Jet finding at the LHC era

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:0803.0678 + works in preparation

Plan

- Foreword: why jets? what are they?

  introducing the basic concepts

- Part 0: recent progress: building a solid toolkit

  jet definitions meeting the fundamental requirements

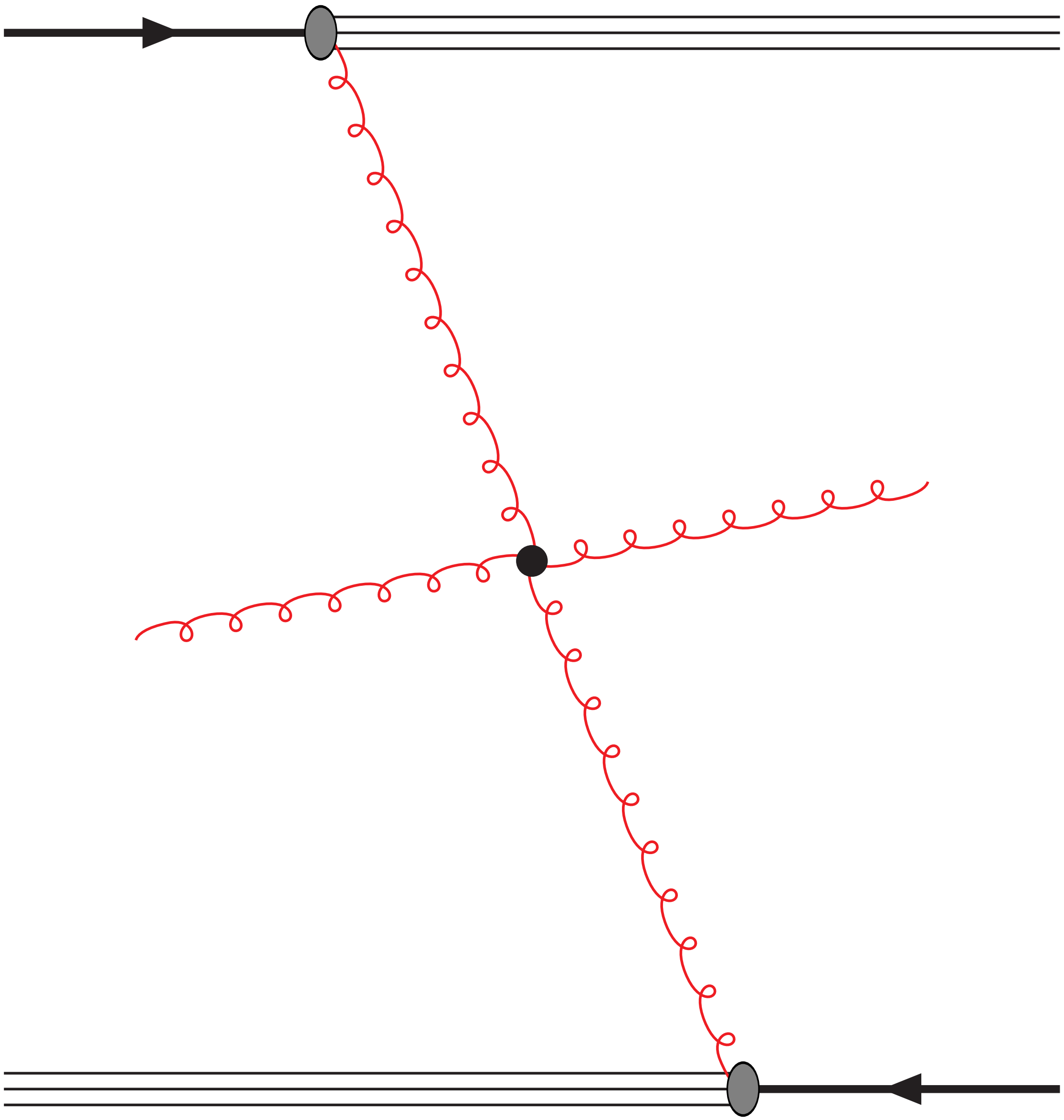
- Part 1: jets in pp collisions
  - Choosing the adapted jet definition

    which jet algorithm is best suited?
  - Subtracting pileup background using jet areas
    - defining areas
    - analytic control
    - using them for pileup subtraction

- Part 2: jets in heavy-ion collisions
  - subtraction subtleties
  - preliminary results

Foreword: why jets? what are they?

General (over)simplified picture



Hard scattering $(2 \to n)$

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

$gg \to gg$, $gg \to ggg$,

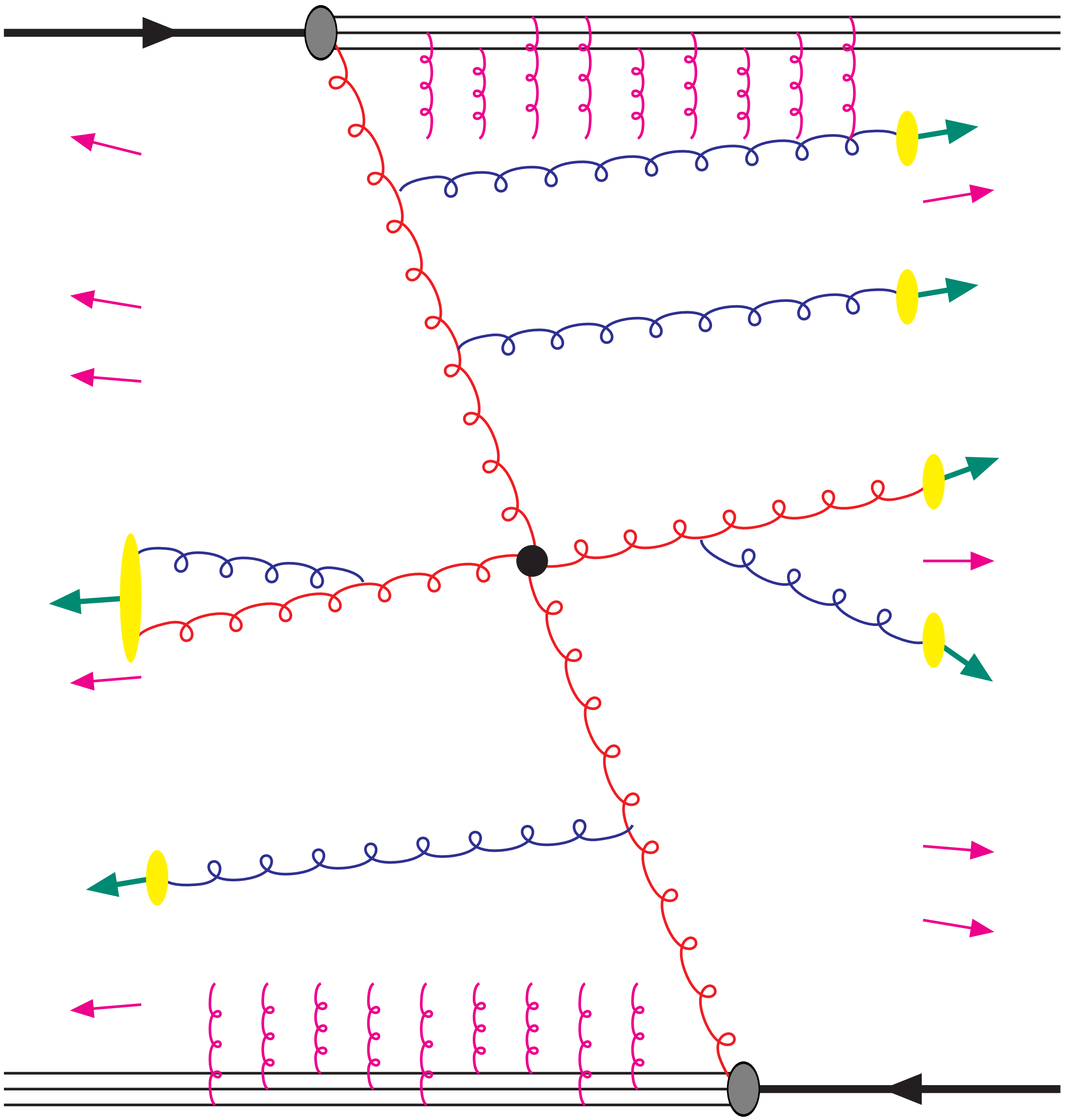
$gg \to gggg$,

$gg \to H \to b\bar{b}$,

$gg \to t\bar{t} \to \mu\nu_\mu b\bar{b}q\bar{q}$,

$gg \to Z' \to q\bar{q}$, ...

General (over)simplified picture



Hard scattering ($2 \to n$)

Parton level

$\approx$ resummed collinear div.

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

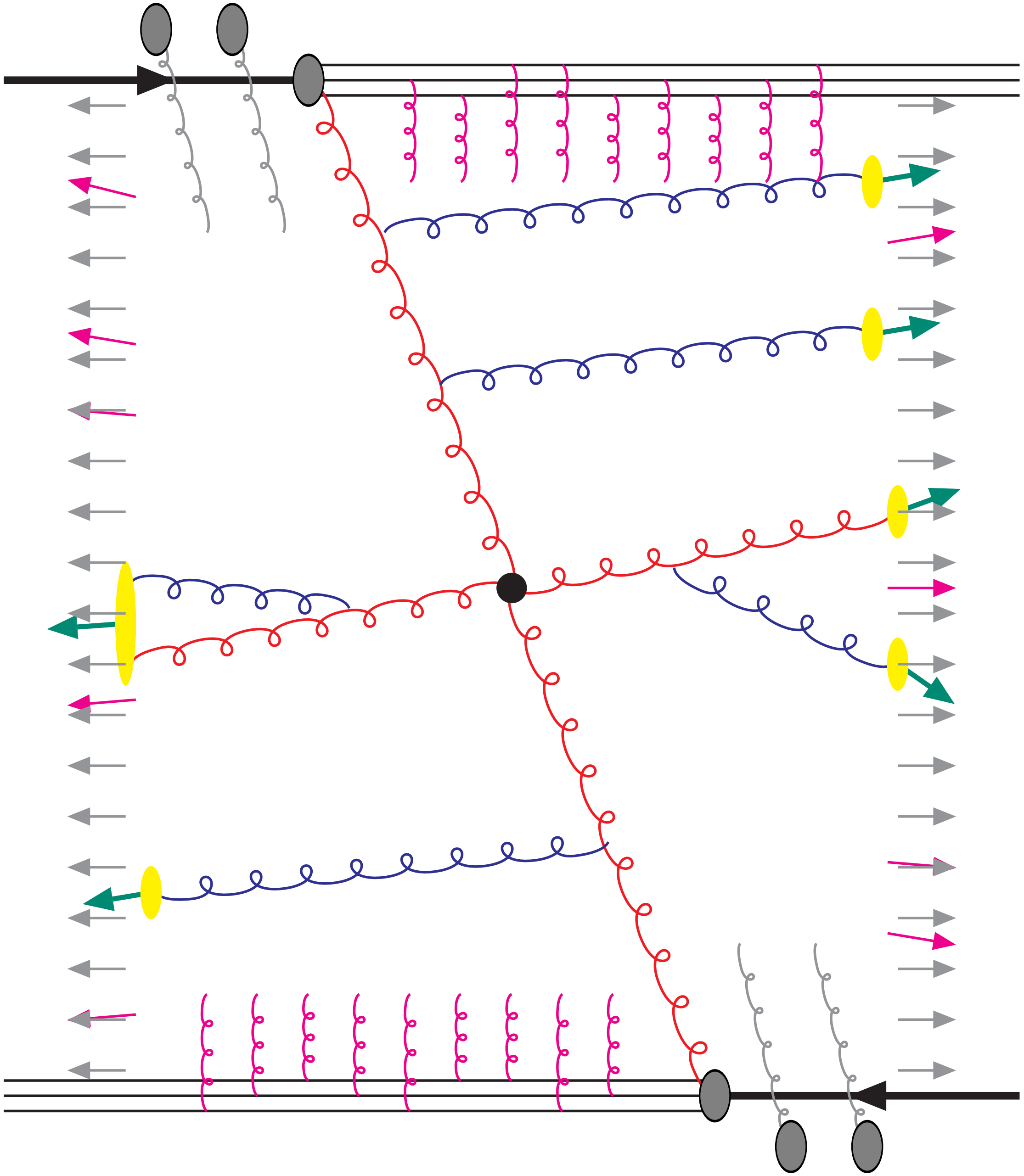
Hadron level: hadronisation

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

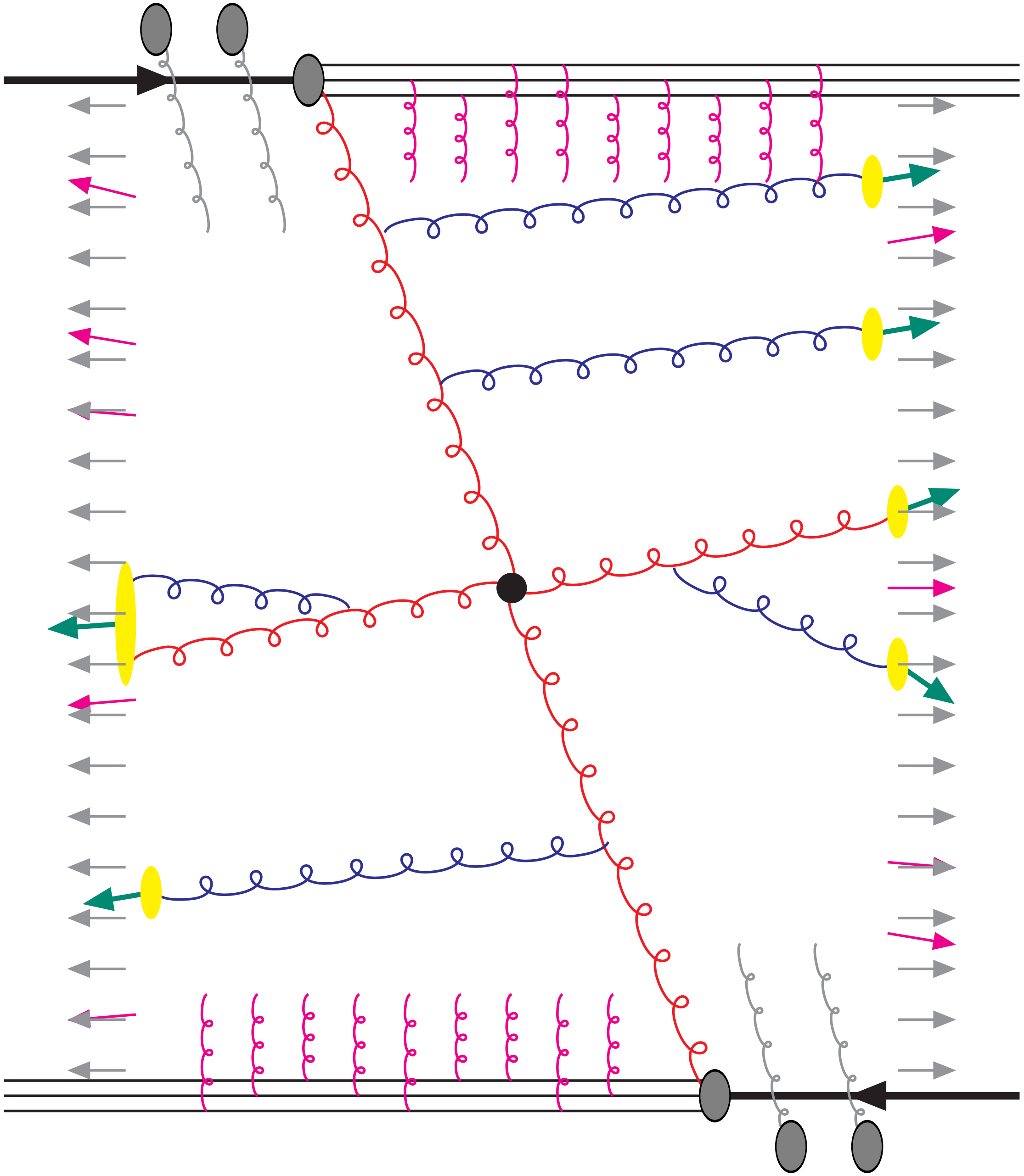
beam remnants interactions

$\Rightarrow$ 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 $d_{\min} > R$)

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

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

$^{(b)}$ **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

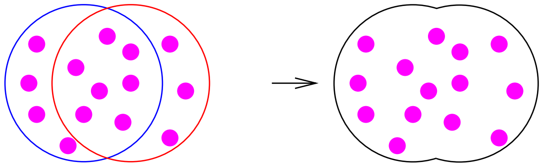
---

How the cone works...

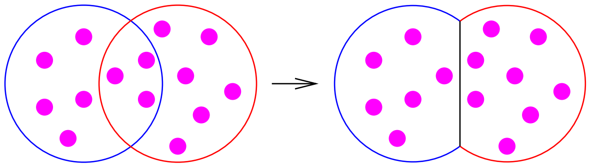
- Seeded (iterative) approaches: iterate from an initial position until stable
  - seed = initial particle
  - seed = midpoint between stable cones found at first step

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

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



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

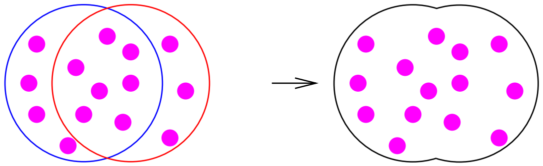


How the cone works...

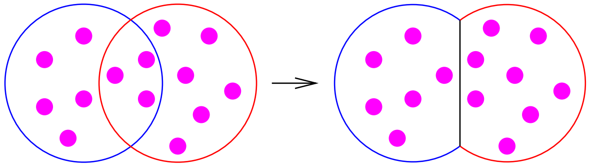
- Seeded (iterative) approaches: iterate from an initial position until stable
  - seed = initial particle
  - seed = midpoint between stable cones found at first step

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:

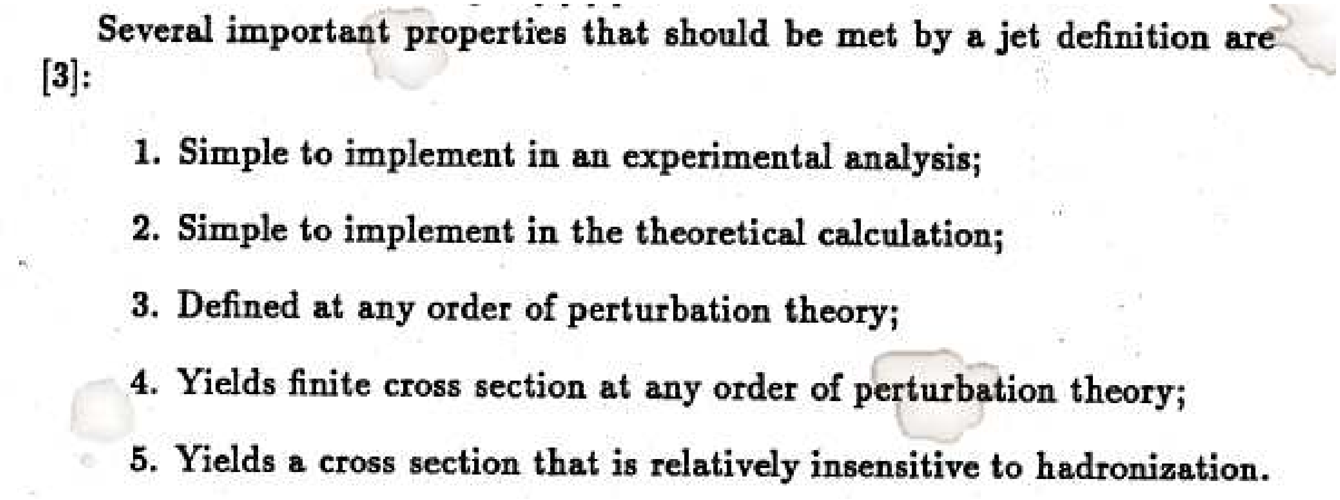
- CDF JetClu
- CDF MidPoint
- D0 (run II) Cone
- PxCone
- ATLAS Cone
- CMS Iterative Cone
- PyCell/CellJet
- GetJet

Part 0

21st century: towards a solid toolkit

1990: fixing the rules

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)

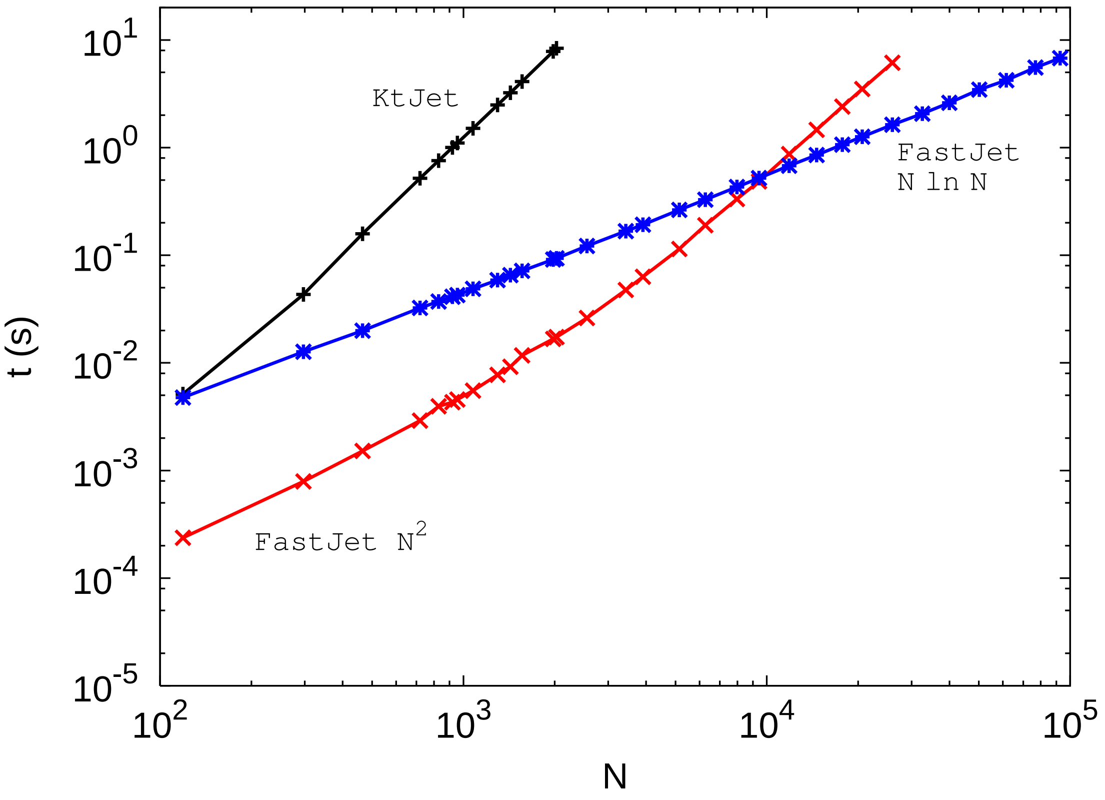
and experimentalists (*e.g.* fast enough, not much UE sensitivity)

Speed improvement

[M. Cacciari, G. Salam, 06]

Speeding up the k_t and Cam/Aachen algorithms

- using computational-geometry techniques: $\mathcal{O}\left(N^3\right) \rightarrow \mathcal{O}\left(N \log N\right)$
- `C++` implementation in `FastJet`

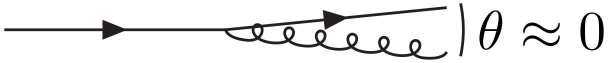


QCD divergences

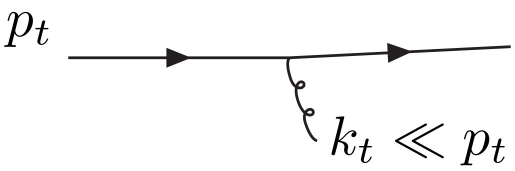
QCD probability for gluon bremsstrahlung at angle θ and $\perp$-mom. k_t:

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

Two divergences:



Collinear



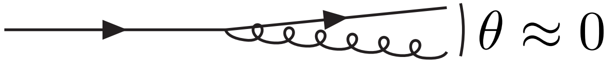
Soft

QCD divergences

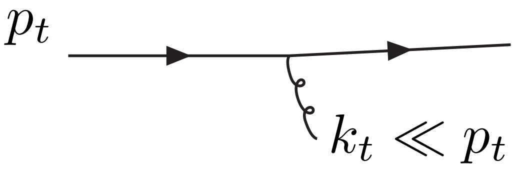
QCD probability for gluon bremsstrahlung at angle θ and $\perp$-mom. k_t:

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

Two divergences:



Collinear

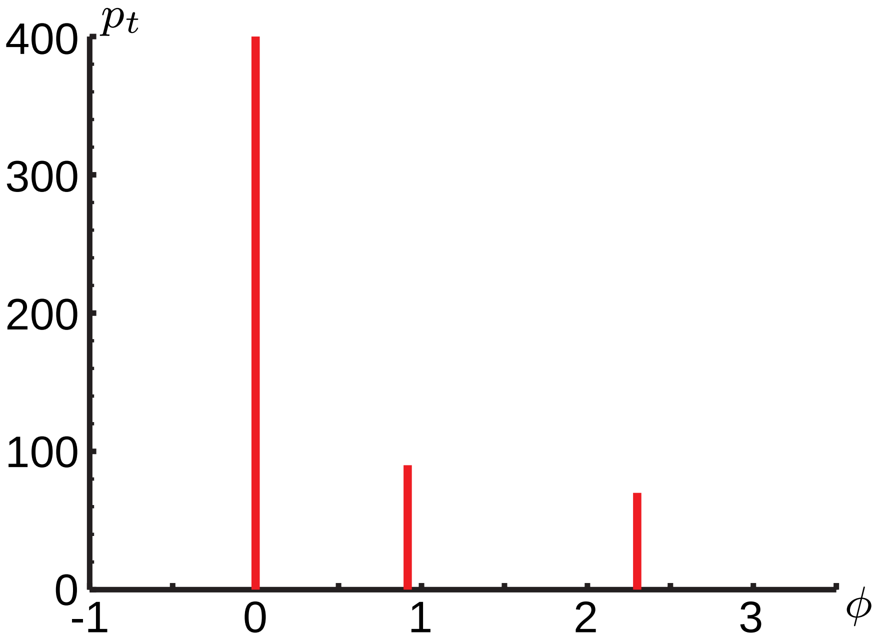


Soft

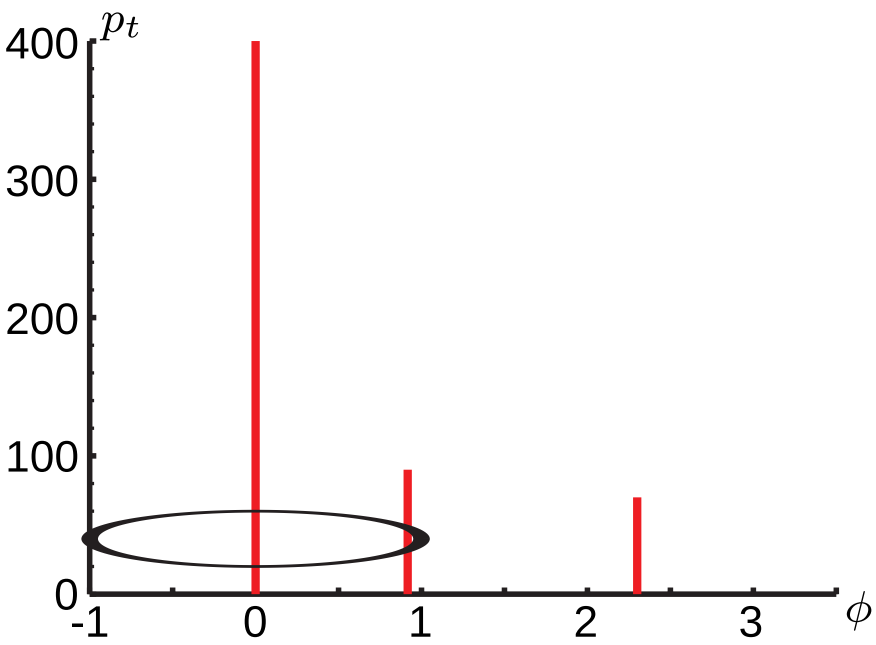
For pQCD to make sense, the (hard) jets should not change when

- 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

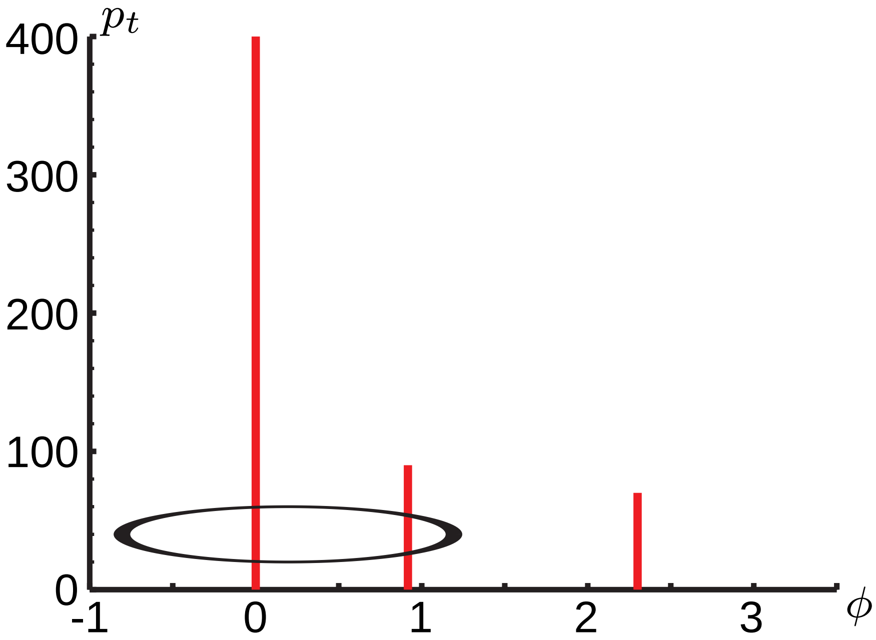
IR unsafety of the Midpoint alg



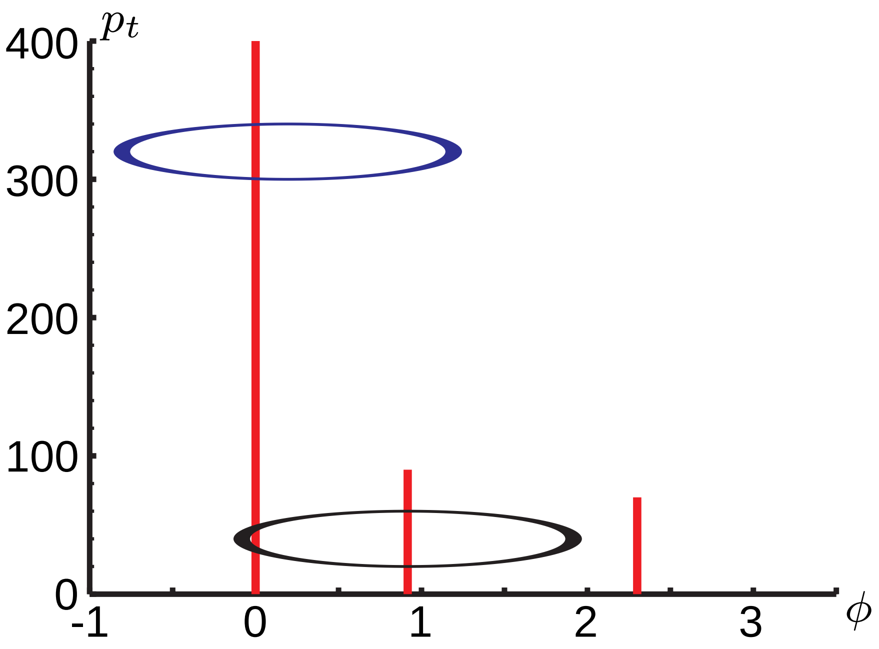
IR unsafety of the Midpoint alg



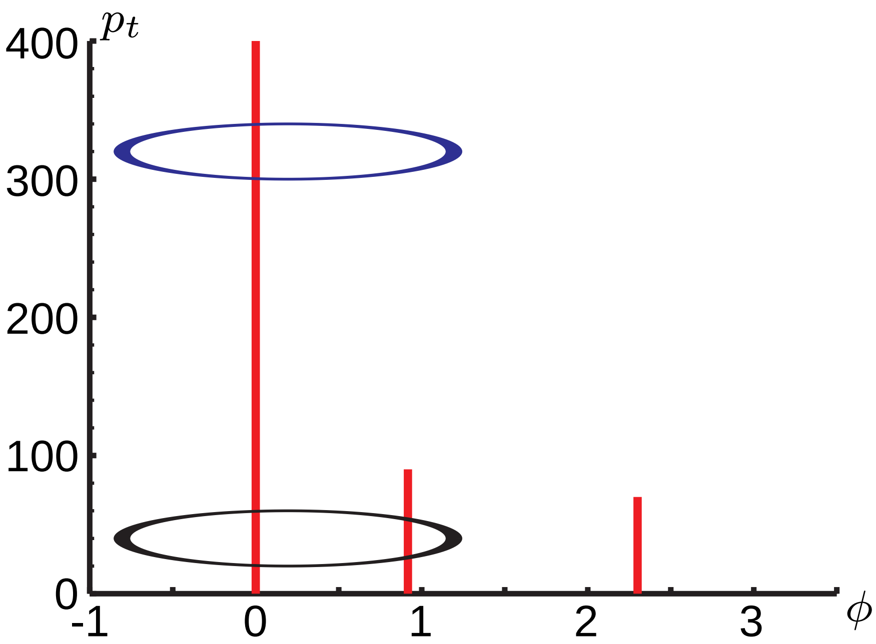
IR unsafety of the Midpoint alg



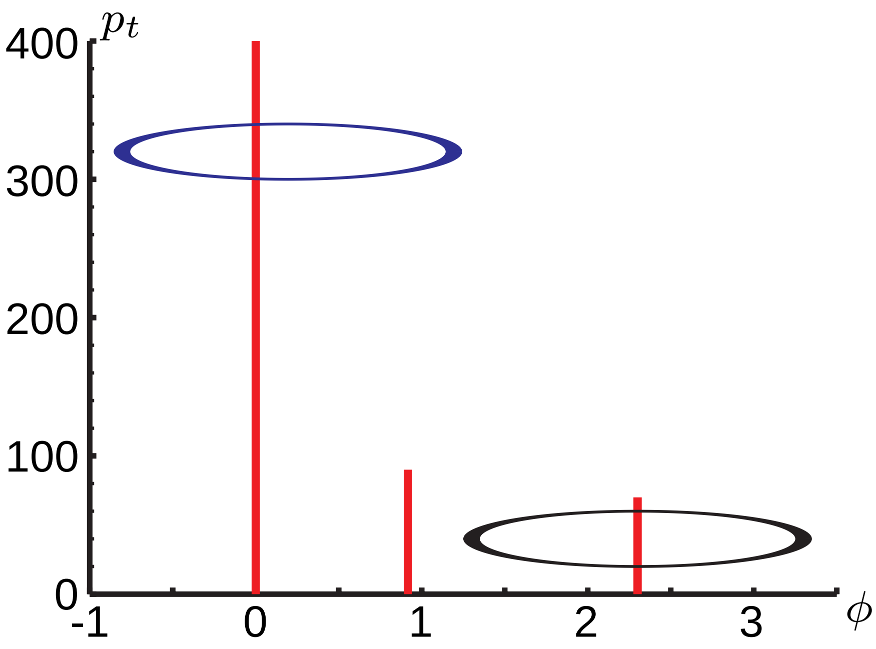
IR unsafety of the Midpoint alg



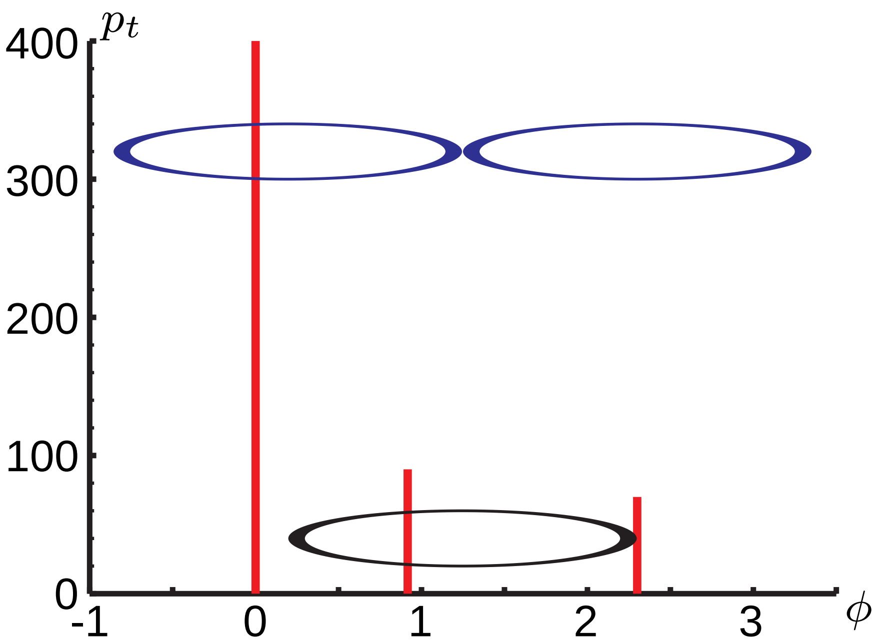
IR unsafety of the Midpoint alg



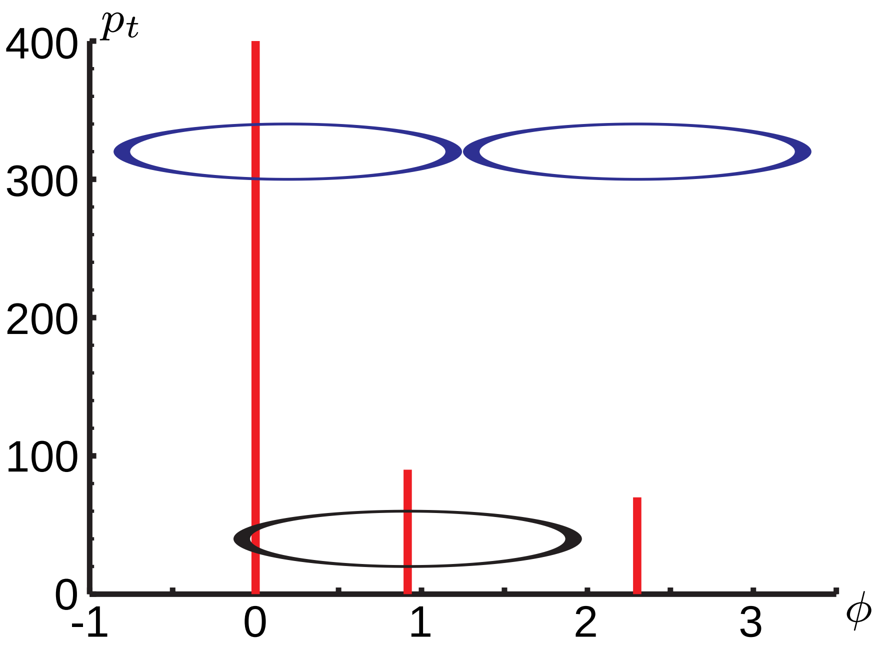
IR unsafety of the Midpoint alg



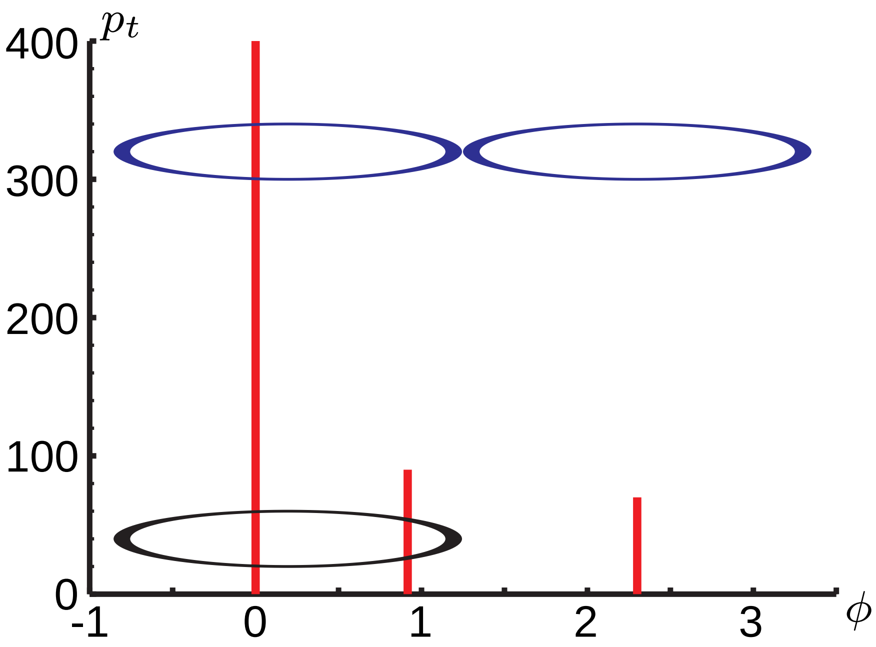
IR unsafety of the Midpoint alg



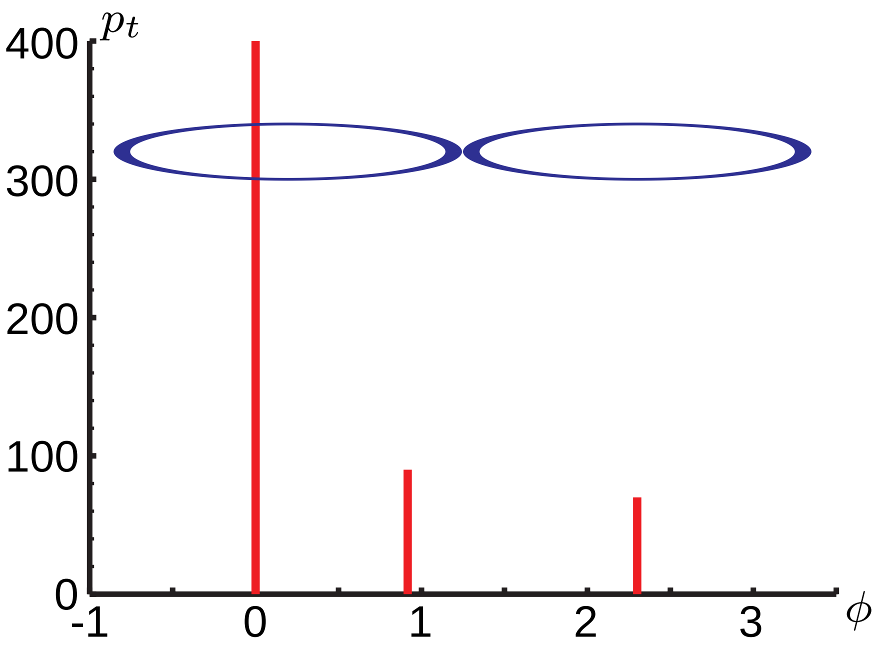
IR unsafety of the Midpoint alg

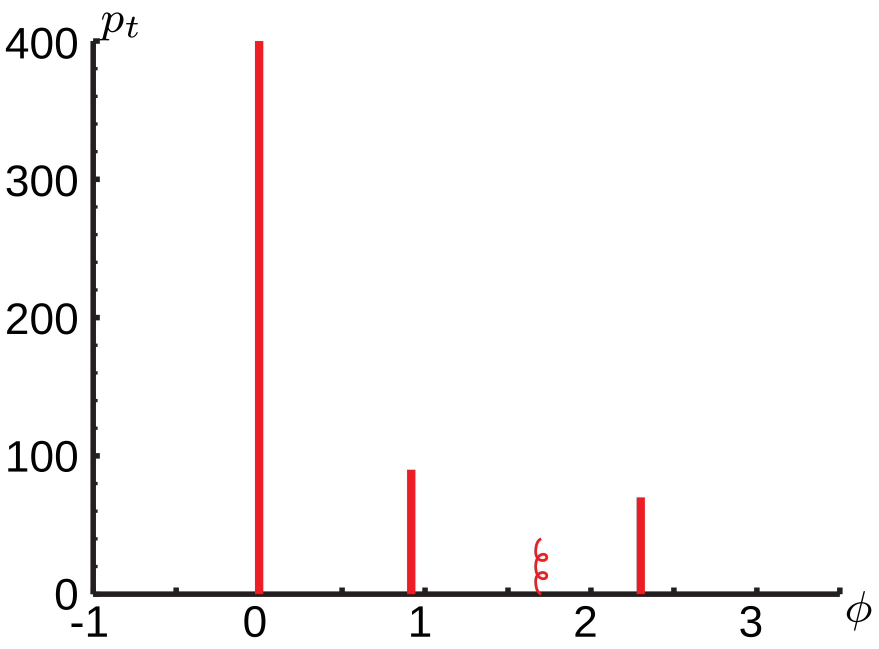


IR unsafety of the Midpoint alg

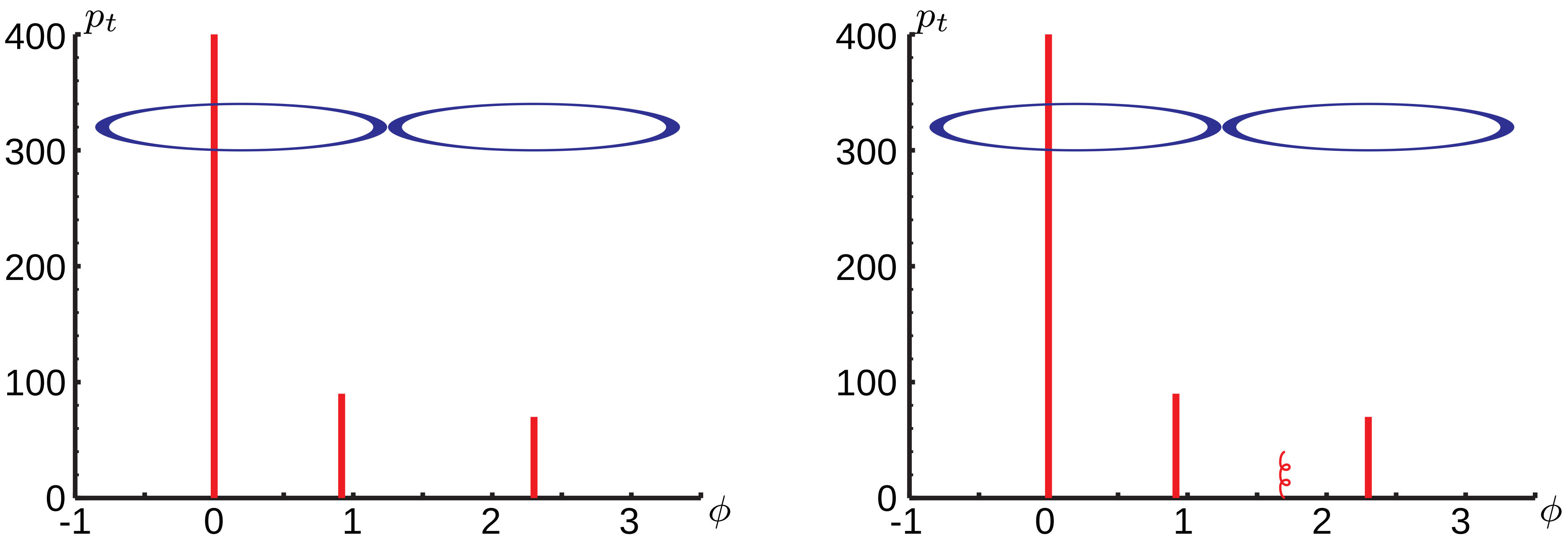


IR unsafety of the Midpoint alg

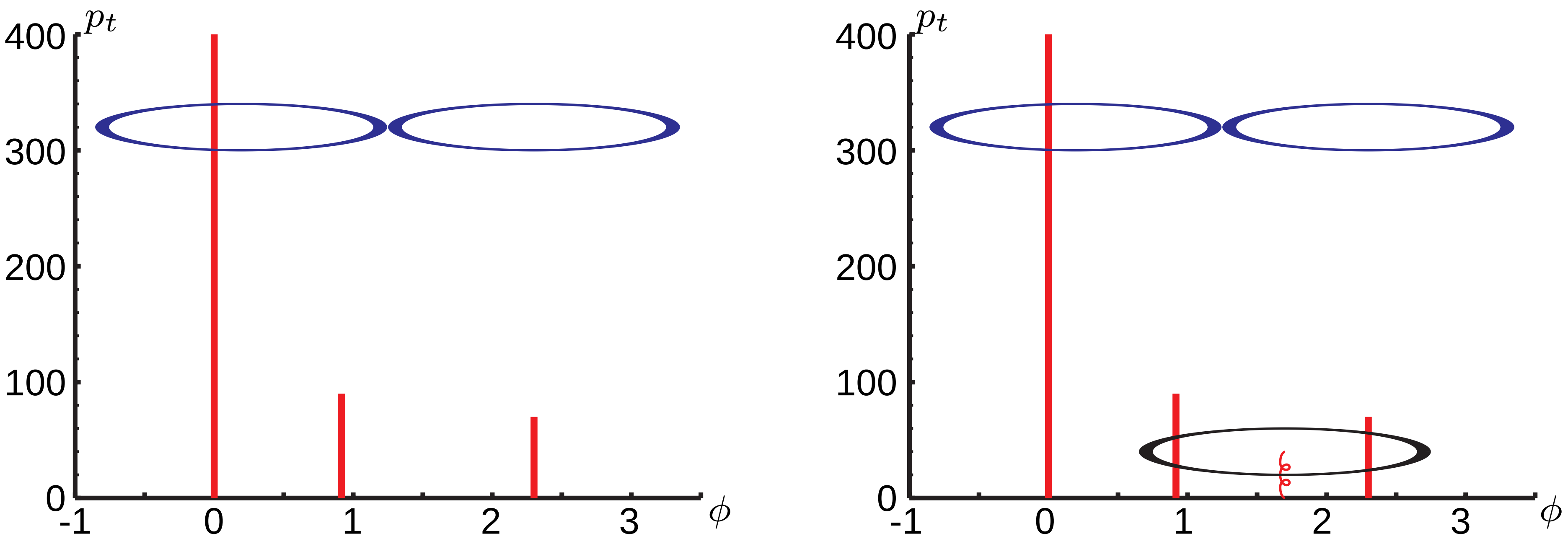




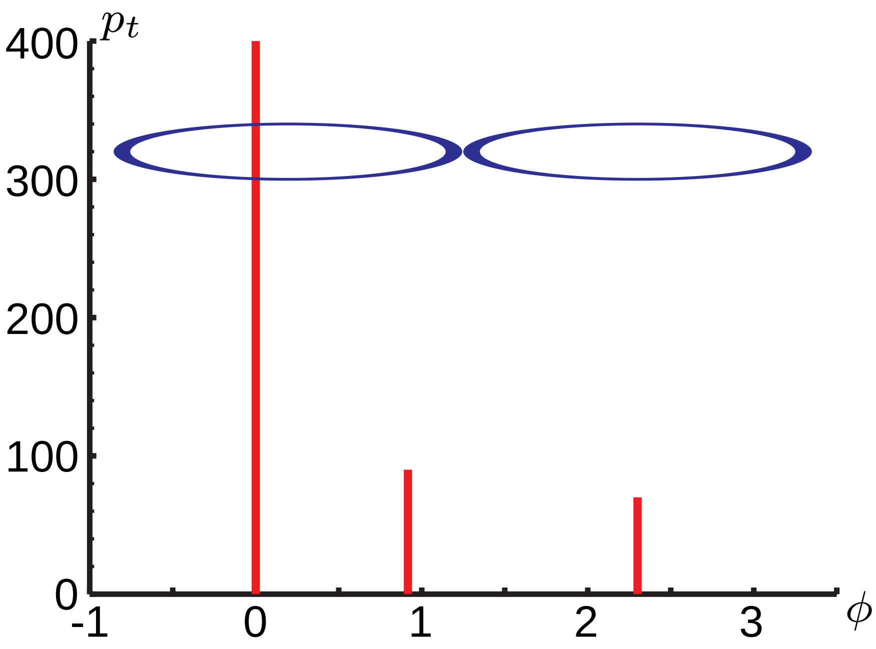
IR unsafety of the Midpoint alg

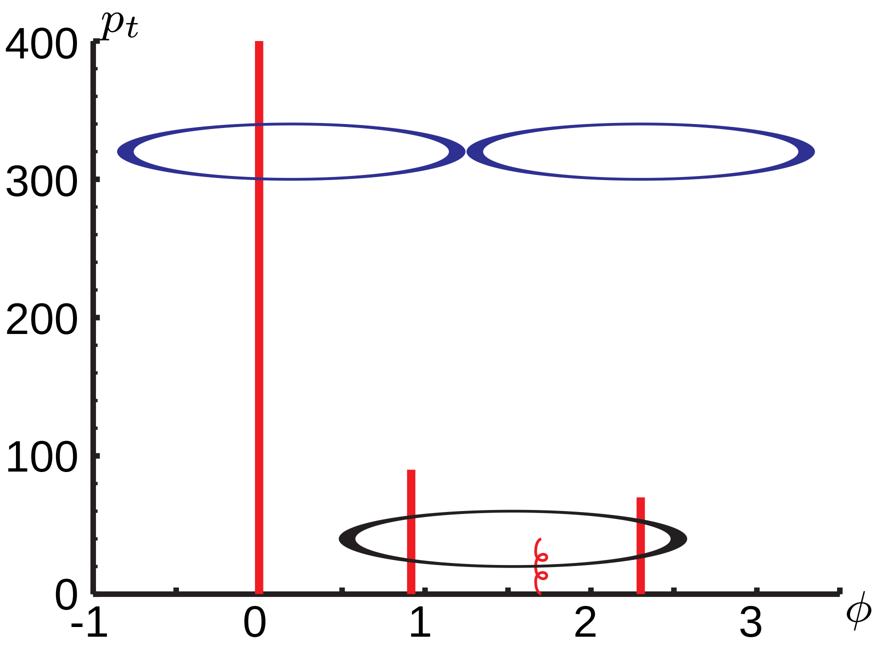


IR unsafety of the Midpoint alg

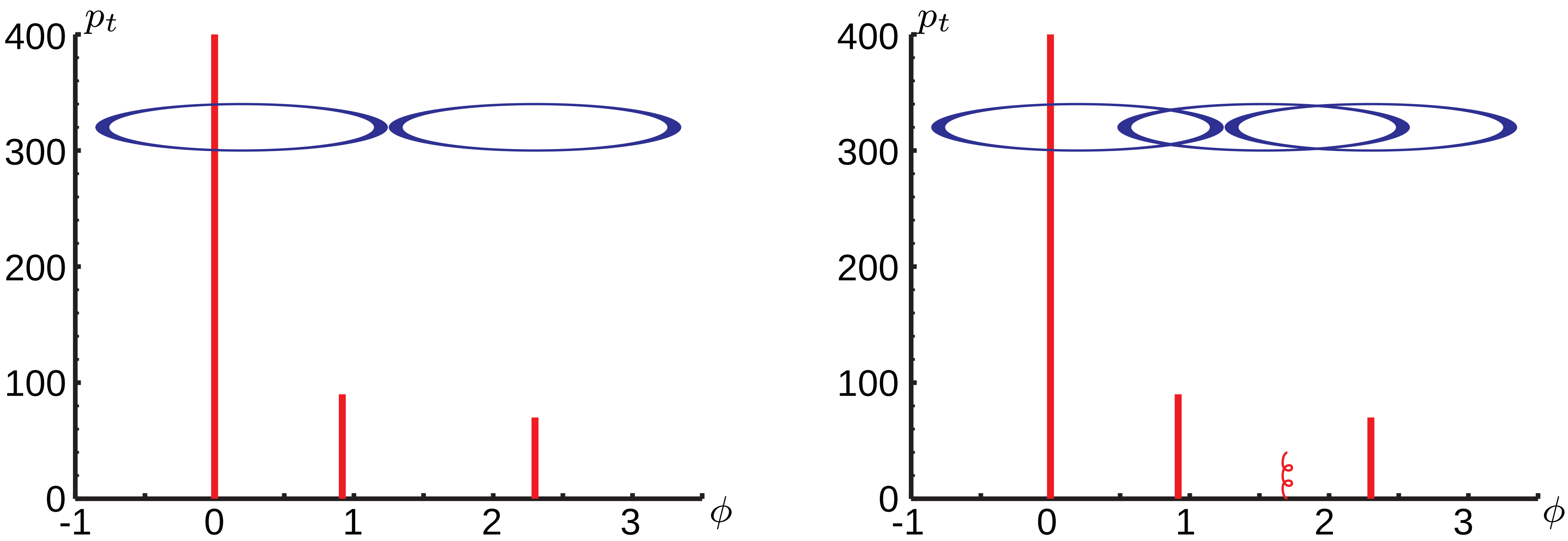


IR unsafety of the Midpoint alg





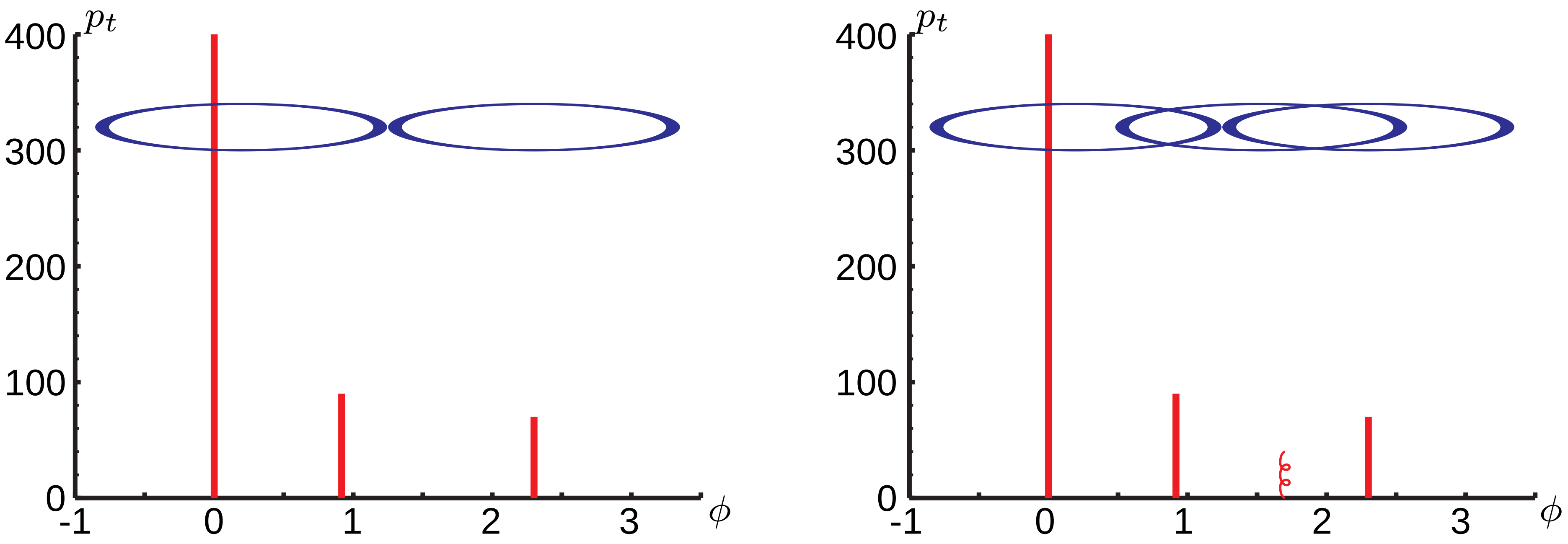
IR unsafety of the Midpoint alg



Stable cones:

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

IR unsafety of the Midpoint alg



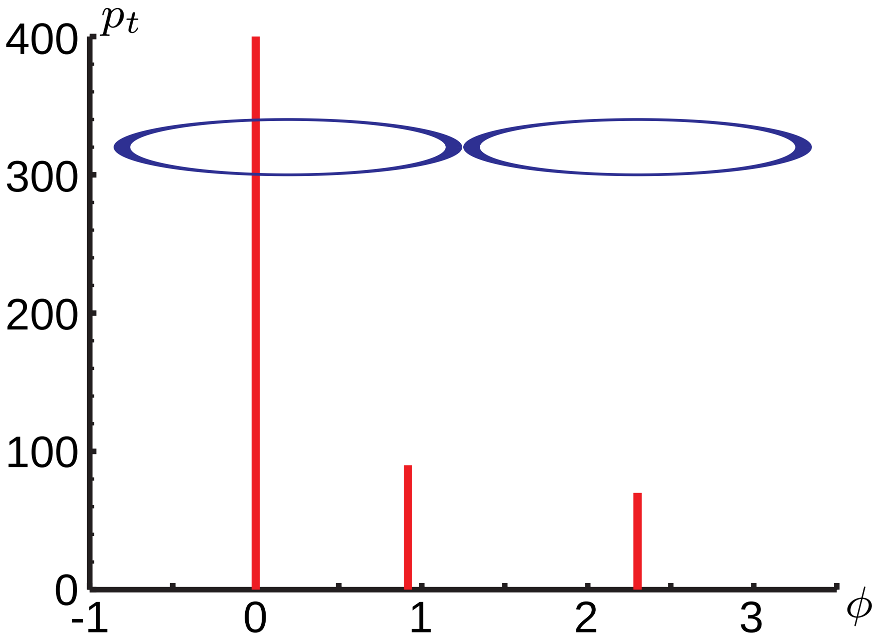
Stable cones:

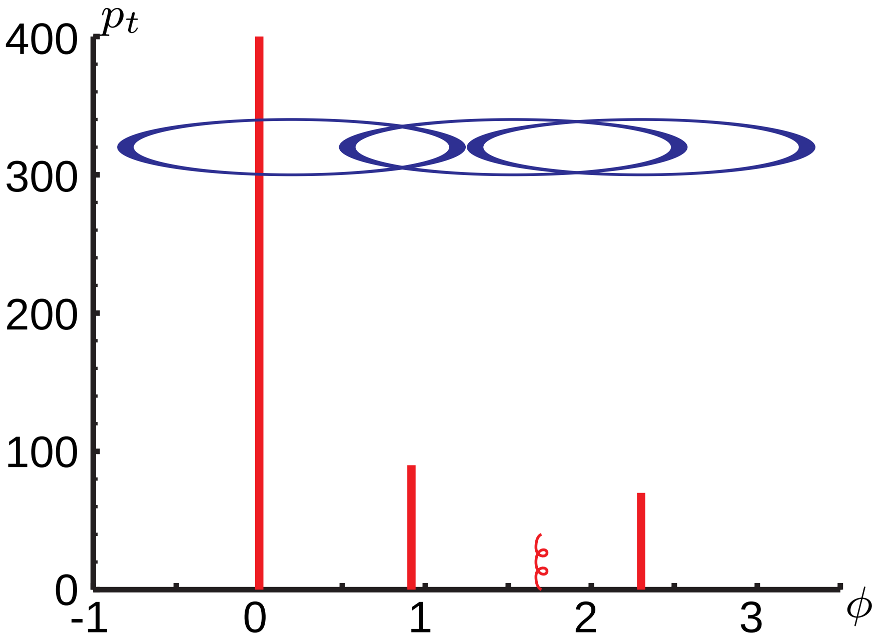
Midpoint:	{1,2} & {3}	{1,2} & {3} & {2,3}

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 ⟶ 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

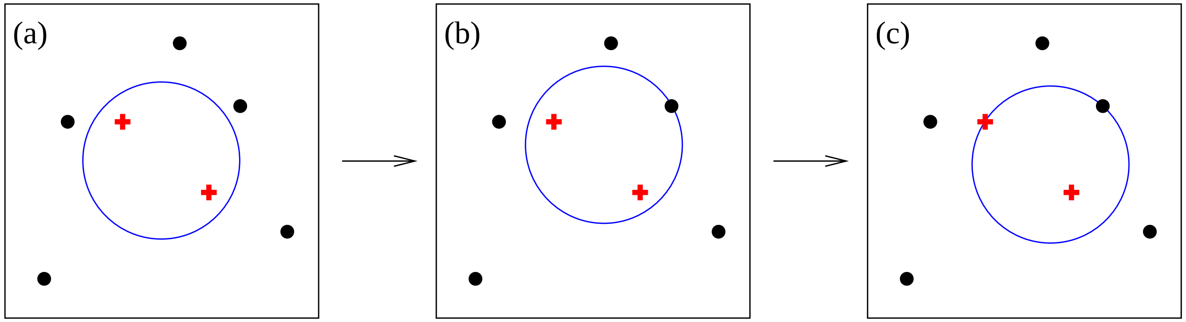
Solution: SISCone

- 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

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: $\mathcal{O}\left(N^2\log(N)\right)$

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

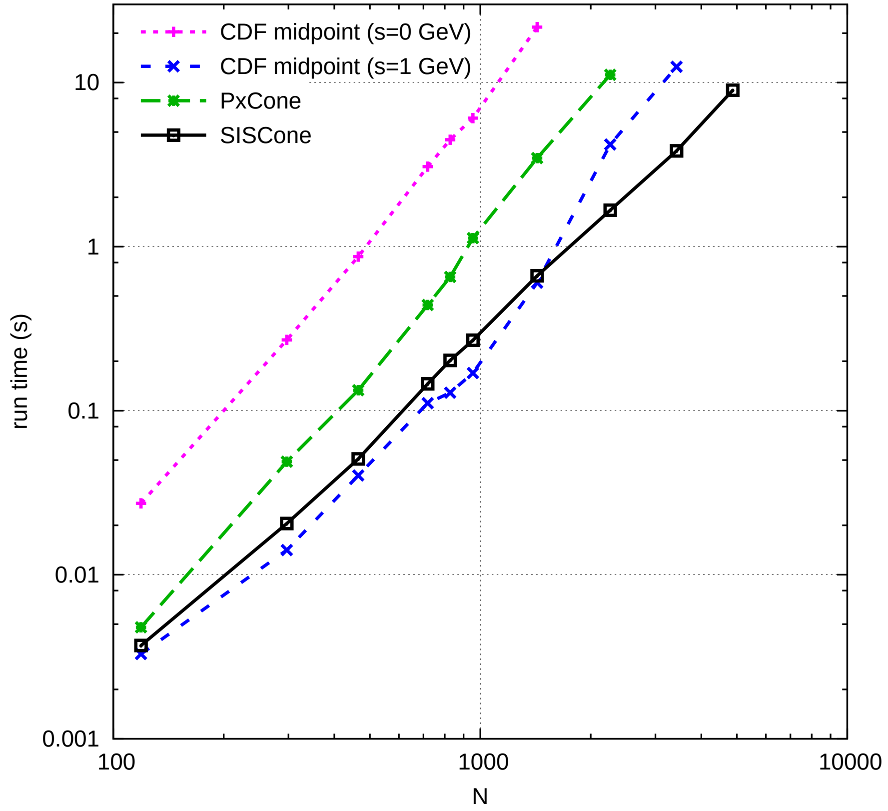
- Complexity: $\mathcal{O}\left(N^3\right)$, with improvements: $\mathcal{O}\left(N^2\log(N)\right)$

$\longrightarrow$ 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
[M.Cacciari, G.Salam, G.S.]; http://www.lpthe.jussieu.fr/∼salam/fastjet

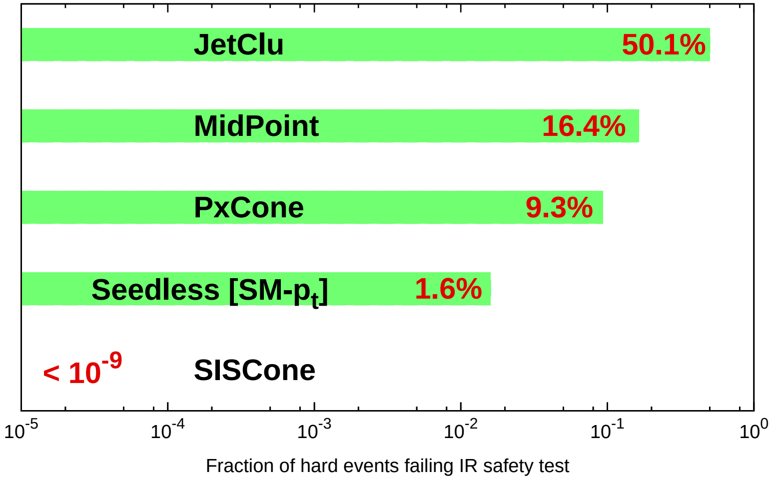
Physical impact

Execution timings:



Random hard & soft partons

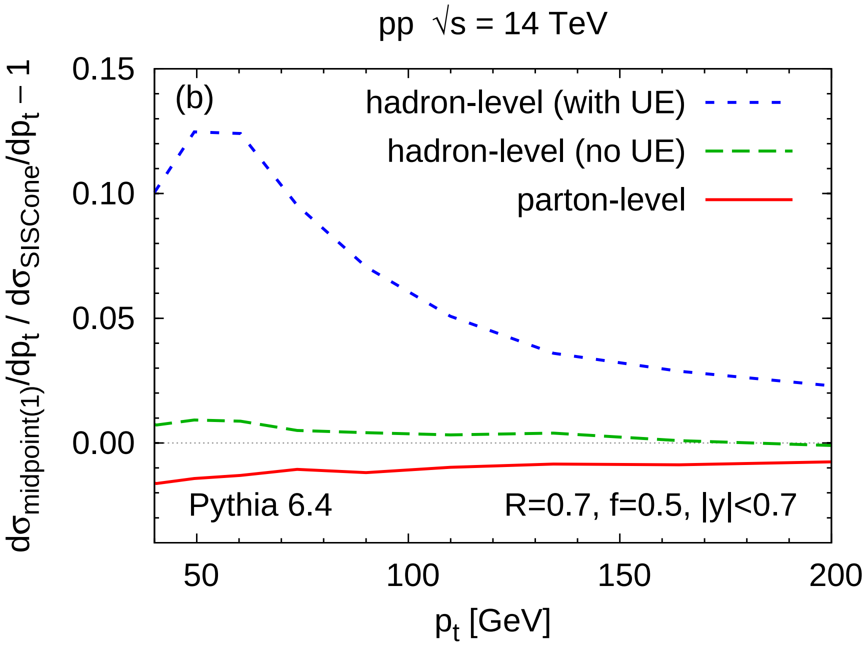
fraction with "hard $\neq$ hard+soft"



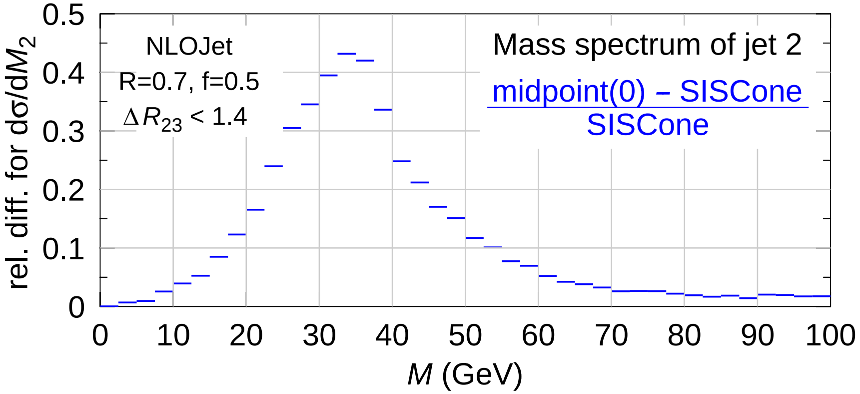
- at least as fast as midpoint cones
- IR safe
  - JetClu,ATLAS cone: 50% failure
  - MidPoint: 15% failure

Physical impact (2)

Inclusive (midpoint/SISCone-1)

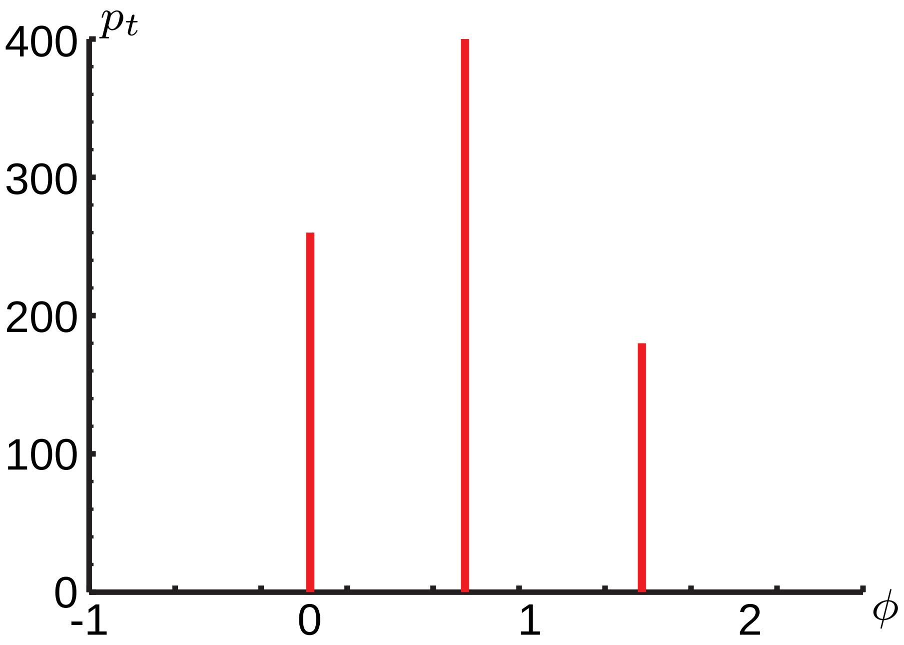


Masses in 3-jet events

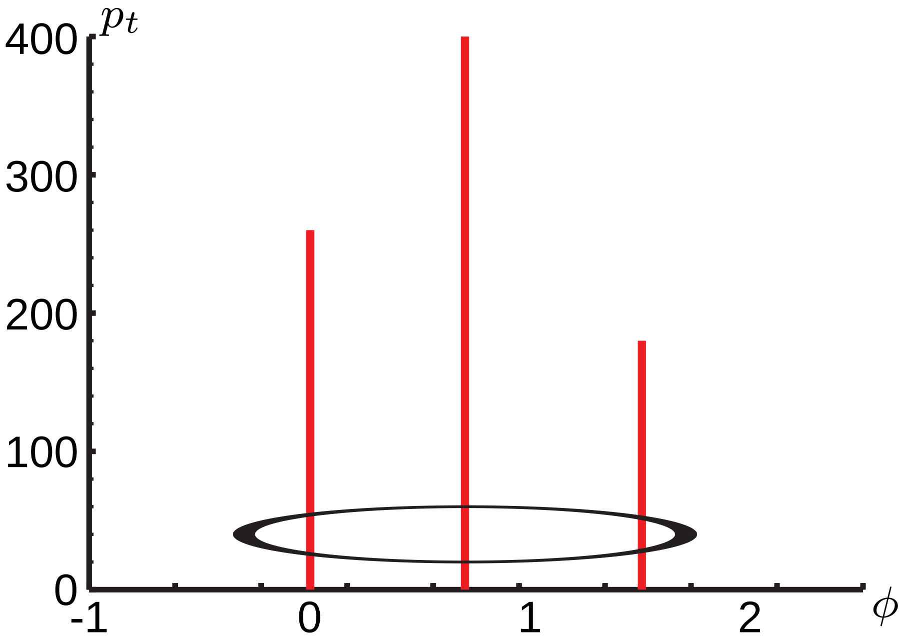


- Inclusive cross-section:
  - Effect of a few percents
  - Less sensitivity to the UE
- More exclusive processes: effects $\sim 45\%$ (Important for LHC!)

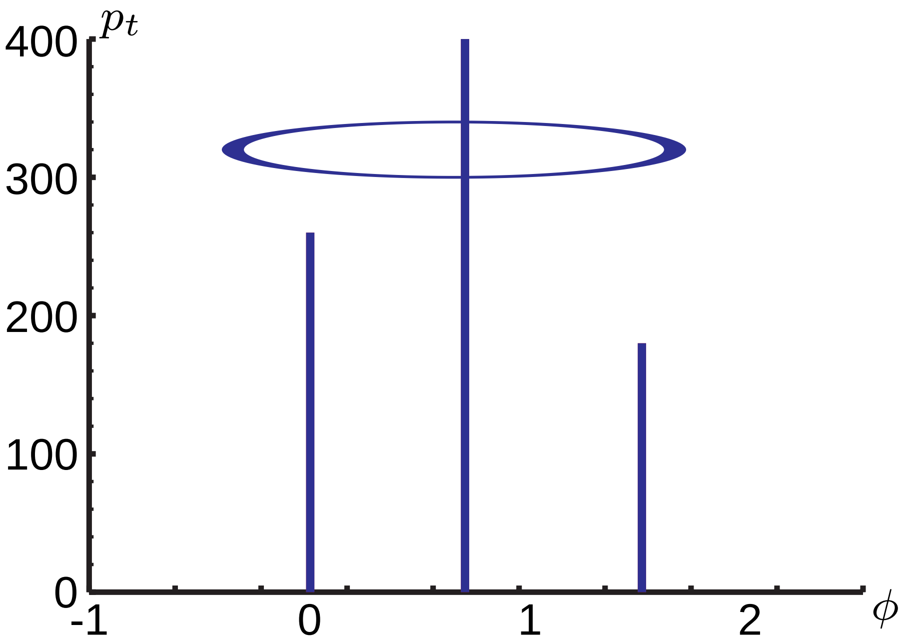
Coll. unsafety of the iterative cone



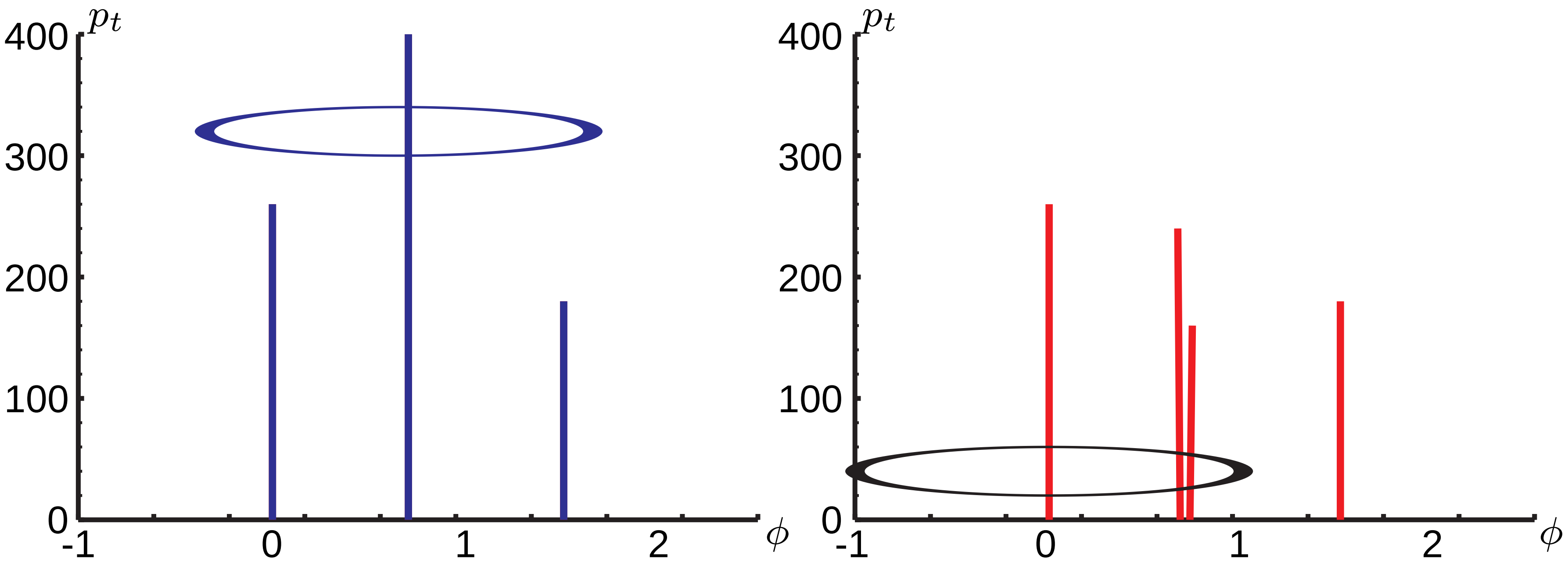
Coll. unsafety of the iterative cone



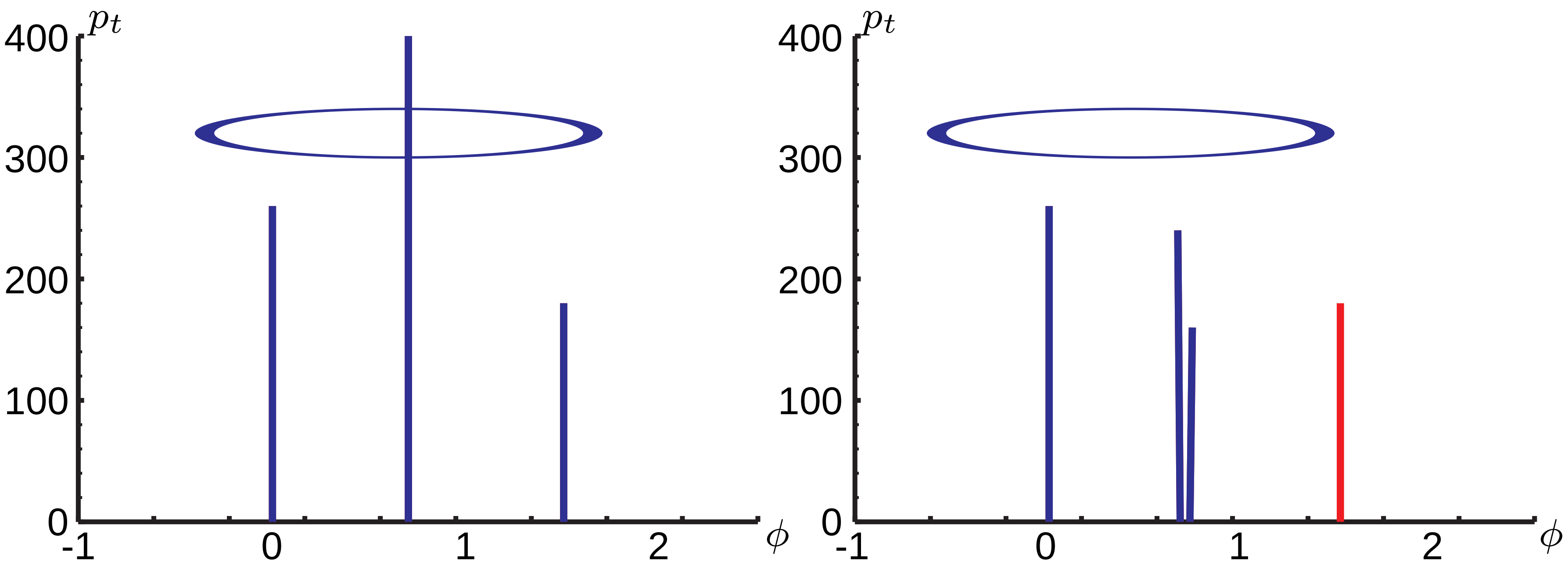
Coll. unsafety of the iterative cone



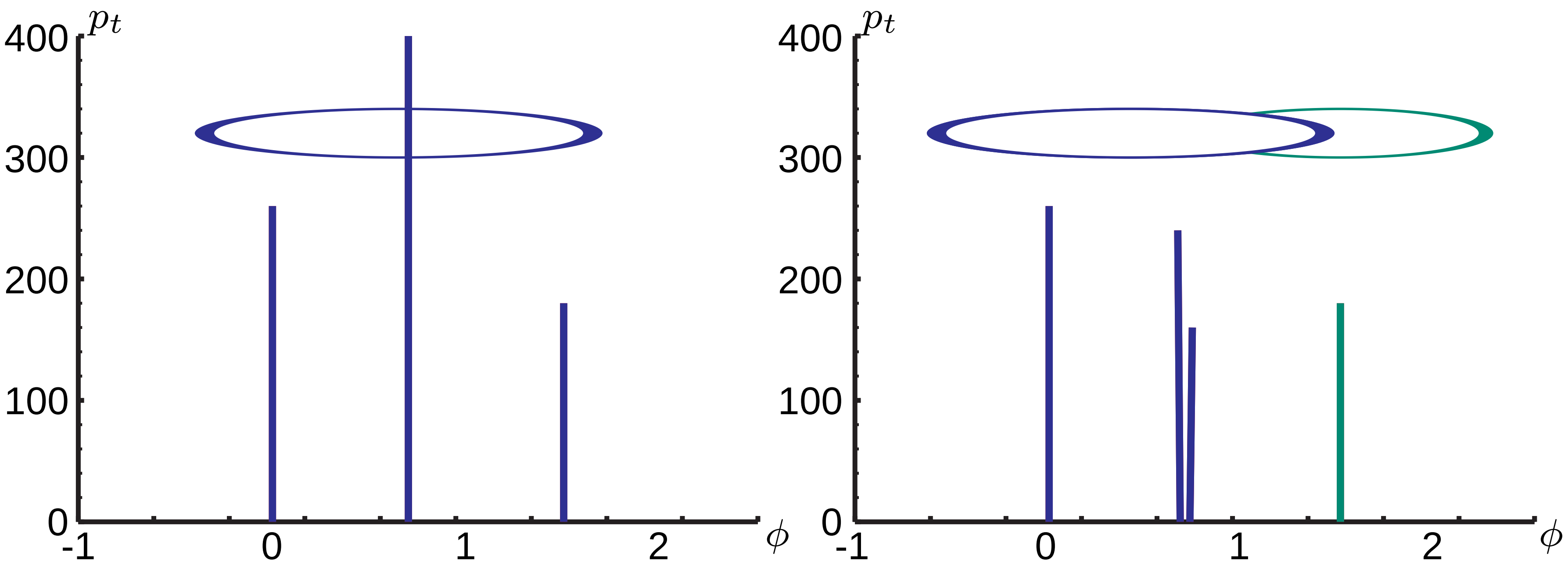
Coll. unsafety of the iterative cone



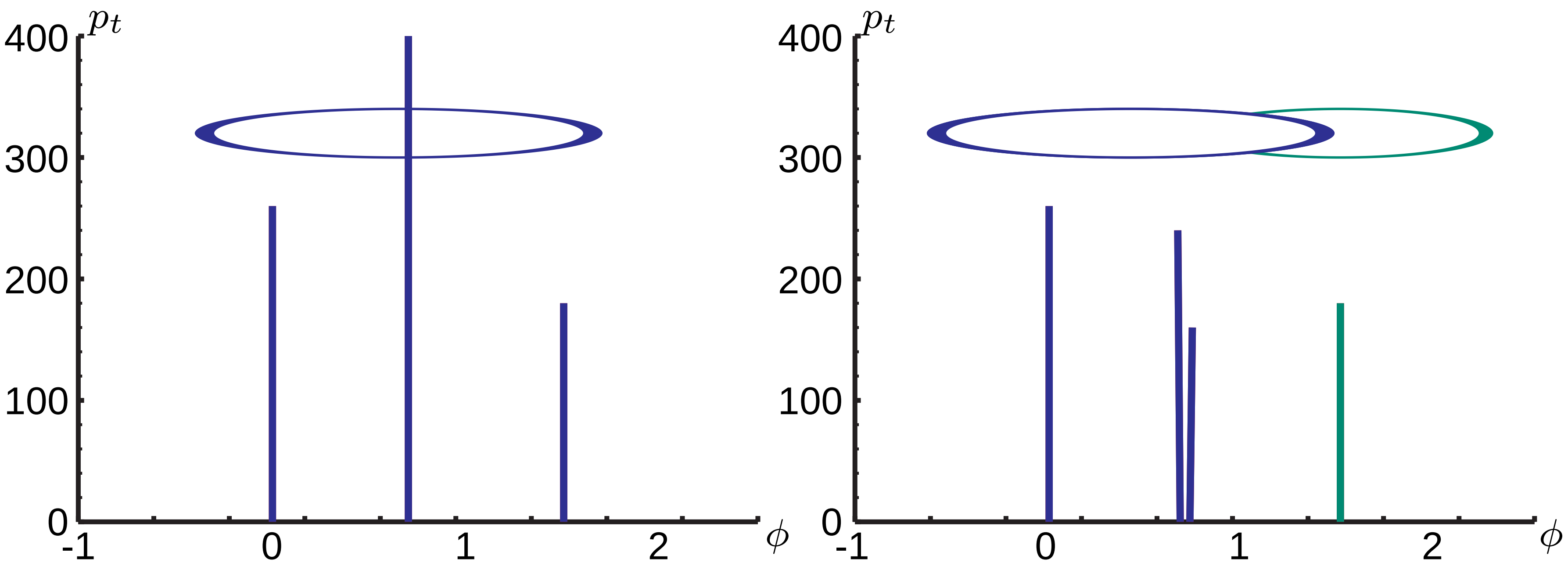
Coll. unsafety of the iterative cone



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**

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

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:

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

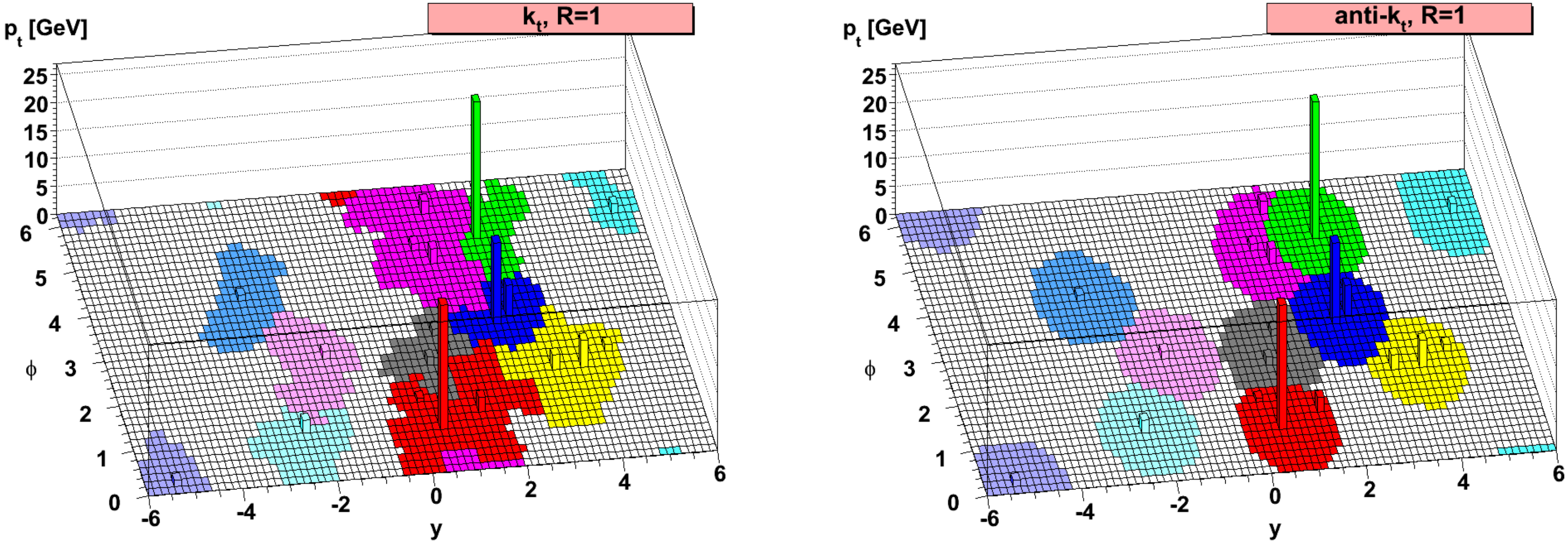
- $p = 1$: k_t algorithm
- $p = 0$: Aachen/Cambridge algorithm
- $p = -1$: anti-k_t algorithm [M.Cacciari, G.Salam, G.S.,JHEP 04 (08) 063]

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

anti-k_t

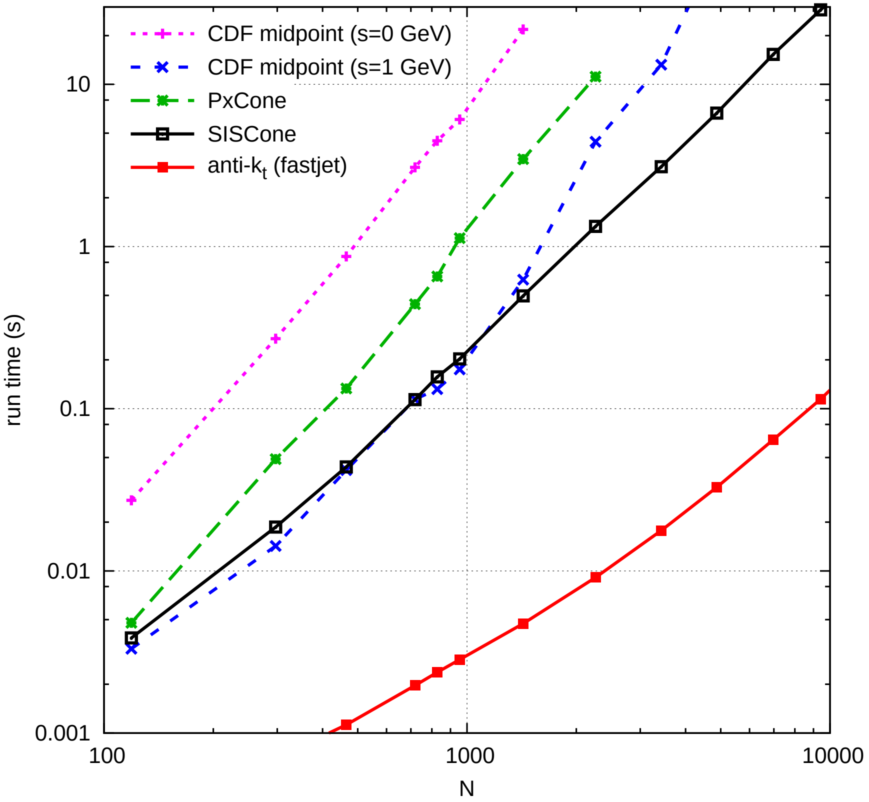
Hard event + homogeneous soft background



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])

IRC safety is mandatory

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

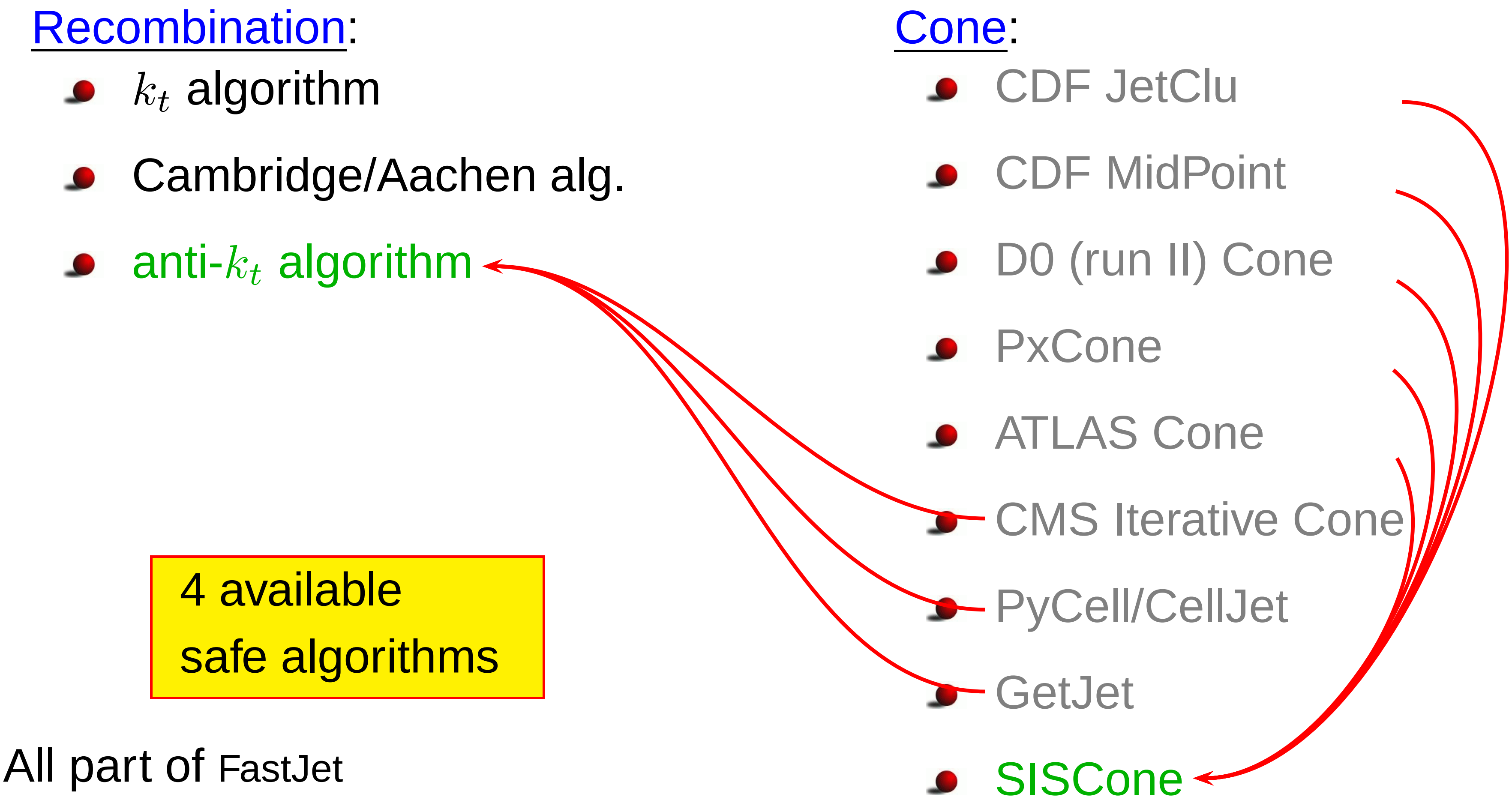
Observable	1st miss cones at	Last meaningful order
Inclusive jet cross section	NNLO	NLO
3 jet cross section	NLO	LO (NLO in NLOJet)
$W/Z/H$ + 2 jet cross sect.	NLO	LO (NLO in MCFM)
jet masses in 3 jets	LO	none (LO in NLOJet)

$\Rightarrow$ The IRC-unsafety issue will matter at LHC

+ We do not want the theoretical efforts to be wasted

Note: 1 order worse for JetClu of the ATLAS Cone!

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 part of FastJet

[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

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

Part 1

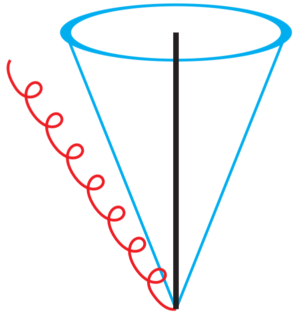
Jets in pp collisions

Choosing the adapted jet definition

Underlying physics

Competition between

- catching perturbative radiation



Out-of-jet radiation:

$$\sim \int_R \frac{d\theta}{\theta} \sim \log(1/R)$$

- not catching soft background radiation

Soft contents $\sim$ jet area $\sim R^2$

more detailed analytic computation in progress...
only numerical results at present time

Sample processes to study

We analyse 3 processes:

- $Z' \to q\bar{q} \to$2 jets: (ficticious narrow Z')

  simple environment: identify 2 jets and reconstruct $M_{Z'}$
  source of monochromatic quark jets
  scale dependence: mass of the Z' between 100 GeV and 4 TeV

- $H \to gg \to$2 jets: (ficticious narrow Higgs)

  simple environment: identify 2 jets and reconstruct M_H
  source of monochromatic gluon jets
  scale dependence: mass of the Higgs between 100 GeV and 4 TeV

- $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

figure of merit for quality measure

Measure of the jet reconstruction efficiency

- Forget about measures related to parton-jet matching, use the reconstructed mass peak
- Forget about fits depending on the shape of the peak
- we shall maximise the signal over background ratio ($S/\sqrt{B}$):

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

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

- it intuitively does what it should
- for a constant background,

$$\frac{Q^w_{f=z}(JA_1,R_1)}{Q^w_{f=z}(JA_2,R_2)} = \frac{B_{JA_1,R_1}}{B_{JA_2,R_2}} = \left[\frac{(S/\sqrt{B})_{JA_1,R_1}}{(S/\sqrt{B})_{JA_2,R_2}}\right]^{-2}$$

smaller width $\equiv$ better signal-to-background ratio

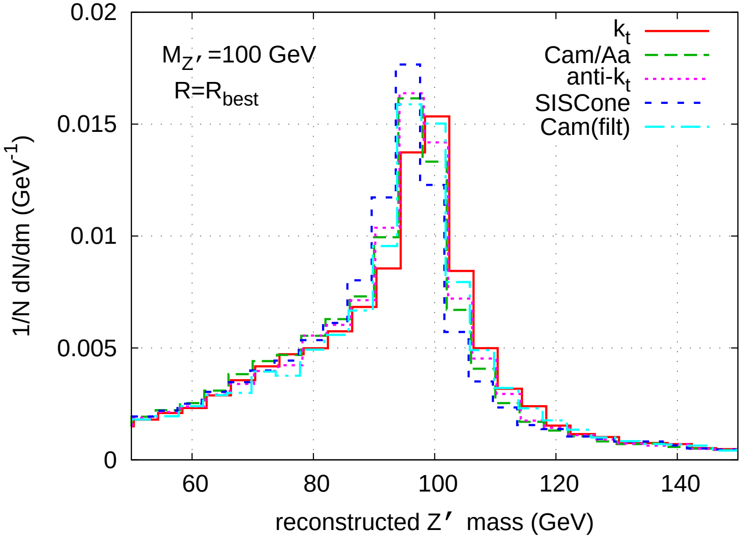
- we can associate an effective luminosity ratio

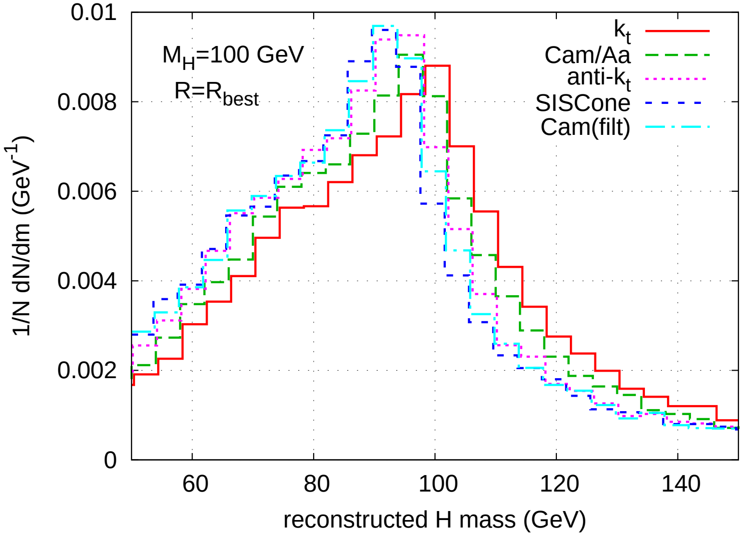
$$\rho_{\mathcal{L}} = \frac{\mathcal{L}_1}{\mathcal{L}_2} = \left[\frac{(S/\sqrt{B})_{JA_1,R_1}}{(S/\sqrt{B})_{JA_2,R_2}}\right]^{2} = \frac{Q^w_{f=z}(JA_2,R_2)}{Q^w_{f=z}(JA_1,R_1)}$$

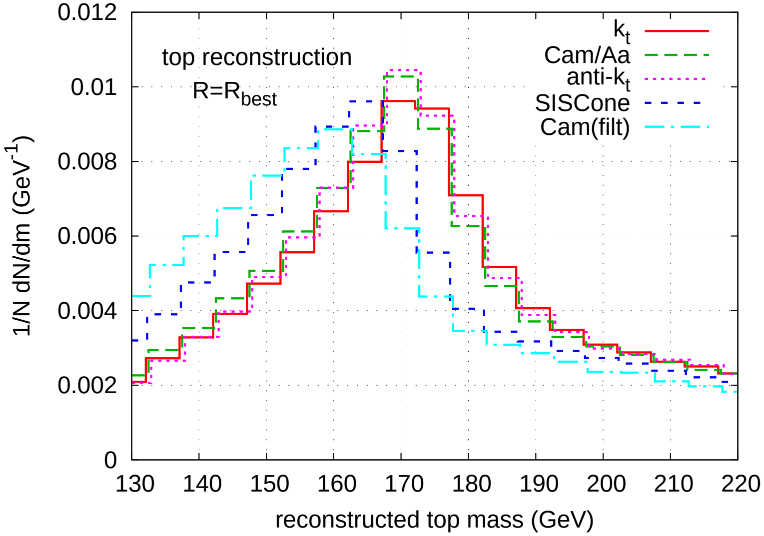
e.g. if $Q^w_{f=z}(JA_2,R_2) = 2Q^w_{f=z}(JA_1,R_1)$, (JA_2,R_2) will need twice the luminosity of (JA_1,R_1) to achieve the same discriminative power.

results (1)

we see peaks...

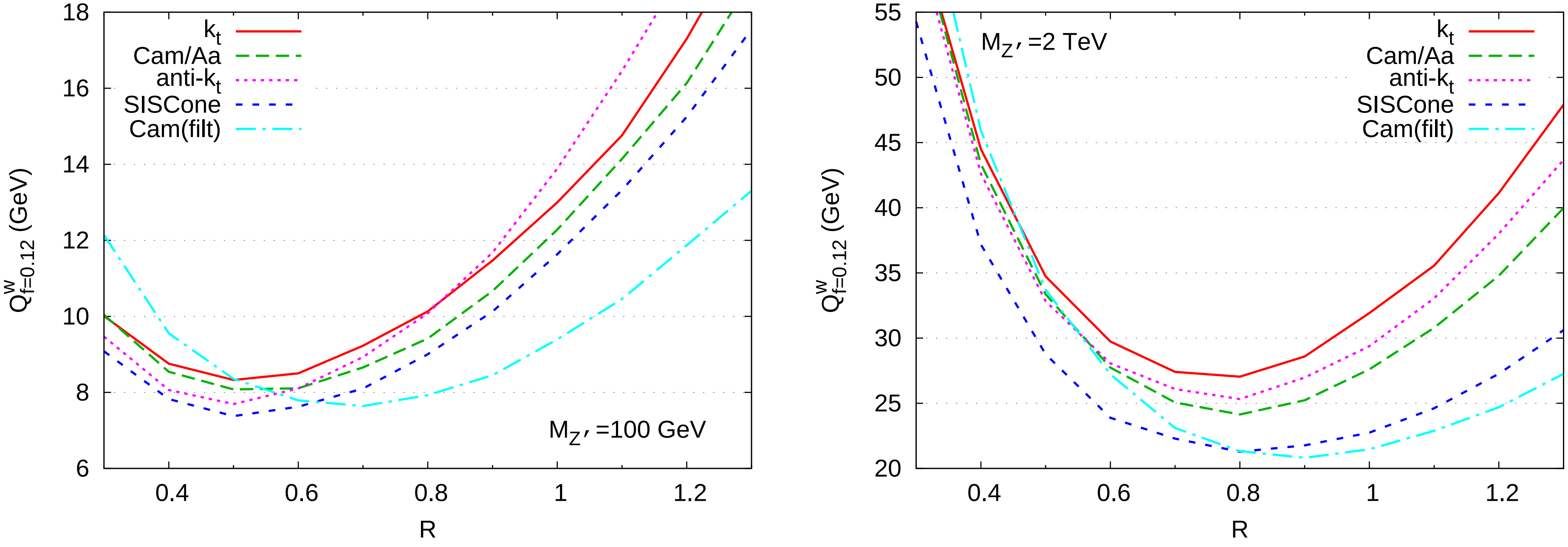






results (1)

Message 1: there is a strong R dependence



At 100 GeV,

using $R = 0.8$ instead of $R = 0.5$ means a discr. power loss of 20% ($\rho_{\mathcal{L}} \approx 0.8$)

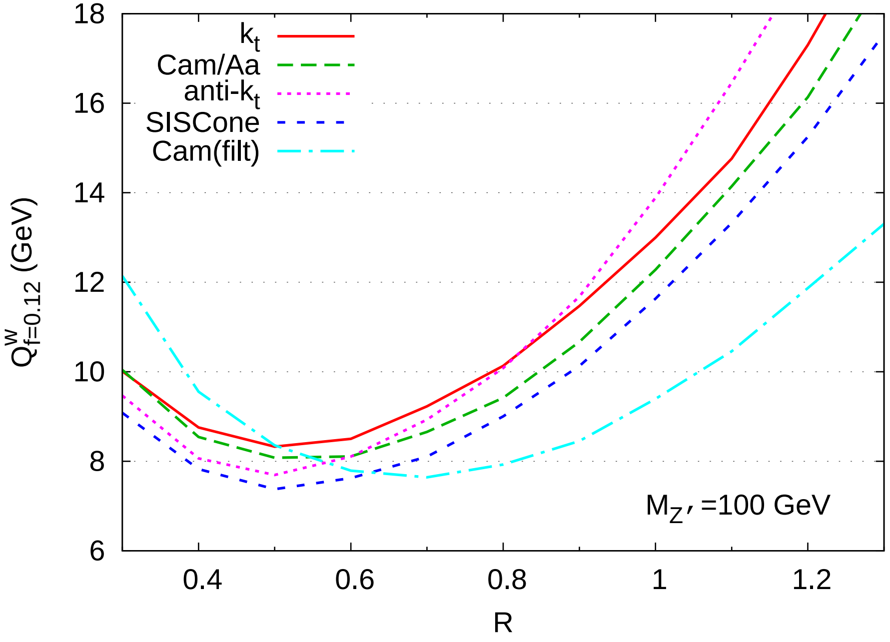
$R = 1$ 40% ($\rho_{\mathcal{L}} \approx 0.6$)

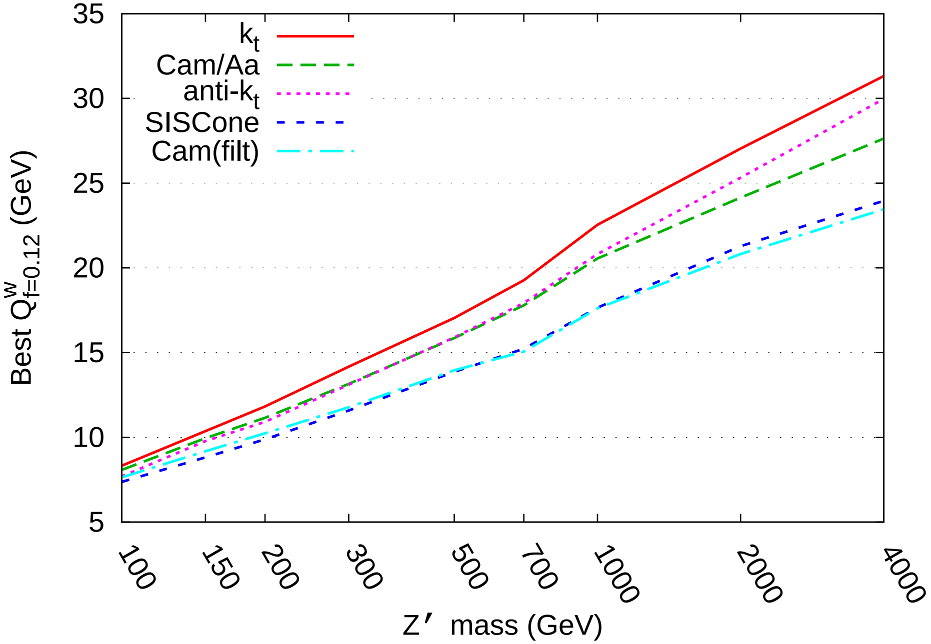
At 2 TeV,

using $R = 0.5$ instead of $R = 0.8$ means a discr. power loss of 20% ($\rho_{\mathcal{L}} \approx 0.8$)

results (1)

Message 2: SISCone and Cam+filt do a slightly better job





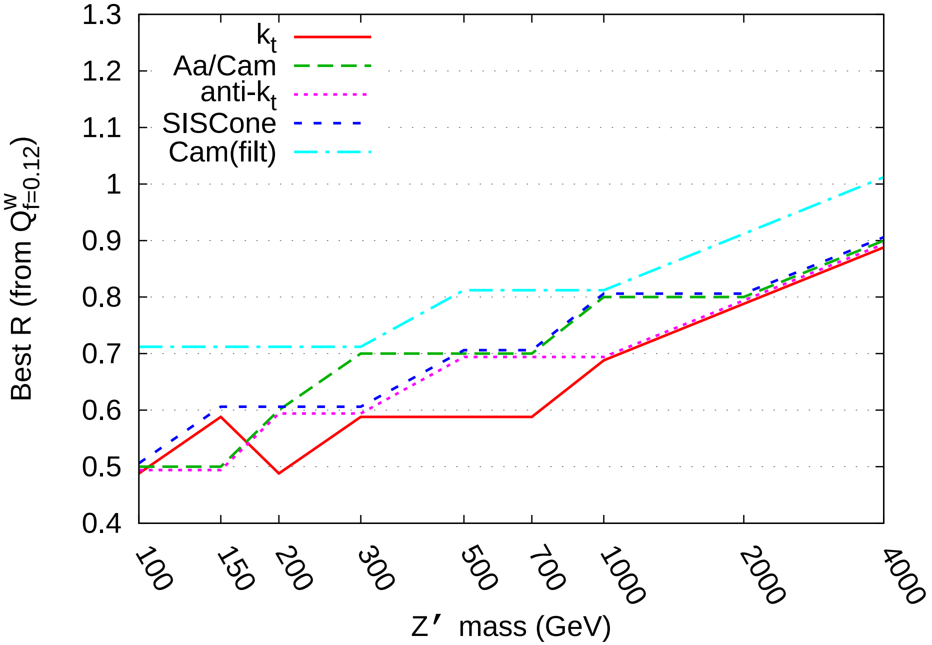
Using k_t instead of SISCone means a discr. power loss of

15% at 100 GeV ($\rho_{\mathcal{L}} \approx 0.85$)

20% at 2 TeV ($\rho_{\mathcal{L}} \approx 0.8$)

results (1)

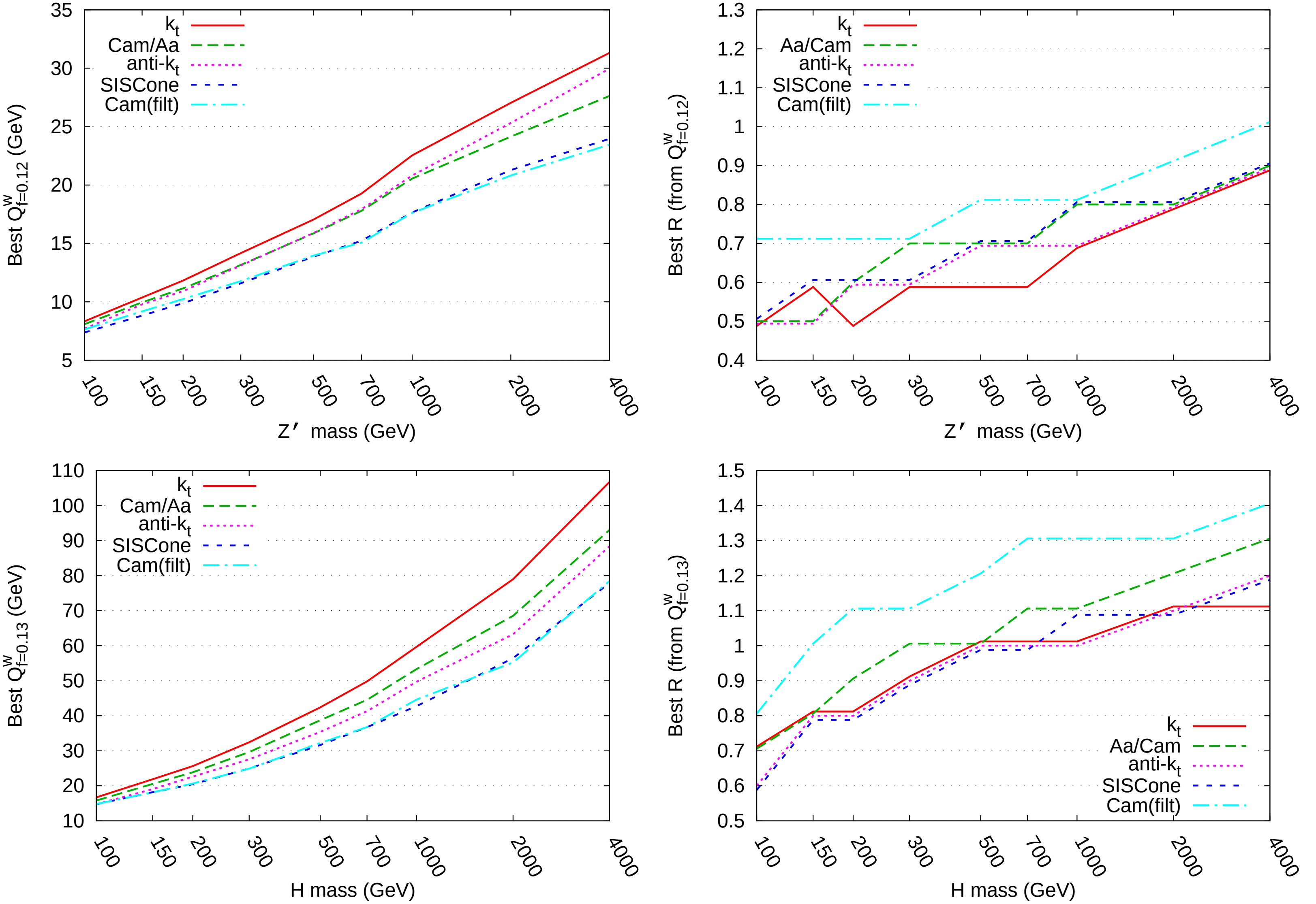
Message 3: The parameters vary with the scale



The preferred value for R increase with the mass scale (typically like $\log(M)$)

results (1)

Message 4: same for the gluon jets, though with a larger R



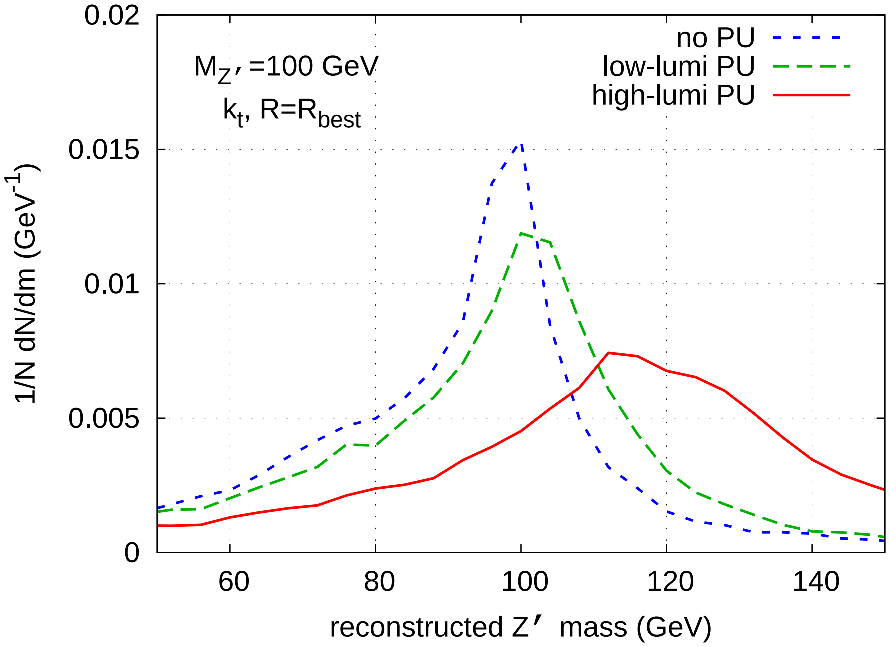
Part 1

Jets in pp collisions

Subtracting pileup background using jet areas

Need for subtraction

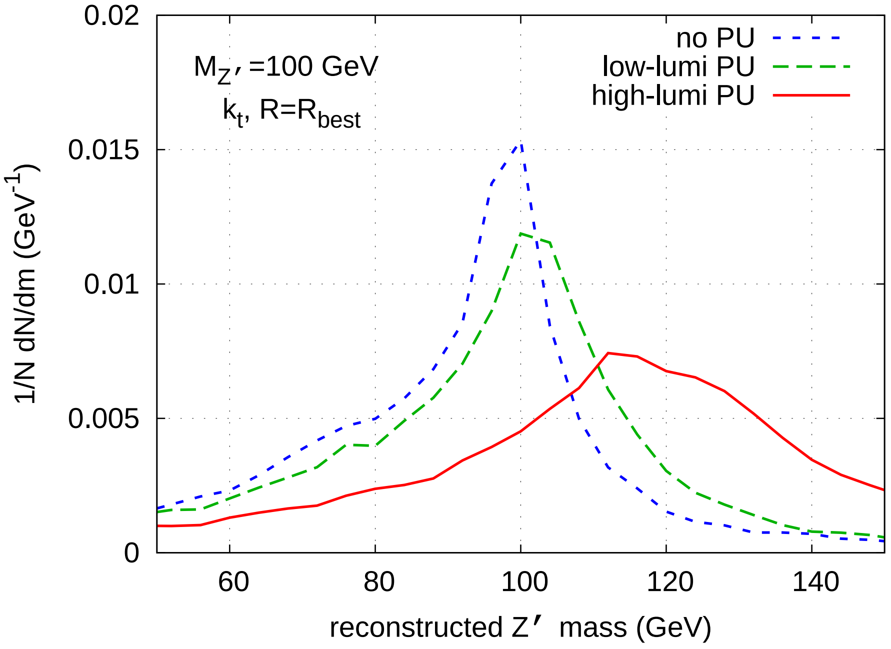
Pileup $\approx$ uniform soft background that shifts jets to higher p_t



... that needs to be subtracted!

Need for subtraction

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



... that needs to be subtracted!

$\Rightarrow$ Using jet areas!

Area definition

- Idea: add soft particles (ghosts) and look in which jets they are caught

jet area = region where it catches ghosts

Area definition

- Idea: add soft particles (ghosts) and look in which jets they are caught

jet area = region where it catches ghosts

- 2 definitions
  - Passive area

    add one ghost and look where it ends. repeat to cover the (y, ϕ) plane
  - Active area

    add a large amount of ghosts and cluster everything

    also gives purely ghosted jets

Area definition

- Idea: add soft particles (ghosts) and look in which jets they are caught

jet area = region where it catches ghosts

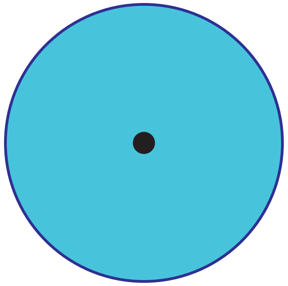
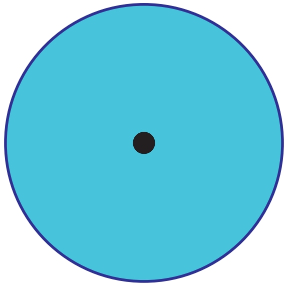
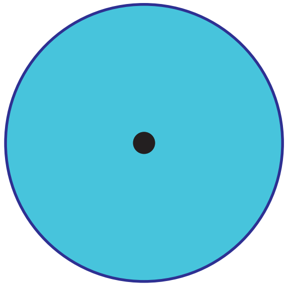
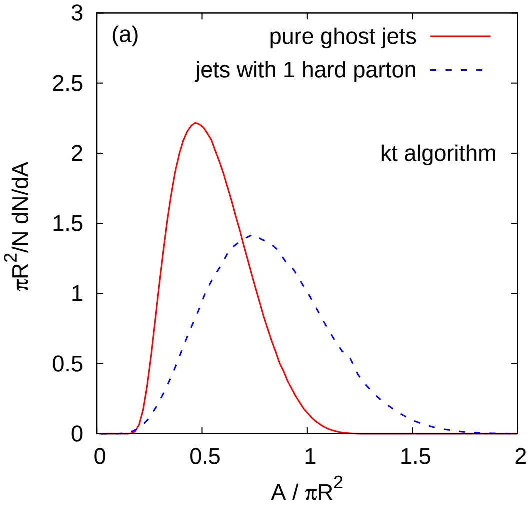
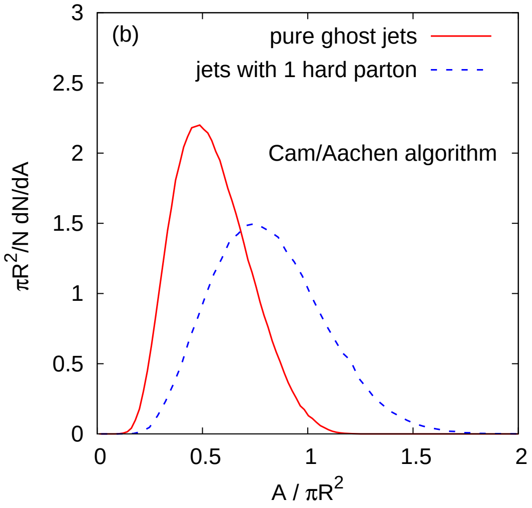
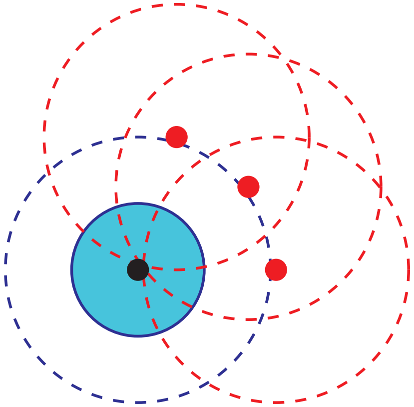
- 2 definitions
  - Passive area

    add one ghost and look where it ends. repeat to cover the (y, ϕ) plane
  - Active area

    add a large amount of ghosts and cluster everything

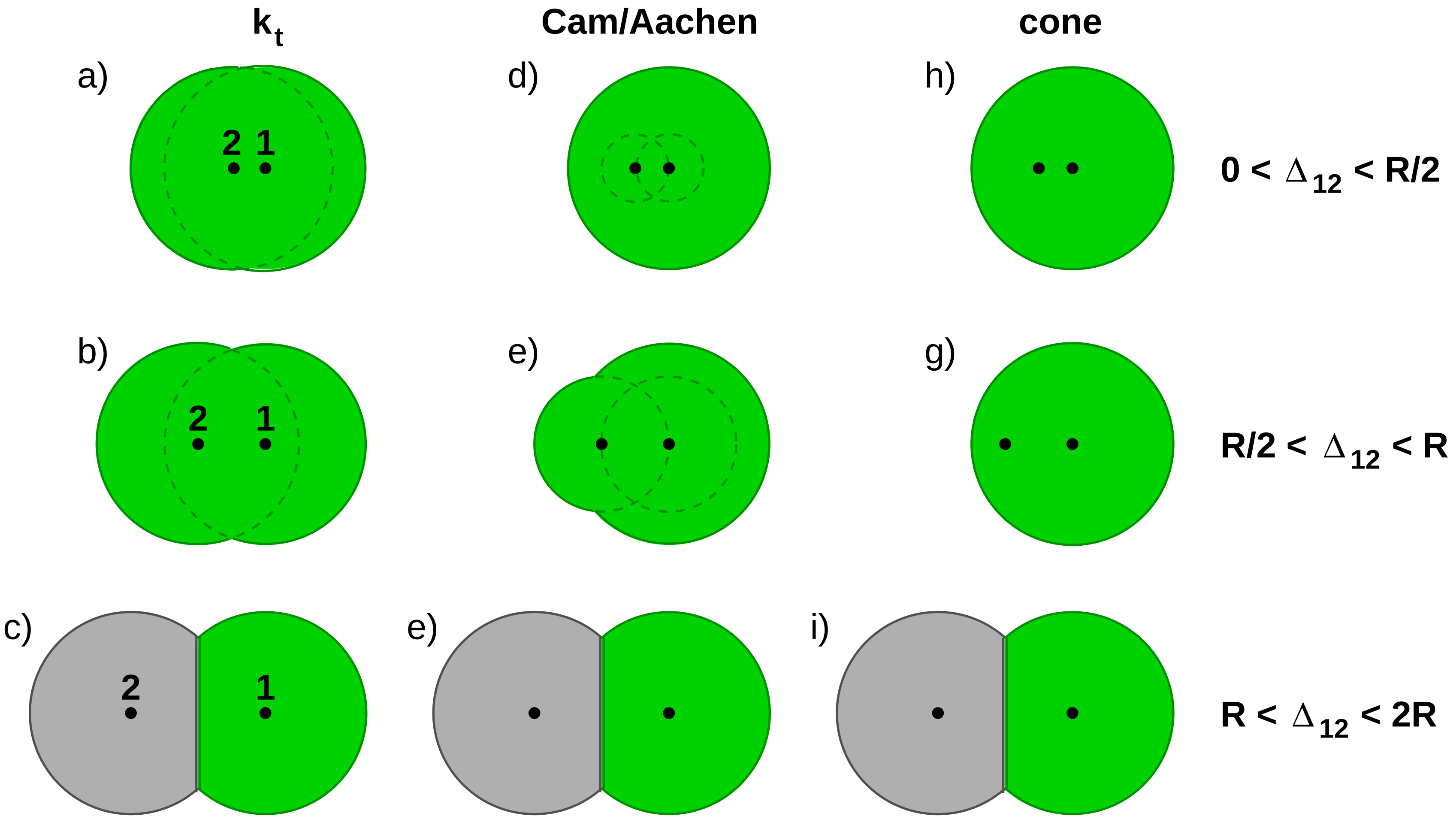
    also gives purely ghosted jets
- Both definitions agree for dense events
- Both practical and tractable analytically

Examples: 1-particle cases

	k_t	Aac/Cam	cone
Passive	 πR^2	 πR^2	 πR^2
Active	 $\frac{A_{\text{hard}}}{\pi R^2} \approx 0.812 \pm 0.277$ $\frac{A_{\text{ghost}}}{\pi R^2} \approx 0.554 \pm 0.174$	 $\frac{A_{\text{hard}}}{\pi R^2} \approx 0.814 \pm 0.261$ $\frac{A_{\text{ghost}}}{\pi R^2} \approx 0.551 \pm 0.176$	 $\frac{A_{\text{hard}}}{\pi R^2} = 0.25$

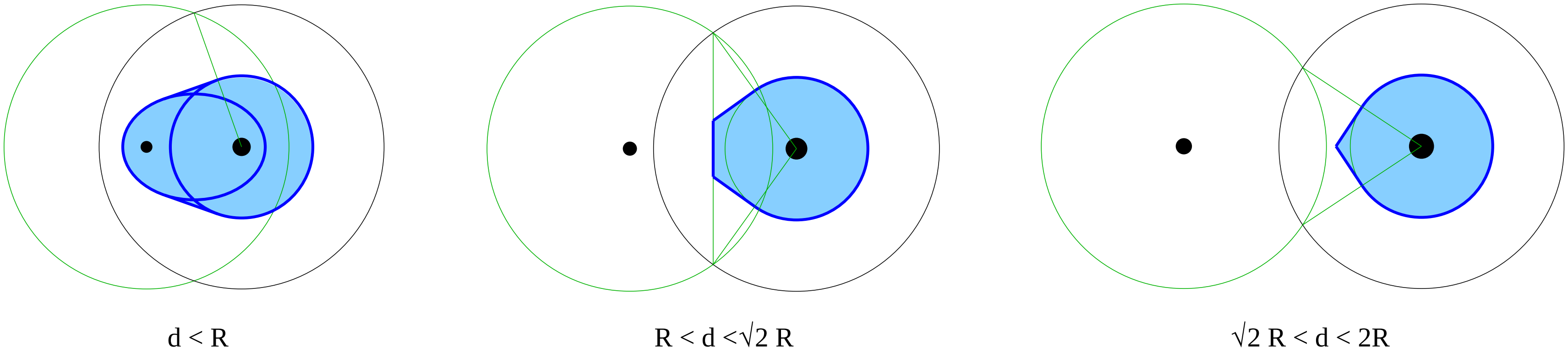
2-particle cases

Passive area: 1 hard particle + 1 soft



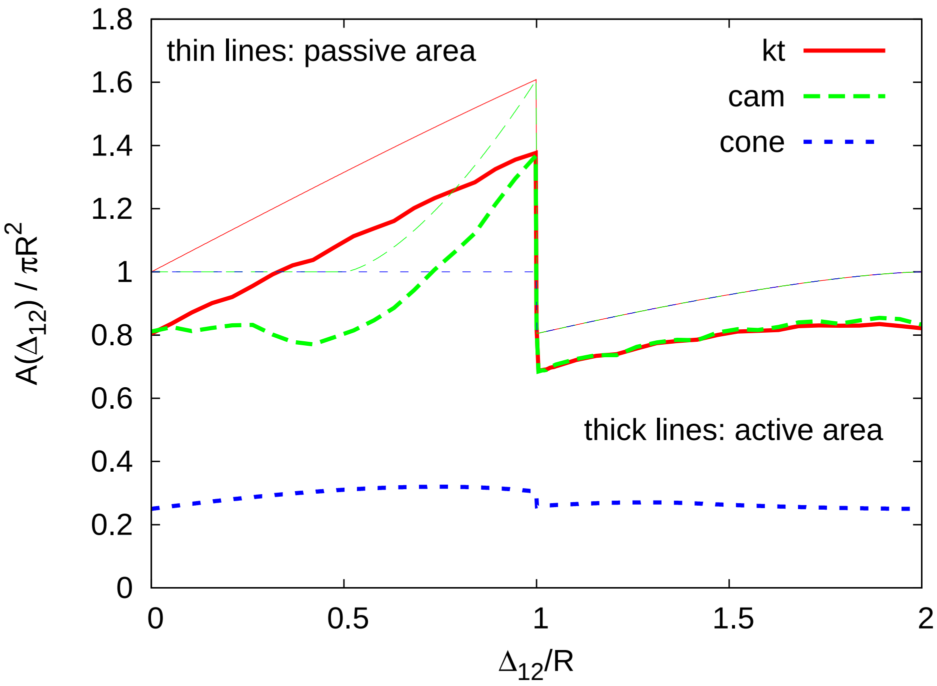
2-particle cases

Active area: 1 hard particle + 1 soft: analytic result for cone only



Alltogether, we have:

- Area $\neq$ cst. πR^2
- Δ_{12} dependence under control



Area scaling violations

QCD probability of emitting a small-angle soft gluon:

$$\frac{dP}{d\Delta_{12} dp_{t,2}} = C_{F,A} \frac{2\alpha_s}{\pi} \frac{1}{\Delta_{12}} \frac{1}{p_{t,2}}$$

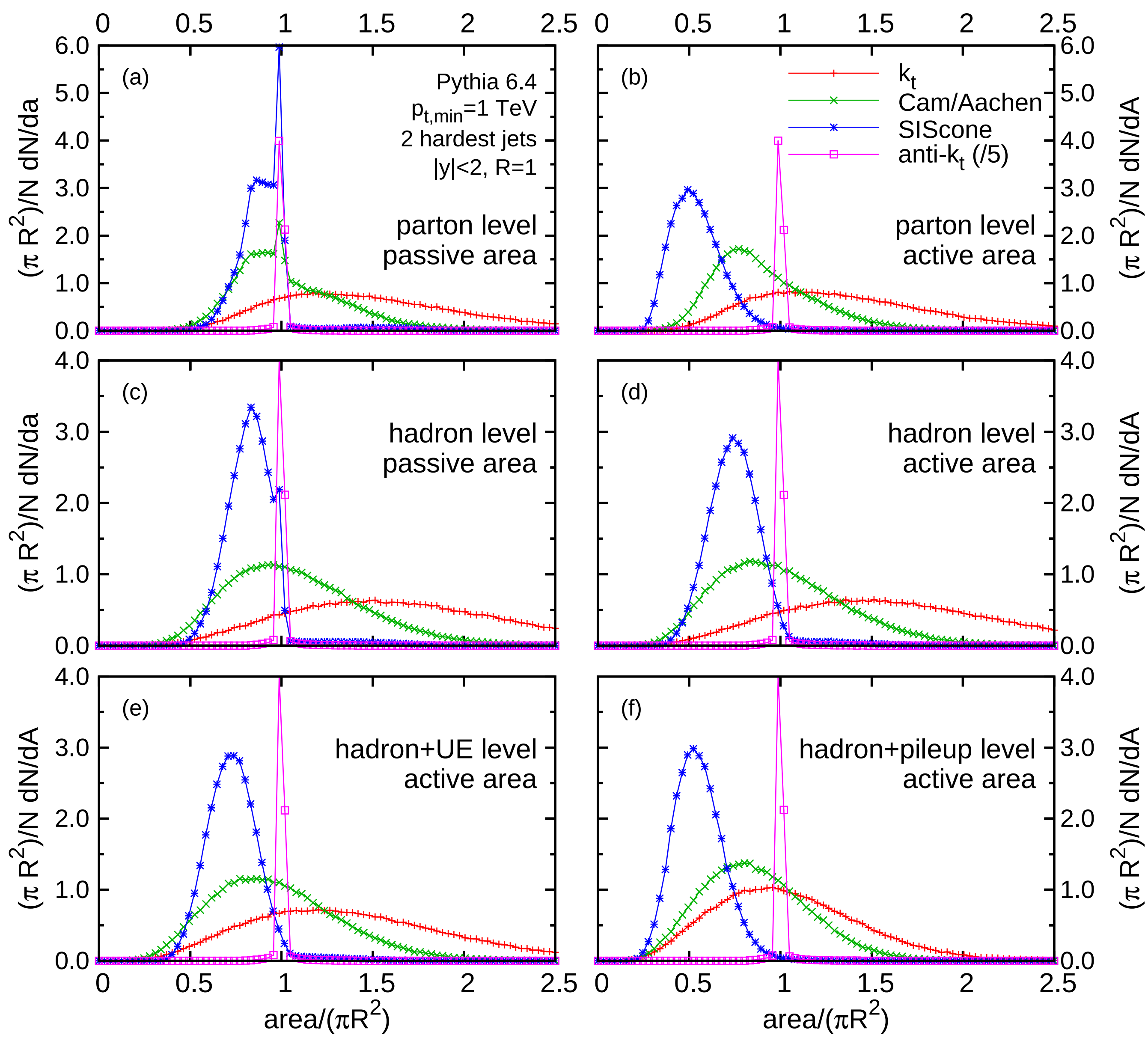
Hence the average area is

$$\begin{aligned} \langle \mathcal{A}(p_{t,1}, R) \rangle &= \mathcal{A}_{\text{1hard}}(R) + \int d\Delta \, dp_{t,2} \frac{dP}{d\Delta_{12} dp_{t,2}} \left[\mathcal{A}_{\text{hard+1 soft}}(\Delta, R) - \pi R^2 \right] \\ &= \frac{C_{F,A}}{\pi b_0} \log\left(\frac{\alpha_s(Q_0)}{\alpha_s(Rp_t)} \right) \pi R^2 \, d \end{aligned}$$

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

- Scaling viloation
- gluon > quark
- with know LO anomalous dimension

Area histograms



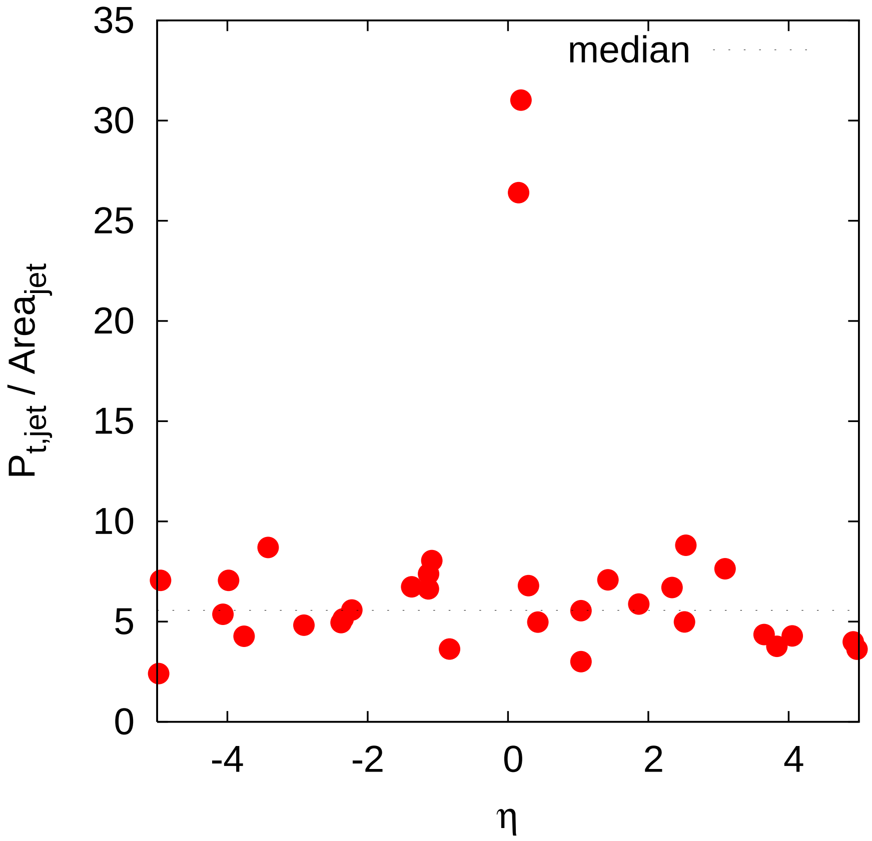
Pileup subtraction

Basic idea:

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

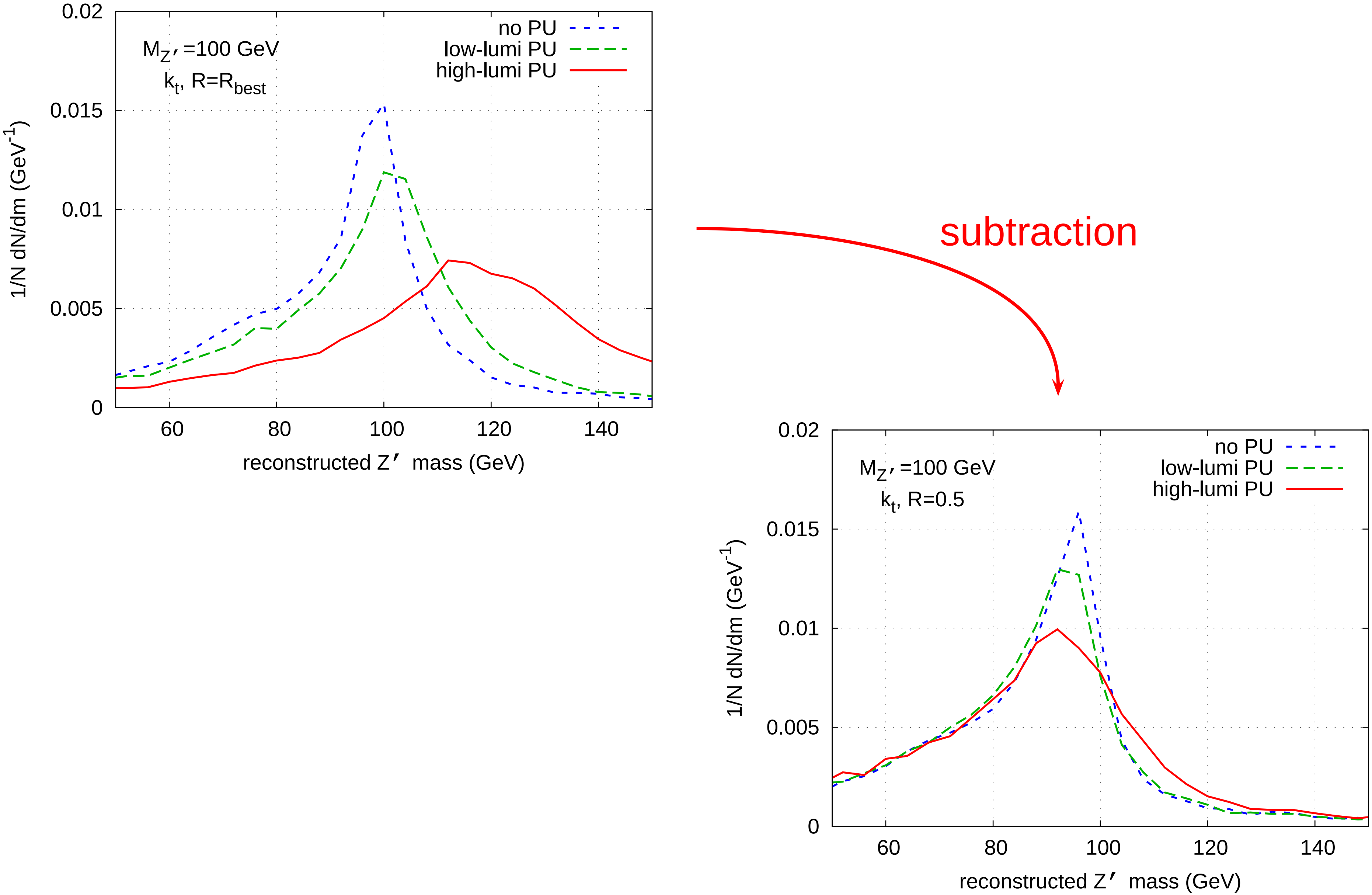
- Jet area:
  - region where the jet catches infinitely soft particles
  - tractable analytically
- Pileup density per unit area: ρ_{pileup}

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



implemented in FastJet
on an event-by-event basis

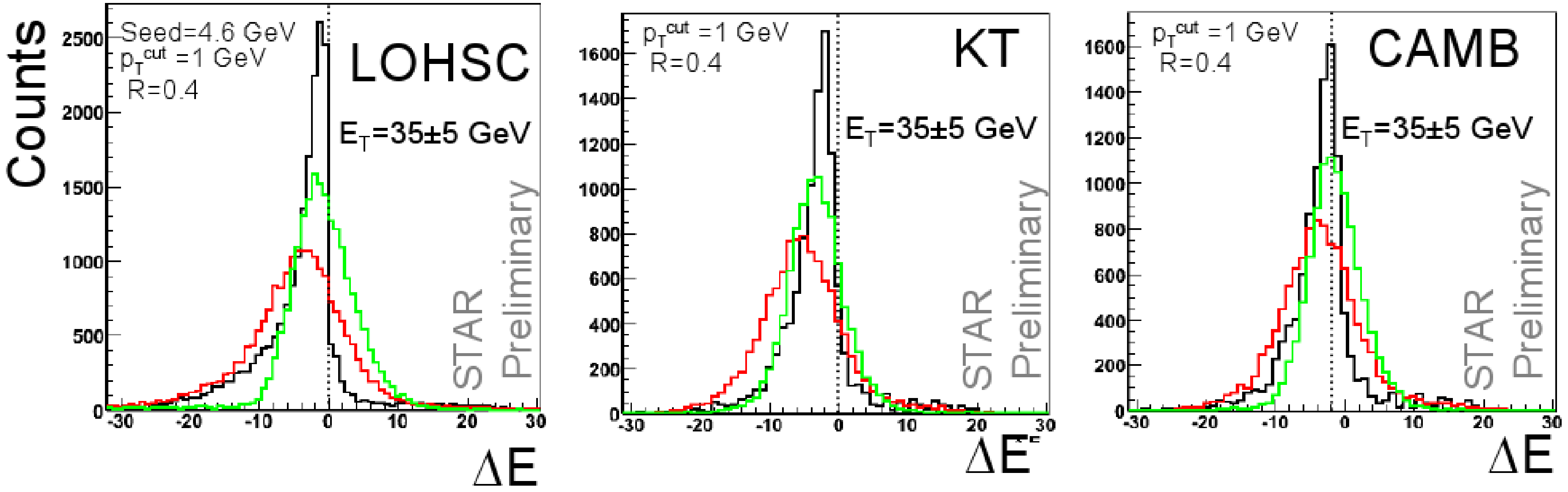
Subtraction recipe



Part 3: heavy-ion background:
subtracting more complex background

Jets at RHIC

[S. Salur, J. Putschke, ... (STAR)]



Red curves: $\Delta E = E^{\text{PyEmbed}} - E^{\text{PyDet}}$

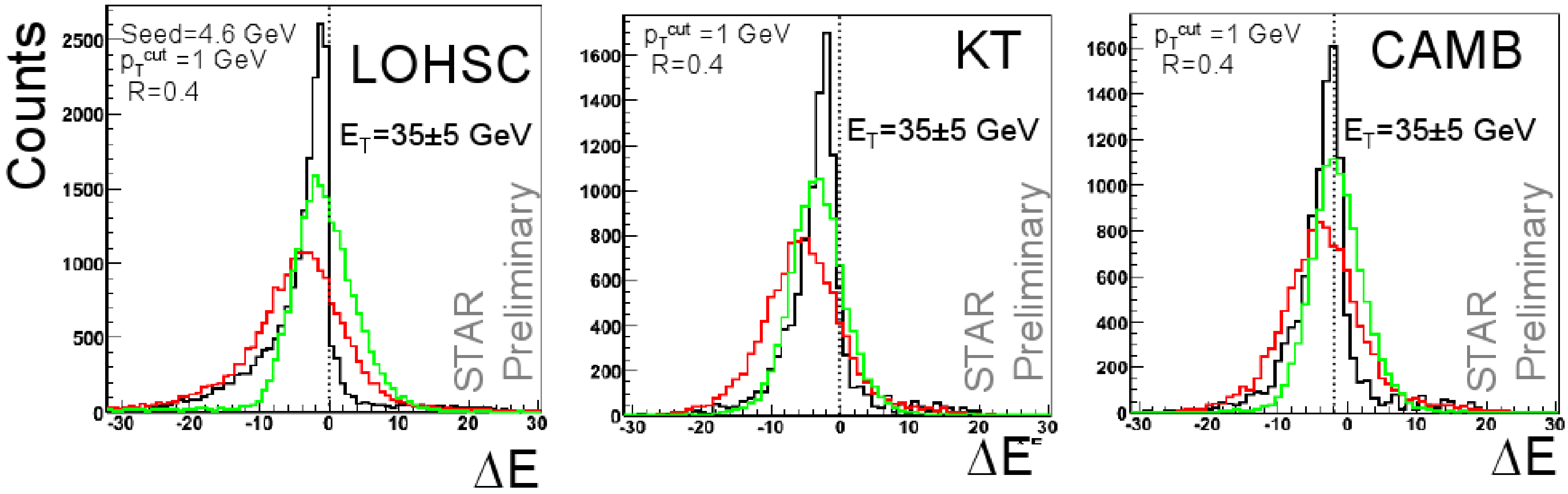
with PyDet $\equiv$ Pythia pp + detector effects

PyEmbed $\equiv$ same with real AuAu event added

$\Rightarrow$ measure of the subtraction efficiency

Jets at RHIC

[S. Salur, J. Putschke, ... (STAR)]



Red curves: $\Delta E = E^{\text{PyEmbed}} - E^{\text{PyDet}}$

with PyDet $\equiv$ Pythia pp + detector effects

PyEmbed $\equiv$ same with real AuAu event added

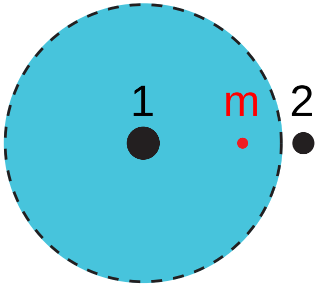
$\Rightarrow$ measure of the subtraction efficiency

Work under progress: removing the last few GeV shift in ΔE

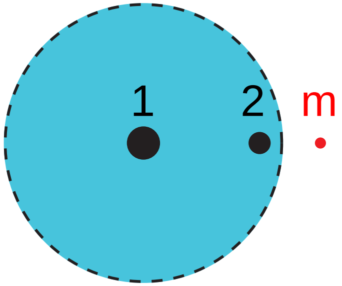
Back-reaction & HI UE

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



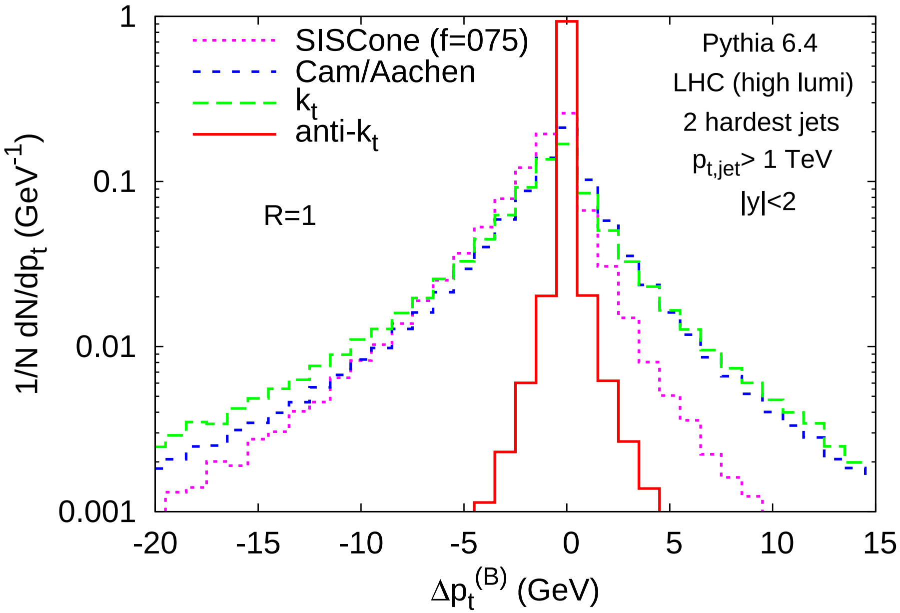
  - loss: p_2 lost when adding p_m



Back-reaction & HI UE

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



Back-reaction & HI UE

Additional soft background has 2 effects:

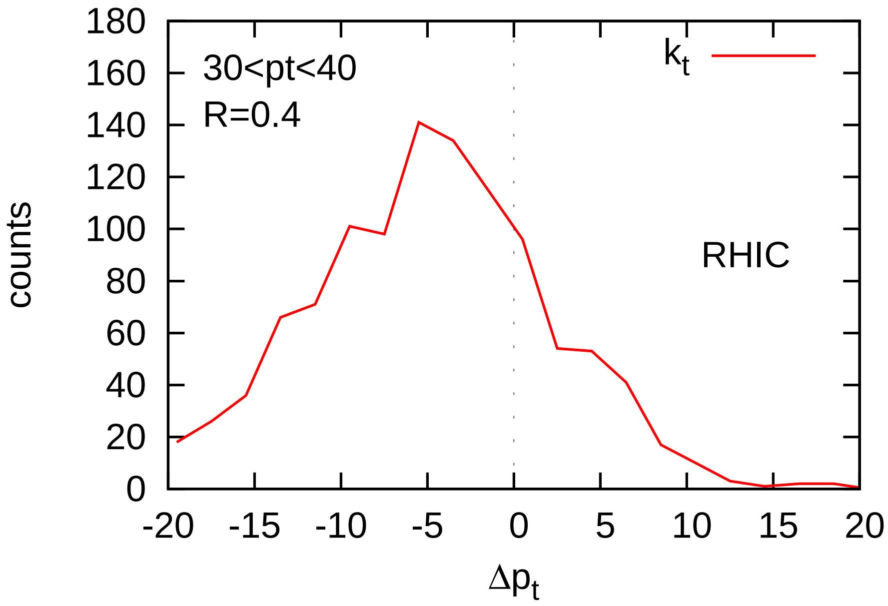
- Throw soft particles in the hard jet: dealt with by subtraction
- Modify the hard scattering (back-reaction)
  - can be pointlike or diffuse
  - tractable analytically (similar to areas)
  - $k_t \gtrsim$ Cambridge > SISCone $\gg$ anti-k_t

+ For heavy-ion collisions: fluctuating underlying event background

$\longrightarrow$ median estimation of ρ might oversimplified

Results (3)

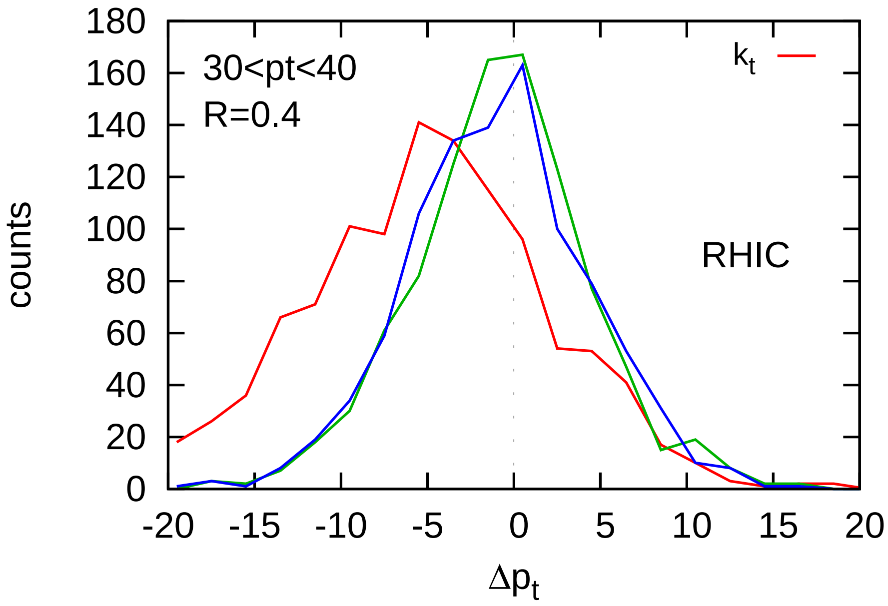
Work under progress: test subtraction using
Pythia pp hard event + HYDJET AA background



Similar shifts than STAR

Results (3)

Work under progress: test subtraction using
Pythia pp hard event + HYDJET AA background



Target: reach a precision of 1-2 GeV by

- carefully tuning the algorithm to reduce UE sensitivity and back-reaction
- carefully tuning the subtraction to deal with the fluctuating background

Conclusions

- Use IRC-safe algorithms!
- Jet-finding in pp at the LHC
  - SISCone and Cam+filt. do a slightly better job
  - strong R dependence: important to choose R_{best}
  - R_{best} increases with the scale
  - same for quark and gluon jets, larger R_{best} for gluons

  $\Rightarrow$ flexibility in jet finding at the LHC
- Subtraction using jet areas
  - Jet areas: clearly defined, analytic control
  - Simple systematic pileup subtraction
- Jet-finding in AA at RHIC and the LHC
  - First measurement at RHIC
  - Work under progress: improve subtraction down to 1-2 GeV