Progress in defining jets for the LHC

Grégory Soyez

Brookhaven National Laboratory

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

arXiv:0704:0292, arXiv:0802:1188, arXiv:0802:1189, arXiv:0810.1304

Plan

- Foreword: why jets? what are they?

  introducing the basic terminology/concepts

- Part 1: building solid jet definitions

  new algorithms to meet the fundamental requirements

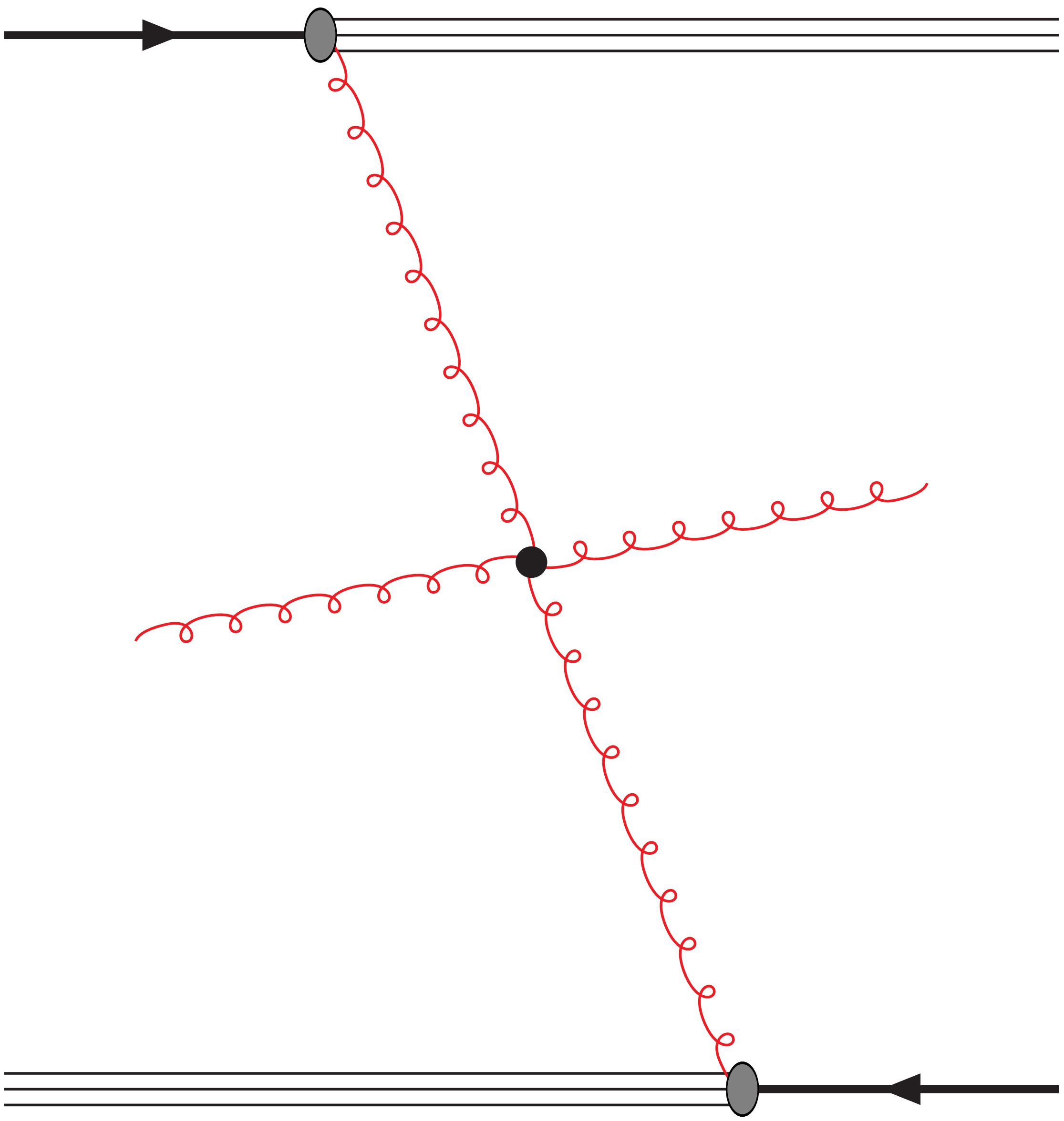
- Part 2: optimizing jets in pp collisions

  which jet algorithm is best suited?

  - how to quantify the reconstruction efficiency
  - Results without pileup
  - Results with pileup (subtraction)

Foreword: why jets? what are they?

General (over)simplified picture

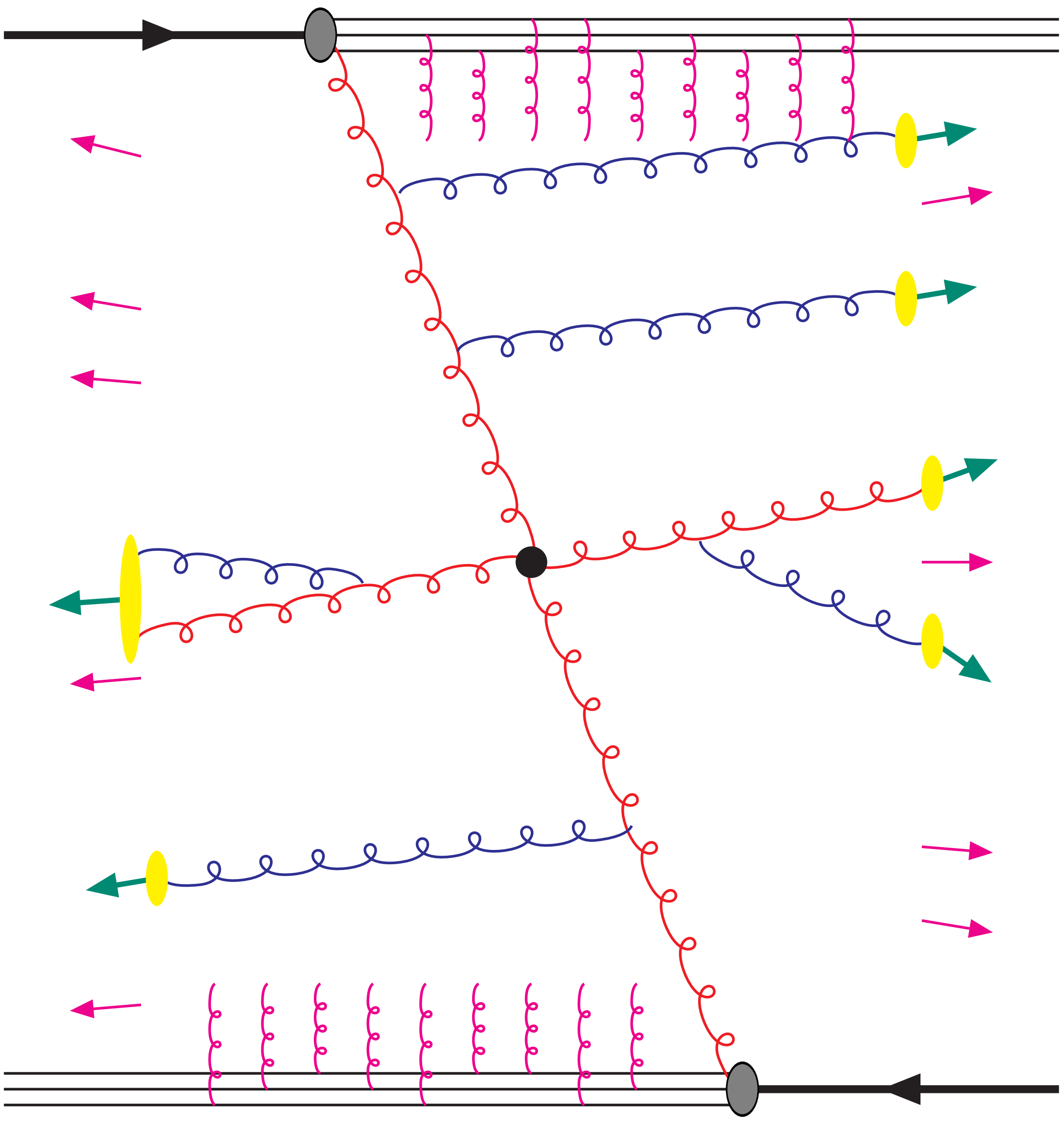


Hard scattering ($2 \to n$)

computed exactly at $\mathcal{O}(\alpha_s^p)$

$gg \to gg$, $gg \to ggg$,
$gg \to gggg$,
$gg \to H \to b\bar{b}$,
$gg \to t\bar{t} \to \mu\nu_\mu b\bar{b}q\bar{q}$,
$gg \to Z' \to q\bar{q}$, ...

General (over)simplified picture



Hard scattering $(2 \to n)$

Parton level

$\approx$ resummed collinear div.

$\sum_i \alpha_s^i \log^i(p_t^2/\mu^2)$

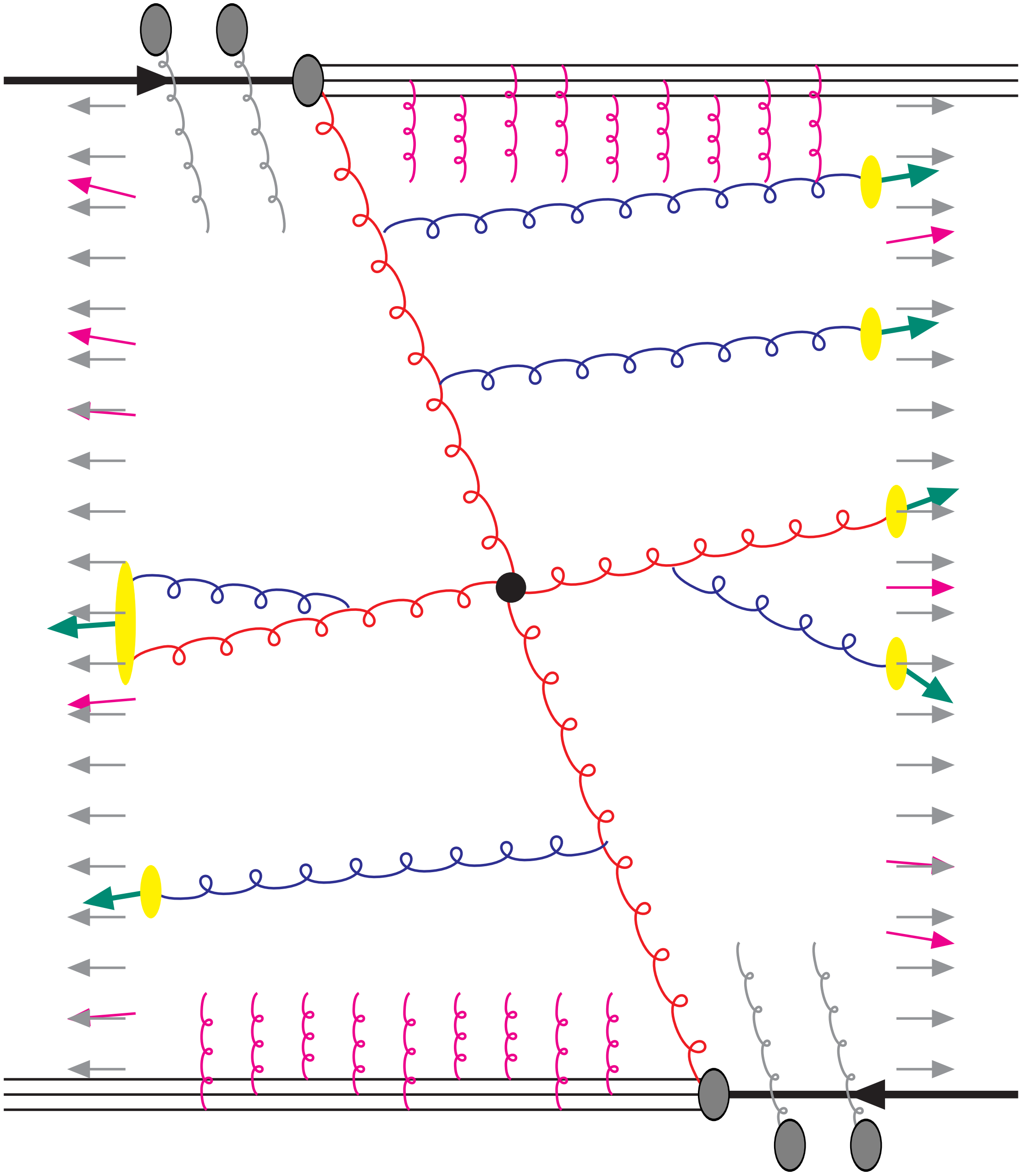
Hadron level: hadronisation

Underlying event

beam remnants interactions

$\Rightarrow$ soft background

General (over)simplified picture



Hard scattering ($2 \to n$)

Parton level

$\approx$ resummed collinear div.

$\sum_i \alpha_s^i \log^i(p_t^2/\mu^2)$

Hadron level: hadronisation

Underlying event

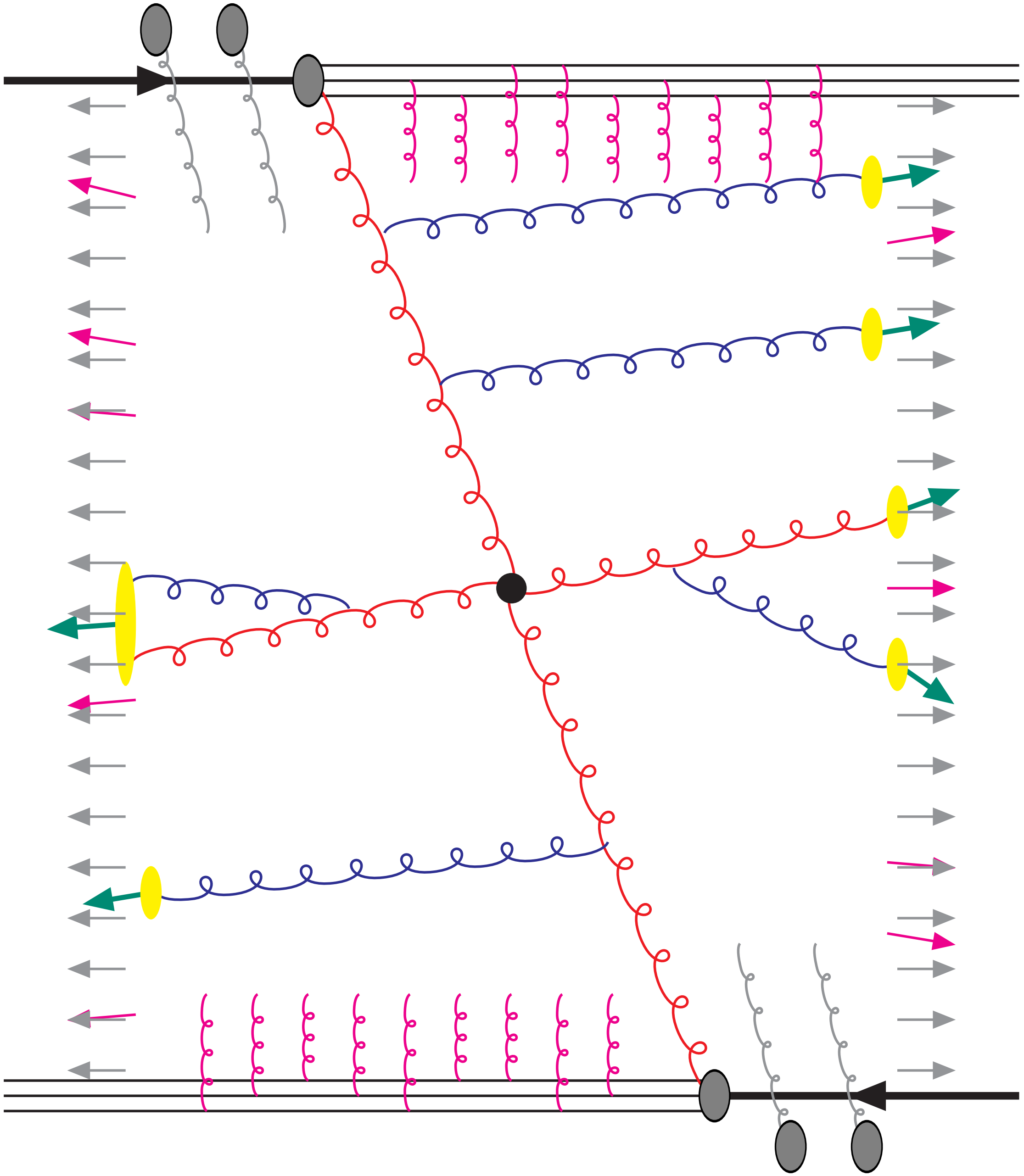
beam remnants interactions

$\Rightarrow$ soft background

Pileup

$\approx$ uniform soft background

General (over)simplified picture



Hard scattering ($2 \to n$)

Parton level

$\approx$ resummed collinear div.

$\sum_i \alpha_s^i \log^i(p_t^2/\mu^2)$

Hadron level: hadronisation

Underlying event

beam remnants interactions

$\Rightarrow$ soft background

Pileup

$\approx$ uniform soft background

"Jets" $\equiv$ hard partons

Parton ambiguous

$\Rightarrow$ multiple jet definitions

Two classes of algorithms

Class 1: recombination	Cass 2: cone
Successive recombinations of the "closest"$^{(a)}$ pair of particle	find directions of energy flow $\equiv$ stable cones$^{(b)}$
Nice perturbative behaviour	Small sensitivity to soft radiation (UE,PU)
Often used in $e^\pm e^\pm$, $e^\pm p$	Often used in pp

$^{(a)}$ Distance: (stop when $d_{\min} > R$)

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

$$\text{Aachen/Cam.:} \quad d_{i,j} = \Delta\phi_{i,j}^2 + \Delta y_{i,j}^2$$

$^{(b)}$ **stable cones** (radius R) such that:
the total momentum of its contents points in the direction of its centre

How the cone works...

- Seeded (iterative) approaches: iterate from an initial position until stable
  - seed = initial particle
  - seed = midpoint between stable cones found at first step
- One has to deal with overlapping stable cones: 2 subclasses

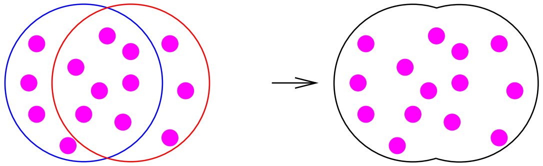
---

How the cone works...

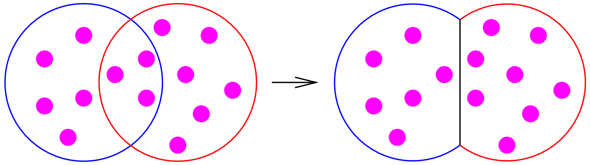
- Seeded (iterative) approaches: iterate from an initial position until stable
  - seed = initial particle
  - seed = midpoint between stable cones found at first step

Class 2(a): cone with split-merge (ex.: JetClu, Atlas, MidPoint):

$\tilde{p}_{t,\text{shared}} > f\tilde{p}_{t,\text{min}}$



$\tilde{p}_{t,\text{shared}} \leq f\tilde{p}_{t,\text{min}}$

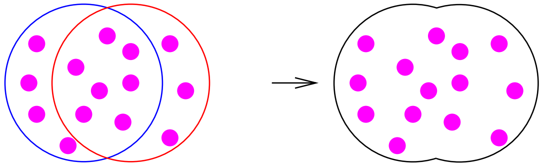


How the cone works...

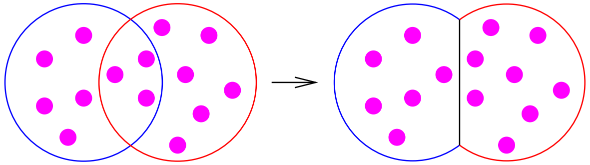
- Seeded (iterative) approaches: iterate from an initial position until stable
  - seed = initial particle
  - seed = midpoint between stable cones found at first step

Class 2(a): cone with split-merge (ex.: JetClu, Atlas, MidPoint):

$\tilde{p}_{t,\text{shared}} > f\tilde{p}_{t,\text{min}}$



$\tilde{p}_{t,\text{shared}} \leq f\tilde{p}_{t,\text{min}}$



Class 2(b): cone with progressive removal (ex.: Iterative Cone)

- iterate from the hardest seed
- remove the stable cone as a jet and start again

Idea: “regular/circular” jets

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

Part 1
21st century: towards a solid toolkit

SNOWMASS accords, Tevatron 1990 (i.e. old!):

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

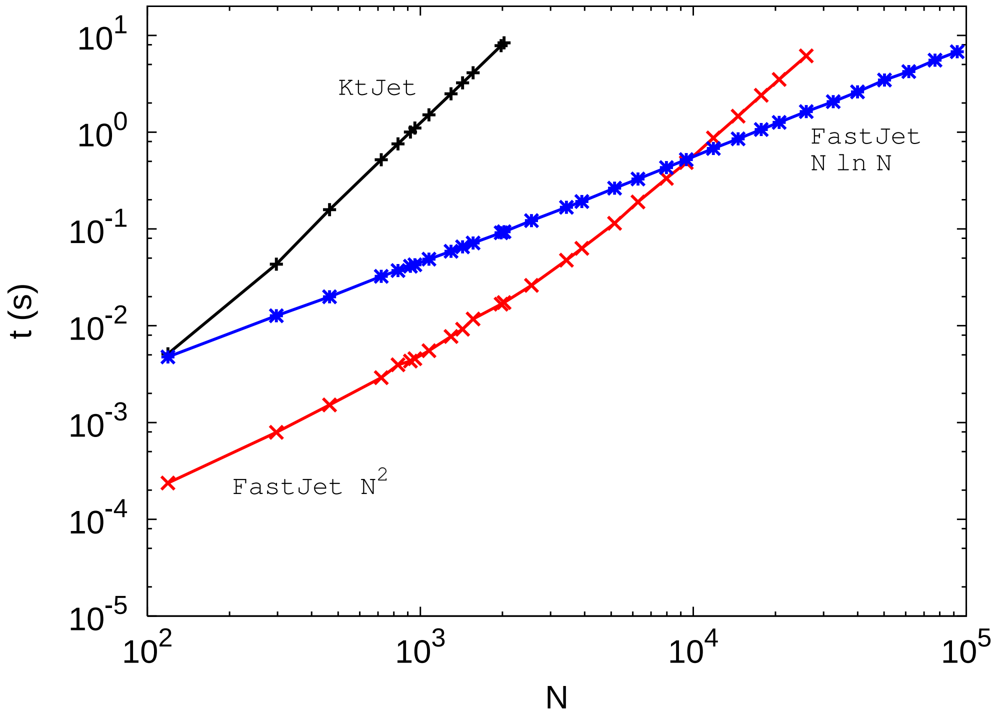
1. Simple to implement in an experimental analysis;
2. Simple to implement in the theoretical calculation;
3. Defined at any order of perturbation theory;
4. Yields finite cross section at any order of perturbation theory;
5. Yields a cross section that is relatively insensitive to hadronization.

i.e. usable by theoreticians (*e.g.* finite perturbative results)
and experimentalists (*e.g.* fast enough, not much UE sensitivity)

Speed improvement

[M. Cacciari, G. Salam, 06]

- Speeding up the k_t and Cam/Aachen algorithms
  using computational-geometry techniques: $\mathcal{O}\left(N^3\right) \to \mathcal{O}\left(N \log N\right)$



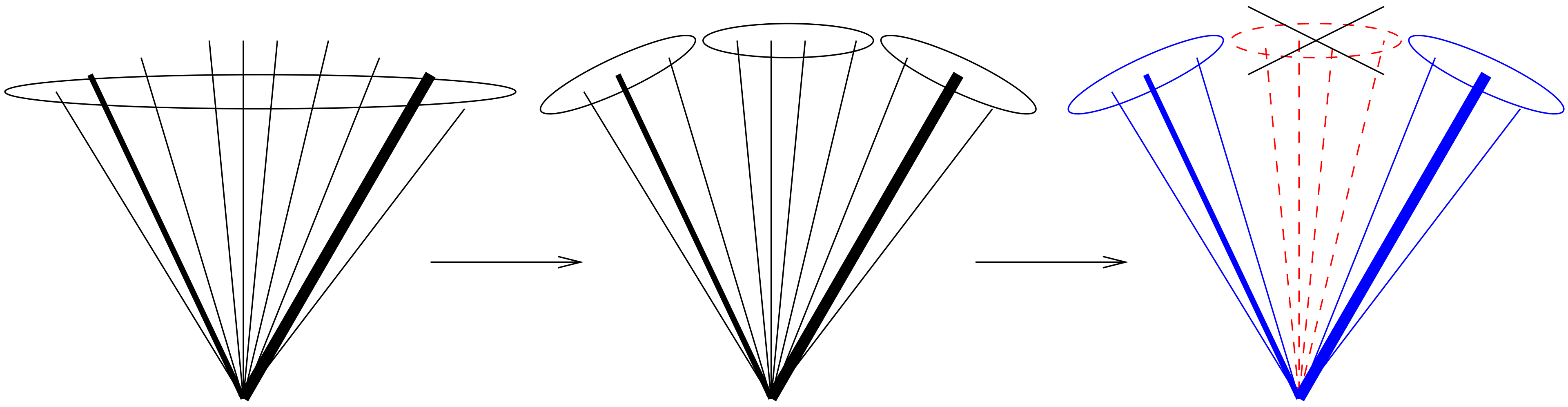
- C++ implementation in `FastJet`
  `http://www.fastjet.fr` (M. Cacciari, G. Salam, G.S.)

Filtering using jet substructure

More refined clustering ("2^{nd} generation of algorithms")

Cambridge+Filtering algorithm:

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



Filtering using jet substructure

More refined clustering ("2^{nd} generation of algorithms")

Cambridge+Filtering algorithm:

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

Aim: remove the soft background

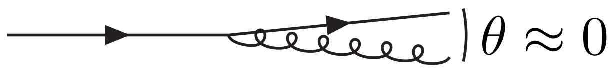
Properties:

- 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: $R_{\text{sub}} = R/2$, $n_{\text{sub}} = 2$

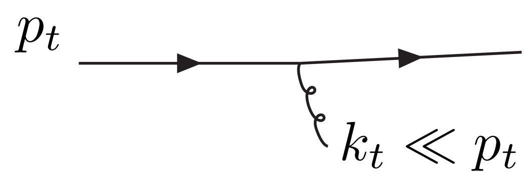
QCD probability for gluon bremsstrahlung at angle θ and $\perp$-mom. k_t:

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

Two divergences:



Collinear

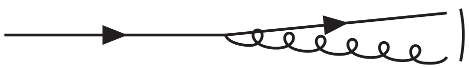


Soft

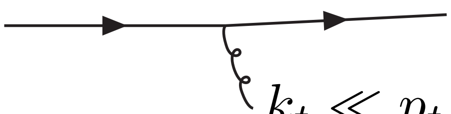
QCD probability for gluon bremsstrahlung at angle θ and $\perp$-mom. k_t:

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

Two divergences:



Collinear

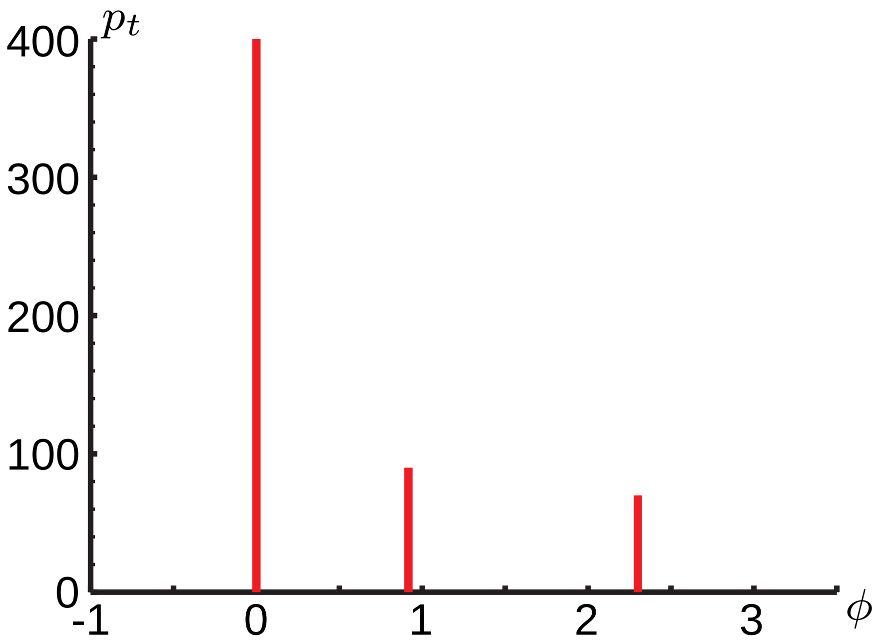


Soft

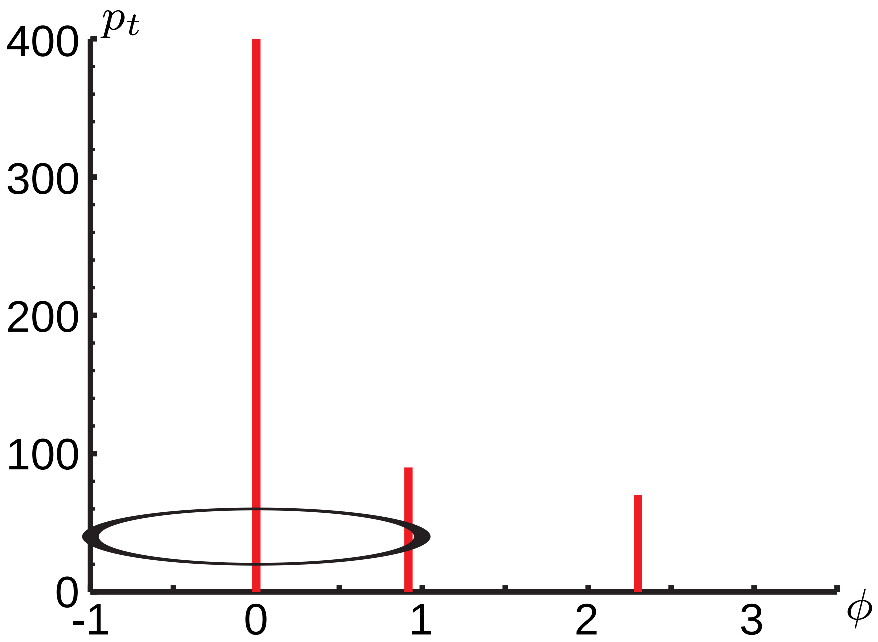
For pQCD to make sense, the (hard) jets should not change when

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

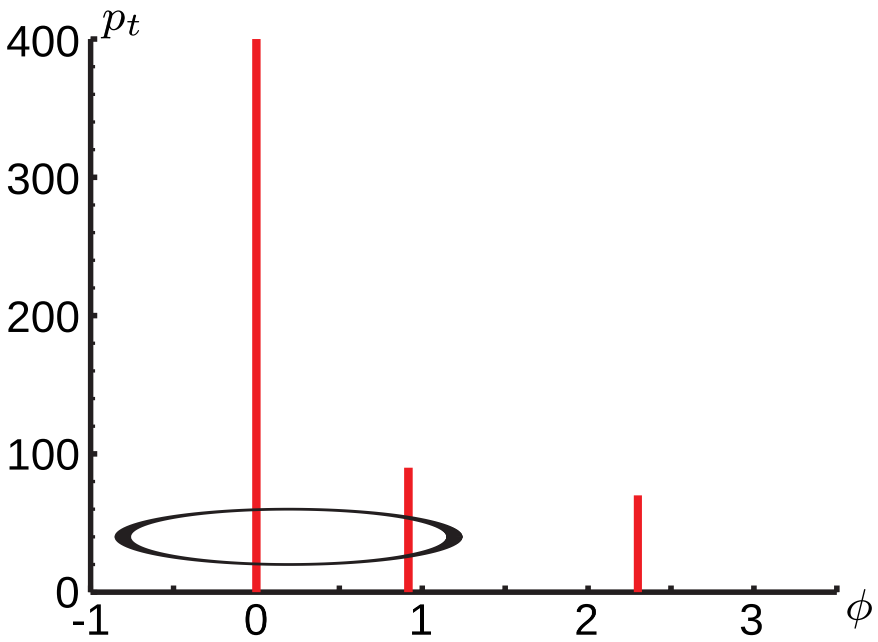
IR unsafety of the Midpoint alg



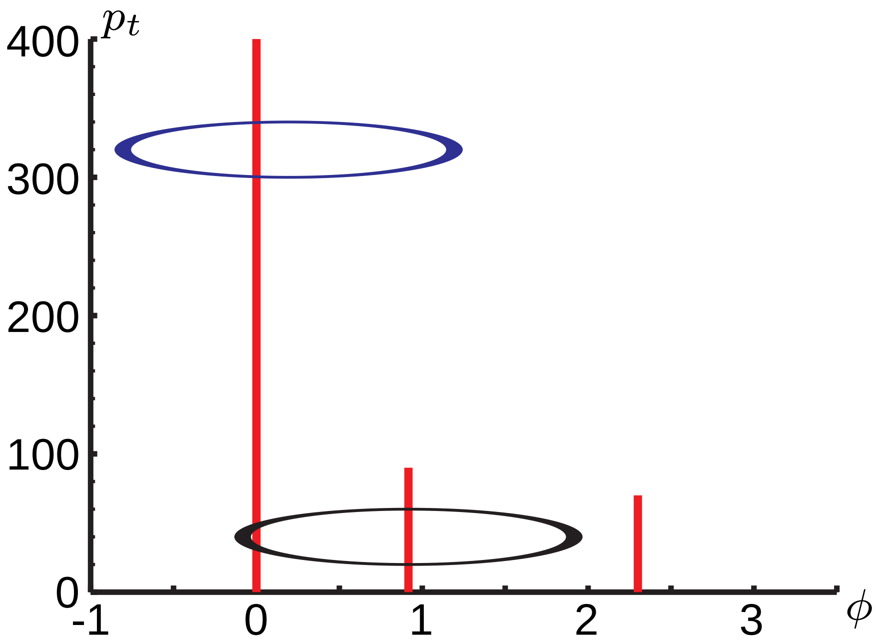
IR unsafety of the Midpoint alg



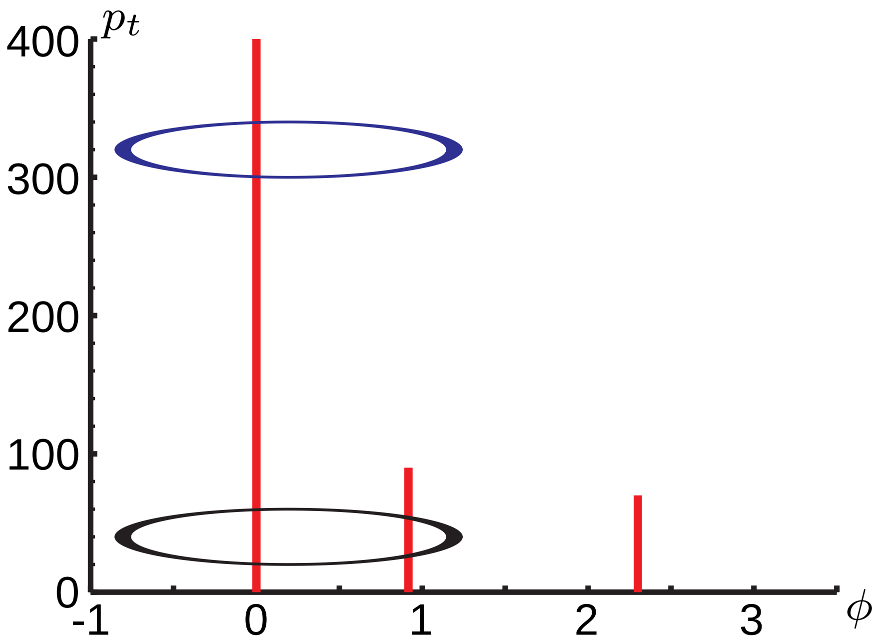
IR unsafety of the Midpoint alg



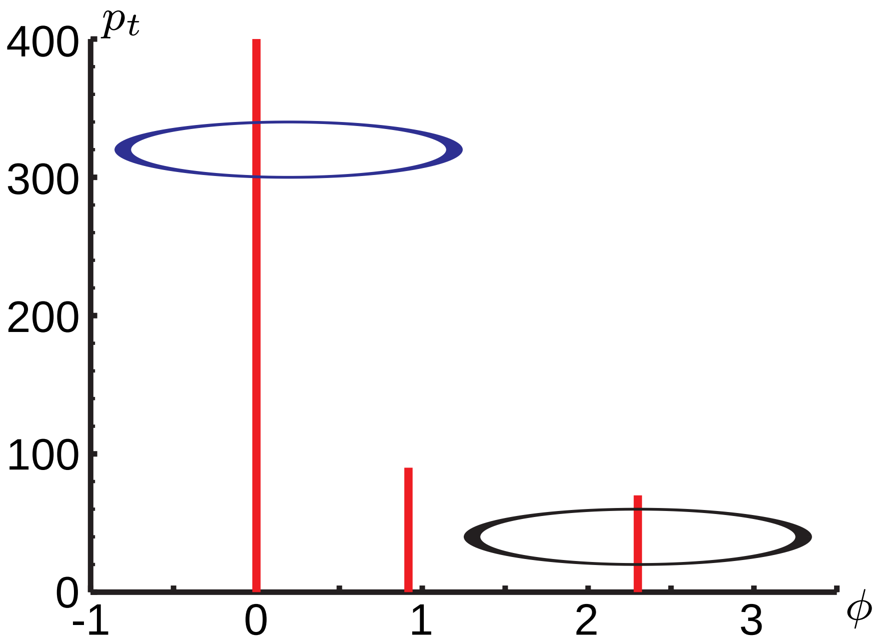
IR unsafety of the Midpoint alg



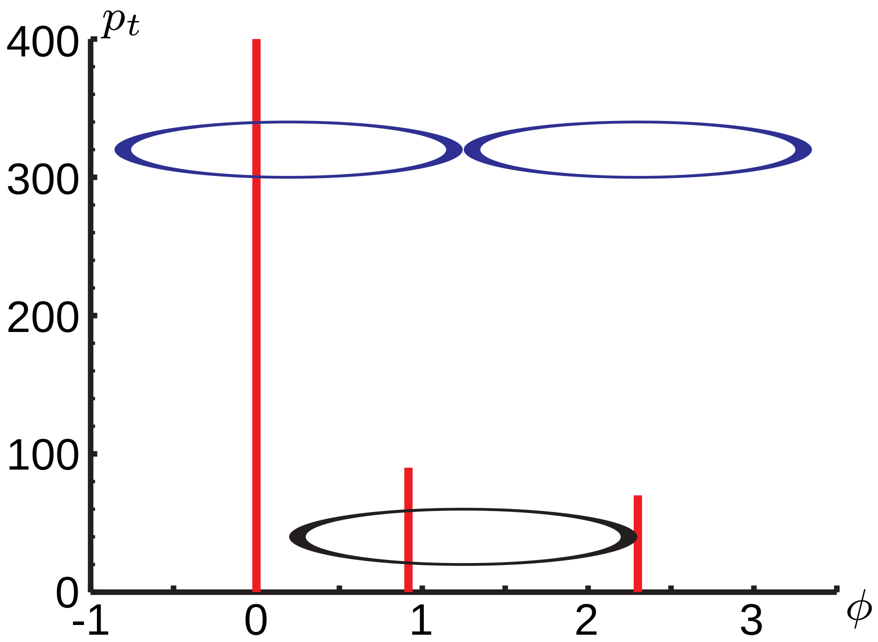
IR unsafety of the Midpoint alg



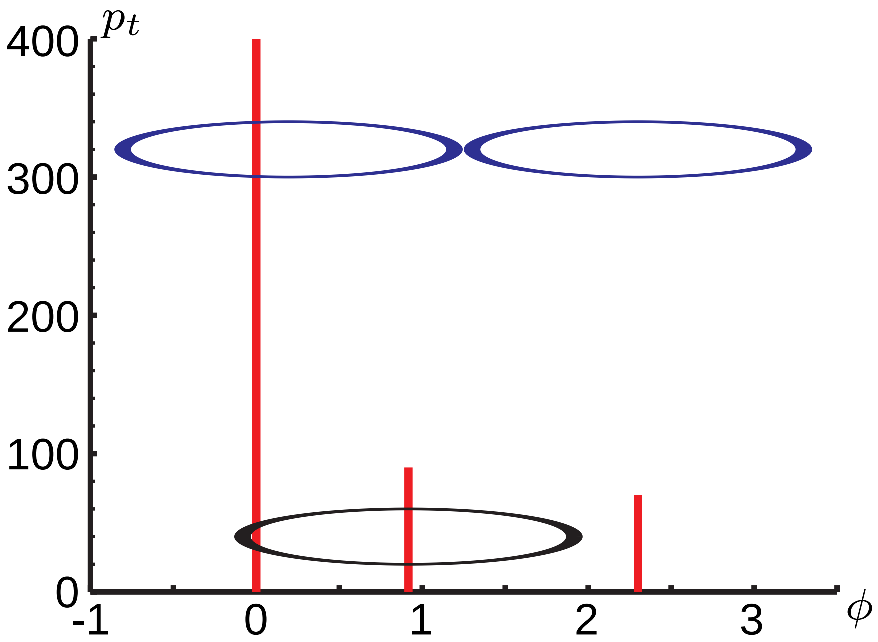
IR unsafety of the Midpoint alg



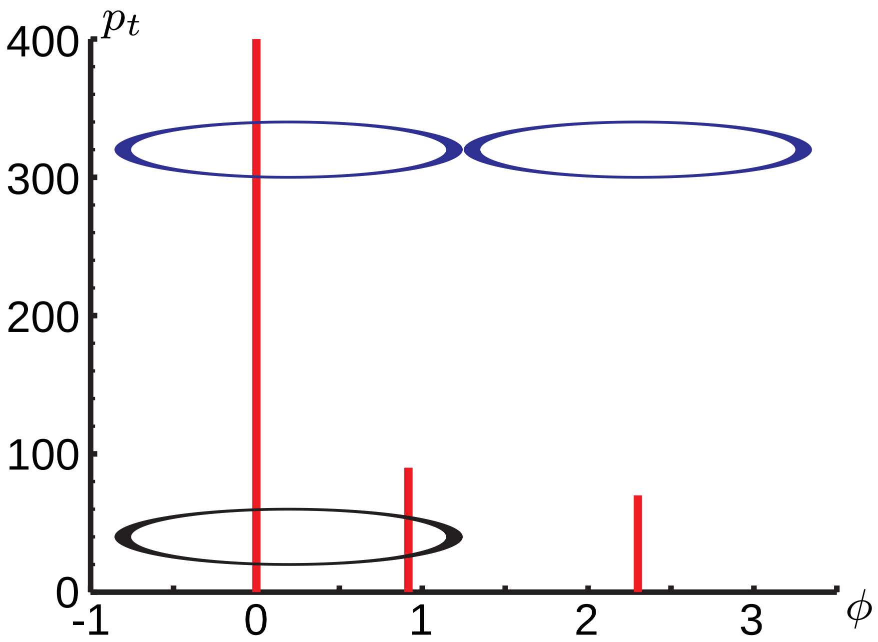
IR unsafety of the Midpoint alg



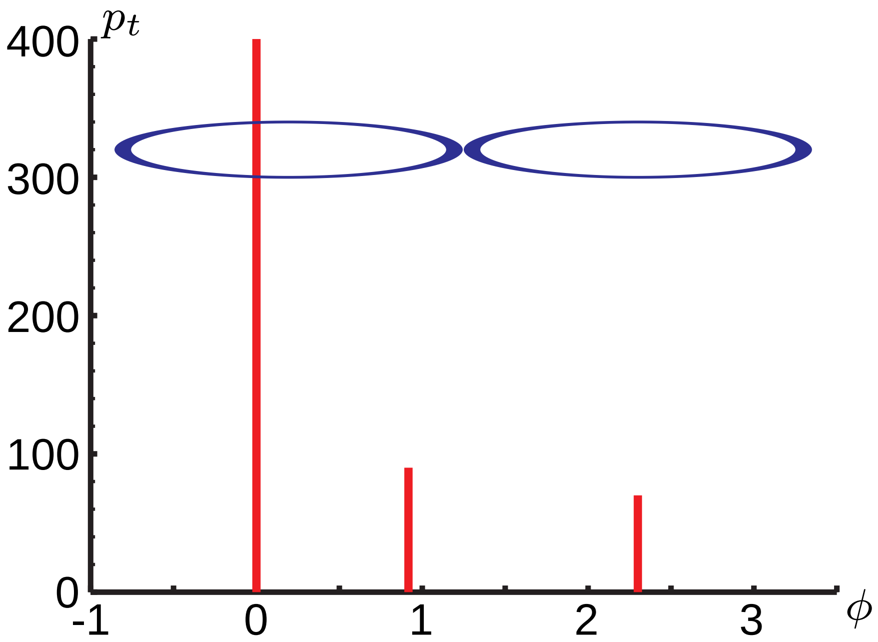
IR unsafety of the Midpoint alg

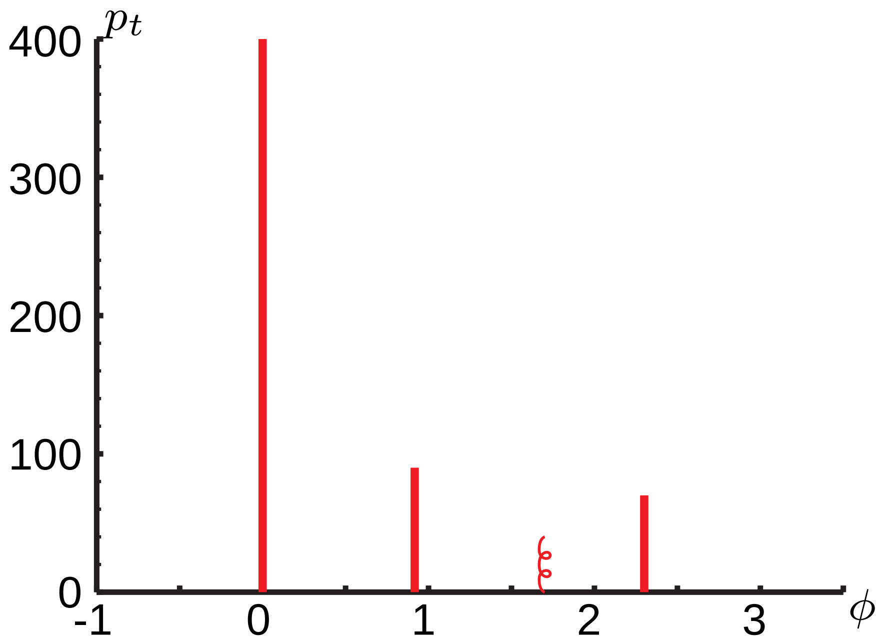


IR unsafety of the Midpoint alg

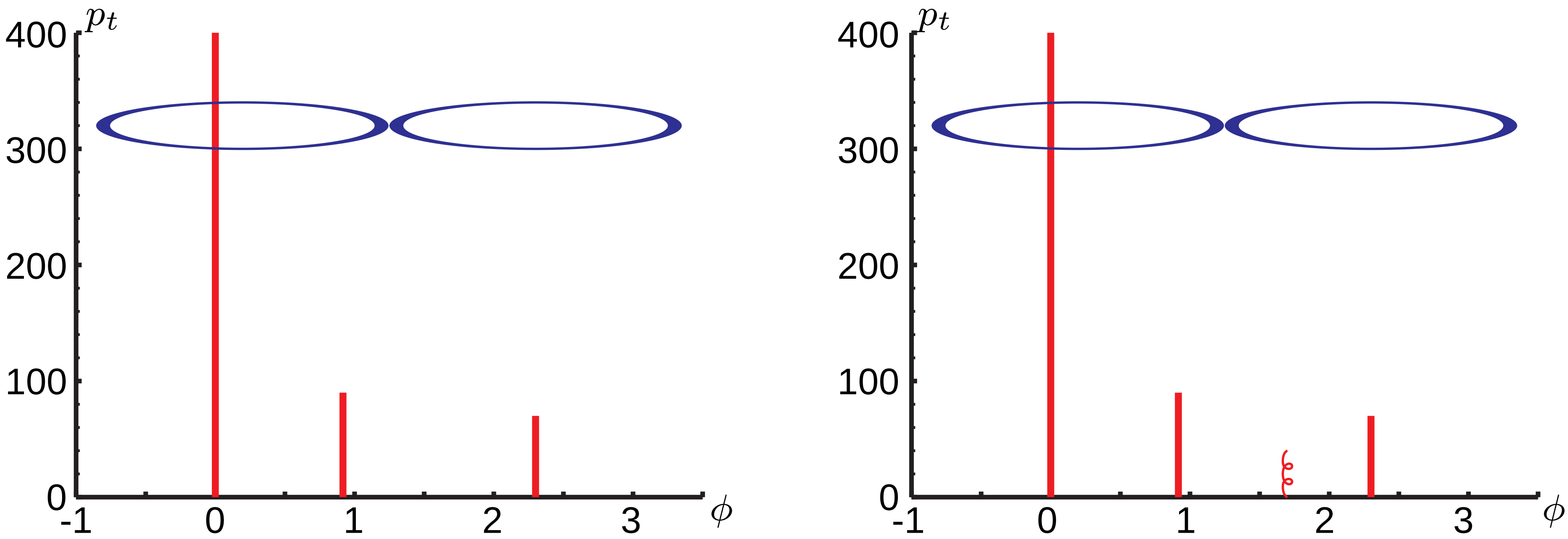


IR unsafety of the Midpoint alg

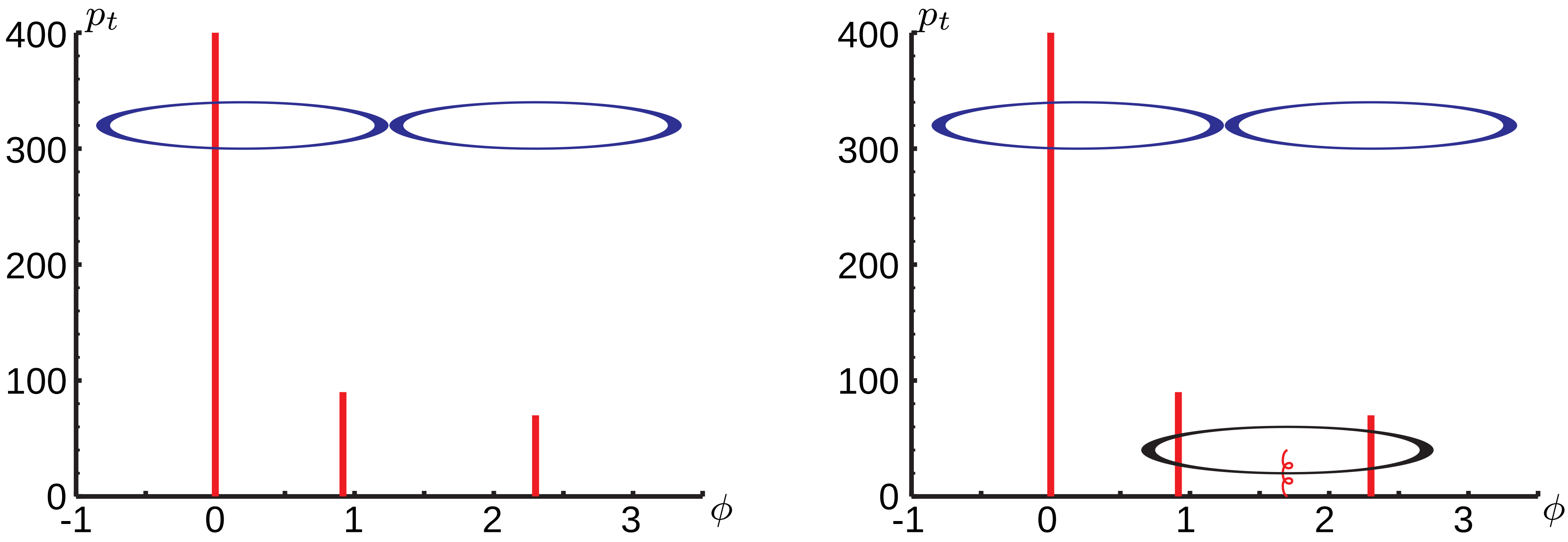




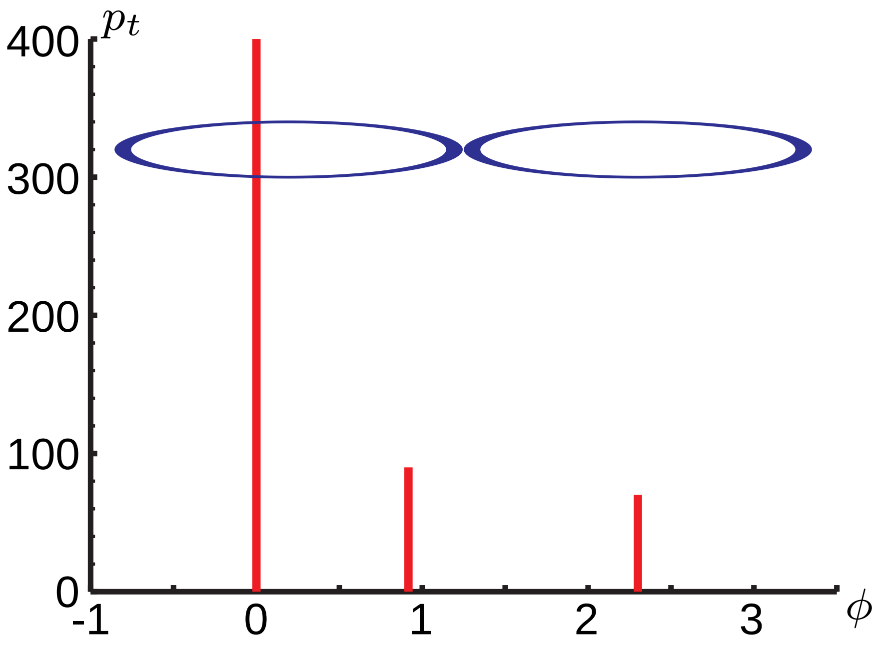
IR unsafety of the Midpoint alg

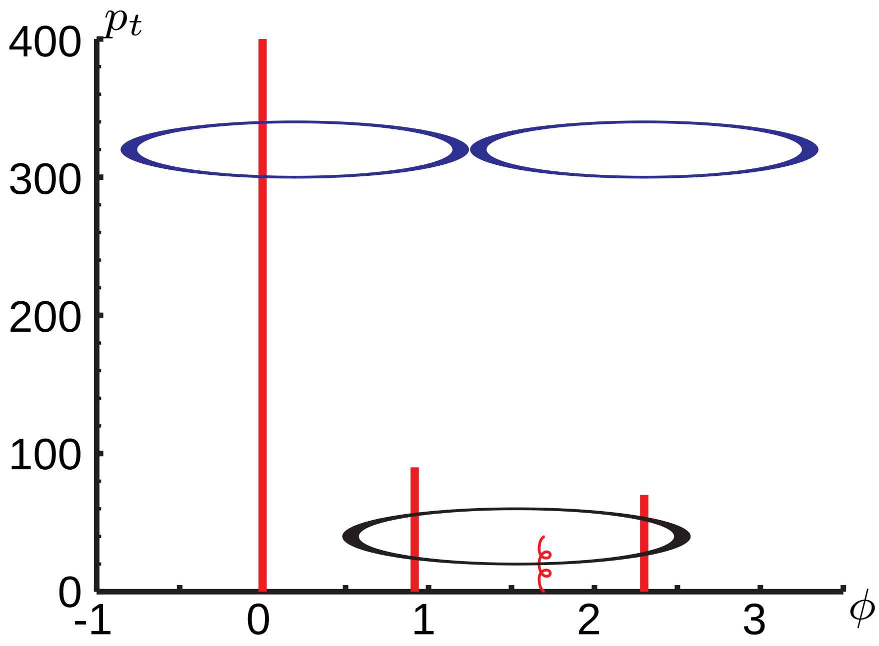


IR unsafety of the Midpoint alg

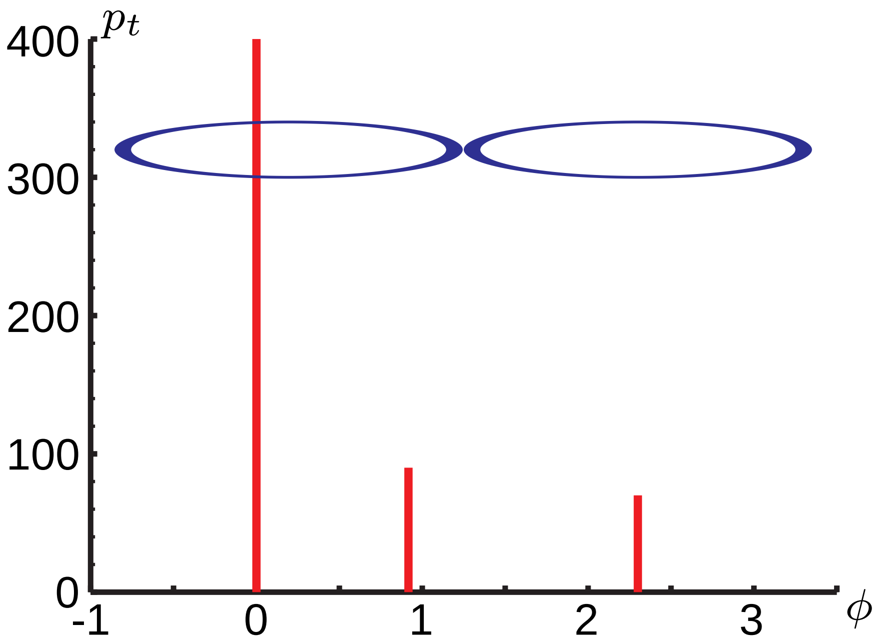


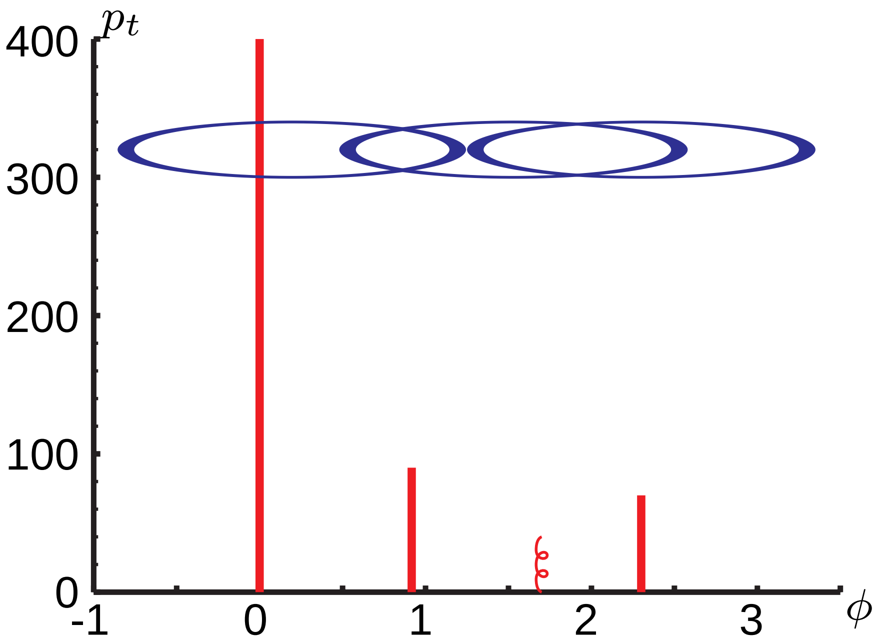
IR unsafety of the Midpoint alg





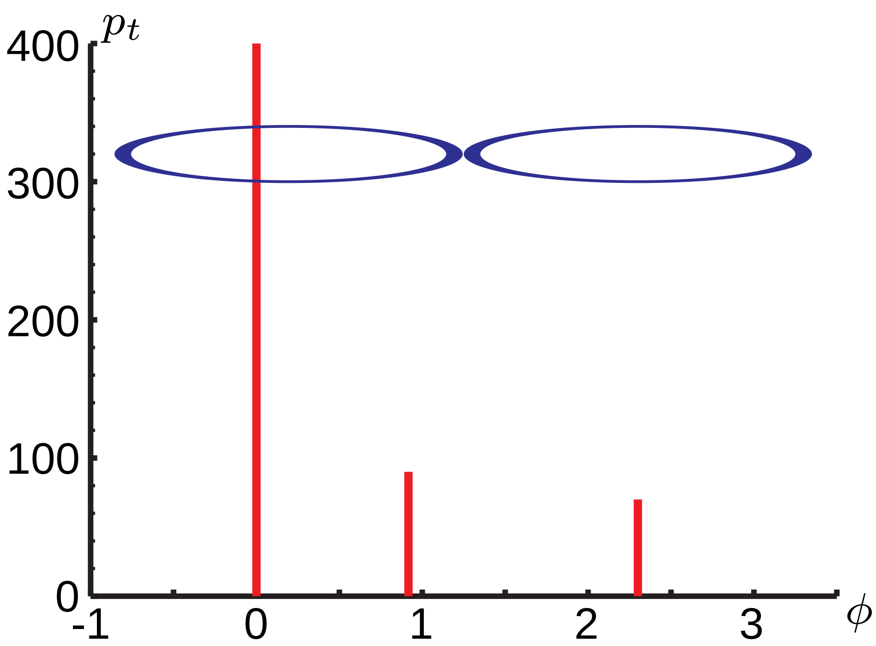
IR unsafety of the Midpoint alg

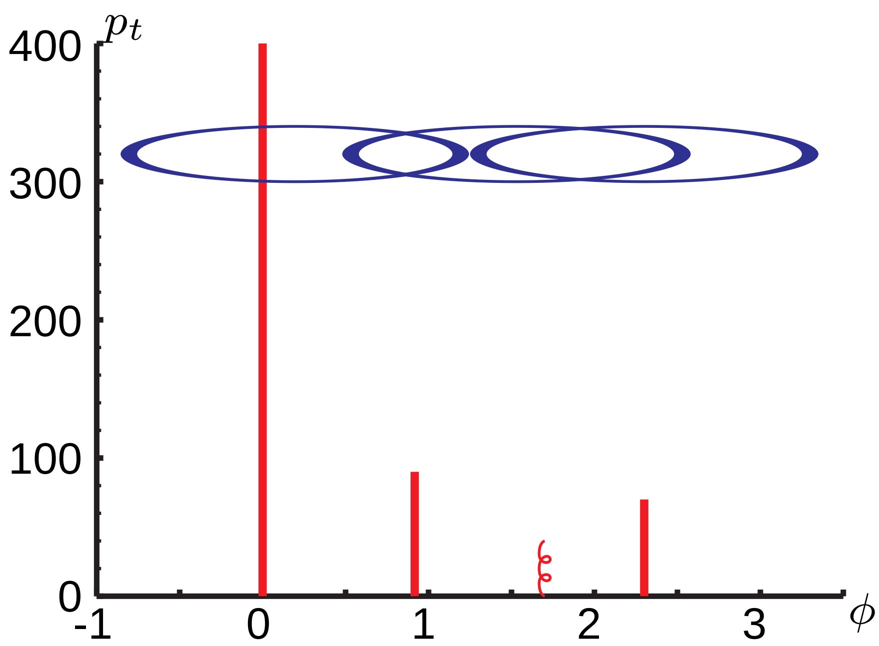




Stable cones:

Midpoint: {1,2} & {3} | {1,2} & {3} & {2,3}





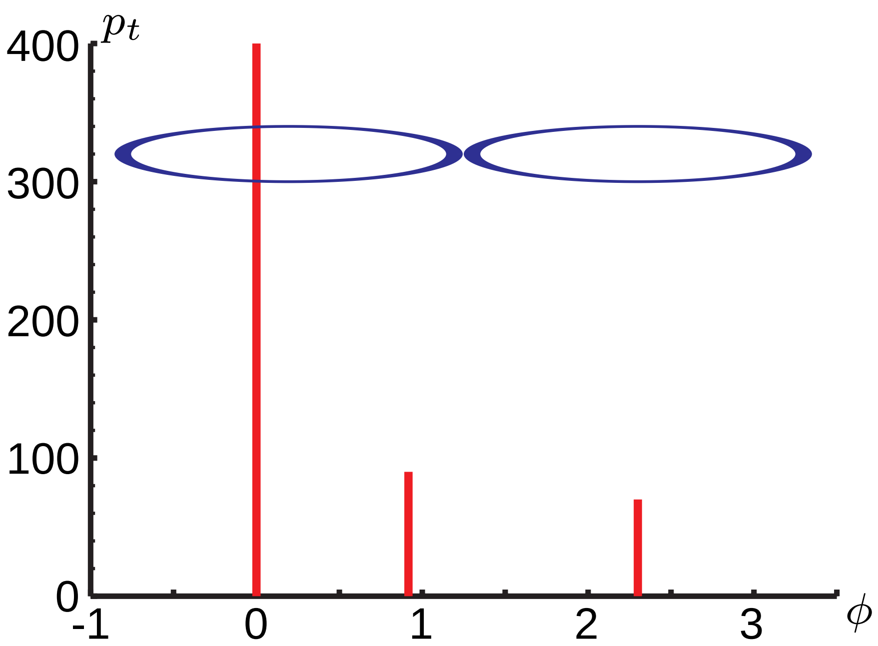
Stable cones:

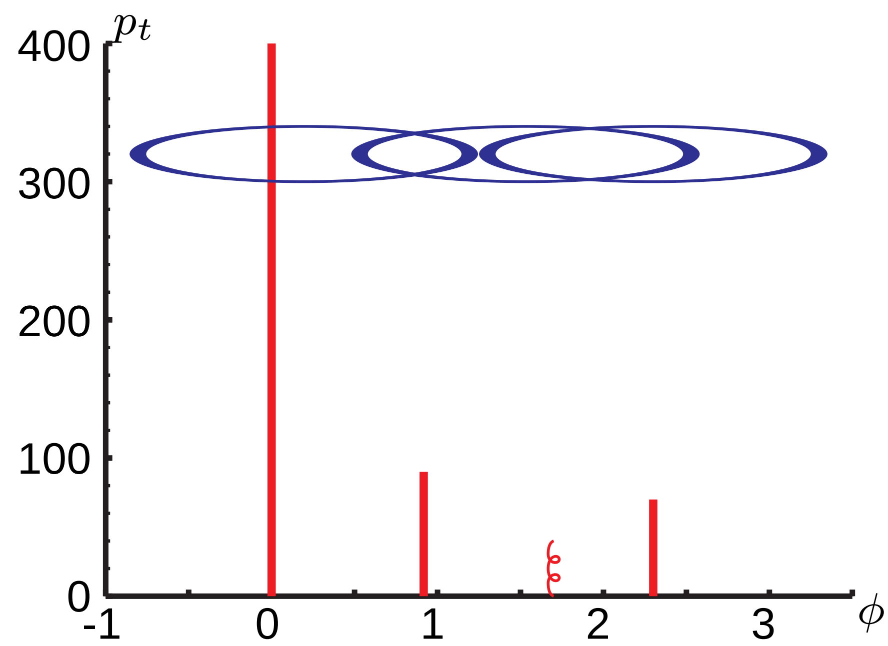
Midpoint:	{1,2} & {3}	{1,2} & {3} & {2,3}

Jets: ($f = 0.5$)

Midpoint:	{1,2} & {3}	{1,2,3}

IR unsafety of the Midpoint alg





Stable cones:

Midpoint:	{1,2} & {3}	{1,2} & {3} & {2,3}
Seedless:	{1,2} & {3} & {2,3}	{1,2} & {3} & {2,3}

Jets: ($f = 0.5$)

Midpoint:	{1,2} & {3}	{1,2,3}
Seedless:	{1,2,3}	{1,2,3}

Stable cone missed $\longrightarrow$ IR unsafety of the midpoint algorithm

Solution: SISCone

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

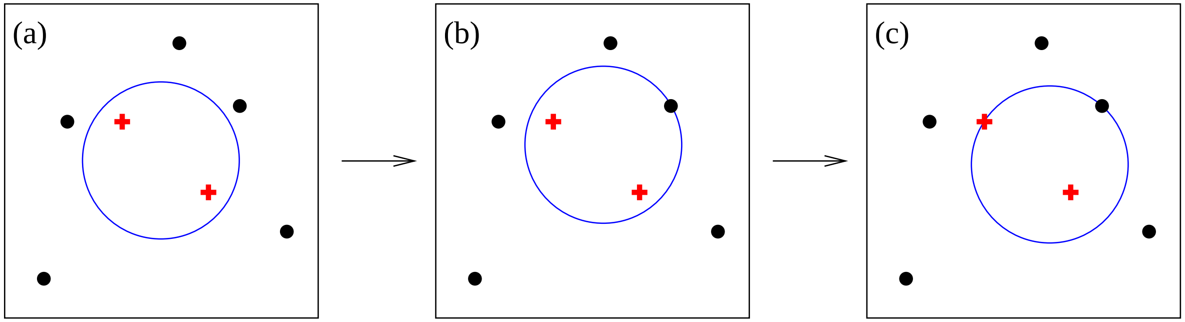
Solution: SISCone

- 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: $\mathcal{O}\left(N^3\right)$

Solution: SISCone

- 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 direction) 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

Solution: SISCone

- 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

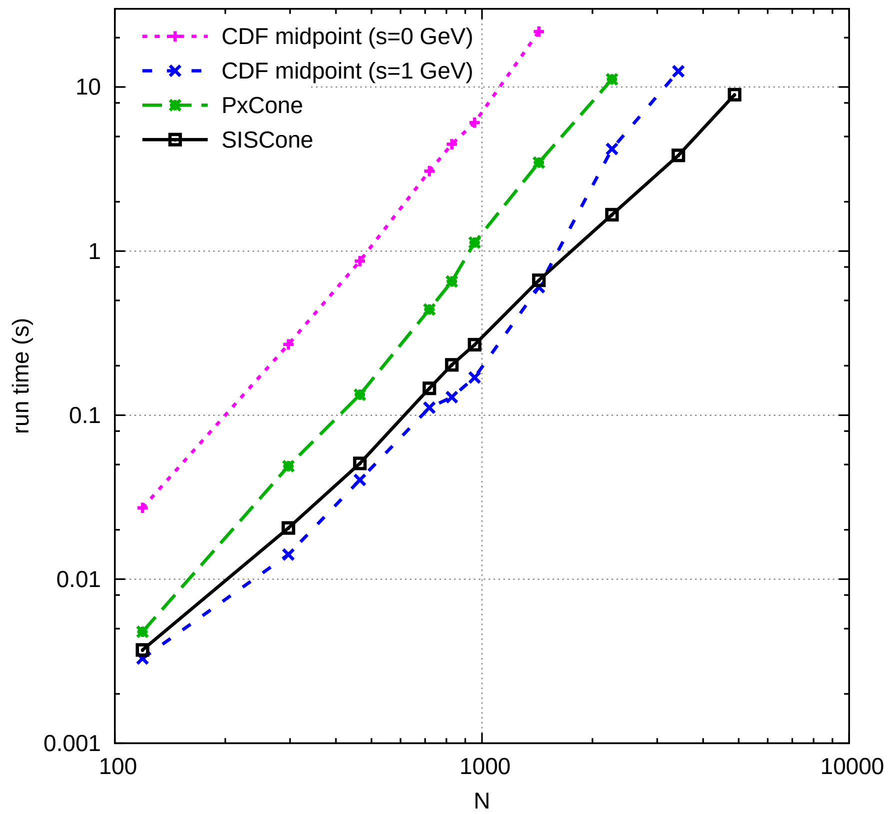
- Complexity: $\mathcal{O}\left(N^3\right)$, with improvements: $\mathcal{O}\left(N^2\log(N)\right)$

$\longrightarrow$ 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

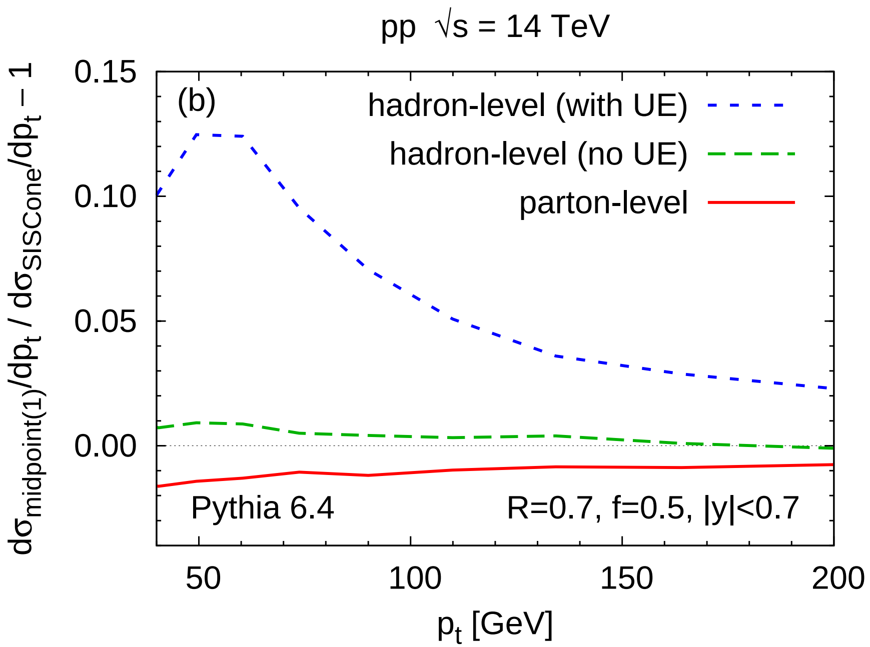
Physical impact

Execution timings:



- faster than midpoint without seed threshold
- at least as fast as as midpoint with seed threshols

(Midpoint-SISCone)/SISCone

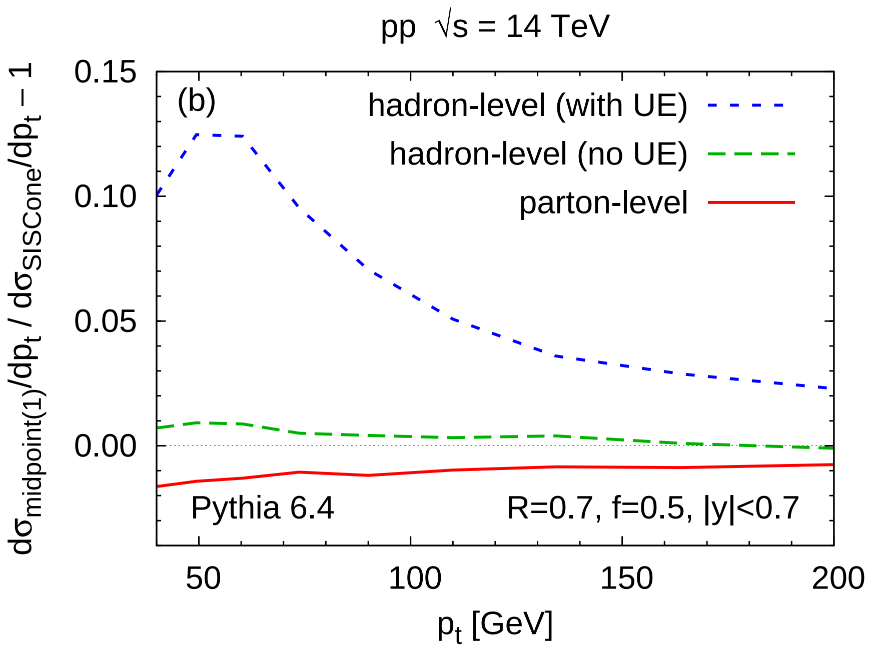


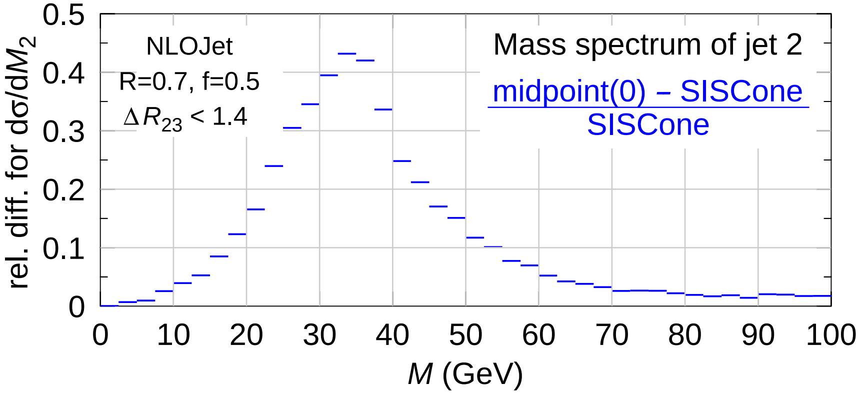
Inclusive cross-section:

- effect of a few %
- less UE sensitivity

Physical impact (2)

(Midpoint-SISCone)/SISCone





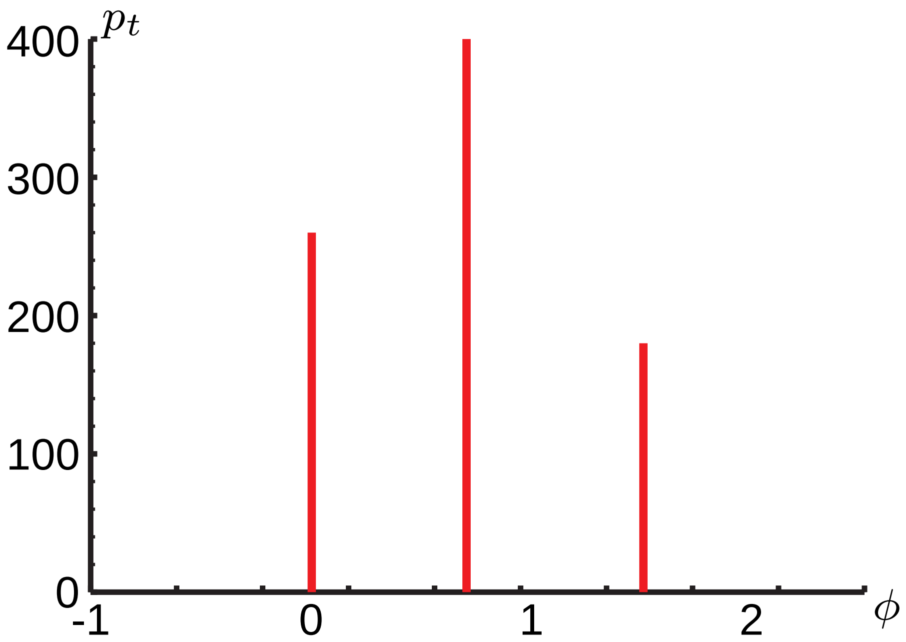
Inclusive cross-section:

- effect of a few %
- less UE sensitivity

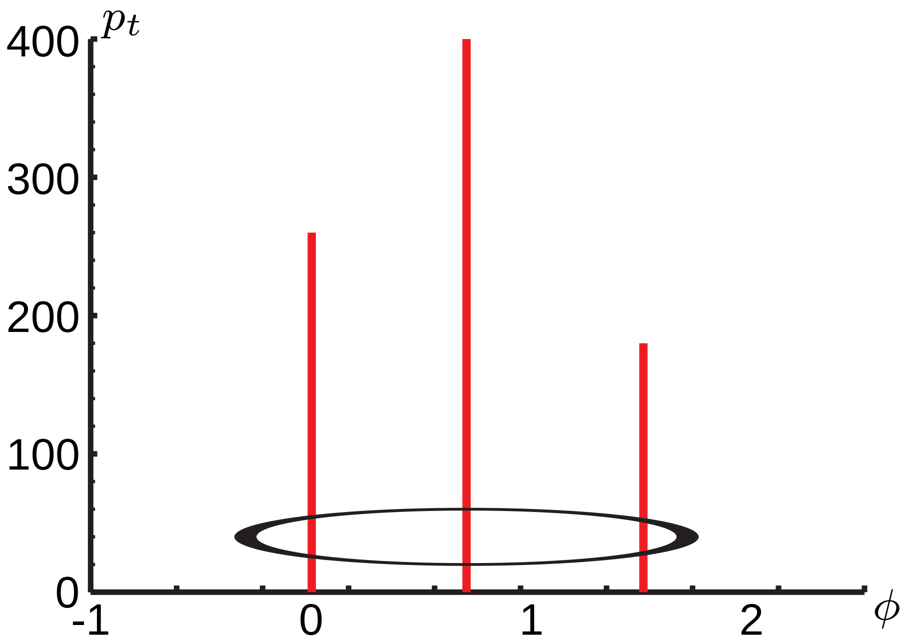
Masses in 3-jet events:

- effects $\sim 45\%$
- Important for LHC!

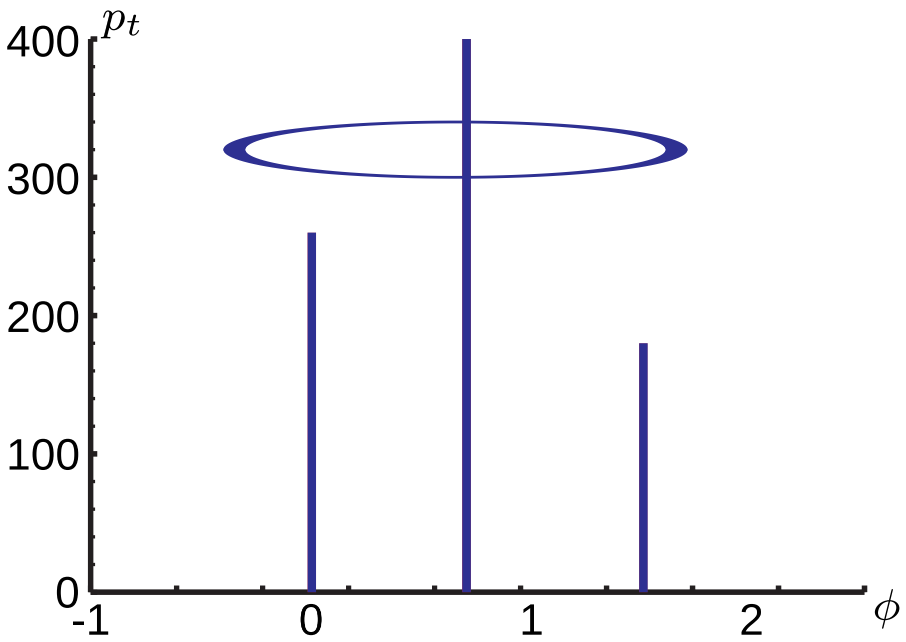
Coll. unsafety of the iterative cone



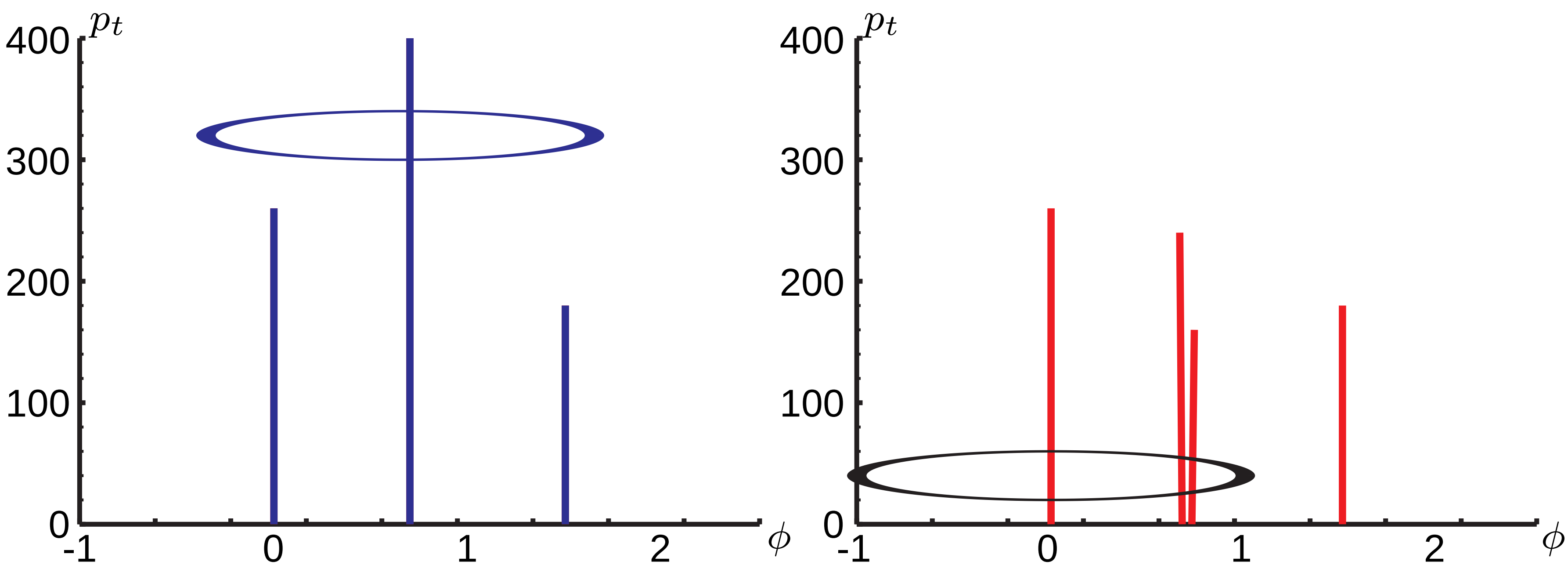
Coll. unsafety of the iterative cone



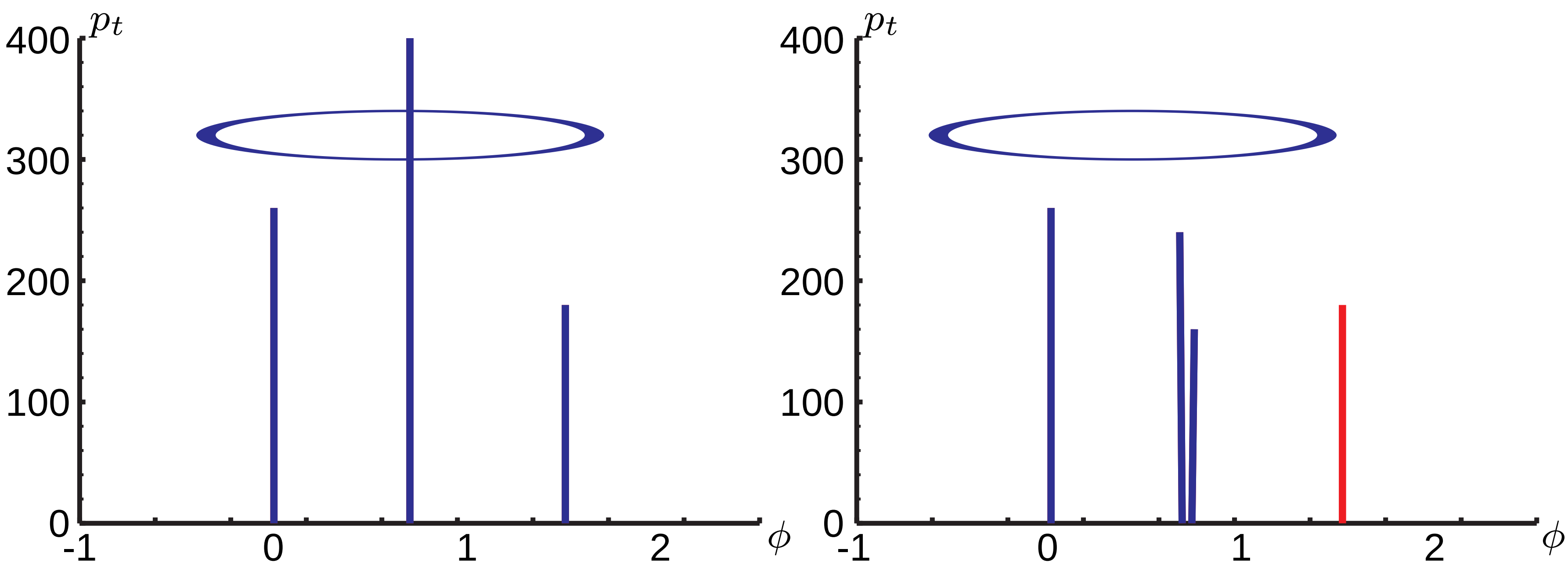
Coll. unsafety of the iterative cone



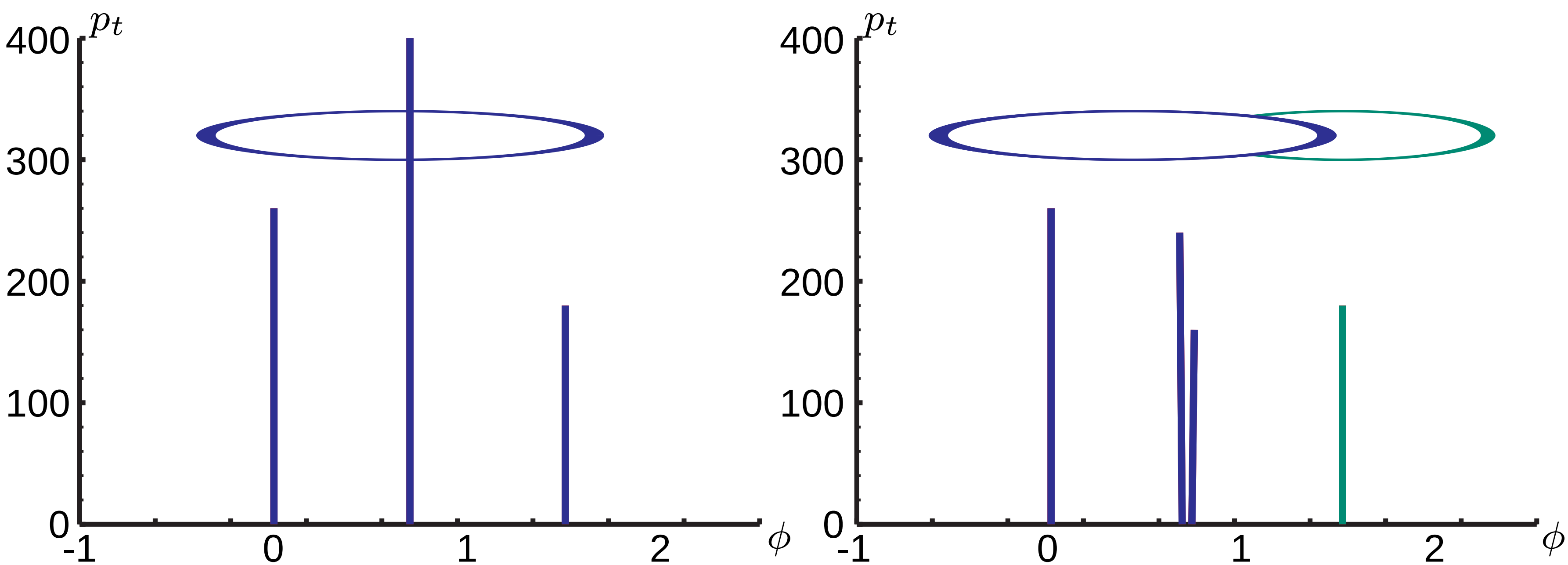
Coll. unsafety of the iterative cone



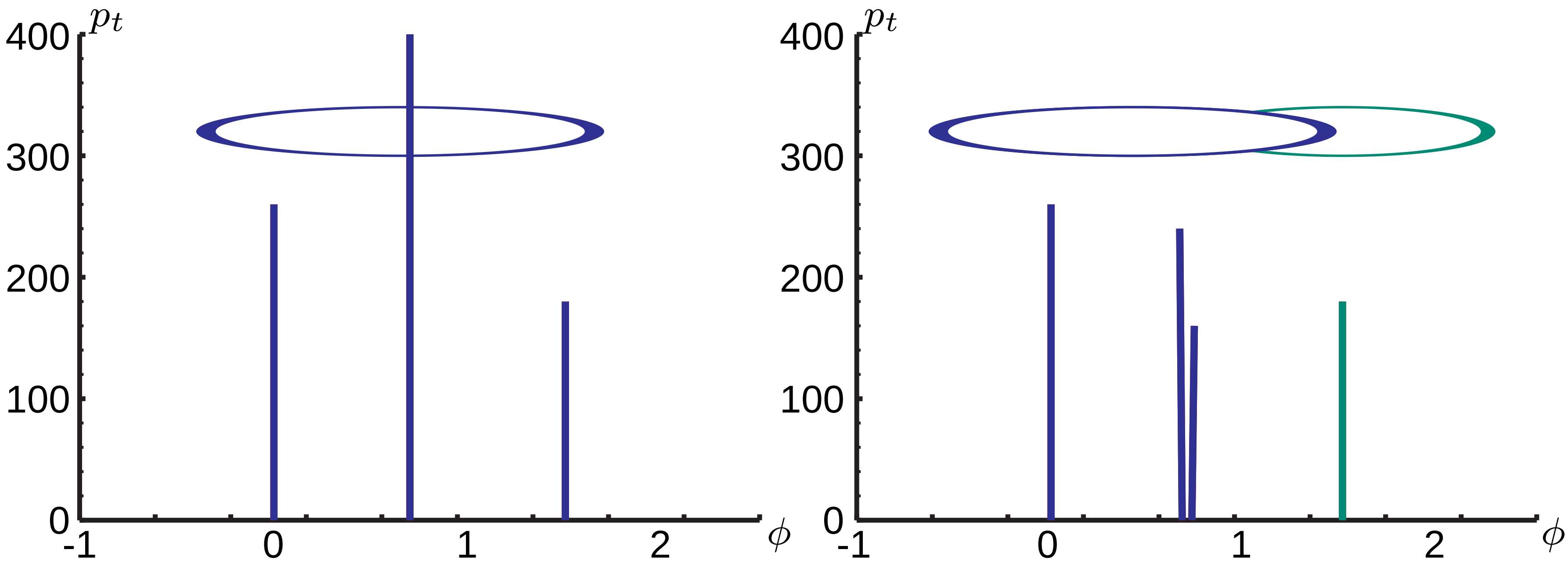
Coll. unsafety of the iterative cone



Coll. unsafety of the iterative cone



Coll. unsafety of the iterative cone



- Before collinear spliting: 1 jet
- After collinear spliting: 2 jets

$\longrightarrow$ collinear unsafety of the iterative cone algorithm

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.,JHEP 04 (08) 063]

Anti-k_t

Come back to recombination-type algorithms:

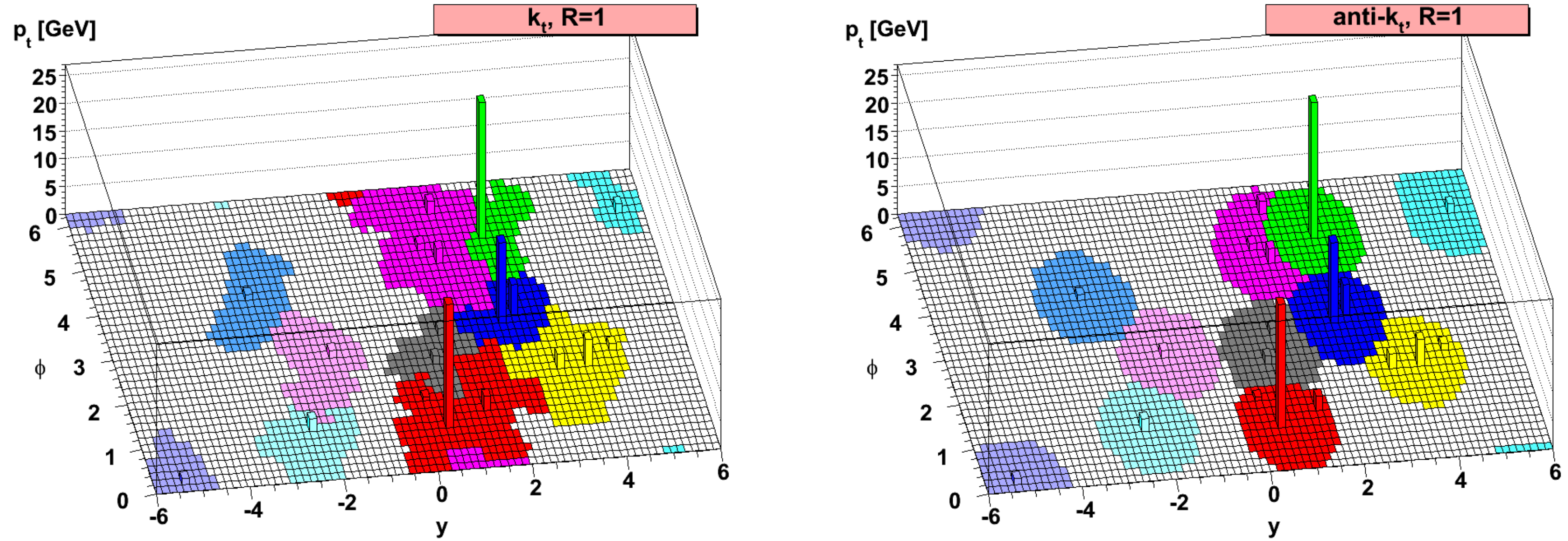
$$d_{ij} = \min(k_{t,i}^{2p}, k_{t,j}^{2p})\ (\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2)$$

- $p = 1$: k_t algorithm
- $p = 0$: Aachen/Cambridge algorithm
- $p = -1$: anti-k_t algorithm [M.Cacciari, G.Salam, G.S.,JHEP 04 (08) 063]

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

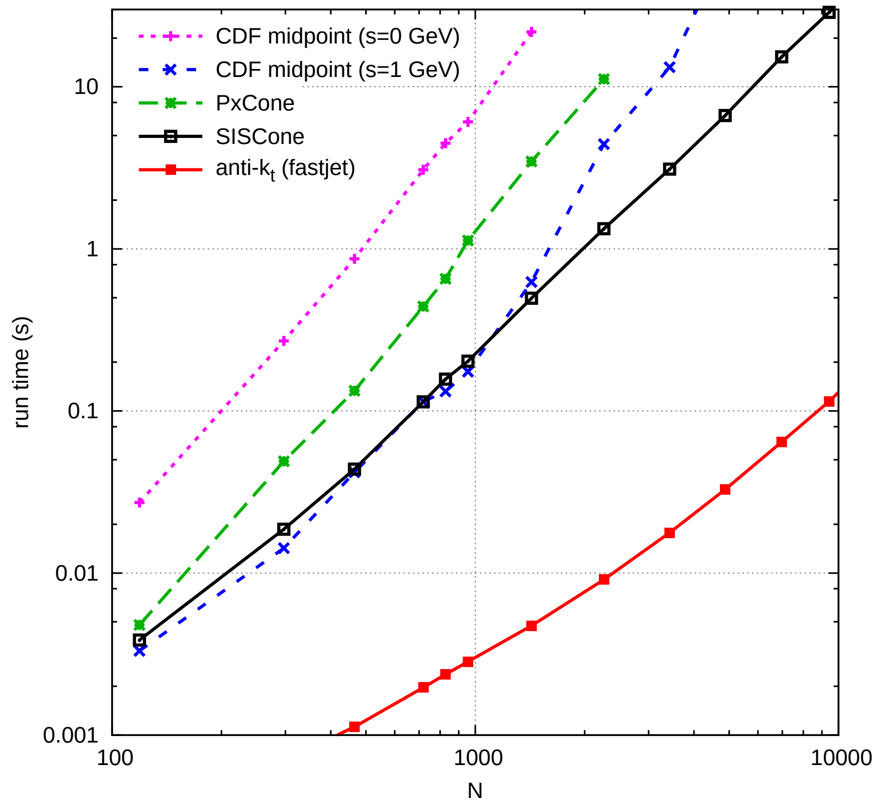
Hard event + homogeneous soft background



anti-k_t is soft-resilient

more later in this talk...

Execution timings:



As fast as the (fast) k_t ([M. Cacciari, G. Salam, 06])

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

Part 2
Jets in pp collisions
(a) Choosing the adapted jet definition

[M.Cacciari,J.Rojo,G.Salam,G.S., arXix:0810.1304]

Sample processes to 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:

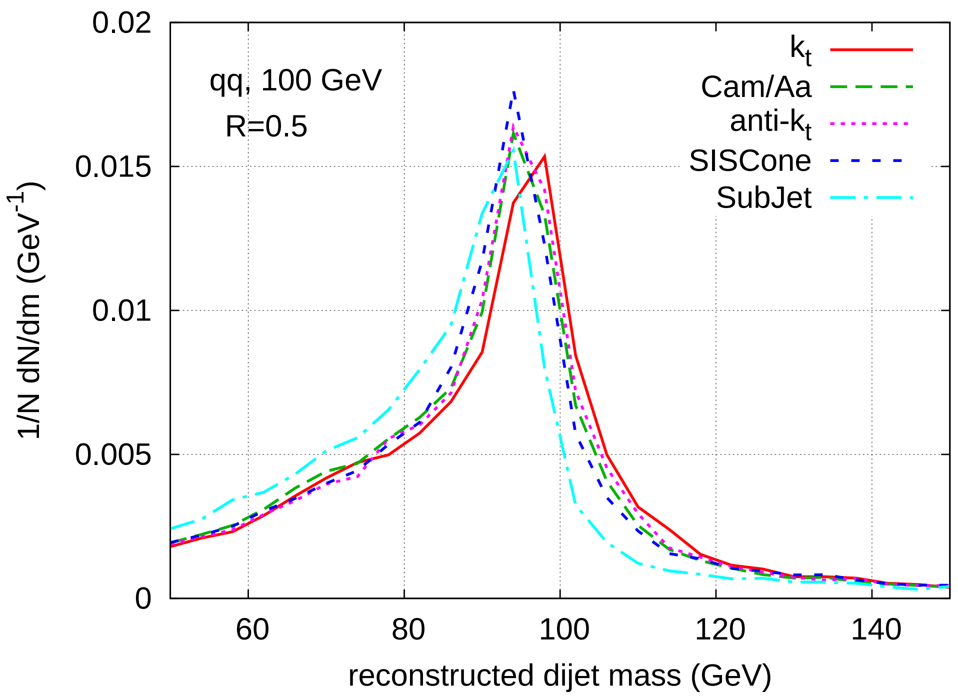
  complex environment: identify 6 jets and reconstruct 2 top
  balance between reconstruction efficiency and identification

with

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

Histograms

- We reconstruct histograms



- How can we quantify the reconstruction efficiency?

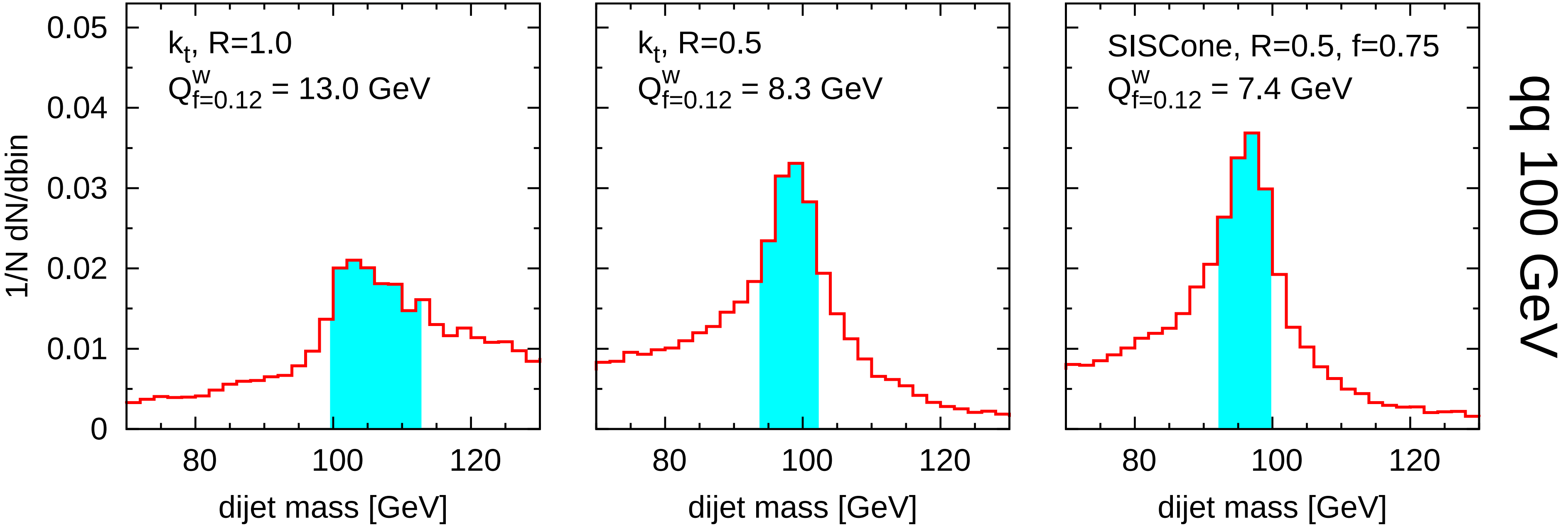
Measure of the jet reconstruction efficiency

- Forget about measures related to parton-jet matching,
  $\longrightarrow$ use the reconstructed mass peak
- Forget about fits depending on the shape of the peak

$\Rightarrow$ maximise the signal over background ratio ($S/\sqrt{B}$):

figure of merit for quality measure

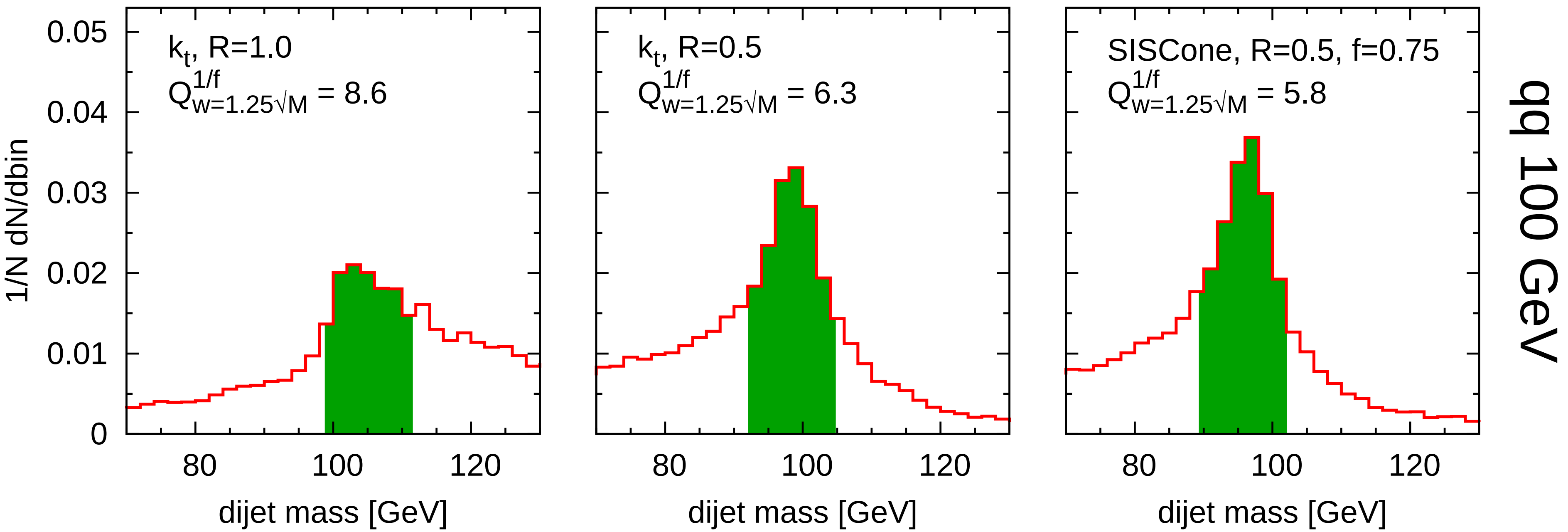
$Q^w_{f=z}(JA,R) =$ minimal width of a window containing a fraction $f = z$ of the events



Fixed signal, minimal width(background)

figure of merit for quality measure

$$Q_f^{w=x\sqrt{M}}(JA, R) = \text{(1/) maximal number of events in a window of width } x\sqrt{M}$$



Maximal signal, fixed width(background)

figure of merit for quality measure

- it intuitively does what it should

figure of merit for quality measure

- it intuitively does what it should
- relates to a signal significance(assuming constant background)

$$\frac{\Sigma(\mathrm{JD}_1)}{\Sigma(\mathrm{JD}_2)} = \left[\frac{N_{\text{signal}}}{\sqrt{N_{\text{bkg}}}}\right]_{\mathrm{JD}} = \sqrt{\frac{Q^w_{f=z}(\mathrm{JD}_2)}{Q^w_{f=z}(\mathrm{JD}_1)}} = \frac{Q_f^{w=x\sqrt{M}}(\mathrm{JD}_2)}{Q_f^{w=x\sqrt{M}}(\mathrm{JD}_1)}$$

minmimal $Q \equiv$ better signal-to-background ratio

figure of merit for quality measure

- it intuitively does what it should
- relates to a signal significance(assuming constant background)

$$\frac{\Sigma(\mathrm{JD}_1)}{\Sigma(\mathrm{JD}_2)} = \left[\frac{N_{\text{signal}}}{\sqrt{N_{\text{bkg}}}}\right]_{\mathrm{JD}} = \sqrt{\frac{Q^{w}_{f=z}(\mathrm{JD}_2)}{Q^{w}_{f=z}(\mathrm{JD}_1)}} = \frac{Q_f^{w=x\sqrt{M}}(\mathrm{JD}_2)}{Q_f^{w=x\sqrt{M}}(\mathrm{JD}_1)}$$

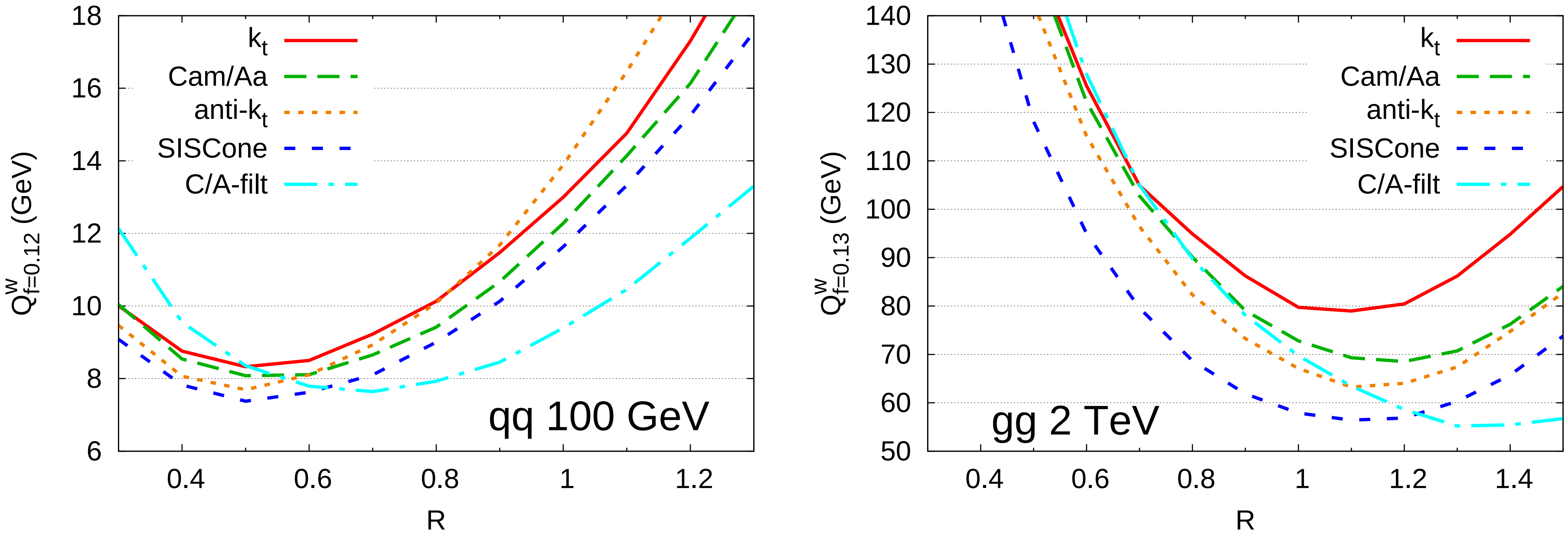
  minmimal $Q \equiv$ better signal-to-background ratio

- we can associate an effective luminosity ratio

$$\rho_{\mathcal{L}}(\mathrm{JD}_2/\mathrm{JD}_1) = \frac{\mathcal{L}\text{ needed with }\mathrm{JD}_1}{\mathcal{L}\text{ needed with }\mathrm{JD}_2} = \left[\frac{\Sigma(\mathrm{JD}_1)}{\Sigma(\mathrm{JD}_2)}\right]^2$$

  e.g. $\rho_{\mathcal{L}} = 2 \equiv \mathrm{JD}_1$ has $\sqrt{2}$ the significance of JD_2
  $\equiv \mathrm{JD}_2$ requires 2 times the integrated luminosity
  to achieve the same discriminative power.

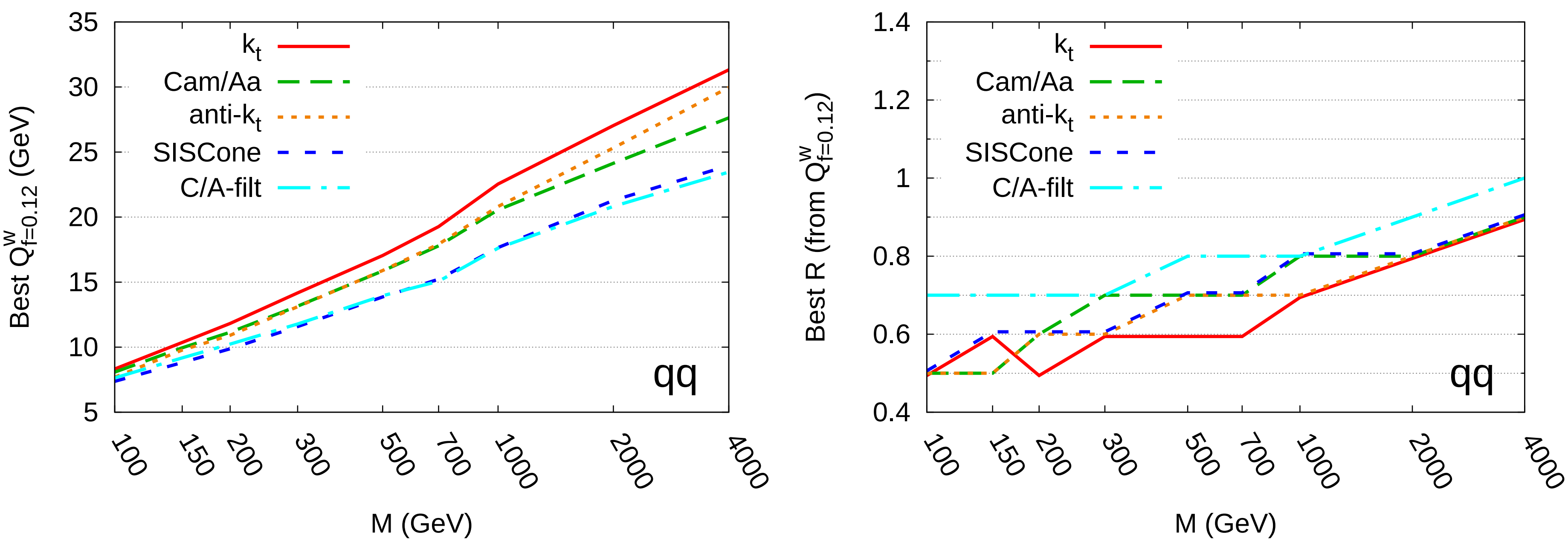
Examples: best quality measures



Allows to

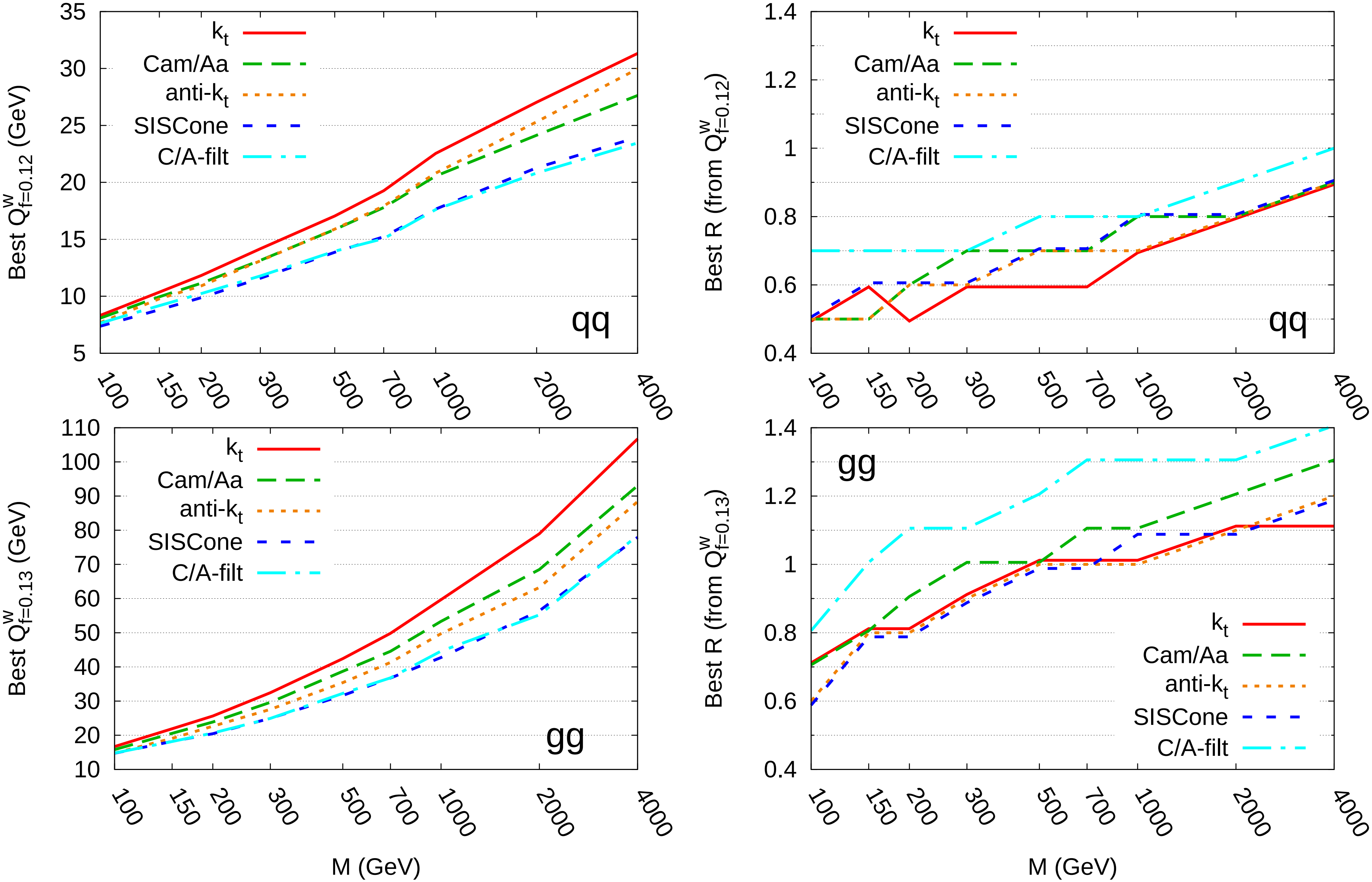
- extract the best radius R_{best}
- compare the different algorithm

Best choices



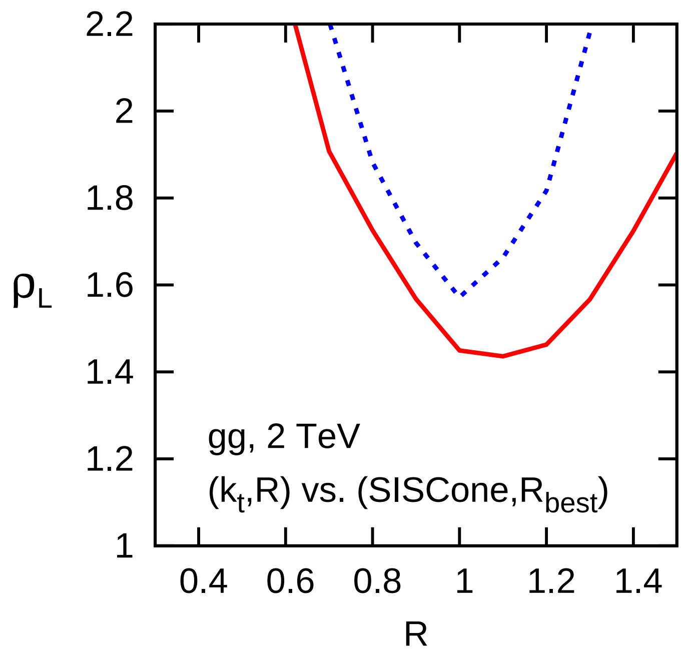
- SISCone and Cam+filtering perform better
- R_{best} strongly depends on the mass

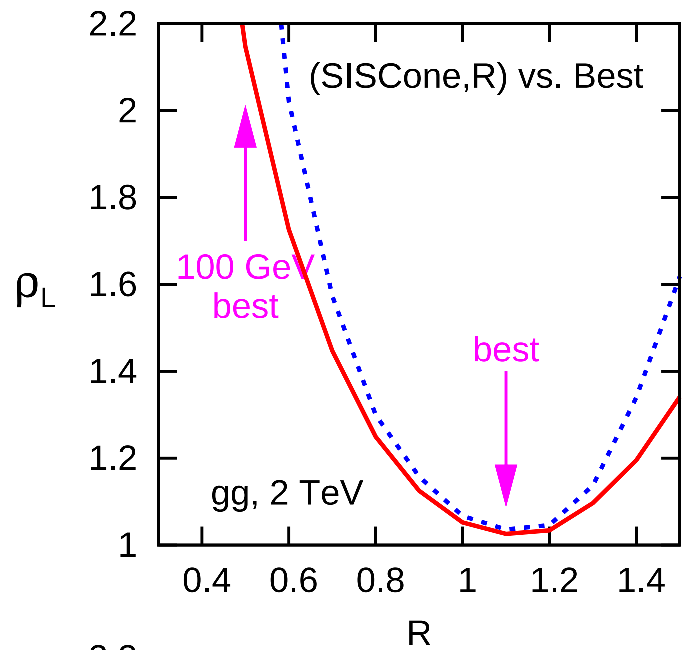
Quarks vs. gluons

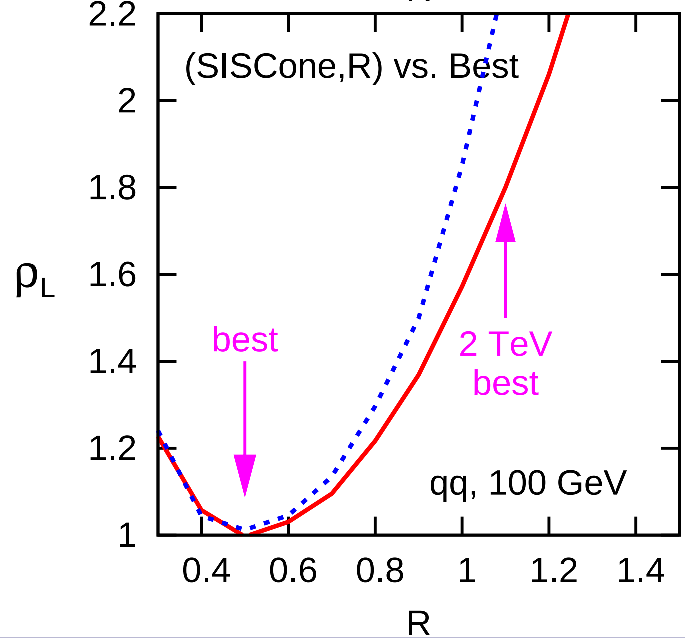


Same conclusions for gluon jets with slightly larger R

Luminosity ratios







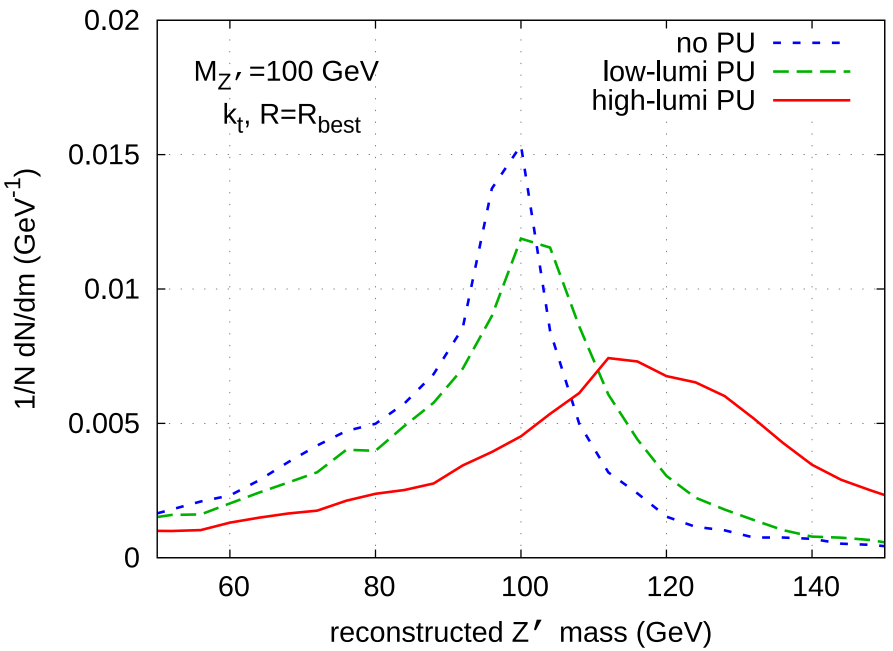
Mandatory at the LHC:
Not choosing the best alg.
AND R can be very costly
for new discoveries

Note: typical choice, $R \sim 0.5$

Part 2
Jets in pp collisions
(b) pileup effect (jet areas & subtraction)

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]

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

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

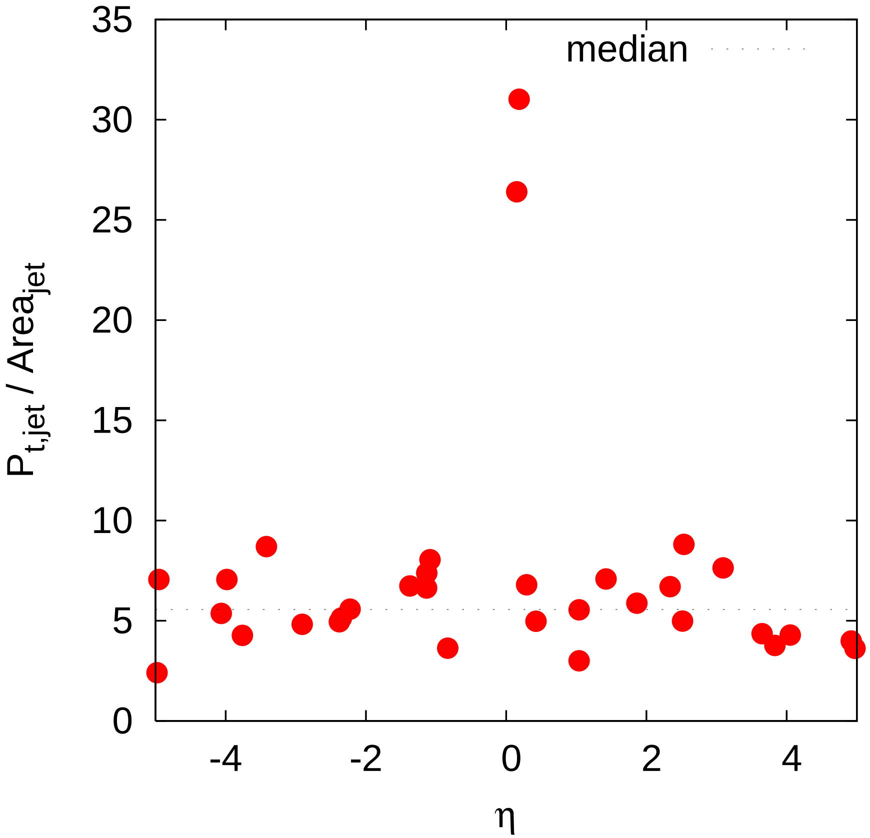
Pileup subtraction

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

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

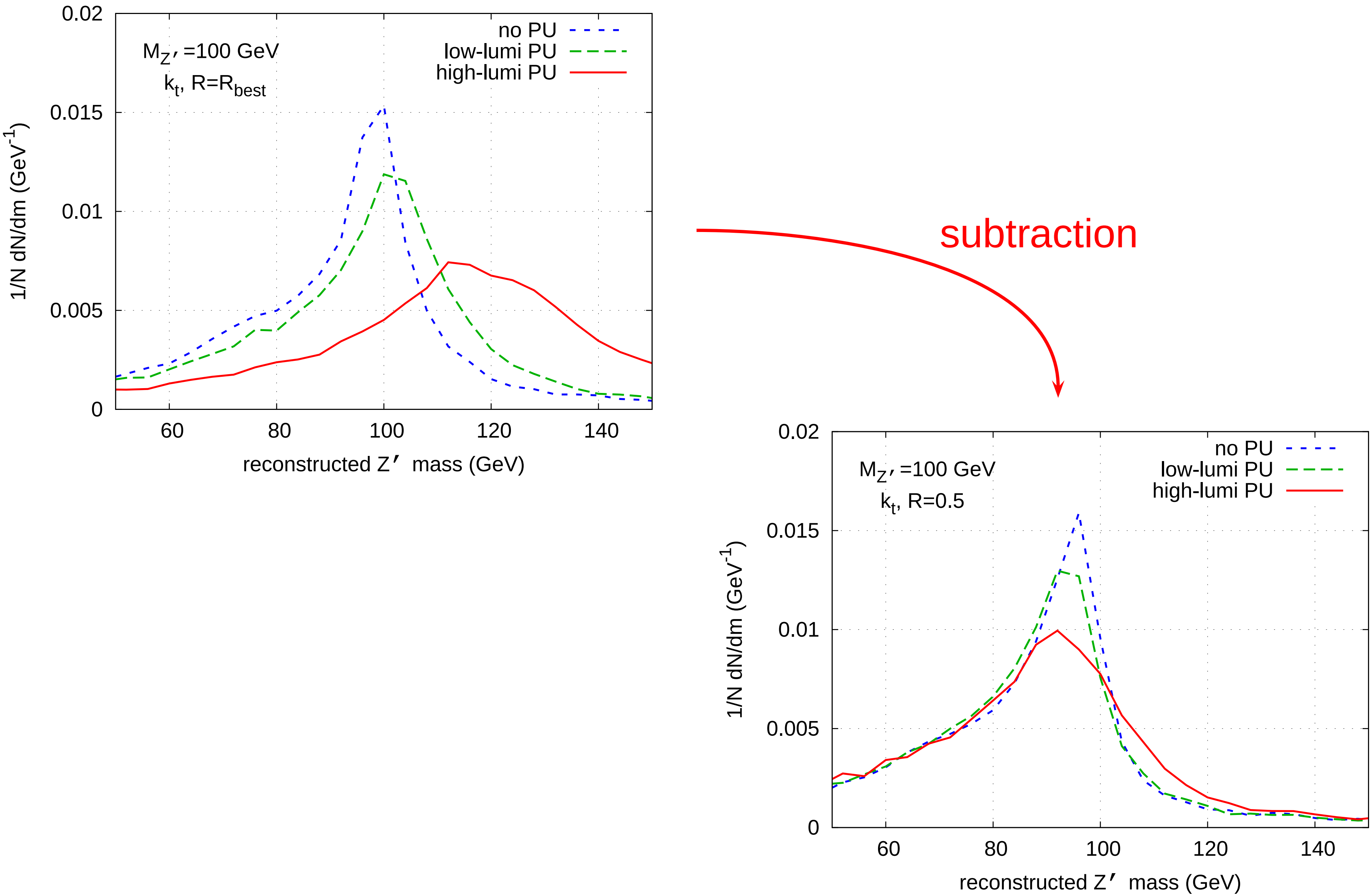
- Jet area: [M.Cacciari, G.Salam, G.S., 08]
  - region where the jet catches infinitely soft particles (active/passive)
  - tractable analytically in pQCD
- Pileup density per unit area: ρ_{pileup}

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

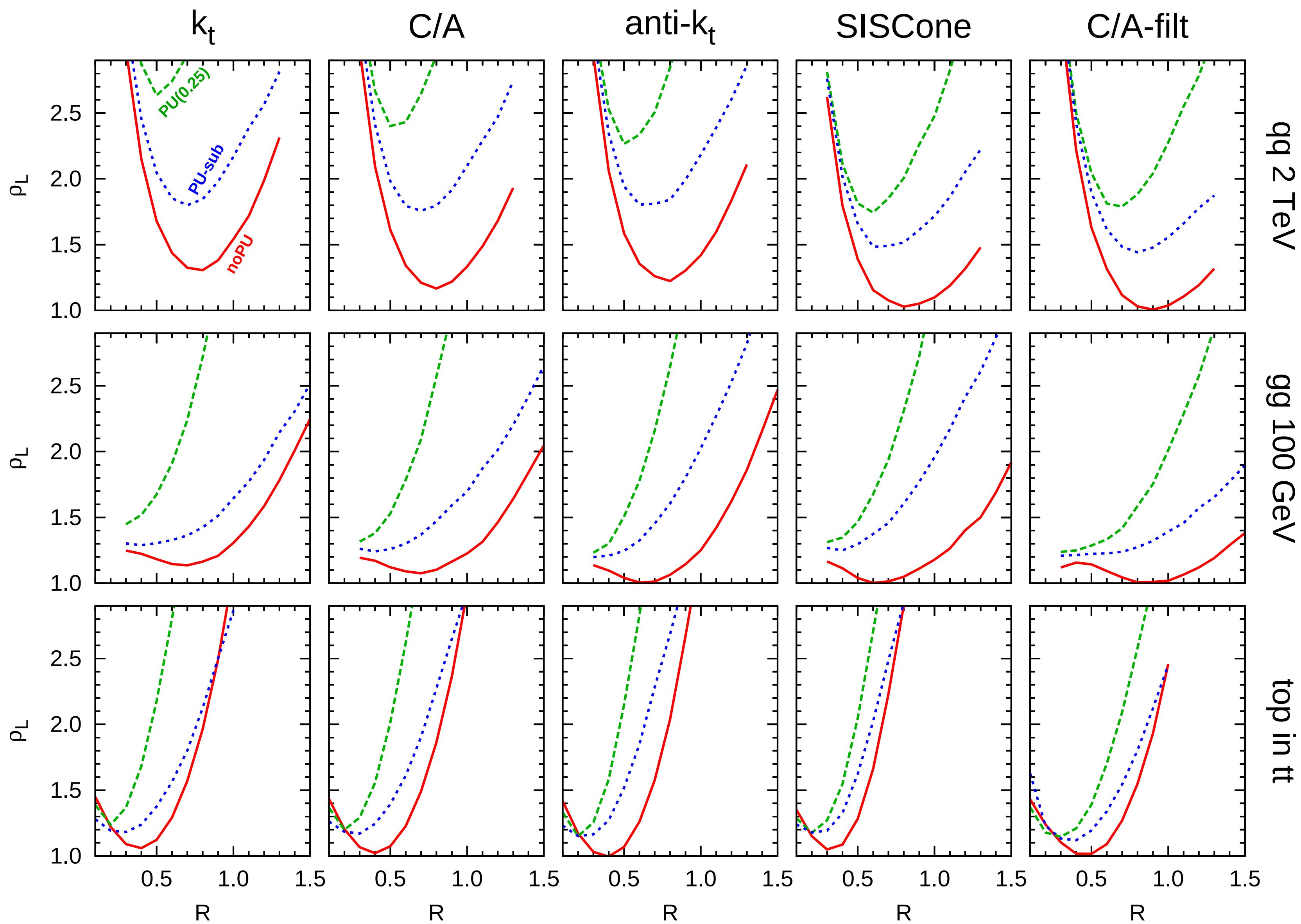


implemented in FastJet
on an event-by-event basis

Subtraction at work



PU effects summary

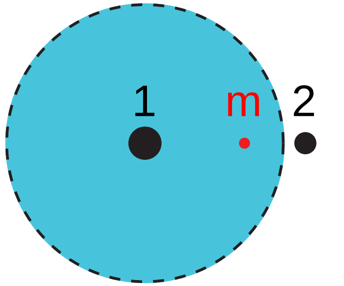


Subtraction ⇒ (i) large improvement, (ii) $R_{\text{best}} \sim$ unchanged

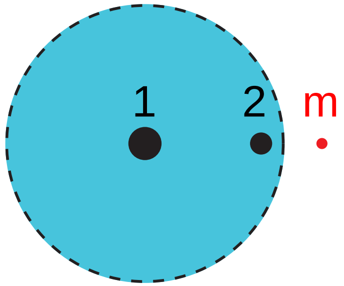
Back-reaction

Additional soft background has 2 effects:

- Throw soft particles in the hard jet: dealt with by subtraction
- Modify the hard scattering (back-reaction)
  - can be pointlike or diffuse
  - gain: p_2 gained when adding p_m

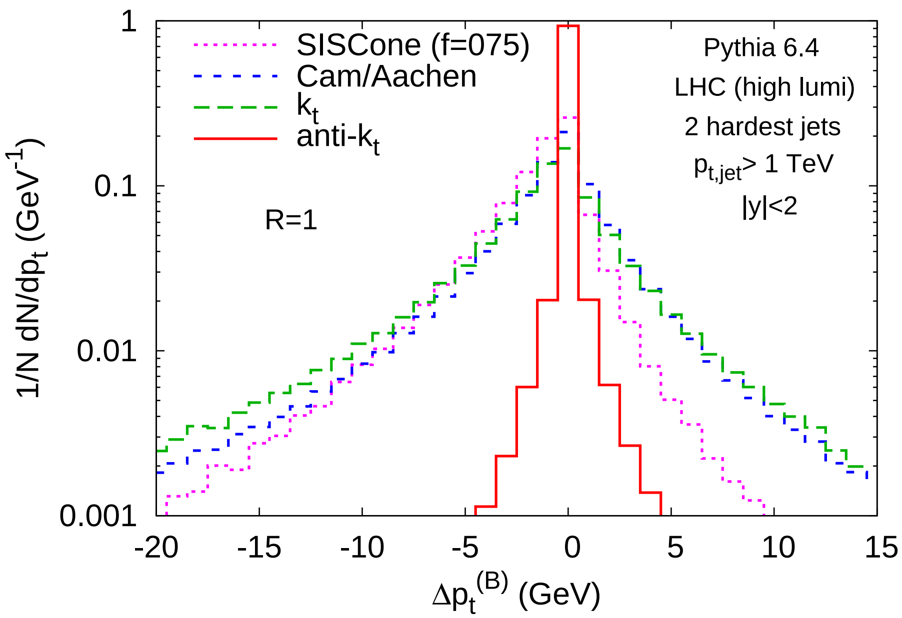


  - loss: p_2 lost when adding p_m



Additional soft background has 2 effects:

- Throw soft particles in the hard jet: dealt with by subtraction
- Modify the hard scattering (back-reaction)
  - can be pointlike or diffuse
  - tractable analytically (similar to areas)
  - $k_t \gtrsim$ Cambridge > SISCone $\gg$ anti-k_t



Conclusions

Message 1: IRC safety is mandatory

Midpoint and the iterative cone IR or Collinear unsafe (at $\mathcal{O}(\alpha_s^4)$)

Observable	1st miss cones at	Last meaningful order
Inclusive jet cross section	NNLO	NLO
3 jet cross section	NLO	LO (NLO in NLOJet)
$W/Z/H$ + 2 jet cross sect.	NLO	LO (NLO in MCFM)
jet masses in 3 jets	LO	none (LO in NLOJet)

+ We do not want the theoretical efforts to be wasted

- Note: 1 order worse for JetClu of the ATLAS Cone!
- All IRC-safe algorithms available from FastJet (http://www.fastjet.fr)

Conclusions

Message 2: flexibility in jet finding at the LHC

- Optimal jet definition (see also http://quality.fastjet.fr)
  - $R_{\text{best}} \sim 0.5$ at 100 GeV, $R_{\text{best}} \sim 1$ at 1 TeV
  - important to choose R_{best}, SISCone and Cam+filt. slightly better
  - same for quark and gluon jets, larger R_{best} for gluons
  - TODO: understand this analytically/ improve clustering (e.g. filtering)

- Pileup subtraction using jet areas
  - Jet areas: clearly defined, analytic control
  - Simple systematic pileup subtraction
  - Same conclusions as without pileup
  - TODO: deal with fluctuating backgroubd (e.g. heavy ions)