2013 TAE School
September 2013, Benasque

QCD & Jets & MC Modeling

Matteo Cacciari
LPTHE Paris

Lecture 3

Many thanks to Guenther Dissertori, Rikkert Frederix, Fabio Maltoni, Paolo Nason, Gavin Salam, Gregory Soyez, Maria Ubiali, and probably others, from whose talks/lectures I have drawn inspiration, as well as extracted many slides

Outline

- Some basics of QCD
- Initial state
  - PDFs
- Hard scattering (and more)
  - higher order calculations and generators
    - Parton shower MCs
    - Merging
- Final state
  - Jets algorithms and jet areas
  - Jets as tools (jet substructure)

Lecture 1

Lectures 2 and 3

[Subdivision in parts actually quite unreliable. Length/depth of descriptions varies quite a lot]

Tools

Background characterisation and subtraction

Mass reconstruction

Remove soft contamination from a hard jet

Tag heavy objects originating the jet

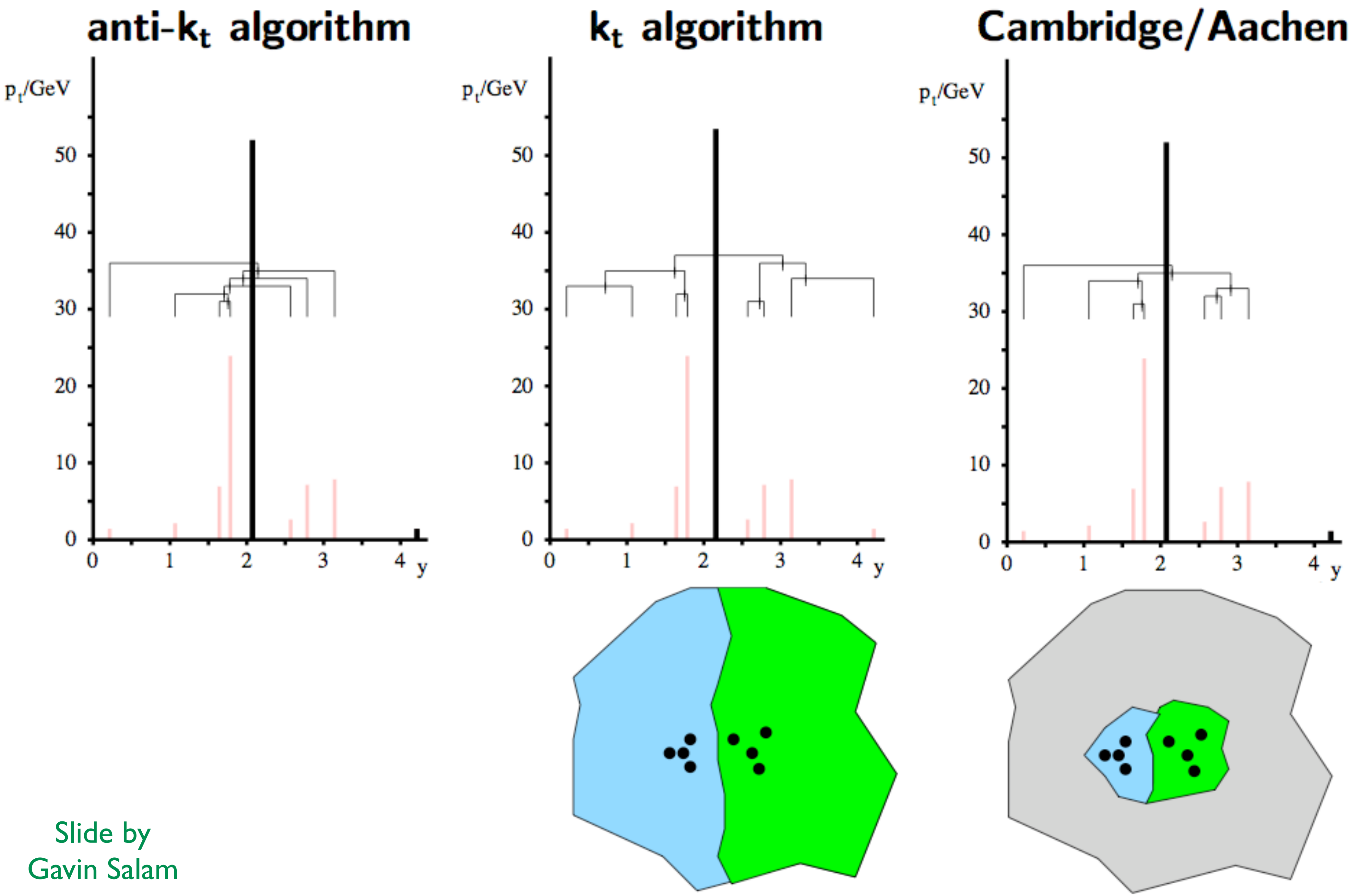
Eventually leading to ‘**third-generation**’ jet algorithms

Nomenclature

- **Groomer**
  - procedure that always returns an output jet
    (e.g. it only subtracts uncorrelated 'UE/pileup' radiation from it)
- **Tagger**
  - procedure that might not return an output jet
    (e.g. it either tags a heavy particle originating the jet or returns zero)

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

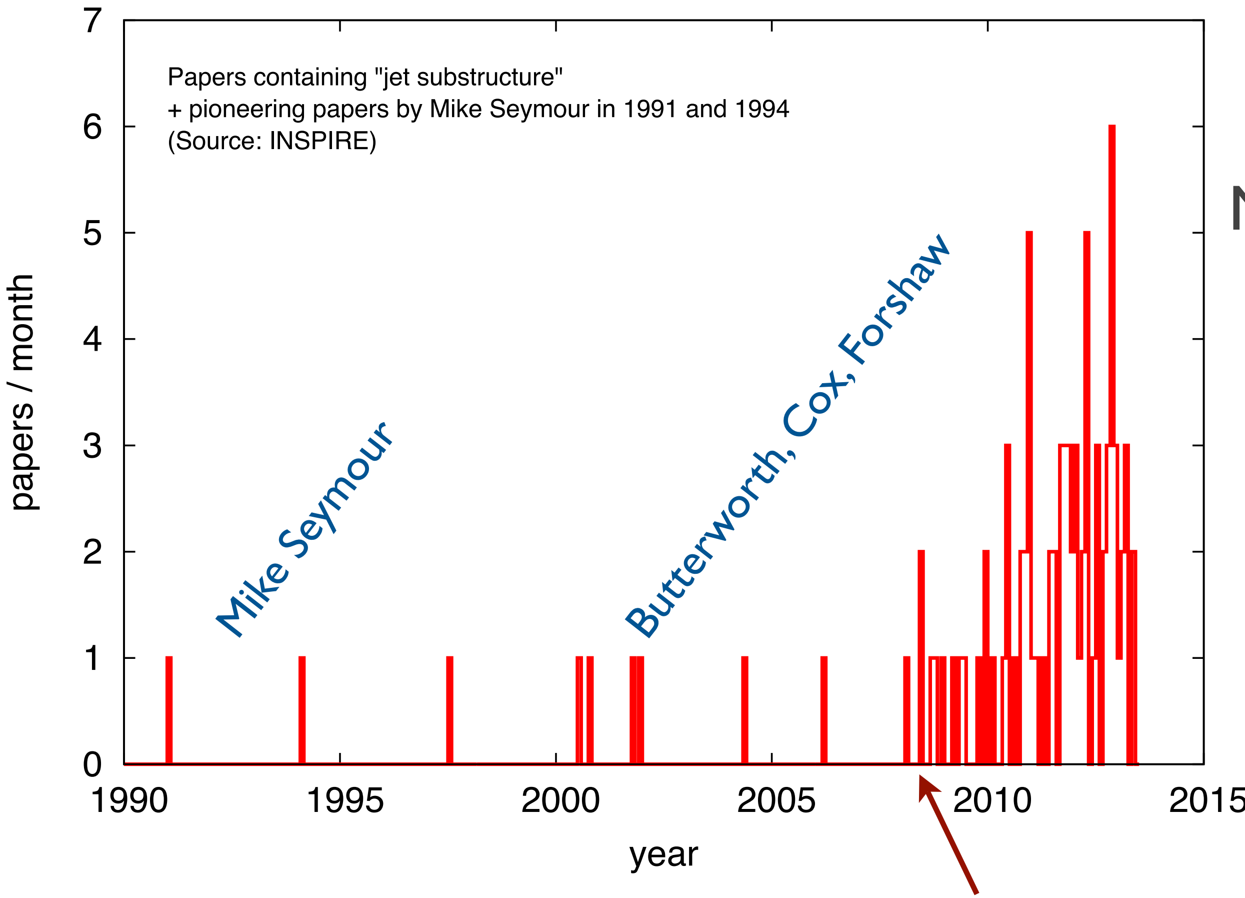
Hierarchical substructure



Slide by
Gavin Salam

‘Jet substructure’ papers in INSPIRE

Number of papers containing the words ‘jet substructure’



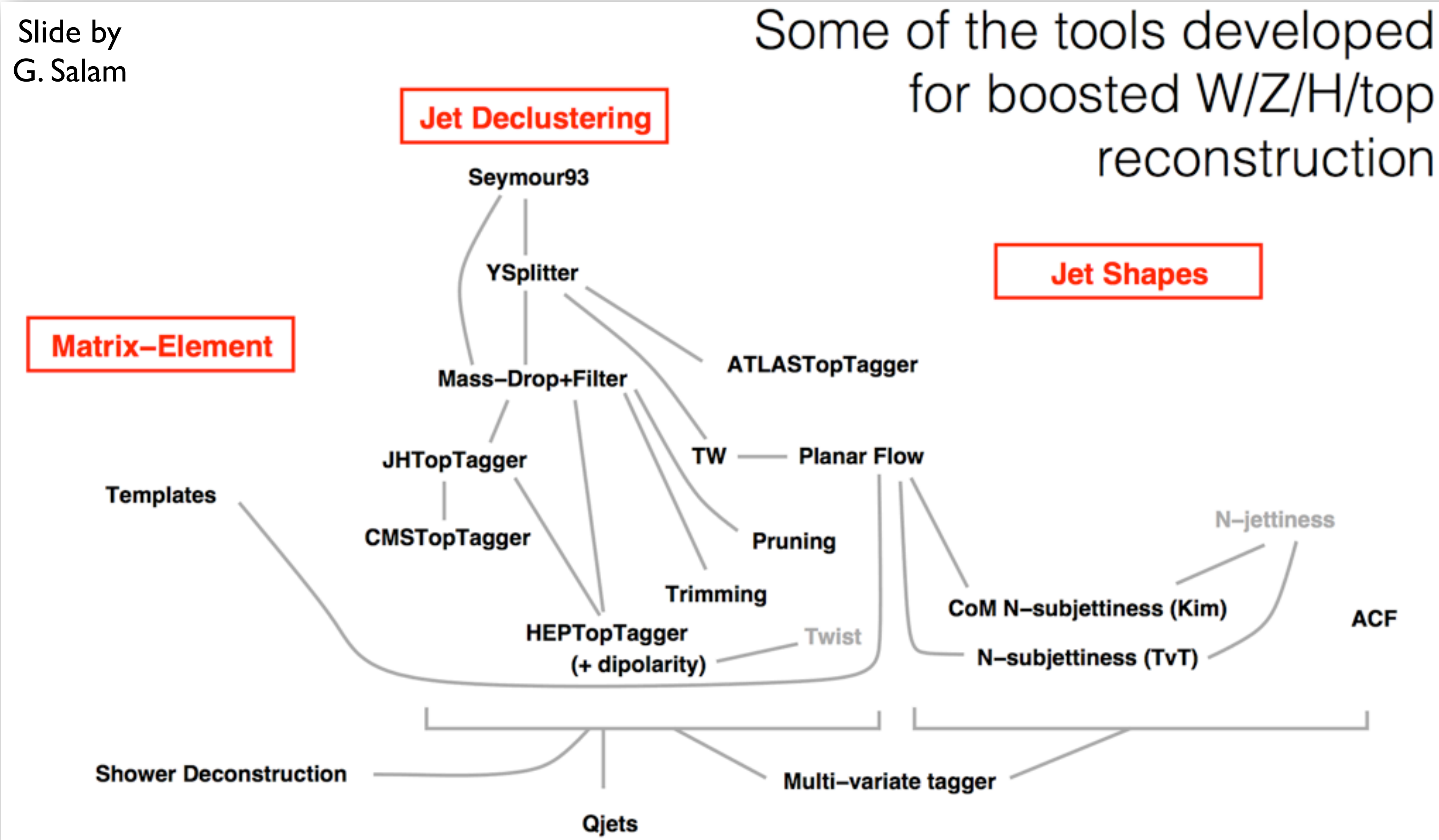
More than 100 papers since 2008

(+ some background noise)

Pioneered by M. Seymour in the early ‘90s, rebooted by BDRS paper

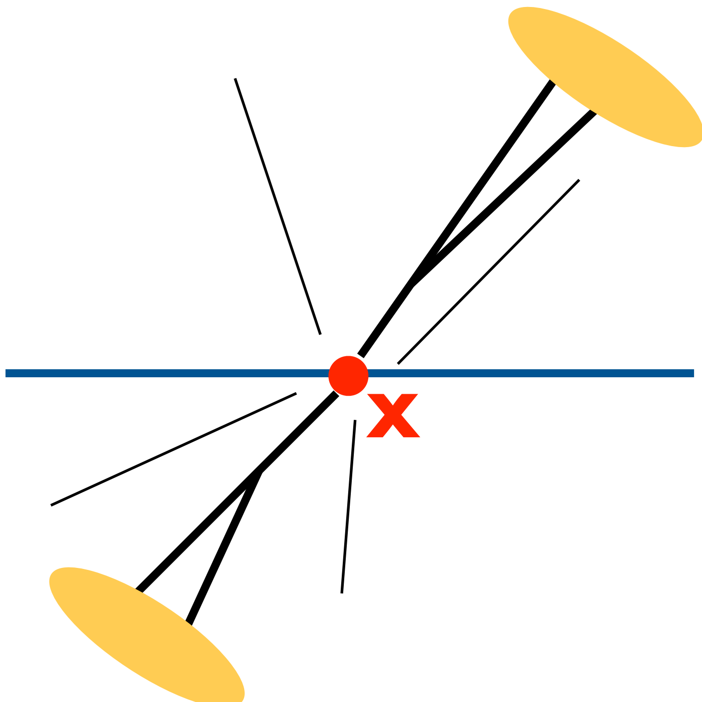
15. **Jet substructure as a new Higgs search channel at the LHC.**
Jonathan M. Butterworth, Adam R. Davison (University Coll. London), Mathieu Rubin, Gavin P. Salam (Paris, LPTHE).
Published in **Phys.Rev.Lett. 100 (2008) 242001**
e-Print: **arXiv:0802.2470 [hep-ph]**

The jet substructure maze



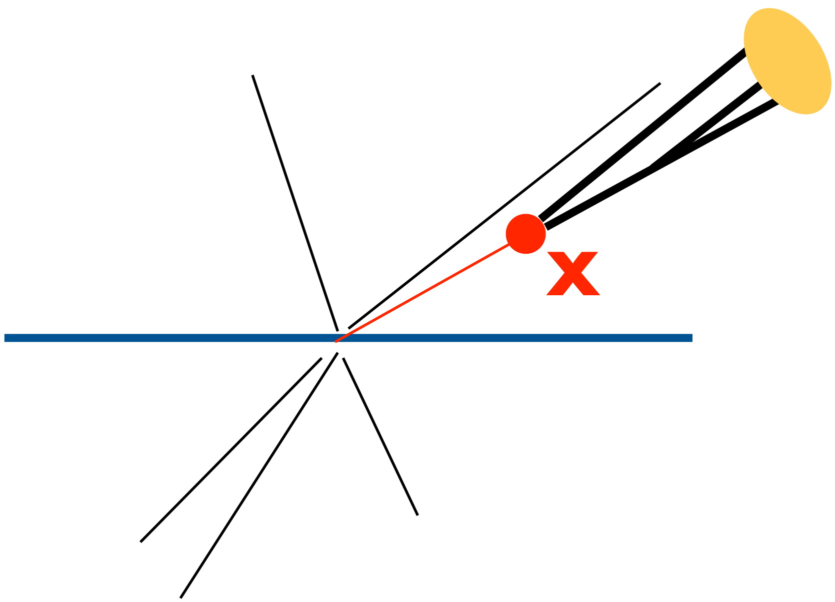
Apologies for missing or misplaced items or links

Why boosted objects



Heavy particle X at **rest**

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

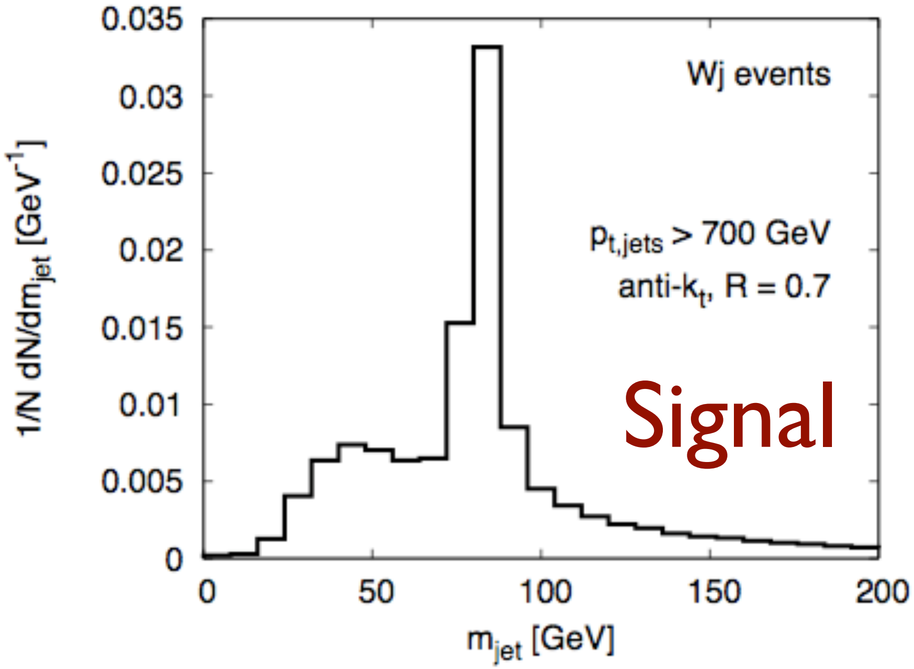


Boosted heavy particle X

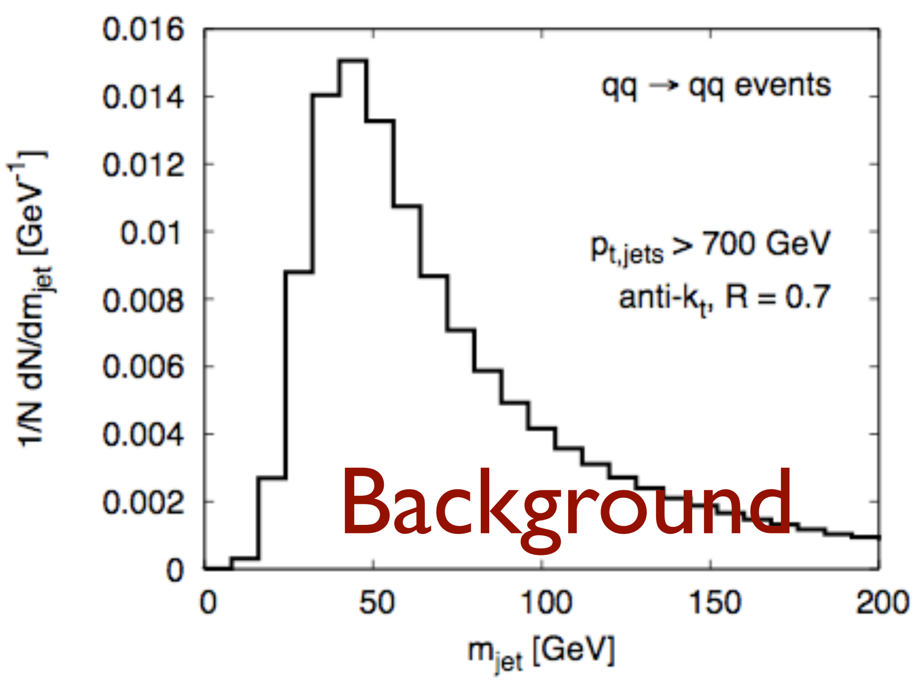
Cross section very much reduced, but acceptance better and some backgrounds smaller/ reducible

Mass of a single jet

G. Salam



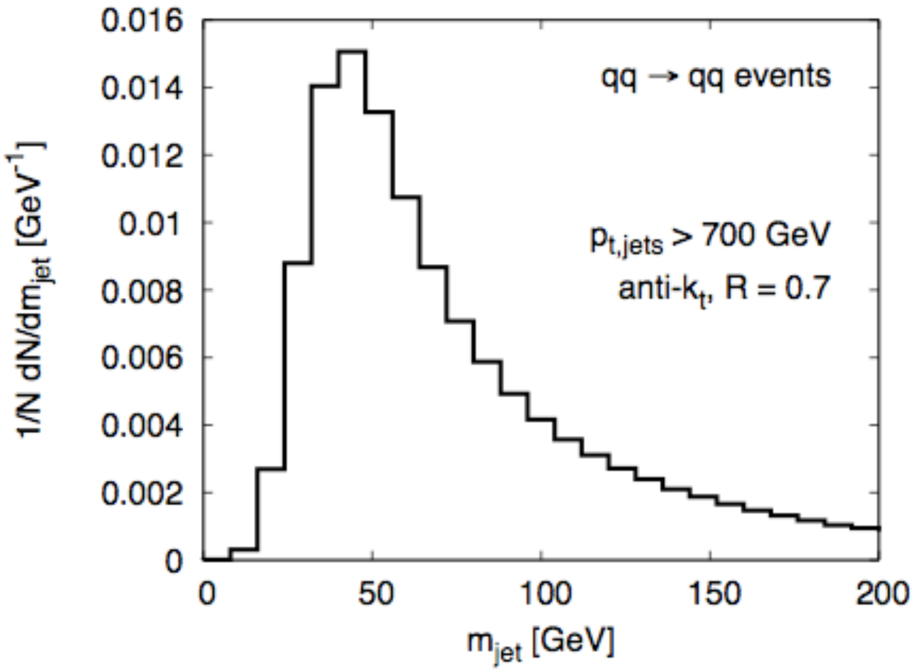
A heavy object decaying into a single jet naturally gives it a mass...



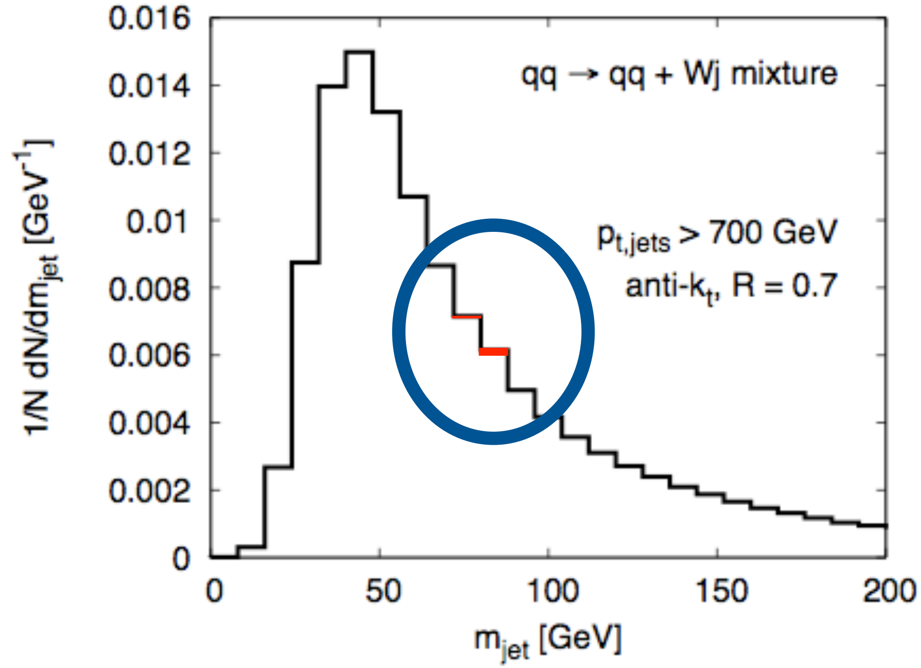
... but pure QCD jets can be massive too:

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

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



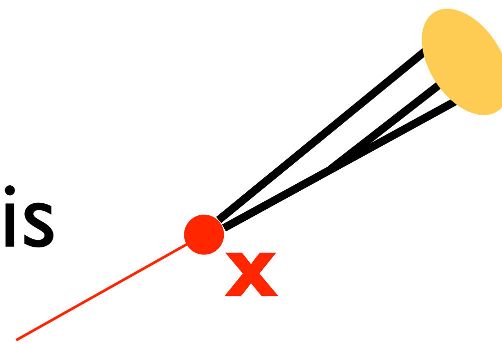
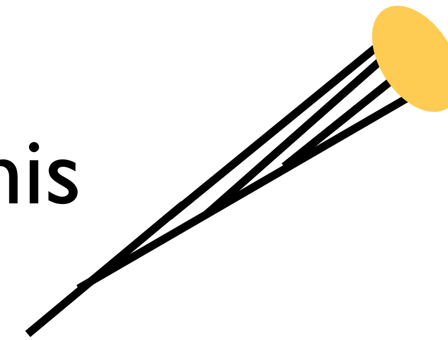
Background only



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

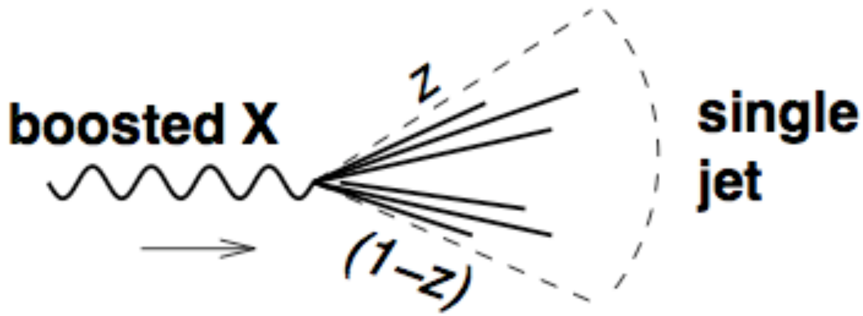
How to tell this  from this  ?

Decay of a heavy (boosted) object

Light parton fragmentation

Why substructure

Scales: m ~ 100 GeV, p_t ~ 500 GeV



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

- 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

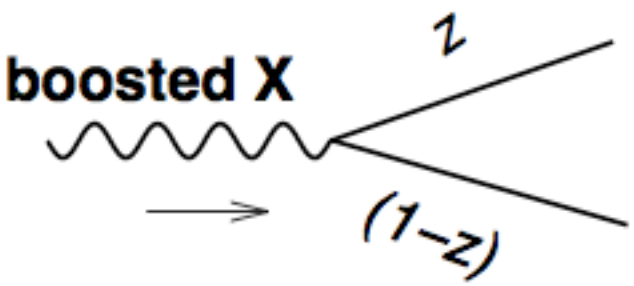
- Use large R, get a single jet : background large
- Use small R, resolve the jets : what is the right scale?
- Let an algorithm find the 'right' substructure

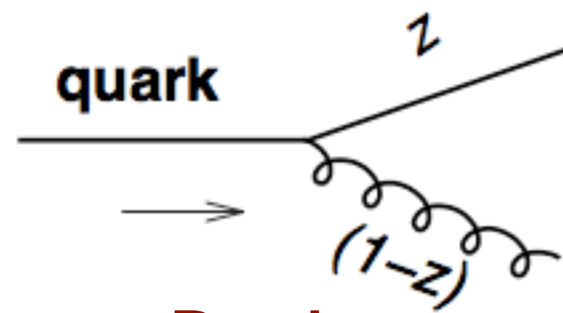
Why jet substructure

- The substructure of a jet (i.e. the ability to further resolve smaller components) can be exploited to
  - *tag a particular structure inside the jet, i.e. a massive particle*
    - *Examples: Higgs (2-prongs decay), top (3-prongs decay)*
  - *remove background contamination from the jet or its components*
    - *Examples: filtering, trimming, pruning*

In the following I'll be mainly illustrating the BDRS tagger/filter as a pedagogical example, and also list other approaches

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





($p_{tj} = zp_t$)

($p_{ti} = (1\text{-}z)zp_t$)

Signal:

$$P(z) \sim \text{constant}$$

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**

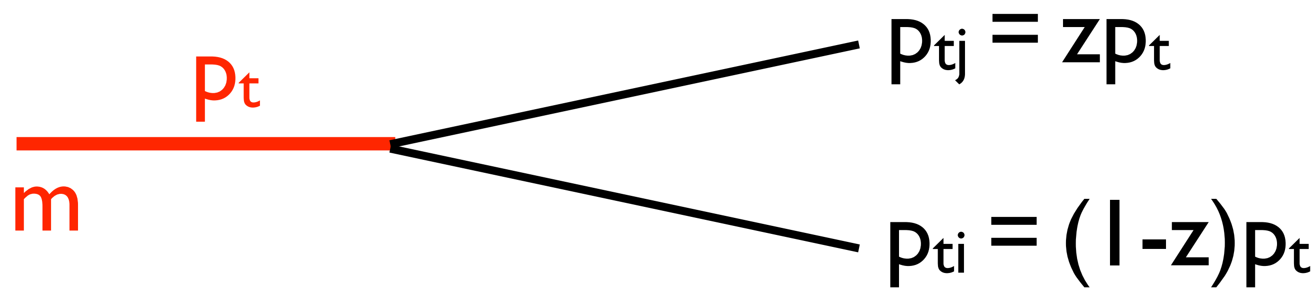
Potential tagger: asymmetric splitting

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

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, therefore 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

The strategy

A generic substructure approach to tagging will

- Cluster initially with a large R, so as to collect all the decay products of a boosted heavy particle into a single jet
- Decluster this jet into subjets, using some conditions to decide when to stop the declustering (i.e. find the 'relevant splitting'), possibly including kinematical cuts to reduce the QCD background.
  - *The stopping criterion automatically finds the 'right size' for the distance between the two prongs of the heavy particle decay*
    - Alternatively to declustering, one can employ one of the jet-shapes based tagging methods, i.e. N-subjettiness ratios
- Optionally add a final 'cleaning' procedure to remove as much as possible spurious soft/background radiation

Generic tagging/grooming

Fat-jet finding

Often anti-kt, R ≈ 1

↓ large p_t, large mass fat-jet, signal or background

Tagging step

↓ signal jet candidate, still background-contaminated

grooming step

↓

final candidate, potentially with little background contamination

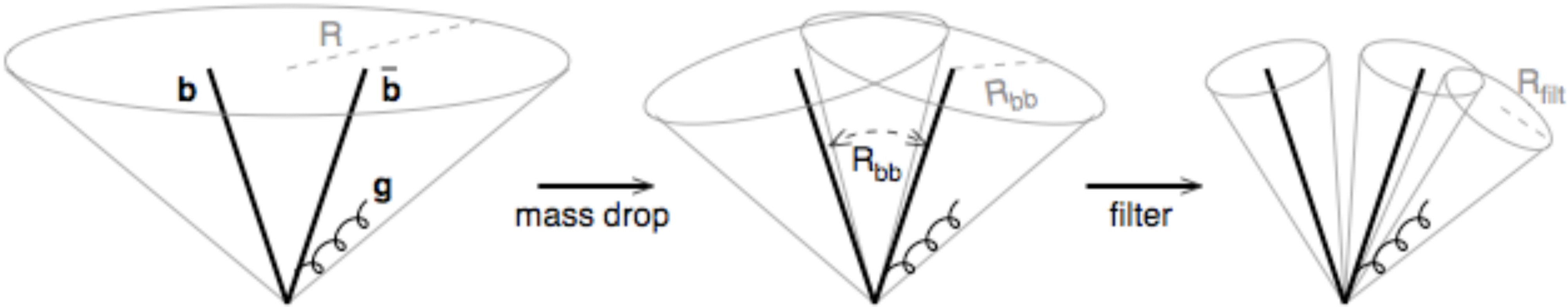
Note that in some taggers (i.e. pruning) the tagging and grooming steps are not explicitly factorised

Also, some tools may actually not follow rigidly this scheme

The BDRS tagger

These ideas led to the first 'modern' implementation of a boosted tagger

15. **Jet substructure as a new Higgs search channel at the LHC.**
Jonathan M. Butterworth, Adam R. Davison (University Coll. London), Mathieu Rubin, Gavin P. Salam (Paris, LPTHE).
Published in **Phys.Rev.Lett. 100 (2008) 242001**
e-Print: **arXiv:0802.2470 [hep-ph]**



It's a two-prongs tagger for boosted Higgs, which

- Uses the **Cambridge/Aachen** algorithm (see why in the next slide)
- 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

Why C/A and not k_t

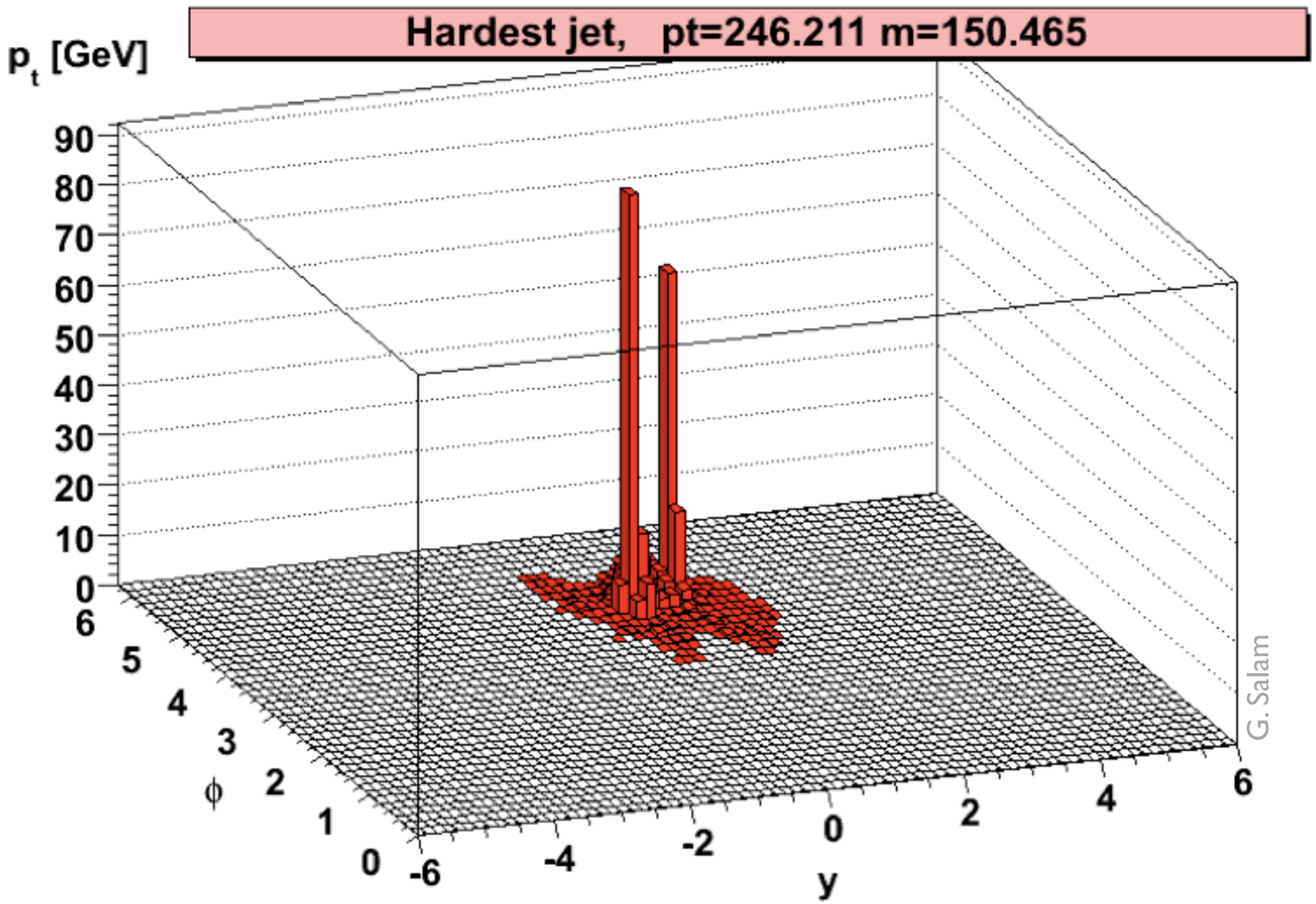
While k_t gives a 'natural relevant splitting' at its first declustering, there are a number of reasons why Cambridge/Aachen has been preferred

- k_t's 'relevant subjets' tend to include more soft radiation than needed, eventually leading to poor resolution (large areas and fluctuations)
- The angle-based clustering distance of Cambridge/Aachen ensures that at the relevant splitting the radii of the jets of the two prongs are similar to the distance between the two prongs themselves. This ensures that, because of angular ordering, these jets contain essentially all the radiations emitted by the decay products of the heavy particles (b quarks, in the case of BDRS)
- Cambridge/Aachen allows one to obtain naturally clustering sequences for any R with a single run, which is useful in the filtering step

Boosted Higgs tagger

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

Butterworth, Davison, Rubin, Salam, 2008



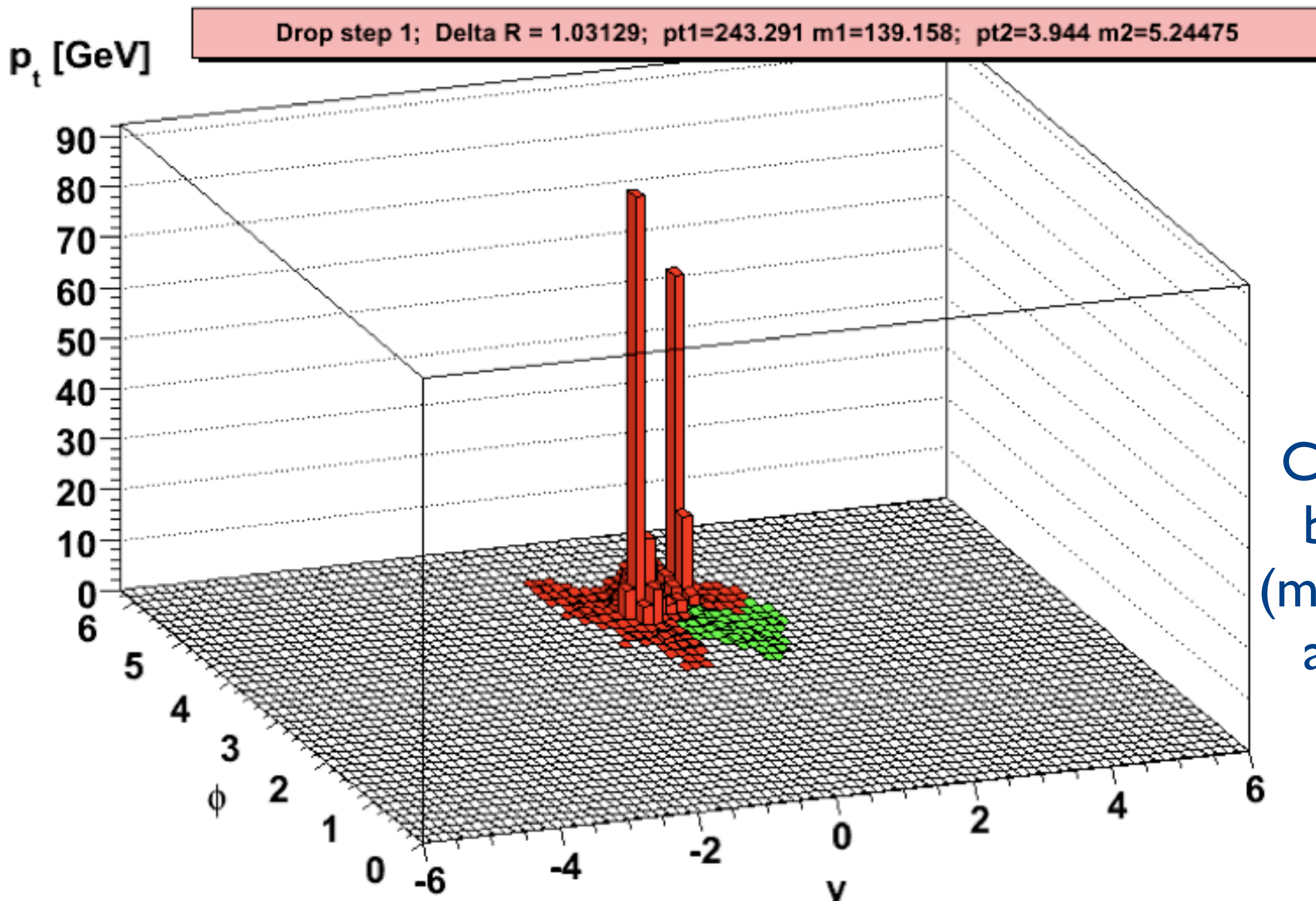
Start with the hardest jet

Use C/A with large R=1.2

m_j = 150 GeV

Boosted Higgs tagger

pp →ZH → ννbb



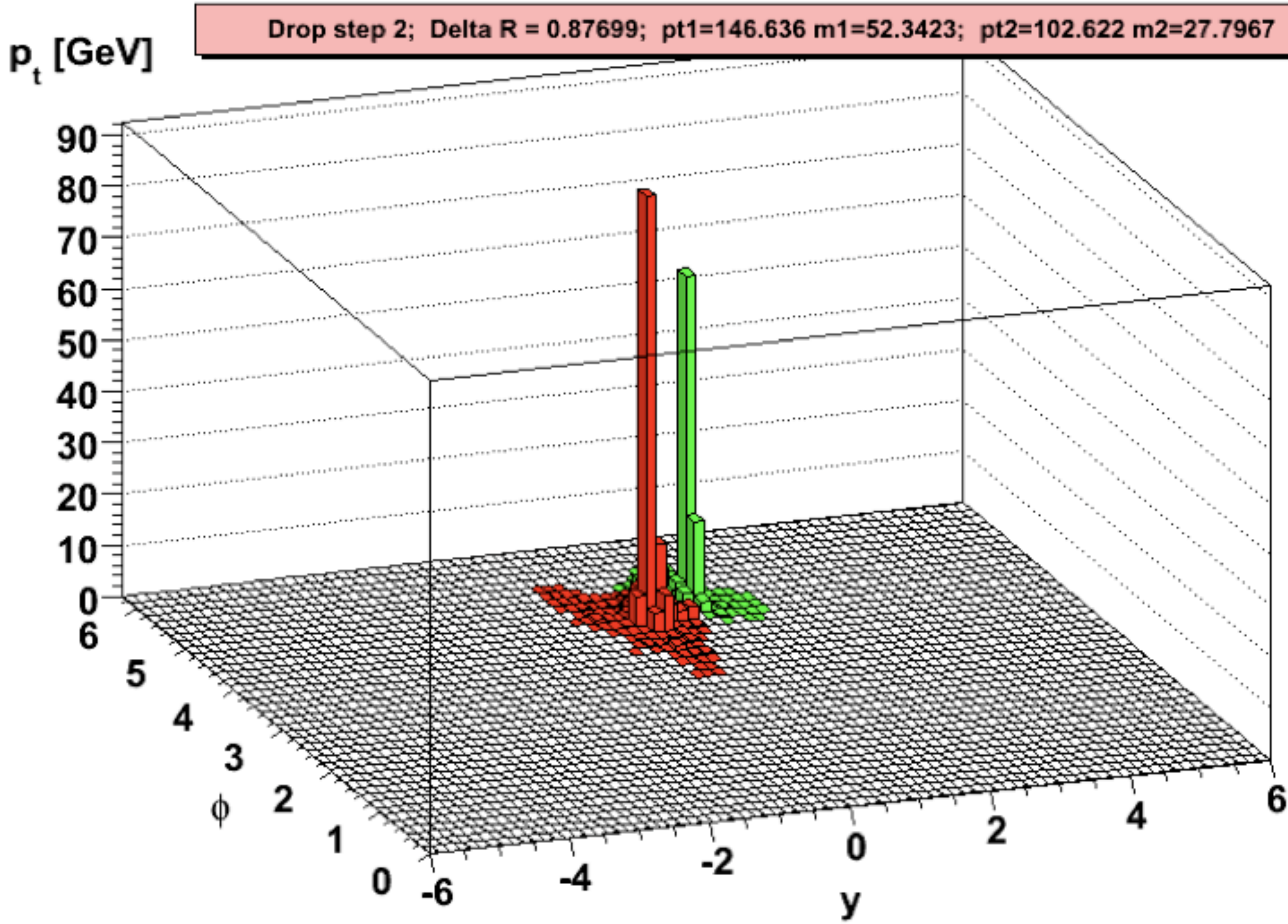
Undo last step of clustering

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

Boosted Higgs tagger

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 μ = 0.67 and y_{cut} = 0.09]

Jet substructure as filter

The ***jet substructure*** can be exploited to help **removing contamination** from a soft background

- Jet 'filtering' — Butterworth, Davison, Rubin, Salam, 2008
  Break jet into subjets at distance scale R_{filt}, retain n_{filt} hardest subjets
- Jet 'trimming' — Krohn, Thaler, Wang, 2009
  Break jet 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 $\min(p_{t1},p_{t2}) < \varepsilon_{prune}\,(p_{t1}+p_{t2})$

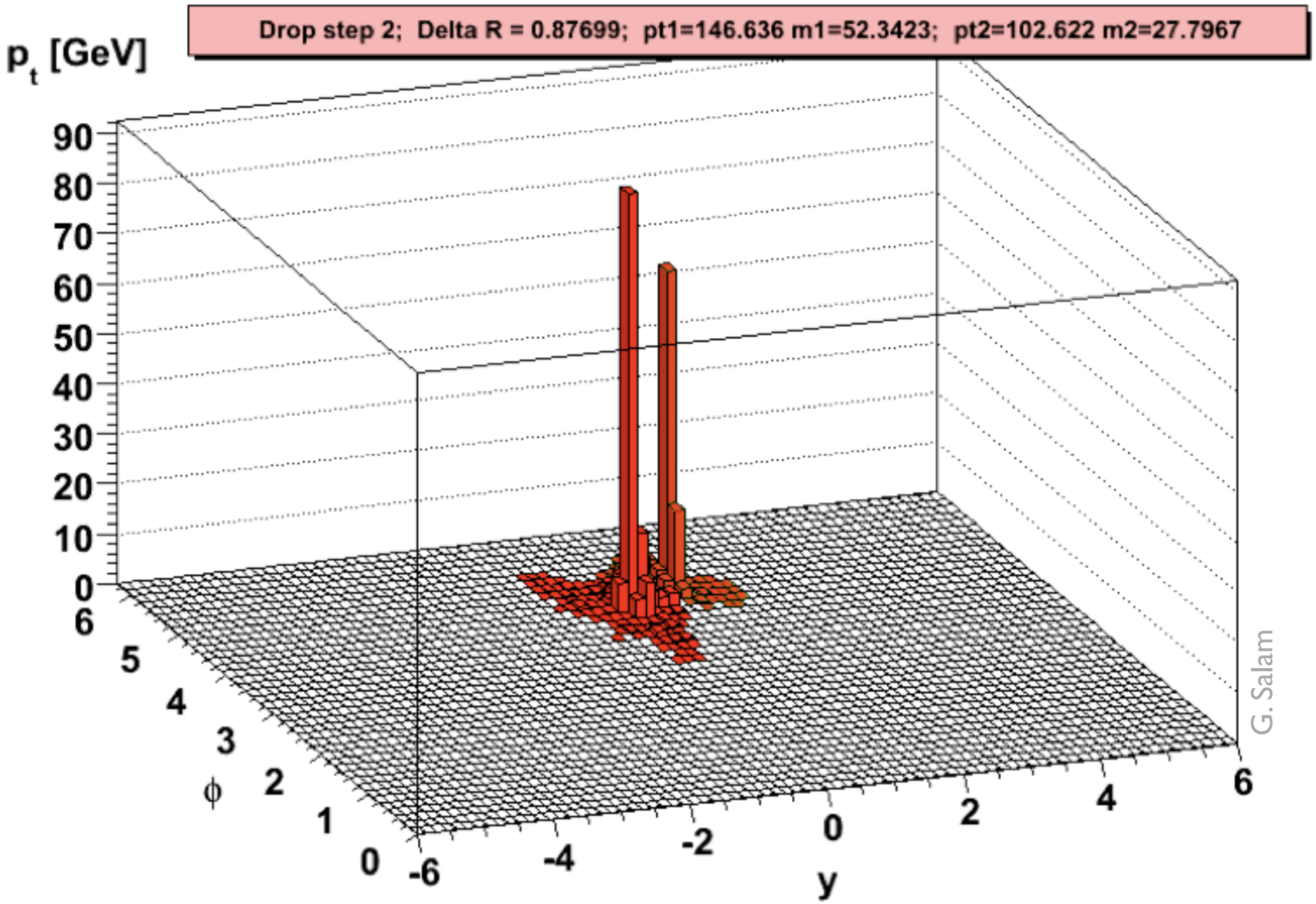
Aim: limit sensitivity to background while retaining bulk of perturbative radiation

Filtering, trimming and pruning are in the end effectively quite similar. These and similar tools are collectively called **groomers**

[Note that trimming and pruning are often also used/denoted as taggers with the meaning that they cut the background more than the signal]

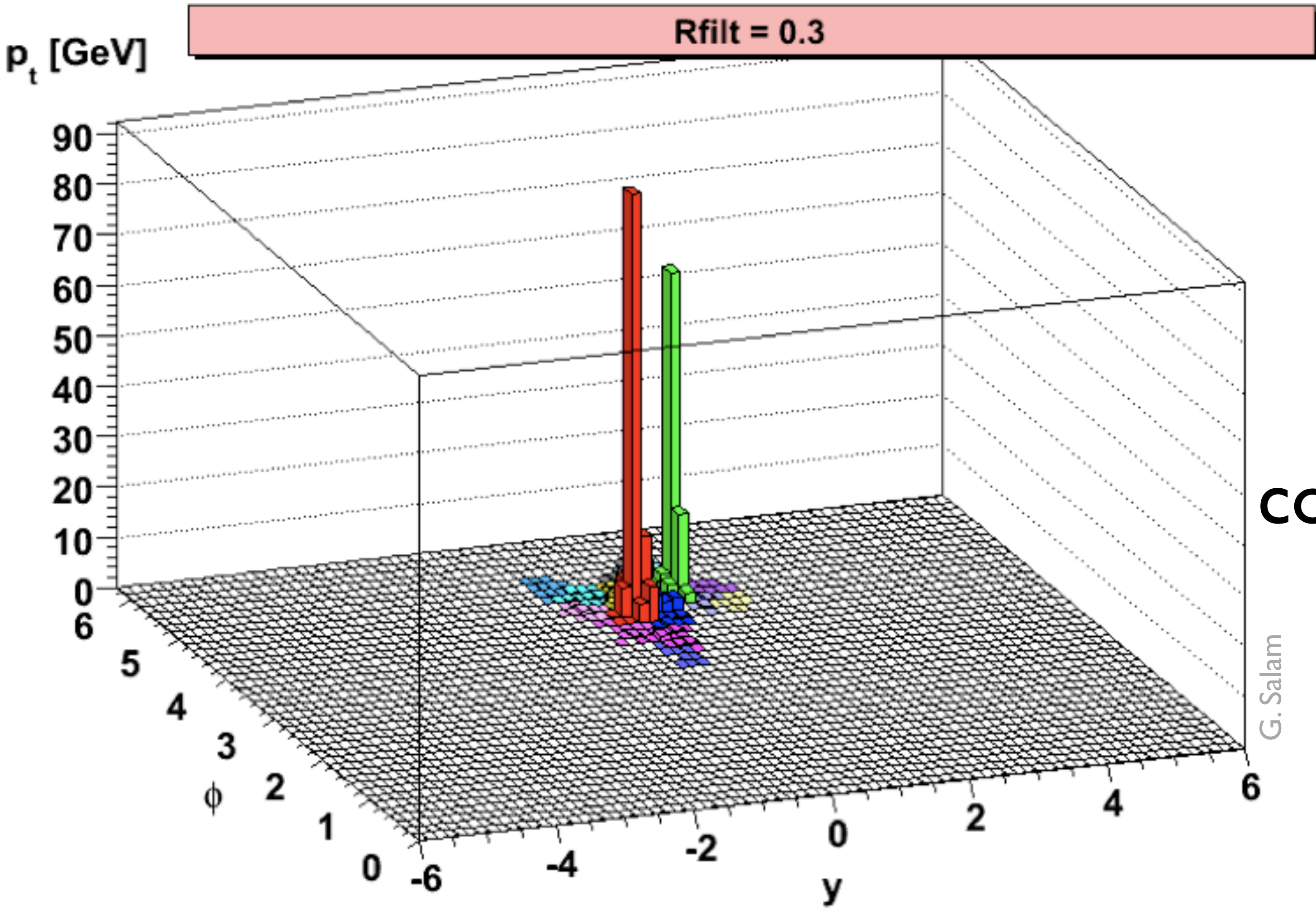
Filtering in action

Butterworth, Davison, Rubin, Salam, arXiv:0802.2470



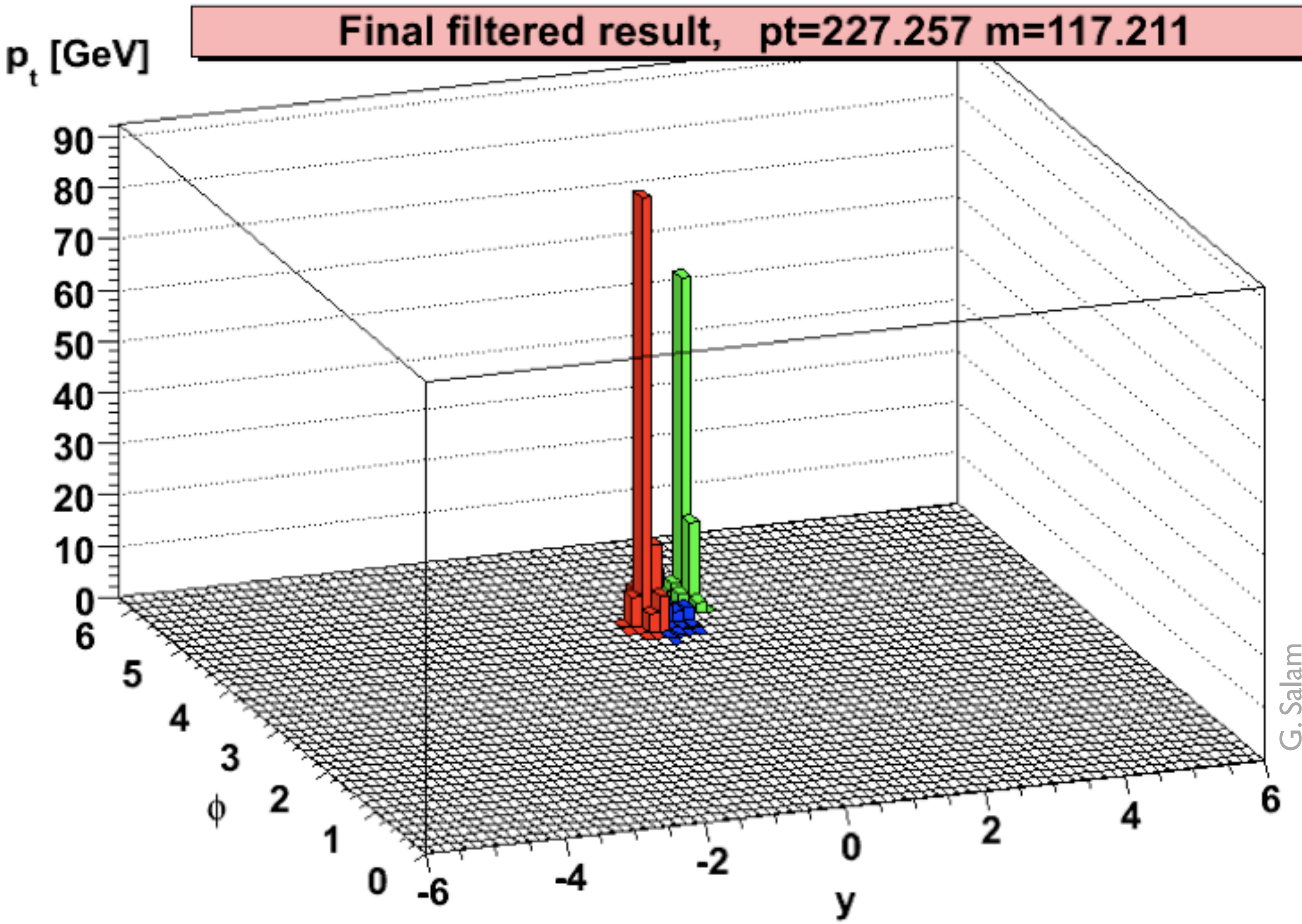
Start with a jet

Filtering in action



Recluster the contituents with R_{filt}

Filtering in action



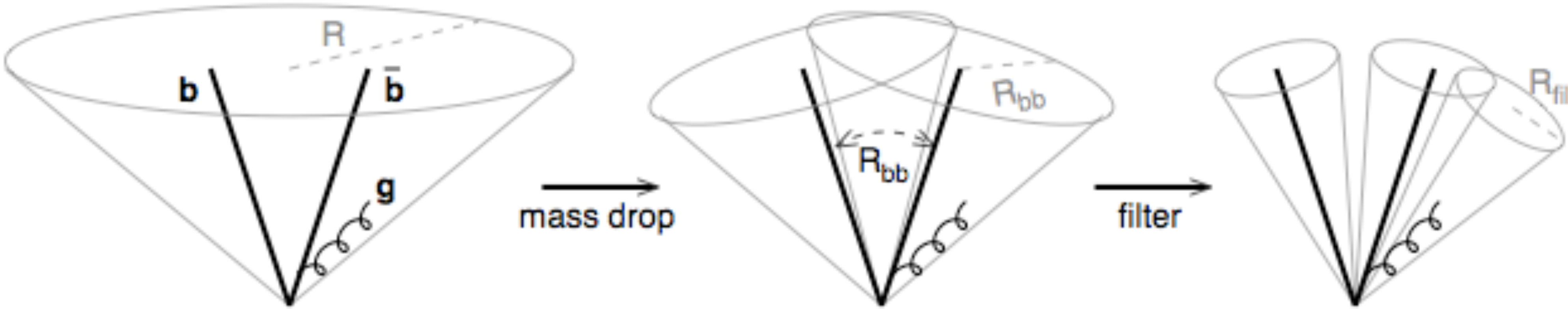
Only keep the n_{filt} hardest jets

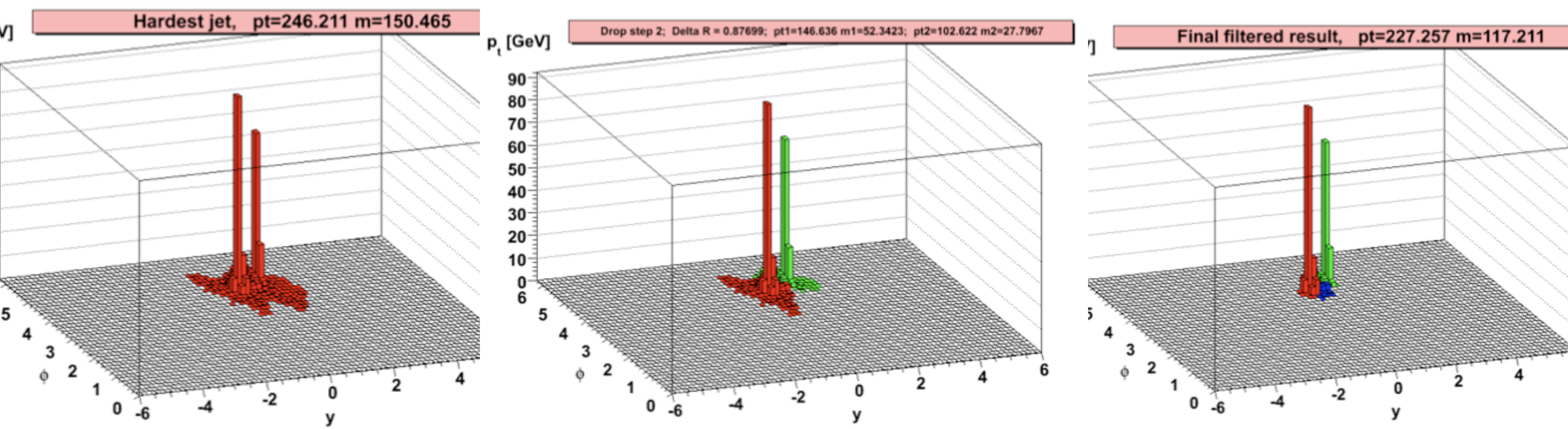
The low-momentum stuff surrounding the hard particles has been removed

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

Boosted Higgs analysis

Butterworth, Davison, Rubin, Salam, 2008





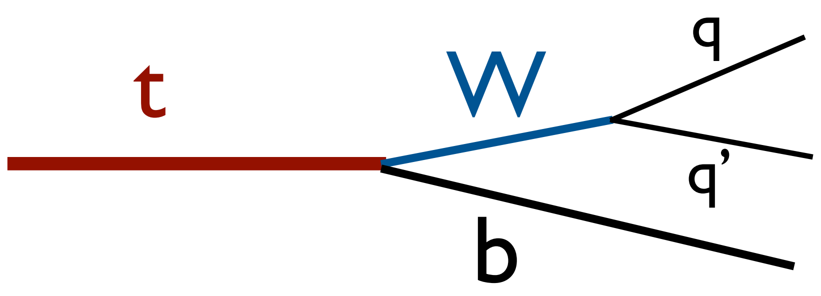
Cluster with a large R

Undo the clustering into subjets, until a large mass drop is observed

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

Top tagging

In order to tag a (boosted) top one must now identify 3-prongs structures originating from the top decay



Simplest approach: iterate declustering of k_t jets, beyond a first 'relevant splitting'

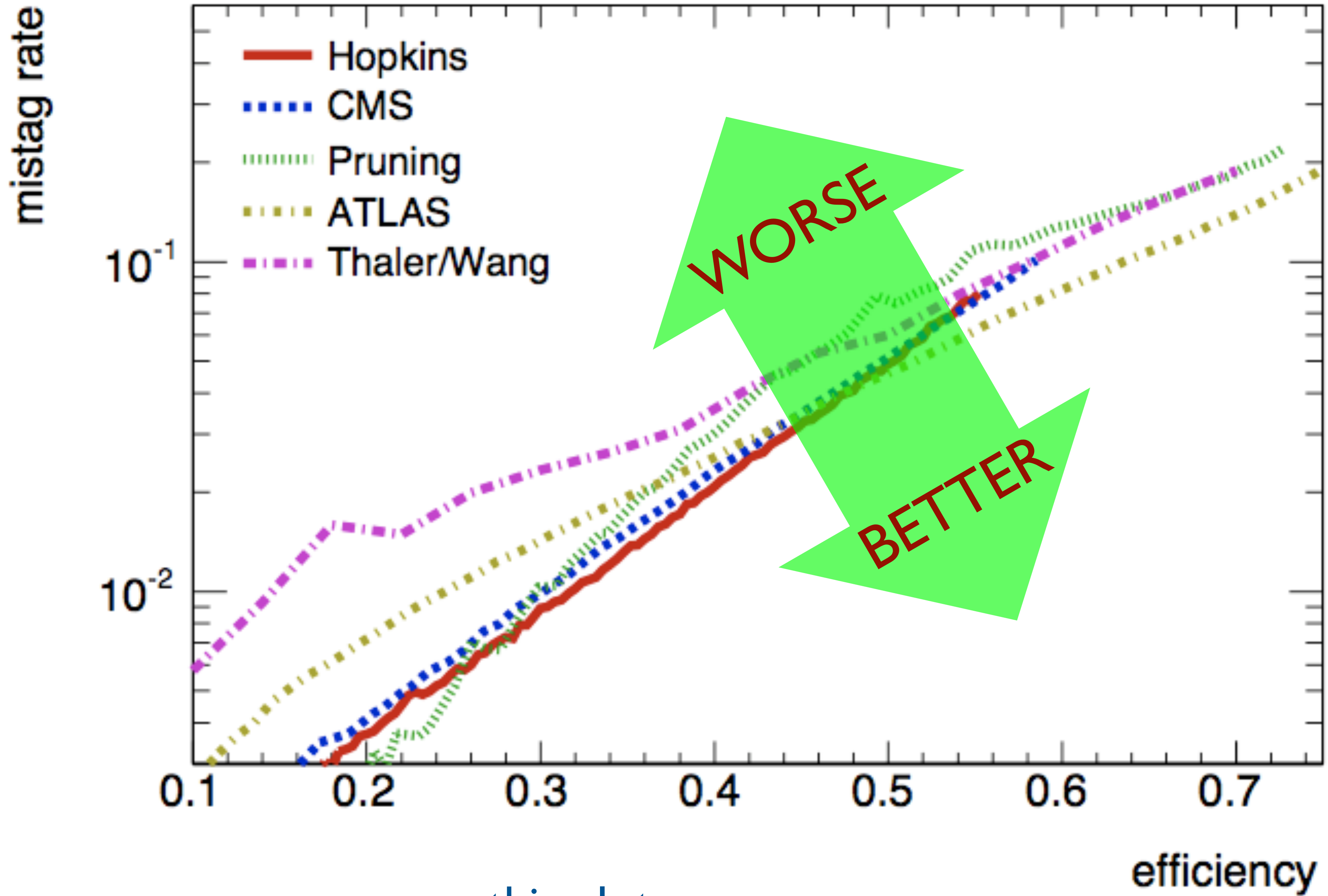
Early examples (2008):

- *ATLAS top tagger: put cuts on jet mass and d_{ij} scale*
- *Thaler-Wang: decluster to exactly 2 or 3 jets, put cuts on jet mass*

Many more top taggers after these

Comparison of top taggers

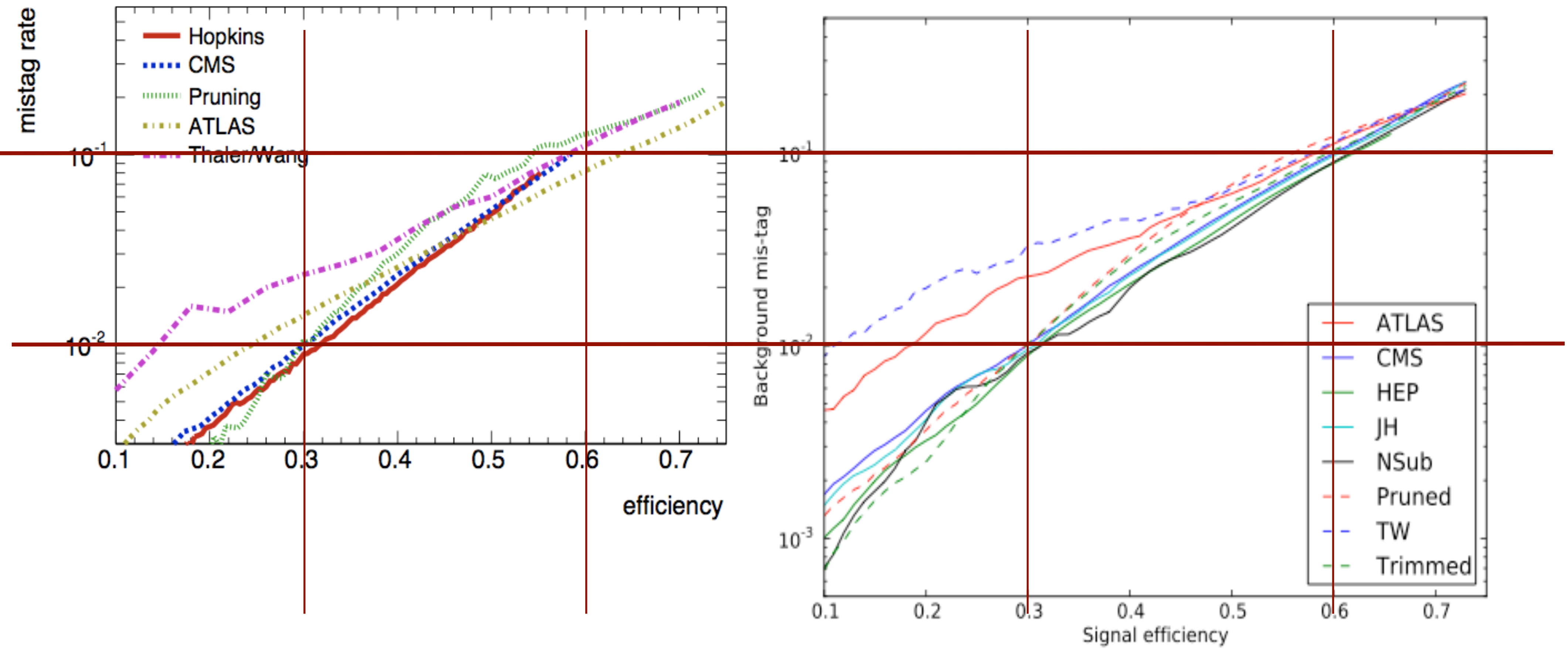
Boost 2010 proceedings, arXiv:1012.5412



Even more curves now on this plot

Comparison of top taggers

Boost **2010** proceedings, arXiv:1012.5412 Boost **2011** proceedings, arXiv:1201.0008



Law of diminishing returns: improvement has become very hard

Concluding remarks

- Proper (IRC-safe) definition of jet algorithms and efficient implementations have allowed for the study and exploitation of jet substructure properties, leading to ***taggers*** and ***groomers***

- Many new physics search strategies based on jet substructure are being explored and commissioned right now at the LHC
    - *As soon as more data (and more boosted particles) are available, we should see the first results from them*

- Many of these tools are mature and are being refined, but one can hope that new radical ideas will fuel another 'revolution'