~~*Jet algorithms*~~

Modern jet definitions (for hadronic colliders)

Grégory Soyez

Brookhaven National Laboratory

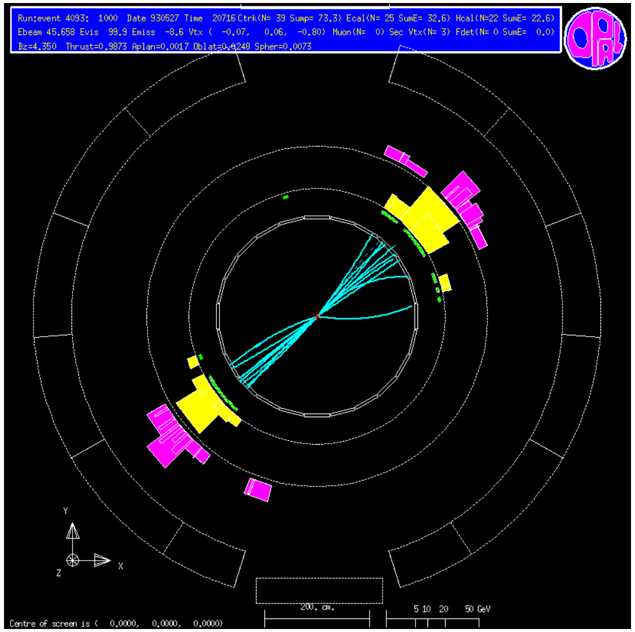
Low-x workshop — September 9-13 2009 — Ischia, Italy

Foreword: what are jets?

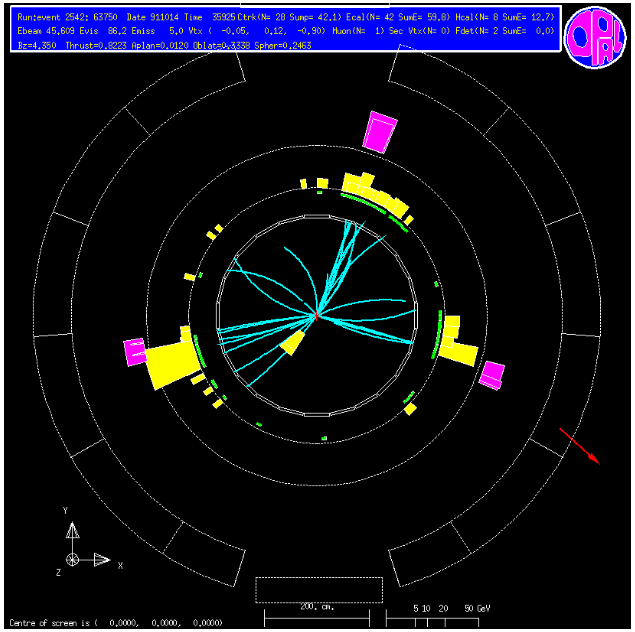
QCD collinear divergence $\longrightarrow$ collimated showers

$$\mathrm{d}P \propto \alpha_s \frac{\mathrm{d}\theta}{\theta}$$

Example: LEP (OPAL) events



2 jets



3 jets

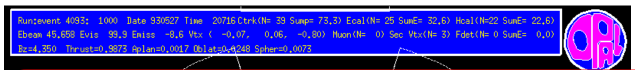
Idea: jet $\equiv$ collimated shower $\simeq$ initial hard partons

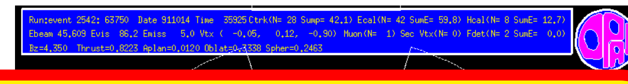
Foreword: what are jets?

QCD collinear divergence $\longrightarrow$ collimated showers

$\theta \qquad \mathrm{d}P \propto \alpha_s \frac{\mathrm{d}\theta}{\theta}$

Example: LEP (OPAL) events





In practice: use of a jet definition

particles $\{p_i\}$ $\xrightarrow[\text{definition}]{\text{jet}}$ jets $\{j_k\}$

2 jets

3 jets

Idea: jet $\equiv$ collimated shower $\simeq$ initial hard partons

Plan

- Jet algorithms and jet definitions
  - e^+e^- collisions: the most simple case
  - pp collisions: most simple casejets in hadronic environments
    $\Rightarrow$ modern jet algorithms
- More advanced topics: how to better use the tools we have?
  - jet areas: tool for pileup subtraction
  - new generation of algorithms
  - Optimal choice (for kinematic reconstructions)

Foreword: an illustrative example

Example process to illustrate various effects:

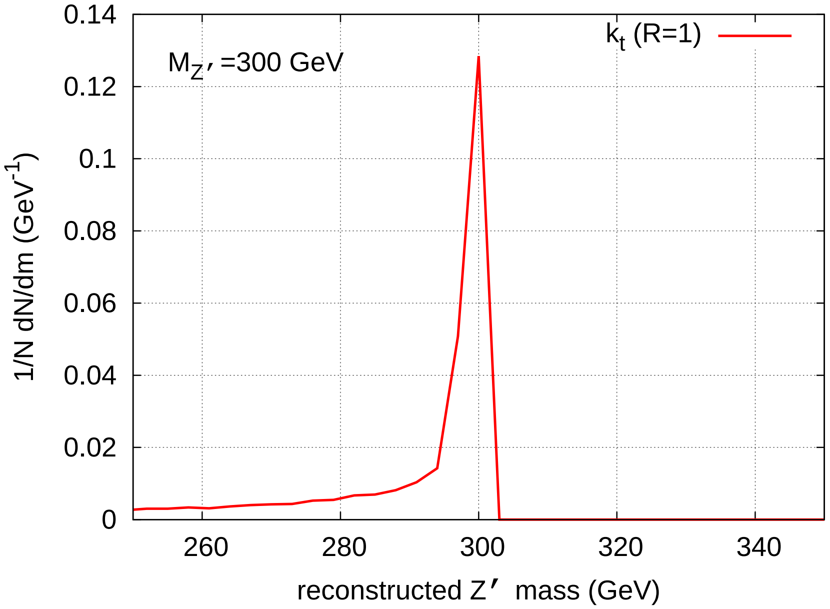
$$Z' \to q\bar{q} \to 2 \text{ jets}$$

- $M_{Z'} = 300$ GeV (width < 1 GeV)
- Reconstruction method:
  - get the 2 hardest jets: j_1 and j_2
  - reconstruct the Z': $m_{Z'} = (j_1 + j_2)^2$
- Look how the mass peak is reconstructed

e^+e^-: recombination algorithms

Recipe:

- find the pair (i,j) with the smallest d_{ij} distance
- recombine i and j
- repeat (until objects more than R apart)



Distance: k_t algorithm

$$d_{ij} = \min(E_i^2, E_j^2)[1 - \cos(\theta_{ij})]$$

- $d_{ij} \to 0$ for soft and collinear splittings

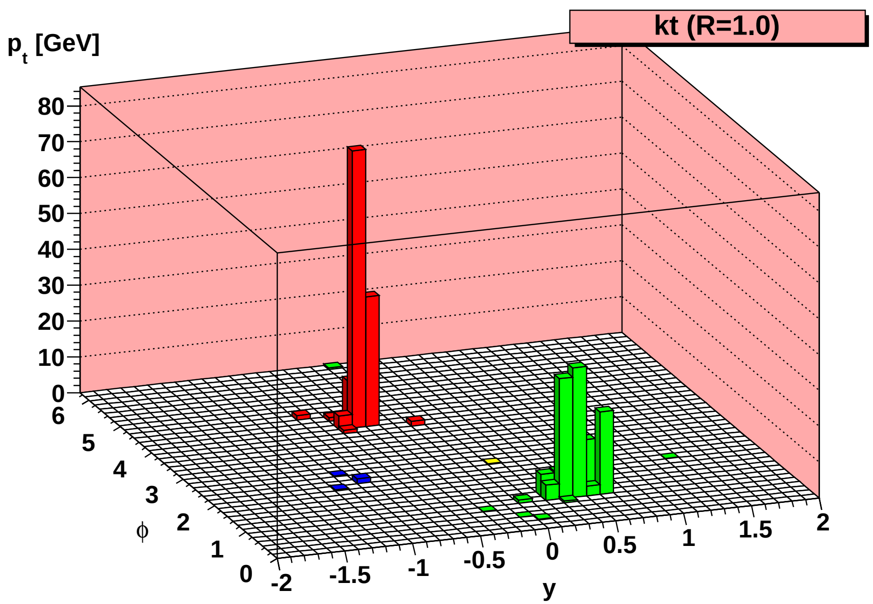
$$\mathrm{d}P = \alpha_s \frac{\mathrm{d}\theta}{\theta} \frac{\mathrm{d}p_t}{p_t}$$

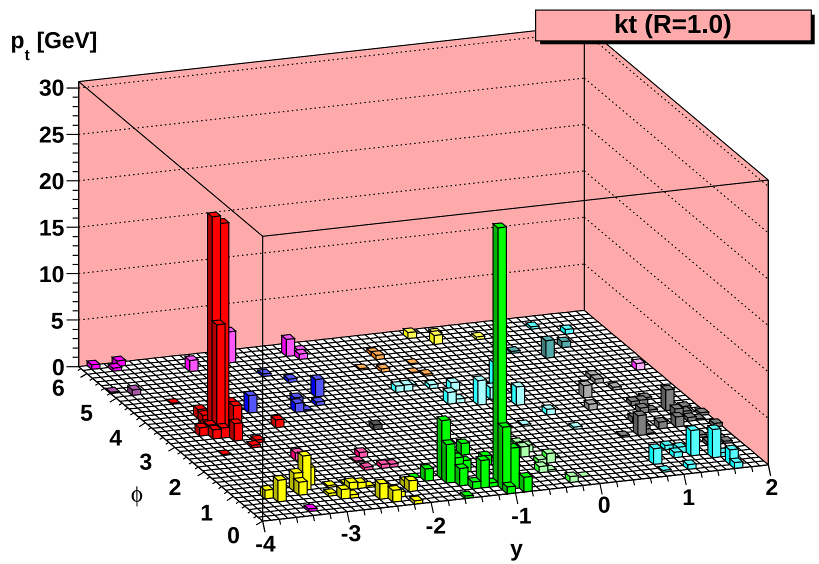
- without the prefactor: Cambridge/Aachen

pp collisions: jets in hadronic environments

Initial-state radiation + underlying event ⇒ noisier environment

e^+e^- $\qquad\qquad\qquad$ pp

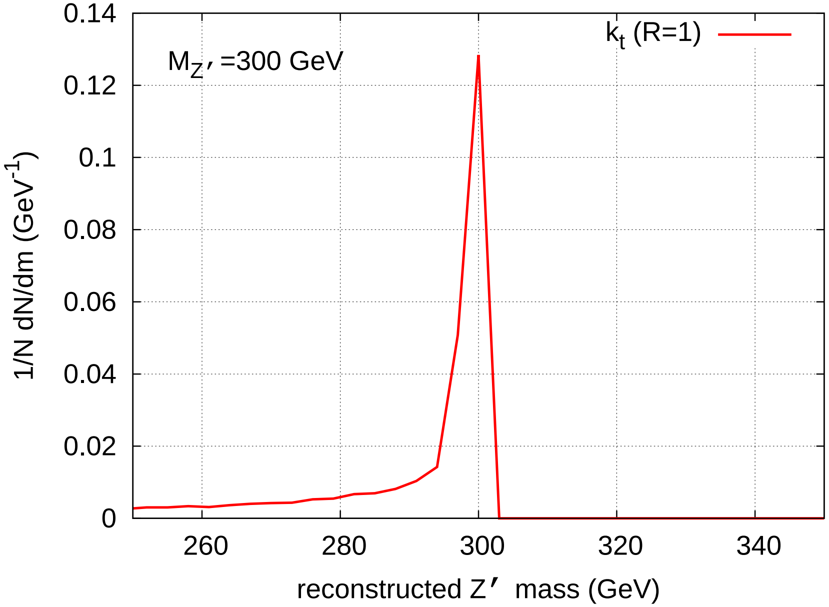




$$d_{ij} = \min(E_i^2, E_j^2)[1 - \cos(\theta_{ij})] \qquad d_{ij} = \min(k_{t,i}^2, k_{t,j}^2)(\Delta y_{ij}^2 + \Delta\phi_{ij}^2)$$

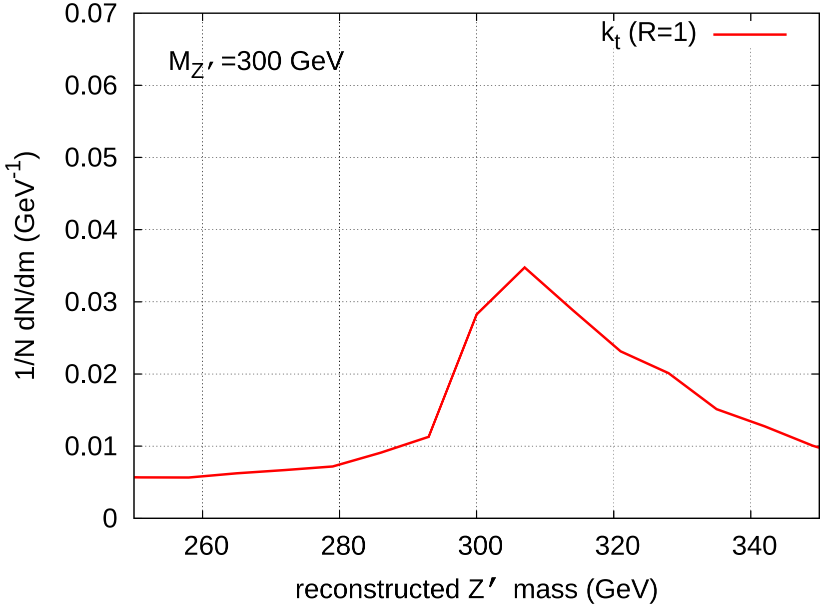
Our example: moving to jets in pp

e^+e^-



width = 0.9 GeV

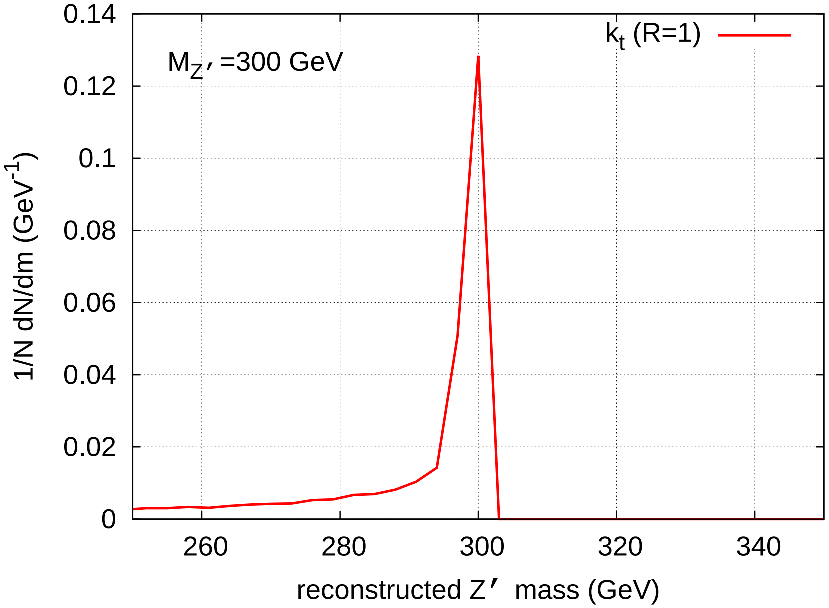
pp



width = 19.7 GeV

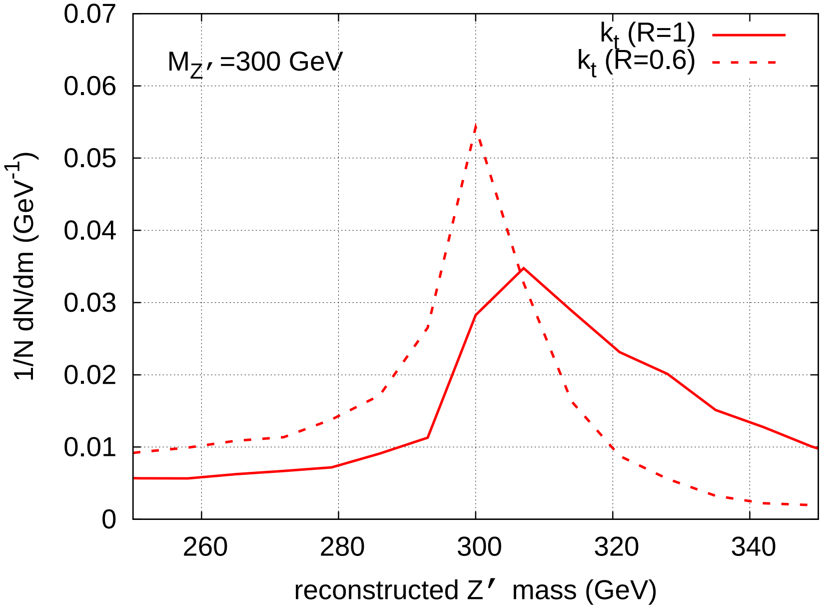
Our example: moving to jets in pp

e^+e^-



width = 0.9 GeV

pp



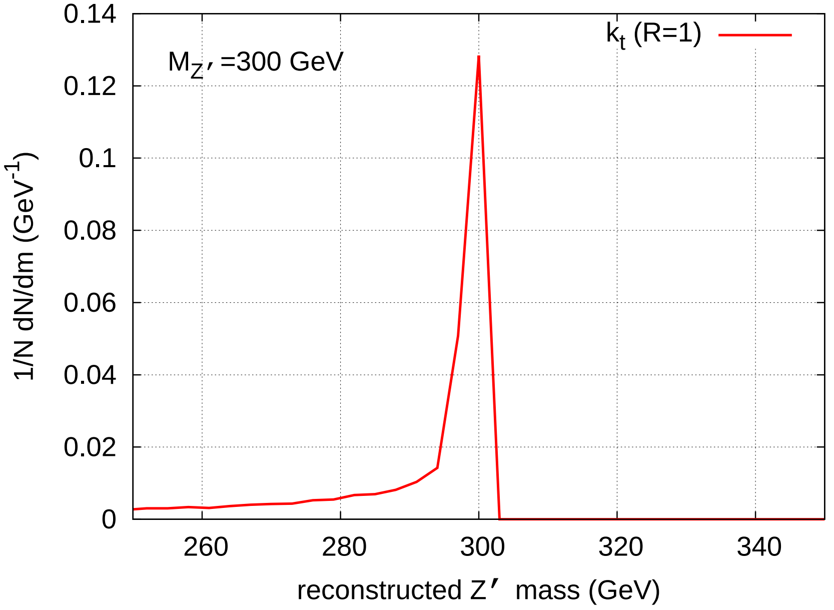
width = 19.7 GeV

width = 14.2 GeV

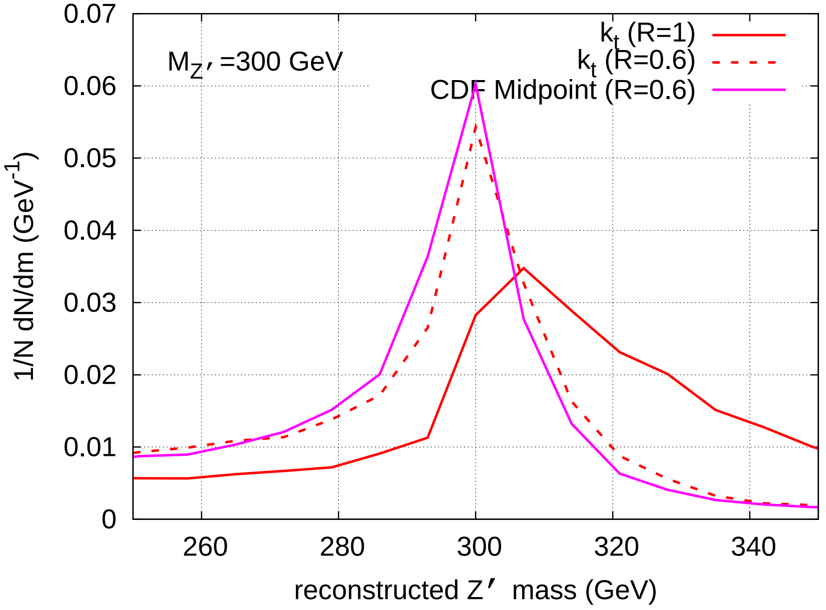
- Reduce R

Our example: moving to jets in pp

e^+e^-



pp



width = 0.9 GeV

width = 19.7 GeV

width = 14.2 GeV

width = 11.8 GeV

- Reduce R
- Use the cone algorithm

The cone algorithm

- Idea: find directions of energy flow
- Stable cone (fixed radius R):
  sum of all constituents points in the direction of the centre
- Search: iterate from initial directions (seeds)
- Stable cones $\rightarrow$ jets: deal with overlaps
  - Solution 1: split/merge depending on the amount of overlap
    Examples: CDF JetClu, CDF MidPoint, D0 runII, ATLAS Cone
  - Solution 2: progressive removal starting from the hardest seed
    Examples: CMS Iterative Cone
    Benchmark: circular/rigid jets

Constraints

1990: SNOWMASS accords – constraints to fulfil

Several important properties that should be met by a jet definition are [3]:

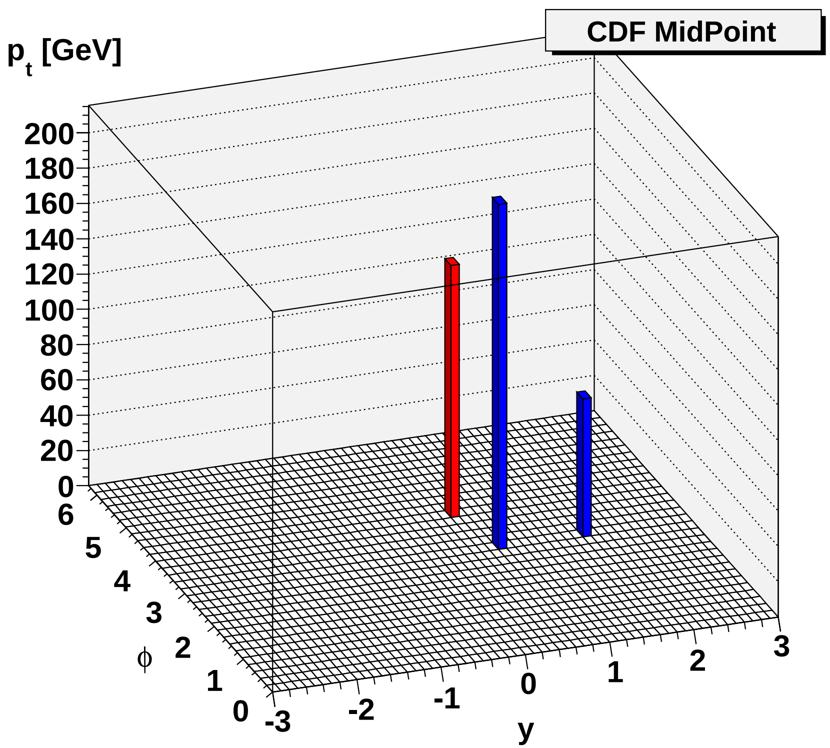
1. Simple to implement in an experimental analysis;
2. Simple to implement in the theoretical calculation;
3. Defined at any order of perturbation theory;
4. Yields finite cross section at any order of perturbation theory;
5. Yields a cross section that is relatively insensitive to hadronization.

Infrared and collinear (IRC) safety:

- ✓ recombination algorithms: Ok
- ✗ Cone: problem inherent to the “seeded” search of stable cones

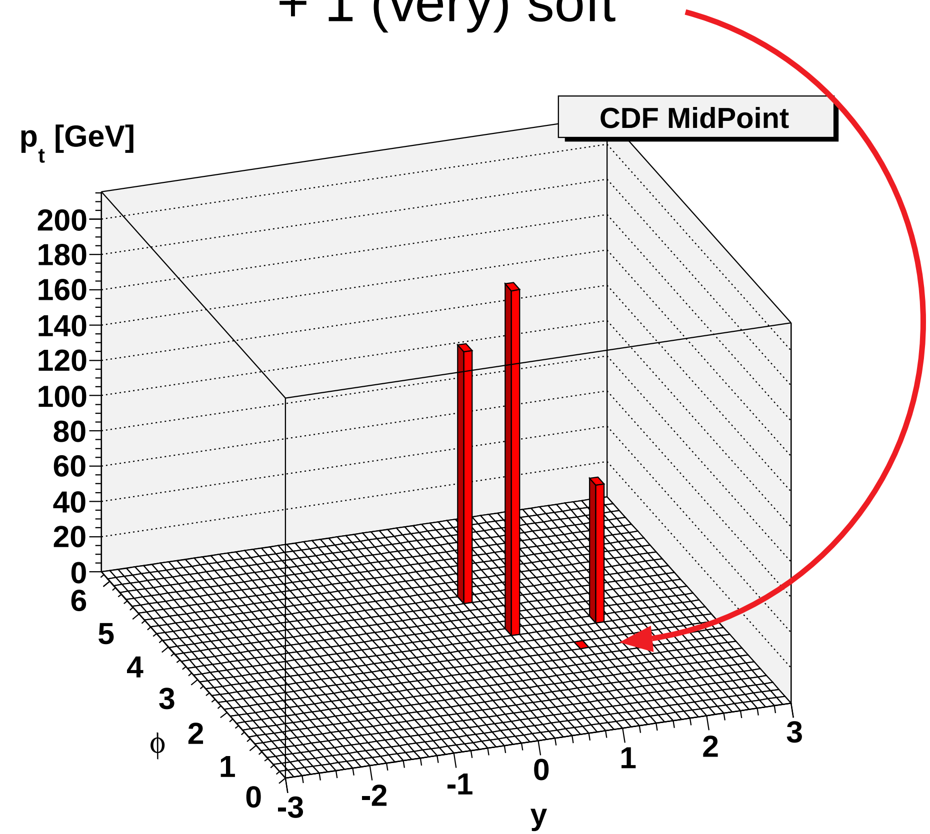
Example: infrared unsafety

3 hard particles



1 jet

+ 1 (very) soft



1 jet

leads to ∞ in perturbative jet cross-sections

$\Rightarrow$ problem with the “MidPoint Cone” (CDF,D0) and the ATLAS cone

New algorithms

- Cone with split–merge (CDF JetClu, CDF/D0 Midpoint, ATLAS cone)
  - Idea: Find an efficient seedless implementation
    that provably identifies all stable cones
  - Solution: SISCone: Seedless Infrared-Safe Cone

[G.Salam, GS, 07]

- Cone with progressive removal (CMS Iterative Cone)
  - Idea: keep the rigidity of the jets + restore IRC-safety
  - Solution: anti-k_t:

$$d_{ij} = \min(k_{t,i}^{-2}, k_{t,j}^{-2})(\Delta y_{ij}^2 + \Delta\phi_{ij}^2)$$

[M.Cacciari, G.Salam, GS, 08]

Impact of these new algorithms

Unsafe

- CMS Iterative Cone
- ATLAS Cone
- CDF JetClu
- CDF MidPoint
- D0 run II Cone

Safe

- k_t
- Cambridge/Aachen
- anti-k_t
- SISCone

Impact

Observable	Last meaningful order: MidPoint/CMS	Last meaningful order: JetClu/ATLAS
Inclusive jet cross sect.	NLO	LO (NLOJet: NLO)
3 jet cross section	LO	none (NLOJet: NLO)
$W/Z/H$ + 2 jet x-sect.	LO	none (MCFM: NLO)
jet masses in 3 jets	none	none (NLOJet: LO)

!

Huge effort (~ 100 M$) to compute processes in pQCD

anti-k_t adopted as default by ATLAS and CMS

Impact of these new algorithms

Unsafe	Safe
CMS Iterative Cone	k_t
ATLAS Cone	Cambridge/Aachen
CDF JetClu	anti-k_t
CDF MidPoint	SISCone
D0 run II Cone	

Impac

er

TLAS

et: NLO)

Jet: NLO)

FM: NLO)

Jet: LO)

```
#--------------------------------------------------------------------------
#                         FastJet release 2.4
#          Written by M. Cacciari, G.P. Salam and G. Soyez
#                         http://www.fastjet.fr
#--------------------------------------------------------------------------
```

All those algorithms (and much more) implemented (efficiently) in FastJet

! Huge effort (~ 100 M\$) to compute processes in pQCD

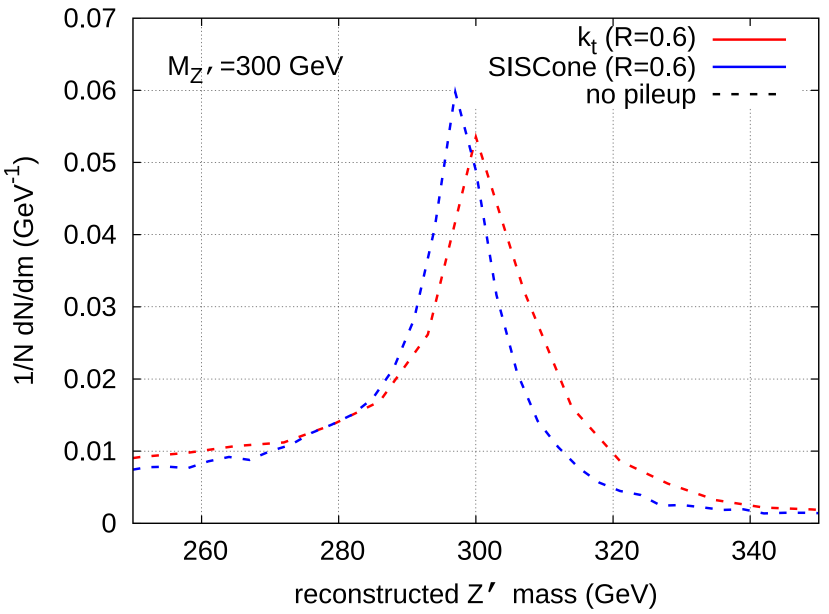
anti-k_t adopted as default by ATLAS and CMS

We (finally) have a good set of tools

Can we do better?

Example 1: Pileup effect

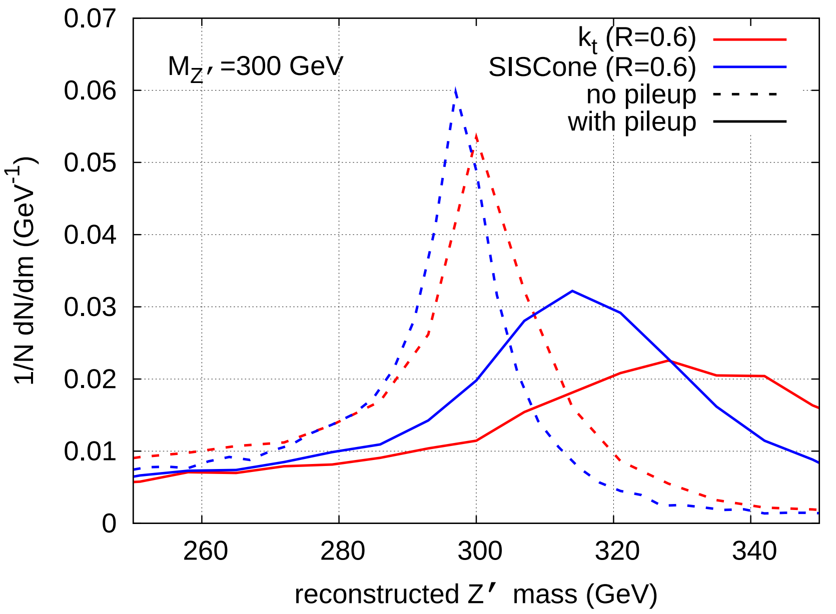
No pileup



width = 14.2 GeV

width = 11.6 GeV

With pileup



width = 29.5 GeV

width = 21.0 GeV

- ✗ shifted towards larger masses
- ✗ width increased

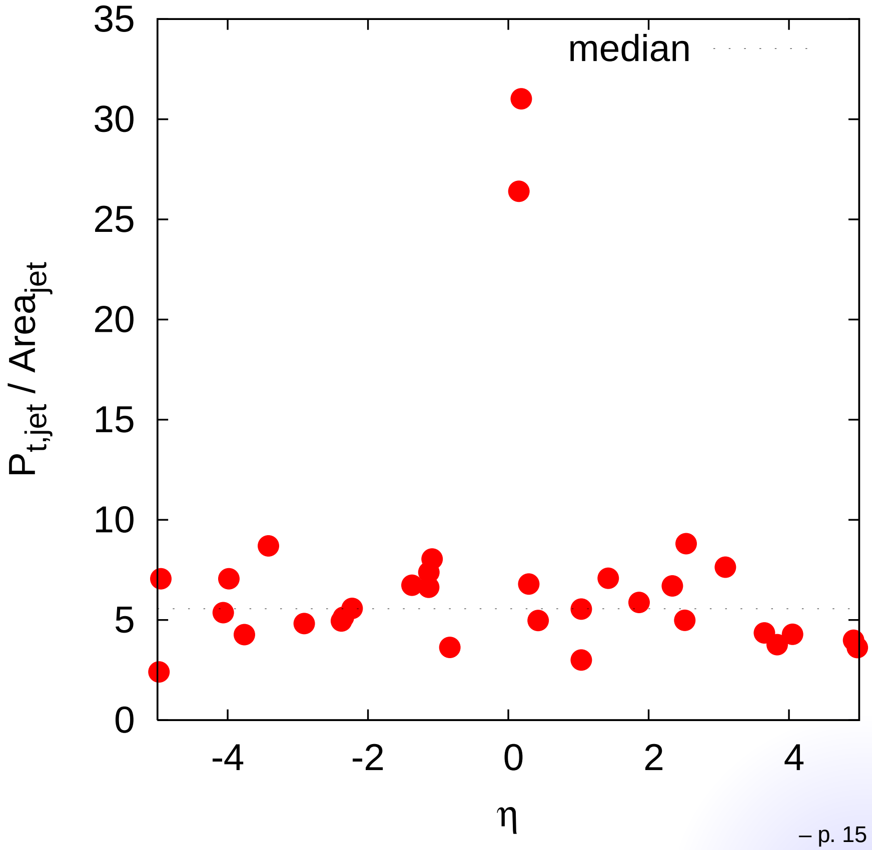
Pileup subtraction *(for uniform backgrounds)*

Basic idea: [M.Cacciari, G.Salam, 08]

$$p_{t,\mathrm{subtracted}} = p_{t,\mathrm{jet}} - \rho_{\mathrm{pileup}} \times \mathrm{Area}_{\mathrm{jet}}$$

- Jet area: [M.Cacciari, G.Salam, G.S., 08]
  - region where the jet catches infinitely soft particles (active/passive)
  - analytic control and understanding in pQCD
- Pileup density per unit area: ρ_{pileup}

  e.g. estimated from the median of $p_{t,\mathrm{jet}}/\mathrm{Area}_{\mathrm{jet}}$



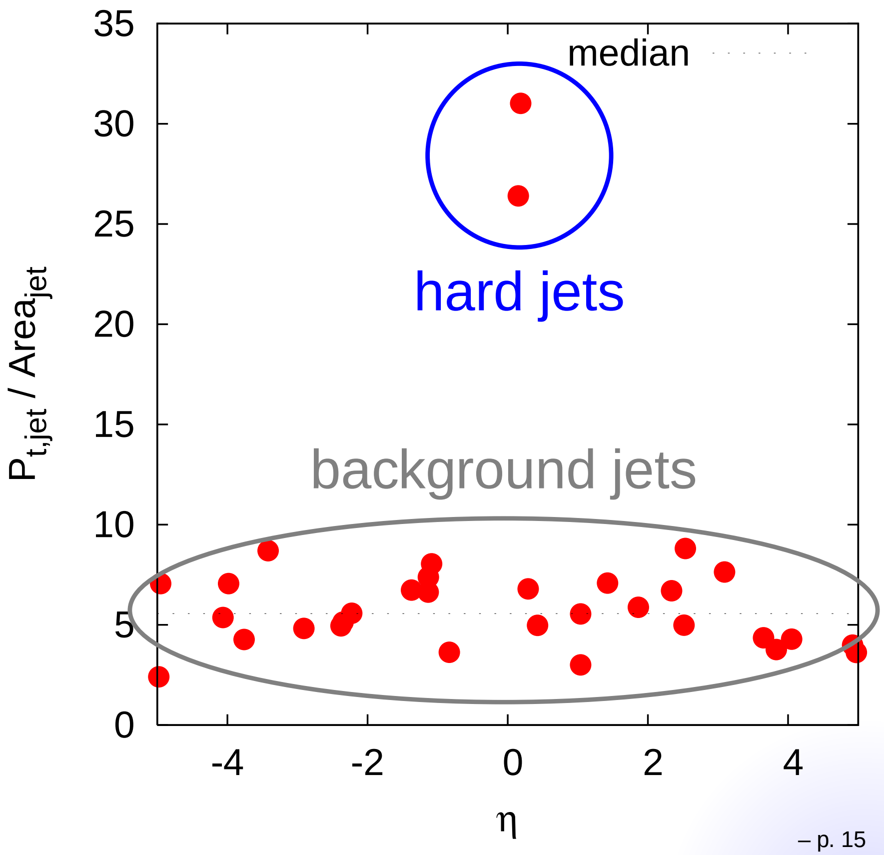
Pileup subtraction *(for uniform backgrounds)*

Basic idea: [M.Cacciari, G.Salam, 08]

$$p_{t,\mathrm{subtracted}} = p_{t,\mathrm{jet}} - \rho_{\mathrm{pileup}} \times \mathrm{Area}_{\mathrm{jet}}$$

- Jet area: [M.Cacciari, G.Salam, G.S., 08]
  - region where the jet catches infinitely soft particles (active/passive)
  - analytic control and understanding in pQCD
- Pileup density per unit area: ρ_{pileup}

  e.g. estimated from the median of $p_{t,\mathrm{jet}}/\mathrm{Area}_{\mathrm{jet}}$



Pileup subtraction *(for uniform backgrounds)*

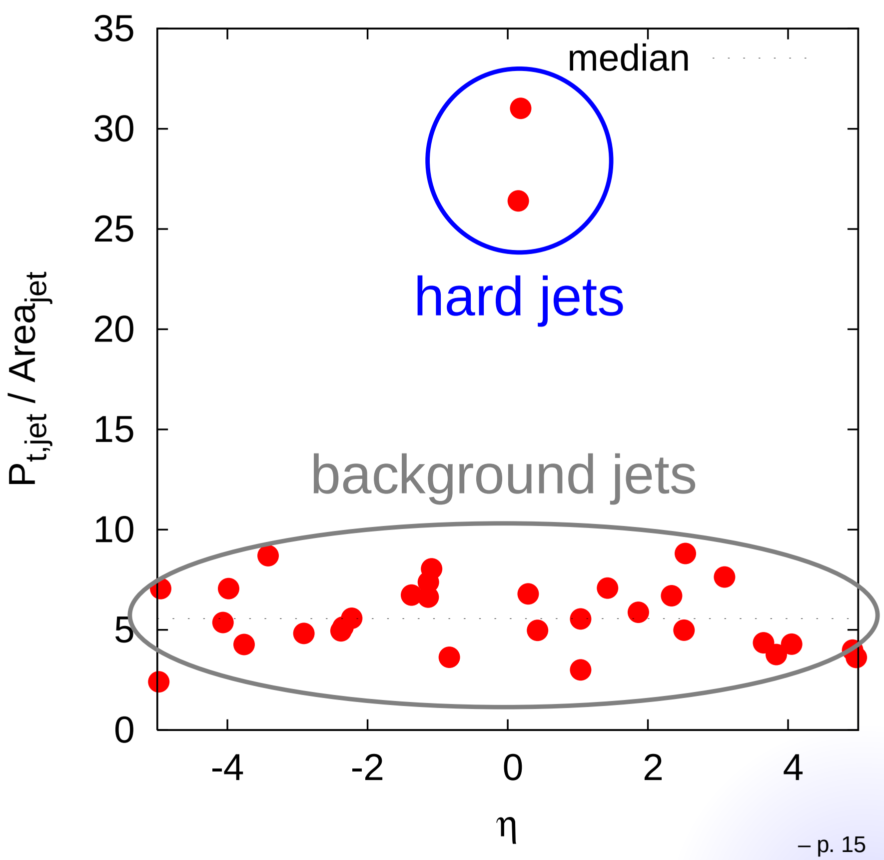
Basic idea: [M.Cacciari, G.Salam, 08]

$$p_{t,\text{subtracted}} = p_{t,\text{jet}} - \rho_{\text{pileup}} \times \text{Area}_{\text{jet}}$$

- Jet area: [M.Cacciari, G.Salam, G.S., 08]
  - region where the jet catches infinitely soft particles (active/passive)
  - analytic control and understanding in pQCD
- Pileup density per unit area: ρ_{pileup}

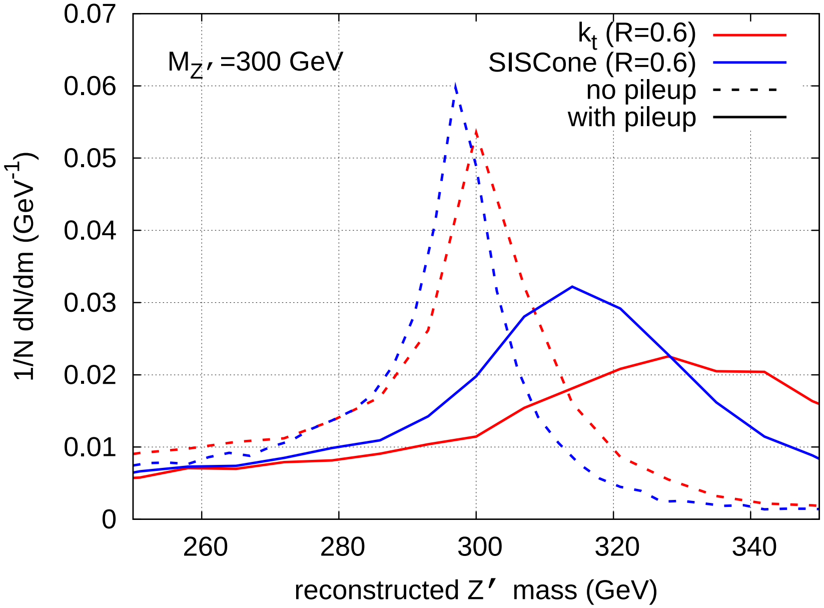
  e.g. estimated from the median of $p_{t,\text{jet}}/\text{Area}_{\text{jet}}$

implemented in FastJet
on an event-by-event basis



Our example: subtracting pileup

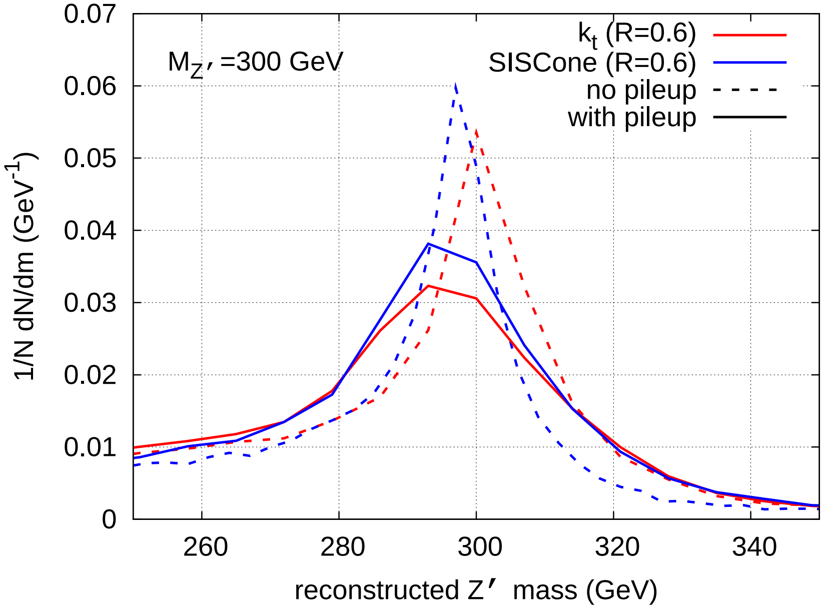
Pileup unsubtracted



width = 29.5 GeV

width = 21.0 GeV

pileup subtracted



width = 21.0 GeV

width = 17.7 GeV

✓ position reasonnable

✓ dispersion reduced (thanks to the event-by-event approach)

✓ used by STAR for the first jet analysis in heavy-ions

Example 2: new generation of algorithms

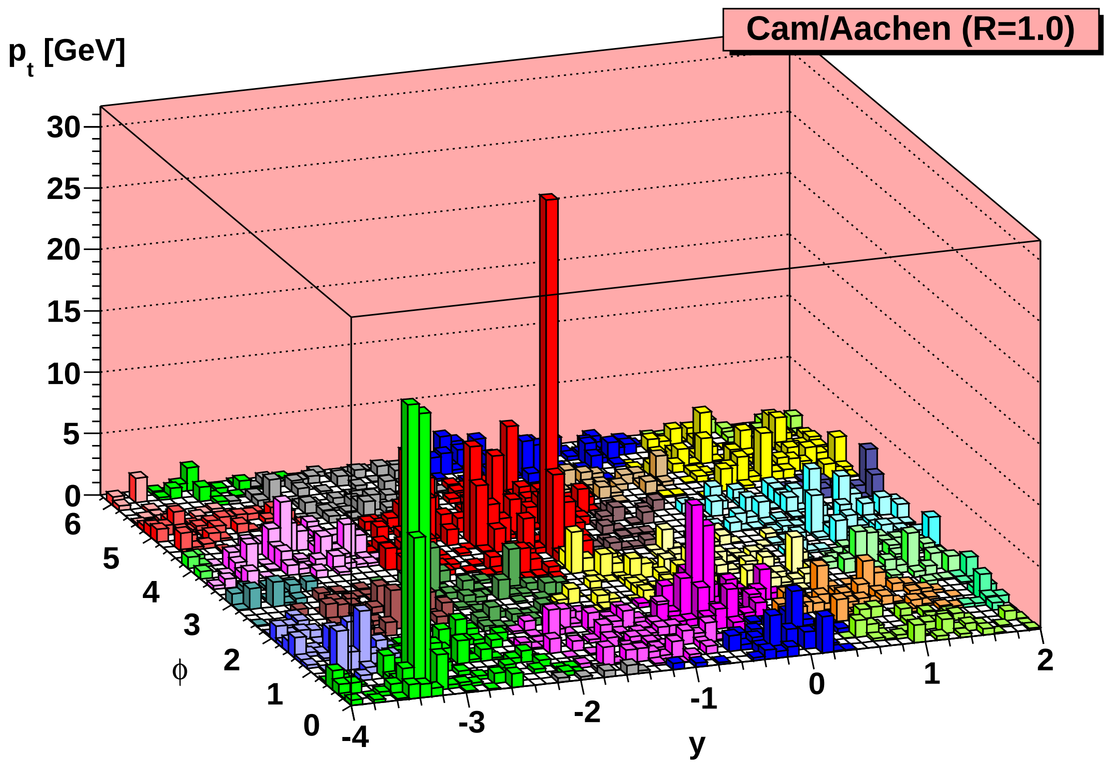
- Jet substructure
  - Filtering: reduce sensitivity to the UE

    [J. Butterworth, A. Davison, M. Rubin, G. Salam, 08]
  - Boosted objets tagging: Higgs, top

    [J. Butterworth, A. Davison, M. Rubin, G. Salam, 08]
    [D. Kaplan, K. Rehermann, M. Schwartz, B. Tweedie, 08]
- Parameter optimisation
  - kinematical reconstruction:
    optimal R larger for heavier objects, and for gluon (vs. quark) jets [M. Cacciari, J. Rojo, G. Salam, G.S., 08]
  - Variable R

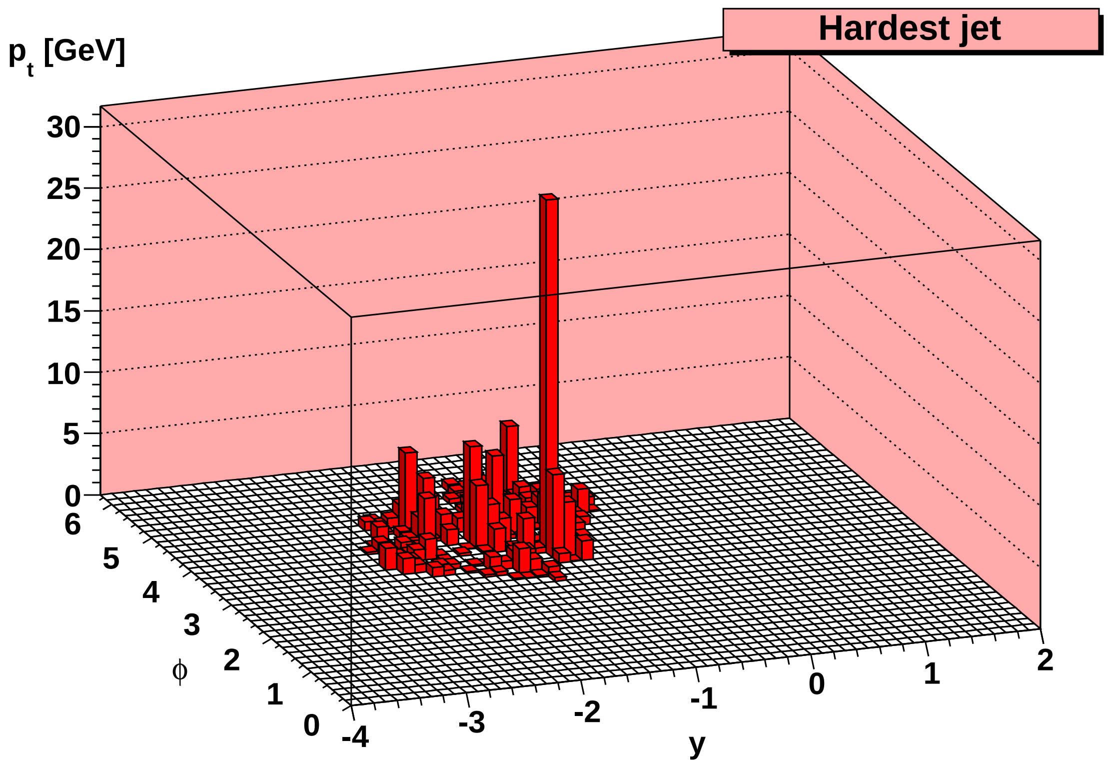
    [D. Krohn, J. Thaler, L-T. Wang, 09]

Filtering



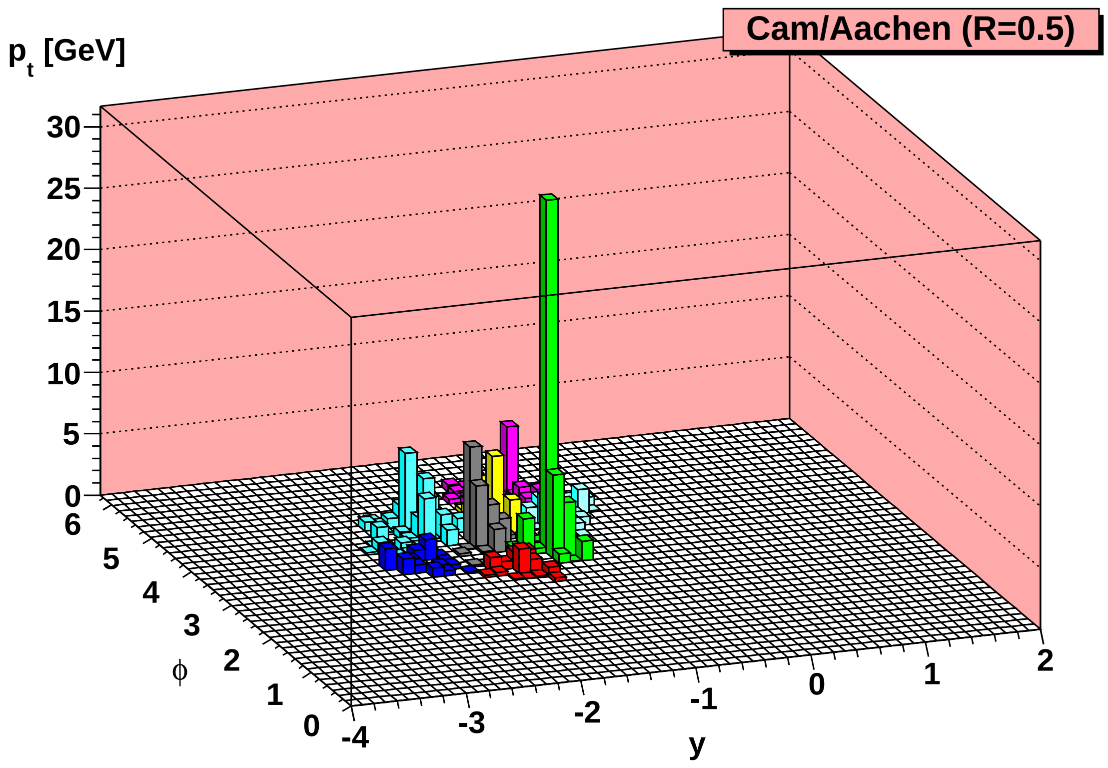
- cluster with Cambridge/Aachen(R)

Filtering



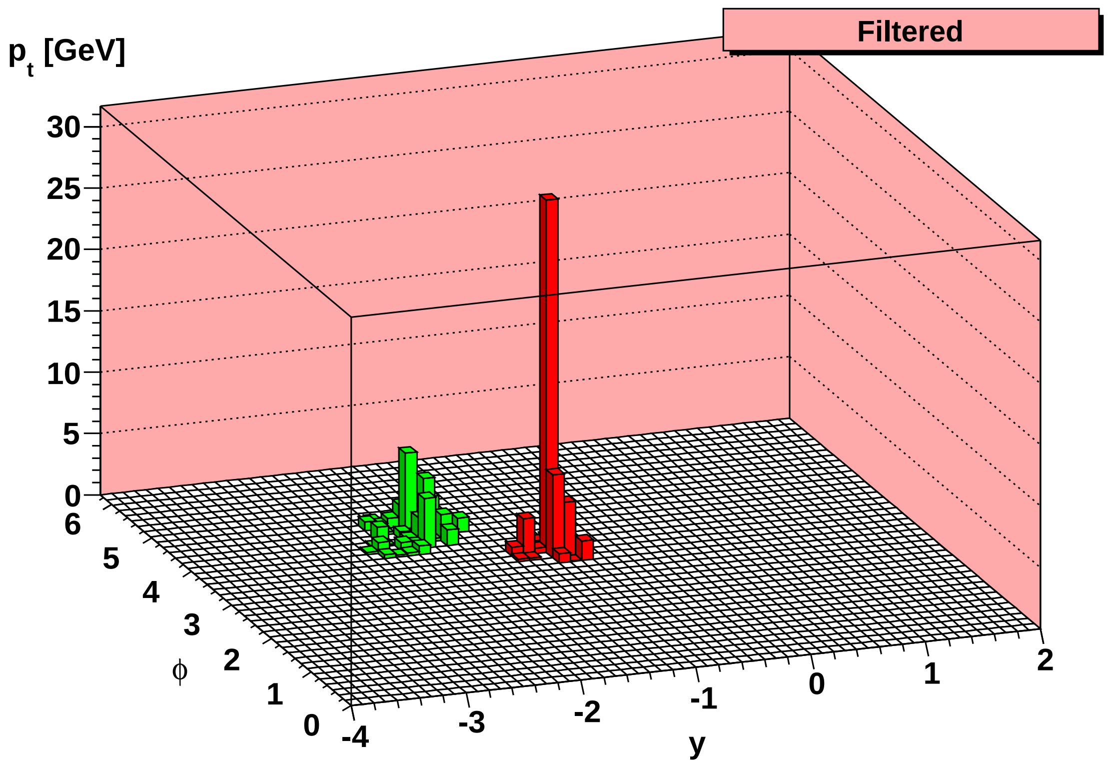
- cluster with Cambridge/Aachen(R)
- for each jet

Filtering



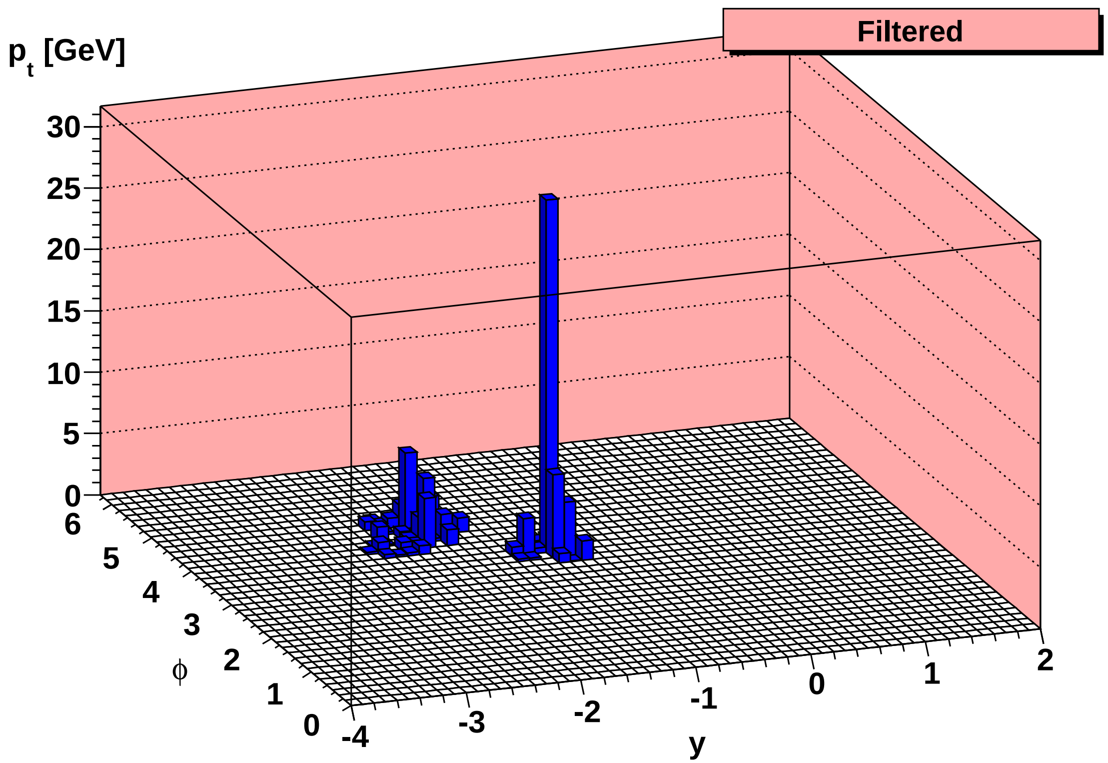
- cluster with Cambridge/Aachen(R)
- for each jet
  - recluster with Cambridge/Aachen(R/2)

Filtering



- cluster with Cambridge/Aachen(R)
- for each jet
  - recluster with Cambridge/Aachen(R/2)
  - keep the 2 hardest subjets

Filtering



- cluster with Cambridge/Aachen(R)
- for each jet
  - recluster with Cambridge/Aachen(R/2)
  - keep the 2 hardest subjets

Idea:

✓ keep perturb. radiation

✓ remove UE

- Proven useful for boosted jet $H \to b\bar{b}$ tagging

  [J.Butterworth, A.Davison, M.Rubin, G.Salam, 08]

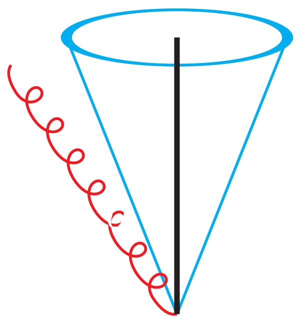
- Proven useful for kinematic reconstructions

  [M.Cacciari, J.Rojo, G.Salam, GS, 08]

Optimisation: underlying idea

Competition between

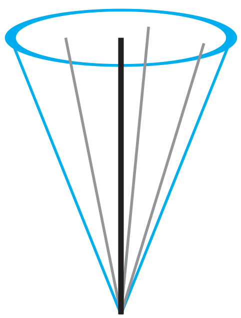
- catching perturbative radiation



Out-of-cone radiation:

$$\langle \delta p_t \rangle \propto -\int_R \frac{d\theta}{\theta} \sim -\log(1/R)$$

- not catching soft background radiation (underlying event)



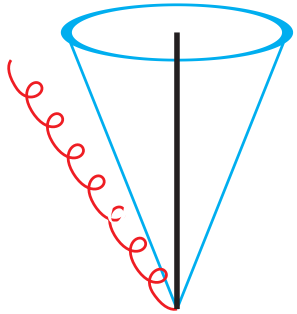
$\langle \delta p_t \rangle \sim$ Soft contents $\propto$ jet area $\sim R^2$

the coefficients depend on the algorithm

Optimisation: underlying idea

Competition between

- catching perturbative radiation



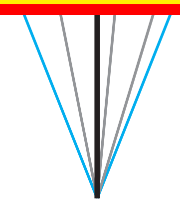
Out-of-cone radiation:

$$\langle \delta p_t \rangle \propto -\int_R \frac{d\theta}{\theta} \sim -\log(1/R)$$

- nc

Study the width of the peak in dijet reconstructions

Search optimal definition (algo + R!) as a fct. of M



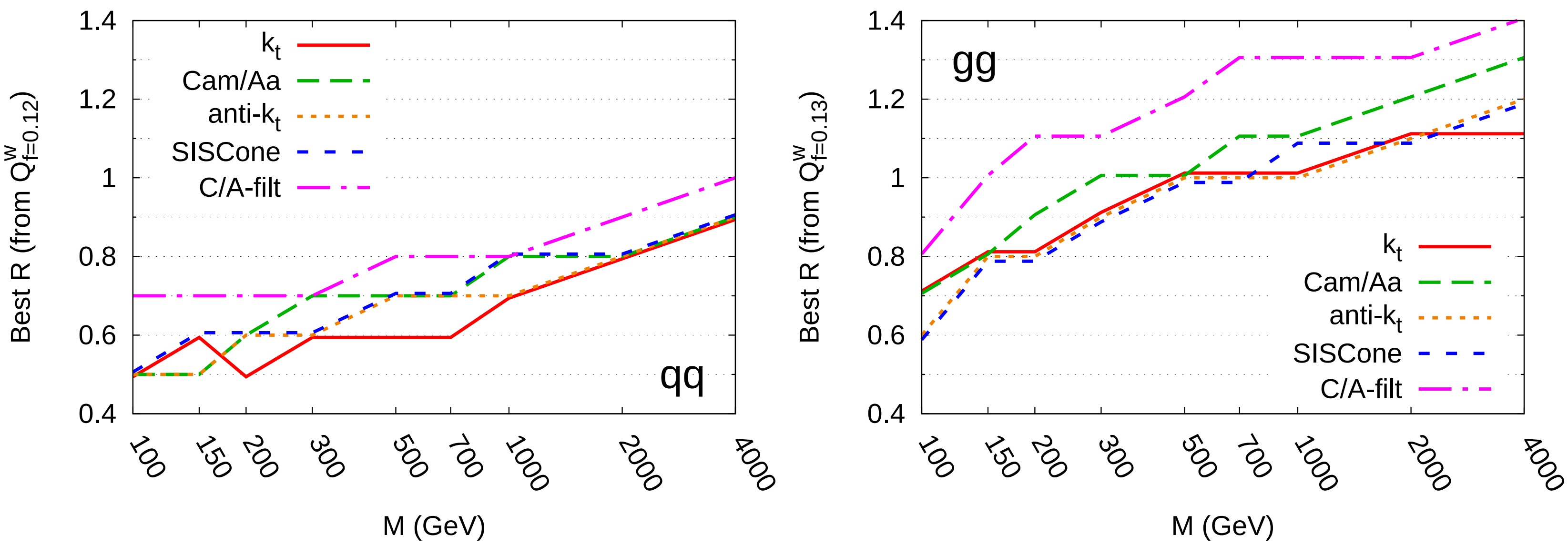
$\langle \delta p_t \rangle \sim$ Soft contents $\propto$ jet area $\sim R^2$

the coefficients depend on the algorithm

Optimisation: results

[[M.Cacciari, J.Rojo, G.Salam, GS, 08]

- SISCone and C/A+filt. do slightly better than k_t, C/A or anti-k_t
- $M \nearrow \Rightarrow R_{\text{best}} \nearrow$ (and $R_{\text{best}}(g) > R_{\text{best}}(q)$)



- Suggests flexibility in the algorithm choice at the LHC
  Could cost a factor 2 in integrated lumi for discoveries

Conclusions

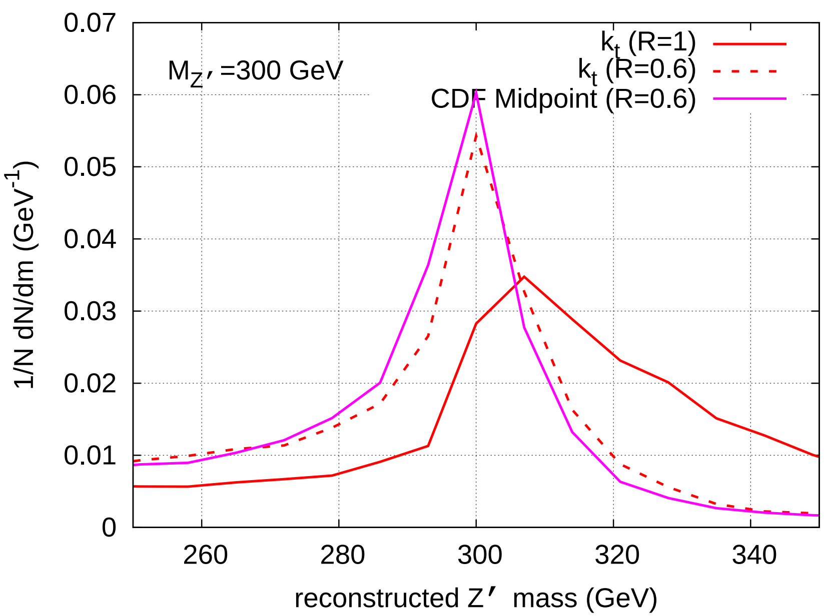
- Cone algorithm: use and infrared-and-collinear-safe one

ATLAS Cone CDF/D0 MidPoint	$\Big\}\longrightarrow$	SISCone	$\surd$ fast $\surd$ safe
CMS It. Cone	$\longrightarrow$	anti-k_t	$\surd$ fast $\surd$ safe

- Background subtraction: use jet areas
  - properly defined, under analytic control
  - simple and generic subtraction method
- Future: new generation of algorithms
  - Solid set of algorithms $\Rightarrow$ play to do a better job
  - use the *jet substructure*
  - optimise the parameters

Our example: safe jet definitions

1990

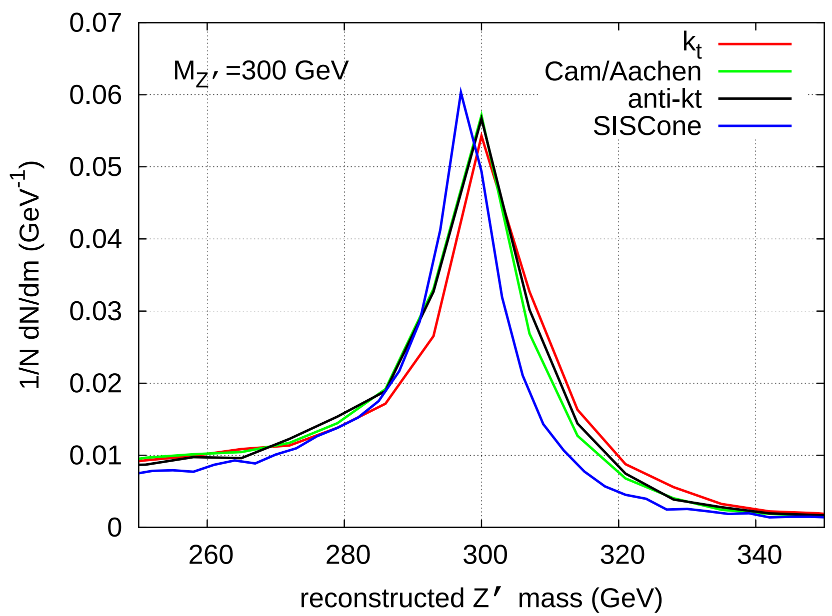


width = 19.7/14.2 GeV

width = 11.8 GeV

For the width
SISCone slightly preferred

2009



width = 14.2 GeV

width = 12.2 GeV

width = 13.1 GeV

width = 11.6 GeV

Jet areas

[M.Cacciari, G.Salam, GS, 08]

Area $\equiv$ region where the jet catches soft particles

- Recipe: add infinitely soft particles (aka ghosts) and see in which jet they are clustered
- 2 methods:
  - Passive area: add one ghost at a time and repeat many times
  - Active area: add a set of ghosts and cluster once
- Idea: ghost $\approx$ background particle

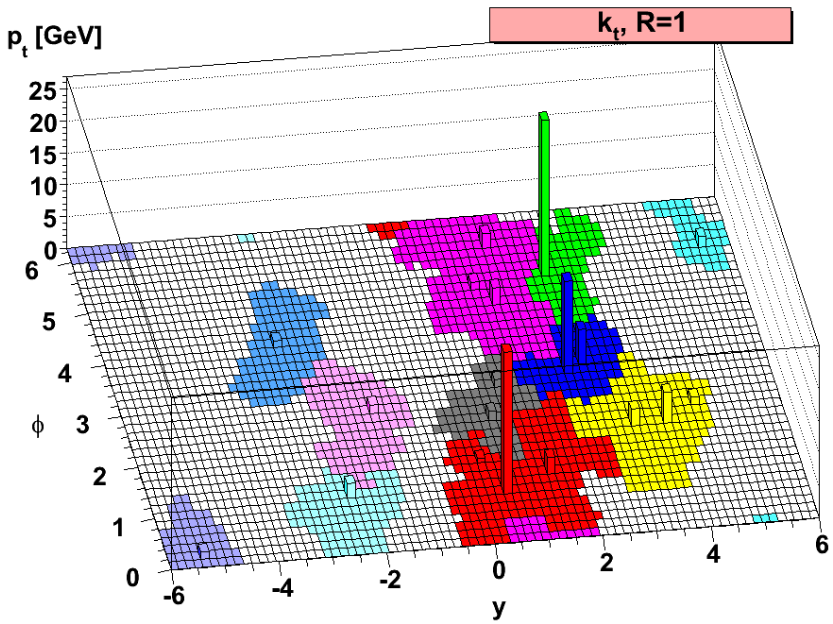
  $\Rightarrow$ active area $\approx$ uniform background

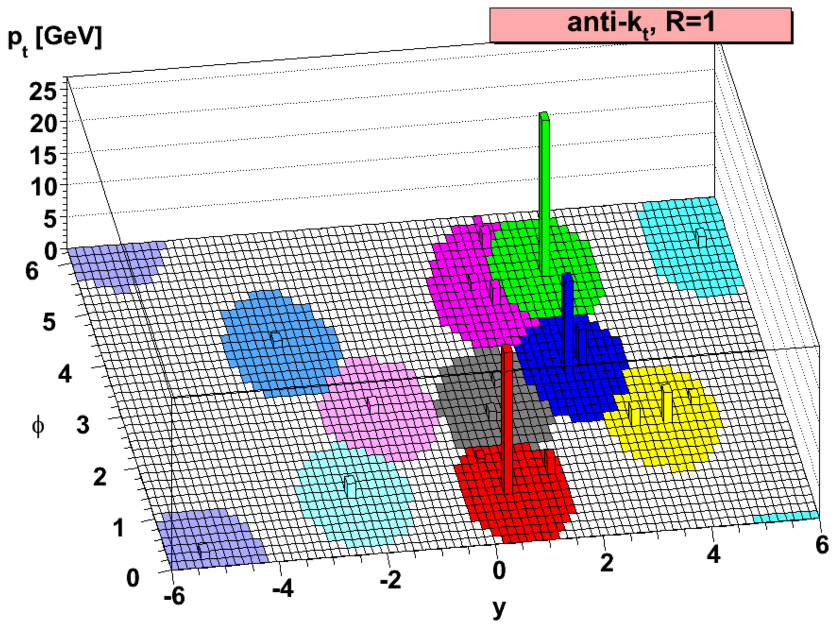
  passive area $\approx$ pointlike background
- Notes:
  - passive = active for large multiplicities
  - require an IR-safe algorithm!

Jet area: examples

Example: active area for a simple event

k_t anti-k_t





one ghost at every grid cell

Note: analytic control

Example: perturbative expansion of areas (at order α_s)

$$\langle \mathcal{A}(p_t, R)\rangle = \mathcal{A}_0 + \frac{C_{F,A}}{b_0 \pi} \pi R^2 \, d \, \log \left(\frac{\alpha_s(Q_0)}{\alpha_s(R p_t)} \right)$$

- area $\neq \pi R^2$, area $\neq$ const.
- coefficients computable

	$\mathcal{A}_0/(\pi R^2)$		d	
	passive	active	passive	active
k_t	1	0.81	0.56	0.52
Cam/Aachen	1	0.81	0.08	0.08
anti-k_t	1	1	0	0
SISCone	1	1/4	-0.06	0.12

- $Q_0 \equiv$ IR regulator $\propto$ background density