Collider Phenomenology 2011
Cambridge, April 2011

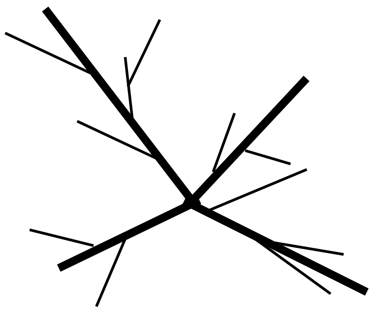
Jet definitions and data

Matteo Cacciari
LPTHE - Paris 6,7 and CNRS

Many thanks to Gavin Salam and Gregory Soyez for the extensive ongoing collaboration on this topic

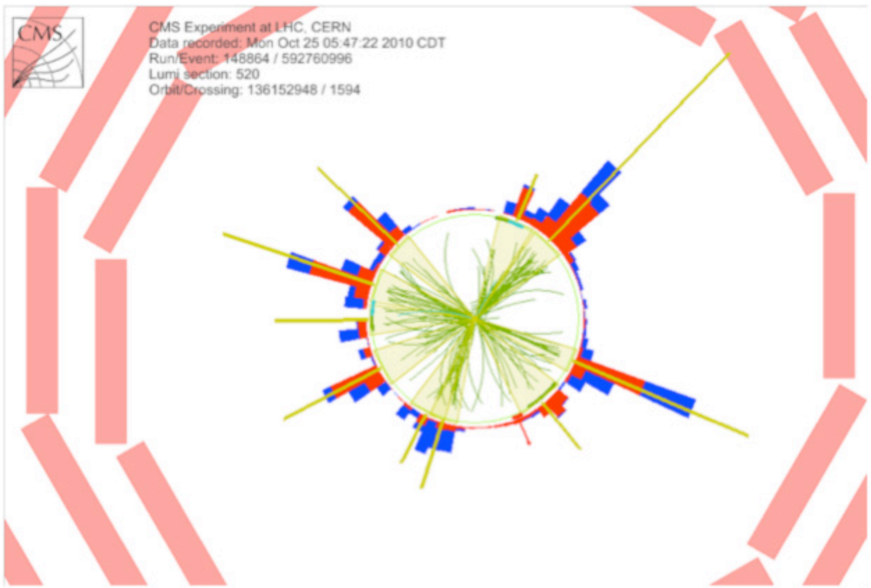
Taming reality

Multileg + PS



??

QCD predictions



Real data

Jets

One purpose of a ‘jet clustering’ algorithm is to **reduce the complexity** of the final state, simplifying many hadrons to **simpler objects** that one can hope to **calculate**

Jet Definition

A jet algorithm

$$\{p_i\} \rightarrow \{j_k\}$$

particles, calo cells, ...   jets

+

its **parameters** (e.g. R)

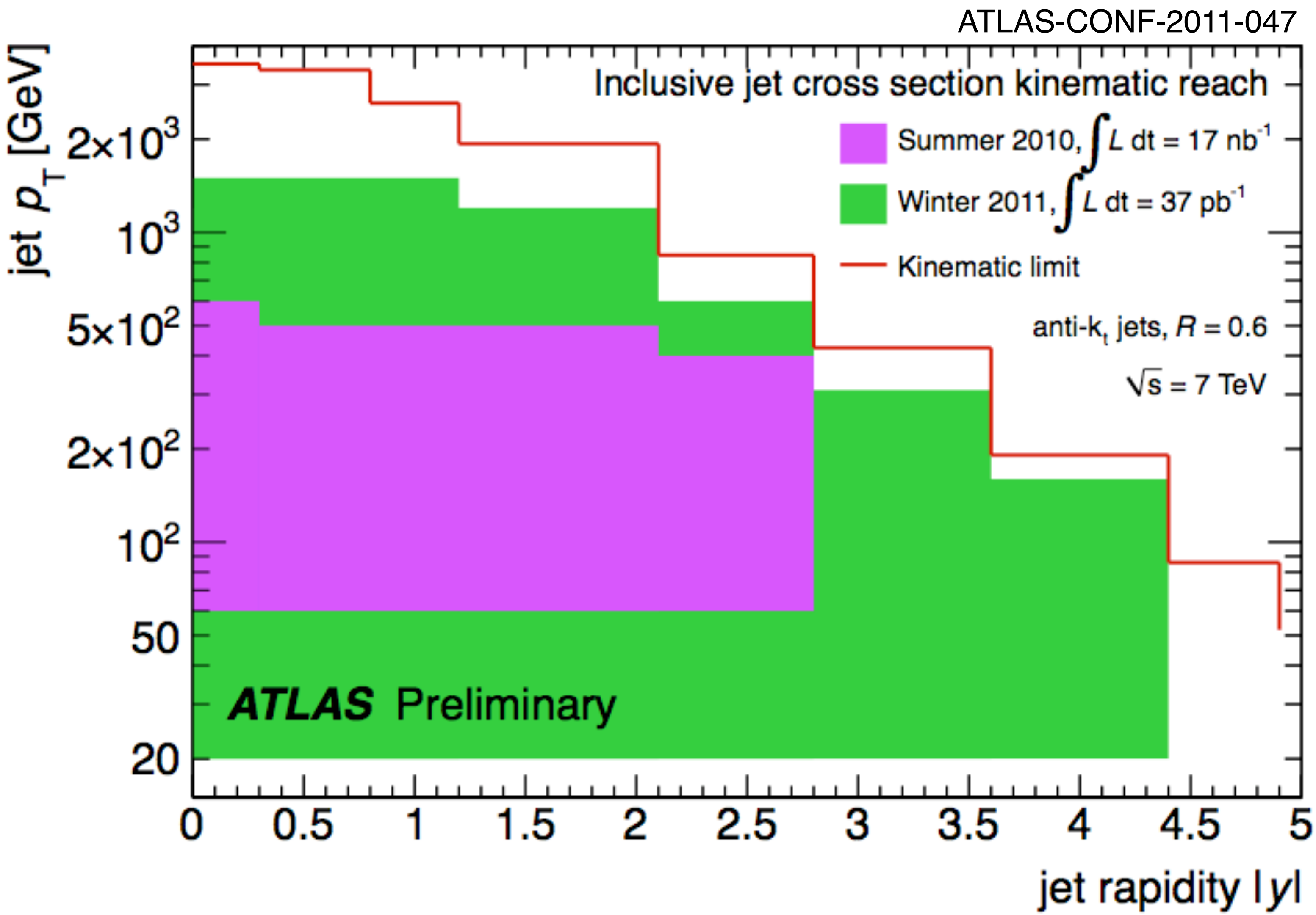
+

a **recombination scheme**

=

a **Jet Definition**

Phase space of jet data

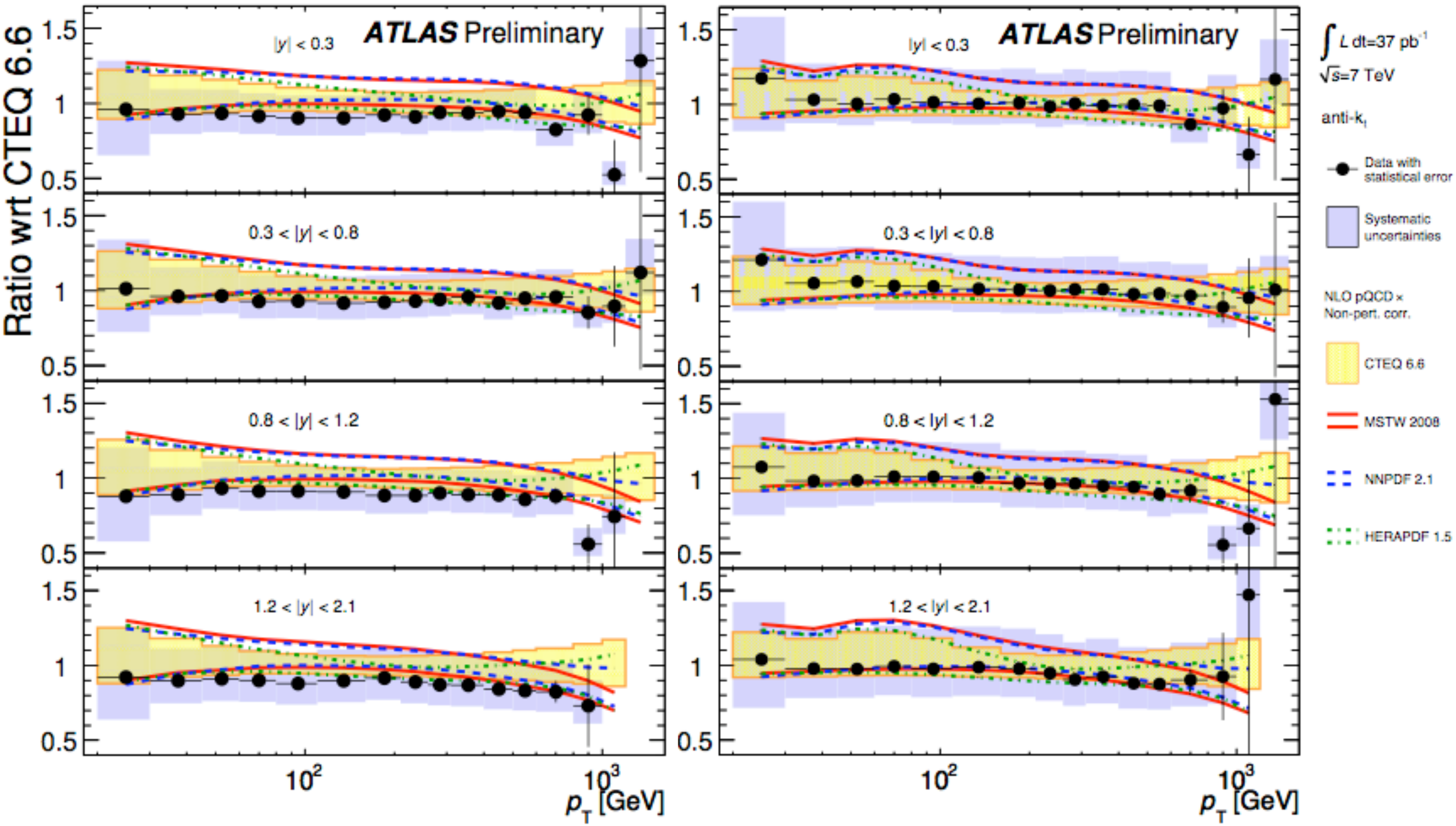


ATLAS inclusive jets: central rapidity

anti-k_t

ATLAS-CONF-2011-047

R = 0.4　　　　R = 0.6

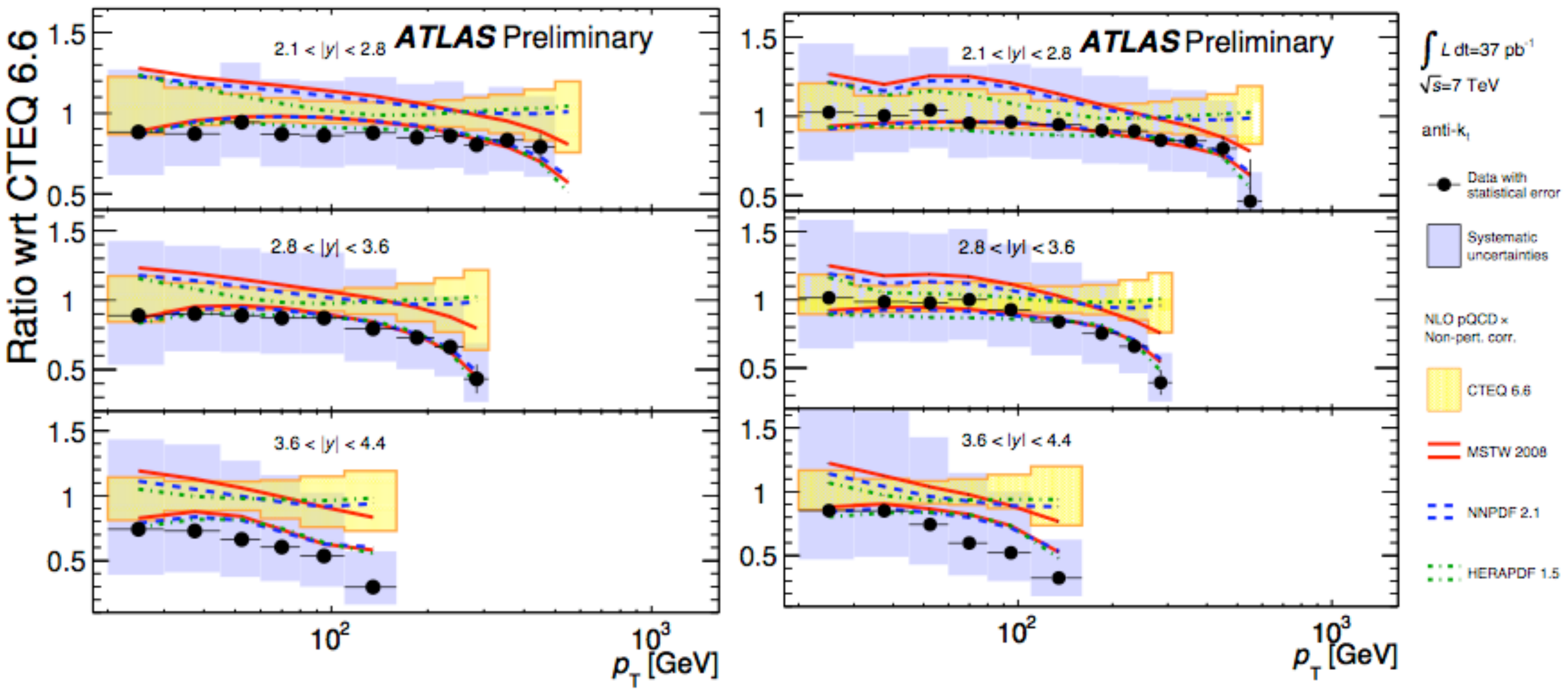


ATLAS inclusive jets: forward rapidity

anti-k_t

ATLAS-CONF-2011-047

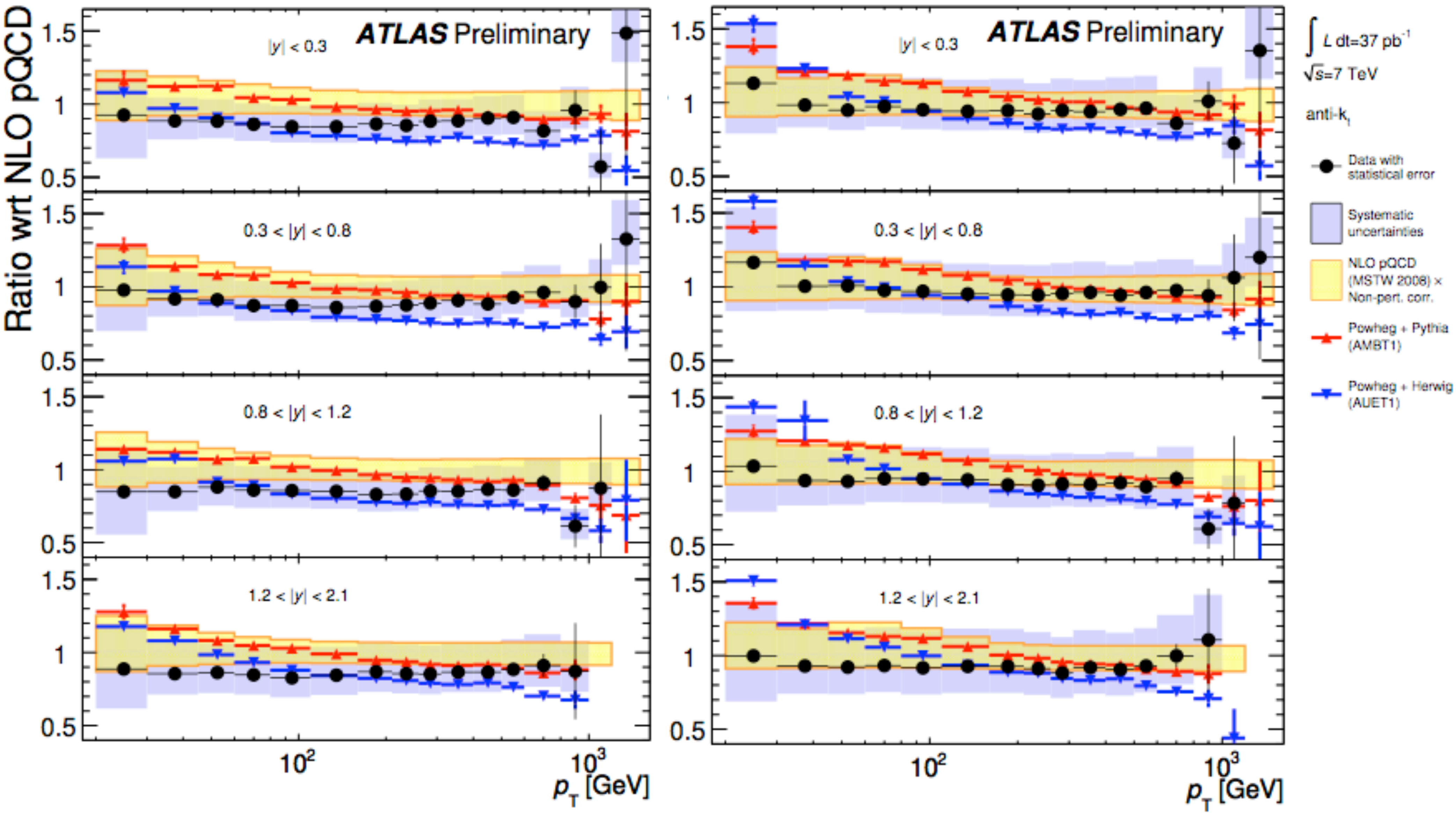
R = 0.4 R = 0.6



ATLAS inclusive jets: central rapidity

anti-k_t ATLAS-CONF-2011-047

R = 0.4 R = 0.6

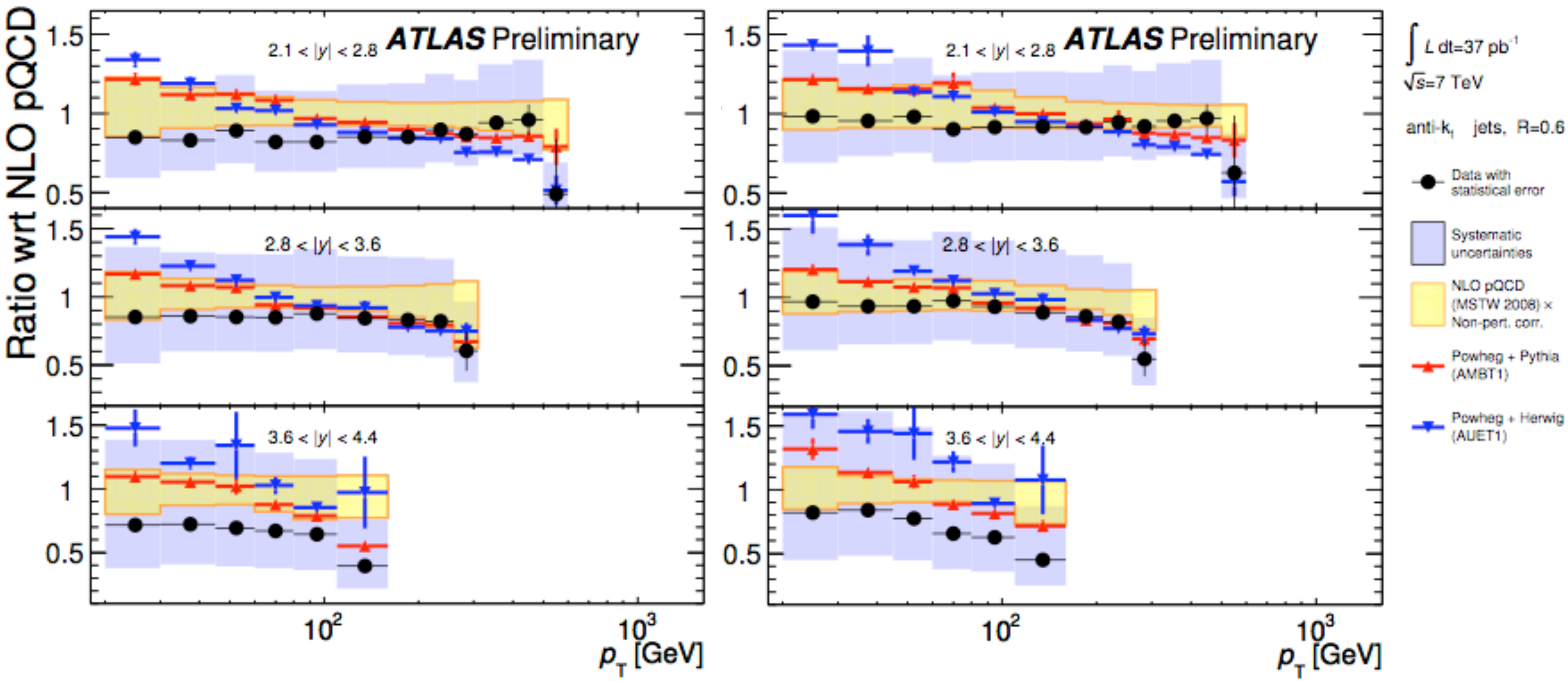


ATLAS inclusive jets: forward rapidity

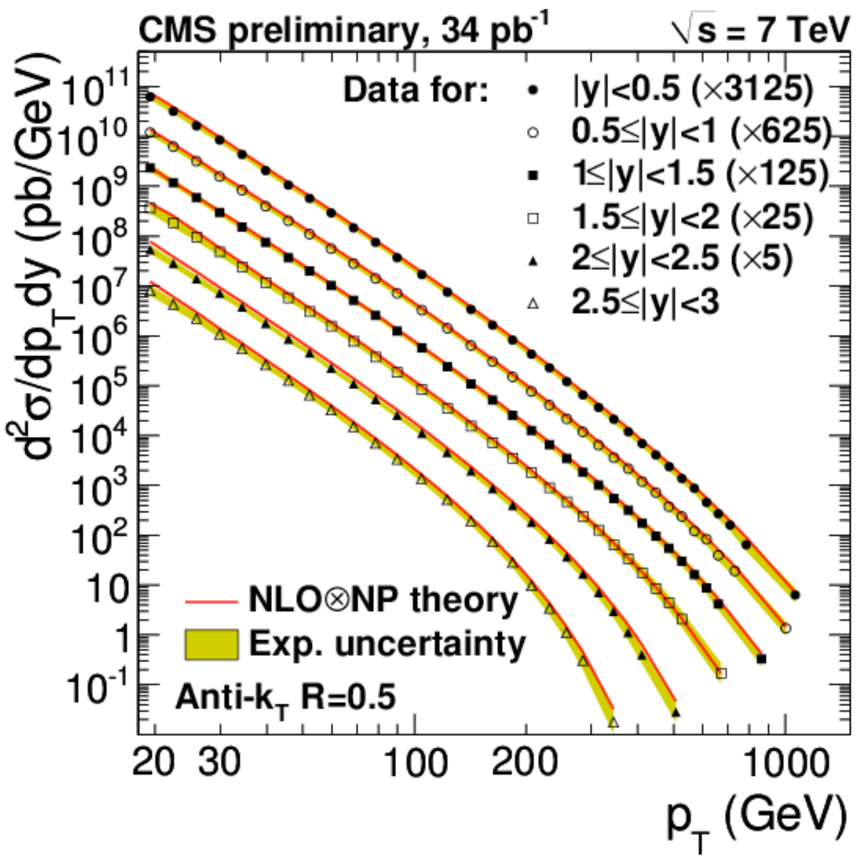
anti-k_t

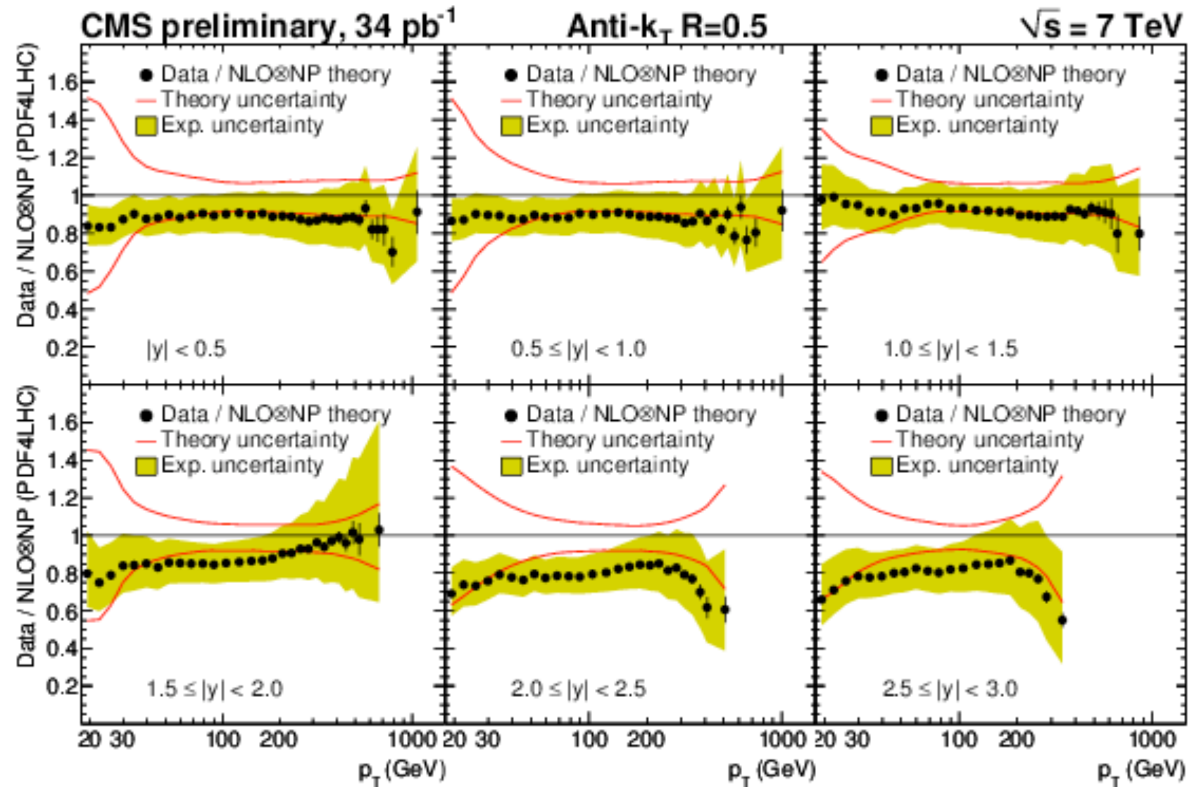
ATLAS-CONF-2011-047

R = 0.4 R = 0.6



CMS inclusive jets



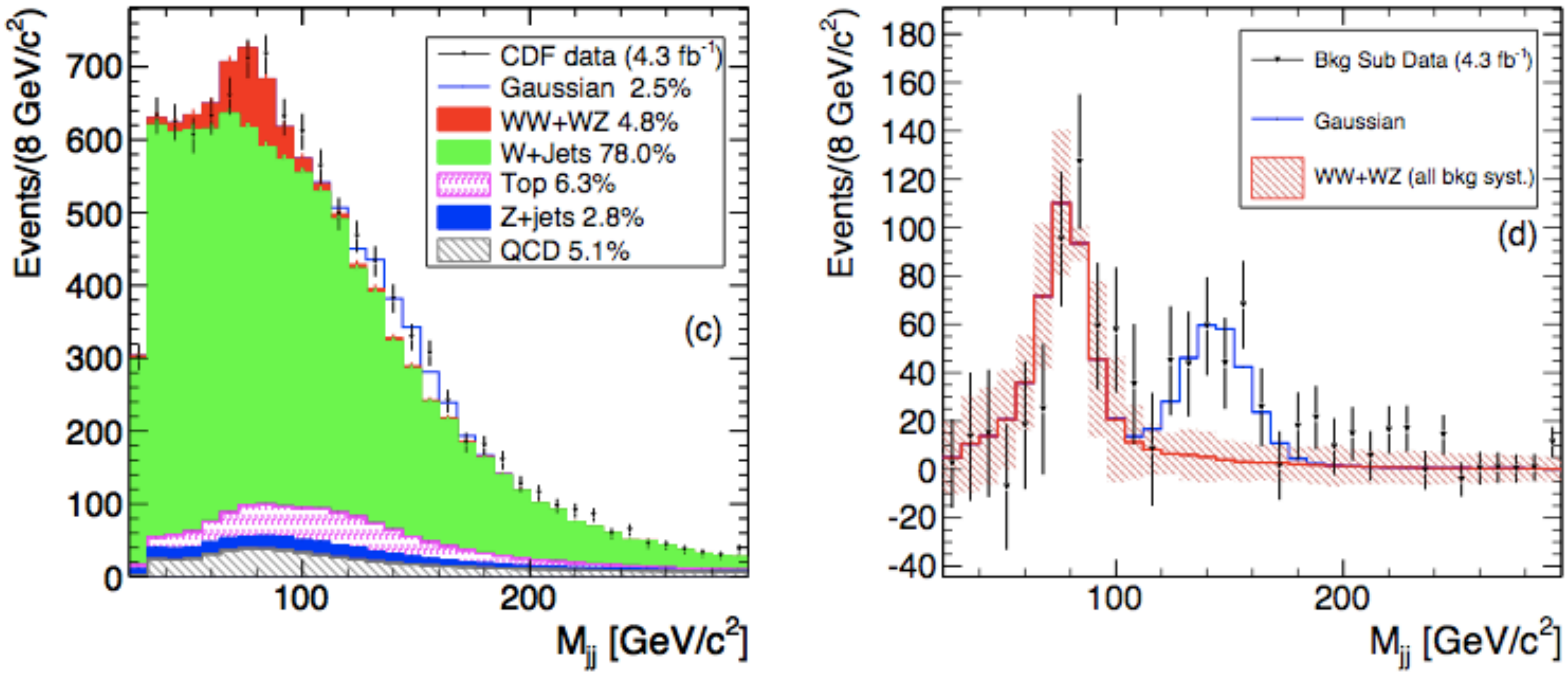


- CMS's standard jet definitions are anti-k_t with R=0.5 and R=0.7 (ensuring that the overlap with ATLAS is zero!)
- Results with R=0.5 appear similar to ATLAS's 0.4/0.6
- R=0.7 would again probably not be significantly different in this case

CDF Wjj excess

Speaking of jet definitions.....

arXiv:1104.0699



Background evaluated with JetClu (infrared-unsafe already at the 2+1level):
LO W+2j already tricky, no NLO calculation formally possible

R=0.4 v. R=0.5 v. R=0.6

- ATLAS reports fairly good agreement of data with NLO calculation, especially with HERAPDF and MRSTW2008 PDFs
  - It also reports some differences between POWHEG predictions with PYTHIA and HERWIG parton showers
  - the two jet definitions used (R=0.4 and R=0.6) give quite similar results: no additional discriminating power from these two choices (or at least not yet)
- CMS adds R=0.5 anti-k_t jets: no surprises there either

Where does R matter?

Jets as tools

Background characterisation and subtraction

Mass reconstruction

Remove soft contamination from a hard jet

Tag heavy objects originating the jet

Jets as tools

Background characterisation and subtraction

Mass reconstruction

Remove soft contamination from a hard jet

Tag heavy objects originating the jet

R-dependent effects

Perturbative radiation: $$\Delta p_t \simeq \frac{\alpha_s(C_F, C_A)}{\pi} p_t \ln R$$

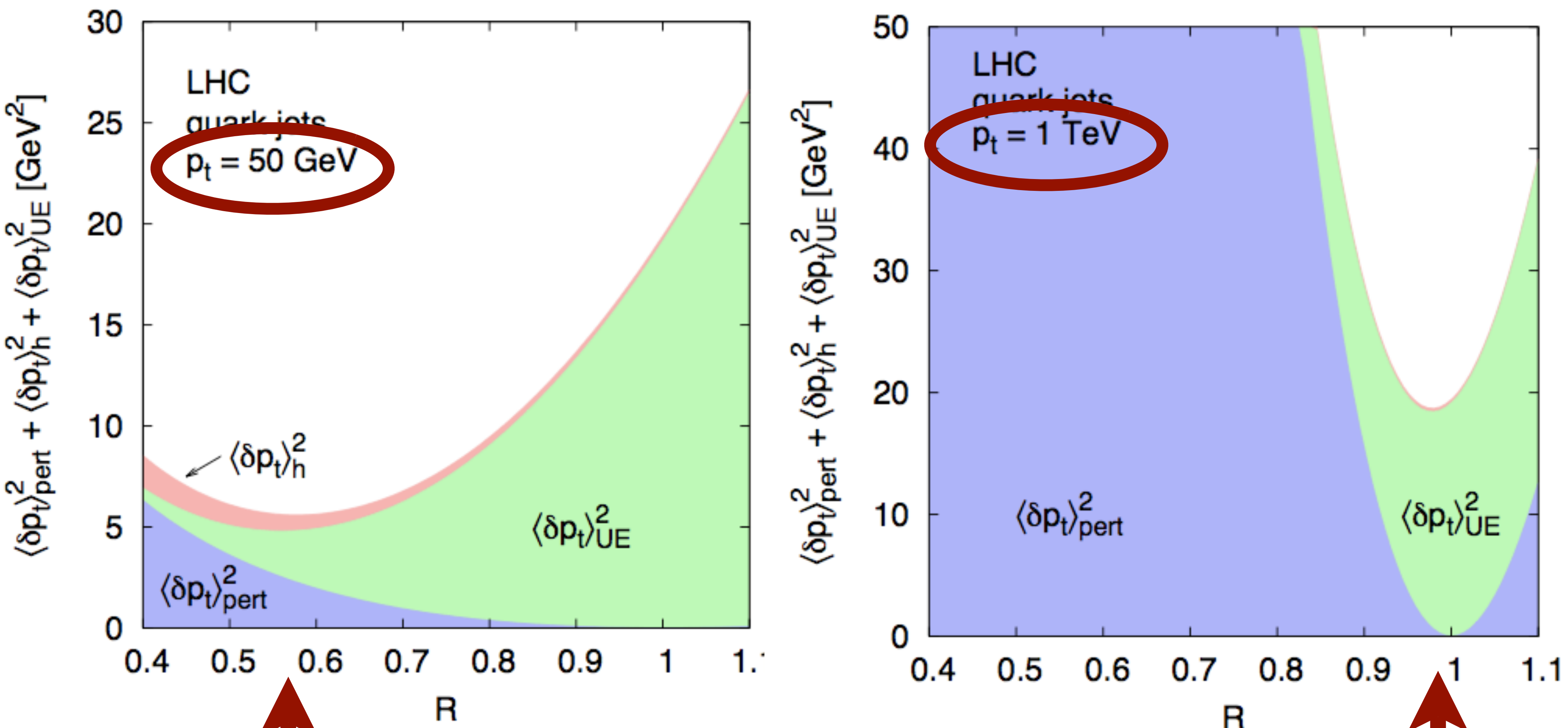
Hadronisation: $$\Delta p_t \simeq \frac{(C_F, C_A)}{R} \times 0.4 \text{ GeV}$$

Underlying Event: $$\Delta p_t \simeq \frac{R^2}{2} \times (2.5 — 15 \text{ GeV})$$

Tevatron LHC

Analytical estimates:
Dasgupta, Magnea, Salam, arXiv:0712.3014
G. Soyez, arXiv:1006.3634

Minimize $\Sigma(\Delta p_t)^2$

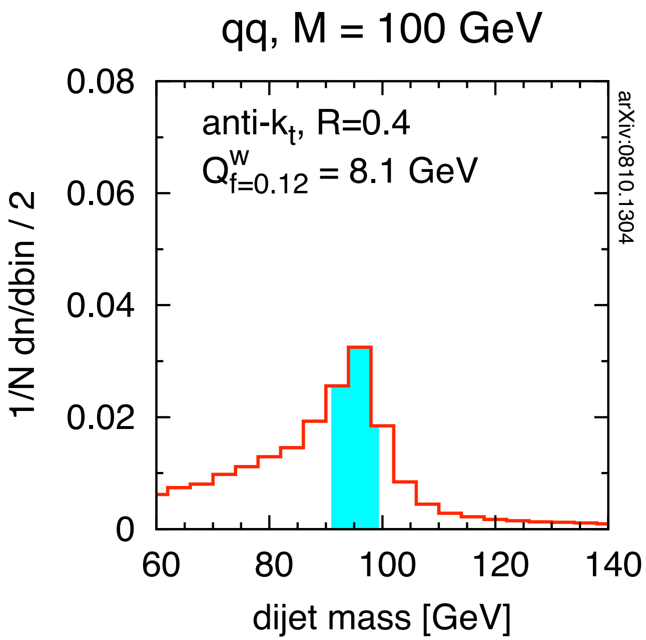


Best R

Best R

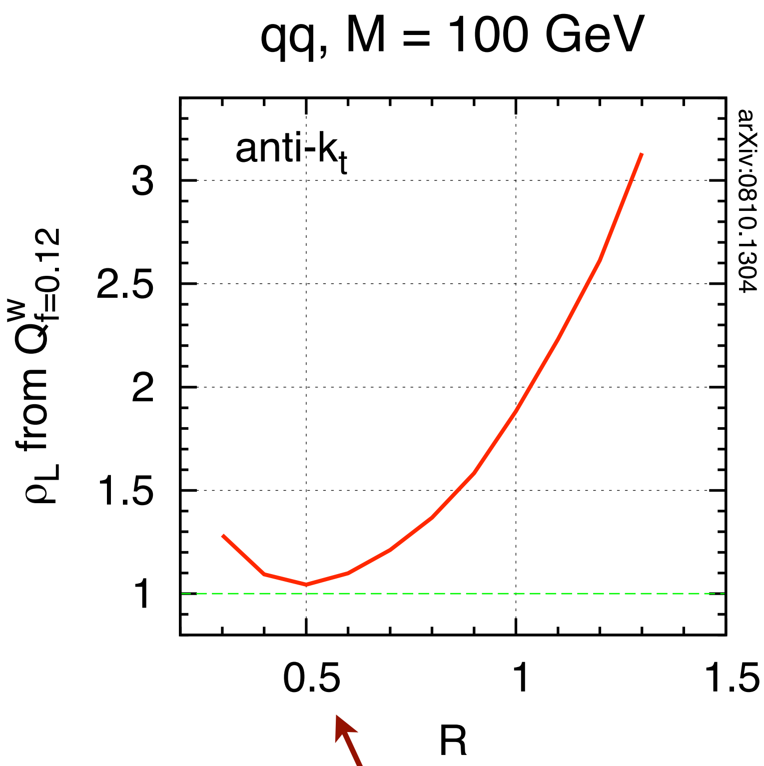
Dasgupta, Magnea, Salam, arXiv:0712.3014

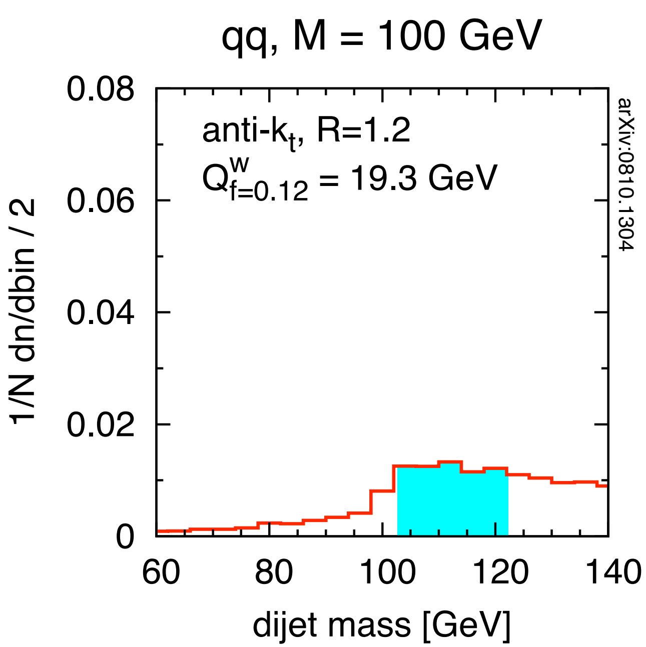
100 GeV qq mass peak



R = 0.4

All R

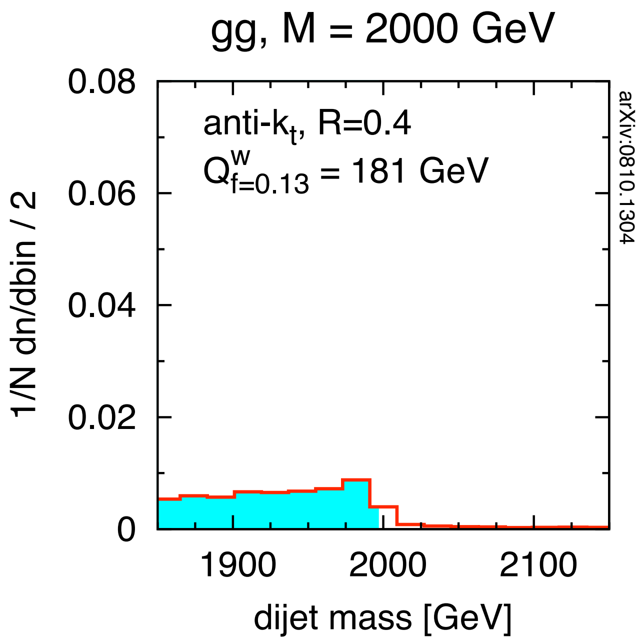




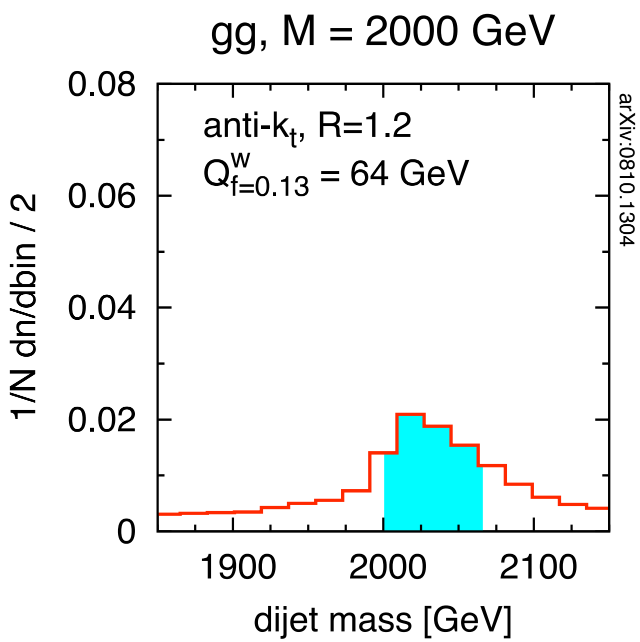
R = 1.2

Best R

2 TeV gg mass peak

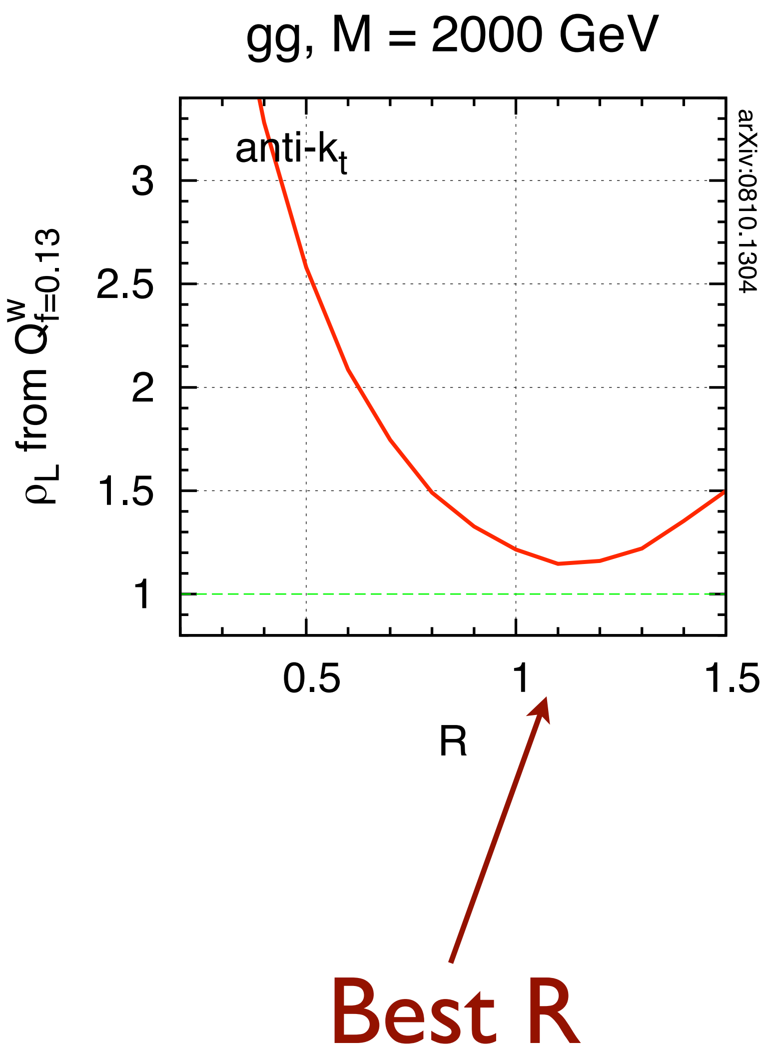


R = 0.4



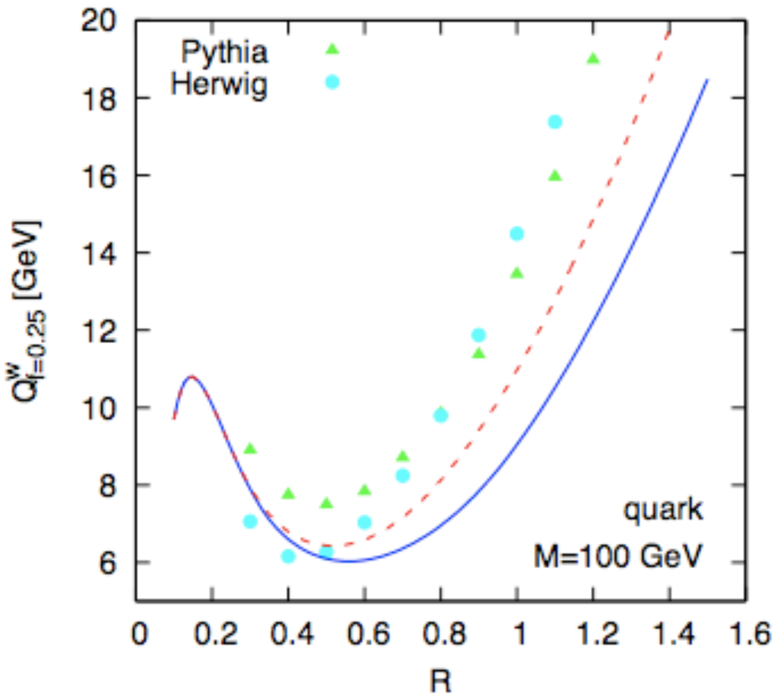
R = 1.2

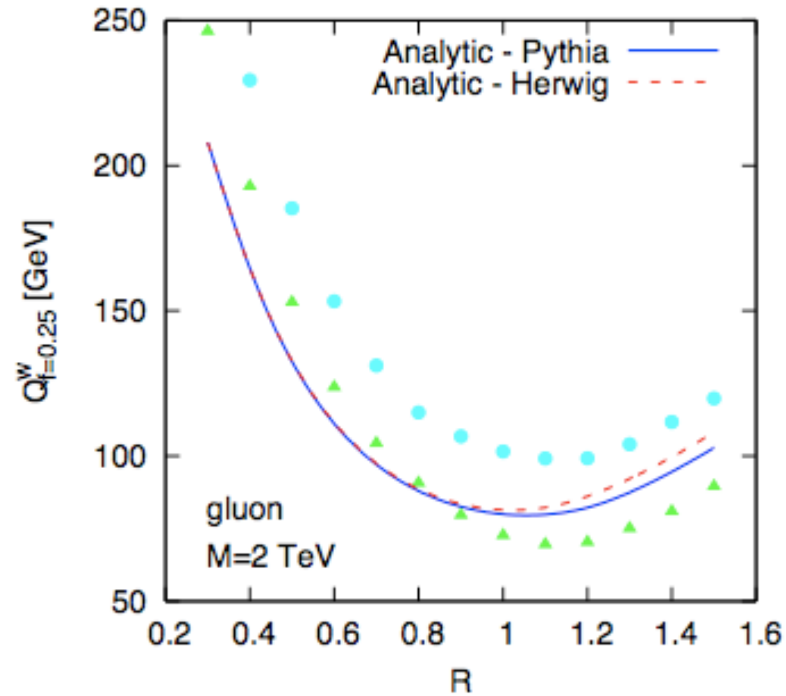
All R

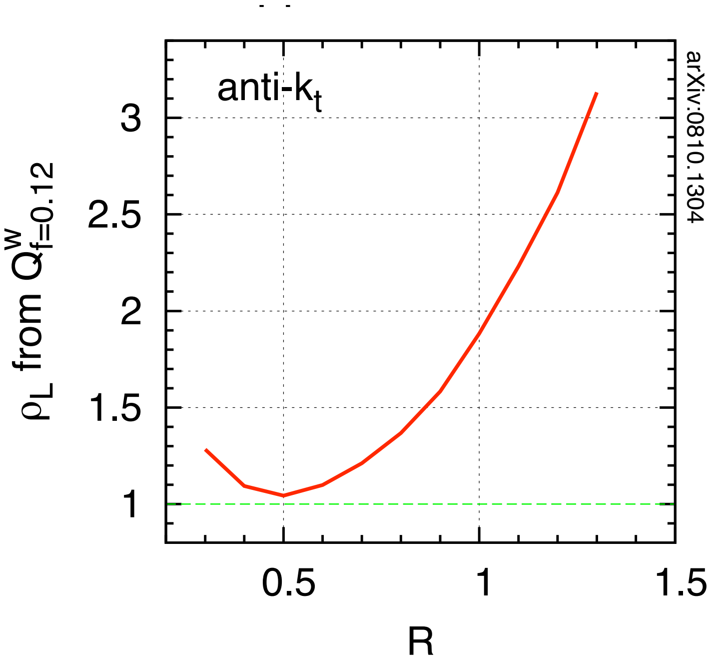


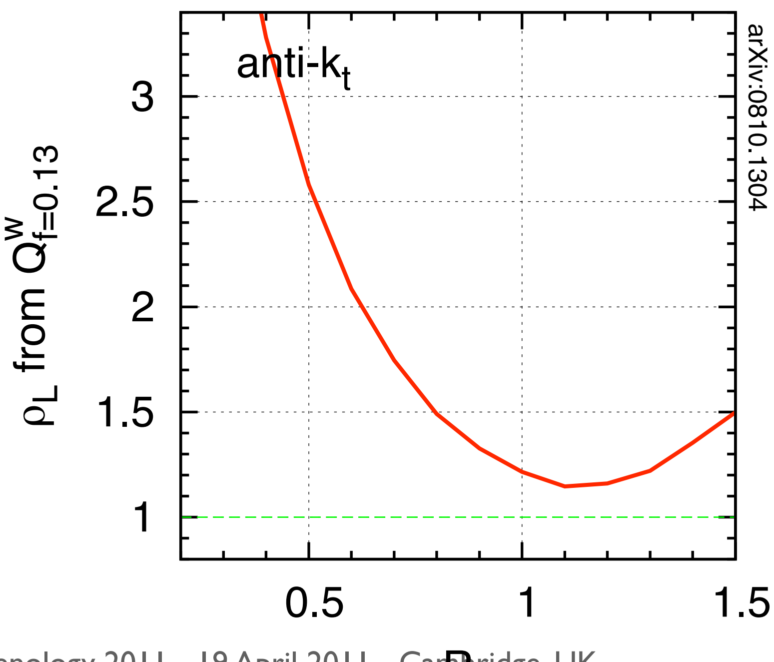
Analytical estimates for best R

G. Soyez, arXiv:1006.3634





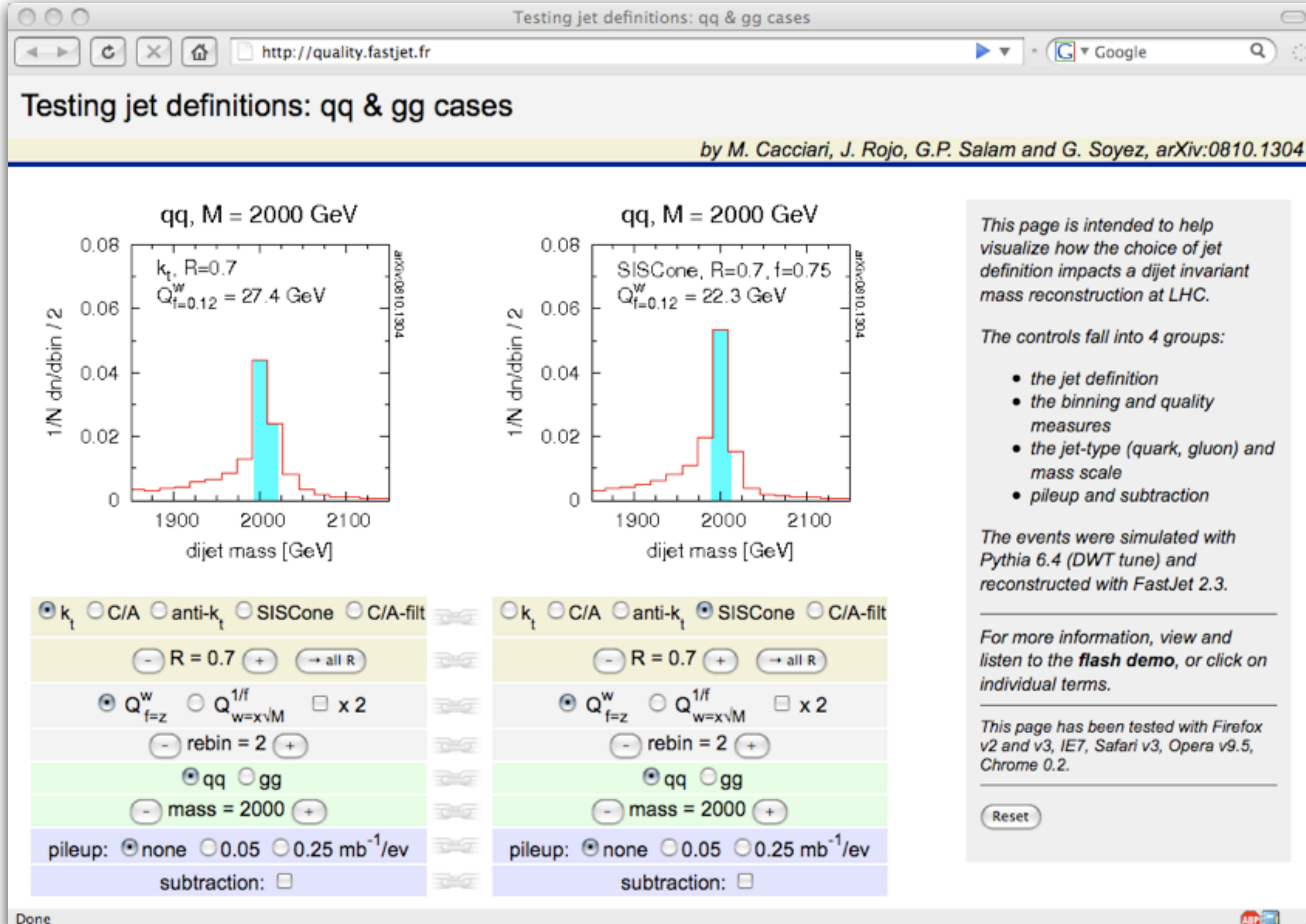




Good agreement with Monte Carlo results

http://quality.fastjet.fr

Do it yourself



Not only anti-k_t

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

p = 1 k_t algorithm

S. Catani, Y. Dokshitzer, M. Seymour and B. Webber, Nucl. Phys. B406 (1993) 187
S.D. Ellis and D.E. Soper, Phys. Rev. D48 (1993) 3160

p = 0 Cambridge/Aachen algorithm

Y. Dokshitzer, G. Leder, S.Moretti and B. Webber, JHEP 08 (1997) 001
M.Wobisch and T.Wengler, hep-ph/9907280

p = -1 **anti-k_t algorithm**

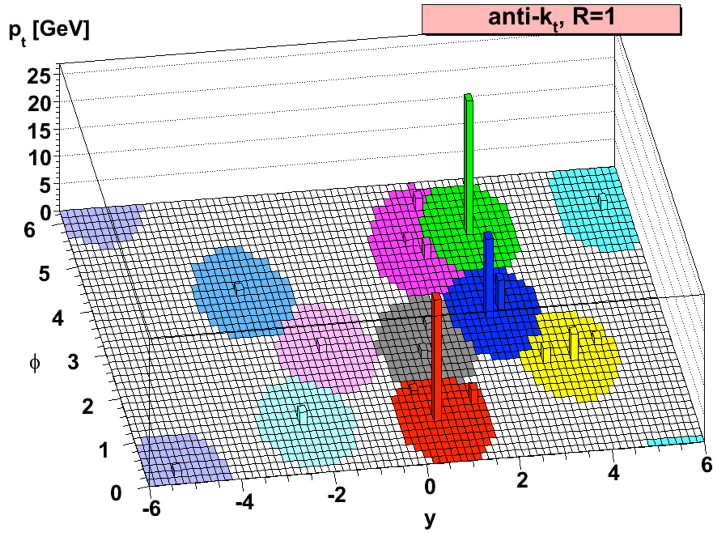
MC, G. Salam and G. Soyez, arXiv:0802.1189

NB: in anti-kt pairs with a **hard** particle will cluster first: if no other hard particles are close by, the algorithm will give **perfect cones**

Quite ironically, a sequential recombination algorithm is the 'perfect' cone algorithm

Why anti-k_t, and why it's not enough

- anti-k_t was quickly adopted by all the LHC collaborations because of a number of useful properties, chief among them the regularity of its hard jets



- It also has other desirable characteristics, for instance the very limited back-reaction, useful in HI studies.

- However, it also has drawbacks, for instance a distribution of areas that make it unsuitable for background estimation and the absence of a hierarchical substructure

Let us see these properties in some detail, and argue that it is important to be able to go beyond the anti-k_t 0.4 - 0.7 jet definitions

Jets as tools

Background characterisation and subtraction

Mass reconstruction

Remove soft contamination from a hard jet

Tag heavy objects originating the jet

How are the hard jets modified by the background?

Susceptibility (how much bkgd gets picked up)

Resiliency (how much the original jet changes)

Hard jets and background

MC, Salam, arXiv:0707.1378
MC, Salam, Soyez, arXiv:0802.1188

Modifications of the hard jet

$$p_{t,jet}^{AA} = p_{t,jet}^{pp} + \boxed{\rho A_{jet} \pm \sigma \sqrt{A_{jet}}} + \boxed{\Delta p_t^{BR}}$$

hard jet — background — back-reaction

background: 'susceptibility'

back-reaction: 'resiliency'

Background subtraction

A fairly uniform soft background can be **subtracted** from a jet momentum by doing

$$p_{\mu,jet}^{sub} = p_{\mu,jet}^{raw} - \rho A_{\mu,jet}$$

jet area: susceptibility to contamination

characterizes the background

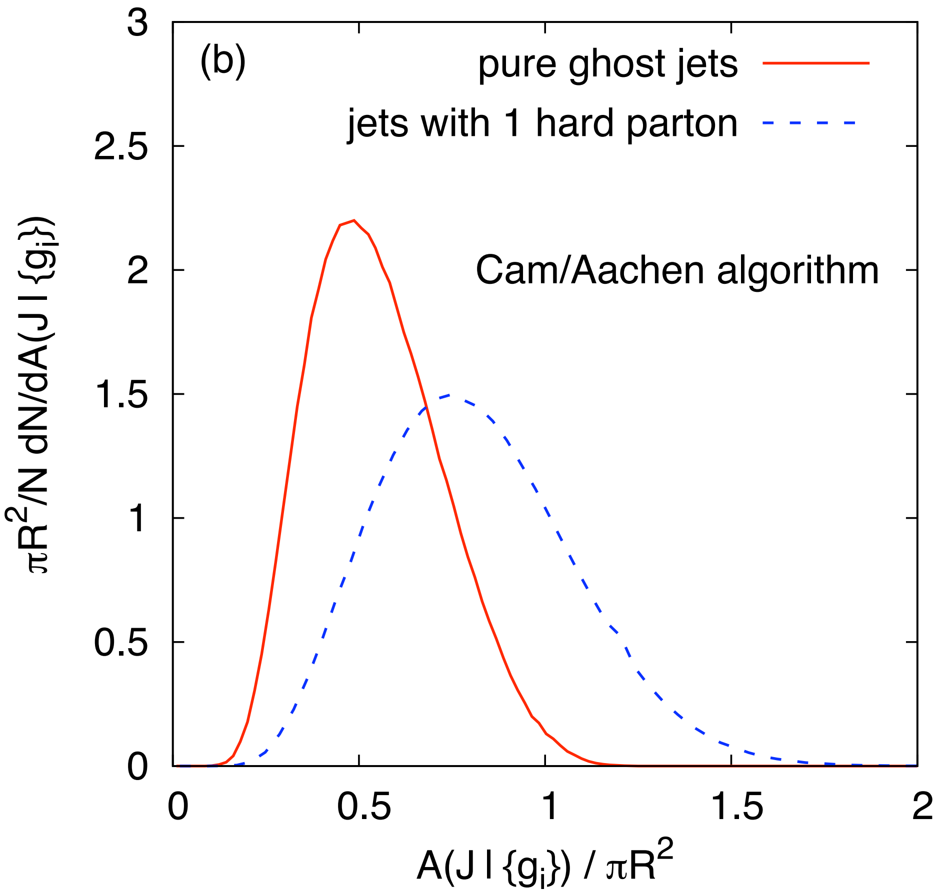
where

$$\rho \equiv \underset{\text{(over a single event)}}{\text{median}} \left[\left\{ \frac{p_t^{jet}}{\text{Area}_{jet}} \right\} \right]$$

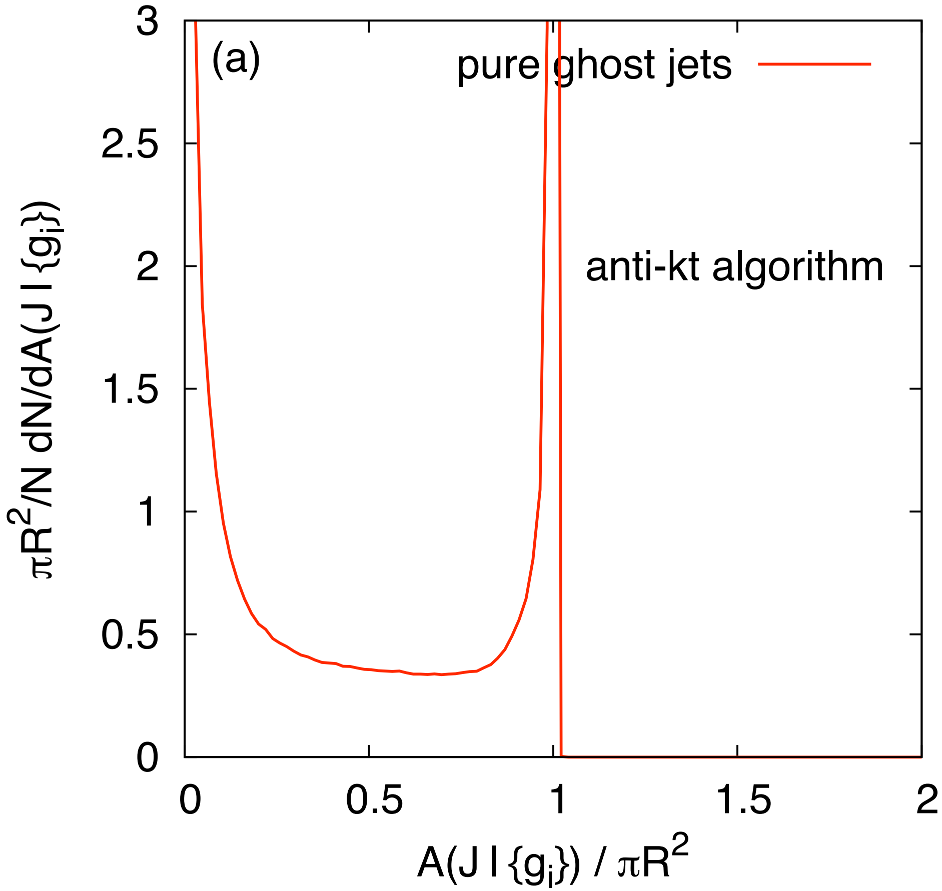
MC, Salam, 2007

Area distributions

k_t and Cam/AA



anti-k_t



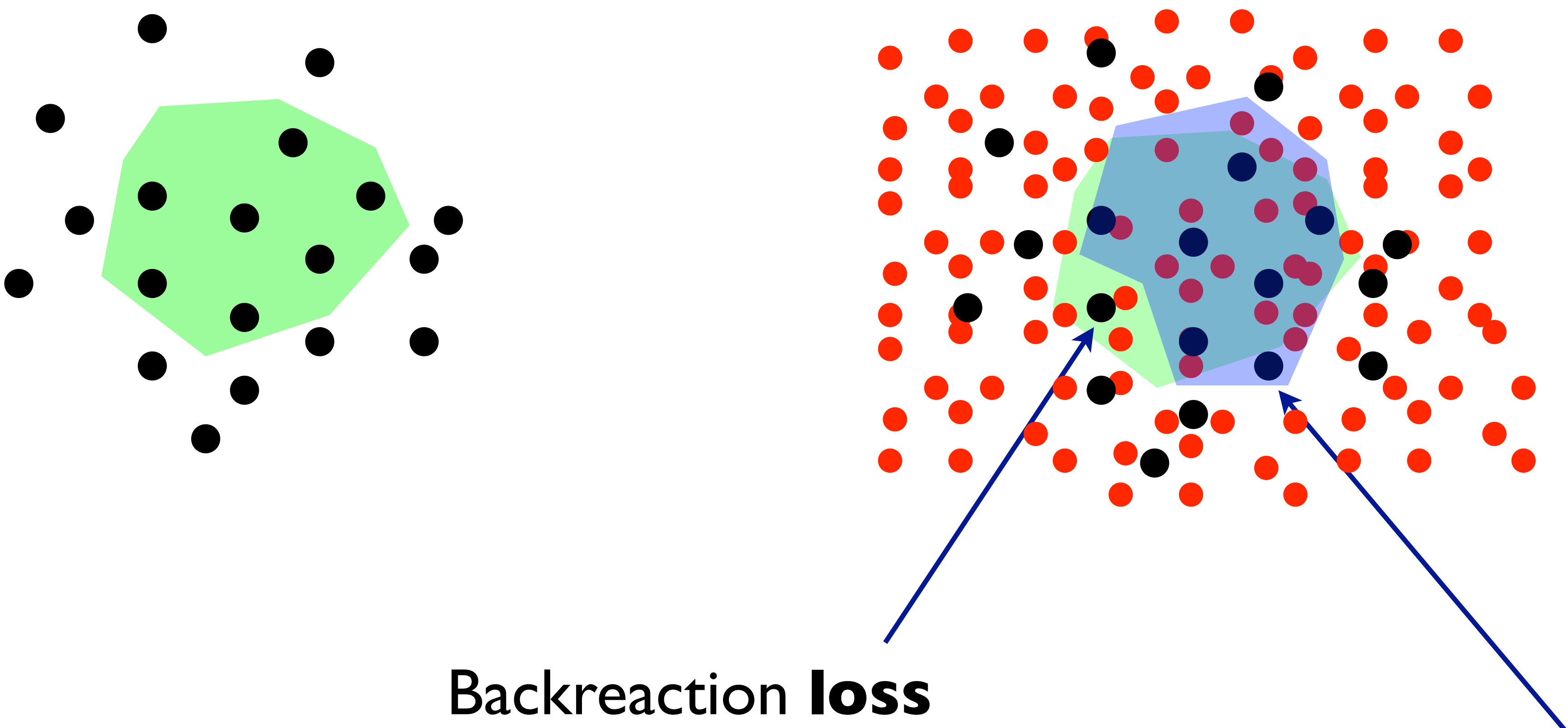
For a roughly uniformly soft background, anti-k_t gives many small jets and many large ones

⇒ not appropriate for estimating a ‘typical’ p_t/area and hence measure ρ

Resiliency: backreaction

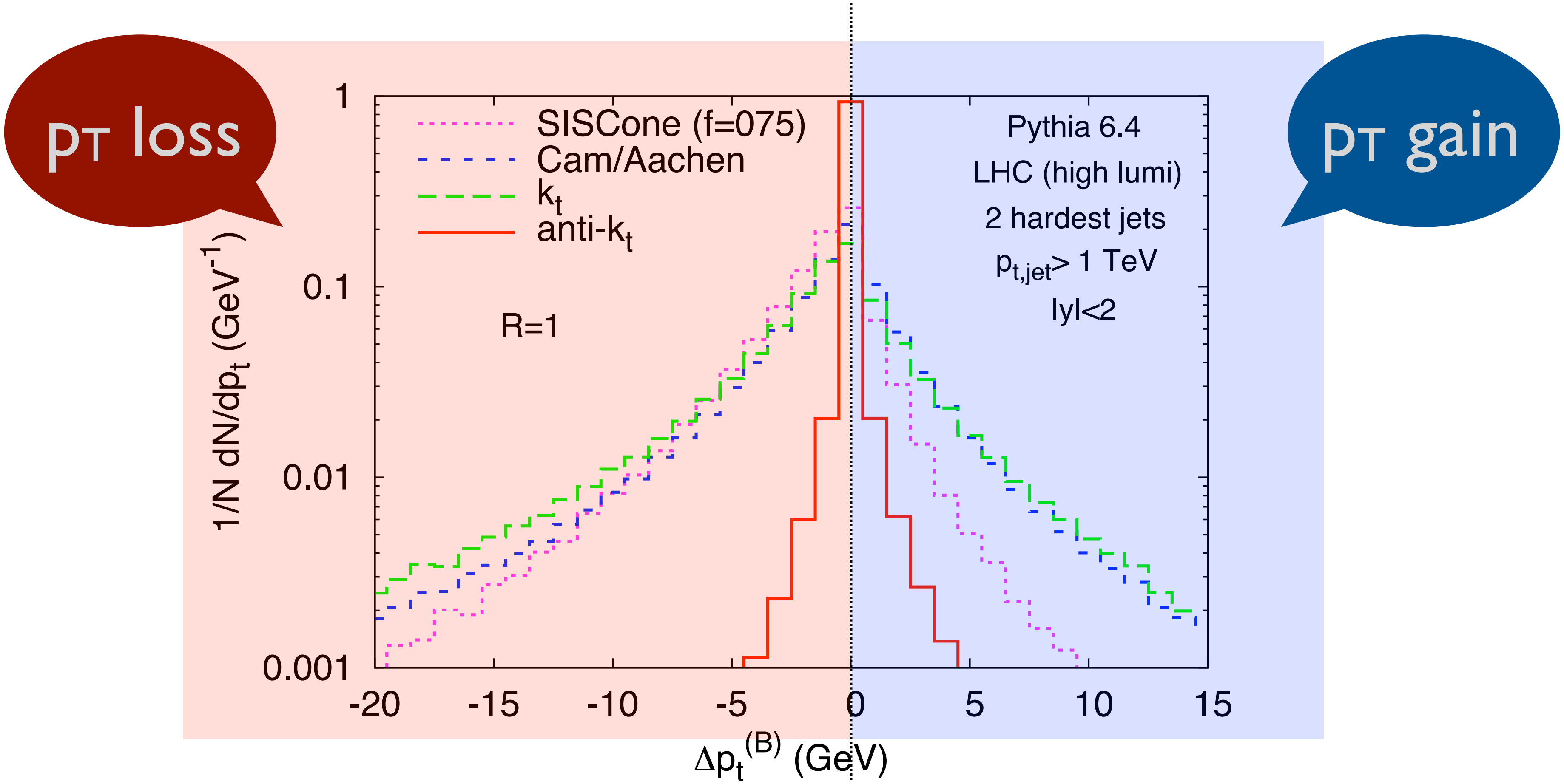
"How (much) a jet changes when immersed in a background"

Without background | **With background**



Resiliency: backreaction

MC, Salam, Soyez, arXiv:0802.1188

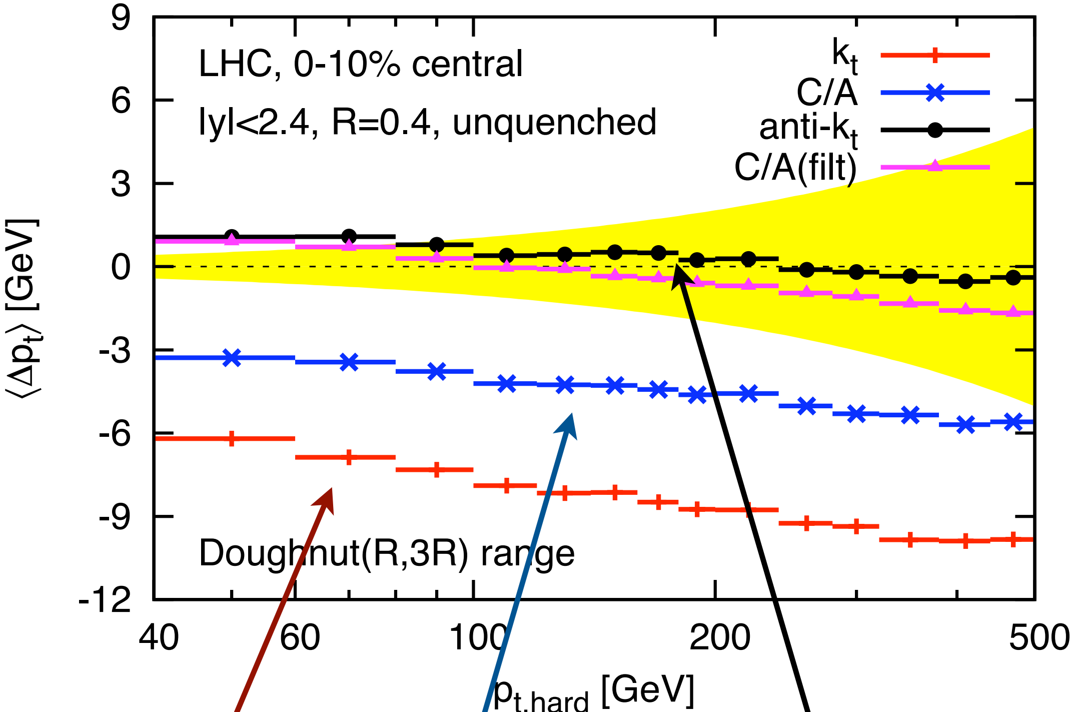


Anti-k_t jets are much more resilient to changes from background immersion

Jet reconstruction in Heavy Ions

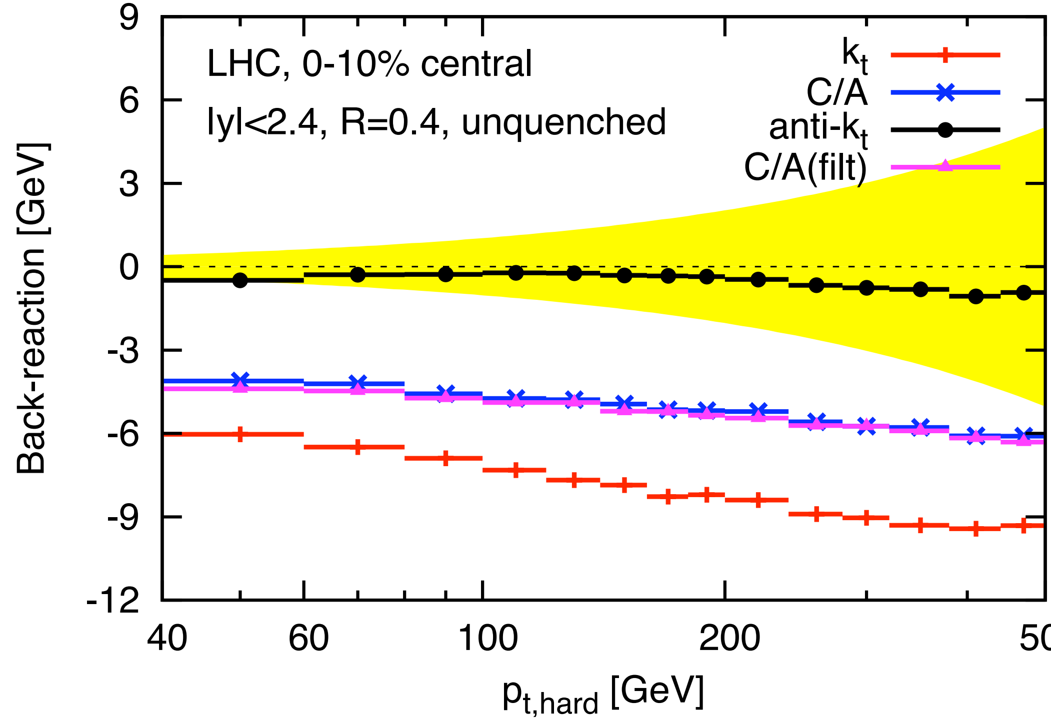
$$\Delta p_t = p_t^{subtracted} - p_t^{true}$$

Total error in p_t reconstruction



k_t C/A anti-k_t

Error due to backreaction



The backreaction contribution explains the mis-reconstruction

anti-k_t provides a bias-free reconstruction

Jets as tools

Background characterisation and subtraction

Mass reconstruction

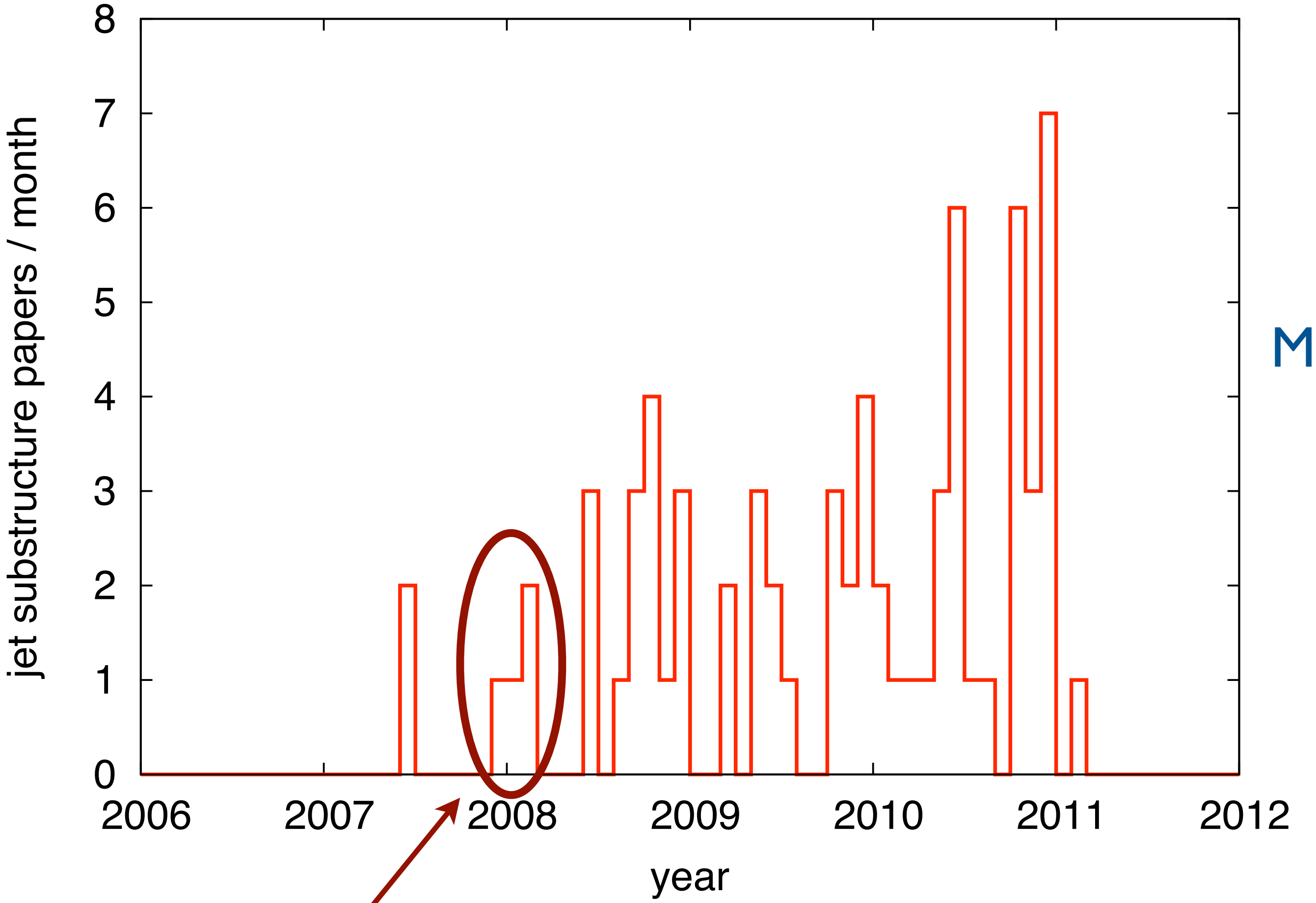
Remove soft contamination from a hard jet

Tag heavy objects originating the jet

Eventually leading to '**third-generation**' jet algorithms

'Jet substructure' papers in SPIRES

Number of papers containing the words 'jet substructure' and 'LHC'



More than 70 papers after 2008.
Essentially none before

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]**

Jet substructure as tagger

Studying the ***jet substructure***

(i.e. the subjets obtained by undoing the clustering of a sequential recombination algorithm)

can lead to **identification capabilities** of specific objects

(as opposed to 'standard' QCD background)

- Boosted Higgs tagger — Butterworth, Davison, Rubin, Salam, 2008
- Boosted top tagger — Kaplan, Rehermann, Schwartz, Tweedie, 2008; Thaler, Wang, 2008; G. Broojmans, ATLAS 2008
- Moderately boosted top and Higgs tagger — Plehn, Salam, Spannowsky, 2009
- + others

Common feature: start with a 'fat jet', decluster it and check if it contains a complex 'hard' substructure

More top taggers

[1 jet ≳ 2 partons]
└[An example]

Tagging boosted top-quarks

Many papers on top tagging in '08–'11: jet mass + something extra.

Questions

- What efficiency for tagging top?
- What rate of fake tags for normal jets?

Rough results for top quark with $p_t \sim 1$ TeV

	"Extra"	eff.	fake
[from T&W]	just jet mass	50%	10%
Brooijmans '08	3,4 k_t subjets, d_{cut}	45%	5%
Thaler & Wang '08	2,3 k_t subjets, z_{cut} + various	40%	5%
Kaplan et al. '08	3,4 C/A subjets, $z_{cut} + \theta_h$	40%	1%
Ellis et al. '09	C/A pruning	10%	0.05%
ATLAS '09	3,4 k_t subjets, d_{cut} MC likelihood	90%	15%
Chekanov & P. '10	Jet shapes	60%	10%
Almeida et al. '08–'10	Template + shapes	13%	0.02%
Thaler & v Tilburg '10	Subjettiness	40%	2%
Plehn et al. '09–'10	C/A MD, θ_h/Dalitz [busy evs, $p_t \sim 300$]	35%	2%

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)

Cambridge/Aachen with filtering

Most of the taggers/filters algorithm use Cambridge/Aachen for the clustering sequence that must then be deconstructed

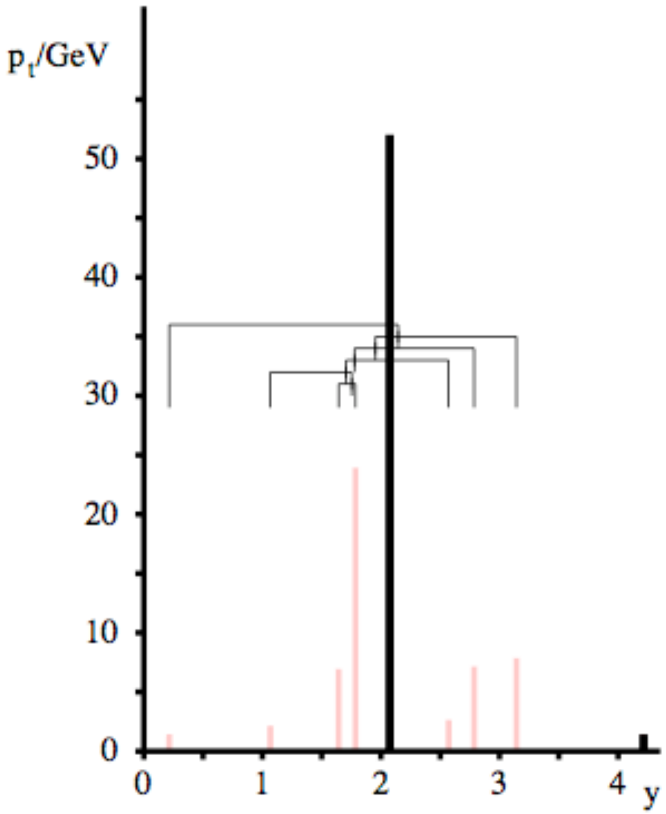
Butterworth, Davison, Rubin, Salam, arXiv:0802.2470

- Cluster with C/A and a given R
- Undo the clustering of each jet down to subjets with radius $x_{filt}R$
- Retain only the n_{filt} hardest subjets

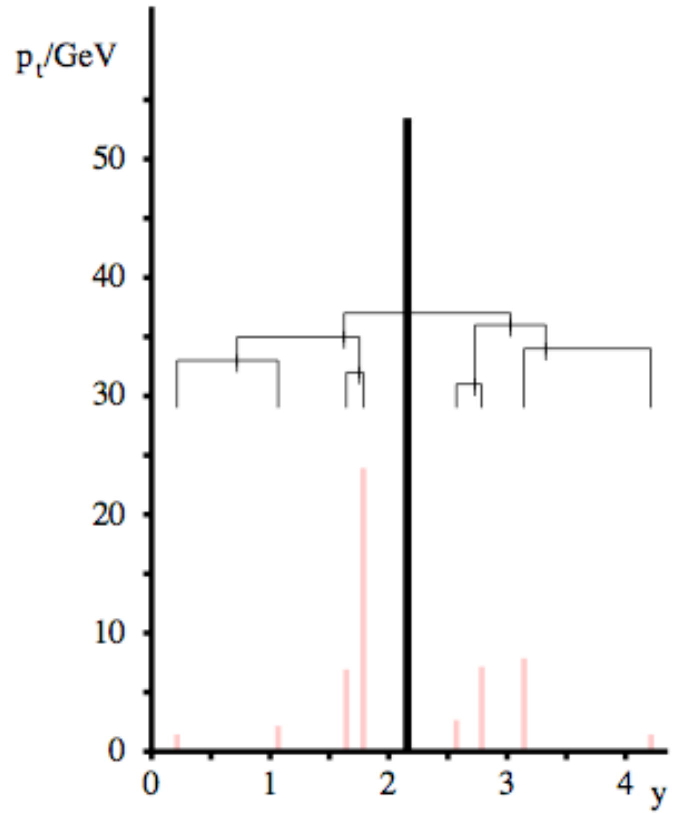
Reason:
meaningfulness of the sequence and
smaller natural sensitivity to soft background

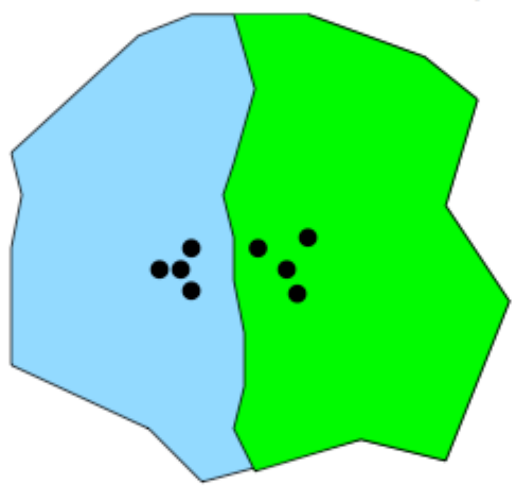
Hierarchical substructure

anti-k_t algorithm

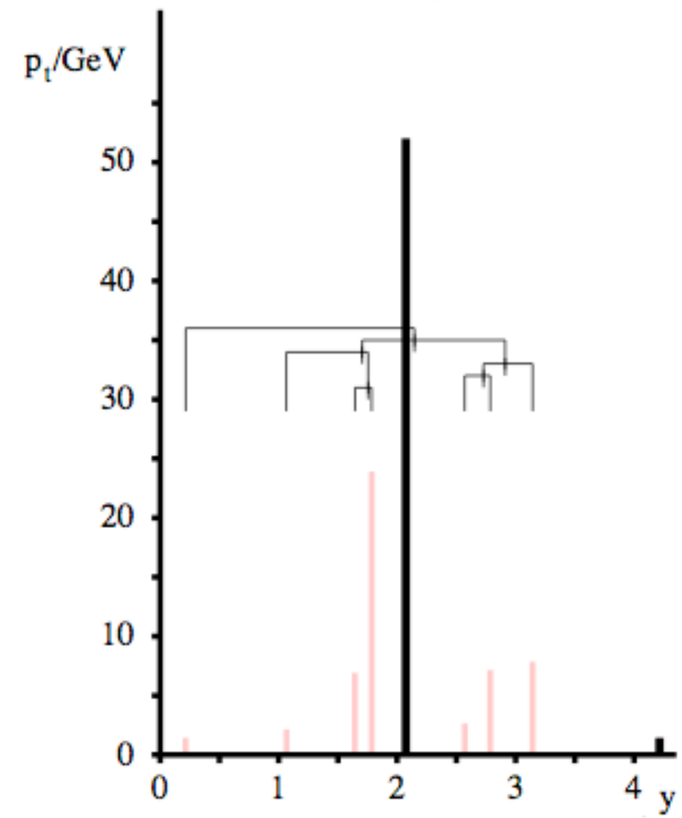


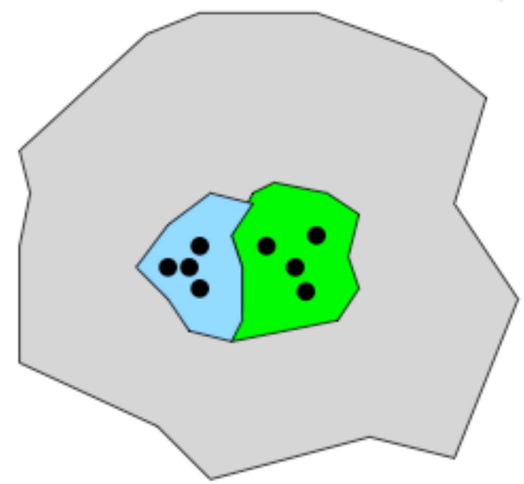
k_t algorithm





Cambridge/Aachen





Slide by
Gavin Salam

The IRC safe algorithms

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

Summary

- At least four IRC-safe algorithms exist: k_t, Cambridge/Aachen, SISCone, anti-k_t (the default algorithm of all LHC Collaborations), all available in FastJet
- Each different **jet definition** provides a handle to a different kind of physics: it is fundamental that flexibility to use the tool most adapted to the task at hand is retained
- Much of future work/progress is probably going to be in substucture studies, i.e. third generation algorithms, which have seen an impressive development in the past two-three years
- Beware: advertisement ahead....

www.fastjet.fr

FastJet version 3, with many improvements and focus on substructure tools, is in the works

FastJet

The FastJet package, written by Matteo Cacciari, Gavin Salam and Gregory Soyez, provides a fast implementation of the longitudinally invariant kt [1,2] longitudinally invariant inclusive Cambridge/Aachen [3,4] and anti-kt [7] jet finders and a uniform interface to external jet finders (notably SISCone [5]) via a plugin mechanism. It also includes tools for calculating jet areas [8] and performing background (pileup/UE) subtraction [9].

Native jet-finding is based on the geometrical methods described in Phys. Lett. B **641** (2006) 57 [hep-ph/0512210] and [6].

***** Alpha preview for 3.0 series: fastjet-3.0alpha2, released March 10, 2011 (release notes)**
*** It includes a preliminary subset of the features planned for inclusion in the 3.0 series.

***** NEW: Current version: fastjet-2.4.3, released April 14, 2011 (release notes) *****

Main new features in the 2.4.x series

- Addition of several new pp algorithms: D0 run II cone, ATLAS Cone, TrackJet, CMS Iterative cone (all as plugins) and generalised k_t (native).
- Introduction of e^+e^- algorithms: k_t and generalised k_t (native), as well as Jade, Cambridge and spherical SISCone (plugins).
- Also: new way of accessing jet substructure, some improvements in tools related to background estimation, facilities for easier implementation of new sequential-recombination algorithms.
- Backwards compatibility note: for a number of jet definitions, certain misleading default values have been removed from the constructor.

Additionally, version 2.4.3 fixes an incorrect return sign of PseudoJet::delta_phi_to(...) and eliminates NaNs in Voronoi areas related to rounding errors for particles on a grid layout. For more details, see the release notes.

FastJet resources

- **Main page**
- Download v2.4.3
- Manual
- Doxygen
- All releases
- Jet algorithm list
- Quick start guide
- Tools *(devel)*
- FAQ

Alpha release

- Download 3.0alpha2
- Manual for 3.0alpha2
- Doxygen for 3.0alpha2

Extra material

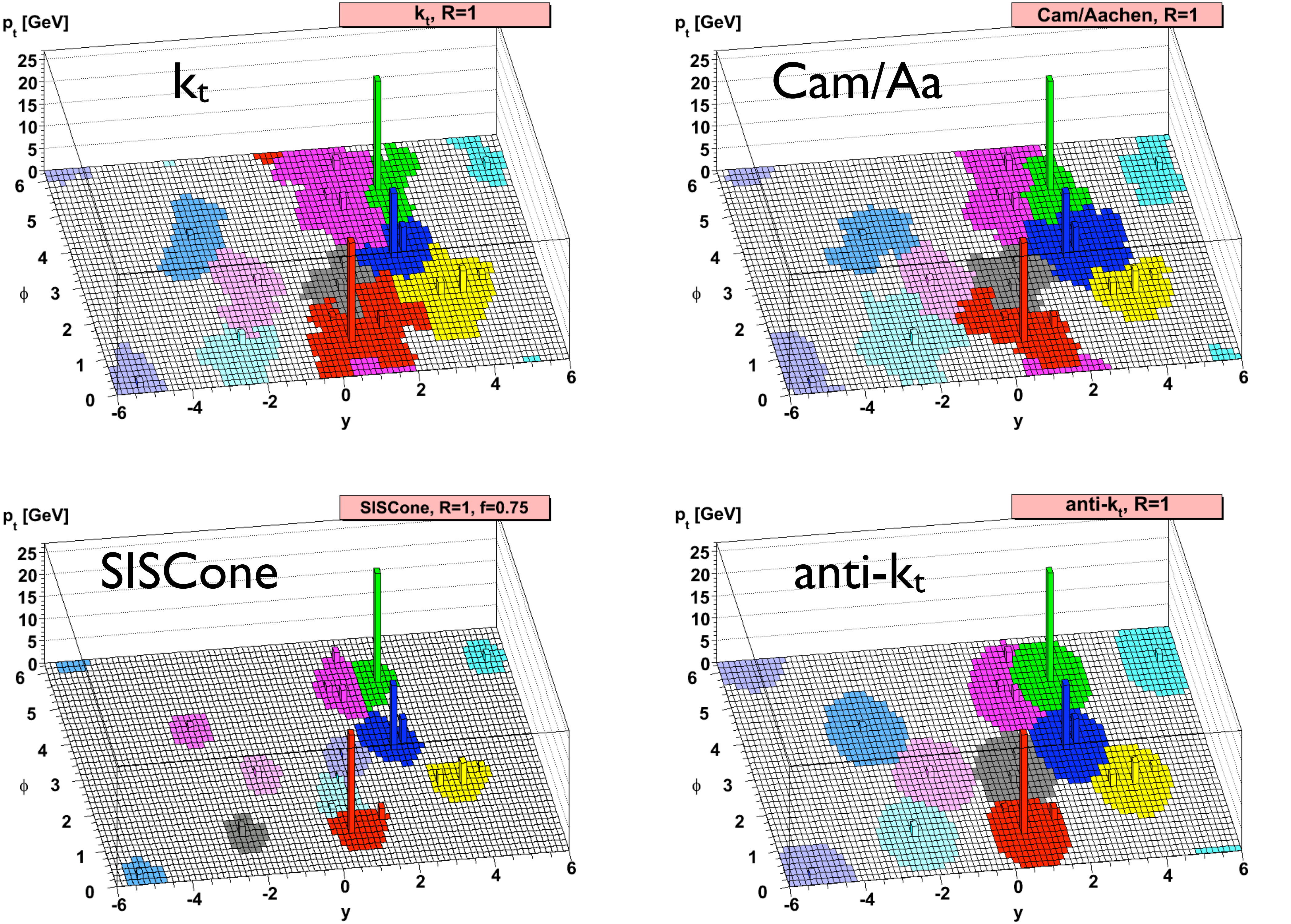
The IRC safe algorithms

k_t	SR $d_{ij} = \min(k_{ti}^2, k_{tj}^2)\Delta R_{ij}^2/R^2$ hierarchical in rel p_t	Catani et al '91 Ellis, Soper '93	NlnN
Cambridge/ Aachen	SR $d_{ij} = \Delta R_{ij}^2/R^2$ hierarchical in angle	Dokshitzer et al '97 Wengler, Wobish '98	NlnN
anti-k_t	SR $d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2})\Delta R_{ij}^2/R^2$ gives perfectly conical hard jets	MC, Salam, Soyez '08 (Delsart, Loch)	$N^{3/2}$
SISCone	Seedless iterative cone with split-merge gives 'economical' jets	Salam, Soyez '07	$N^2 \ln N$

We call these algs '**second-generation**' jet algorithms

All are available in FastJet, http://fastjet.fr

(As well as many IRC unsafe ones)



k_t, R=1
k_t
Cam/Aachen, R=1
Cam/Aa
SISCone, R=1, f=0.75
SISCone
anti-k_t, R=1
anti-k_t
p_t [GeV]
φ
y

Jet trimming

Krohn, Thaler, Wang, 2009

1. Cluster all cells/tracks into jets using any clustering algorithm. The resulting jets are called the seed jets.

2. Within each seed jet, recluster the constituents using a (possibly different) jet algorithm into subjets with a characteristic radius R_{sub} smaller than that of the seed jet.

3. Consider each subjet, and discard the contributions of subjet i to the associated seed jet if $p_{Ti} < f_{\mathrm{cut}} \cdot \Lambda_{\mathrm{hard}}$, where f_{cut} is a fixed dimensionless parameter, and Λ_{hard} is some hard scale chosen depending upon the kinematics of the event.

4. Assemble the remaining subjets into the trimmed jet.

Different condition for retaining jets (p_T-cut rather than n_{filt} hardest) with respect to filtering

In FastJet (v3 only)

```
Filter trimmer(JetDefinition(cambridge_algorithm,xfilt*R),
               SelectorPtMin(fcut*Lambdahard));

PseudoJet trimmed_jet = trimmer(jet);
```

Jet pruning

S. Ellis, Vermilion, Walsh, 2009

0. Start with a jet found by any jet algorithm, and collect the objects (such as calorimeter towers) in the jet into a list L. Define parameters $D_{\rm cut}$ and $z_{\rm cut}$ for the pruning procedure.

1. Rerun a jet algorithm on the list L, checking for the following condition in each recombination $i, j \to p$:

$$z = \frac{\min(p_{Ti}, p_{Tj})}{p_{Tp}} < z_{\rm cut} \quad \text{and} \quad \Delta R_{ij} > D_{\rm cut}.$$

   This algorithm must be a recombination algorithm such as the CA or k_T algorithms, and should give a "useful" jet substructure (one where we can meaningfully interpret recombinations in terms of the physics of the jet).

2. If the conditions in 1. are met, do not merge the two branches 1 and 2 into p. Instead, discard the softer branch, i.e., veto on the merging. Proceed with the algorithm.

3. The resulting jet is the *pruned jet*, and can be compared with the jet found in Step 0.

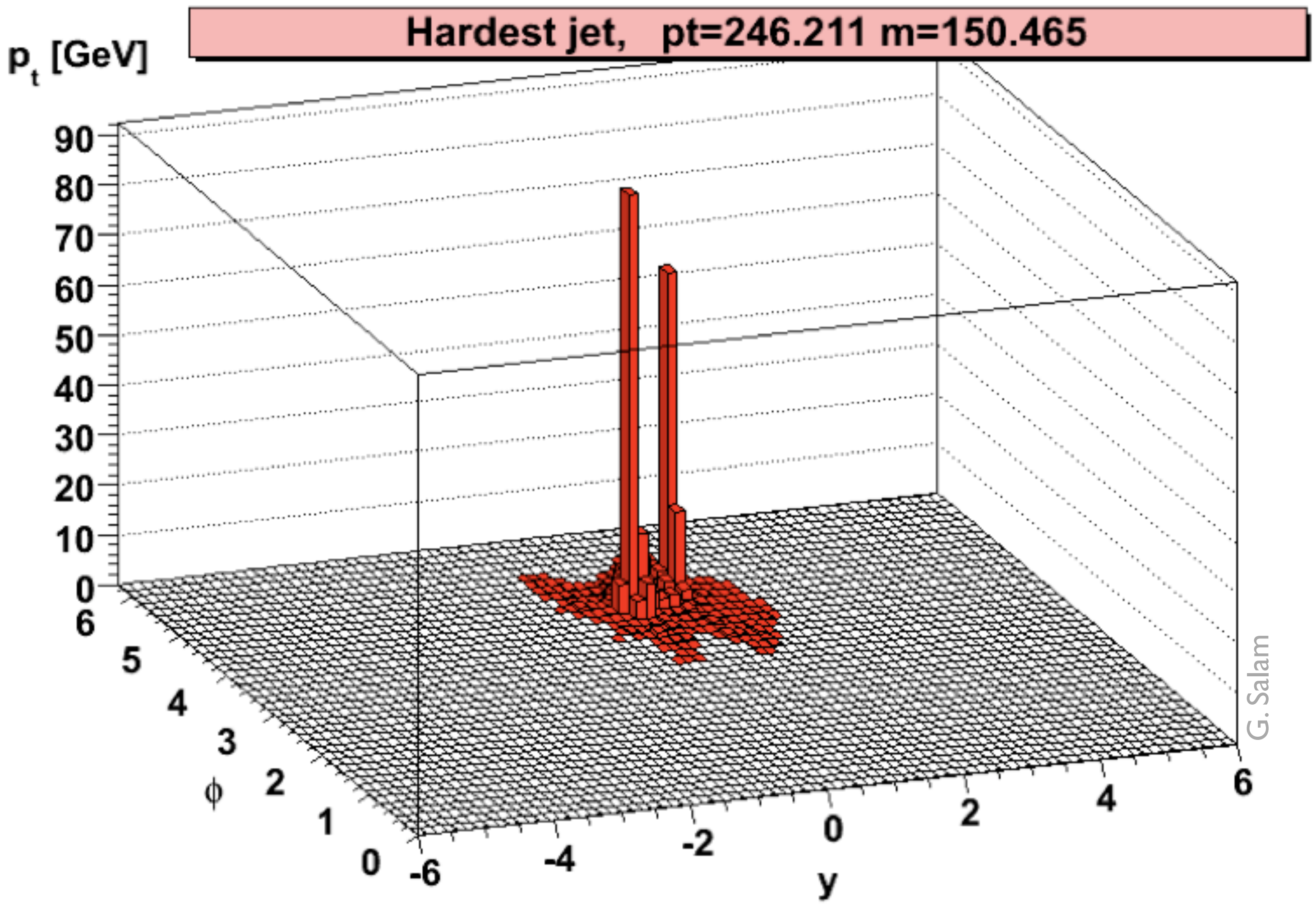
True in general for substructure studies

Exclude soft stuff and large angle recombinations from clustering

Boosted Higgs tagger

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

Butterworth, Davison, Rubin, Salam, 2008



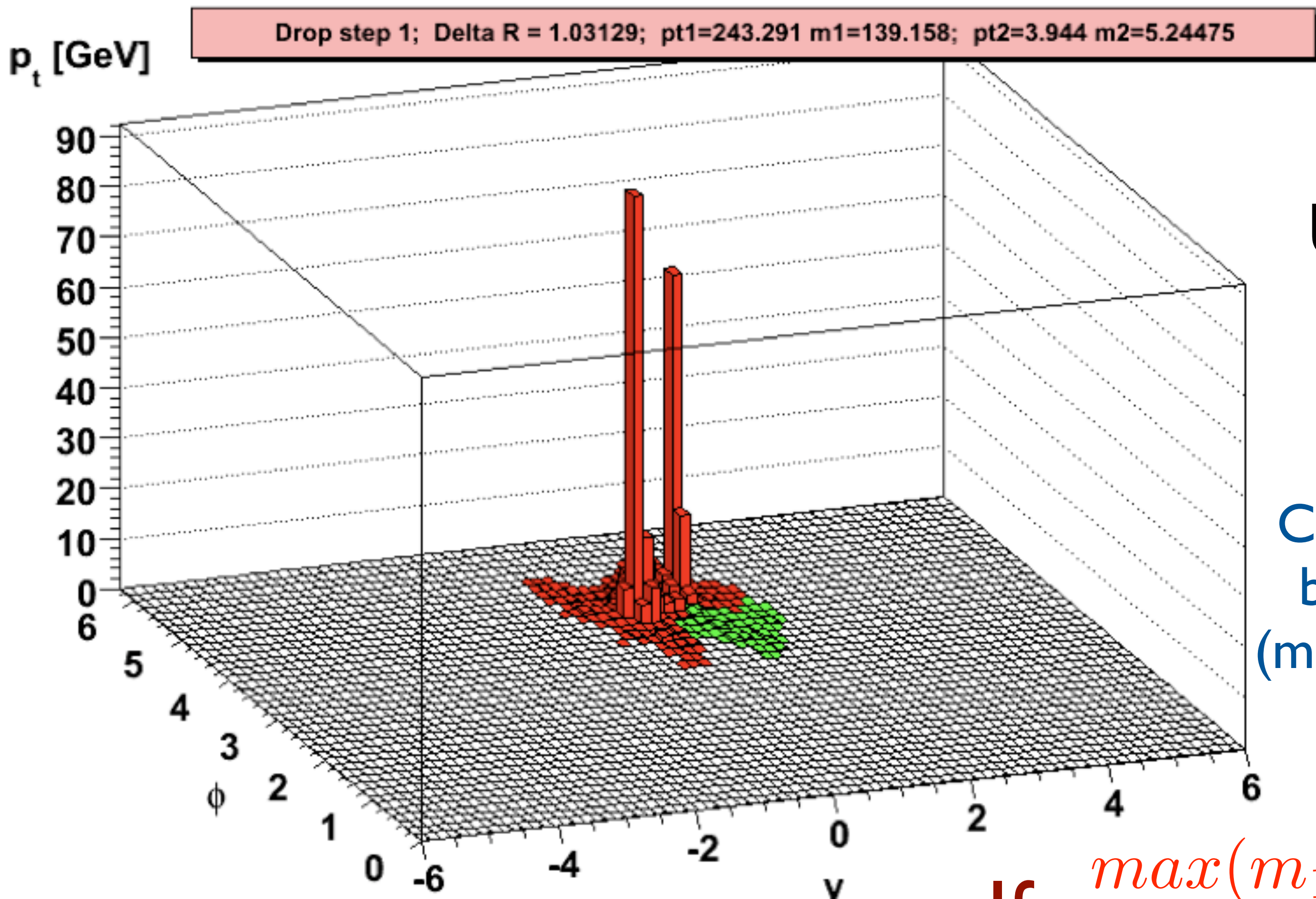
Start with the hardest jet

Use C/A with large R=1.2

m = 150 GeV

Boosted Higgs tagger

pp → ZH → ννbb



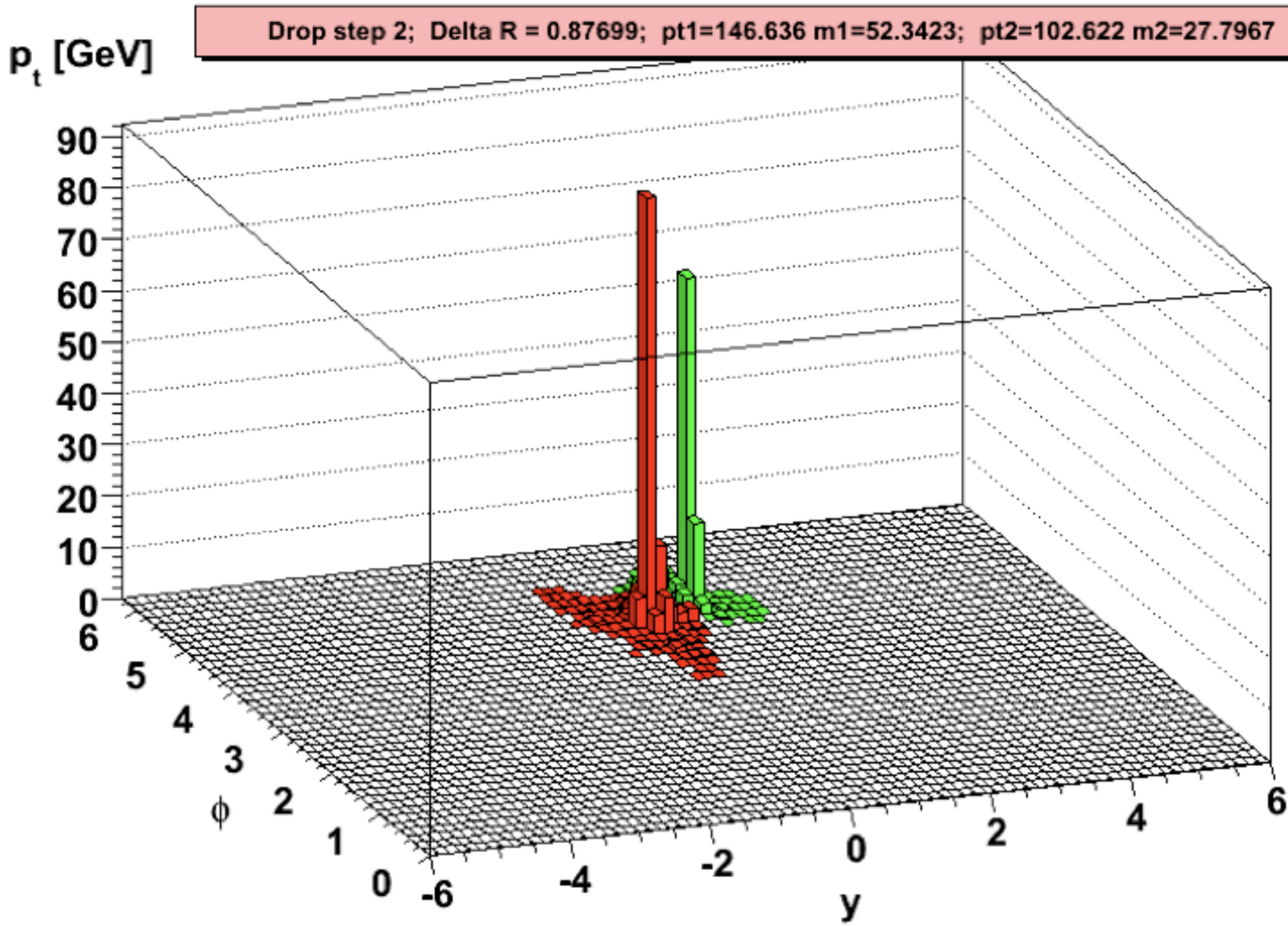
Undo last step of clustering

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

If $\frac{max(m_1, m_2)}{m} > \mu$ repeat

Boosted Higgs tagger

pp → ZH → ννbb



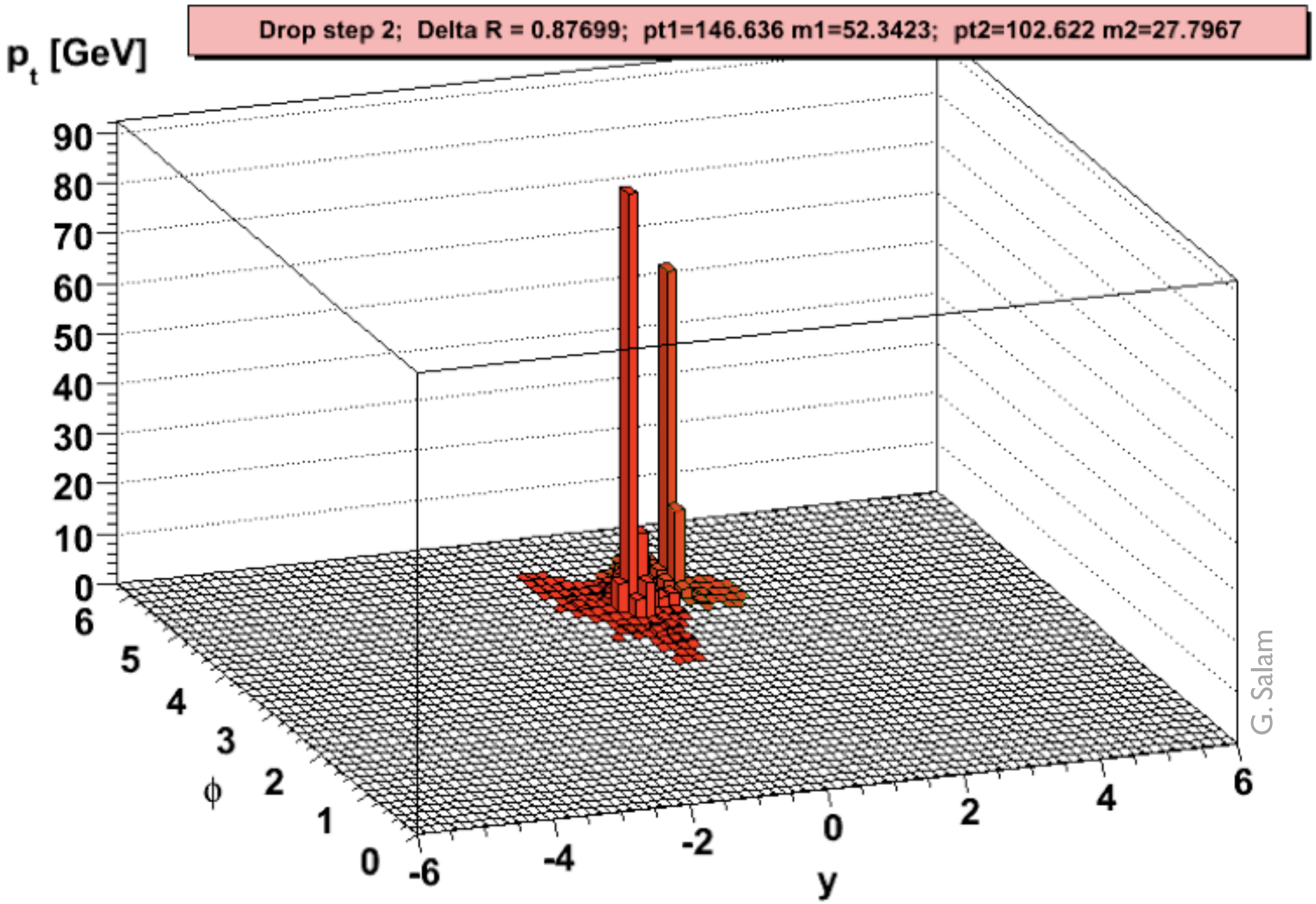
m_1 = 52 GeV, m_2 = 28 GeV

Stop when a large mass drop is observed

(and recombine these two jets)

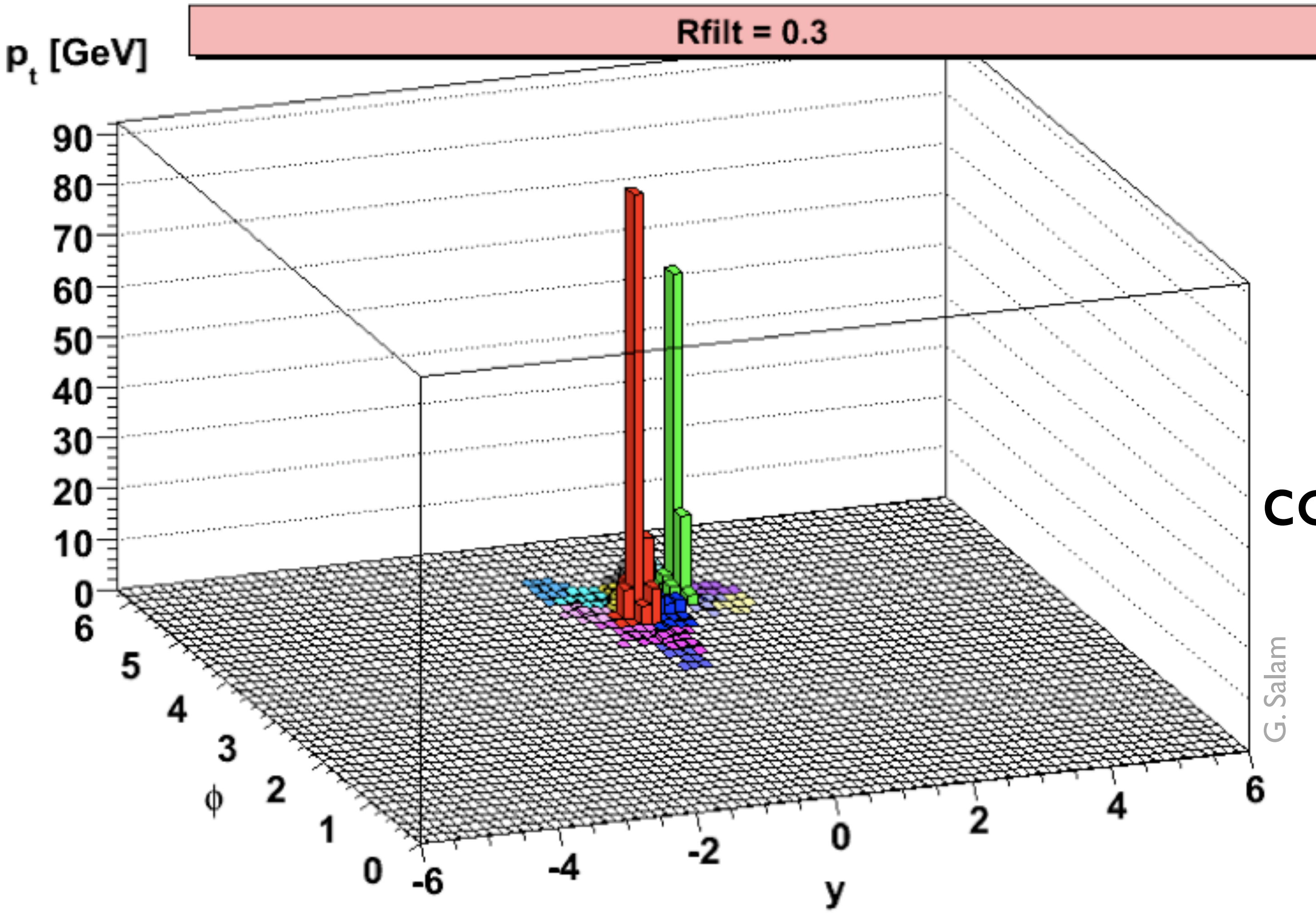
Filtering in action

Butterworth, Davison, Rubin, Salam, arXiv:0802.2470



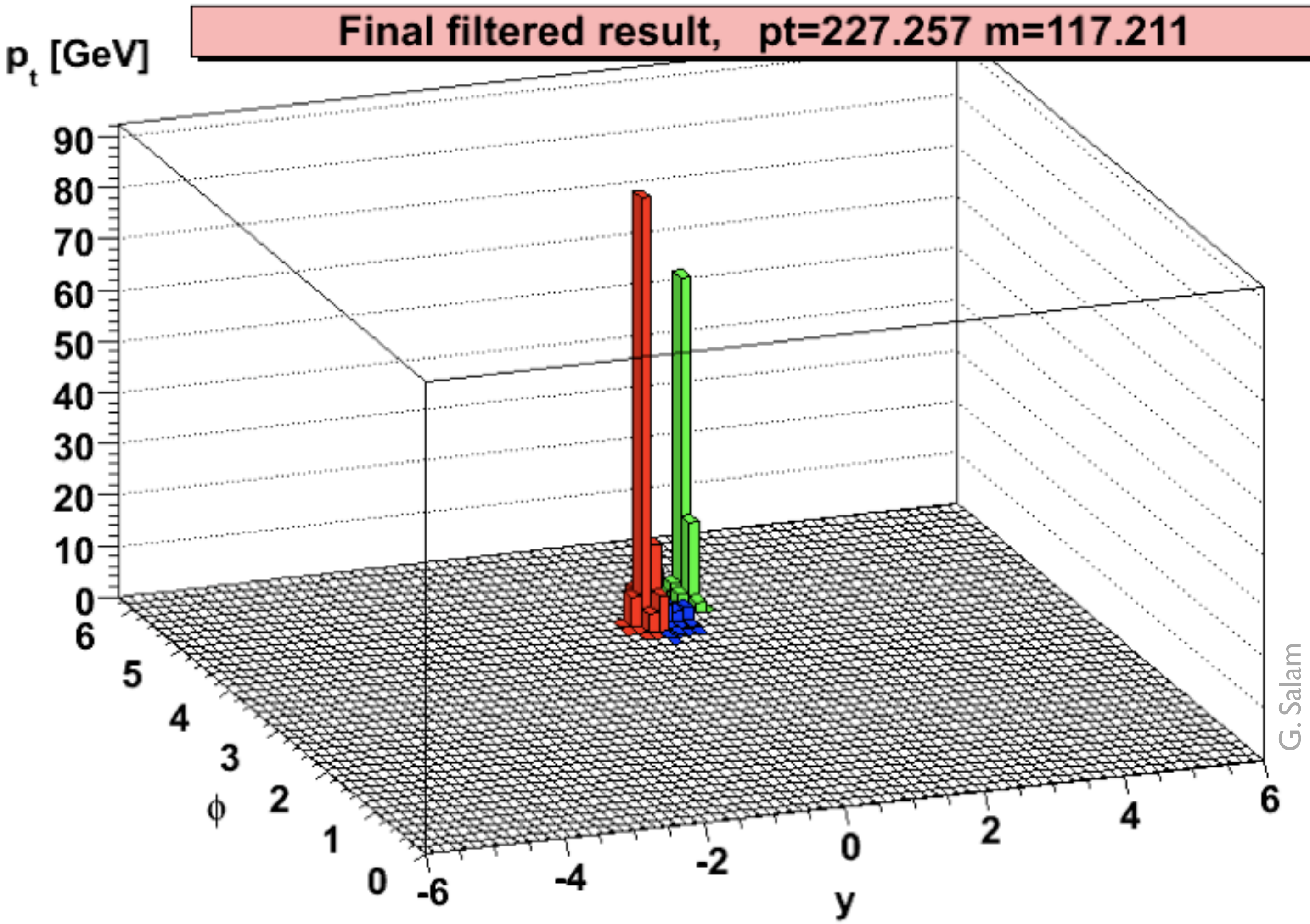
Start with a jet

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