Defining jets at the LHC

Grégory Soyez

Brookhaven National Laboratory

in collaboration with G. Salam, M. Cacciari and J. Rojo

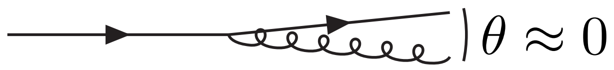
arXiv:0704:0292, arXiv:0802:1189, arXiv:0810.1304

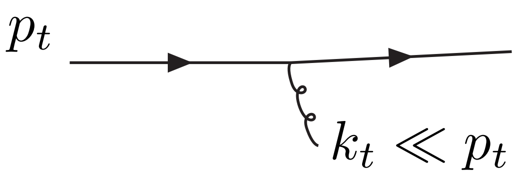
Unavoidable theory

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 QCD?

Lot of QCD in the final-state at the LHC

- QCD studies: *e.g.* $t \to bq\bar{q}$
- new physics:
  - $H \to b\bar{b}$
  - $Z' \to q\bar{q}$
  - SUSY $\to$ QCD
- backgrounds: *e.g.* $gg \to b\bar{b}$, ...

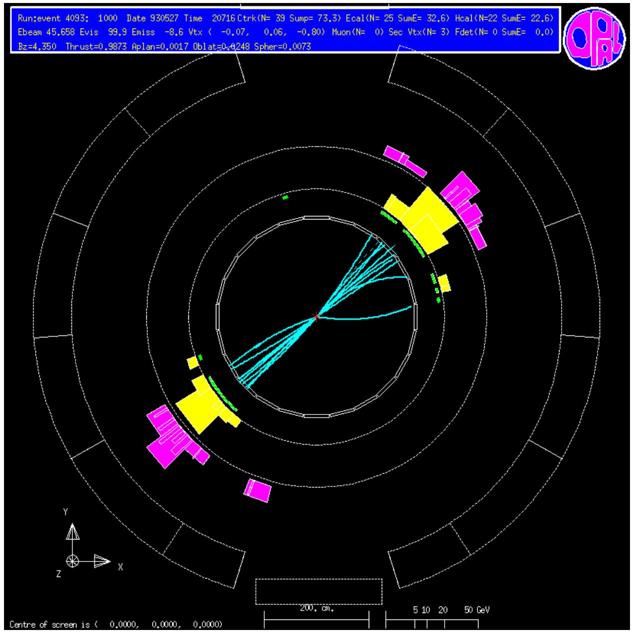
$\Rightarrow$ Huge effort to compute multileg/NLO processes in pQCD ($\sim$ 100 M$)

This talk: how to avoid wasting that effort, in the optimal way

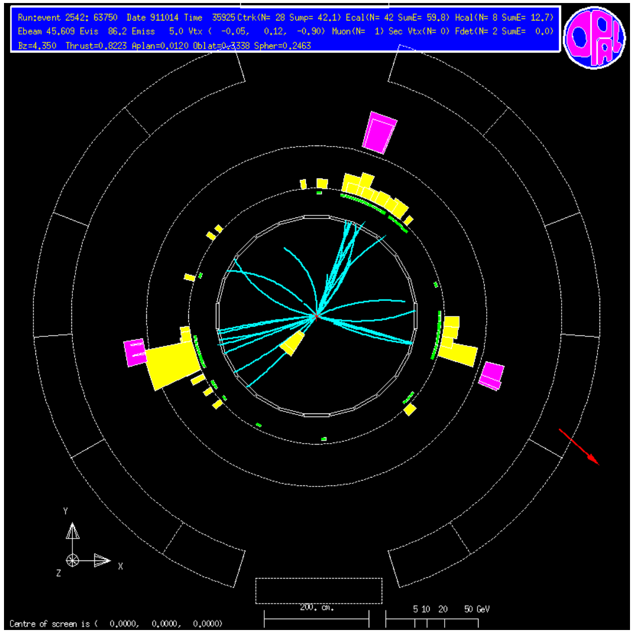
Motivation: why jets

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

Example: LEP (OPAL) events



2 jets



3 jets

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

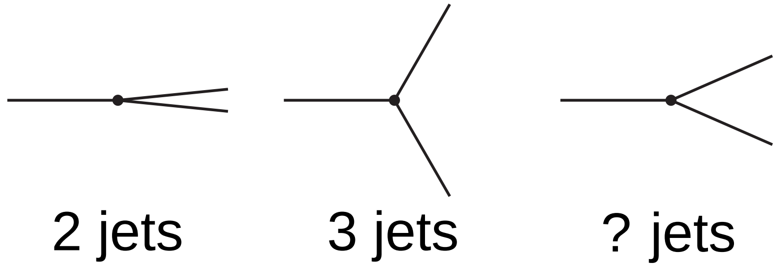
Motivation: why jets

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

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

BUT

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



$\Rightarrow$ Different jet definitions

20th century jet finders

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 finders

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 finders

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

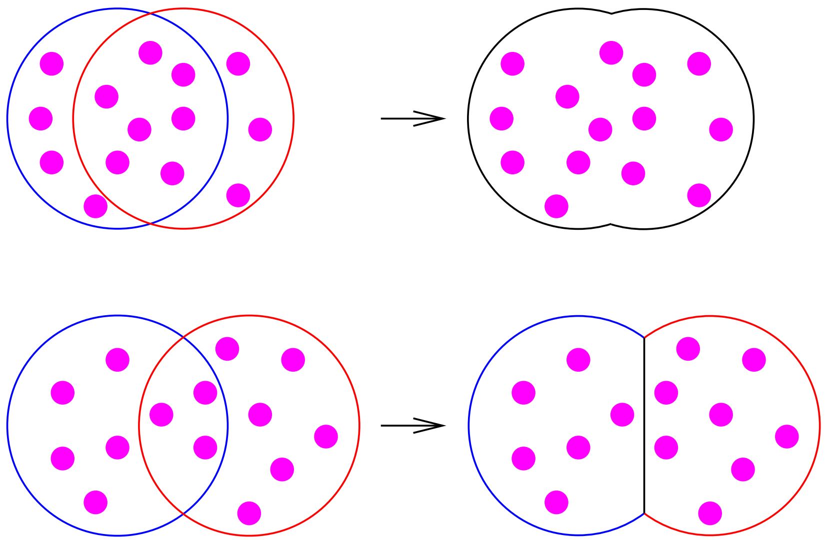
Cone:

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

20th century jet finders

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 finders

Cone with progressive removal

Successively

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

Basic property:

hard circular jets

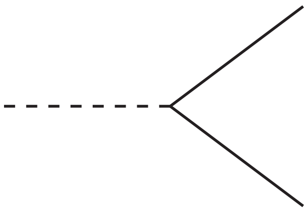
Cone:

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

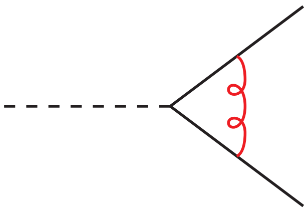
21st century: how does that picture change?

QCD divergences

Ingredient: QCD soft and collinear divergencies

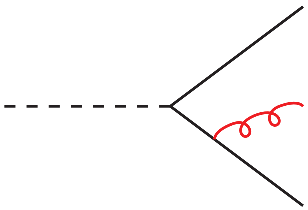


LO



NLO(virt)

∞



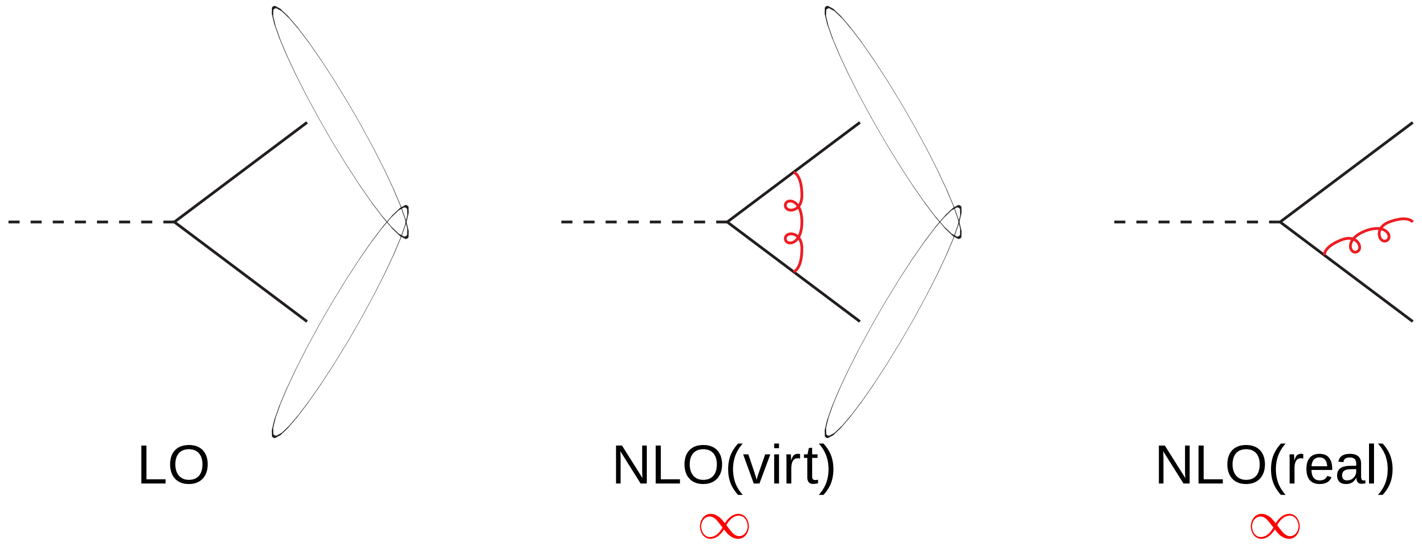
NLO(real)

∞

- ∞ (from soft gluons) cancel

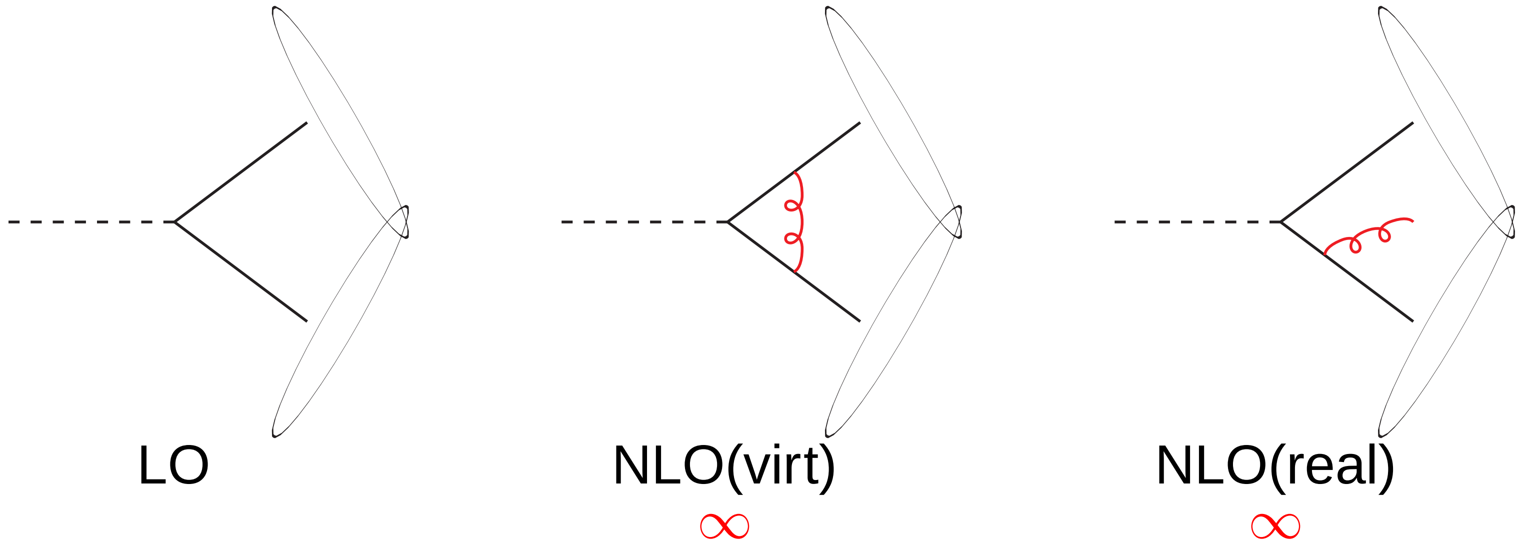
QCD divergences

Ingredient: QCD soft and collinear divergencies



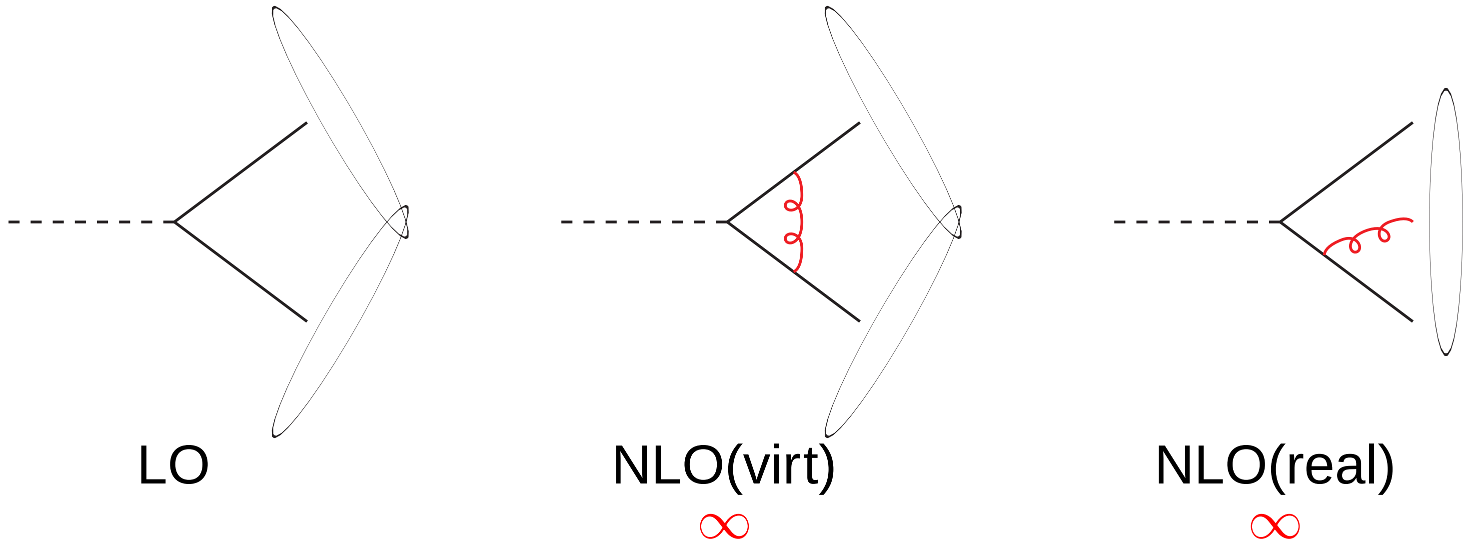
- Consider an extra (NLO) soft gluon
- Assume LO gives 2 jets $\Rightarrow$ NLO(virt) gives 2 jets

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 $\Rightarrow \infty$ cancel $\Rightarrow$ finite jet cross-section

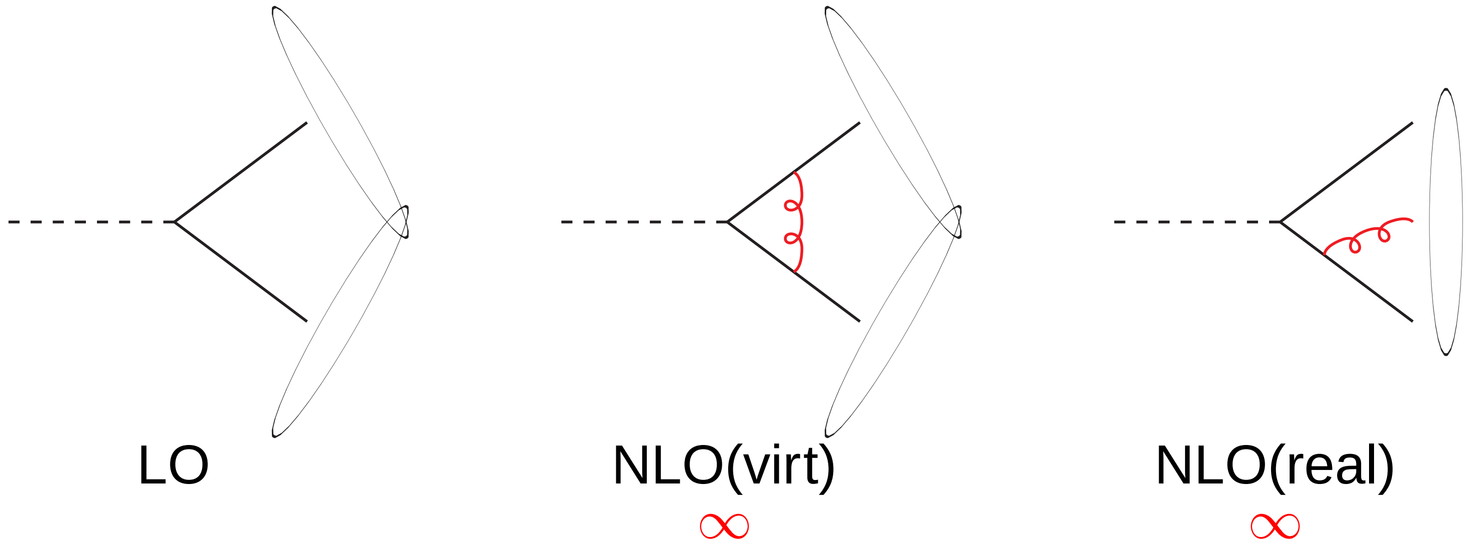
Ingredient: QCD soft and collinear divergencies



- 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 cross-section

QCD divergences

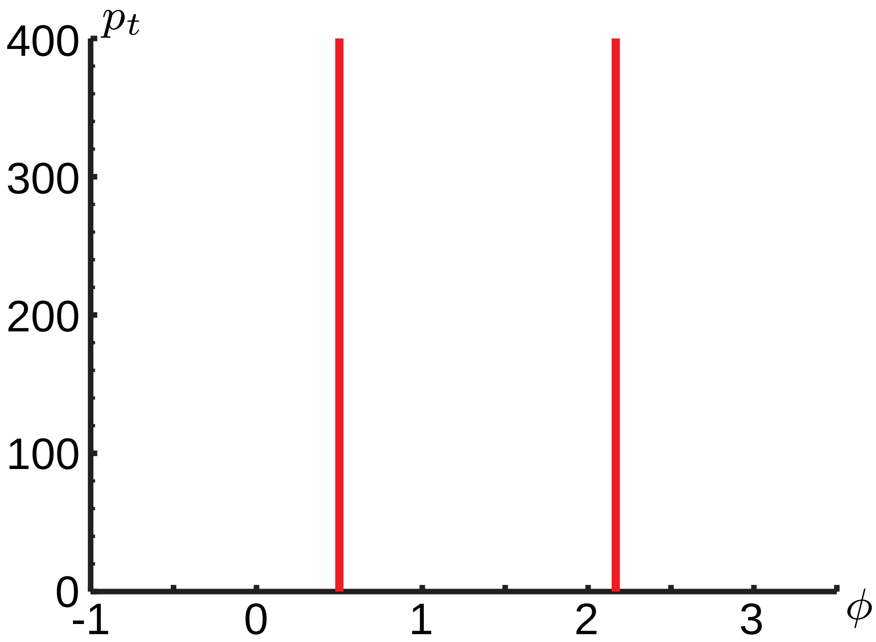
Ingredient: QCD soft and collinear divergencies

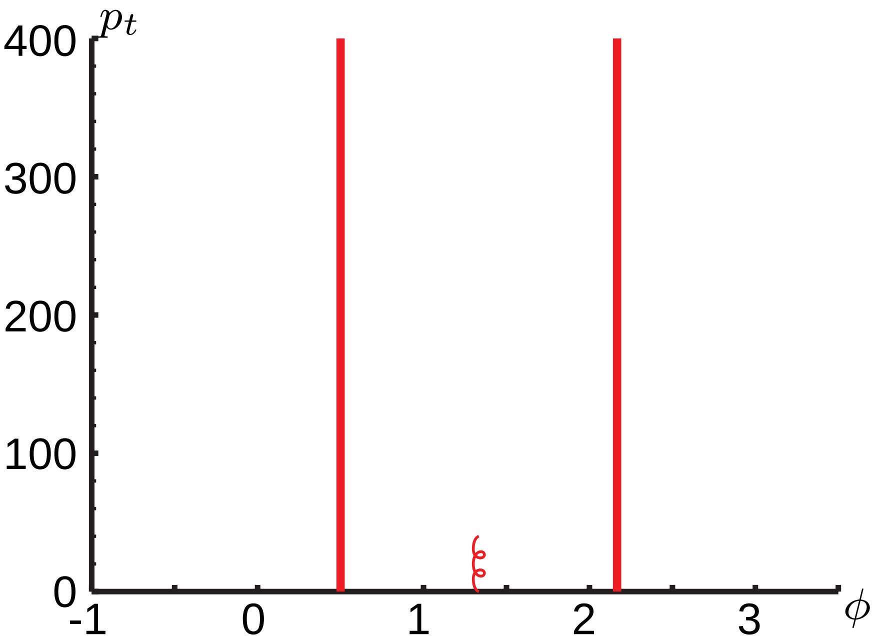


For pQCD to make sense, the (hard) jets should not change when

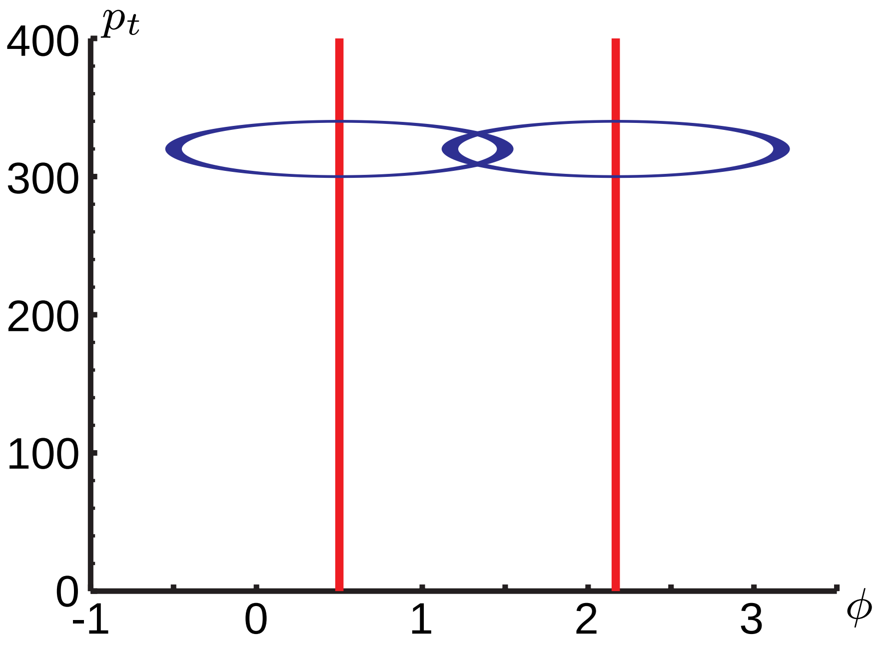
- one has a soft emission *i.e.* adds a very soft gluon
- one has a collinear splitting
  i.e. replaces one parton by two at the same place (η, ϕ)

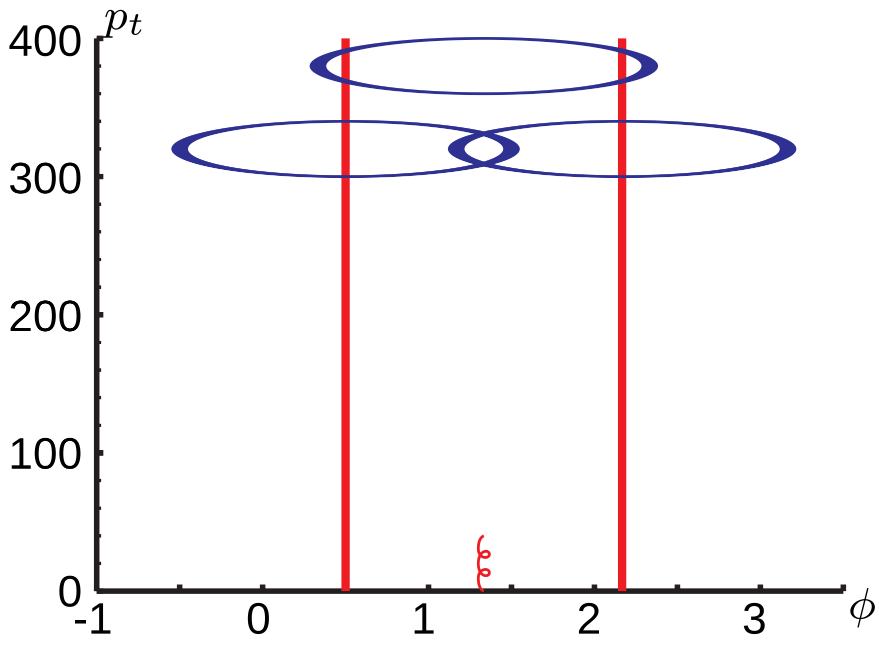
IR (un)safety? JetClu and Atlas Cone





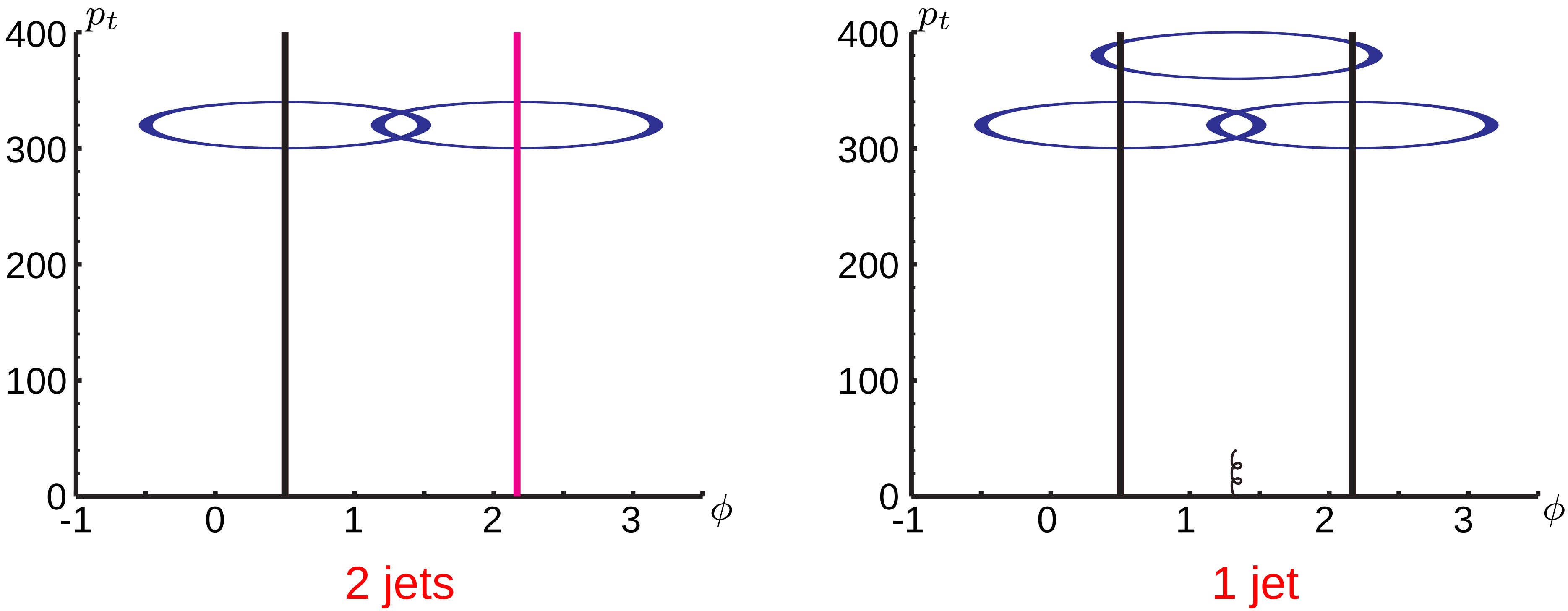
IR (un)safety? JetClu and Atlas Cone





Stable cones found

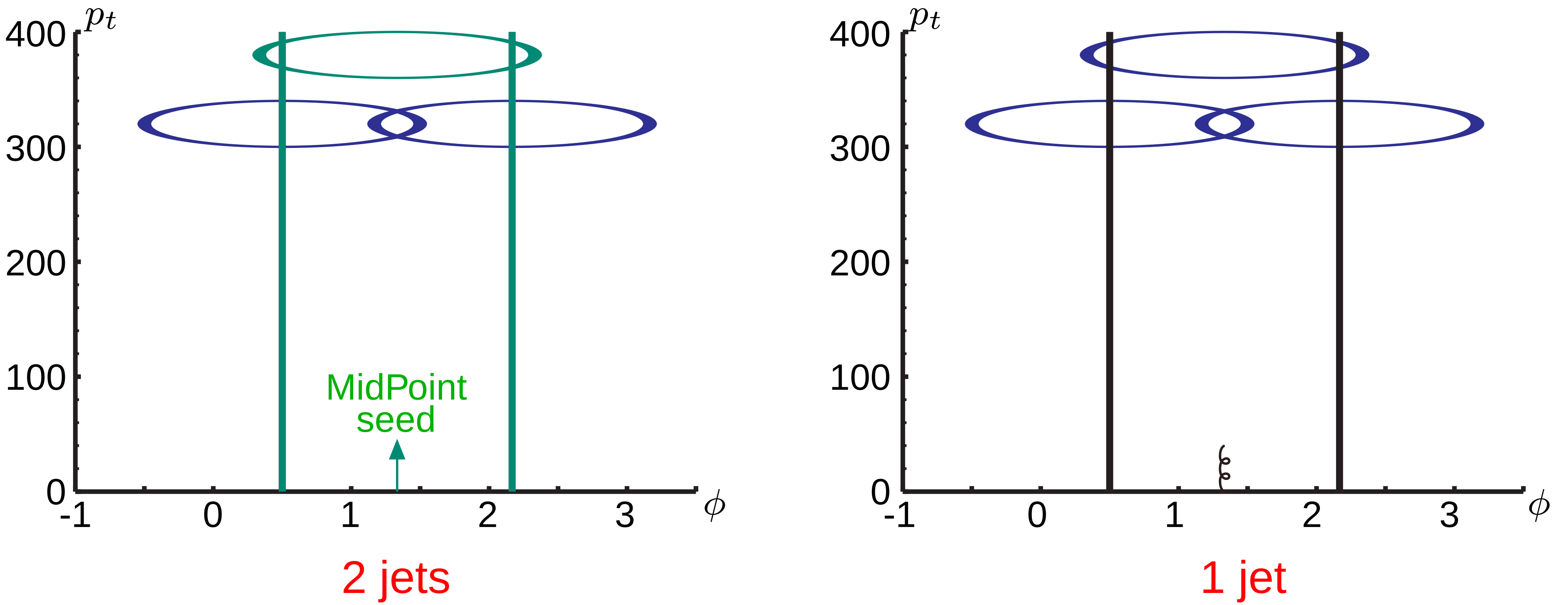
IR (un)safety? JetClu and Atlas Cone



A soft gluon changed the number of jets

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

IR (un)safety? JetClu and Atlas Cone



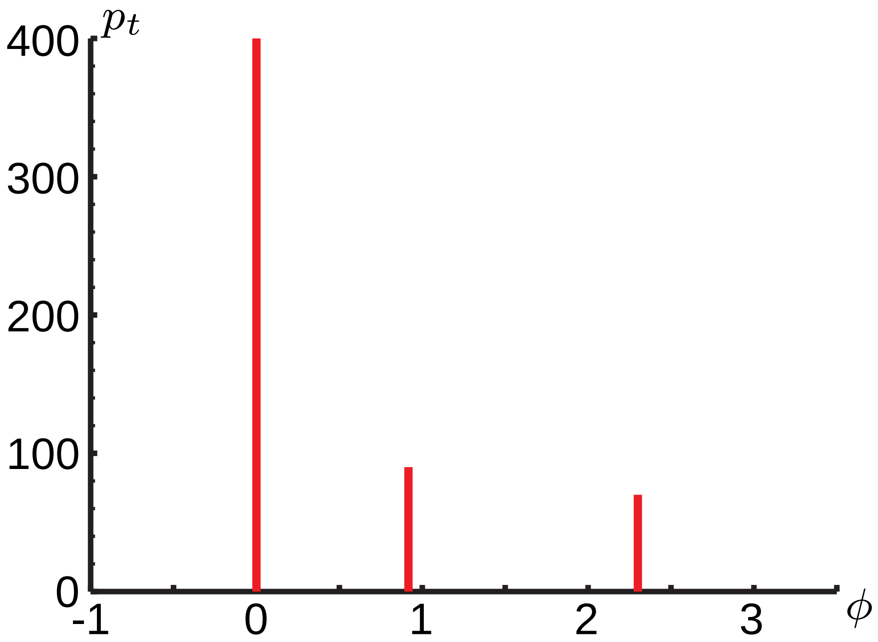
A soft gluon changed the number of jets

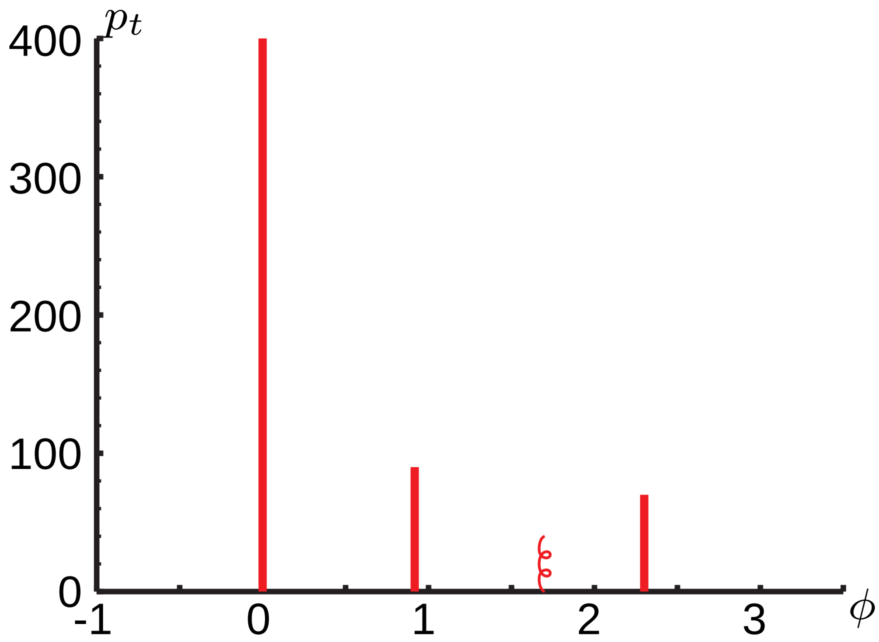
$\Rightarrow$ IR unsafety of JetClu and the ATLAS Cone

Fixed by MidPoint

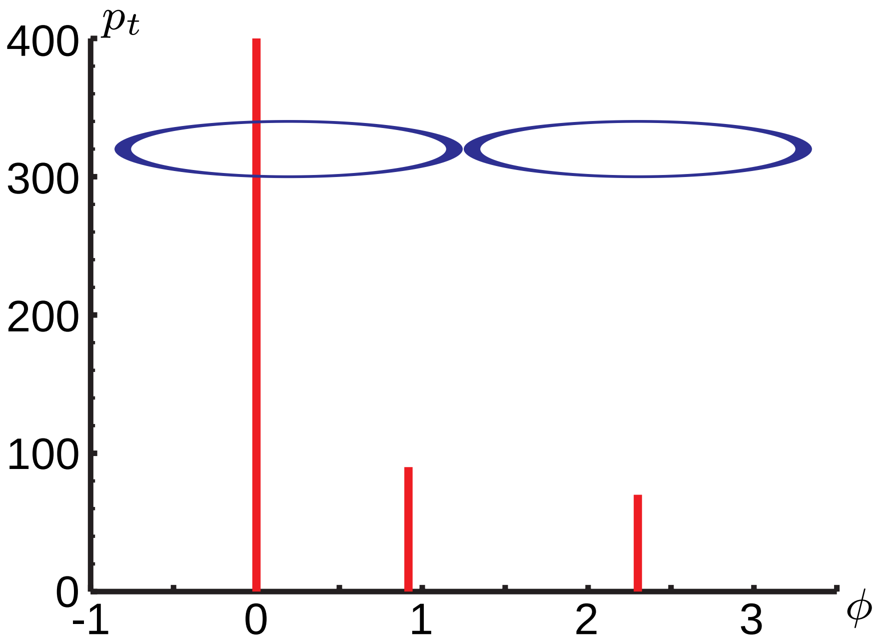
[Blazey *et al.*, 00]

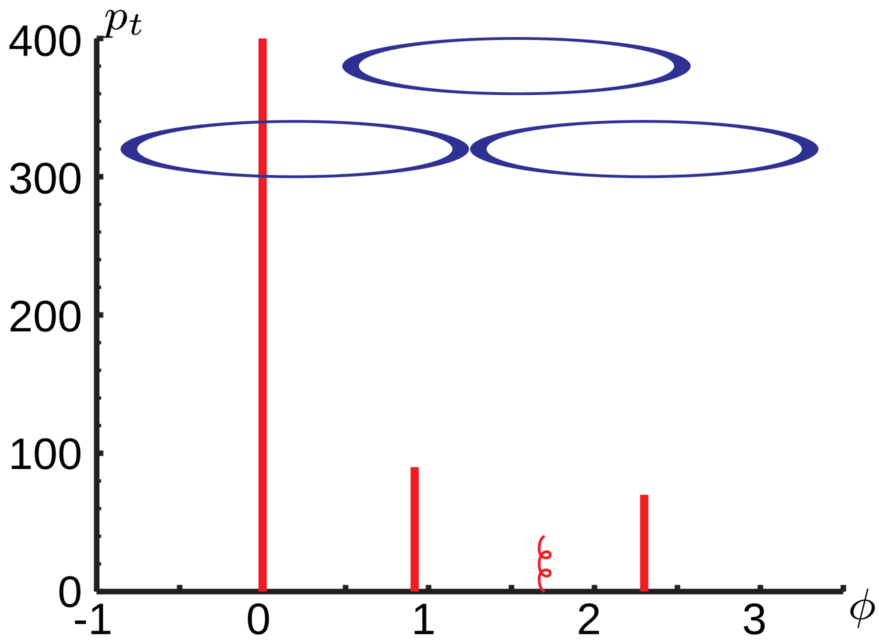
IR (un)safety? MidPoint





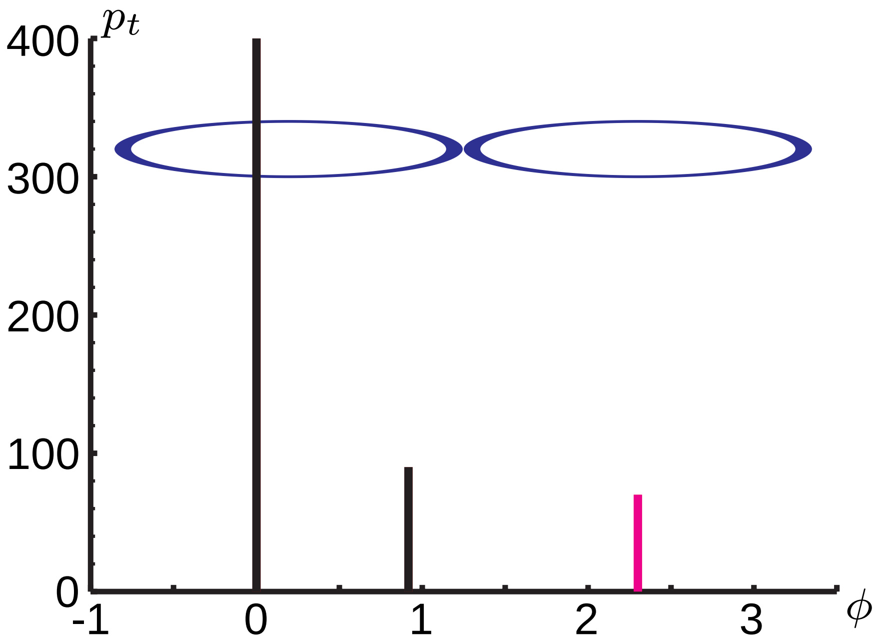
IR (un)safety? MidPoint



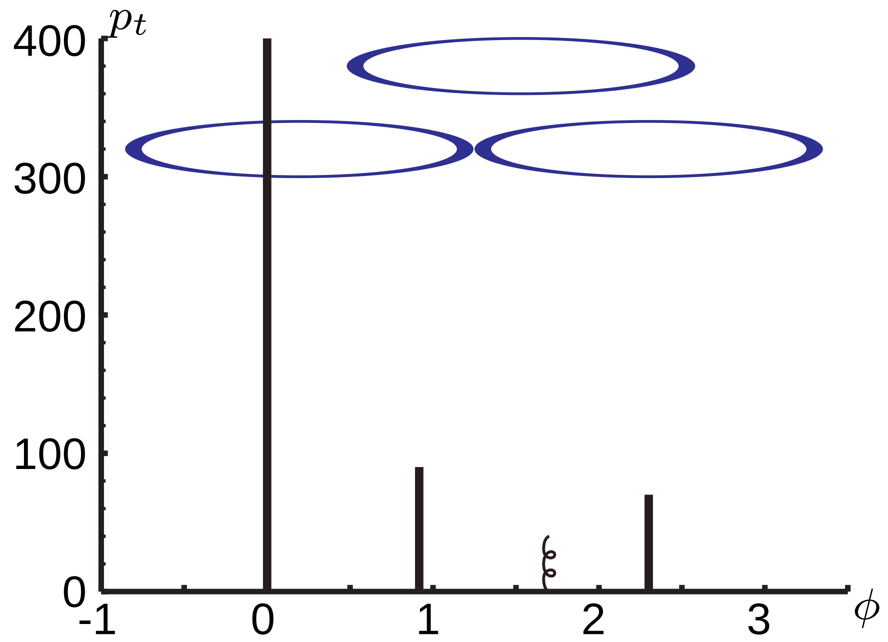


Stable cones found

IR (un)safety? MidPoint



2 jets

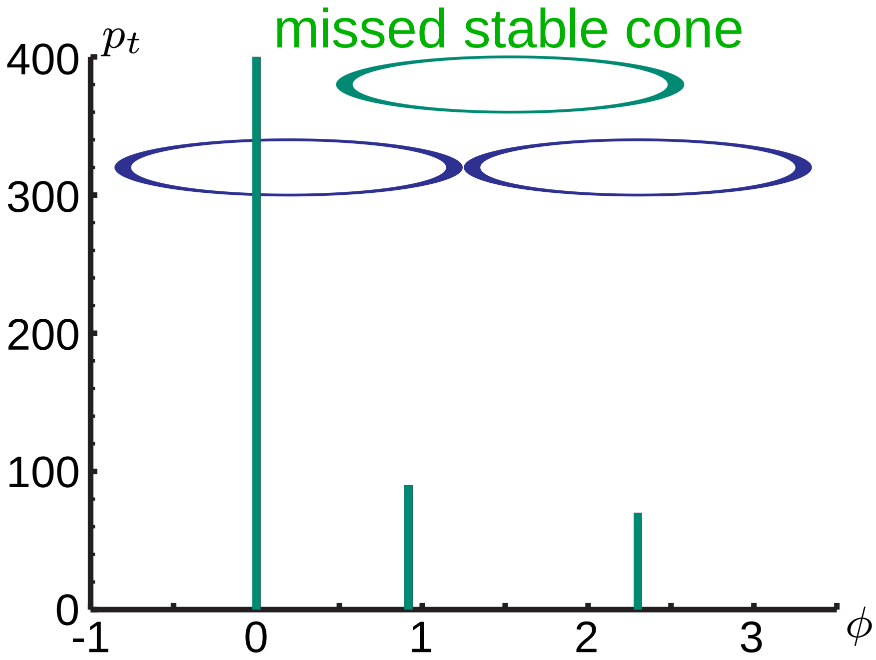


1 jet

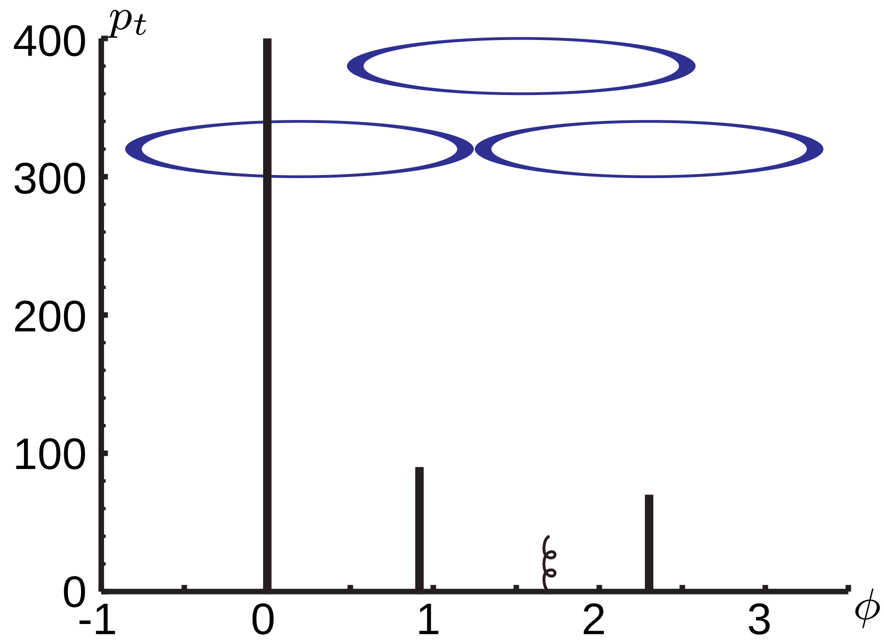
A soft gluon changed the number of jets

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

IR (un)safety? MidPoint



2 jets



1 jet

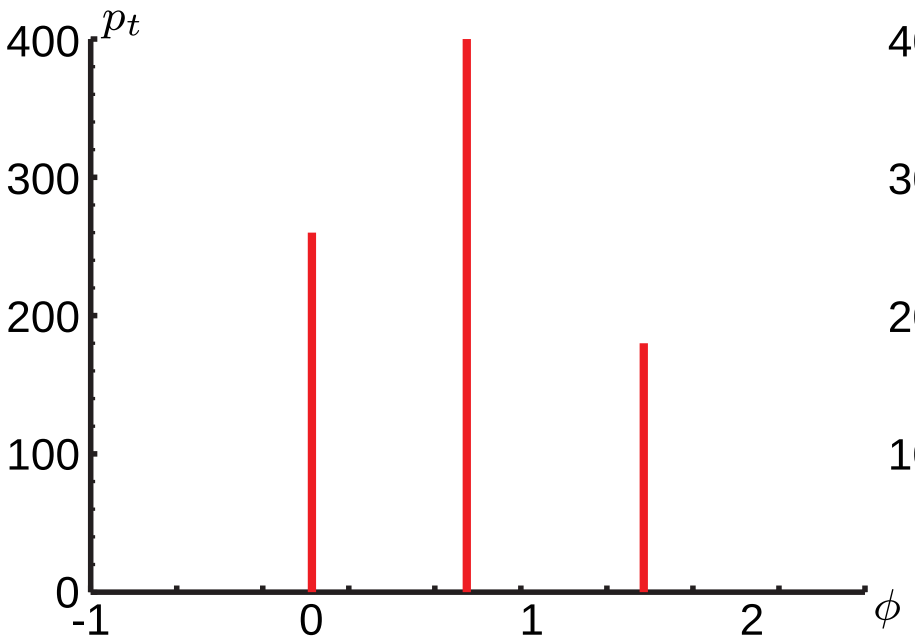
Solution: be sure to find **all** stable cones

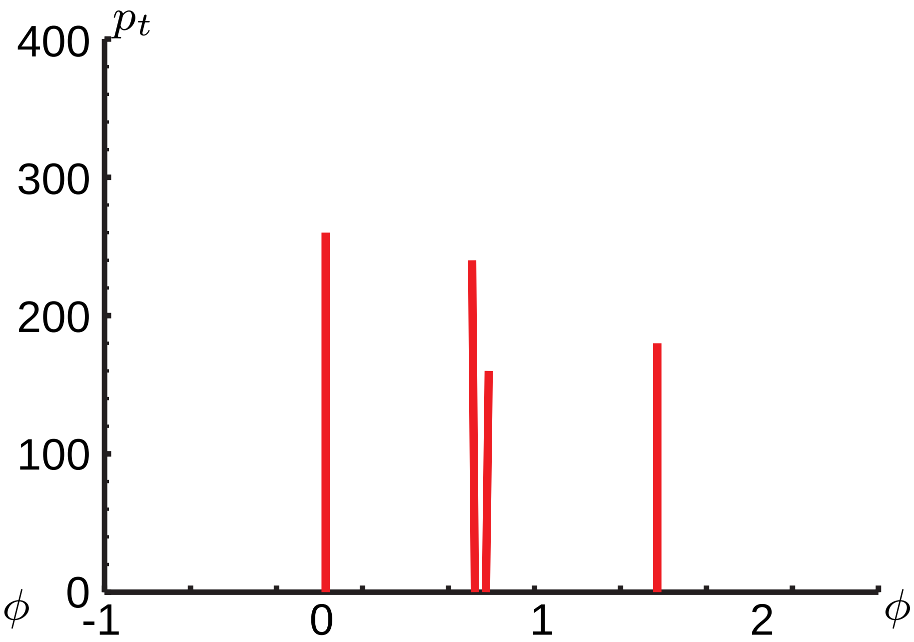
> SISCone: Seedless Infrared-Safe Cone algorithm
> `http://projects.hepforge.org/siscone`

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

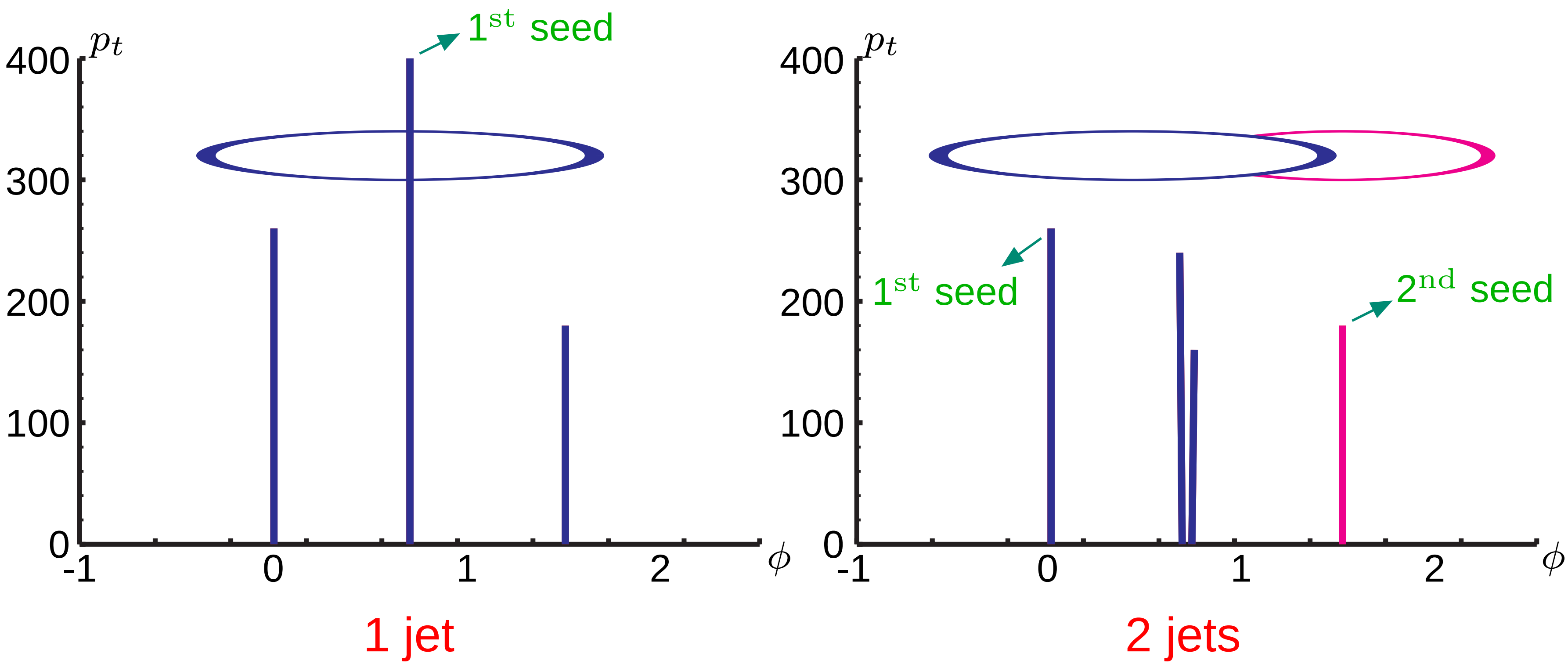
Idea: enumerate enclosures by enumerating pairs of particles

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
- $p = 0$: Cambridge/Aachen 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
- $p = 0$: Cambridge/Aachen 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

4 available
safe algorithms

All accessible from **FastJet**

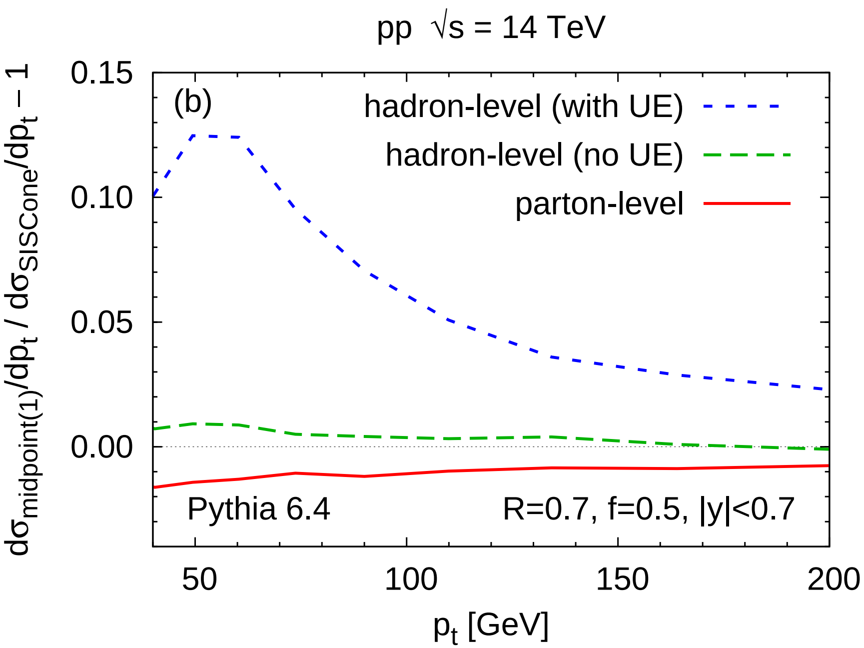
http://www.fastjet.fr [M.Cacciari, G.Salam, G.S.]

Physical impact

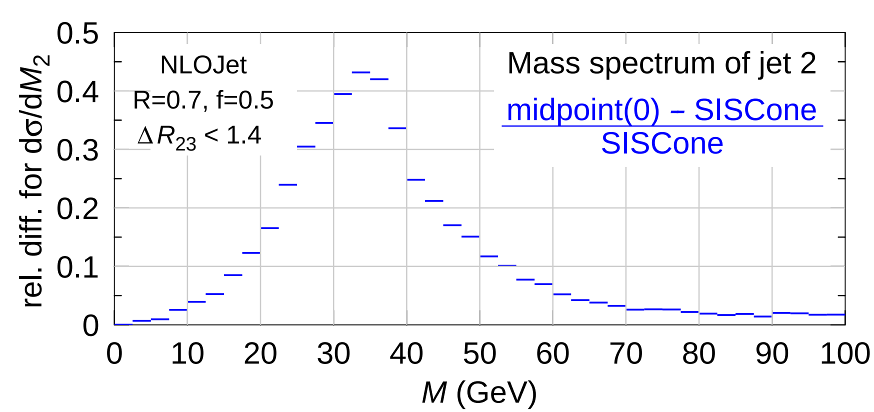
MidPoint/CMS iterative cone unsafe at $\mathcal{O}(\alpha_s^4)$ (or $\mathcal{O}(\alpha_{\rm ew}\alpha_s^3)$)

Physical observable	IRC-safe until		
	JetClu/ATLAS cone	MidPoint/CMS it. cone	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

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

Physical observable	IRC-safe until		
	JetClu/ATLAS cone	MidPoint/CMS it. cone	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 theoretical effort to compute multileg/NLO processes

That can be wasted by using unappropriate tools.

Note:

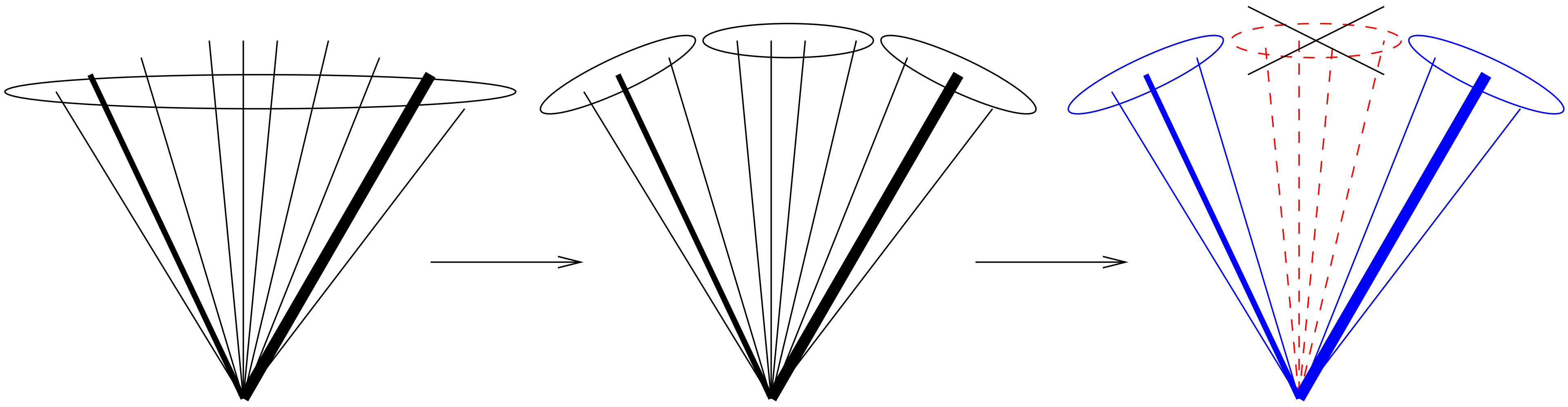
- arXiv:0903.0814: $W+2$ jets vs. LO QCD using CDF JetClu
- arXiv:0903.1748: $Z+2$ jets vs. NLO QCD using the D0runII cone
- arXiv:0903.1801: $Z+2$ jets vs. NLO QCD using the CMS iterative cone

Filtering using jet substructure

More refined clustering ("$3^{\rm rd}$ generation of algorithms")

Cambridge+Filtering algorithm:

- Cluster with Cambridge/Aachen and radius R
- For each jet, recluster it with Cambridge/Aachen and radius $R_{\rm sub}$ keep only $n_{\rm sub}$ hardest sub-jets of the initial jet



Filtering using jet substructure

More refined clustering ("$3^{\rm rd}$ generation of algorithms")

Cambridge+Filtering algorithm:

- Cluster with Cambridge/Aachen and radius R
- For each jet, recluster it with Cambridge/Aachen and radius $R_{\rm sub}$
  keep only $n_{\rm sub}$ hardest sub-jets of the initial jet

Aim: remove the soft background

Properties:

- Proven to improve jet reconstruction, in $H \to b\bar{b}$
  [J.Butterworth, A.Davison, M.Rubin, G.Salam, 08]
- Additional parameters that deserve appropriate studies
- We will use the simplest choice: $R_{\rm sub} = R/2$, $n_{\rm sub} = 2$

jet definitions = algorithm + parameters

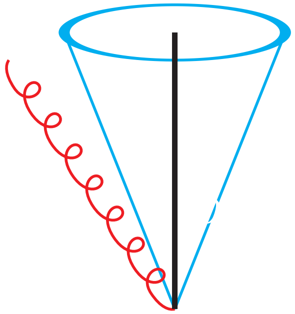
Algorithm: k_t***, Cam./Aachen, anti-***k_t***, SISCone, filtering?***
+ *parameters:* mainly R

Which one to choose?

Underlying idea

Competition between

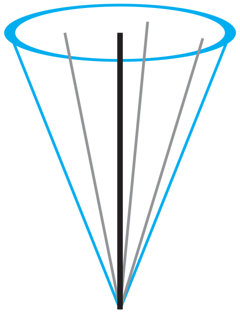
- catching perturbative radiation



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

Jet optimisation study

We analyse 3 processes typical of kinematic reconstructions:

- $Z' \to q\bar{q} \to$2 jets and $H \to gg \to$2 jets:

  simple environment: identify 2 jets and reconstruct $M_{Z',H}$
  source of monochromatic quark/gluon jets
  scale dependence: mass of the Z'/H varied between 100 GeV and 4 TeV
  ficticious narrow Z', H

- $t\bar{t} \to W^+bW^-\bar{b} \to q\bar{q}bq\bar{q}\bar{b} \to$6 jets:

  complex environment: identify 6 jets and reconstruct 2 top
  balance between reconstruction efficiency and identification

with

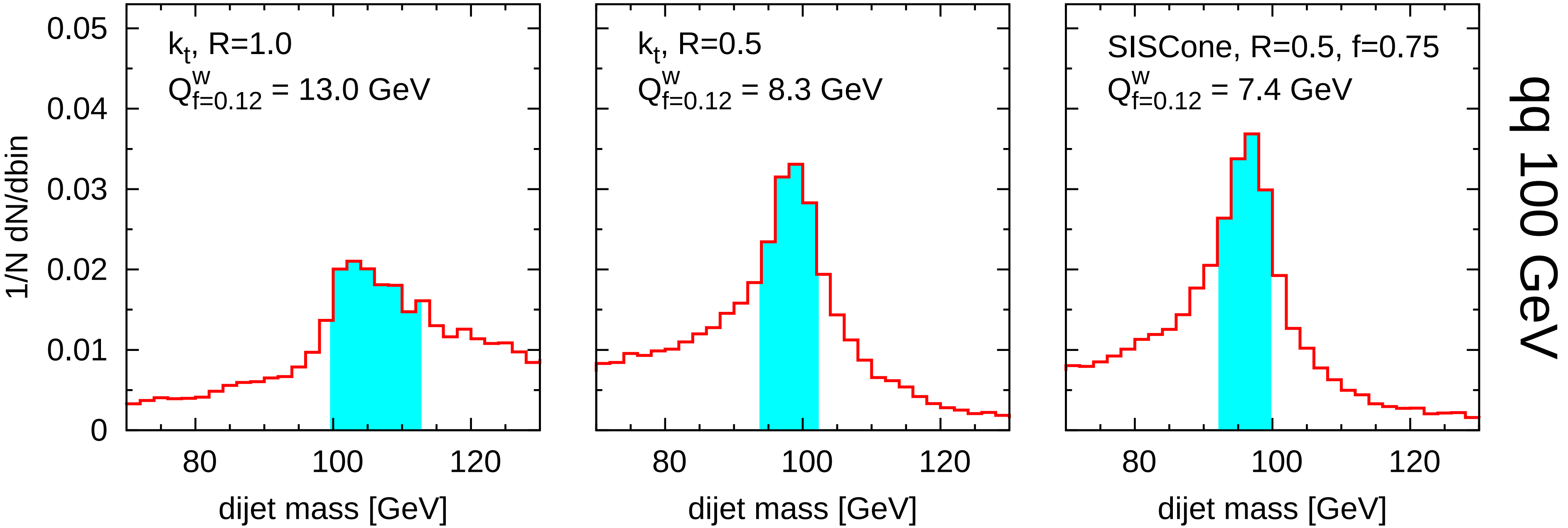
- the 5 IRC-safe algorithms: k_t, Cambridge, anti-k_t, SISCone, Cam+filtering
- jet radius varied between 0.1 and 1.5

Reconstruction quality (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

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



Reconstruction quality (2)

Assuming a constant background,

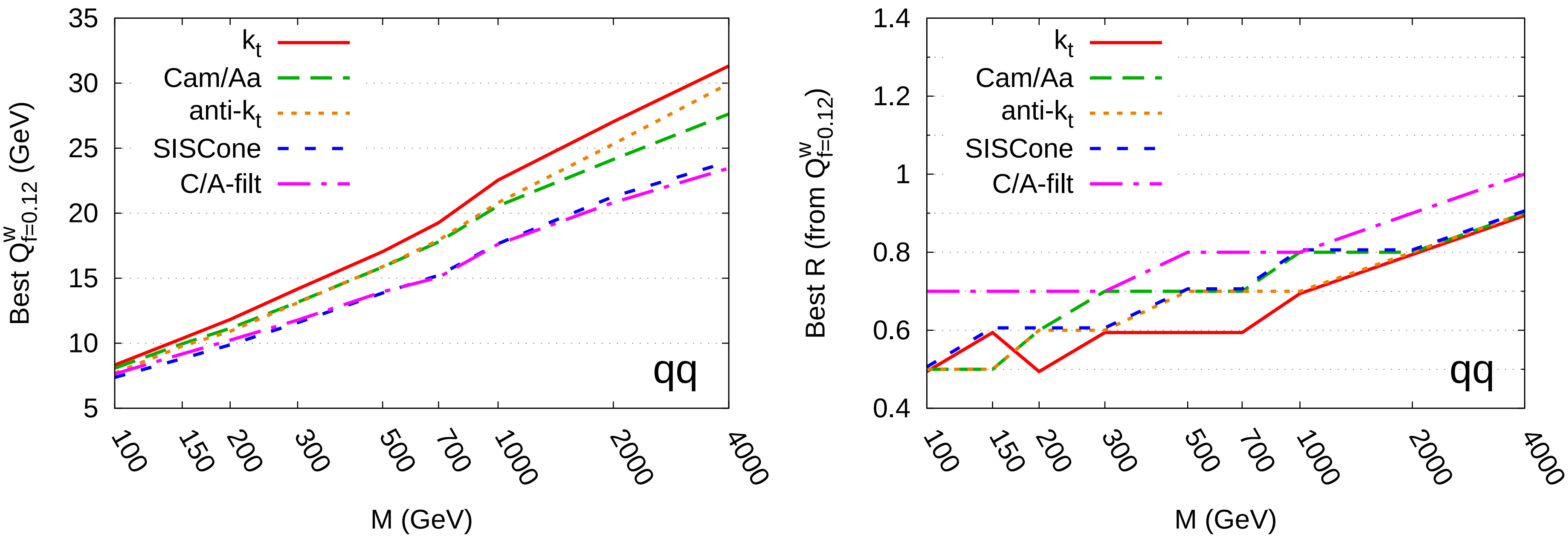
quality measure $\longrightarrow$ effective luminosity ratio

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

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

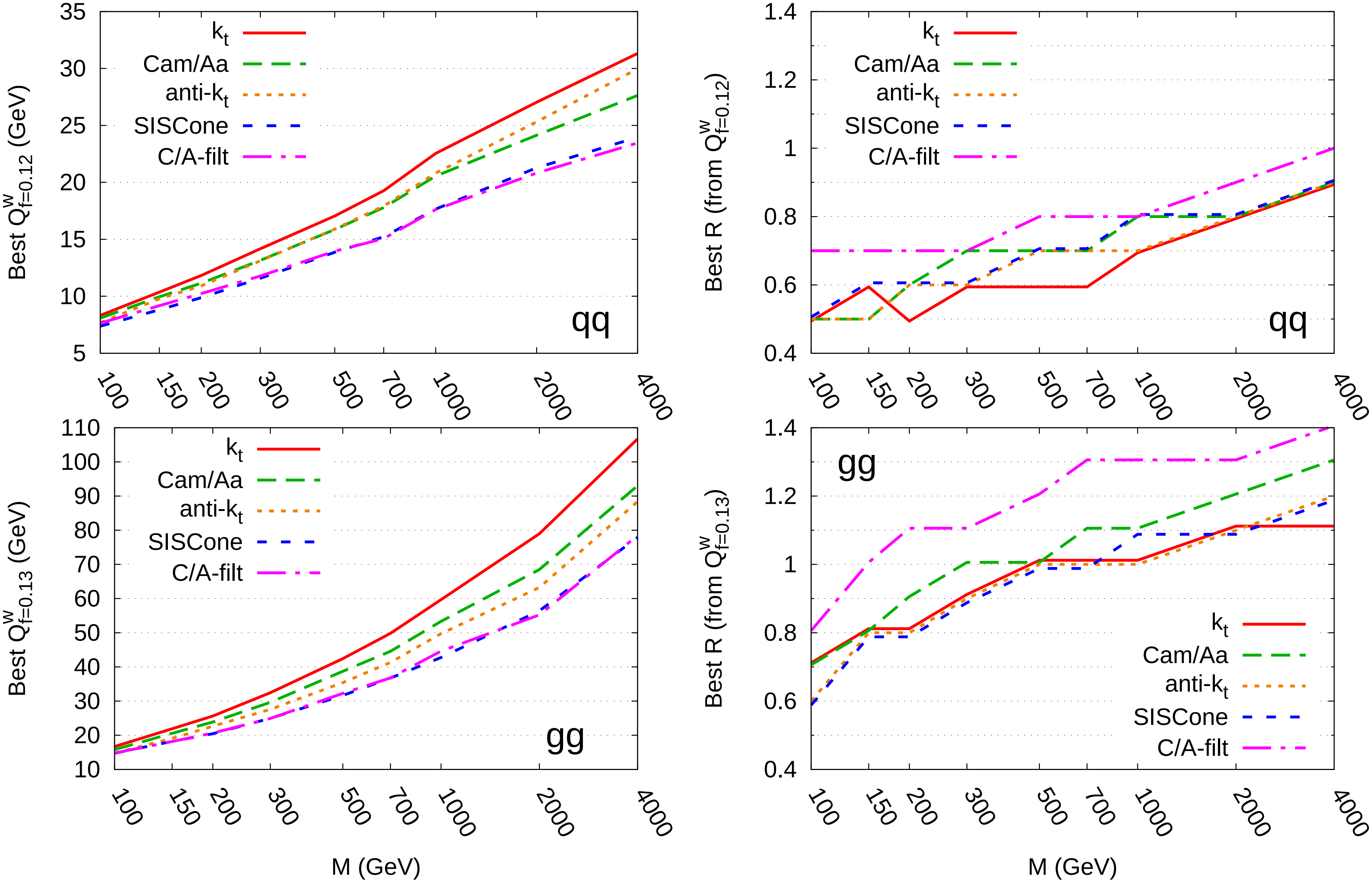
$\Leftrightarrow$ JD_2 requires 2 times the integrated luminosity of JD_1 to achieve the same discriminative power.

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



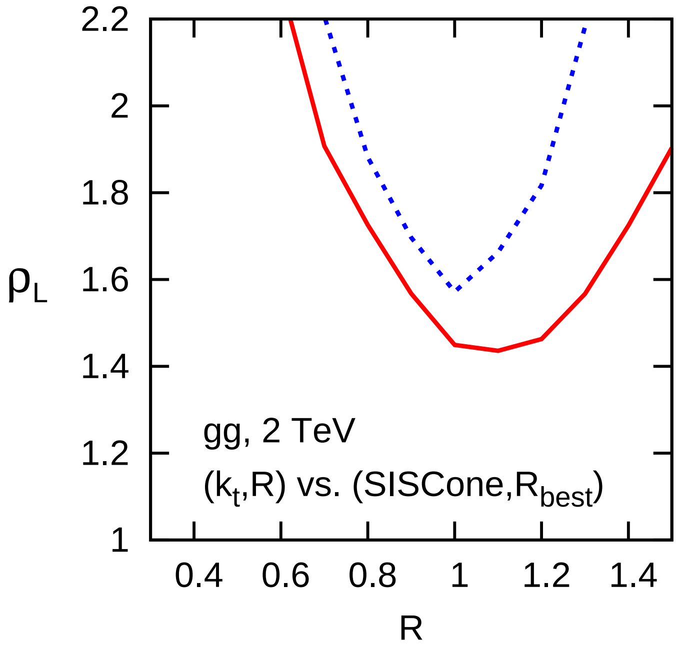
- SISCone and Cam+filtering perform better
- R_{best} strongly depends on the mass

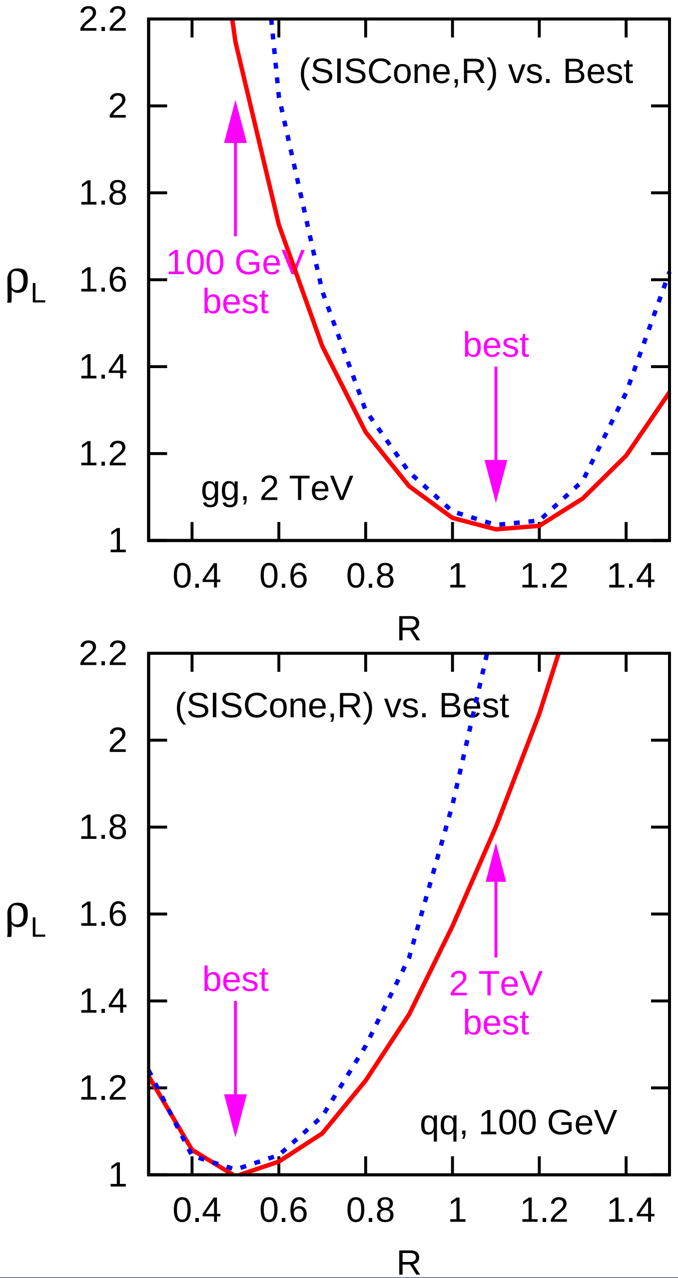
Quarks vs. gluons



Same conclusions for gluon jets with slightly larger R

Luminosity ratios





Mandatory at the LHC:
Not choosing the best alg.
AND R can be very costly
for new discoveries

Conclusions

- New jet algorithms available

  ~~JetClu~~, ~~ATLAS Cone~~, ~~MidPoint~~ → SISCone: ✓ as fast, ✓ IRC safe

  ~~Iterative Cone~~ → Anti-k_t: ✓ regular, ✓ IRC safe

  important for precision at the LHC!

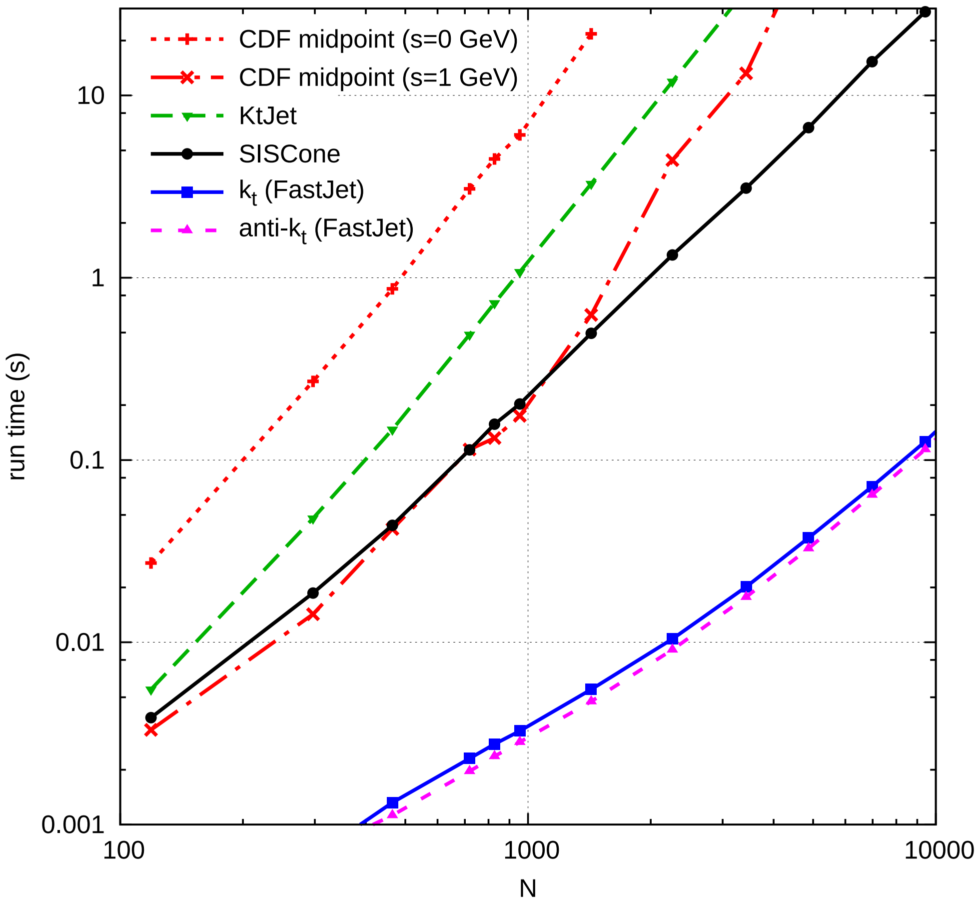
- Early discovery may rely on an optimal choice of a jet definition
  - SISCone and Camb/Aachen+filtering slightly preferred
  - Strong dependence on R: larger scales $\leftrightarrow$ larger R

Available from FastJet (or SpartyJet)

Check http://quality.fastjet.fr for interactive plots

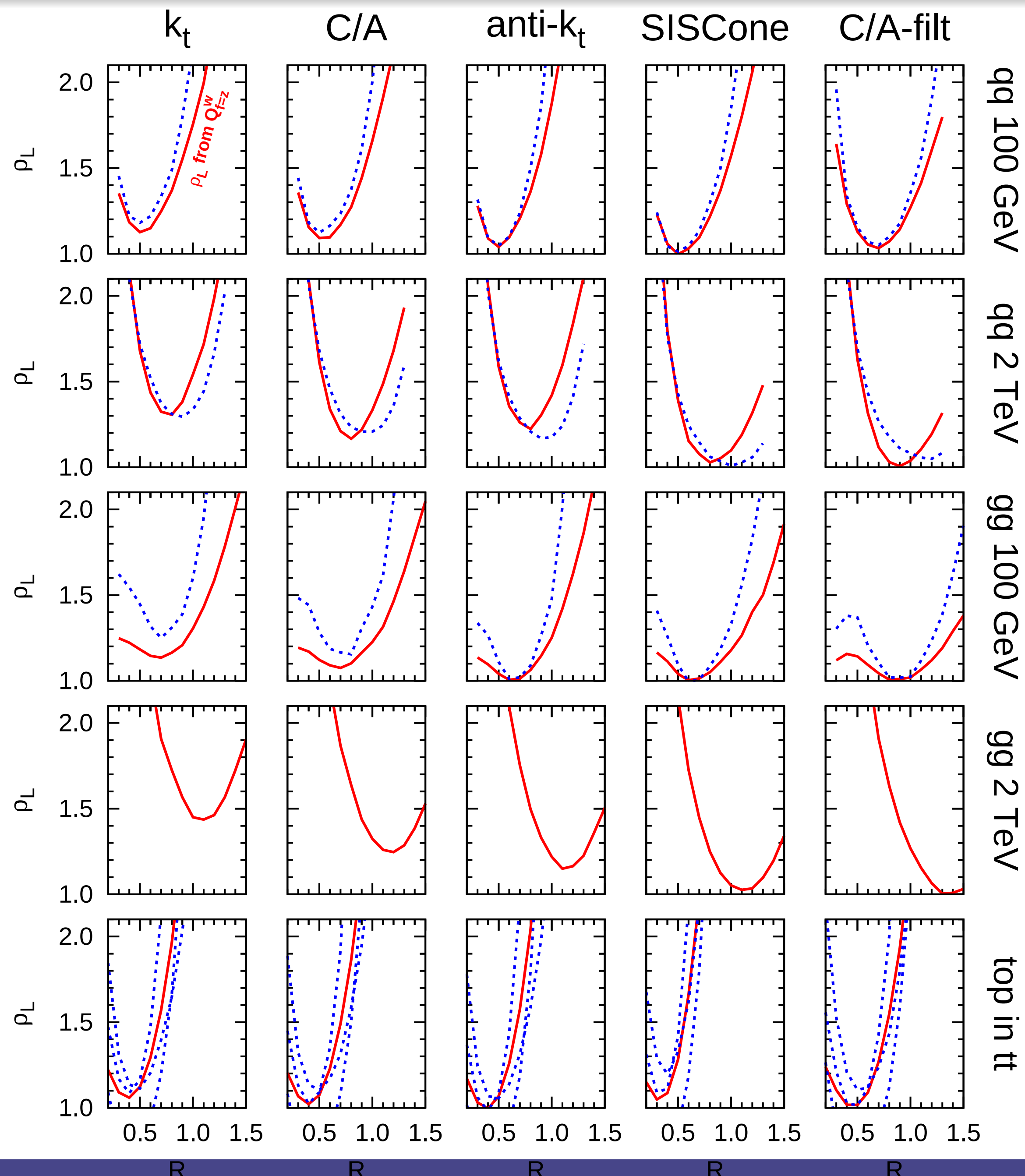
Backup slides

Speed



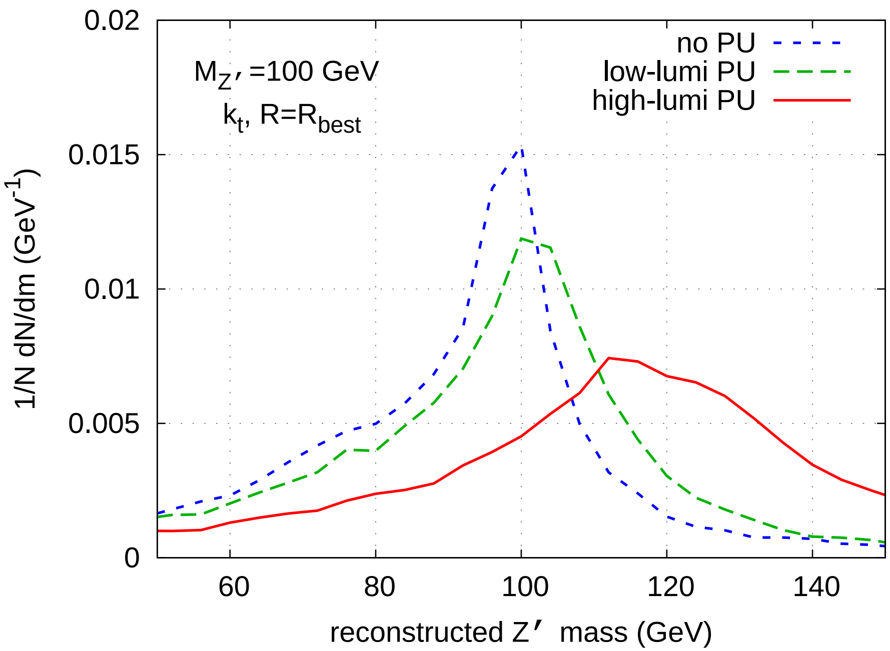
- Recombination algorithms very fast [M. Cacciari, G. Salam, 06]
- SISCone not slower than Midpoint (even with a 1 GeV seed threshold)

Optimisation - more results



Need for subtraction

Pileup $\approx$ uniform soft background that shifts jets to higher p_t



... that needs to be subtracted!

$\Rightarrow$ Using jet areas!

Pileup subtraction

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

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

Pileup subtraction

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)
  - tractable analytically in pQCD

    Example: area corrections from QCD radiation

$$\langle \mathcal{A}(p_{t,1}, R)\rangle = \mathcal{A}_{1\ \text{parton}}(R) + \frac{C_{F,A}}{\pi b_0} \log\left(\frac{\alpha_s(Q_0)}{\alpha_s(Rp_t)}\right) \pi R^2\, d$$

- area $\neq \pi R^2$
- area scaling violations

d	passive	active
k_t	0.5638	0.519
Cam	0.07918	0.0865
SISCone	-0.06378	0.1246
anti-k_t	0	0

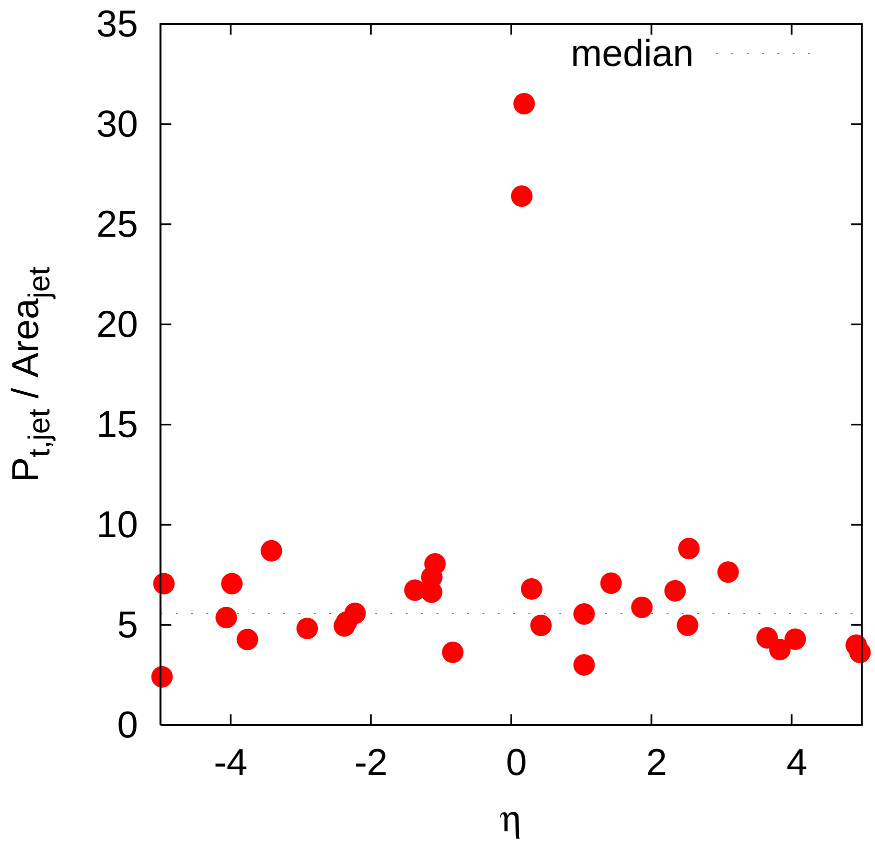
Pileup subtraction

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

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

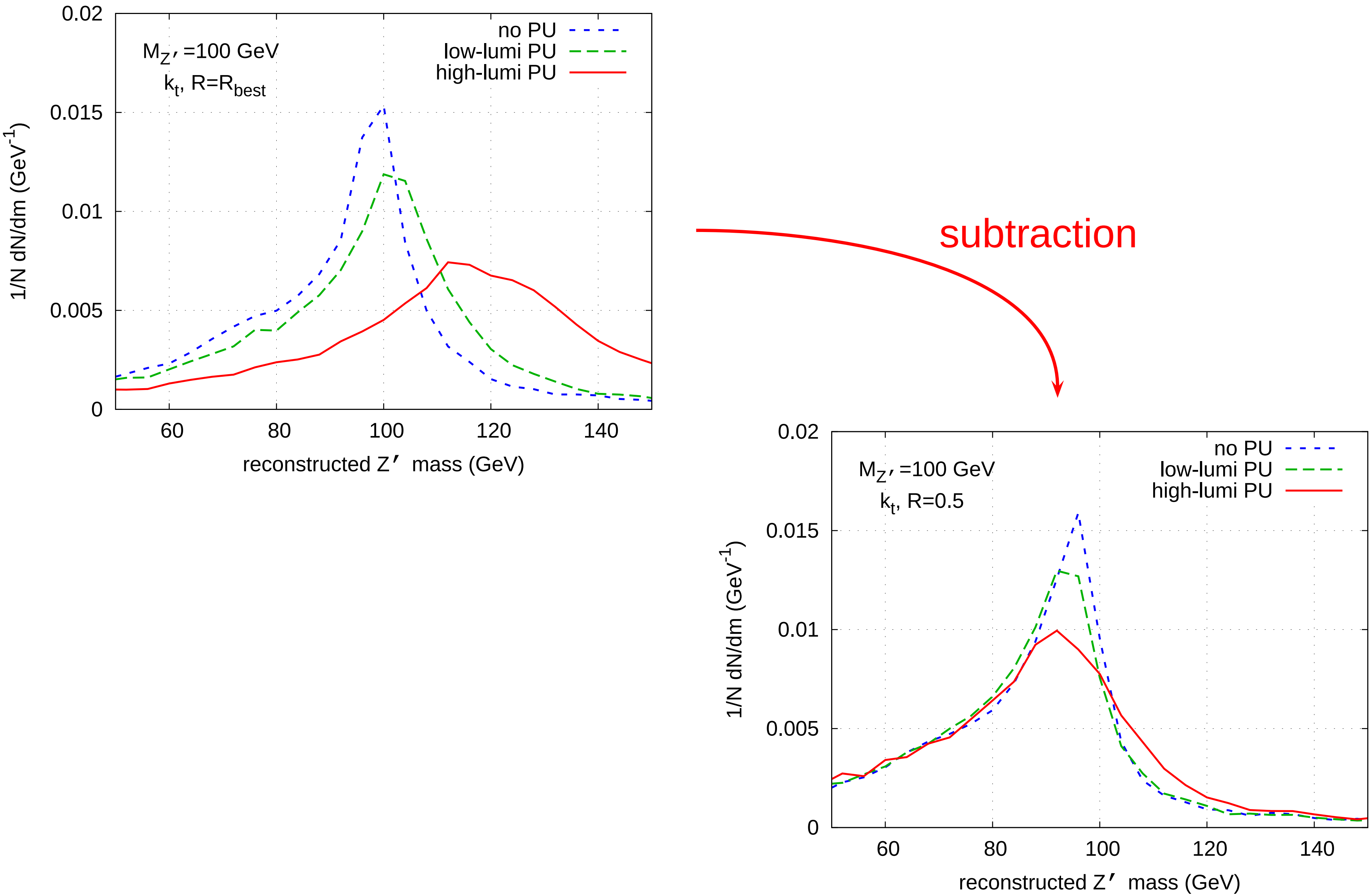
- Jet area: [M.Cacciari, G.Salam, G.S., 08]
  - region where the jet catches infinitely soft particles (active/passive)
  - tractable analytically in pQCD
- Pileup density per unit area: ρ_{pileup}

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



implemented in FastJet
on an event-by-event basis

Subtraction at work



Optimisation - pileup