CMS JetMET Workshop
Vienna, 25-27 August 2014

Considerations about jet substructure and pileup subtraction techniques

Matteo Cacciari (LPTHE Paris)

My own work done in collaboration with Gavin Salam and Grégory Soyez.
Many parts of the talk also borrowed from them

Recent events

Two recent events collected most of what I am going to discuss today:

- Pileup Mitigation Workshop at CERN in May
  - https://indico.cern.ch/event/306155/
- BOOST14 in London last week
  - https://indico.cern.ch/event/302395/

I'll be cherry-picking material from these two events
(with a very personal, and non exhaustive, choice of topics)

[Apologies to those among you who were at both events]

Outline

- Substructure and Taggers
- Pileup subtraction
  - SoftKiller

(in most cases, a small subset of ongoing activity)

Recent progress

Perhaps best visualized by the increased number of FastJet Contrib projects

June 2013

Version 1.005 of FastJet Contrib is di

Package	Version
GenericSubtractor	1.2.0
JetFFMoments	1.0.0
VariableR	1.0.1
Nsubjettiness	1.0.2
EnergyCorrelator	1.0.1
ScJet	1.1.0

August 2014

Version 1.014 of FastJet Contrib is distril

Package	Version
ConstituentSubtractor	1.0.0
EnergyCorrelator	1.0.1
GenericSubtractor	1.2.0
JetCleanser	1.0.1
JetFFMoments	1.0.0
JetsWithoutJets	1.0.0
Nsubjettiness	2.1.0
RecursiveTools	1.0.0
ScJet	1.1.0
SoftKiller	1.0.0
SubjetCounting	1.0.1
VariableR	1.1.1

Slide by G. Salam

FastJet Contrib

The large increased in the number of projects hosted by FastJet Contrib tells us various things:

- A lot of activity is going on
- FJ Contrib is catching on as a repository for jet-related software
  - People are appreciating:
    - The usefulness of a single repository, with uniform build system, etc
    - The added value in having public, properly versioned and stable implementations of old and new ideas
      - Very easy to test them immediately!

http://fastjet.sourceforge.org/contrib

Public code

"It can be very hard to document properly all the details of even a simple analysis"

Andy Buckley at BOOST14, advocating the use of RIVET

The same holds for all code, jet algorithms being no exception

If the code is public, there is no ambiguity: the code IS the algorithm (and, sometimes, contains surprises...)

FastJet 3.1.0-beta.1

The first beta of FastJet 3.1 was released a few days ago

15 August 2014: fastjet-3.1.0-beta.1 (manual, doxygen, fjcore). Main improvements relative to version 3.0.6:

- Significant speed improvements (x1.5-10) for large N (few thousand-10^5)
- FASTJET_VERSION_NUMBER preprocessor symbol, for easy in-code version testing
- New JetDefinition::operator(): jets = jet_def(particles);
- Native particle-mass support in PU estimation (rho_m())
- Subtraction can use rho_m (set_use_rho_m()) and force m>0 (set_safe_mass())
- New Recluster class, serving as base for Filter
- New RectangularGrid class, base for GridMedianEstimator & GridJetPlugin
- Fixed long-standing issue with coincident points in NlnN strategies
- Other small additions and changes include:
    - Selector::sum(particles) to get 4-vector sum of particles that pass selector
    - pruned_jet.structure_of() has Rcut() and zcut() functions
    - easy copying of Recombiner info: jet_def1.set_recombiner(jet_def2)
- Various bug-fixes, build-system tweaks (e.g. default -O2 instead of -O3)

FastJet 3.1.0-beta.1

The first beta of FastJet 3.1 was released a few days ago

15 August 2014: fastjet-3.1.0-beta.1 (manual, doxygen, fjcore). Main improvements relative to version 3.0.6:

- Significant speed improvements (x1.5-10) for large N (few thousand-10^5)
- FASTJET_VERSION_NUMBER preprocessor symbol, for easy in-code version testing
- New JetDefinition::operator(): jets = jet_def(particles);
- Native particle-mass support in PU estimation (rho_m())
- Subtraction can use rho_m (set_use_rho_m()) and force m>0 (set_safe_mass())
- New Recluster class, serving as base for Filter
- New RectangularGrid class, base for GridMedianEstimator & GridJetPlugin
- Fixed long-standing issue with coincident points in NlnN strategies
- Other small additions and changes include:
    - Selector::sum(particles) to get 4-vector sum of particles that pass selector
    - pruned_jet.structure_of() has Rcut() and zcut() functions
    - easy copying of Recombiner info: jet_def1.set_recombiner(jet_def2)
- Various bug-fixes, build-system tweaks (e.g. default -O2 instead of -O3)

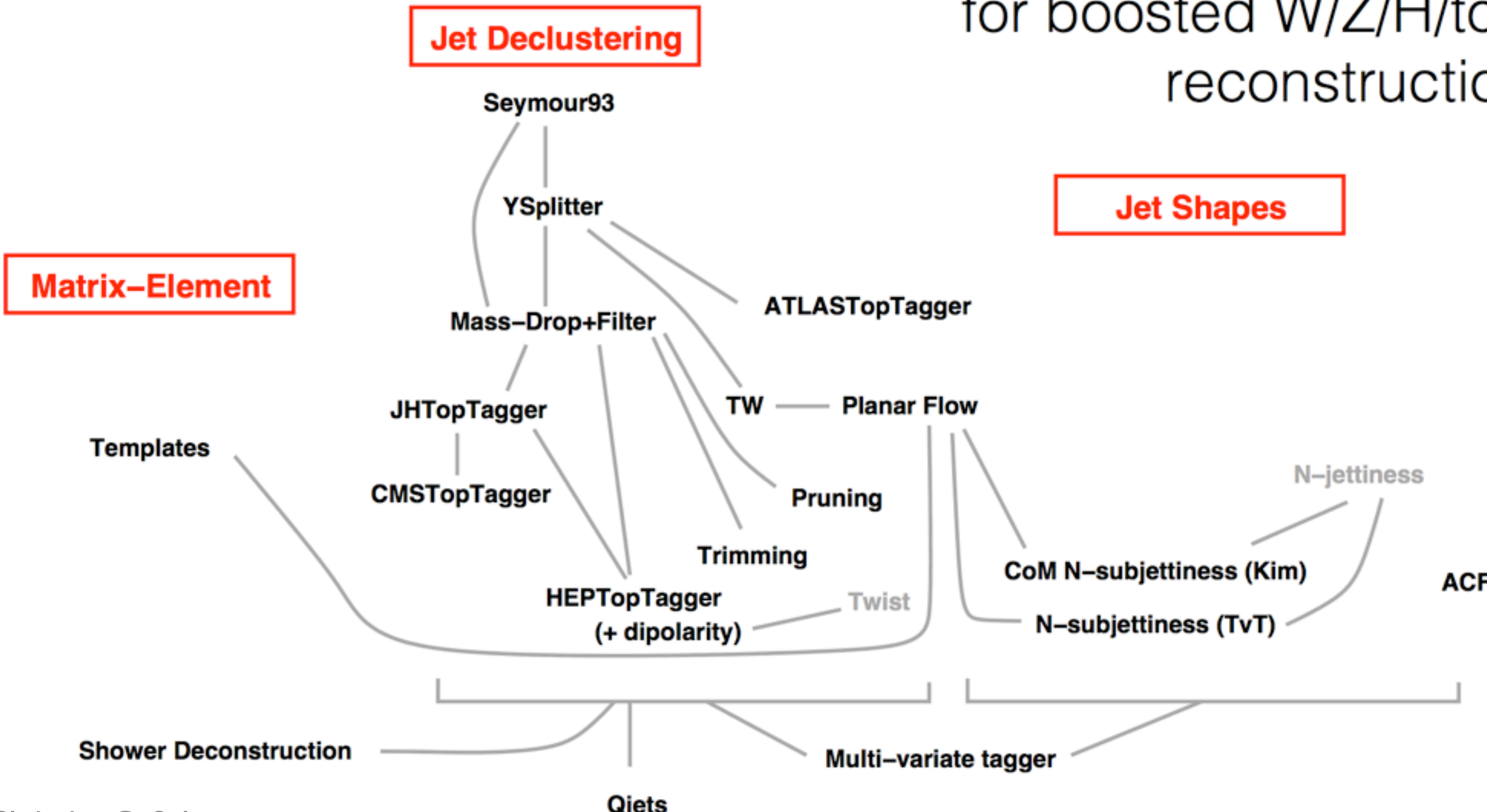
speed

pileup

substr.

The jet substructure maze

Some of the tools developed for boosted W/Z/H/top reconstruction



Slide by G. Salam

Substructure measurements

What is available

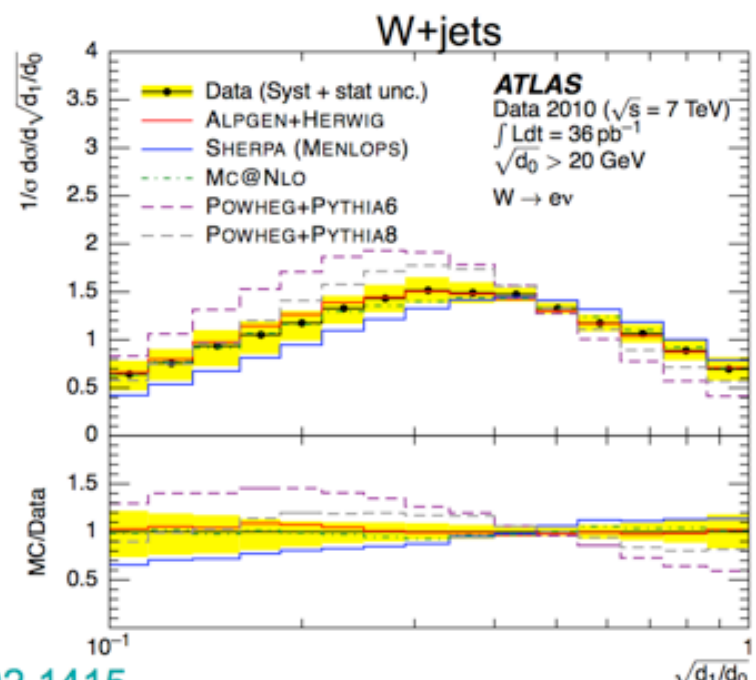
Reference	Final state	Jets, p_T (GeV)	Jet substructure observables
1101.0070 ATLAS 3/pb	incl. jets	q/g-jets (AK6), $30<p_T<600$	integral jet shape, differential jet shape
1204.3170 CMS 36/pb	incl. jets	q/g-jets (AK7), $20<p_T<1000$ q/g-jets (AK5), $50<p_T<1000$	integral jet shape, differential jet shape charged hadron multiplicity and width
1307.5749 ATLAS 1.8/fb	ttbar	q-jets (AK4), $30<p_T<150$ b-jets (AK4), $30<p_T<150$	integral jet shape, differential jet shape
1109.5816 ATLAS 36/pb	incl. jets	q/g-jets (AK6), $25<p_T<500$	charged hadron fragmentation function, p_T^{rel} and radial density
QCD-10-041 CMS 36/pb	dijets	q/g-jets (KT6), $97<p_T<1032$	subjet multiplicities and p_T^{rel}
1302.1415 ATLAS 36/pb	W+jets	q-jets (KT6), no p_T cut	k_T splitting scales
1203.4606 ATLAS 35/pb	incl. jets	q/g-jets (AK10, CA12), $200<p_T<600$	jet mass, split/filtered jet mass, k_T splitting scales, N-subjettiness ratios
1303.4811 CMS 5/fb	dijets W/Z+jets	q/g-jets (AK7), $220<p_T<1500$ q-jets (AK7, CA8, CA12), $125<p_T<450$	jet mass, pruned jet mass, trimmed jet mass, filtered jet mass
1206.5369 ATLAS 35/pb	incl. jets	q/g-jets (AK6, AK10), $p_T>300$	jet mass, jet width, eccentricity, planar flow, angularity

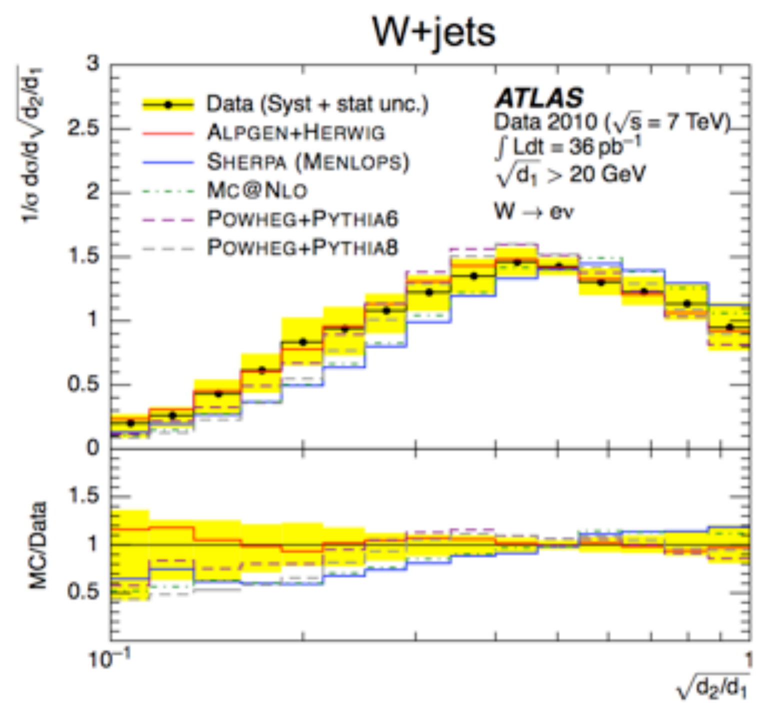
14 19 Aug 2014 Andreas Hinzmann University of Zurich[UZH]

Substructure measurements

k_T splitting scale

- Squared splitting scale at k_T-algorithm step: $d_{ij} = \min(p_{Ti}^2, p_{Tj}^2)\frac{\Delta R_{ij}^2}{R^2}$
- Look at last (hardest) clustering steps $d_{iB} = p_{Ti}^2$
- Probability for QCD emission with hardness $\sqrt{d_{k+1}}$ given previous emission of scale $\sqrt{d_k}$ is $\sqrt{d_{k+1}/d_k}$





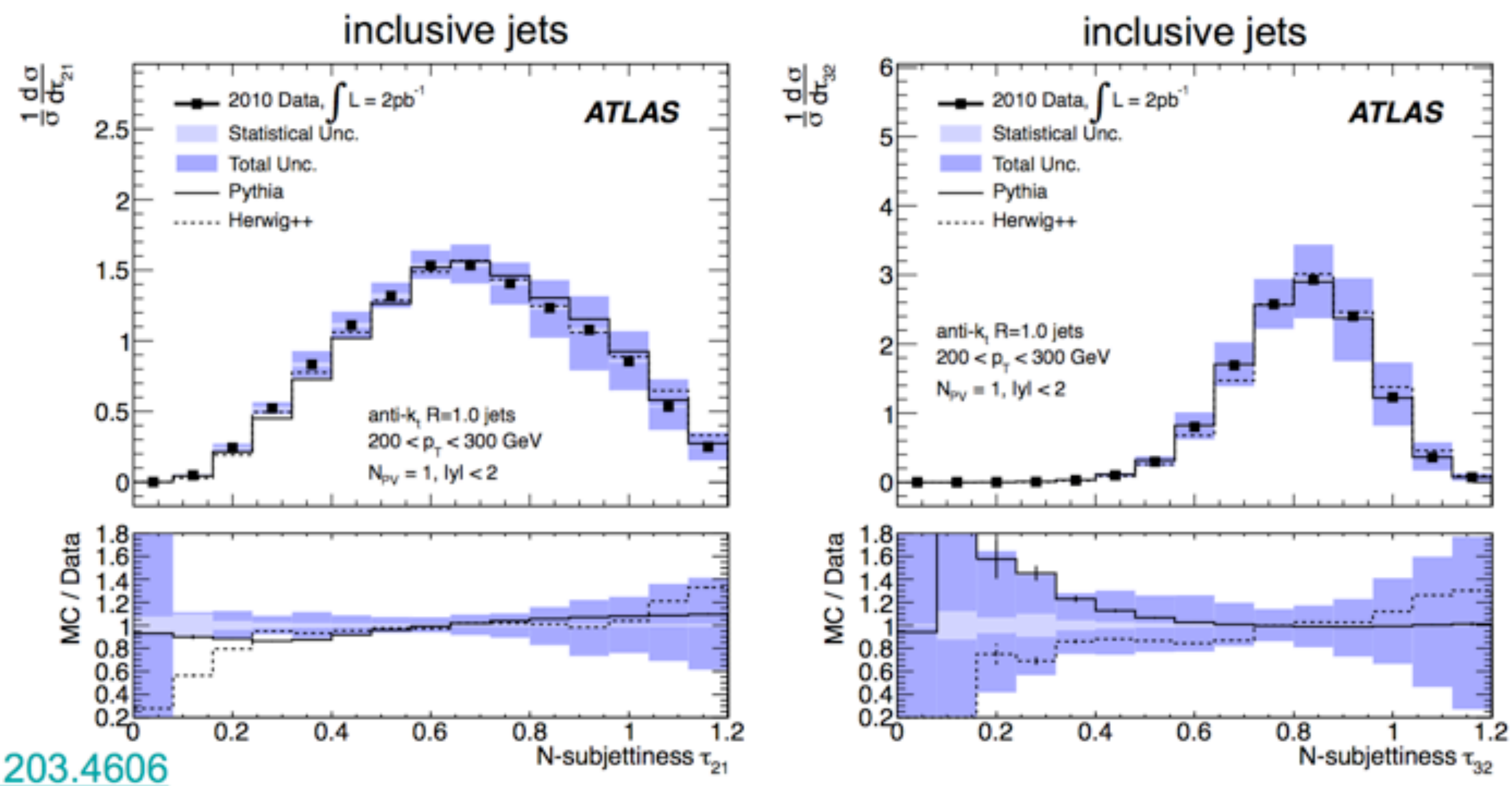
1302.1415

20 19 Aug 2014 Andreas Hinzmann University of Zurich[UZH]

Substructure measurements

N-subjettiness

- N-subjettiness $\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \times \min(\delta R_{1,k}, \delta R_{2,k}, \ldots, \delta R_{N,k})$ $\quad d_0 = \sum_k p_{T,k} R$
- Ratio τ_2/τ_1 discriminates compatibility with 2 subjet axes rather than 1



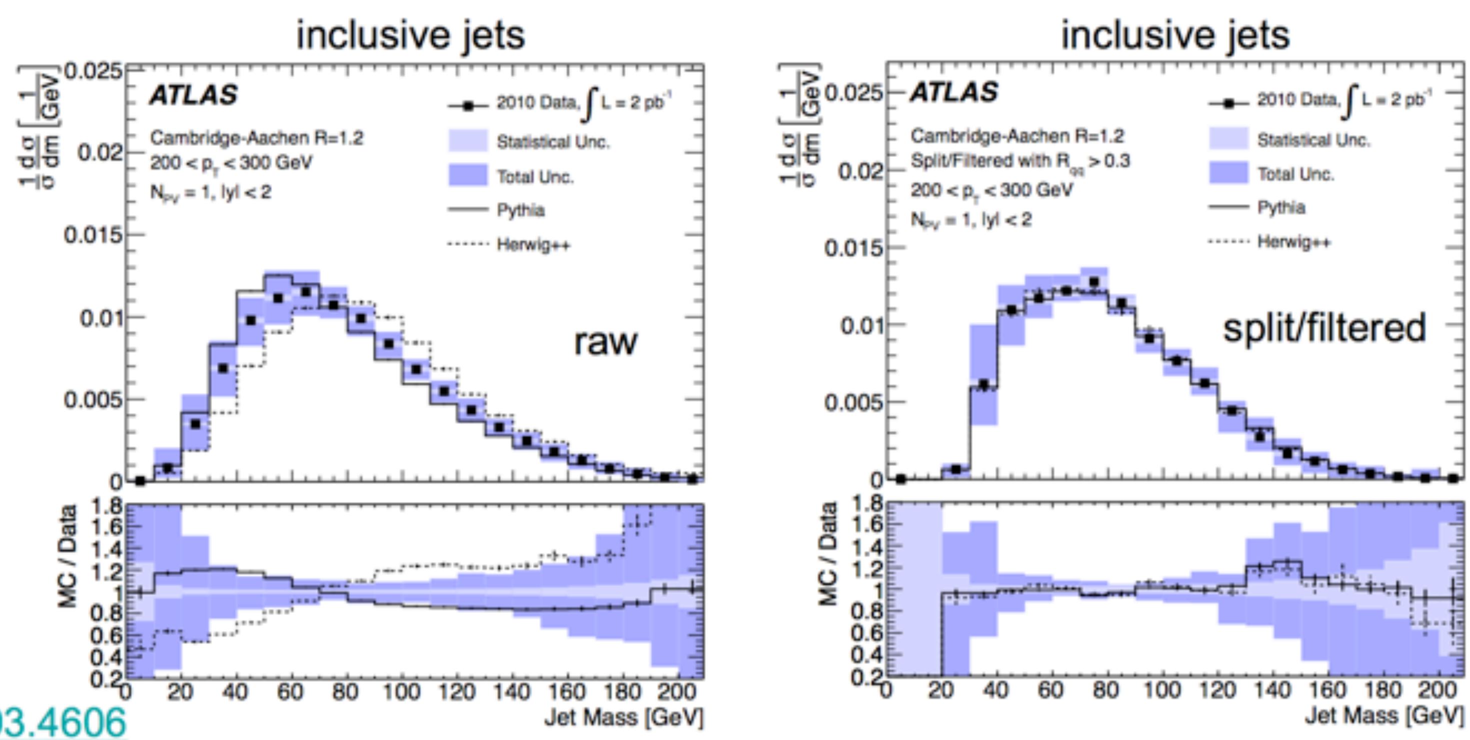
1203.4606

21 19 Aug 2014 Andreas Hinzmann University of Zurich

Substructure measurements

Jet mass – 1

- Raw jet mass difficult to model, generators disagree
- Groomed jet masses (here split/filtered) agree better between generators and with data



1203.4606

22 19 Aug 2014 Andreas Hinzmann University of Zurich

Substructure measurements

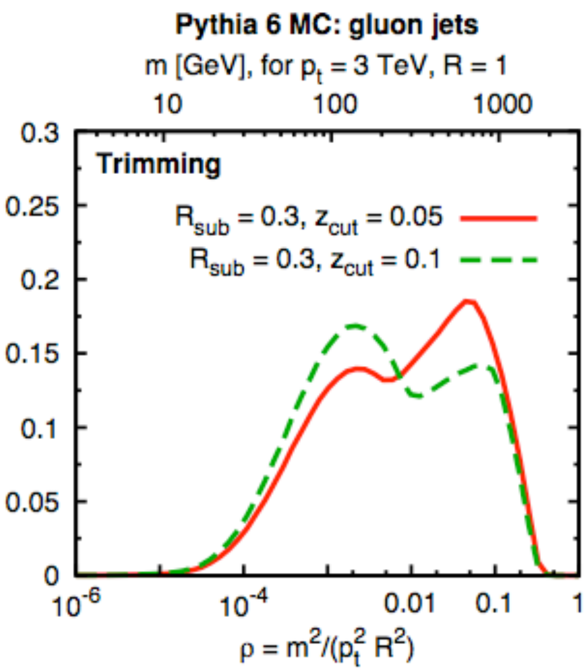
- **Take-home message**
  - Things 'generally' work well
  - Non-negligible residual uncertainties from Monte Carlo modeling
    - Need to design variables with well understood sensitivity to non-perturbative physics
    - Need to properly assess and quote systematic uncertainties

Recent progress in taggers/groomers

A lot of the recent activity has been centred on analytical understanding of existing taggers/groomers
[though not exclusively, new developments are also taking place -- see e.g. next slide and Tilman's talk]

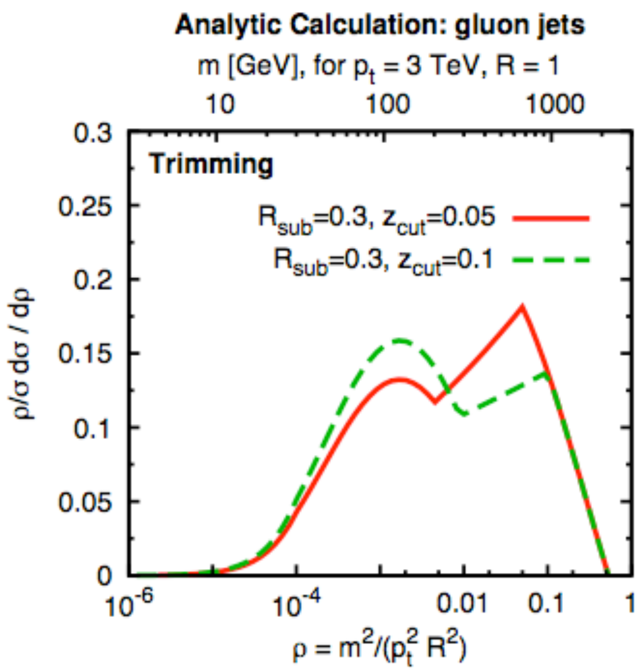
Dasgupta, Fregoso, Marzani, Salam, 2013

Monte Carlo



Analytic

(resummed pQCD)



- Analytical understanding of 'kinks' in distributions
- Check of Monte Carlo predictions
- Other analytical investigations: Rubin 2010 (filtering), Walsh, Zuberi 2011 (jet substructure with SCET), Feige Schwartz, Stewart, Thaler 2012 (N-subjettiness), Dasgupta, Marzani, Powling 2013 (groomed jet mass), ...

Soft Drop declustering

Larkoski, Marzani, Soyez, Thaler, 2014

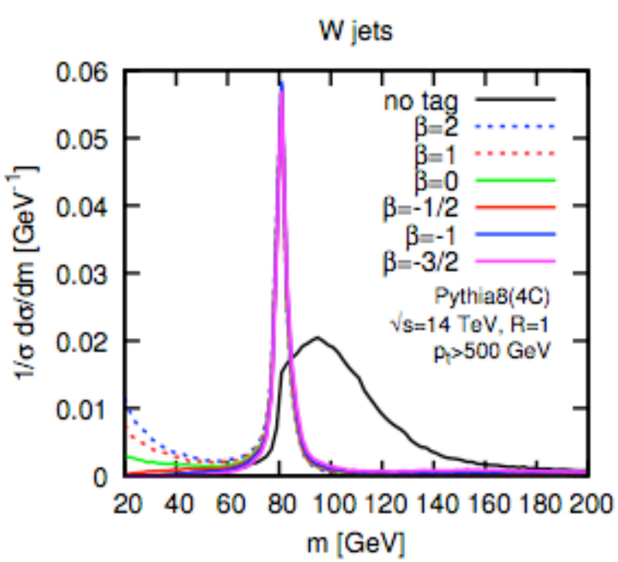
Decluster and drop softer constituent unless

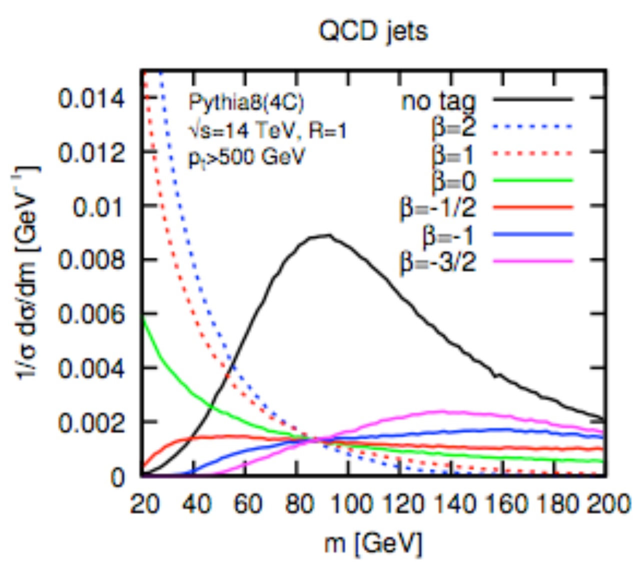
$$\text{Soft Drop Condition:} \quad \frac{\min(p_{T1}, p_{T2})}{p_{T1}+p_{T2}} > z_{\text{cut}} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

i.e. remove wide-angle soft radiation from a jet

The paper contains

✓ analytical calculations and comparisons to Monte Carlos
✓ study of effect of non-perturbative corrections
✓ performance studies

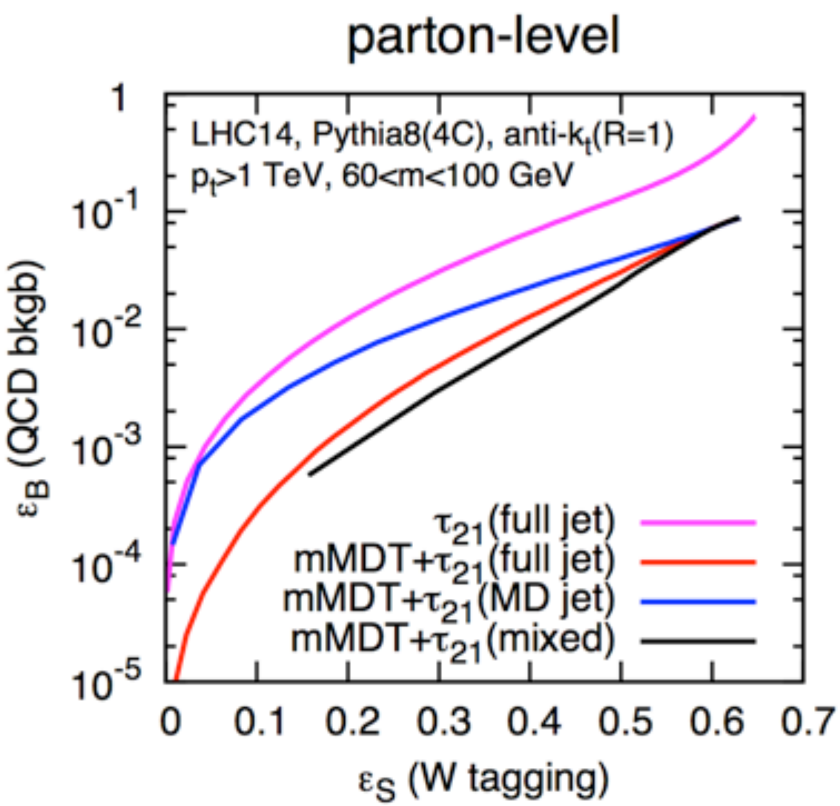




Example of SoftDrop performance when used as a boosted W tagger

Taggers performance

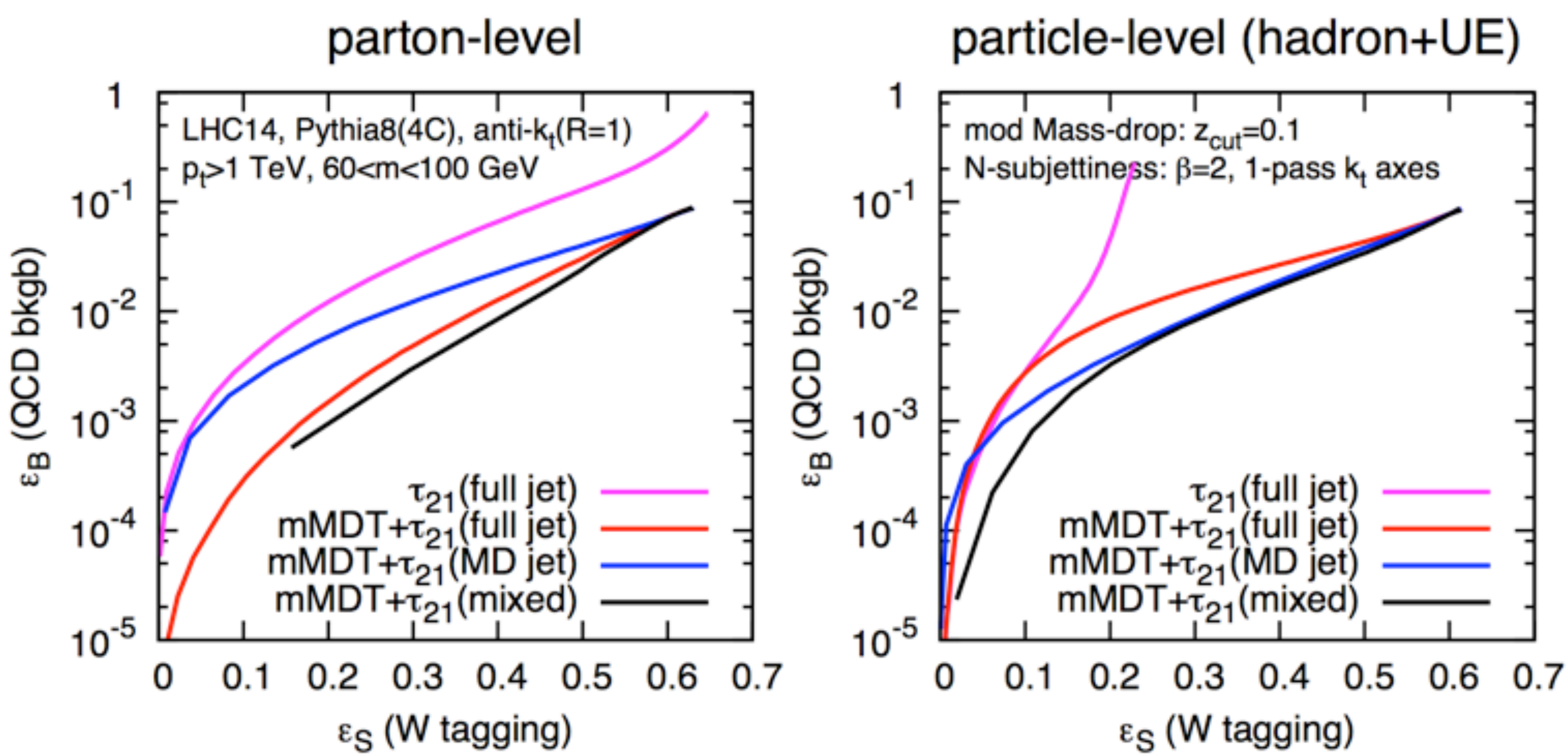
Example: mMDT + N-subjettiness



- Combining helps!
- Various options for τ_{21}
  - τ_2 and τ_1 from the full jet
  - τ_2 and τ_1 from MD'd jet
  - τ_2 from MD, τ_1 from full
- mixed case most efficient
  - τ_1 from MD: 2-prongs resolved
  - τ_2 from full: reach large angles

Taggers performance

Example: mMDT + N-subjettiness



- Non-perturbative effects can change the picture quite drastically
- using mass-drop everywhere (*i.e.* grooming) limits NP effects

Taggers: take-home messages

- Many different options
- Performance validated by measurements
- Combinations of different kinds of taggers (e.g. prongs-based + radiation-based) brings improvements [see Tilman's talk]
- Correlations and dependence on non-perturbative effects still to be properly assessed/understood

Recent progress in pileup removal

A lot of recent activity:

- CMS Voronoi method (Lai, unpubl.)
- Cleansing (Krohn, Schwartz, Low, Wang, 1309.4777)
- corrJVF (ATLAS-PHYS-PUB-2014-001)
- Constituent Subtraction (Berta, Spousta, Miller, Leitner, 1403.3108)
- NpC (MC, Salam, Soyez, 1404.7353)
- PUPPI (Bertolini, Harris, Low, Tran, 1407.6013)
- SoftKiller (MC, Salam, Soyez, 1407.0408)
- ...

Various methods reviewed and compared at the CERN pileup workshop in May
https://indico.cern.ch/event/306155/

Pileup subtraction methods

Full jet/Observable level

- Determination of *susceptibility to contamination* of each specific observable needed

- Basic example: transverse momentum
  $\mathbf{p_t^{sub} = p_t^{raw} - \rho A}$ (MC, Salam 0707.1378)

- Other examples:
  - Analytical calculations of susceptibility for selected jet shapes (Sapeta et al. 1009.1143, Alon et al. 1101.3002)
  - Moments of jet fragmentation functions (MC, Quiroga, Salam, Soyez, 1209.6086)
  - Generic (numerical) approach to susceptibility determination for any shape (Soyez et al, 1211.2811)

Subjet/particle level

Pileup subtraction methods

Full jet/Observable level

- Determination of *susceptibility to contamination* of each specific observable needed

- Basic example: transverse momentum
  $\mathbf{p_t^{sub} = p_t^{raw} - \rho A}$ (MC, Salam 0707.1378)

- Other examples:
  - Analytical calculations of susceptibility for selected jet shapes (Sapeta et al. 1009.1143, Alon et al. 1101.3002)
  - Moments of jet fragmentation functions (MC, Quiroga, Salam, Soyez, 1209.6086)
  - Generic (numerical) approach to susceptibility determination for any shape (Soyez et al, 1211.2811)

Subjet/particle level

- The event is modified before calculating observables (jets, shapes, etc). Corrections applied to subjets or even to particles

- Examples (warning: shaky classification, to be refined)
  - Cleansing (Krohn, Schwartz, Low, Wang, 1309.4777)
  - corrJVF (ATLAS-PHYS-PUB-2014-001)
  - NpC (MC, Salam, Soyez, 1404.7353)

  - CMS Voronoi method (Lai, unpubl.)
  - Constituent Subtraction (Berta, Spousta, Miller, Leitner, 1403.3108)
  - PUPPI (Bertolini, Harris, Low, Tran, 1407.6013)
  - SoftKiller (MC, Salam, Soyez, 1407.0408)
  - ...

Pileup subtraction methods

Full jet/Observable level

- **Pros**:
  - Subtraction is unbiased by construction
  - Not too sensitive to detector effects (works at the jet/subjet level)
- **Cons**:
  - Need to cluster (e.g. to calculate areas), hence time-consuming

Subjet/particle level

Pileup subtraction methods

Full jet/Observable level

- **Pros**:
  - Subtraction is unbiased by construction
  - Not too sensitive to detector effects (works at the jet/subjet level)
- **Cons**:
  - Need to cluster (e.g. to calculate areas), hence time-consuming

Subjet/particle level

- **Pros**:
  - Often no need to cluster
  - Dispersion usually reduced
  - If left with a 'subtracted' event, one can then calculate any observable
- **Cons**:
  - Potentially sensitive to detector effects
  - Potentially biased (but can usually be tuned)

Area-median

- **Working hypothesis**: pileup energy flow is distributed sufficiently uniformly over the event

- Estimate **pileup transverse momentum density ρ**, using measurements of energy flow in patches of given size
  - Possibly rescale ρ as a function of rapidity and azimuth

- Calculate **area A_μ** of each jet

- Subtract pileup contamination using

$$p_\mu^{\text{jet,sub}} = p_\mu^{\text{jet,full}} - \rho A_\mu$$

(This method can be adapted to jet shapes)

Jet-shape subtraction

The **p_T^{raw}-ρA** technique (also called **area/median**) only corrects a jet's transverse momentum

Each jet shape has its own specific sensitivity to background contamination. **How to correct them?**

- One option is to study analytically each shape [Sapeta et al. 1009.1143, Alon et al. 1101.3002].
  Can be time consuming and cumbersome

- Alternatively, determine **numerically** the *susceptibility* of any IRC-safe jet shape to contamination [Soyez et al. 1211.2811] (this generalises the jet area)

Numerical jet shape correction

Numerical derivative w.r.t. ghosts momenta

Ghosts area

Jet shape as a function of the jets's constituents momenta

$$V_{\rm jet}^{[n]} \equiv A_g^n \frac{d^n}{dr_{t,g}^n} V(\{p_i\}_{\rm jet})$$

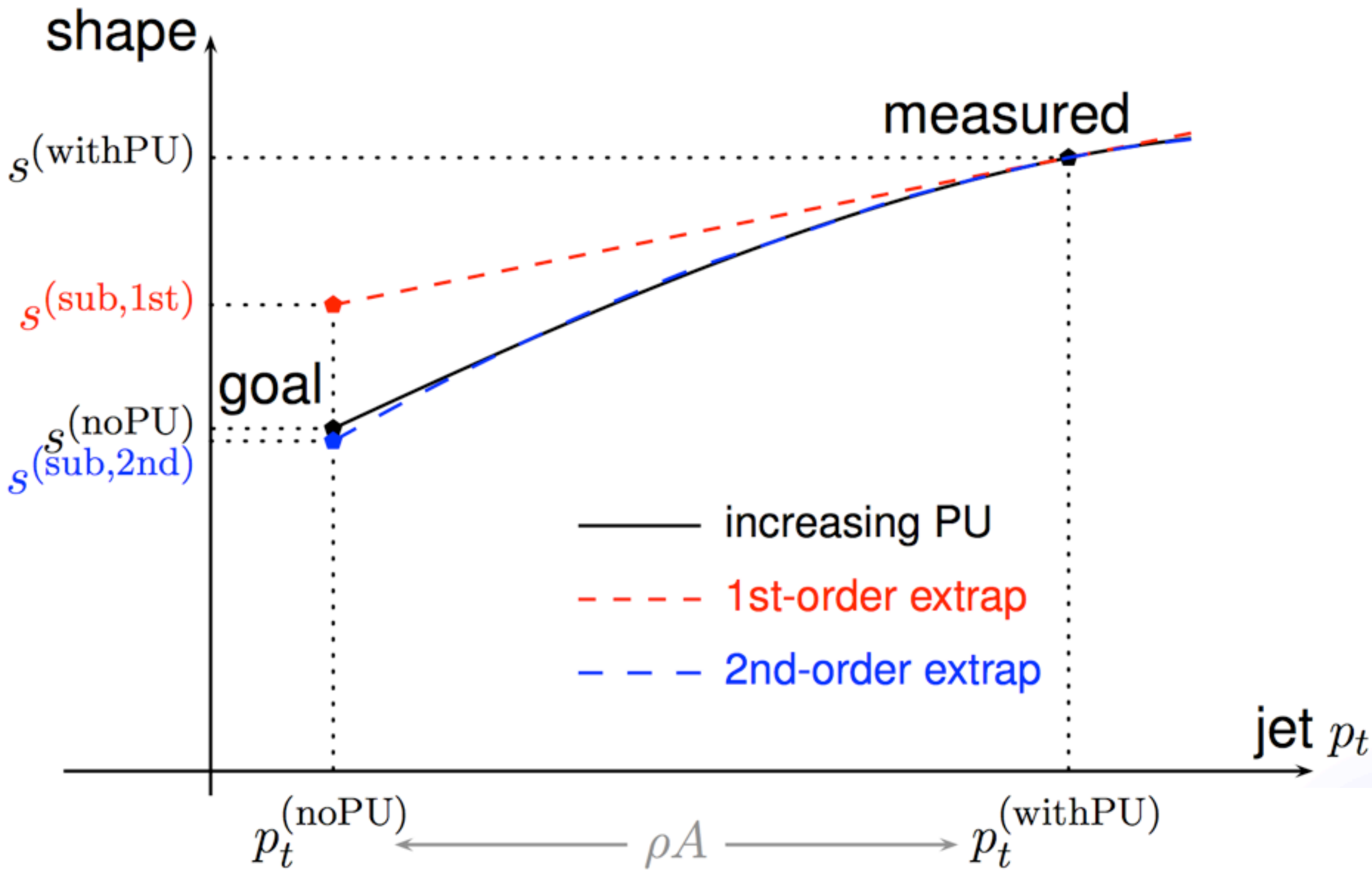
Pileup momentum density

$$V_{\rm jet,sub} = V_{\rm jet} - \rho V_{\rm jet}^{[1]} + \frac{1}{2}\rho^2 V_{\rm jet}^{[2]} + \cdots$$

Numerical derivative w.r.t. ghosts momenta

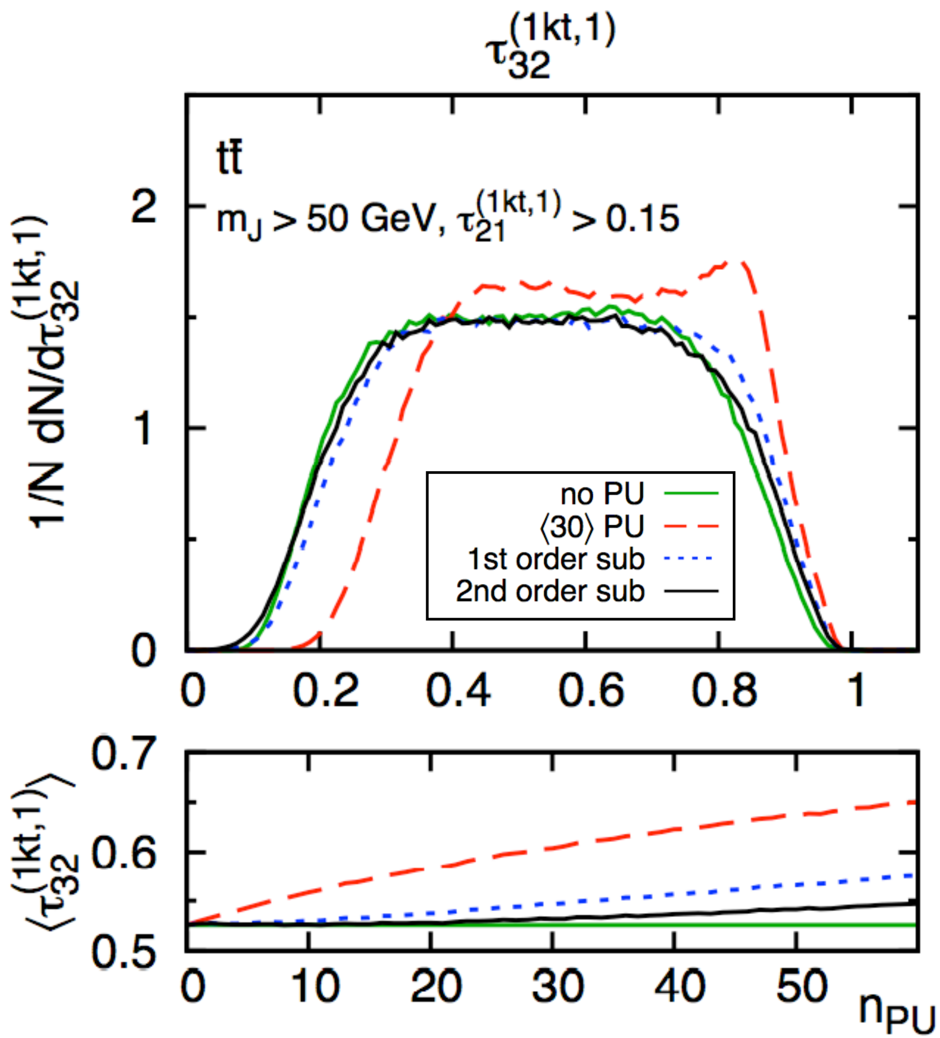
This procedure generalises the transverse momentum correction to any jet shape

Numerical jet shape correction



Numerical jet shape correction

Example: τ_{32} correction and top tagging [Soyez et al. 1211.2811]

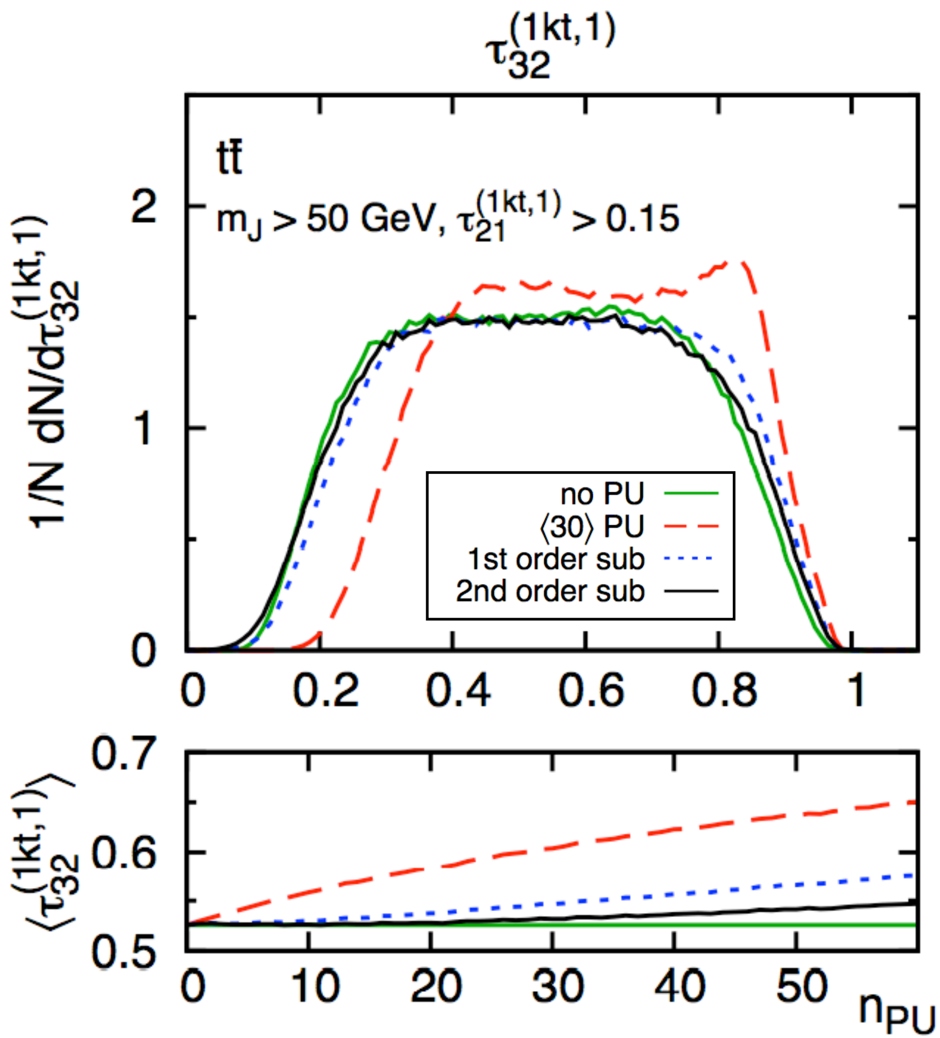


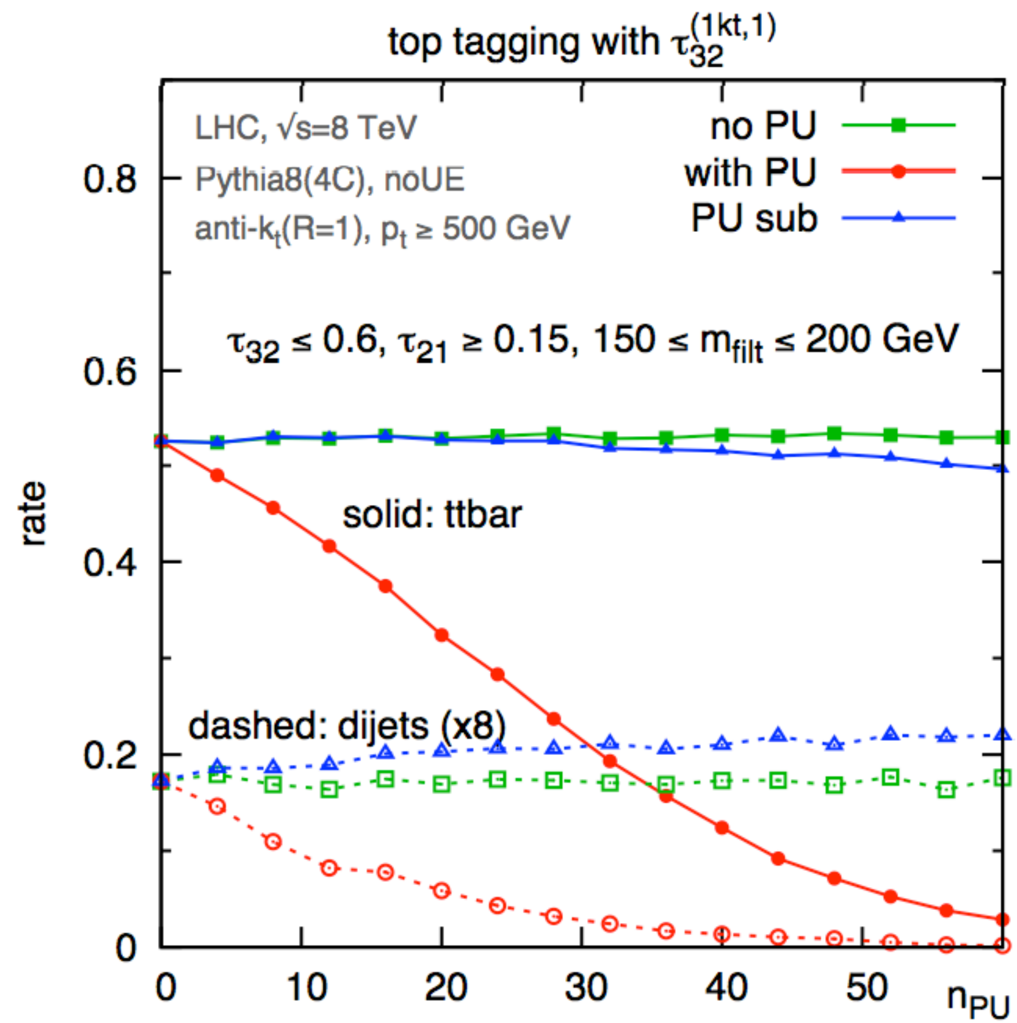
- Original distribution reproduced after pileup subtraction

Numerical jet shape correction

Example: τ_{32} correction and top tagging

- [Soyez et al. 1211.2811]

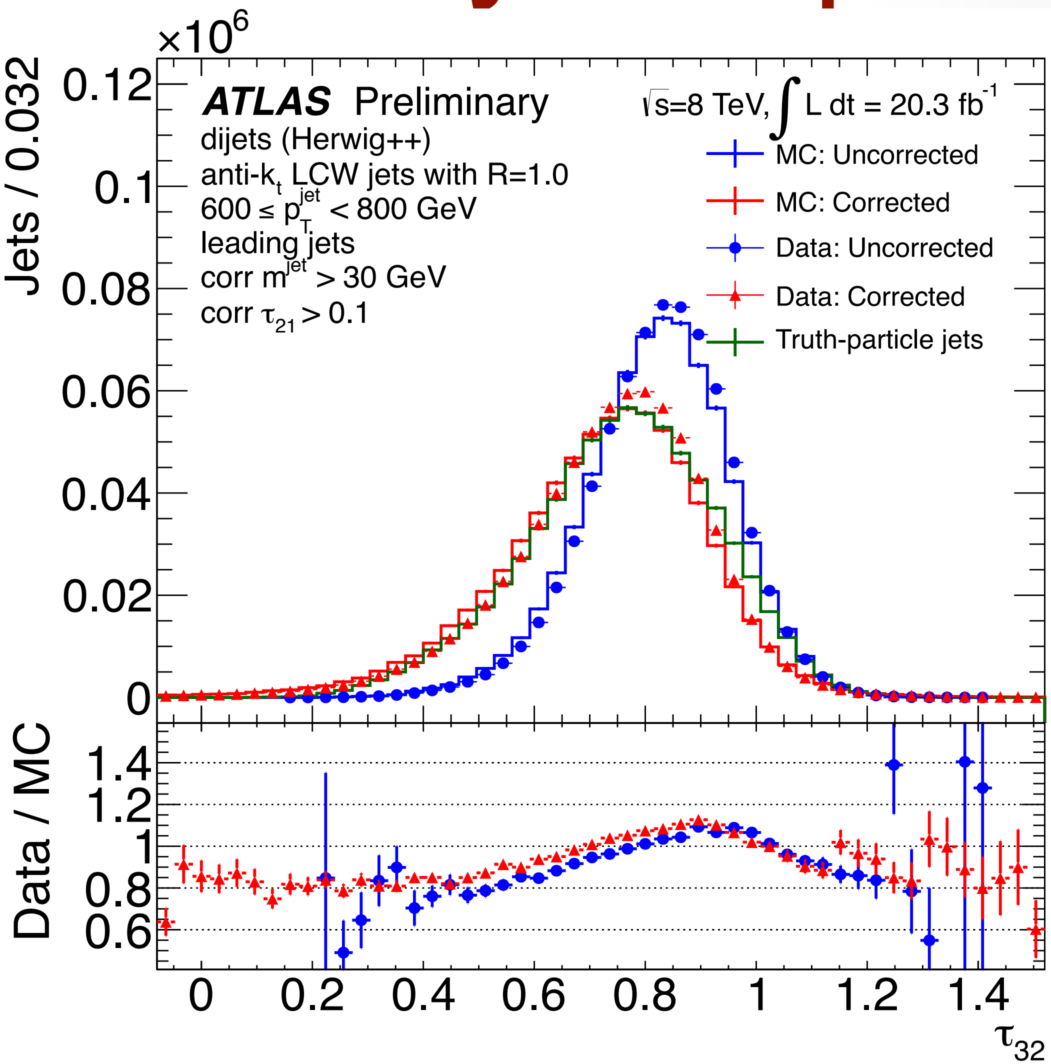




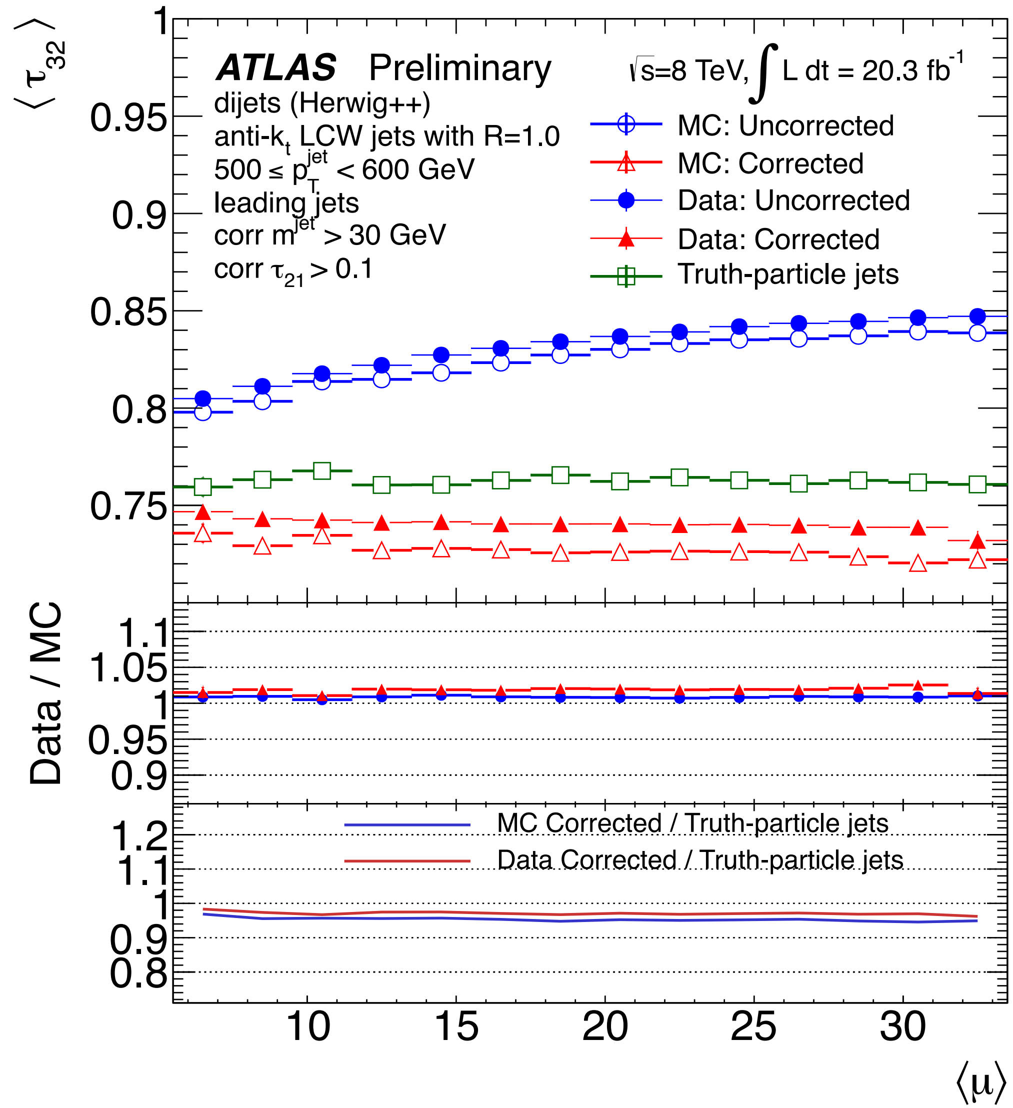
- Original distribution reproduced after pileup subtraction
- **Tagging rates independent of amount of pileup after correction of the jet shapes involved in the tagging**

ATLAS-CONF-2013-085

Jet shape subtraction in ATLAS



τ_{32} distribution



$<\tau_{32}>$ as a function of the average number of pileup collisions

Neutral proportional to Charged

arXiv:1404.7353

- **Working hypotheses:**
  1. one can **detect all charged** particles from pileup, and therefore **measure** $p_\mu^{jet,chg\text{-}PU}$
  2. the momentum from the unseen **neutral** component of pileup is **proportional** to the measured **charged** one, i.e. there exists a **fixed charged fraction** $\gamma_0 = p^{chg\text{-}PU}/p^{PU}$

- Then, two options:
  - Use **full** event, and subtract pileup as

$$p_\mu^{\rm jet,sub} = p_\mu^{\rm jet,full} - \frac{1}{\gamma_0} p_\mu^{\rm jet,chg\text{-}PU}$$

  - Use **CHS** (= charged hadron subtracted) event, i.e. without charged particles from pileup (technically, scaled by ε << 1) and subtract as

$$p_\mu^{\rm jet,sub} = p_\mu^{\rm jet,CHS} - (1-\gamma_0)\frac{1}{\gamma_0}\frac{p_\mu^{\rm jet,rescaled\text{-}chg\text{-}PU}}{\epsilon}$$

When can one expect things to work well?

- For **area-median**, if point-to-point pileup energy-flow fluctuations are moderate (since ρ is estimated globally)
- For **NpC**, if energy flows from neutral and charged pileup particles are really spatially well correlated

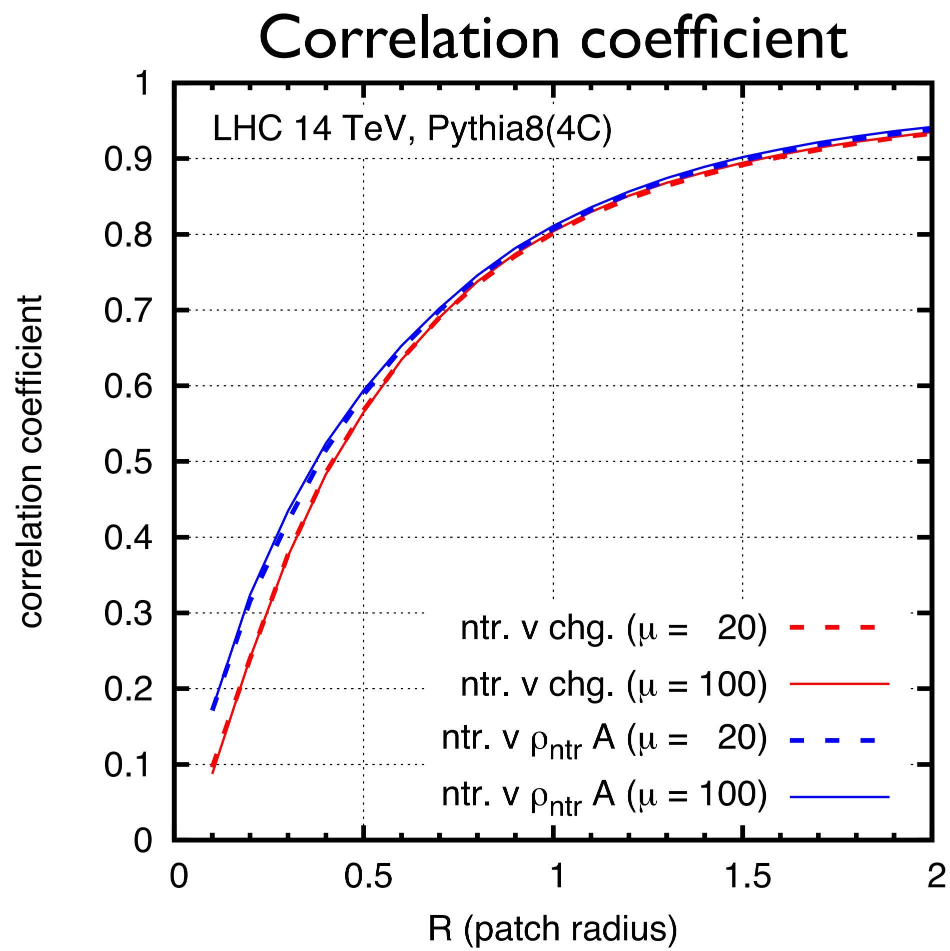
Which one wins?

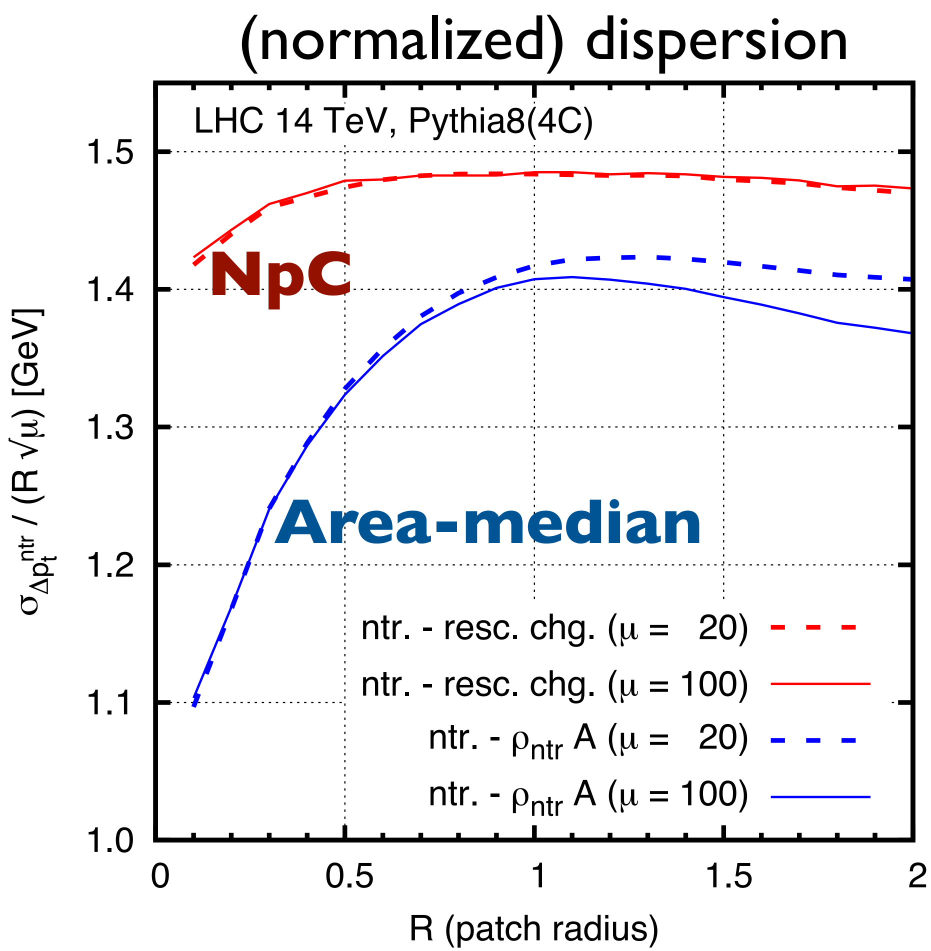
Check how well each method estimates the neutral component of pileup transverse momentum

1. **Estimate p_t^{ntr}** in a patch of radius R using
   - either $\rho^{ntr}A$
   - or $p_t^{chg}(1-\gamma_0)/\gamma_0$

2. **Determine quality of estimation** by looking
   - at the correlation coefficient between p_t^{ntr} and $\rho^{ntr}A$ or p_t^{chg}
   - at the dispersion of the "misestimations": $p_t^{ntr} - \rho^{ntr}A$ or $p_t^{ntr} - p_t^{chg}(1-\gamma_0)/\gamma_0$

area-median v. NpC



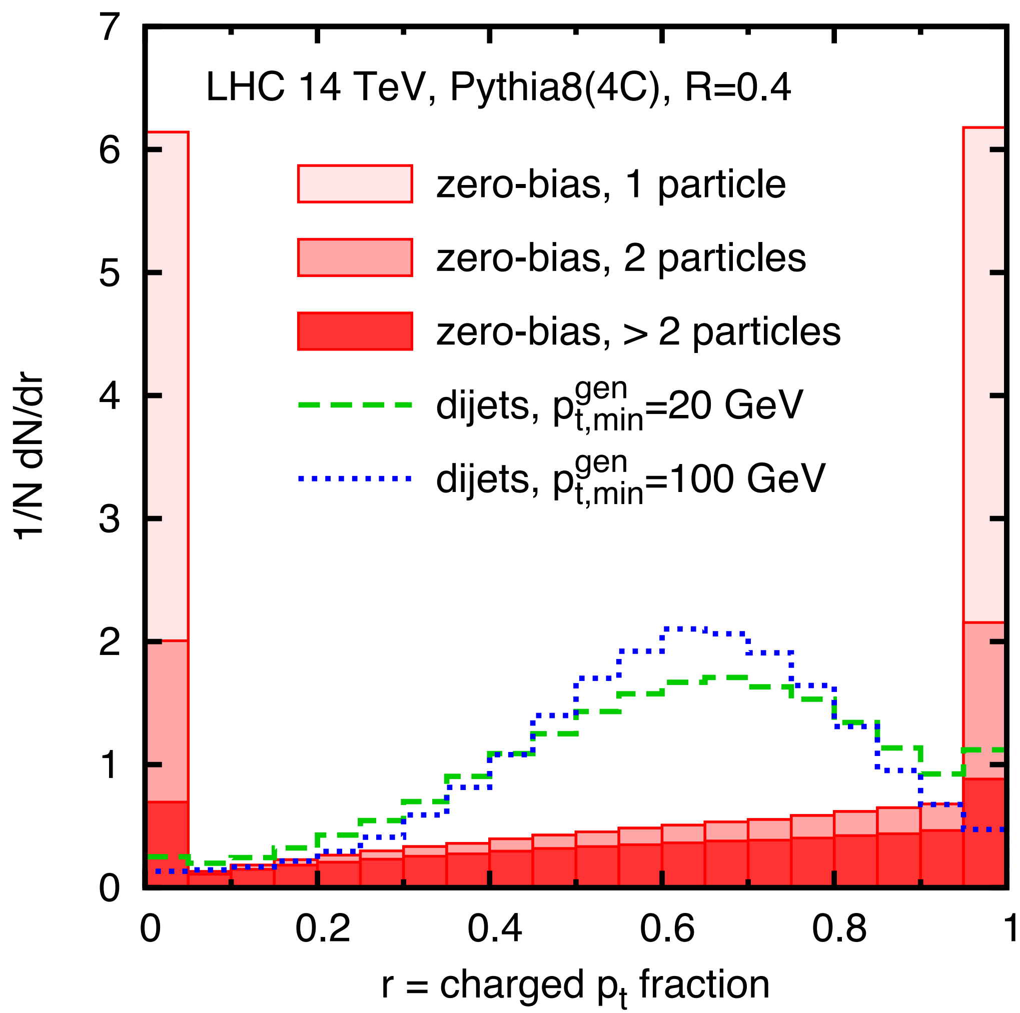


NpC is no better than area-median at estimating the neutral component of pileup p_t^{ntr} in a patch

In fact, area-median is slightly better at all values of R, and especially at small R (< 0.5)

Why is NpC no better?

Short answer: because local correlation between neutral and charged is not that great for pileup



Charged pt fraction in

- a patch of pileup of radius 0.4
- anti-kt jet with R=0.4

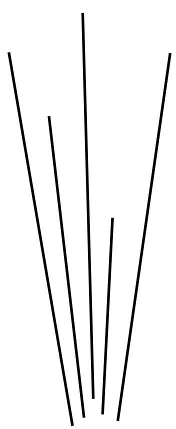
Most of the time, the pileup patch contains only a few charged or a few neutral particles from a given PU vertex, and the charged fraction is **not** peaked around a single value as in the case of the hard jets

Why is NpC no better?

Marginally longer answer: because decays of pileup particles tend to take place at large angle

hard jet

$p_t \gg \Lambda_{\mathrm{QCD}}$



hadronisation

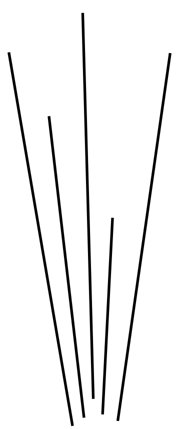
Charged and neutral energy tend to go in the same direction

Why is NpC no better?

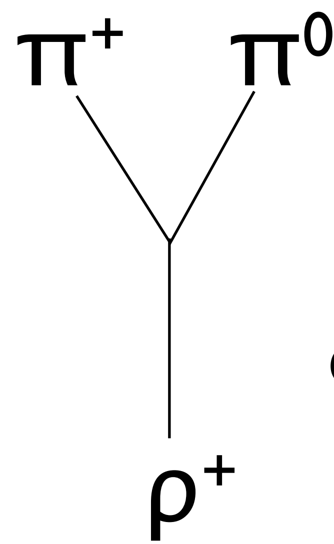
Marginally longer answer: because decays of pileup particles tend to take place at large angle

hard jet

$p_t \gg \Lambda_{\mathrm{QCD}}$



hadronisation



one step of hadronisation

pileup

$p_t \simeq 0.5 \text{ - } 2 \text{ GeV}$

Opening angle ~ $2m_\rho/p_{t,\rho}$ ~ 1

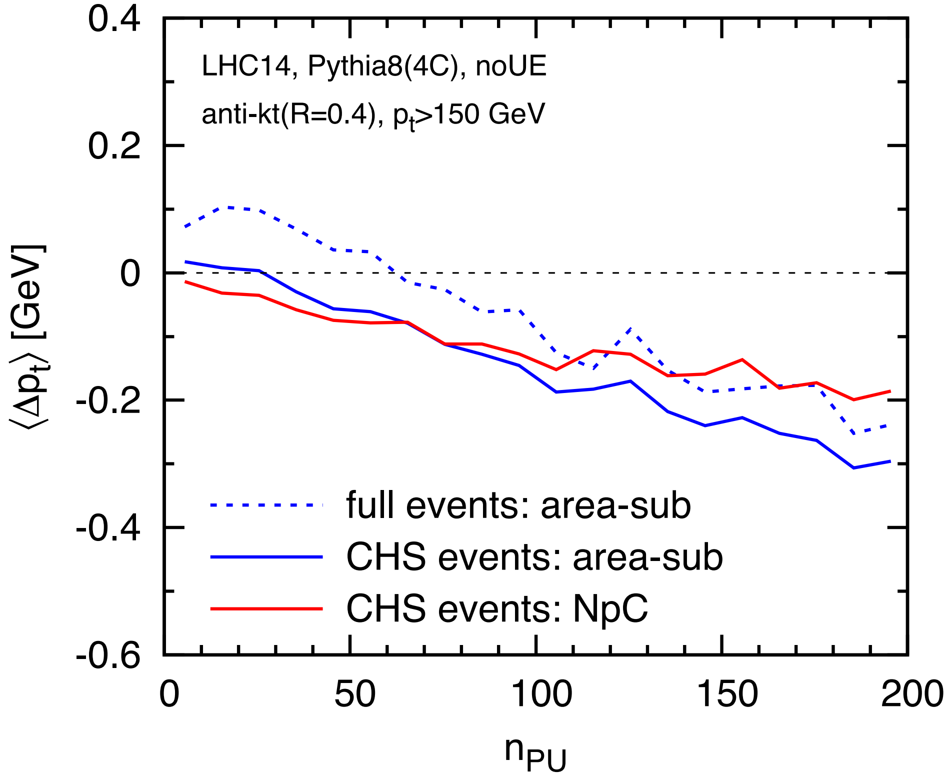
Charged and neutral energy tend to go in the same direction

Emissions at large angles break local correlation

Performance of NpC

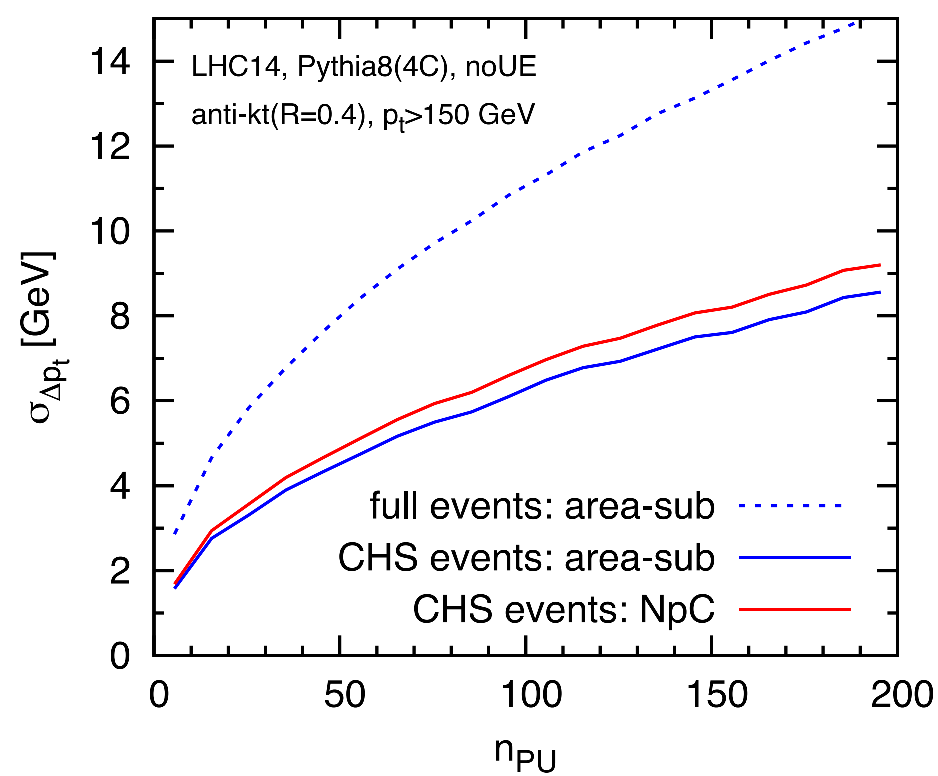
Pileup subtraction in dijet production
Observable: jet p_t

Offset



Offset limited to
a few hundred MeV

Dispersion



As expected,
$\sigma^{\mathrm{CHS,NpC}} \sim \sigma^{\mathrm{CHS,area}} < \sigma^{\mathrm{full,area}}$

Take-home message

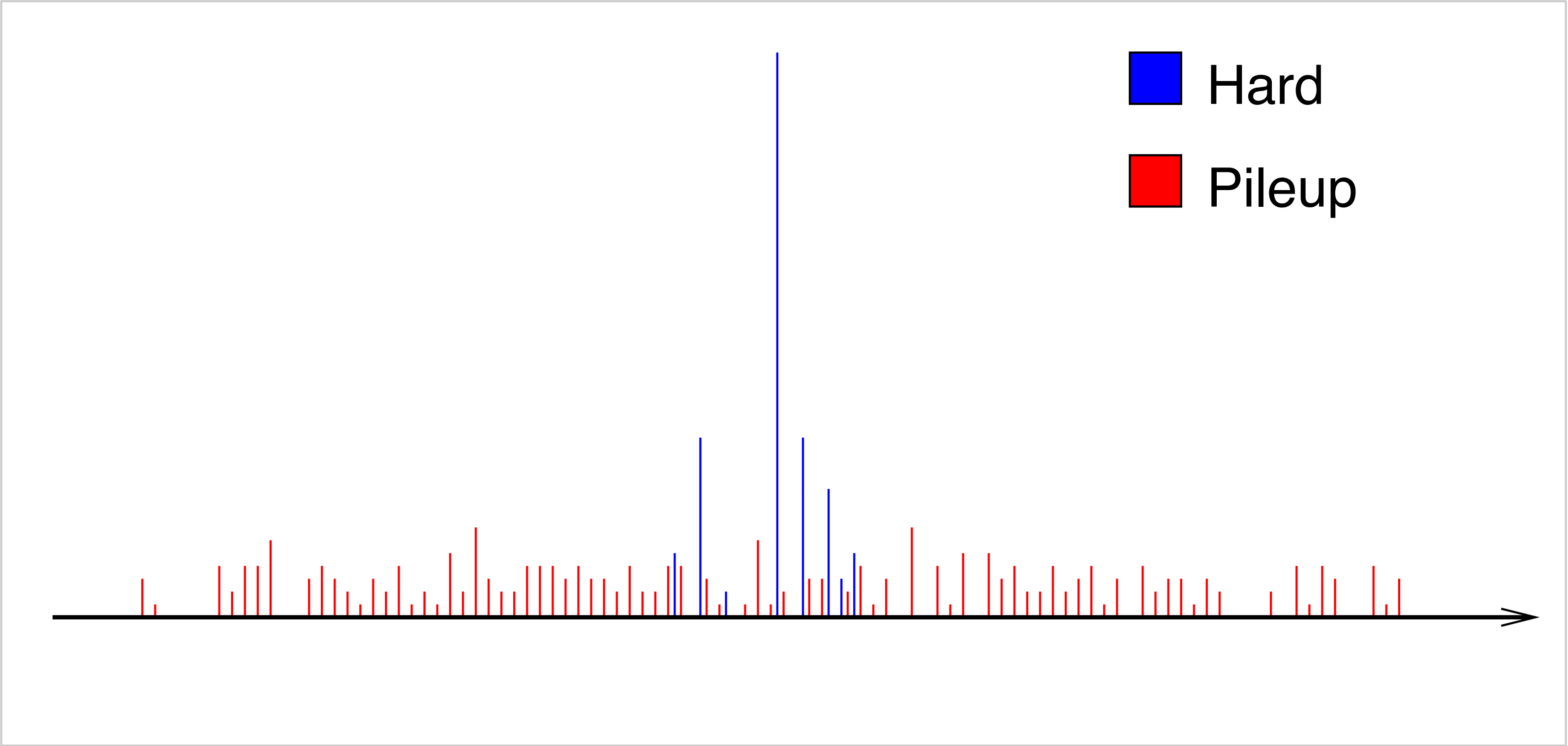
- **Area-median** and **NpC** perform similarly
  - Perhaps contrary to intuition, NpC does not work better, because large angle emissions of low-pt pileup particle tend to destroy local correlation between neutral and charged particles
  - The dispersion of both methods scales with $\sqrt{N_{PU}}$

SoftKiller

arXiv:1407.0408

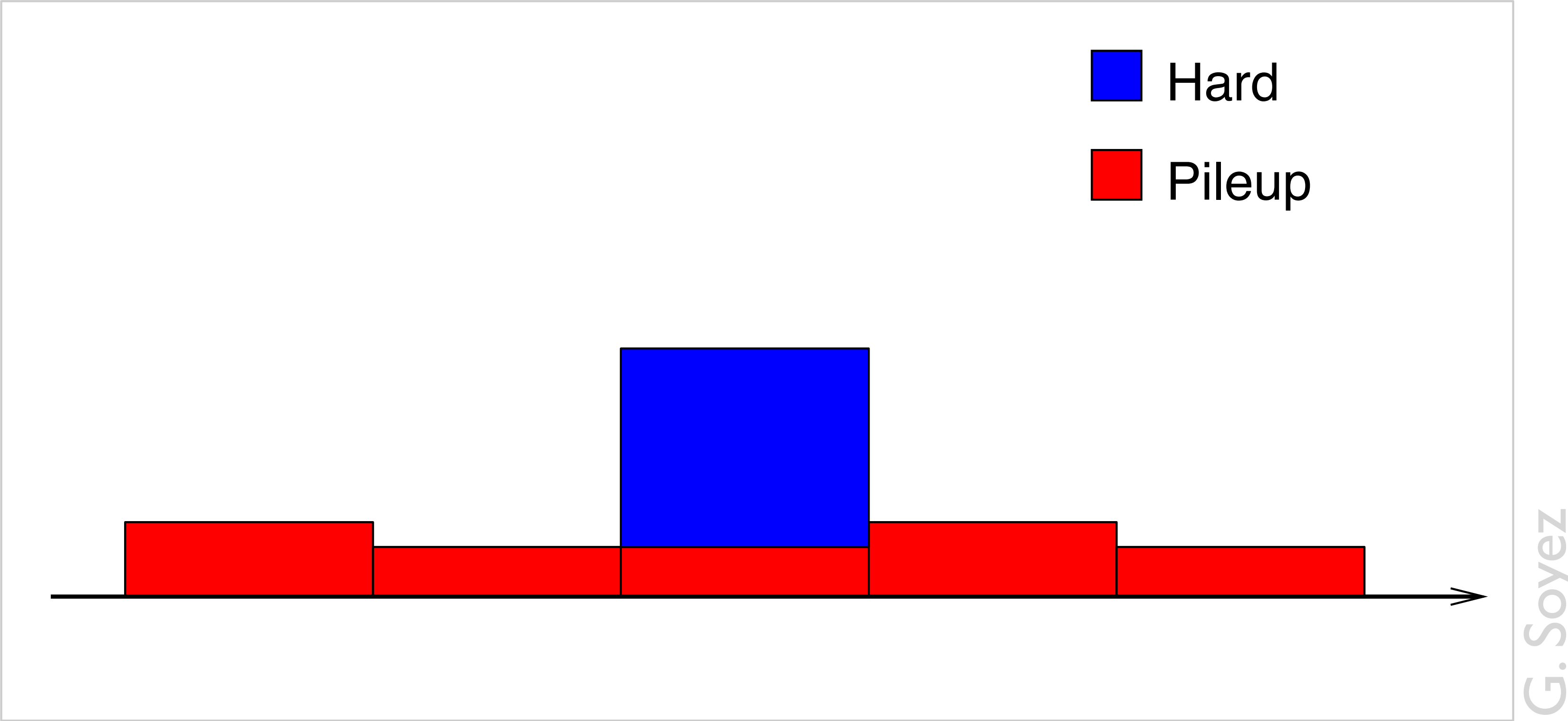
- A very simple, and fast, event-level pileup removal method
- Essentially a p_T cut with a dynamical, event-by-event-determined threshold

An event: particle level

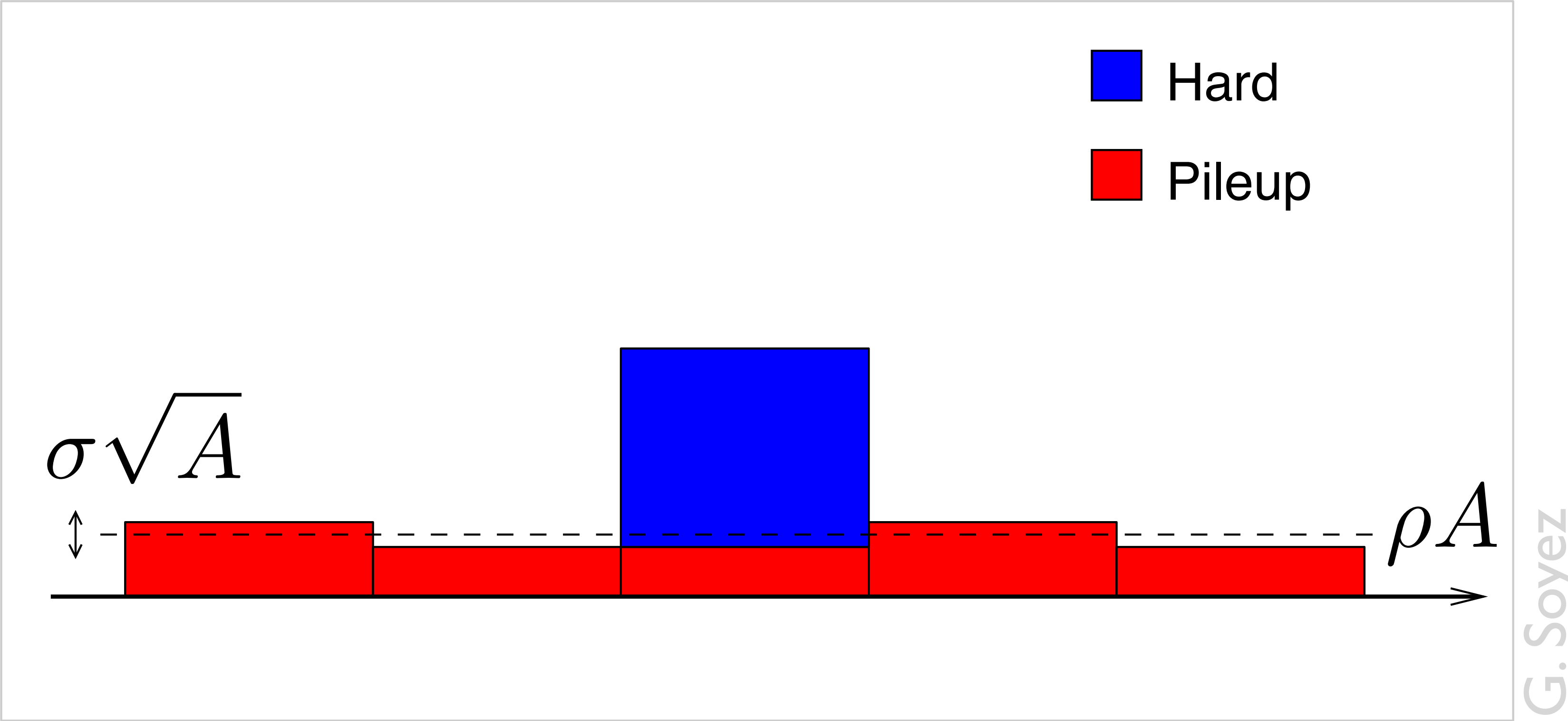


G. Soyez

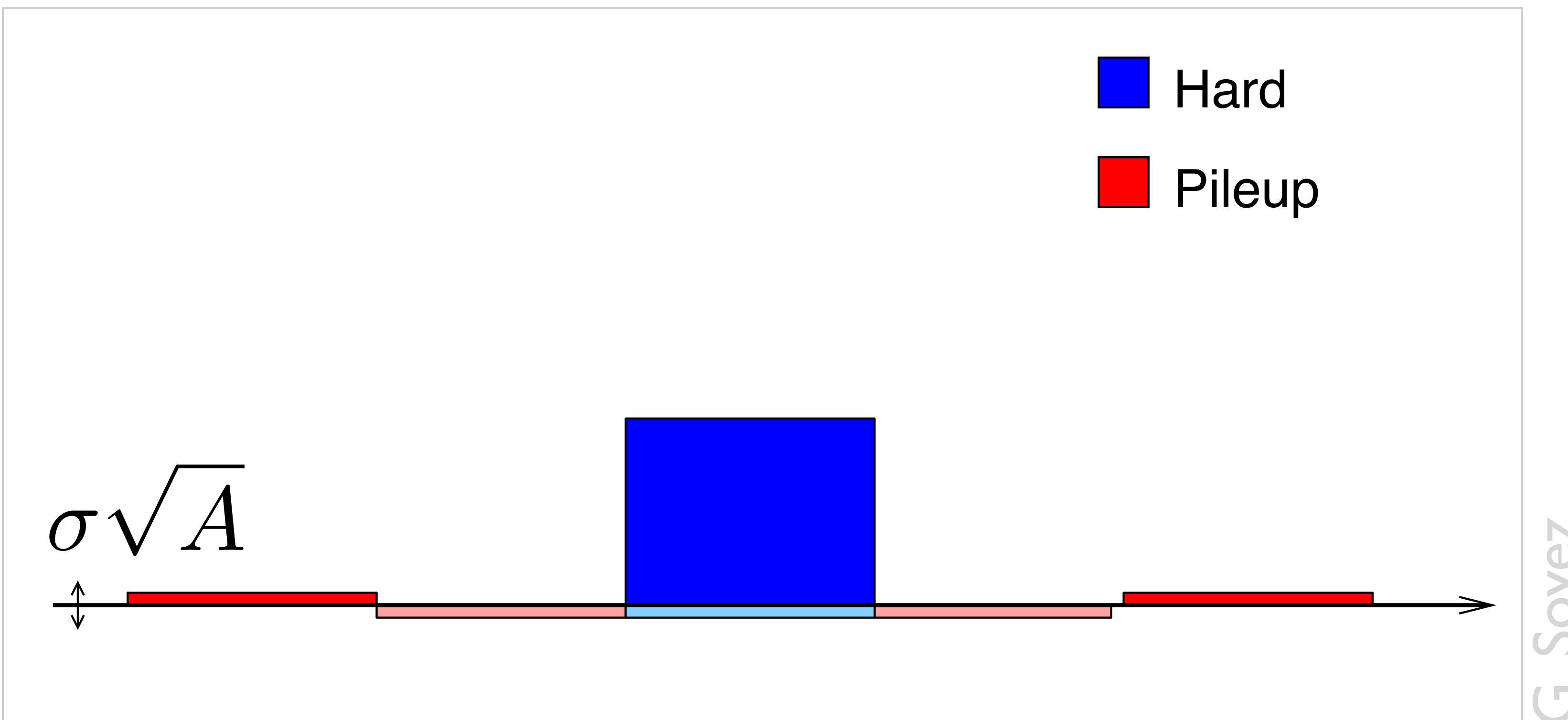
An event: jet level



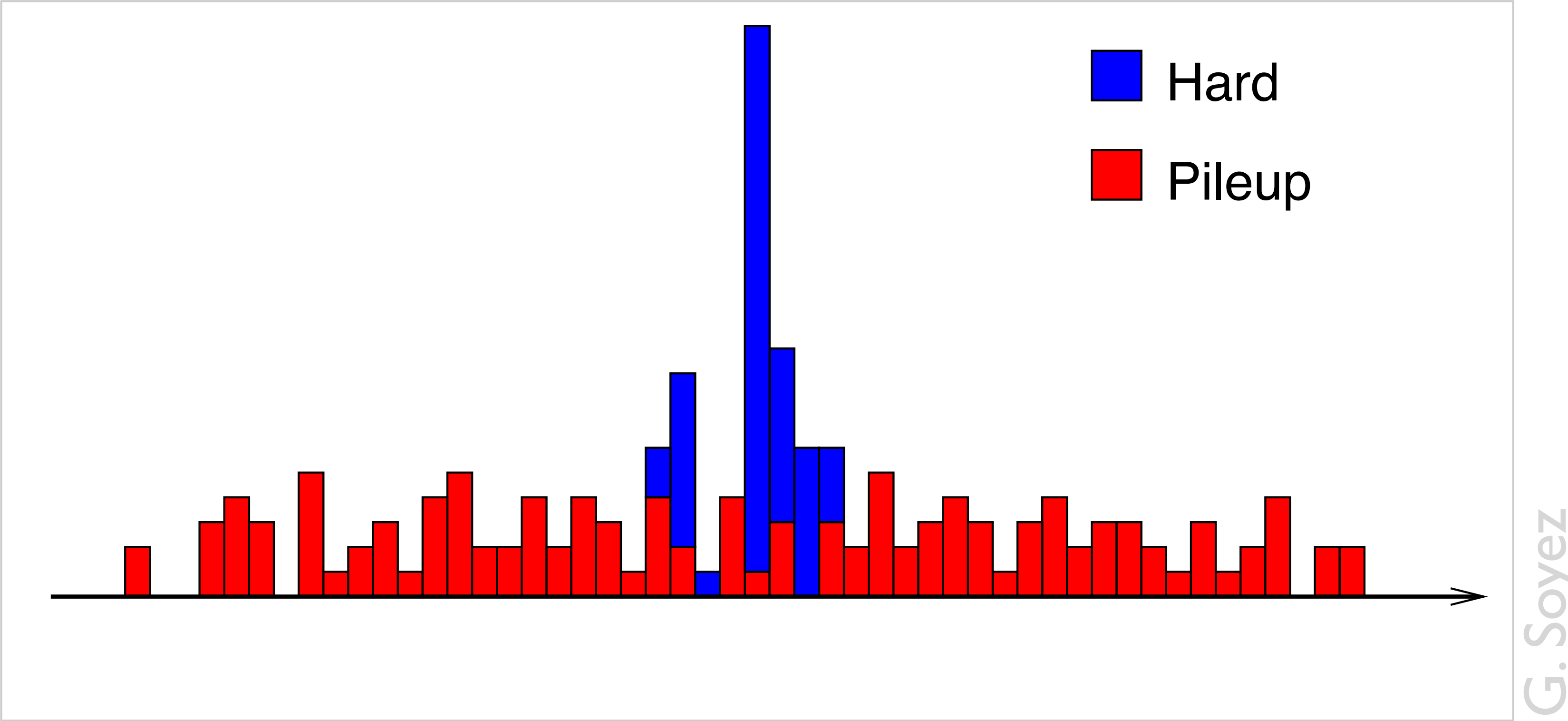
Background determination



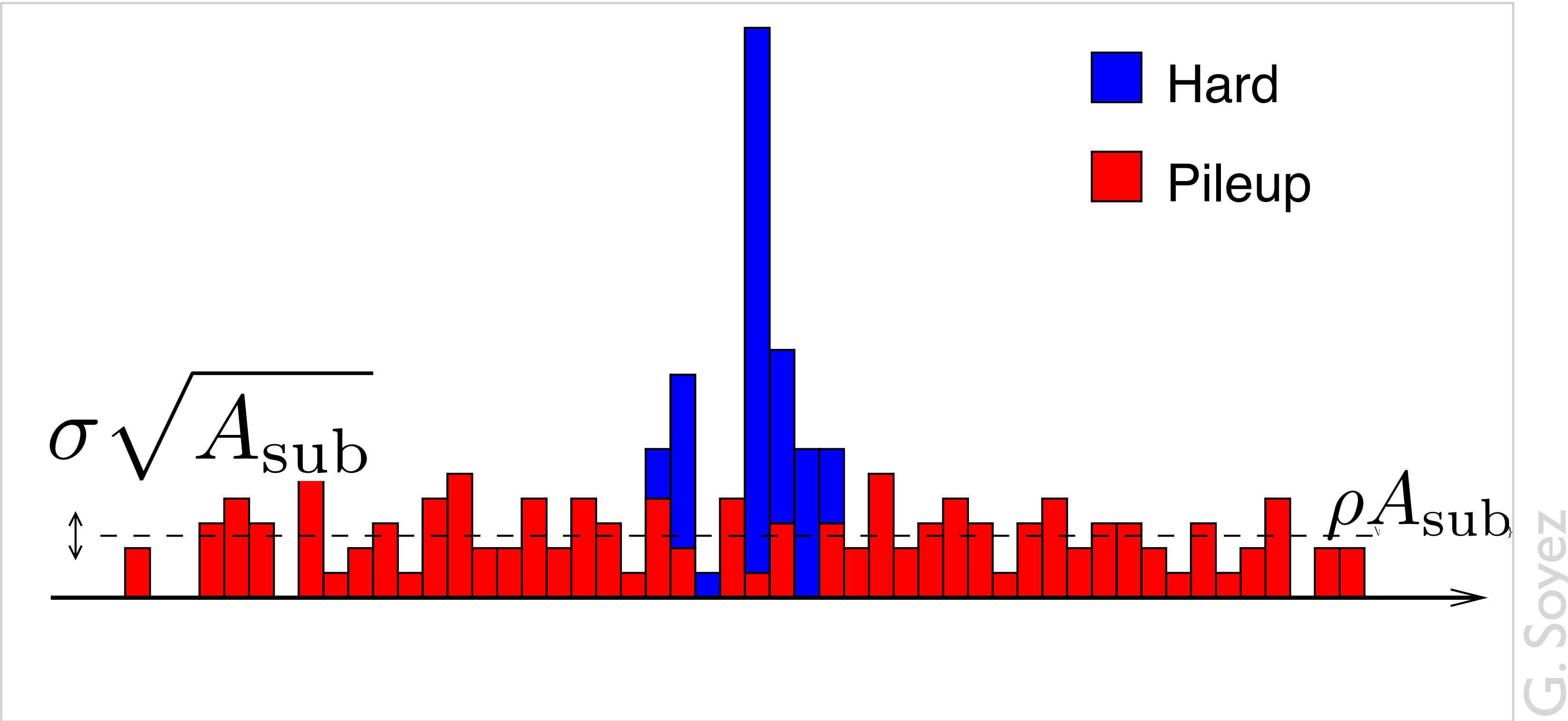
Background subtraction



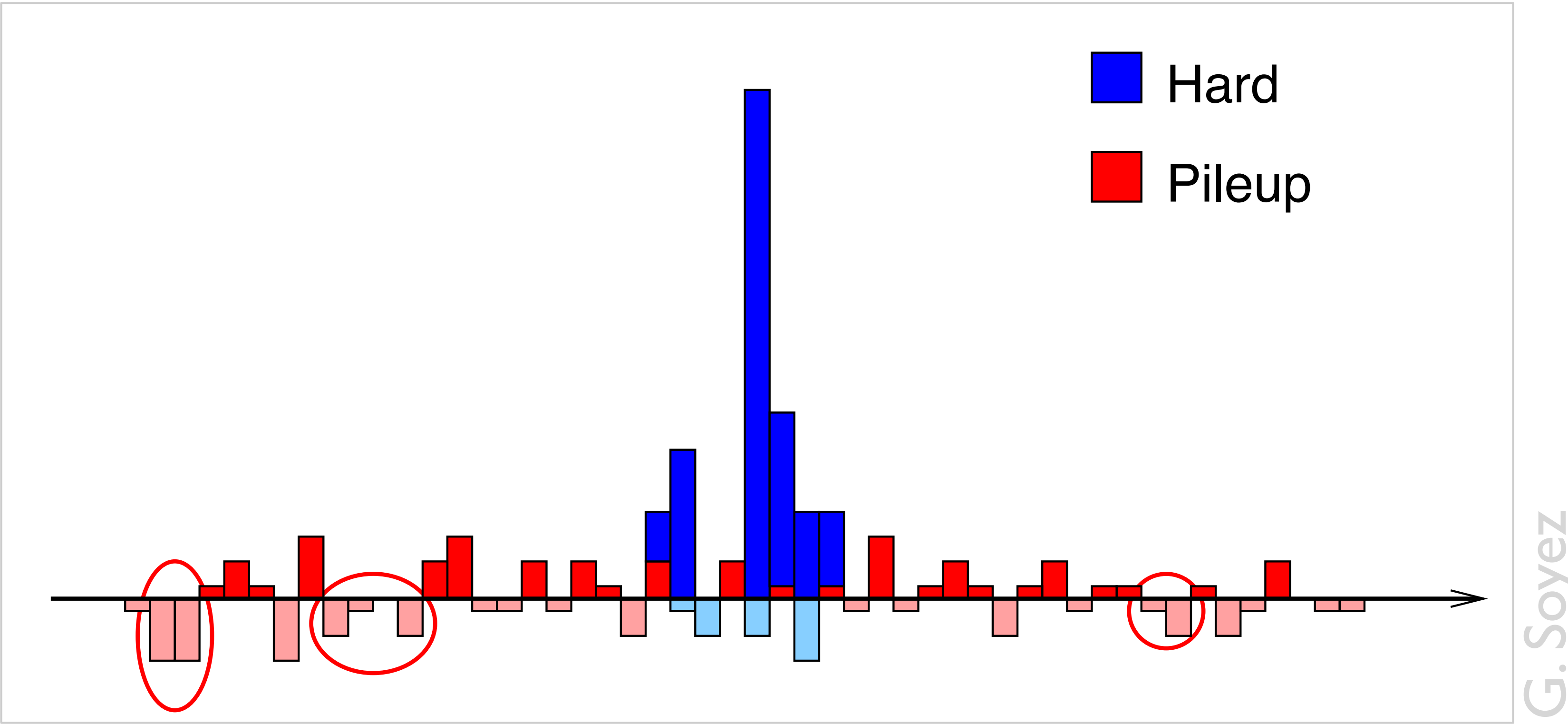
Use subjets instead



Use subjets instead

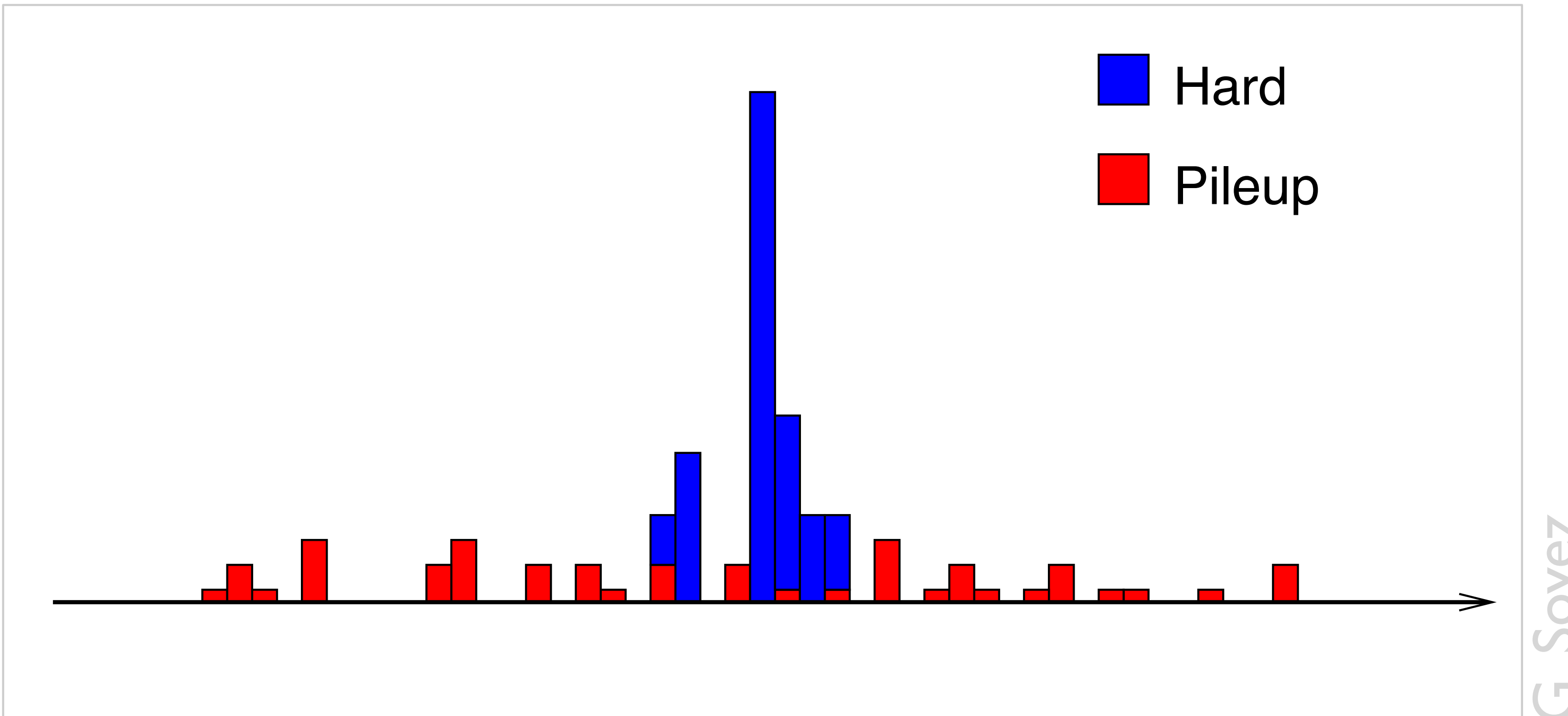


Use subjets instead



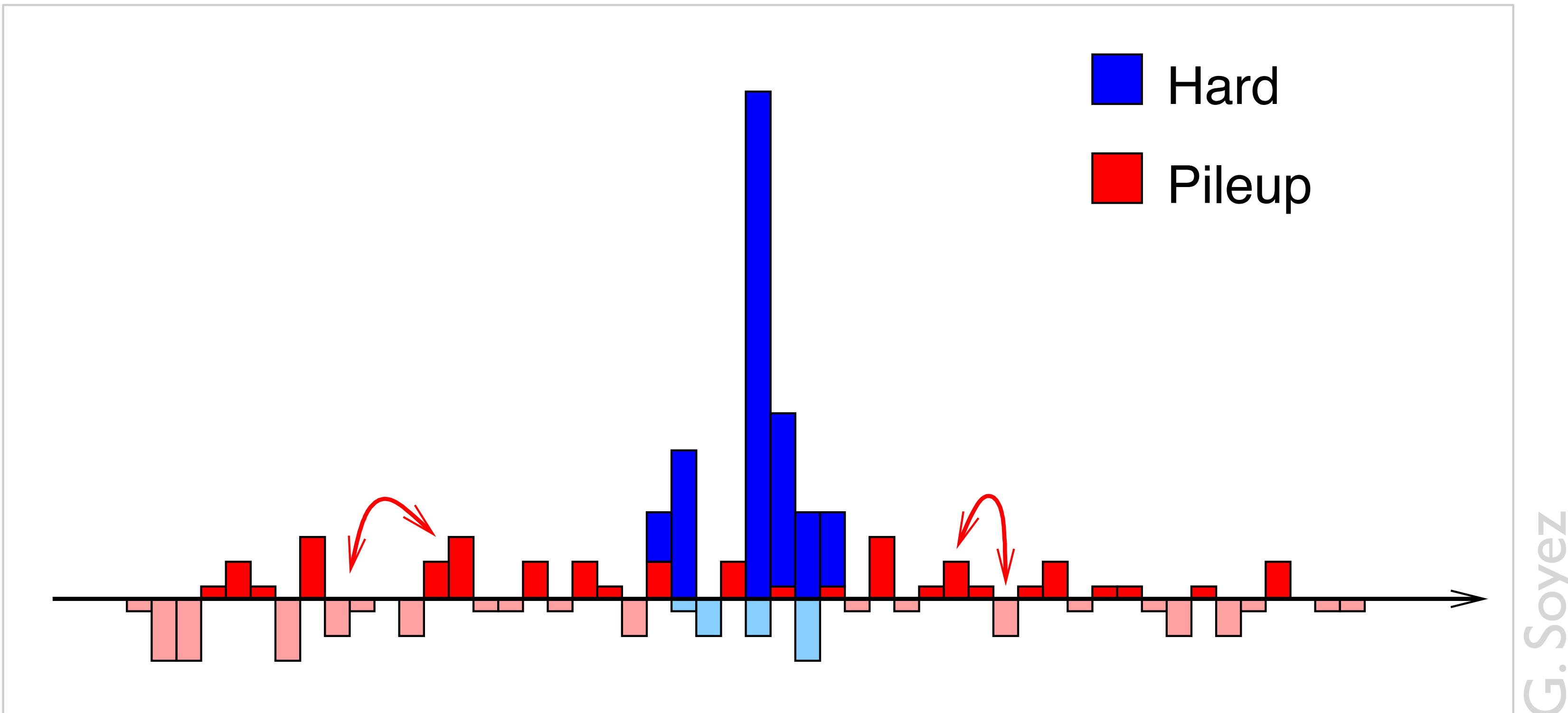
Same average (~ 0) and dispersion on the scale of the full jet, but some subjets are negative

Use subjets instead



Most naive noise reduction approach: cut away the negative part.
Biased.

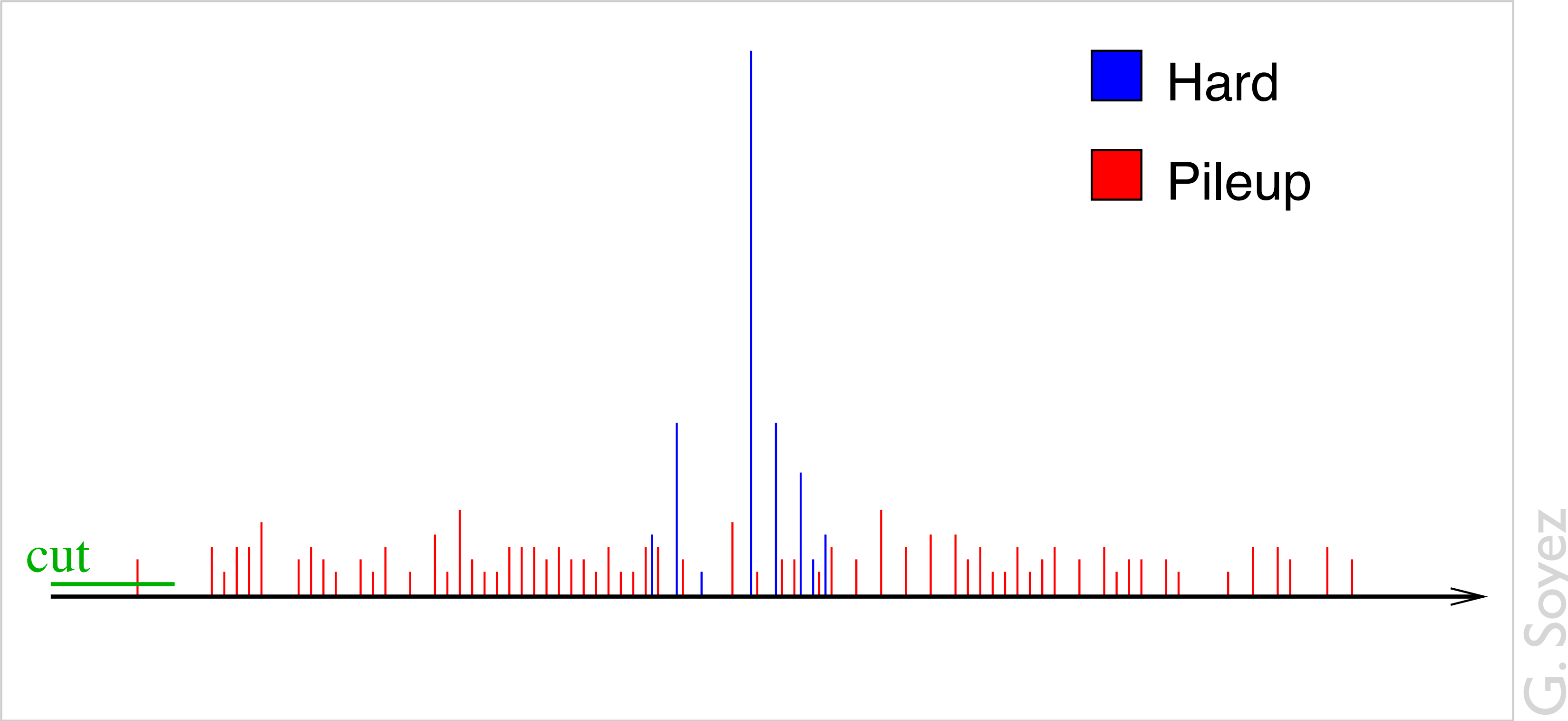
Use subjets instead



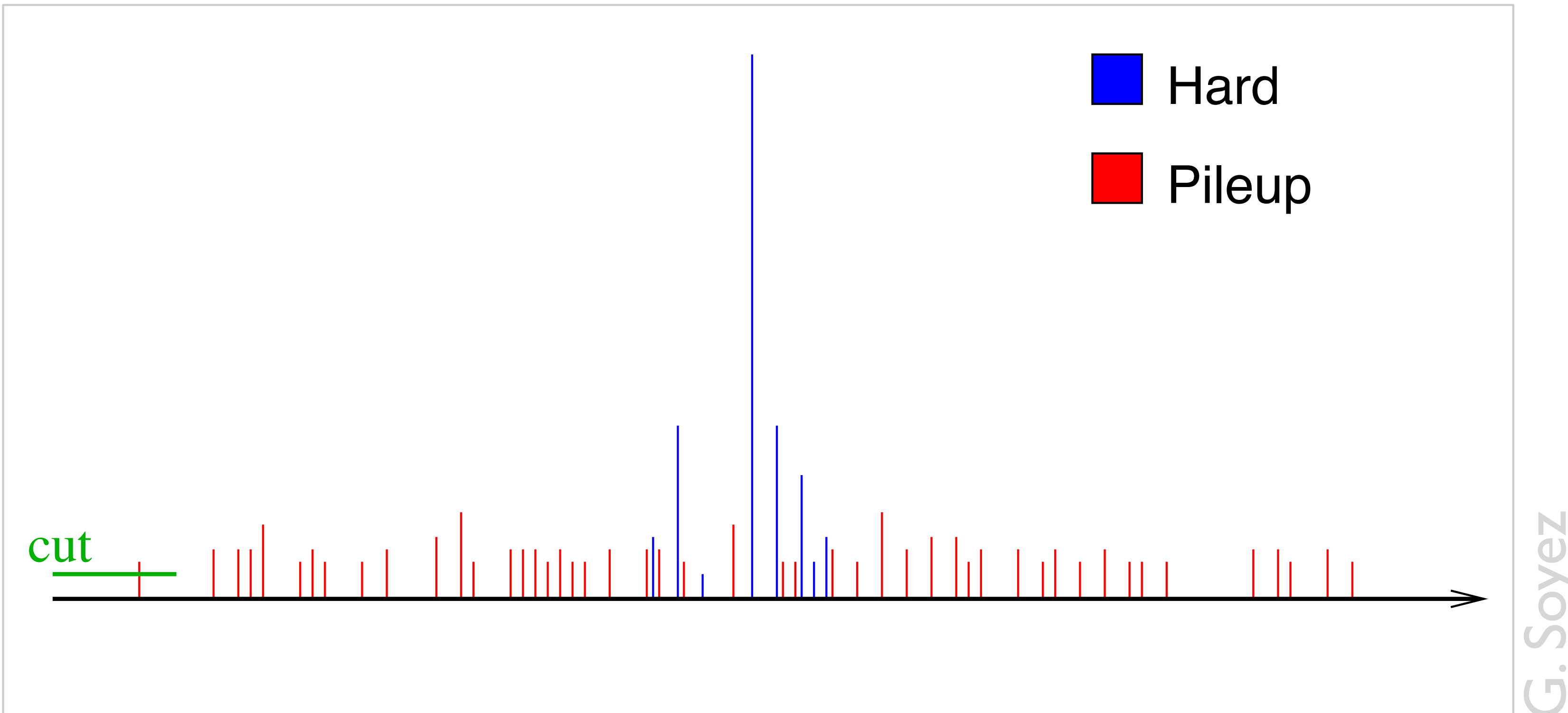
Recover local unbiasedeness by rebalancing negative energy flow into neighbouring areas

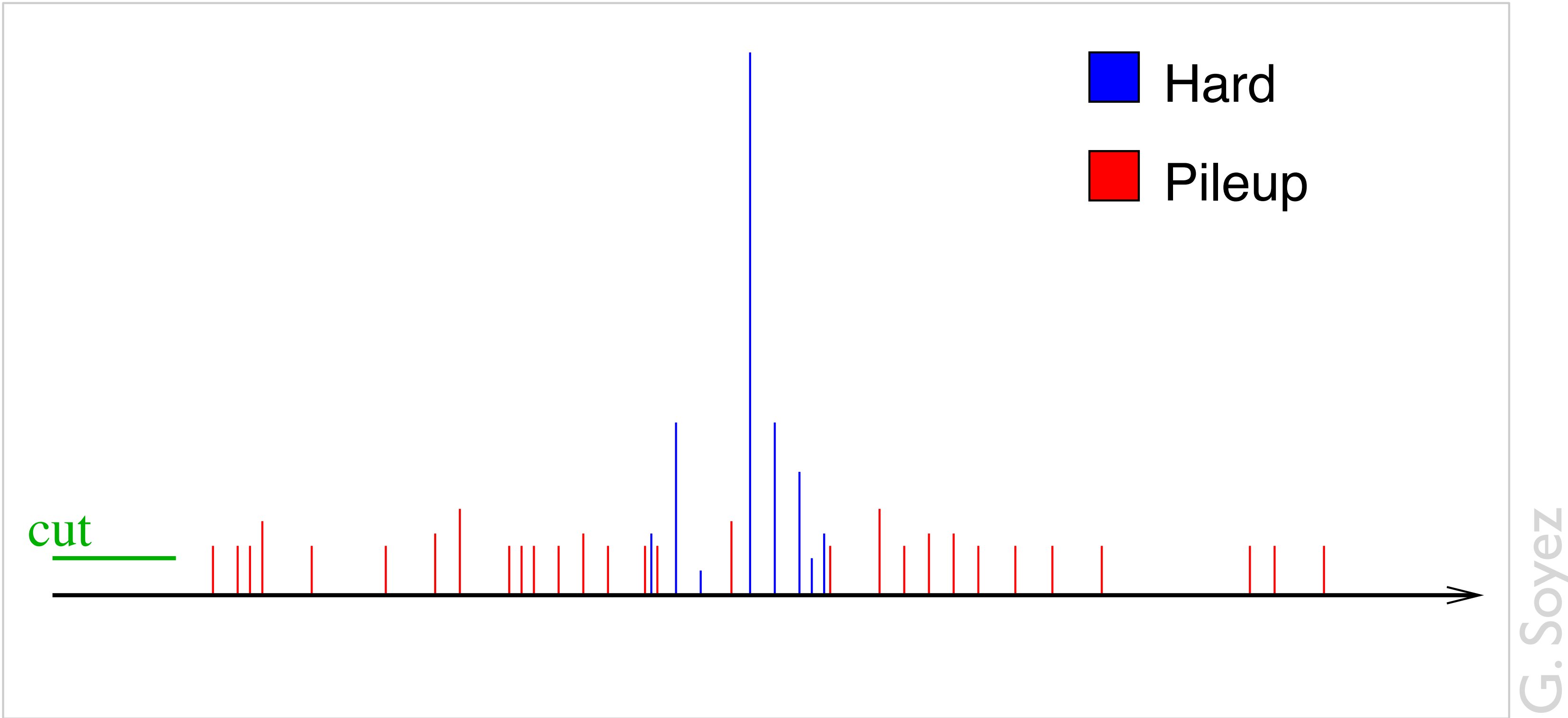
SoftKiller

Act directly on particles rather than subjets

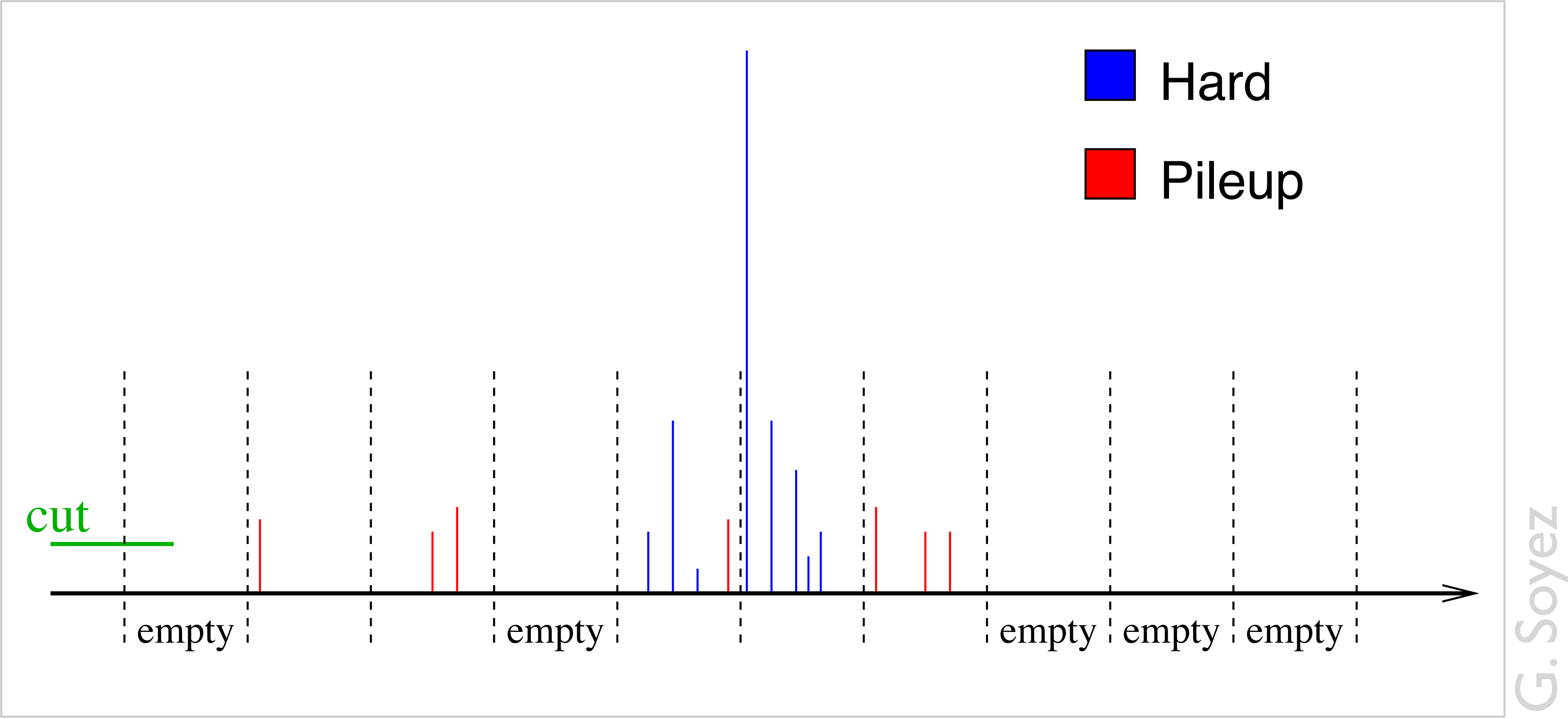


Progressively remove softest particles from event, **until ρ = 0**


Hard
Pileup
cut
G. Soyez


Hard
Pileup
cut
G. Soyez

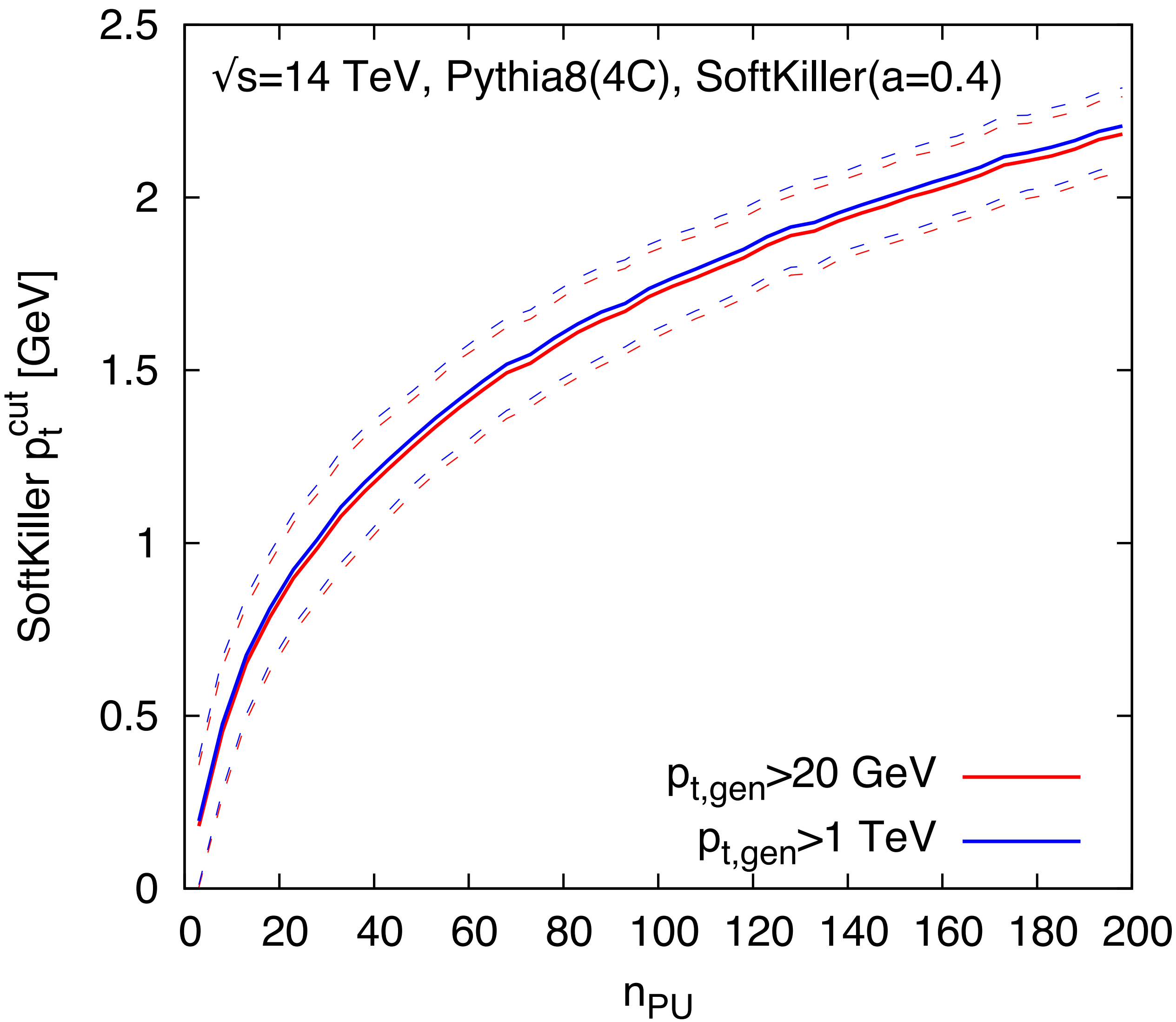
$$p_t^{\rm cut} = \underset{i \in \rm patches}{\rm median} \{p_{ti}^{\max}\}$$



Half of the event is empty ⇒ **ρ = 0** (because it’s the median)

NB. SK needs tuning of the size of the patches used to calculate ρ.
0.4 was found to be a good choice for R=0.4 jets

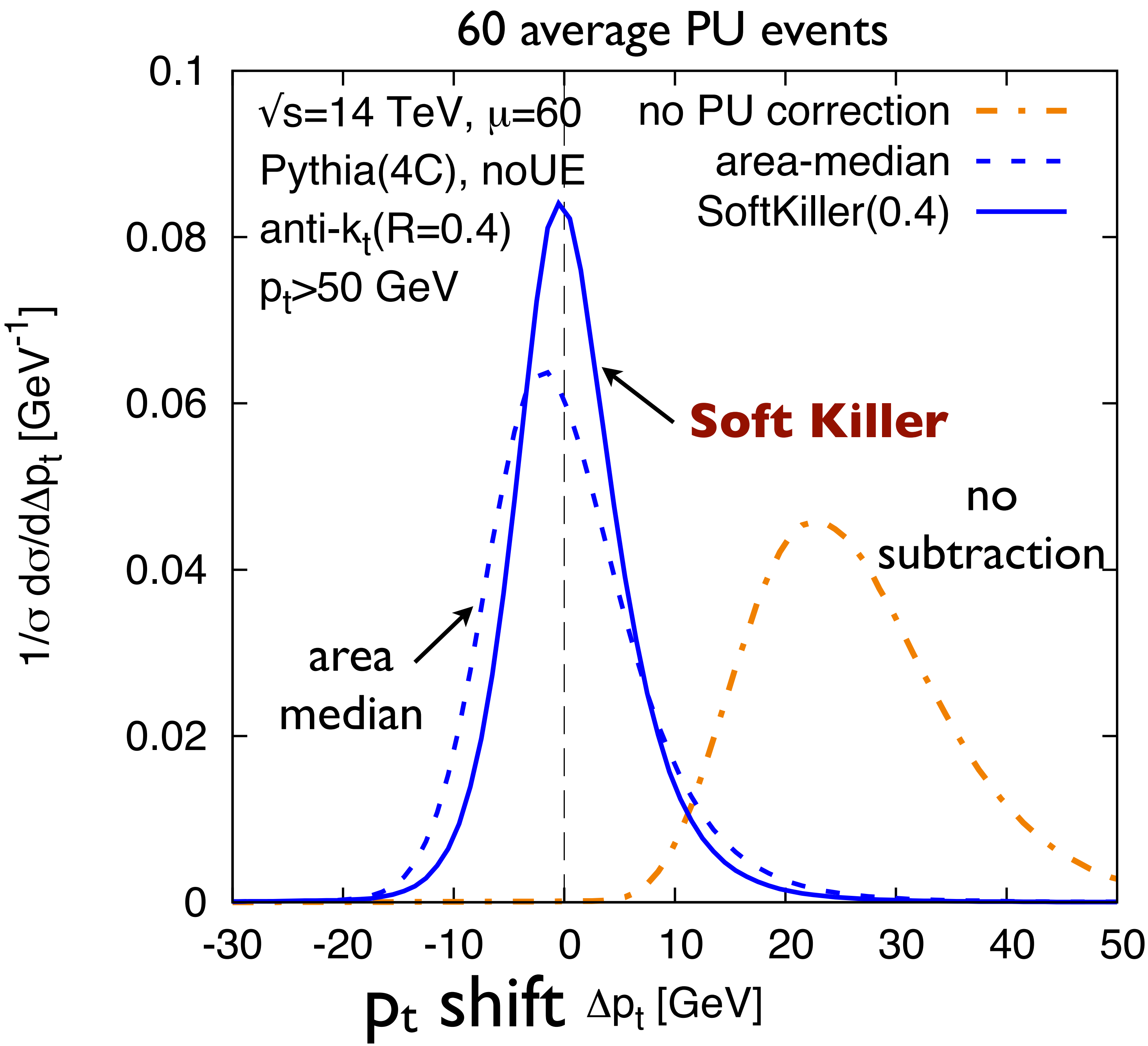
SoftKiller performance: p_t cut



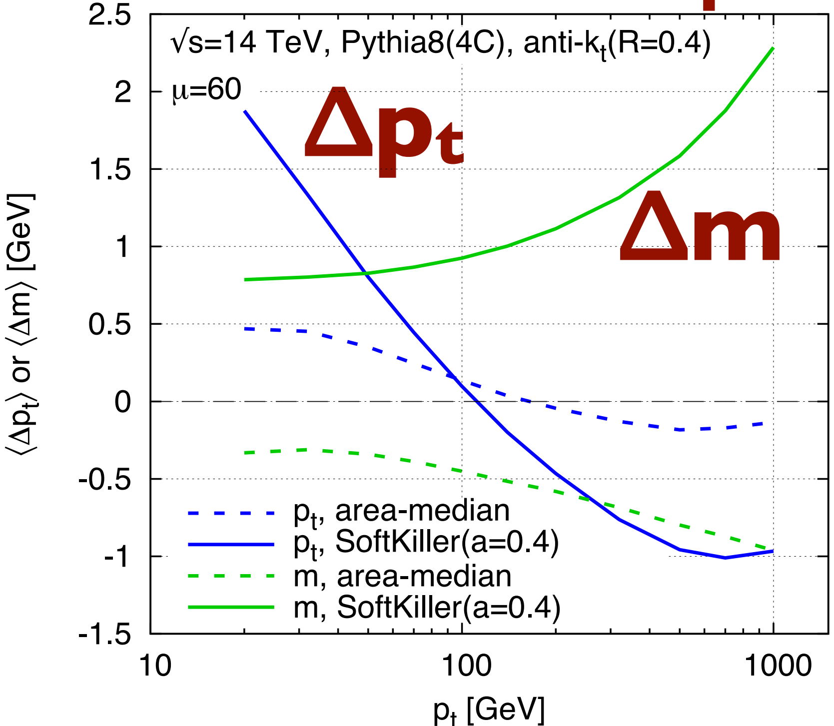
The dynamically determined p_t^{cut} grows with n_{PU}, and is stable with respect to the jet p_t

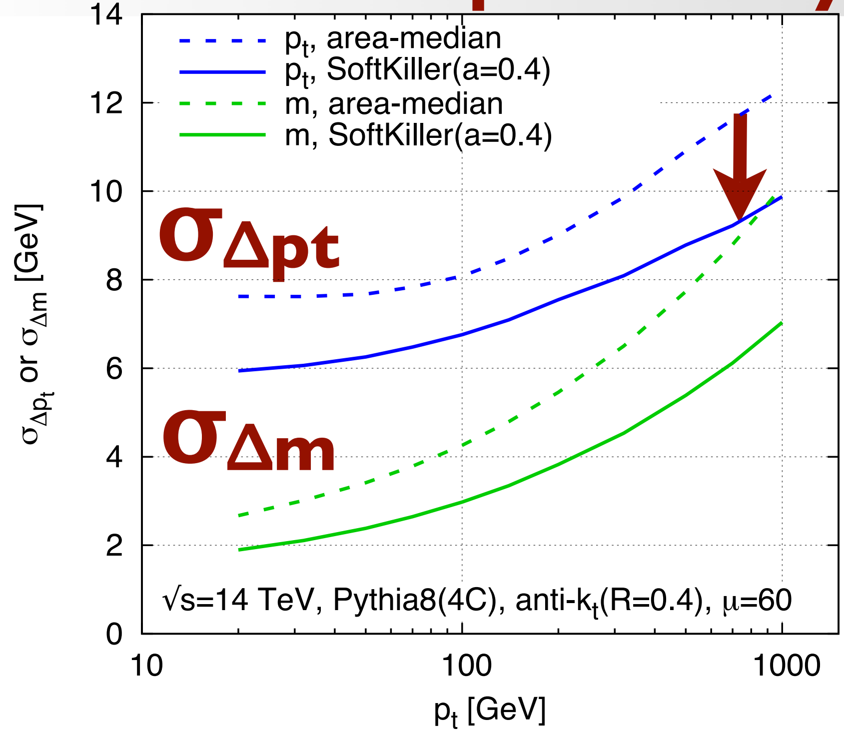
(One could parametrize this cut as a function of n_{PU} and still achieve reasonable performance)

SoftKiller performance: p_t correction



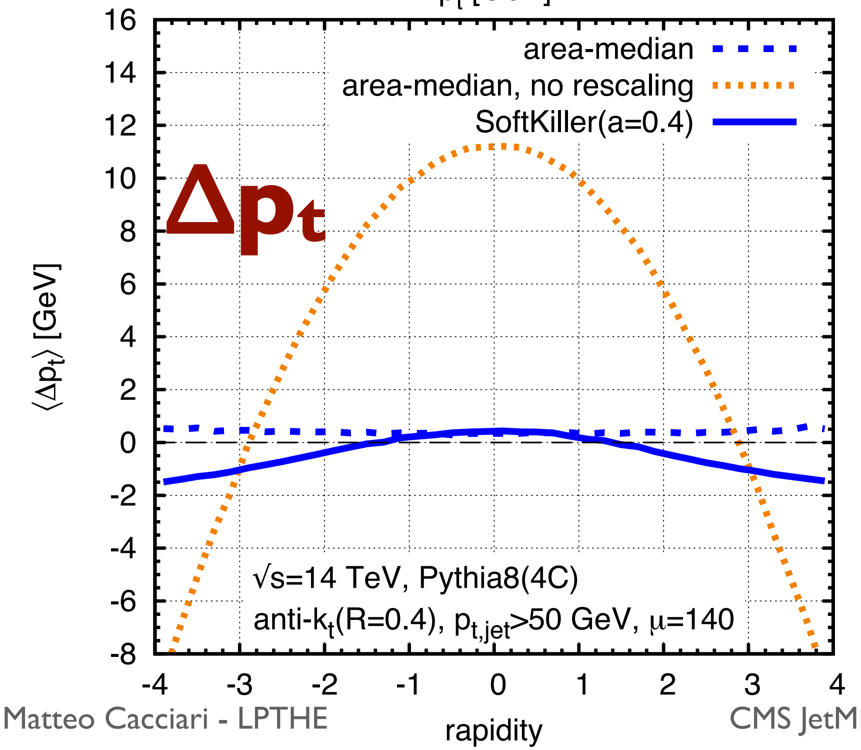
SoftKiller performance: p_t and y dep.

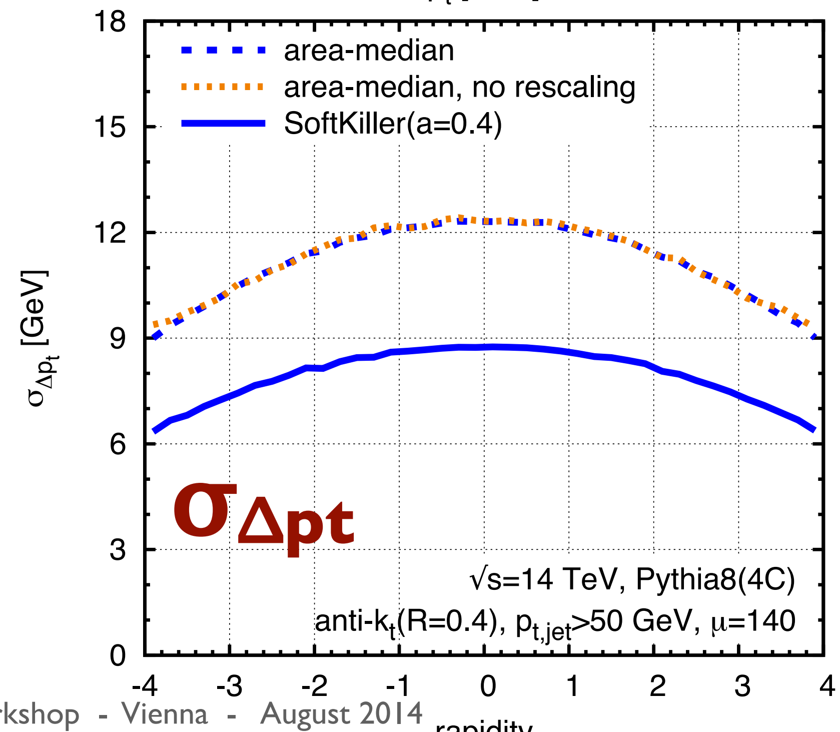




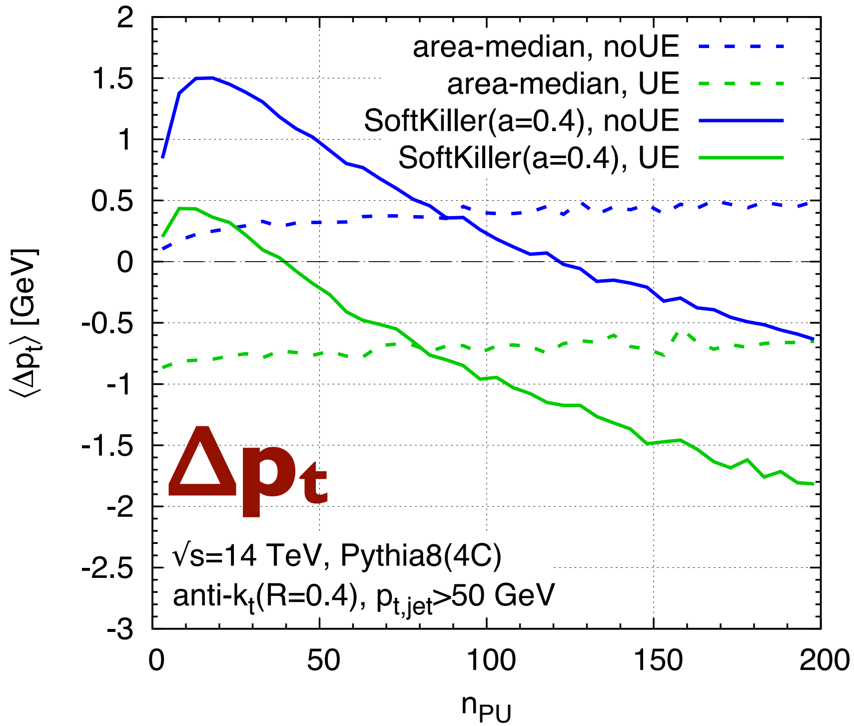
area-median

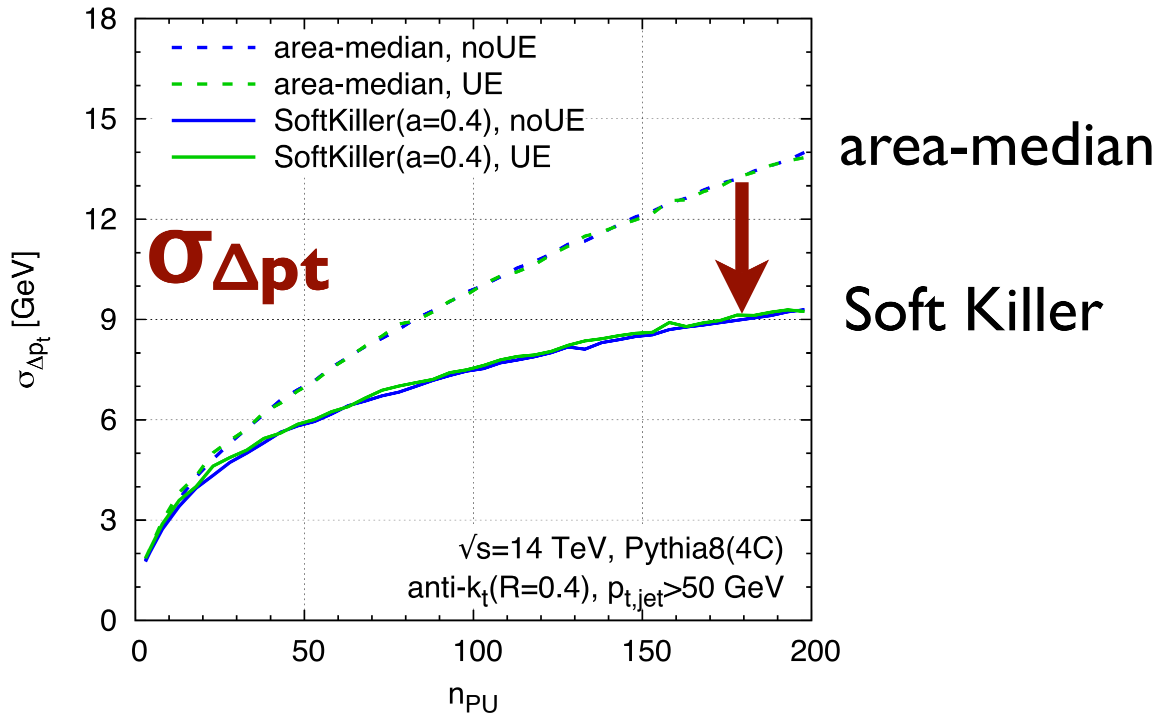
Soft Killer





SoftKiller performance: n_{PU} dep.

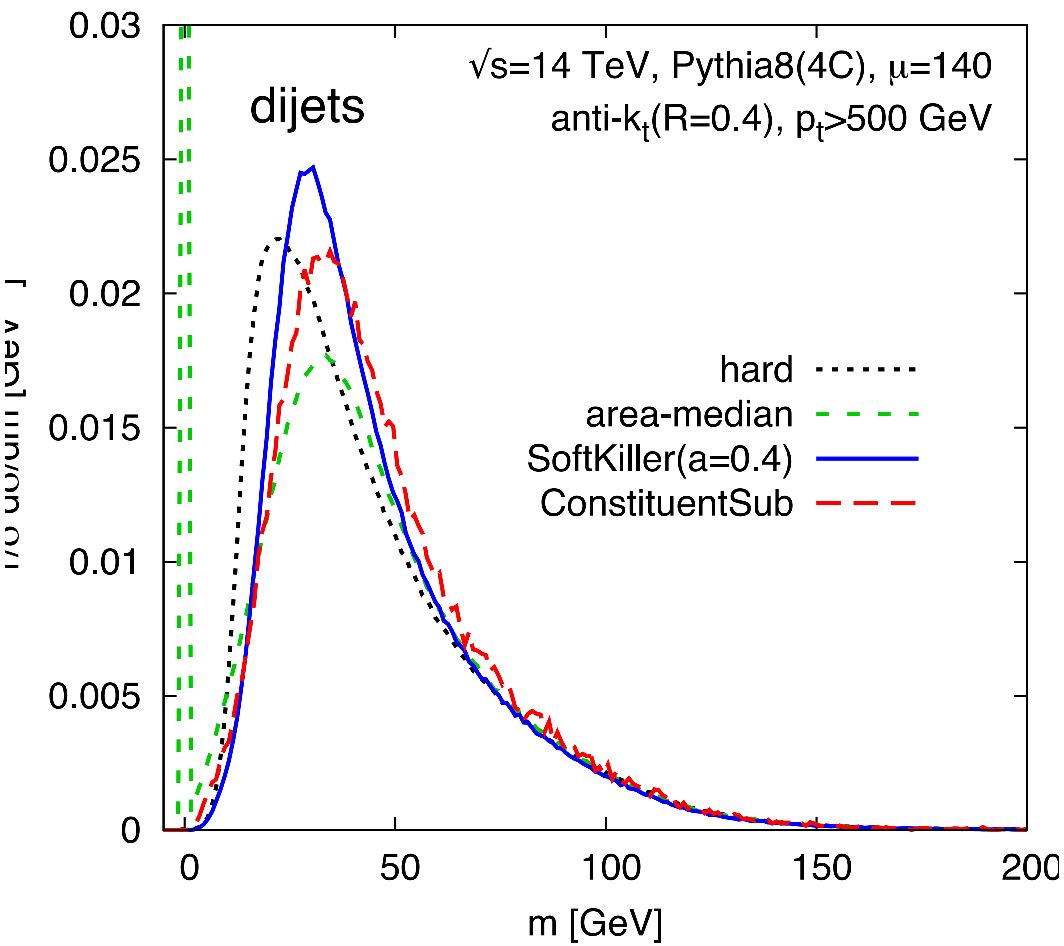




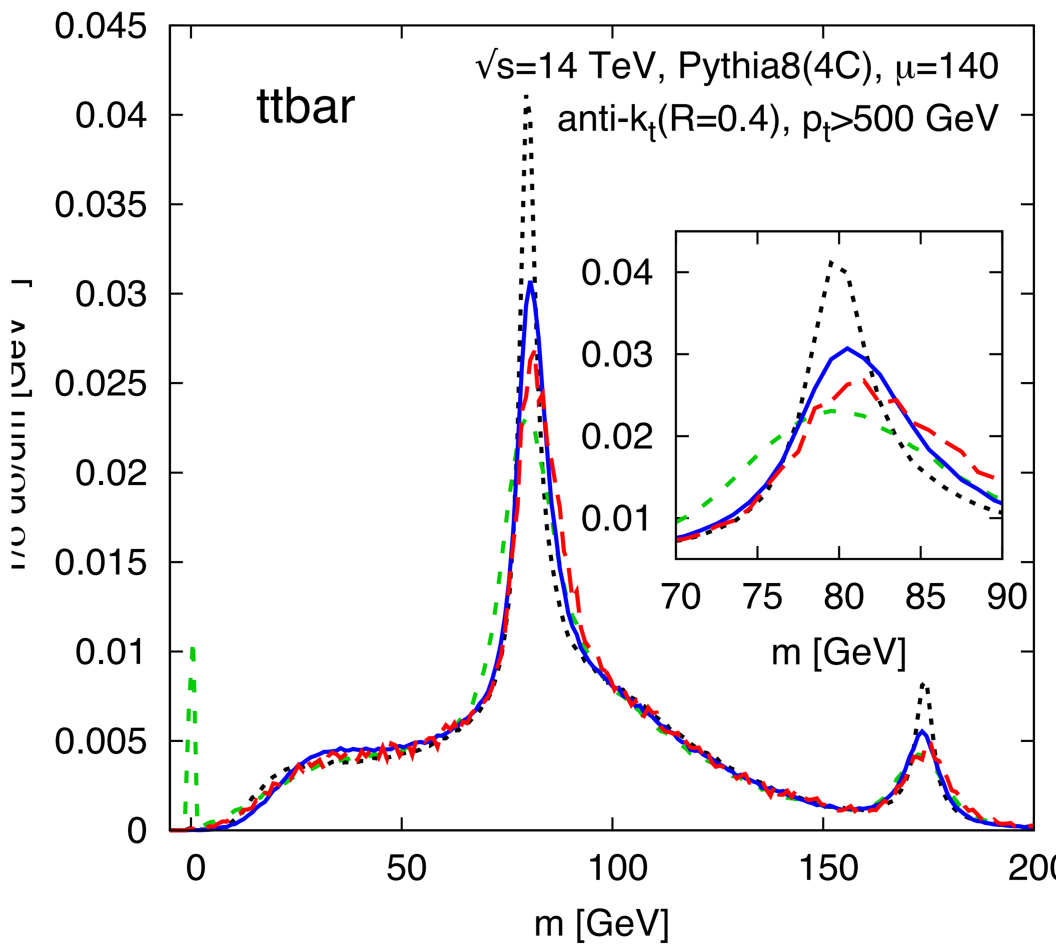
Bias under control (i.e. <4% for p_t = 50 GeV) up to n_{PU}=200, dispersion reduced (beating $\sqrt{n_{PU}}$ scaling)

Soft Killer performance

Jet mass distributions, 140 average PU events, R=0.4

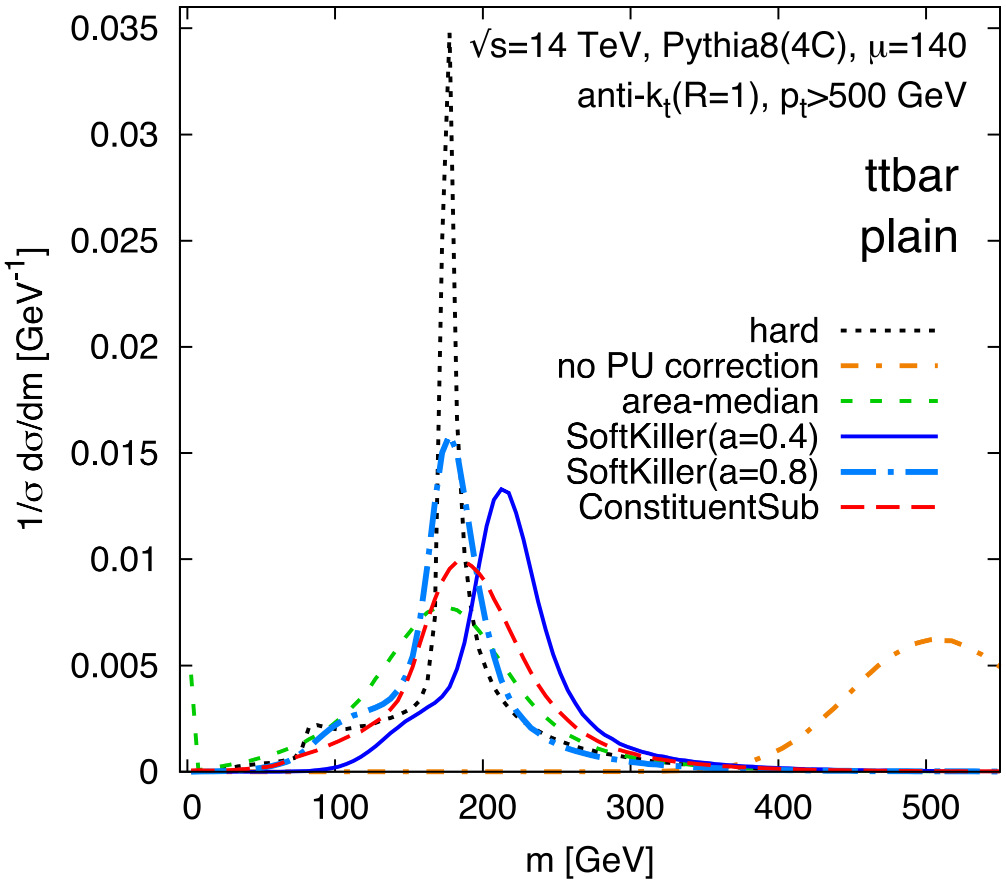


dijet events

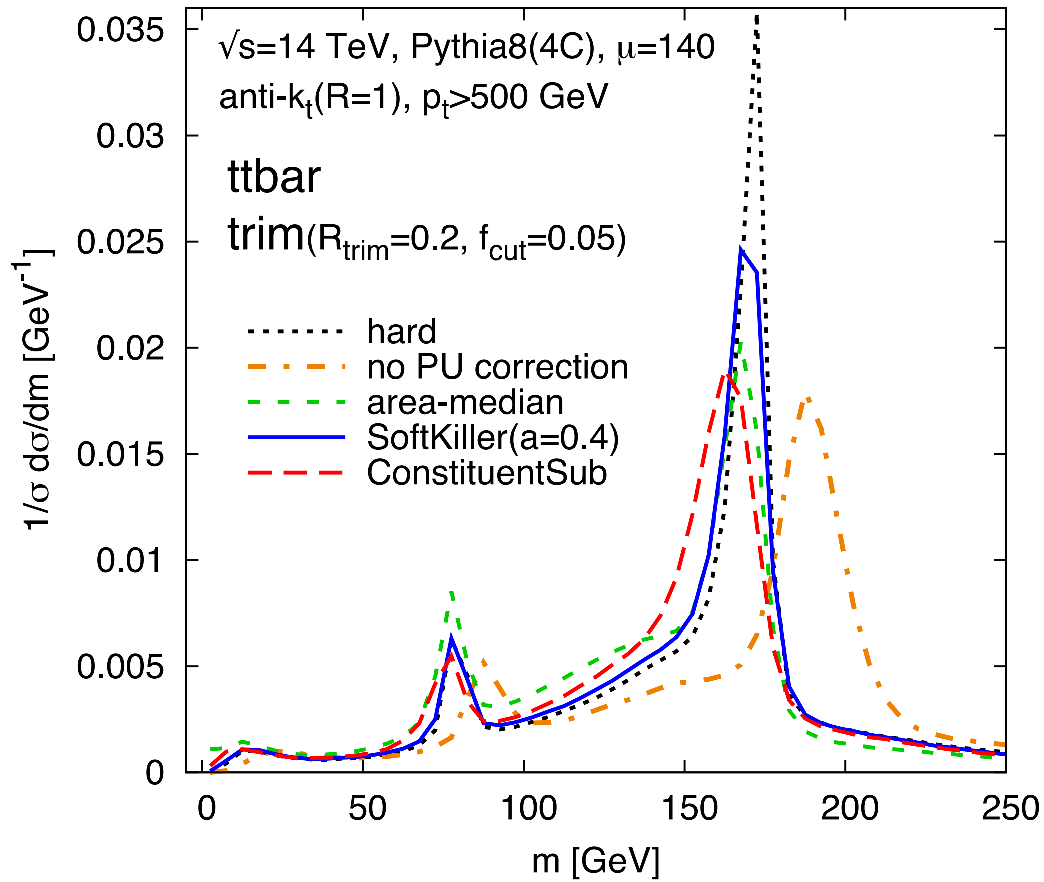


ttbar events

SoftKiller with trimming in ttbar events, R=1

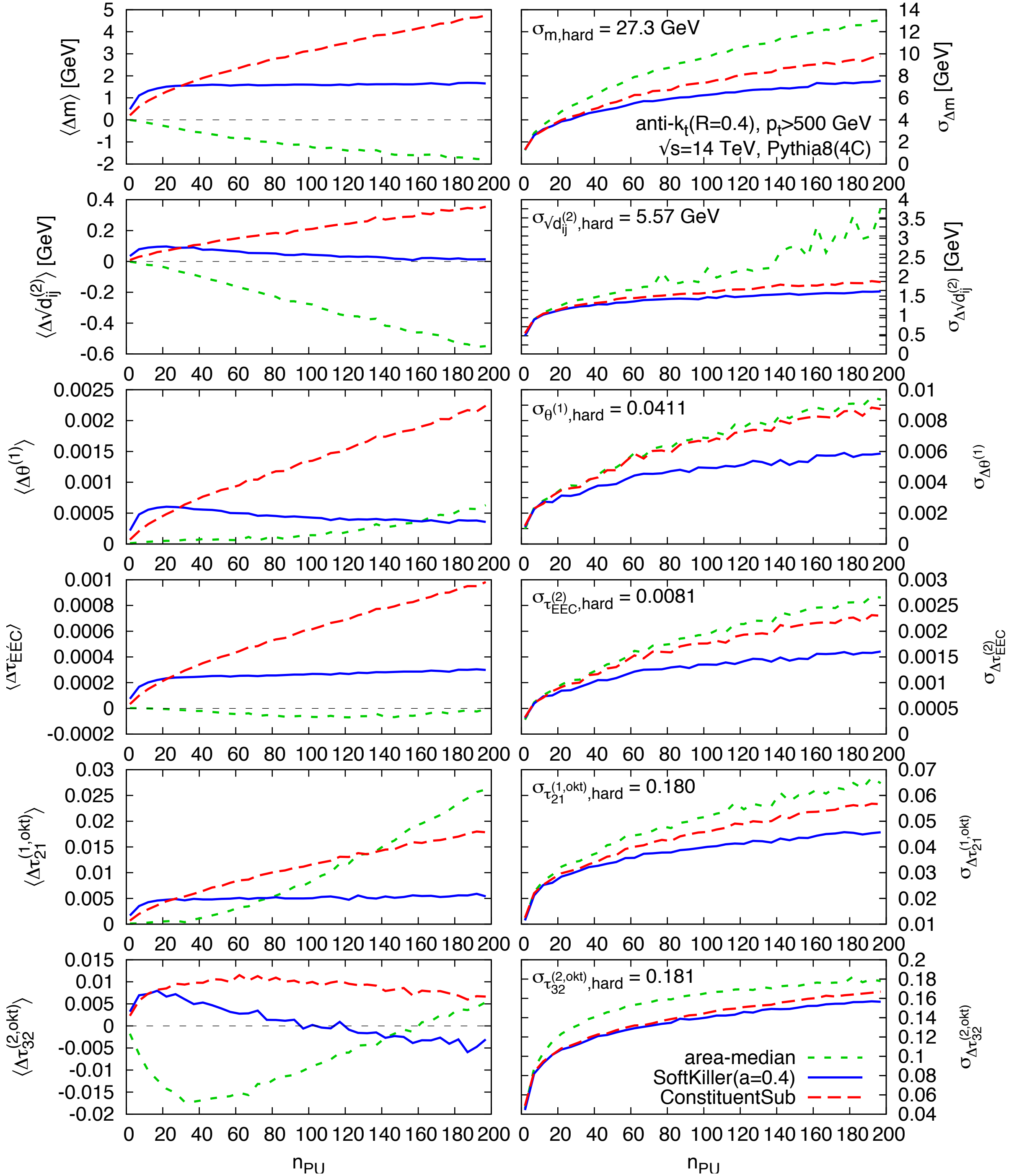


No trimming



Trimming (R_{sub}=0.2, f=0.05)

Soft Killer performance



Many jet shapes:

(pt>500 GeV, anti-kt R=0.4)

- jet mass
- kt clustering scale
- jet width (= broadening, = girth)
- energy-energy correlation moment
- τ_{21} and τ_{32} N-subjettiness ratios

- Biases under control
- Dispersions smaller than with other methods

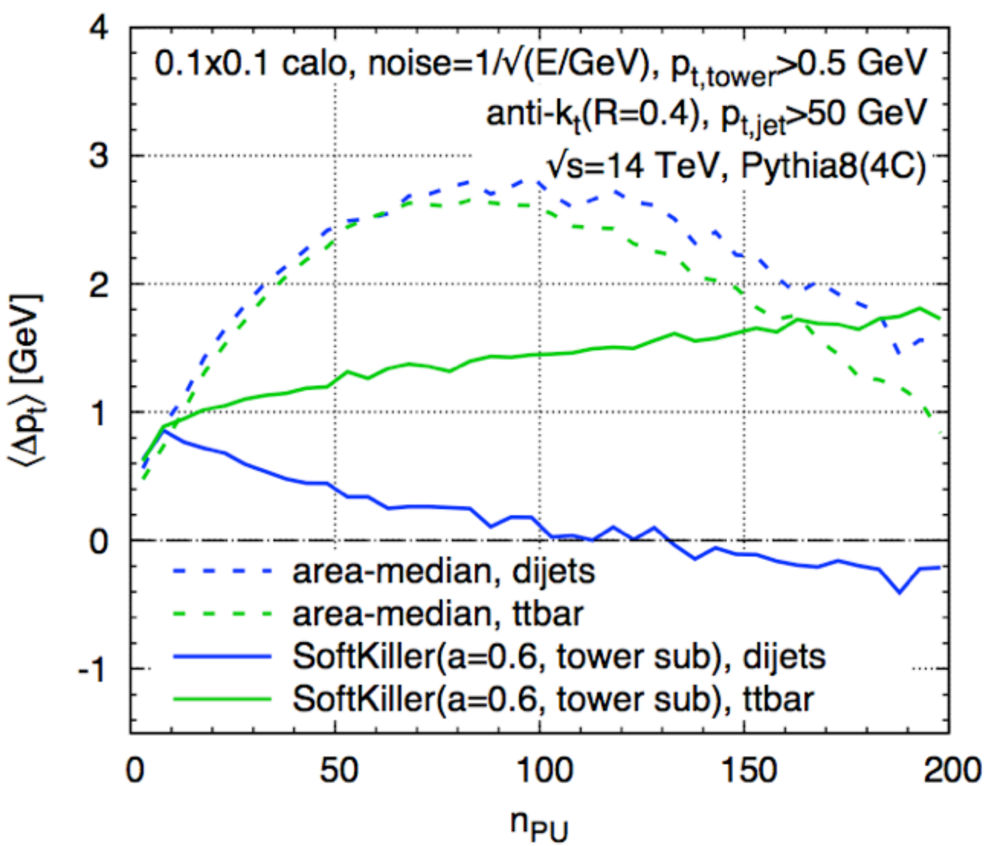
SoftKiller with a calorimeter

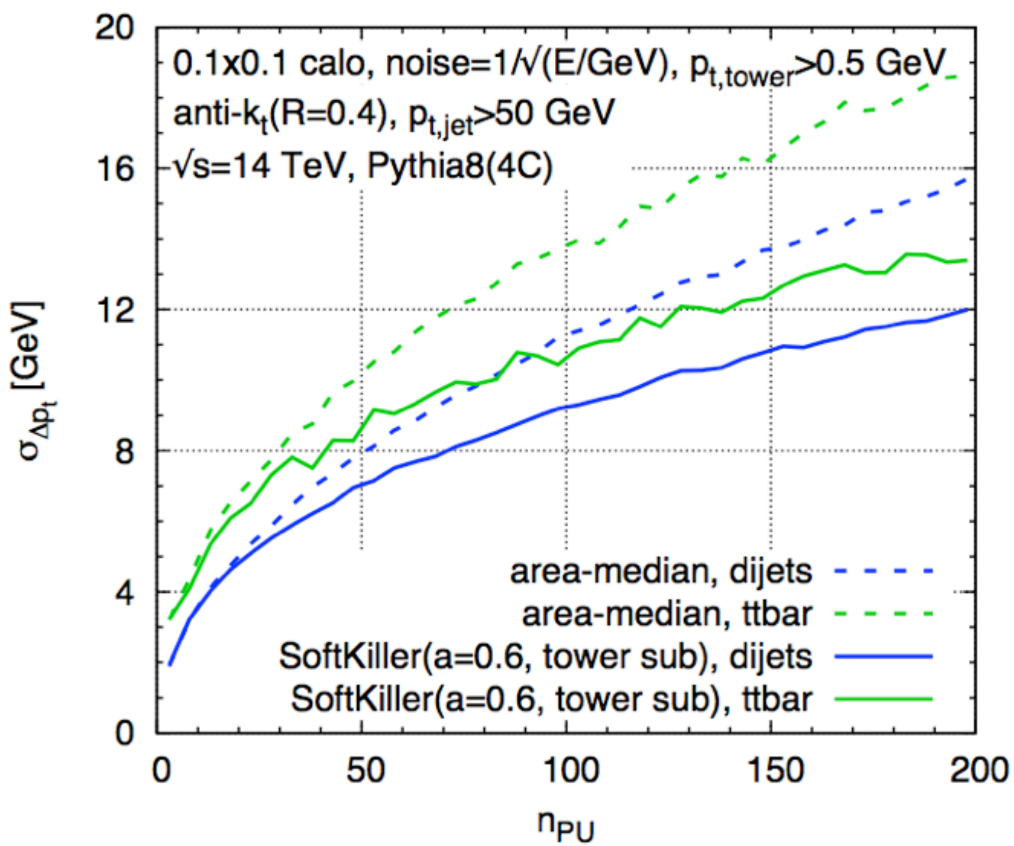
Two steps:

towers area-median subtraction, followed by soft killing

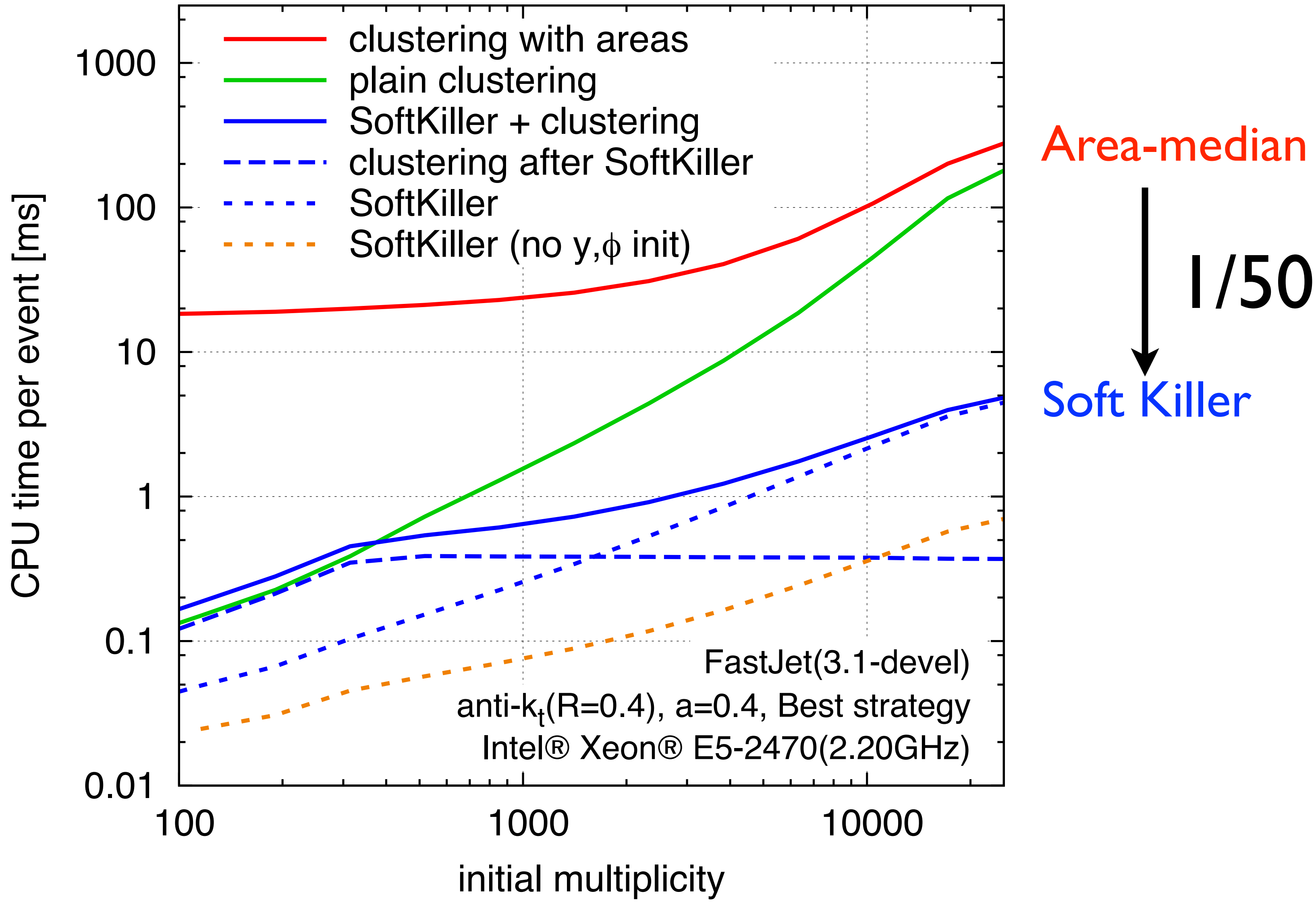
$$p_t^{\text{tower,sub}} = \max\left(0,\ p_t^{\text{tower}} - \rho A^{\text{tower}}\right)$$

$$p_t^{\text{cut,sub}} = \underset{i \in \text{patches}}{\text{median}} \left\{ p_{ti}^{\text{tower,sub, max}} \right\}$$





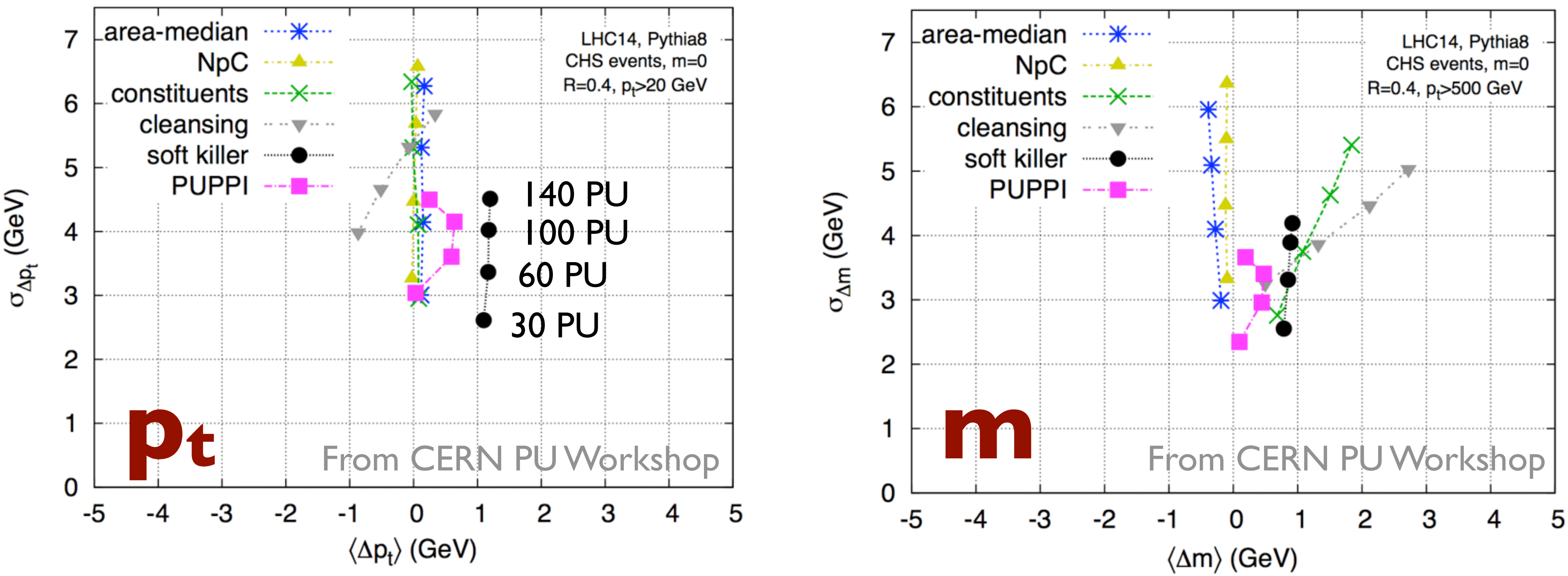
SoftKiller speed



SK is very fast (no clustering is involved).

Almost two orders of magnitude faster that area-median.

Fast enough for a trigger?



- Subjet/particle-based background subtraction methods tend to perform better in terms of dispersion than full jet-based ones
  - can be made reasonably unbiased and robust
  - can be fast
  - allow one to calculate any observable
- Many tools are already public and available in FastJet Contrib

Final considerations

- Many tools have become available in the past few years: taggers, groomers, 'subtractors'

- 'Power use' usually implies a combination of them
(eg. to tag efficiently one needs to groom, radiation-based taggers are affected by pileup that needs to be subtracted, etc)

- Given the variety and complexity of the tools, I'll never emphasize sufficiently the need to use public and validate codes to avoid ambiguities