Pileup subtraction and rapidity dependence

Gavin Salam
with Matteo Cacciari & Gregory Soyez

CERN, Princeton & LPTHE/CNRS (Paris)

CERN, 11 November 2010

Subtracting noise from jets

$$p_{t,jet}^{\mathrm{subtracted}} = p_{t,jet} - \rho \times A_{jet}$$

$$A_{jet} = \text{jet area}$$

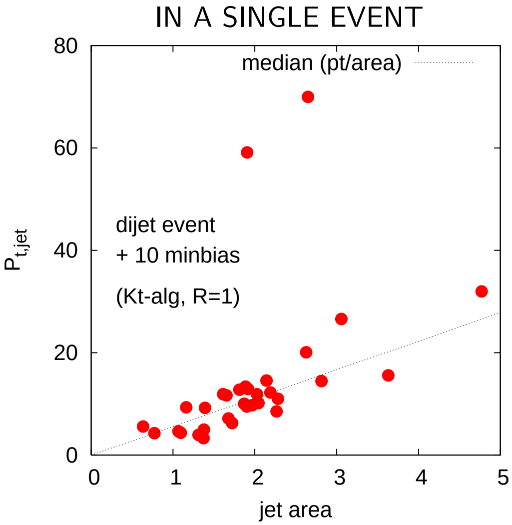
$$\rho = p_t \text{ per unit area from pileup}$$

(or “background”)

This procedure is intended to be common to pp ($\rho \sim 1{-}2$ GeV), pp with pileup ($\rho \sim 2 - 15$ GeV) and Heavy-Ion collisions ($\rho \sim 100 - 300$ GeV)

As proposed so far: jet-by-jet area determination,
event-by-event ρ determination

Event-by-event ρ (background) estimation



Most jets in event are "background"

Their p_t is correlated with their area.

Estimate ρ:

$$\rho \simeq \underset{\{jets\}}{\text{median}} \left[\frac{p_{t,jet}}{A_{jet}} \right]$$

Median limits bias from hard jets

Cacciari & GPS '07

A slide from Philipp

INTERNAL CMS PLOT NOT SHOWN IN PUBLIC VERSION OF TALK

A slide from Philipp

INTERNAL CMS PLOT NOT SHOWN IN PUBLIC VERSION OF TALK

Issues: 1) drop-off at $\eta \gtrsim 2.5$, 2) spike at $\eta \simeq 2.5$

1. First study drop-off (smooth rapidity dependence)

The elements of this study

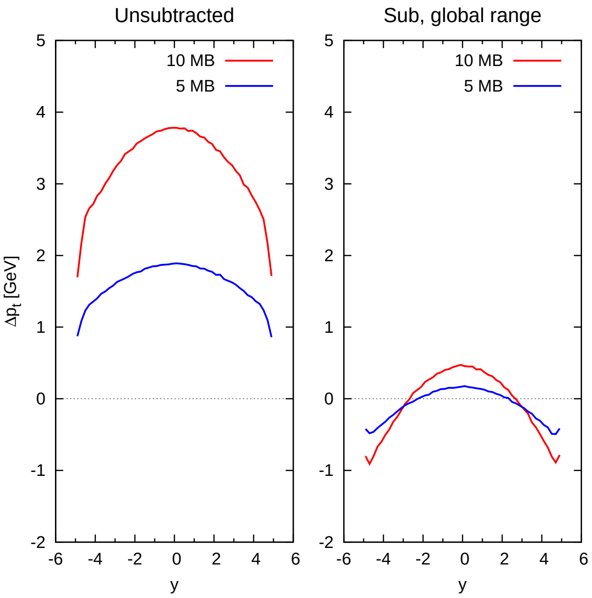
The quantity we look at is

$$\Delta p_t \equiv p_t^{\text{pileup,sub}} - p_t^{\text{no pileup,sub}}$$

perfect subtraction will give $\Delta p_t = 0$

- Pythia DW tune, LHC 7 TeV, *particles*
- FastJet 2.5-devel
- anti-k_t with $R = 0.5$ for main jets
- k_t with $R = 0.6$ for ρ determination
- particles, ghosts and jets up to $y = 5$
- jets (no pileup) with $20 < p_t < 40$ GeV
- global ρ estimation range extends to $y = 5$

Results for 5 & 10 min-bias (MB) pileup



Subtraction works to within 1 GeV.

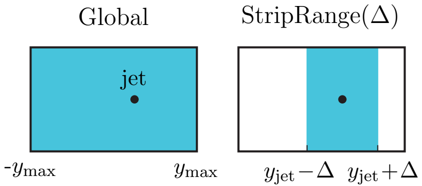
But there is some residual rapidity dependence.

This is because minimum-bias p_t flow has rapidity dependence.

Local estimate for ρ?

So far: take all jets with $|y| < 5$ to estimate single ρ for whole event.

Alternative: **for each jet, estimate ρ in its local neighbourhood**, using "local range"



Useful also for heavy-ion collisions, cf Cacciari, Rojo, GPS & Soyez '10

What to watch out for? Local range must be large enough

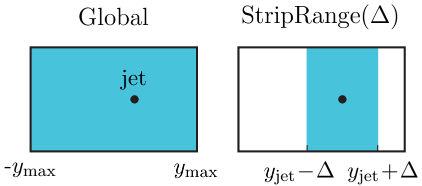
- to have statistically meaningful measurement of ρ
- for occasional hard jets not to bias median

In what follows we'll use strip range with $\Delta = 1.2$

Local estimate for ρ?

So far: take all jets with $|y| < 5$ to estimate single ρ for whole event.

Alternative: **for each jet, estimate ρ in its local neighbourhood**, using "local range"



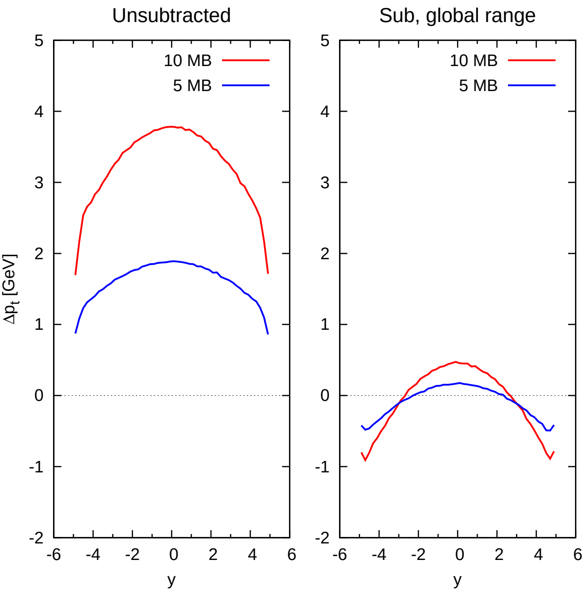
Useful also for heavy-ion collisions, cf Cacciari, Rojo, GPS & Soyez '10

What to watch out for? Local range must be large enough

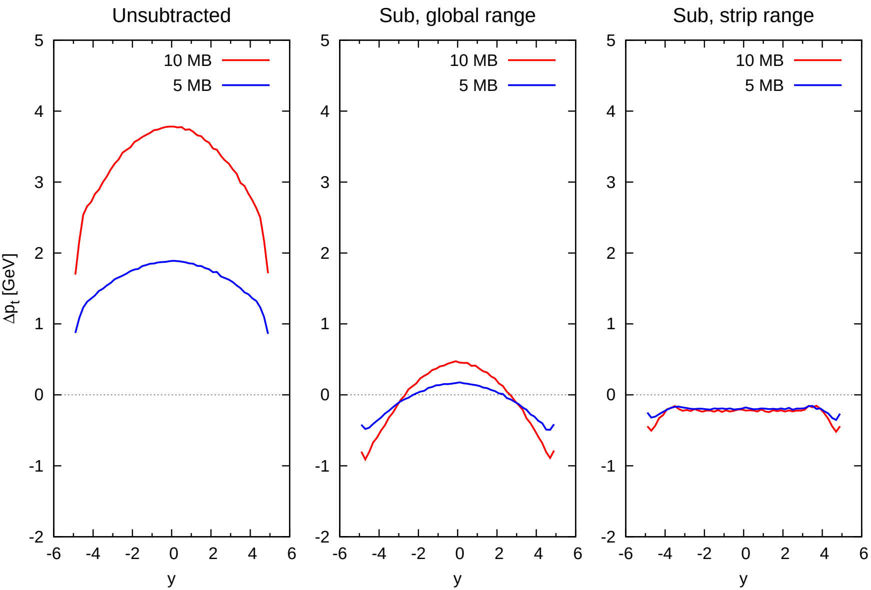
- to have statistically meaningful measurement of ρ
- for occasional hard jets not to bias median

In what follows we'll use strip range with $\Delta = 1.2$

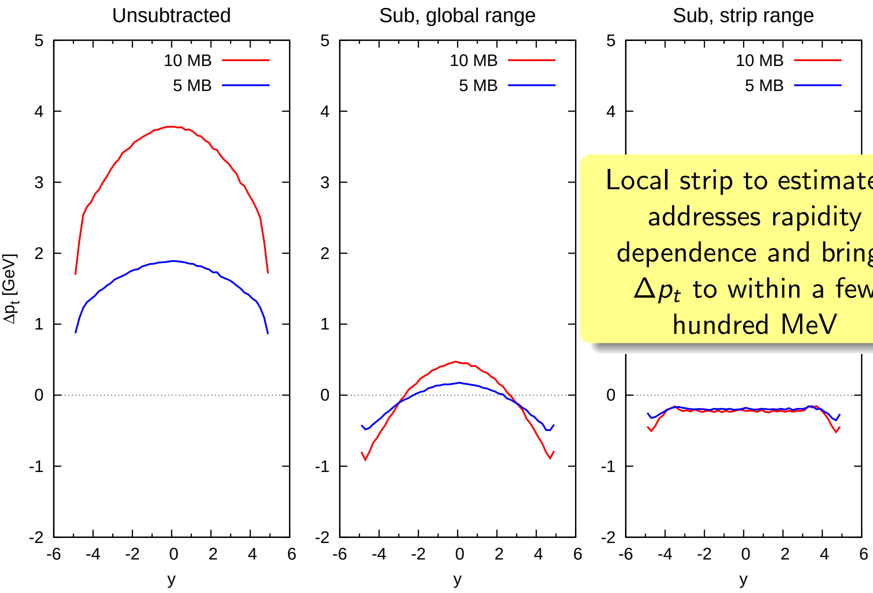
Impact of switch to local range



Impact of switch to local range



Impact of switch to local range



Local strip to estimate ρ addresses rapidity dependence and brings Δp_t to within a few hundred MeV

2. Rapidity “ears”

The spikes or “ears”

INTERNAL CMS PLOT

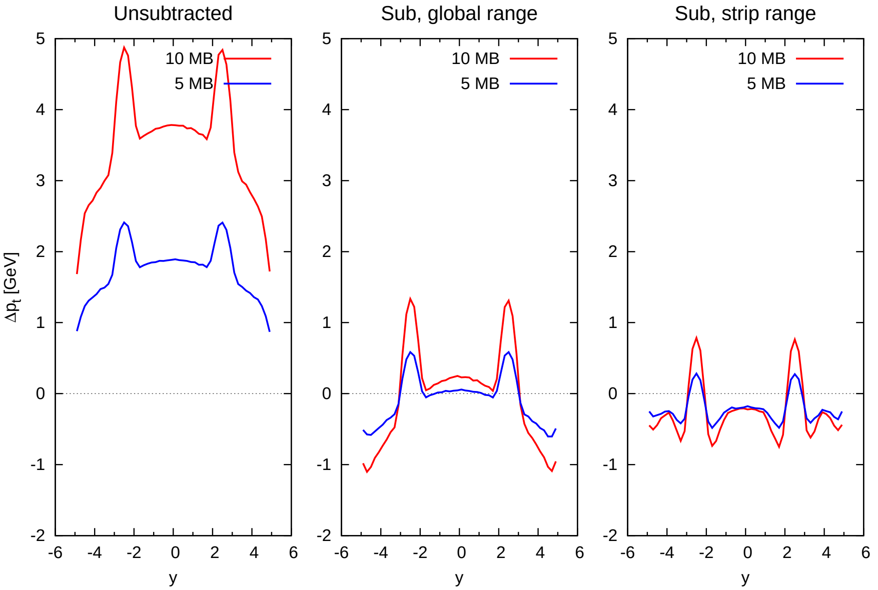
What’s going on here?

Min-bias physics doesn’t have structures like this.

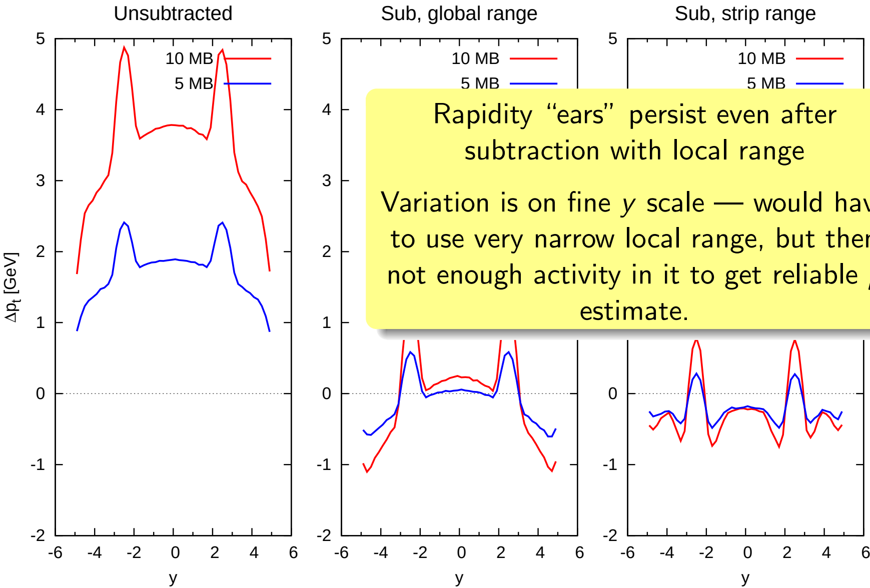
We mimic such an effect by

doubling the p_t of every particle with $\mathbf{p_t} < 2$ GeV & $2.3 < |\mathbf{y}| < 2.7$

Our simulated rapidity "ears"



Our simulated rapidity "ears"



Rapidity "ears" persist even after subtraction with local range

Variation is on fine y scale — would have to use very narrow local range, but then not enough activity in it to get reliable ρ estimate.

Encoding rapidity structure

We use "ghosts" to measure jet areas $\equiv$ susceptibility of jet to contamination from noise.

"Ears" $\equiv$ greater susceptibility to contamination from a given amount of actual (particle-level) noise.

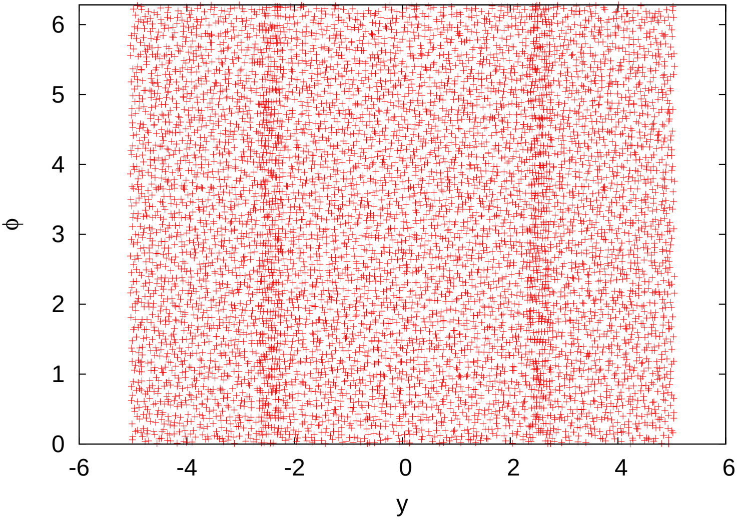
Two ways of accounting for this:

- Associate "weight" w_i with each ghost i. Detector regions with more noise-sensitivity $\to$ larger w_i. Then use "weighted area" for ρ estimation and subtraction:

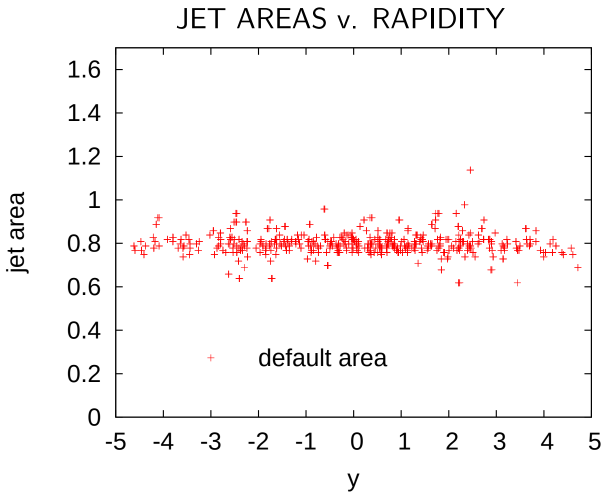
$$\text{weighted area} = A_{\text{ghost}} \times \sum_{\text{ghost } i \in \text{jet}} w_i$$

- Conceptually related: add extra ghosts ($\equiv$ enhanced area) in relevant part of $y - \phi$. **This is what we use in next slides**.
  Specifically, a crude first try: double up the ghosts for $2.3 < |y| < 2.7$

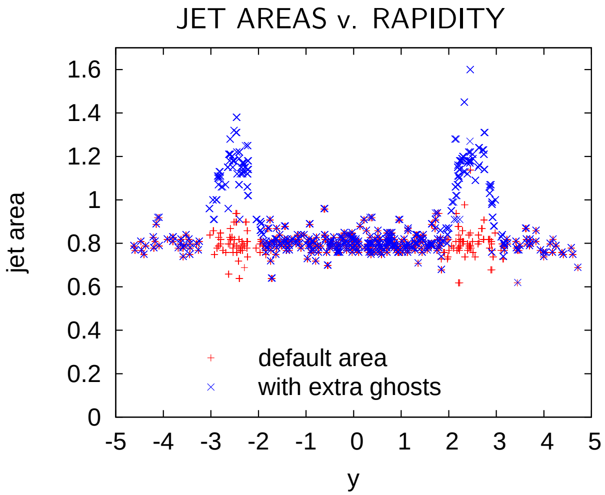
The (extra) ghosts



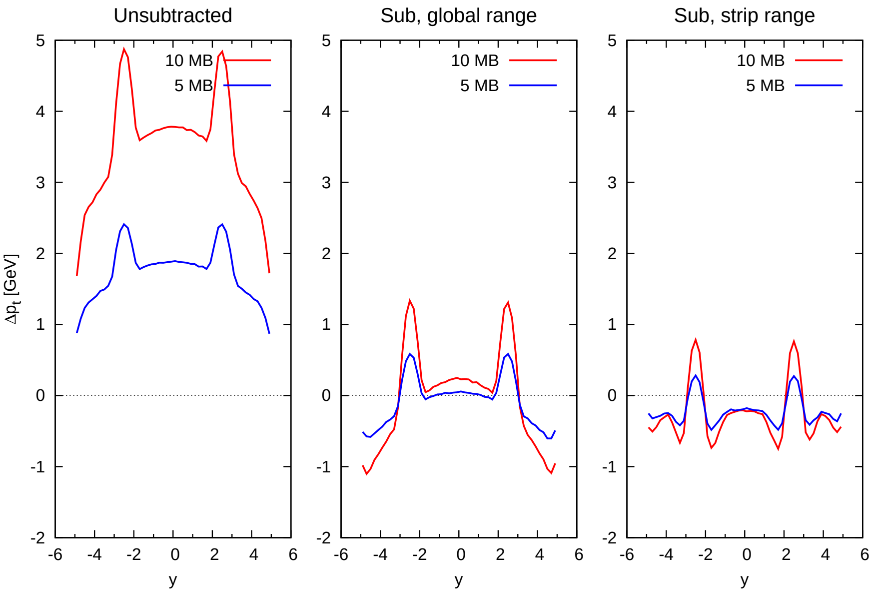
Areas of hard jets



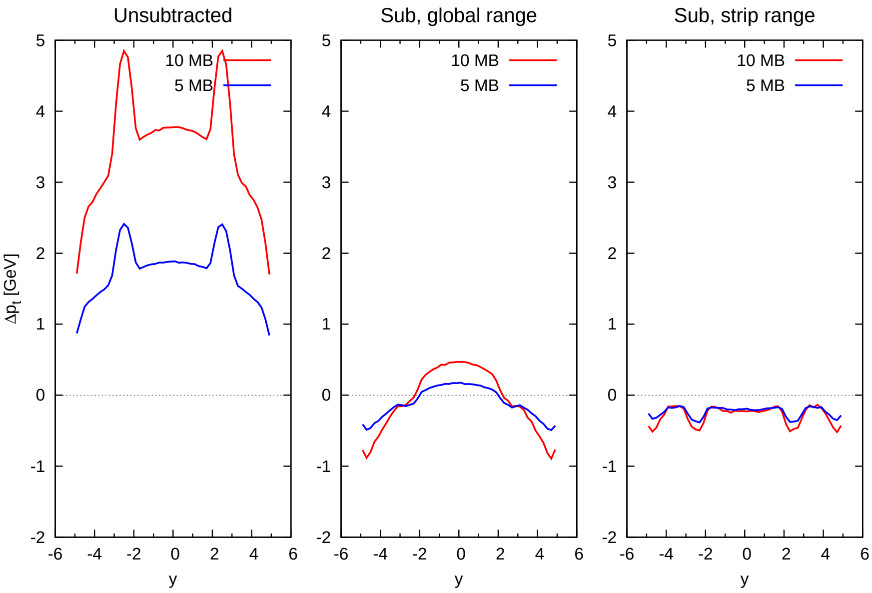
Areas of hard jets



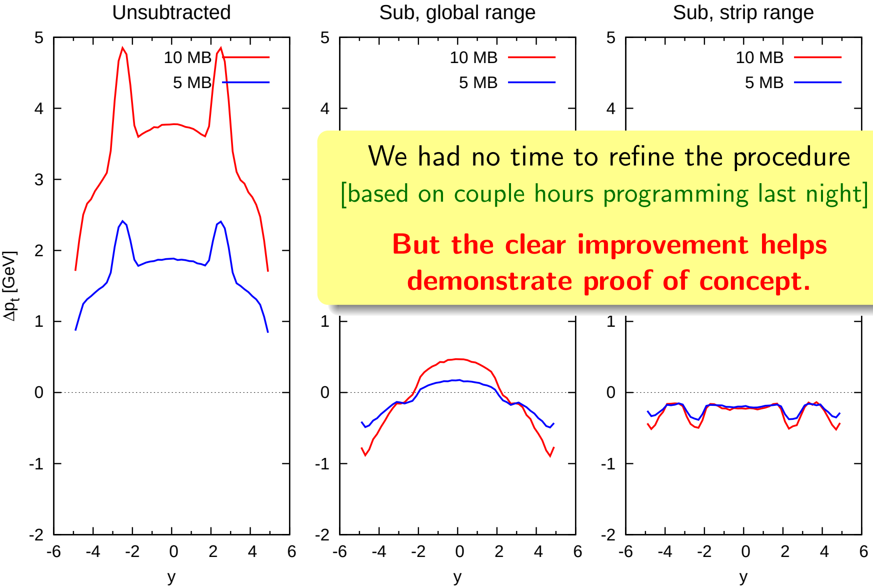
Subtraction without extra ghosts



Subtraction with extra ghosts



Subtraction with extra ghosts



We had no time to refine the procedure

[based on couple hours programming last night]

But the clear improvement helps demonstrate proof of concept.

Comments

- Smooth rapidity dependence can be addressed by local ρ determination
- "Ears" can be cured by small adaptation of area/median subtraction method

But is this a detector issue better resolved "upstream"?

More generally: There are two separate parts to PU subtraction problem

- The jet's sensitivity to contamination (area)
- Determining the level of the contamination (ρ)

Different methods can also be mixed. E.g.

- With track primary vertex info, you could choose to deduce "neutral" ρ from tracks and use that to subtract $A \times \rho^{neutral}$ from particle flow jets

And use median ρ method where you have no tracks?

- With correction $\propto$ number of primary vertices ($\rho \propto n^{PU}$), knowledge of area still important, because not all jets have $A = \pi R^2$

Comments

- Smooth rapidity dependence can be addressed by local ρ determination
- "Ears" can be cured by small adaptation of area/median subtraction method

But is this a detector issue better resolved "upstream"?

More generally: There are two separate parts to PU subtraction problem

- The jet's sensitivity to contamination (area)
- Determining the level of the contamination (ρ)

Different methods can also be mixed. E.g.

- With track primary vertex info, you could choose to deduce "neutral" ρ from tracks and use that to subtract $A \times \rho^{neutral}$ from particle flow jets

  And use median ρ method where you have no tracks?

- With correction $\propto$ number of primary vertices ($\rho \propto n^{PU}$), knowledge of area still important, because not all jets have $A = \pi R^2$

FastJet 2.5/3.0 [in development]

Main aim: make it easier to do advanced things

Will definitely be included

- `PseudoJet` now knows what `ClusterSequence` it belongs to.
  So one can write `jet->area()`, `jet->constituents()`, etc.
  [instead of `cs->area(jet)`, etc.]

Will probably be included

- `PseudoJet` carries `extra_info` (shared pointer derived from `PseudoJet::ExtraInfo`) for cleaner user-information management
- Revamp of background estimation framework, to facilitate use of local ranges (`Selector`, will extend `Range` functionality, `BackgroundEstimator` will replace what's currently in `CSAreaBase`)
- Etc. (e.g. filtering tools, boosted top ID(?), . . .)

EXTRAS

Places with more information

Basic idea of area/median pileup subtraction: arXiv:0707.1378

The systematics behind jet areas: arXiv:0802.1189

The systematics behind the jet/area median ρ estimation method: arXiv:0912.4926

Application to heavy-ion collisions (including local ranges, etc.): arXiv:1010.1759