Jets' Substructure

- Introduction
- Reminder: Jet algorithms
- Substructure Observables
- Pileup and its removal
- Results
- Coffee

Introduction

Jet algorithms collect groups of particles based on a common property usually the direction of motion or small invariant mass w.r.t. a given direction.
At high boosts the decay products of a heavy object, e.g. Z, W, H or t may be contained in a single jet.
One would like to tell such jets from ordinary high p_T QCD ones.
Looking for jet substructure: searching for differences between particles that have been classified as similar. --> sort of contradiction

<> Result is prone to be jet-algorithm dependent

<> Observables must be defined with some care (mass, planarity,...

<> Correct for detector effects

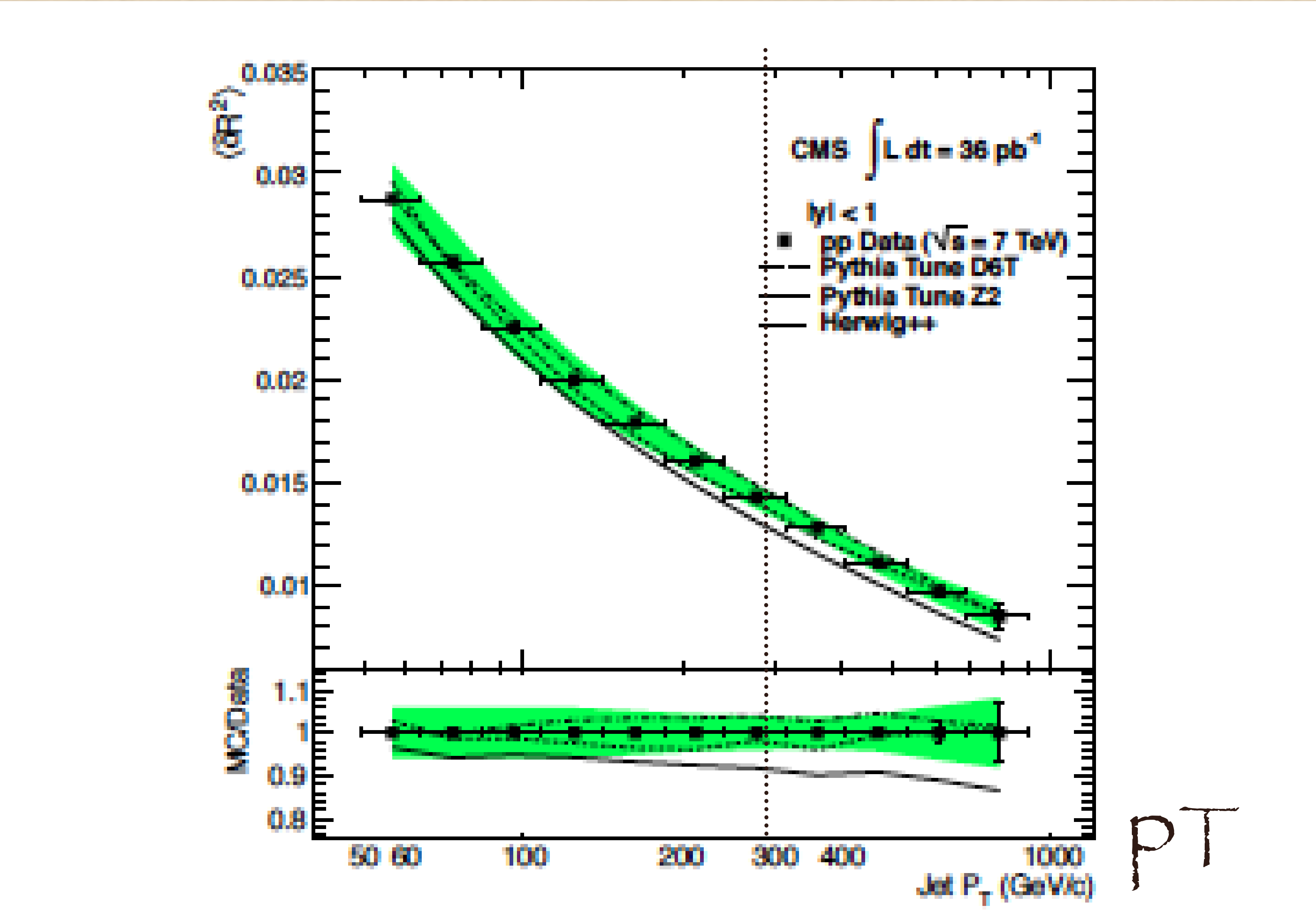
Substructure is studies for highly boosted massive jets ($E/m>2$)

RECYCLE
COMPOST
COMPOSTABLES
PAPER ONLY

Effect of Boost

Jet “width” vs. its momentum

$< \delta R^2 >$



E/M>2

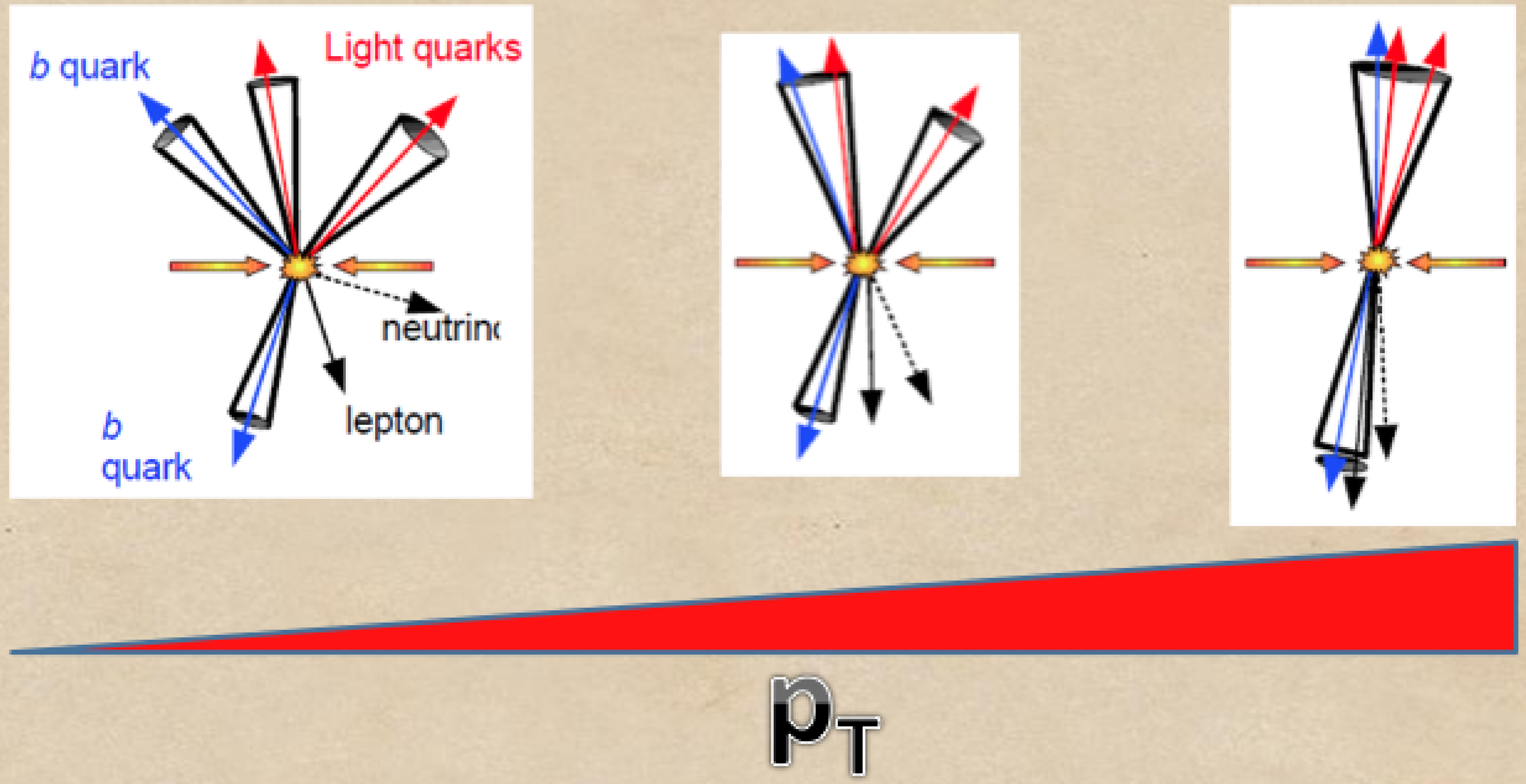
pT

The higher the jets’ pT the narrower the jet is.

Implication: Creation of Massive Jets

Low pT

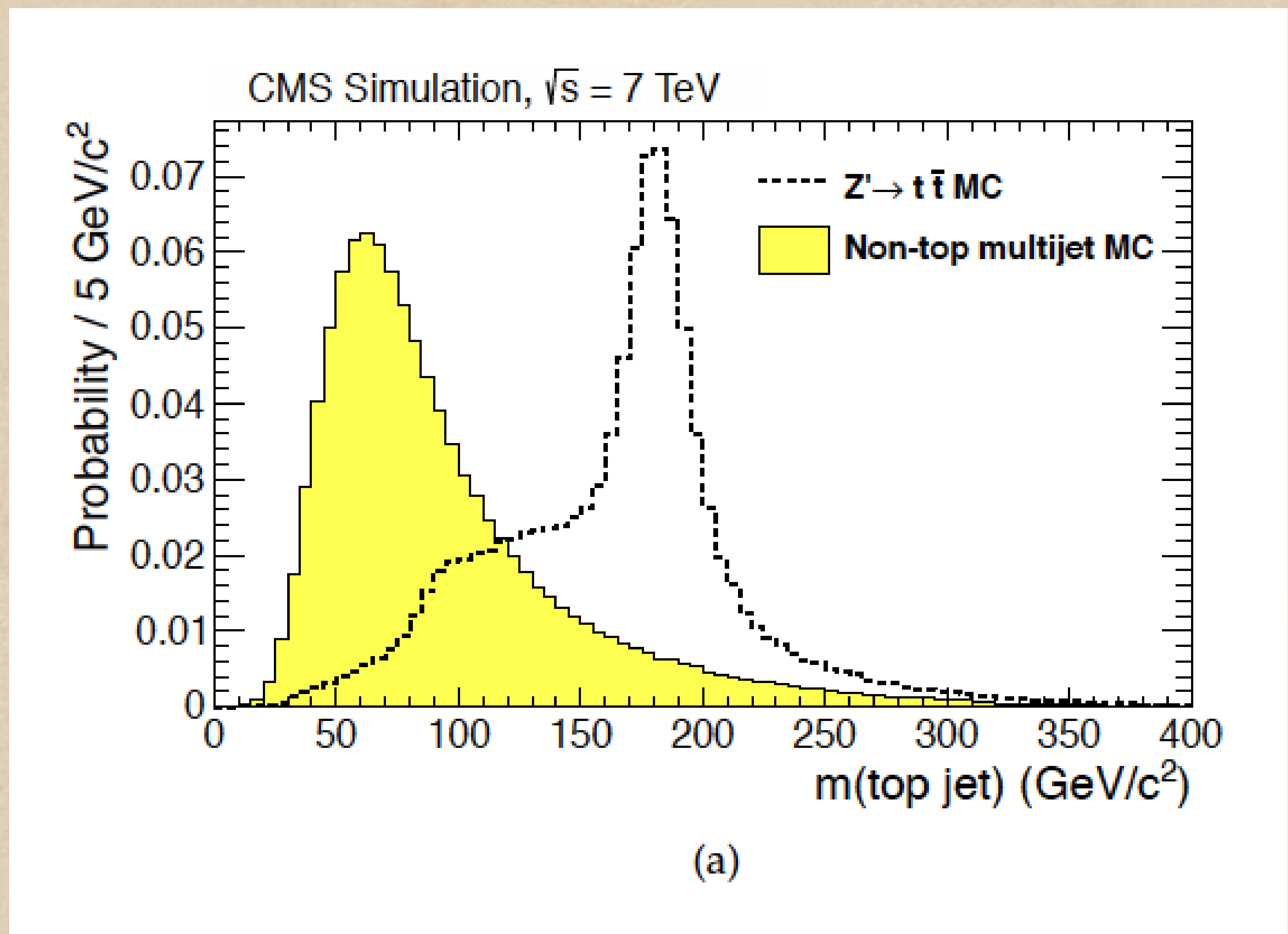
High pT



Highly boosted hadronic Top quark may be fully contained by a single jet

Boosted Top (Simulation)

Top signal is largely enhanced

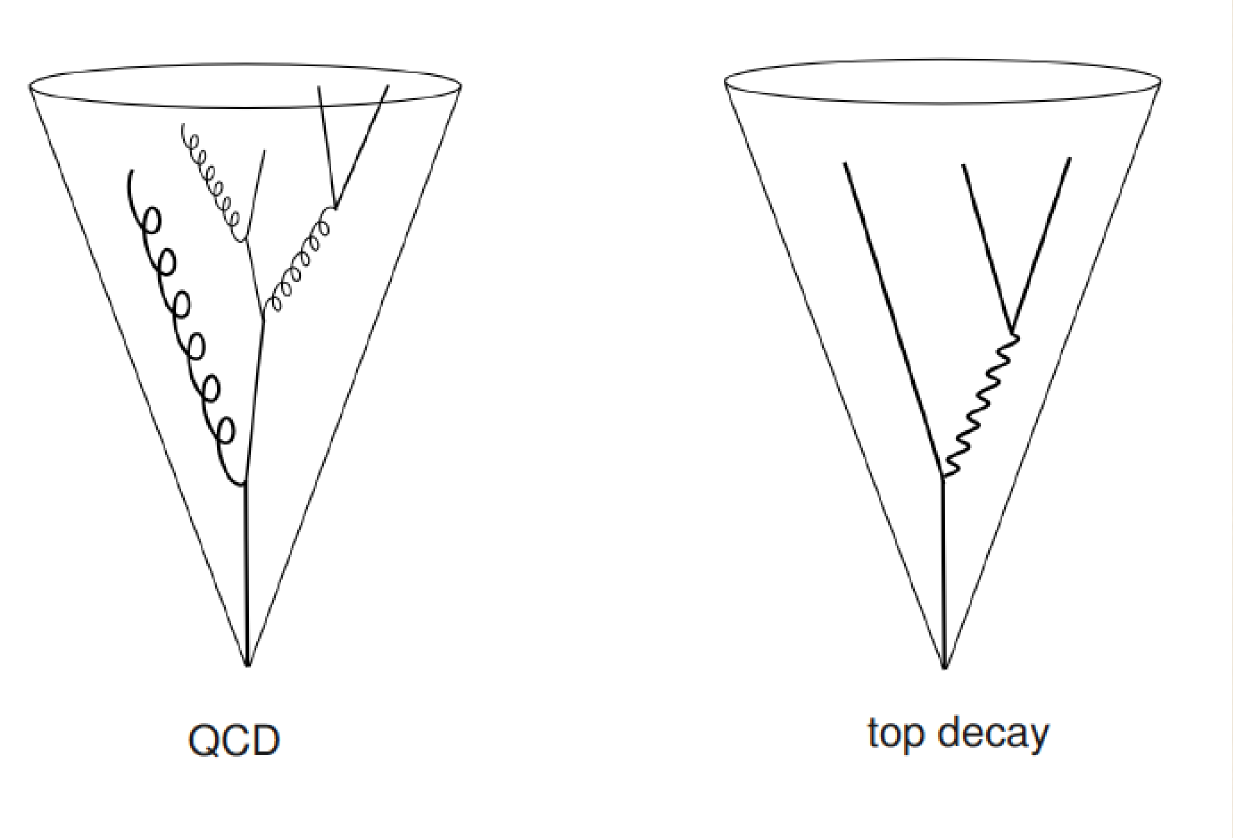


(a)

Not all is lost! Boosted top jets appear as massive jets

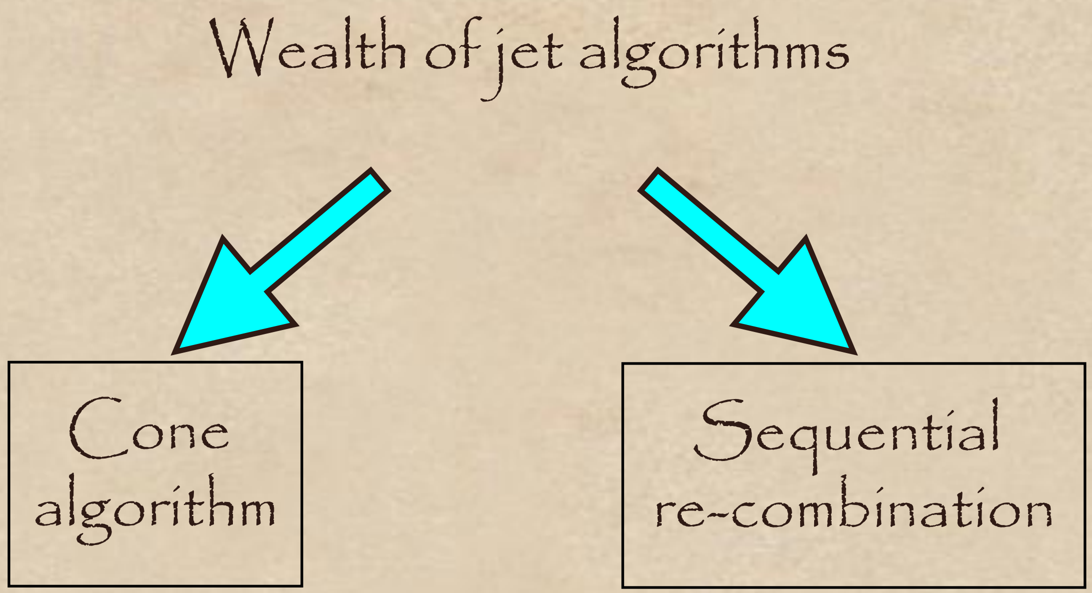
Motivation

- New physics must be characterized by high mass/energy
- Boosted jets are a very likely outcome of BSM physics
- Decayed objects like top, W, Z and Higgs are likely to be contained by a single massive jet



Challenge: Tell the left configuration from the right one

Reminder: Jet Algorithms

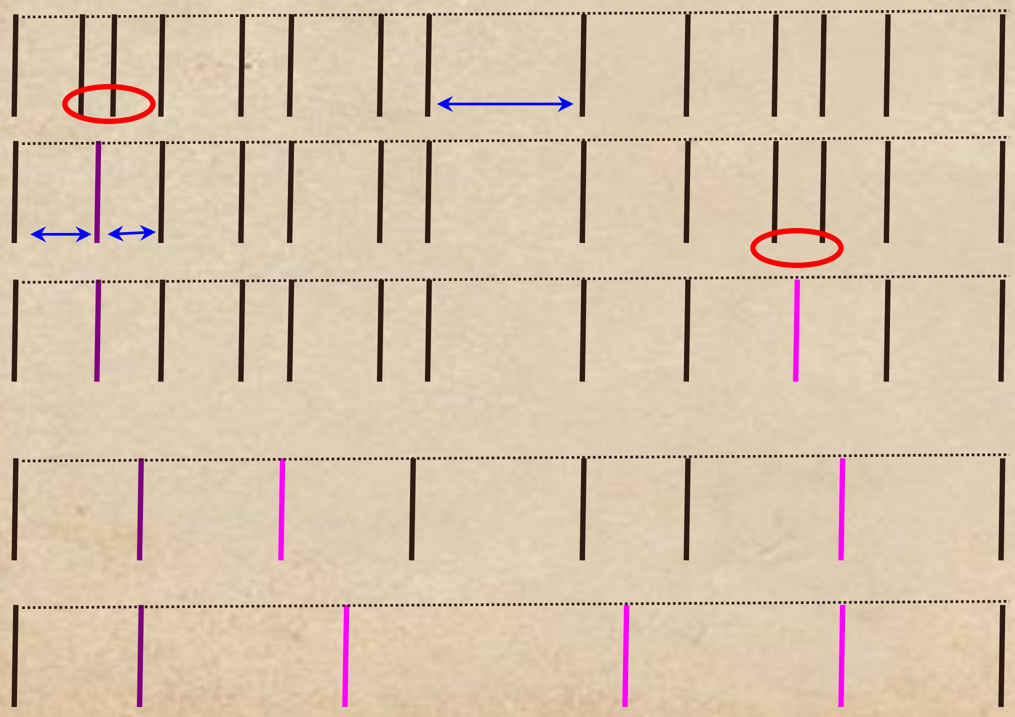


Wealth of jet algorithms

Cone algorithm

Sequential re-combination

Infrared and collinear safety	bad	Good



	p
Kt	1
Cambridge/Aachen	0
Anti Kt	-1

$$d_{ij}=\min\{p_i^{2p},p_j^{2p})(1-\cos\theta_{ij})$$

Observables of Jets' Substructure (incomplete list)

- Energy spread
  - Width
  - Mass
  - Angularity
  - Eccentricity
- Sub-clustering
  - Planar flow
  - Sub-jettiness

Many observables, partially overlapping

Guy Gur-Ari, Michele Papucci, and Gilad Perez. arXiv: 1101.2905 [hep-ph]

Integrated and Differential Jet Shape

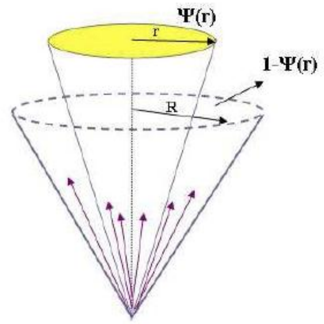
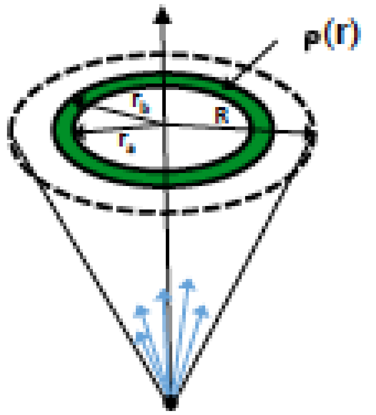


Figure 1: Definition of the integrated jet shape, $\psi(r)$.

$$\psi(r) = \frac{1}{N_{Jets}} \sum_{Jets} \frac{p_T(0,r)}{p_T(0,R)}, \quad 0 < r < R$$

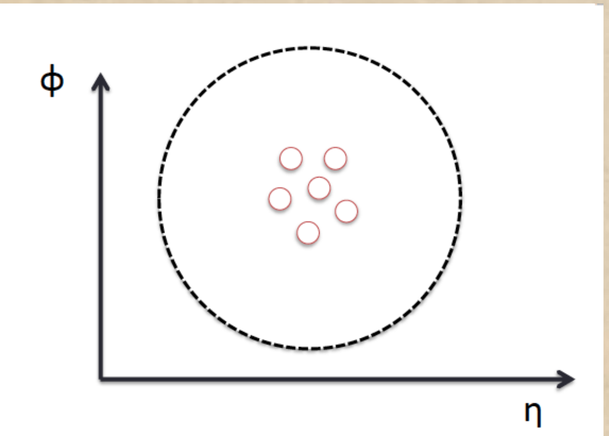


$$\rho(r) = \frac{1}{\Delta r} \frac{1}{N_{\text{jet}}} \sum_{\text{jets}} \frac{p_T(r - \Delta r/2, r + \Delta r/2)}{p_T(0,R)}, \; \Delta r/2 \le r \le R - \Delta r/2,$$

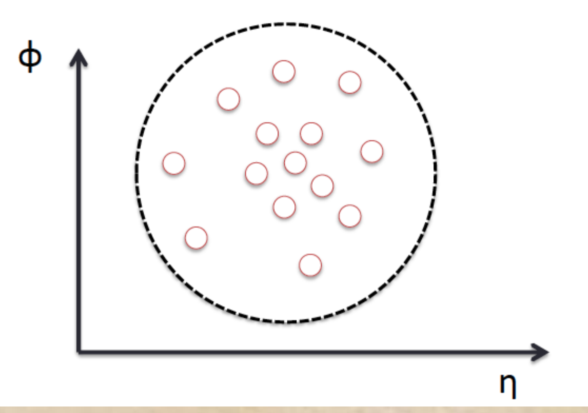
Jets' Width (Girth)

$$W = \frac{\sum_{i=1}^{n} \Delta R^{i} p_T^{i}}{\sum_{i=1}^{n} p_T^{i}}$$

$$g = \sum_{i \in Jet} \frac{p_T^{i}}{p_T^{Jet}} |r_i|$$



Light



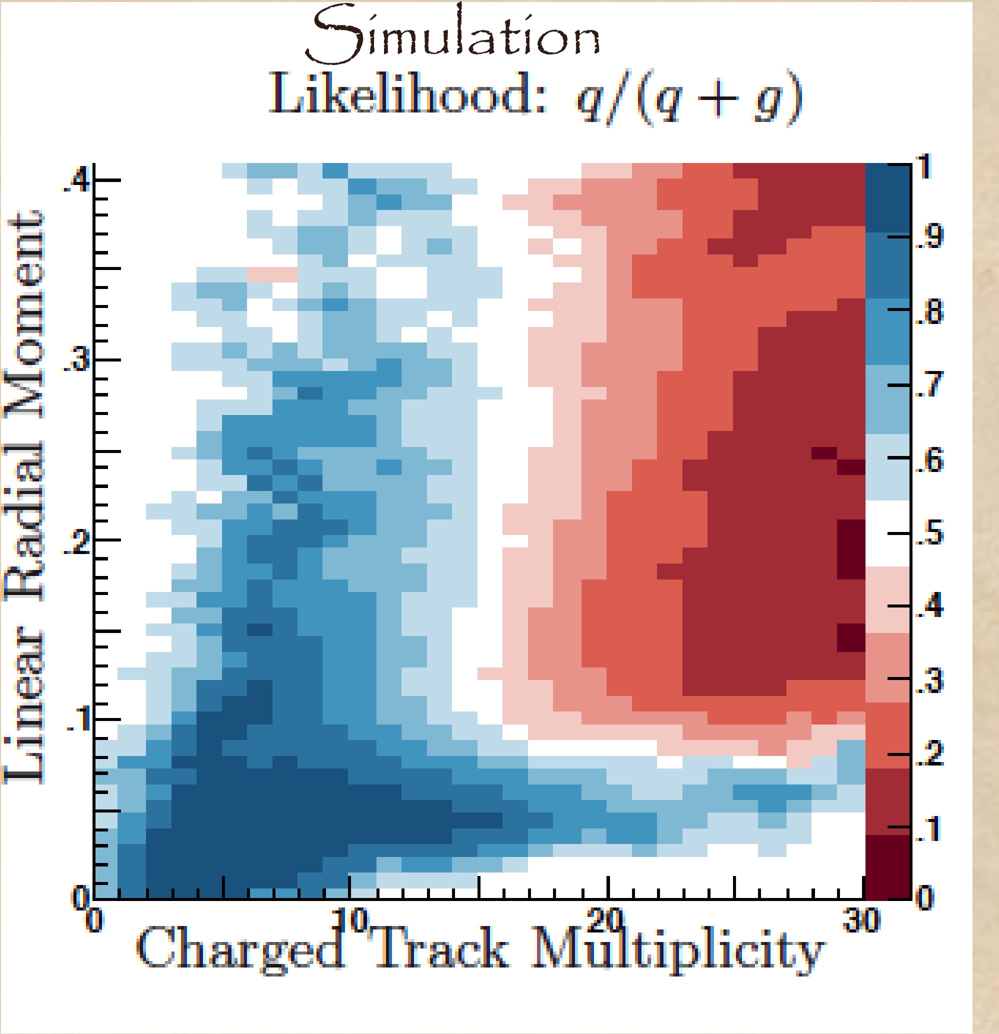
Heavy

Separating Quark Jets from Gluon Jets

Due to their color charge gluon jets are "wider" than quark jets and have higher (charged) particle multiplicity

Linear radial moments (Girth)

$C_A=3; C_F=4/3$



Jason Gallicchio, Matthew D. Schwartz arXiv:1106.3076v2

Girth(width) may help enriching jet sample with gluon/quark jets.

Angularity

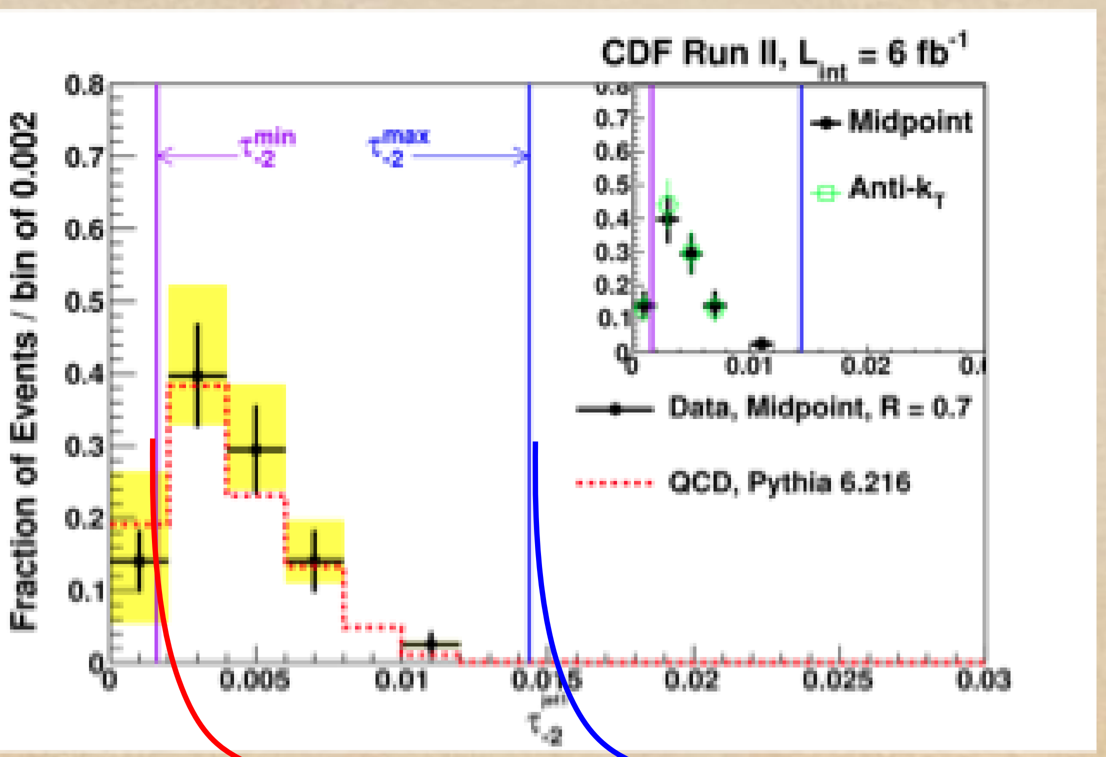
Angularity is defined by:

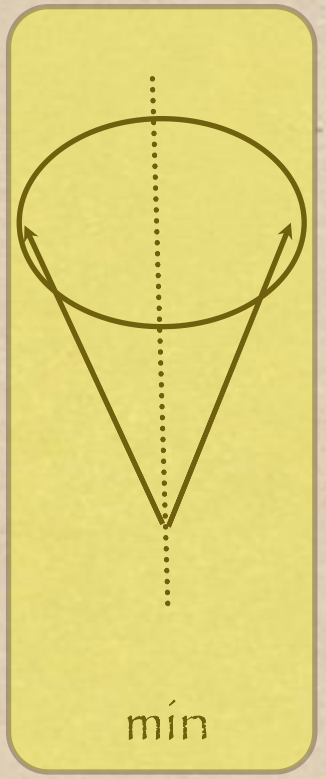
$$\tau_a(R,p_T) = \frac{1}{m_{Jet}} \sum_{i\in Jet} E_i \sin^a\theta_i\,(1-\cos\theta_i)^{1-a} \sim \frac{2^{a-1}}{m_{Jet}} \sum_{i\in Jet} E_i \theta^{2-a}$$

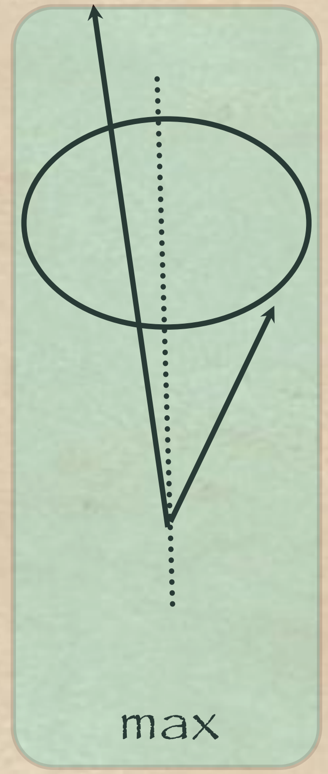
Usually the a=-2 case is used namely:

$$\tau_{-2} \approx \frac{1}{8m_{Jet}} \sum_{i\in Jet} E_i \theta_i^4$$

Angularity manifests itself by having a lower and upper limits. The lower corresponds to symmetric decay while the higher to a very asymmetric one.





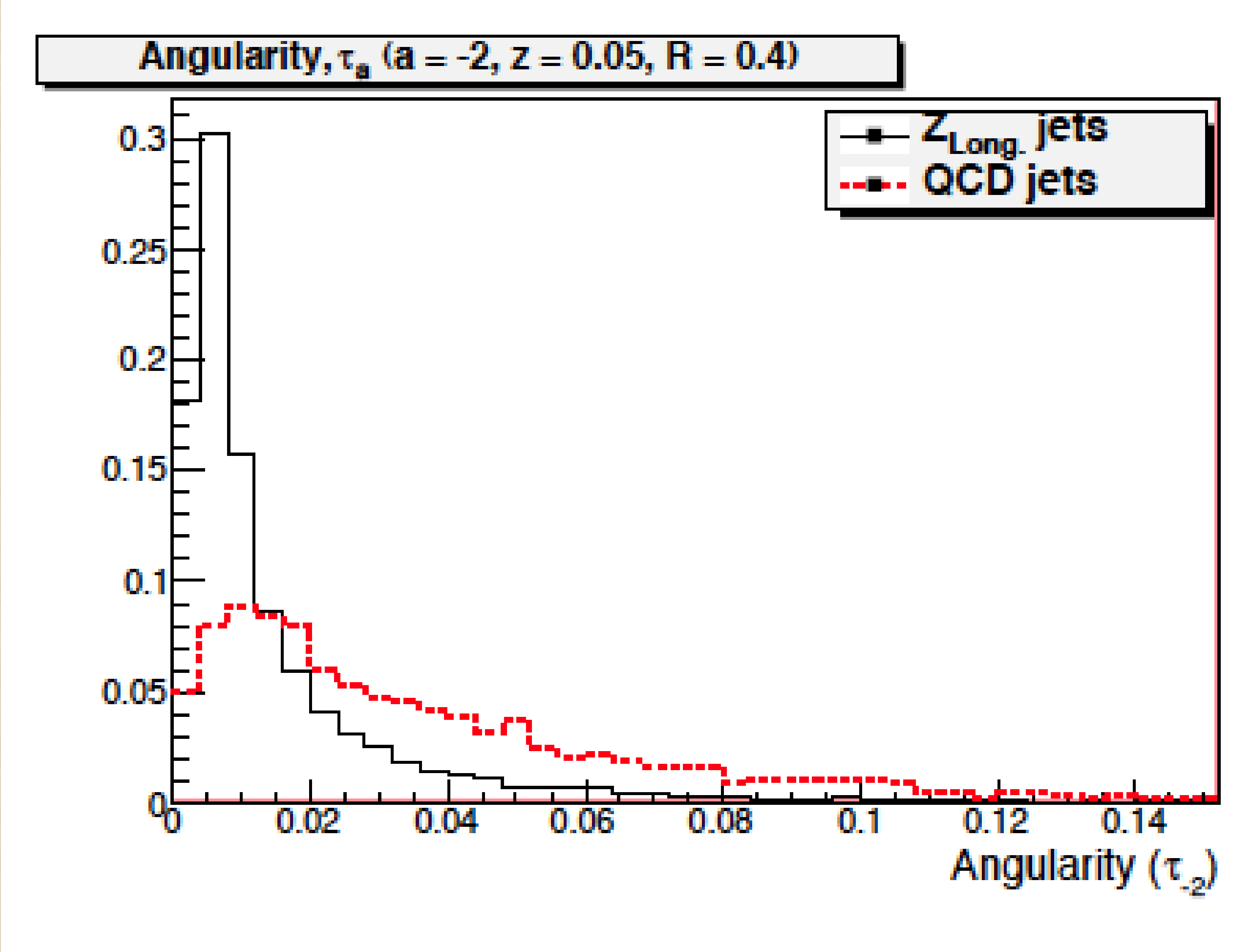


$$\tau_{-2}^{min} \approx \left(\frac{m_{Jet}}{2p_T}\right)^3$$

$$\tau_{-2}^{max} \approx \frac{2^{-3}R^2 m_{Jet}}{p_T}$$

Angularity as a Separator of QCD/Signal(s)

Angularity can tell QCD from boosted V



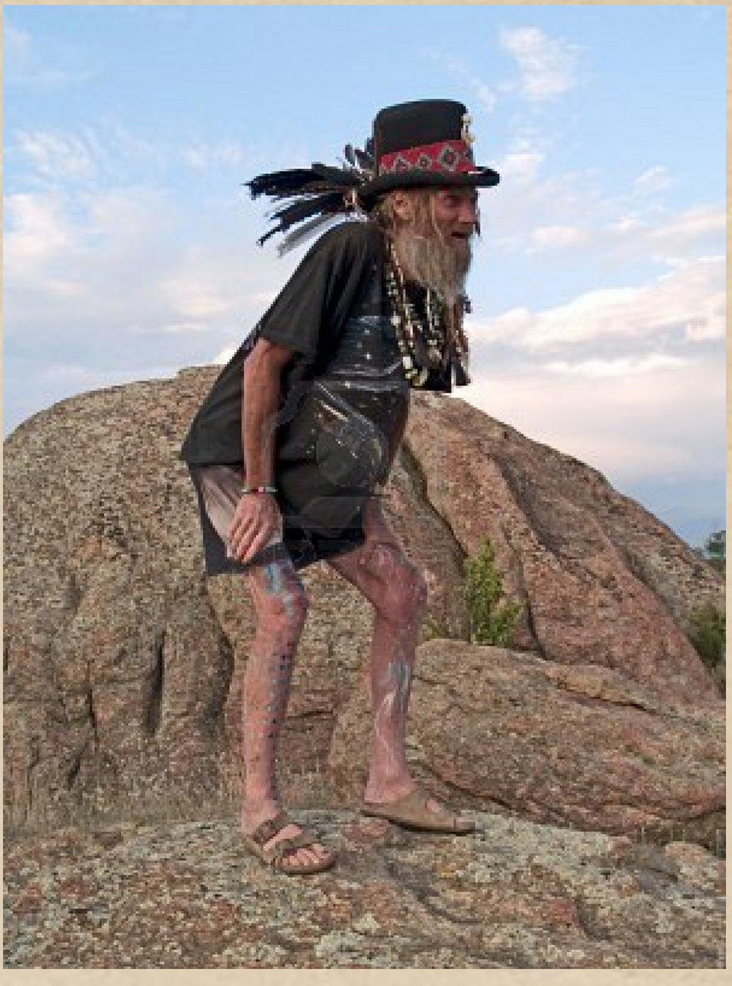
$Z = m_{Jet}/p_T$

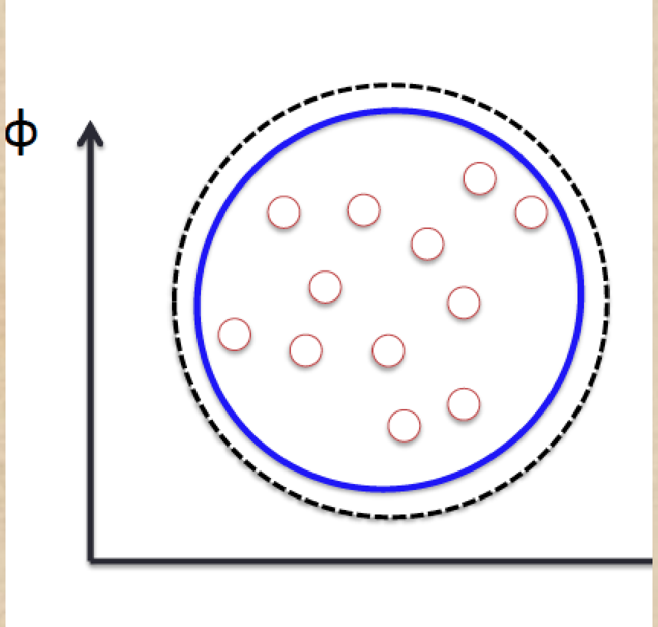
Almeida et al., hep-ph:0807.0234 using MadGraph

Angularity is suggested as a separator between QCD and heavy object jets

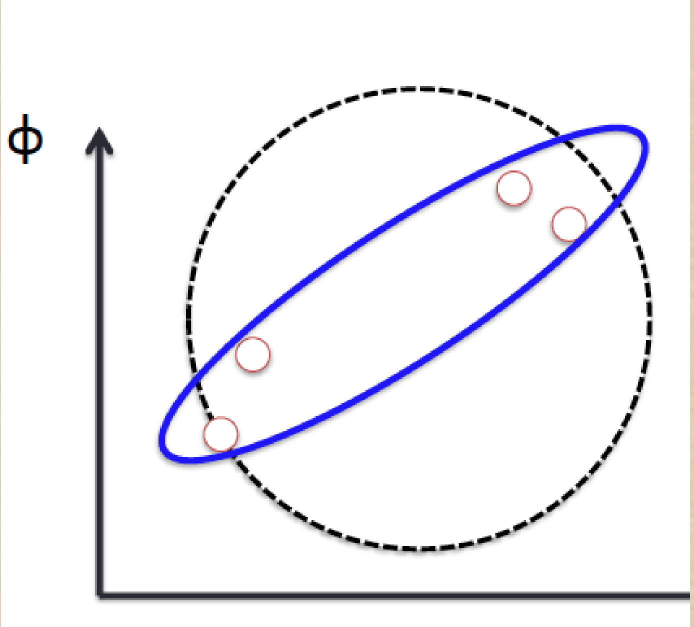
Eccentricity

Eccentricity of jets is defined by $1 - v_{max}/v_{min}$, where v_{max} & v_{min} are the maximal & minimal values of variances of jet constituents along the principle & minor axes.





E~0



E>0

$$\mathcal{E} = 1 - \frac{v_{\min}}{v_{\max}},$$

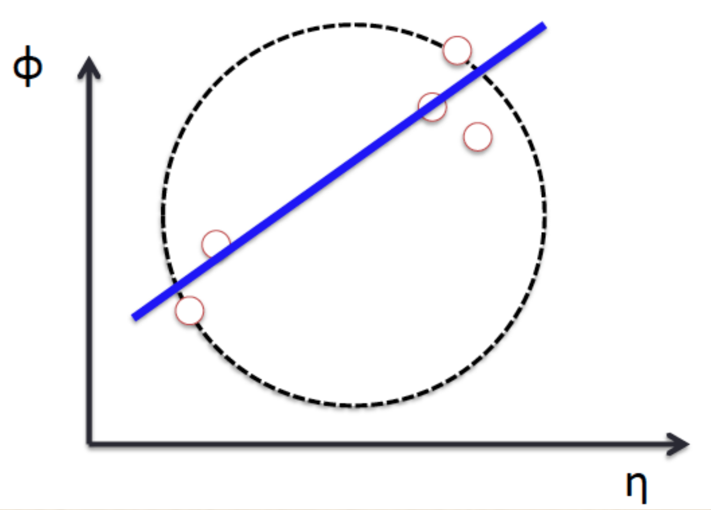
Eccentricity is weakly correlated with mass and width but strongly correlated with the planar flow

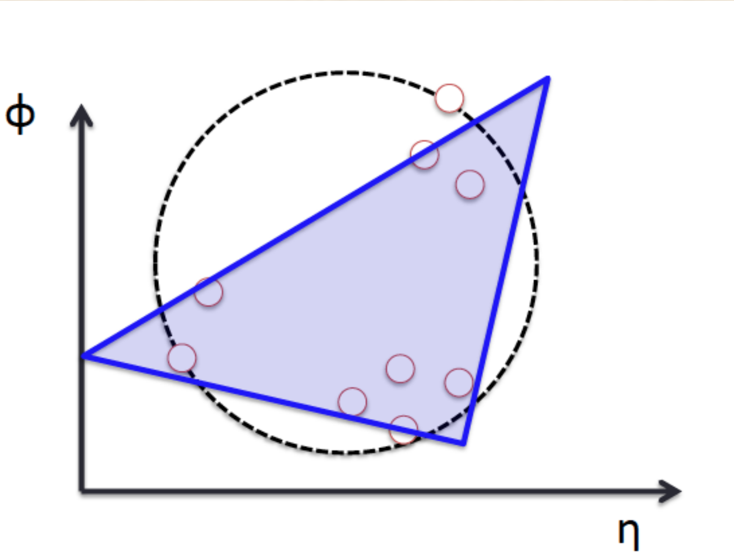
Planar Flow

Planar flow is helpful in separating two-body decays from multi-body decays

$$I_E^{kl} = \sum_i \frac{1}{E_i} p_{i,k} p_{i,l}$$

$$p_l = 4\frac{det(I_E)}{tr(I_E)} = \frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}$$





Two-body decays, e.g. boosted H, W or Z should give very low planar flow while 3-body decays, e.g. top quarks should give relatively high value of planar flow

choose a plane that minimizes the p_T outside, and measures the sum of this p_T.

Planar flow can distinguish between two and three body decays

N-Subjettiness

$$\tau_N \equiv \frac{1}{d_0} \sum_{k=1}^{M} \left(p_{\mathrm{T},k} \times \underbrace{\Delta R_{\mathrm{min},k}}_{\text{distance to nearest subjet}} \right)$$

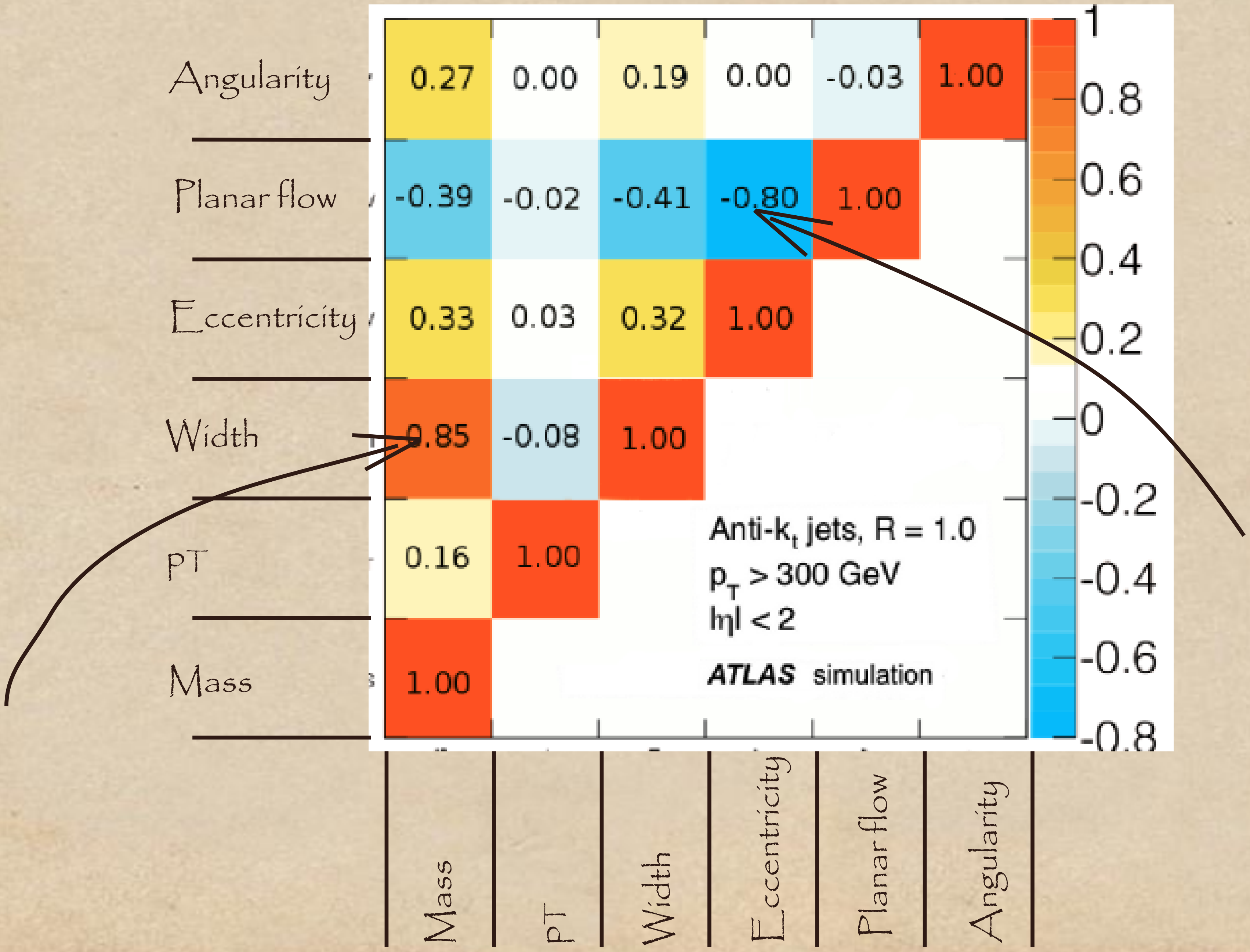
$d_0 = R \times$ sum of p_{T} of all constituents

- τ_N small if hard constituents close to subjets
- $\tau_{N+1}/\tau_N < 1$: structure better described with additional subjet

Jesse Thaler and Ken Van Tilburg arXiv 1011.2268 [hep-ph]

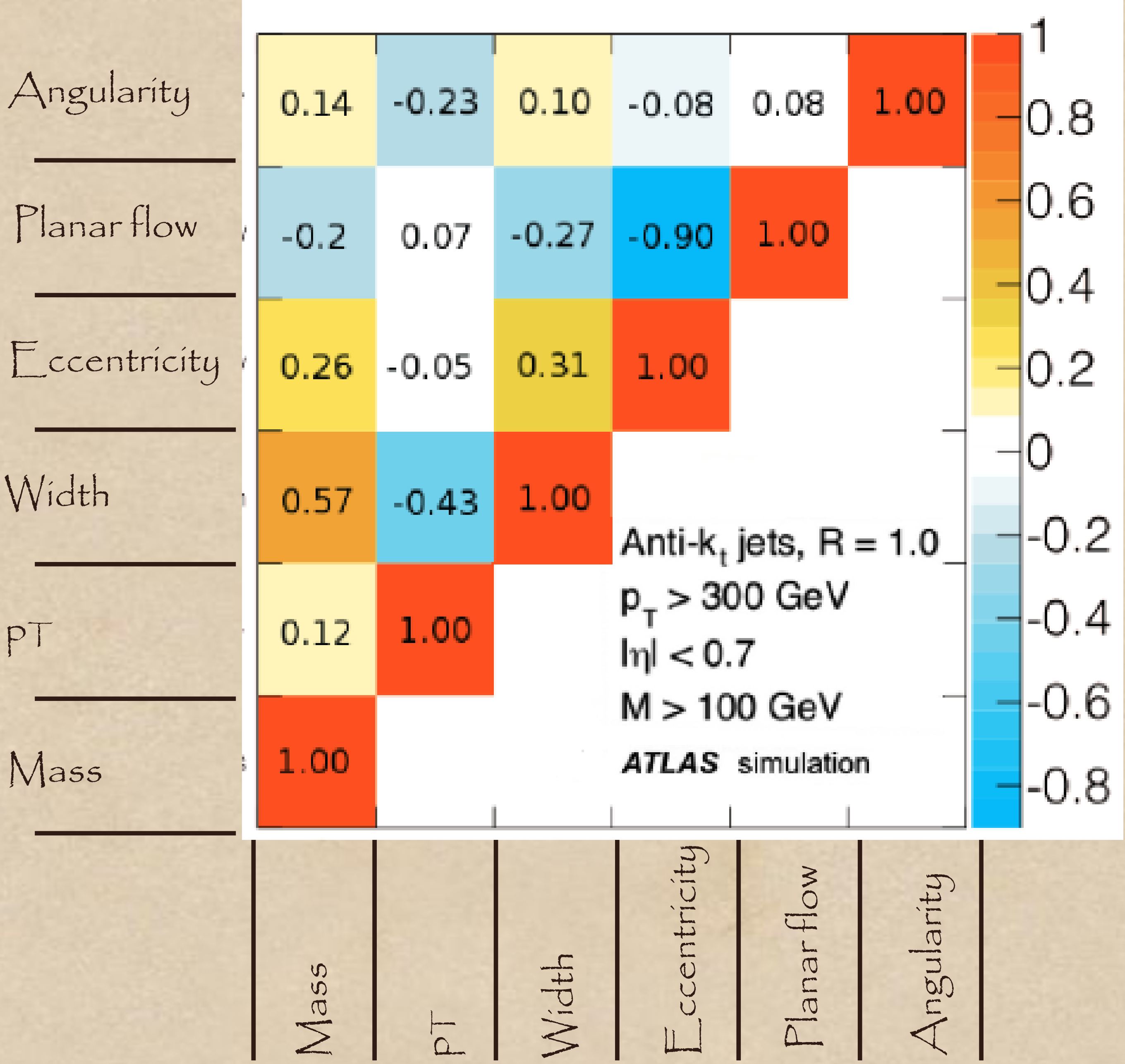
Correlations

The various shape/substructure measures are correlated with each other using MC. The exact level of correlation depends on the MC used, but general picture can be obtained by studying any of the MC. Here, Pythia 6.4 is used:



	Mass	PT	Width	Eccentricity	Planar flow	Angularity
Angularity	0.27	0.00	0.19	0.00	-0.03	1.00
Planar flow	-0.39	-0.02	-0.41	-0.80	1.00	
Eccentricity	0.33	0.03	0.32	1.00		
Width	0.85	-0.08	1.00			
PT	0.16	1.00				
Mass	1.00					

Correlations for Massive Jets



	Mass	PT	Width	Eccentricity	Planar flow	Angularity
Angularity	0.14	-0.23	0.10	-0.08	0.08	1.00
Planar flow	-0.2	0.07	-0.27	-0.90	1.00	
Eccentricity	0.26	-0.05	0.31	1.00		
Width	0.57	-0.43	1.00			
PT	0.12	1.00				
Mass	1.00					

While inspecting high mass (M>100GeV) jets, the correlations between substructure observables are changed.

Variation of Correlation as a Function of the Mass

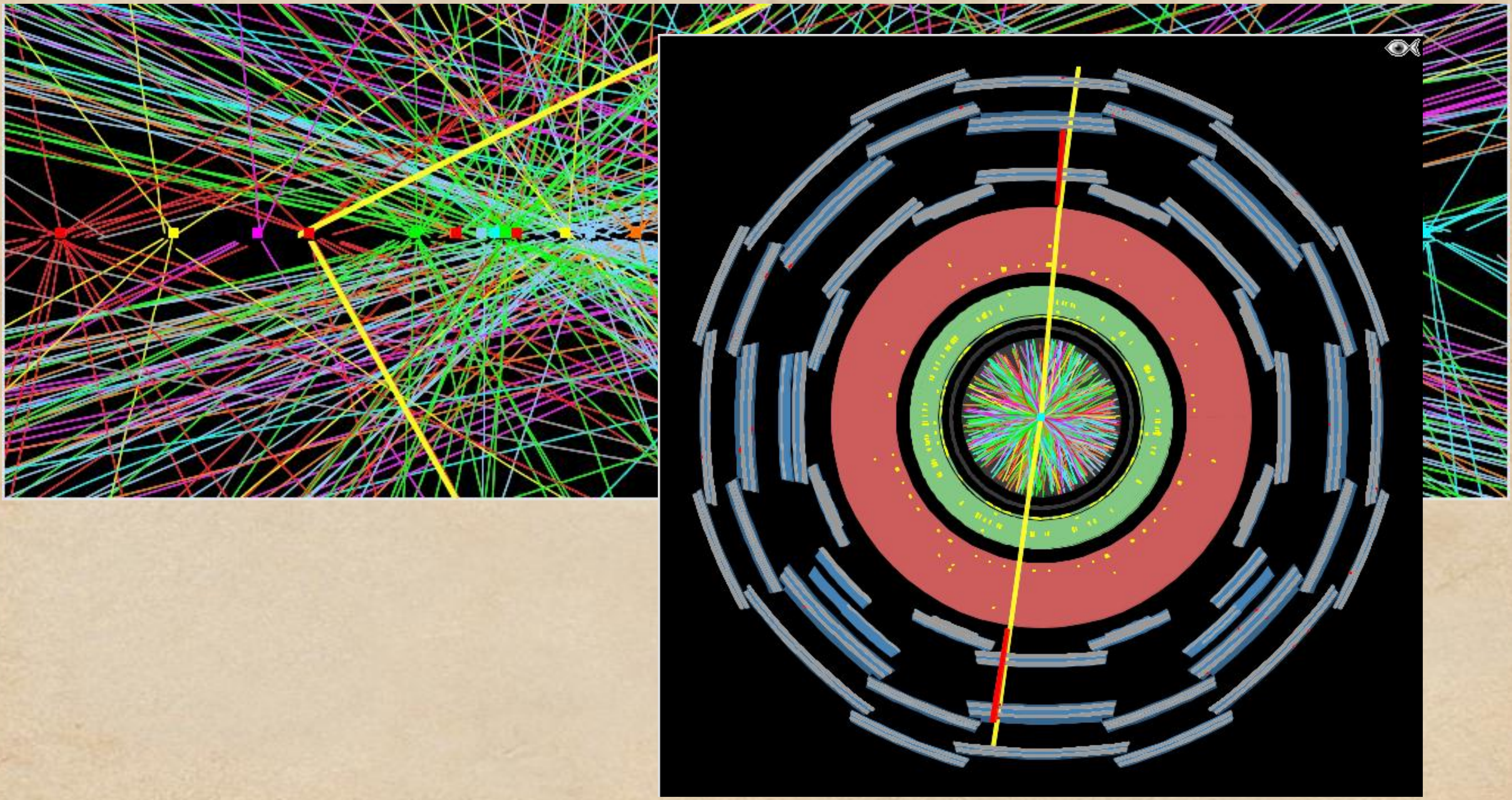
$C_{ij}^{All}-C_{ij}^{Massive}$

	Angularity	**Planar flow**	**Eccentricity**	**p_T**	**Mass**
Angularity	0.13	0.23	0.09	0.08	-0.11
Planar flow	-0.19	-0.09	-0.14	0.10	
Eccentricity	0.07	0.08	0.01		
p_T	0.28	0.35			
Mass	0.04				

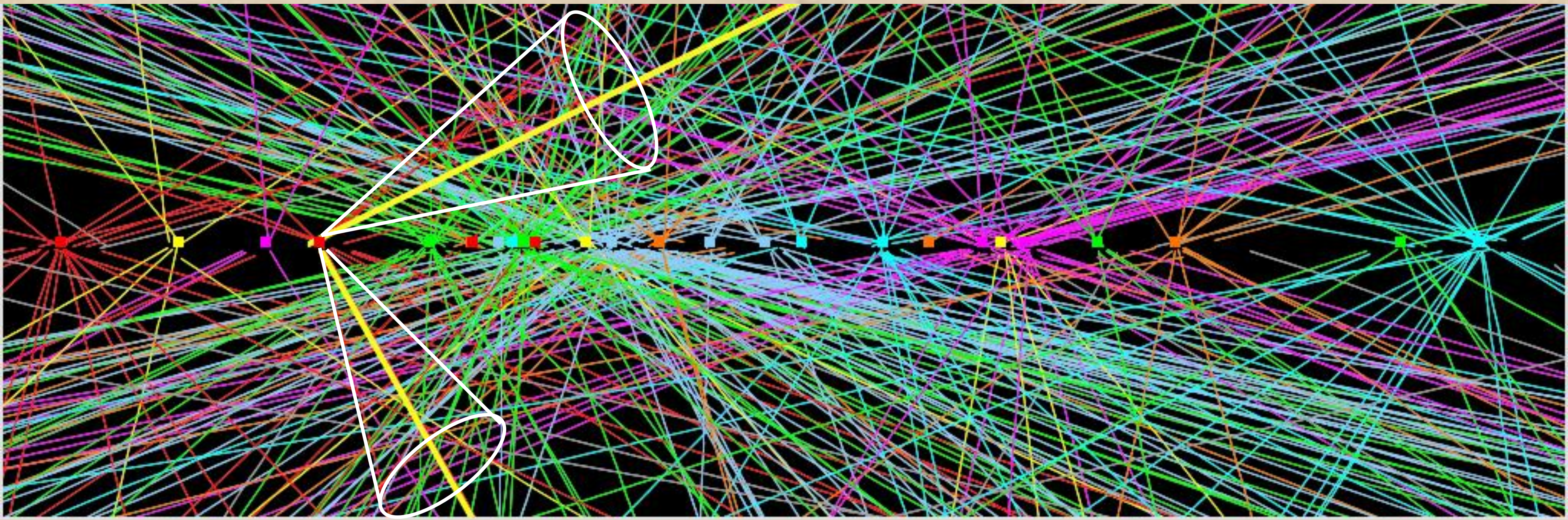
The move from all jets to heavy ones induces changes in the correlations between various observables. The effect is quite small.

The Pileup Problem

“Typical” event



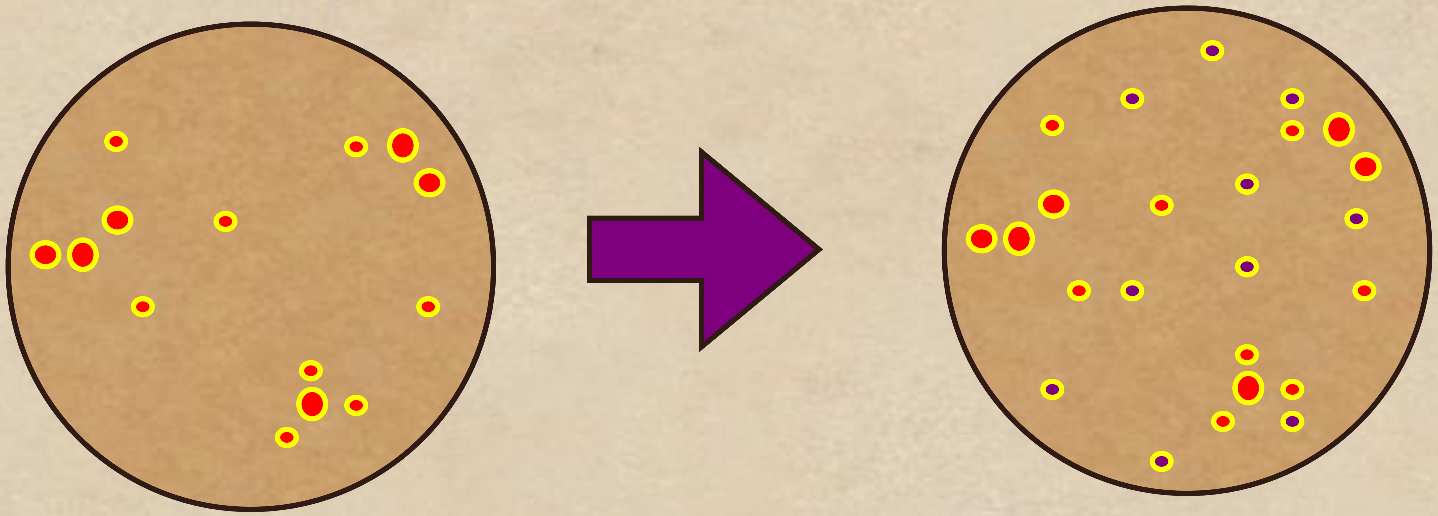
Jets' Contamination



Jets would include significant amount of "incoherent" energy from MI and UE

Need to find a method to remove the excessive contribution

Grooming



Jet grooming: Family of algorithms that seek to get rid of the softer components in and around a jet from UE or pileup and keep the constituents of the hard scatter for further analysis.

- Pruning
- Filtering
- Trimming

brush, clean, coach, comb, curry, dress, drill,educate, lick into shape, make attractive, make presentable, nurture, preen, prep, pretty up,prim, prime, primp, put through grind, put through mill, ready, refine, refresh, rub down,shape up, sleek, slick up, smarten up, spiff up,spruce up, tend, tidy, train, turn out

Ambiguous procedures neighborhood-dependent

Mass Drop/Filtering

(J. Butterworth, A. Davidson, M. Rubin, G. Salam; http://arxiv.org/abs/0802.2470) BDRS

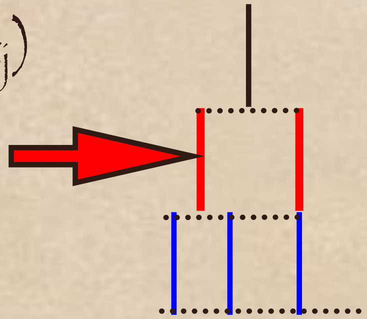
Goal: remove incoherent energy and get the “cleaned” heavy jet

split to 2 partonic constituents

Undo last stage of C/A

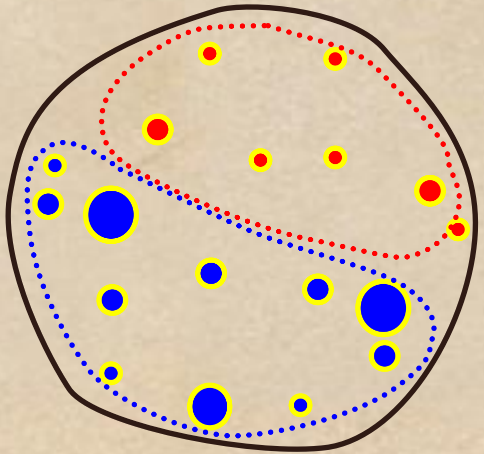
$J(m_j, pt_j)$ $J_1(m_1, pt_1)$ $J_2(m_2, pt_2)$

$m_{j1} > m_{j2}$



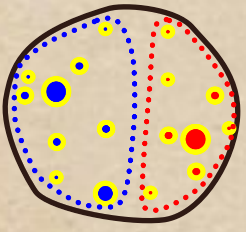
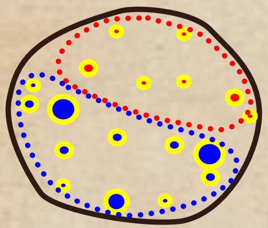
$m_1 > m_2$

But there is another option of split:

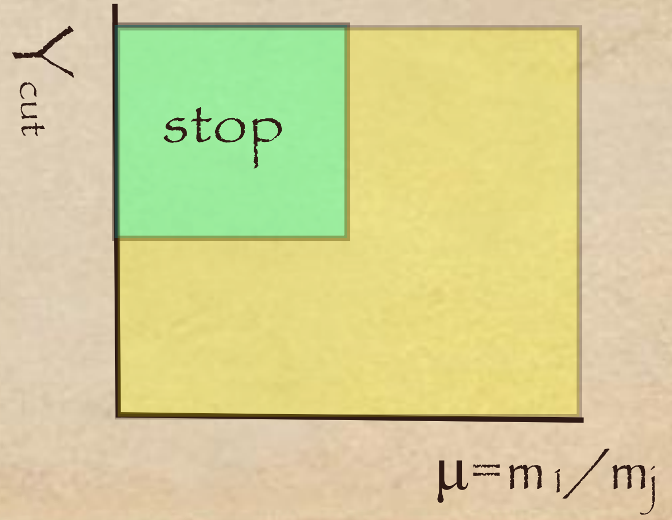


Lump the two partons in one jet(blue) and the incoherent b.g. into the other

Mass Drop Criteria (BDRS)

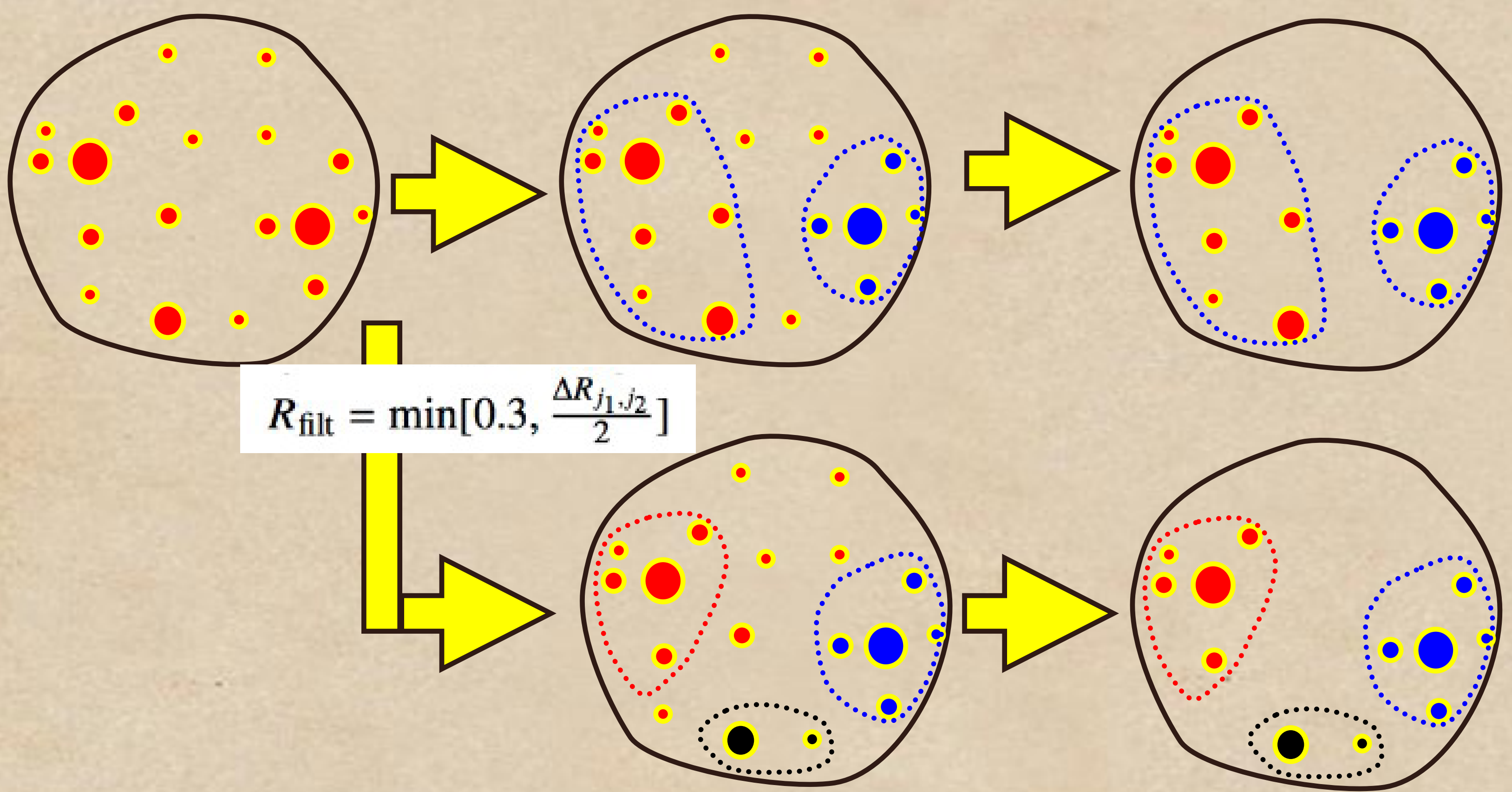
	$\mu = m_1/m_j$	$Y_{cut} = \frac{\min(p_T^2(j1), p_T^2(j2))}{\max(p_T^2(j1), p_T^2(j2))}$	
	Large	Small	Stop
	Small	Large	drop j_2 and use j_1 as new j
	2/3	0.09	

Iteratively removes incoherent energy depositions till system is consistent with the decay of a heavy object into two light ones.

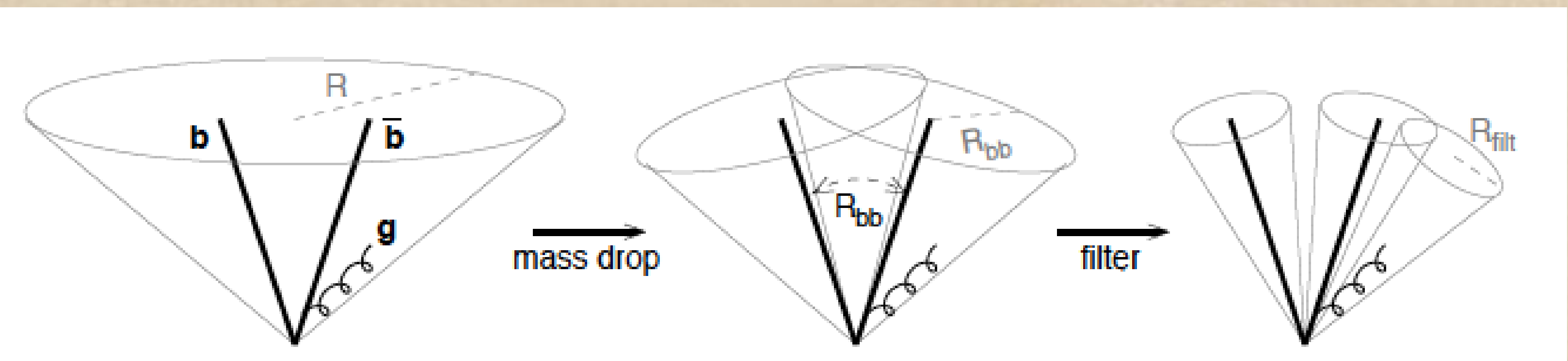


Filtering

The jet is split into sub-jets in an attempt to unveil its inner structure
(2, 3 or more sub-jets)



$$R_{\text{filt}} = \min[0.3, \frac{\Delta R_{j_1,j_2}}{2}]$$

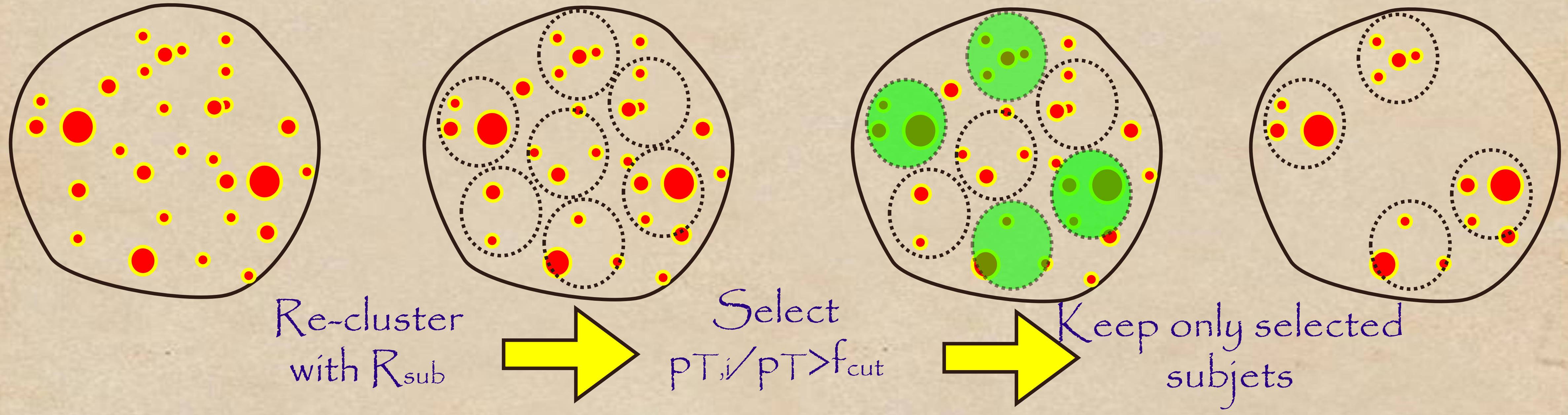


three parameters
μ, Y_{Cut}, R_{Filt}

Trimming

D. Krohn, J. Thaler, L. Wang, http://arxiv.org/abs/0912.1342

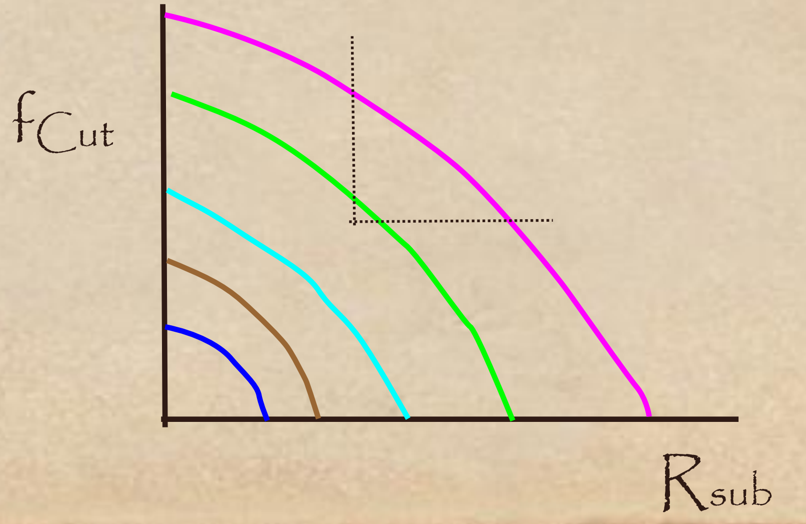
Use a jet algorithm (kt or C/A) to create small subjets of size **R_{sub}** from the constituents of the large-R jet: any subjets failing **$p_{Ti} / p_T < f_{cut}$** are removed



Re-cluster with R_{sub}

Select $p_{T,i}/p_T > f_{cut}$

Keep only selected subjets

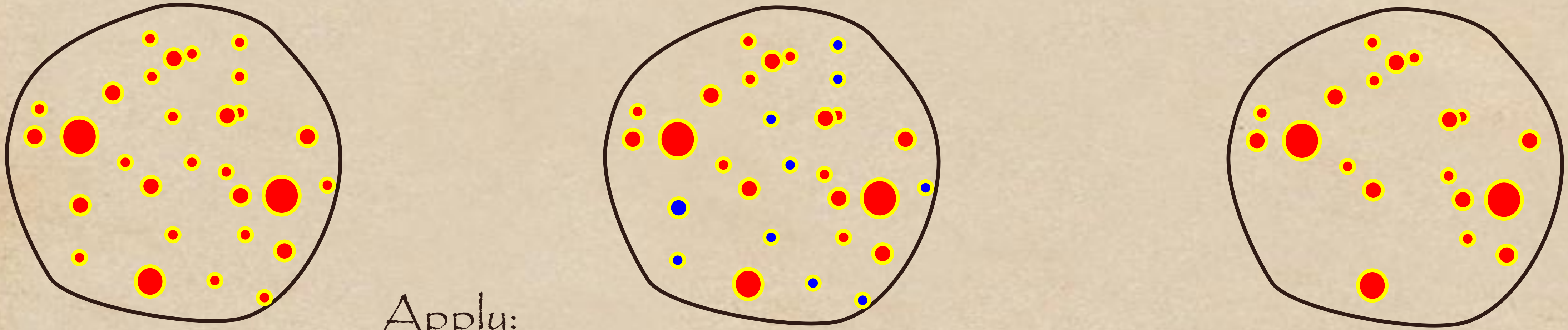


Two free parameters, RSub & fCut

Pruning

S. Ellis, C. Vermilion, J. Walsh, http://arxiv.org/abs/0912.0033

Recombine jet constituents with C/A or kt while vetoing wide angle (R_{cut}) and softer (z_{cut}) constituents.
Does not recreate subjets but prunes at each point in jet reconstruction



Apply:
$p_{T,i}/(p_{T,i}+p_{T,j})>Z_{cut}$
or $R_{ij}>R_{cut}$

$p_{T,i}>\text{alpha}*p_{T,j}$, $\text{alpha}\sim 1$

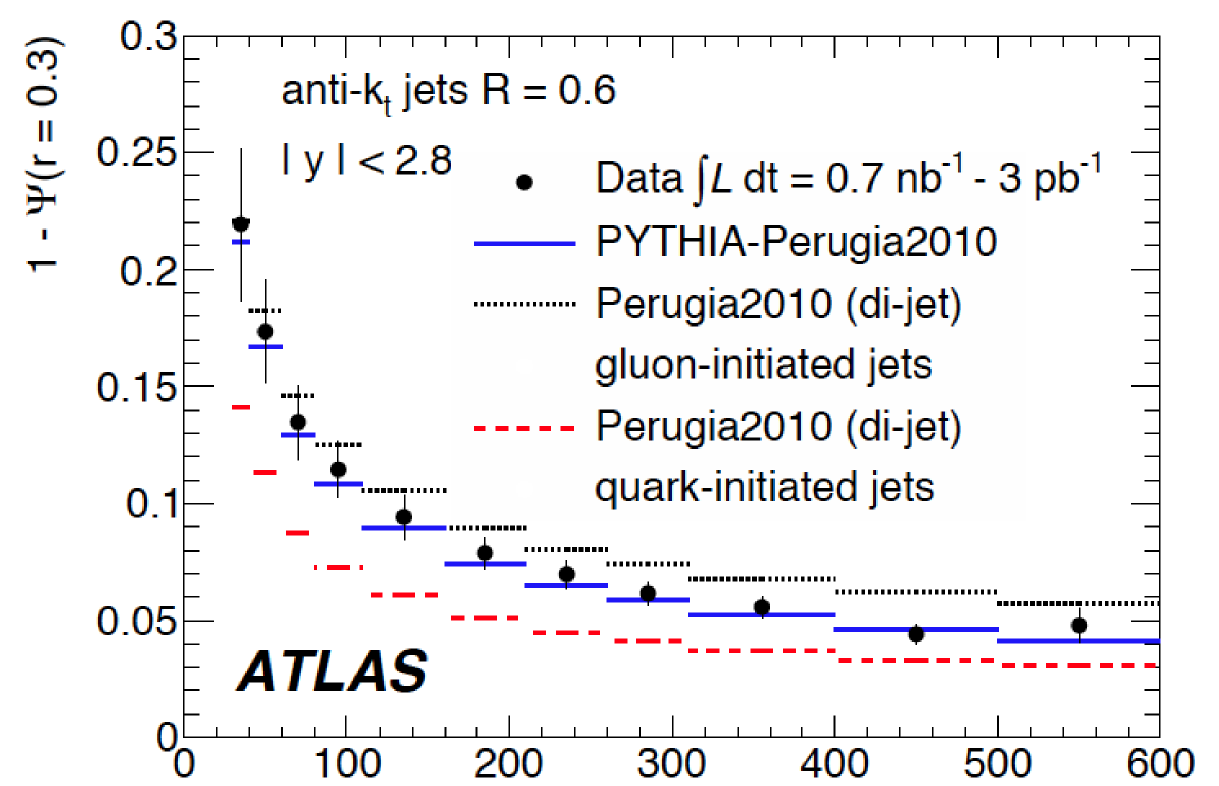
Two parameters
R_{Cut} and Z_{Cut}

Summary of Parameters Studied

Jet Algorithms

Algorithm	**R**	**Grooming**	**Parameters**
Anti-kt	1.0	Pruning	R_{cut}=0.1, 0.2, 0.3; Z_{cut}=0.01, 0.1
Anti-kt	1.2	Trimming	f_{cut}=0.01, 0.03, 0.05; R_{sub}=0.2, 0.3
C/A		M/D+Filtering	mu=0.20, 0.33, 0.67

Integrated and Differential Shapes



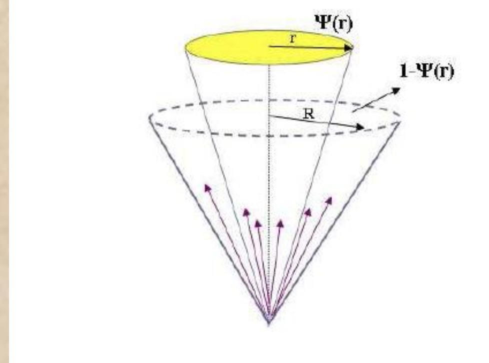
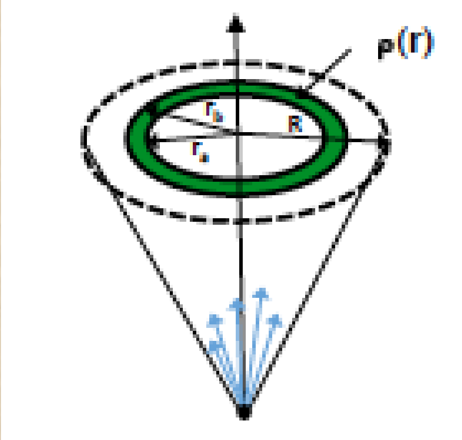
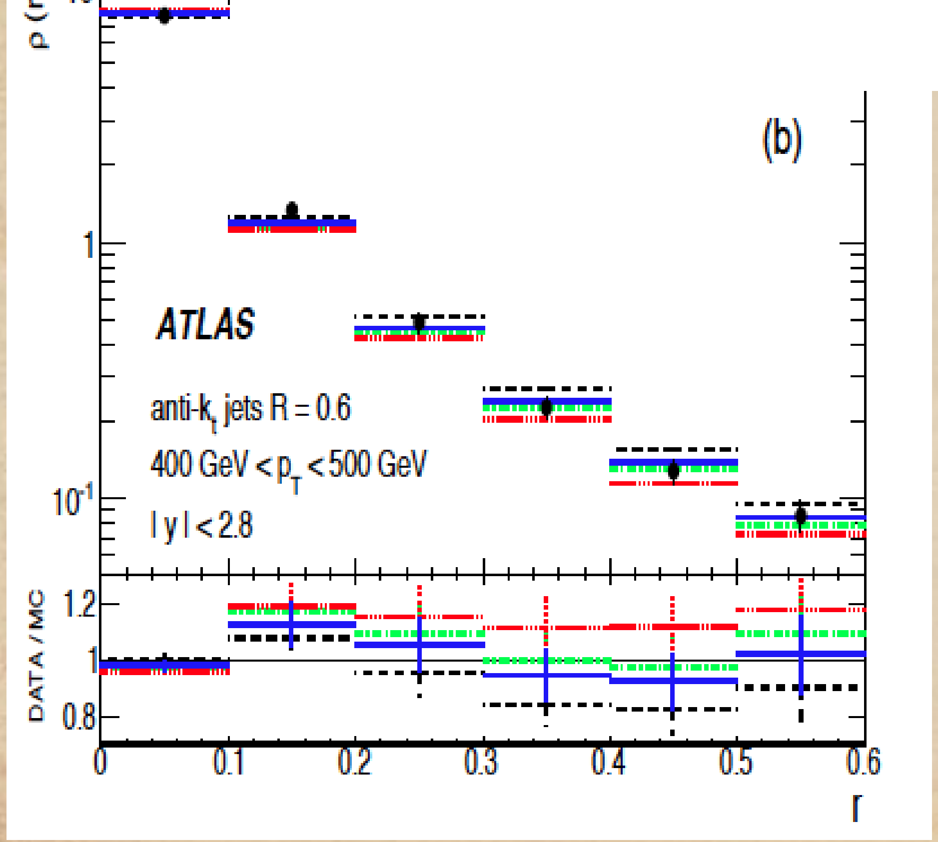
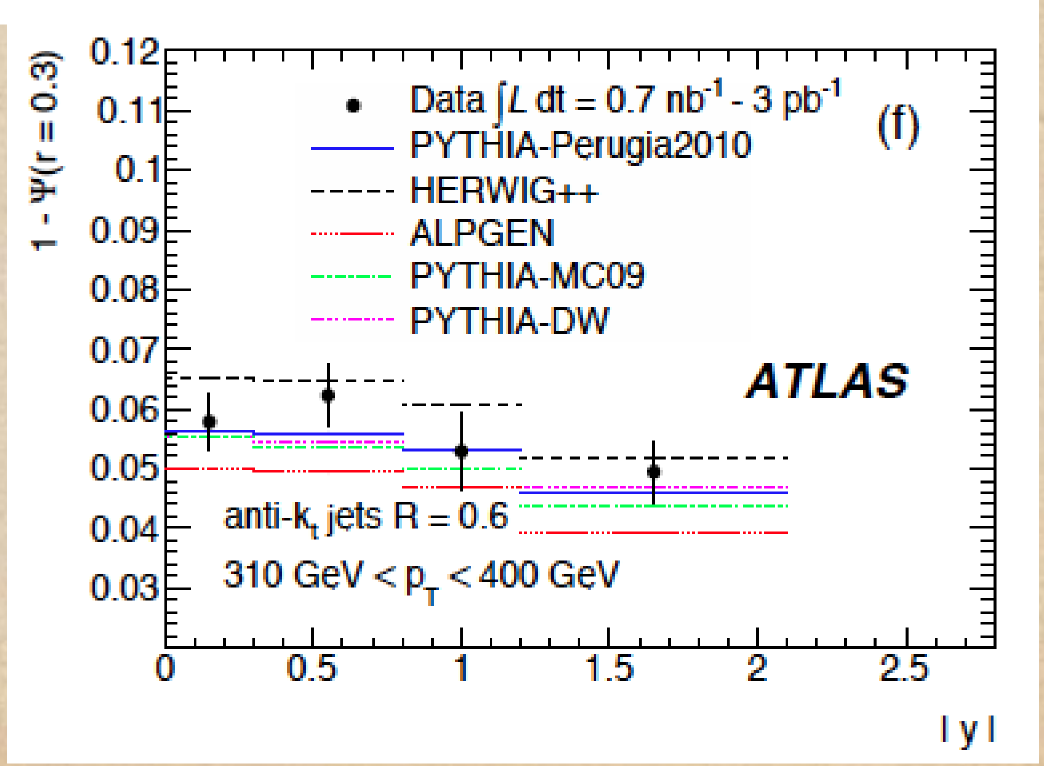


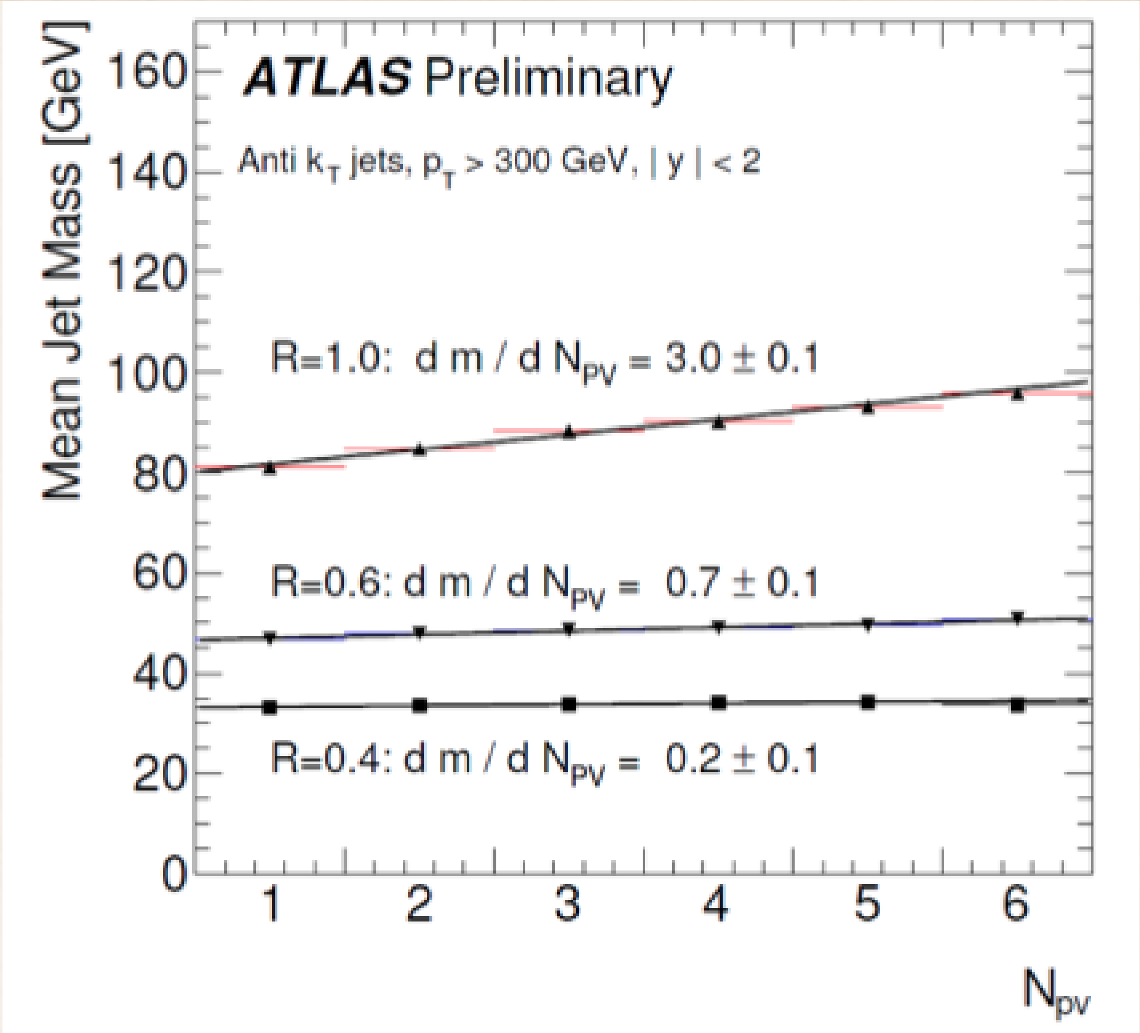
Figure 1: Definition of the integrated jet shape, $\psi(r)$.

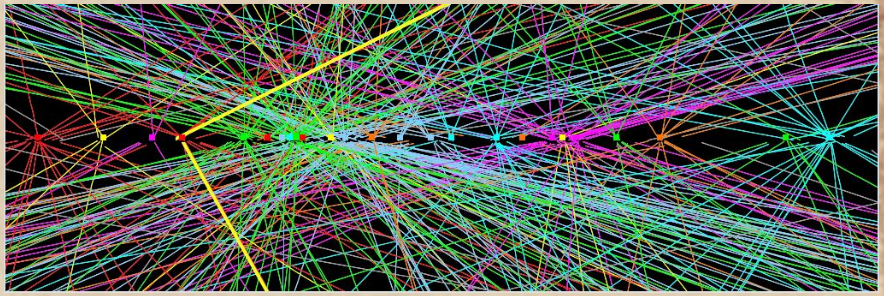






Sanity Check: Pileup Effect on Mass is Scaling with R & N_v

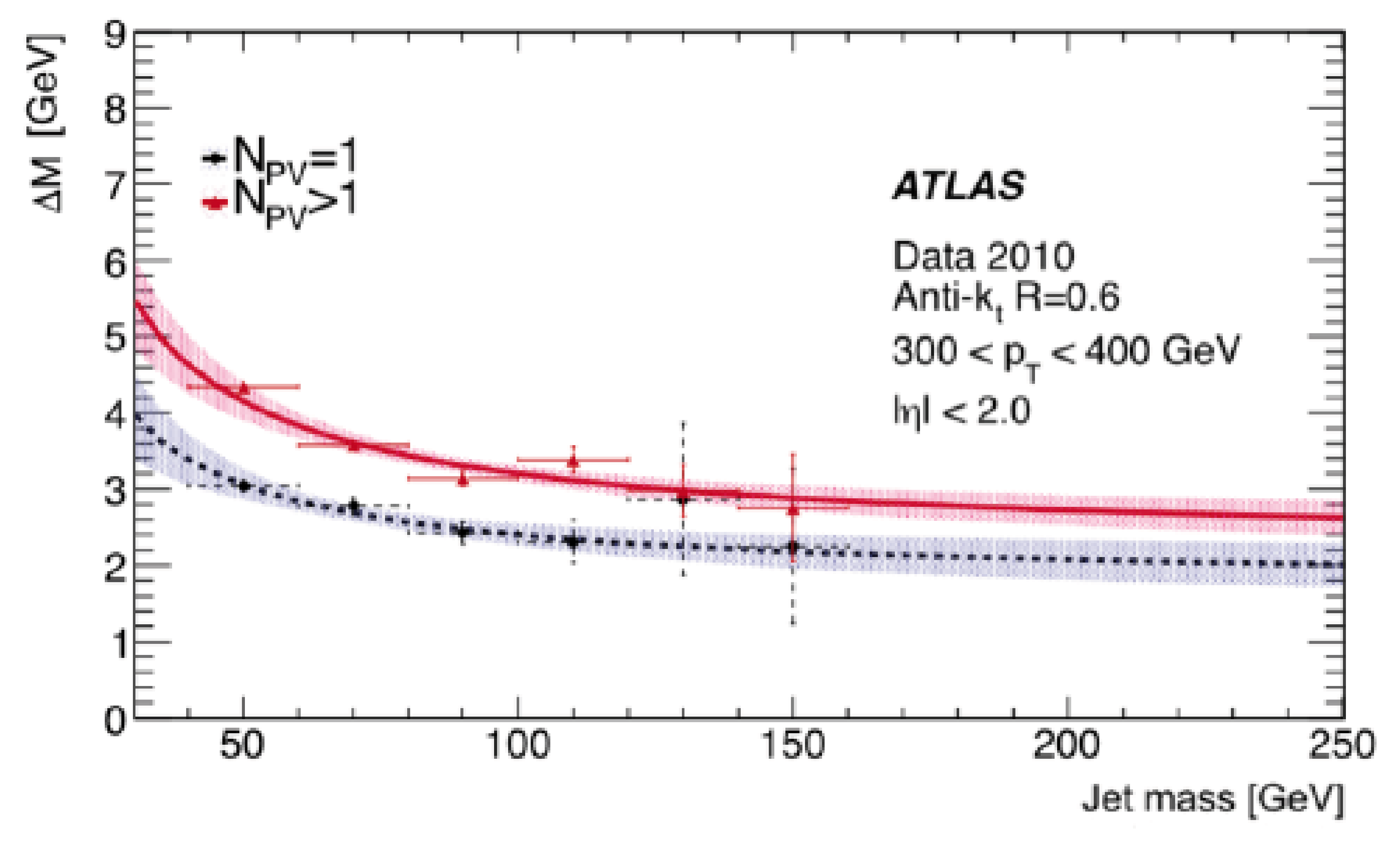




$\Delta m \propto R^4$

As expected: the correction is larger for “wider” jets

Pileup Correction to the Mass Measurement

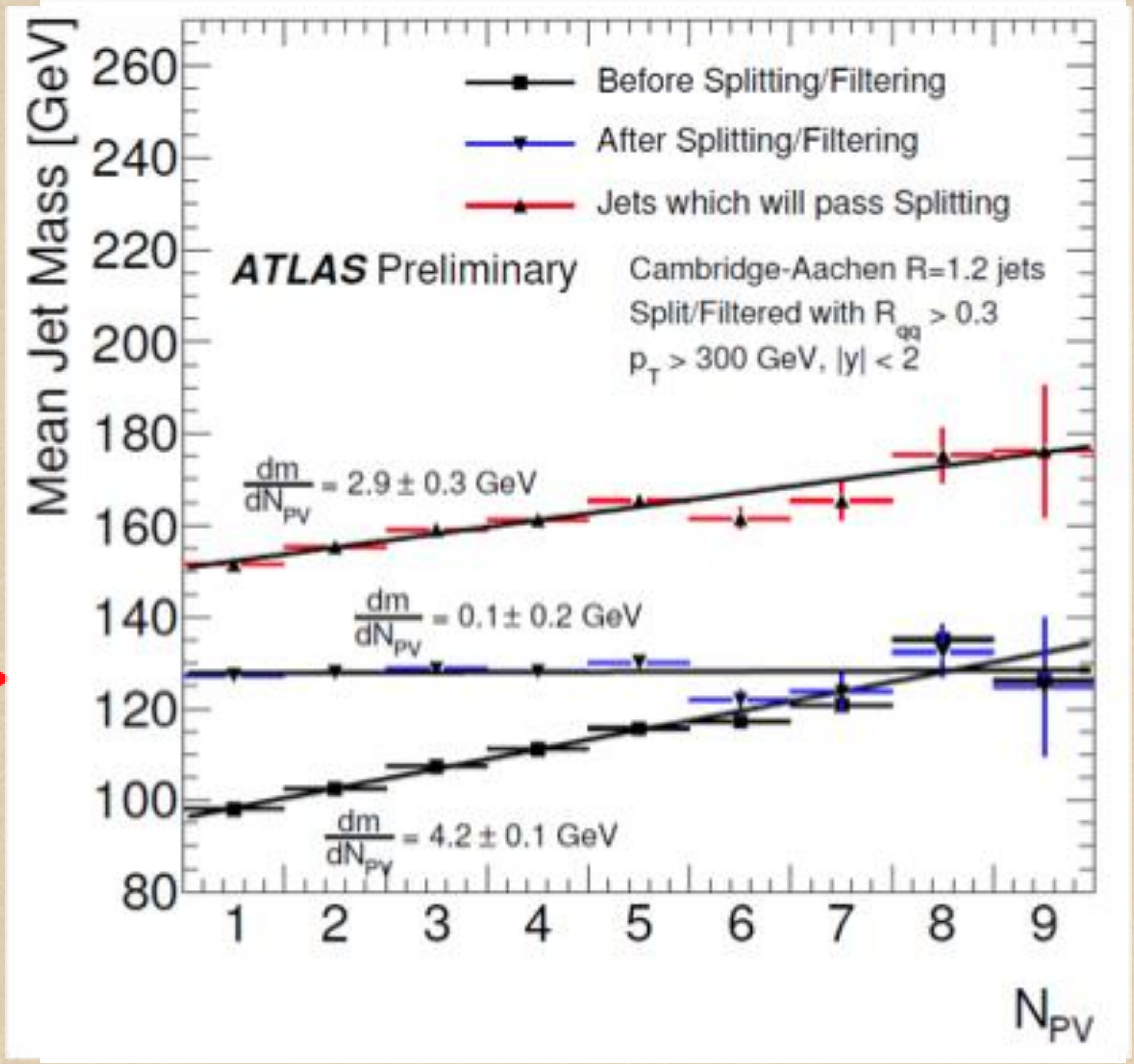


$$\Delta m = p_0 + \frac{p_1}{m}$$

Phenomenological calculations seem to be in agreement with data

R. Alon , E. Duchovni, G. Perez, A.P. Pranko, P. Sinervo. PRD 84 (2011) 114025

Correcting for Pileup by Filtering

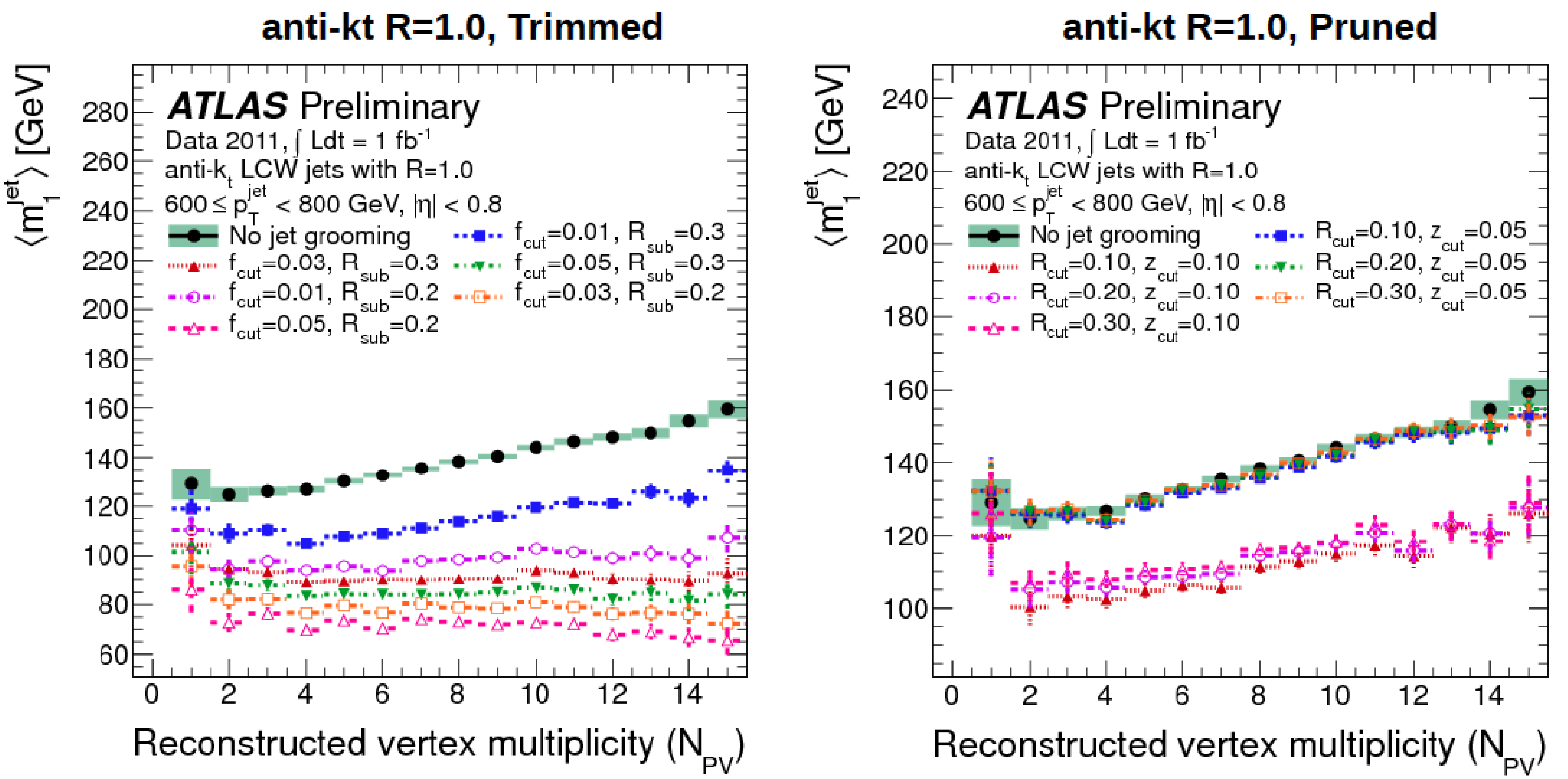


Filtering seems to work very well

Looking at the Performance of Trimming and Pruning

- Mean of mass distribution in data

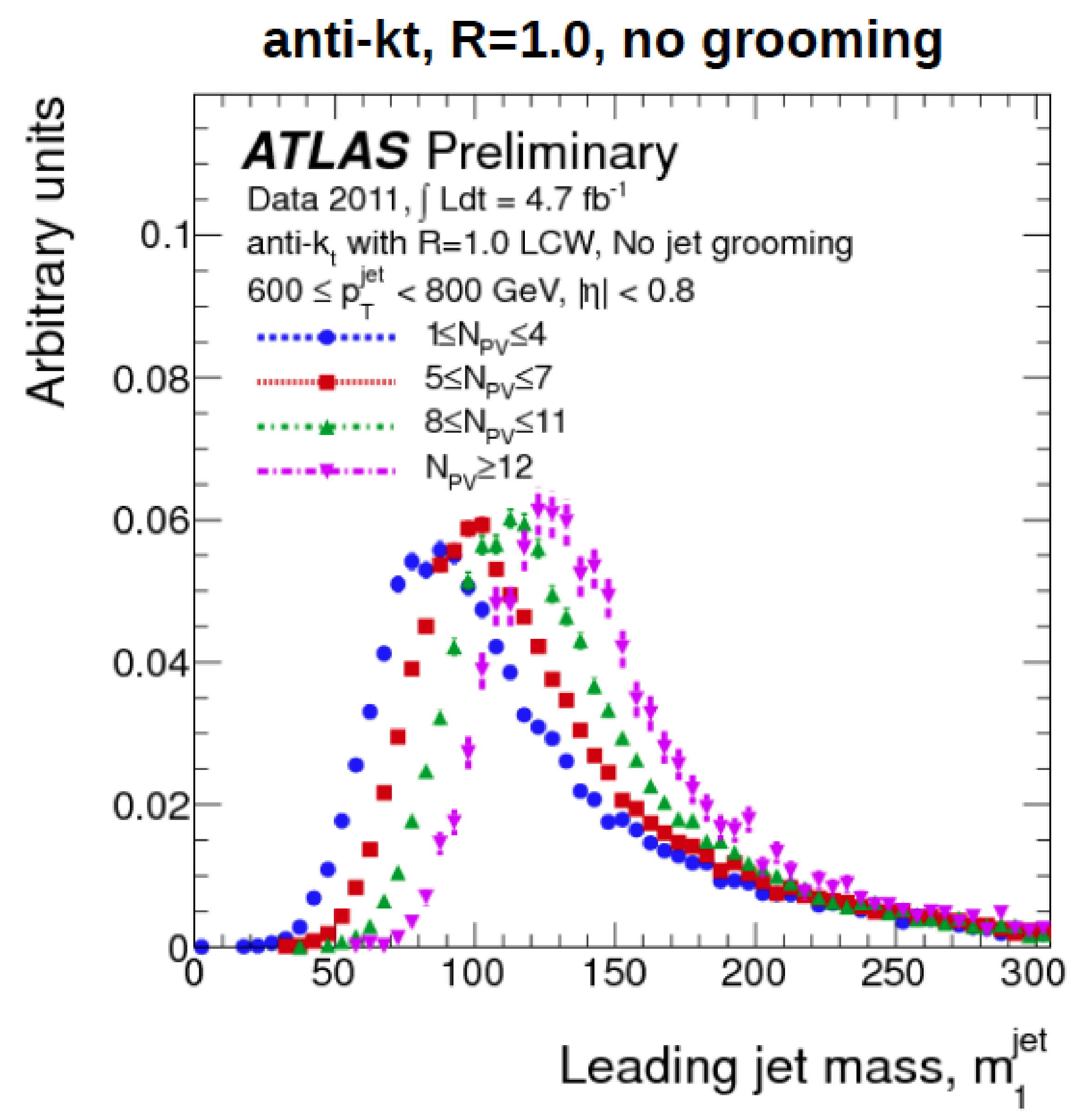
  600 < jet p_T < 800 GeV

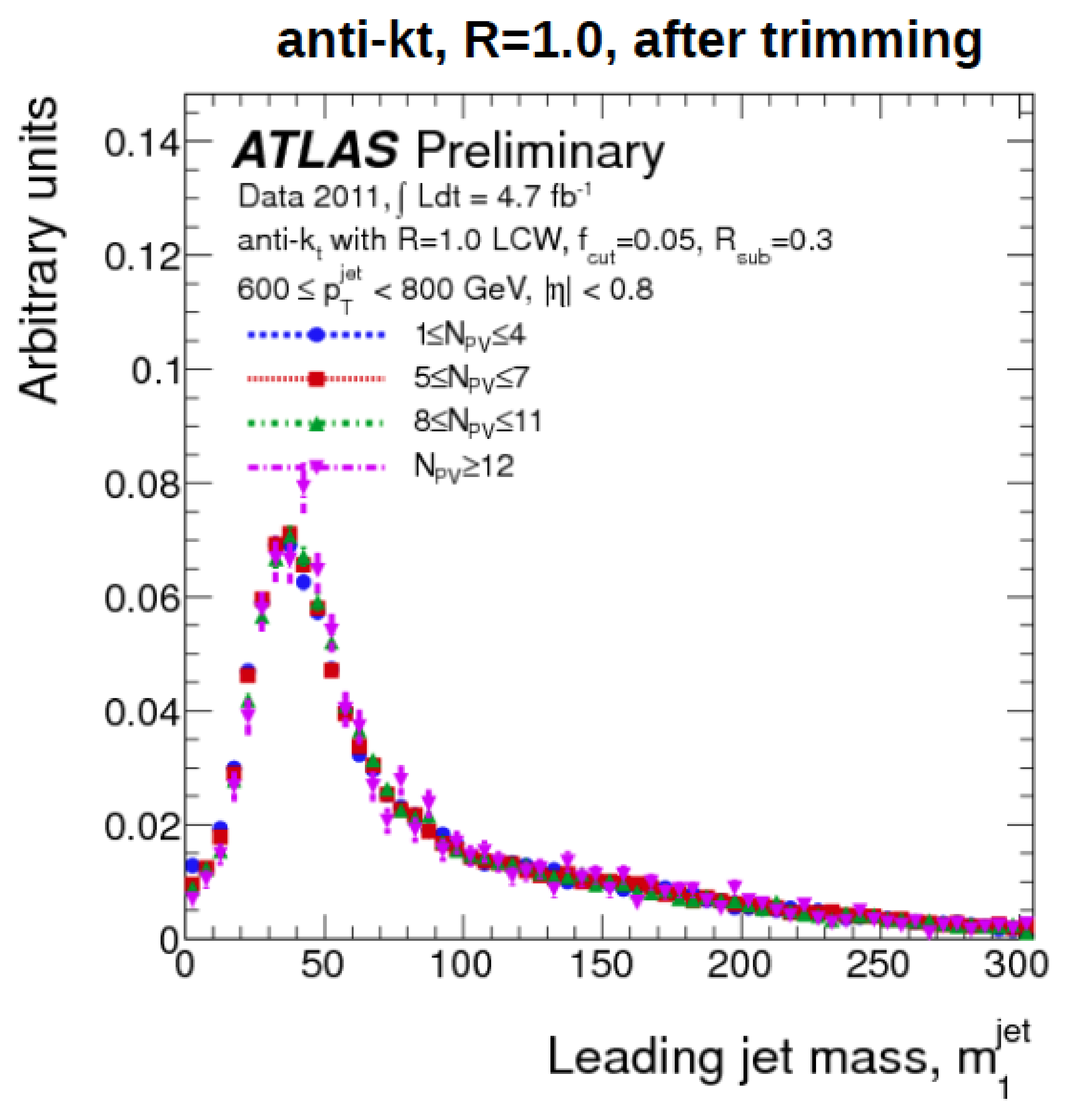


Trimming works quite well

Recommended values are $f_{Cut}=0.05$ & $R_{Sub}=0.3$

Jets' Mass



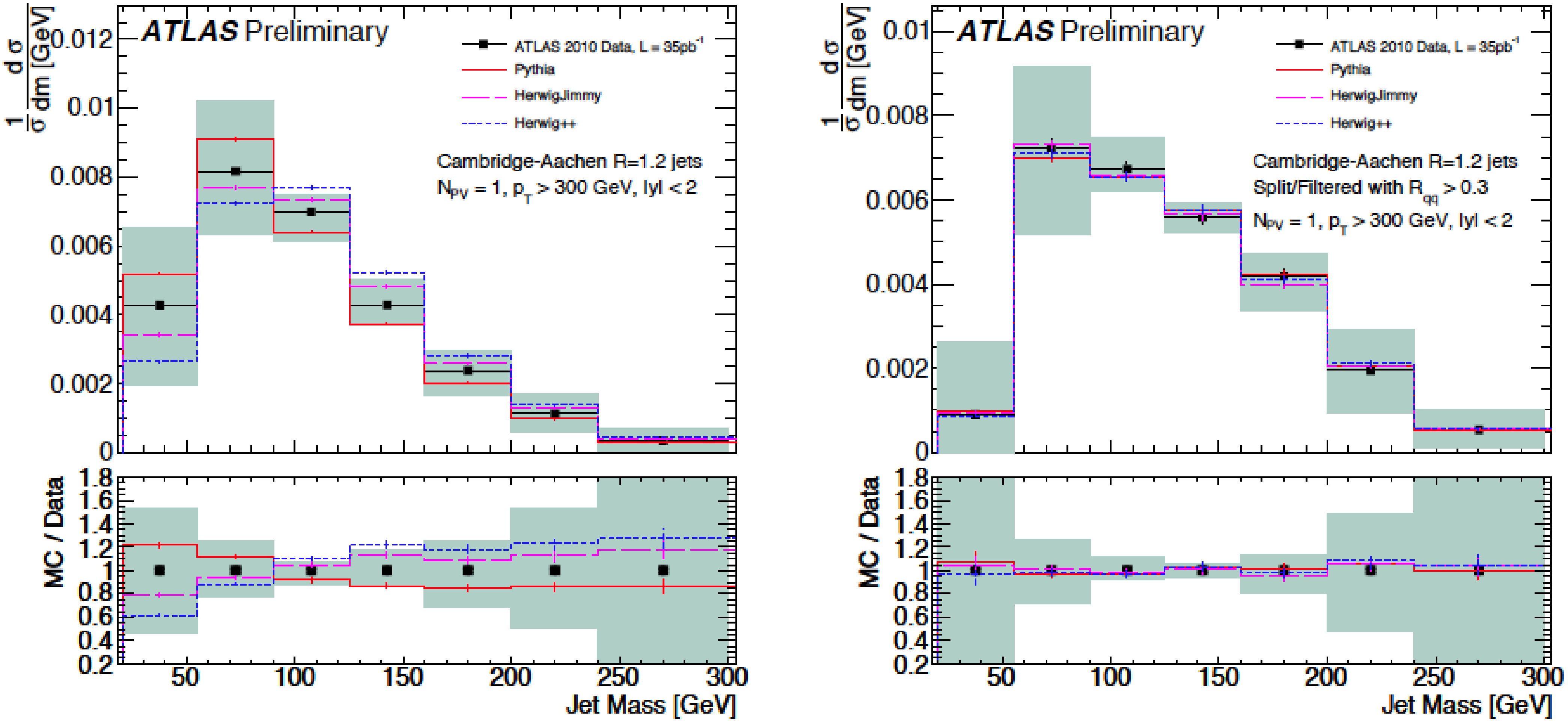


it works well!

Pileup is partially under control

Can now look at the Data

Jets' Mass



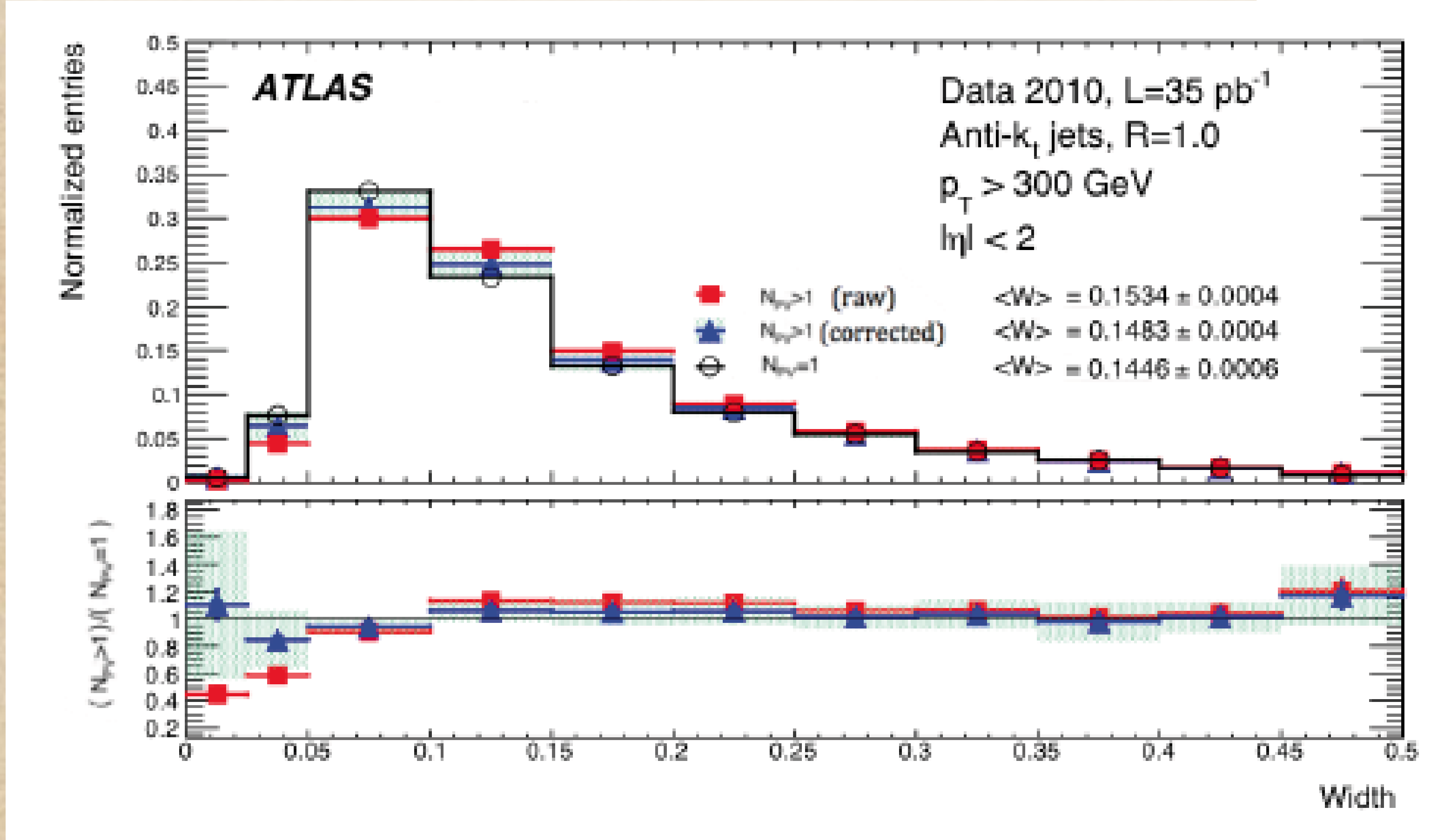
(a) (b)

FIG. 1: Invariant mass spectrum of Cambridge-Aachen jets with $p_T > 300$ GeV and $|y| < 2$ (a) before and (b) after the splitting and filtering procedure has been applied. Both distributions are fully corrected for detector effects, systematic uncertainties are depicted by the shaded band.

Filtering improves the agreement between the data and simulations

Jets' Width

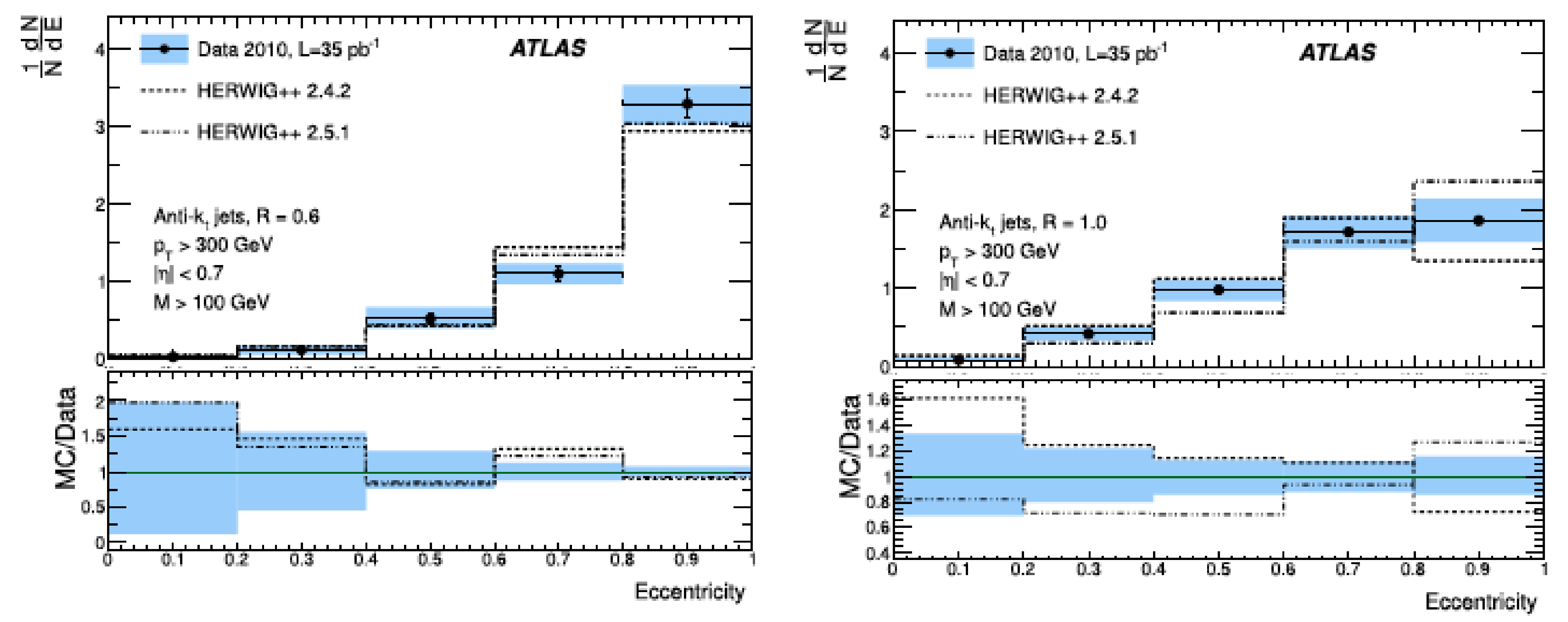
$$W = \frac{\sum_{i=1}^{n} \Delta R^{i} p_T^{i}}{\sum_{i=1}^{n} p_T^{i}}$$



for the record

Jets' Eccentricity

$$\mathcal{E} = 1 - \frac{v_{\min}}{v_{\max}},$$



R=0.6

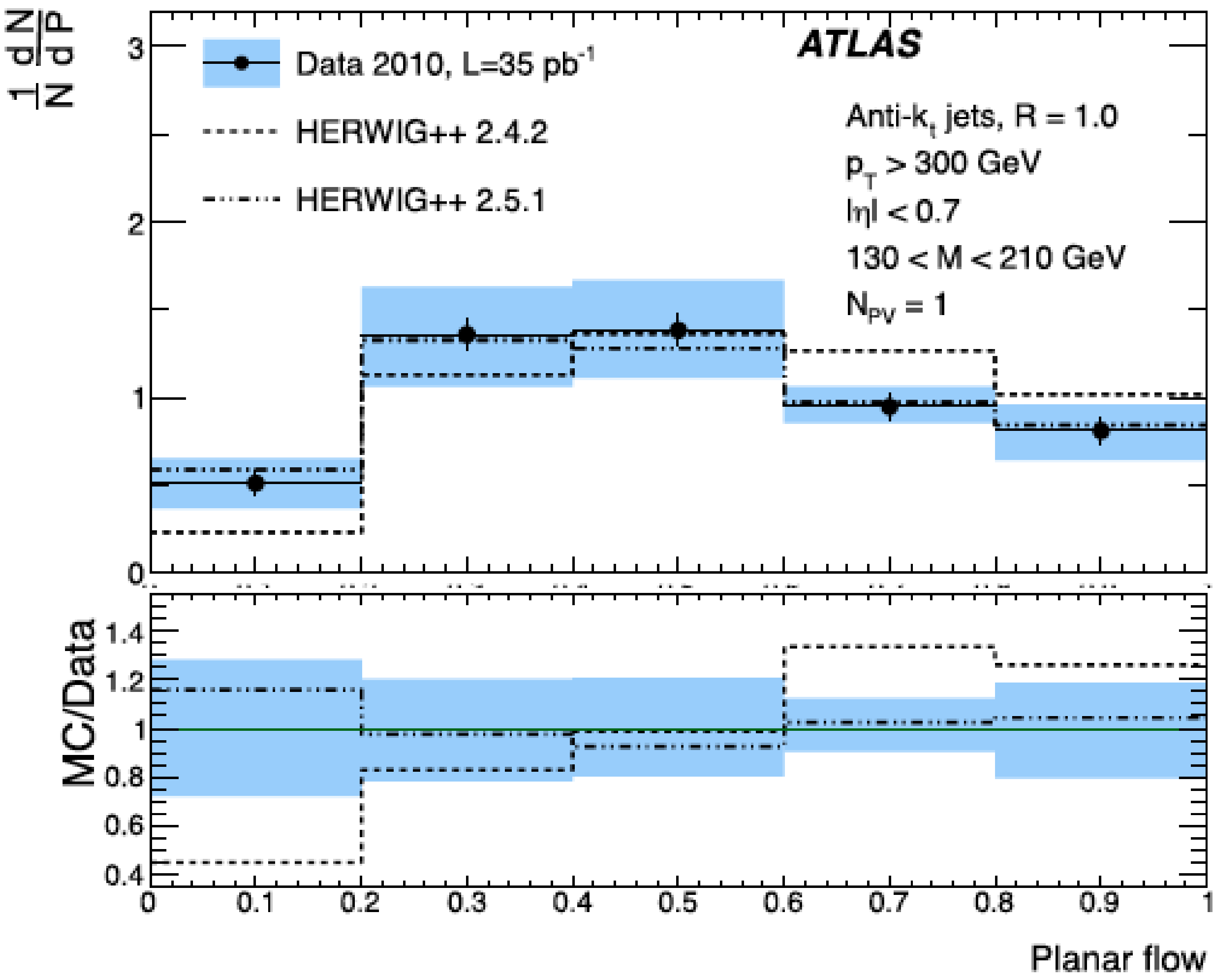
R=1.0

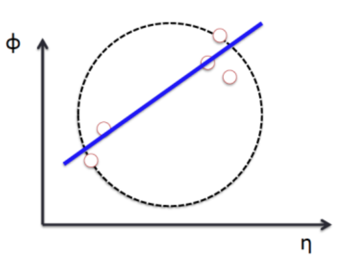
Even the eccentricity is not eccentric

Planar Flow

$$p_l = 4\frac{det(I_E)}{tr(I_E)} = \frac{4\lambda_1\lambda_2}{(\lambda_1+\lambda_2)^2}$$

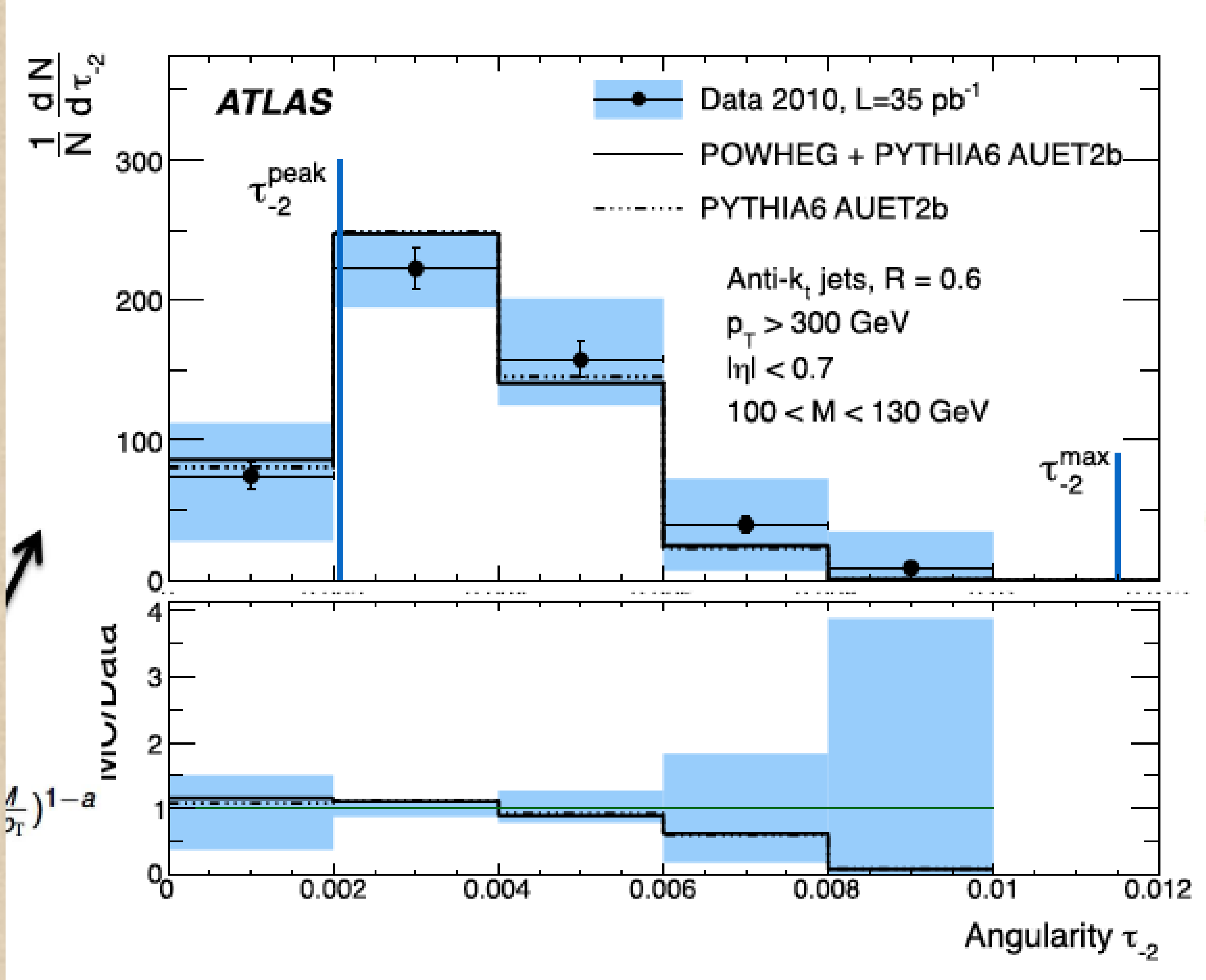
Planar flow





Jets' Angularity

$$\tau_{-2} \approx \frac{1}{8m_{Jet}} \sum_{i \in Jet} E_i \theta_i^4$$



Nice agreement with data and with upper and lower bounds on angularity

N-Sub-Jettiness

$$\tau_N \equiv \frac{1}{d_0}\sum_{k=1}^{M}\left(p_{\mathrm{T},k} \times \underbrace{\Delta R_{\min,k}}_{\text{distance to nearest subjet}}\right)$$

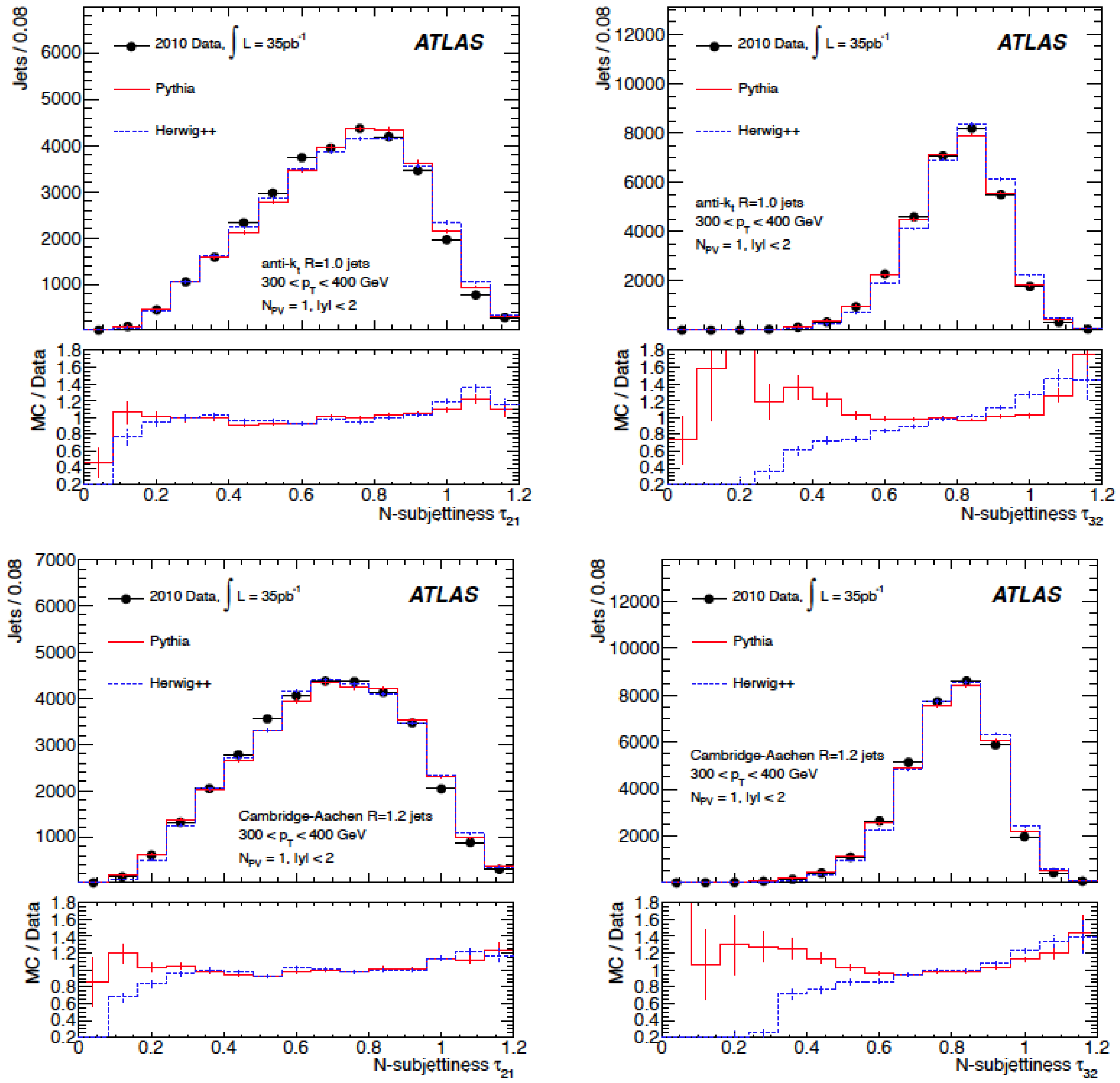
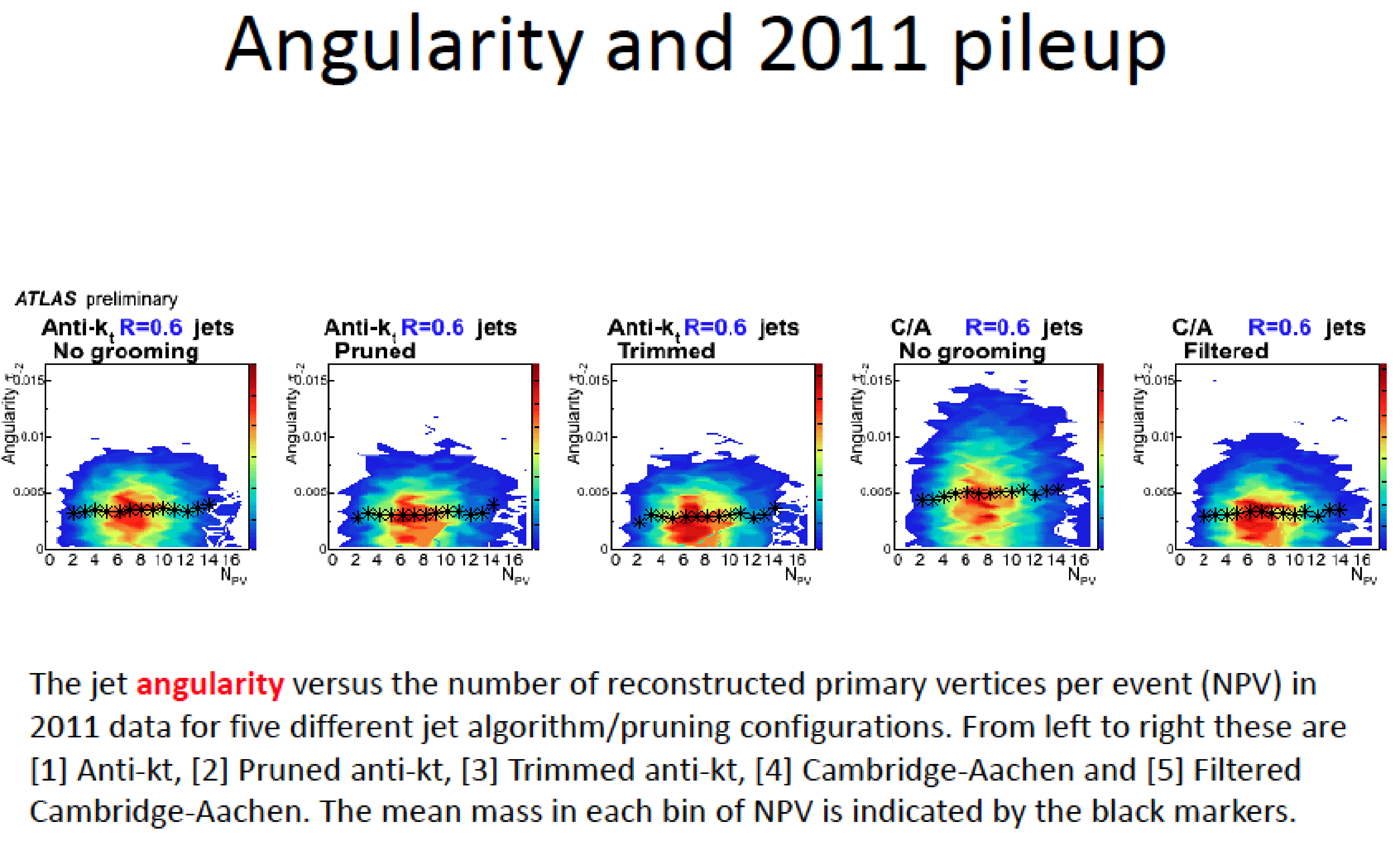


Figure 6. Distributions for τ_{21} (left) and τ_{32} (right) of jets with $|y| < 2.0$ in the 300–400 GeV p_T bin for anti-k_t (top) and Cambridge-Aachen jets (bottom).

Also looks nice

New 2011 Results: Angularity



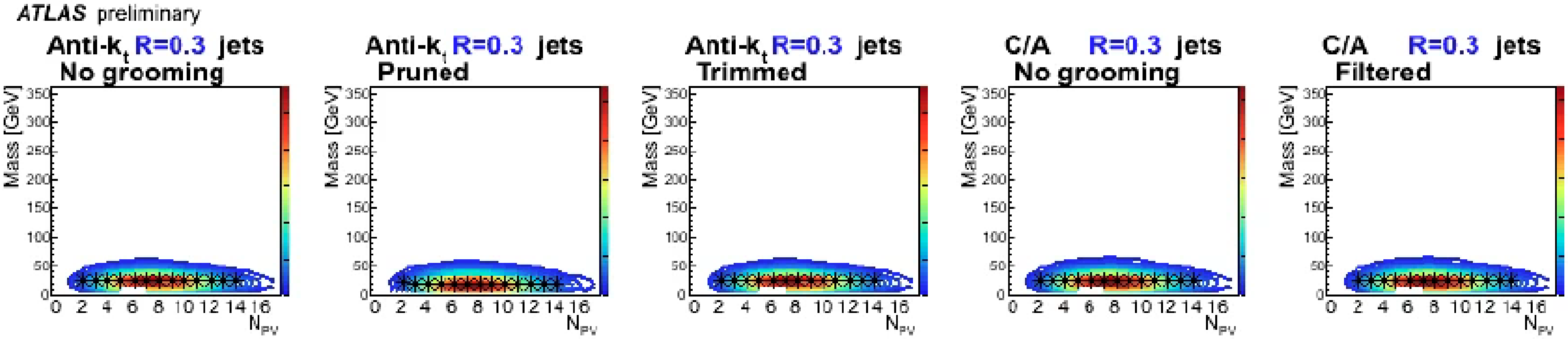
Angularity and 2011 pileup

The jet **angularity** versus the number of reconstructed primary vertices per event (NPV) in 2011 data for five different jet algorithm/pruning configurations. From left to right these are [1] Anti-kt, [2] Pruned anti-kt, [3] Trimmed anti-kt, [4] Cambridge-Aachen and [5] Filtered Cambridge-Aachen. The mean mass in each bin of NPV is indicated by the black markers.

Borrowed from Lily Asquith "Boost2012" talk

New 2011 Results: Mass

Jet mass and 2011 pileup



The jet **mass** versus the number of reconstructed primary vertices per event (NPV) in 2011 data for five different jet algorithm/pruning configurations. From left to right these are [1] Anti-kt, [2] Pruned anti-kt, [3] Trimmed anti-kt, [4] Cambridge-Aachen and [5] Filtered Cambridge-Aachen. As the animation plays, the distance parameter (R) of the jet increases from 0.4 to 1.6. The mean mass in each bin of NPV is indicated by the black markers

Borrowed from Lily Asquith "Boost2012" talk

Usage

- X->tt
- WH
- ZH

others

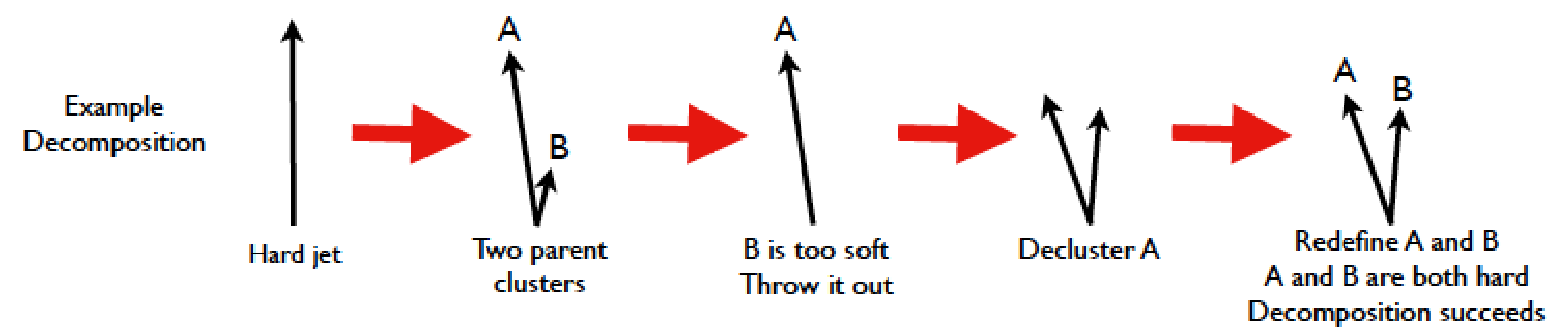
Tagging Top Jets at CMS

Basic requests:
- ★jets with $p_T > 250$ GeV
- ★$|\eta| < 2.5$

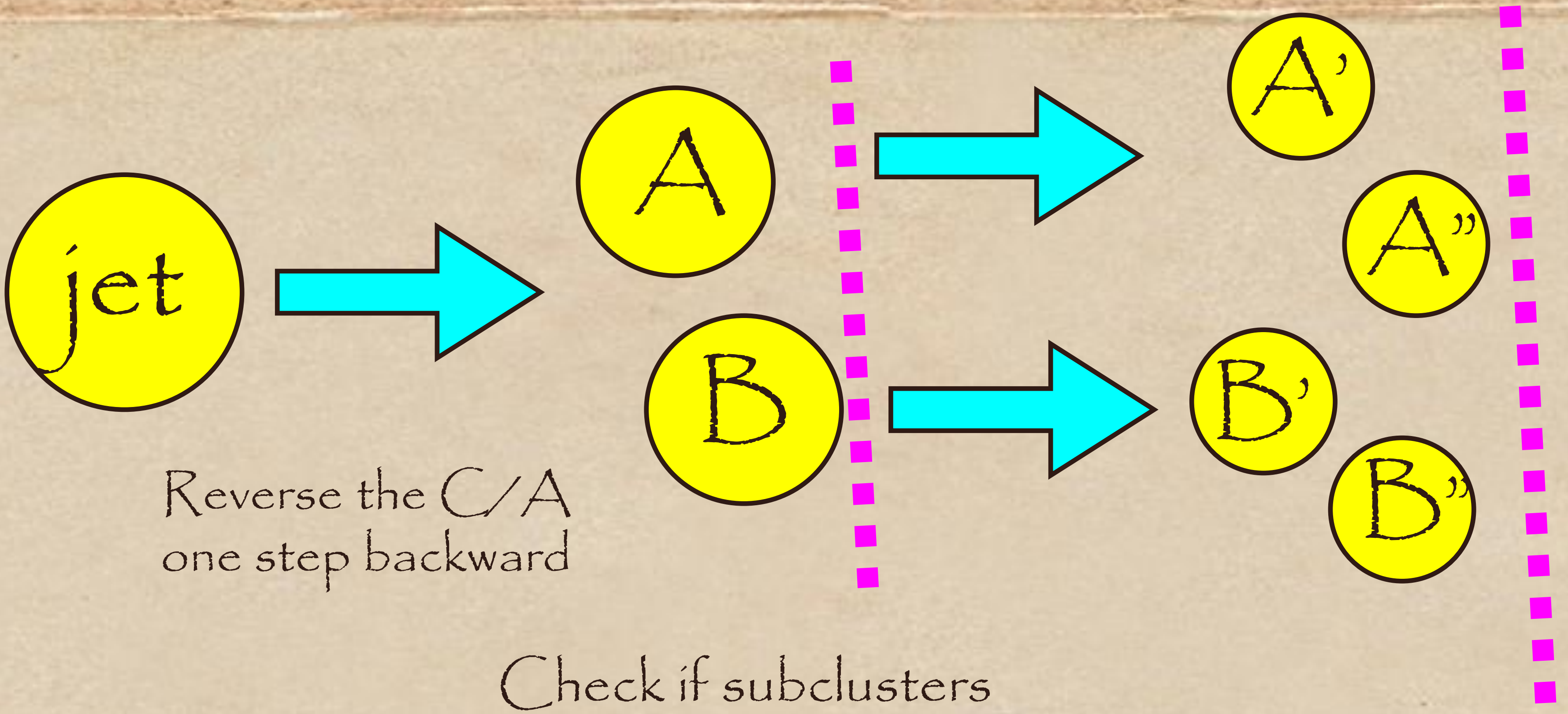
Apply C/A with R=0.8 to define the participating jets

Reverse the clustering process to find the sub-jets

- ★Subjets should:
- ★$p_T^{Subjet} > 0.05 * p_T^{Jet}$
- ★$\Delta R(1,2) > 0.4 - 0.0004 * p_T$
  - ★Iterate rejecting failed subjets and decluster again



Based on the Hopkins Algorithm (Kaplan, Rehermann, Schwartz, Tweedie) (arXiv:0806.0848)



Reverse the C/A
one step backward

Check if subclusters
pass the
requirements

Reverse again the
C/A one step
backward

Check again if
subsubclusters pass
the requirements

Pick up the (a',a'',b',b'')
(A,b',b'')
(A',A'',B) configuration for further study

Require:

- 3 or four subjets,
- minimal $m_{ij}>50$;
- subjets mass~m_{top} [100,250]:

Top Mass Reconstruction

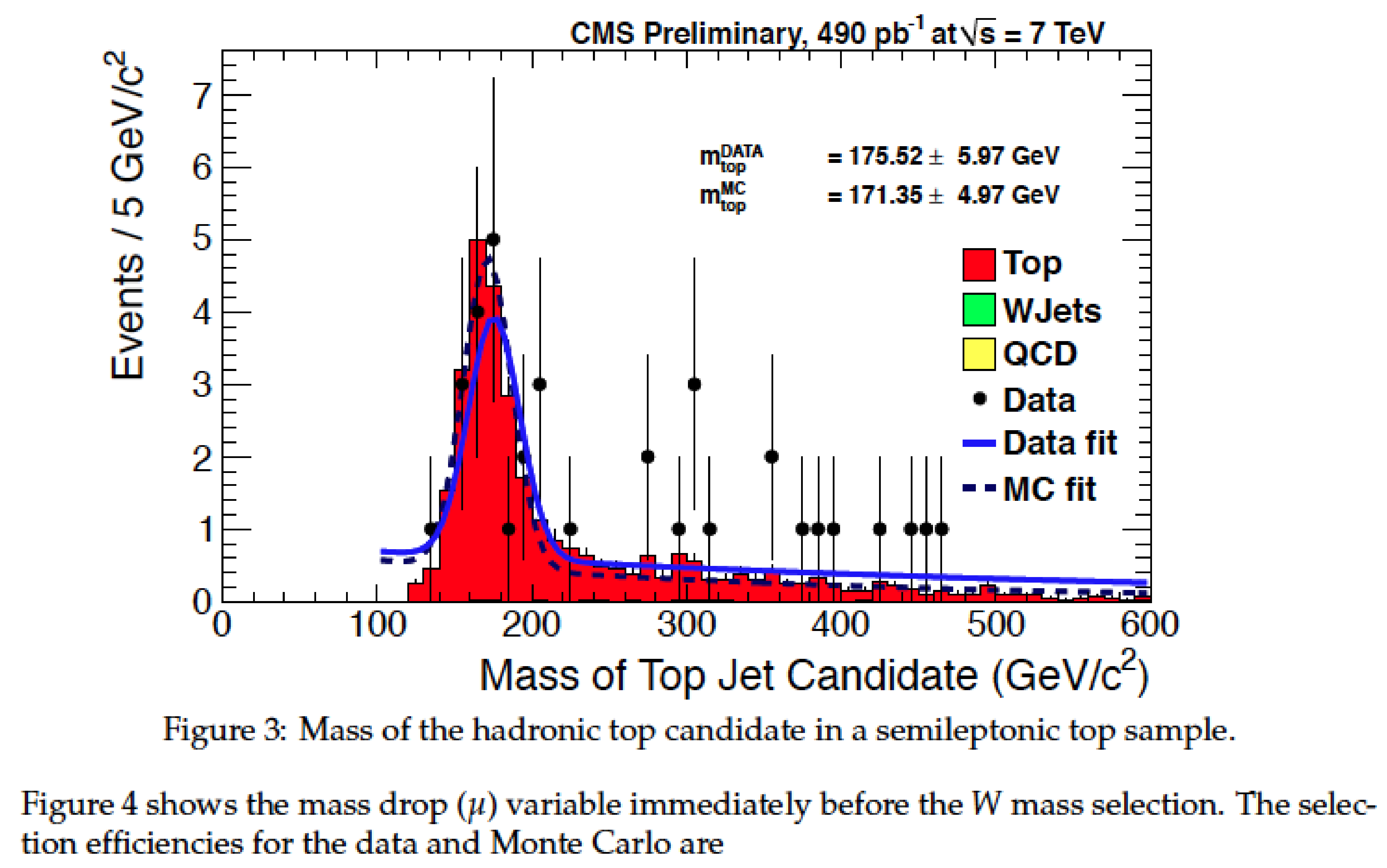


Figure 3: Mass of the hadronic top candidate in a semileptonic top sample.

Figure 4 shows the mass drop (μ) variable immediately before the W mass selection. The selection efficiencies for the data and Monte Carlo are

Top Mistag Rate

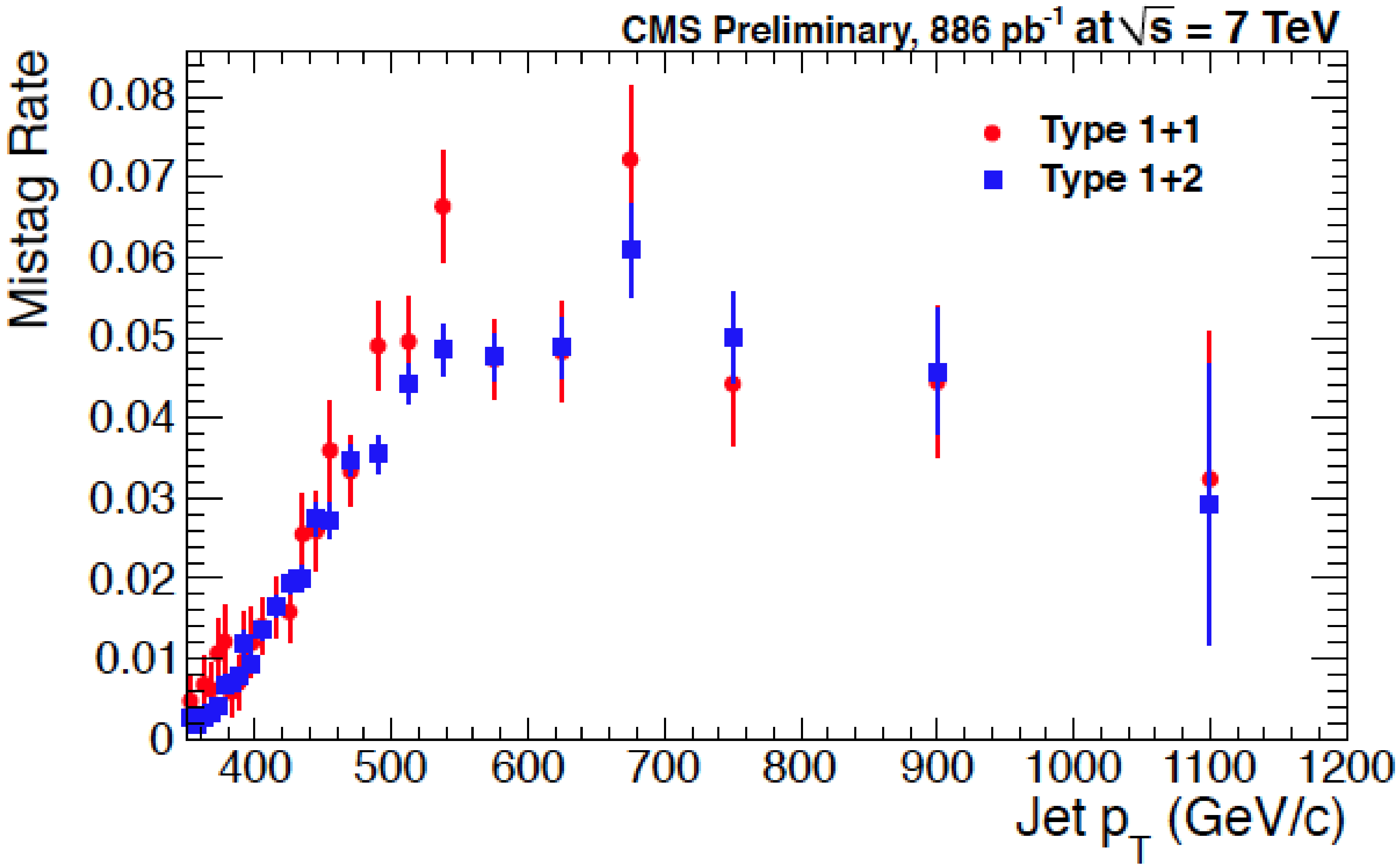


Figure 5: Top tagging mistag rate derived from dijet data (red circles) versus trijet data (blue squares), following the ‘anti-tag and probe’ procedure, as explained in the text. The rate derived from dijet data is applied to the “Type 1 + 1” analysis, whereas the rate derived from trijet data is applied to the “Type 1 + 2” analysis. There is a small (< 5%) contribution from continuum $t\bar{t}$ production that is removed, using the expectation from Monte Carlo.

Mass Bounds on Hypothetical tt Resonance

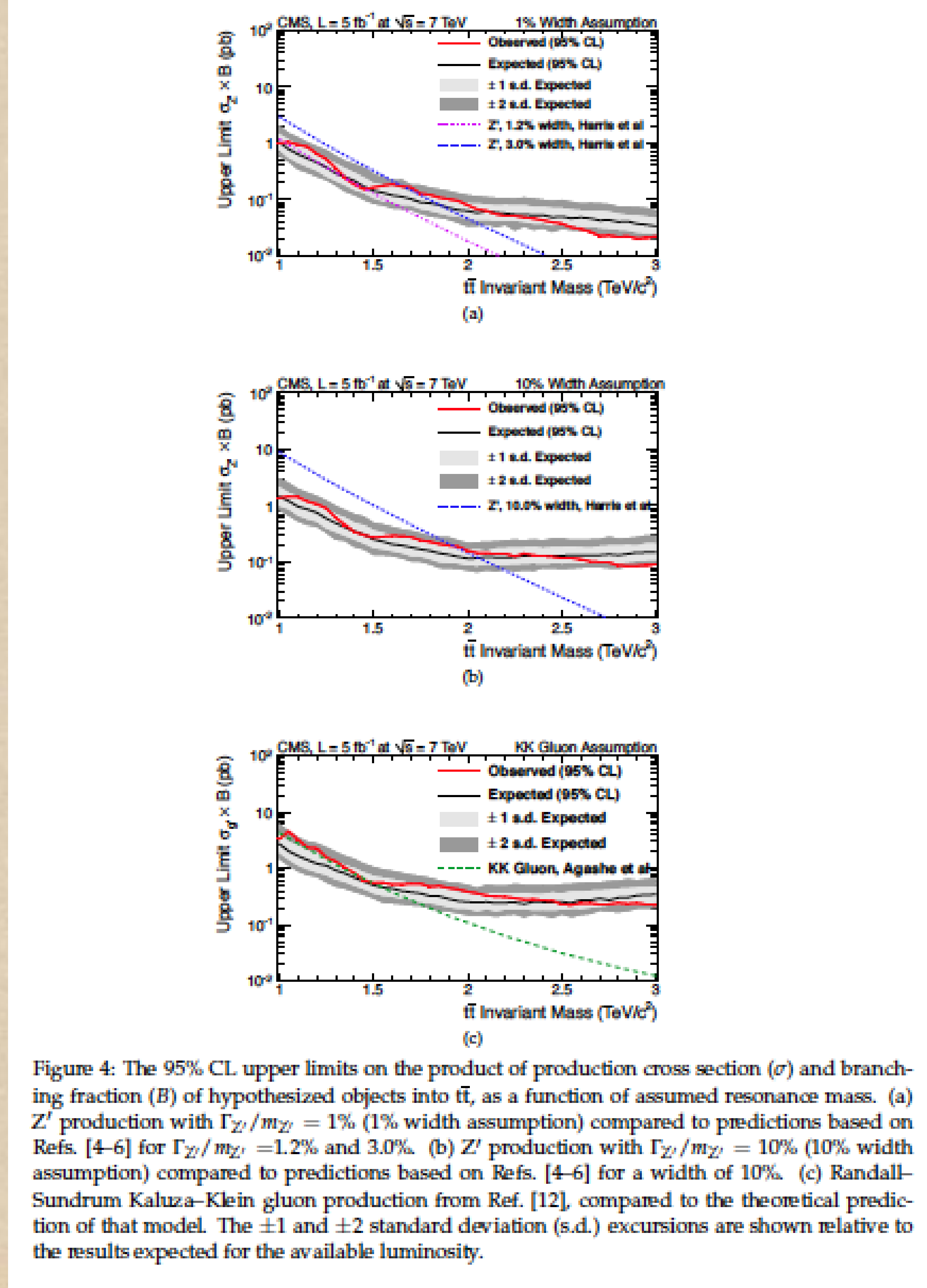
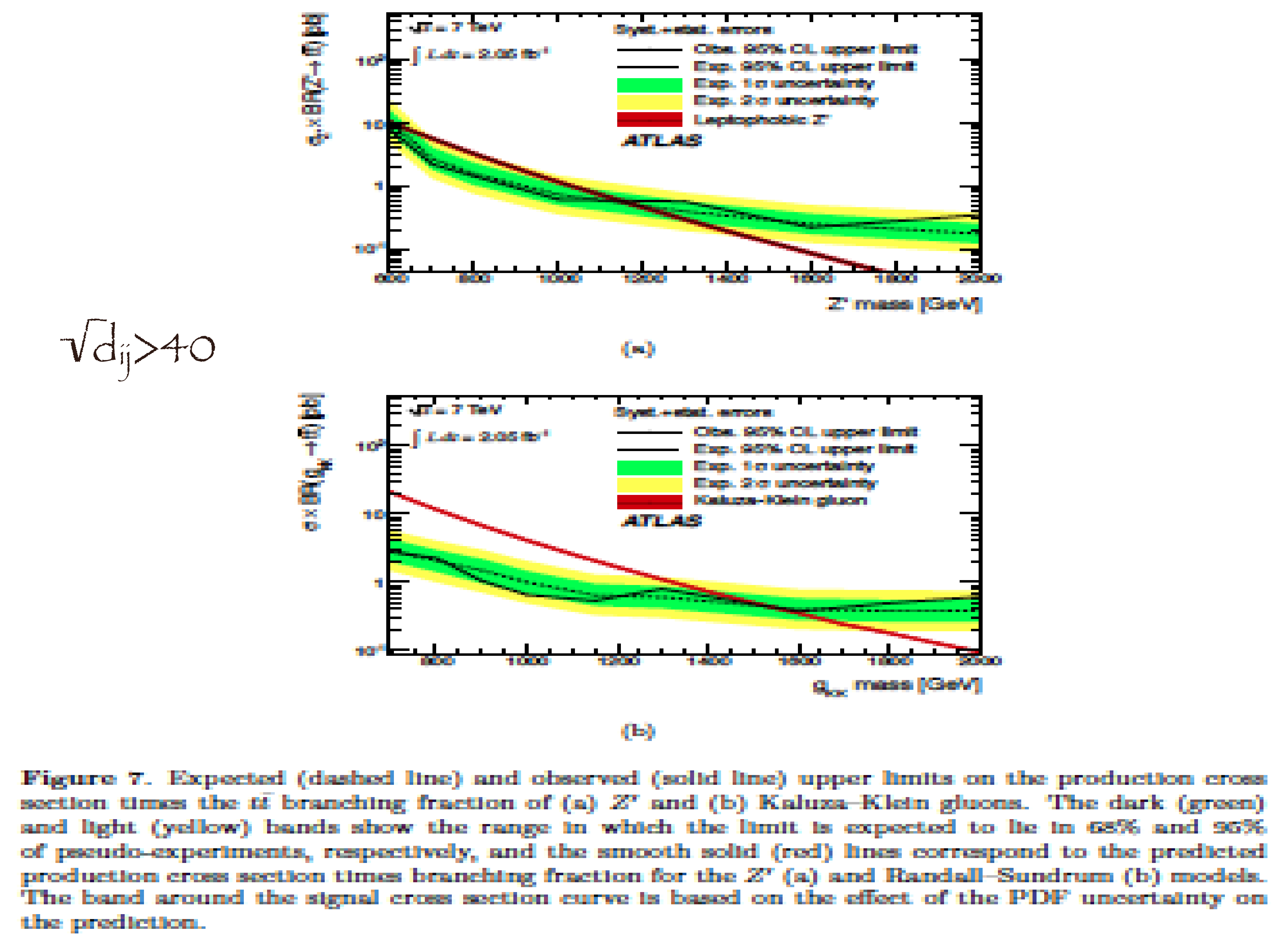


Figure 4: The 95% CL upper limits on the product of production cross section (σ) and branching fraction (B) of hypothesized objects into $t\bar{t}$, as a function of assumed resonance mass. (a) Z′ production with $\Gamma_{Z'}/m_{Z'} = 1\%$ (1% width assumption) compared to predictions based on Refs. [4–6] for $\Gamma_{Z'}/m_{Z'}$ =1.2% and 3.0%. (b) Z′ production with $\Gamma_{Z'}/m_{Z'} = 10\%$ (10% width assumption) compared to predictions based on Refs. [4–6] for a width of 10%. (c) Randall–Sundrum Kaluza–Klein gluon production from Ref. [12], compared to the theoretical prediction of that model. The ±1 and ±2 standard deviation (s.d.) excursions are shown relative to the results expected for the available luminosity.

Text

ATLAS Boosted tt Study



$\sqrt{d_{ij}}>40$

Figure 7. Expected (dashed line) and observed (solid line) upper limits on the production cross section times the $t\bar{t}$ branching fraction of (a) Z' and (b) Kaluza–Klein gluons. The dark (green) and light (yellow) bands show the range in which the limit is expected to lie in 68% and 95% of pseudo-experiments, respectively, and the smooth solid (red) lines correspond to the predicted production cross section times branching fraction for the Z' (a) and Randall–Sundrum (b) models. The band around the signal cross section curve is based on the effect of the PDF uncertainty on the prediction.

Conclusion

Complex issue

Need to correct for pileup and future looks,...

Excellent testing ground for pQCD -
predictions are already confronted with data

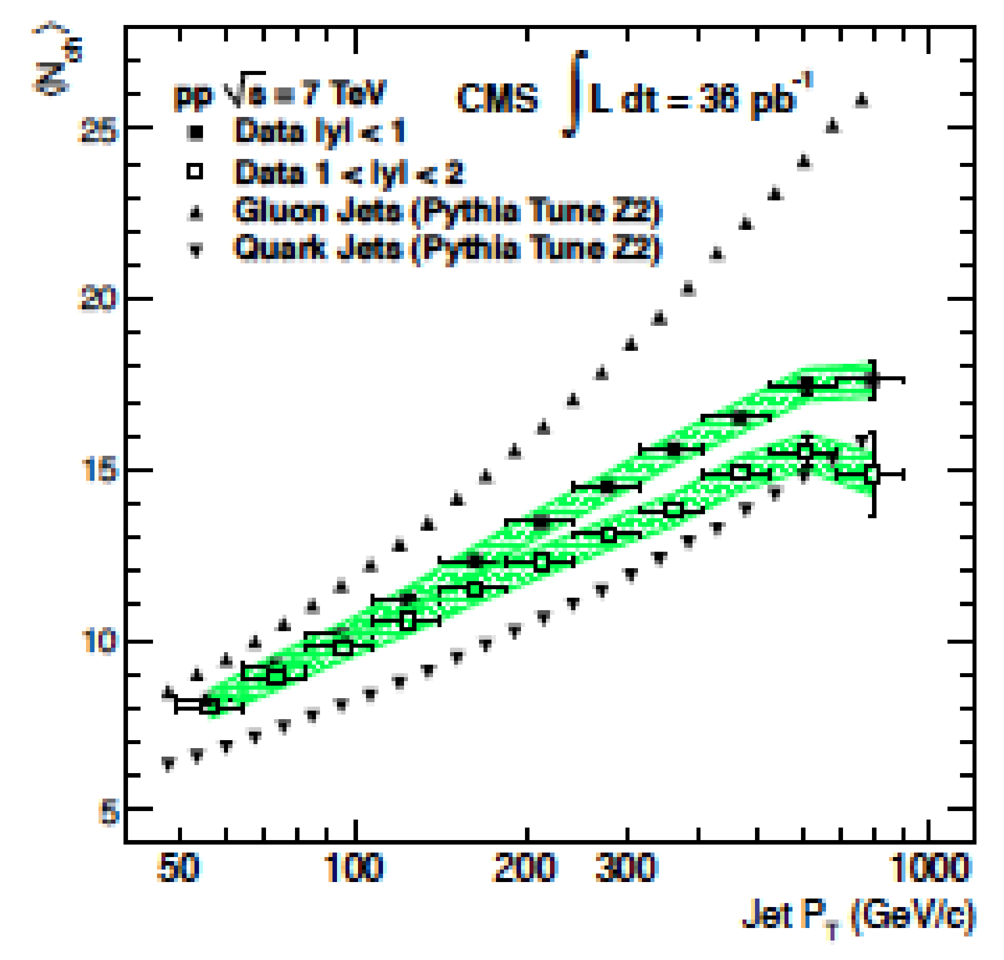
Already used to identify highly boosted top-quark initiated jets

May come handy in the search for Higgs via WH and ZH

Nice topic for talks

Backup

Track-Based Vs Calorimeter-Based



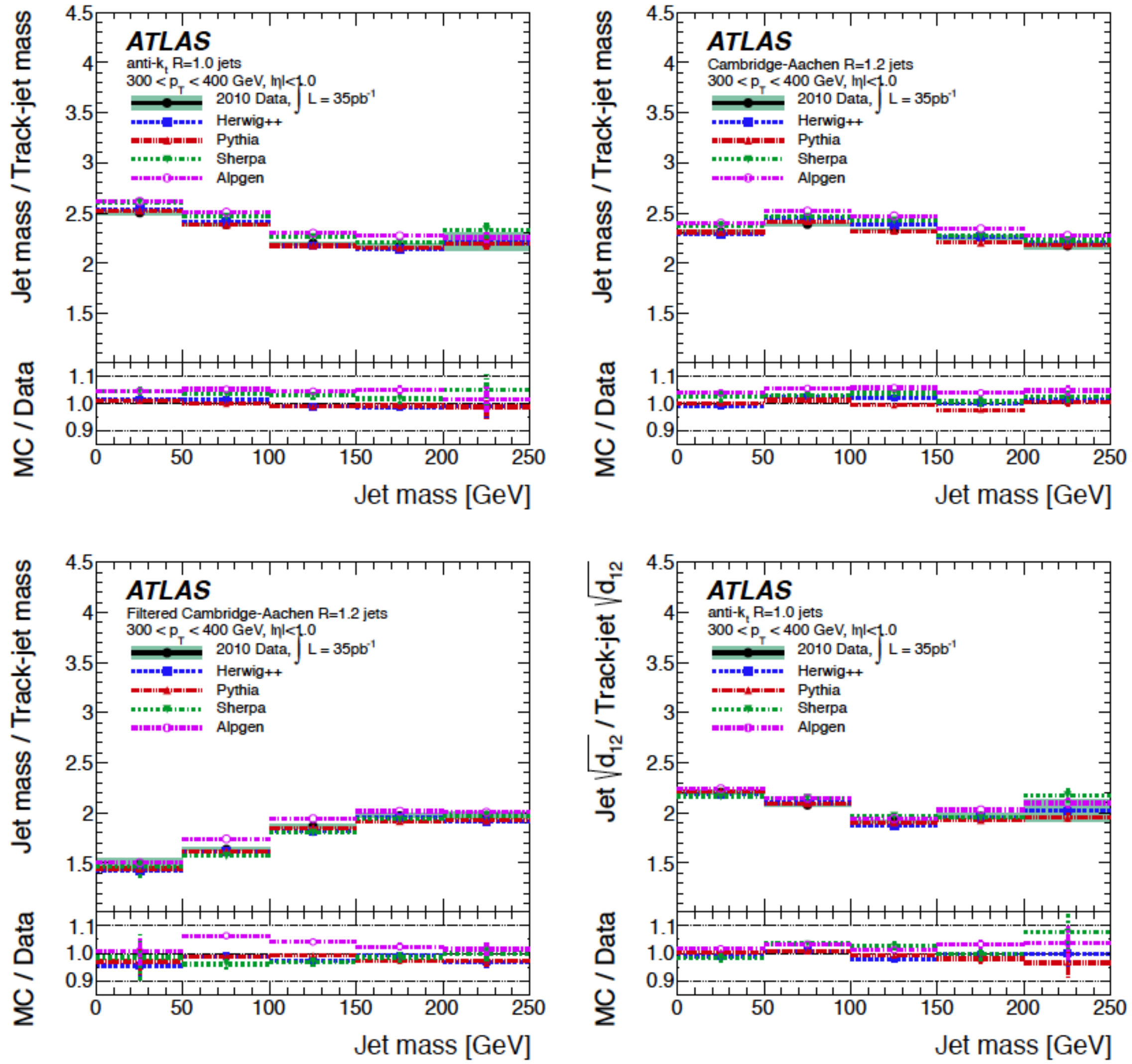
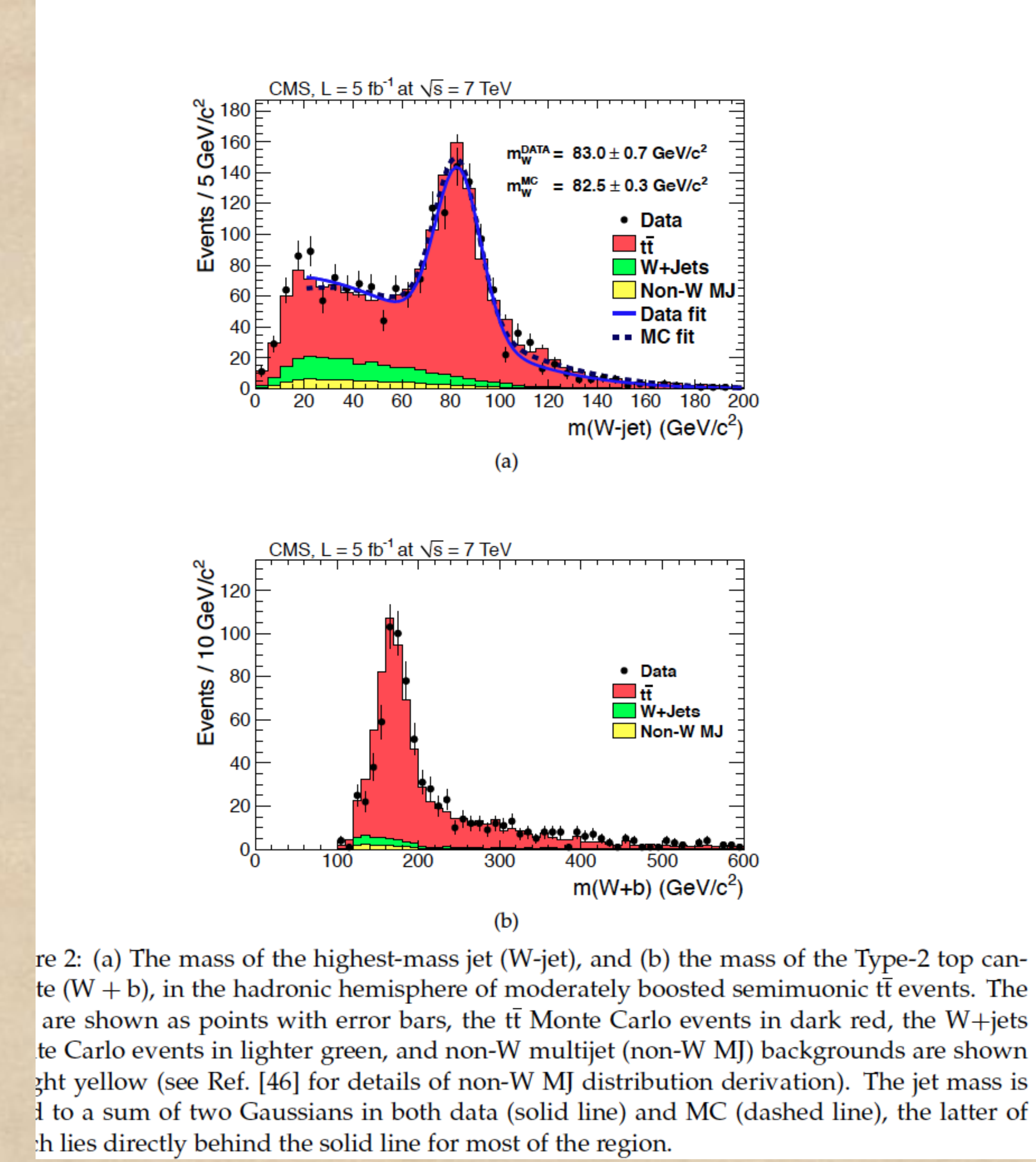


Figure 7. The ratio of a jet property determined by the calorimeter to that determined by tracks versus the calorimeter jet mass for jets with 300–400 GeV in p_T. Shown are the data and a variety of Monte Carlo models. The bottom frame shows the ratio of the Monte Carlo models to data. The top left, top right and bottom left figures show the ratio for jet mass for three different jet algorithms. The bottom-right figure shows the ratios for $\sqrt{d_{12}}$ in anti-k_t jets.

4606

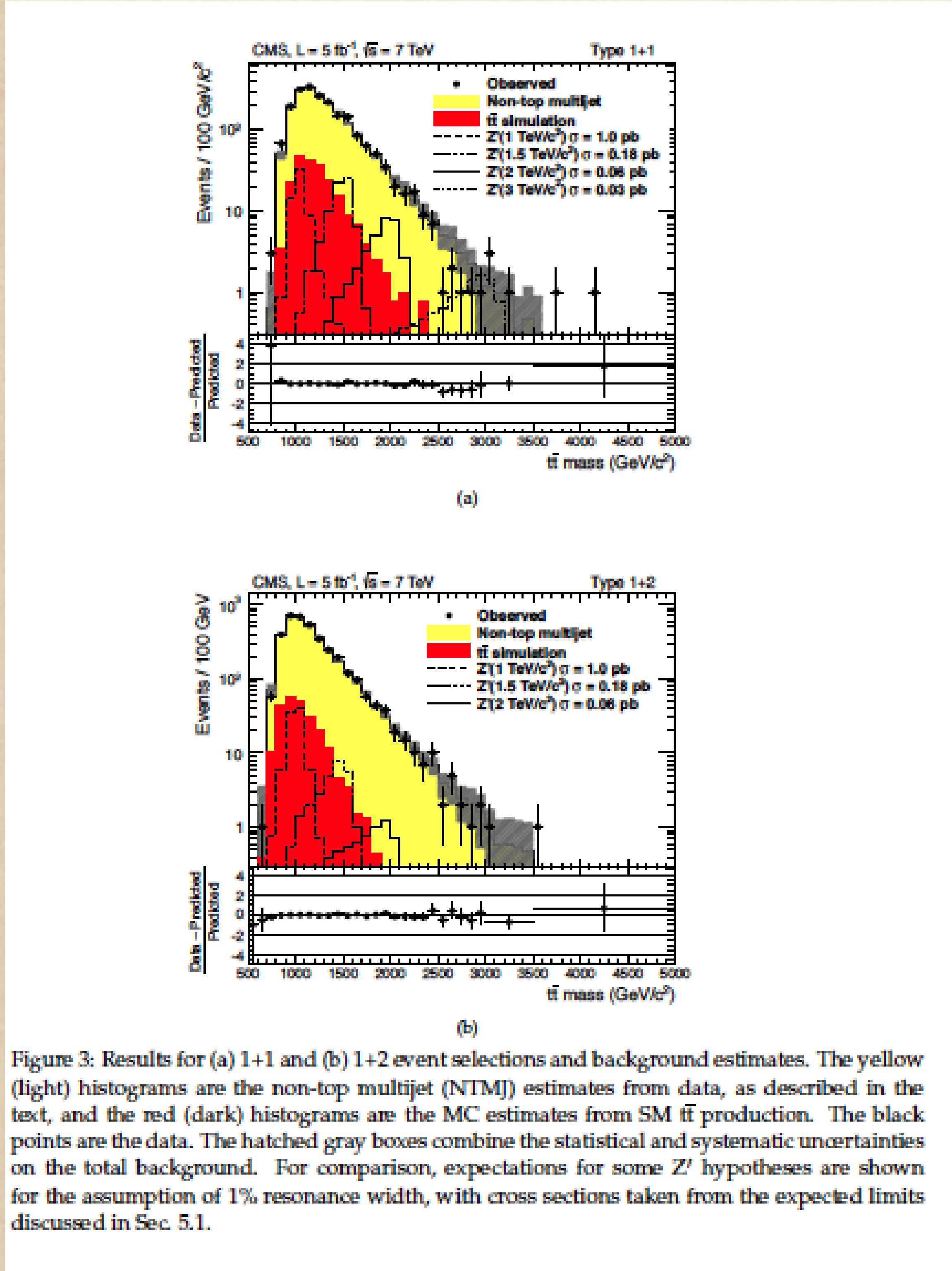
Boosted Top



re 2: (a) The mass of the highest-mass jet (W-jet), and (b) the mass of the Type-2 top can-
te (W + b), in the hadronic hemisphere of moderately boosted semimuonic $t\bar{t}$ events. The
are shown as points with error bars, the $t\bar{t}$ Monte Carlo events in dark red, the W+jets
te Carlo events in lighter green, and non-W multijet (non-W MJ) backgrounds are shown
ght yellow (see Ref. [46] for details of non-W MJ distribution derivation). The jet mass is
l to a sum of two Gaussians in both data (solid line) and MC (dashed line), the latter of
h lies directly behind the solid line for most of the region.

Text

tt Resonance



(a)

(b)

Figure 3: Results for (a) 1+1 and (b) 1+2 event selections and background estimates. The yellow (light) histograms are the non-top multijet (NTMJ) estimates from data, as described in the text, and the red (dark) histograms are the MC estimates from SM $t\bar{t}$ production. The black points are the data. The hatched gray boxes combine the statistical and systematic uncertainties on the total background. For comparison, expectations for some Z' hypotheses are shown for the assumption of 1% resonance width, with cross sections taken from the expected limits discussed in Sec. 5.1.

Text

Jet Shape

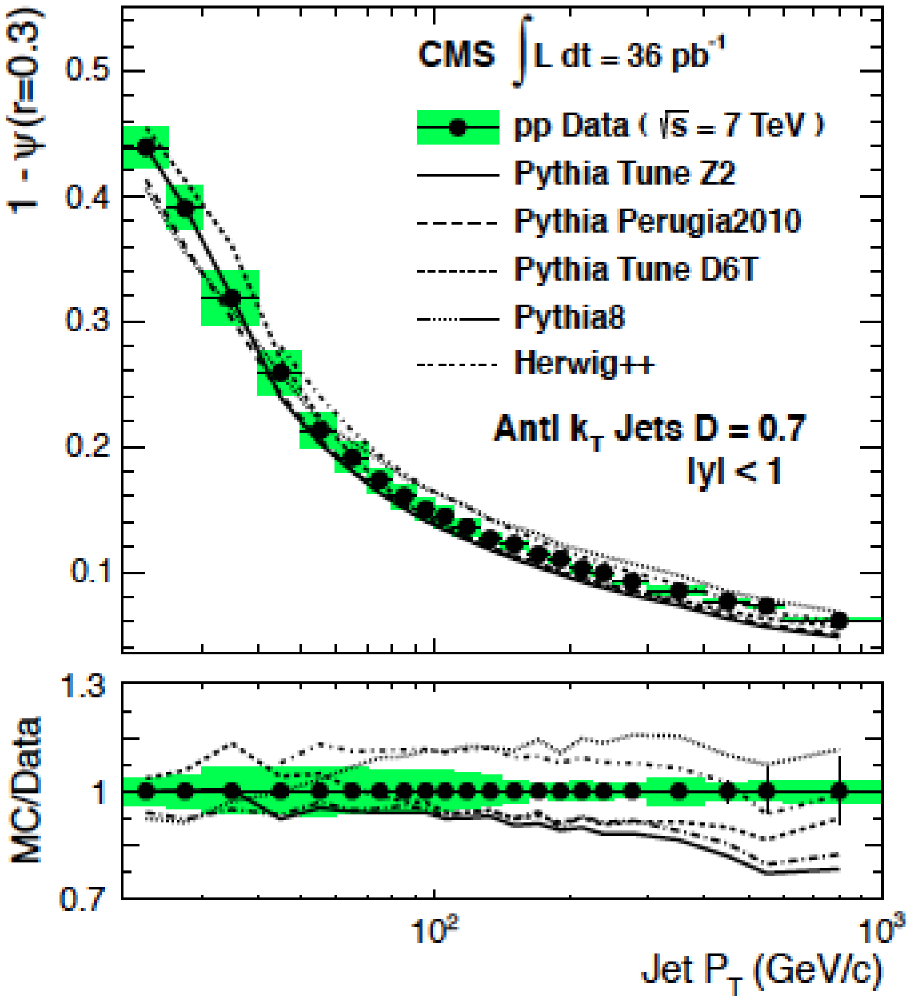
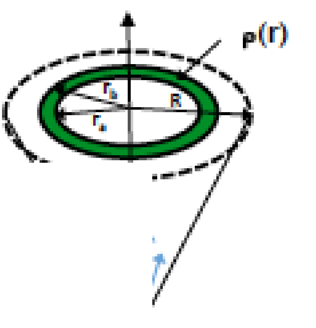
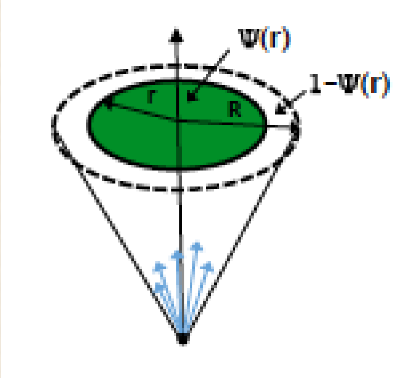


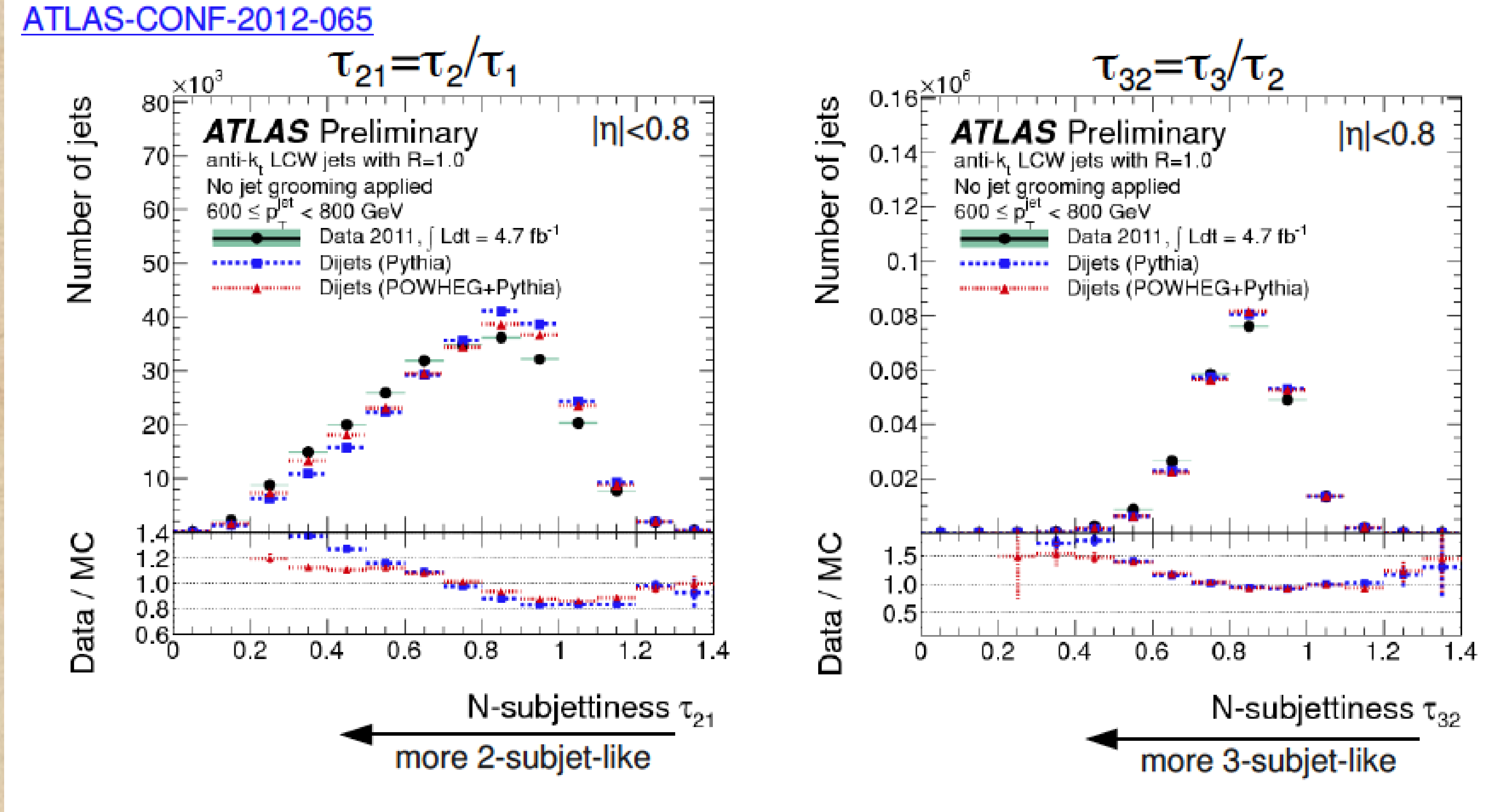
Figure 4: Measured integrated jet shape, $1 - \Psi(r = 0.3)$, as a function of jet p_T in the central rapidity region $|y| < 1$, compared to HERWIG++, PYTHIA8, and PYTHIA6 predictions with various tunes. Statistical uncertainties are shown as uncertainties on the data points and the shaded region represents the total systematic uncertainty of the measurement. Data points are placed at the bin centre; the horizontal bars show the size of the bin. The ratio of each MC prediction to the data is also shown in the lower part of each plot.





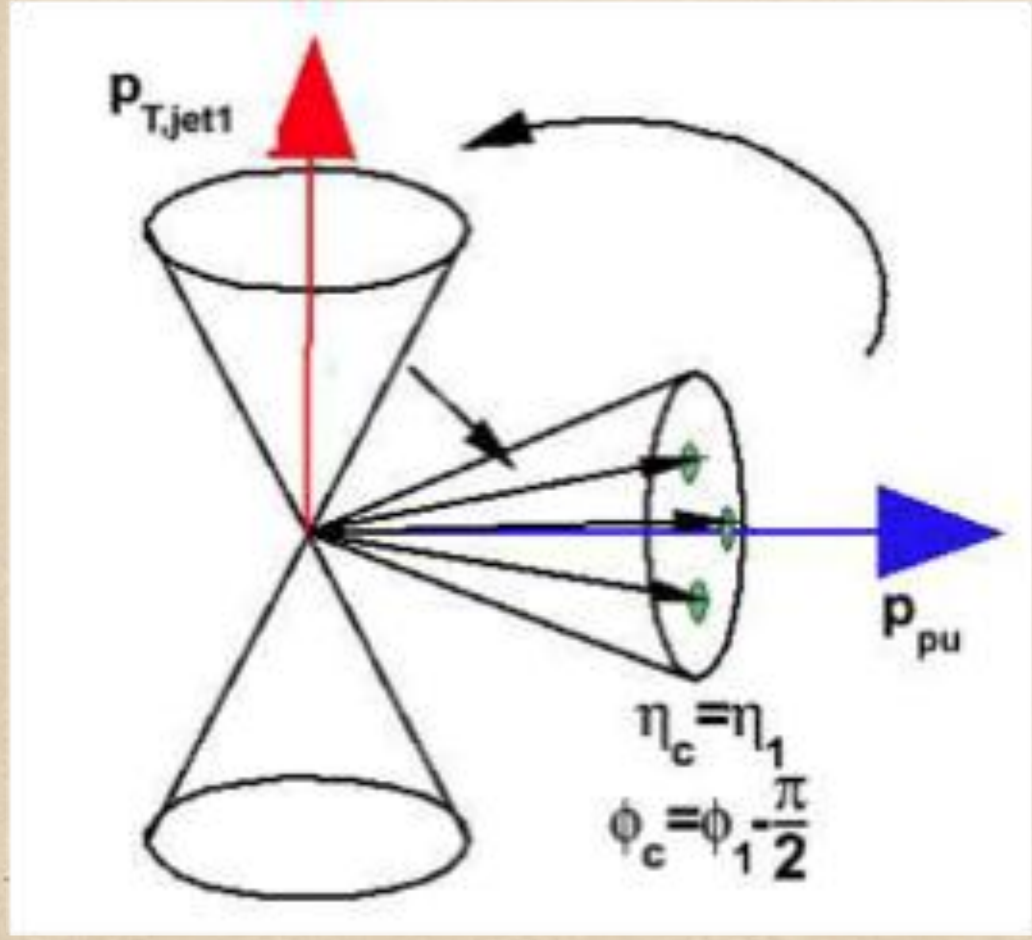
Text

N-Subjettiness of Inclusive QCD jets


ATLAS-CONF-2012-065

Text

Complementary Cone Technique



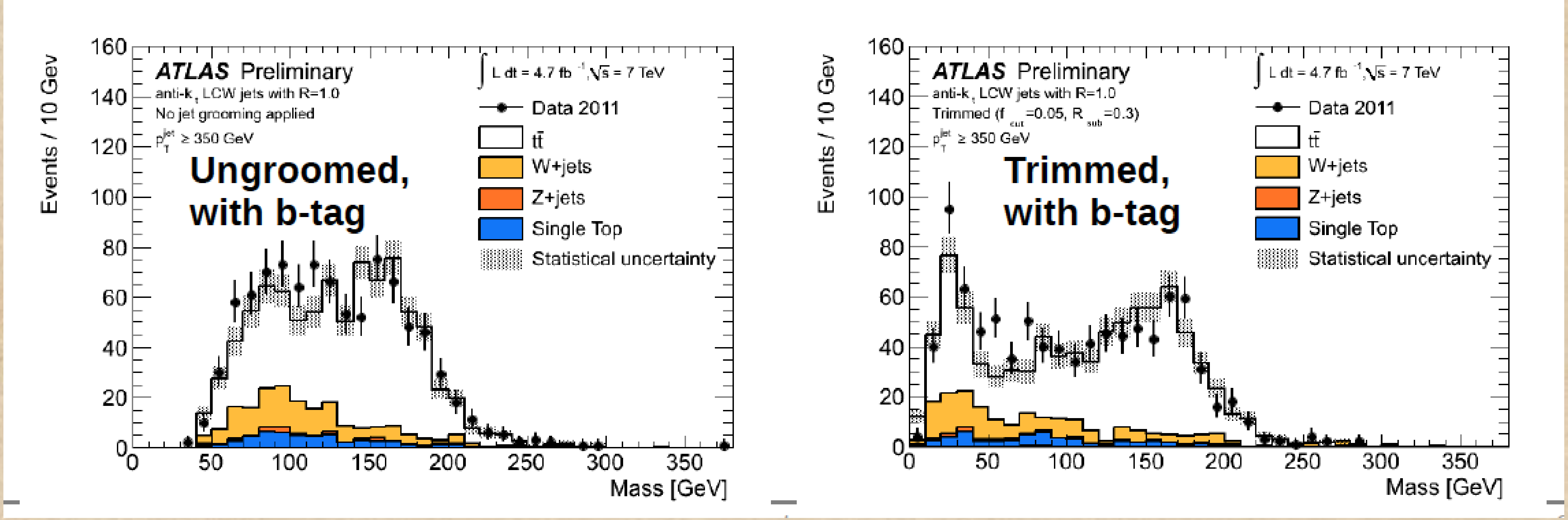
- PRD 84 (2011) 114025
  (R.A., Duchovni, Perez, Pranko, Sinrevo)
- ATL-COM-PHYS-2011-1662
  (Trisha Farooque, University of Toronto)

CMS W-Tagger

Use “pruned” jets.
The total mass of the jet is inside [60,100]
Last Mass Drop is 0.4
2 last jets should be roughly equal in pt

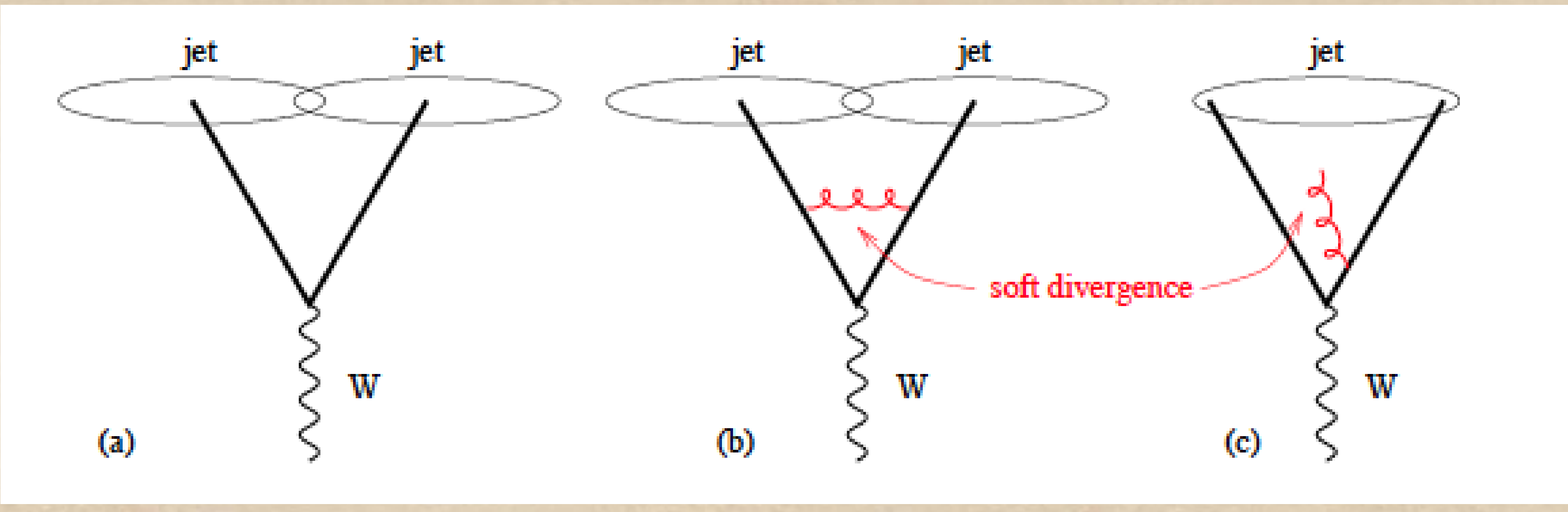
Looking for Boosted Top

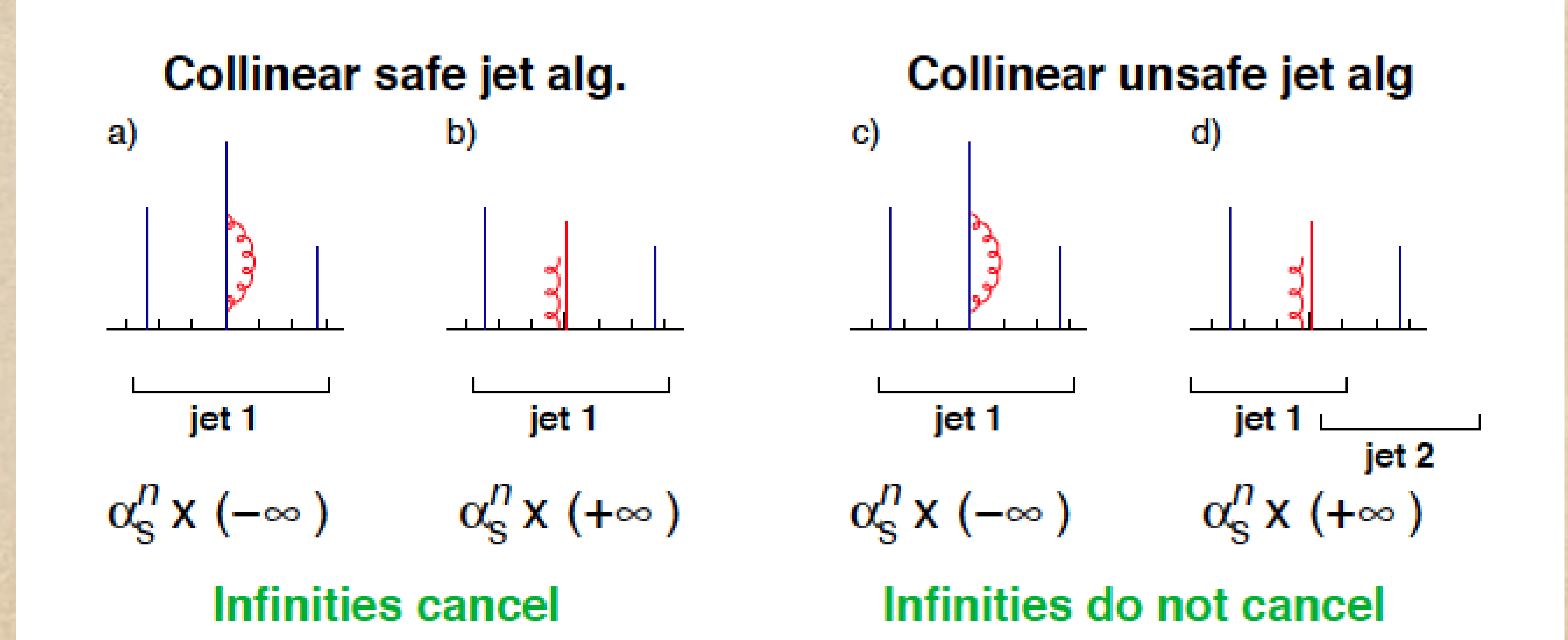
Jet Mass (leading pt=350, anti-kt R=1.0 with b-tag)



Rediscovering the top in its hadronic decay mode

Infra-red and co-linear safety





Collinear safe jet alg.

$\alpha_S^n \times (-\infty)$ $\alpha_S^n \times (+\infty)$

Infinities cancel

Collinear unsafe jet alg

$\alpha_S^n \times (-\infty)$ $\alpha_S^n \times (+\infty)$

Infinities do not cancel