

# Recent Progress in Jet Algorithms and Their Impact in Underlying Event Studies

Matteo Cacciari  
LPTHE - Paris 6,7 and CNRS

Contains work done in collaboration  
with G. Salam, G. Soyez, J. Rojo

# Outline

Infrared safety

Recent set of infrared and collinear safe jet algorithms

Jet Areas

Second-generation jet algorithms allow area definition and measurement

UE level estimation

Knowledge of areas allows estimation of a uniform background

UE/pileup subtraction

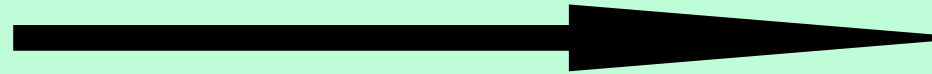
The estimated uniform background might be subtracted from jets

# jet definition

 $\{P_i\}$ 

particles,  
4-momenta,  
calorimeter towers, ....

jet algorithm

 $\{j_k\}$ 

jets

+ parameters (usually at least the radius  $R$ )

+ recombination scheme

Reminder: running a jet definition gives a well defined physical observable,  
which we can measure and, hopefully, calculate

# Two main classes of jet algorithms

## Sequential recombination algorithms

bottom-up approach: combine particles starting from **closest ones** in some distance measure. Repeat until few left: call them jets

Work because of mapping closeness  $\Leftrightarrow$  QCD divergence

Examples:  $k_t$ , Cambridge/Aachen, .... Loved by  $e^+e^-$ , ep and theorists

## Cone algorithms

top-down approach: find coarse regions of energy flow.

How? Find **stable** cones (i.e. their axis coincides with sum of momenta of particles in it)

Work because QCD only modifies energy flow on small scales

Examples: JetClu, MidPoint, ..... Loved by pp and (fewer) theorists

# Jet Algorithm

In order to work properly jet algorithms should be

- **Infrared safe** (for the calculation to converge)
  - soft emission shouldn't change jets
  - collinear splitting shouldn't change jets
- **fast** (if many particles have to be clustered)
  - In order to be realistically applicable at detector level

These (and other) requirements were widely acknowledged as early as 1990  
(Snowmass accord)

Perhaps quite surprisingly, they had not yet been met in 2005

A calorimeter is of course always infrared safe.  
You need these properties not so much to find jets, but to be able to calculate them in pQCD

# A (long!) list of cones (all eventually unsafe)

Les Houches 2007 proceedings, arXiv:0803.0678

‘First-generation’ algorithms

CDF JetClu	$IC_r$ -SM	$IR_{2+1}$
CDF MidPoint cone	$IC_{mp}$ -SM	$IR_{3+1}$
CDF MidPoint searchcone	$IC_{se,mp}$ -SM	$IR_{2+1}$
D0 Run II cone	$IC_{mp}$ -SM	$IR_{3+1}$
ATLAS Cone	IC-SM	$IR_{2+1}$
PxCone	$IC_{mp}$ -SD	$IR_{3+1}$
CMS Iterative Cone	IC-PR	$Coll_{3+1}$
PyCell/CellJet (from Pythia)	FC-PR	$Coll_{3+1}$
GetJet (from ISAJET)	FC-PR	$Coll_{3+1}$

IC = Iterative Cone  
 SM = Split-Merge  
 SD = Split-Drop  
 FC = Fixed Cone  
 PR = Progressive Removal

type of  
algorithm

safety issue

$IR_{n+1}$  : unsafe when a soft particle is added to  
 n hard particles in a common neighbourhood  
 $Coll_{n+1}$  : unsafe when one of n hard particles in  
 a common neighbourhood is split collinearly

# Lessons

- There isn't **one** cone algorithm, but rather many different cones, which can behave quite distinctly from one another
- Essentially all of the cones commonly used are unsafe at some point. The best ones only fail at NNLO (3+1), others already at NLO (2+1)

Examples:	<i>Last meaningful order</i>			Known at
	ATLAS cone [IC-SM]	MidPoint [IC <sub>mp</sub> -SM]	CMS it. cone [IC-PR]	
Inclusive jets	LO	NLO	NLO	NLO (→ NNLO)
W/Z + 1 jet	LO	NLO	NLO	NLO
3 jets	<b>none</b>	LO	LO	NLO [nlojet++]
W/Z + 2 jets	<b>none</b>	LO	LO	NLO [MCFM]
$m_{\text{jet}}$ in $2j + X$	<b>none</b>	<b>none</b>	<b>none</b>	LO

G. Salam

Calculations cost real money:  
 ~ 100 theorists × 15 years ≈ 100 M€

Using unsafe jet tools essentially renders them potentially useless

**Solutions?**

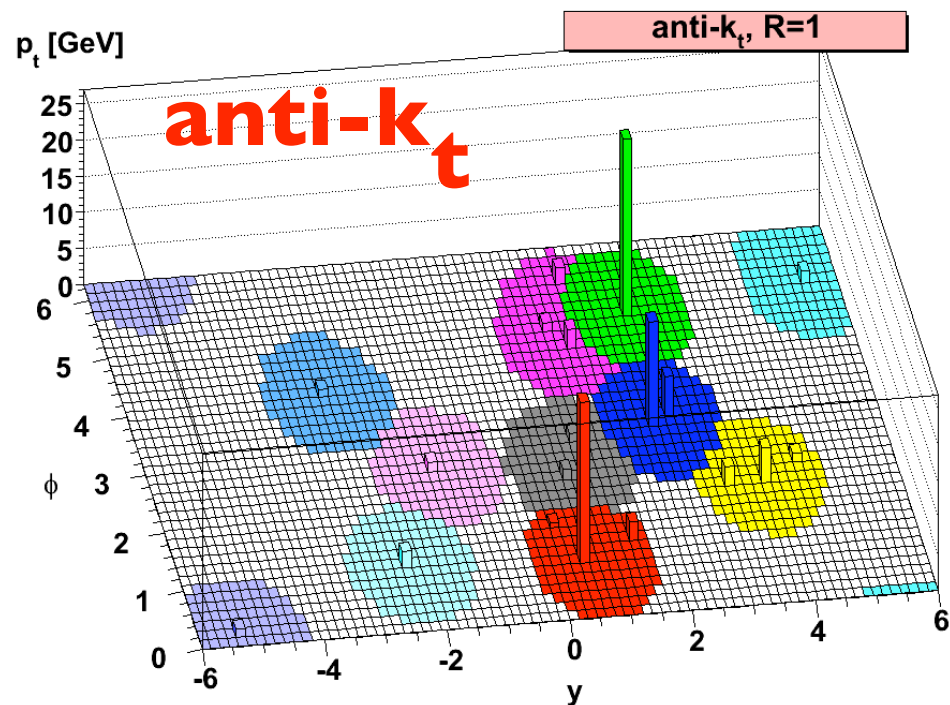
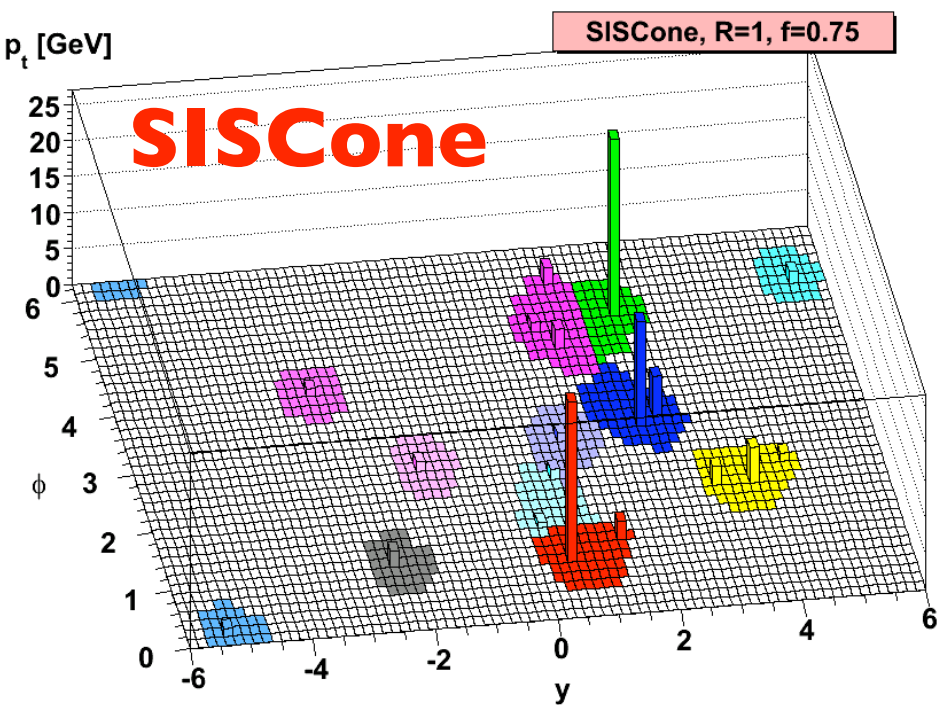
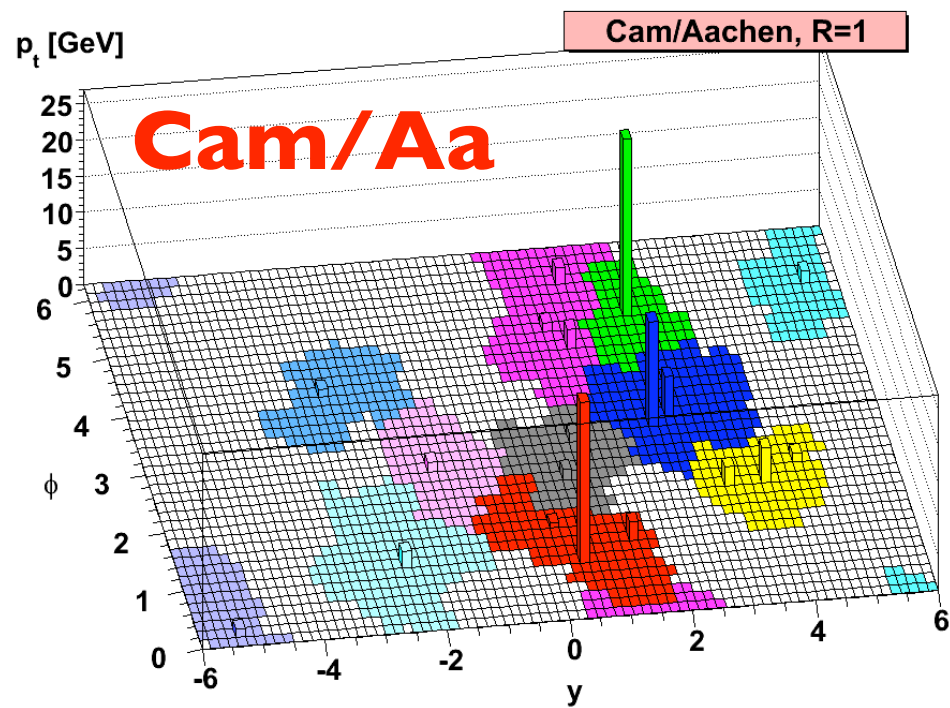
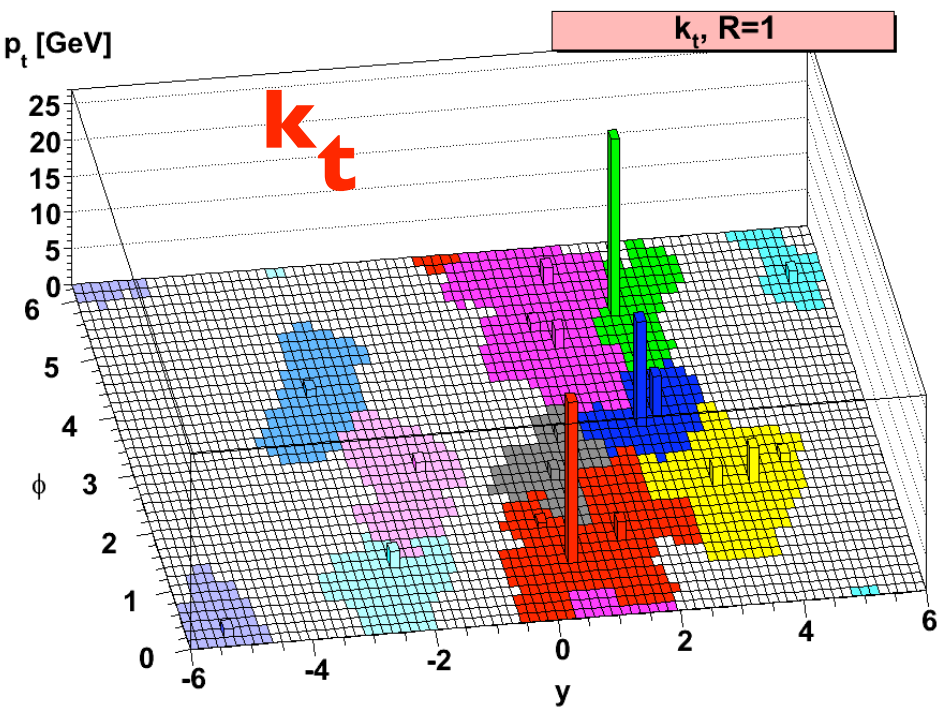
# The IRC safe algorithms

$k_t$	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	$N \ln N$
Cambridge/ Aachen	SR $d_{ij} = \Delta R_{ij}^2 / R^2$ hierarchical in angle	Dokshitzer et al '97 Wengler, Wobish '98	$N \ln N$
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' ones

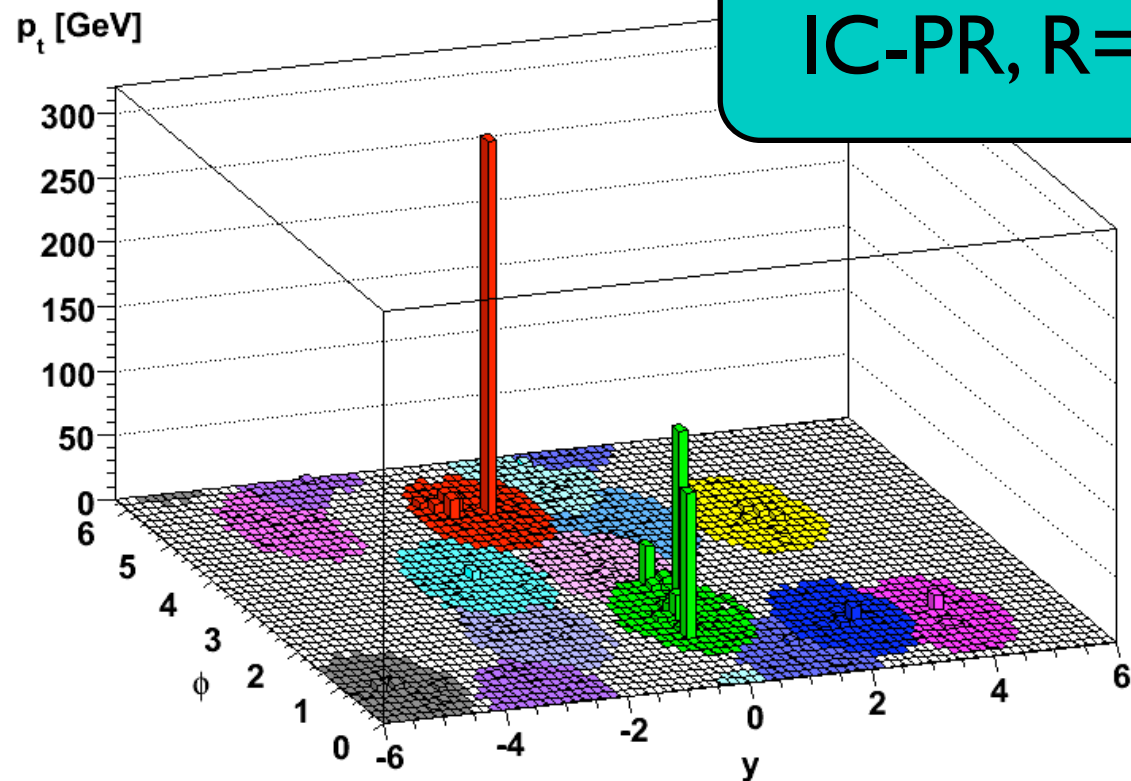
All are available in FastJet, <http://fastjet.fr>

(As well as many IRC unsafe ones)

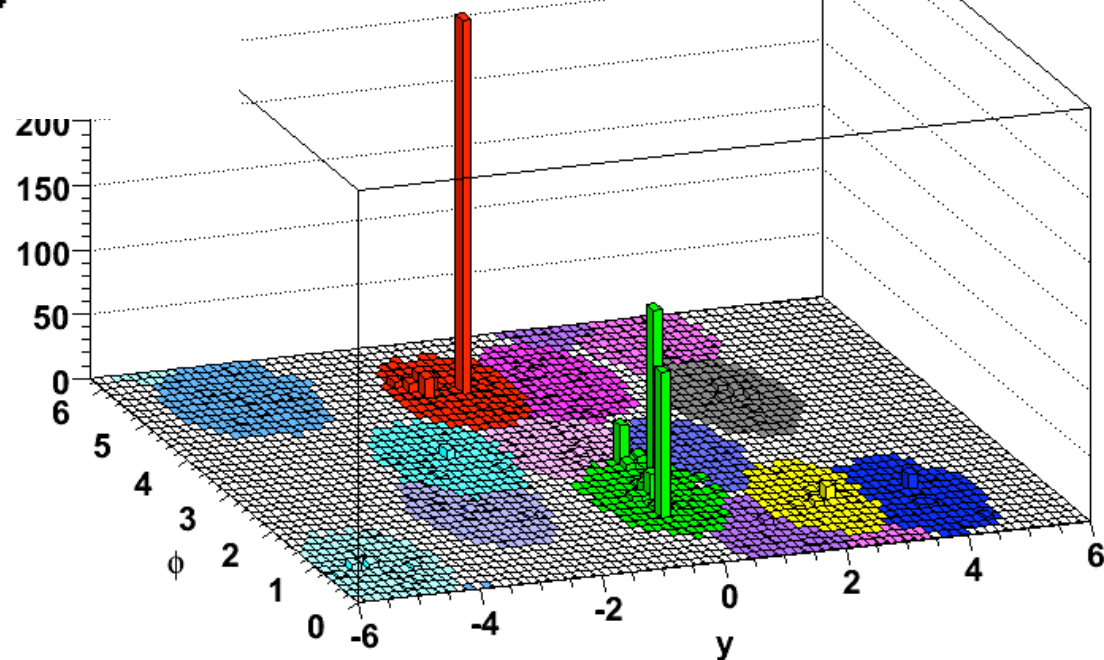


# Example: IC-PR v. anti- $k_t$

IC-PR,  $R=1$



anti- $k_t$ ,  $R=1$



Hardest jets are similarly regular

# Replacements

If you care about IRC safety but don't want to stray too far from algorithms used so far, these are possible replacements:

JetClu  
ATLAS Cone  
MidPoint



SIS Cone

(As fast, but  
IRC safe)

Iterative Cone (PR)



anti- $k_t$

(Gives regular cones too,  
but IRC safe)

In addition,  $k_t$  and Cambridge/Aachen can provide further flexibility

Different algorithms spanning a series of different and complementary characteristics: should be enough for most purposes

One should probably try to concentrate on these, both for analytical understanding and practical use in experiments, rather than using IRC unsafe ones

# Why so many different algorithms?

Not all analyses are alike:

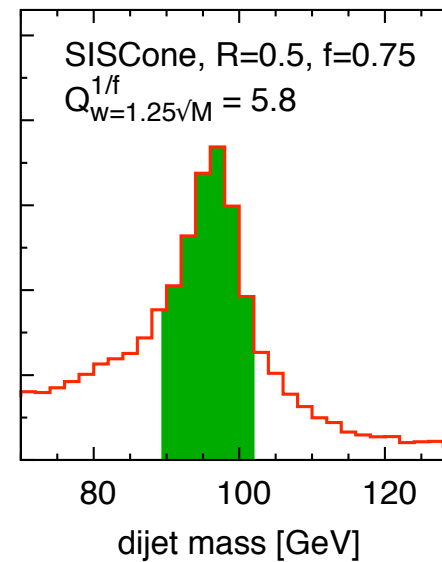
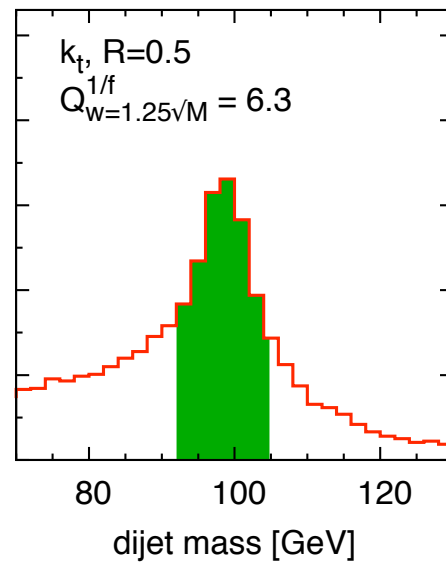
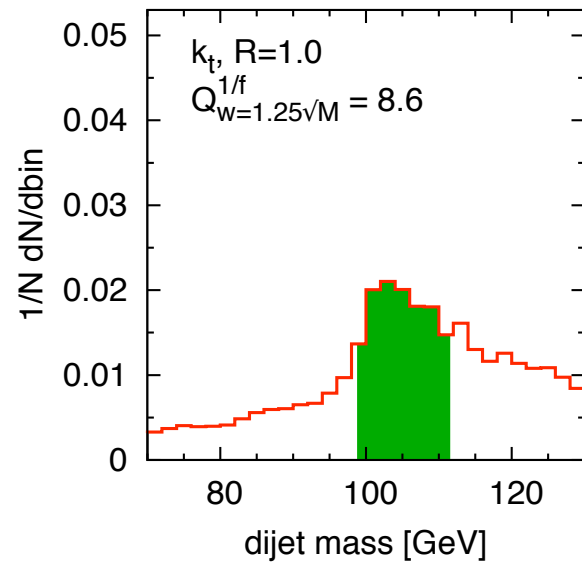
Advantages of **flexibility**

# Reconstruction of a mass peak

MC, Rojo, Salam, Soyez, arXiv:0810.1304

R=1: **BAD**

R=0.5: **BETTER**

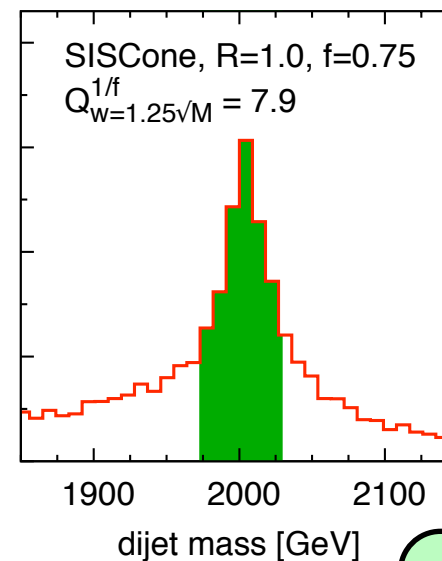
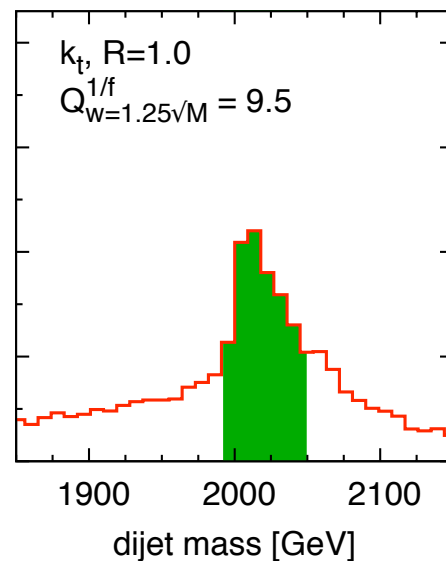
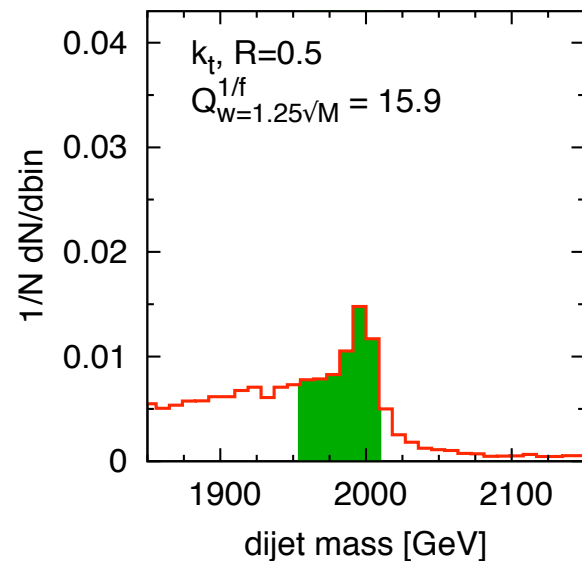


qq 100 GeV

qq jets  
at 100 GeV

R=0.5: **BAD**

R=1: **BETTER**



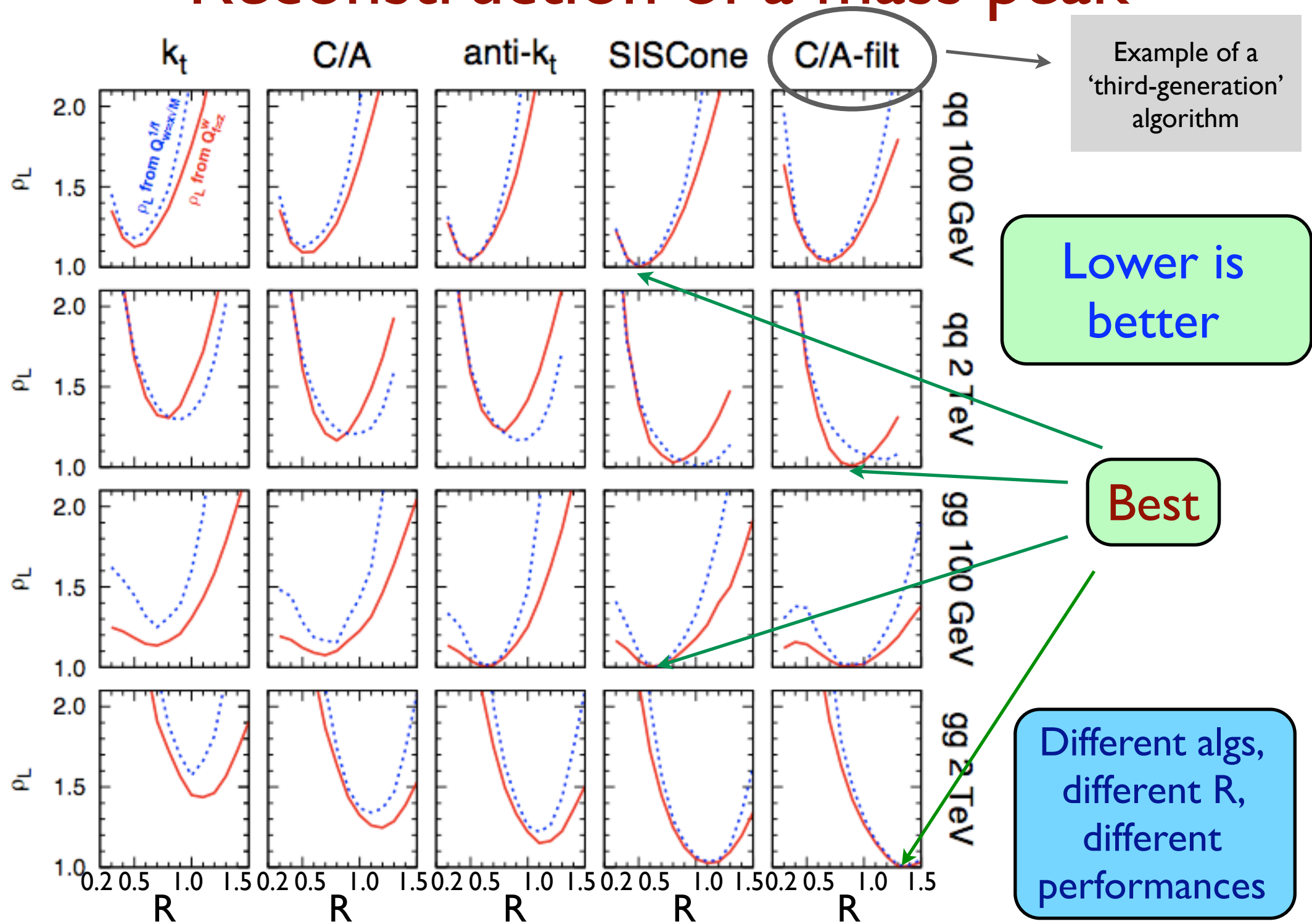
gg 2 TeV

gg jets  
at 2 TeV

Gluons (and heavy objects) prefer larger R

~ 100000 similar plots at  
<http://quality.fastjet.fr>

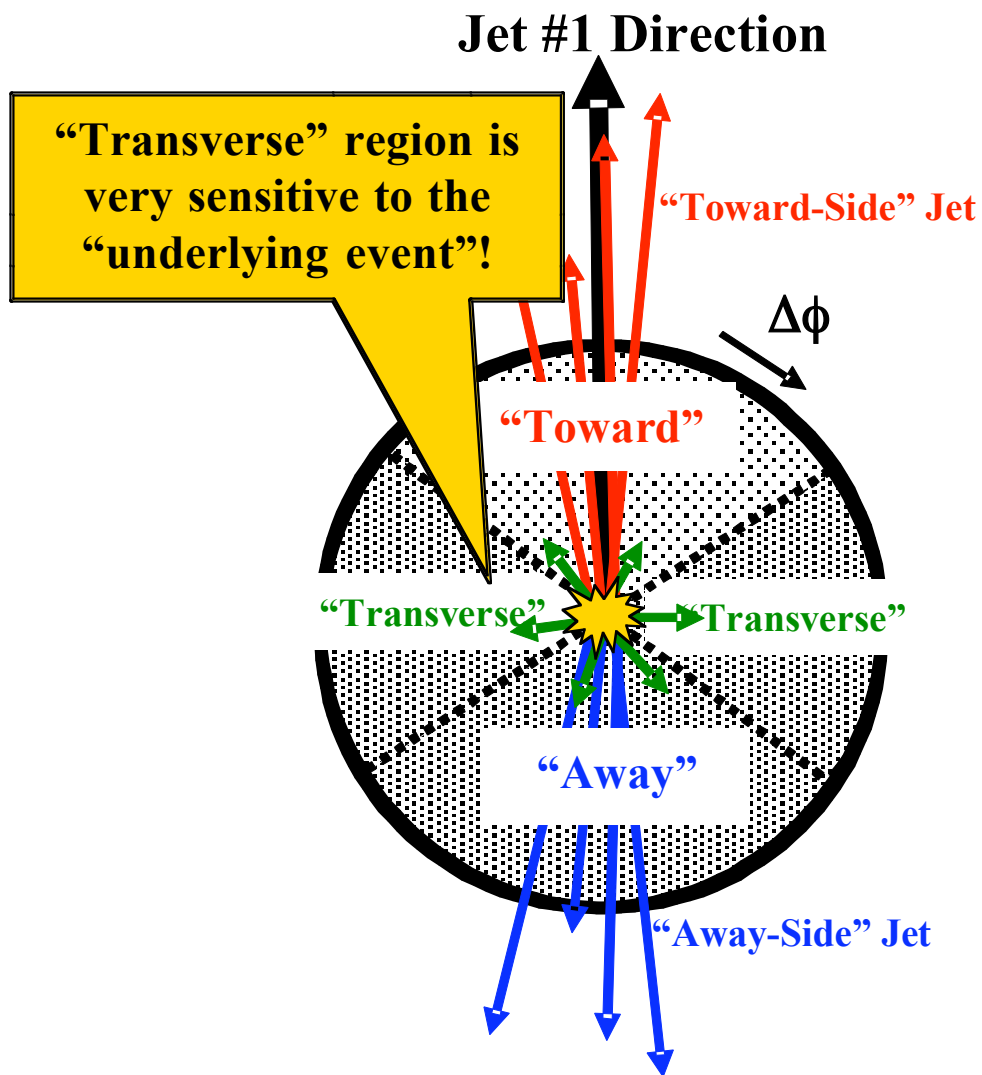
# Reconstruction of a mass peak



**A different use of  
(IRC safe) jets**



# Rick Field's picture



Marchesini-Webber idea:  
look at transverse region to  
measure underlying event



## Topological selection

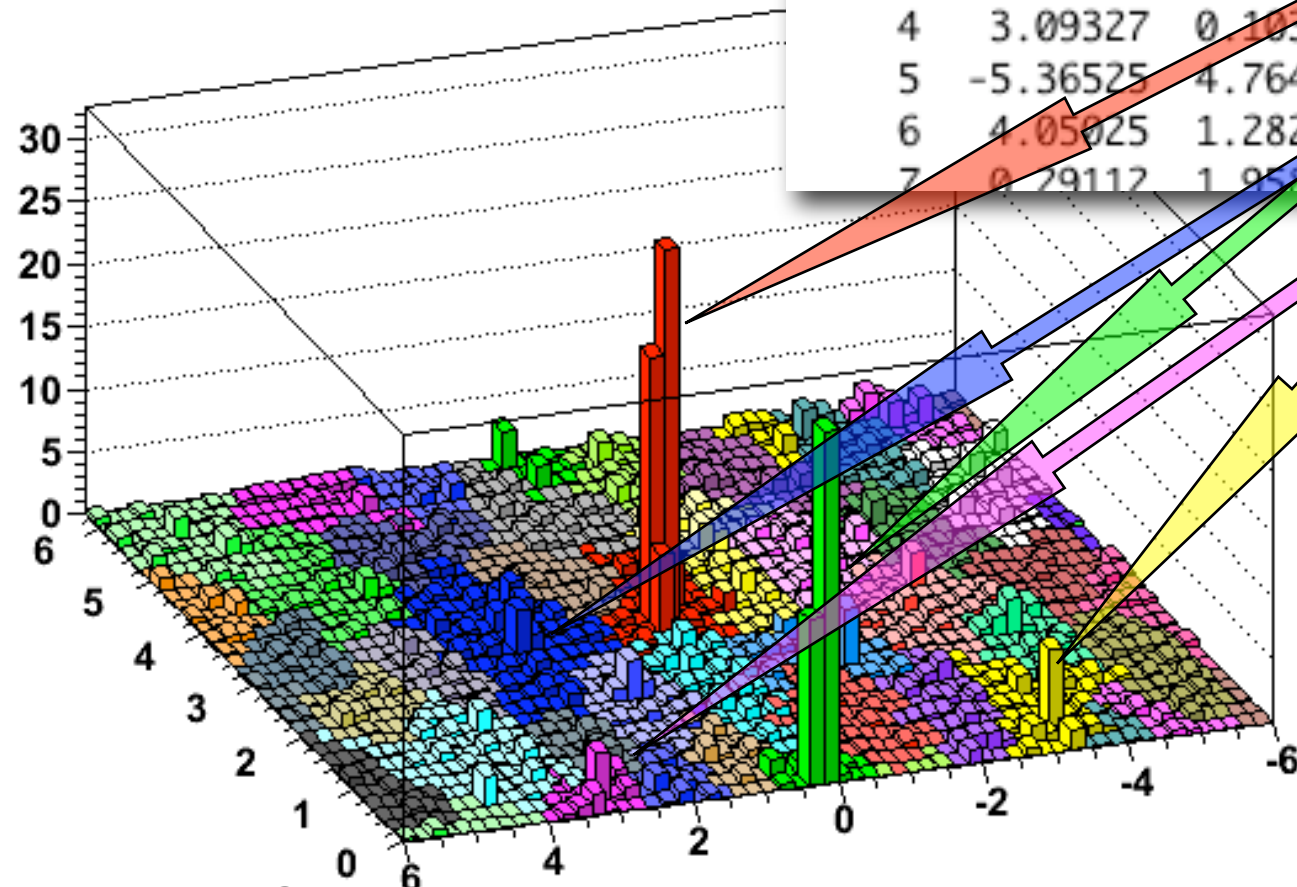
The jets are classified as belonging  
to the noise on the ground of  
their **position**

# The Active Jet Area

IRC safe algs allow the calculation of the **areas** of the jets

```
iev 0 (irepeat 24): number of particles = 1428  
strategy used = NlnN  
number of particles = 9051  
Total area: 76.0265  
Expected area: 76.0265
```

ijet	eta	phi	Pt	area +-	err
0	0.15050	3.24498	69.970	2.625 +-	0.020
1	0.18579	0.13150	59.133	1.896 +-	0.020
2	2.33840	3.23960	31.976	4.749 +-	0.028
3	-3.41796	0.52394	26.595	3.084 +-	0.021
4	3.09327	0.10350	20.072	2.688 +-	0.023
5	-5.36525	4.76491	19.594	2.780 +-	0.012
6	4.05025	1.28278	15.361	3.592 +-	0.028
7	0.29112	1.95735	14.566	2.114 +-	0.018

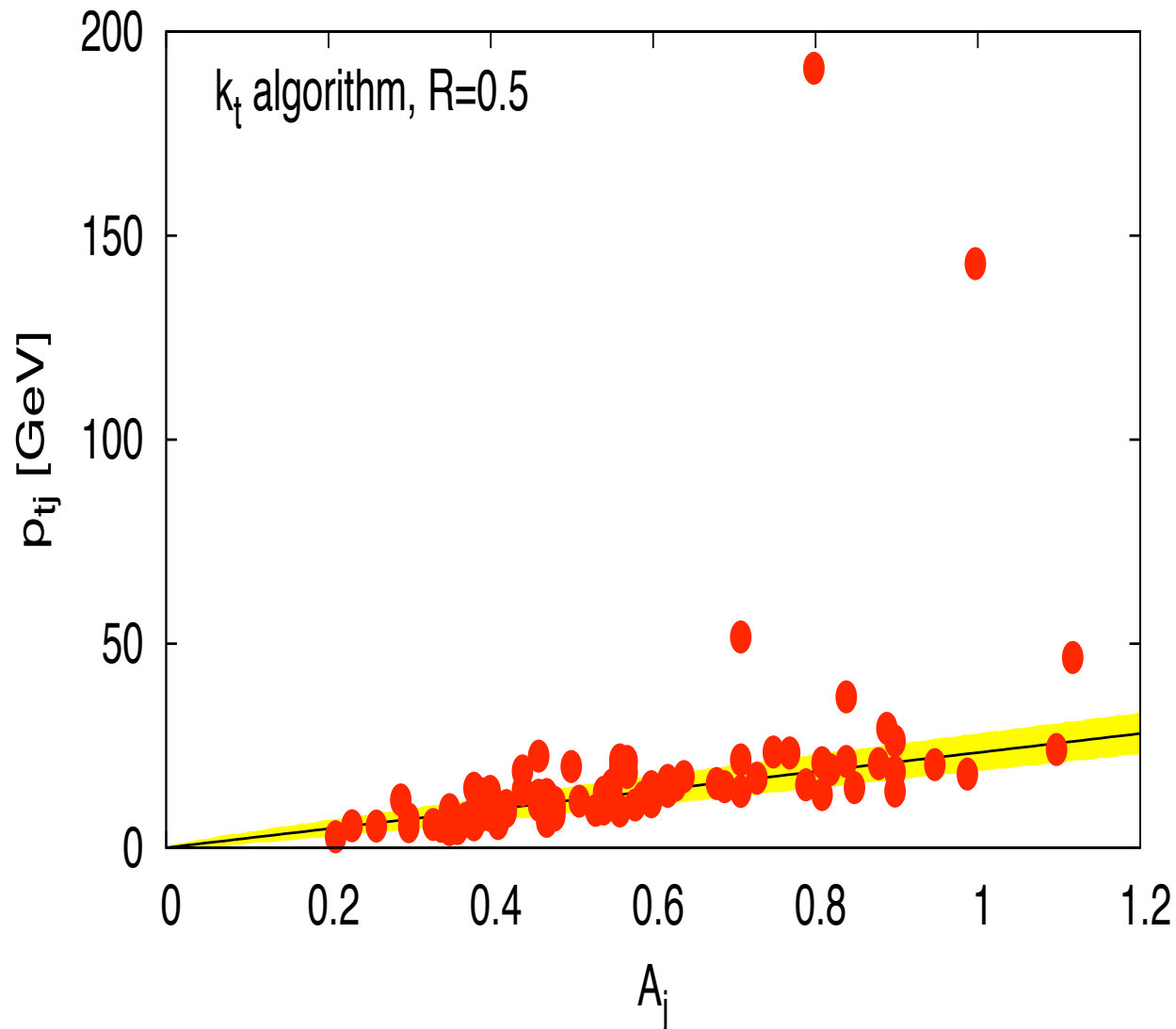


# Area vs. $p_T$

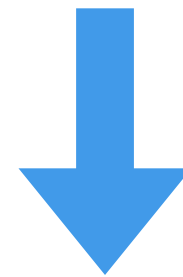
LHC: dijet event + high-lumi pileup

a few hard particles and many softer ones

(a similar picture applies to the Underlying Event)



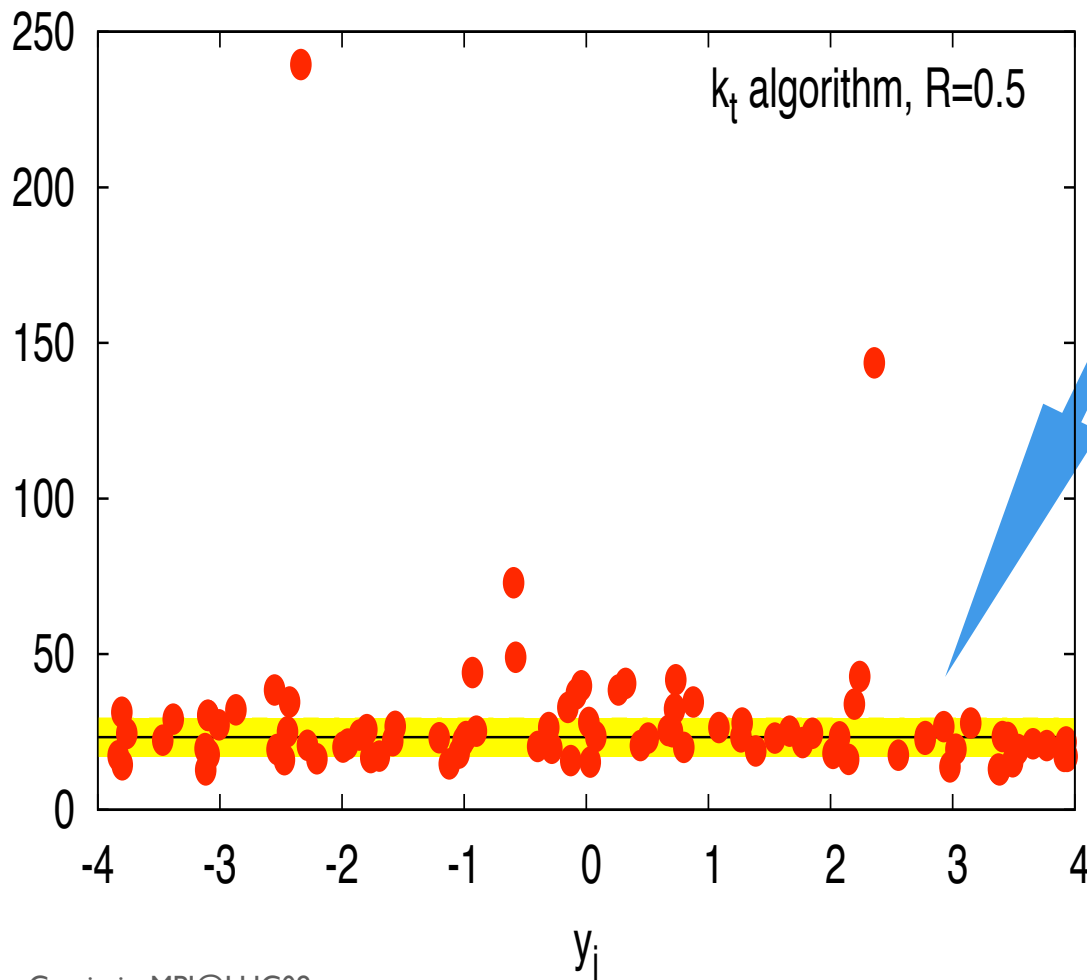
The jets adapt to the surrounding environment



They can have very different areas

# The key observation

$p_T/\text{Area}$  is fairly constant, except for the hard jets



The distribution of background jets establishes its own average momentum density  $\rho$

**(NB. this is true on an event-by-event basis)**

## Dynamical selection

The jets are classified as belonging to the noise on the ground of their **characteristics**

# Extraction of average noise momentum density

MC, Salam, arXiv:0707.1378

$$\rho \equiv \text{median} \left[ \left\{ \frac{p_t^{jet}}{\text{Area}_{jet}} \right\} \right]$$

Taking the median of the distribution is a way to get rid of (part of the) possible bias from the few hard jets

(This can still be tweaked/improved)

For  $\rho$  to be non-zero at the *perturbative* level one would need at least as many hard jets as 'empty jets'.

This can be shown to happen at order  $\alpha_s^n$ , with

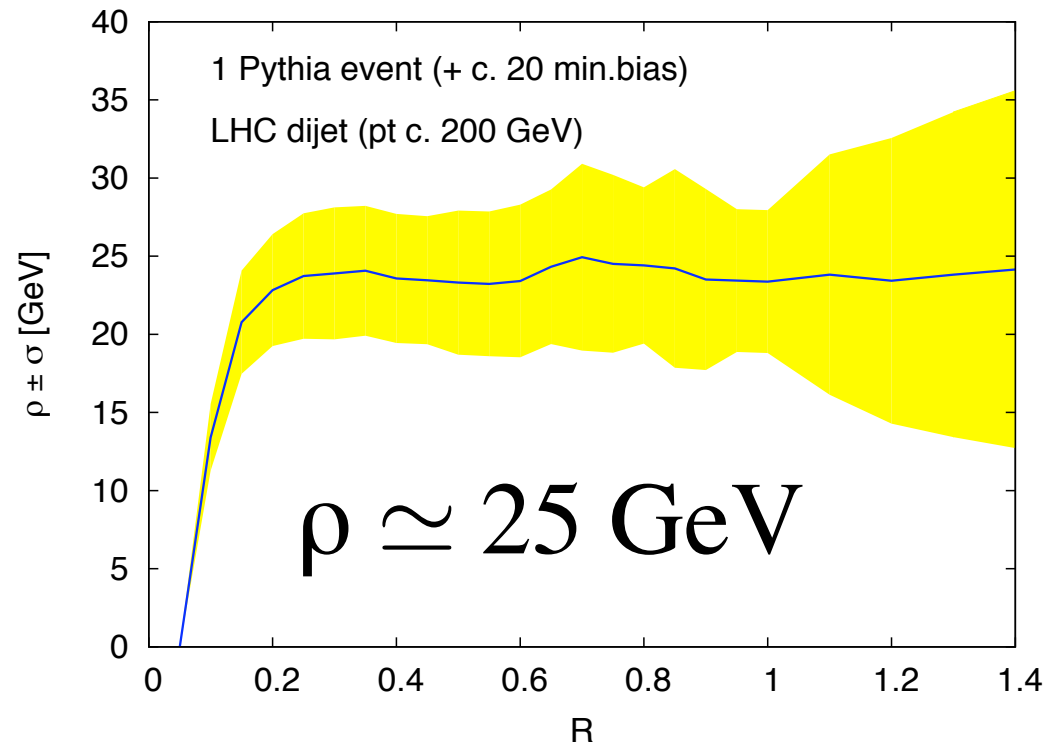
$$n \simeq \frac{4y_{\max}}{(x_{pg} + x_{sp})R^2} \simeq 2.94 \frac{y_{\max}}{R^2}$$

Factors related to typical areas of pure-ghost and single-particle jets

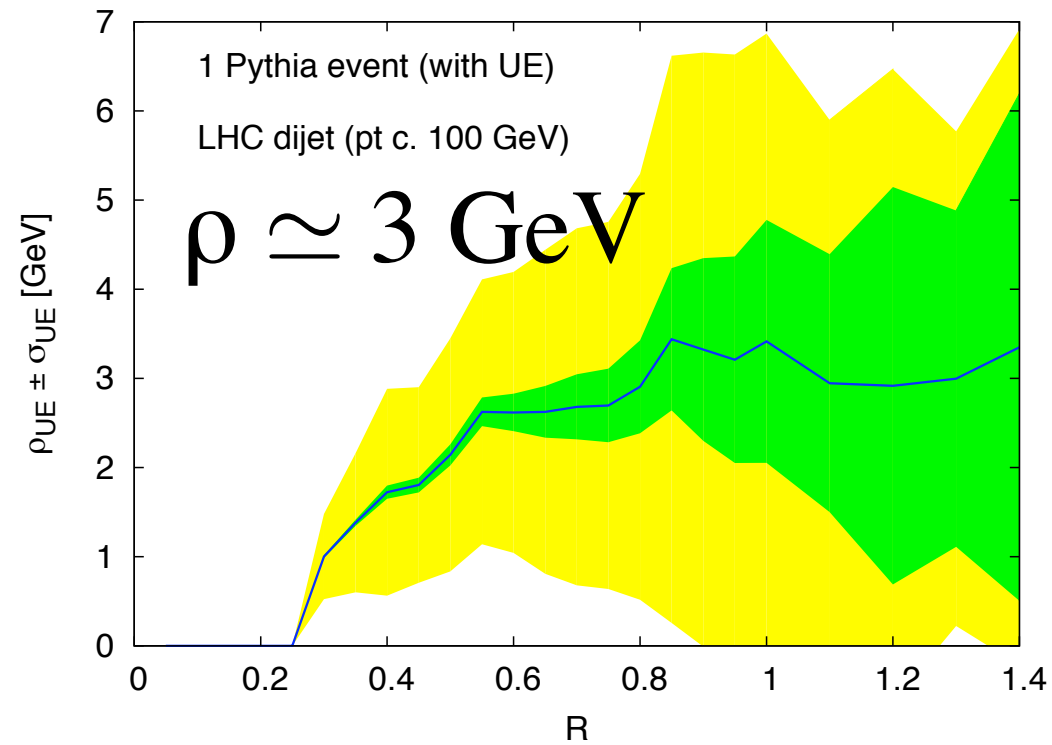
This gives  $n \sim 24\text{--}47$  for  $y_{\max} = 4$  and  $R = 0.5\text{--}0.7$

# Noise levels

## Pileup @ LHC



## UE @ LHC



When 'measuring' the background,  $R$  should be **not too small** (too many empty jets) and **not too large** (too few jets, biased by the hard particles)

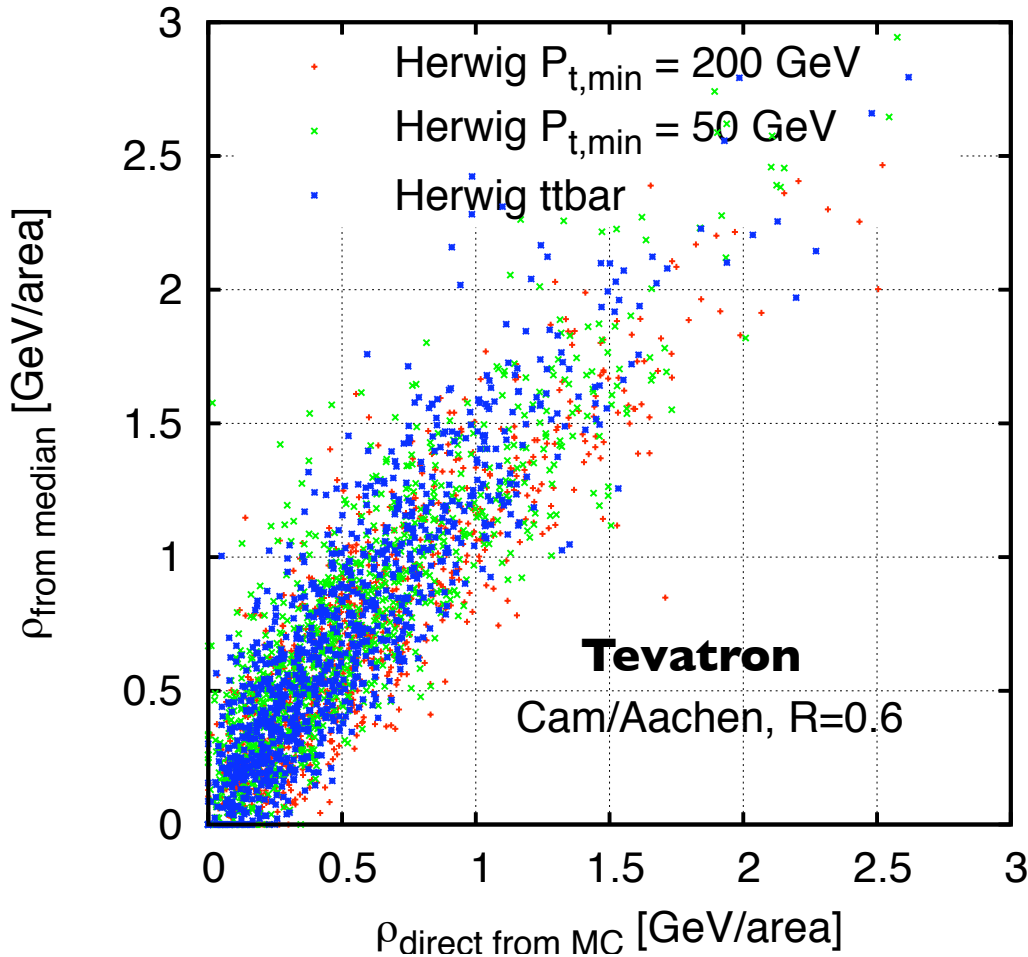
Theoretical estimates and empirical evidence point to  $R \sim 0.5\text{--}0.6$  for Underlying Event measurement (Also, a 'sensible' jet alg like  $k_t$  or Cambridge/Aachen should be used)

# Underlying Event estimation

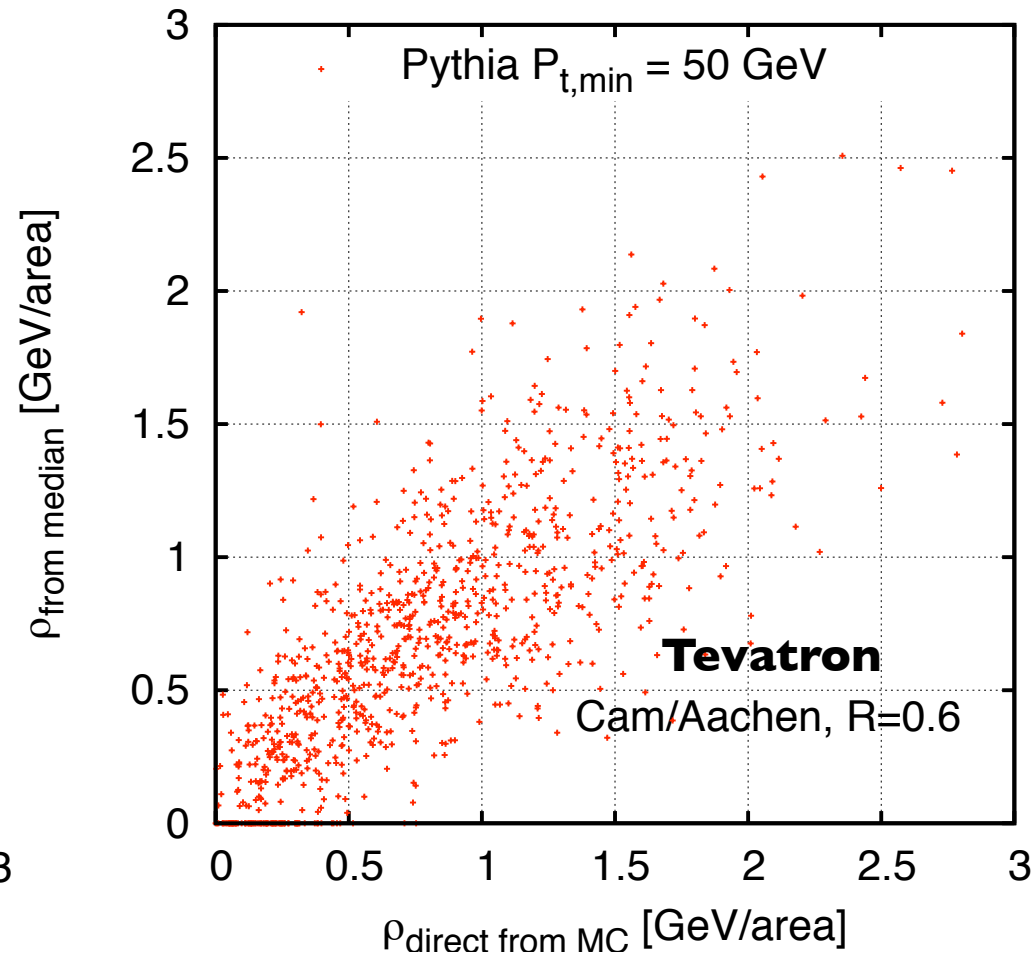
To test the procedure for the Underlying Event, compare the measurement of the background level made with areas with the known amount a Monte Carlo put in

Measurement

## HERWIG



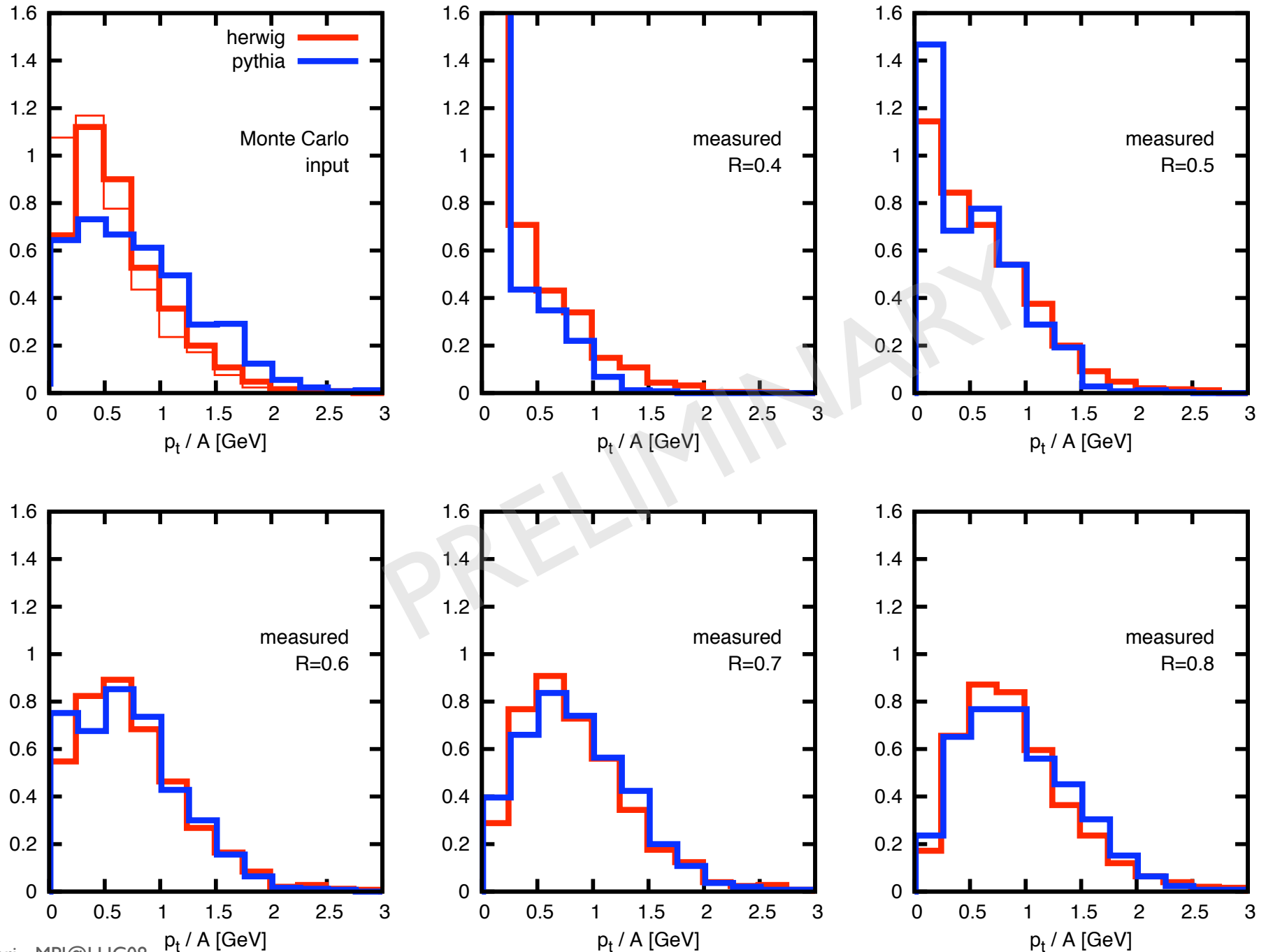
## PYTHIA



Input from Monte Carlo

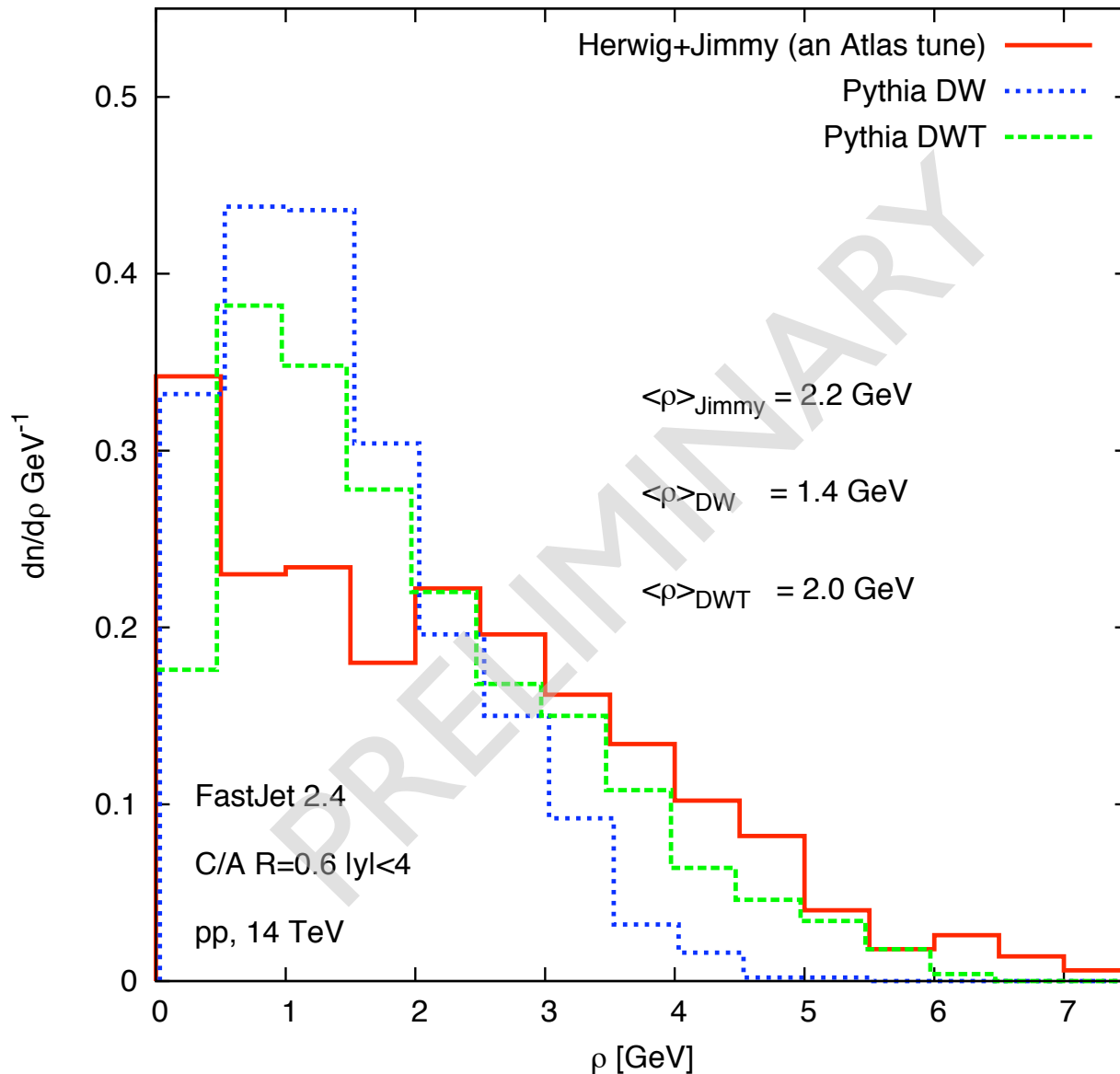
# Underlying Event estimation: Tevatron

## Tevatron



# UE at LHC

Dijet events,  $p_t > 50$  GeV,  $y < |4|$



Comparison of

**HERWIG+Jimmy,**  
**PYTHIA with DW,**  
**PYTHIA with DWT**

# Last but not least: subtraction

Once measured, the background density can be used to **correct** the transverse momentum of the hard jets:

$$p_T^{\text{hard jet, corrected}} = p_T^{\text{hard jet, raw}} - \rho \times \text{Area}_{\text{hard jet}}$$

$\rho$  being calculated on an **event-by-event basis**, this procedure will generally **improve the resolution** of, say, a mass peak

NB. Also be( a)ware of *backreaction*

(immersing a hard jet in a soft background may cause some particles belonging to the hard event to be lost from (backreaction loss) or added to (backreaction gain) the jet).

Small effect for UE, larger for pileup, can be very important for heavy ions.

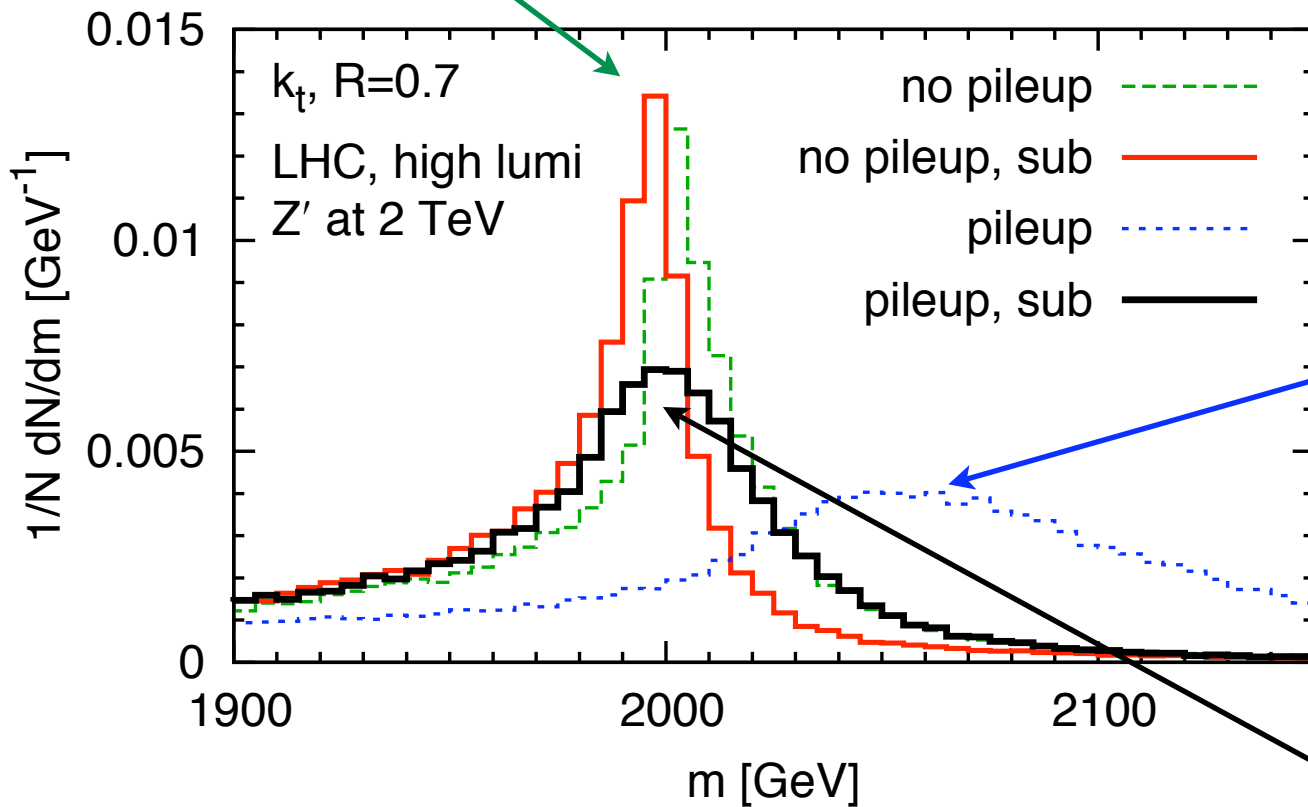
Analytical understanding of this effect available (MC, Salam, Soyez, arXiv:0802.1188)

# Subtraction: big effects with pileup....

Let's 'discover' a leptophobic Z' and measure its mass:

MC simulation:

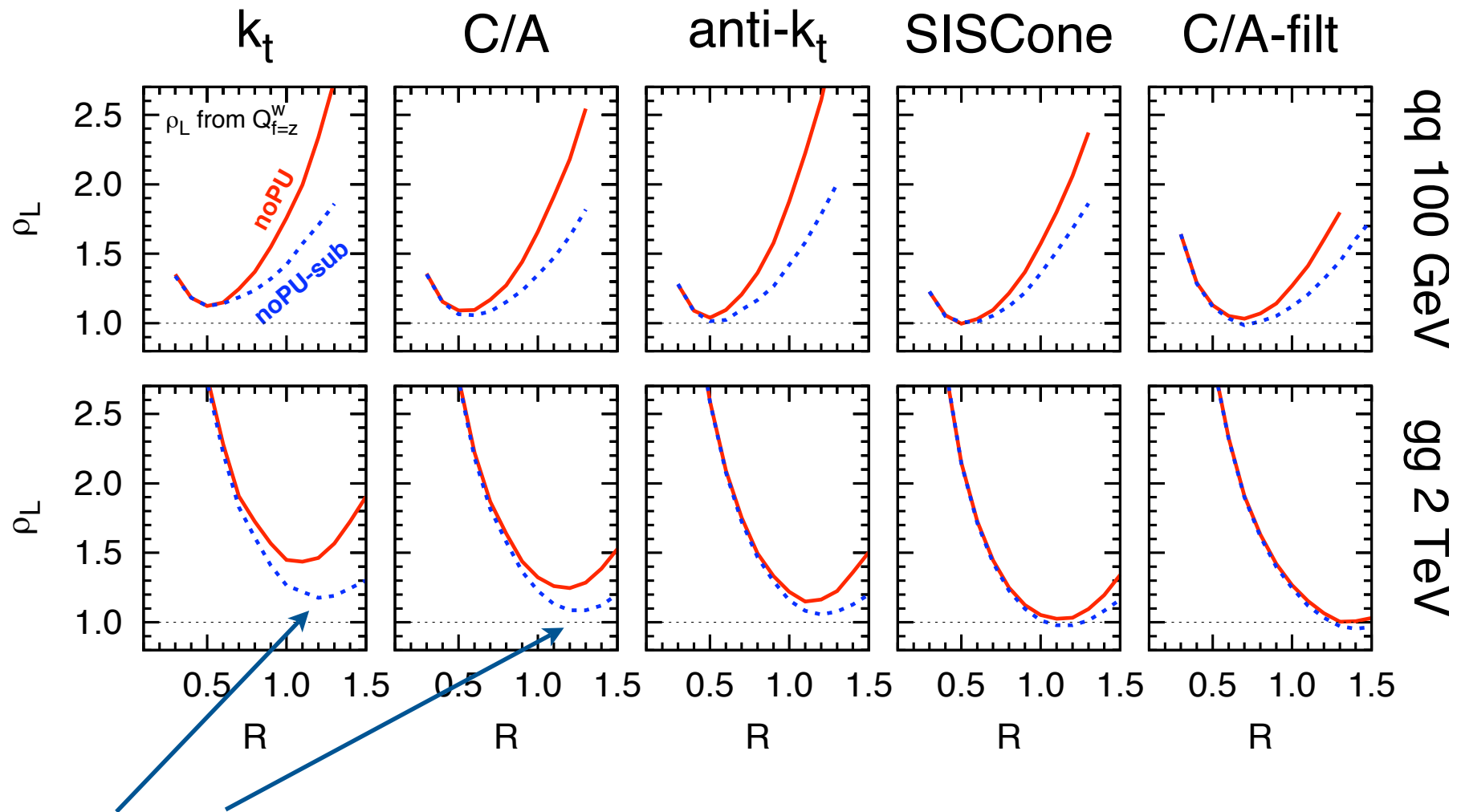
$m = 2000$  GeV, width  $\sim 10$  GeV



Naive measurement with PU:  
 $m \sim 2050$  GeV, width  $\sim 60$  GeV

Measurement after subtraction:  
 $m \sim 2000$  GeV, width  $\sim 25$  GeV

# ...but even UE is there and 'seen'







**Effect of UE subtraction: bad jet definitions are improved.**

**Gain in effective luminosity\* about 20-30%**

(i.e. they were 'bad' due to a large extent to their behaviour with respect to the UE)

\*  $\rho_L$  = factor of luminosity needed to obtain equivalent signal/background significance

# Conclusions

-  An extensive sets of fast, IRC jet algorithms exists, offering replacements for the IRC unsafe ones.
-  They offer ample flexibility in choosing the most effective jet definition for any given analysis.  
'Third-generation' algorithms look promising.
-  They can be used to estimate the level of a uniformly distributed noise, and study its characteristics.
-  They can be used to subtract the noise from the hard jets, improving the quality of kinematical reconstructions.