Optimising jet finding in pp and AA collisions

Grégory Soyez

Brookhaven National Laboratory

C. Buttar *et al.*, arXiv:0803.0678

M. Cacciari, J. Rojo, G.P. Salam, G. Soyez, in prep.

M. Cacciari, J. Rojo, G.P. Salam, G. Soyez, in prep.

Plan

- Foreword: why jets?

  introducing the basic terms

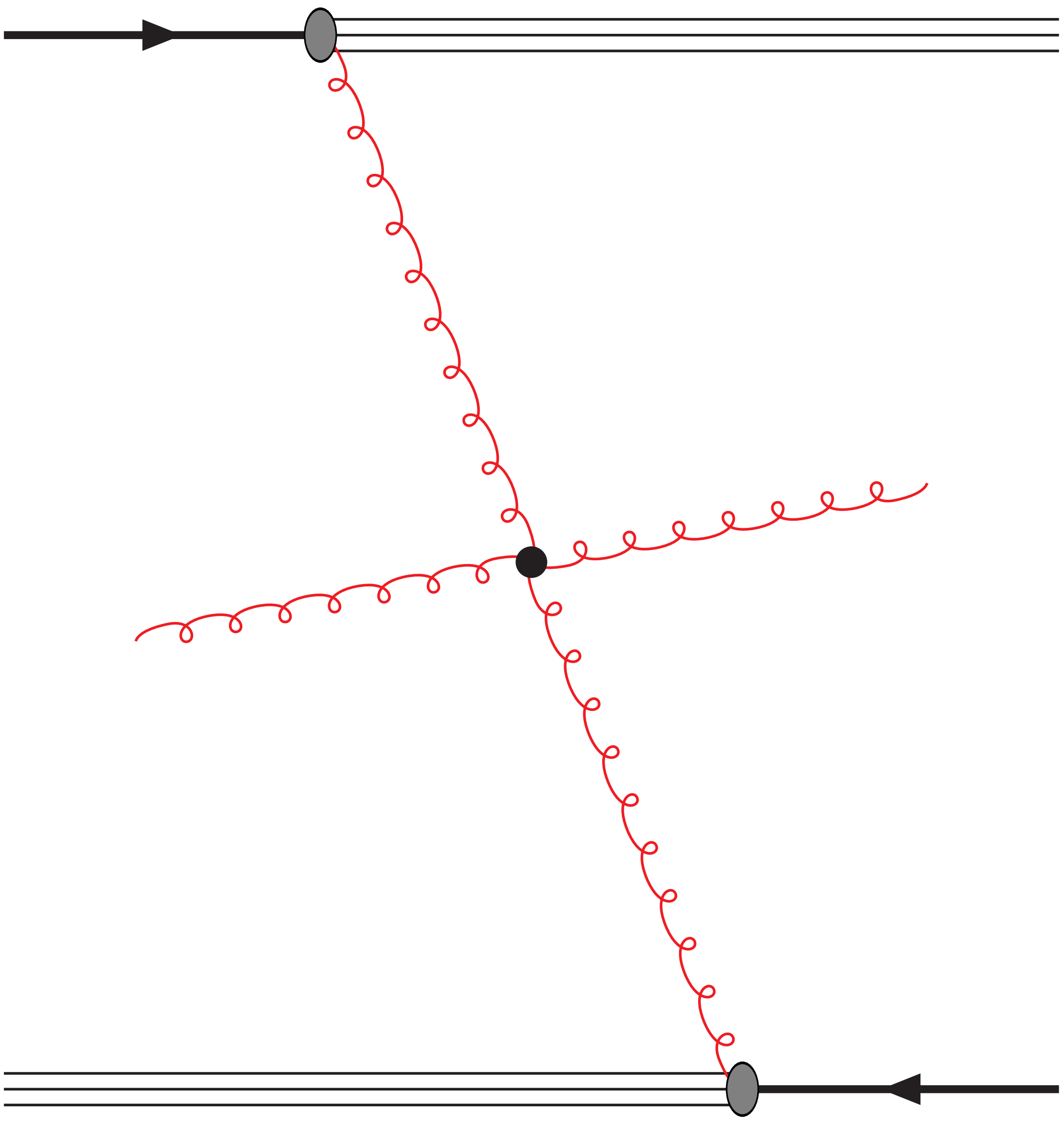
- Part 0: a solid toolkit

  jet definitions meeting the fundamental requirements

- Part 1: “best” jet finder in pp collisions
  - sample processes to study
  - figure of merit for quality measure
  - results (1)
  - analytic insight
- Part 2: when pileup enters the game
  - subtracting soft background using jet areas
  - results (2)
- Part 3: heavy-ion background
  - subtraction subtleties
  - results (3)

Foreword: why jets?

General (over)simplified picture



Perturbative level

Hard scattering $2 \to n$

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

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

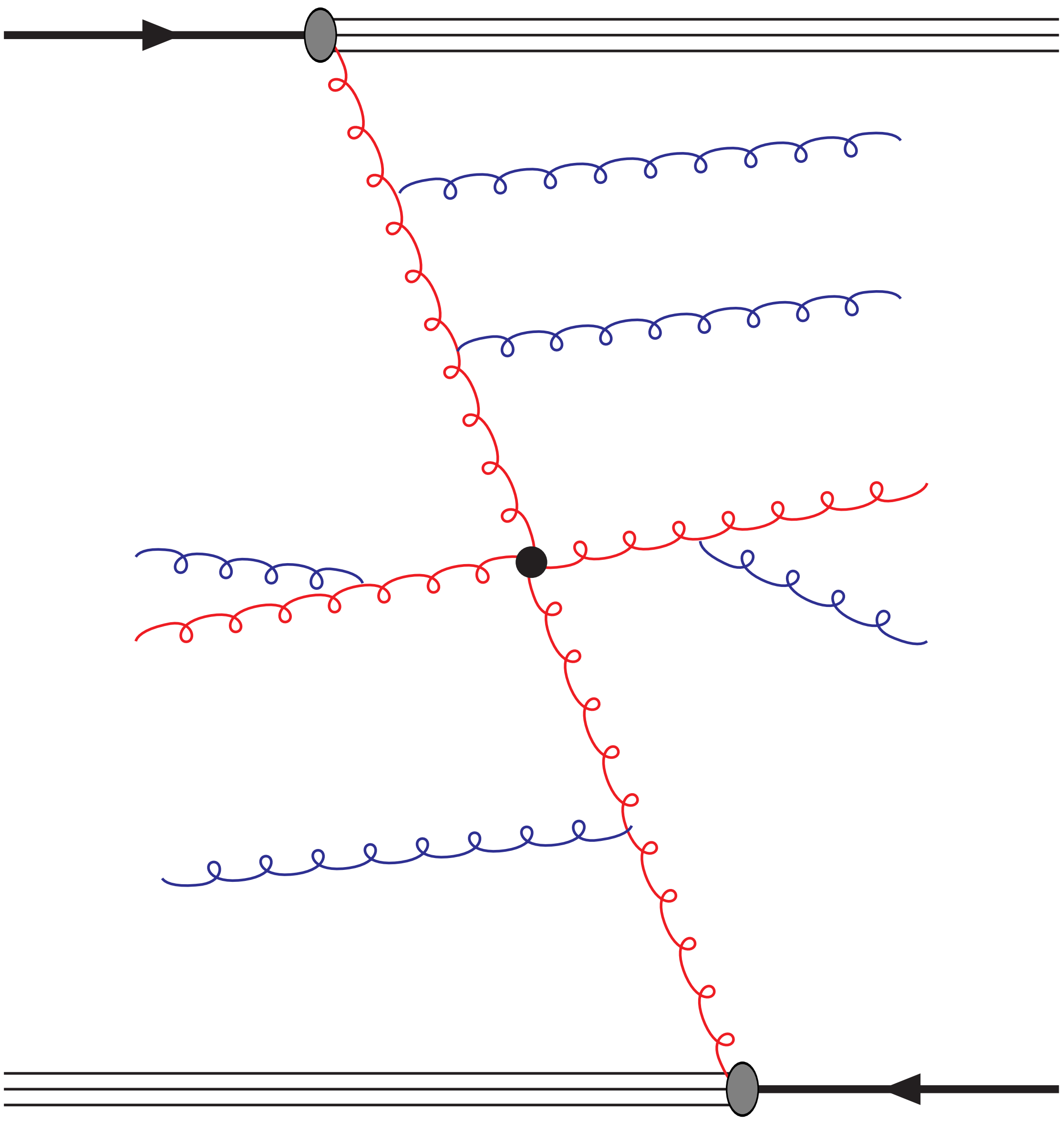
$gg \to gggg$,

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

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

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

General (over)simplified picture

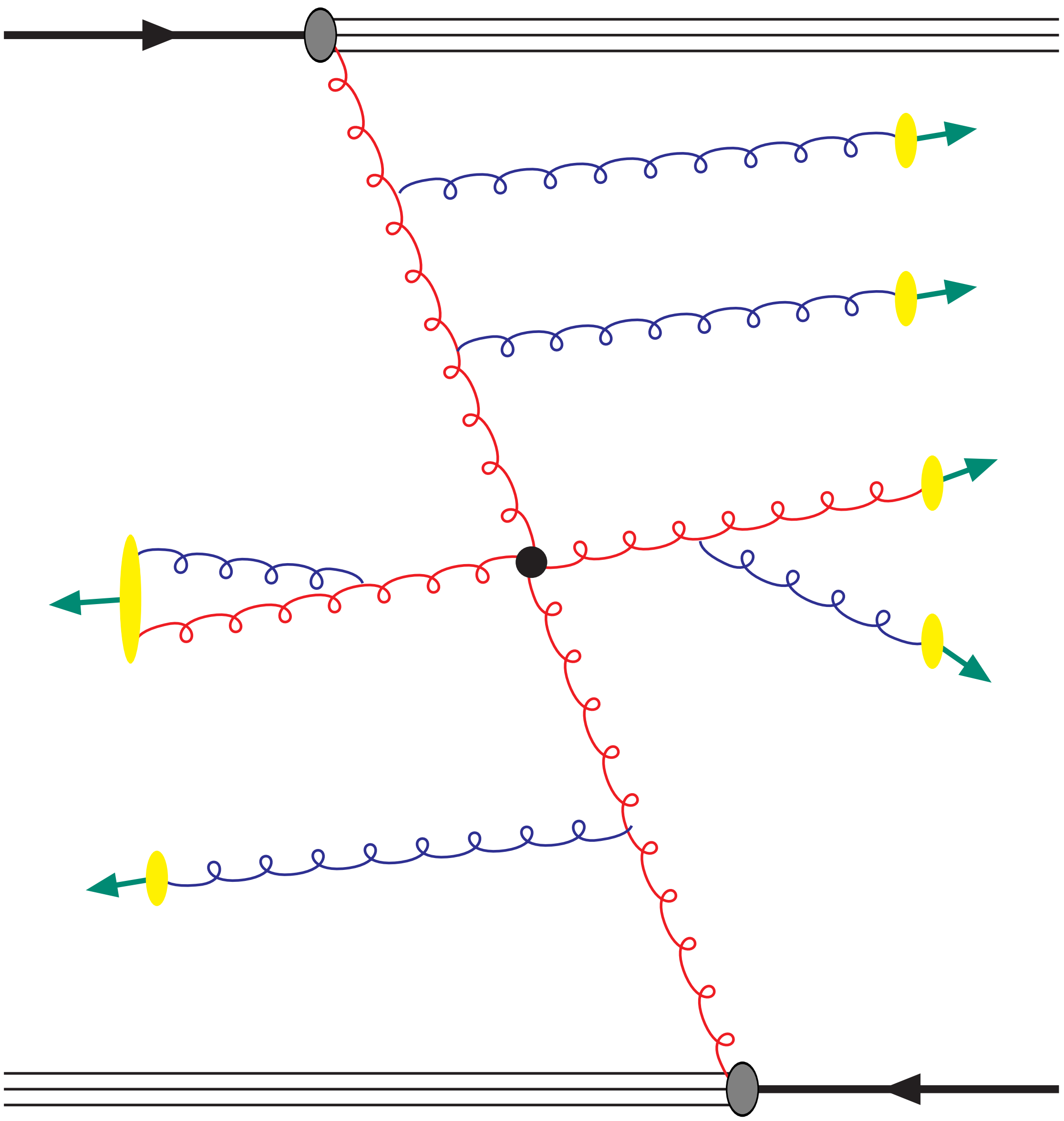


Perturbative level

Parton level

$\approx$ collinear divergences
resummation
$\sum_i \alpha_s^i \log^i(p_t^2/\mu^2)$

General (over)simplified picture



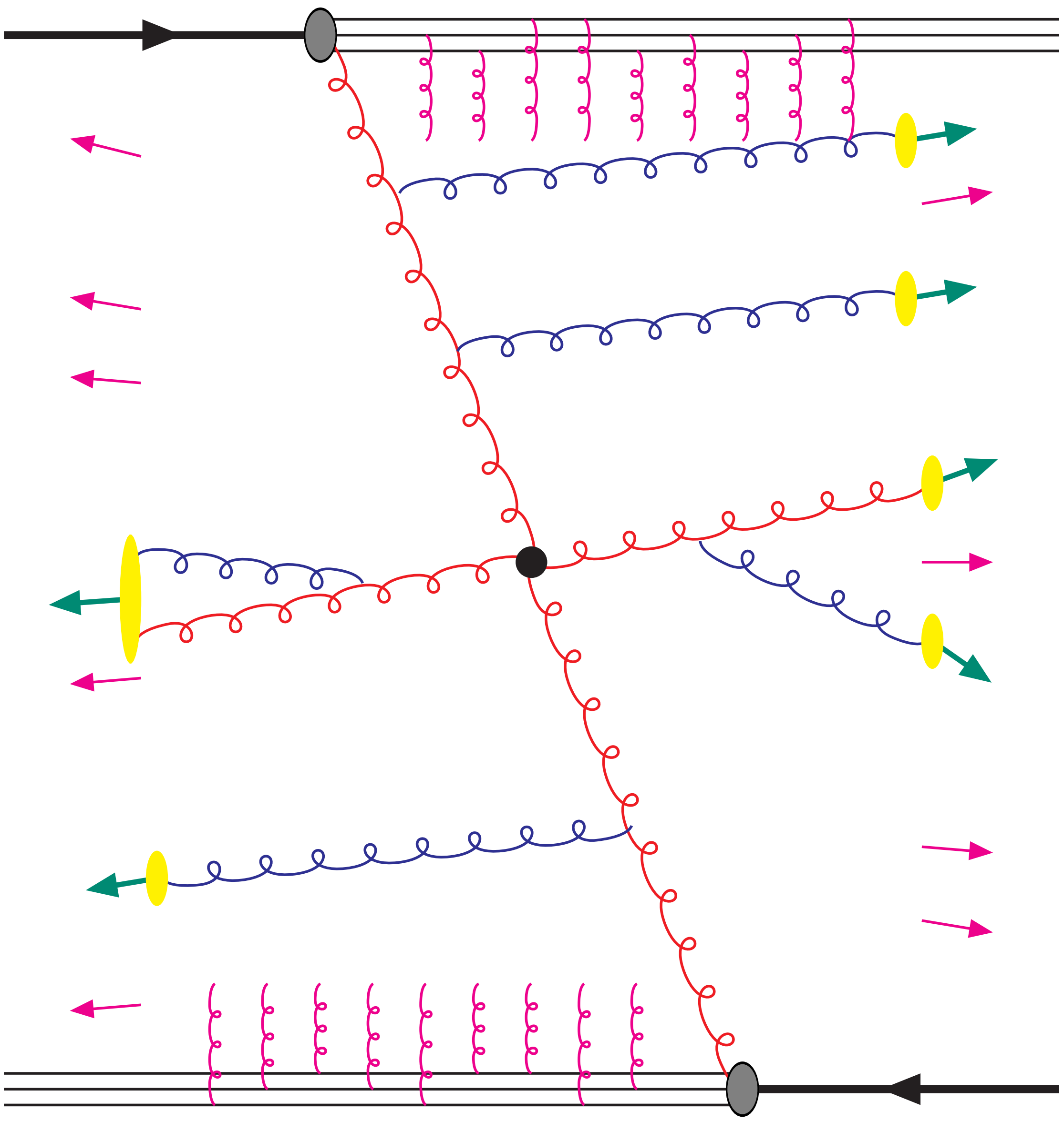
Perturbative level

Parton level

Hadron level

quarks+gluon → hadrons
(various models)

General (over)simplified picture



Perturbative level

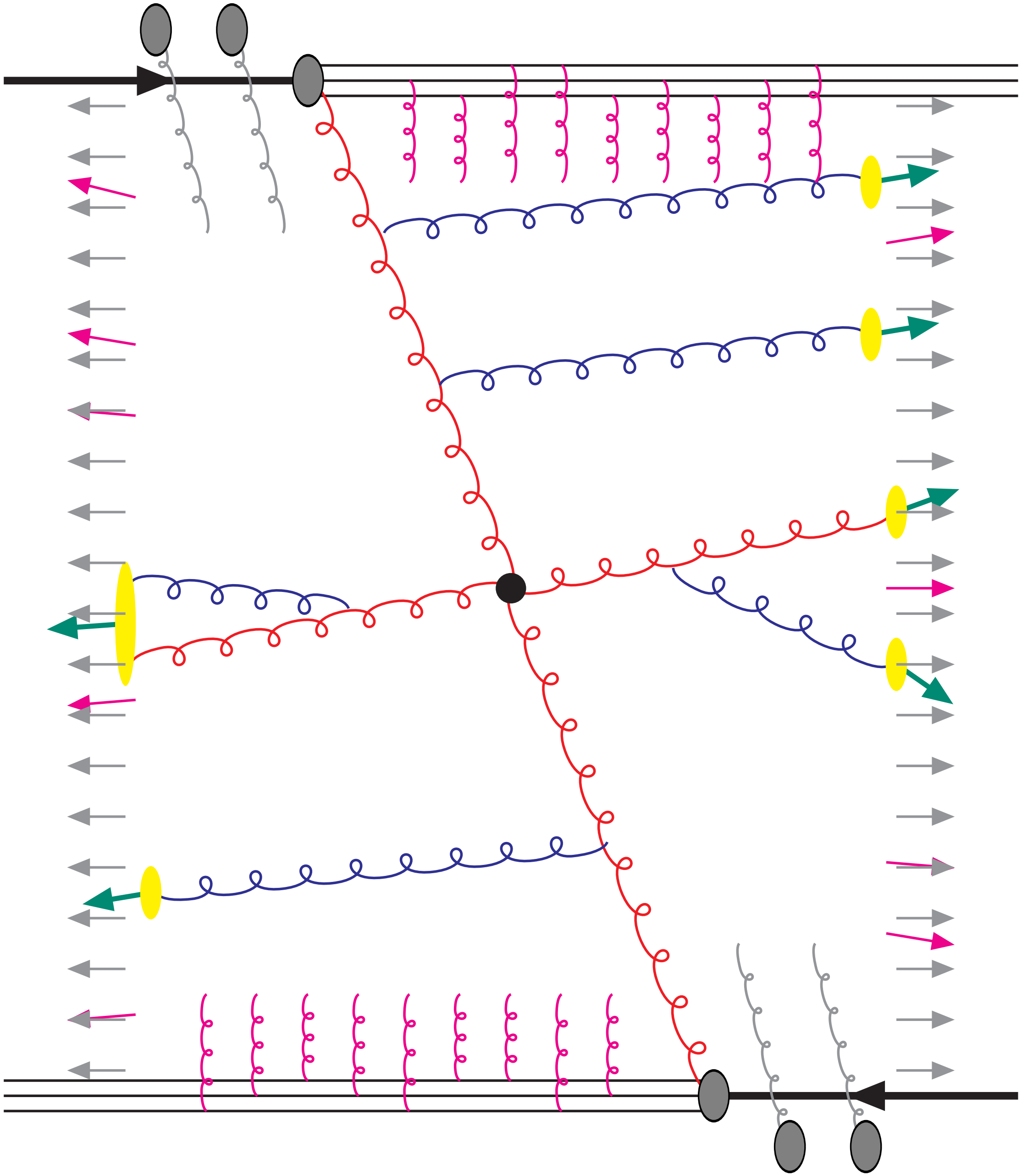
Parton level

Hadron level

+ Underlying event

Multiple interactions
from beam remnants
$\Rightarrow$ soft background

General (over)simplified picture



Perturbative level

Parton level

Hadron level

+ Underlying event

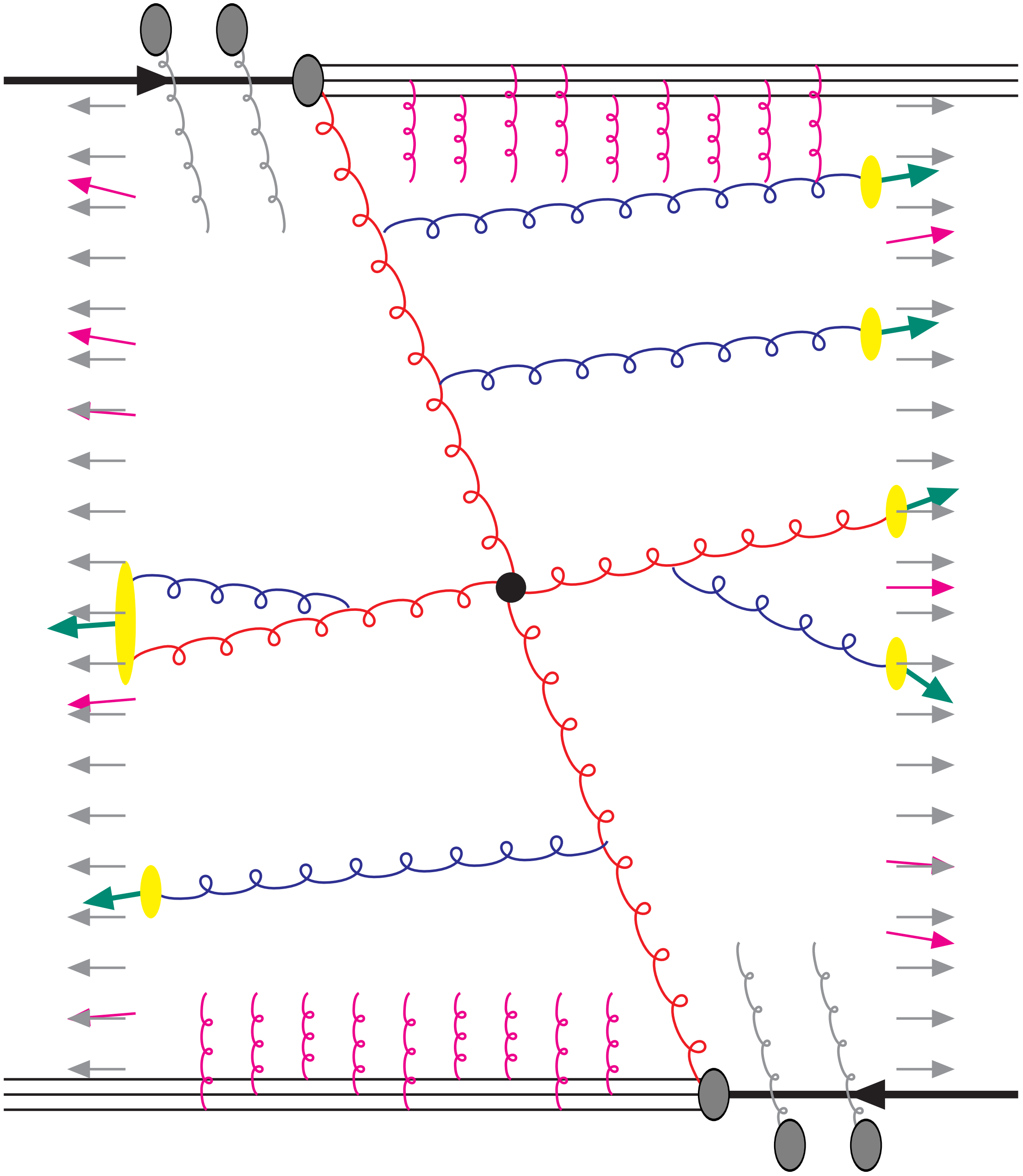
+ Pile up

additionnal pp interactions

$\Rightarrow$ soft background

$\approx$ uniform

General (over)simplified picture



Perturbative level

Parton level

Hadron level

+ Underlying event

+ Pile up

"Jets" $\equiv$

Clusters to access the hard scattering?

Part 0: 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.

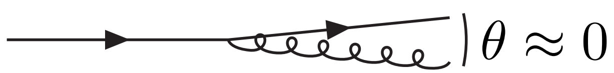
i.e. usable by theoreticians (*e.g.* finite perturbative results)
and experimentalists (*e.g.* fast enough, not much UE sensitivity)

QCD divergences

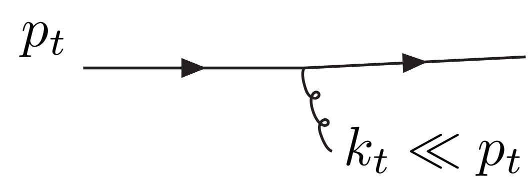
QCD probability for gluon bremsstrahlung at angle θ and $\perp$-mom. k_t:

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

Two divergences:



Collinear



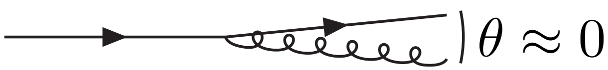
Soft

QCD divergences

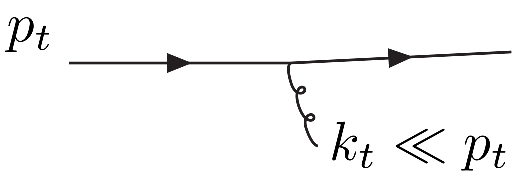
QCD probability for gluon bremsstrahlung at angle θ and $\perp$-mom. k_t:

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

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

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 dive

LHC requires precision in QCD

$\Rightarrow$ IRC safety is mandatory

Huge theoretical effort to compute higher-order processes:

don't waste it $\Rightarrow$ IRC safety is mandatory

For pQC

- one
  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

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

21st century breakthrough

Many important recent developments between 2006 and 2008:

- Speeding up the k_t and Cam/Aachen algorithms

  using computational-geometry techniques: $\mathcal{O}\left(N^3\right) \to \mathcal{O}\left(N \log N\right)$

  [M. Cacciari, G. Salam, 06]

- Existing cones are infrared and/or collinear unsafe
  - Cones with split–merge replaced by SISCone

    the only infrared-and-collinear-safe cone
    geometry technques ⇒ no cost in time ($\mathcal{O}\left(N^3\right) \to \mathcal{O}\left(N^2 \log N\right)$)

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

  - Cones with progressive removal $\longrightarrow$ anti-k_t

    fast recombintion-type algorithm
    hard jets are circular as for cones with progressive removal

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

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 ("$2^{\rm nd}$ generation of algorithms")

Cambridge+Filtering algorithm:

- Cluster with Aachen/Cambridge and radius R
- For each jet, recluster it with Aachen/Cambridge and radius $R_{\rm sub}$
  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: $R_{\rm sub} = R/2$, $n_{\rm sub} = 2$

Part 1: “best” jet finder in pp collisions

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

We need a measure of the jet reconstruction efficiency

- Forget about measures related to parton-jet matching
- use the reconstructed mass peak

figure of merit for quality measure

We need a 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

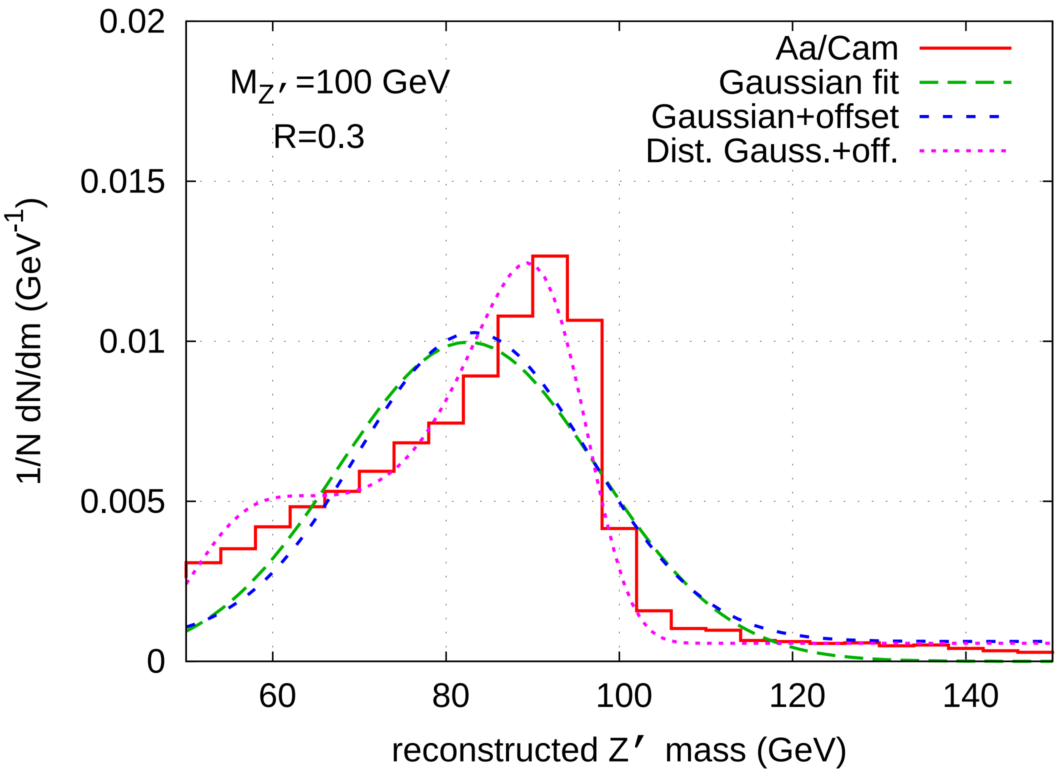


figure of merit for quality measure

We need a 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,

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

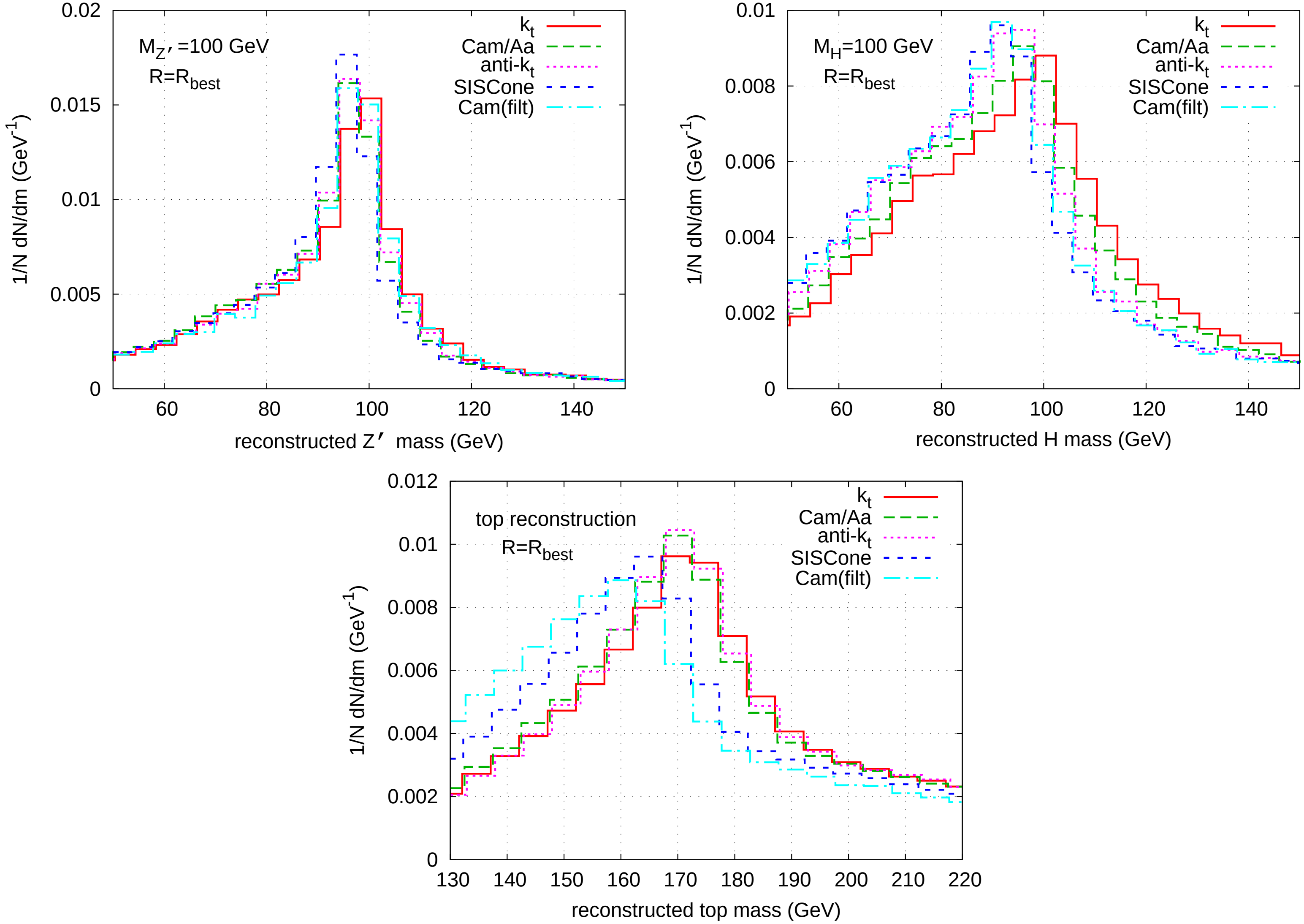
smaller width $\equiv$ better signal-to-background ratio

- 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.

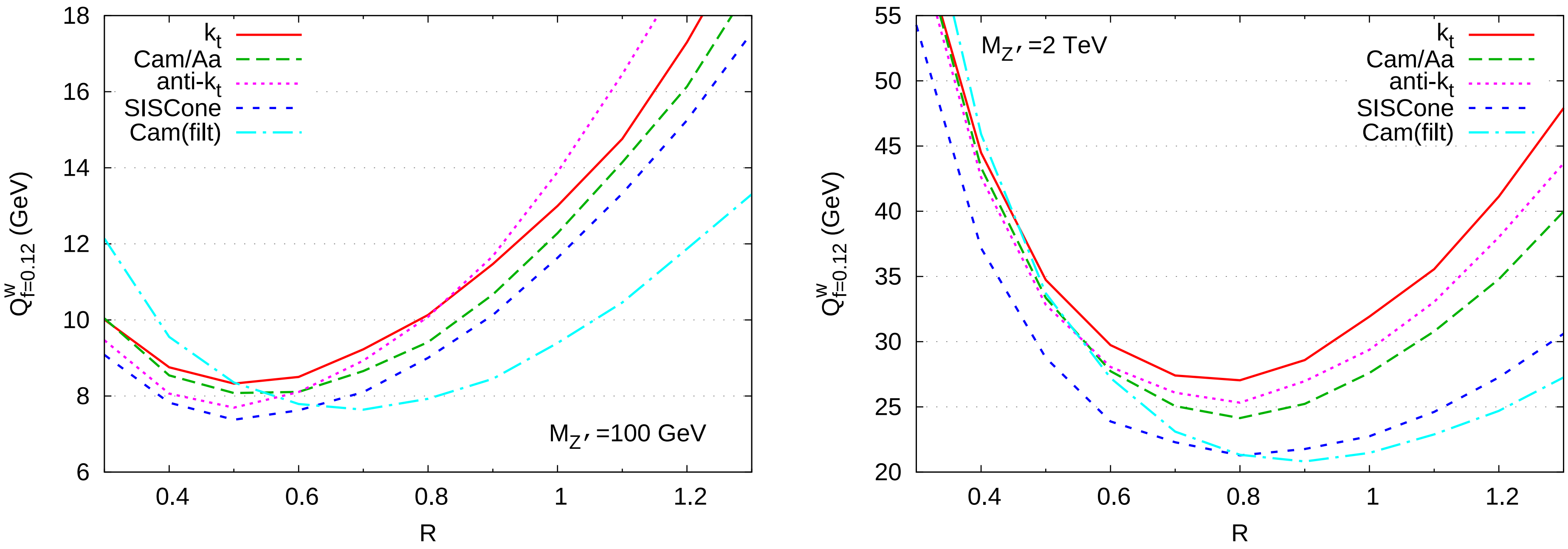
we see peaks...



$M_{Z'}$=100 GeV
R=R_{best}
k_t
Cam/Aa
anti-k_t
SISCone
Cam(filt)
1/N dN/dm (GeV^{-1})
reconstructed Z′ mass (GeV)
M_H=100 GeV
R=R_{best}
reconstructed H mass (GeV)
top reconstruction
R=R_{best}
reconstructed top mass (GeV)

BROOKHAVEN NATIONAL LABORATORY

Message 1: there is a strong R dependence



At 100 GeV,

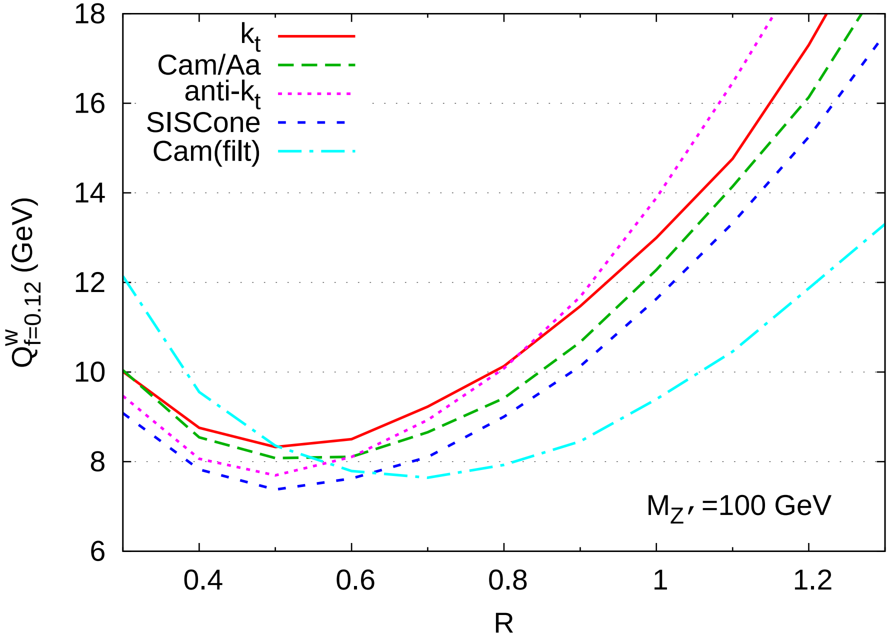
using $R = 0.8$ instead of $R = 0.5$ means a discr. power loss of 20% ($\rho_{\mathcal{L}} \approx 0.8$)

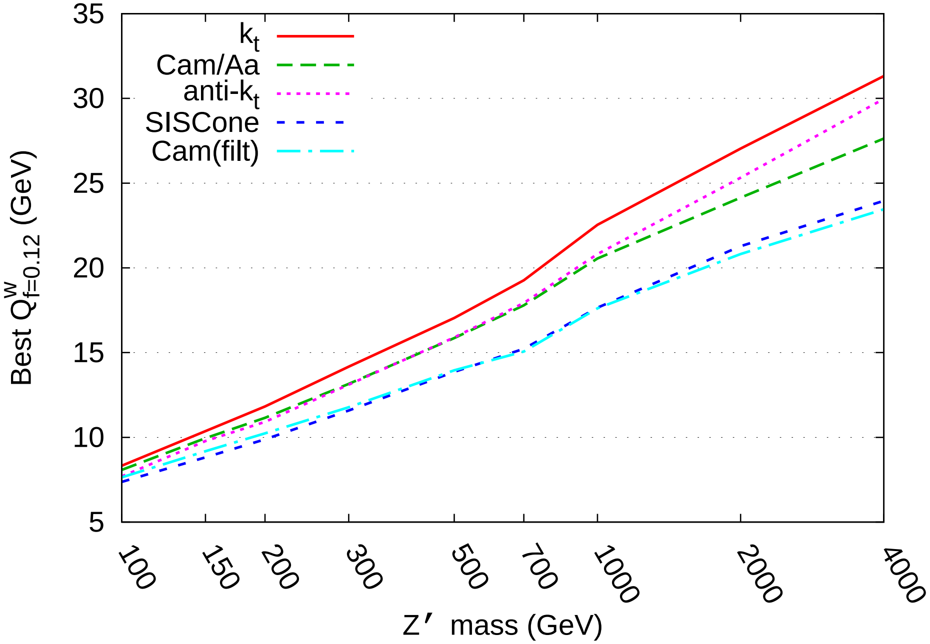
$R = 1$ 40% ($\rho_{\mathcal{L}} \approx 0.6$)

At 2 TeV,

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

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





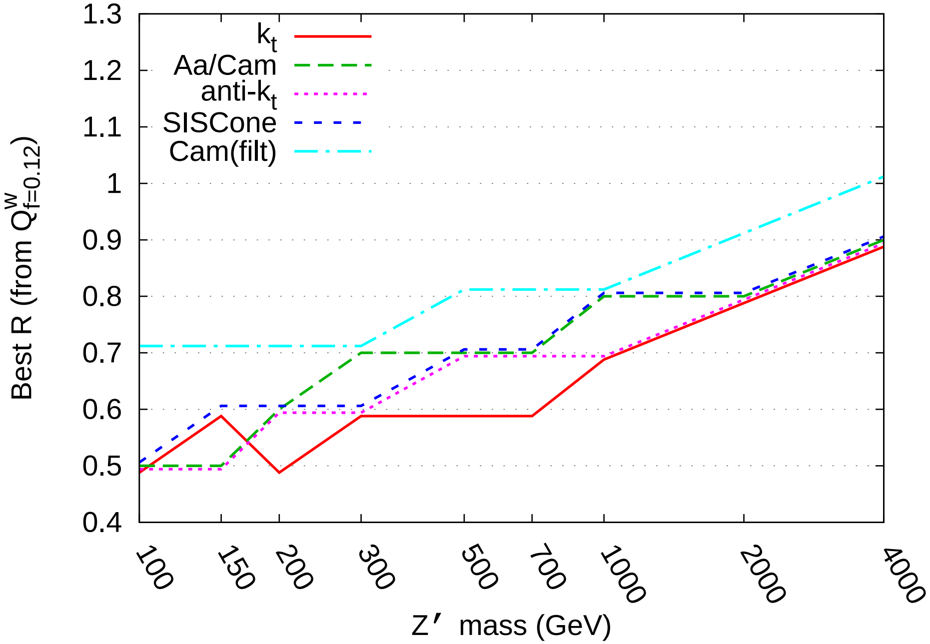
Using k_t instead of SISCone means a discr. power loss of

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

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

results (1)

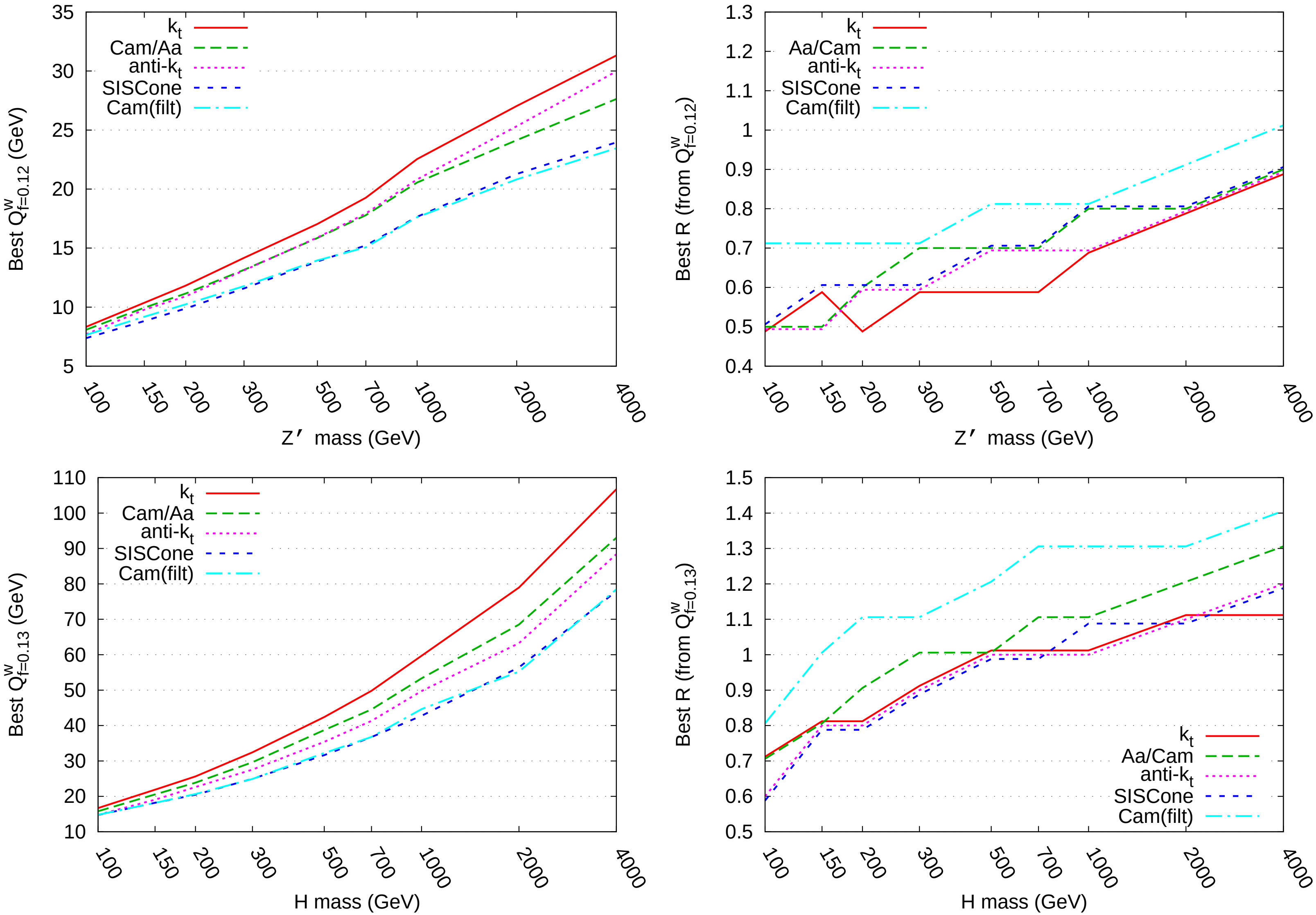
Message 3: The parameters vary with the scale



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

results (1)

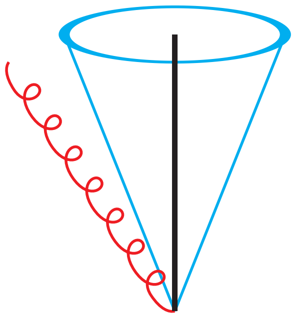
Message 4: same for the gluon jets, though with a larger R



analytic insight

Competition between

- catching perturbative radiation



Out-of-cone radiation:

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

- not catching soft background radiation

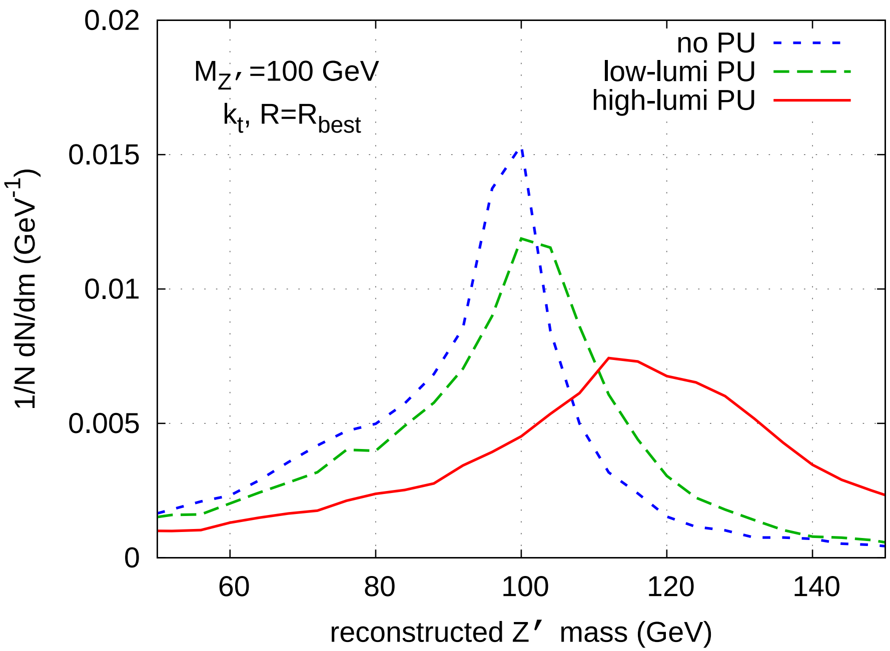
Soft contents $\sim$ jet area $\sim R^2$

more detailed computation in progress...

Part 2: when pileup enters the game

Need for subtraction

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



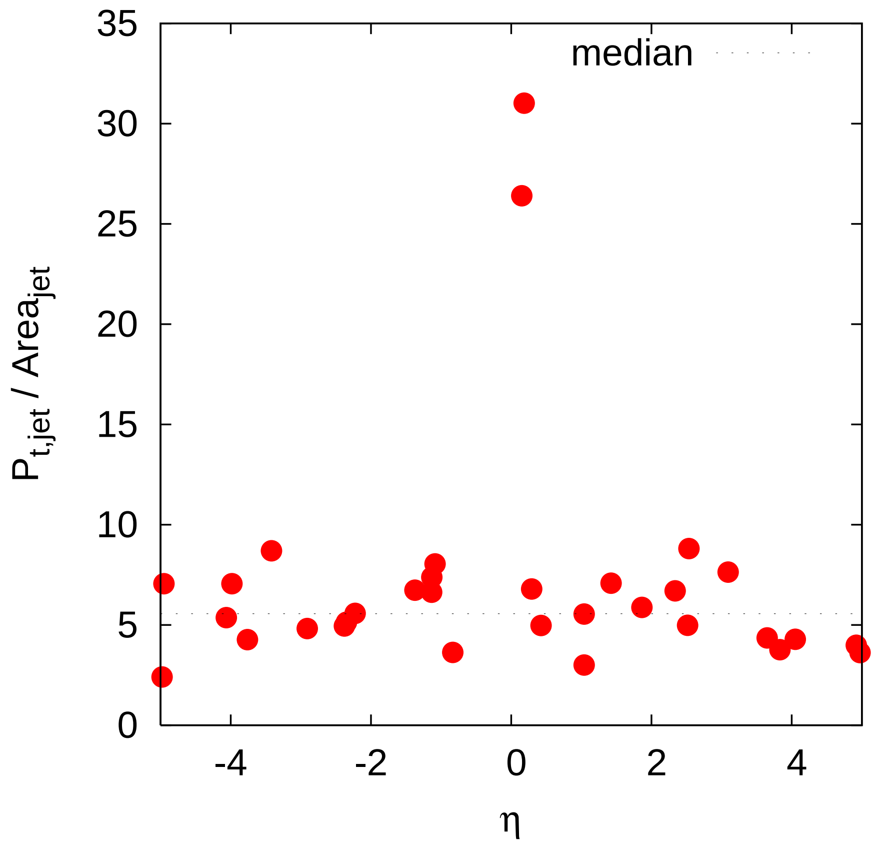
... that needs to be subtracted!

Basic idea:

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

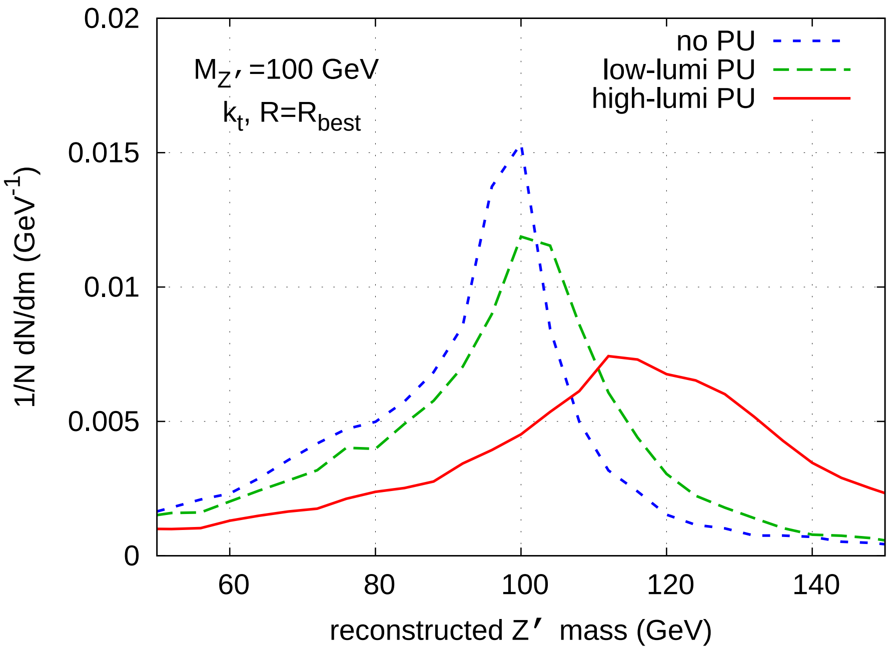
- Jet area:
  - region where the jet catches infinitely soft particles
  - tractable analytically
- Pileup density per unit area: ρ_{pileup}

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

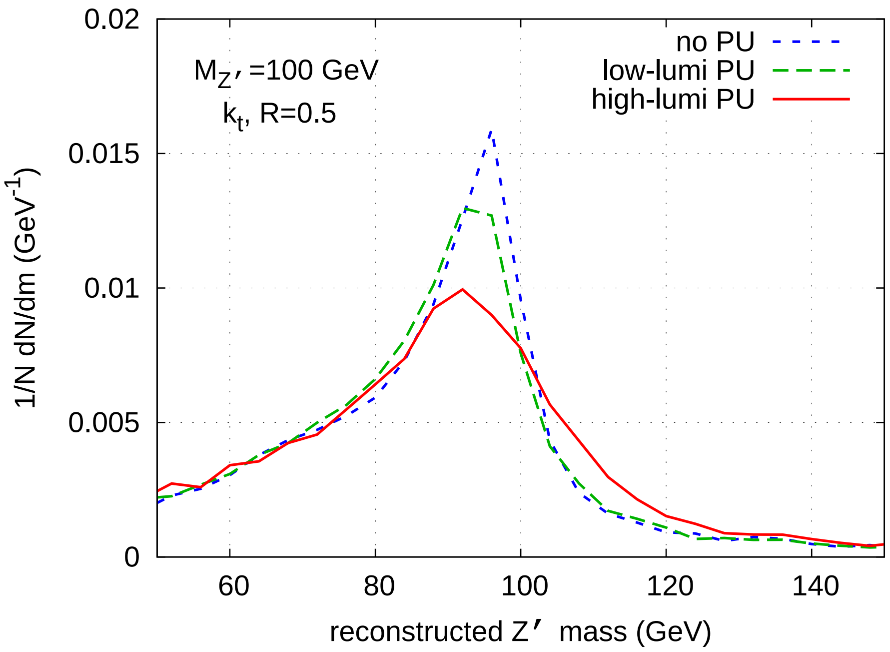


implemented in FastJet
on an event-by-event basis

Subtraction recipe

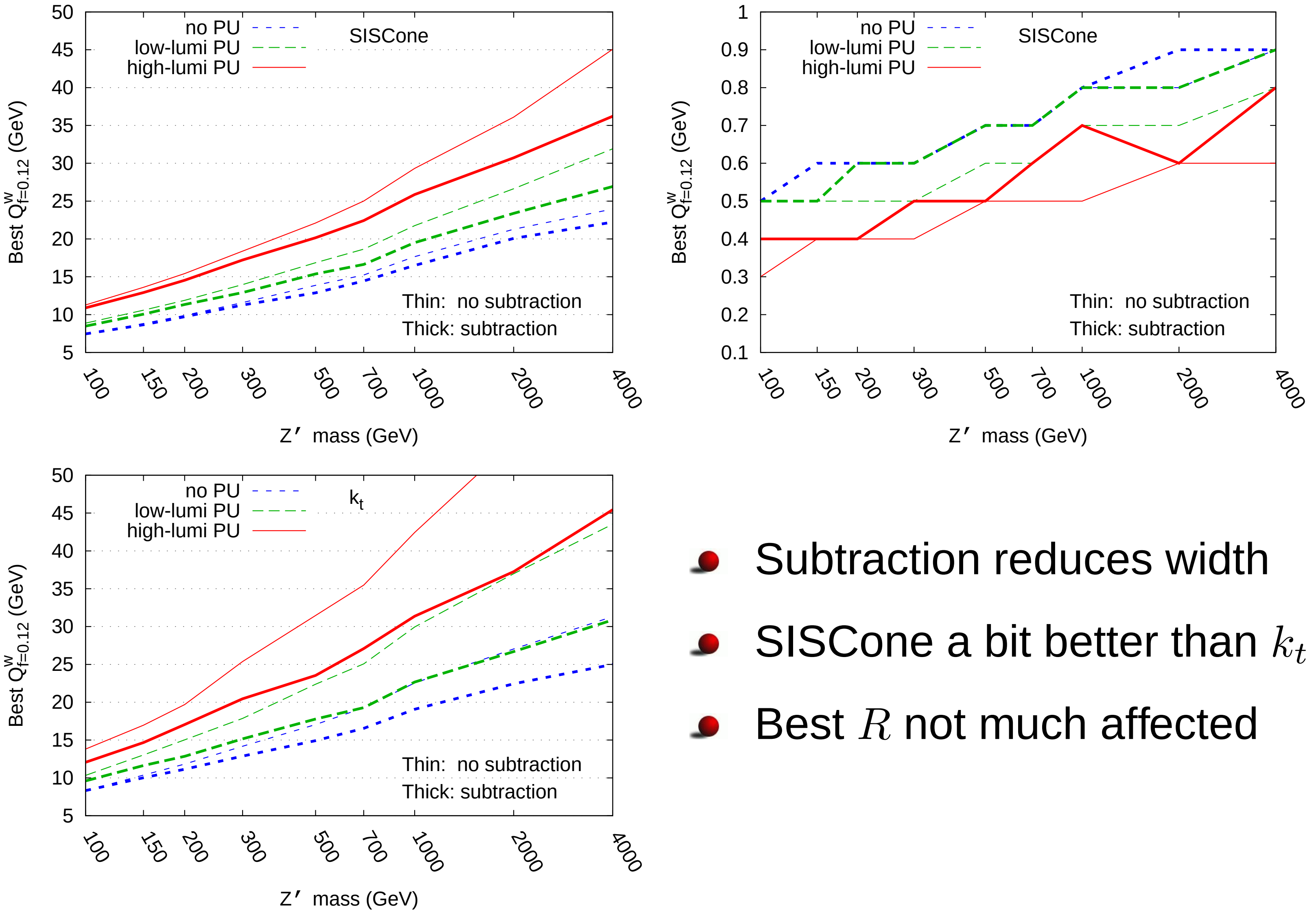


subtraction



results (2)

Message 5: with subtraction, pileup has reasonably little influence



- Subtraction reduces width
- SISCone a bit better than k_t
- Best R not much affected

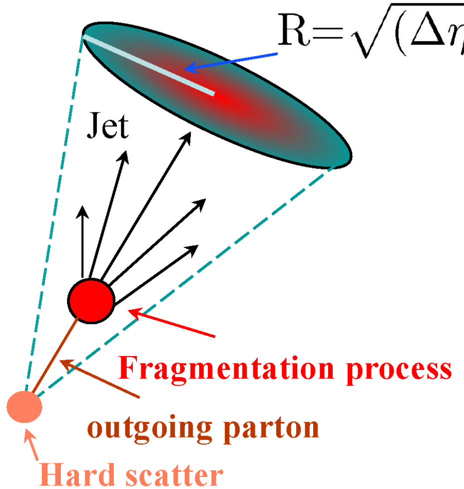
Part 3: heavy-ion background:
subtracting more complex background

Jets at RHIC

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

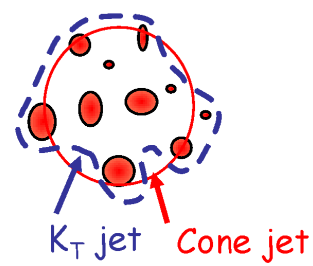
Jet Reconstruction Algorithms:

$R=\sqrt{(\Delta\eta)^2+(\Delta\phi)^2}$



Cone Algorithm

1. Leading Order High Seed Cone (LOHSC)
2. Mid Point Cone: Merging & Splitting



Sequential recombination

3. KT
4. Cambridge/ Aachen

Explore systematics: Use both Clustering & Cone algorithms.

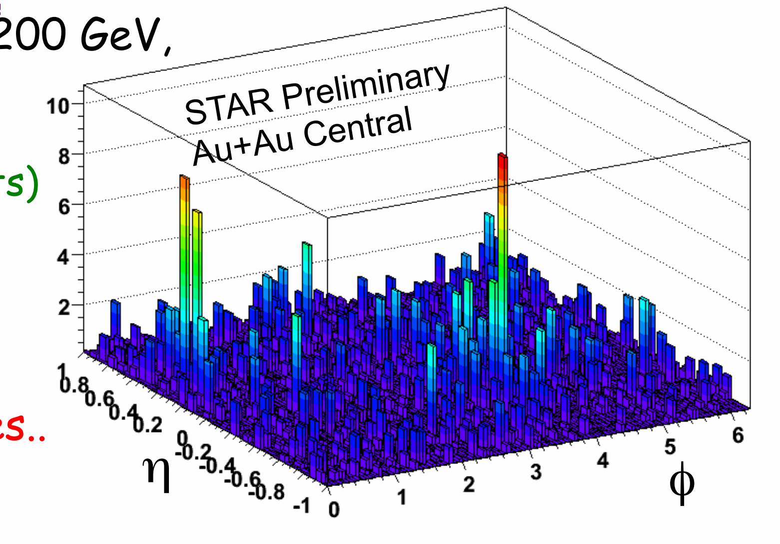
Sevil Salur 6

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

Correction for Heavy-Ion Background

0-10% Most Central Au+Au at $\sqrt{s_{NN}}$= 200 GeV,

- R=0.4, Bkg Energy ~ 40 GeV
- Unmodified (p+p) jets: ~80% of energy within R~0.3 for 50 GeV jet (CDF/D0 Jets)



- Background Estimates:

Assess backgrounds event by event.

1. **Cone:** Look at $\langle p_T \rangle$ out-of jet cones..
   $A=\pi R^2$

2. **Sequential Recombination**: Estimate the active area of each jet by addition of zero energy particles of known density.

$$p_T \text{ (Jet Measured)} \sim p_T\text{(Parton)} + \rho \times A(\text{Jet}) \pm \sigma\sqrt{A(\text{Jet})}$$

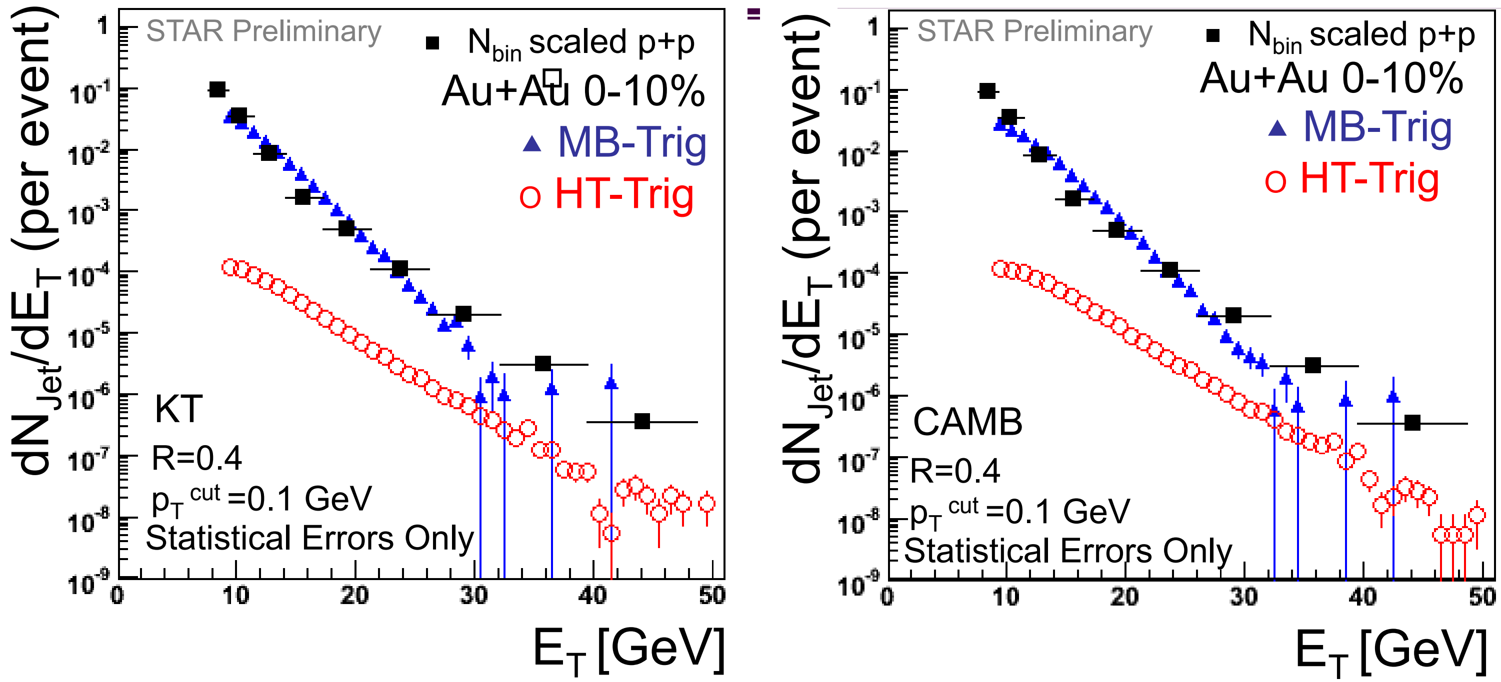
ρ = Diffuse noise, σ=noise fluctuations

M. Cacciari, G. Salam, G. Soyez 0802.1188 [hep-ph]

Reduction of background fluctuations: p_T cuts, limit R.

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

Jets with Sequential Recombination Algorithm



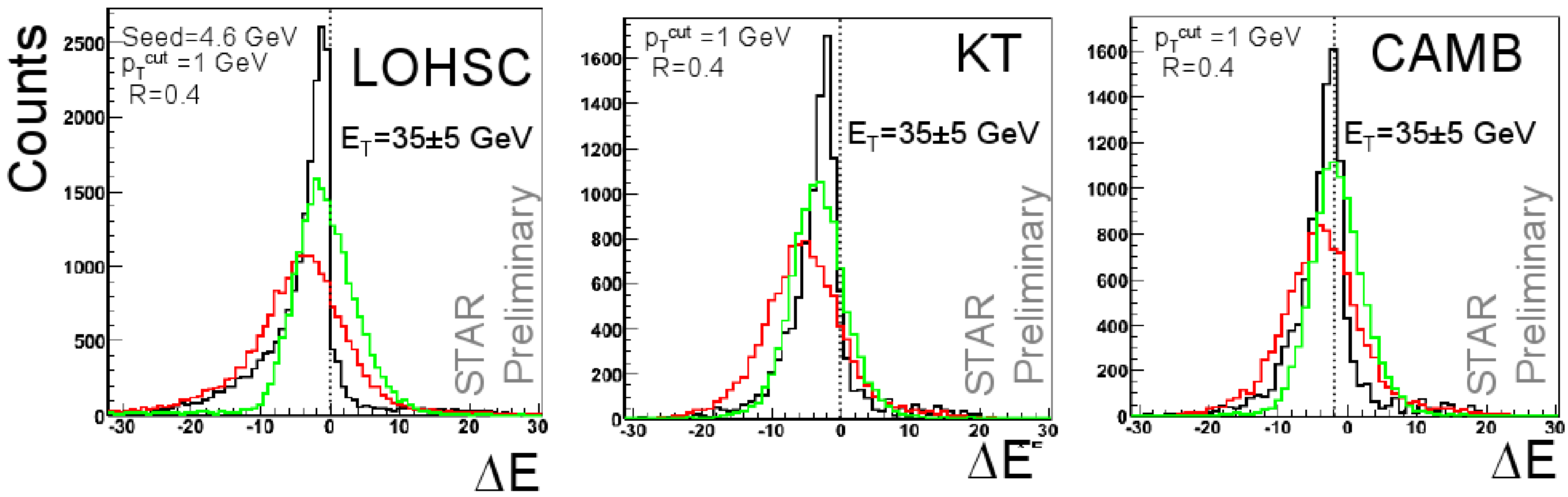
KT & CAMB biases are different wrt. LOHSC due to:

- -- background subtraction algorithm
- -- no seed
- -- low pt cut

Systematic Uncertainity on Normalization: 50%

Good agreement with N_{bin} scaled p+p for unbiased algorithms.

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



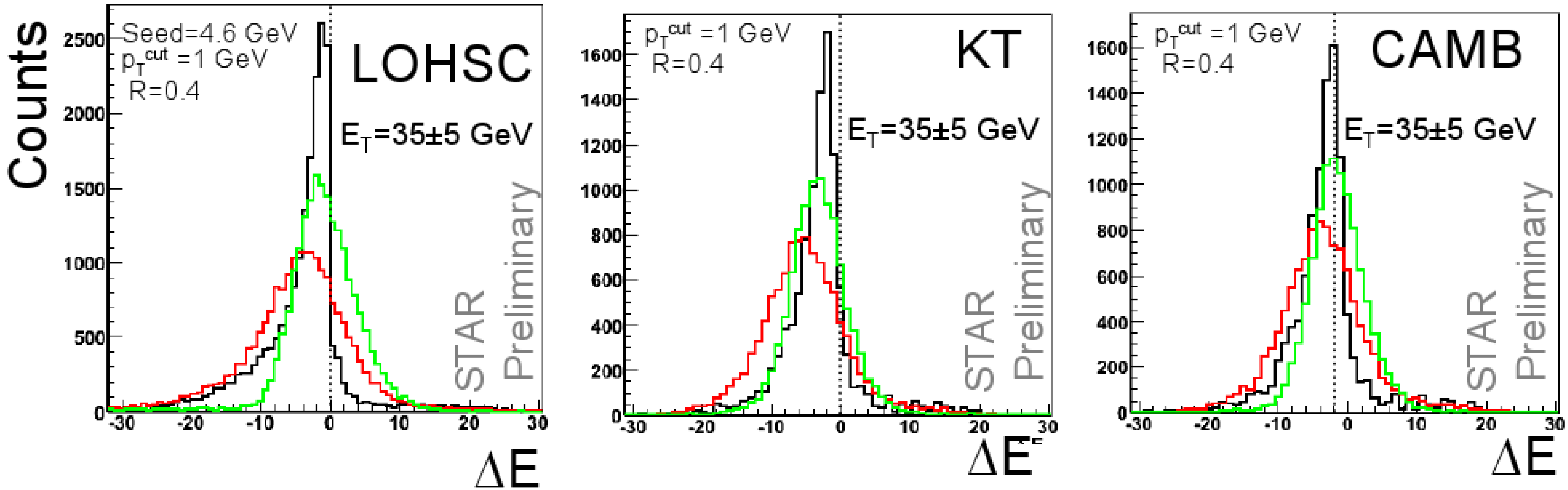
Red curves: $\Delta E = E^{\text{PyEmbed}} - E^{\text{PyDet}}$

with PyDet $\equiv$ Pythia pp + detector effects

PyEmbed $\equiv$ same with real AuAu event added

$\Rightarrow$ measure of the subtraction efficiency

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



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

with PyDet $\equiv$ Pythia pp + detector effects

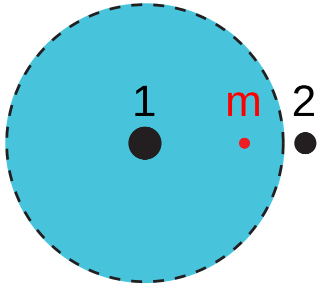
PyEmbed $\equiv$ same with real AuAu event added

$\Rightarrow$ measure of the subtraction efficiency

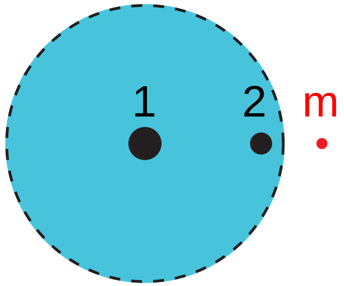
Work under progress: removing the last few GeV shift in ΔE

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

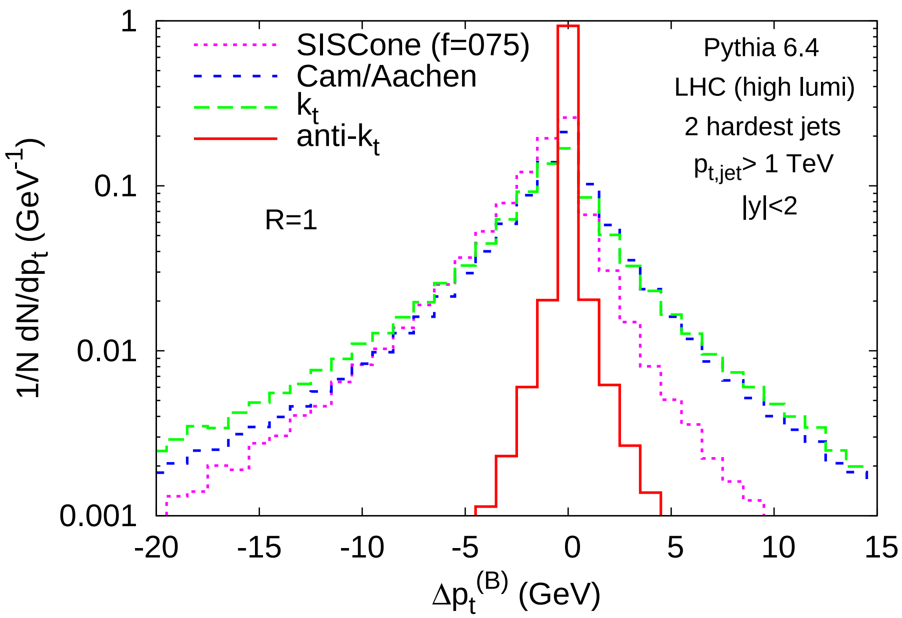


  - loss: p_2 lost when adding p_m



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



Back-reaction & HI UE

Additional soft background has 2 effects:

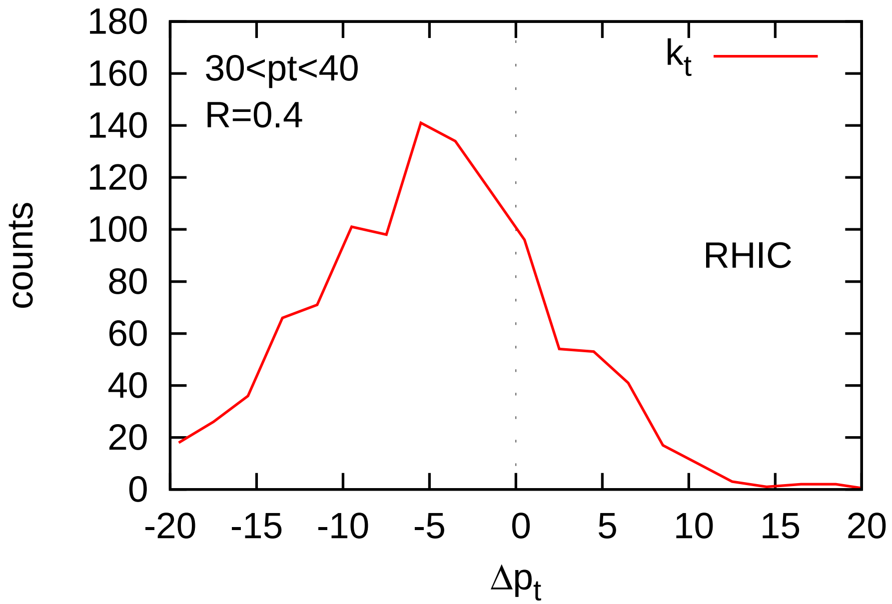
- Throw soft particles in the hard jet: dealt with by subtraction
- Modify the hard scattering (back-reaction)
  - can be pointlike or diffuse
  - tractable analytically (similar to areas)
  - $k_t \gtrsim$ Cambridge > SISCone $\gg$ anti-k_t

+ For heavy-ion collisions: fluctuating underlying event background

$\rightarrow$ median estimation of ρ might oversimplified

Results (3)

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

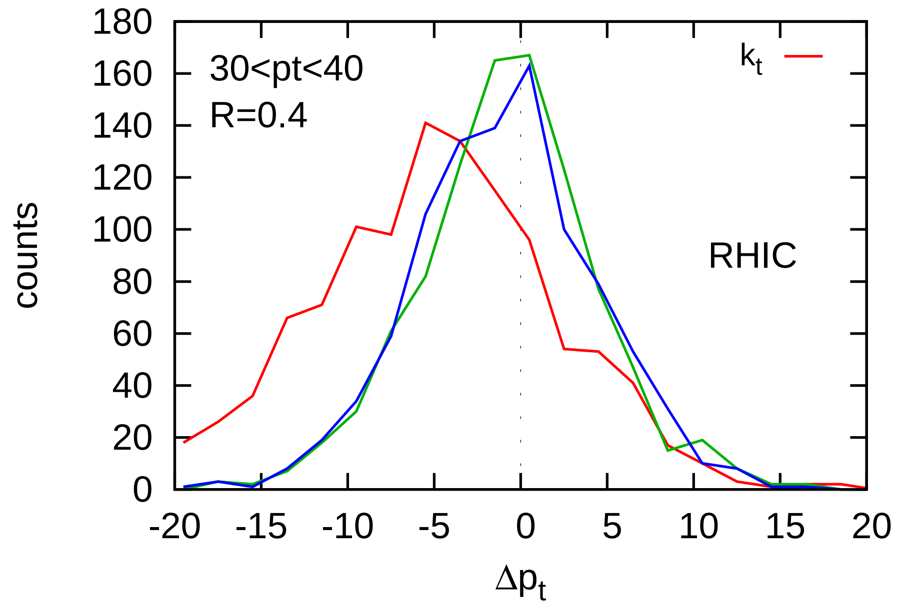


Similar shifts than STAR

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 $R_{\rm best}$
  - $R_{\rm best}$ increases with the scale
  - same for quark and gluon jets, larger $R_{\rm best}$ for gluons
  - with subtraction, pileup has reasonably little influence

  $\Rightarrow$ flexibility in jet physics at the LHC

- Jet-finding in AA at RHIC and the LHC
  - First measurement at RHIC
  - Work under progress: improve subtraction down to 1-2 GeV