HPHD
Paris, Ecole Polytechnique
May 30, 2011

Jet Reconstruction Algorithms in HI Environment: Successes and Failures

Matteo Cacciari
LPTHE - Paris 6,7 and CNRS

In collaboration with G. Salam and G. Soyez

HPHD
Paris, Ecole Polytechnique
May 30, 2011

Jet Reconstruction Algorithms in HI Environment: Successes and ~~Failures~~

Matteo Cacciari
LPTHE - Paris 6,7 and CNRS

In collaboration with G. Salam and G. Soyez

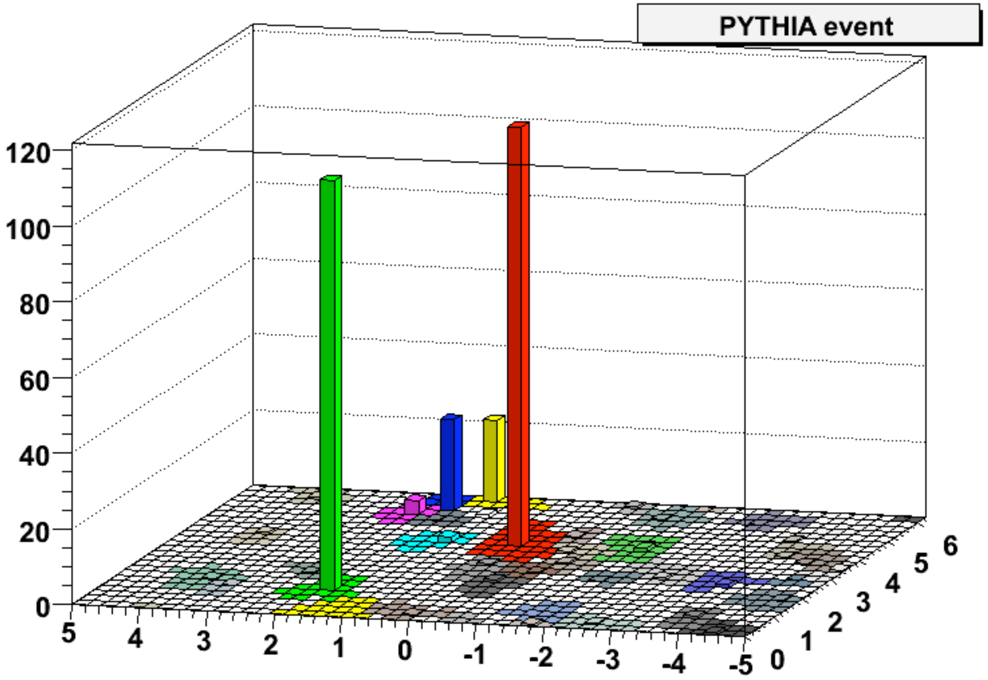
HPHD
Paris, Ecole Polytechnique
May 30, 2011

Jet Reconstruction Algorithms in HI Environment: Successes and ~~Failures~~ Limitations

Matteo Cacciari
LPTHE - Paris 6,7 and CNRS

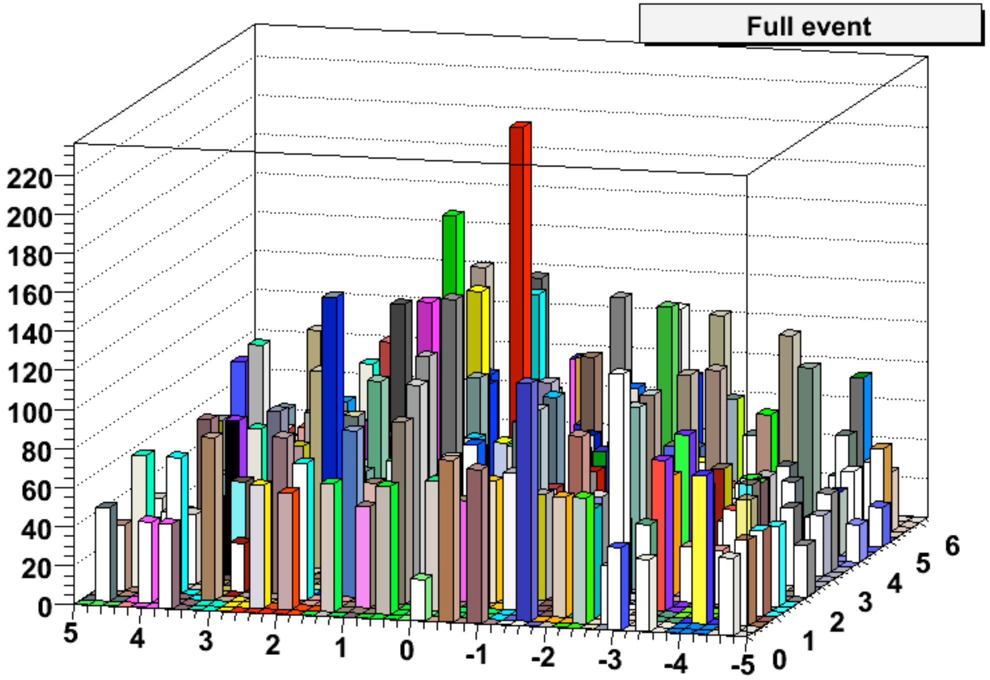
In collaboration with G. Salam and G. Soyez

Hard jets and background



Hard jets

(pp collisions)



Hard jets + background

(AA collisions)

- To what extent can we ‘reconstruct’ the hard jets?
- How are specific observables affected by limitations in the reconstruction?

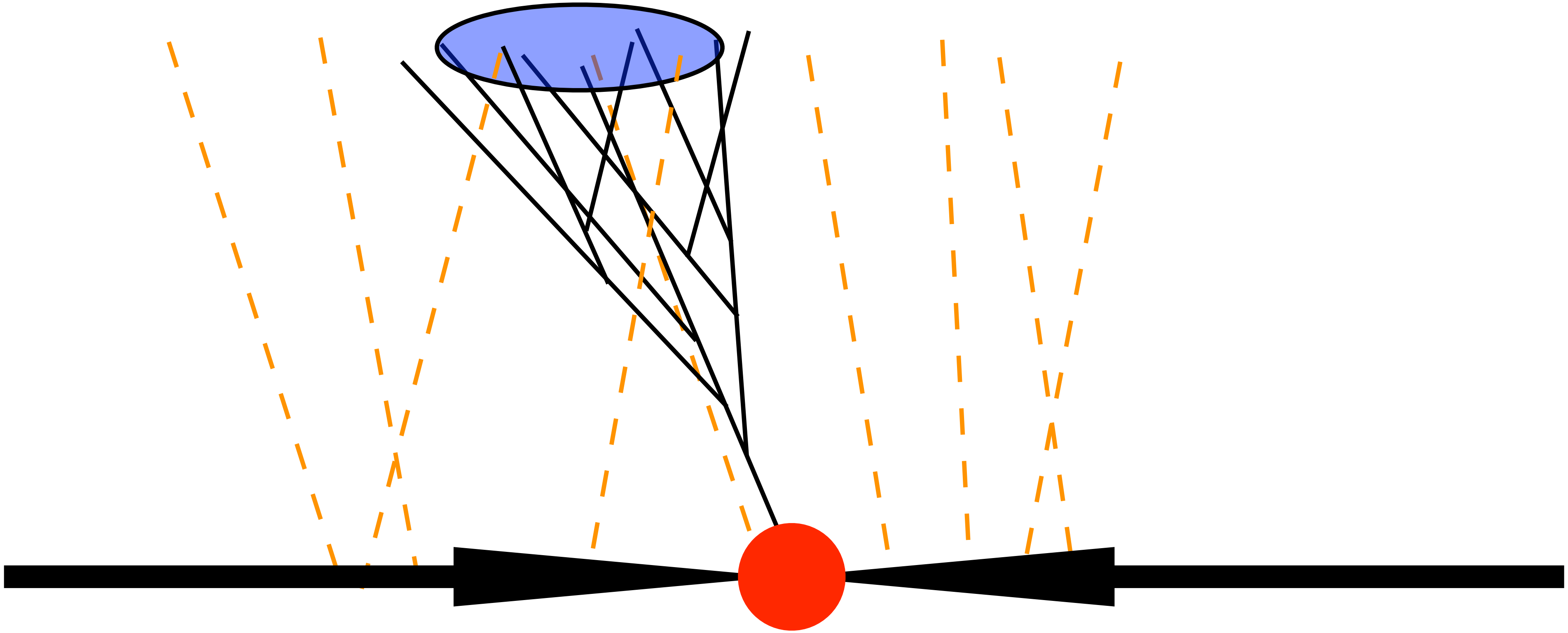
Hard jets and background

How are the hard jets modified by the background?

Susceptibility (how much bkgd gets picked up)

Resiliency (how much the original jet changes)

Susceptibility



The larger a jet is, the more it is contaminated by background radiation

If a jet algorithm does not return jets with a fixed area, this needs to be calculated on a jet-by-jet basis

Susceptibility: jet area

MC, Salam, Soyez, arXiv:0802.1188

Operational definition of **active jet area:**

Add many **ghost-particles** of infinitesimally small momentum to the hard event.
Cluster them together with the real particles,
and count how many on average get clustered within a given jet.

Number of ghosts in jet J

$$A(J\,|\,\{g_i\}) = \frac{N_g(J)}{\nu_g}$$

Active area of a single ghosts configuration

Ghost density

$$A(J) = \lim_{\nu_g \to \infty} \left\langle A(J\,|\,\{g_i\}) \right\rangle_g$$

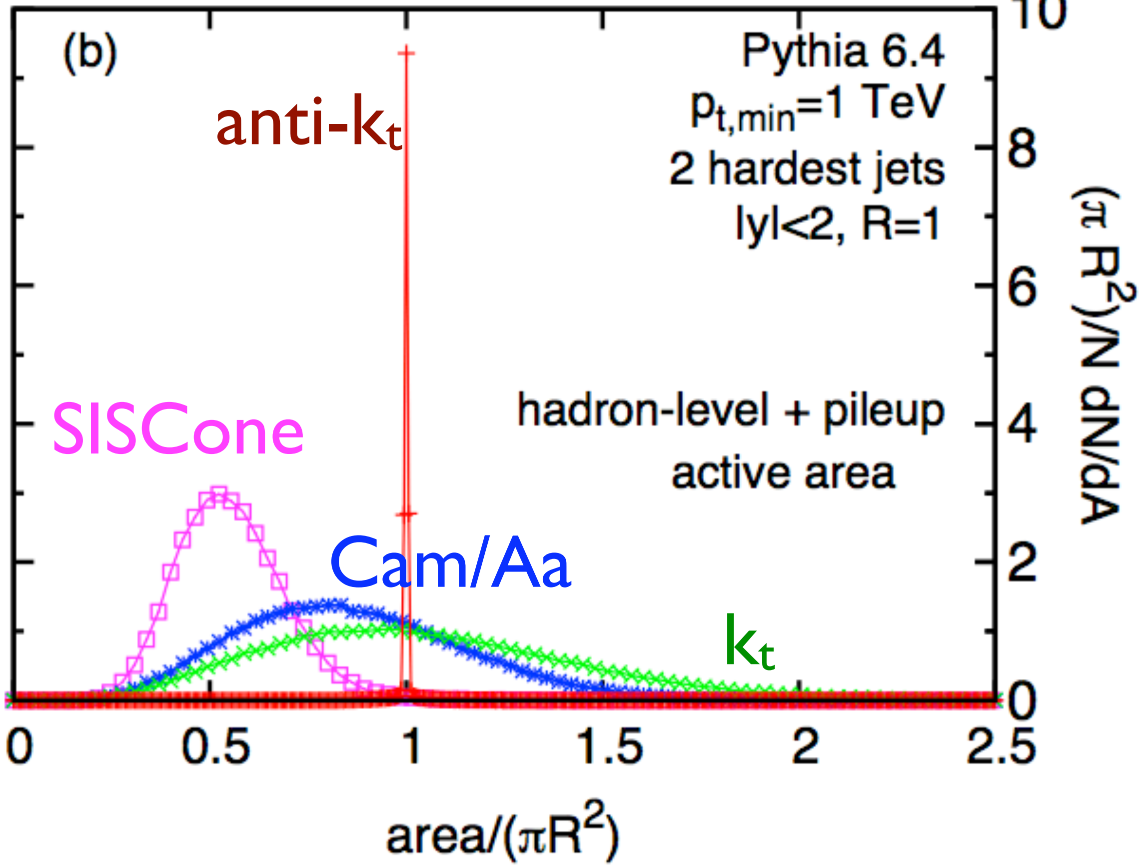
Active area

A jet is not (always) a cone

The typical area of a jet around a jet is **not** necessarily πR^2

1-particle areas	k_t	Cam/Aa	SISCone	anti-k_t
$\langle A\rangle/\pi R^2$	0.81	0.81	1/4	1

Area distribution



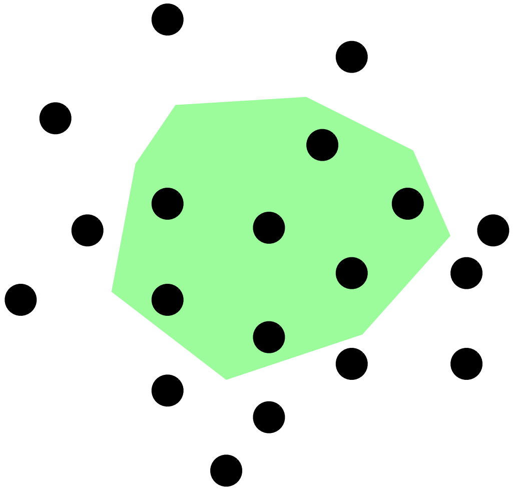
Only anti-k_t has the behaviour one would naively expect

(Note also the small fluctuations of the anti-k_t area. Actually irrelevant, though, if measured jet-by-jet)

Back-reaction

"How (much) a jet changes when immersed in a background"

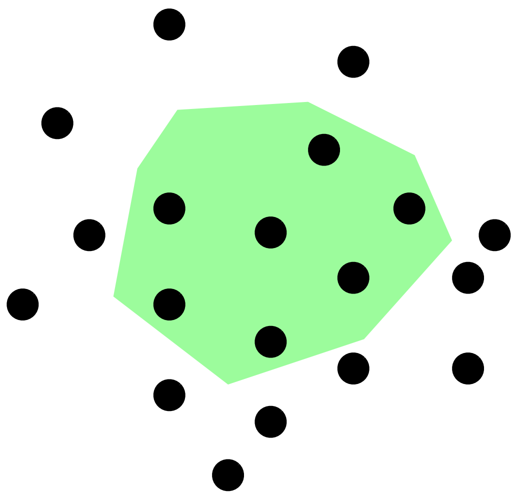
Without background

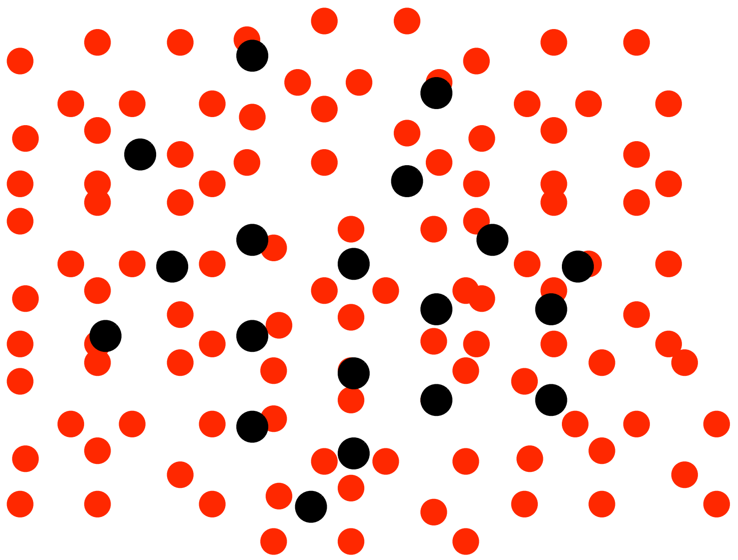


Back-reaction

"How (much) a jet changes when immersed in a background"

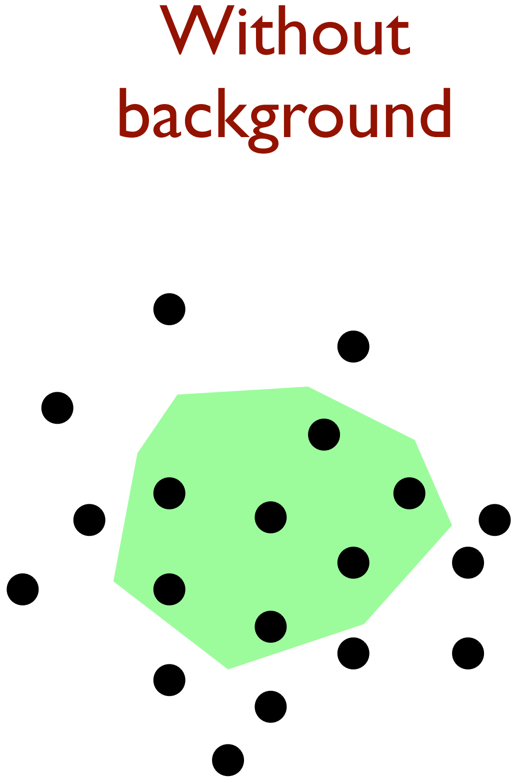
Without background

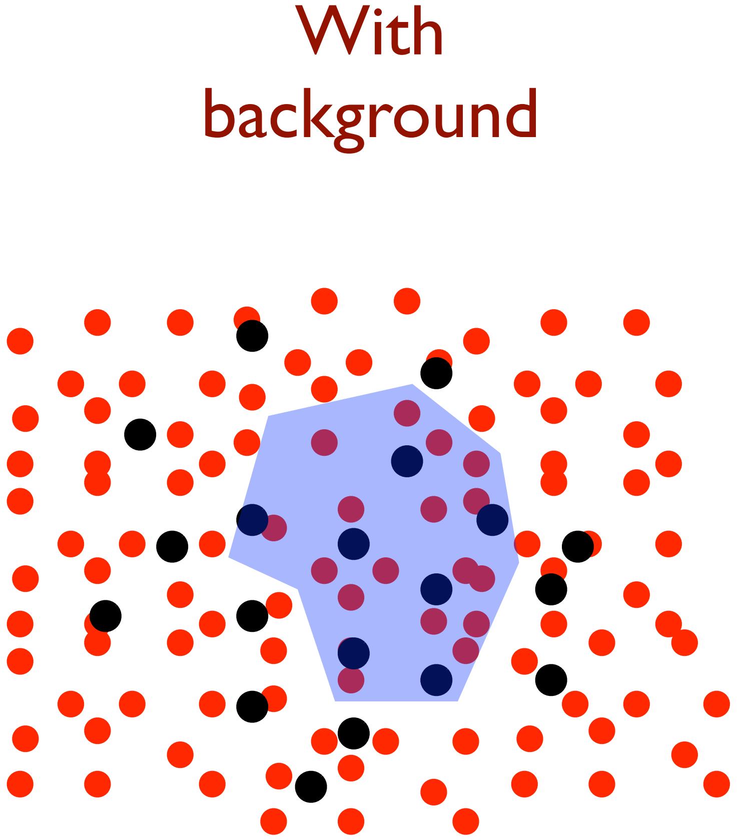




Back-reaction

"How (much) a jet changes when immersed in a background"


Without background

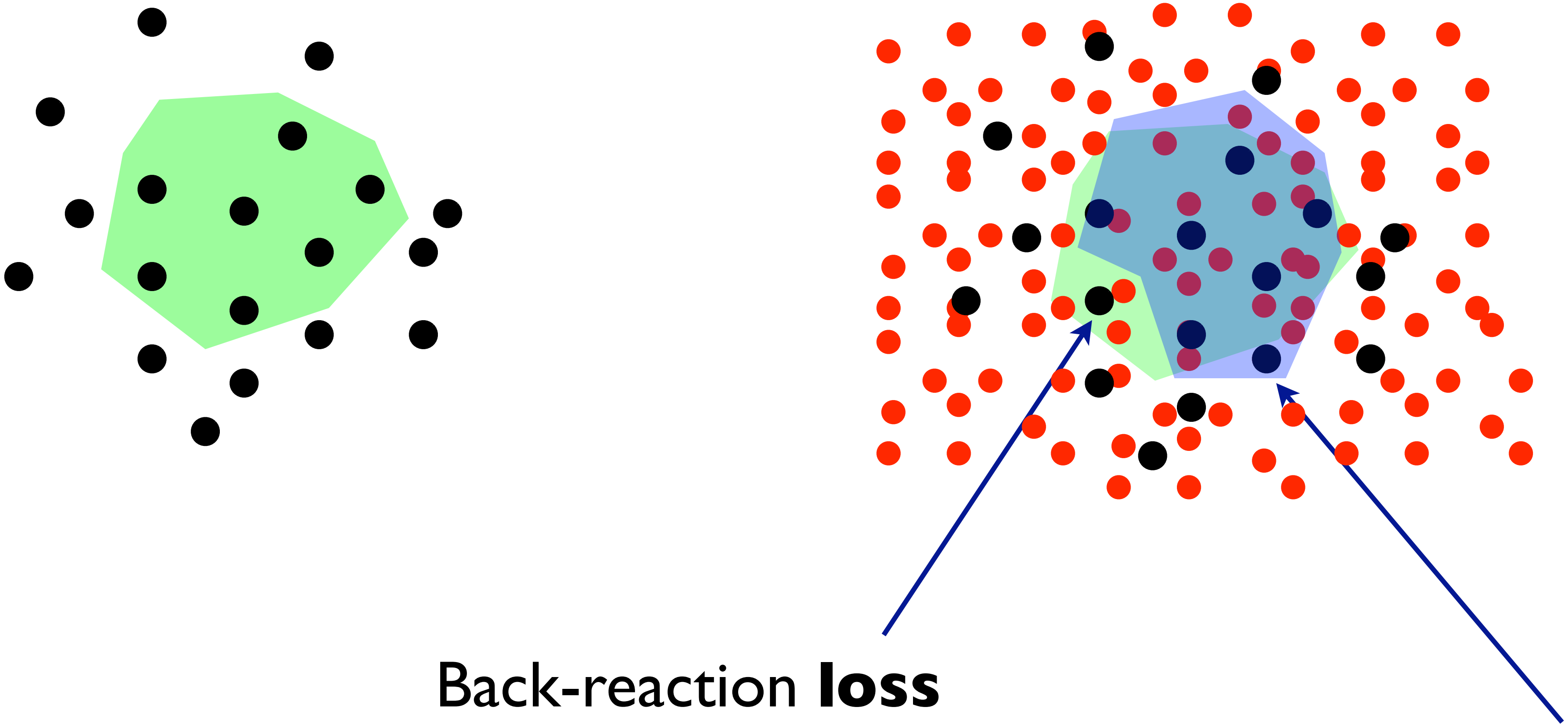

With background

Back-reaction

"How (much) a jet changes when immersed in a background"

Without background

With background

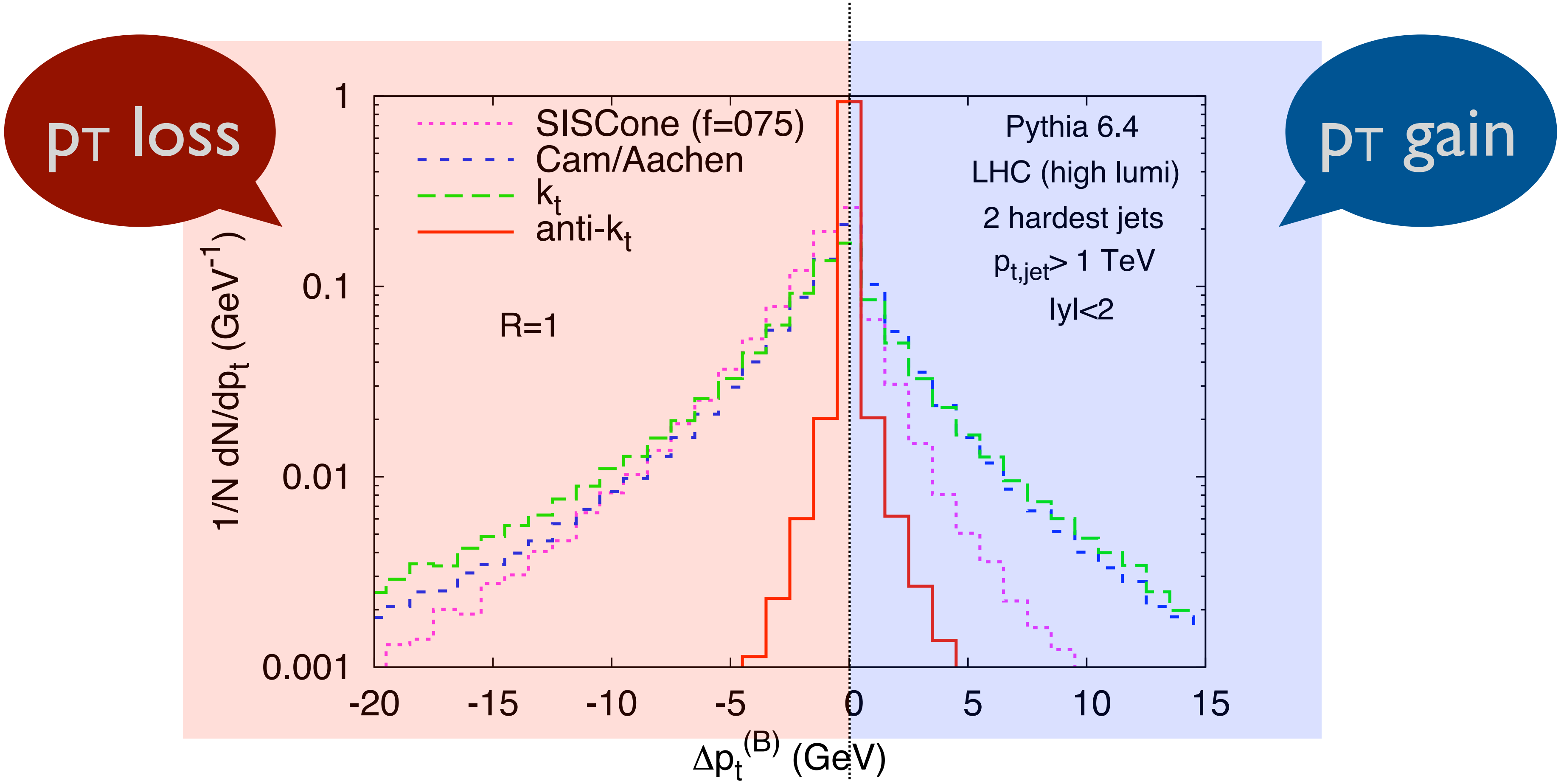


Back-reaction **loss**

Back-reaction **gain**

Back-reaction

MC, Salam, Soyez, arXiv:0802.1188



Anti-k_t jets are much more resilient to changes from background immersion

The IRC safe jet algorithms

k_t	SR $d_{ij} = \min(k_{ti}^2, k_{tj}^2)\Delta R_{ij}^2/R^2$ hierarchical in rel p_t	Catani et al '91 Ellis, Soper '93	NlnN
Cambridge/ Aachen	SR $d_{ij} = \Delta R_{ij}^2/R^2$ hierarchical in angle	Dokshitzer et al '97 Wengler, Wobish '98	NlnN
anti-k_t	SR $d_{ij} = \min(k_{ti}^{-2}, k_{tj}^{-2})\Delta R_{ij}^2/R^2$ gives perfectly conical hard jets	MC, Salam, Soyez '08 (Delsart, Loch)	$N^{3/2}$
SISCone	Seedless iterative cone with split-merge gives 'economical' jets	Salam, Soyez '07	$N^2 \ln N$

We call these algs 'second-generation' ones

All are available in FastJet, http://fastjet.fr

(As well as many IRC unsafe ones)

Cambridge/Aachen with filtering

Butterworth, Davison, Rubin, Salam, arXiv:0802.2470

An example of a **third-generation** jet algorithm

- Cluster with C/A and a given R
- Undo the clustering of each jet down to subjets with radius $\mathbf{x_{filt}R}$
- Retain only the $\mathbf{n_{filt}}$ hardest subjets

Idea: filter out soft background, retain hard core

(for this work we'll be using $x_{filt} = 0.5$, $n_{filt} = 2$)

The IRC safe algorithms

	Speed	Regularity	UE	Backreaction	Hierarchical substructure
k_t	☺ ☺ ☺	☂	☂☂	☁☁	☺ ☺
Cambridge /Aachen	☺ ☺ ☺	☂	☂	☁☁	☺ ☺ ☺
anti-k_t	☺ ☺ ☺	☺ ☺	☁/☺	☺ ☺	✗
SISCone	☺	☁	☺ ☺	☁	✗

Hard jets and background

MC, Salam, arXiv:0707.1378
MC, Salam, Soyez, arXiv:0802.1188

Modifications of the hard jet

$$p_{t,jet}^{AA} = p_{t,jet}^{pp} + \boxed{\rho A_{jet} \pm \sigma \sqrt{A_{jet}}} + \boxed{\Delta p_t^{BR}}$$

hard jet

background

back-reaction

‘susceptibility’

‘resiliency’

Jet reconstruction quality

Reconstruct the momentum the hard jet would have without the background:

MC, Salam, arXiv:0707.1378

$$p^{sub}_{\mu,jet} \equiv p_{\mu,jet} - \rho A_{\mu,jet}$$

(subtracts background,
fluctuations and back-reaction remain)

Jet reconstruction quality

Reconstruct the momentum the hard jet would have without the background:

MC, Salam, arXiv:0707.1378

$$p^{sub}_{\mu,jet} \equiv p_{\mu,jet} - \rho A_{\mu,jet}$$

(subtracts background,
fluctuations and back-reaction remain)

Quality measures

'probe'

Offset $$\langle \Delta p_t \rangle \equiv \langle p_t^{AA,sub} - p_t^{pp,sub} \rangle$$

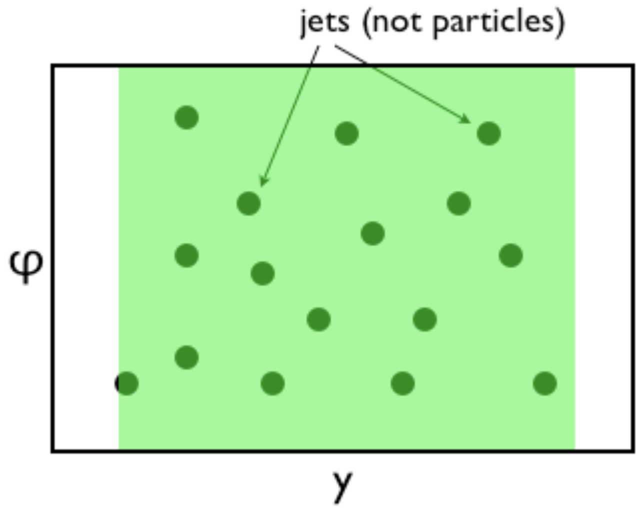
Dispersion $$\sigma_{\Delta p_t} \equiv \sqrt{\langle \Delta p_t^2 \rangle - \langle \Delta p_t \rangle^2}$$

Small offset and dispersion will indicate a good reconstruction

Background determination

In order to subtract the background, one must first determine it

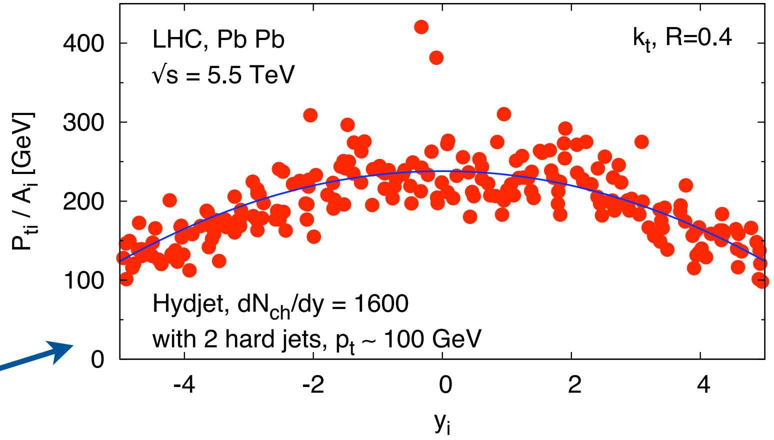
Proposal in 2007 paper (MC, Salam, arXiv:0707.1378)



- either, choose a region in rapidity-azimuth plane where the background is uniform
  - calculate ρ (p_t per unit area) as

$$\rho \equiv \text{median}\left[\left\{\frac{p_t^{jet}}{\text{Area}_{jet}}\right\}\right]$$

- or, account for rapidity dependence of background by fitting a quadratic function to $p_{t,jet}/\text{Area}_{jet}$ distribution



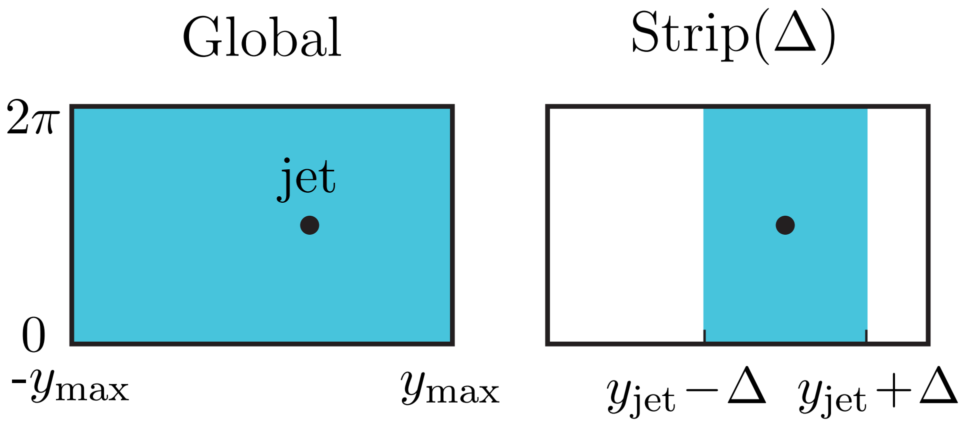
This way to account for rapidity dependence of background turns out to be insufficiently accurate

⇒

Adapt the median method to a varying background

Background determination: the ranges

Ranges can now be *fixed*, or *local* (tied to a jet's position)



Choose a range **such that you expect the background to be uniform within it,** place it **where you need it.**

Use median operation within each local range of interest

A range should be ***not too large*** (to avoid non-uniformity of background) ***nor too small*** (to have sufficient statistics for the median operation).

We find $\text{Area}_{\text{range}} \geq 25R^2$ to be a reasonable lower limit

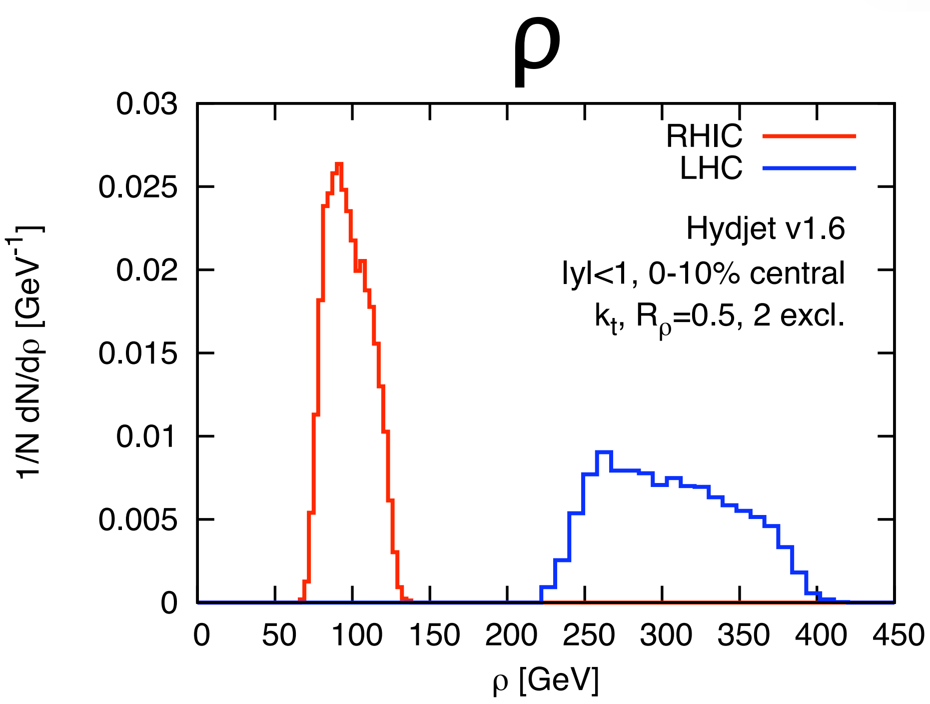
Background determination: the median

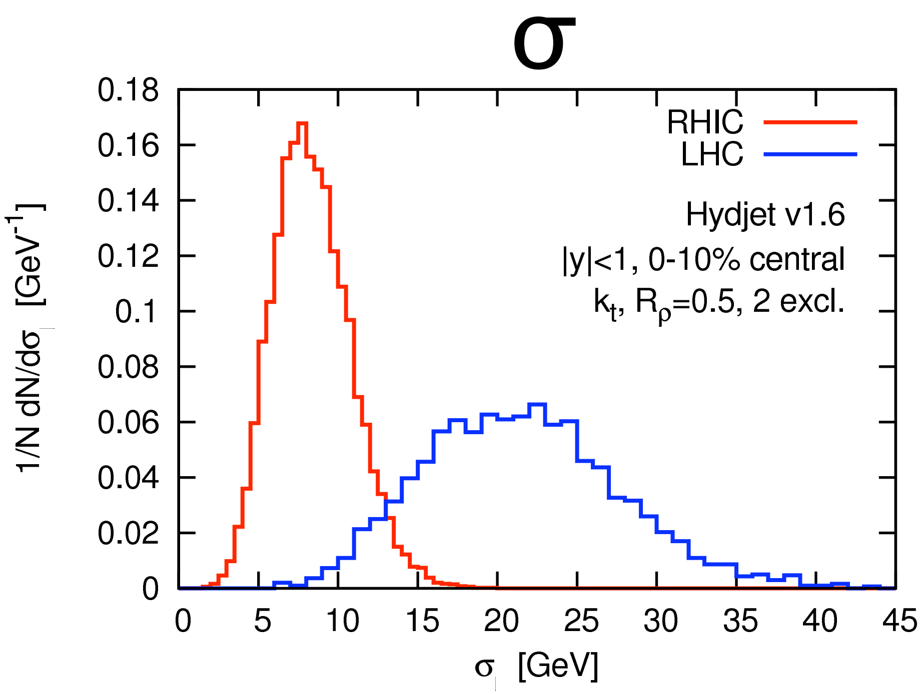
MC, Salam, arXiv:0707.1378

$$\rho \equiv \underset{jet \in range}{\text{median}} \left[\left\{ \frac{p_t^{jet}}{\text{Area}_{jet}} \right\} \right]$$

- •Should be used only with algorithms like k_t or Cambridge/Aachen (but the subtraction can then be performed on jets of any algorithm)
- •Works on **an event-by-event basis** (this removes many fluctuations)
- •One can also explicitly remove the hard(est) jet(s) before taking the median, to reduce a potential bias from the hard jets in the event

The background


ρ


σ

Typical values (depend on model):

Hydjet v1.6	$dN_{ch}/d\eta\|_{\eta=0}$	ρ (GeV) (y=0, 0-10%)	σ (GeV)
RHIC	658 (0-6%)	100	8
LHC 5.5 TeV	1570 (0-10%)	310	21
LHC 2.76 TeV	1400 (0-10%)	210	17

How does the background affect the jet reconstruction?

How do the different algorithms fare?

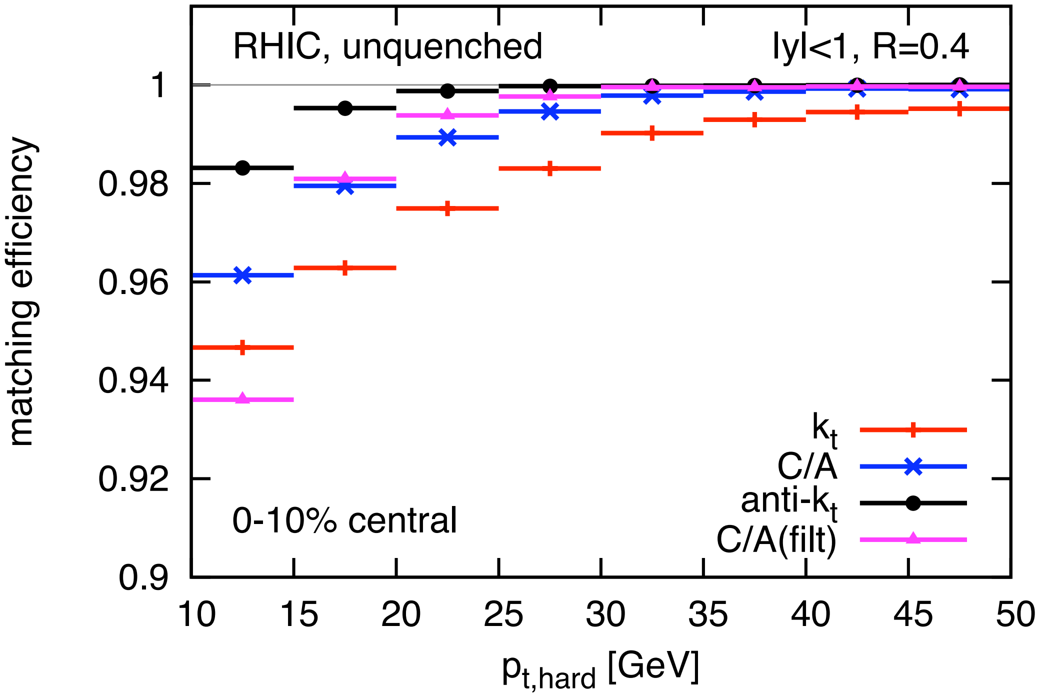
Offset $$\langle \Delta p_t \rangle \equiv \langle p_t^{AA,sub} - p_t^{pp,sub} \rangle$$

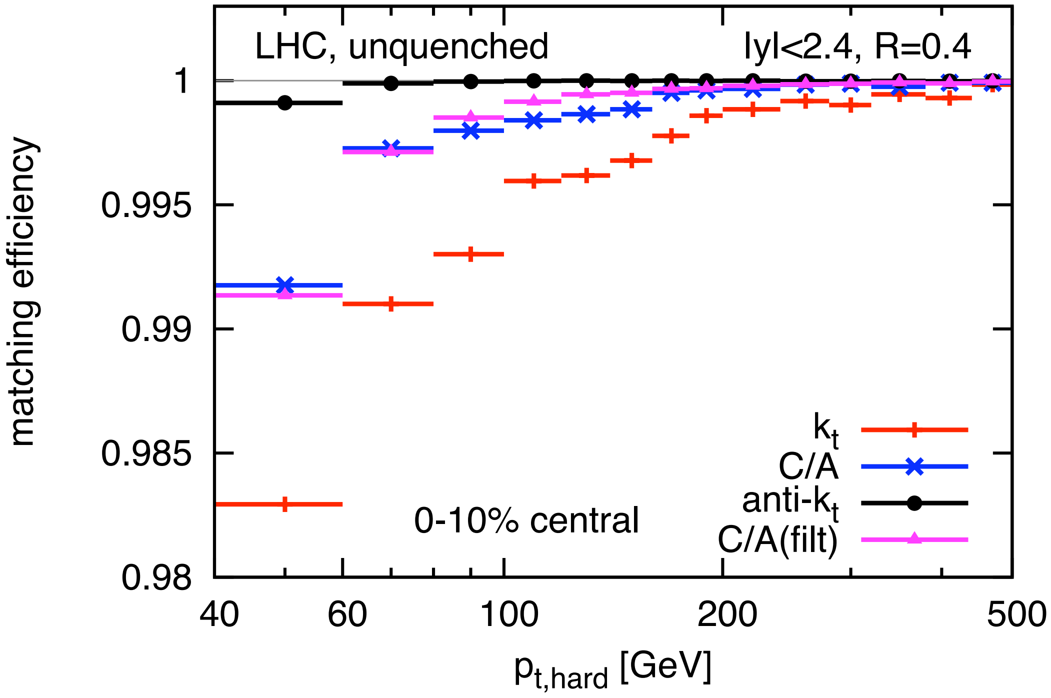
Dispersion $$\sigma_{\Delta p_t} \equiv \sqrt{\langle \Delta p_t^2 \rangle - \langle \Delta p_t \rangle^2}$$

MC, Rojo, Salam, Soyez, 1010.1759

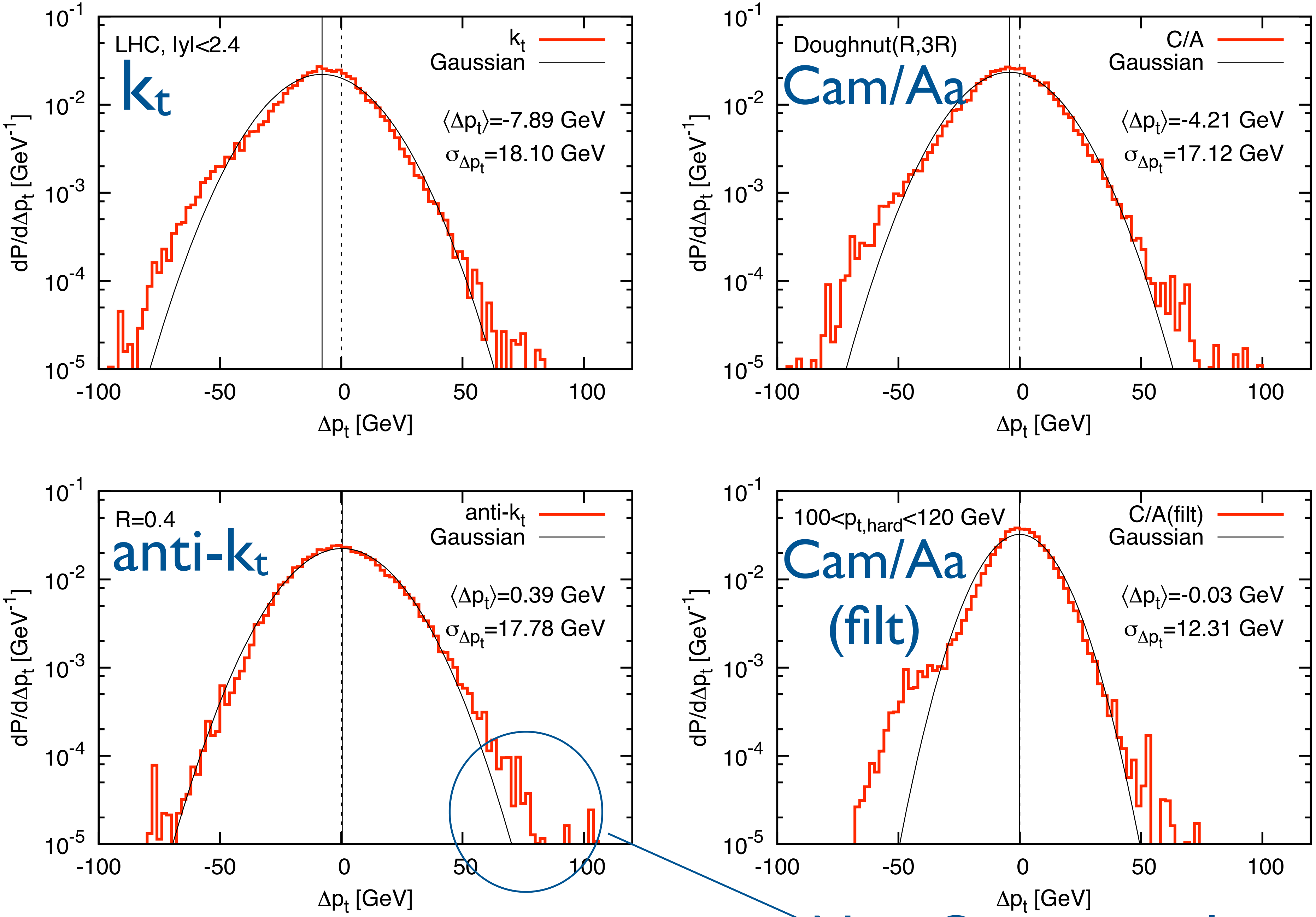
Reconstruction efficiency

A jet reconstructed in the full event is considered **matched** to a hard jet *if the constituents common to both the hard and the full jet make up at least 50% of the transverse momentum of the constituents of the hard jet*





Δp_t distributions in PbPb at LHC



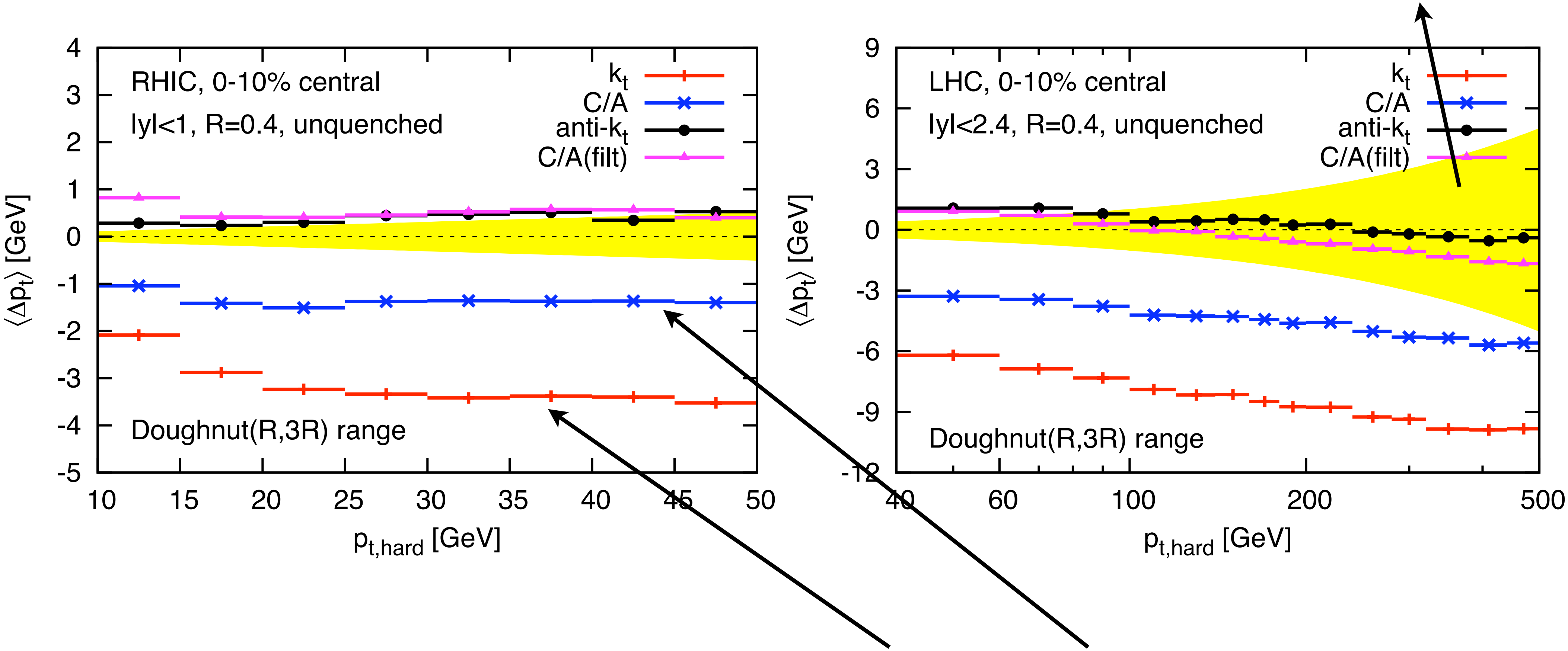
Non-Gaussian tails

<Δp_t>

anti-k_t and C/A(filt) fare best

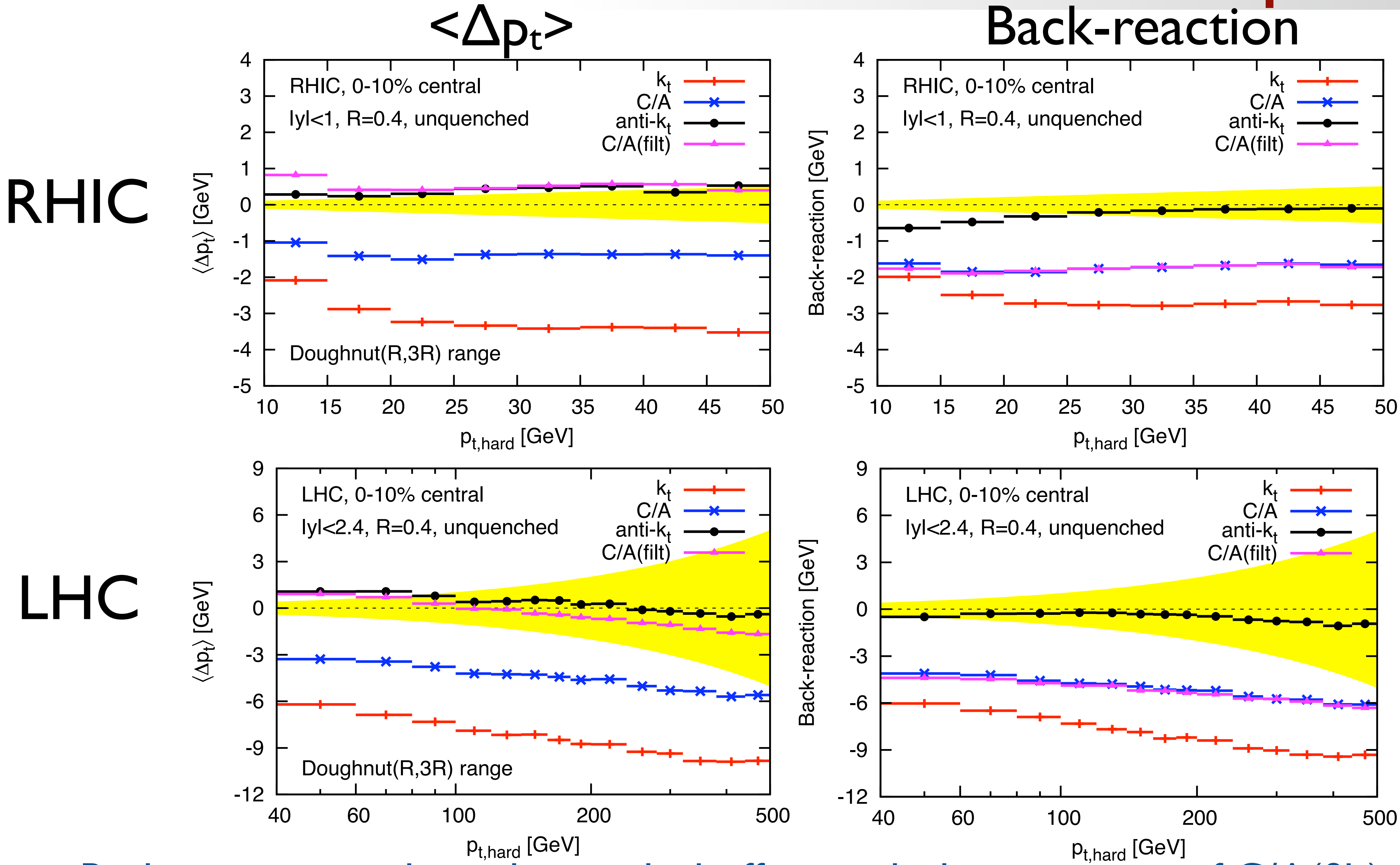
(Results are fairly independent of p_t)

Yellow band:
1% accuracy



The residual offset of k_t and C/A can be interpreted as an effect of the **back-reaction**

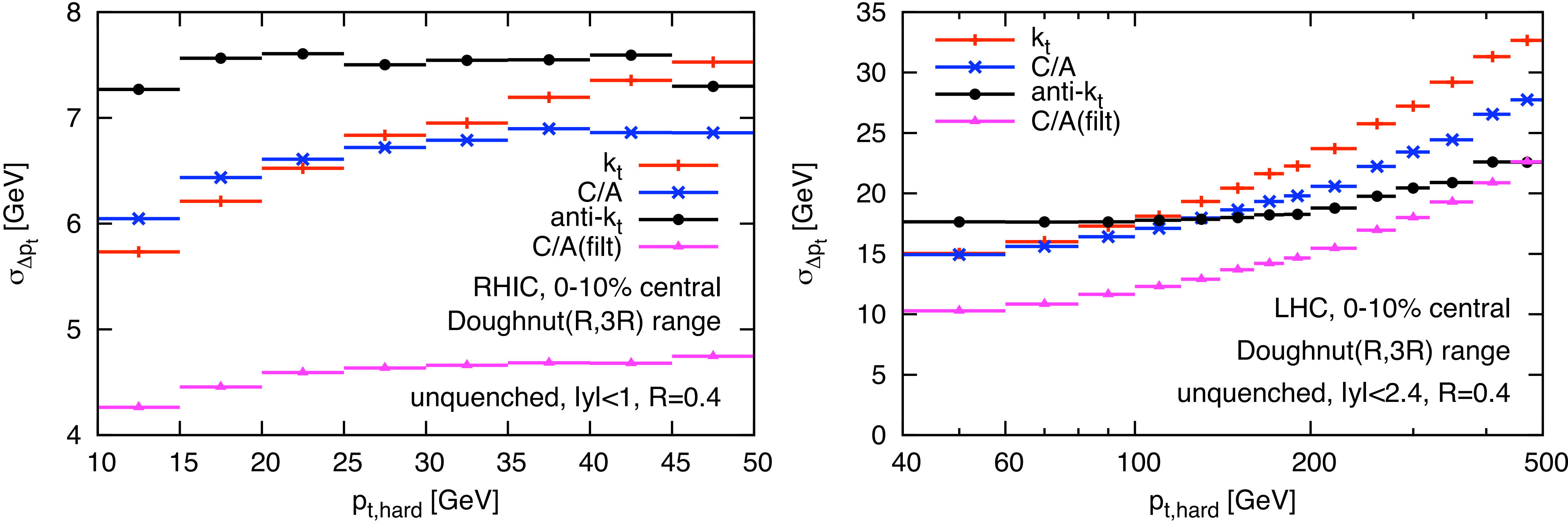
Back-reaction contribution to $<\Delta p_t>$



Back-reaction explains the residual offset, with the exception of C/A(filt)

(accidental compensation of back-reaction and positive offset)

Dispersion of $\Delta p_t = \sigma_{\Delta pt}$



- C/A(filt) markedly better, as a consequence of its smaller effective area
- Dispersions increase at large p_t, probably as a consequence of a larger dispersion of back-reaction
- anti-k_t remains fairly constant ('resiliency'), and eventually becomes better at large p_t

Summary

- Jet area/median method for background determination and subtraction validated in simulated collisions at RHIC and LHC: ***high efficiency, small or almost zero <Δp_t> offset*** (though each different jet algorithm has characteristics which affect the subtraction in specific ways (e.g. back-reaction))

- Irreducible dispersions are left, and may of course play an important role in measurements like the inclusive cross section (fakes rate). Their size also depends on the algorithm used.

- anti-k_t turns out to have the safest smallest offset, filtering algorithms have the smallest dispersion (but may be more affected by quenching)

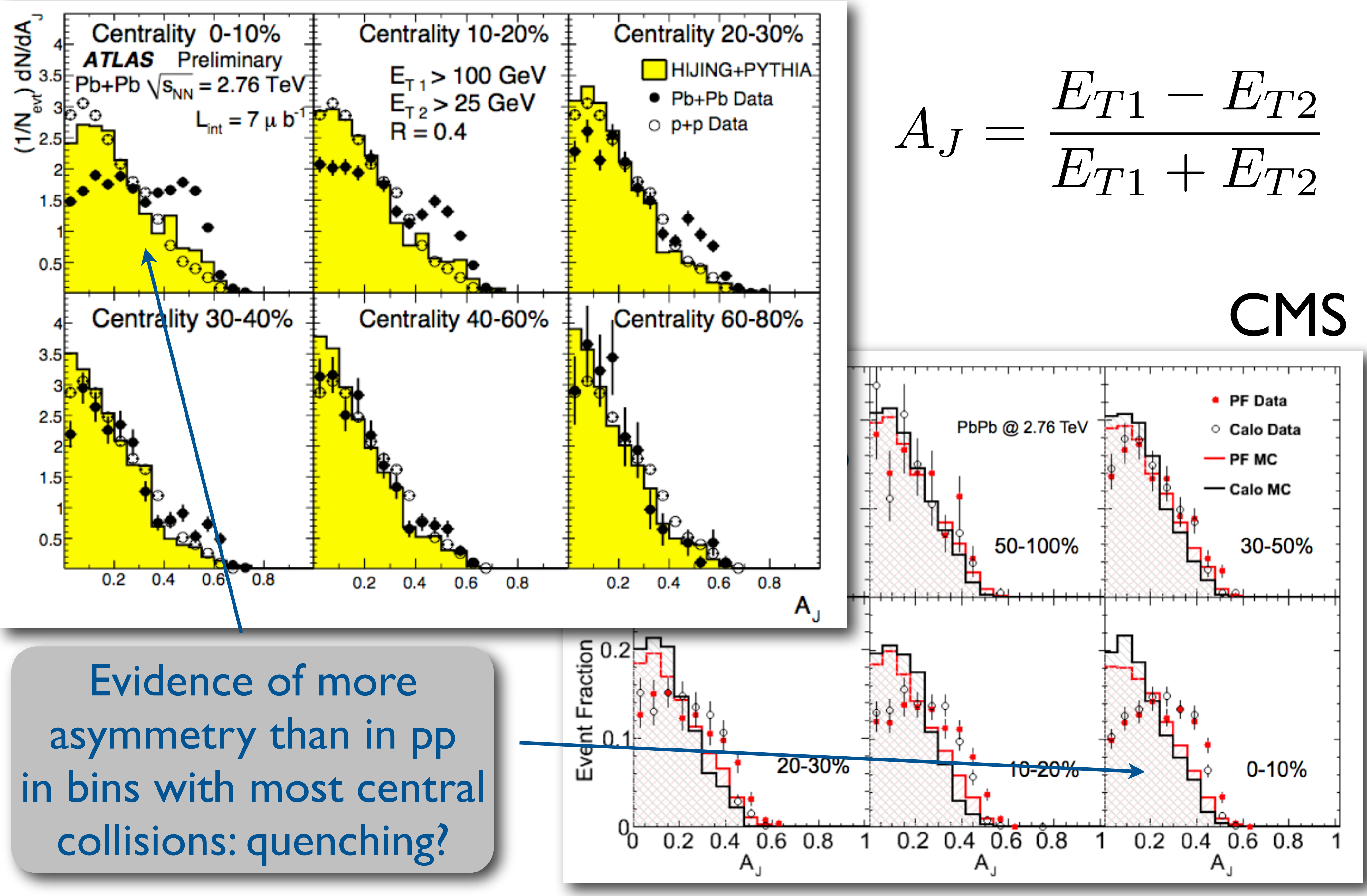
What do we do with this tool?

The LHC dijet asymmetry

ATLAS

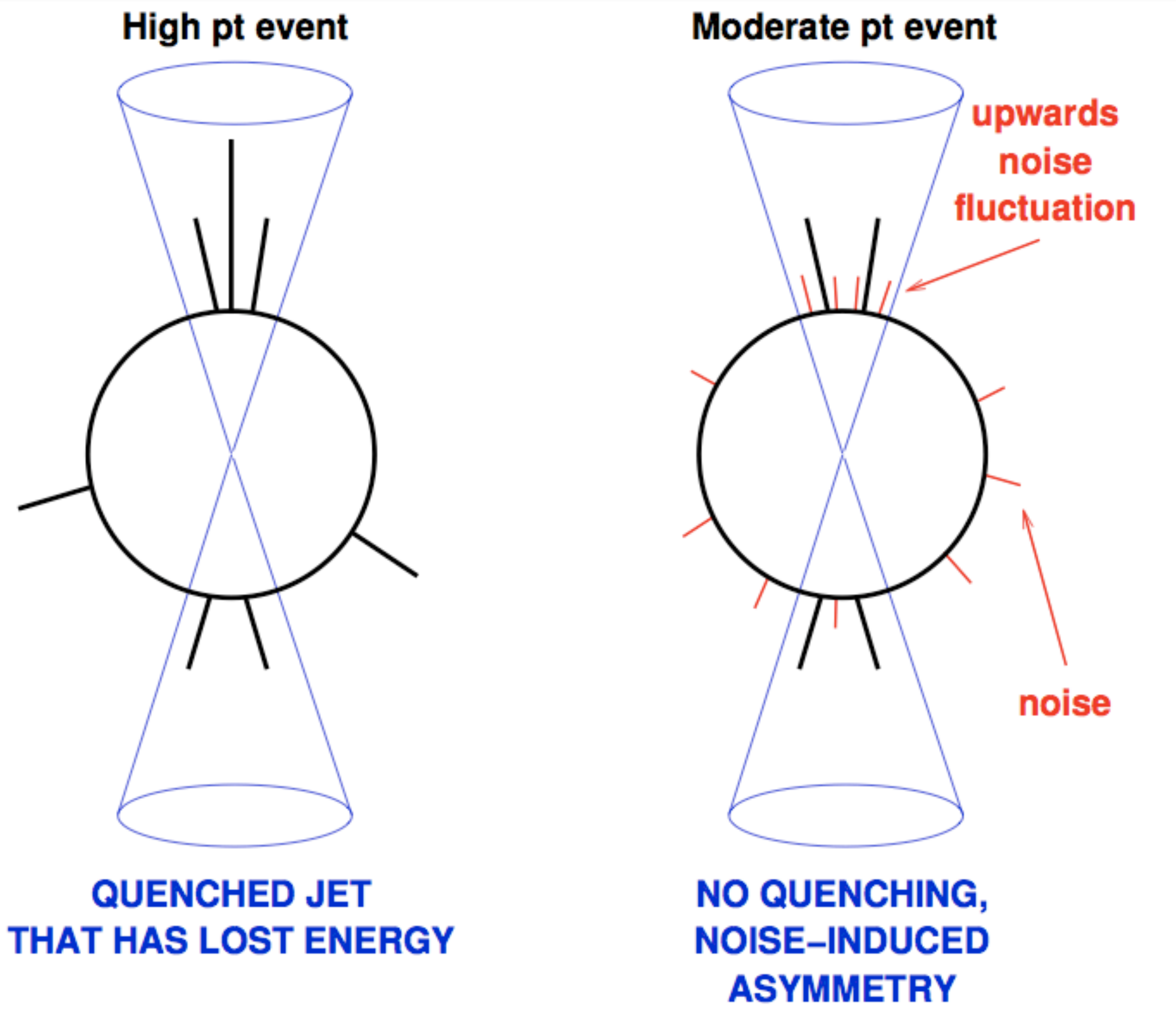
$$A_J = \frac{E_{T1} - E_{T2}}{E_{T1} + E_{T2}}$$

CMS



Evidence of more asymmetry than in pp in bins with most central collisions: quenching?

Quenching v. fluctuations



G. Salam

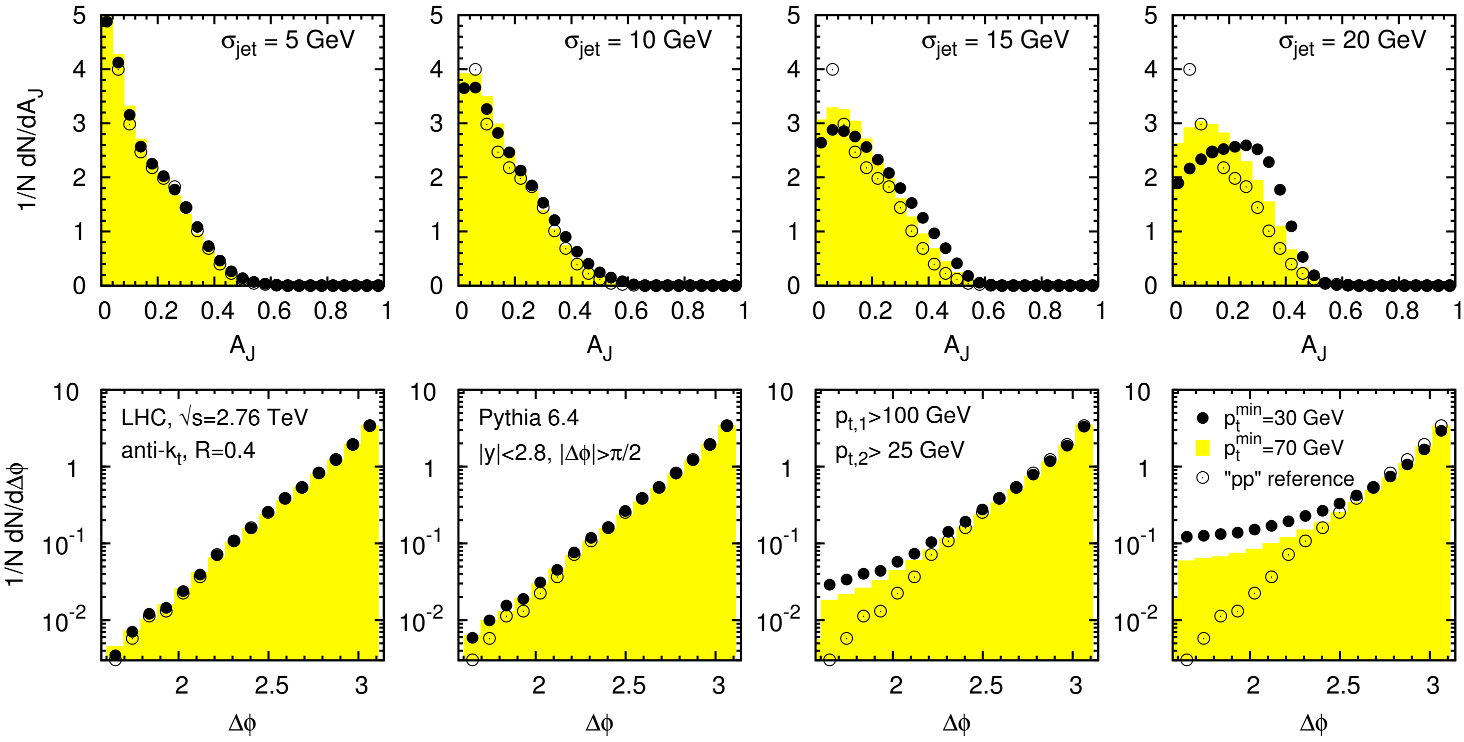
Due to the **steeply falling** p_t spectrum, a **rare upwards fluctuation** at moderate p_t can **contribute significantly** to events at larger p_t

We have seen that residual fluctuations for the anti-k_t algorithm (R=0.4) can be of order 15-20 GeV at the LHC

MC, Salam, Soyez, 1101.2878

Origin of asymmetry?

Add a gaussian smearing to PYTHIA pp jets:
simulates residual fluctuations after subtraction



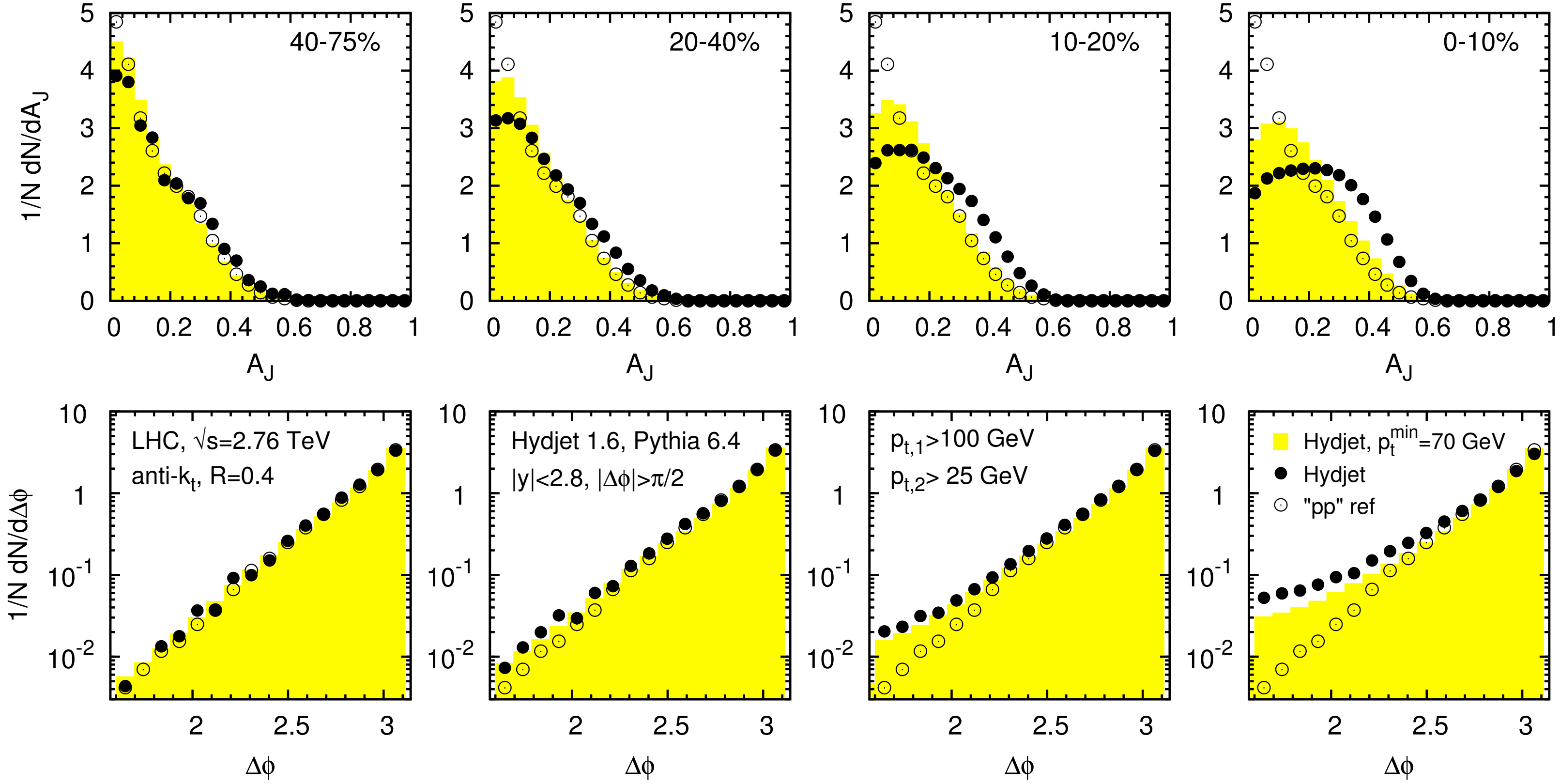
Asymmetry is similar to the one observed by ATLAS and CMS, but no quenching whatsoever is present here

Obviously, the value of σ_{jet} is critical

MC, Salam, Soyez, 1101.2878

HYDJET simulation

Instead of gaussian smearing, full simulation of PbPb events + background subtraction (area/median) + simple calorimeter simulation



Same conclusions:
Asymmetry is similar to the one observed by ATLAS and CMS, but no quenching whatsoever is present here

Note that HYDJET for 0-10% gives **$\sigma_{jet} \approx$ 17 GeV**, but the effects on the asymmetry can be as large as the Gaussian 20 GeV because of non-gaussianities

σ_{jet}

σ_{jet} parametrises the uncertainty left in the knowledge of the p_t of a jet after the background has been subtracted

Its value encompasses both the **physical characteristics of the HI background** and the **procedure used to subtract it**

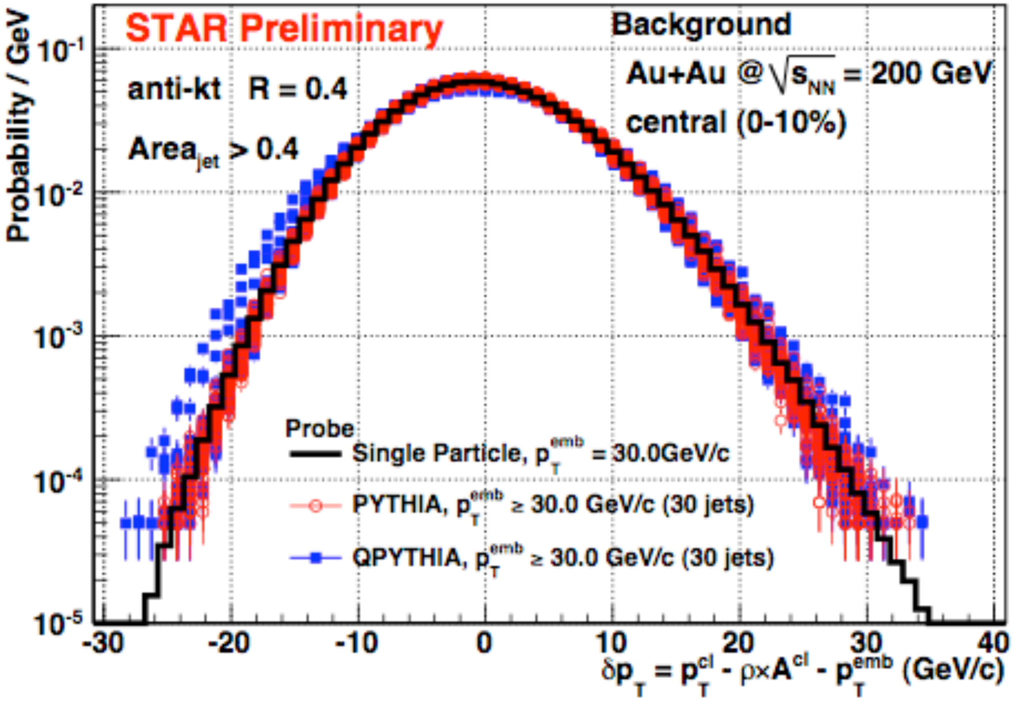
It tells you - quantitatively - how well you are doing

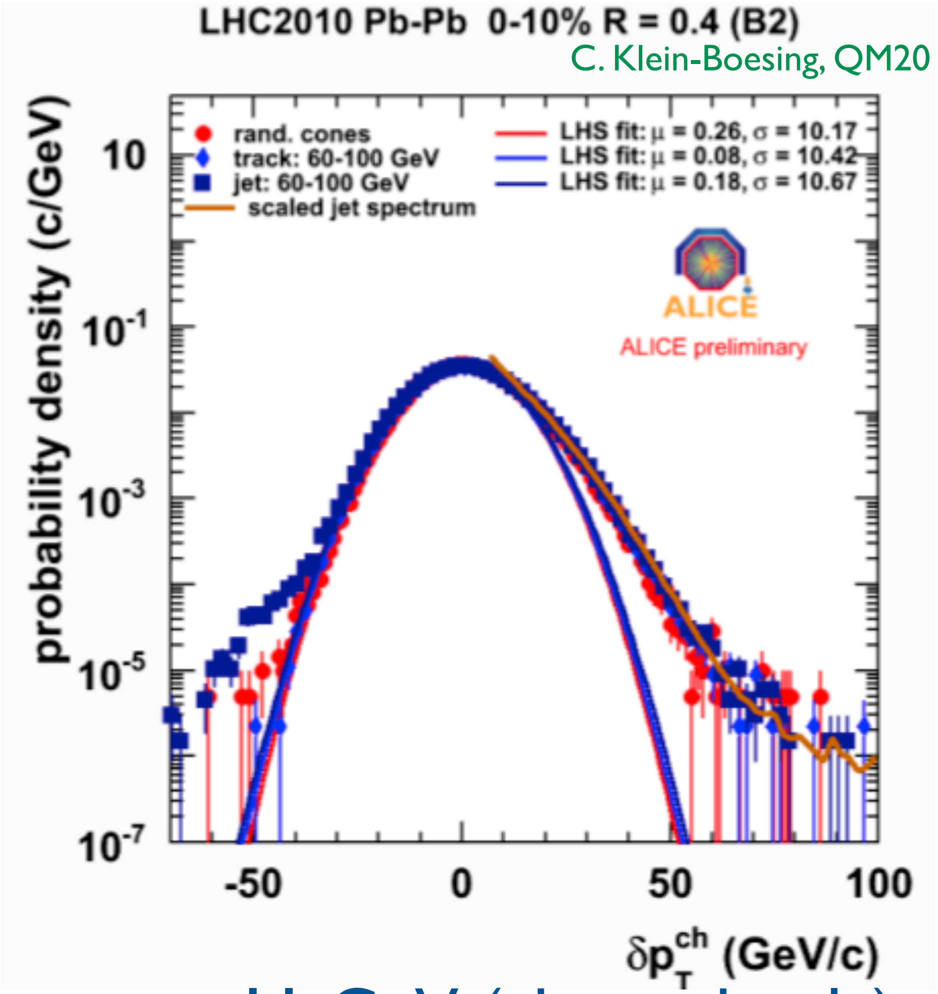
σ_{jet} should probably be the one of the first things one looks at, before any physics analysis is attempted with the reconstructed jets

σ_{jet} from STAR and ALICE

Distribution of $\Delta pt = p_t^{AA}$ - bkgd - p_t^{pp}

H. Caines, QM2011





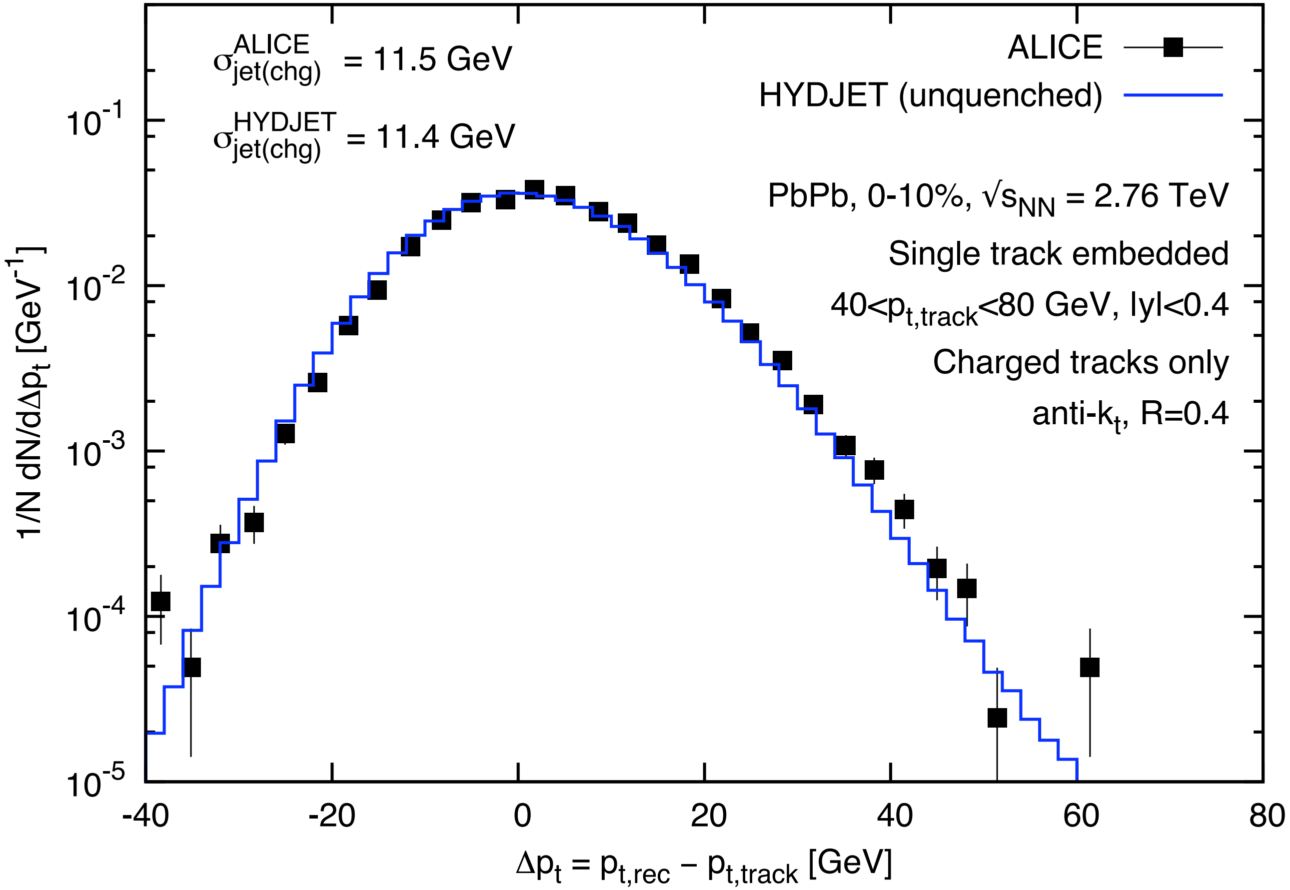
C. Klein-Boesing, QM2011

$\sigma_{jet} \approx$ 6-7 GeV

$\sigma_{jet} \approx$ 11 GeV (charged only)

HYDJET v. ALICE charged tracks jets

MC, Salam, Soyez, 1101.2878

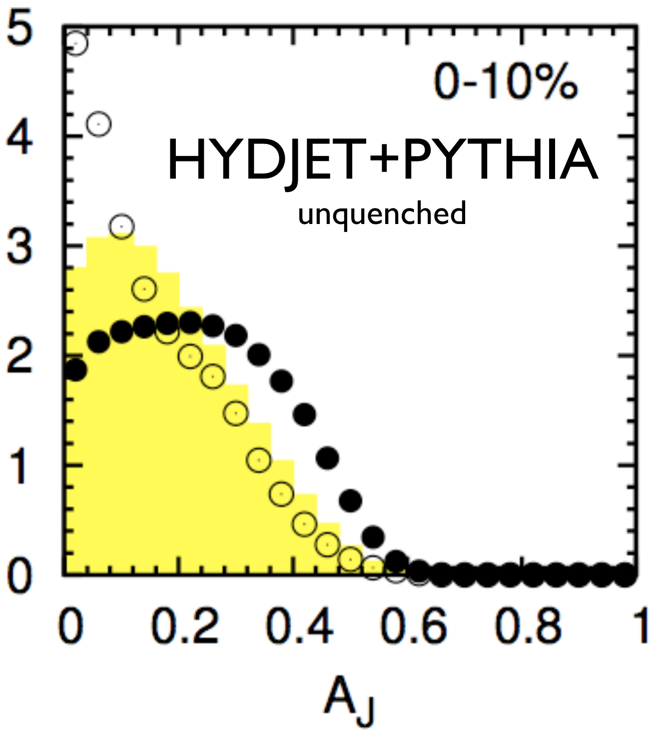


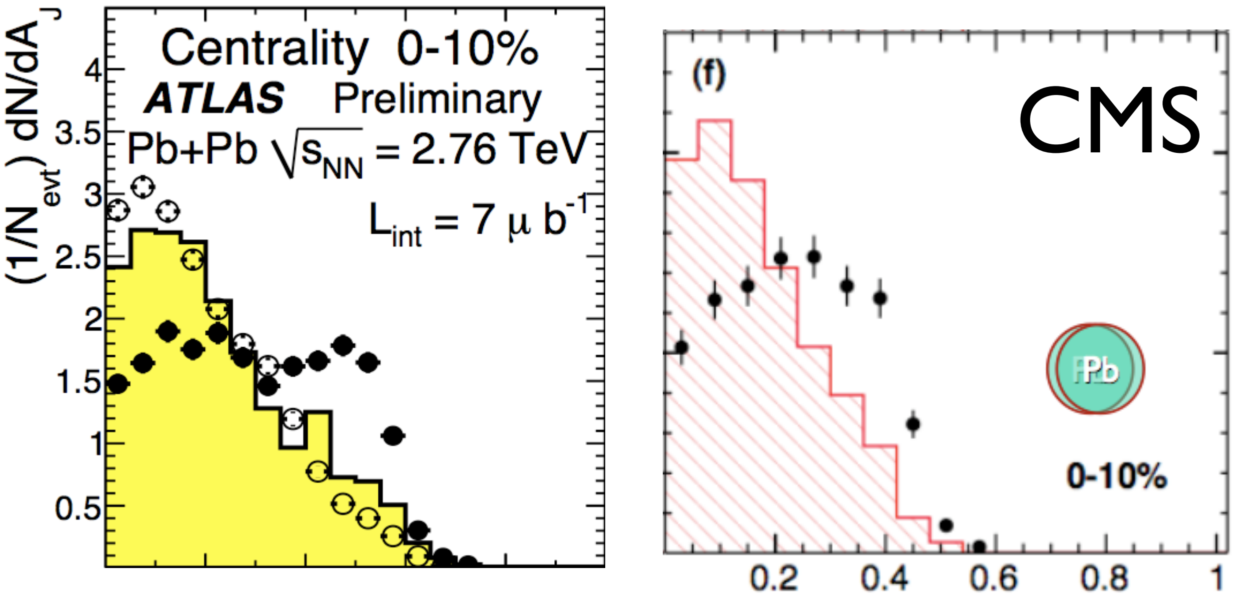
Striking agreement, and $\sigma_{jet(chg)} \approx 11.5$ GeV

Translates to a full $\sigma_{jet} \approx 17$ GeV

These are **real data**.
It seems that HYDJET does a good job in describing the PbPb background characteristics
(as a side note, HYDJET was not even tuned to LHC data)

...SO....





If this is a legitimate effect of fluctuations without quenching...

...what is the contribution of quenching to these measurements?

Conclusions

Effect of residual fluctuations of a non-noise reduction subtraction seems capable of inducing an asymmetry which mimics the one observed by ATLAS and CMS

Does this mean that there is no quenching? No

However, it likely means that in order to make **quantitative statements** about quenching one needs to have better control of background subtraction effects (residual fluctuations and/or biases)

Extra material

Caveats

While jet clustering is a deterministic procedure (though one must still choose a jet definition), background subtraction is less well-determined

A number of not fully clear-cut choices must be made:

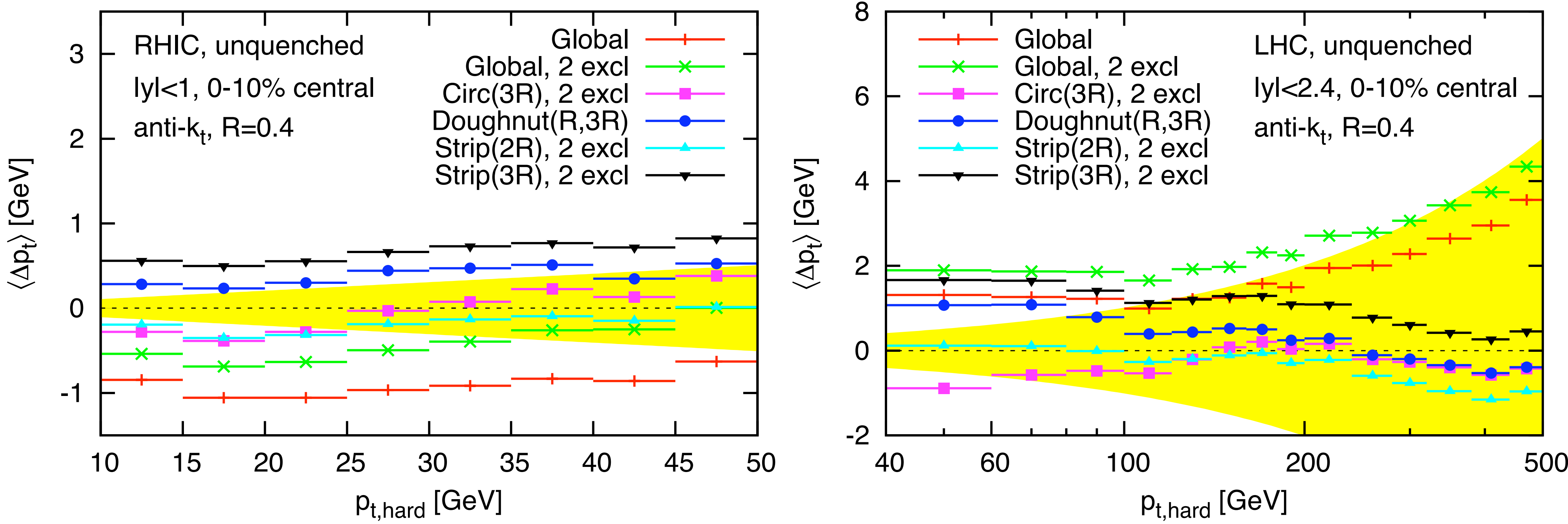
- **Where** to estimate the background (i.e. which range)
- **How** to estimate it (for instance, subtract hard jets?)
- **Which jet algorithm** to use (privilege small bias or small dispersion?)

Making the "proper" choice is as much a matter of
art (i.e. experience) as of science,
and depends on what you want to do

Having many algorithms and techniques at one's disposal
will allow better tuning of procedure with aim

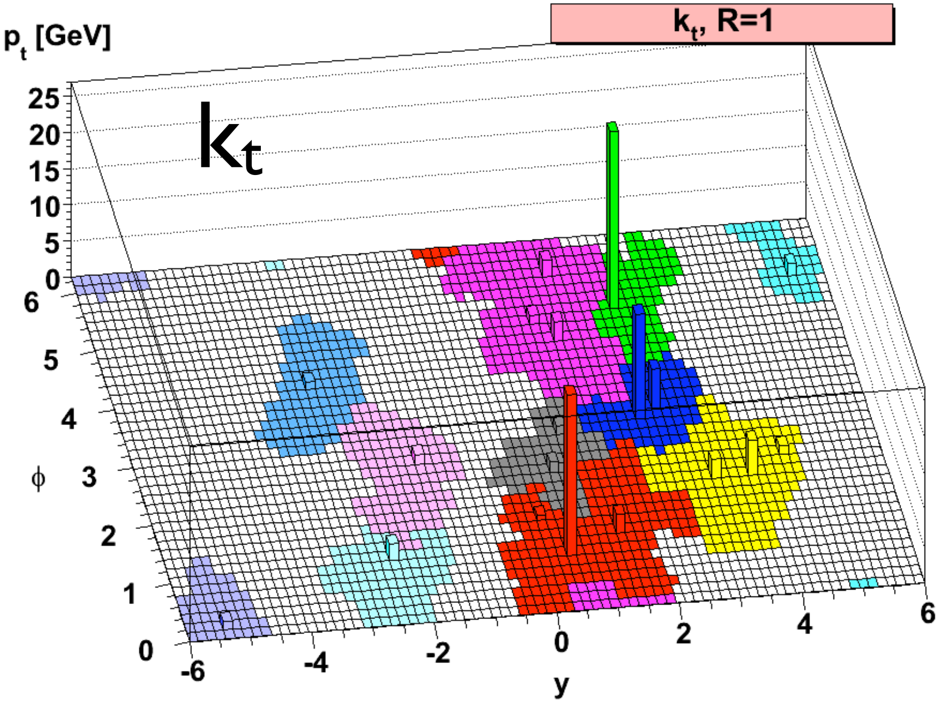
Ranges

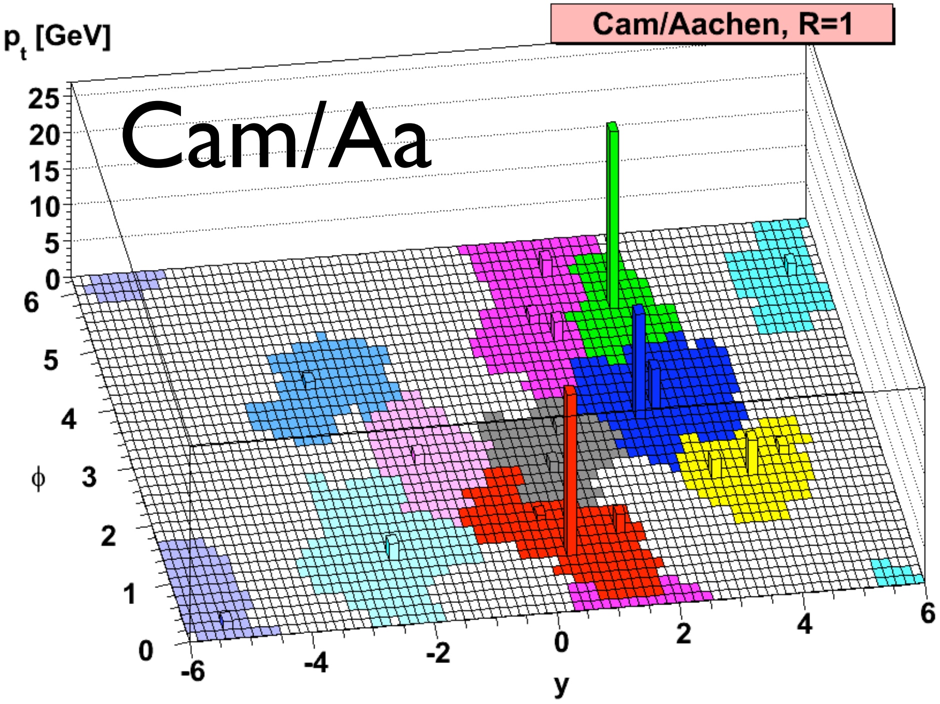
$<\Delta p_t>$

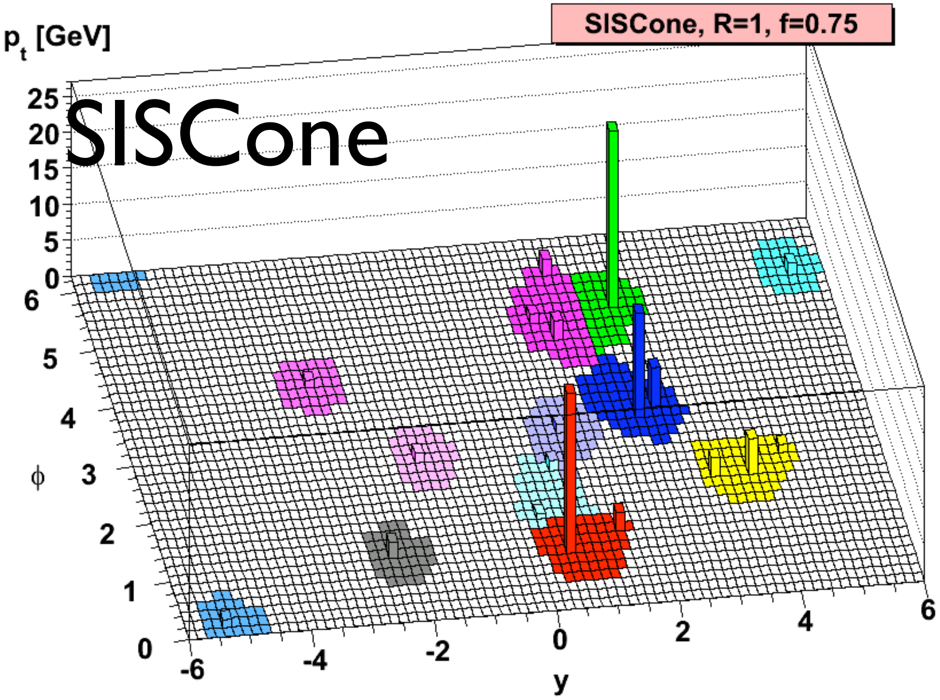


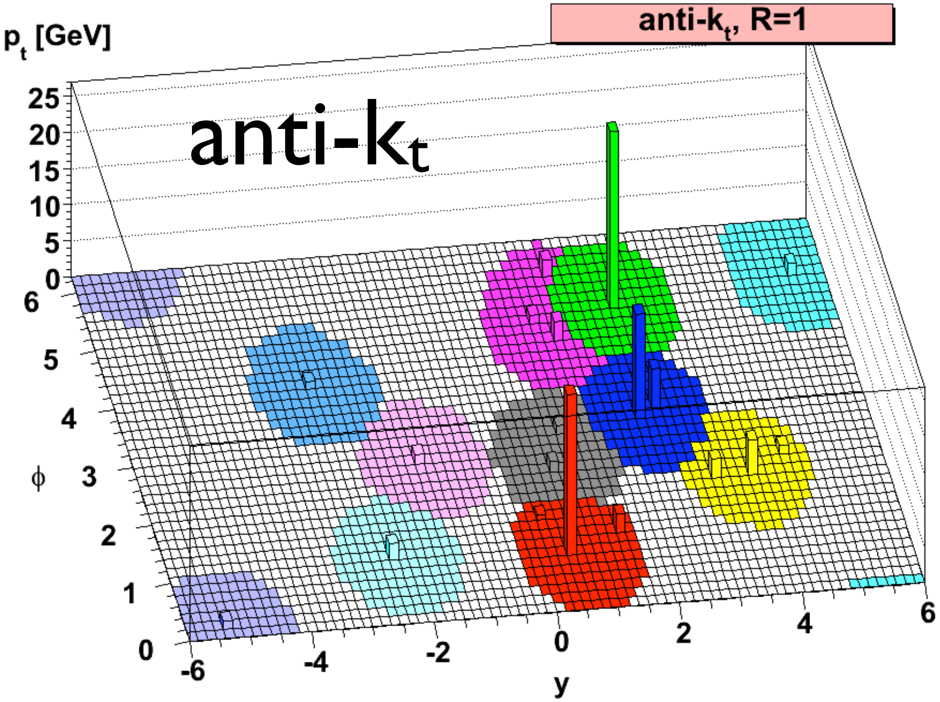
Intrinsic ambiguity mostly of order 1-2 GeV on Δp_t

The local ranges perform similarly, the exclusion of hardest jets helps a little,
the global range also performs fairly well here thanks to the limited rapidity coverage


k_t

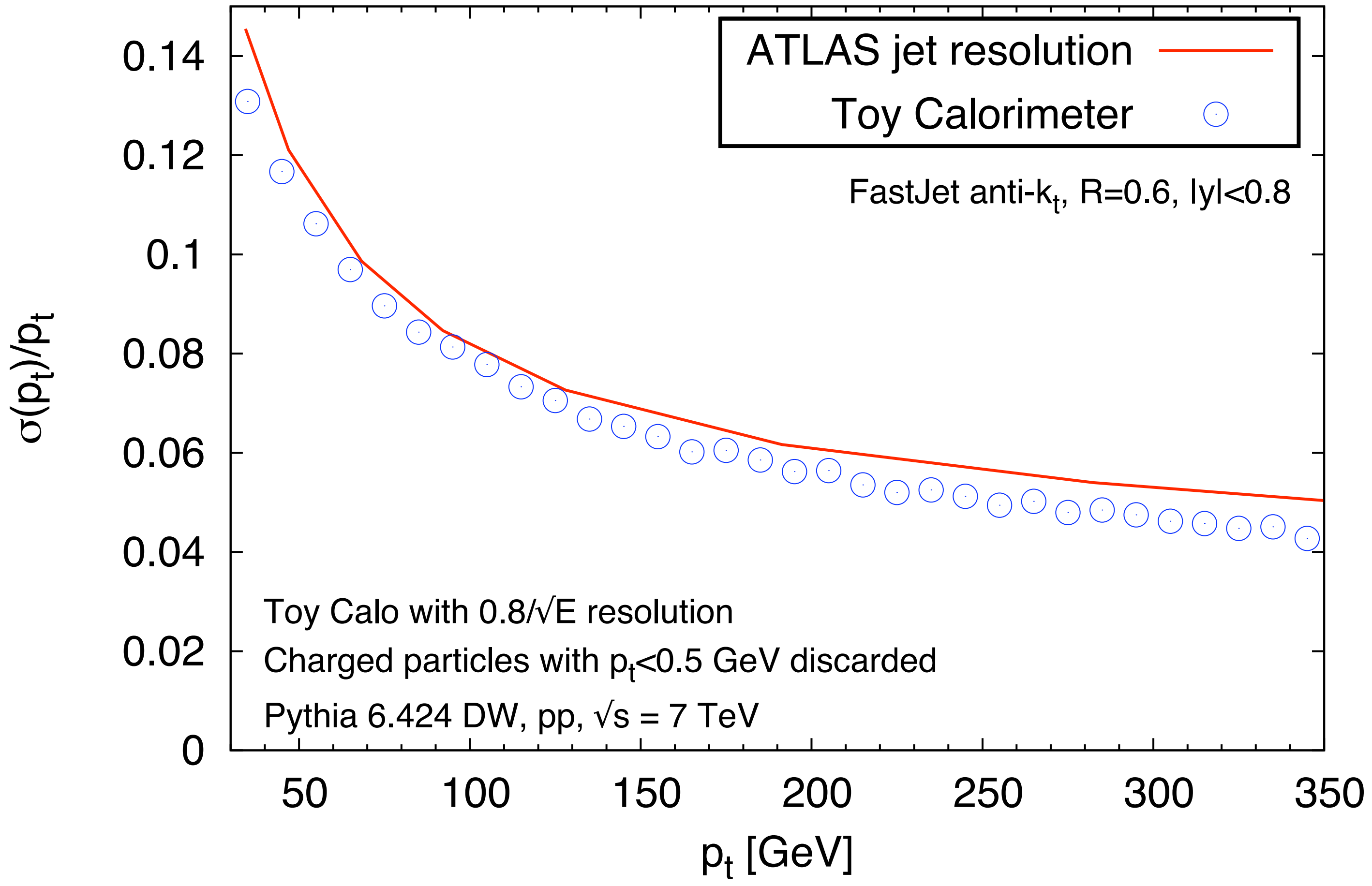

Cam/Aa


SISCone


anti-k_t

Toy calorimeter

Comparison of the resolution from a toy calorimeter and from the full ATLAS simulation

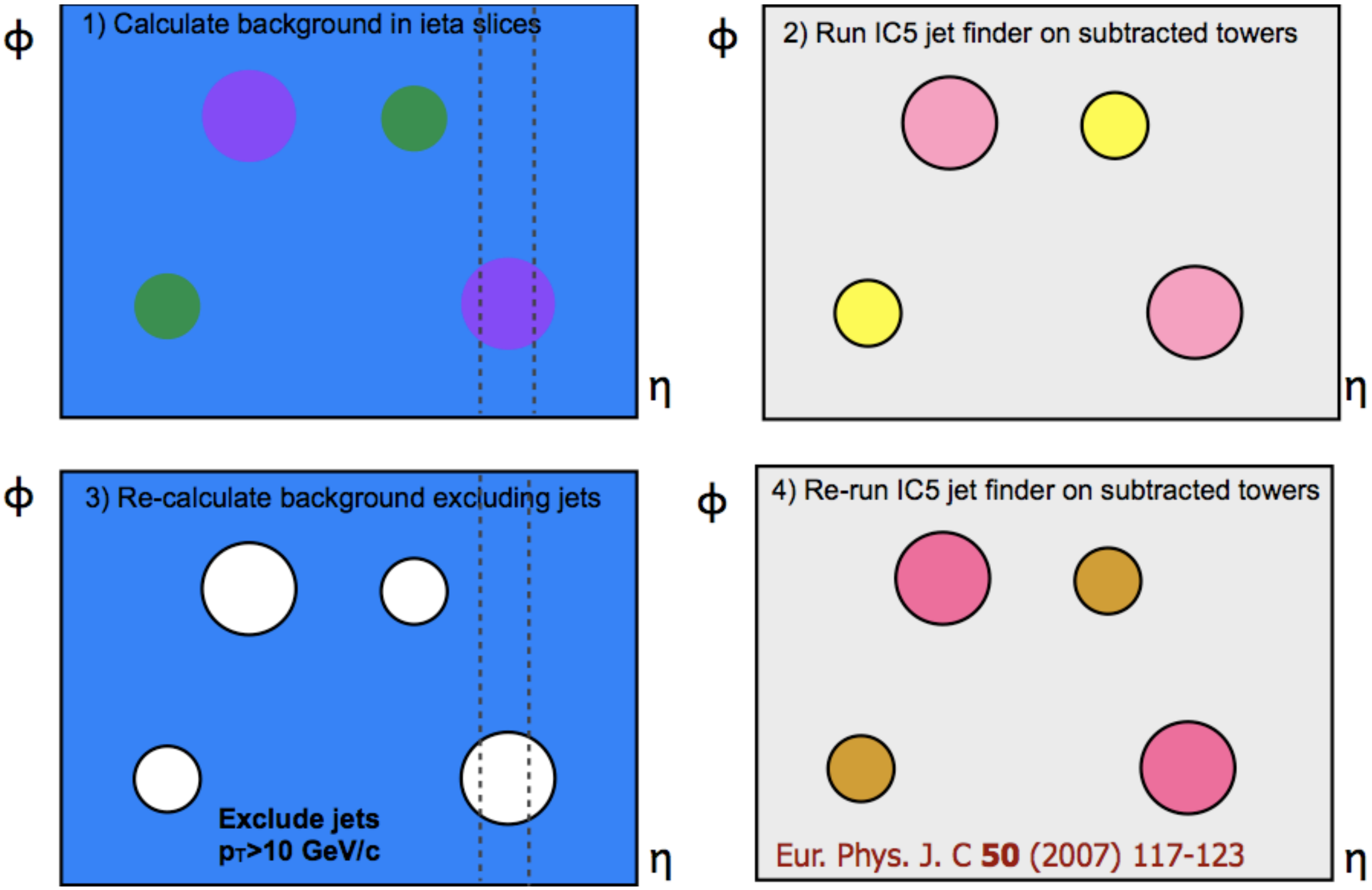


Toy calorimeter slightly better ⇒ no enhancement of fluctuations

Another subtraction technique

Iterative Cone Subtraction (used by CMS)

O. Kodolova et al. EPJC 50 (2007) 117



Eur. Phys. J. C **50** (2007) 117-123

This algorithm contains **noise reduction**: only towers with a positive p_t after subtracting average background + σ are retained

Iterative Cone Subtraction bias

Smaller fluctuations:

MC, Salam, Soyez, 1101.2878

$$\begin{aligned}(\sigma_{\rm jet}^{\rm noise\text{-}suppressed})^2 &= N_{\rm tower}\left[\langle(\delta p_{t,\rm tower}^{\rm noise})^2\rangle - \langle\delta p_{t,\rm tower}^{\rm noise}\rangle^2\right]\\ &= N_{\rm tower}\left[\int_{\sigma_{\rm tower}}^{\infty} dx \frac{(x-\sigma_{\rm tower})^2}{\sqrt{2\pi}\sigma_{\rm tower}} e^{-\frac{x^2}{2\sigma_{\rm tower}^2}} - \langle\delta p_{t,\rm tower}^{\rm noise}\rangle^2\right]\\ &\simeq (0.262\,\sigma_{\rm tower})^2\, N_{\rm tower}\,,\end{aligned}$$

at the price of a potential bias on the jet p_t:

$$\langle\delta p_{t,\rm jet}^{\rm overall}\rangle = \langle\delta p_{t,\rm jet}^{\rm noise}\rangle + \langle\delta p_{t,\rm jet}^{\rm hard}\rangle \simeq (0.0833 - f) N_{\rm tower}\sigma_{\rm tower}$$

f ≈0.1 is the occupancy of a hard perturbative jet ⇒ **large cancellation**

However, what happens to *f* in the case of quenching?

$$\langle\delta p_{t,\rm jet}^{\rm noise}\rangle = N_{\rm tower}\langle\delta p_{t,\rm tower}^{\rm noise}\rangle = N_{\rm tower}\int_{\sigma_{\rm tower}}^{\infty} dx \frac{(x-\sigma_{\rm tower})}{\sqrt{2\pi}\sigma_{\rm tower}} e^{-\frac{x^2}{2\sigma_{\rm tower}^2}} \simeq 0.0833\,\sigma_{\rm tower}\,N_{\rm tower}$$

$$\langle\delta p_{t,\rm jet}^{\rm hard}\rangle \simeq -f N_{\rm tower}\sigma_{\rm tower}$$