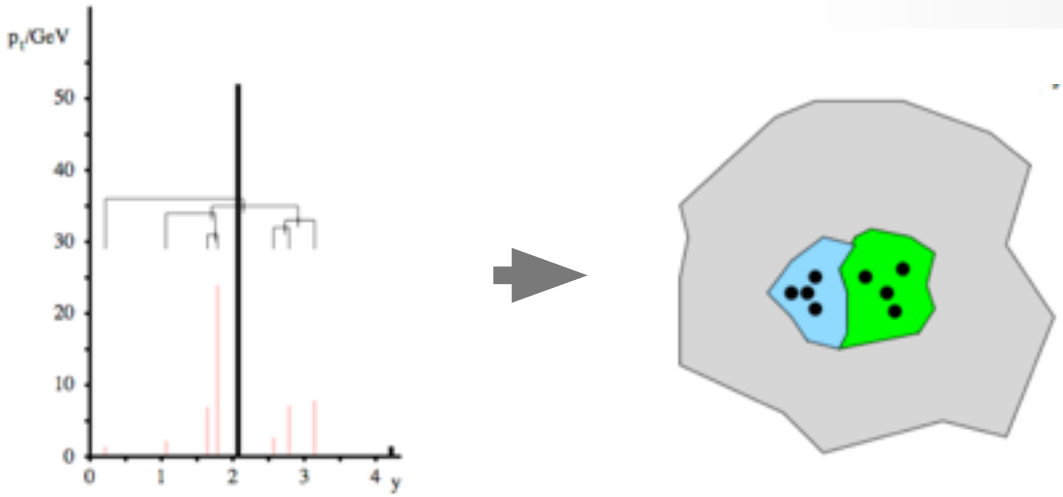


Jet substructure



The past ten years have seen en explosion in jet substructure studies, i.e. **how radiation is arranged within jets**, and what it can tell us

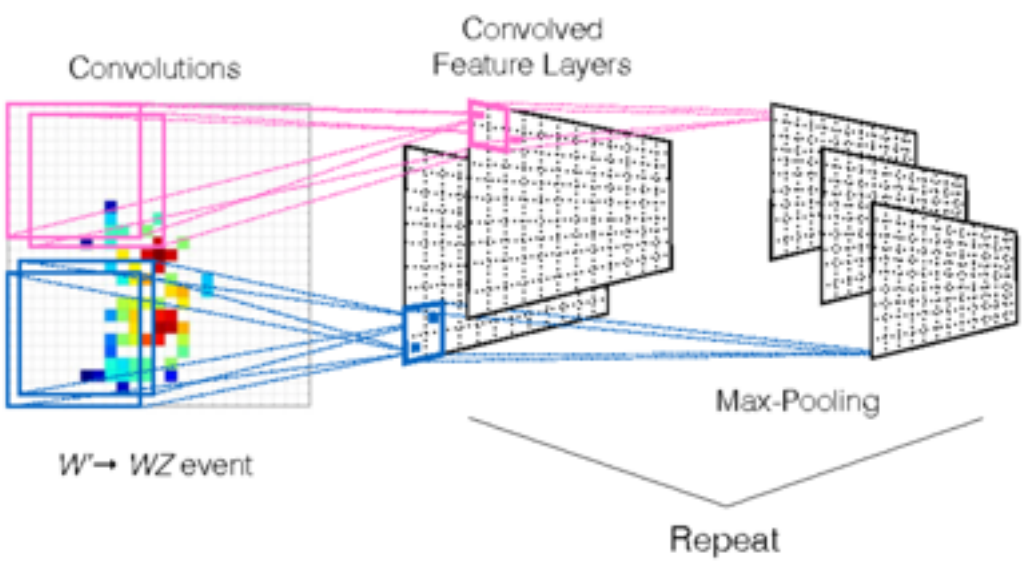
Jet substructure



Jet declustering

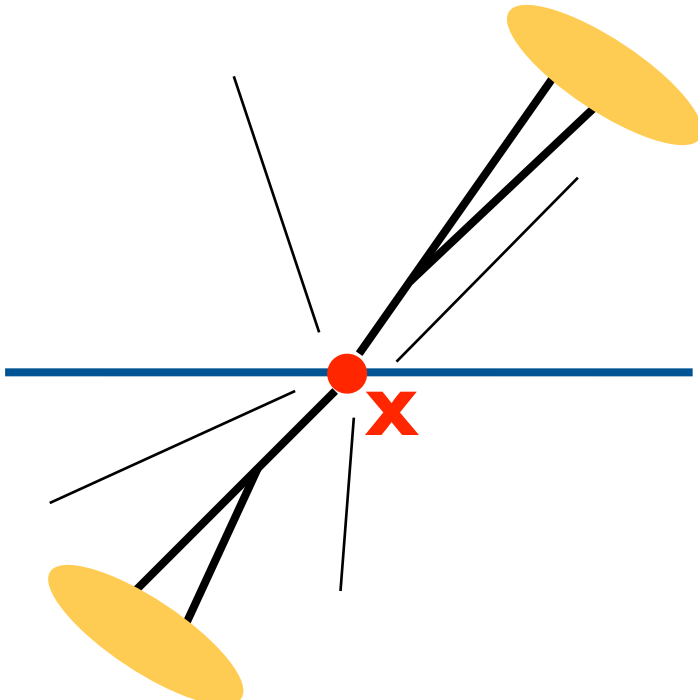
Jet shapes

(calculate a function from radiation distribution

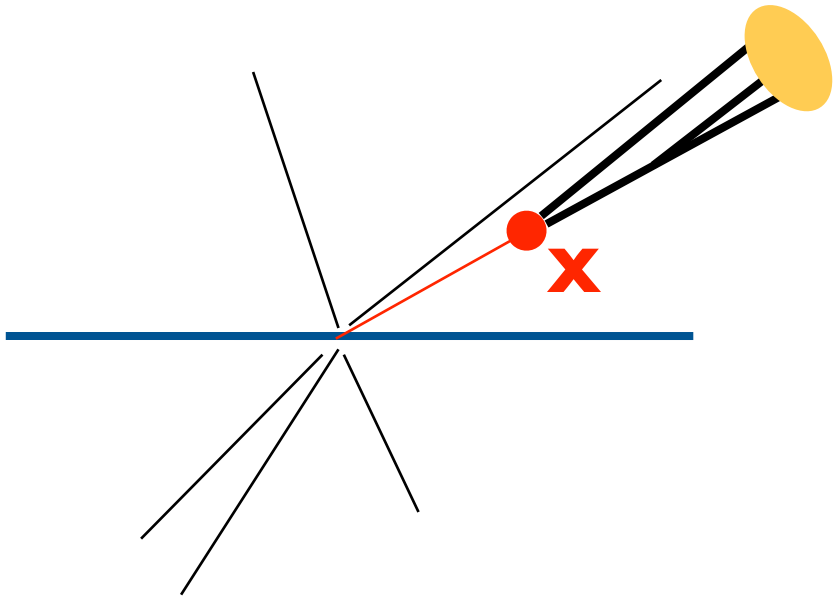


Machine learning

Why boosted objects



Heavy particle X at **rest**



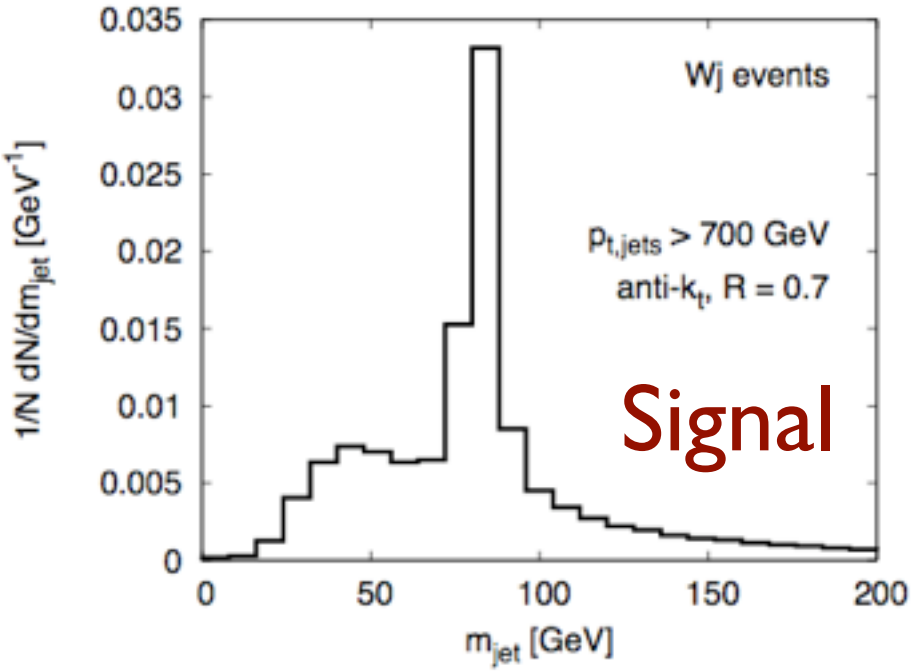
Boosted heavy particle X

Easy to resolve jets and calculate invariant mass, but signal very likely swamped by background (eg H→bb v. tt →WbWb)

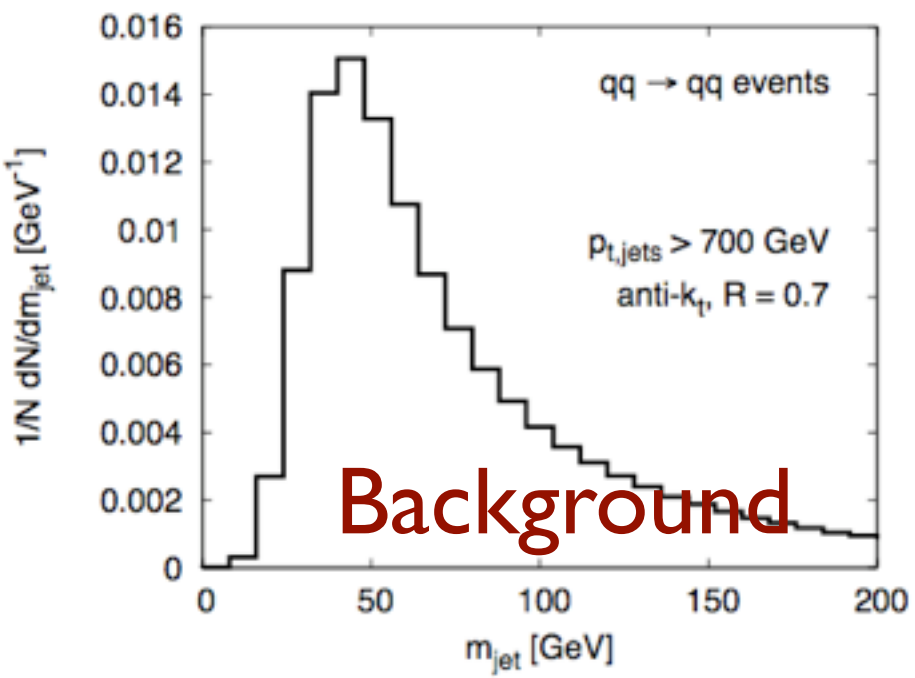
Cross section very much reduced, but acceptance better and some backgrounds smaller/ reducible

Mass of a single jet

G. Salam



A heavy object decaying into a single jet naturally gives it a mass...

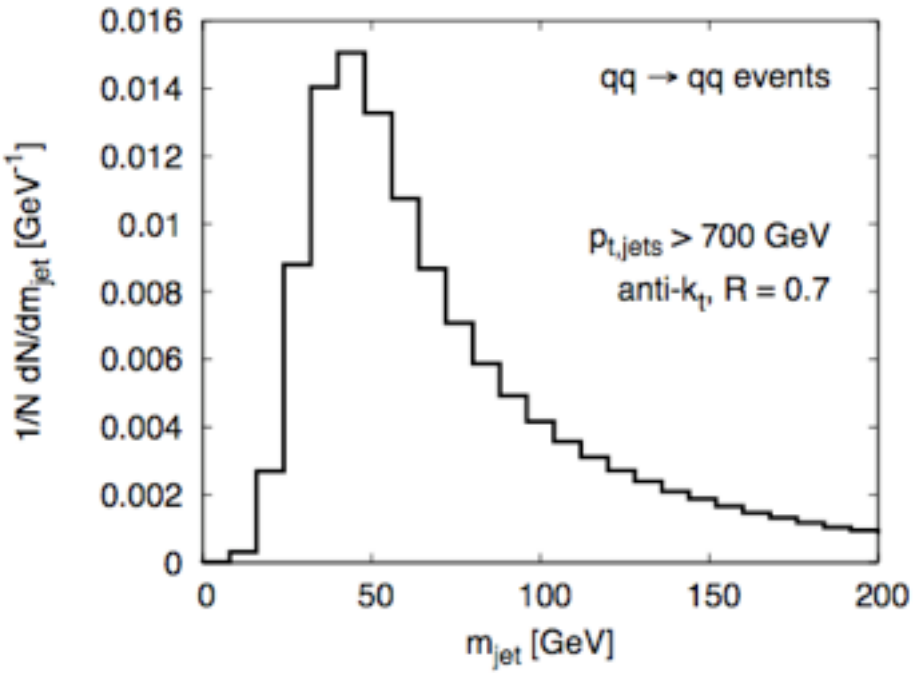


... but pure QCD jets can be massive too:

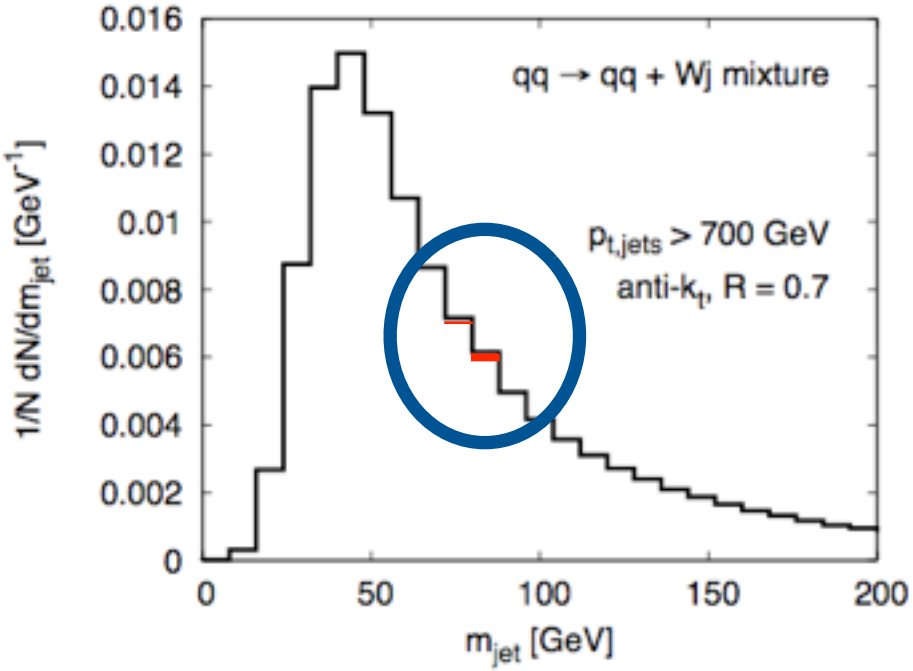
$$\frac{dN}{d\ln m} \sim \alpha_s \ln \frac{p_t R}{m} \times \text{Sudakov}$$

Mass of a single jet

Summing 'signal' and 'background' (with appropriate cross sections) shows how much the background dominates



Background only

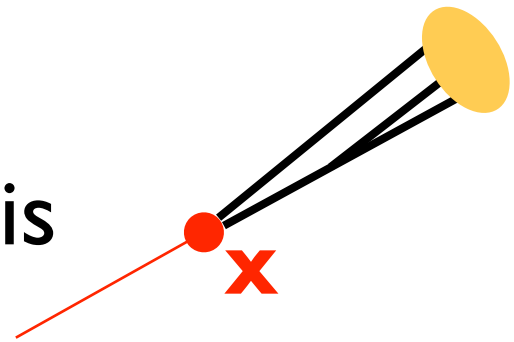


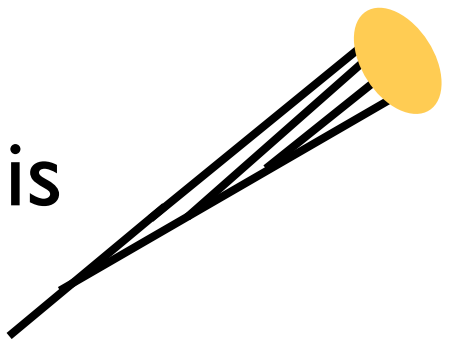
Signal + background

Practically identical

This means that one can't rely on the invariant mass only.
An appropriate strategy must be found to reduce the background and enhance the signal

Tagging

How to tell this 

from this  ?

Decay of a heavy (boosted) object

Light parton fragmentation

Tagging and Grooming

- The substructure of a jet can be exploited to
  - ***tag*** *a particular structure inside the jet, i.e. a massive particle*
    - *First examples: Higgs (2-prong decay), top (3-prong decay)*
  - *remove background contamination from the jet or its components, while keeping the bulk of the perturbative radiation (often generically denoted as* ***grooming****)*
    - *First examples: filtering, trimming, pruning*

Nomenclature

- **Groomer**
  - procedure that always returns an output jet
    (i.e. it only subtracts uncorrelated 'UE/pileup' radiation from it. This is used to "clean" the jets from radiation largely unrelated to the fragmentation of the particle of interest)

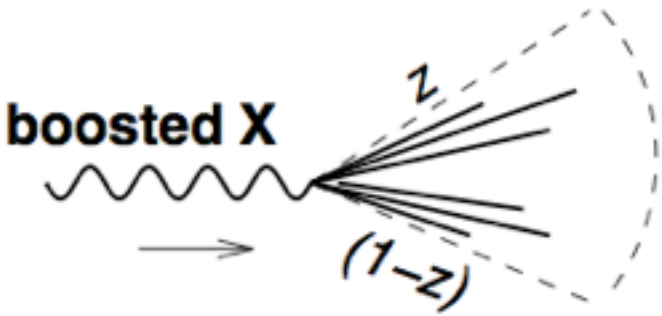
- **Tagger**
  - procedure that might not return an output jet
    (i.e. it either tags a heavy particle originating the jet or returns zero. This is used to identify a specific particle originating the jet.)

In practice, this classification is not always followed.
In some cases it also denoted a 'tagger' a procedure that rejects background jets more often than signal jets

Why substructure

Scales: m ~ 100 GeV, p_t ~ 500 GeV

(e.g. electroweak particle from decay of ~ 1TeV BSM particle)



$$R \gtrsim \frac{m}{p_t} \frac{1}{\sqrt{z(1-z)}}$$

- need small R (< $2m/p_t$ ~ 0.4) to resolve two prongs
- need large R (>~ $3m/p_t$ ~ 0.6) to cluster into a single jet

Possible strategies

- Use large R, get a single jet : background large
- Use small R, resolve the jets : what is the right scale?
  - Also: small jets lead to huge combinatorial issues

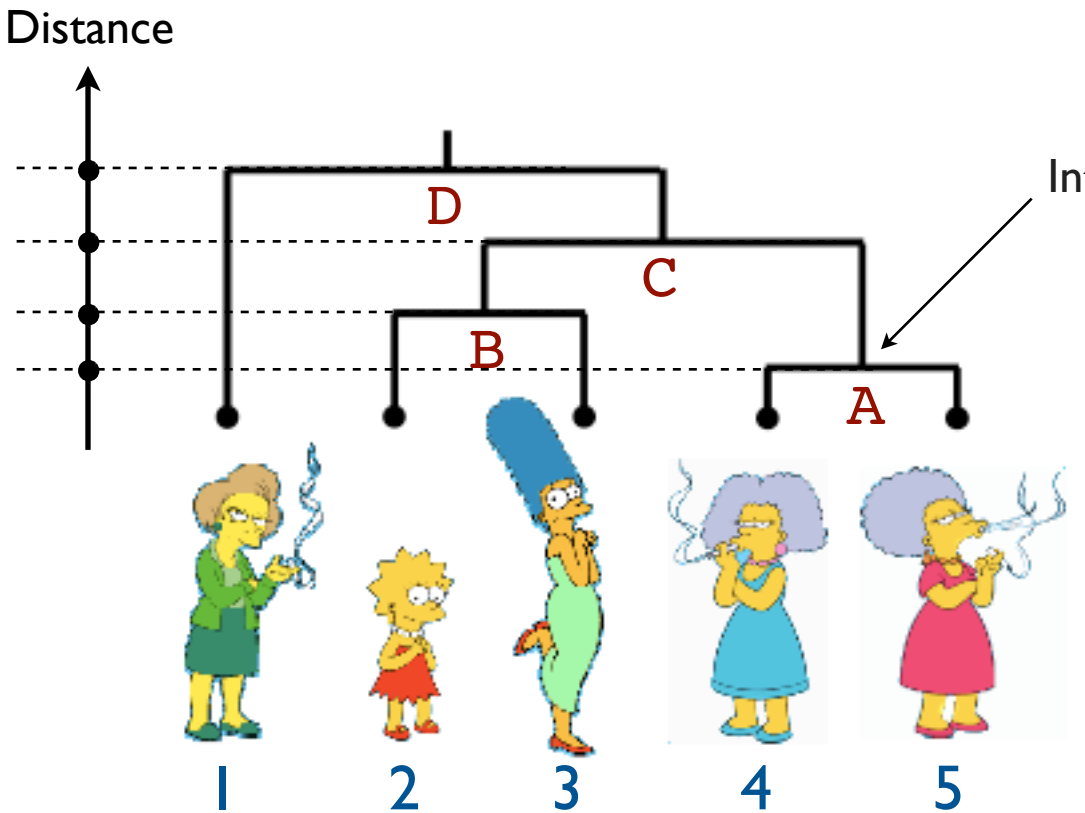
Let an algorithm find the 'right' substructure

What jets to use for substructure?

Different jet algorithms will give different 'pictures' of what's inside a jet

Dendrogram

Used to represent graphically the sequence of clustering steps in a sequential recombination algorithm



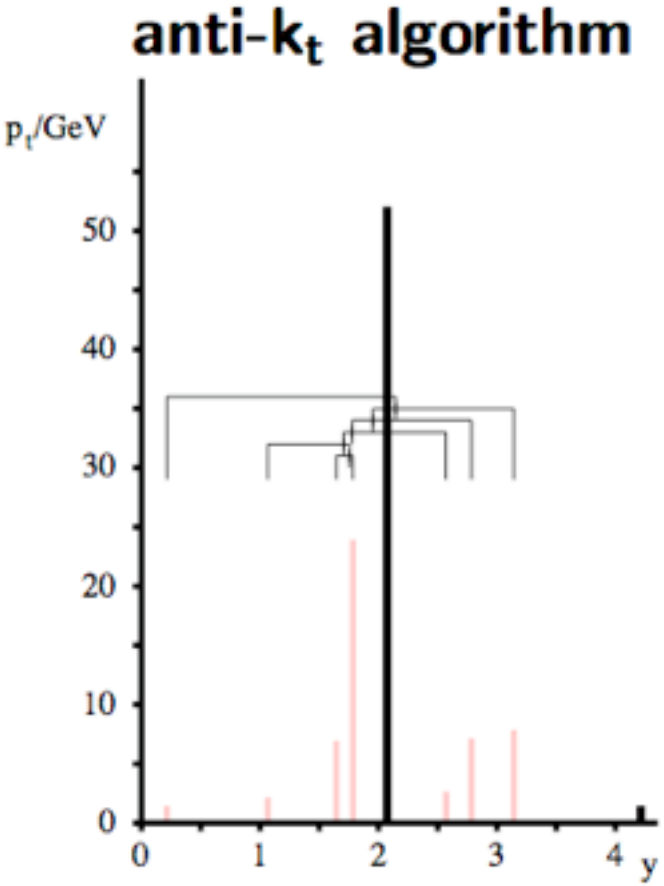
Distance between two objects is given by the **height** of the lowest internal node that they share.

Order of clustering here is A, B, C, D

The **clustering sequence** is 4-5 (A), 2-3 (B), 23-45 (C), 1-2345 (D)

anti-kt

Hierarchical substructure

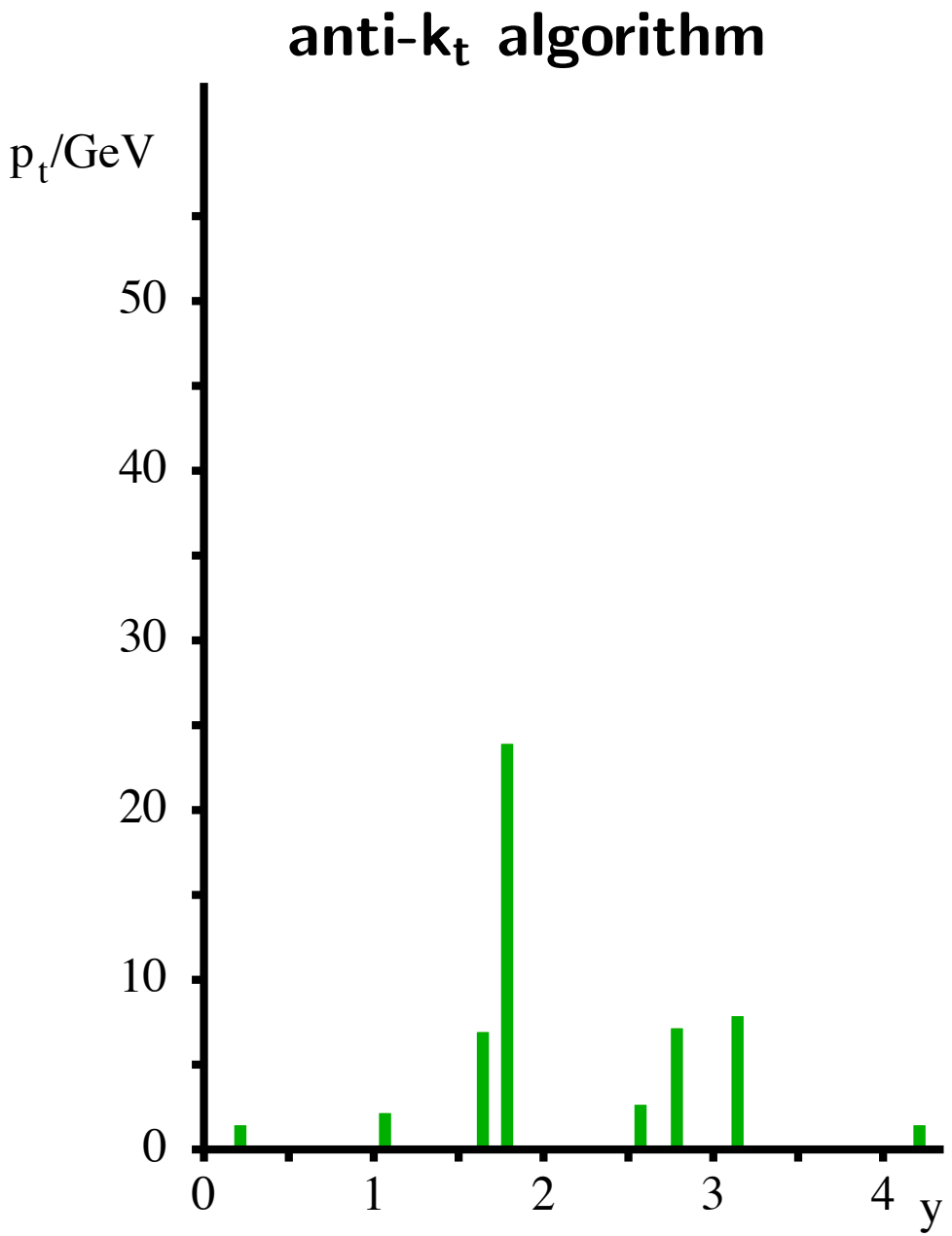


Anti-k_t distance measure

$$d_{ij} = \min\left(\frac{1}{p_{ti}^2}, \frac{1}{p_{tj}^2}\right)\frac{\Delta y^2 + \Delta\phi^2}{R^2}$$

Cluster by merging
to the **hardest/closest** particle

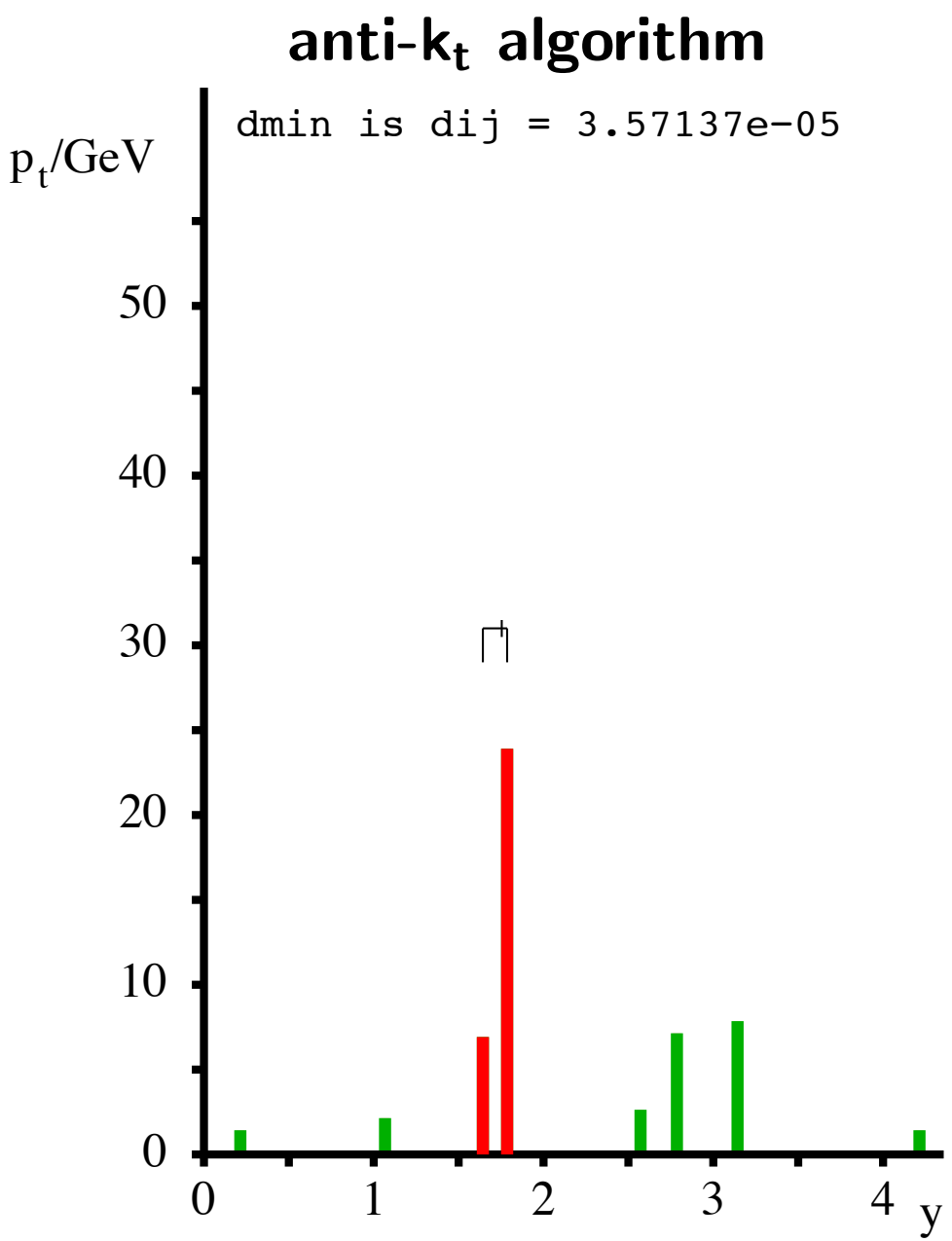
Identifying jet substructure: try out anti-k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

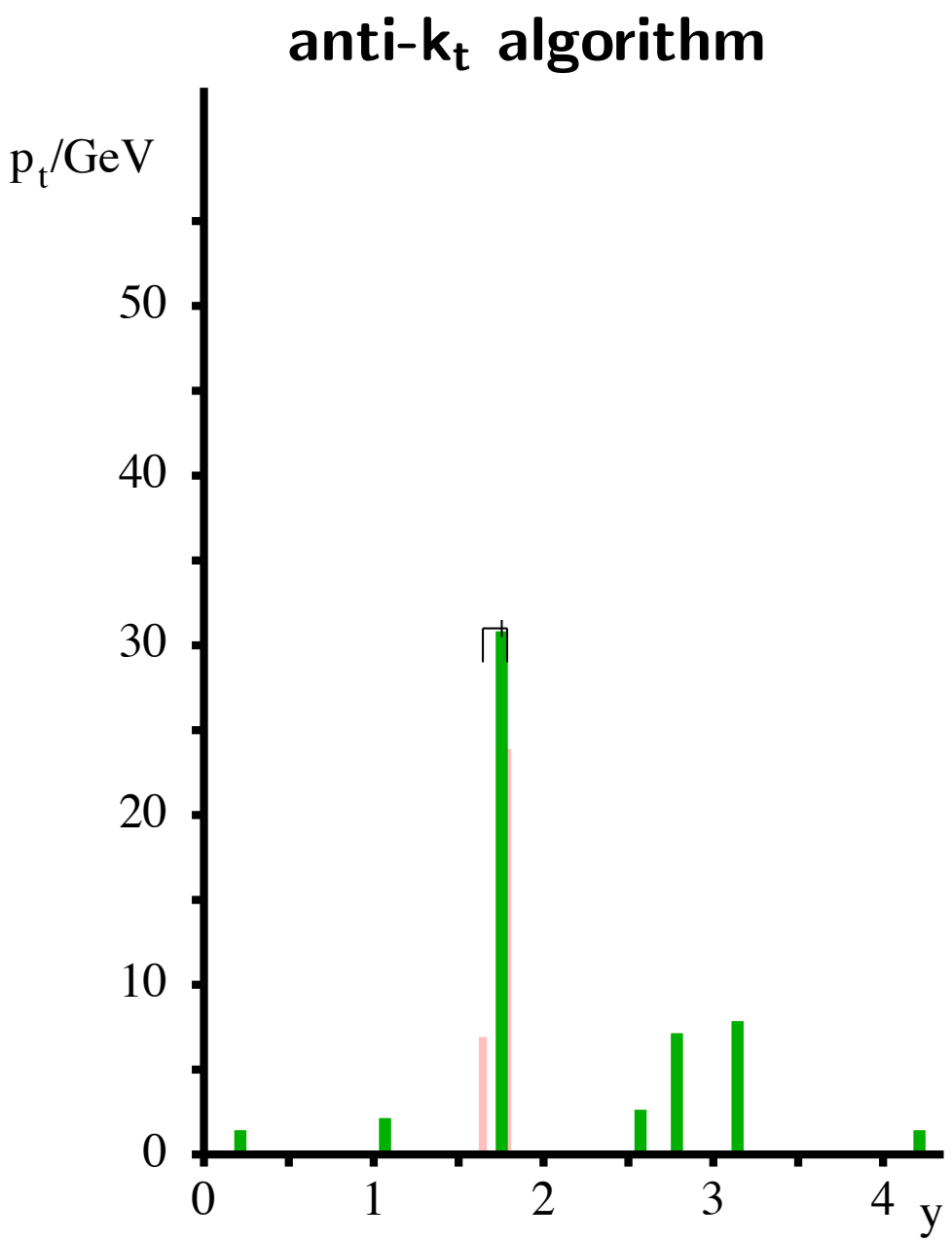
Identifying jet substructure: try out anti-k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

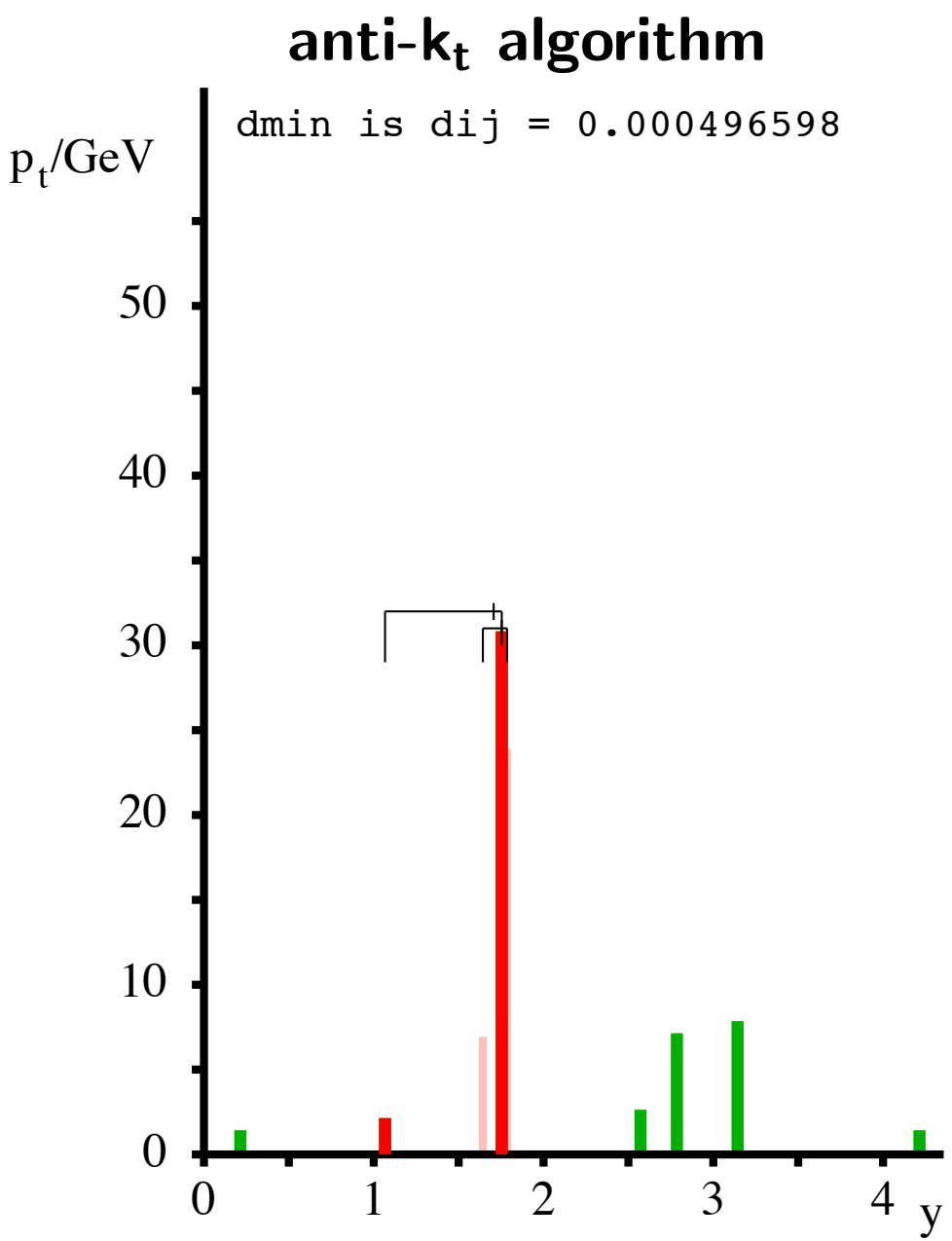
Identifying jet substructure: try out anti-k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

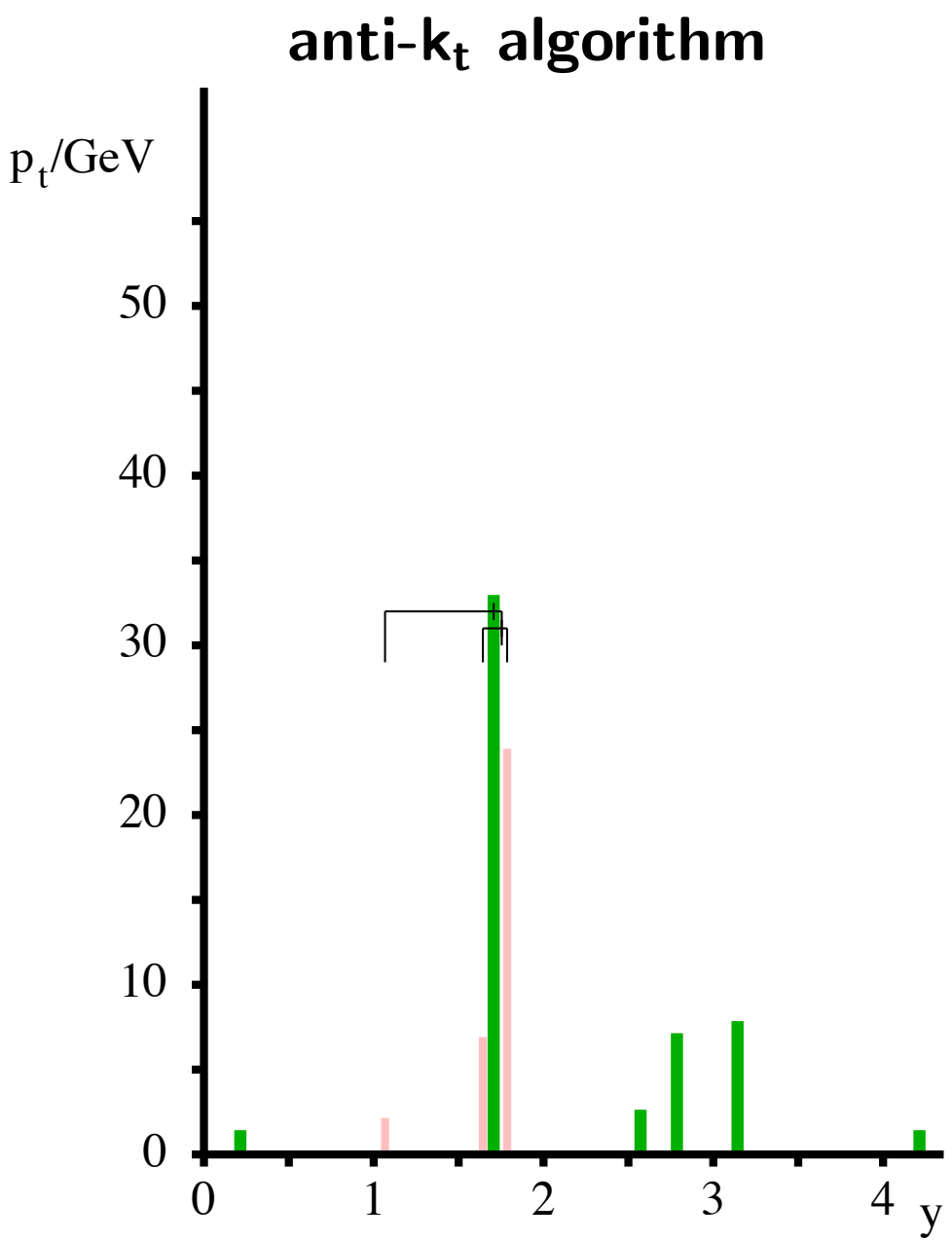
Identifying jet substructure: try out anti-k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

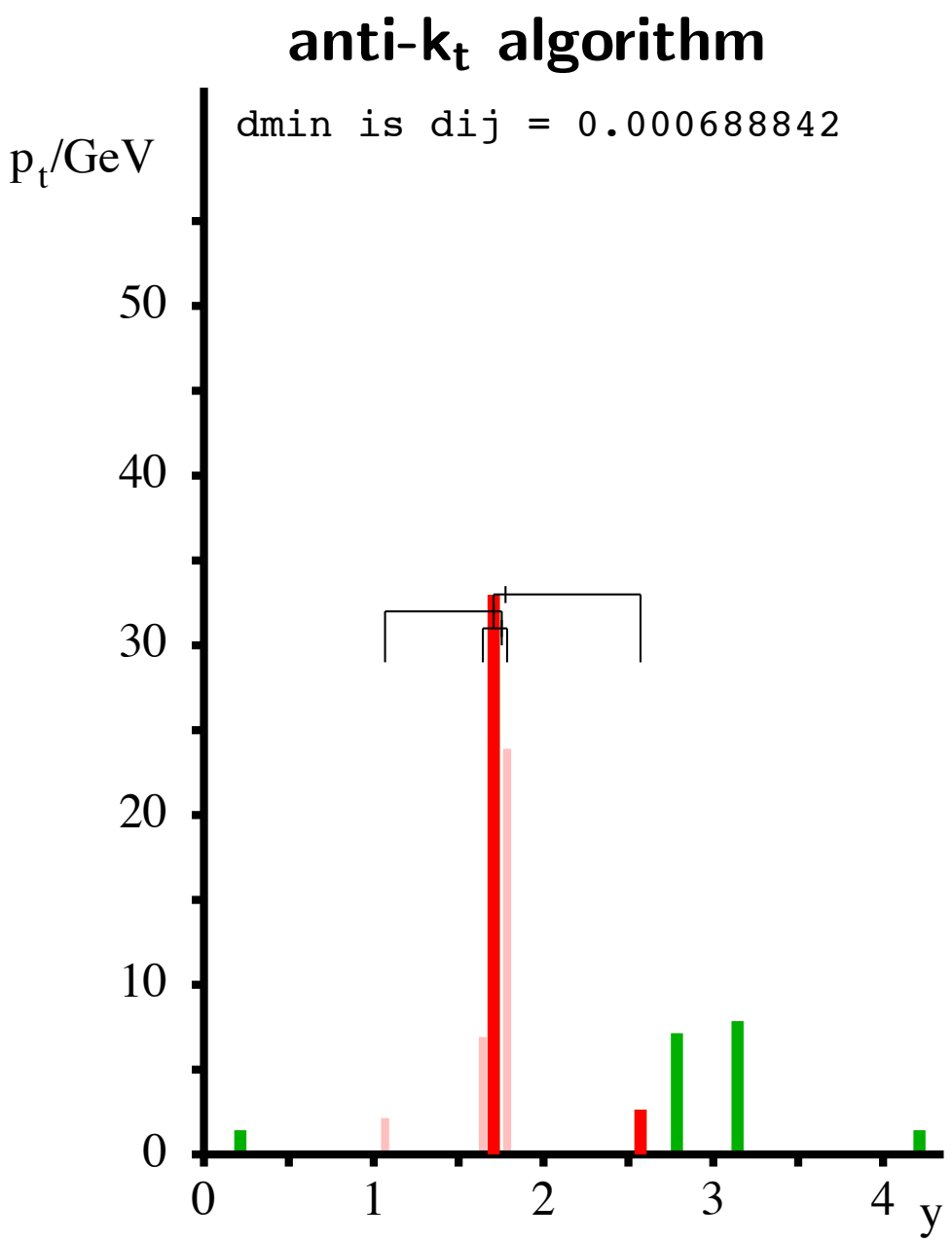
Identifying jet substructure: try out anti-k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

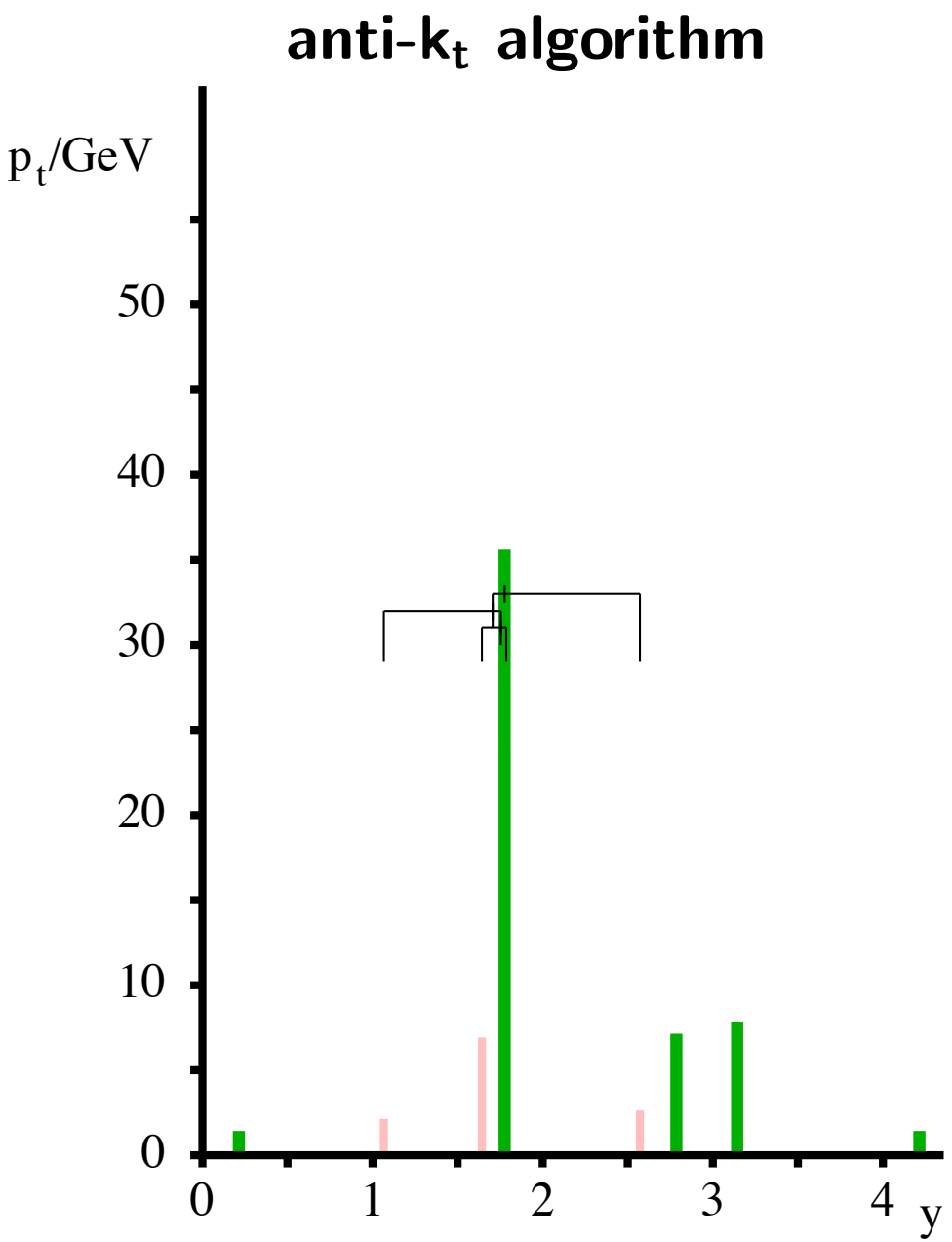
Identifying jet substructure: try out anti-k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

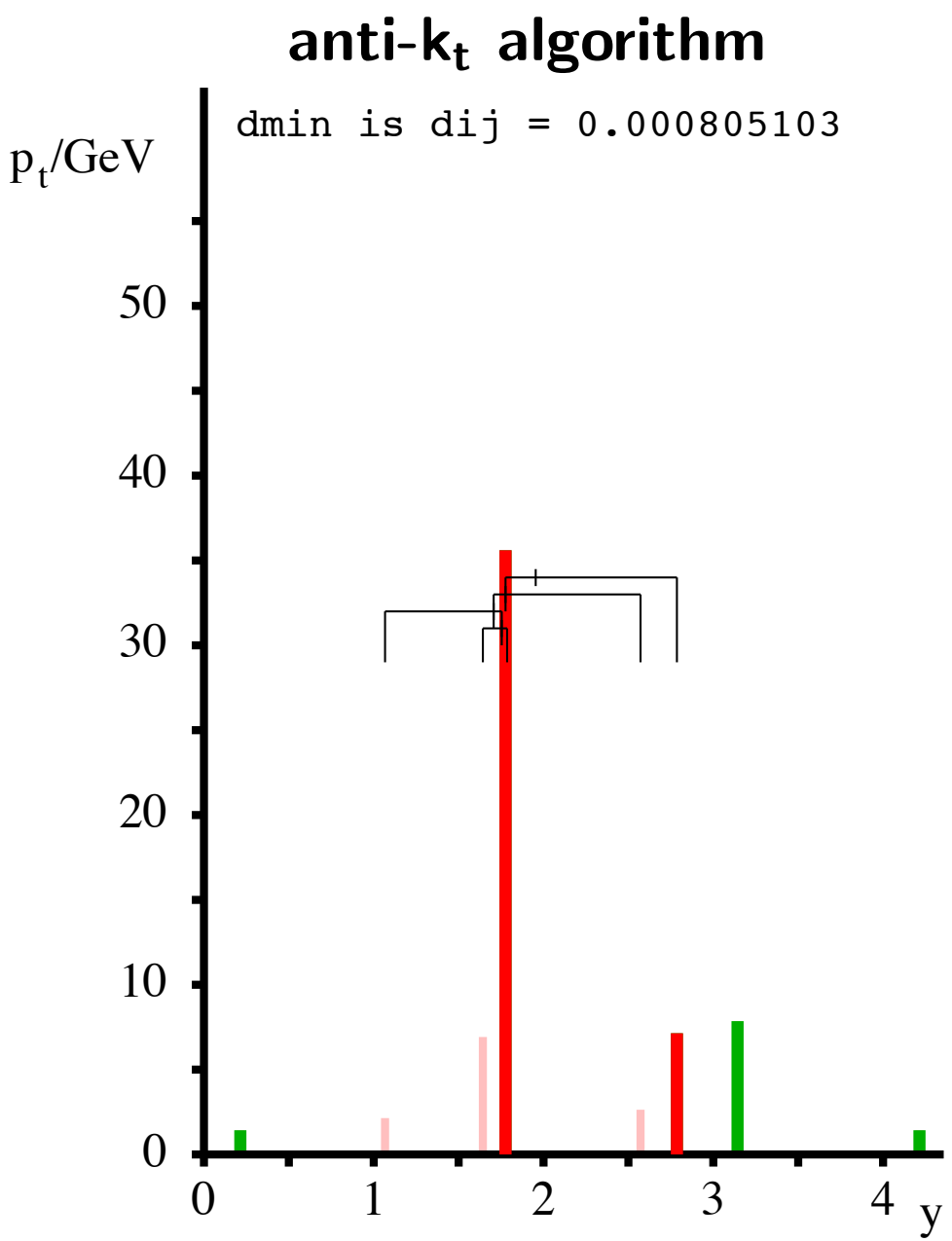
Identifying jet substructure: try out anti-k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Identifying jet substructure: try out anti-k_t

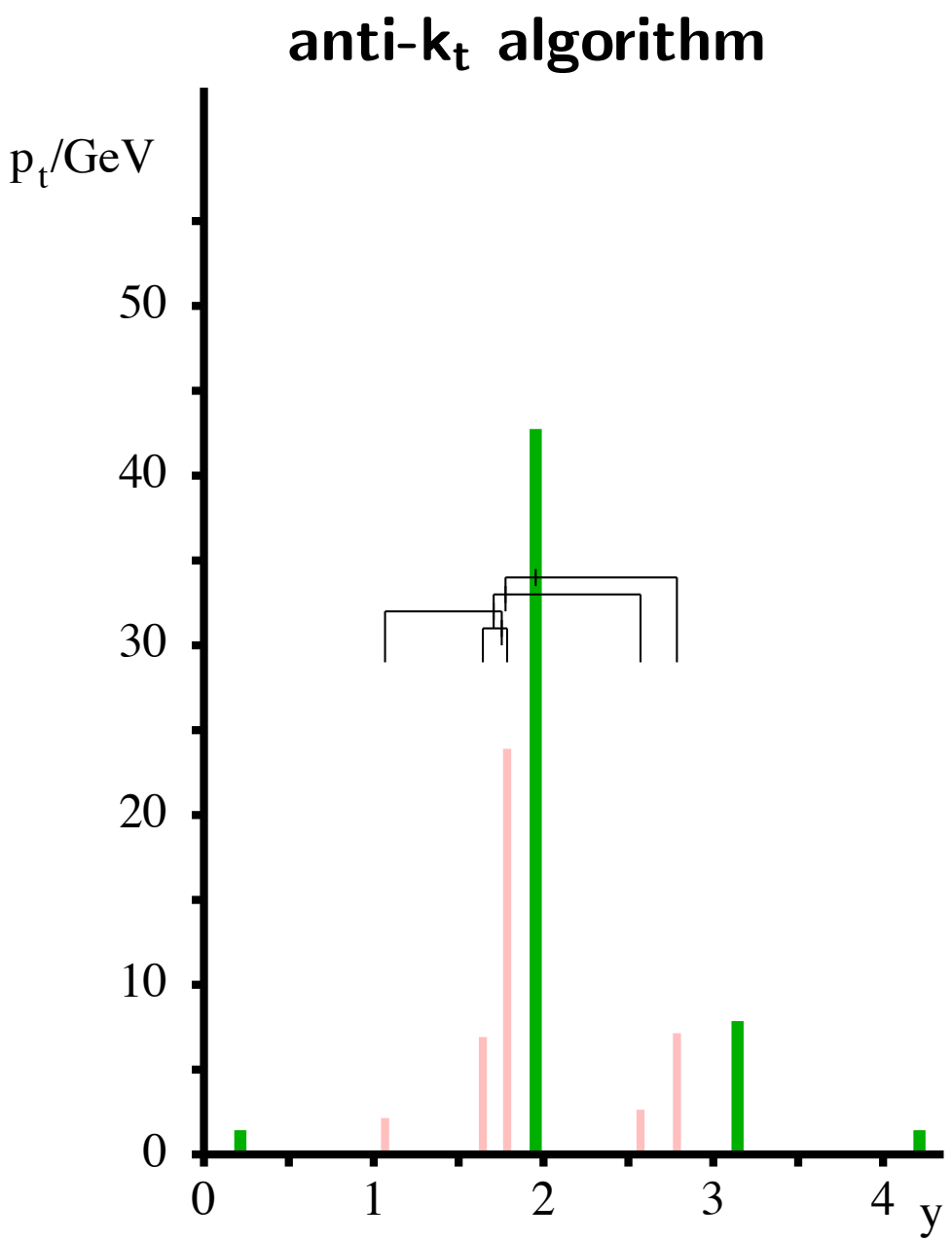


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Identifying jet substructure: try out anti-k_t

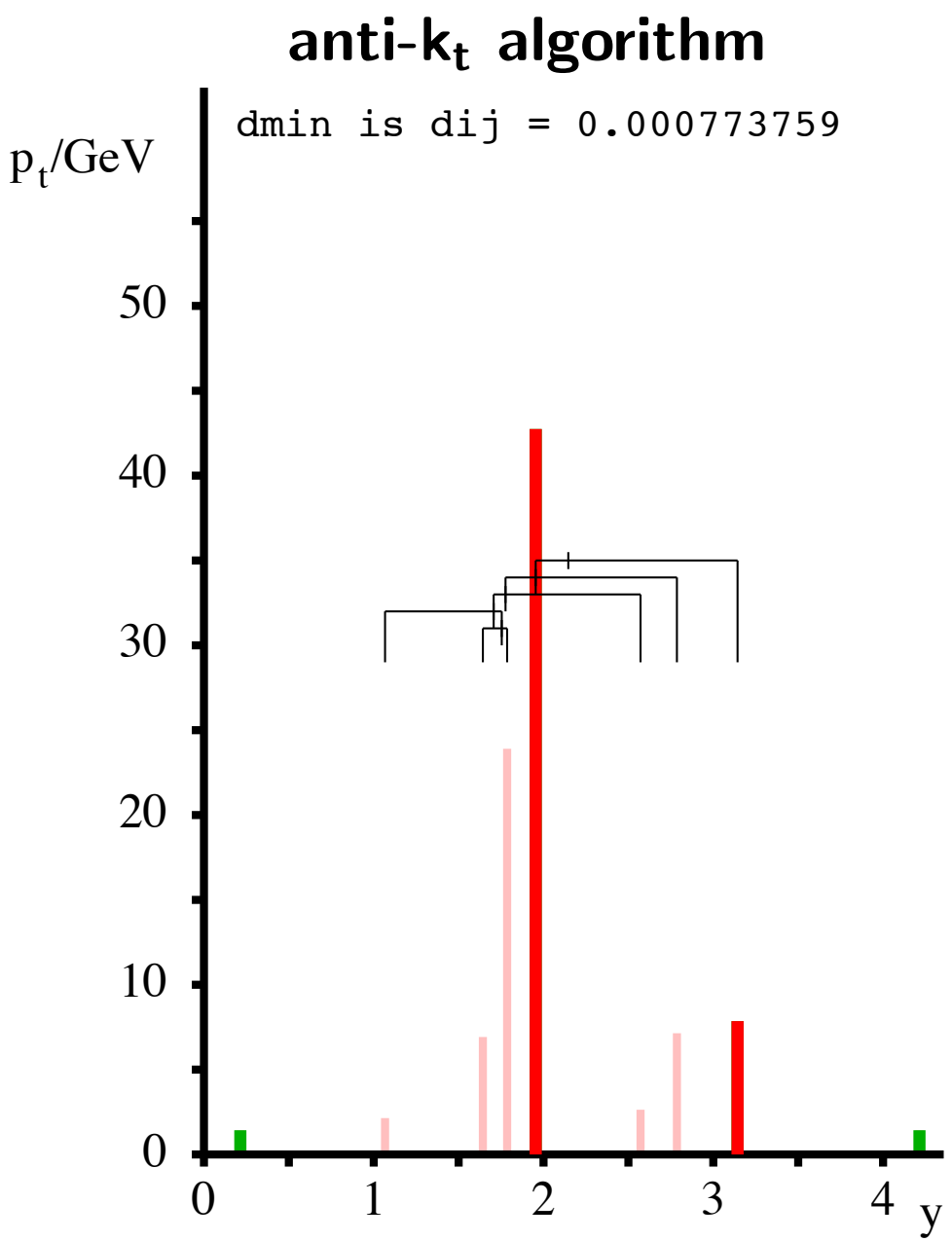


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Identifying jet substructure: try out anti-k_t

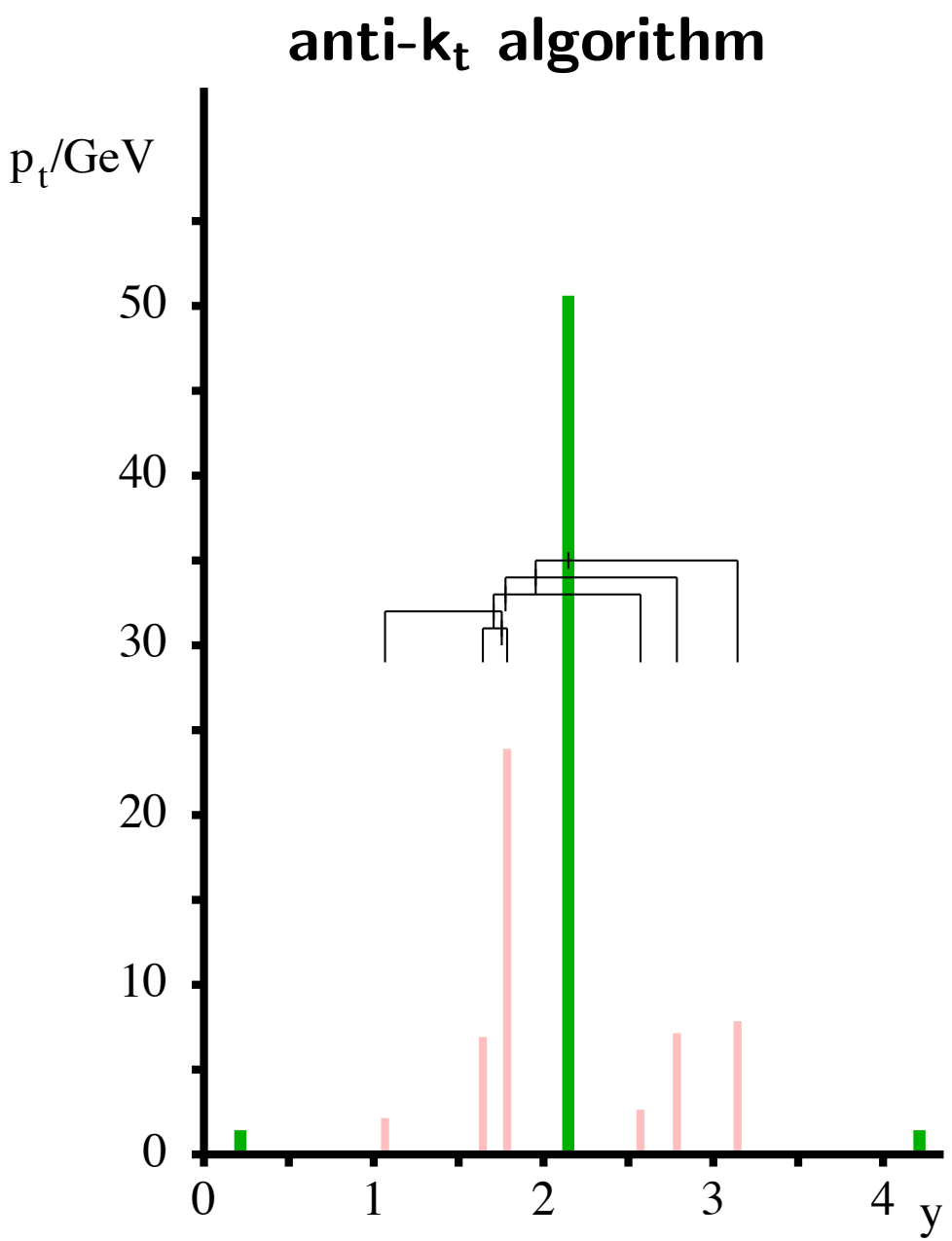


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Identifying jet substructure: try out anti-k_t

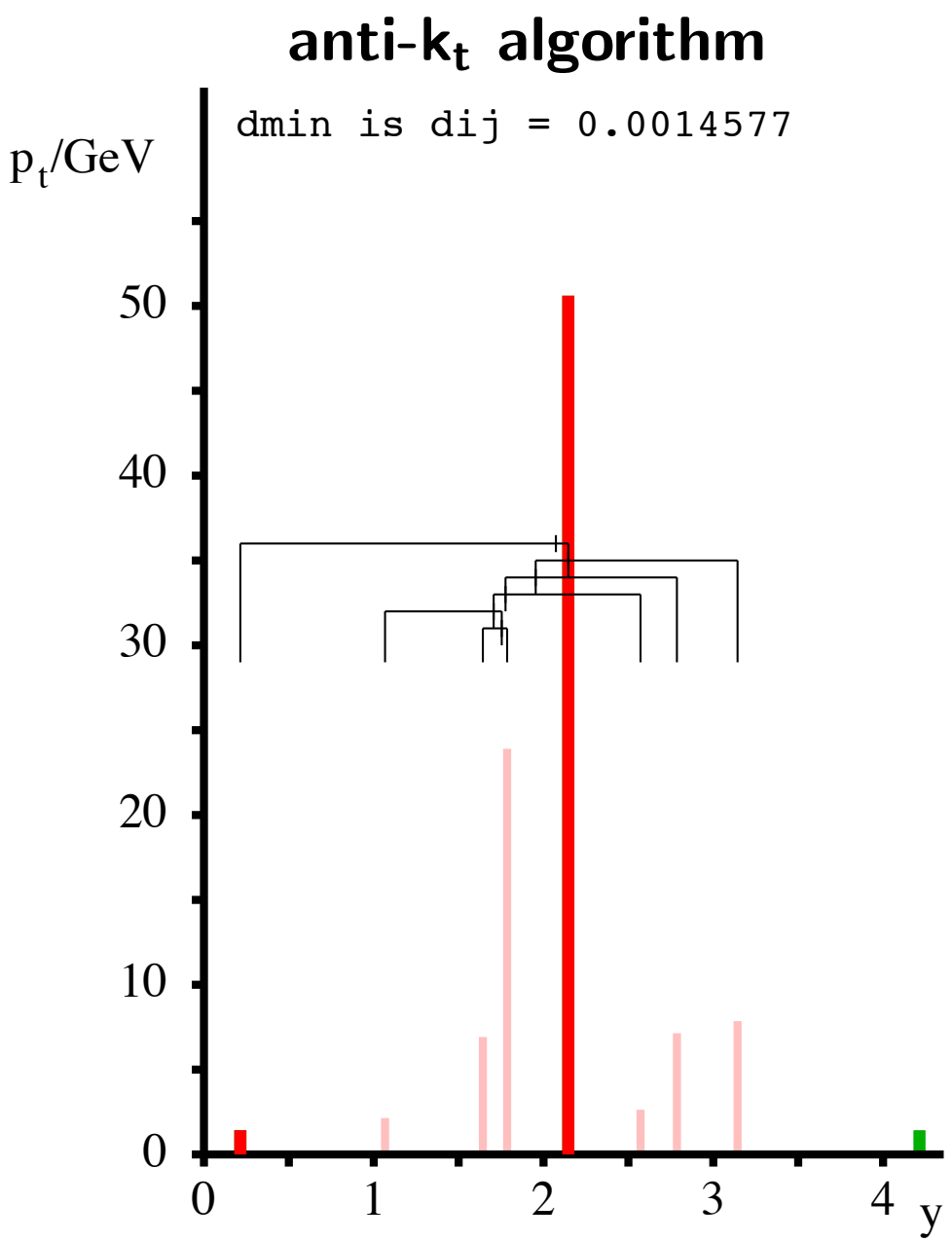


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Identifying jet substructure: try out anti-k_t

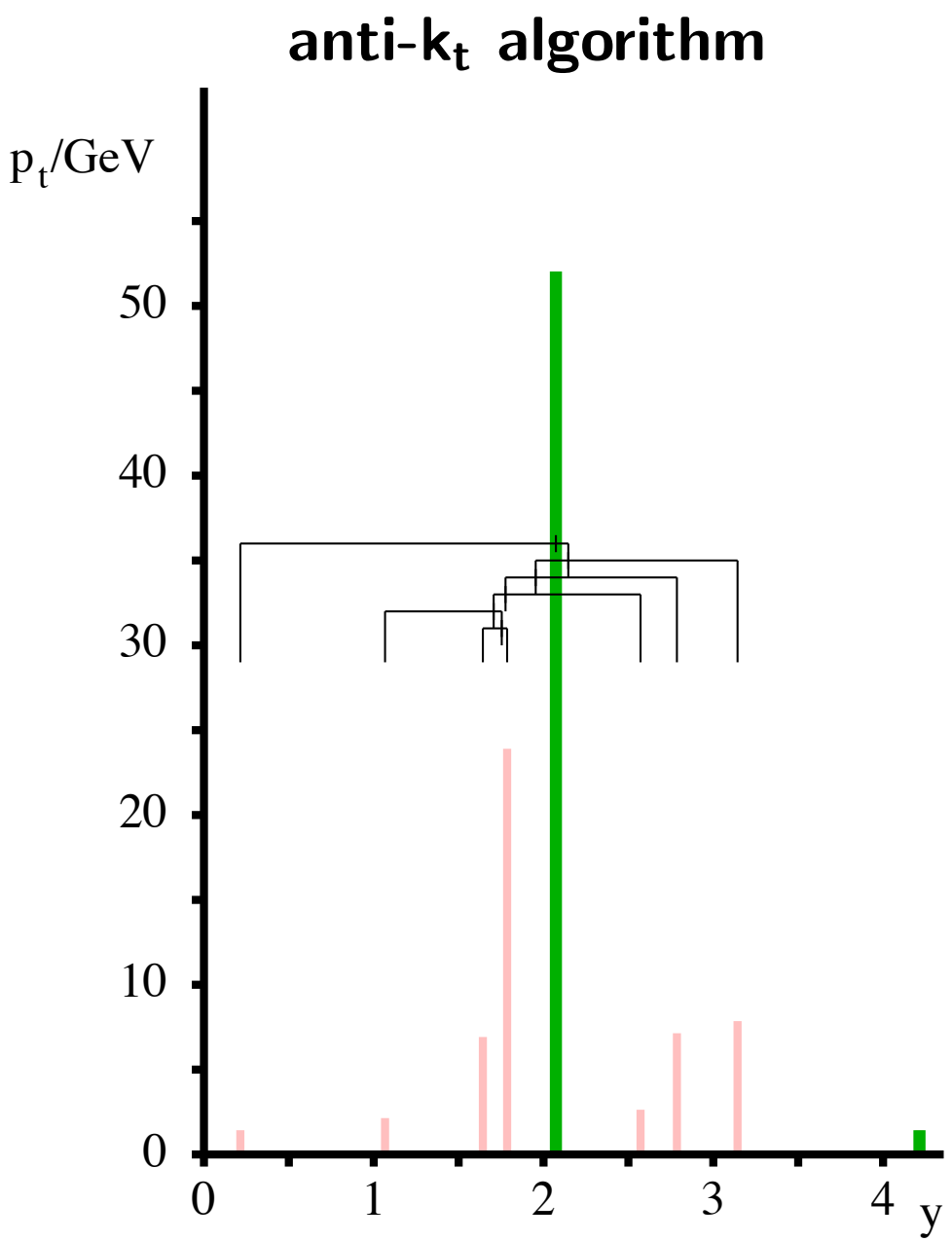


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Identifying jet substructure: try out anti-k_t

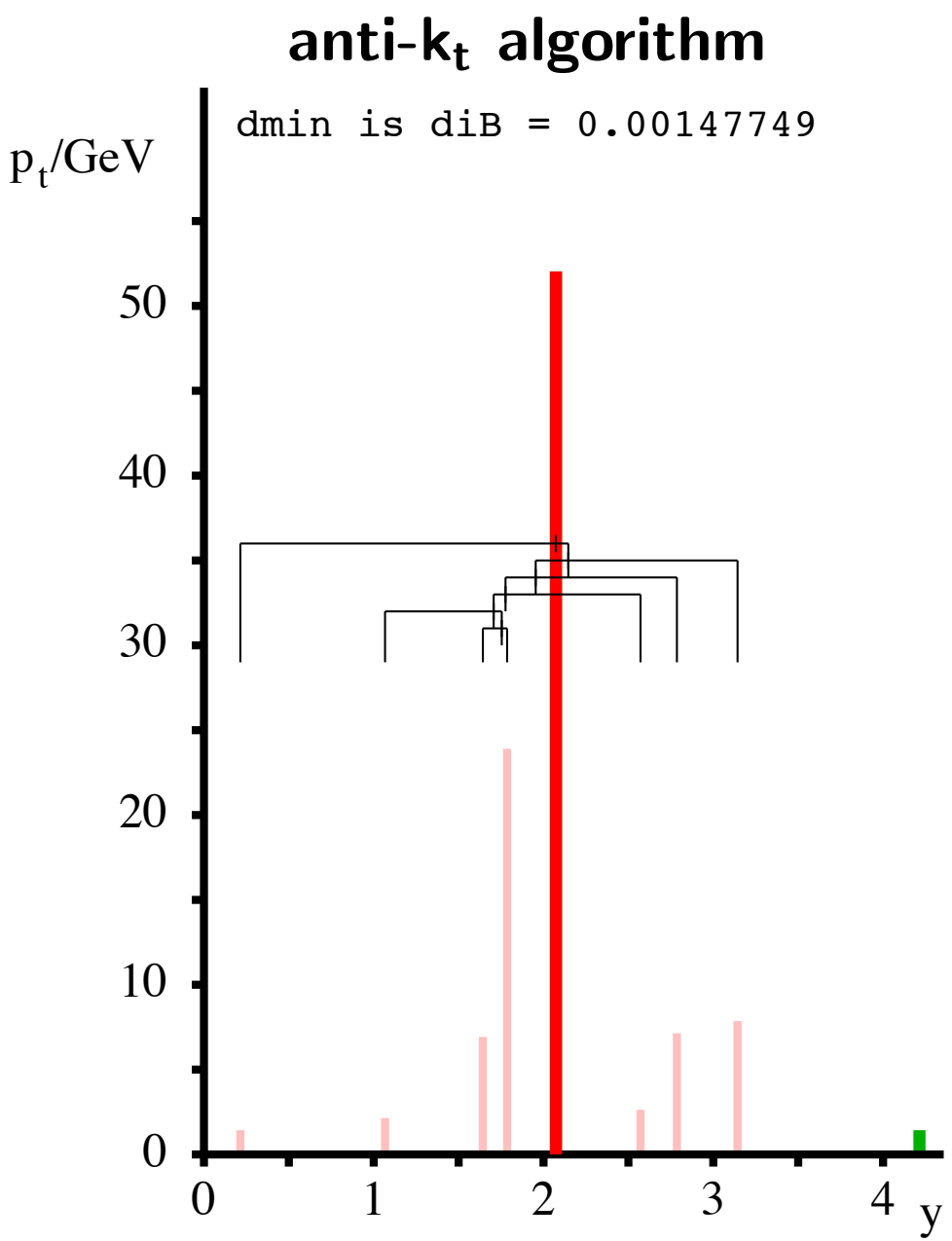


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Identifying jet substructure: try out anti-k_t

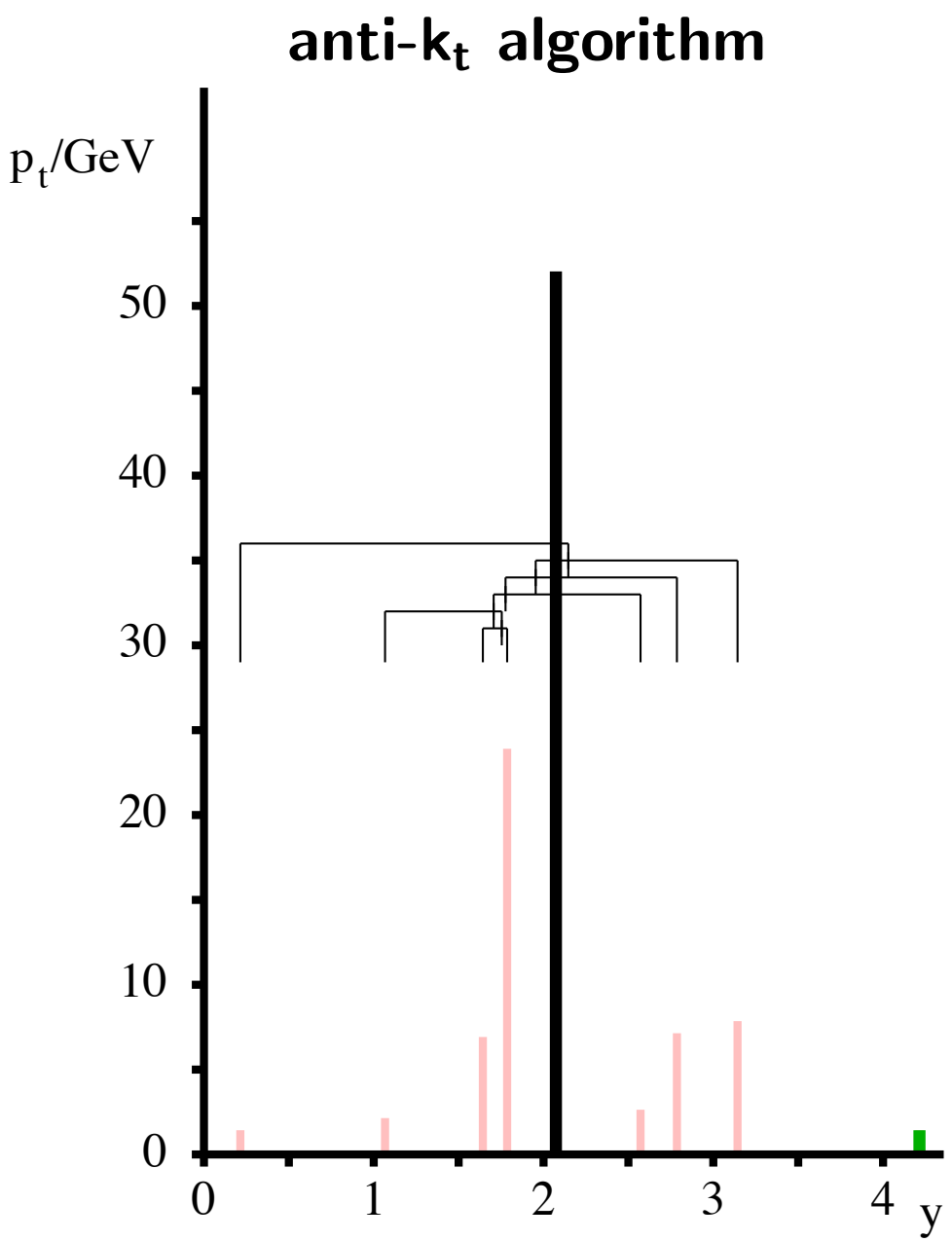


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Identifying jet substructure: try out anti-k_t

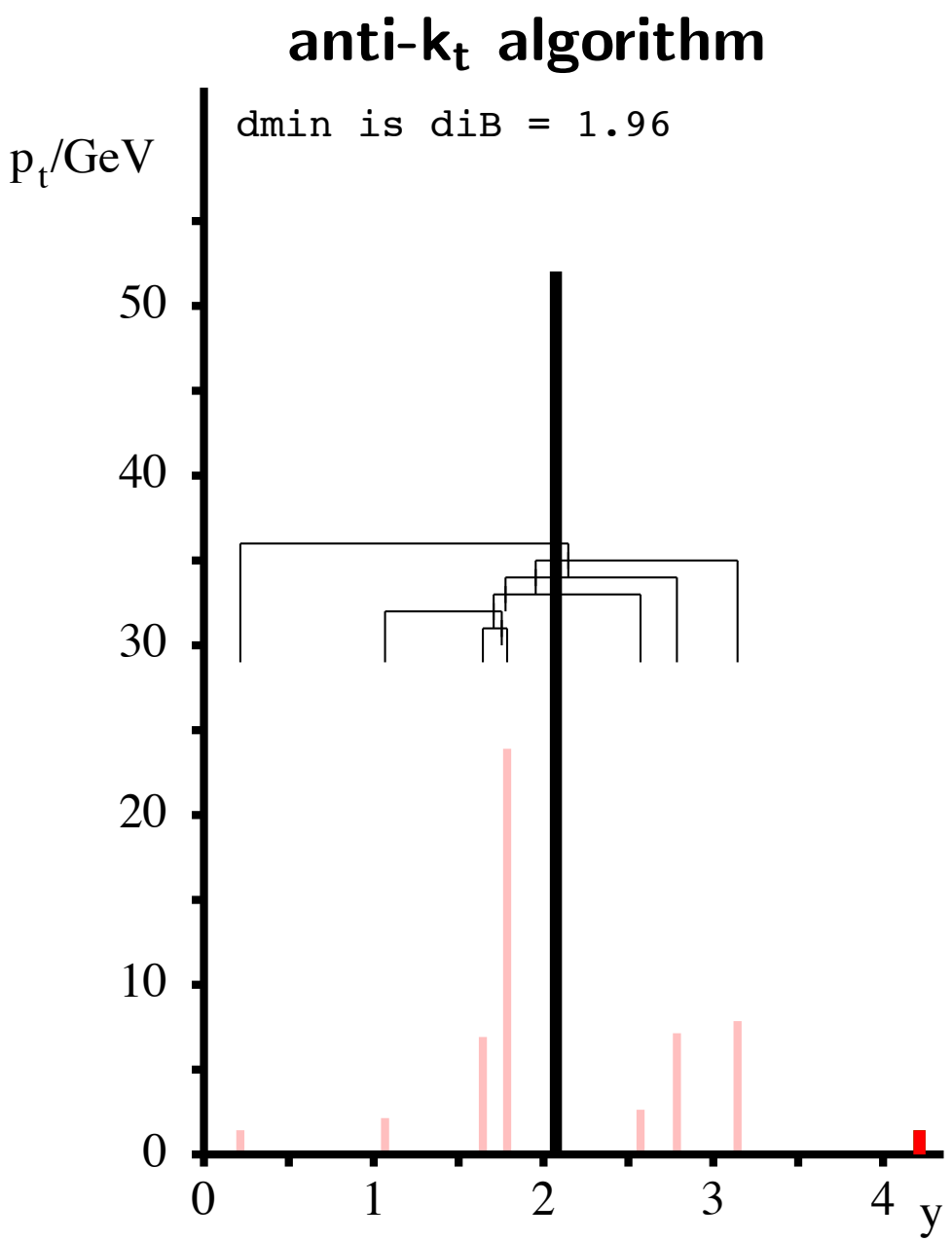


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Identifying jet substructure: try out anti-k_t

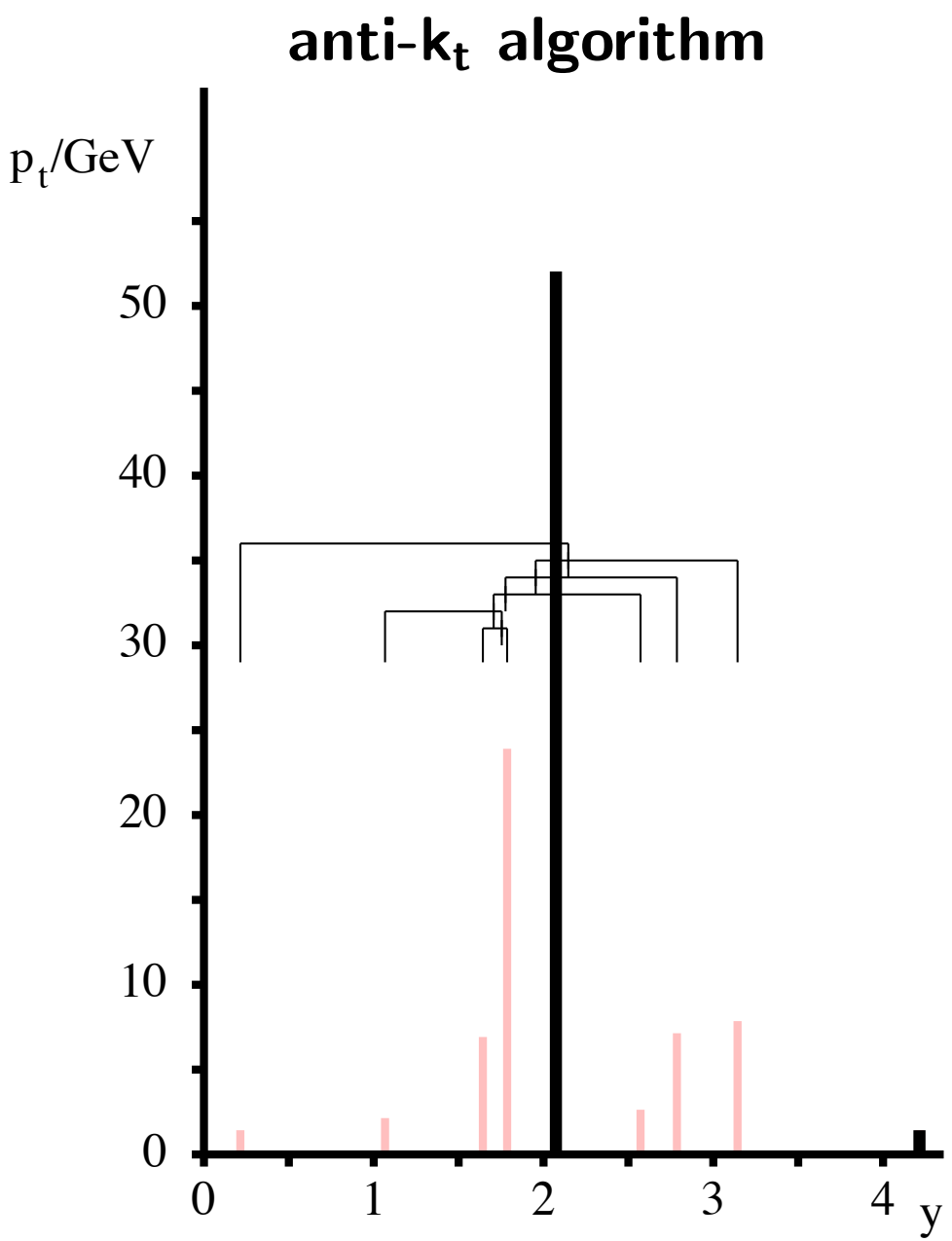


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Identifying jet substructure: try out anti-k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

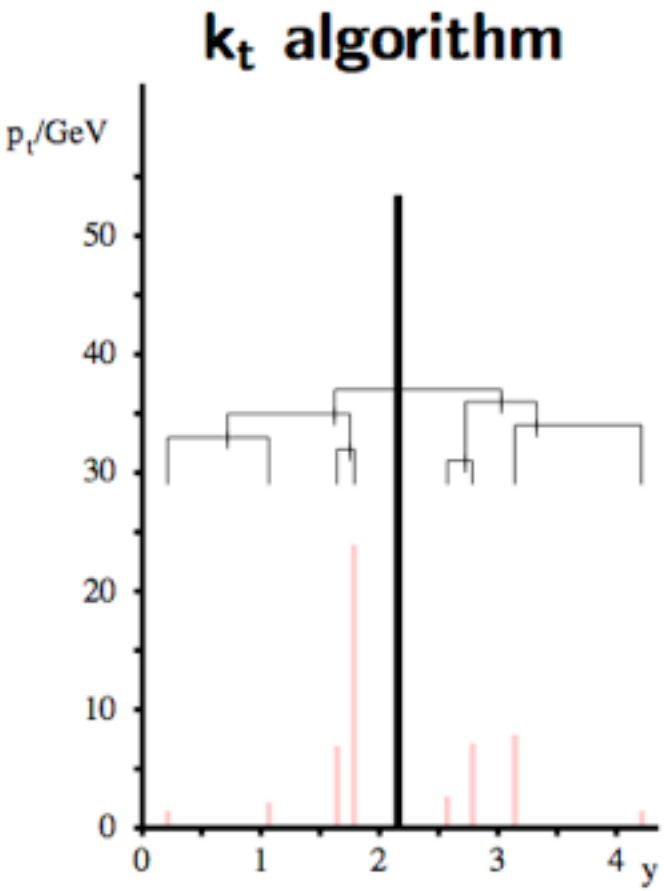
This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Anti-k_t gradually makes its way through the secondary blob $\rightarrow$ no clear identification of substructure associated with 2nd parton.

Second try

kt

Hierarchical substructure

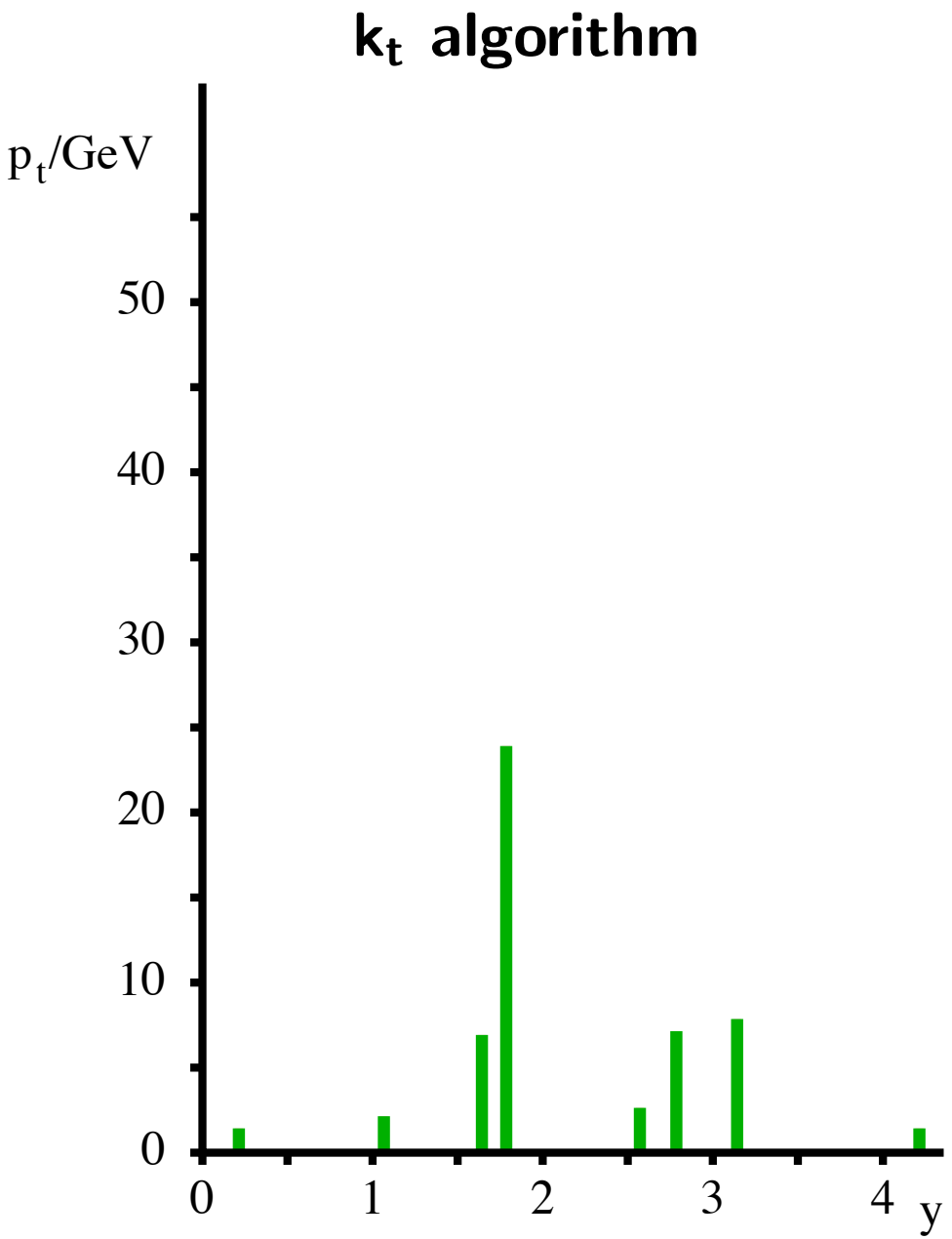


k_t distance measure

$$d_{ij} = \min(p_{ti}^2, p_{tj}^2) \frac{\Delta y^2 + \Delta \phi^2}{R^2}$$

Cluster by merging
the **softest/closest** particles

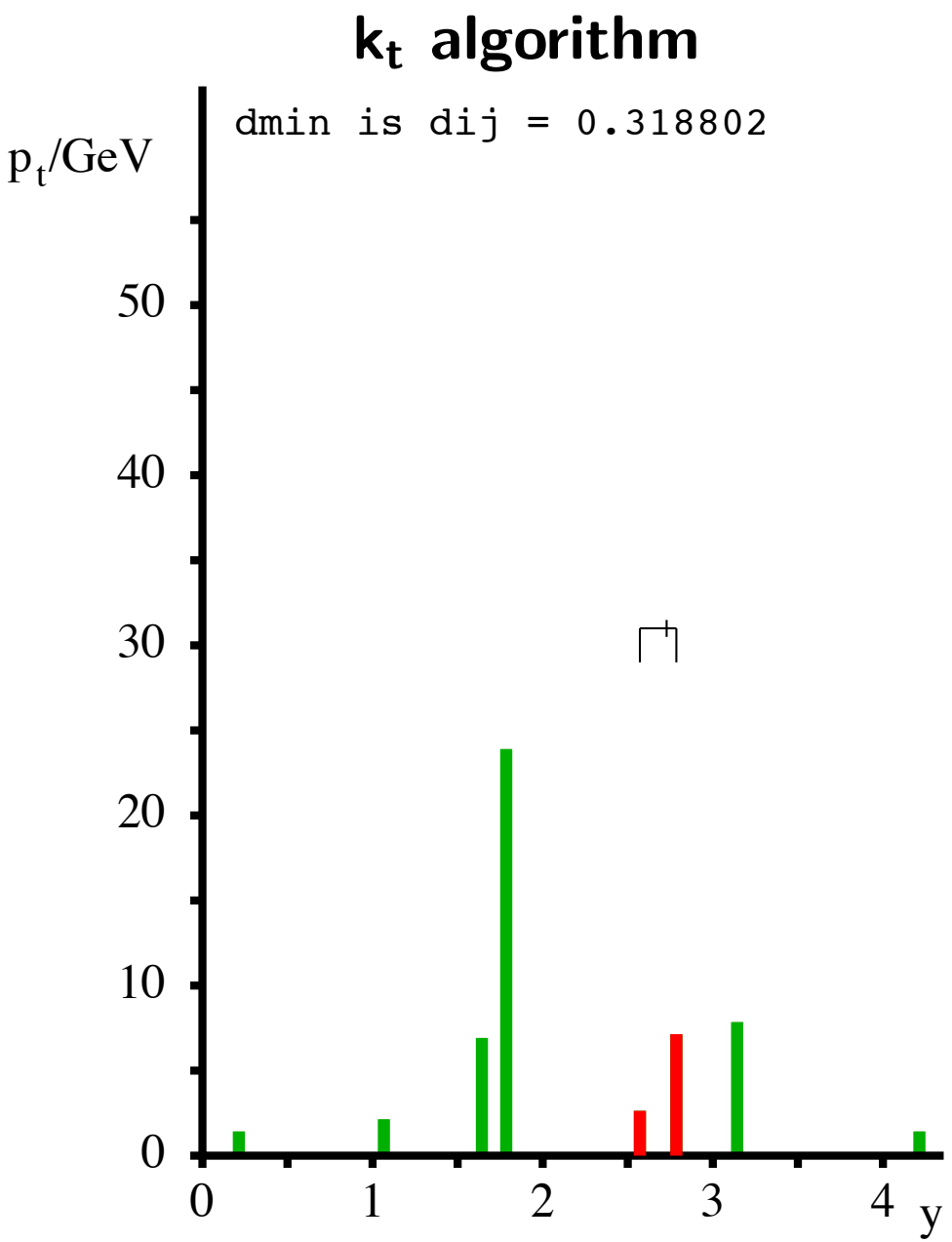
Identifying jet substructure: try out k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

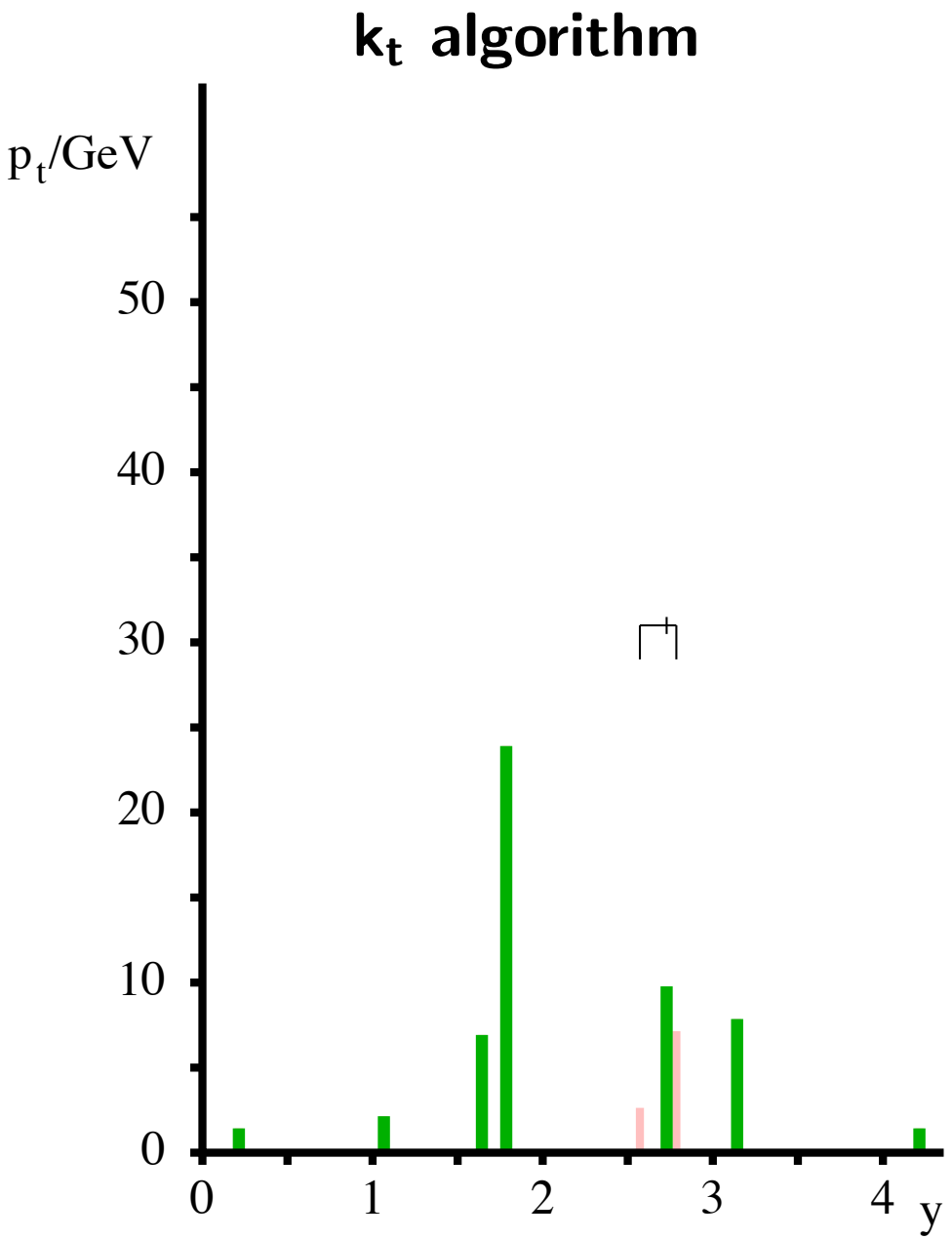
Identifying jet substructure: try out k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

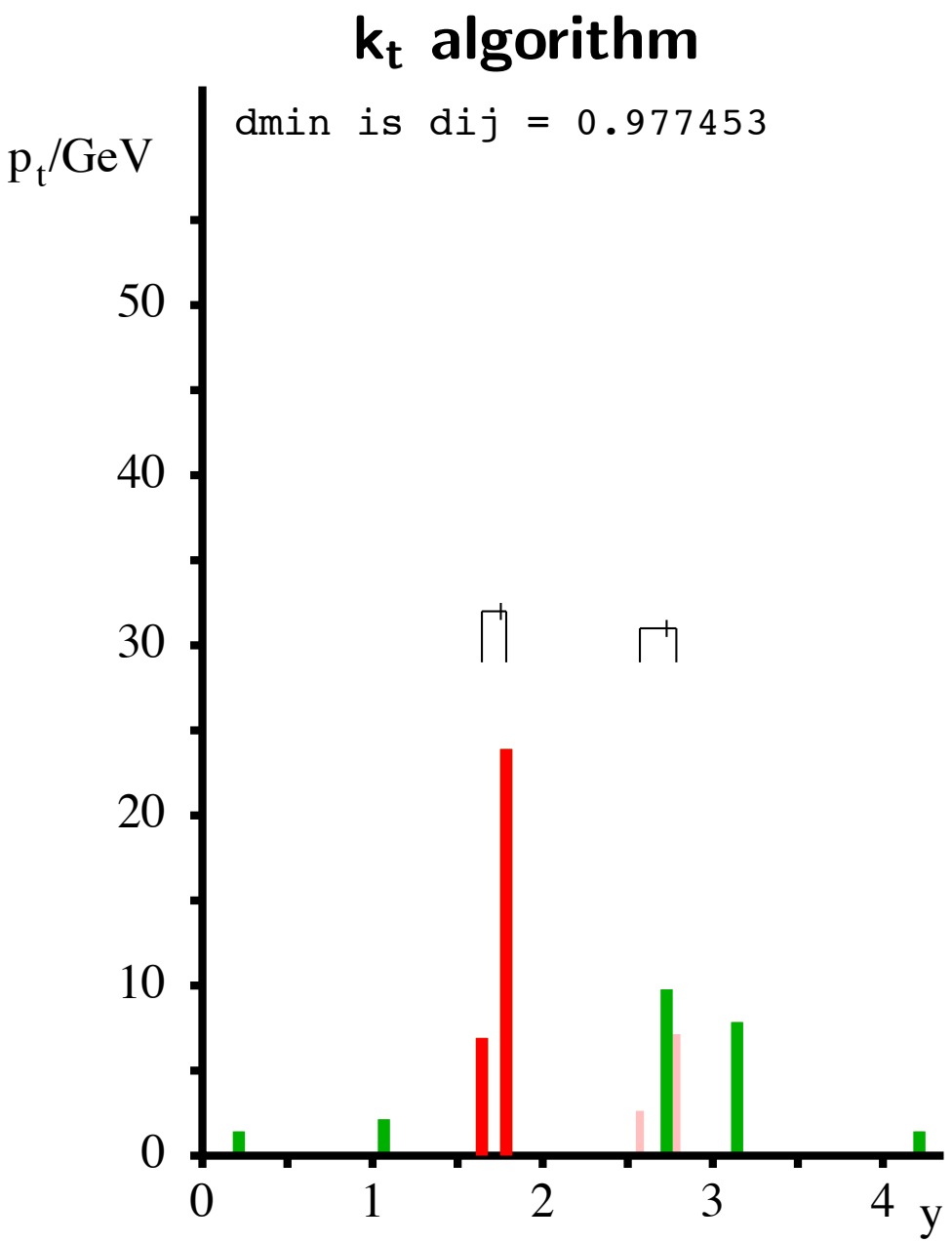
Identifying jet substructure: try out k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

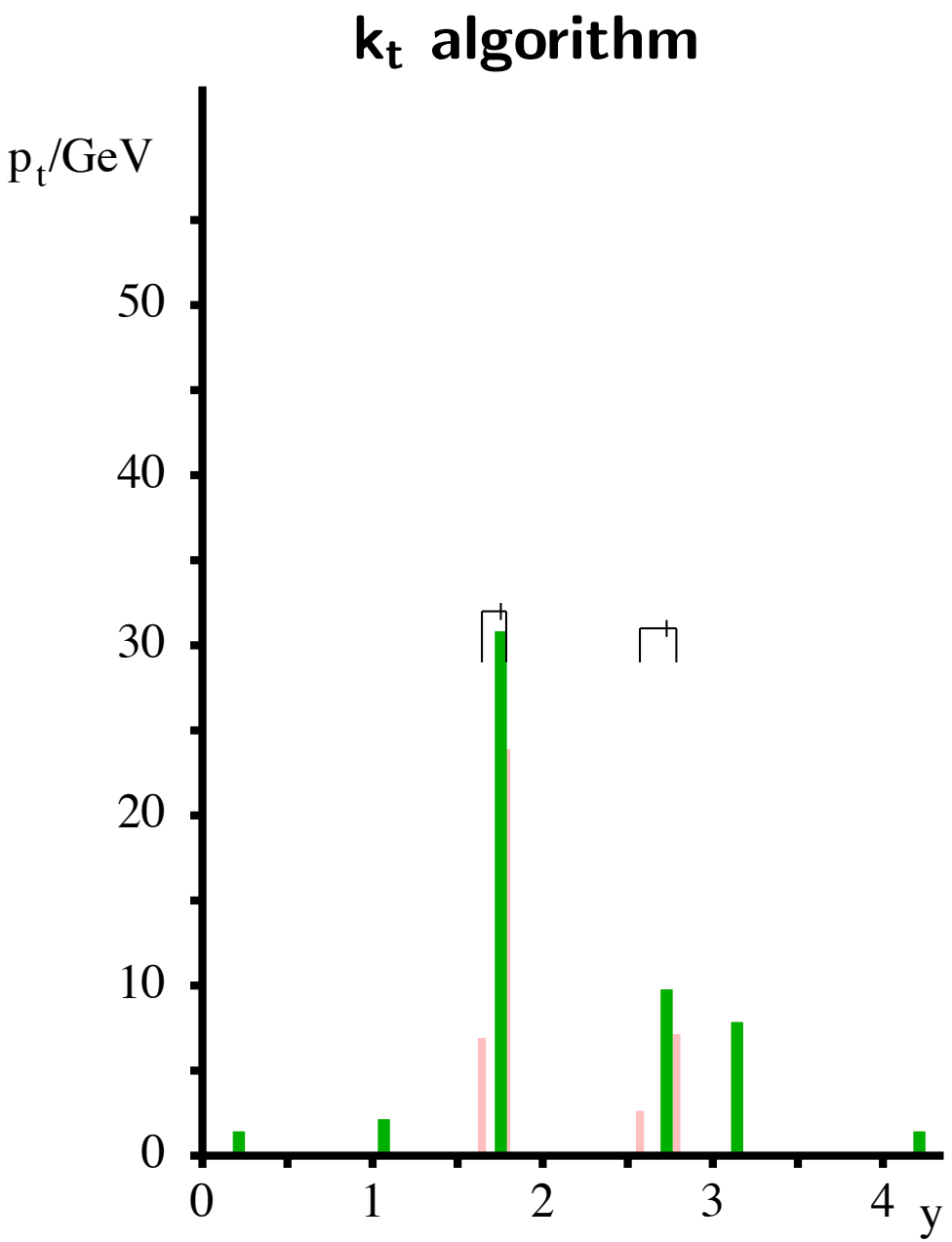
Identifying jet substructure: try out k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

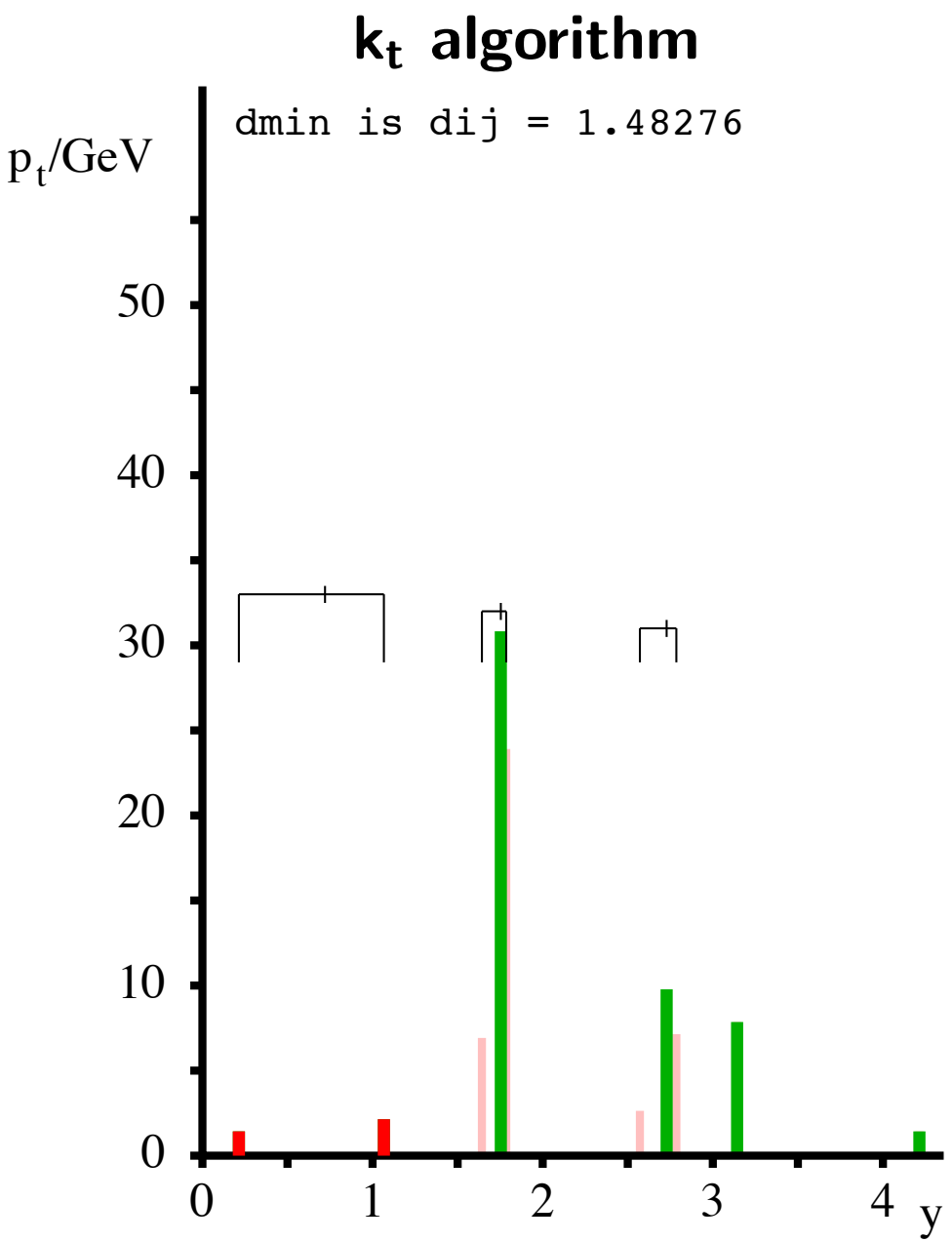
Identifying jet substructure: try out k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

Identifying jet substructure: try out k_t

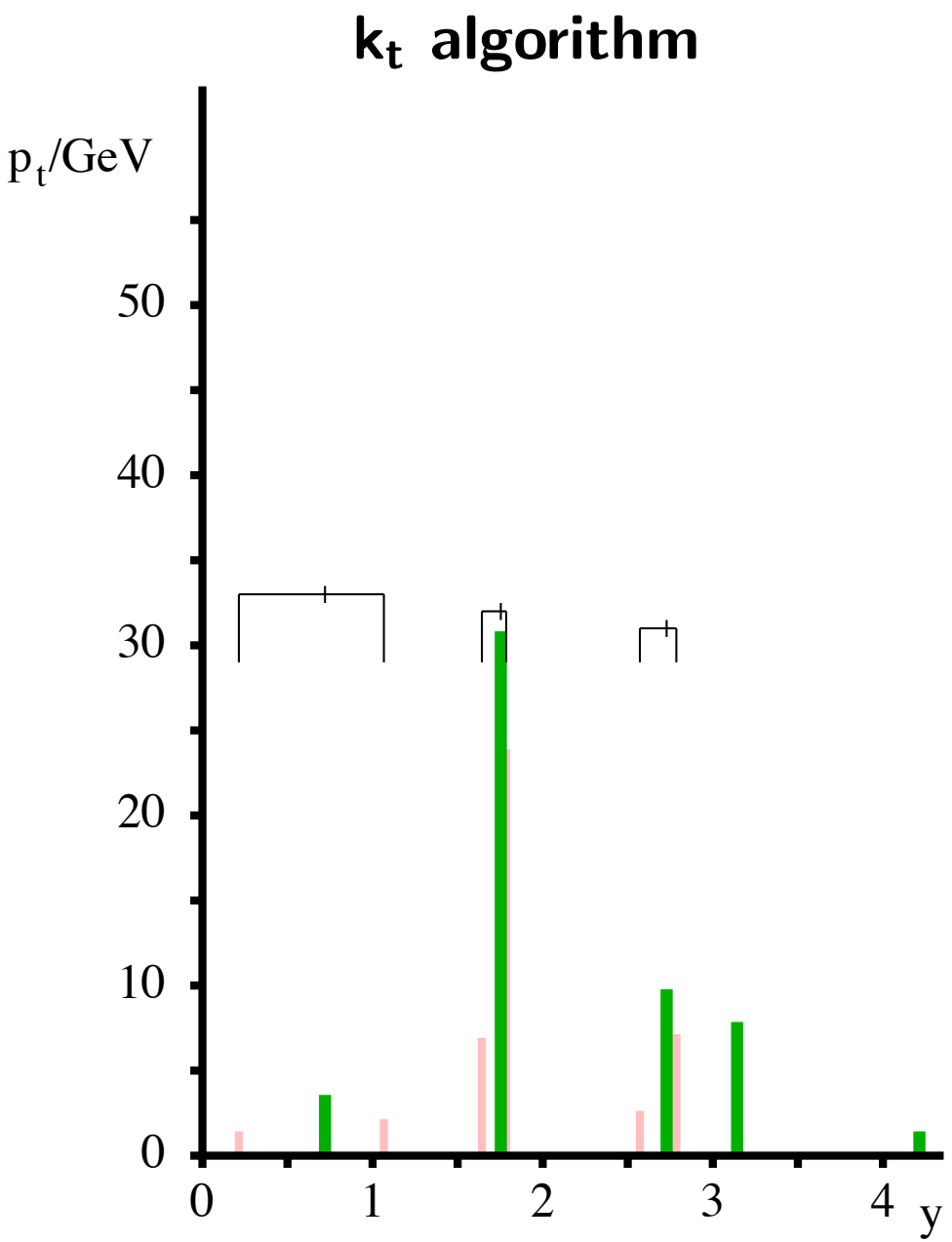


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Identifying jet substructure: try out k_t

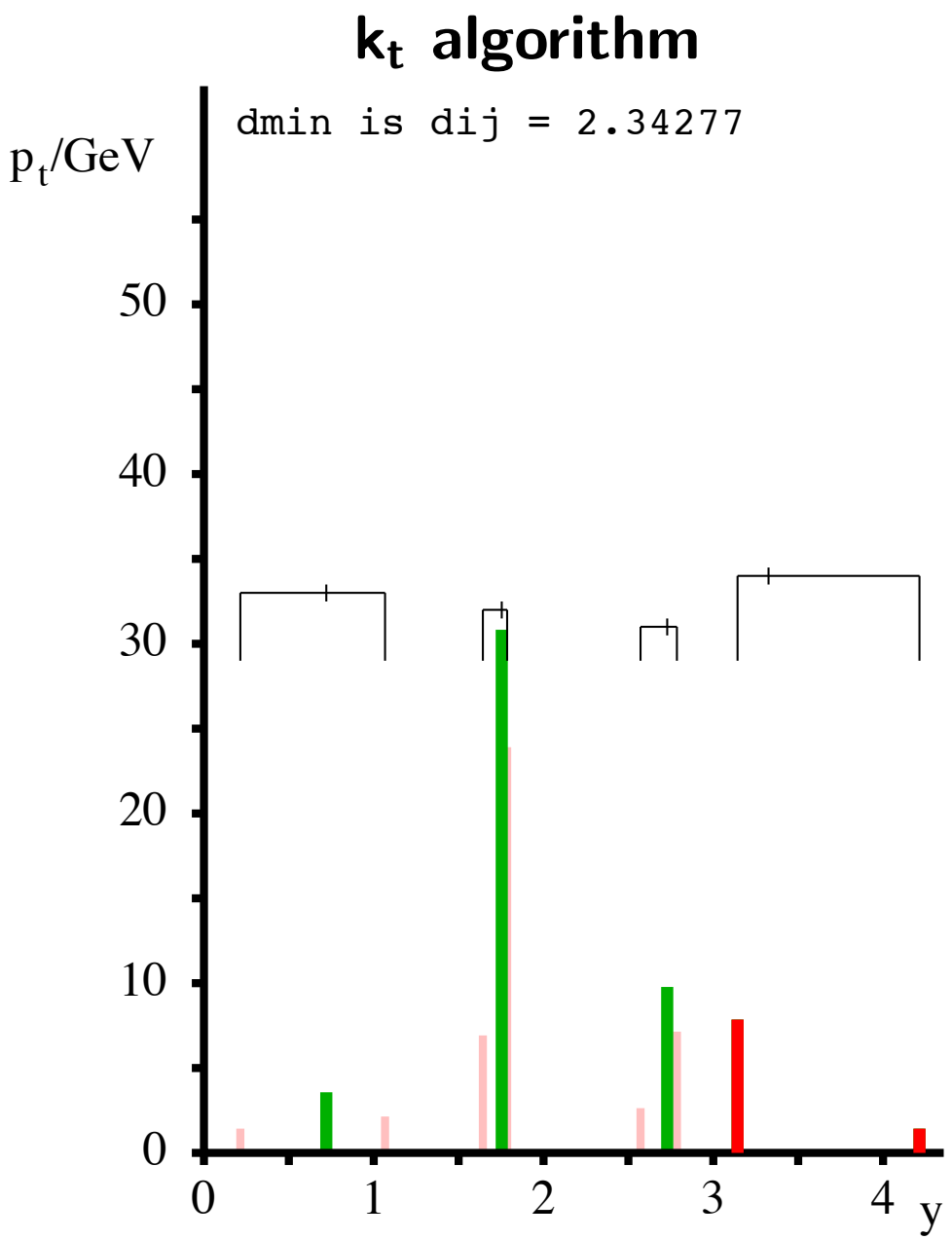


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Identifying jet substructure: try out k_t

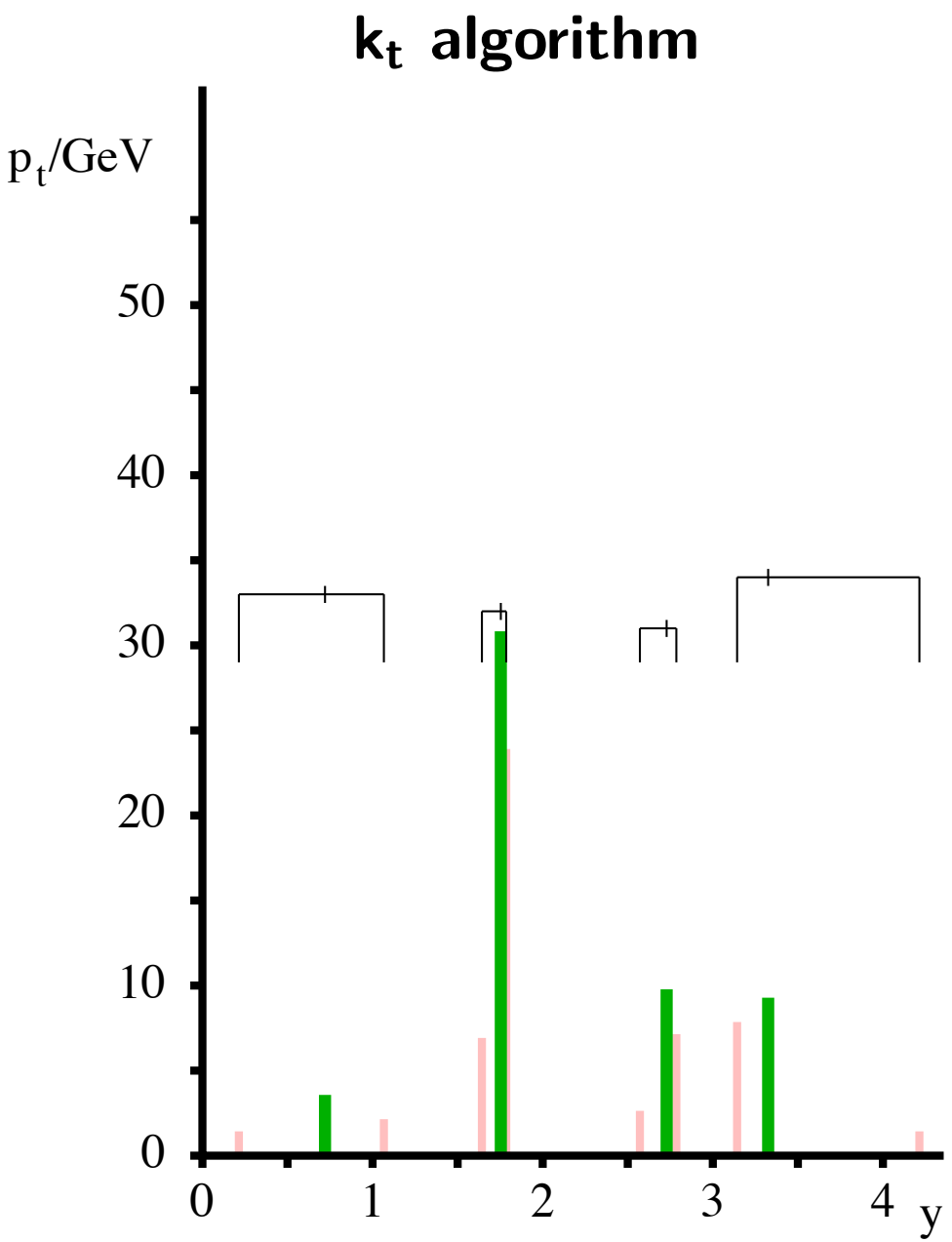


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Identifying jet substructure: try out k_t

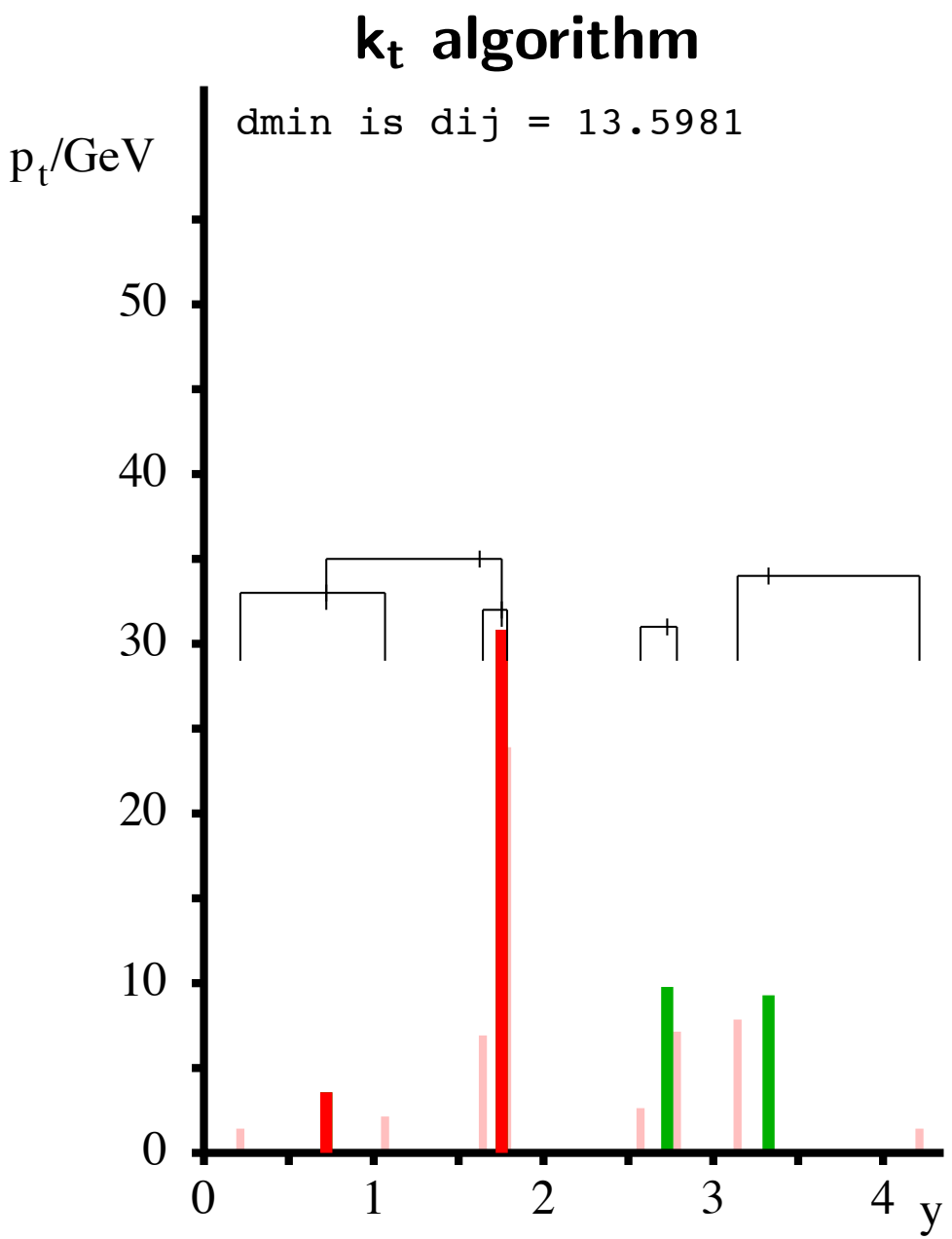


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Identifying jet substructure: try out k_t

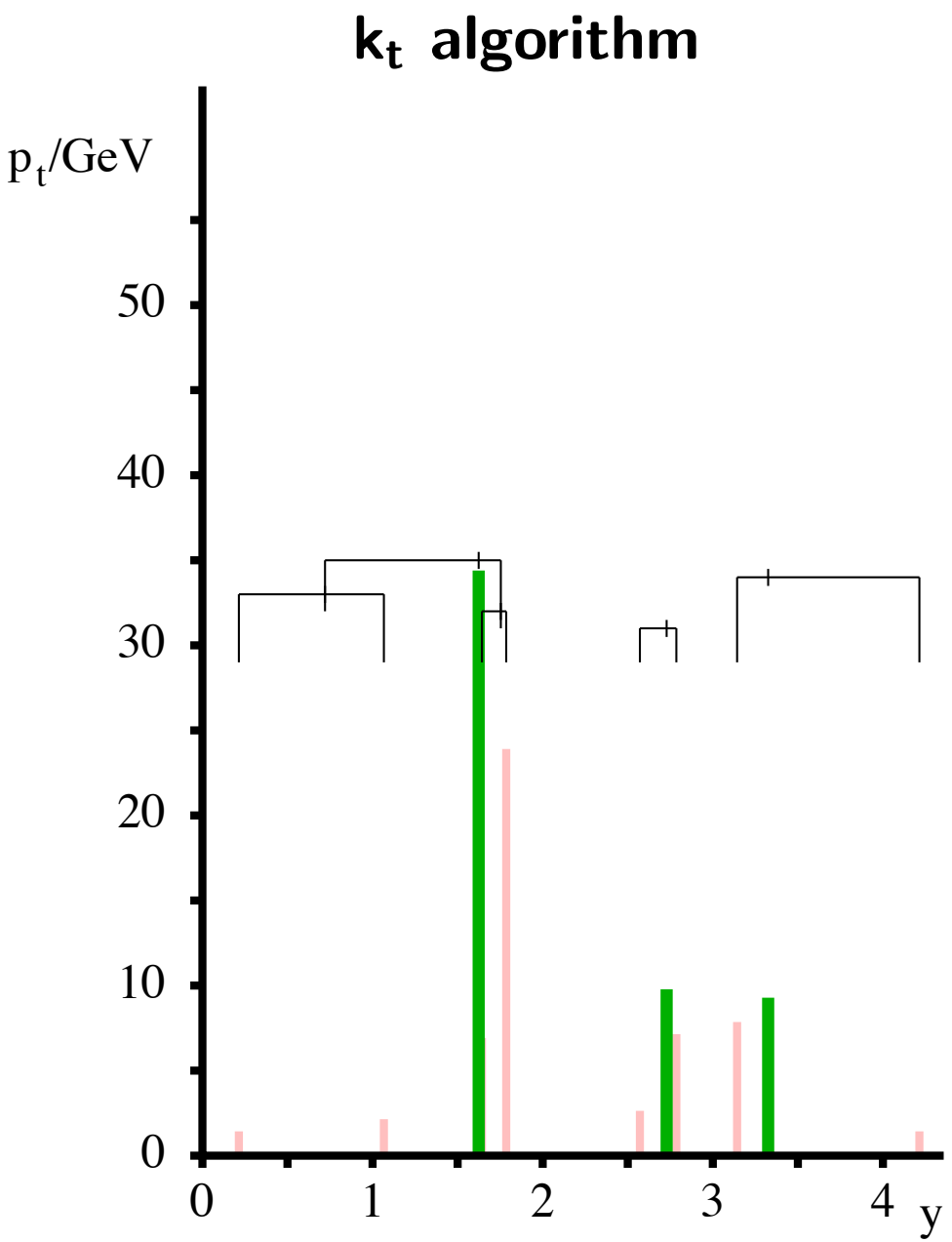


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Identifying jet substructure: try out k_t

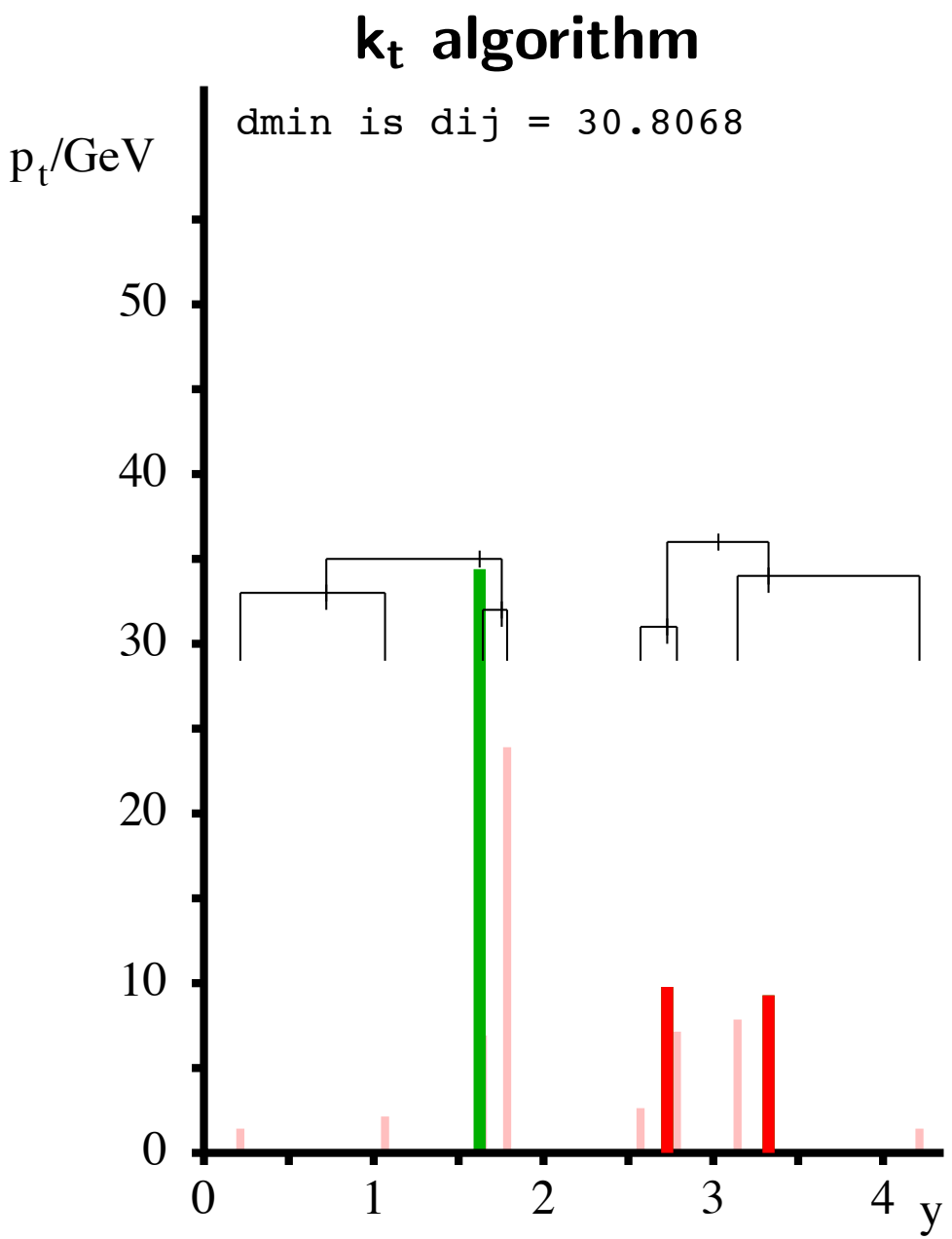


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Identifying jet substructure: try out k_t

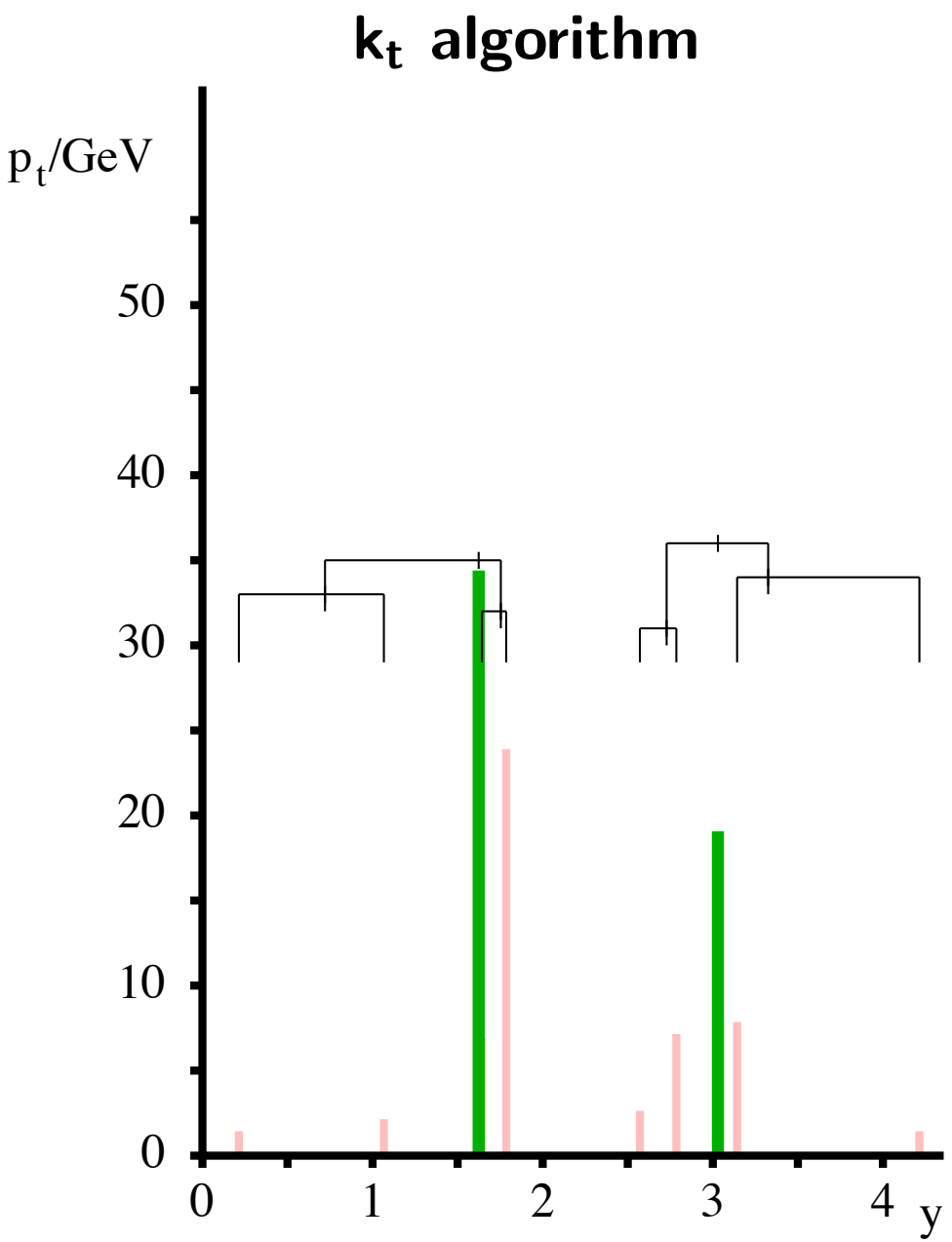


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Identifying jet substructure: try out k_t

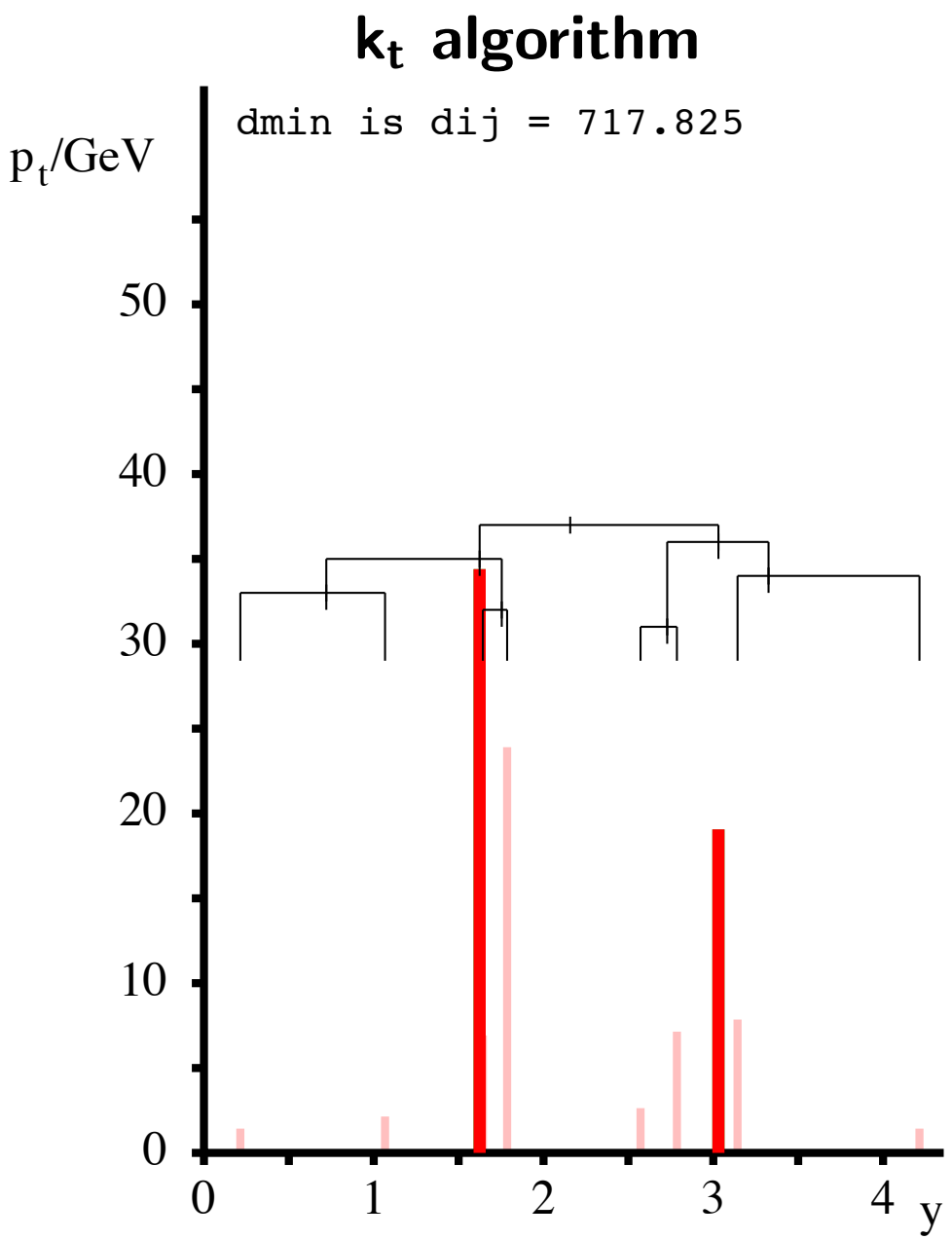


How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Identifying jet substructure: try out k_t



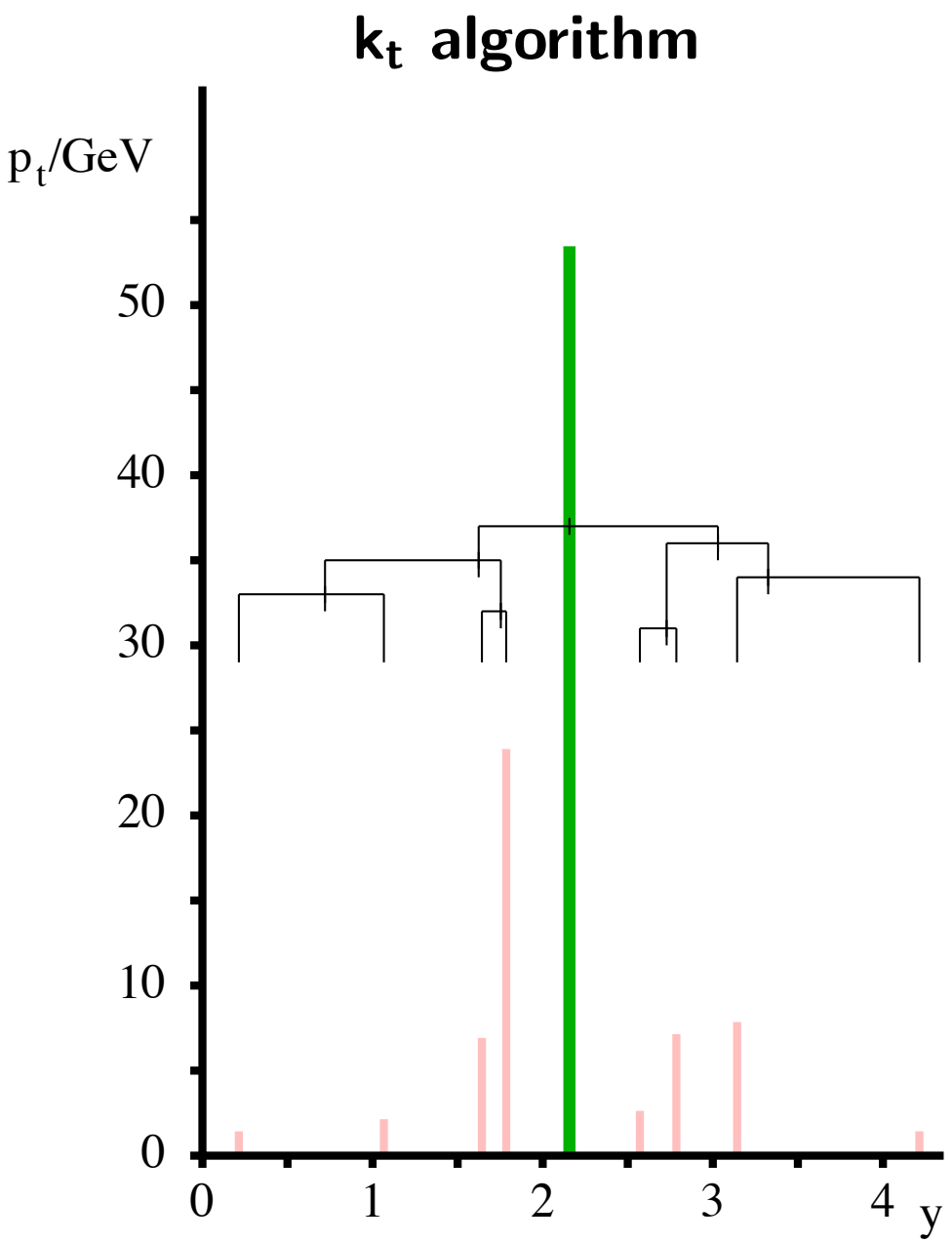
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Its last step is to merge two hard pieces. Easily undone to identify underlying kinematics

Identifying jet substructure: try out k_t



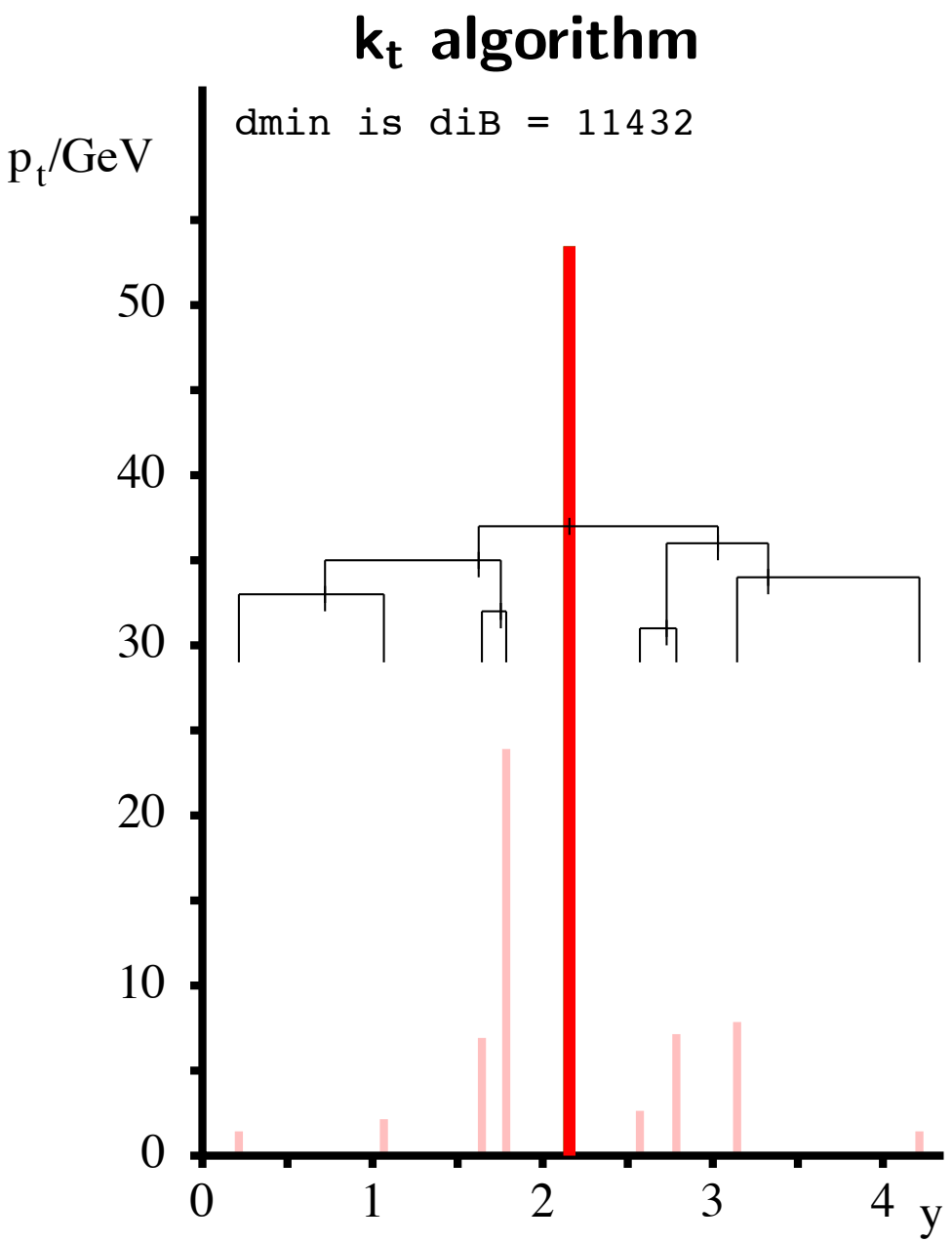
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Its last step is to merge two hard pieces. Easily undone to identify underlying kinematics

Identifying jet substructure: try out k_t



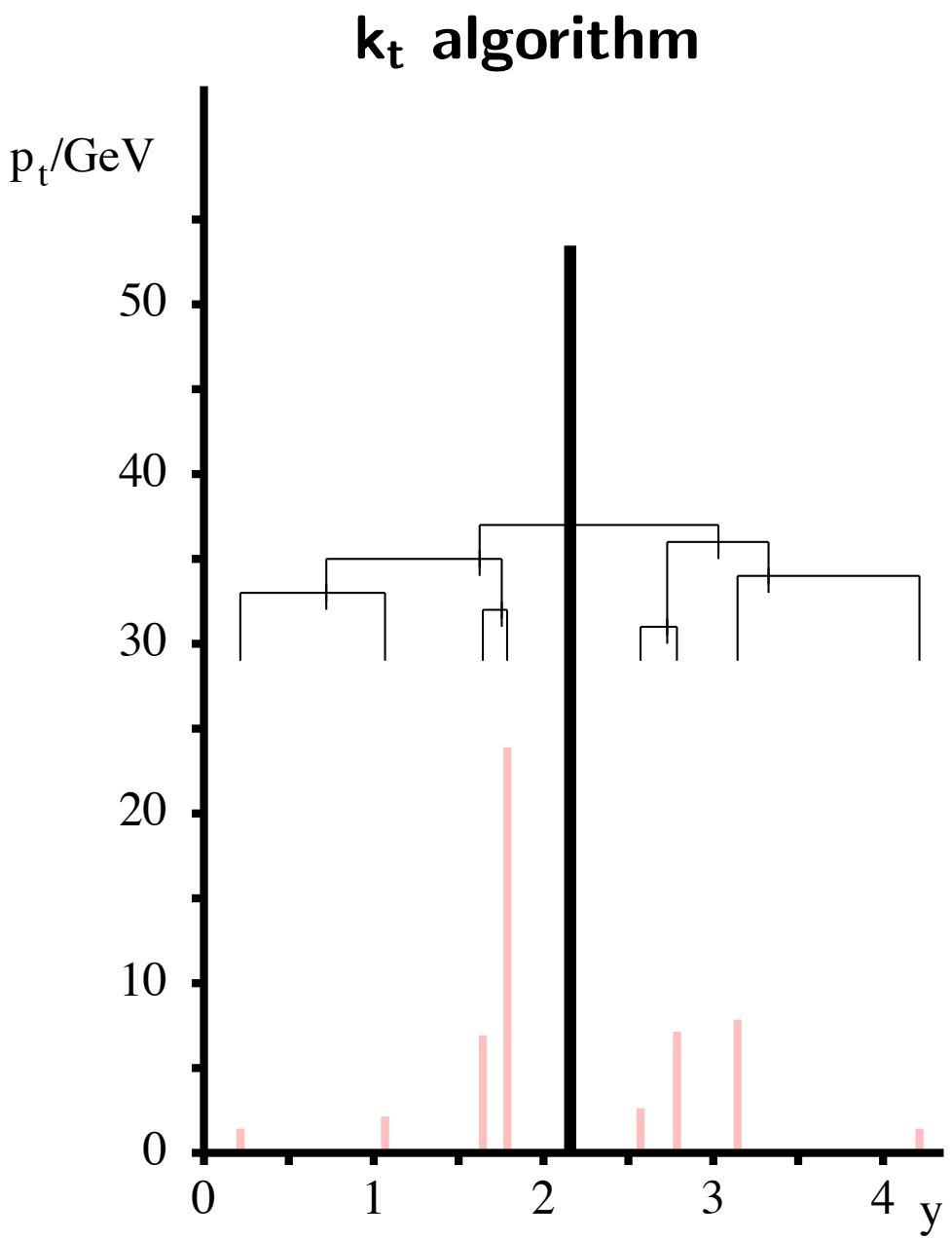
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Its last step is to merge two hard pieces. Easily undone to identify underlying kinematics

Identifying jet substructure: try out k_t



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

This is crucial for identifying the kinematic variables of the partons in the jet (e.g. z).

k_t clusters soft "junk" early on in the clustering

Its last step is to merge two hard pieces. Easily undone to identify underlying kinematics

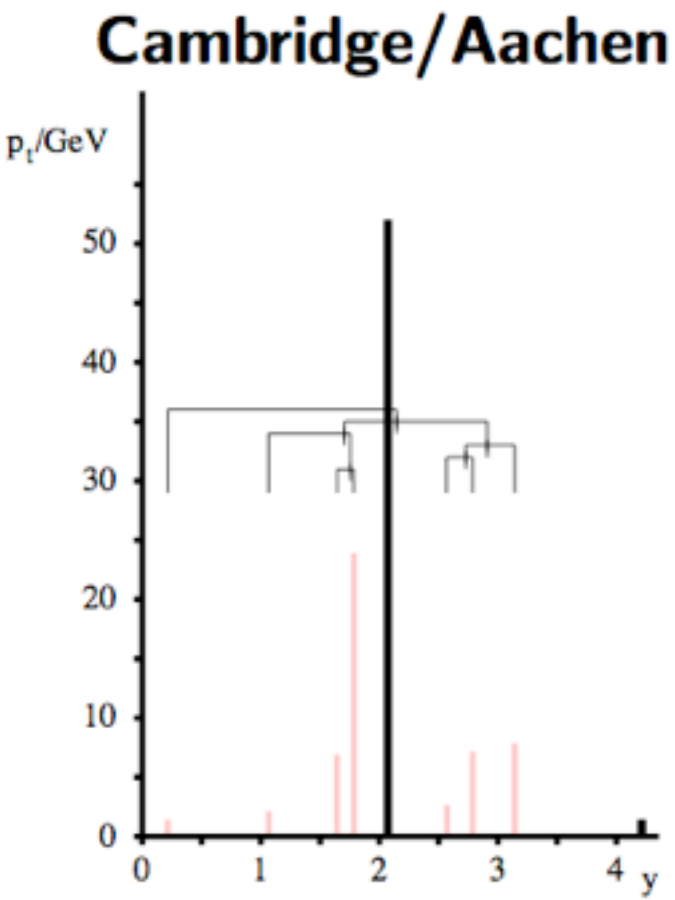
This meant it was the first algorithm to be used for jet substructure.

Seymour '93

Butterworth, Cox & Forshaw '02

Cambridge/Aachen

Hierarchical substructure

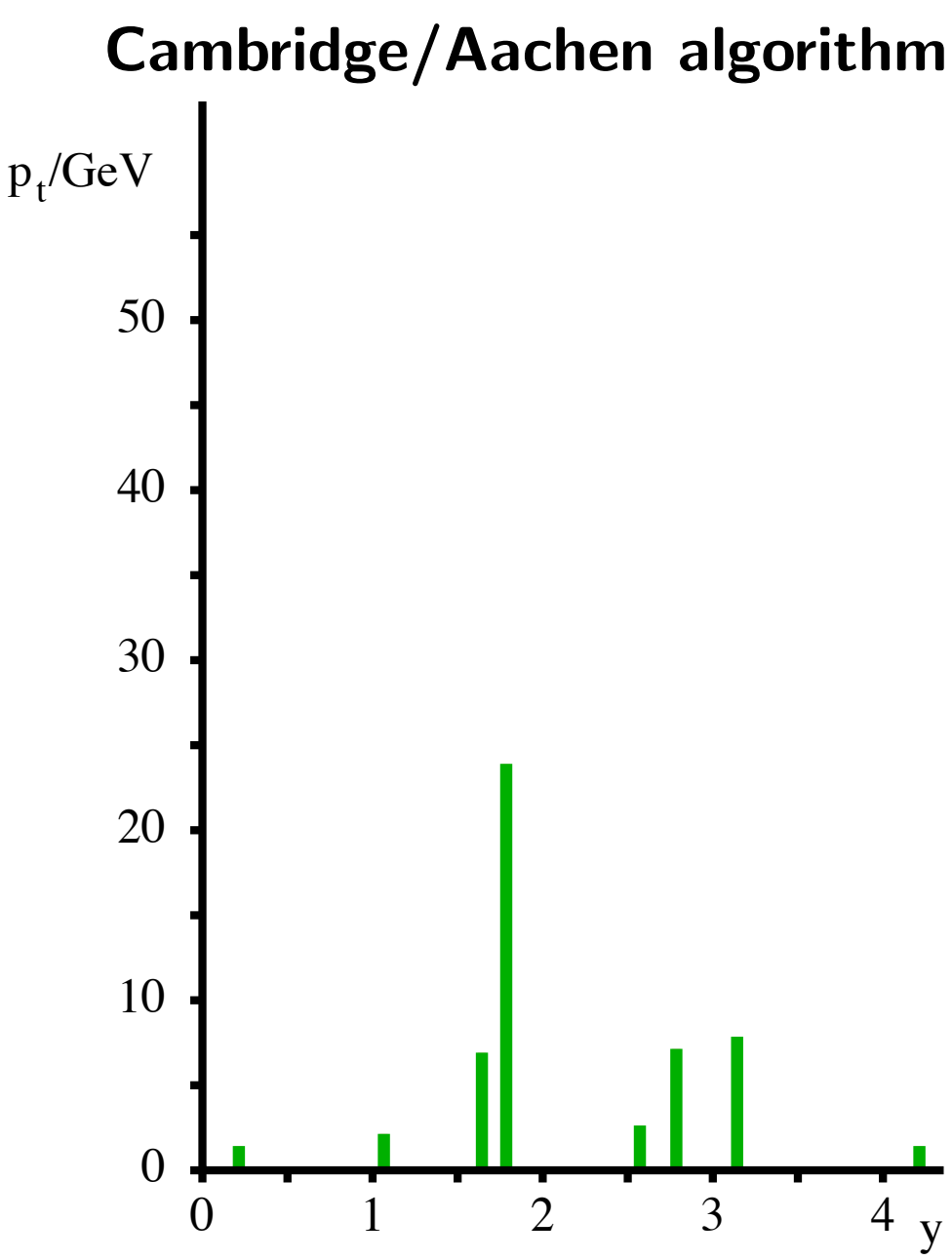


C/A distance measure

$$d_{ij} = \frac{\Delta y^2 + \Delta \phi^2}{R^2}$$

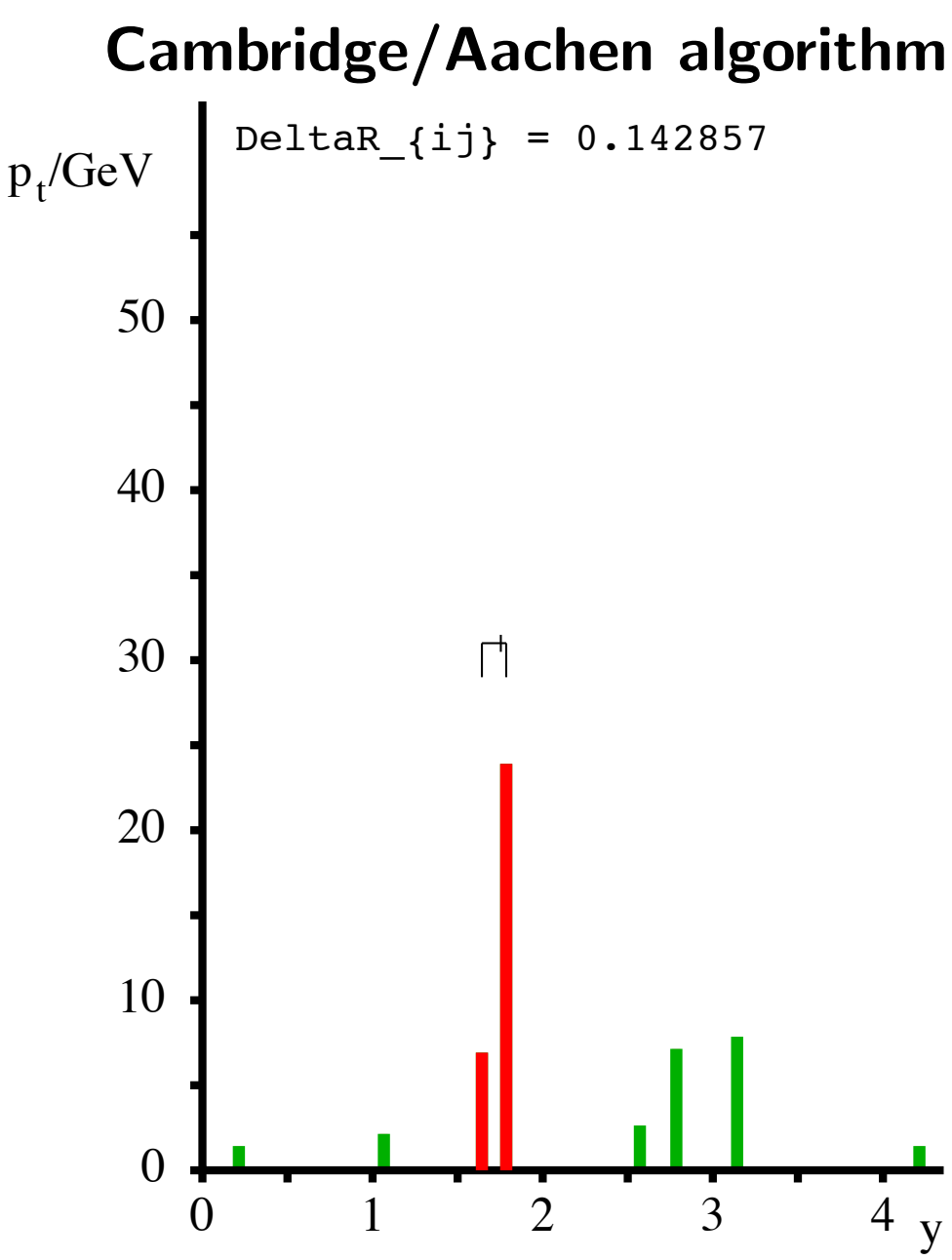
Cluster by merging the **closest** particles

Identifying jet substructure: Cam/Aachen



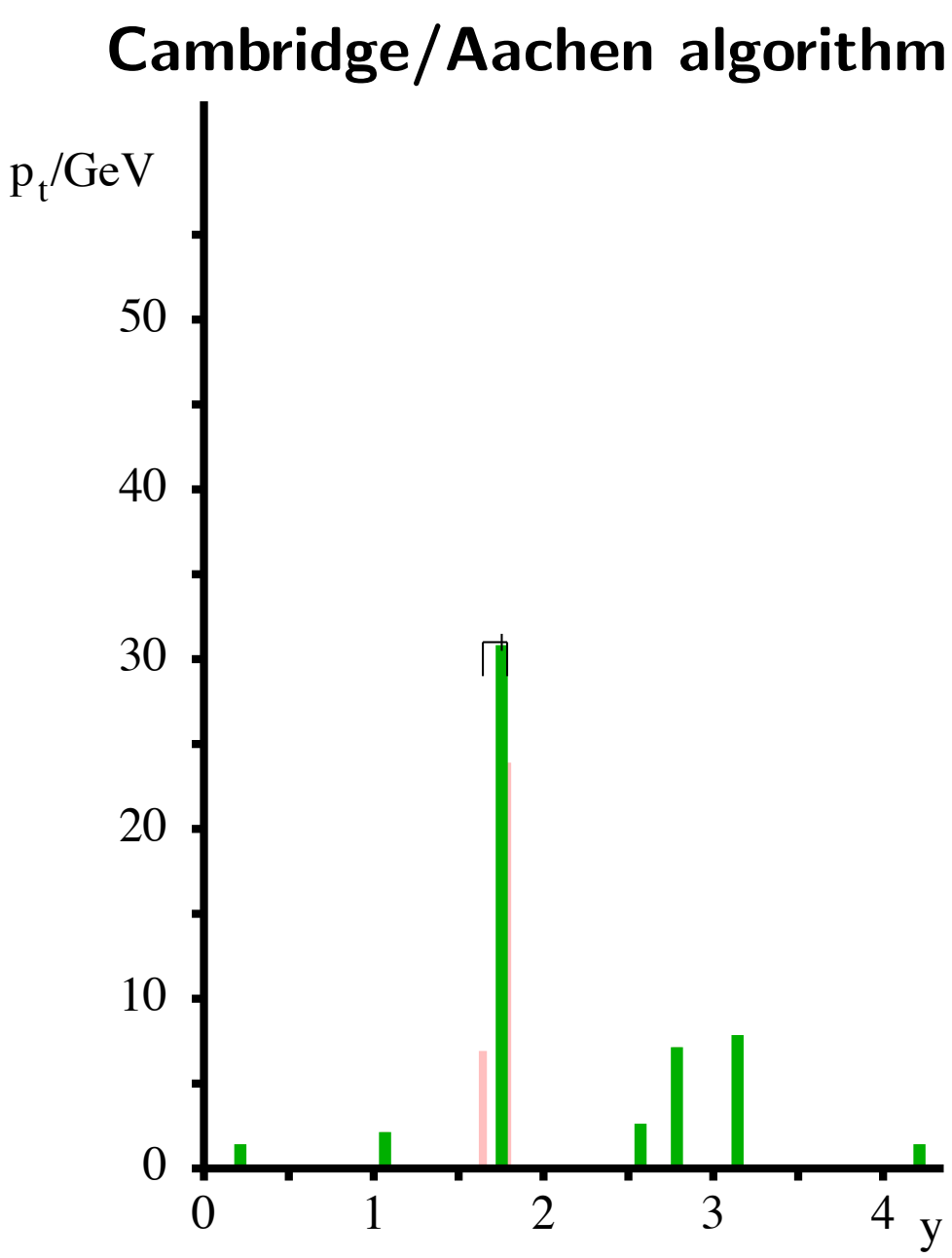
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

Identifying jet substructure: Cam/Aachen



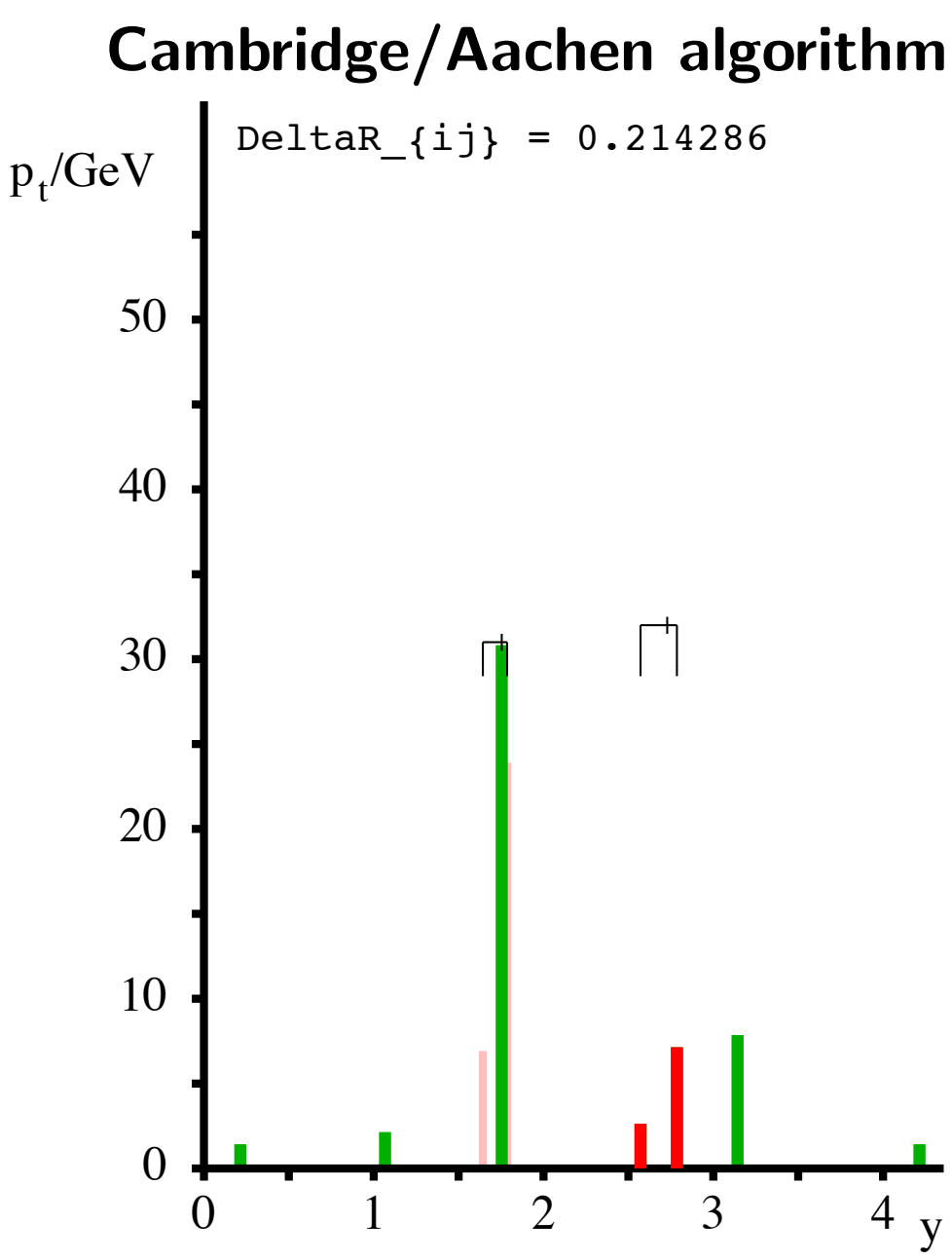
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

Identifying jet substructure: Cam/Aachen



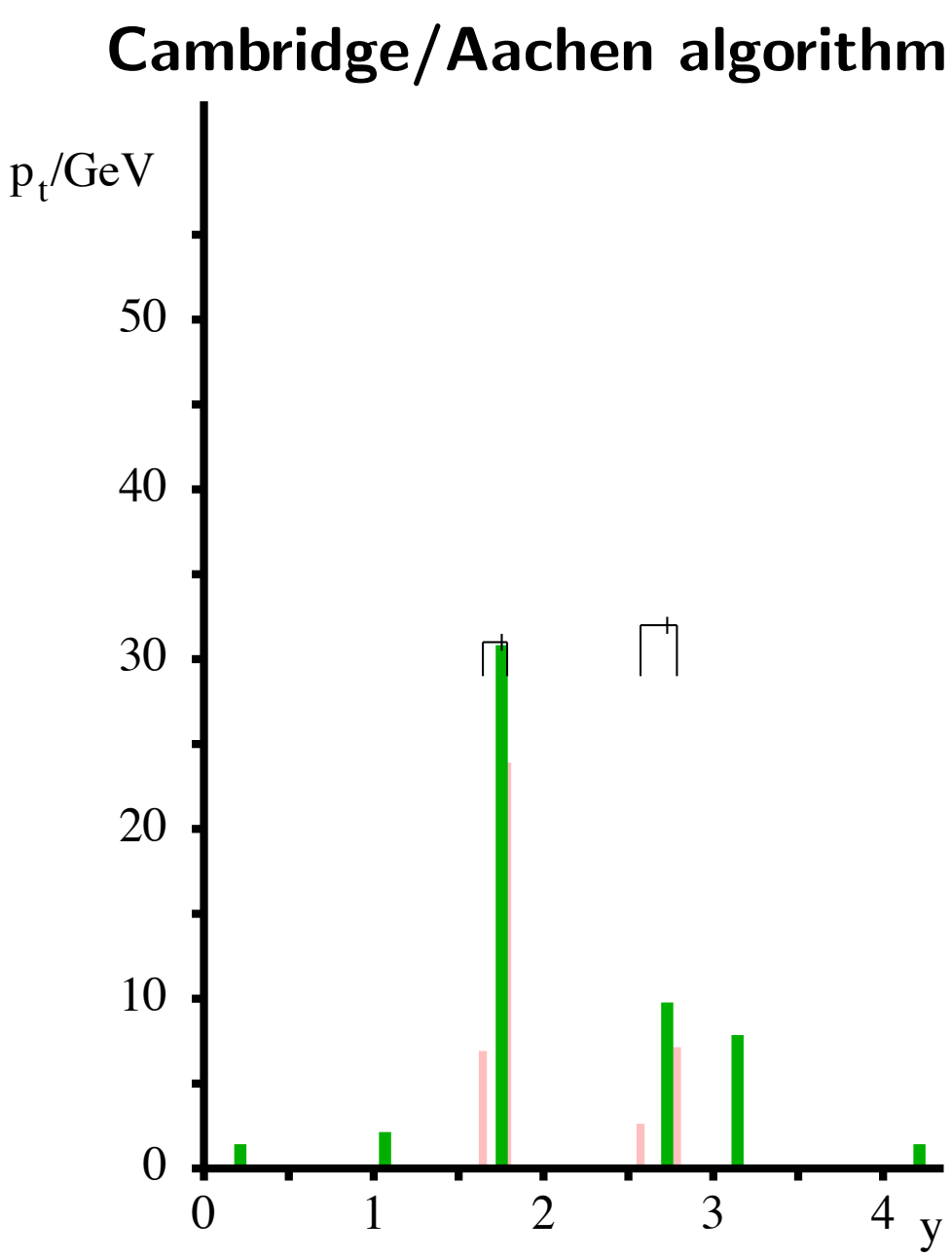
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

Identifying jet substructure: Cam/Aachen



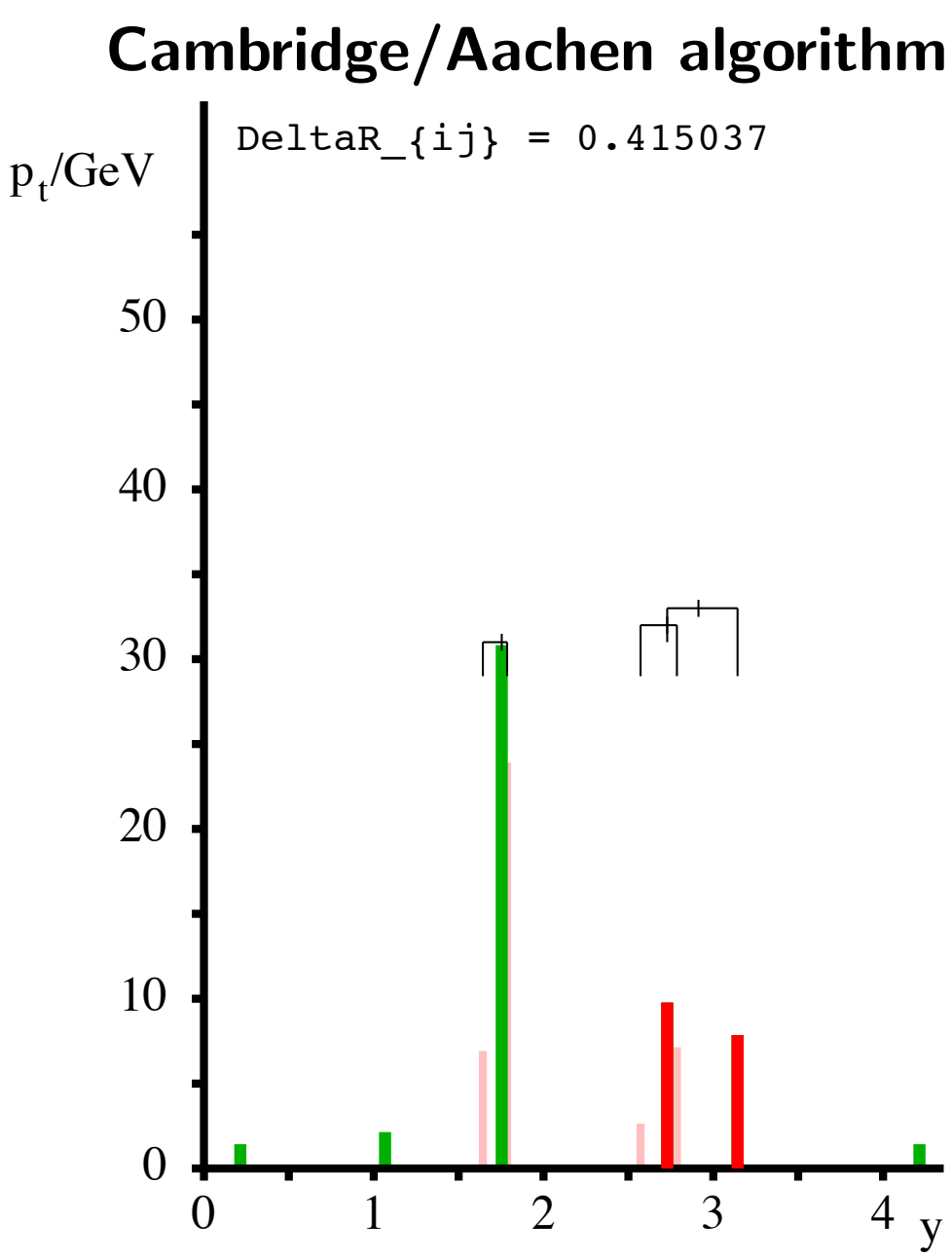
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

Identifying jet substructure: Cam/Aachen



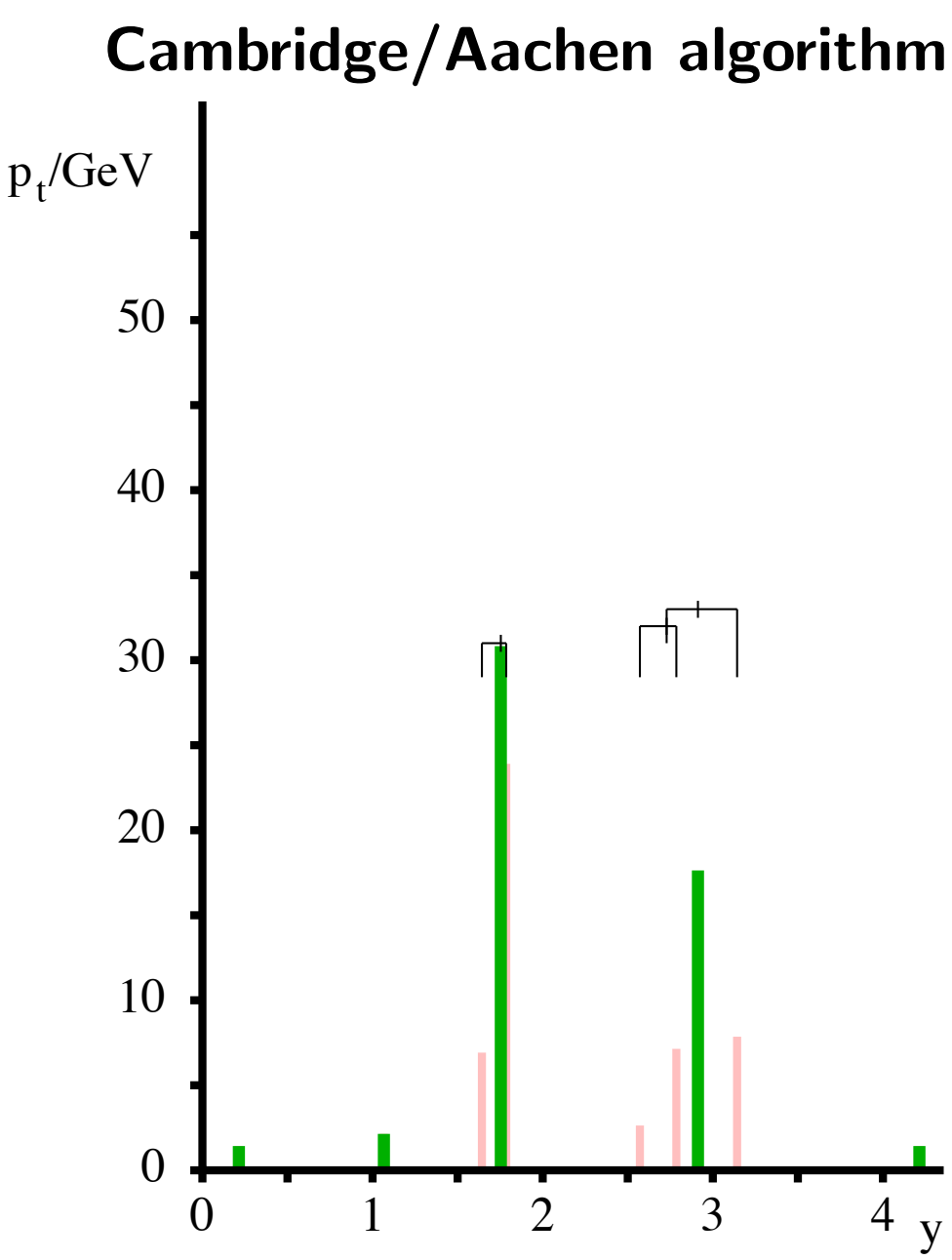
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

Identifying jet substructure: Cam/Aachen



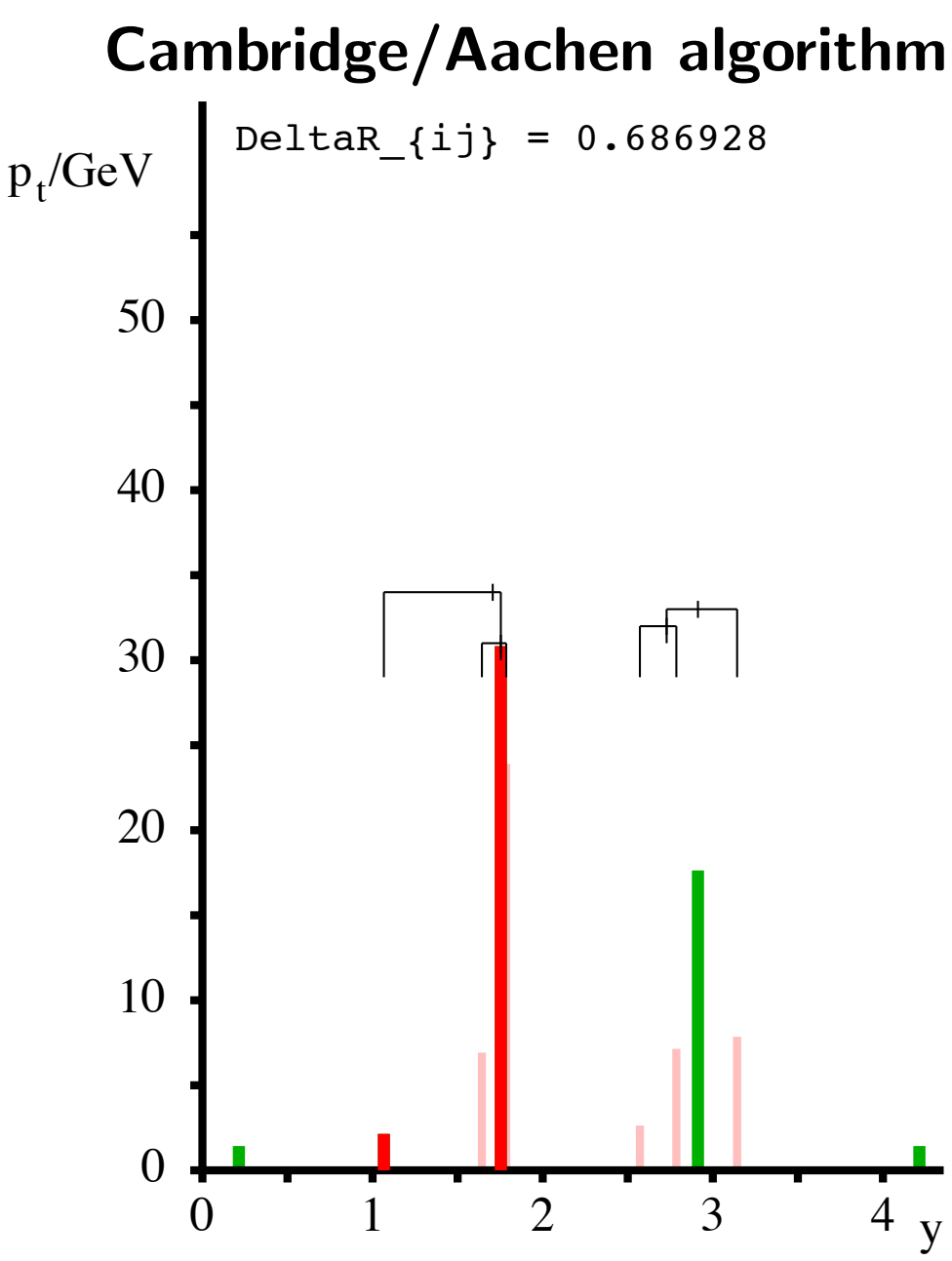
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

Identifying jet substructure: Cam/Aachen



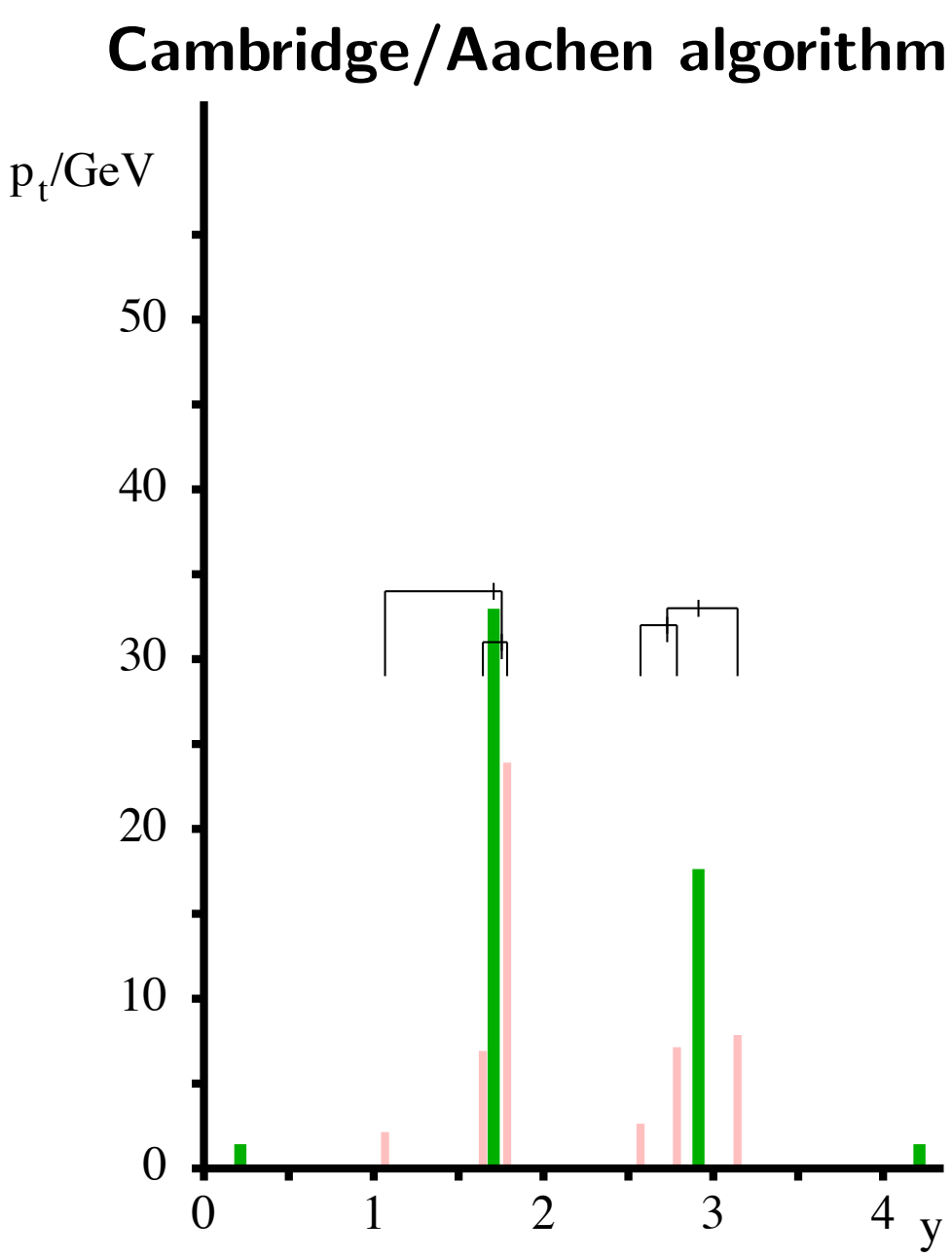
How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

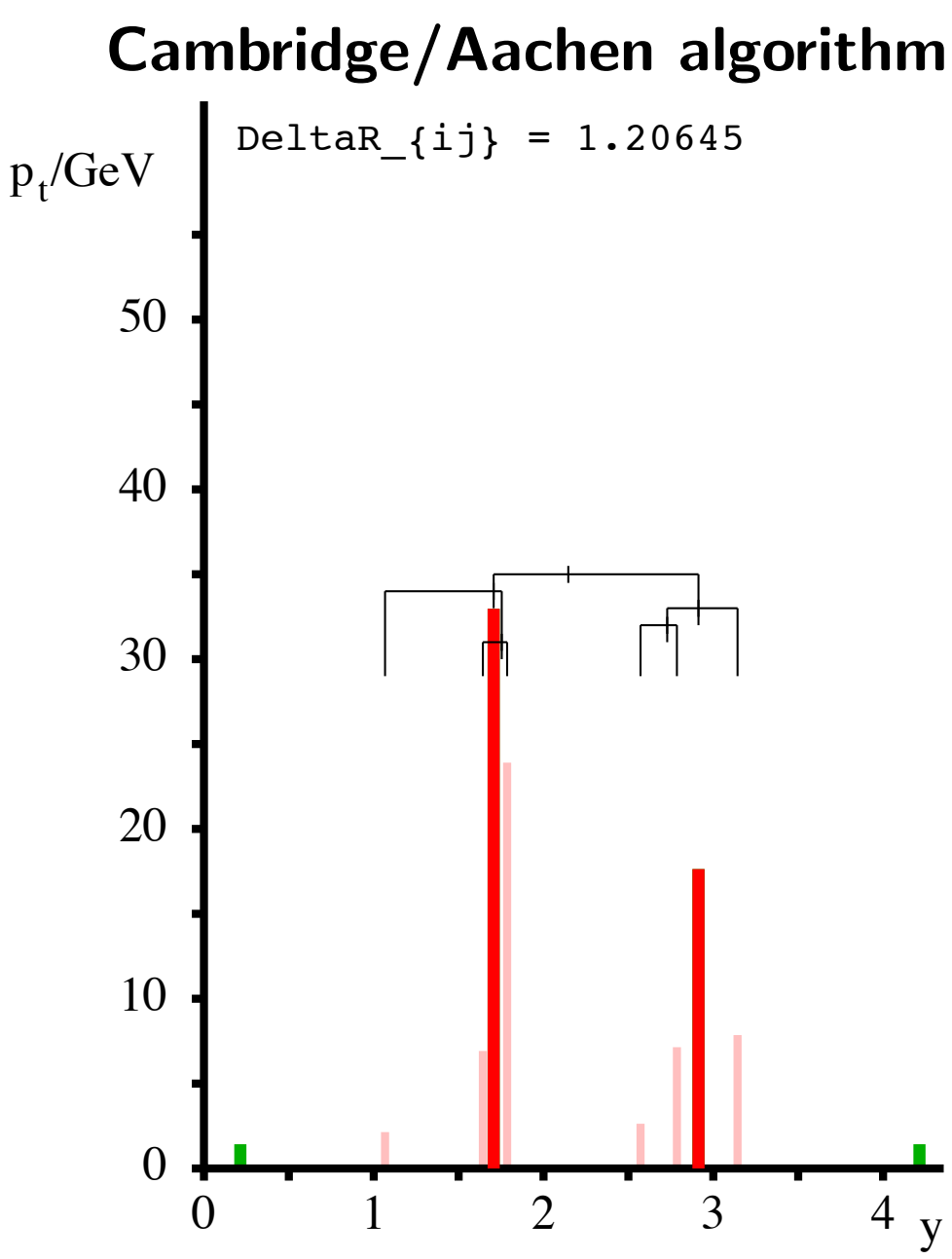
Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

C/A identifies two hard blobs with limited soft contamination

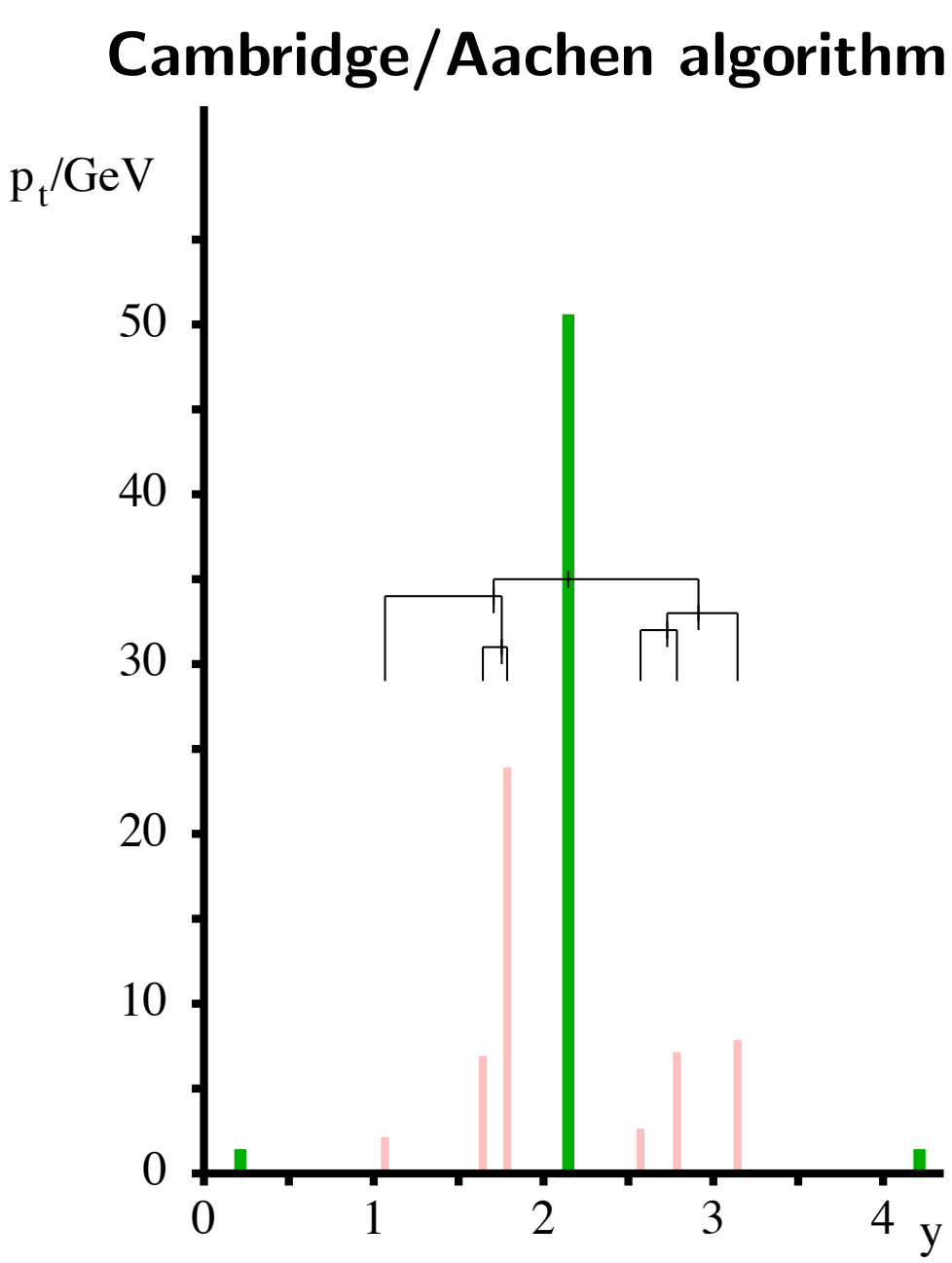
Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

C/A identifies two hard blobs with limited soft contamination, joins them

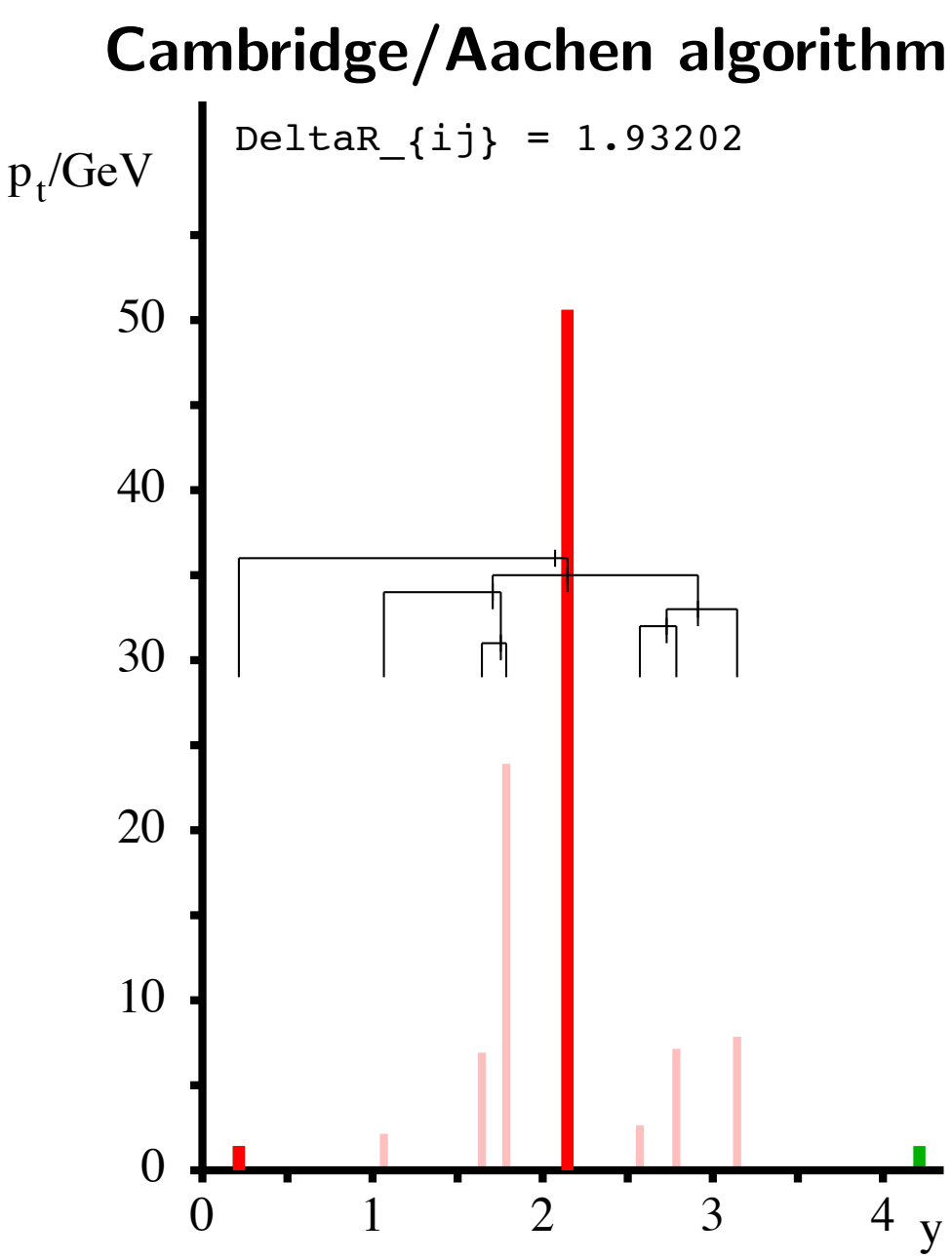
Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

C/A identifies two hard blobs with limited soft contamination, joins them

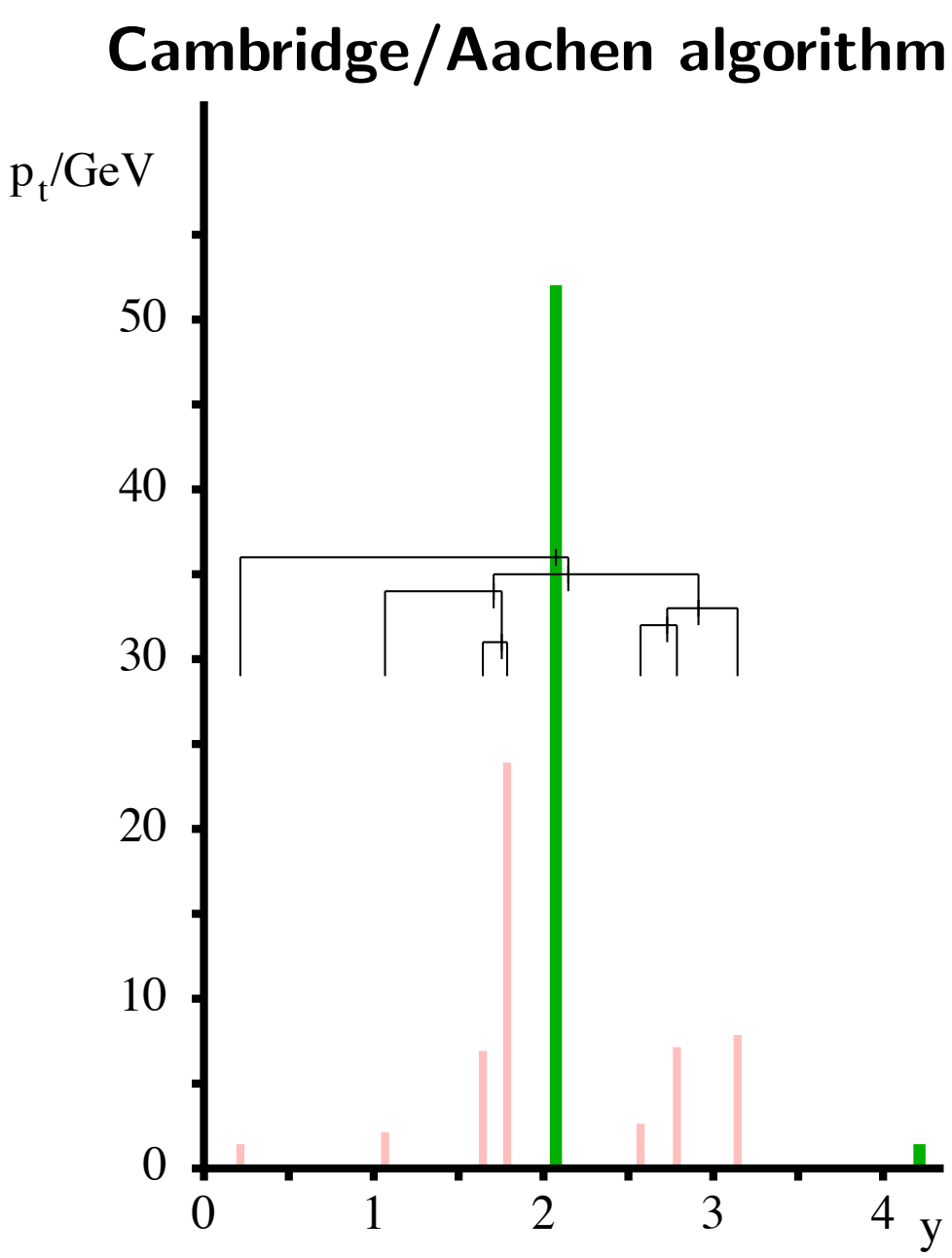
Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

C/A identifies two hard blobs with limited soft contamination, joins them, and then adds in remaining soft junk

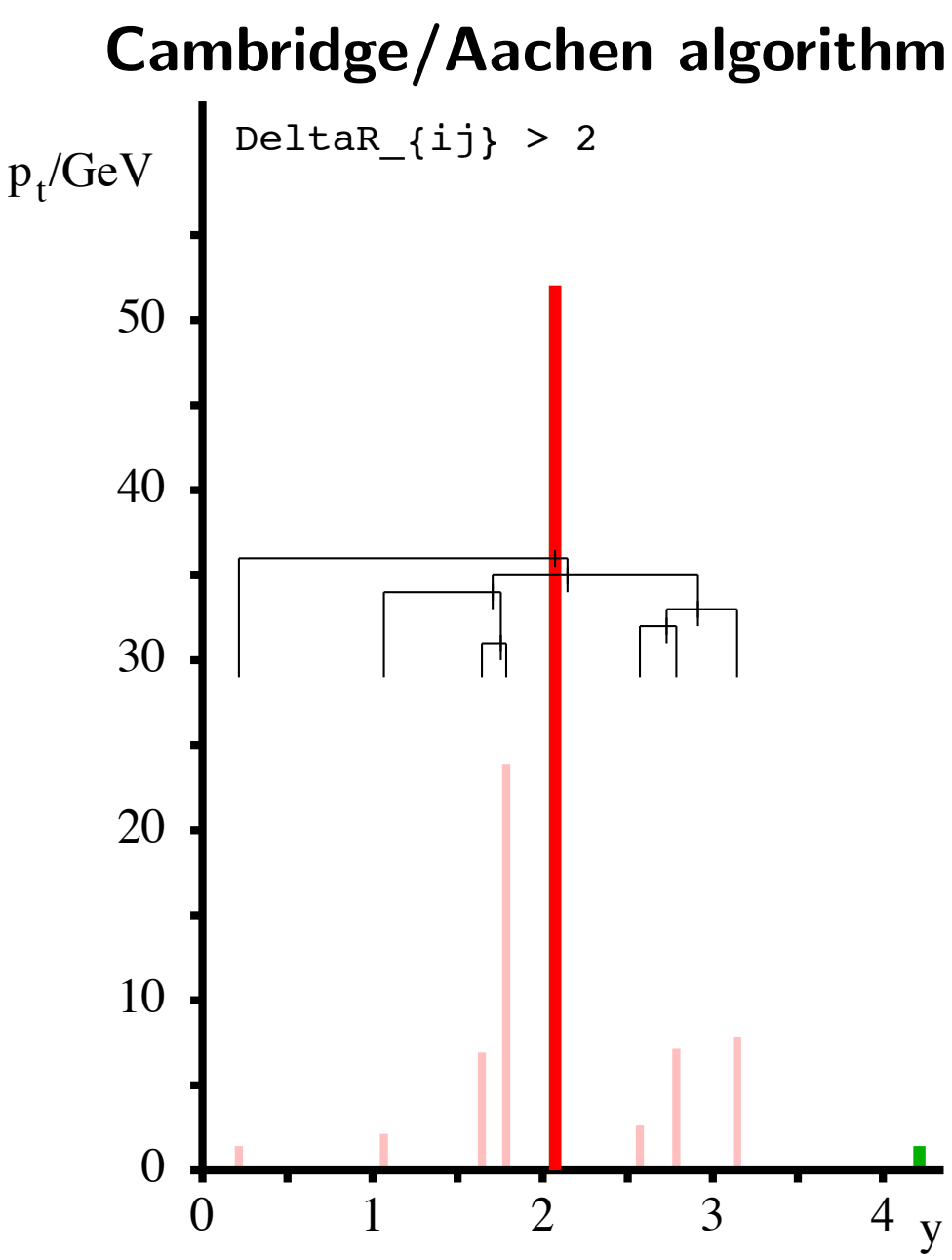
Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

C/A identifies two hard blobs with limited soft contamination, joins them, and then adds in remaining soft junk

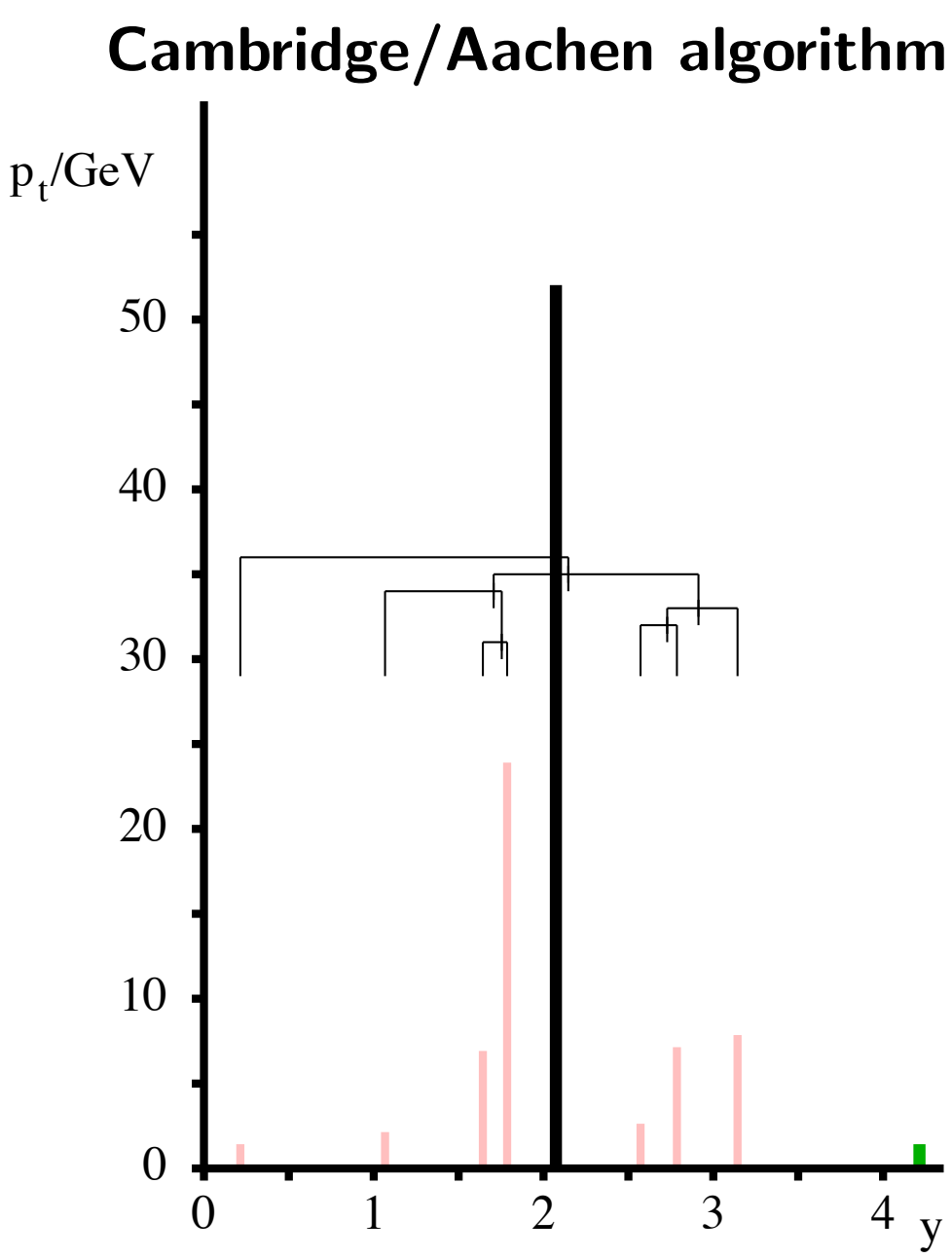
Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

C/A identifies two hard blobs with limited soft contamination, joins them, and then adds in remaining soft junk

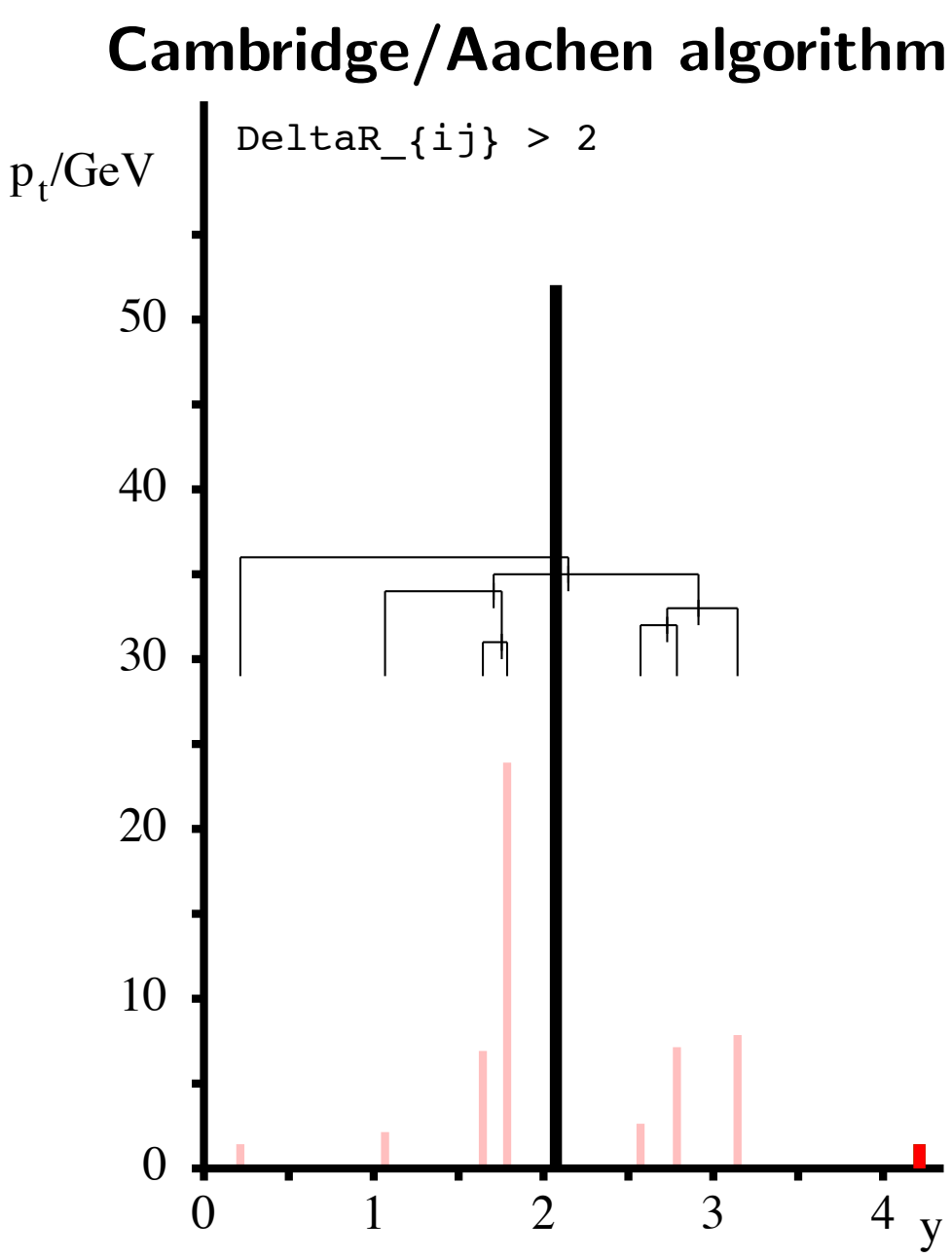
Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

C/A identifies two hard blobs with limited soft contamination, joins them, and then adds in remaining soft junk

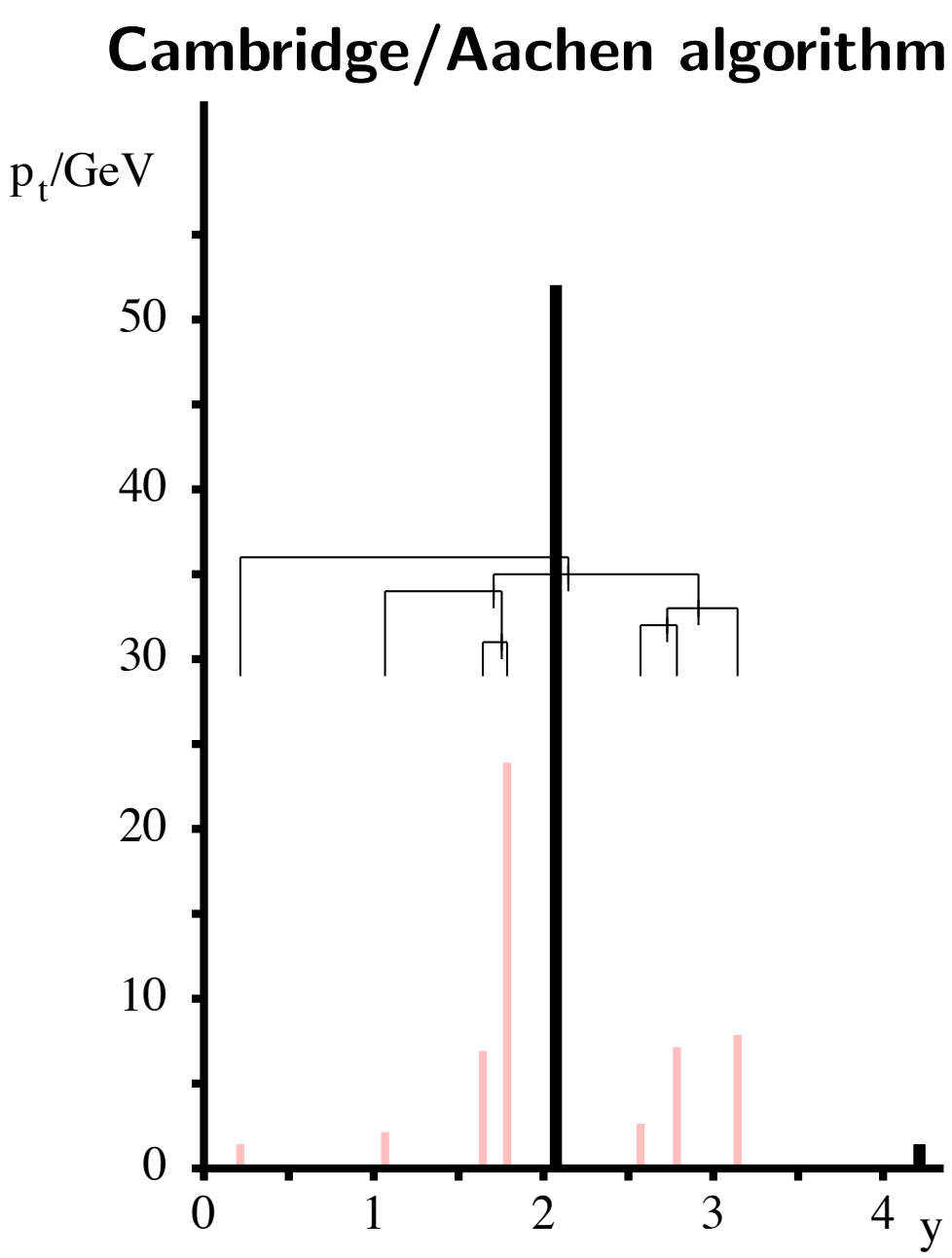
Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

C/A identifies two hard blobs with limited soft contamination, joins them, and then adds in remaining soft junk

Identifying jet substructure: Cam/Aachen



How well can an algorithm identify the "blobs" of energy inside a jet that come from different partons?

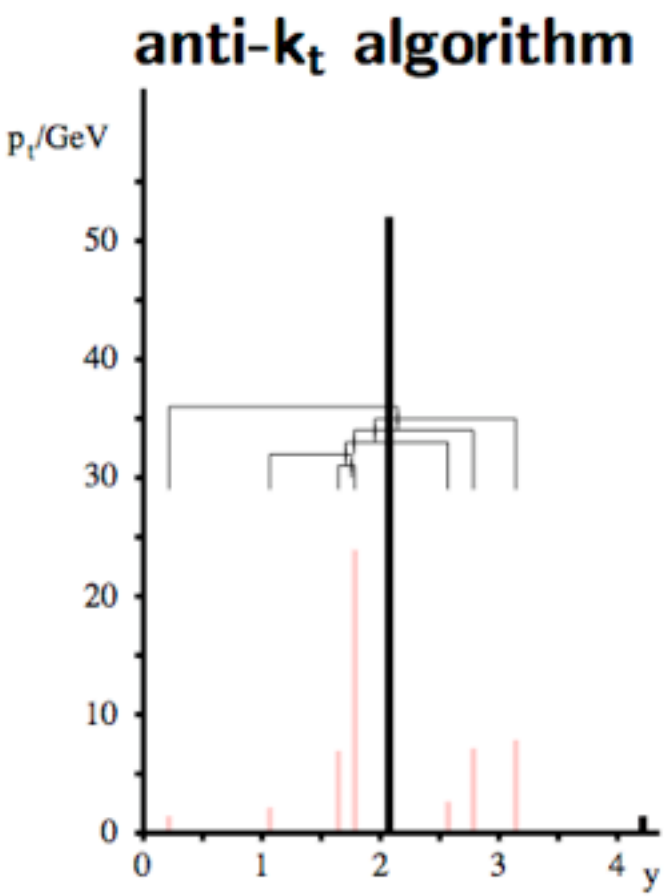
C/A identifies two hard blobs with limited soft contamination, joins them, and then adds in remaining soft junk

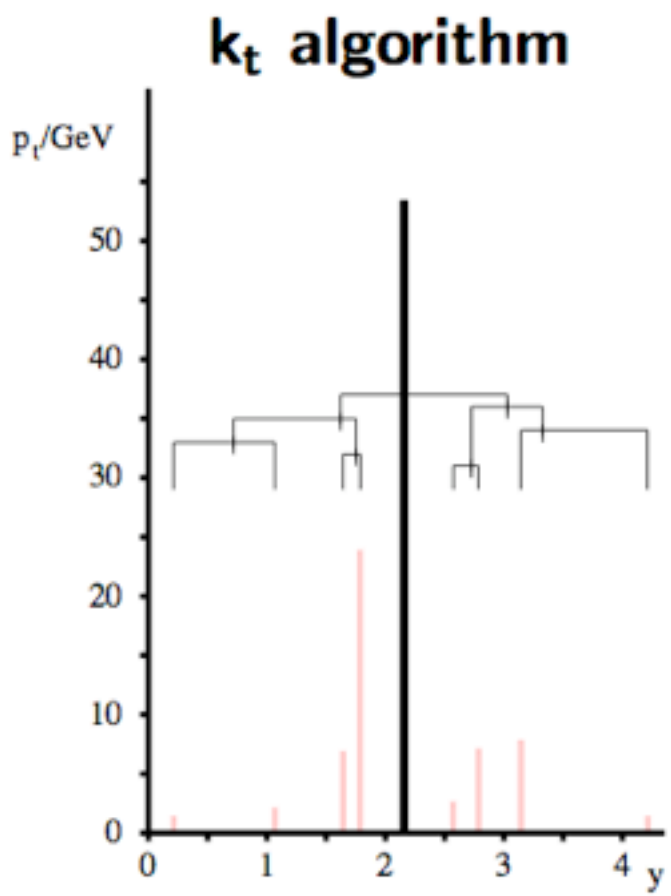
The interesting substructure is buried inside the clustering sequence — it's less contamined by soft junk, but needs to be pulled out with special techniques

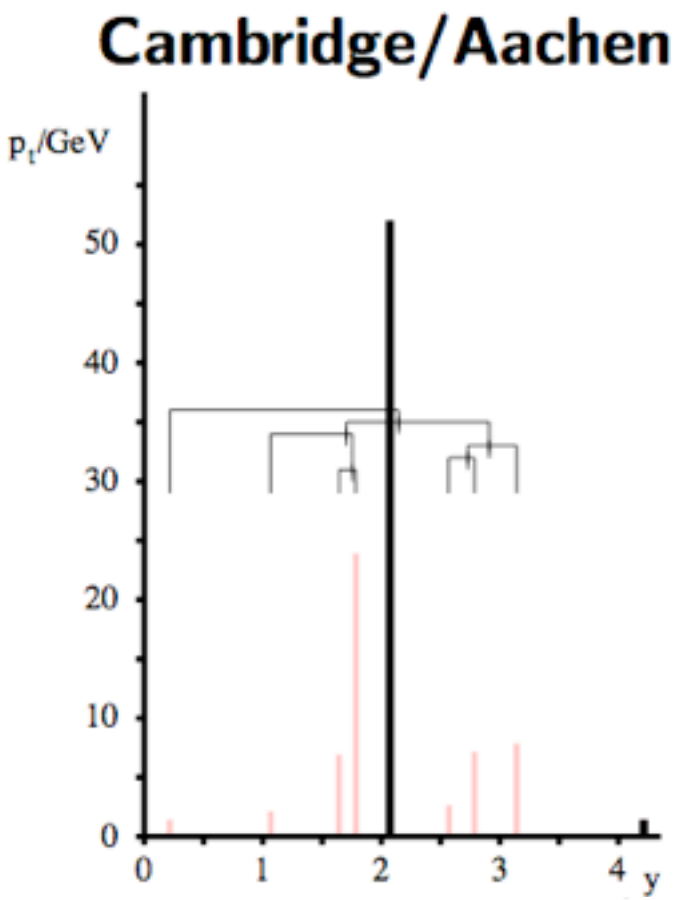
Butterworth, Davison, Rubin & GPS '08
Kaplan, Schwartz, Reherman & Tweedie '08
Butterworth, Ellis, Rubin & GPS '09
Ellis, Vermilion & Walsh '09

Slide by
Gavin Salam

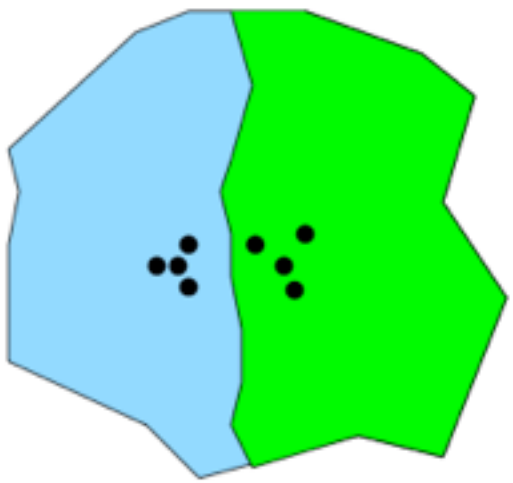
Hierarchical substructure

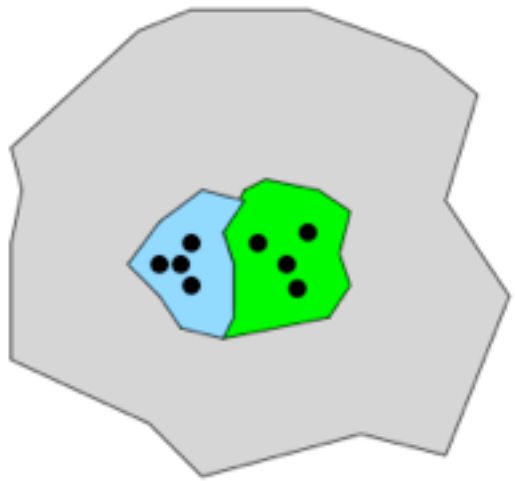






Undo the last clustering step(s) →





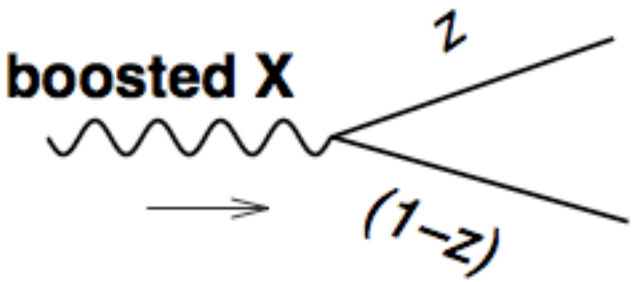
The IRC safe algorithms

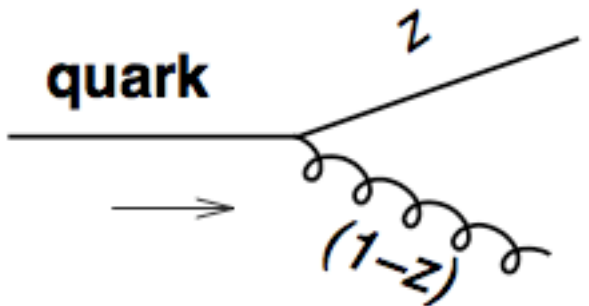
	Speed	Regularity	UE contamination	Backreaction	Hierarchical substructure
k_t	☺☺☺	☂	☂☂	☁☁	☺☺
Cambridge /Aachen	☺☺☺	☂	☂	☁☁	☺☺☺
anti-k_t	☺☺☺	☺☺	☁/☺	☺☺	✘
SISCone	☺	☁	☺☺	☁	✘

Array of tools with different characteristics.
Pick the right one for the job

QCD v. heavy decay

A possible approach for reducing the QCD background is to identify the two prongs of the heavy particle decay, and put a cut on their momentum fraction





Signal:

$$P(z) \sim 1$$

Will split mainly **symmetrically**

Background:

$$P(z) \sim \frac{1+z^2}{1-z} \qquad P(z) \sim \frac{1+(1-z)^2}{z}$$

Will split mainly **asymmetrically**

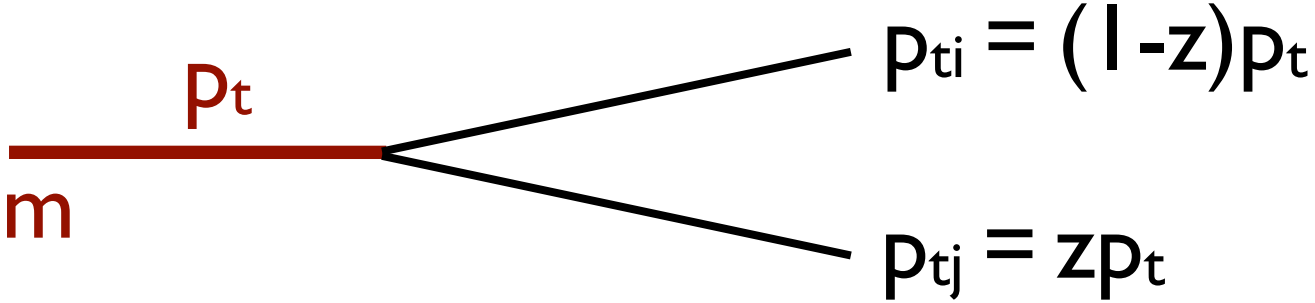
Potential tagger: asymmetric splitting

Possibly implemented via a cut on

$$y = min(p_{ti}^2, p_{tj}^2)\frac{\Delta R_{ij}^2}{m^2} \simeq \frac{min(p_{ti}, p_{tj})}{max(p_{ti}, p_{tj})}$$

Splittings and distances

Quasi-collinear splitting ($p_{tj} < p_{ti}$)



Invariant mass: $$m^2 \simeq p_{ti} p_{tj} \Delta R_{ij}^2 = (1-z) z p_t^2 \Delta R_{ij}^2$$

k_t distance: $$d_{ij} \overset{(p_{tj} < p_{ti})}{=} z^2 p_t^2 \Delta R_{ij}^2 \simeq \frac{z}{1-z} m^2$$

For a given mass, the **background** will have smaller distance d_{ij} than the signal, i.e. it will tend to **cluster earlier** in the k_t algorithm

Potential tagger: last clustering in k_t algorithm

This is where the hierarchy of the k_t algorithm becomes relevant.
QCD radiation is clustered first, and only at the end the symmetric, large-angle splittings due to decays are reclustered

Alternative algorithms

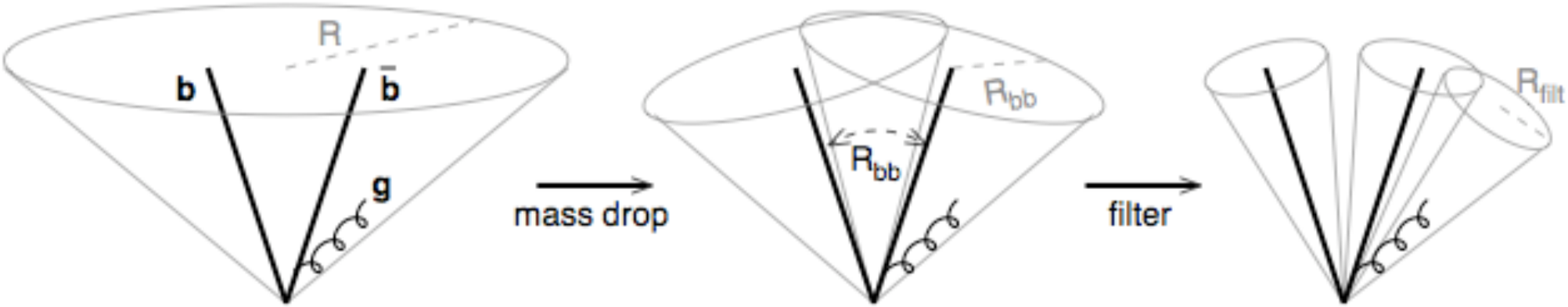
- Suppose that for some reasons (which will become clearer later) one does not with to use the k_t algorithm
  - *One must then find a way to determine what the **relevant splitting** (i.e. the one due to the decay, not to QCD radiation) is.*

A possible approach is to use a Mass-Drop requirement:
the clustering is **progressively undone**, and a splitting is the relevant one if both subjects are much less massive than their combination

$pp \to ZH \to \nu\bar{\nu}b\bar{b}$

The BDRS tagger/groomer

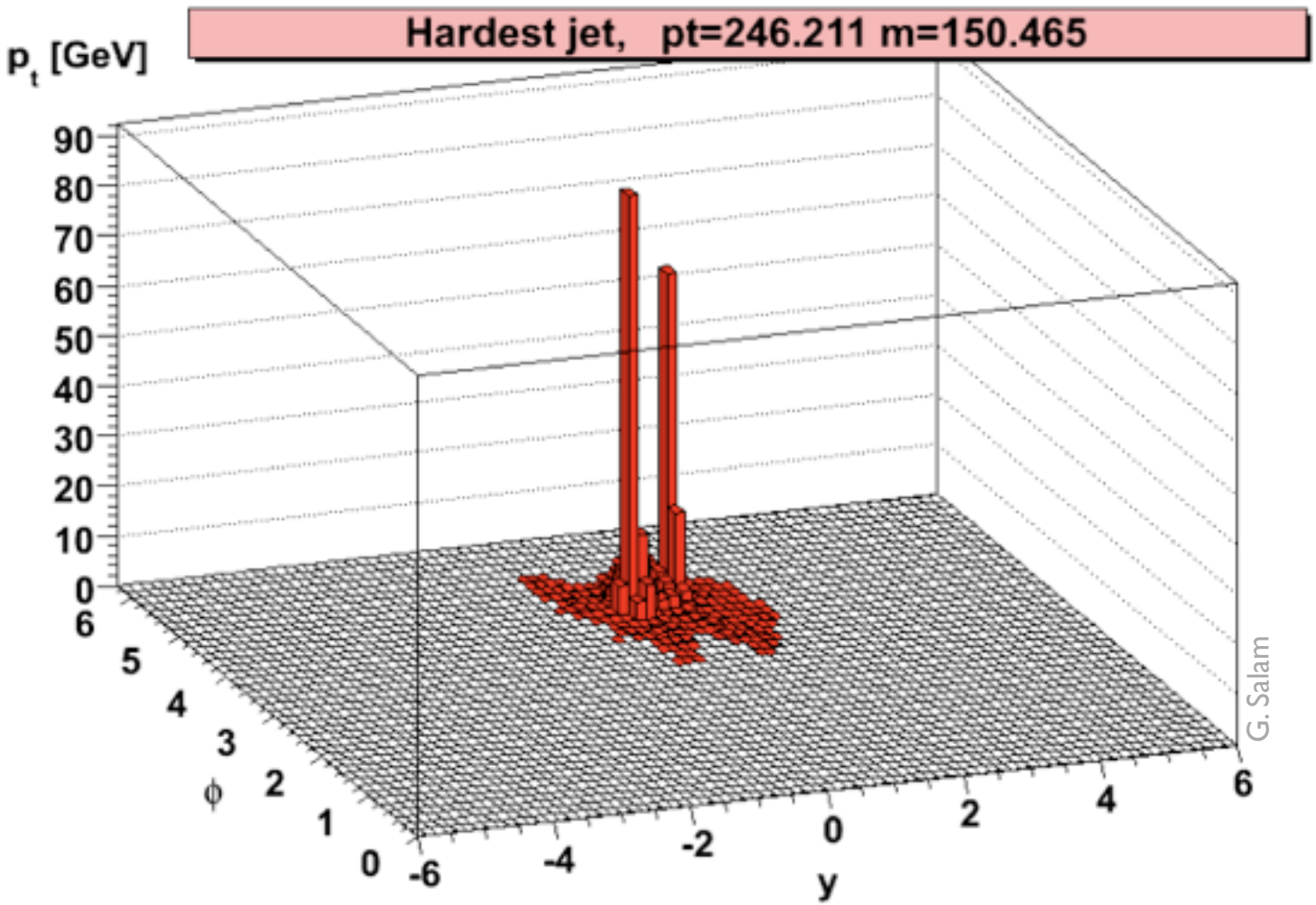
Butterworth, Davison, Rubin, Salam, 2008



- A two-prong tagger/groomer for boosted Higgs, which
  - Uses the **Cambridge/Aachen** algorithm (because it's 'physical')
  - Employs a **Mass-Drop** condition, as well as an **asymmetry cut** to find the **relevant splitting** (i.e. '**tag**' the heavy particle)
  - Includes a post-processing step, using 'filtering' (introduced in the same paper) to clean as much as possible the resulting jets of UE contamination ('**grooming**')

BDRS: tagging

$pp \to ZH \to \nu\bar{\nu}b\bar{b}$



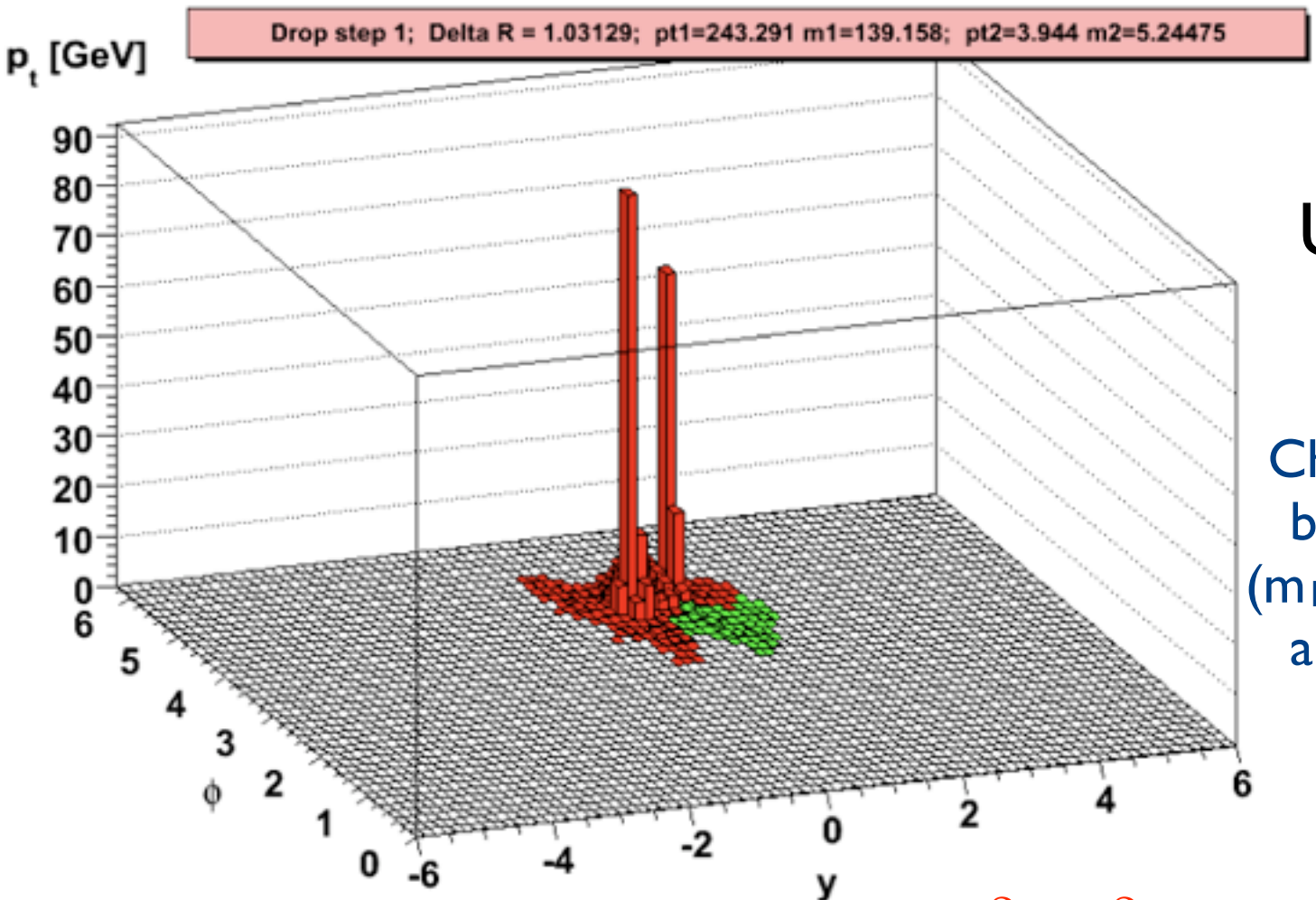
Start with the hardest jet

Use C/A with large R=1.2

m_j = 150 GeV

BDRS: tagging

pp →ZH → ννbb



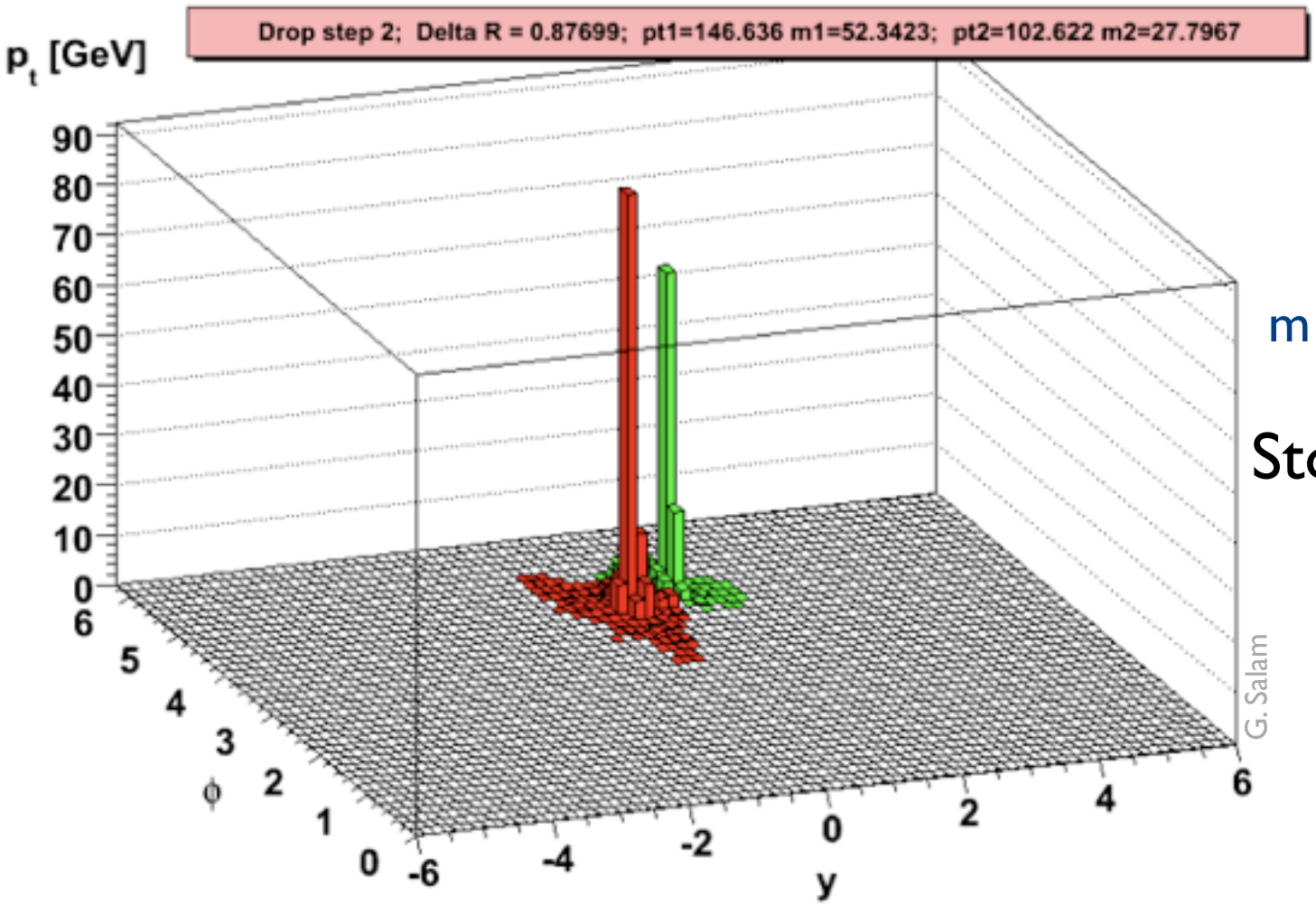
Undo last step of clustering

Check how the mass splits between the two subjets (m_1 = 139 GeV, m_2 = 5 GeV) and how asymmetric the splitting is

If $\frac{max(m_1, m_2)}{m_j} > \mu$ or $\frac{min(p_{t1}^2, p_{t2}^2)}{m_j^2} \Delta R_{12}^2 < y_{cut}$ repeat

BDRS: tagging

pp →ZH → ννbb



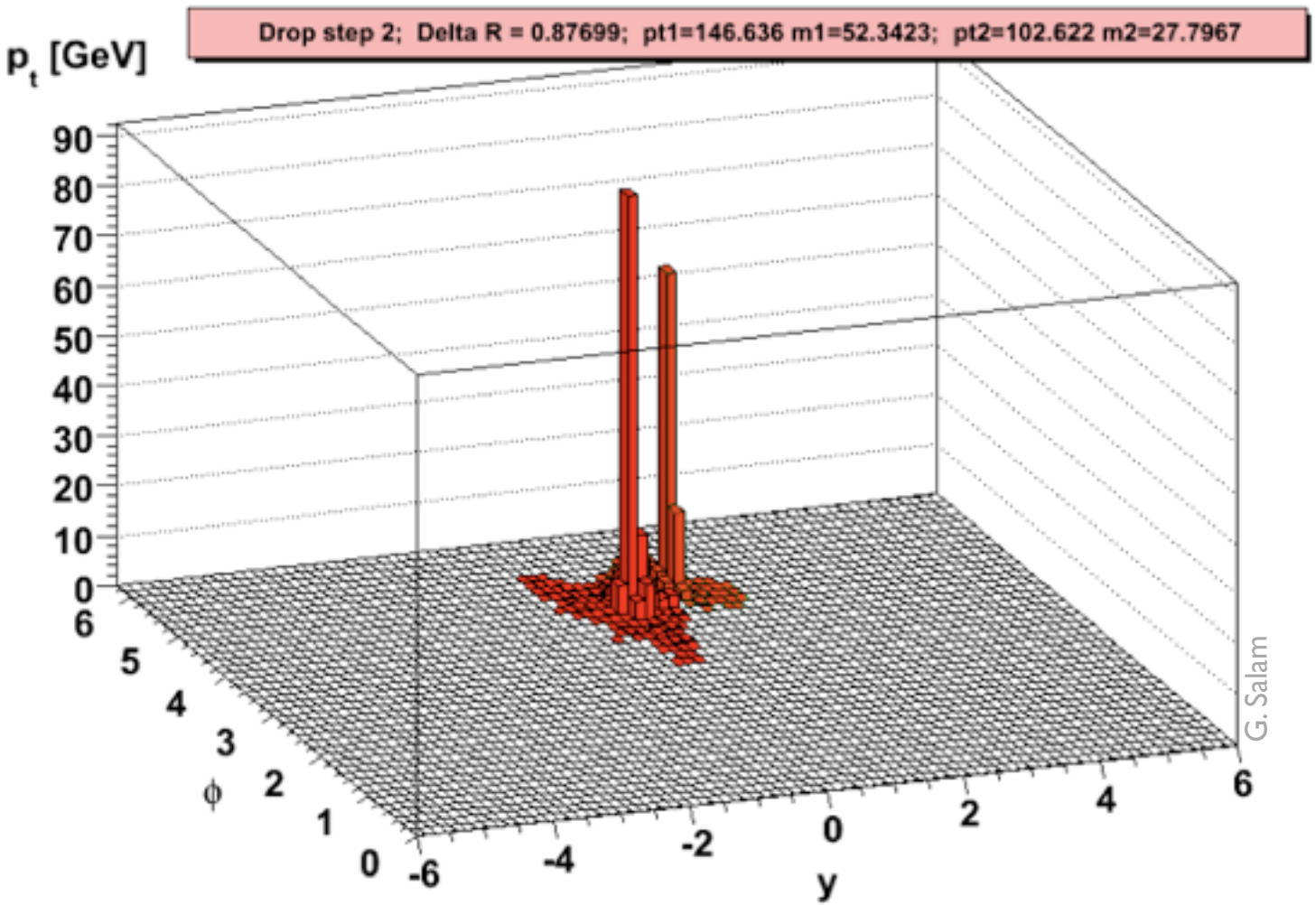
m_1 = 52 GeV, m_2 = 28 GeV

Stop when a large mass drop is observed
(and **recombine** these two jets)

[NB. Parameters used $\mu = 0.67$ and $y_{cut} = 0.09$]

BDRS: filtering

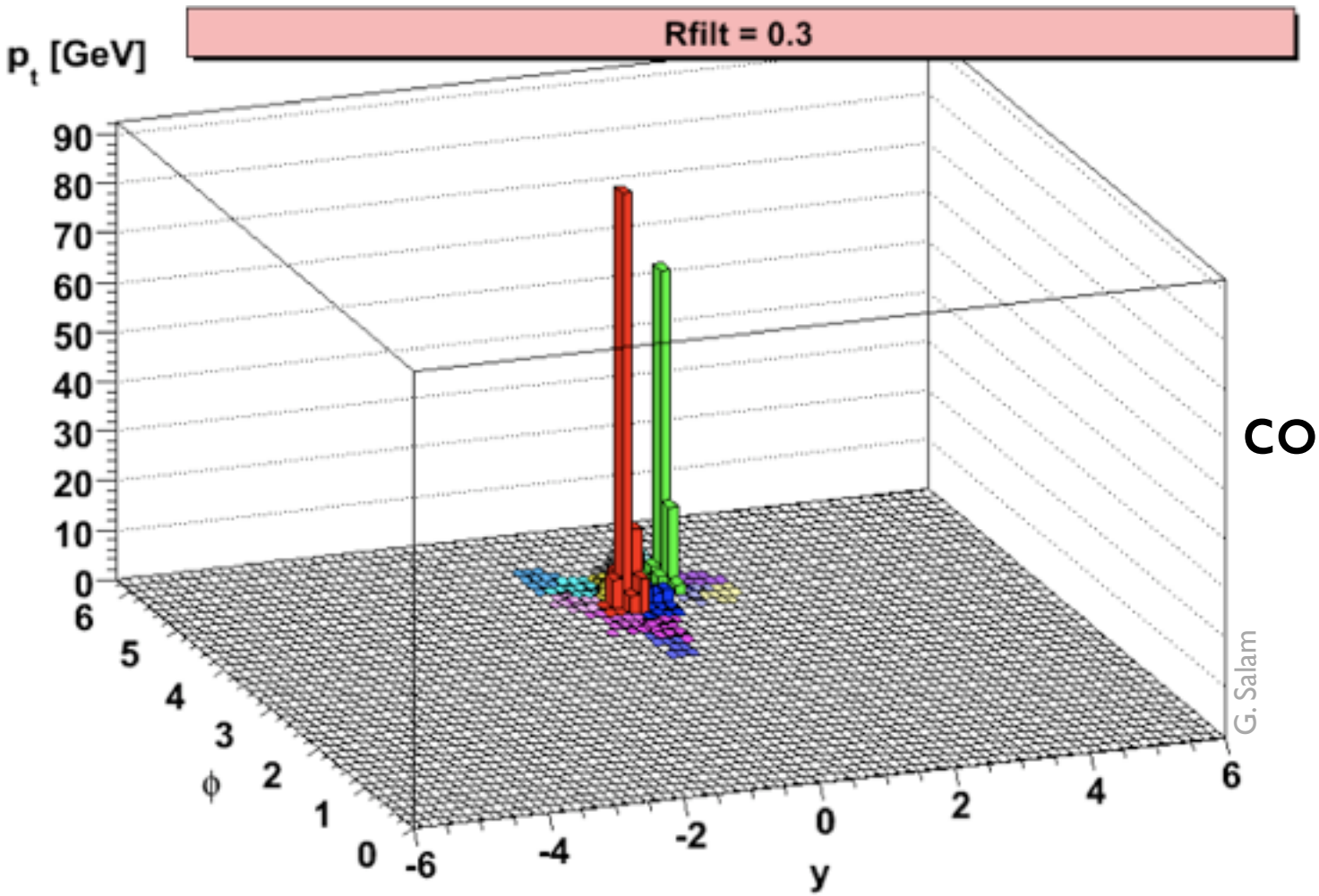
$pp \to ZH \to \nu\nu bb$



Start with the recombined jet

BDRS: filtering

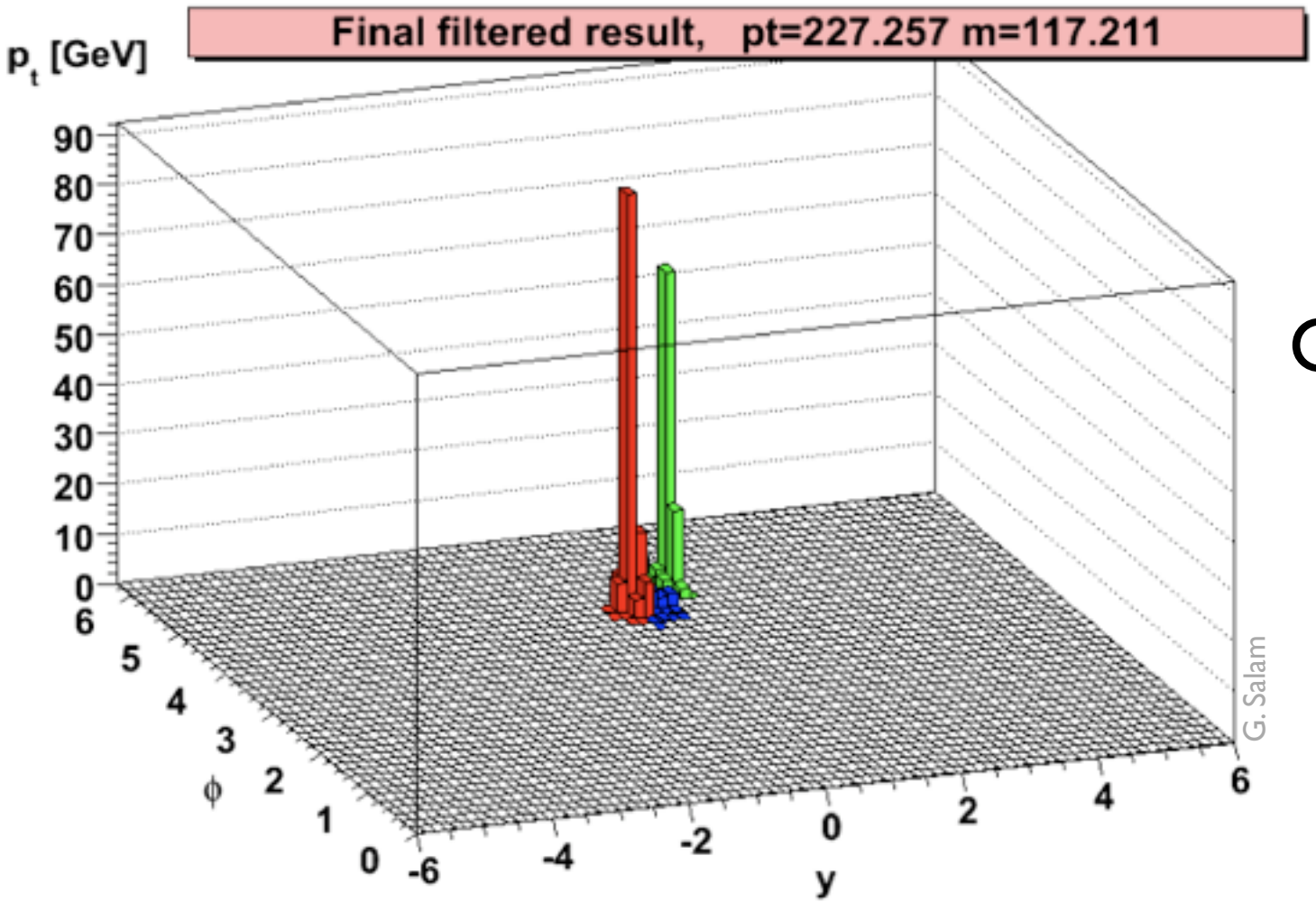
$pp \to ZH \to \nu\nu bb$



Recluster the contituents with R_{filt}

BDRS: filtering

$pp \to ZH \to \nu\nu bb$



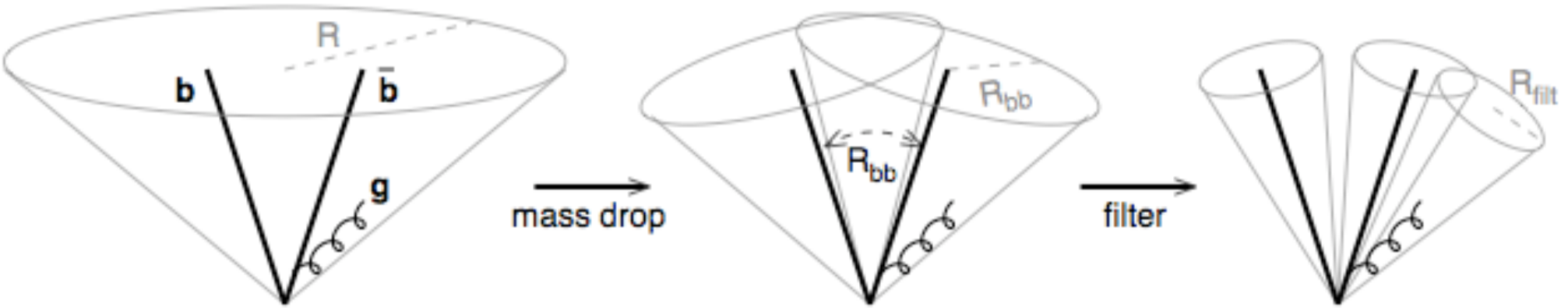
Only keep the n_{filt} hardest jets

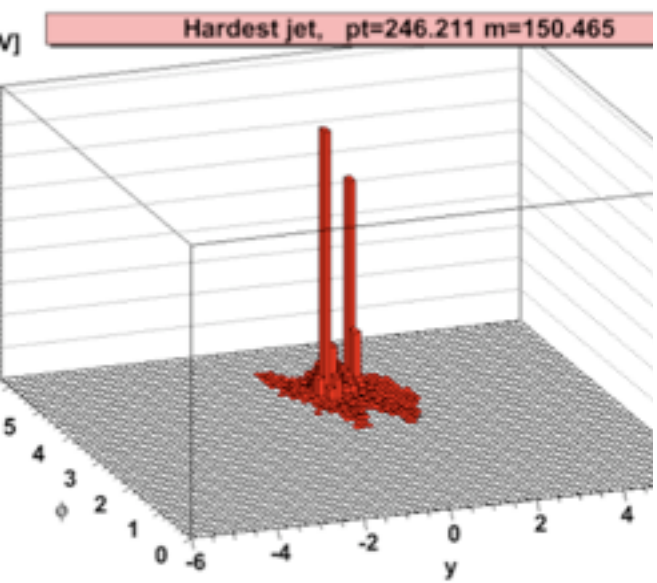
The low-momentum stuff surrounding the hard particles has been removed

$pp \to ZH \to \nu\bar{\nu}b\bar{b}$

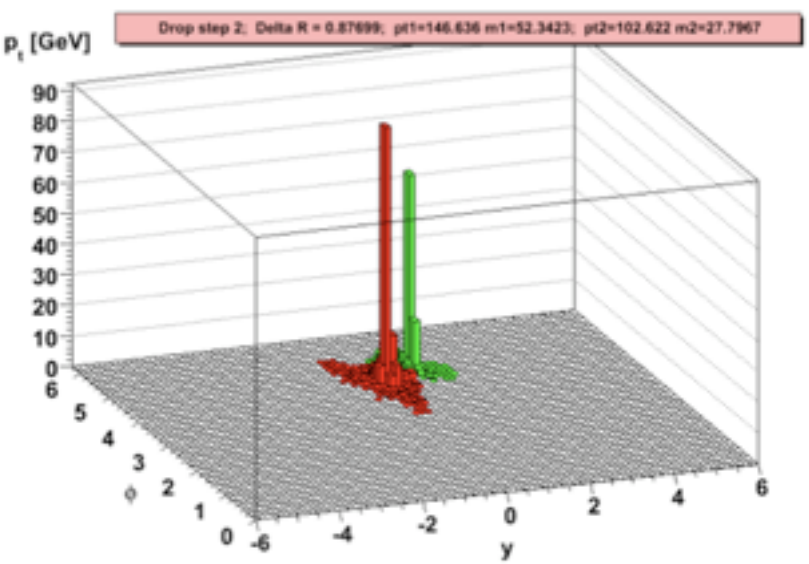
Visualisation of BDRS

Butterworth, Davison, Rubin, Salam, 2008

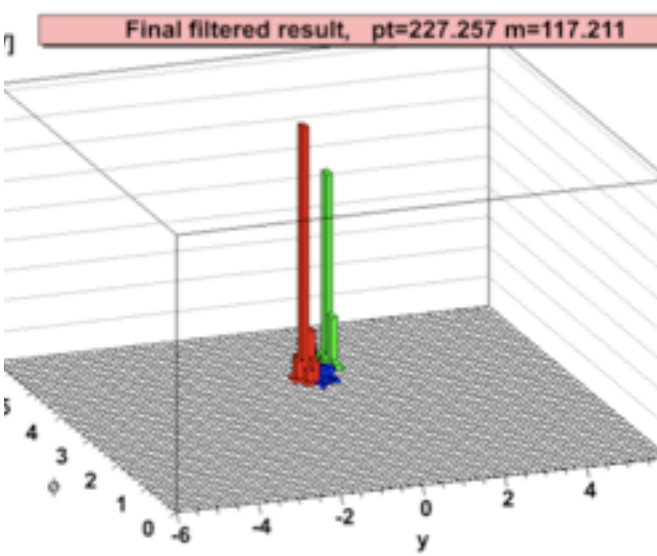




Cluster with a large R



Undo the clustering into subjets, until a large asymmetry/mass drop is observed: tagging step



Re-cluster with smaller R, and keep only 3 hardest jets: grooming step

First taggers/groomers

- **Mass Drop + Filtering** — Butterworth, Davison, Rubin, Salam, 2008

  Decluster with mass drop and asymmetry conditions
  Recluster constituents into subjets at distance scale R_{filt}, retain n_{filt} hardest subjets

- **Jet 'trimming'** — Krohn, Thaler, Wang, 2009

  Recluster constituents into subjets at distance scale R_{trim},
  retain subjets with $p_{t,subjet} > \varepsilon_{trim}\ p_{t,jet}$

- **Jet 'pruning'** — S. Ellis, Vermilion, Walsh, 2009

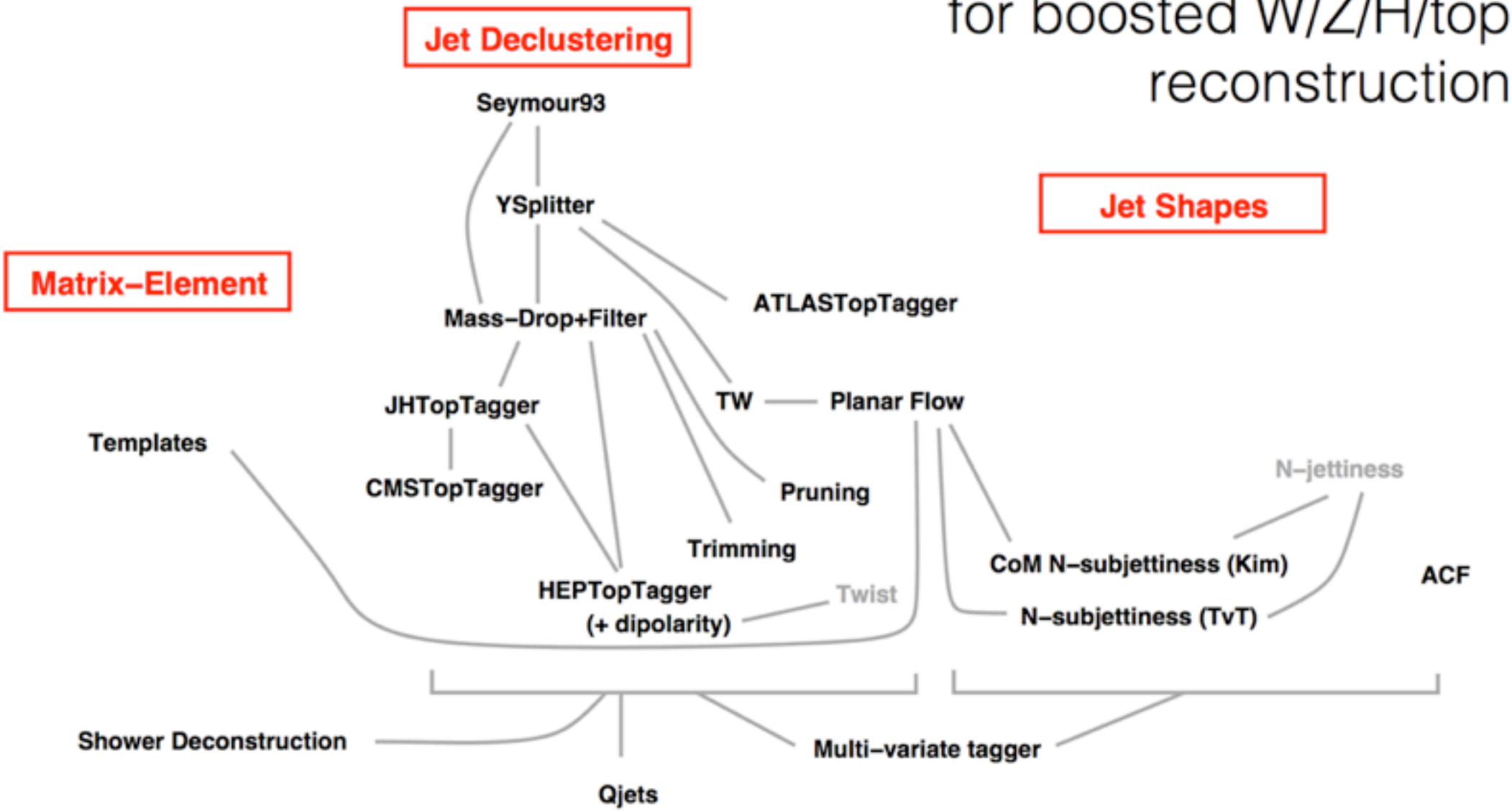
  While building up the jet, discard softer subjets when $\Delta R > R_{prune}$
  and $\min(p_{t1},p_{t2}) < \varepsilon_{prune}\ (p_{t1}+p_{t2})$

Aim: limit contamination from QCD background while retaining bulk of perturbative radiation

Trimming and pruner are a priori groomers, but can become taggers when combined with an invariant mass window test
(if you can groom everything then there's no heavy particle in the jet)

The jet substructure maze

Some of the tools developed for boosted W/Z/H/top reconstruction



Slide by G. Salam, now a few years old

Soft Drop declustering

Larkoski, Marzani, Soyez, Thaler, 2014

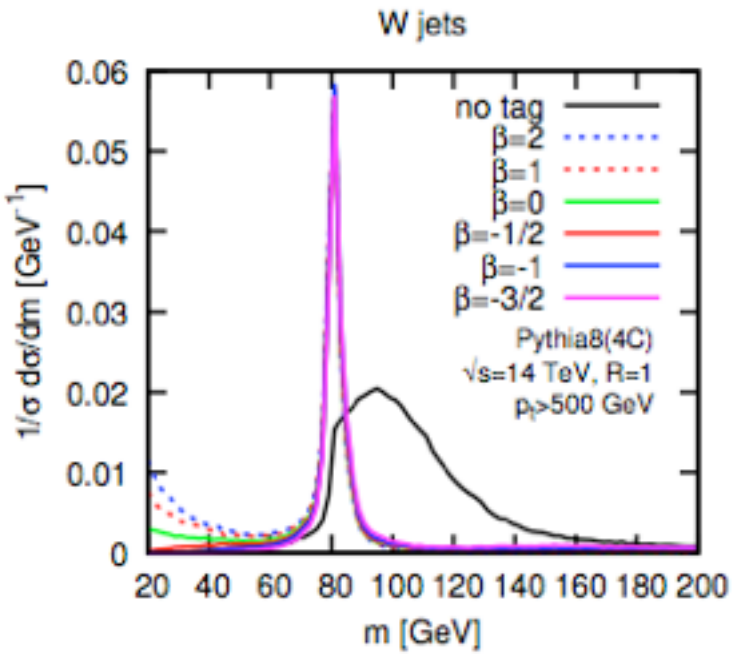
Decluster and drop softer constituent unless

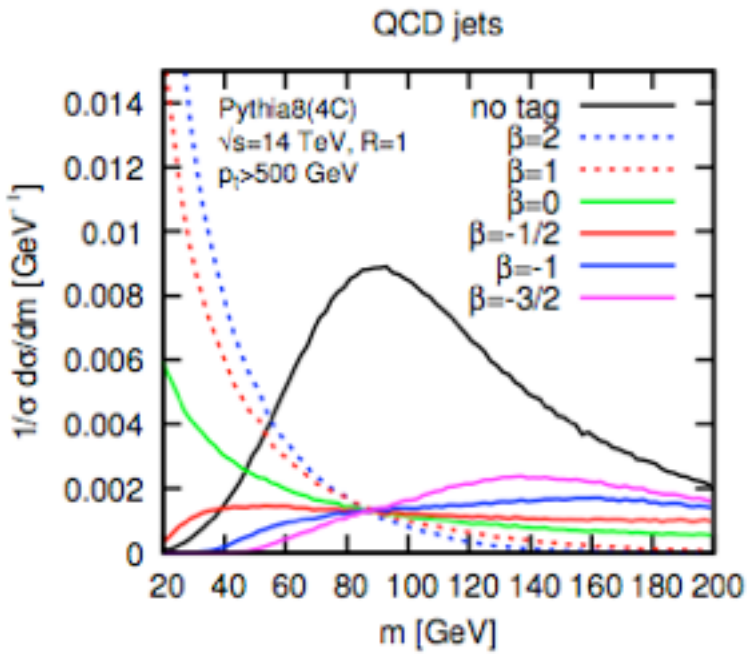
$$\text{Soft Drop Condition:} \quad \frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0} \right)^{\beta}$$

i.e. remove wide-angle soft radiation from a jet

The paper contains

- ✓ analytical calculations and comparisons to Monte Carlos
- ✓ study of effect of non-perturbative corrections
- ✓ performance studies





Example of SoftDrop performance when used as a boosted W tagger

Alternatives to hierarchical substruct.

- If what we are interested in is the structure of the constituents of a jet, the "jet" itself is not the most important feature.
- A different algorithm, or simply the study of the constituents in a certain patch will also do. Selected alternatives are:
  - Use of jet-shapes to characterise certain features
    - e.g. *N-subjettiness*: how many subjets a jets appears to have

      Thaler, van Tilburg, 2011
  - Alternative ways of clustering
    - e.g. *Qjets*: the clustering history not deterministic, but controlled by random probabilities of merging. Can be combined with, e.g. pruning

      Ellis, Hornig, Roy, Krohn, Schwartz, 2012
  - Use information from matrix element
    - e.g. *shower deconstruction*: use analytic shower calculations to estimate probability that a certain configuration comes from signal or from background

      Soper, Spannowsky, 2011
  - Use event shapes mimicking jet properties
    - e.g. *JetsWithoutJets*, mimicking trimming

      Bertolini, Chen, Thaler, 2013

N-subjettiness

Thaler, van Tilburg, 2010

$$\tau_N^{(\beta)} = \sum_i p_{Ti} \min \left\{ R_{1,i}^{\beta}, R_{2,i}^{\beta}, \ldots, R_{N,i}^{\beta} \right\}$$

Sum over constituents of a jet

Distances to axes of N subjets

τ_N measures departure from N-parton energy flow:

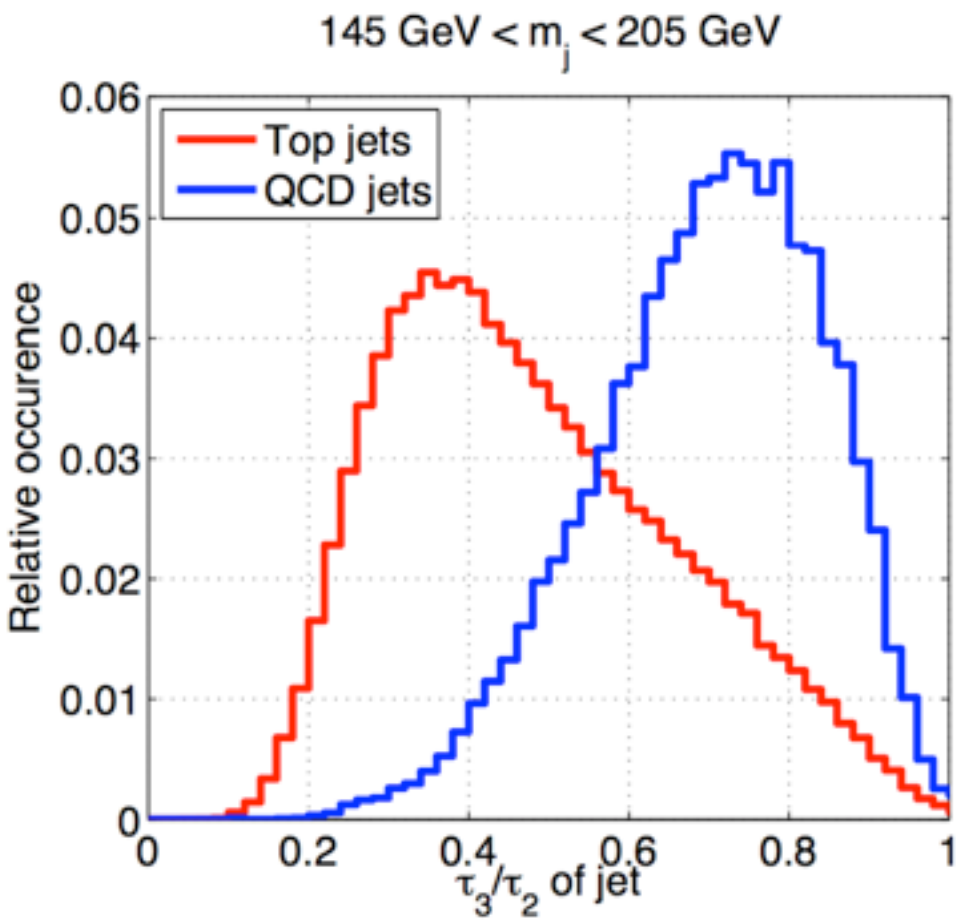
if a jet has N subjets, τ_{N-1} should be much larger than τ_N

N-subjettiness

Thaler, van Tilburg, 2010

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

A jet with a **small** $\tau_{N,N-1}$ is more likely to have N than N-1 subjets



(from 1011.2268, with β=1)

Larkoski, Salam, Thaler 2013

Energy correlation functions

Probes of N-prong structures without requiring identification of subjets

$$\mathrm{ECF}(N,\beta) = \sum_{i_1<i_2<\ldots<i_N\in J} \left(\prod_{a=1}^{N} p_{T i_a}\right) \left(\prod_{b=1}^{N-1}\prod_{c=b+1}^{N} R_{i_b i_c}\right)^{\beta}$$

Angular (y-φ) distances between constituents

ECF(N+1) is zero if there are only N particles

More generally, *if there are N subjets one expects ECF(N+1) to be much smaller than ECF(N)*

[because radiation will be mainly soft/collinear to subjets]

Larkoski, Salam, Thaler 2013

Discriminators

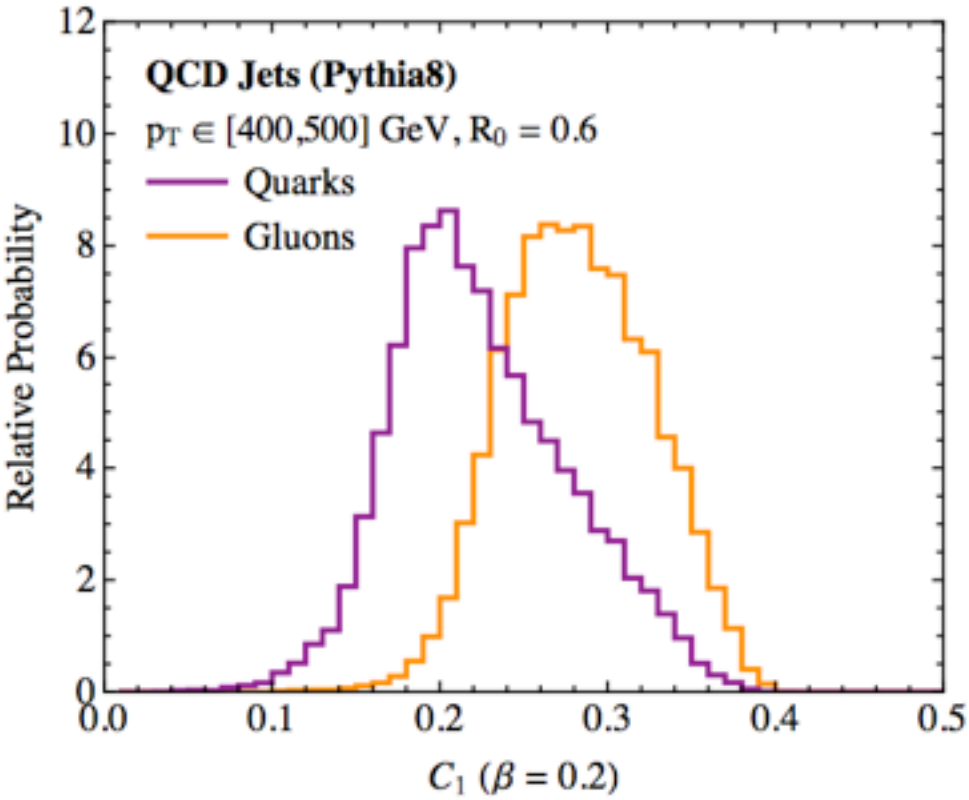
$$r_N^{(\beta)} \equiv \frac{\mathrm{ECF}(N+1,\beta)}{\mathrm{ECF}(N,\beta)}$$

small for N prongs:
if N hard partons, small if radiation only soft-collinear

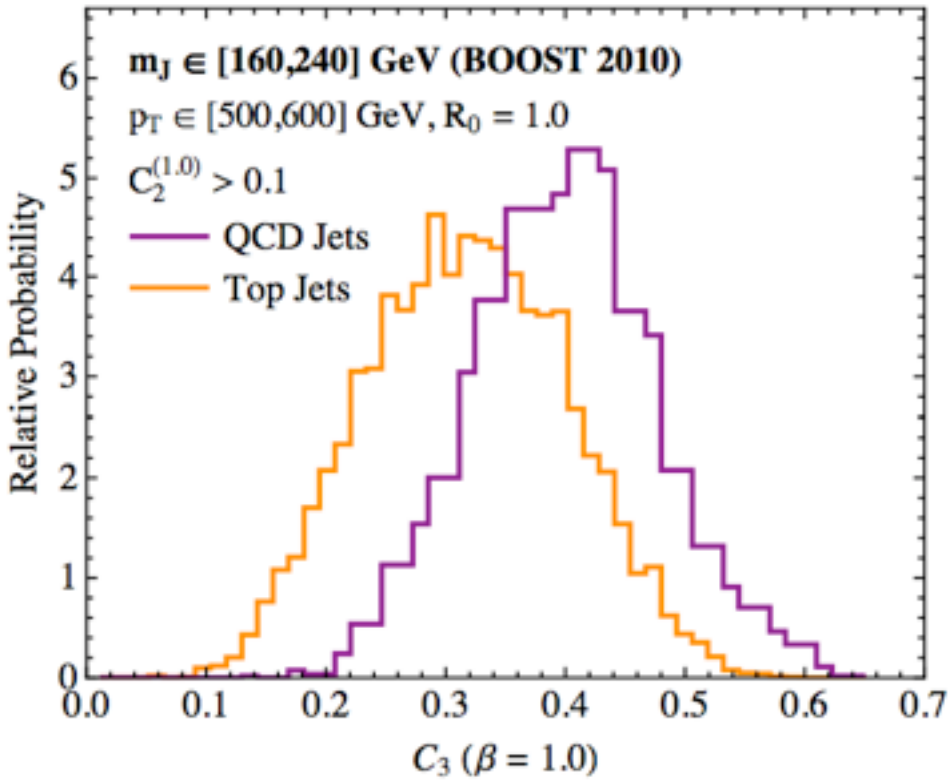
$$C_N^{(\beta)} \equiv \frac{r_N^{(\beta)}}{r_{N-1}^{(\beta)}} = \frac{\mathrm{ECF}(N+1,\beta)\,\mathrm{ECF}(N-1,\beta)}{\mathrm{ECF}(N,\beta)^2}$$

A jet with a **small** C_N is more likely
to have N prongs and at most soft/coll radiation

C_1
quark-gluon discriminator



C_3
top tagging



Note different values of β
(chosen to maximise discriminating power)

Larkoski, Moult, Neill, 2014

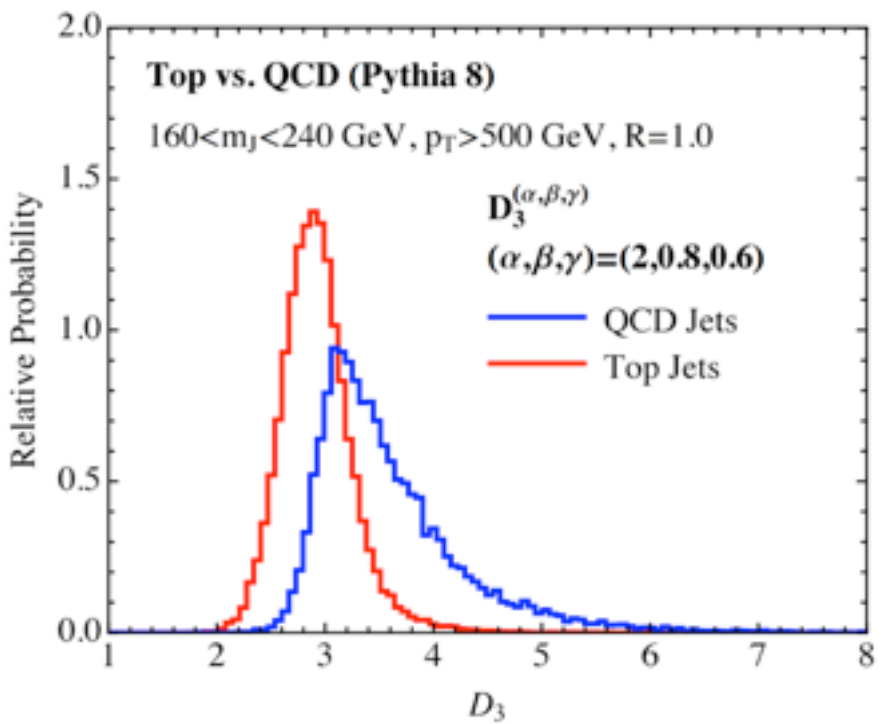
The D functions are variations of the C ones

Instead of $C_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^2}$ $C_3^{(\beta)} = \frac{e_4^{(\beta)} e_2^{(\beta)}}{(e_3^{(\beta)})^2}$

define $D_2^{(\beta)} = \frac{e_3^{(\beta)}}{(e_2^{(\beta)})^3}$ $D_3^{(\alpha,\beta,\gamma)} = \frac{e_4^{(\gamma)} \left(e_2^{(\alpha)}\right)^{\frac{3\gamma}{\alpha}}}{\left(e_3^{(\beta)}\right)^{\frac{3\gamma}{\beta}}} + x \frac{e_4^{(\gamma)} \left(e_2^{(\alpha)}\right)^{\frac{2\gamma}{\beta}-1}}{\left(e_3^{(\beta)}\right)^{\frac{2\gamma}{\beta}}} + y \frac{e_4^{(\gamma)} \left(e_2^{(\alpha)}\right)^{\frac{2\beta}{\alpha}-\frac{\gamma}{\alpha}}}{\left(e_3^{(\beta)}\right)^{2}}$

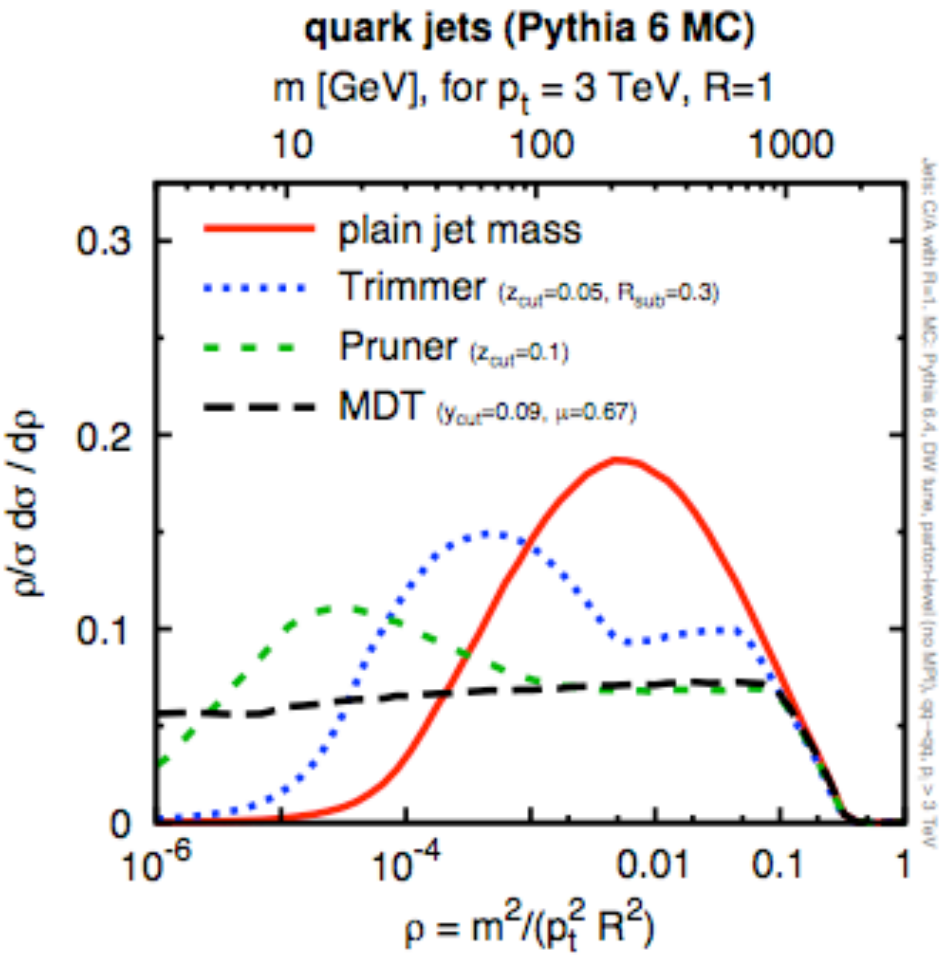
Attempt to improve the discriminating power, and to account for different regions of phase space of radiation

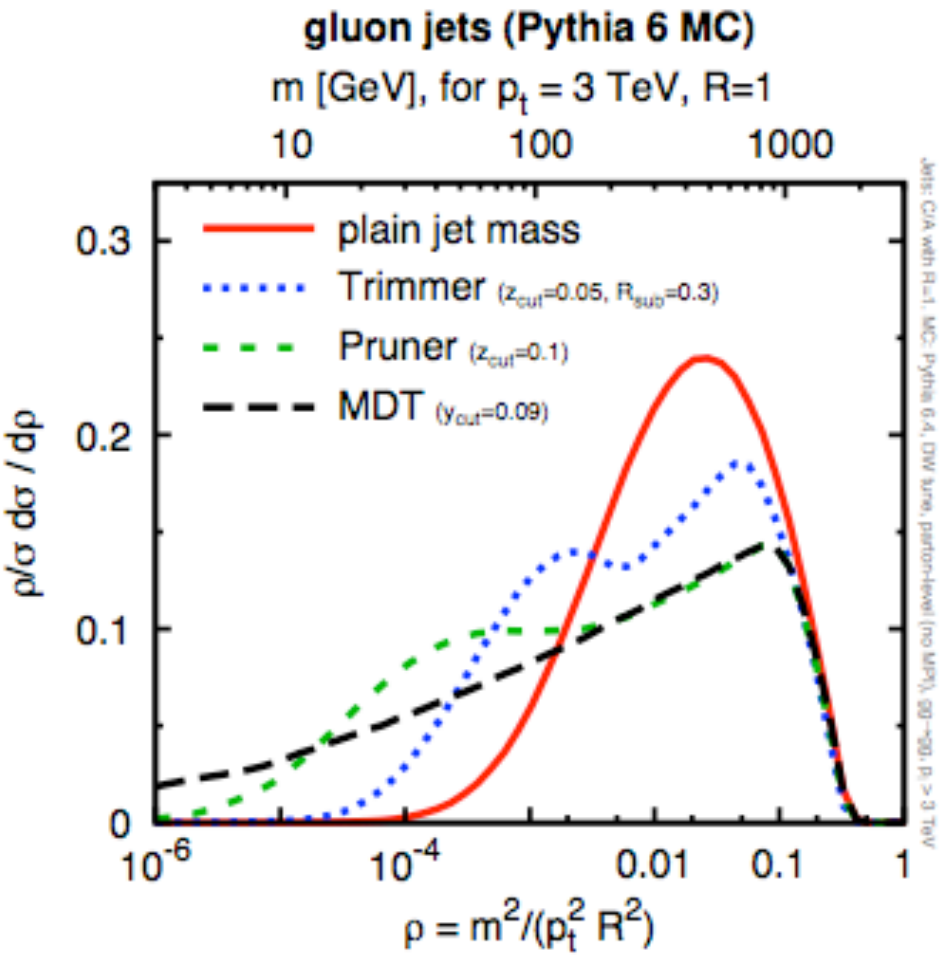
[also, gives an idea of increasing 'sophistication', or complexification]



Robustness of substructure tools

Dasgupta, Fregoso, Marzani, Salam, 2013



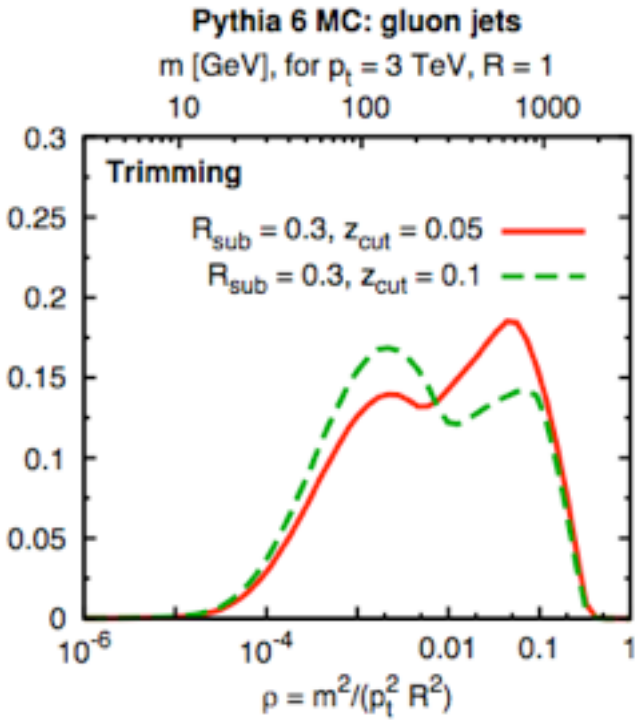


Tools that are considered (or can be seen in Monte Carlo tests) to behave 'similarly' could cease to do so in different parameter regions

Analytic calculations of jet substructure

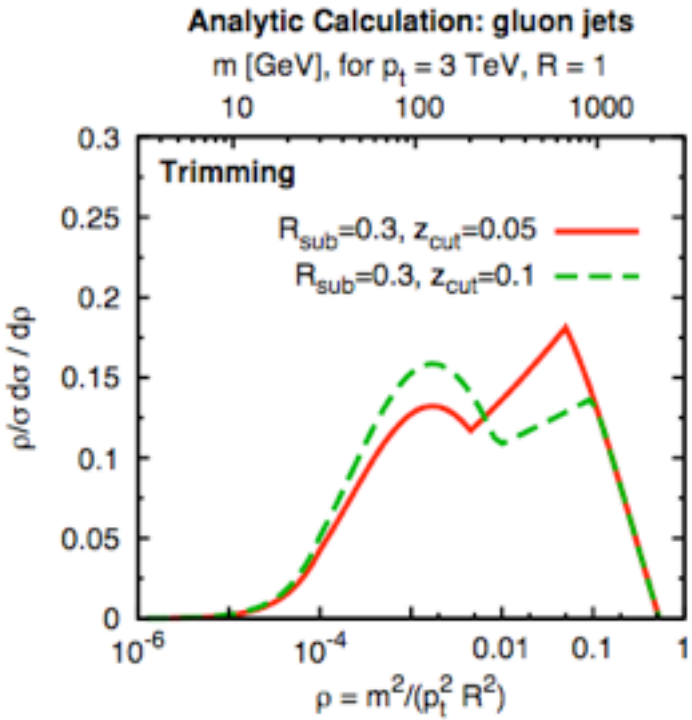
Dasgupta, Fregoso, Marzani, Salam, 2013

Monte Carlo



Analytic

(resummed pQCD)



- Analytical understanding of 'kinks' in distributions
- Check of Monte Carlo predictions
- Other analytical investigations: Rubin 2010 (filtering), Walsh, Zuberi 2011 (jet substructure with SCET), Feige Schwartz, Stewart, Thaler 2012 (N-subjettiness), Dasgupta, Marzani, Powling 2013 (groomed jet mass), ...

$$\frac{1}{\sigma}\frac{d\sigma}{dm^2}^{(\text{trim, LO})} = \frac{\alpha_s C_F}{\pi}\int_0^1 dz\, p_{gq}(z)\int\frac{d\theta^2}{\theta^2}\,\delta\left(m^2 - z(1-z)p_t^2\theta^2\right)\times$$
$$\times\left[\Theta\left(z - z_{\text{cut}}\right)\Theta\left(1 - z - z_{\text{cut}}\right)\Theta(\theta^2 - R_{\text{sub}}^2) + \Theta(R_{\text{sub}}^2 - \theta^2)\right]\Theta\left(R^2 - \theta^2\right)$$

Recap of Lecture 2

The big news of the past few years has been the development of taggers and groomers using properties of jet substructure, through

- declustering
- jet shapes
- direct analysis of images (machine learning)

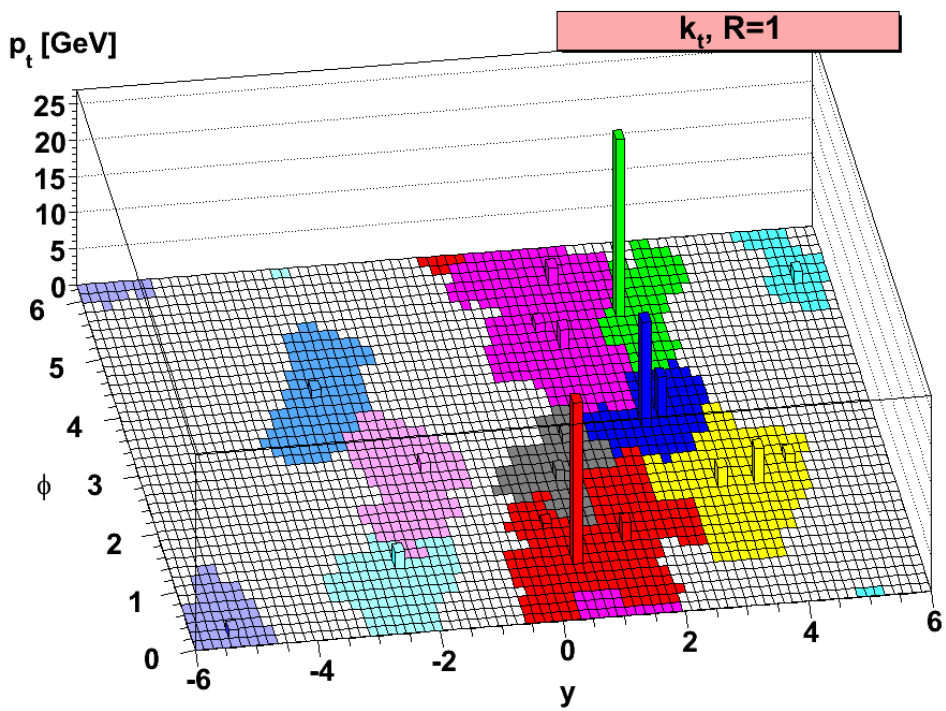
These techniques have been commissioned by experimental collaborations proven their worth in 'Standard Model' analyses. They are now being implemented in BSM searches

Bonus: Jet areas

A jet's area is **defined** as the extent of the region where infinitesimally soft particles get clustered into the jet

More in details, a jet's **active area** is the extent of the region where a distribution of infinitesimally soft particles, that can also cluster among themselves, is clustered into the jet

A jet's active area, represented by the coloured regions, measures a jet's **susceptibility to contamination** from soft particles like underlying event and pileup



Jets do not necessarily have cone-shaped profiles

From jet 'reach' to jet areas

MC, Salam, Soyez, 0802.1188

Not one, but three **definitions** of a jet's size:

- **Passive** area

  Place a single very soft particle (a '**ghost**') in the event, measure the extent of the region where it gets clustered within a given jet

  Reach of jet for **pointlike** radiation

- **Active** area

  Fill the events with many very soft particles ('**ghosts**'), cluster them together with the hard ones, see how many get clustered within a given jet

  Reach of jet for **diffuse** radiation

- **Voronoi** area

  Sum of areas of intersections of Voronoi cells of jet constituents with circle of radius R centred on each constituent

  Coincides with passive area for k_t algorithm

(In the large number of particles limit all areas converge to the same value)

Passive Area

Add a **single** ghost* particle to the event.
Move it around.
Check if it gets clustered in a given jet J.

$$a(J) \equiv \int dy\, d\phi\, f(g(y,\phi),J) \qquad f(g,J) = \begin{cases} 1 & g \in J \\ 0 & g \notin J \end{cases}$$

* ghost particle: particle with infinitesimally small momentum with respect to all other particles in the event (in practice, O(10^{-100} GeV))

Active Area

Add **many** ghost particles in random configurations to the event.
Cluster many times. *Allow ghosts to cluster among themselves too.*
Count how many ghosts on average get clustered into a given jet J.

$$A(J \mid \{g_i\}) = \frac{N_g(J)}{\nu_g} \equiv A_g N_g(J)$$

- $A(J \mid \{g_i\})$: Active area of jet J for a single ghosts configuration
- $N_g(J)$: Number of ghosts in jet J
- ν_g: Ghost density
- A_g: Area of a single ghost

$$A(J) = \lim_{\nu_g \to \infty} \langle A(J \mid \{g_i\}) \rangle_g$$

- $A(J)$: Active area of jet J

Jet areas: the single hard particle case

It is worth noting that, for a jet made of a **single hard particle**, while **passive** areas are indeed **πR^2, active** areas are **not**

Active areas	k_t	Cam/Aa	SISCone	anti-k_t
$\langle A \rangle / \pi R^2$	0.81	0.81	1/4	1

Only anti-k_t has the behaviour one would naively expect, i.e. area = πR^2

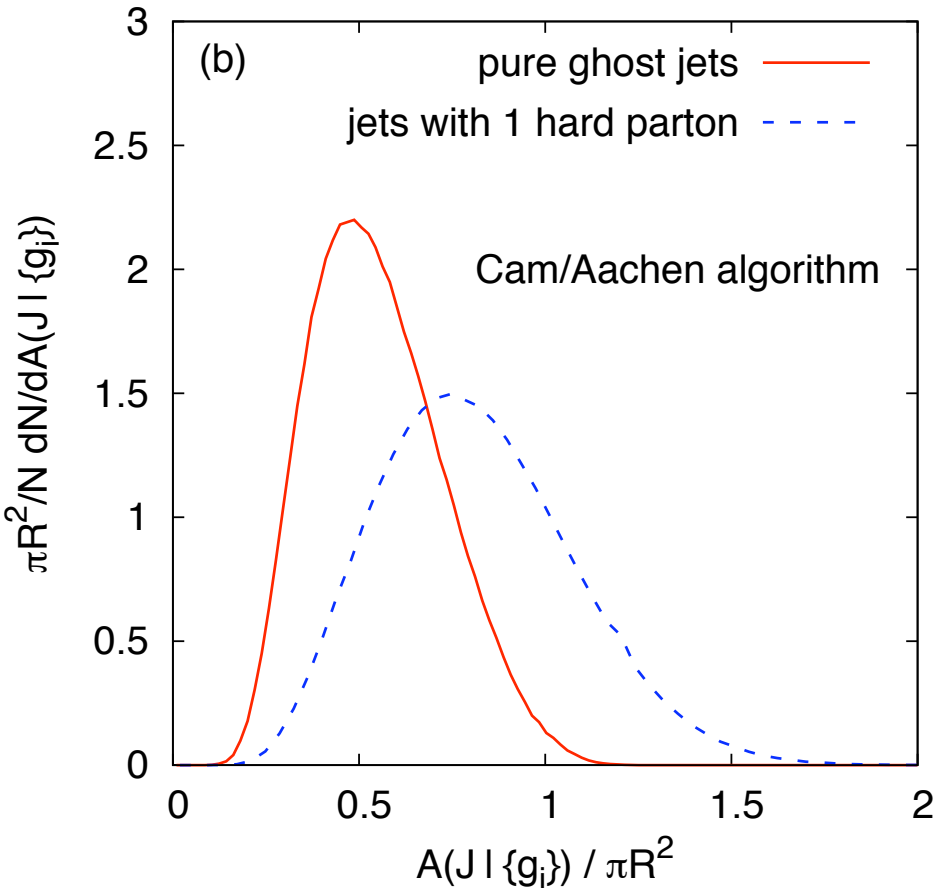
Jet areas: the pure ghost jet case

Jets composed exclusively of ghosts can also be clustered, and their active area measured

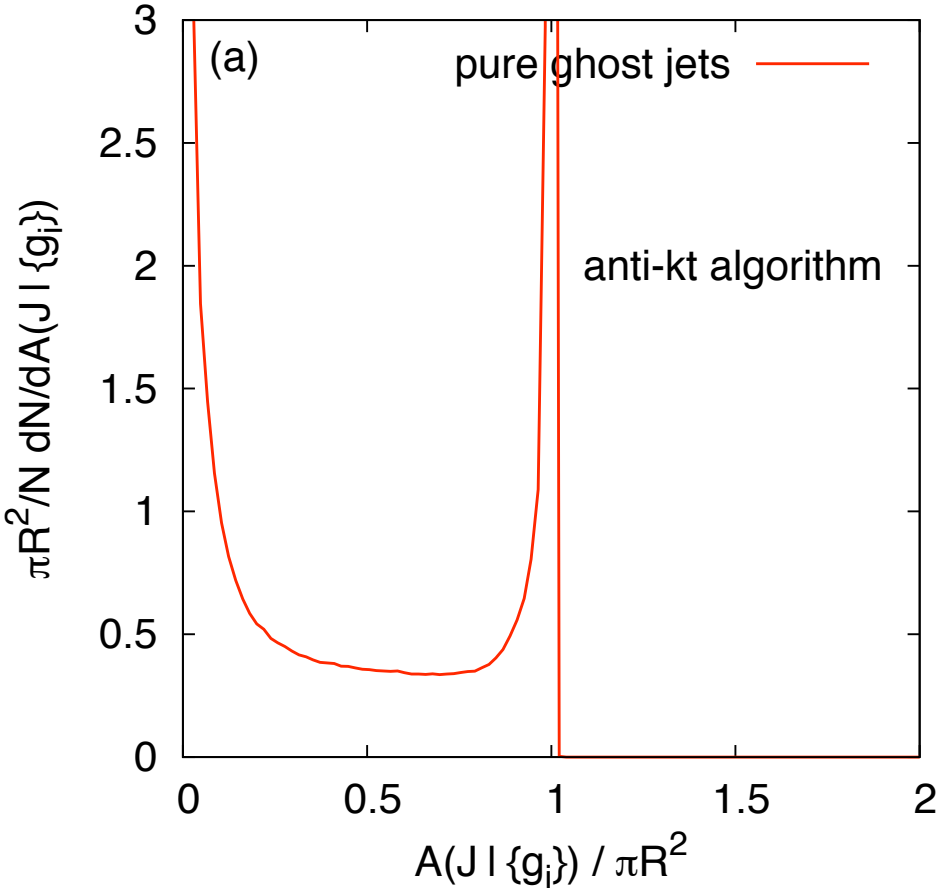
Active areas	k_t	Cam/Aa
$\langle A \rangle / \pi R^2$	0,554	0,551

Active area distributions

k_t and Cam/AA



anti-k_t



For a roughly uniformly soft background, anti-k_t gives many small jets and many large ones (you can't fill a plane with circles!)

Active area: analytical results?

Can these active area distributions be calculated analytically?

At least the averages?

At least for pure ghost jets?

A challenge

Can we predict analytically the active area distribution or, at least, its average, for the **self-clustering of a large number of 'particles'**?

Consider a simplified toy-model: 1-dimensional clustering, and an "euclidean" recombination scheme: the recombination of two 'particles' is simply the mid-point of their coordinate.

N 'particles' are distributed randomly over a length L, such that

$$a \equiv L/N \ll 2R \ll L$$

where R is the 'radius' parameter of the Cambridge-like clustering algorithm

Upon clustering, we get n_J jets, with number of constituents n_C.
The average 'area' (over many events) of a jet will be $\langle A \rangle = a \langle n_C \rangle$

Can we calculate analytically $\lim_{a/2R \to 0, 2R/L \to 0} \frac{\langle A \rangle}{2R}$?

(In this toy model, or in a similarly simple one. What about in higher dimensions?)