What (optimal) definition for jets at the LHC?

Grégory Soyez

CERN

In collaboration with Gavin Salam, Matteo Cacciari and Juan Rojo

IFAE, Barcelona — April 30 2010

Plan

- Jet algorithms and jet definitions
  - basic ideas: why jets? recombinations and cones
  - failures of the 20$^{\rm th}$-century cone algorithms
  - new algorithms without the failures

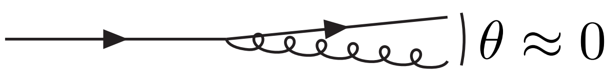
- Optimal choice: which jet definition to use?
  - Framework: how to quantify reconstructino performances
  - Monte-Carlo simulations
  - Analytic approach

Unavoidable theory

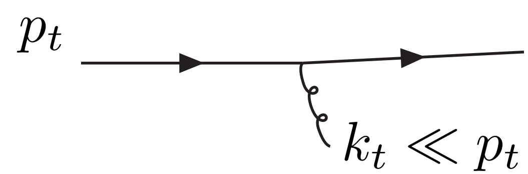
QCD probability for gluon emission (angle θ and $\perp$-mom. k_t):

$$dP \propto \alpha_s \frac{d\theta}{\theta} \frac{dk_t}{k_t}$$

Two divergences:



collinear



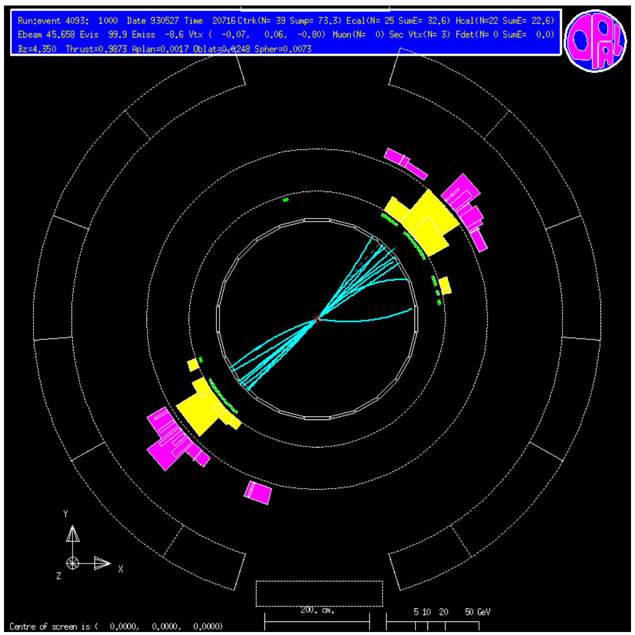
soft

Divergences cancelled by virtual corrections

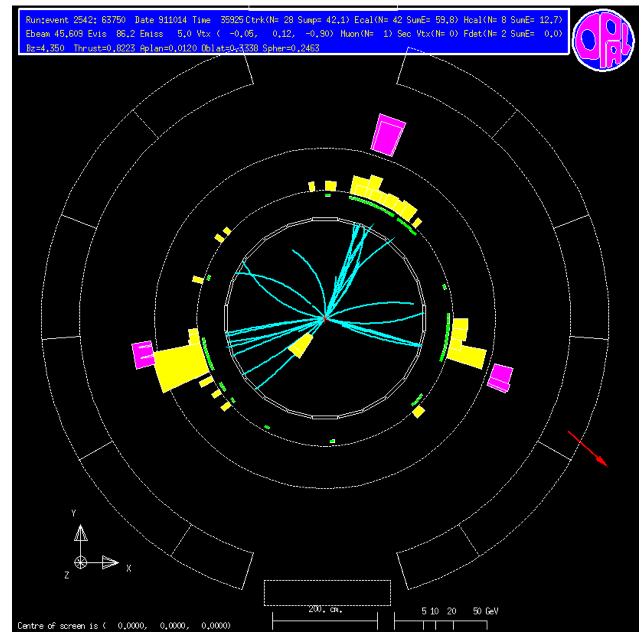
Motivation: why jets

Collinear divergence $\Rightarrow$ QCD produces "jetty" showers

Example: LEP (OPAL) events



2 jets



3 jets

"Jets" $\equiv$ bunch of collimated particles $\cong$ hard partons

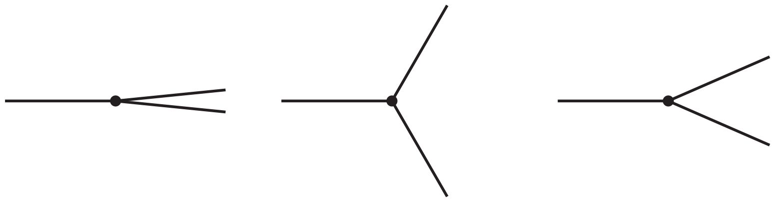
Motivation: why jets

Collinear divergence ⇒ QCD produces “jetty” showers

“Jets” ≡ bunch of collimated particles ≅ hard partons

BUT

- a “parton” is an ambiguous concept (NLO)
- “collinear” has some arbitraryness



2 jets 3 jets ? jets

Motivation: why jets

Collinear divergence $\Rightarrow$ QCD produces “jetty” showers

“Jets” $\equiv$ bunch of collimated particles $\cong$ hard partons

BUT

- a
- “c

In practice: use of a jet definition

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

Jet algorithm: the recipe (insufficient!)

Jet definition: algorithm + the parameters

20th century jet algorithms

Recombination:

- k_t algorithm
- Cambridge/Aachen alg.

Cone:

- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet

20th century jet algorithms

Recombination:

- k_t algorithm
- Cambridge/Aachen alg.

Idea: undo the showering

Successively

- find the closest pair of particles
- recombine them

Distance:

k_t:

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

Cam/Aachen:

$$d_{i,j} = \Delta\phi_{i,j}^2 + \Delta y_{i,j}^2$$

stop at a distance R

20th century jet algorithms

Idea: dominant flow of energy

Stable cone (radius R):
sum of particles in the cone points towards the cone centre

All these are iterative cones:

- start from a seed
- iterate until stable

seeds = {particles, midpoints}

Jet $\equiv$ stable cone
modulo overlapping

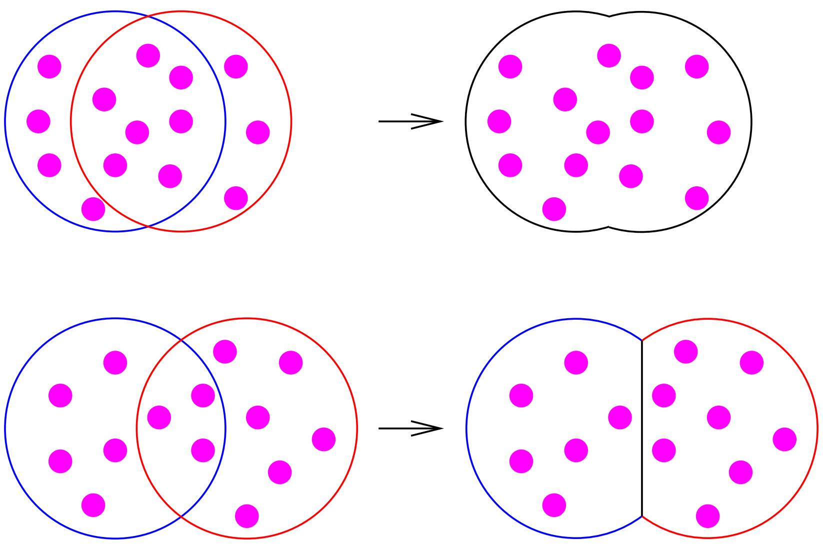
Cone:

- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet

20th century jet algorithms

Cone with split-merge

Split/merge if the overlap is smaller/larger than a threshold f



Cone:

- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet

20th century jet algorithms

Cone with progressive removal

Successively

- iterate from hardest particle
- call that a jet (remove particles)

Basic property:

hard circular jets

Cone:

- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet

20th century jet algorithms

Recombination:

- k_t algorithm
- Cambridge/Aachen alg.

✓perturbative behaviour

Cone:

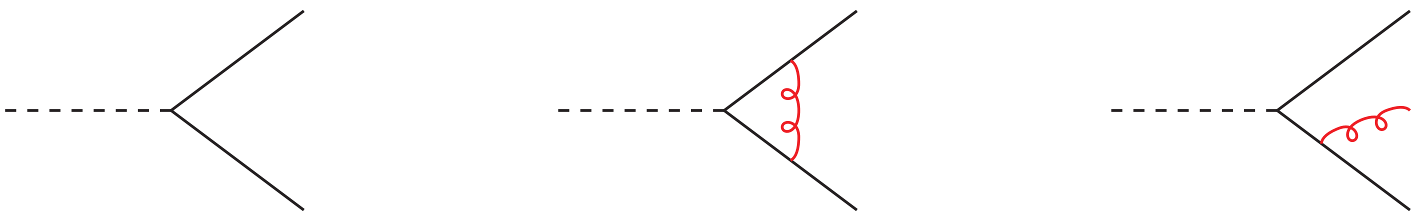
- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet

✓UE sensitivity

21st century: how does that picture change?

QCD divergences

Ingredient: QCD soft and collinear divergencies



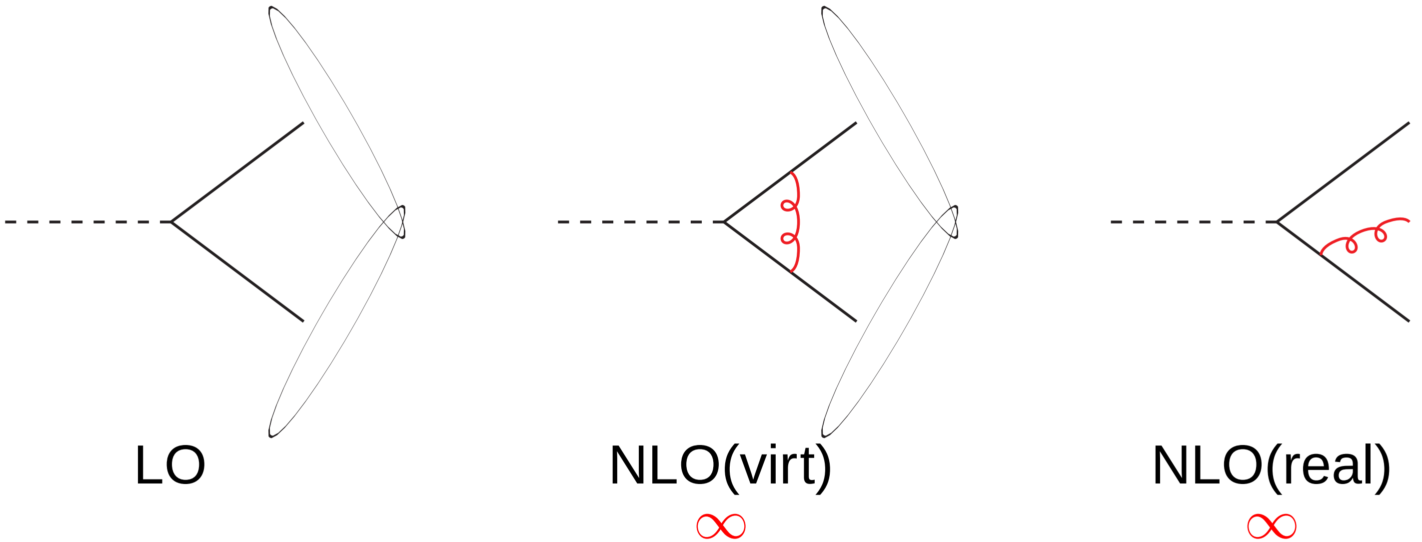
LO NLO(virt) NLO(real)

∞ ∞

- ∞ (from soft gluons) cancel (inclusive x-section)

QCD divergences

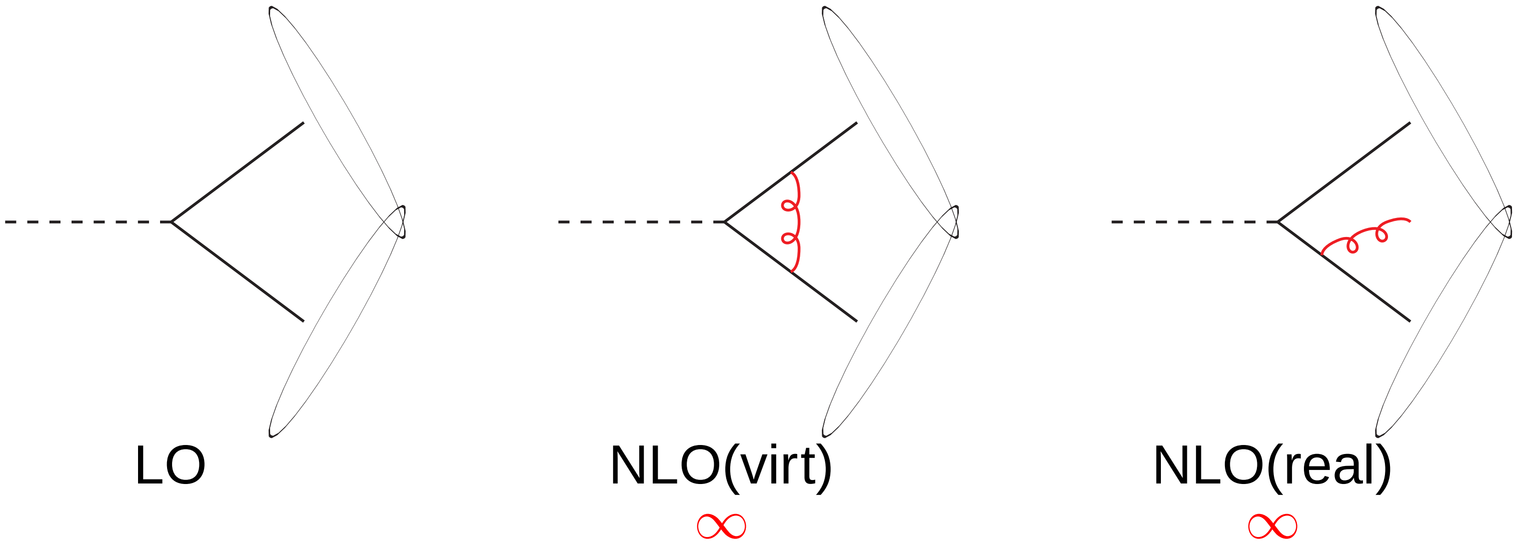
Ingredient: QCD soft and collinear divergencies



- Consider an extra (NLO) soft gluon
- Assume LO gives 2 jets $\Rightarrow$ NLO(virt) gives 2 jets

QCD divergences

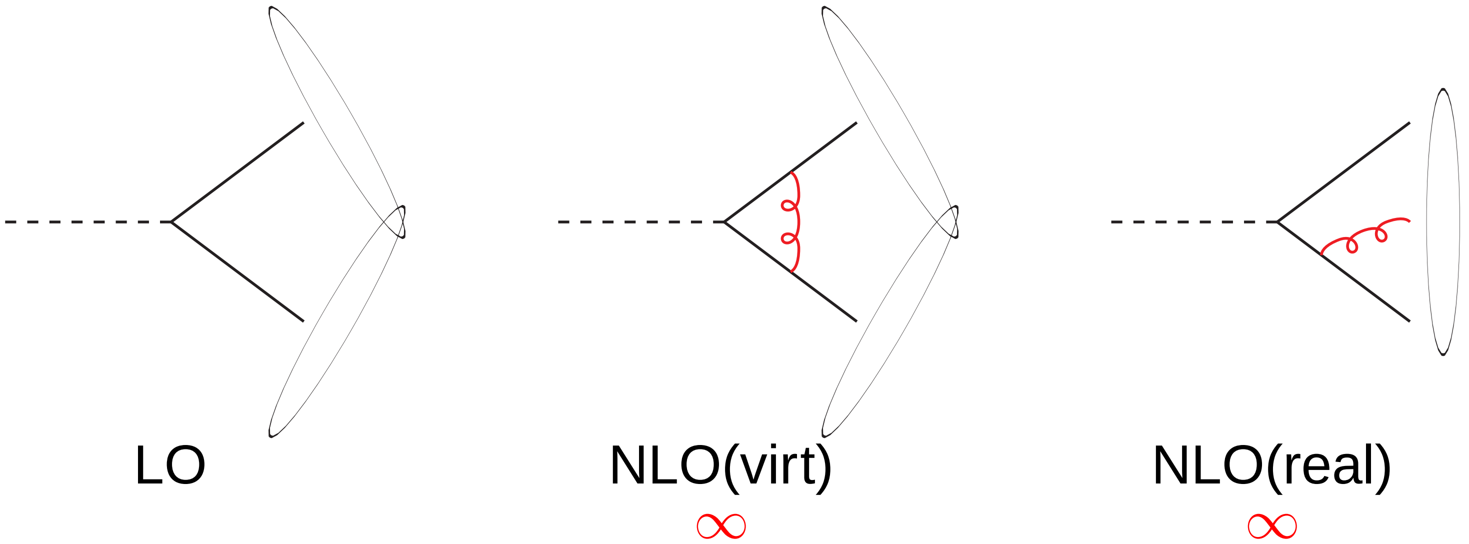
Ingredient: QCD soft and collinear divergencies



- Consider an extra (NLO) soft gluon
- Assume LO gives 2 jets $\Rightarrow$ NLO(virt) gives 2 jets
- NLO(real) gives 2 jets $\Rightarrow$ ∞ cancel $\Rightarrow$ finite jet cross-section

QCD divergences

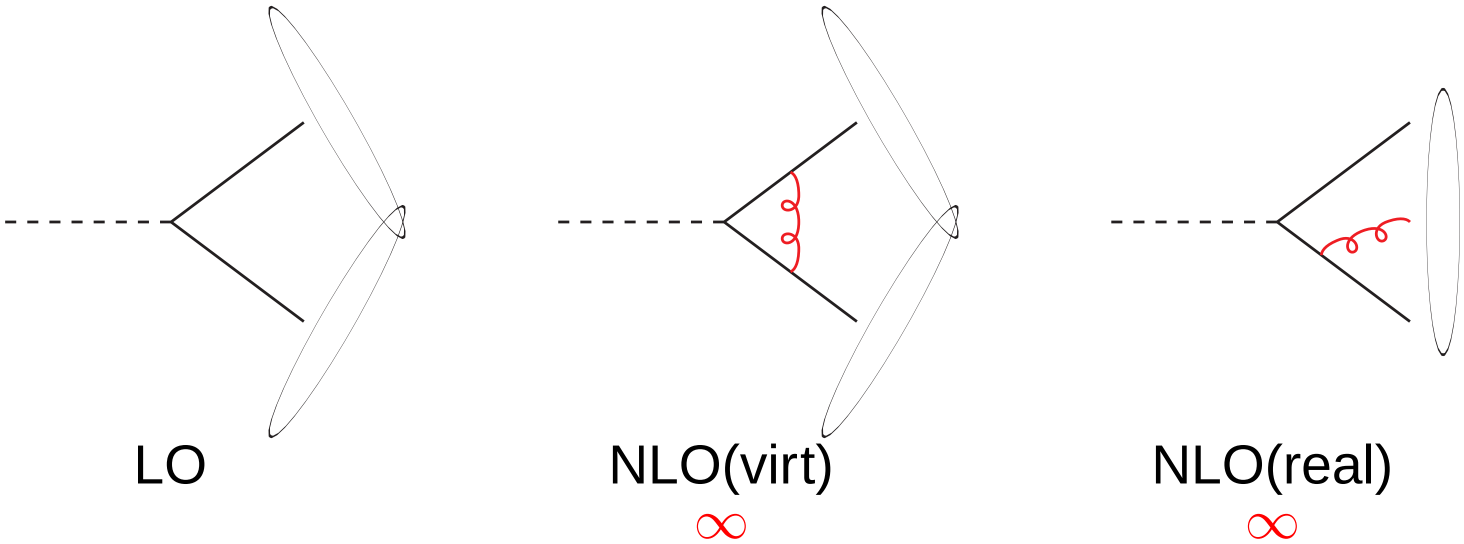
Ingredient: QCD soft and collinear divergencies



- Consider an extra (NLO) soft gluon
- Assume LO gives 2 jets $\Rightarrow$ NLO(virt) gives 2 jets
- NLO(real) gives 2 jets $\Rightarrow$ ∞ cancel $\Rightarrow$ finite jet cross-section
  NLO(real) gives 1 jets $\Rightarrow$ ∞ do not cancel $\Rightarrow$ infinite jet x-section

QCD divergences

Ingredient: QCD soft and collinear divergencies

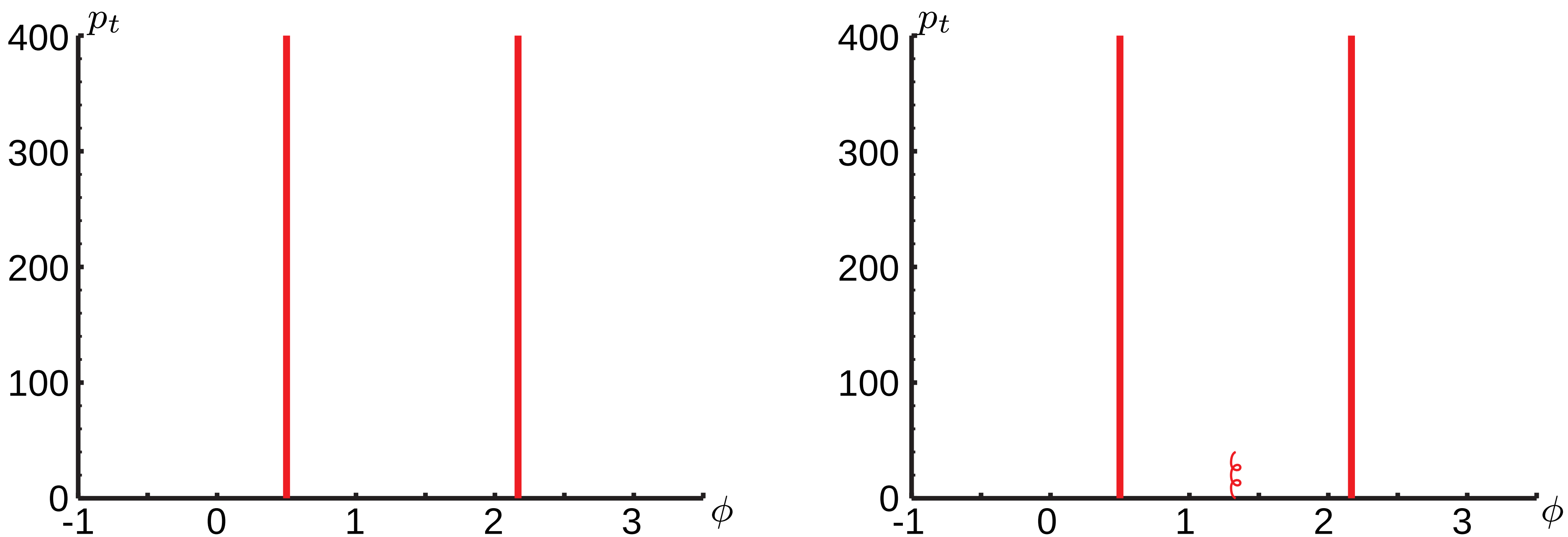


For pQCD to make sense, the (hard) jets should not change when

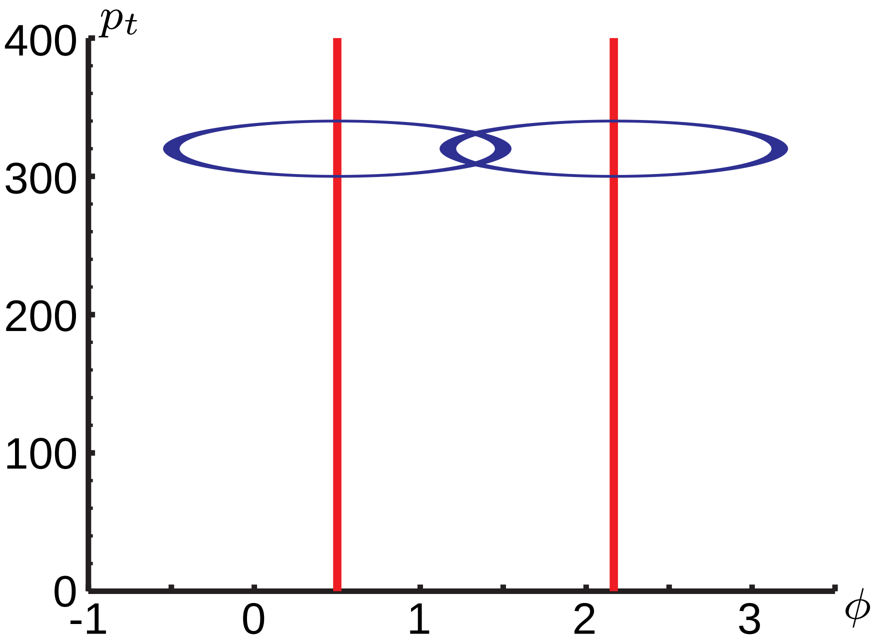
- one has a soft emission *i.e.* adds a very soft gluon
- one has a collinear splitting
  i.e. replaces one parton by two at the same place (η, ϕ)

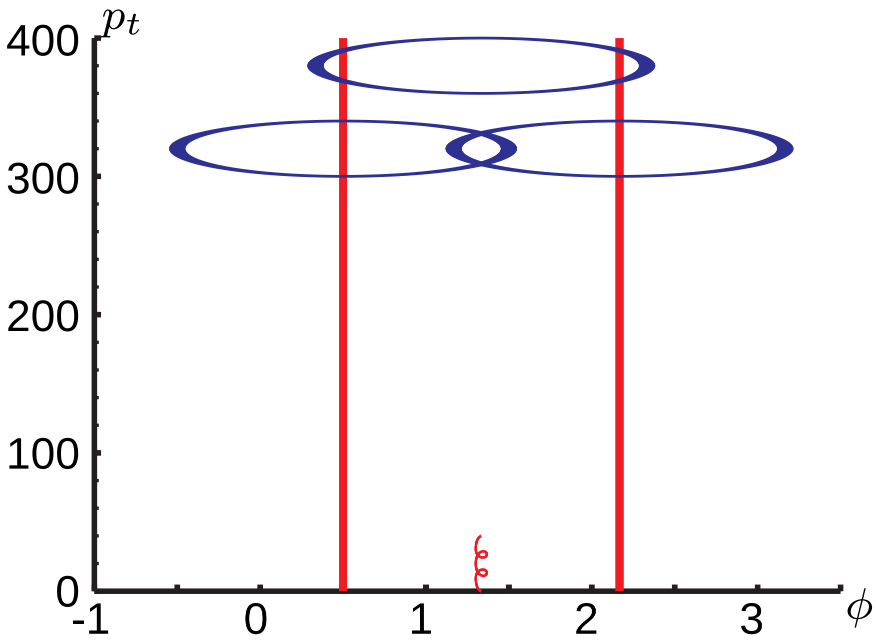
[SNOWMASS Accords, Fermilab, 1990]

IR (un)safety? JetClu and Atlas Cone



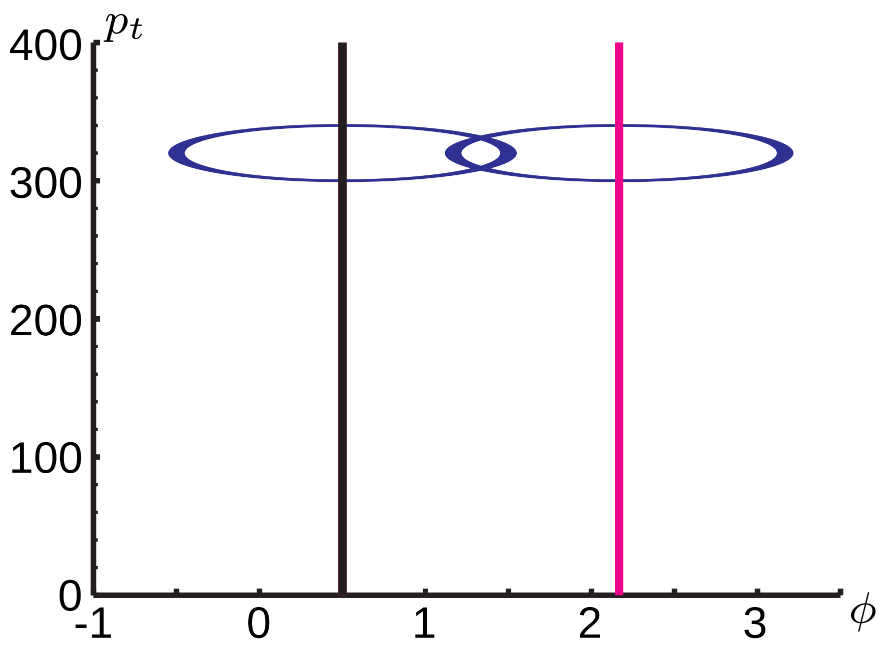
IR (un)safety? JetClu and Atlas Cone



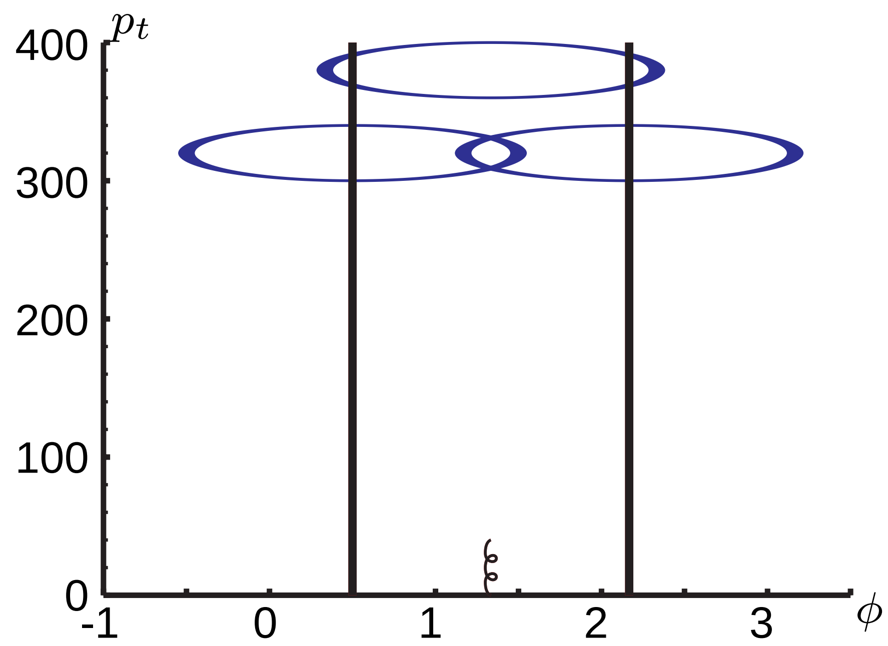


Stable cones found

IR (un)safety? JetClu and Atlas Cone



2 jets

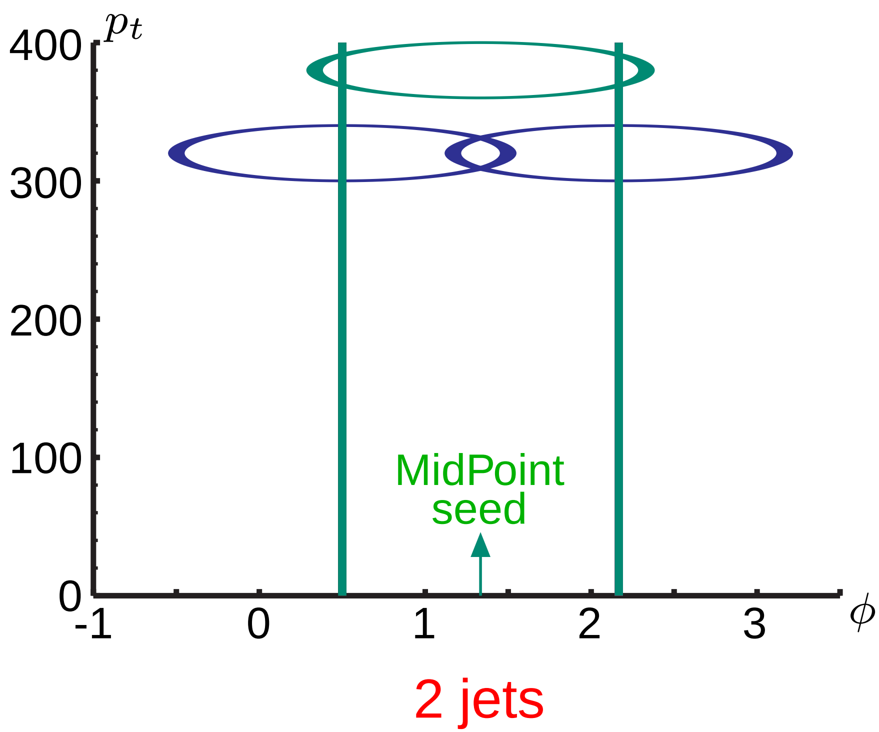


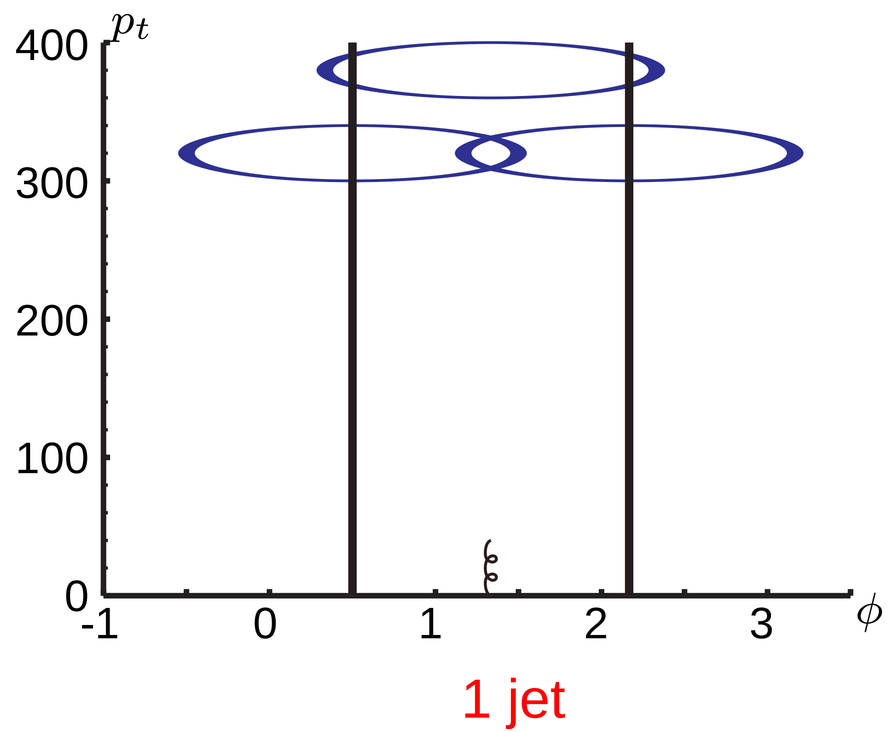
1 jet

A soft gluon changed the number of jets

$\Rightarrow$ IR unsafety of JetClu and the ATLAS Cone

IR (un)safety? JetClu and Atlas Cone





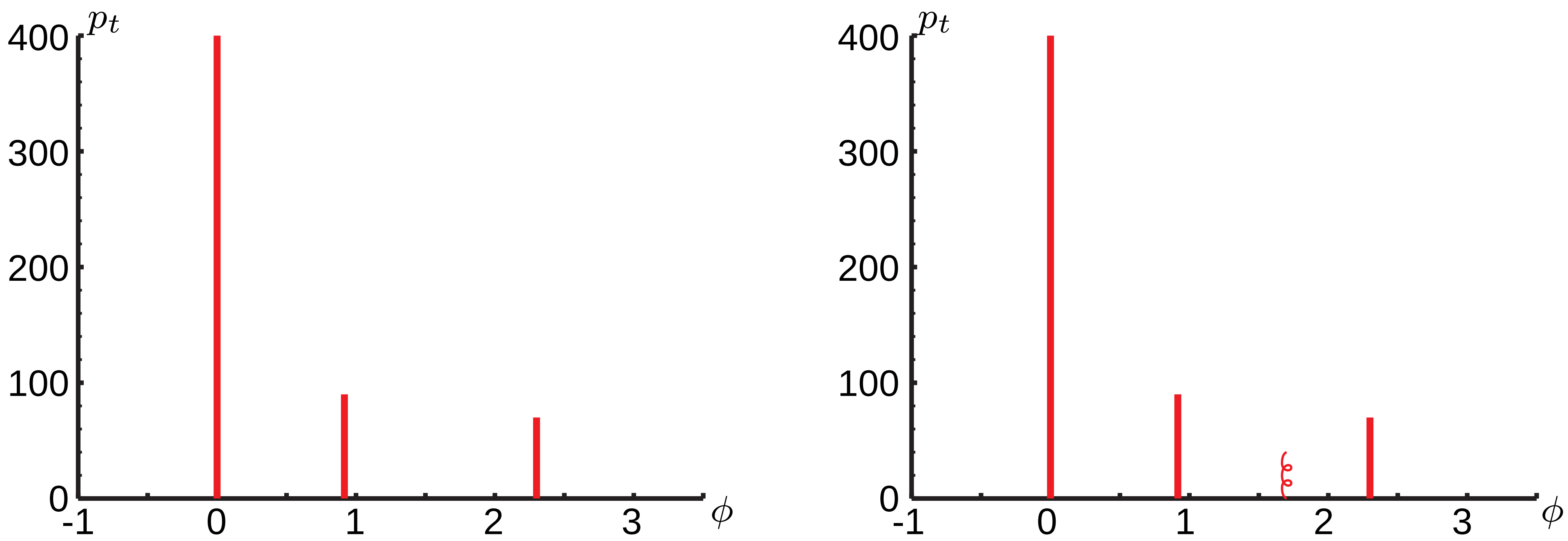
A soft gluon changed the number of jets

$\Rightarrow$ IR unsafety of JetClu and the ATLAS Cone

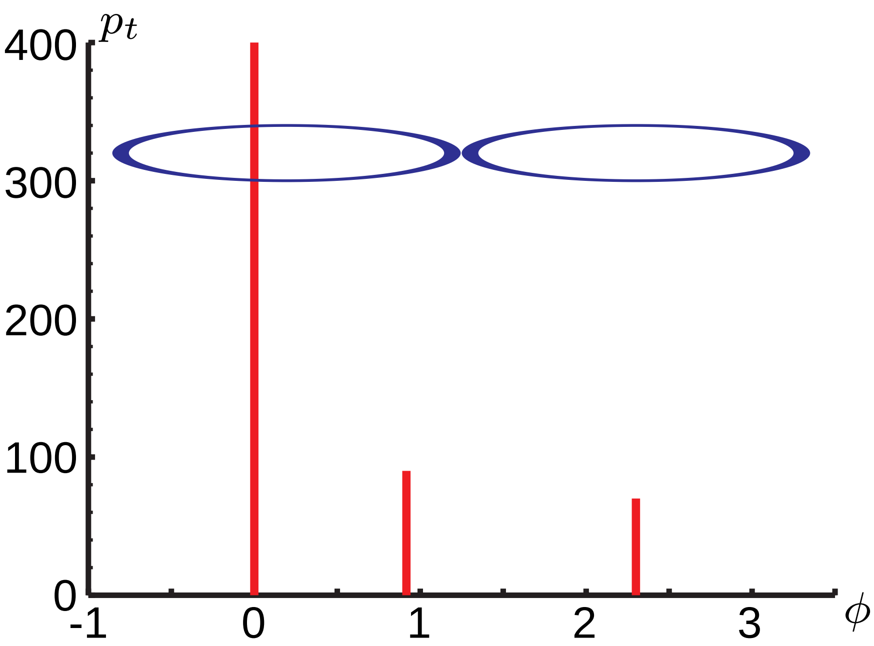
Fixed by MidPoint

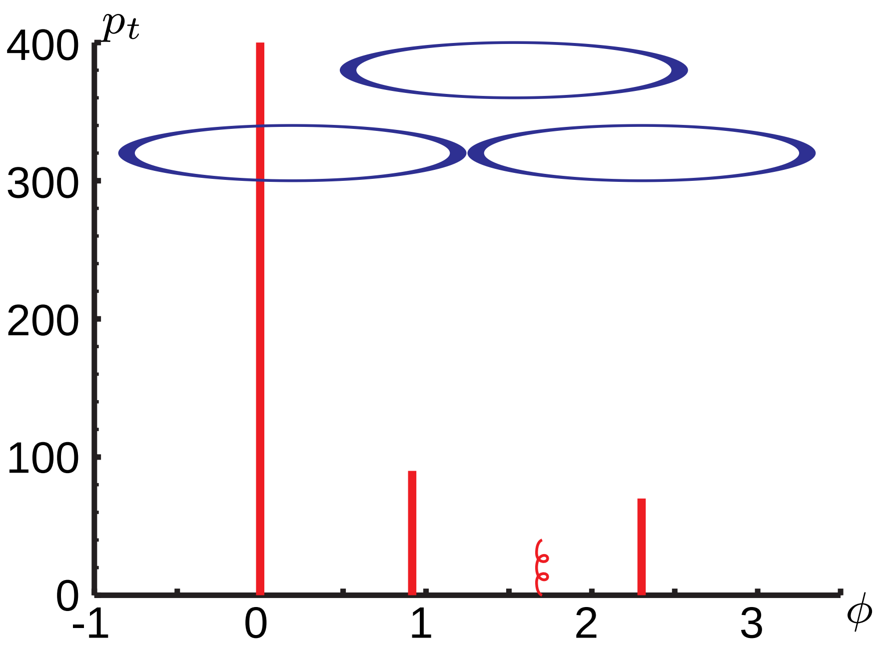
[Blazey *et al.*, 00]

IR (un)safety? MidPoint



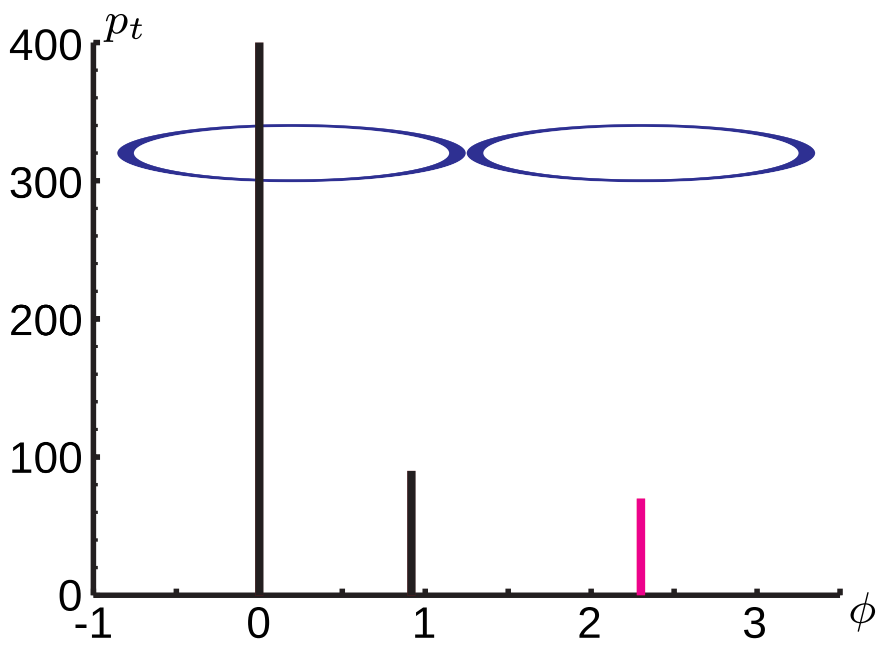
IR (un)safety? MidPoint



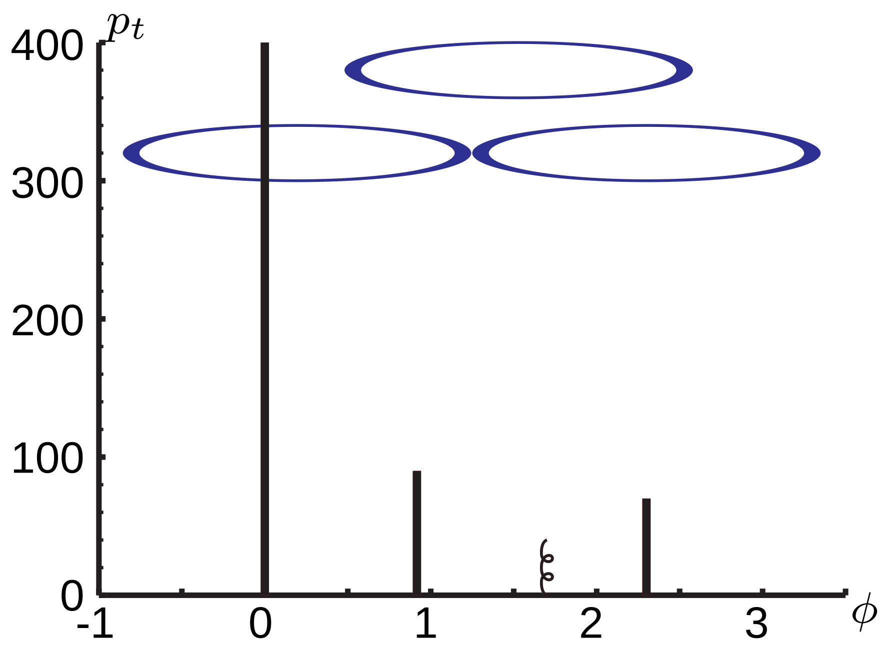


Stable cones found

IR (un)safety? MidPoint



2 jets

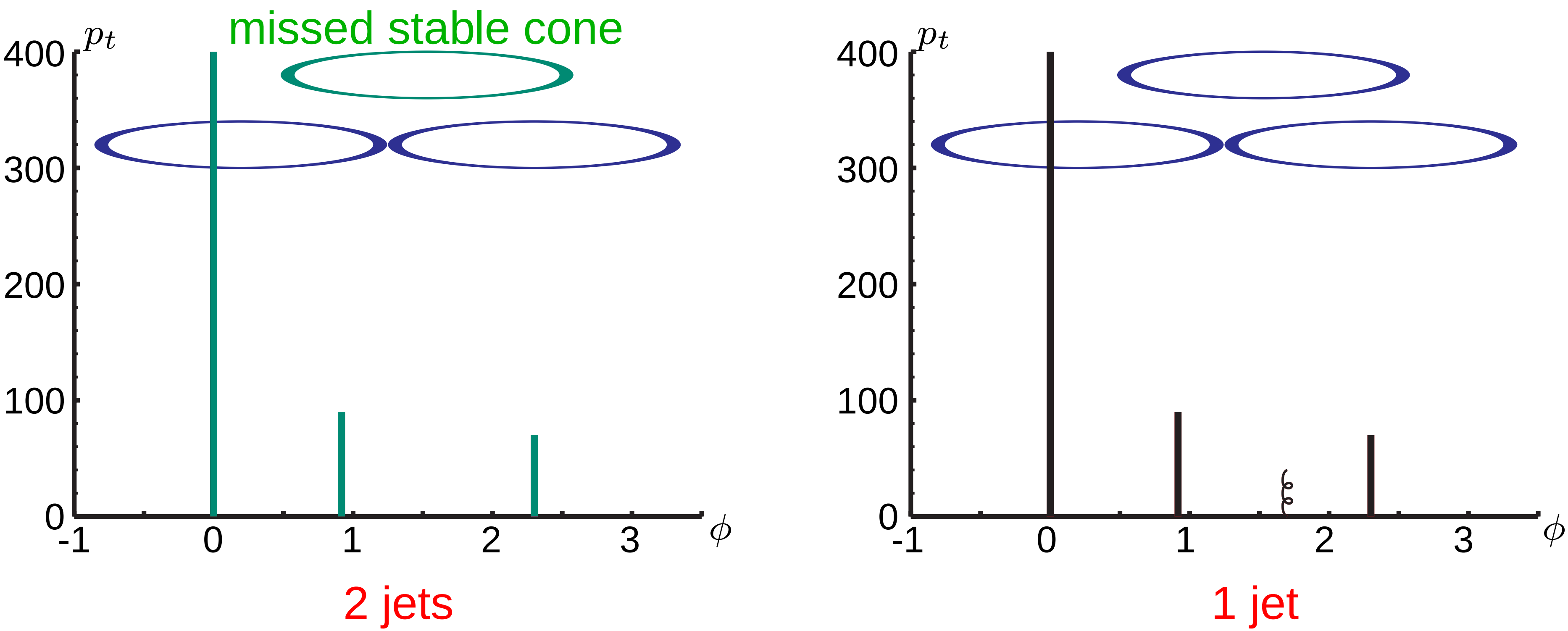


1 jet

A soft gluon changed the number of jets

$\Rightarrow$ IR unsafety of MidPoint (1 order in α_s later than JetClu)

IR (un)safety? MidPoint



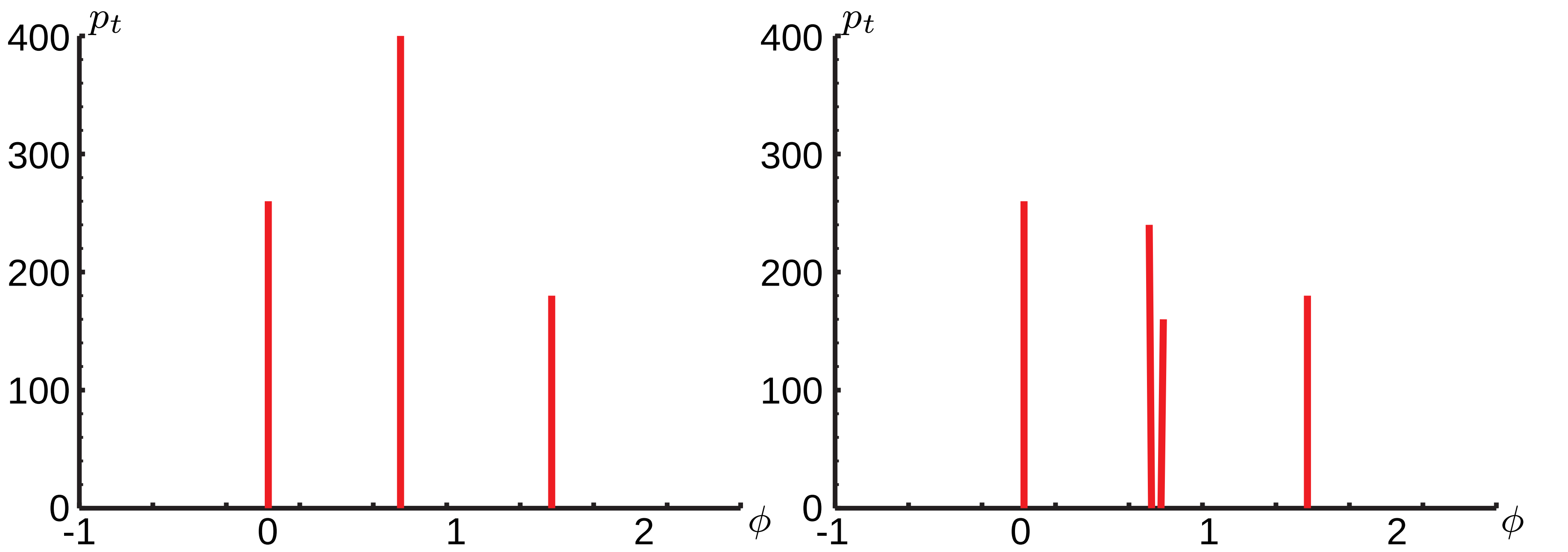
Solution: be sure to find **all** stable cones

SISCone: Seedless Infrared-Safe Cone algorithm
`http://projects.hepforge.org/siscone`

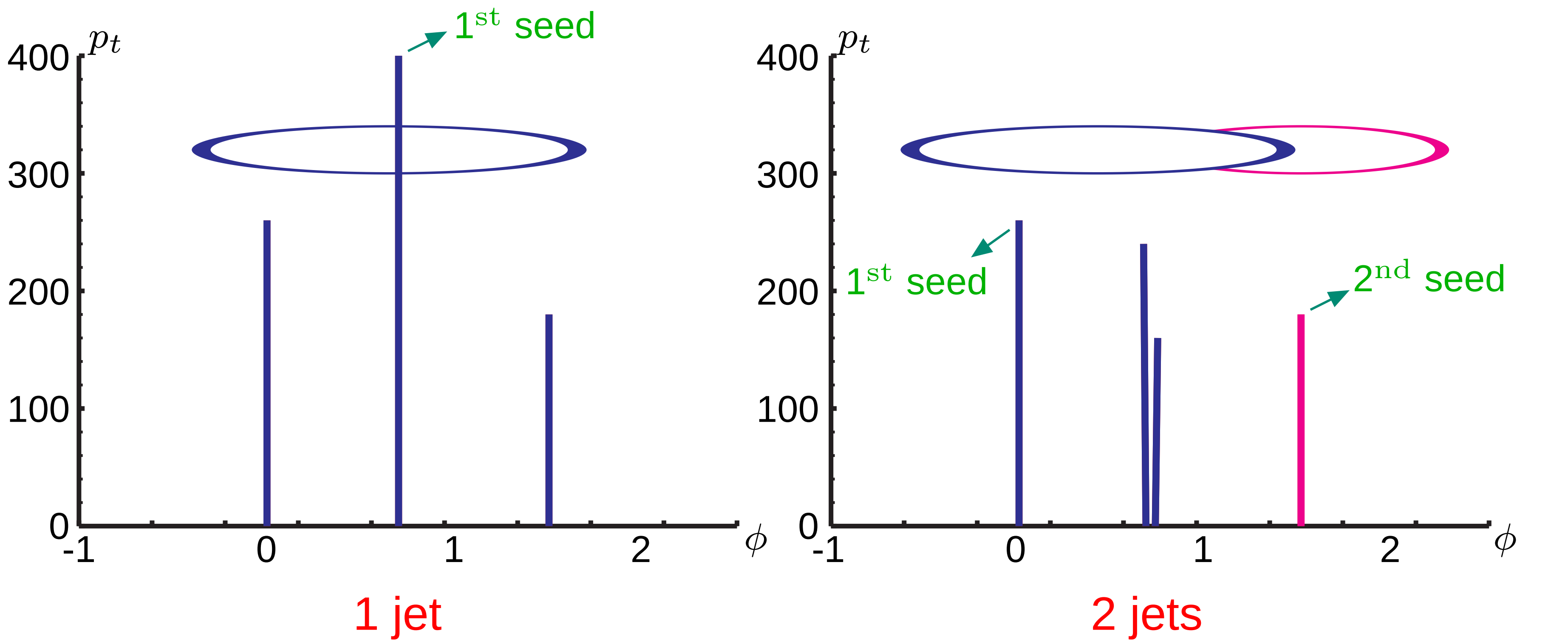
[G.Salam, G.S., 07]

Idea: enumerate enclosures by enumerating pairs of particles

Collinear (un)safety? the CMS iterative cone



Collinear (un)safety? the CMS iterative cone



A colinear splitting changed the number of jets

$\Rightarrow$ Collinear unsafety of the CMS iterative cone

Anti-k_t

Come back to recombination-type algorithms:

$$d_{ij} = \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \left(\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2\right)$$

- $p = 1$: k_t algorithm
- $p = 0$: Aachen/Cambridge algorithm

Anti-k_t

Come back to recombination-type algorithms:

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

- $p = 1$: k_t algorithm
- $p = 0$: Aachen/Cambridge algorithm
- $p = -1$: anti-k_t algorithm [M.Cacciari, G.Salam, G.S., 08]

Why should that be related to the iterative cone ?!?

- "large $k_t \Rightarrow$ small distance"
  i.e. hard partons "eat" everything up to a distance R
  i.e. circular/regular jets, jet borders unmodified by soft radiation
- infrared and collinear safe

21st century jet finders

Recombination:

- k_t algorithm
- Cambridge/Aachen alg.
- anti-k_t algorithm

Cone:

- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet
- SISCone

21st century jet finders

Recombination:

- k_t algorithm
- Cambridge/Aachen alg.
- anti-k_t algorithm

Cone:

- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet
- SISCone

4 available
safe algorithms

21st century jet finders

Recombination:

- k_t algorithm
- Cambridge/Aachen alg.
- anti-k_t algorithm

Cone:

- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- SISCone

```
#--------------------------------------------------------------------------
#                         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

Physical impact

JetClu/ATLAS cone	unsafe at $\mathcal{O}(\alpha_s^3)$ (or $\mathcal{O}(\alpha_{\rm ew}\alpha_s^2)$)
MidPoint/CMS it. cone	unsafe at $\mathcal{O}(\alpha_s^4)$ (or $\mathcal{O}(\alpha_{\rm ew}\alpha_s^3)$)

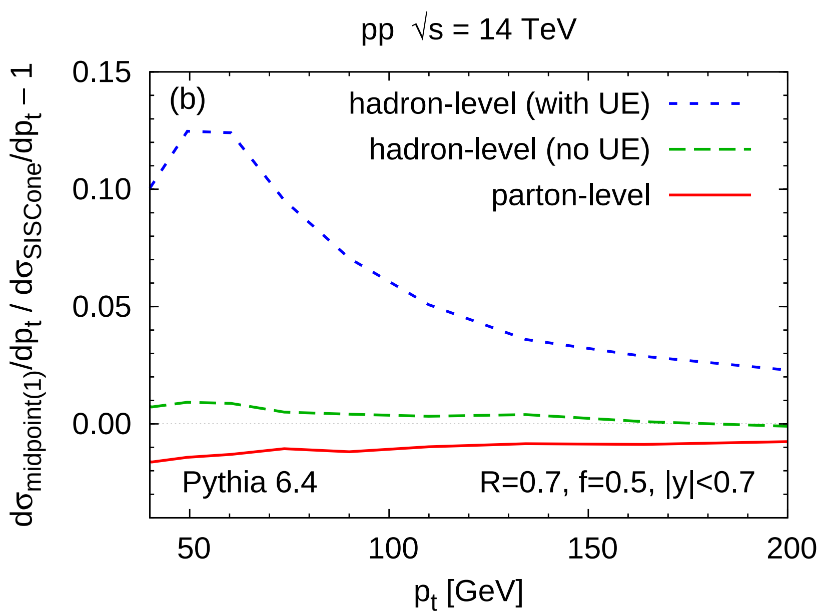
Physical impact

JetClu/ATLAS cone unsafe at $\mathcal{O}(\alpha_s^3)$ (or $\mathcal{O}(\alpha_{\rm ew}\alpha_s^2)$)

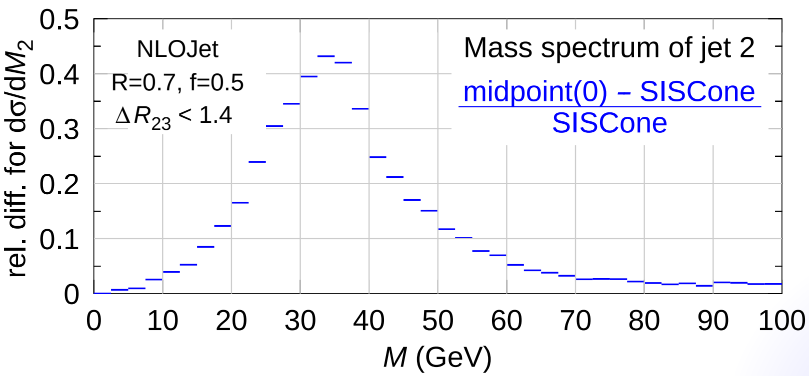
MidPoint/CMS it. cone unsafe at $\mathcal{O}(\alpha_s^4)$ (or $\mathcal{O}(\alpha_{\rm ew}\alpha_s^3)$)

	IRC-safe until		
Physical observable	JetClu/ATLAS c.	MidPoint/CMS it. c.	SISCone/recomb.
Inclusive jet cross section	LO	NLO	any
3-jet cross section	**none**	LO	any
$W/Z/H$ + 2 jet cross sect.	**none**	LO	any
jet masses in 3 jets	**none**	**none**	any

Example: (Midpoint-SISCone)/SISCone



- Incl. cross-section: a few %
- Masses in 3-jet events: $\sim 45\%$



Physical impact

JetClu/ATLAS cone unsafe at $\mathcal{O}(\alpha_s^3)$ (or $\mathcal{O}(\alpha_{\rm ew}\alpha_s^2)$)

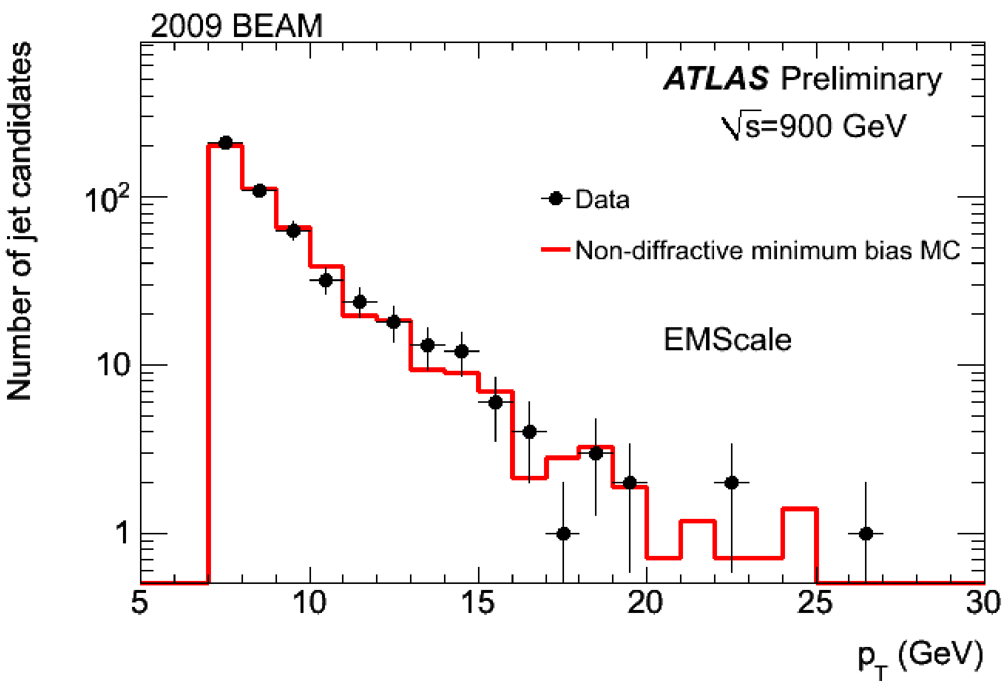
MidPoint/CMS it. cone unsafe at $\mathcal{O}(\alpha_s^4)$ (or $\mathcal{O}(\alpha_{\rm ew}\alpha_s^3)$)

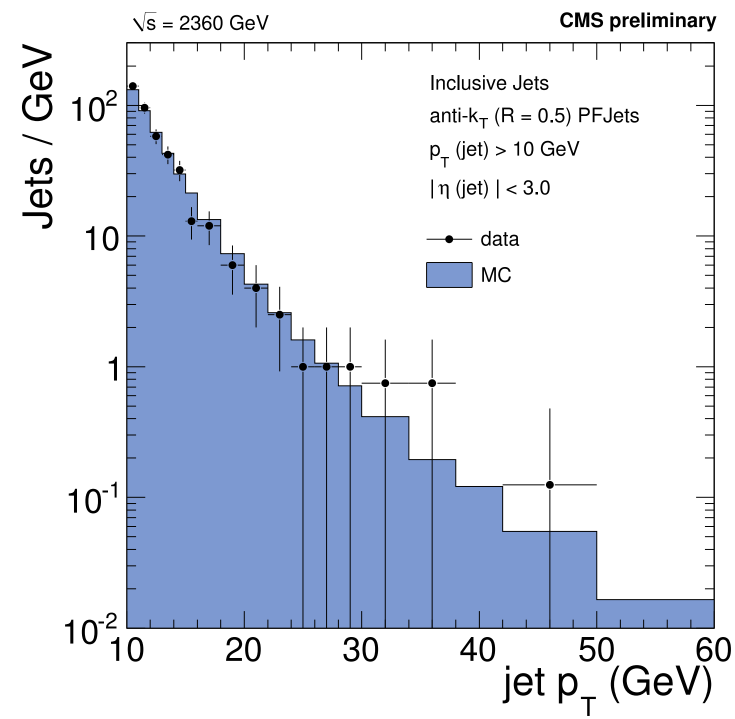
Physical observable	IRC-safe until		
	JetClu/ATLAS c.	MidPoint/CMS it. c.	SISCone/recomb.
Inclusive jet cross section	LO	NLO	any
3-jet cross section	**none**	LO	any
$W/Z/H$ + 2 jet cross sect.	**none**	LO	any
jet masses in 3 jets	**none**	**none**	any

- Huge effort (~ 50 M€) to compute processes in pQCD
- Mandatory at the LHC:
  precise background, many multi-jet processes

Anti-k_t at the LHC

FastJet + anti-k_t default choice for ATLAS and CMS



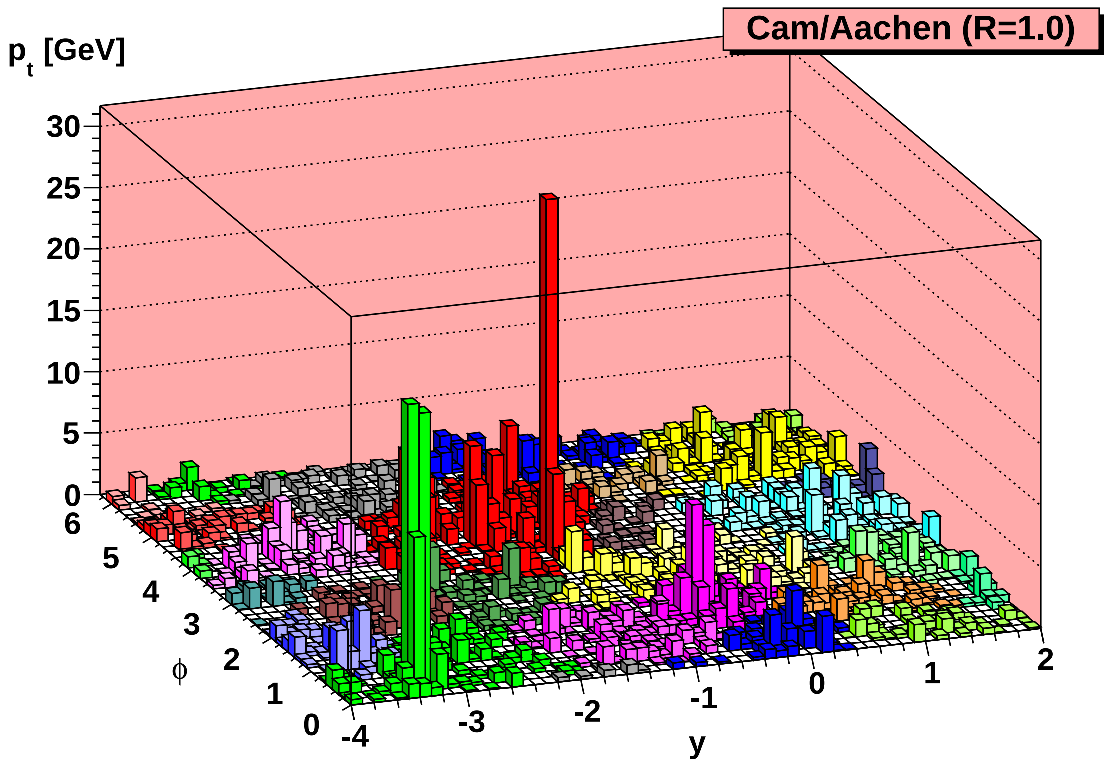


ATLAS, 900 GeV, MB events
(anti-k_t, R=0.6)

CMS, 2.36 TeV, MB events
(anti-k_t, R=0.5)

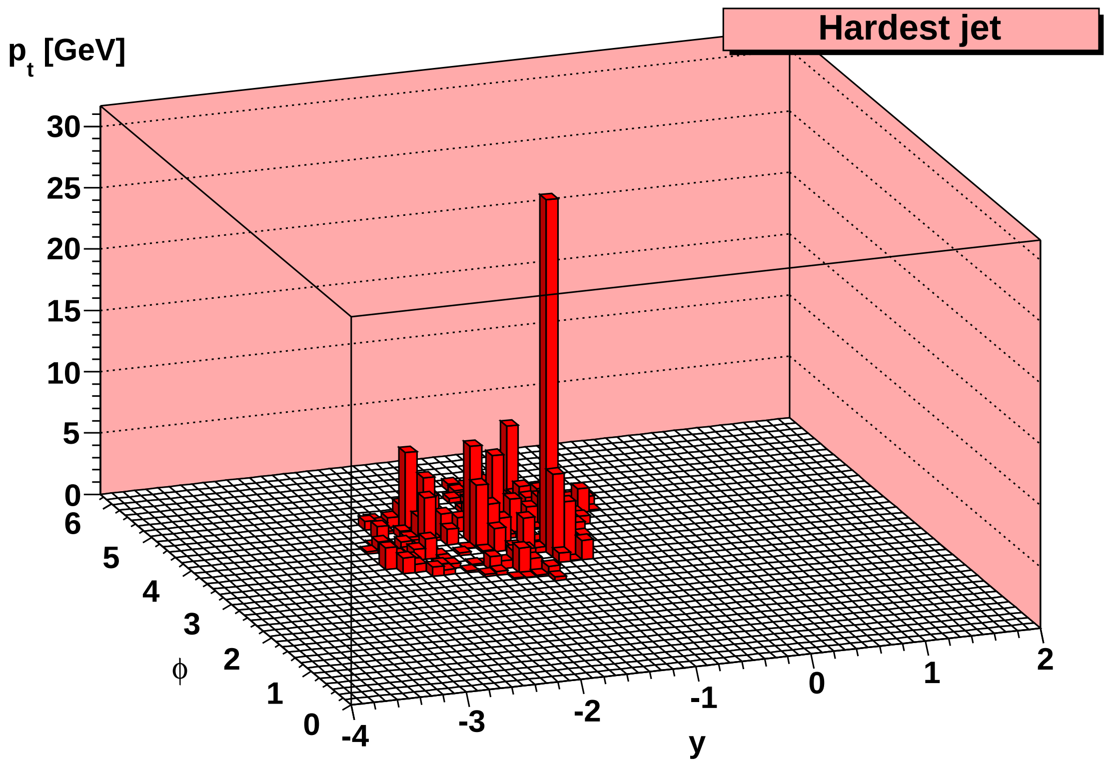
Digression: jet filtering

Filtering



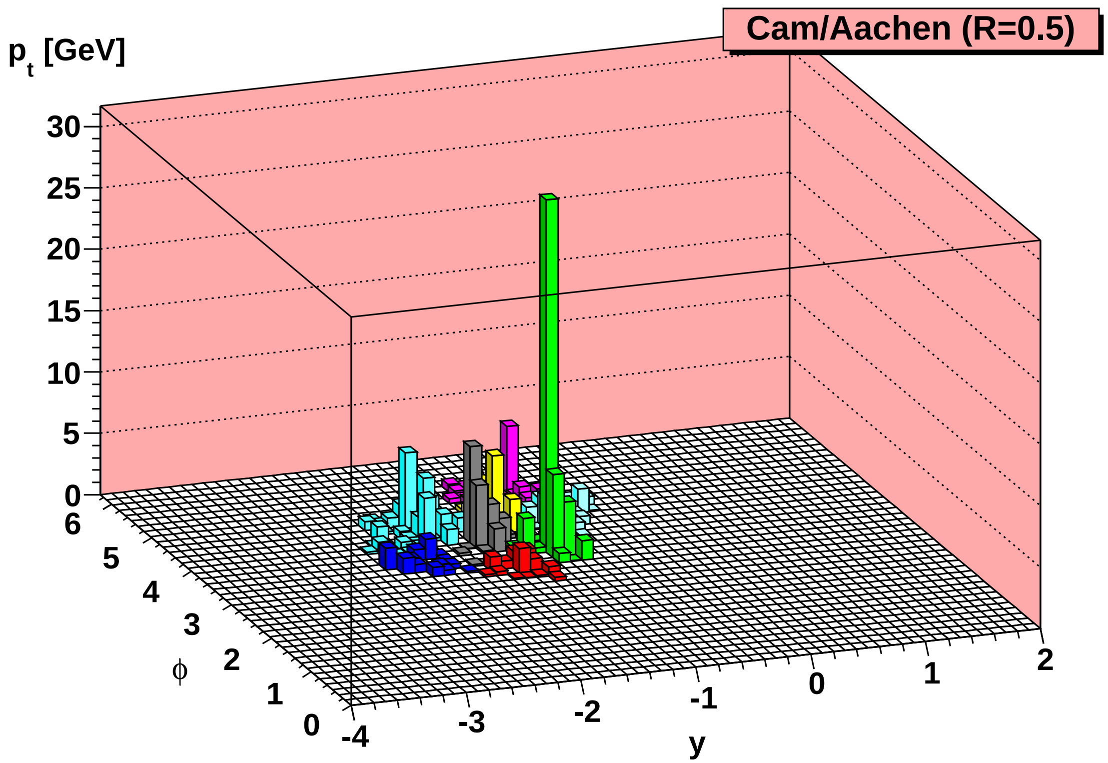
- cluster with Cambridge/Aachen(R)

Filtering



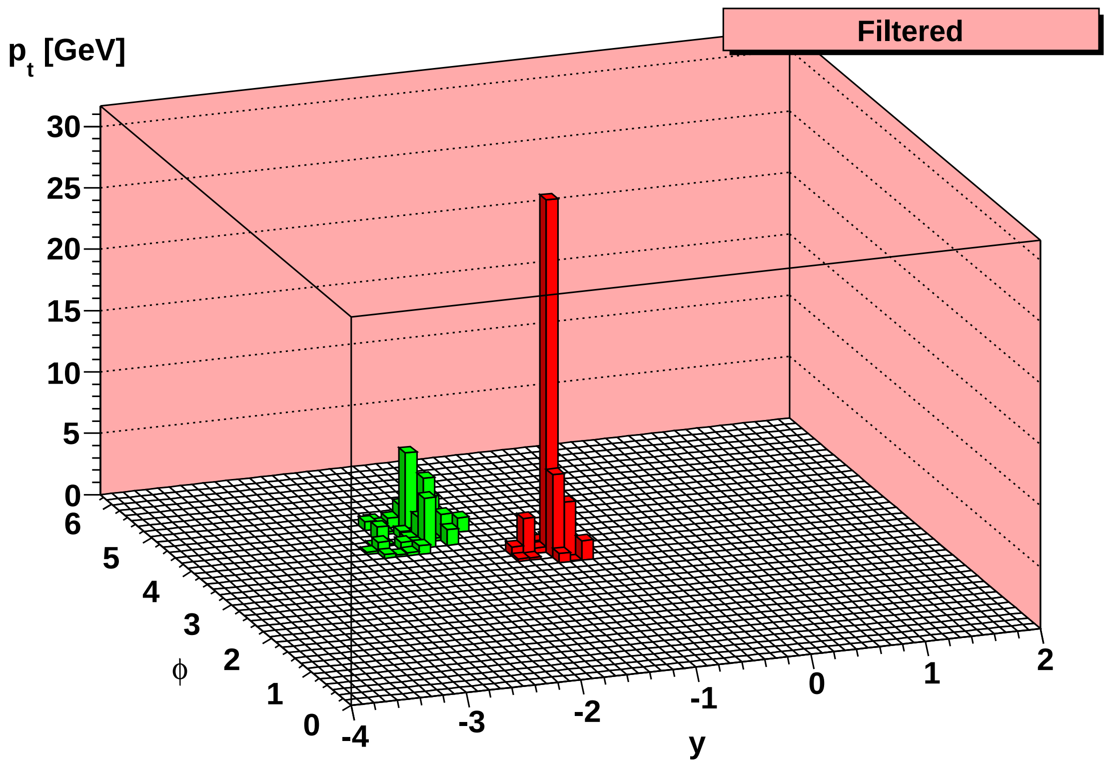
- cluster with Cambridge/Aachen(R)
- for each jet

Filtering



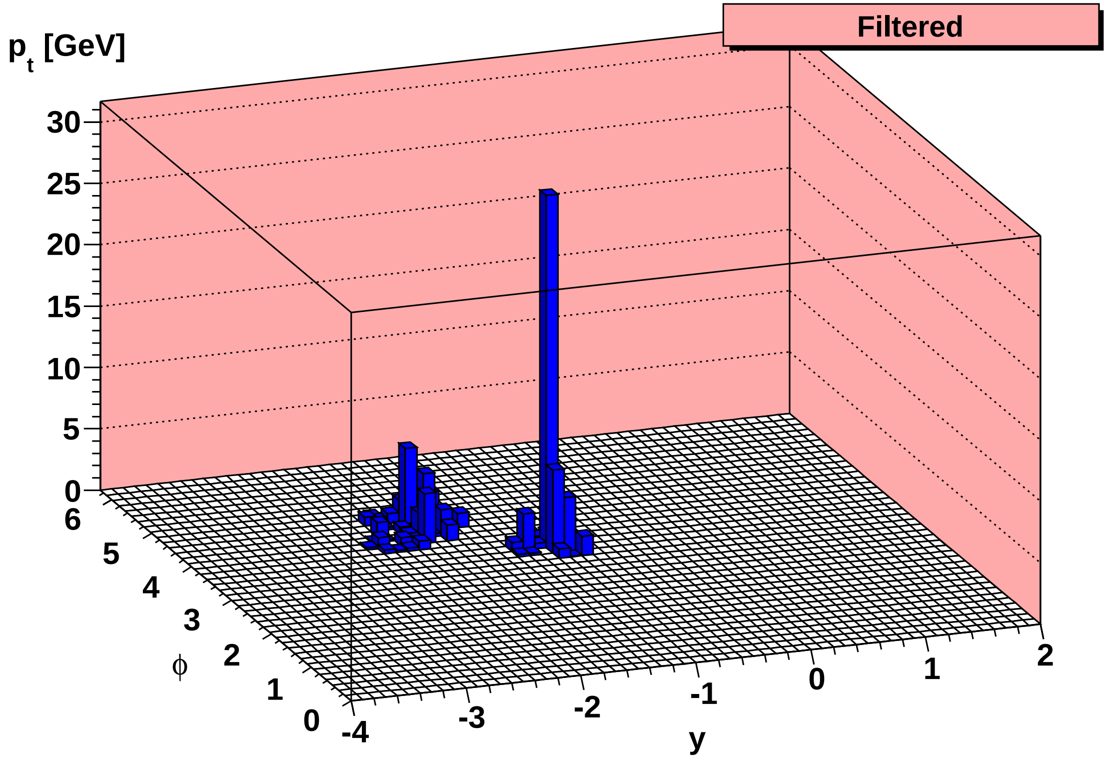
- cluster with Cambridge/Aachen(R)
- for each jet
  - recluster with Cambridge/Aachen(R/2)

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

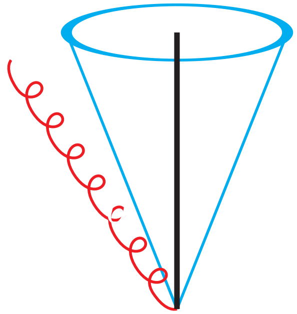
  [M.Cacciari, J.Rojo, G.Salam, GS, 08]

Jet definition optimisation

Optimisation: underlying idea

Competition between

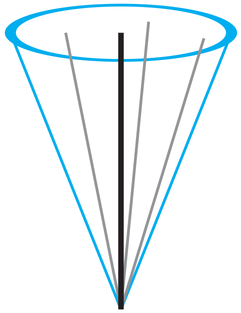
- catching perturbative radiation



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)



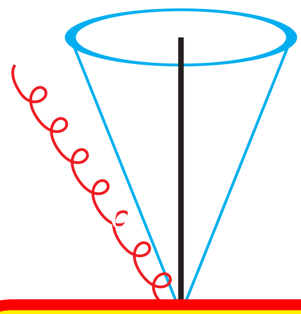
$\langle \delta p_t \rangle \sim$ Soft contents $\propto$ jet area $\sim R^2$

the coefficients depend on the algorithm

Optimisation: underlying idea

Competition between

- catching perturbative radiation

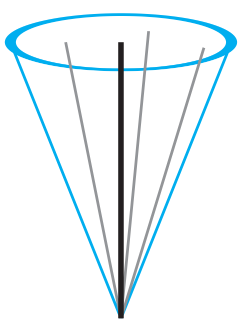


Out-of-cone radiation:

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

What is the optimal jet definition (algo+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: dijet reconstruction

Example process to illustrate various effects:

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

- $M_{Z'}$ can be varied (between 100 GeV and 4 TeV)
- Also valid for $H \to gg$ to study gluon jets
- Reconstruction method:
  - get the 2 hardest jets: j_1 and j_2
  - reconstruct the Z': $m_{Z'} = (j_1 + j_2)^2$

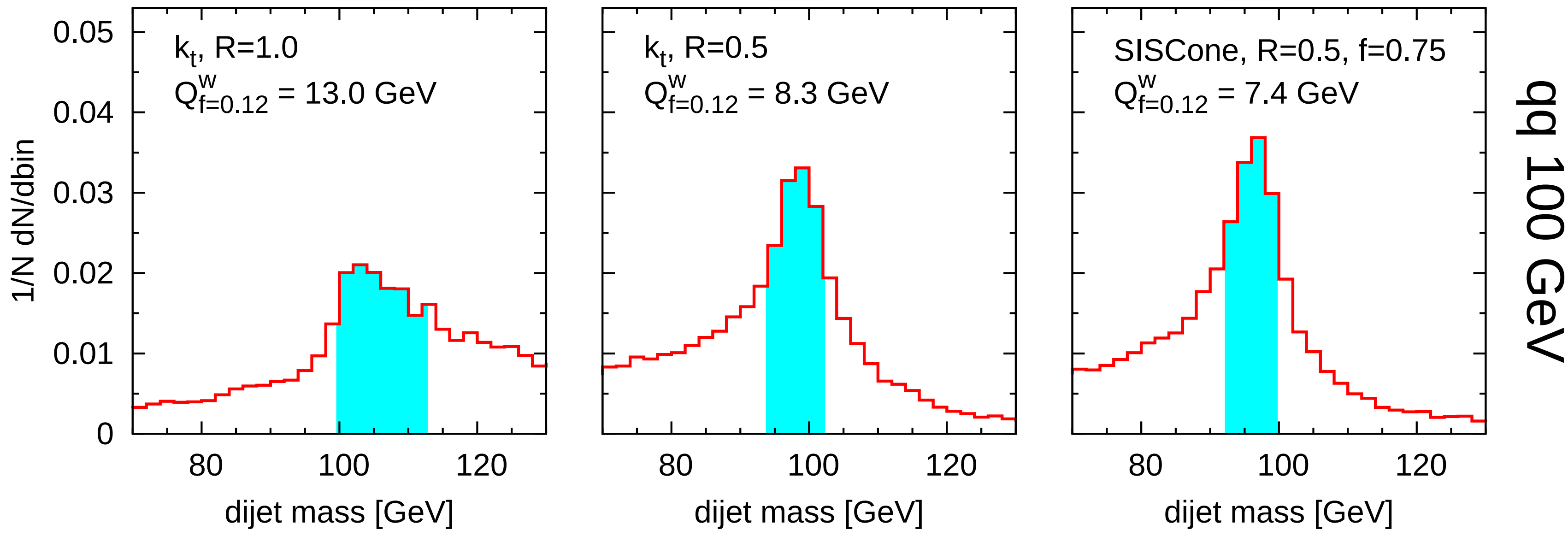
  Look how the mass peak is reconstructed
- Also $t\bar{t}$ with full hadronic decay for multijet tests

Optimisation: quality measure (1)

Measure of the jet reconstruction efficiency:

- Forget about measures related to parton-jet matching
- Forget about fits depending on the shape of the peak

⇒ maximise the signal over background ratio ($S/\sqrt{B}$)
a narrower peak is better.

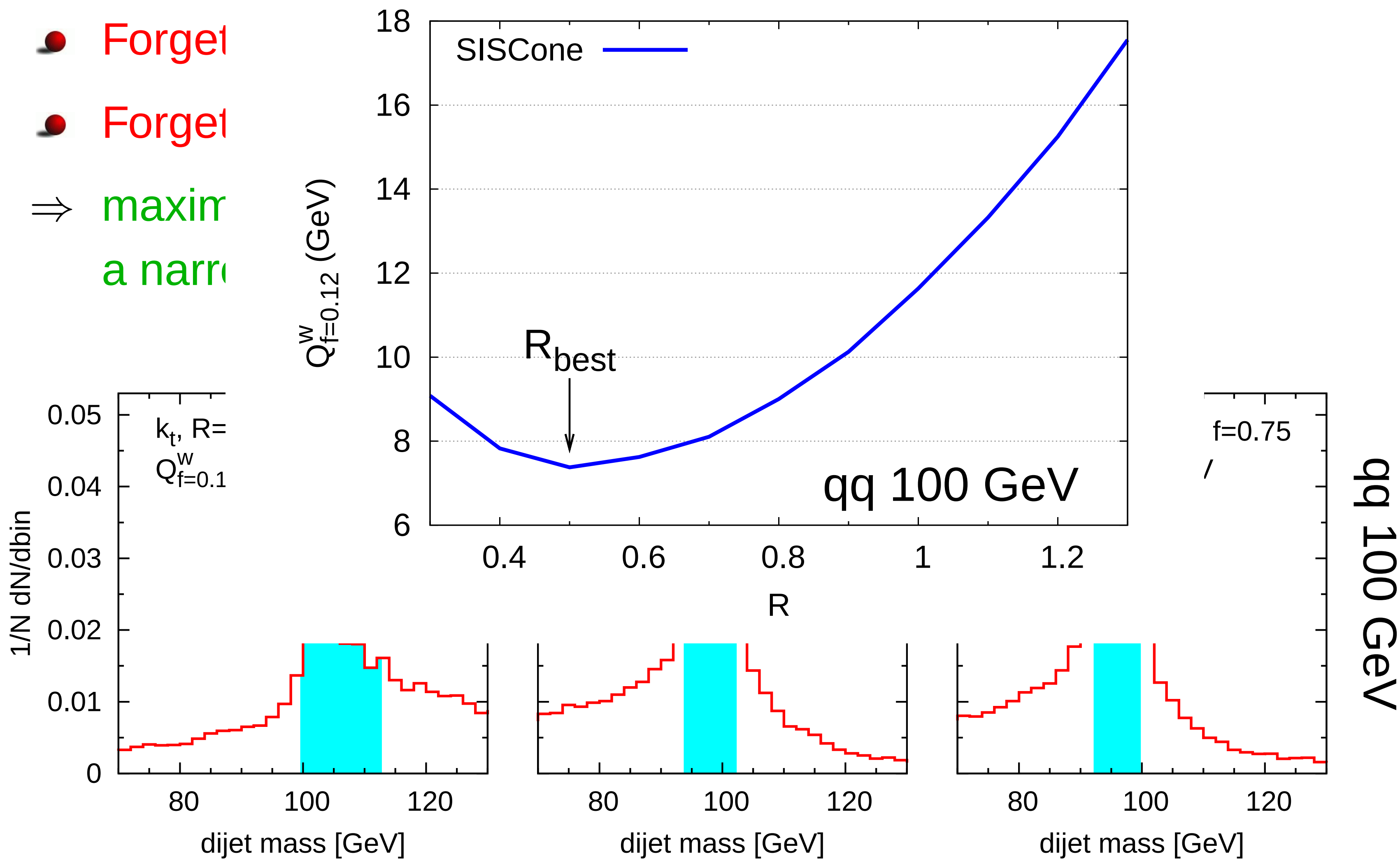


Optimisation: quality measure (1)

Measure of the jet reconstruction efficiency:

- Forget
- Forget

⇒ maxim
a narr



Optimisation: quality measure (2)

Assuming a constant background,

quality measure $\longrightarrow$ effective luminosity ratio

$$\rho_{\mathcal{L}}(\mathrm{JD}_2/\mathrm{JD}_1) = \frac{\mathcal{L} \text{ needed with } \mathrm{JD}_2}{\mathcal{L} \text{ needed with } \mathrm{JD}_1} = \frac{Q^w_{f=z}(\mathrm{JD}_2)}{Q^w_{f=z}(\mathrm{JD}_1)}$$

e.g. $\rho_{\mathcal{L}}(\mathrm{JD}_2/\mathrm{JD}_1) = 2$

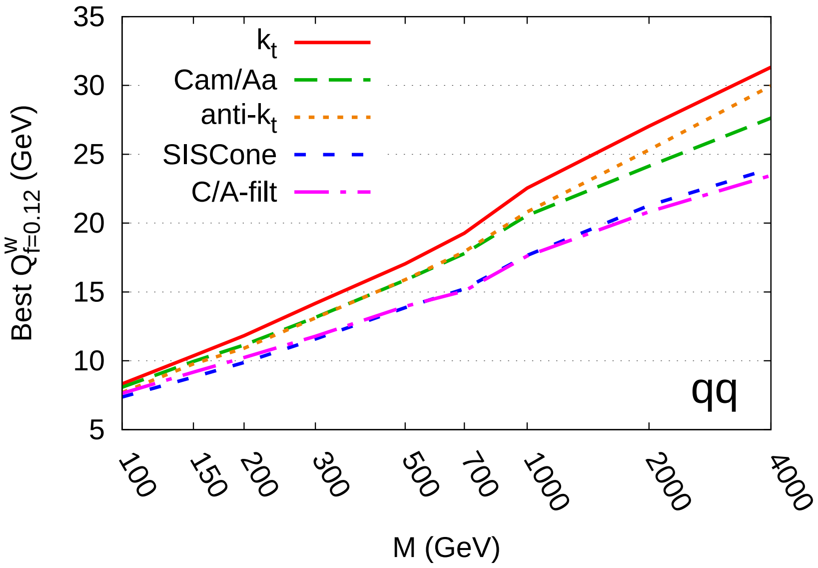
$\Leftrightarrow$ JD_2 requires 2 times the integrated luminosity of JD_1 to achieve the same discriminative power.

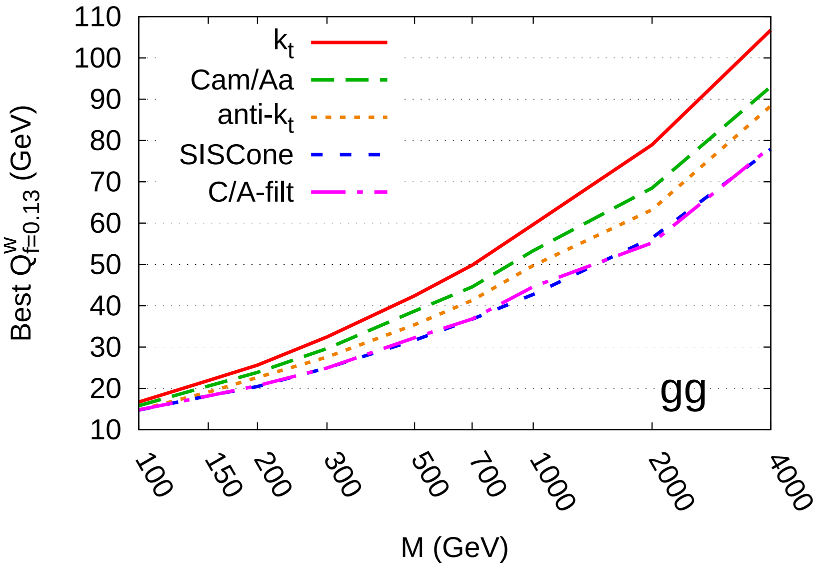
Note: results cross-checked with 2 different definitions of the quality measure

Optimisation: best definition

[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

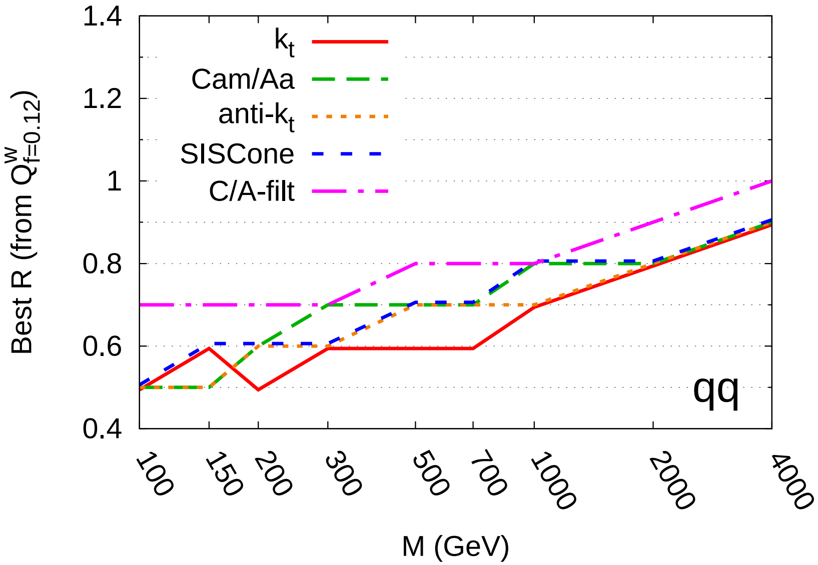


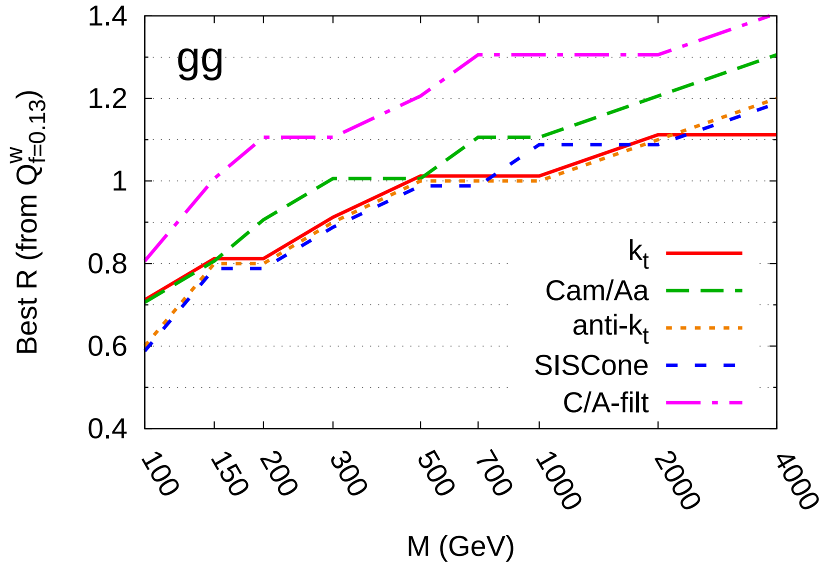


Optimisation: best definition

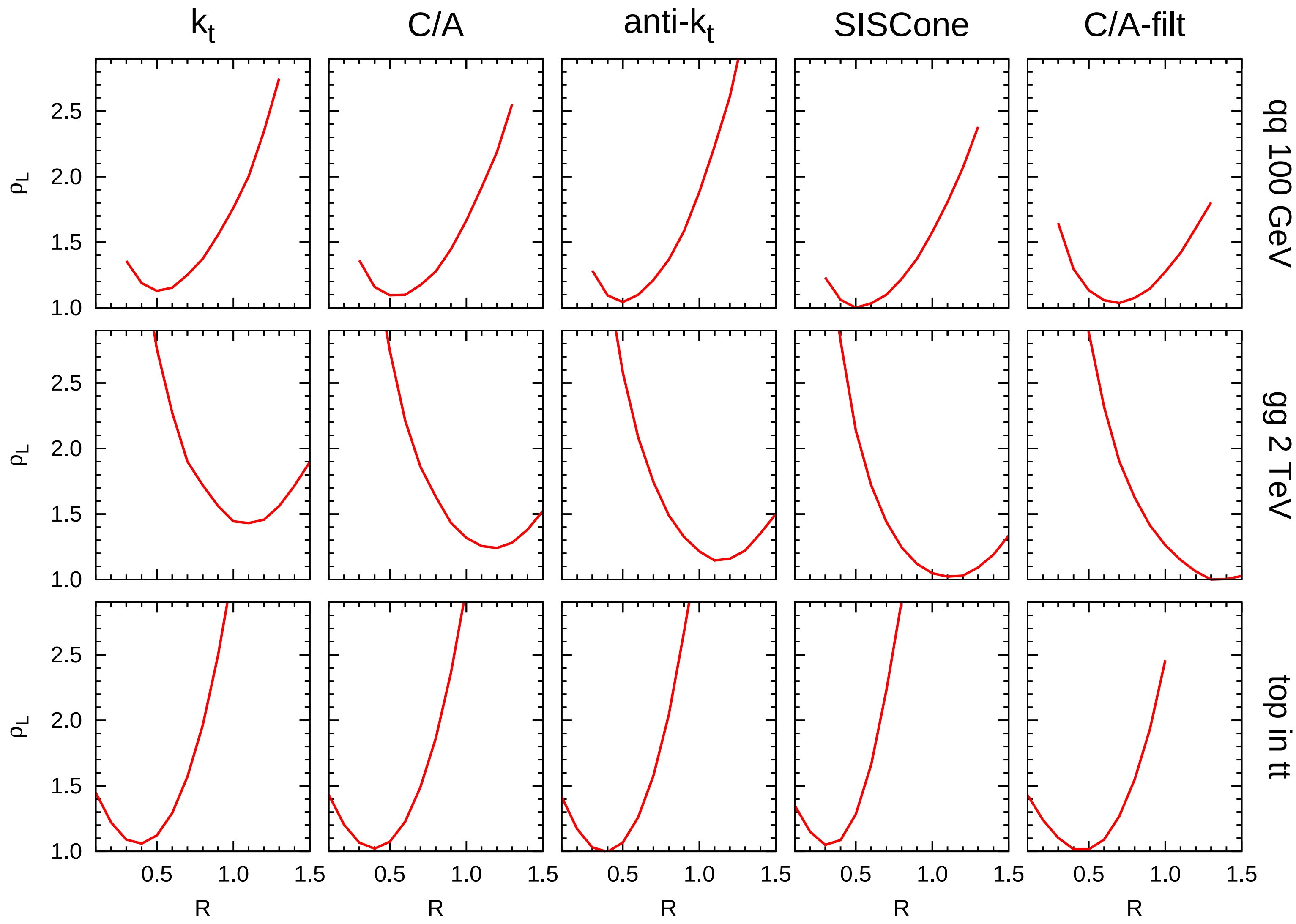
[M.Cacciari, J.Rojo, G.Salam, GS, 08]

- 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 $R_{\text{best}}(g) > R_{\text{best}}(q)$)



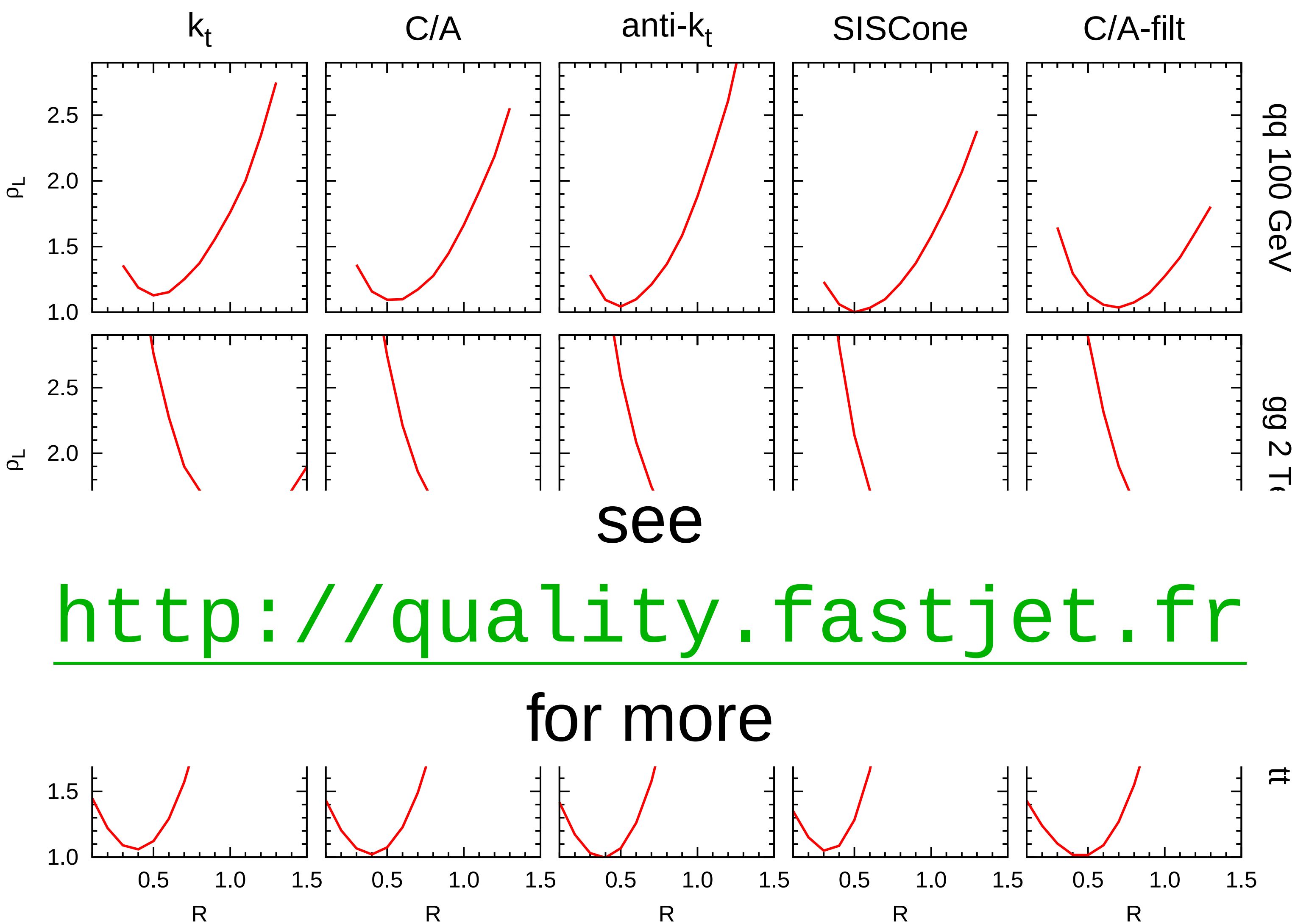


Optimisation: consequences



- single jet definition for all processes may cost a factor ~ 2 in $\mathcal{L}_{\rm int}$
- The choice of R plays a particularly important role

Optimisation: consequences



see

http://quality.fastjet.fr

for more

- single jet definition for all processes may cost a factor ~ 2 in $\mathcal{L}_{\mathrm{int}}$
- The choice of R plays a particularly important role

Analytic computation of the optimal radius

[GS, in prep]

Approach

- Spectrum $\frac{dP}{d\delta m}$ ($\delta m = m_{\rm rec} - M$): convolute different pieces
  - final-state gluon emission
  - initial-state gluon emission (+ PDF effects)

$$\frac{dP_{i,f}}{d\delta m} = \frac{K_{i,f}}{|\delta m| \; \log(M/(2\Lambda_{i,f}))} \left[\frac{\log(\delta m/\Lambda_{i,f})}{\log(M/(2\Lambda_{i,f}))}\right]^{K_{i,f}-1}$$

with $K_f = \frac{C_R}{\beta_0 \pi} 2\pi \log(2/R)$, $K_i = \frac{C_R}{\beta_0 \pi} \pi R^2$.

Approach

- Spectrum $\frac{dP}{d\delta m}$ ($\delta m = m_{\rm rec} - M$): convolute different pieces
  - final-state gluon emission
  - initial-state gluon emission (+ PDF effects)

$$\frac{dP_{i,f}}{d\delta m} = \frac{K_{i,f}}{|\delta m| \, \log(M/(2\Lambda_{i,f}))} \left[\frac{\log(\delta m/\Lambda_{i,f})}{\log(M/(2\Lambda_{i,f}))}\right]^{K_{i,f}-1}$$

with $K_f = \frac{C_R}{\beta_0 \pi} 2\pi \log(2/R)$, $K_i = \frac{C_R}{\beta_0 \pi} \pi R^2$.

  - UE:

$$\frac{dP_{UE}}{d\delta m} = \frac{1}{\Gamma(\alpha)} \frac{\delta m^{\alpha-1}}{\delta m_0^\alpha} e^{-\delta m/\delta m_0}$$

Reproduce average shift and dispersion

- fluctuations of the UE from one event to another ($\propto R^2$)
- fluctuations of the UE inside an event ($\propto R$)
- fluctuations of the jet area ($\propto R^2$)

Approach

- Spectrum $\frac{dP}{d\delta m}$ ($\delta m = m_{\rm rec} - M$): convolute different pieces
  - final-state gluon emission
  - initial-state gluon emission (+ PDF effects)

$$\frac{dP_{i,f}}{d\delta m} = \frac{K_{i,f}}{|\delta m|\,\log(M/(2\Lambda_{i,f}))}\left[\frac{\log(\delta m/\Lambda_{i,f})}{\log(M/(2\Lambda_{i,f}))}\right]^{K_{i,f}-1}$$

  with $K_f = \frac{C_R}{\beta_0\pi}\,2\pi\log(2/R)$, $K_i = \frac{C_R}{\beta_0\pi}\,\pi R^2$.

  - UE:

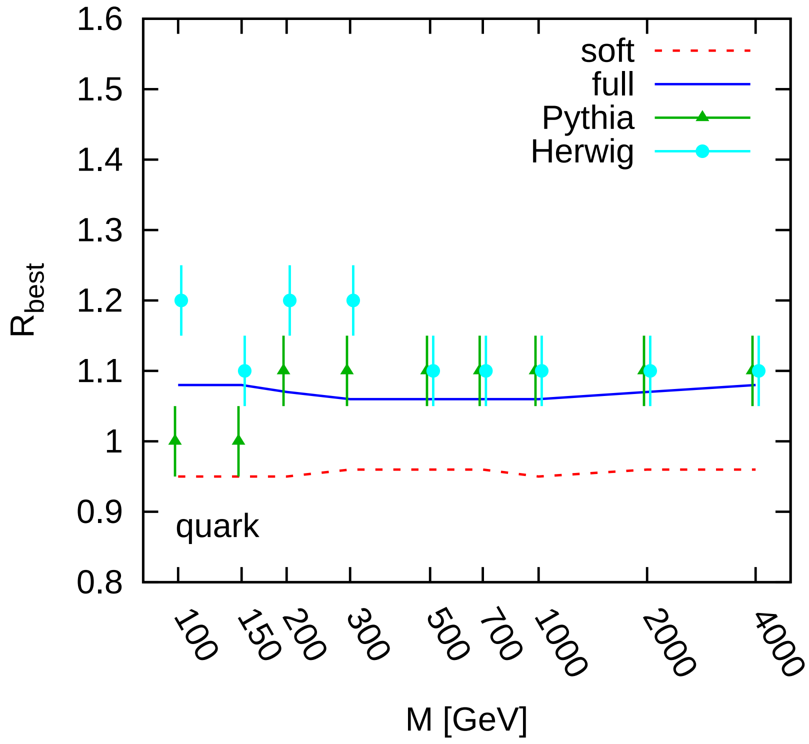
$$\frac{dP_{UE}}{d\delta m} = \frac{1}{\Gamma(\alpha)}\,\frac{\delta m^{\alpha-1}}{\delta m_0^{\alpha}}\,e^{-\delta m/\delta m_0}$$

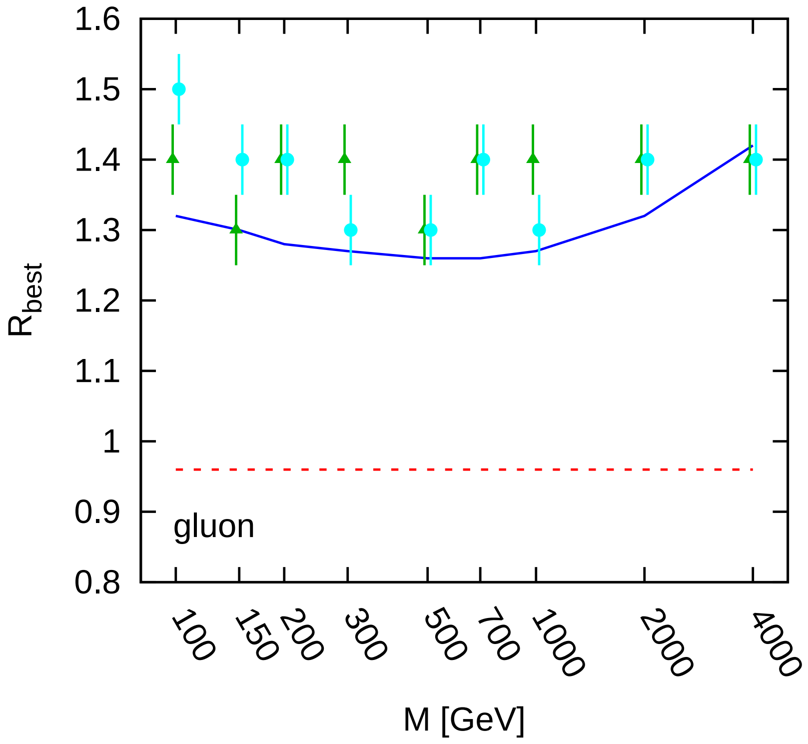
  Reproduce average shift and dispersion
    - fluctuations of the UE from one event to another ($\propto R^2$)
    - fluctuations of the UE inside an event ($\propto R$)
    - fluctuations of the jet area ($\propto R^2$)

- Compute the quality measure, then $R_{\rm best}$

R_{best} *at parton-level*

Just use ISR and FSR radiation

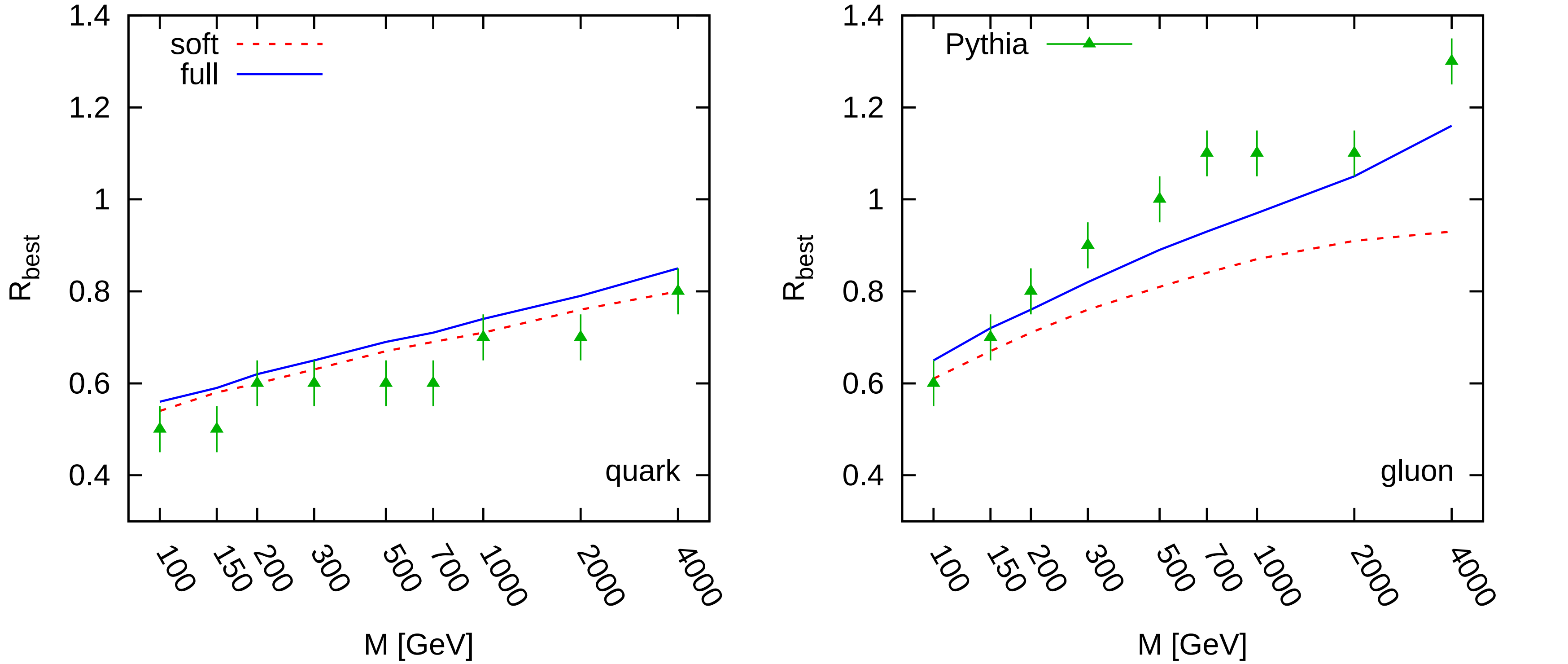




'full' = 'soft' + PDF effects

R_{best} *for full events*

Also add UE contribution



'full' = 'soft' + PDF effects

Summary (1)

Message #1:

Use infrared-and-collinear-safe algorithms

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

Important to benefit fully from pQCD multilegs/multiloops calculations

Summary (2)

Message #2:

keep some flexibility in the jet definition choice

- optimisation $\longrightarrow$ luminosity gains for LHC searches
- the optimal parameter R varies with the process!
- analytic control on its way

backup slides

The SISCone search for stable cones

- Solution: use a seedless approach, find ALL stable cones
- Naive approach: check stability of each subset of particle

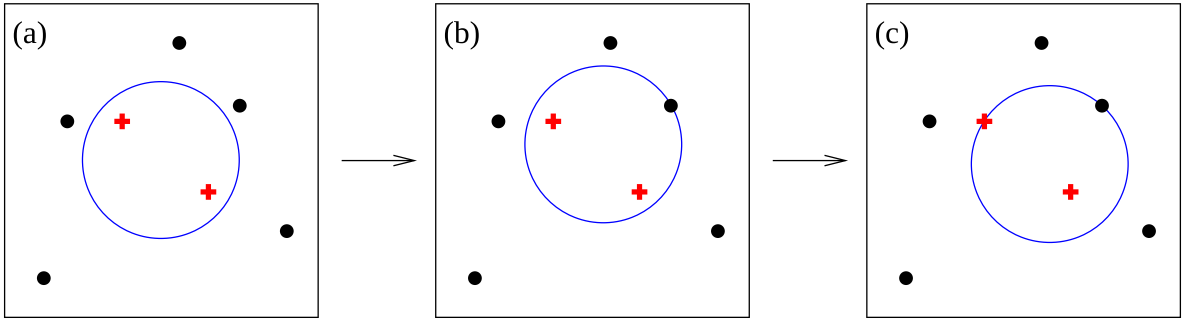
The SISCone search for stable cones

- Solution: use a seedless approach, find ALL stable cones
- Naive approach: check stability of each subset of particle
  Complexity is $\mathcal{O}\left(N2^N\right)$
  $\Rightarrow$ definitely unrealistic: 10^{17} years for $N = 100$
- Midpoint complexity: $\mathcal{O}\left(N^3\right)$

The SISCone search for stable cones

- Solution: use a seedless approach, find ALL stable cones
- Midpoint complexity: $\mathcal{O}\left(N^3\right)$

Idea: use geometric arguments



- Each enclosure can be moved (in any dir.) until it touches a point
- ... then rotated until it touches a second one

$\Rightarrow$ Enumerate all pairs of particles
with 2 circle orientations and 4 possible inclusion/exclusion
$\longrightarrow$ find all enclosures

The SISCone search for stable cones

- Solution: use a seedless approach, find ALL stable cones
- Midpoint complexity: $\mathcal{O}\left(N^3\right)$

Idea: use geometric arguments

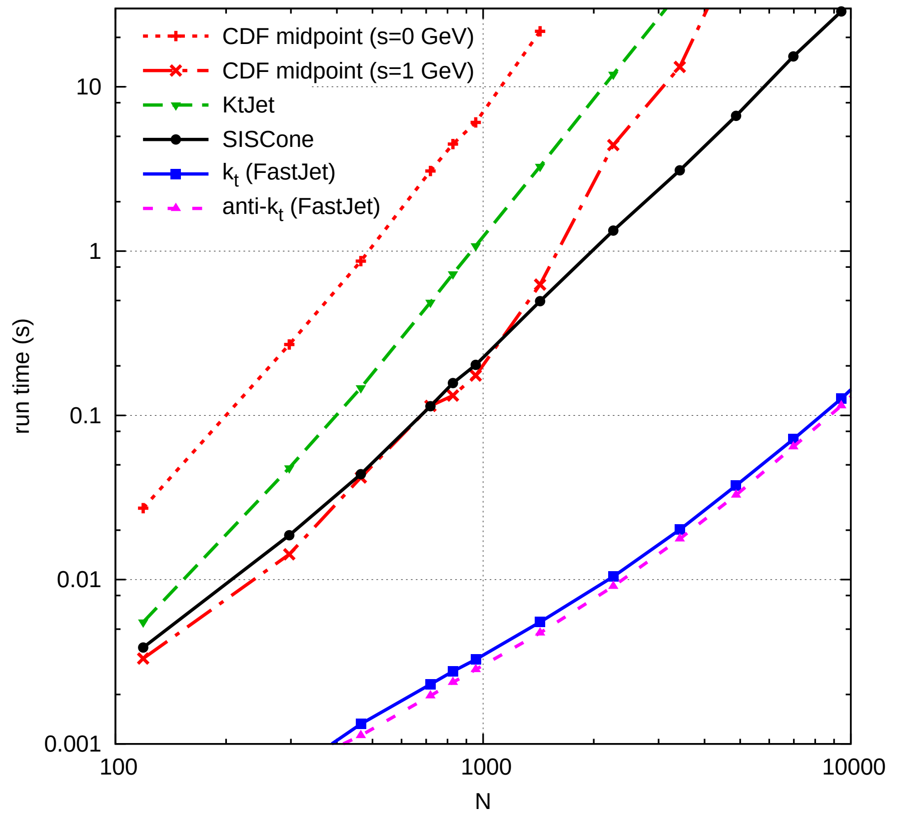
$\Rightarrow$ Enumerate all pairs of particles
with 2 circle orientations and 4 possible inclusion/exclusion
$\longrightarrow$ find all enclosures

- Complexity: $\mathcal{O}\left(N^3\right)$, with improvements: $\mathcal{O}\left(N^2\log(N)\right)$

$\longrightarrow$ C++ implementation: Seedless Infrared-Safe Cone algorithm (SISCone)
G.Salam, G.S., JHEP 04 (2007) 086; http://projects.hepforge.org/siscone

NB.: also available from FastJet
[M.Cacciari, G.Salam, G.S.]; http://www.fastjet.fr

Algorithm timings



- Recombination algorithms very fast

[M. Cacciari, G. Salam, 06]

- SISCone not slower than Midpoint (even with a 1 GeV seed threshold)