Results Using Jet Substructure in ATLAS

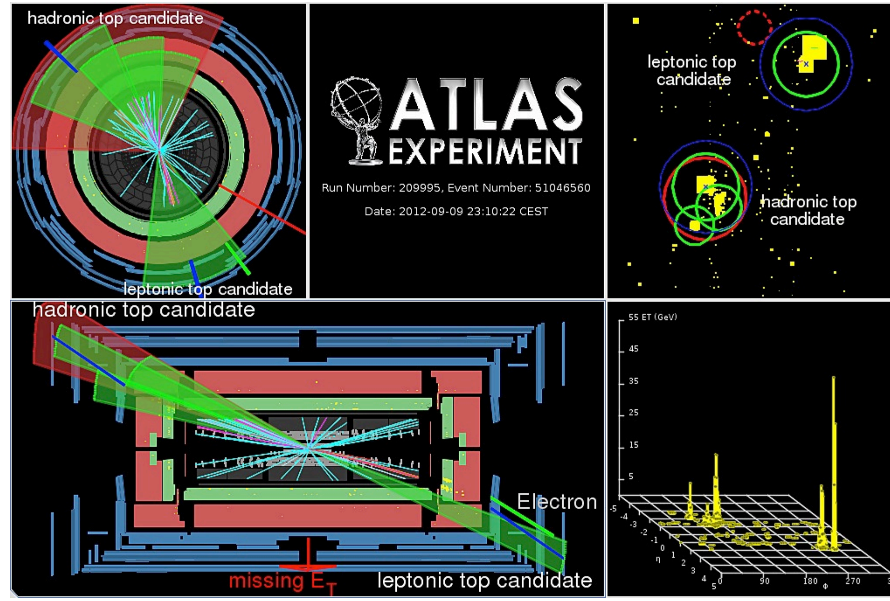
QCD@LHC 2013

Frank Merritt
University of Chicago
September 4, 2013

The two primary goals of substructure analysis

Substructure

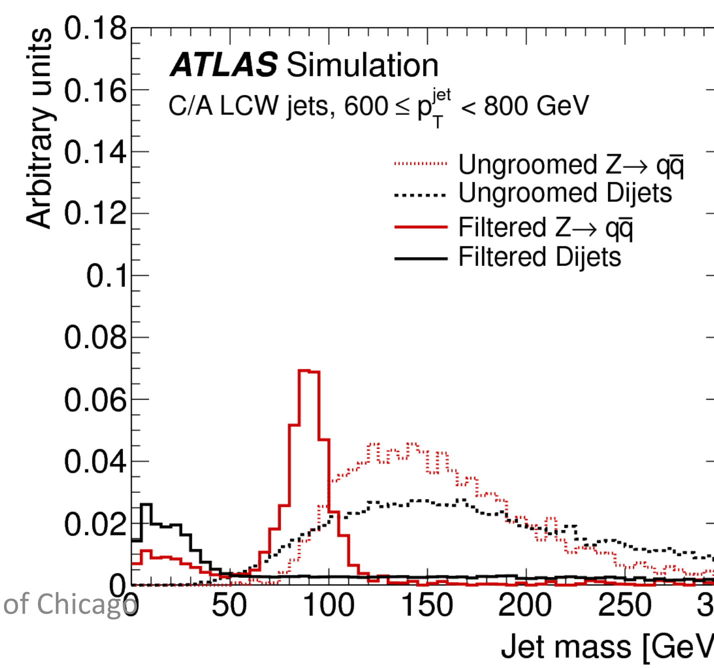
Identifying and analyzing the subjets of highly boosted systems in order to identify the constituent decay products.



R6: fig13a

Jet grooming

Improving resolution (especially of mass) of highly boosted systems



R1: fig29b

Substructure variables

- **Jet mass:** Defined from the 4-vector sum of the constituents of the jet. Mass of large-R jet is critical; masses of subjets are important both for substructure analysis and for filtering.

- **Kt splitting scales:** Defined by reclustering the constituents using the kt (or C/A) clustering algorithm. At each step,

$$\sqrt{d_{ij}} = \min(p_{Ti}, p_{Tj}) \times \Delta R_{ij} \approx m_P / 2$$

Backgrounds are more likely to give small values of $\sqrt{d_{12}}$ (corresponding to an apparently highly asymmetric decay).

- **N-subjettiness:** This approximately measures the reduction factor in subjet size at each step of the declustering. For N subjets,

$$\tau_N = \frac{1}{d_0} \sum_k p_{Tk} \times \min(\delta R_{1k}, \delta R_{2k}, \ldots, \delta R_{Nk}) \quad \text{where} \quad d_0 = \sum_k p_{Tk} \times R$$

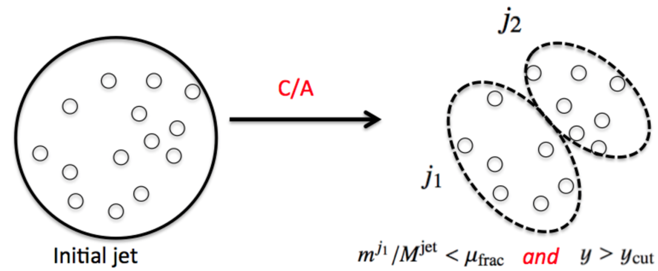
where R is the size parameter of the jet algorithm. The changes in τ_i during the first few steps of declustering are especially important for determining substructure. These are defined by

$$\tau_{12} = \tau_2 / \tau_1 \quad \text{and} \quad \tau_{32} = \tau_3 / \tau_2$$

A small value of τ_{ij} indicates that this step gives a significant improvement in representing the substructure. A value near 1 indicates little improvement.

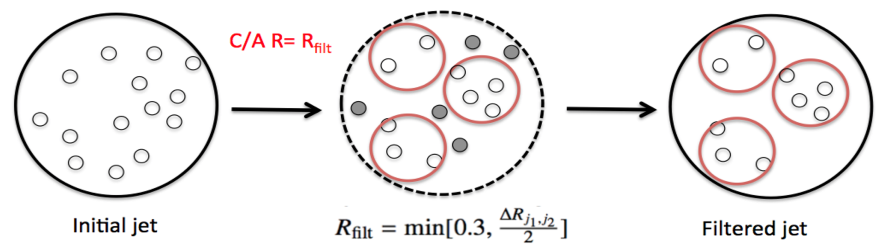
Jet Substructure techniques

Mass-drop and symmetry: Iteratively undo last stage of kt clustering until $\max(m_{j1},m_{j2})<\mu\bullet M_{j1+j2}$ with $\mu=0.67$. Also require $[\min(p_T^{j1}, p_T^{j1})xR_{12}/m_{jet}]^2>y_{cut}\approx0.09$. Discard remaining constituents.

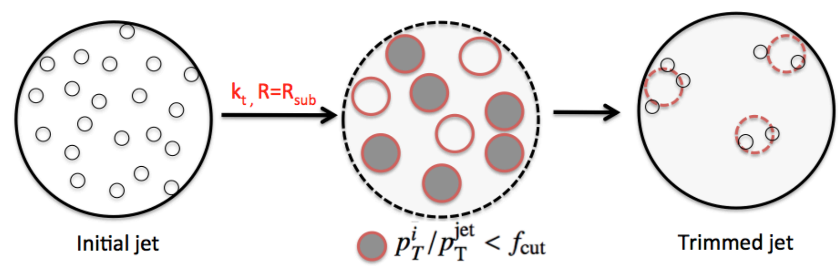


R1: fig 3+4

Filtering: Recluster constituents j_1, j_2 with C/A, $R=\min[0.3, \Delta R_{j1j2}]$. Drop all constituents outside the 3 hardest.



Trimming: Recluster constituents with kt to create subjets of size R_{sub}=0.2 . Remove any with $p_{Ti}<p_T^{jet}\bullet f_{cut}$ with f_{cut} ≈0.03.



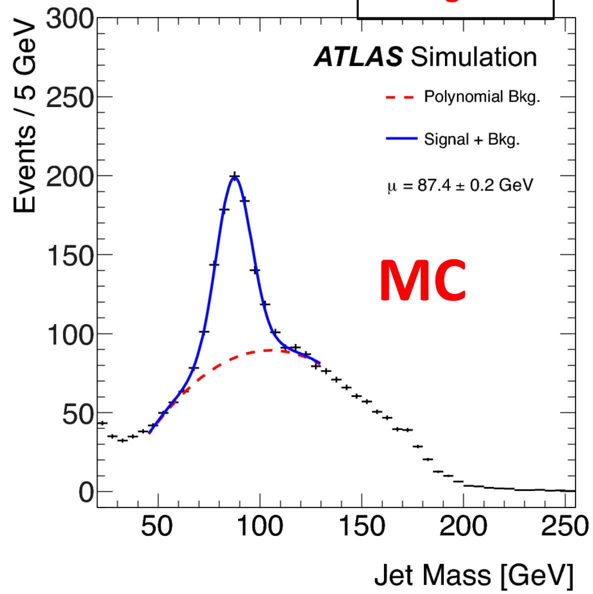
(Specific choices of parameters will vary with kinematics and specific final state.)

Verification of mass scale with hadronic W decays from top events

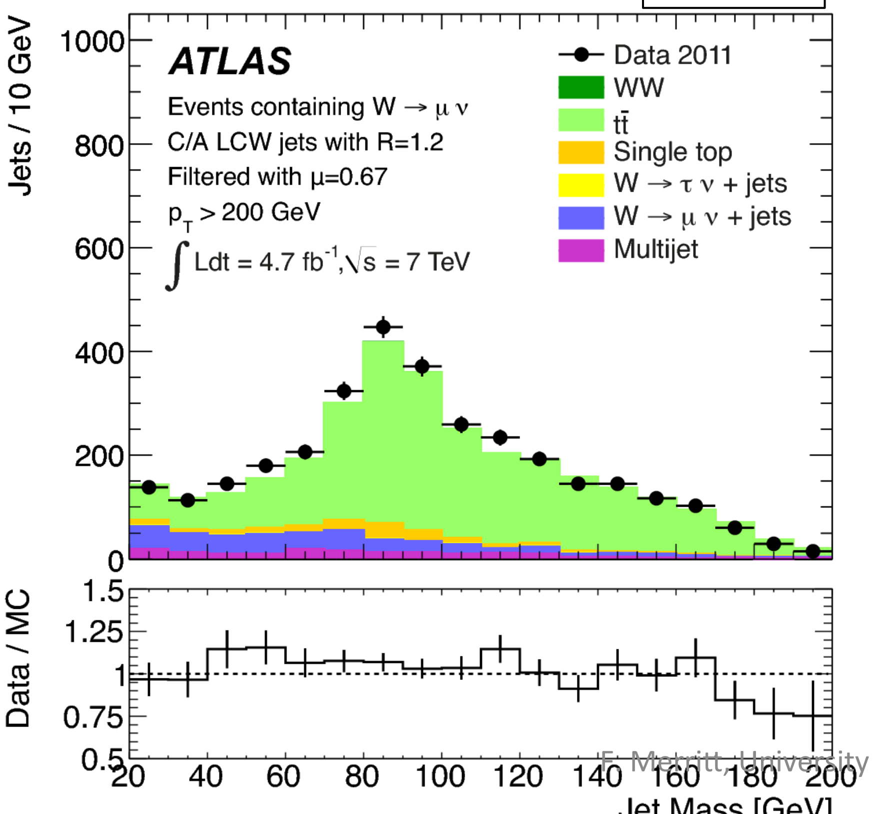
Below: Comparison of reconstructed mass (data vs MC) of hadronically decaying Ws from t-tbar events. Due to low pt threshold, typically only 2 subjets are contained.

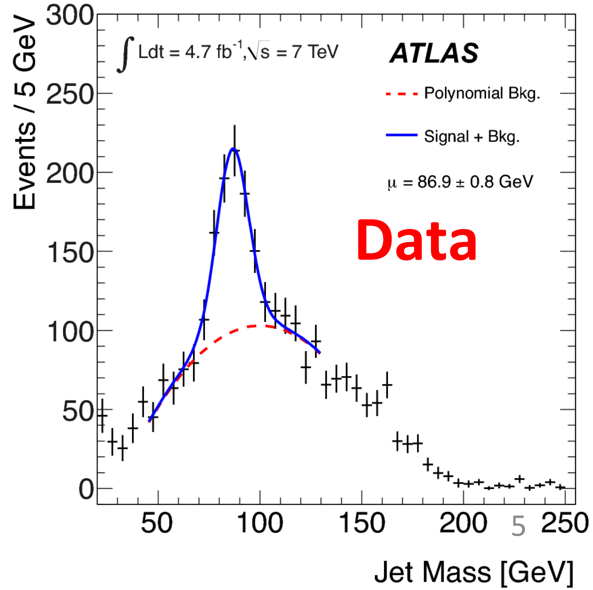
Right: The figures compare mass fits to data and MC. The ratio of the two fits agree to within better than 1%.

R1: fig 12



R1: fig 11a



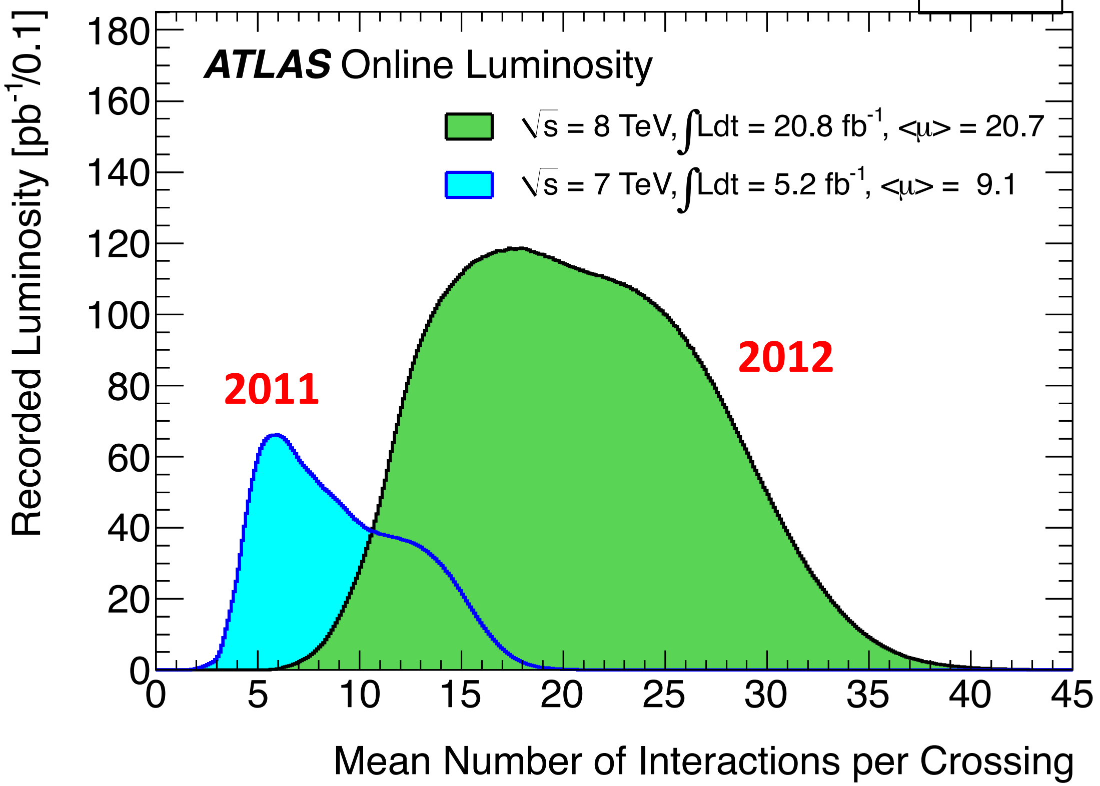


Pile-up in Run I

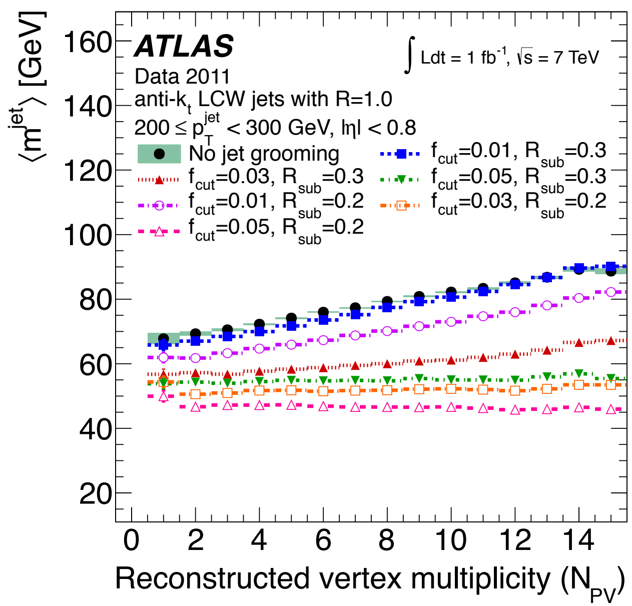
Pile-up is expected to **double** or **triple** in Run II., compared to 2012. This greatly distorts the reconstructed mass and substructure parameters, especially for large-R jets.

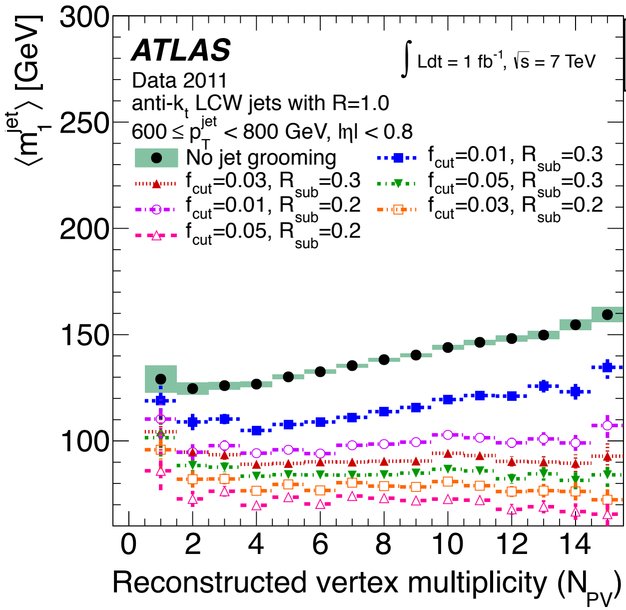
Grooming techniques have been developed (and are being improved) to restore the correct parameters and masses of the large-R jet and its subjets.

R2: fig 1

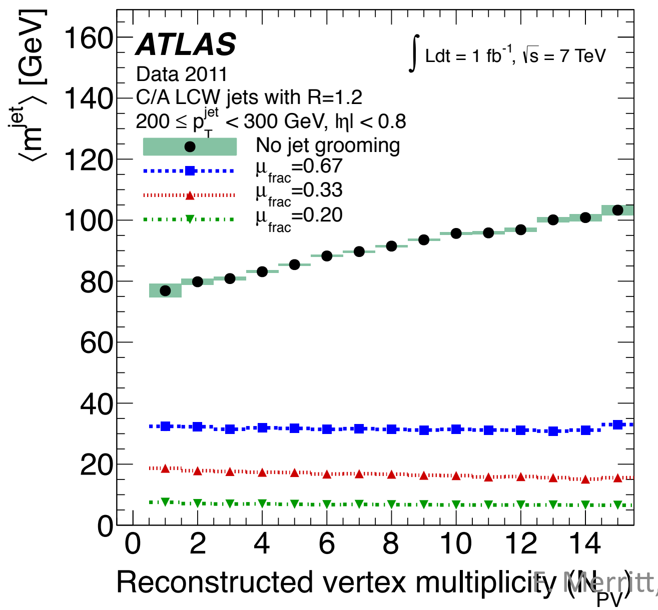


Effect of pile-up and grooming as a function of N_{PV}





R1:fig 18

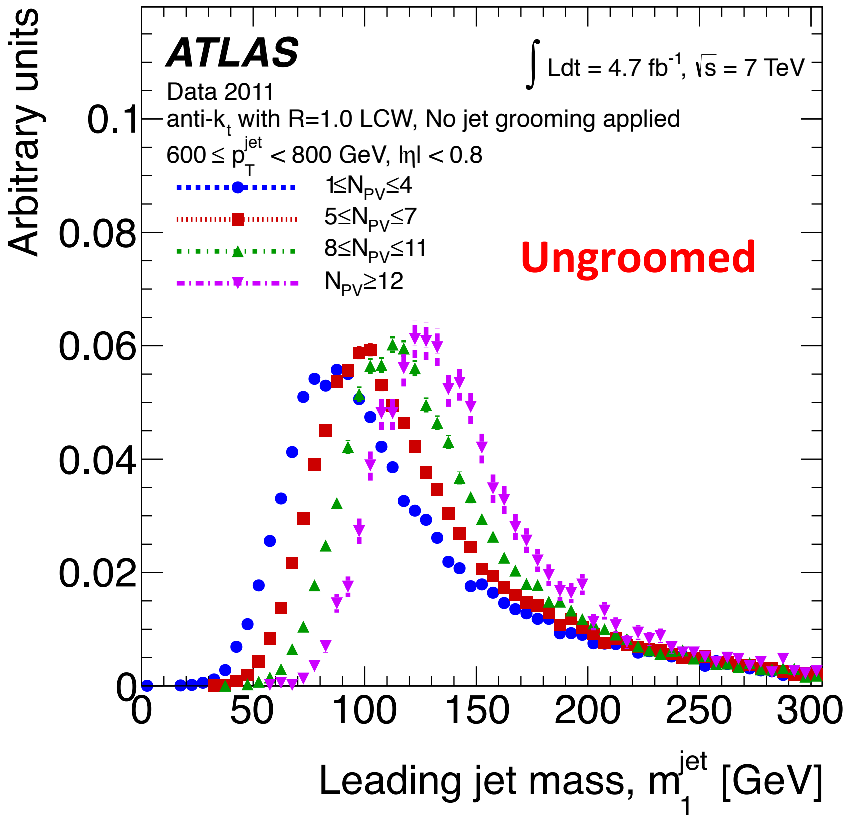


Top: mass of trimmed anti-kt jets with R=1.0 in the momentum ranges 200-300 (left) 600-800 (right) GeV as a function of N_{PV} for various trimming parameters. The optimal parameters (lower green curve) are $\mathbf{f_{cut}=0.05, R_{sub}=0.3}$.

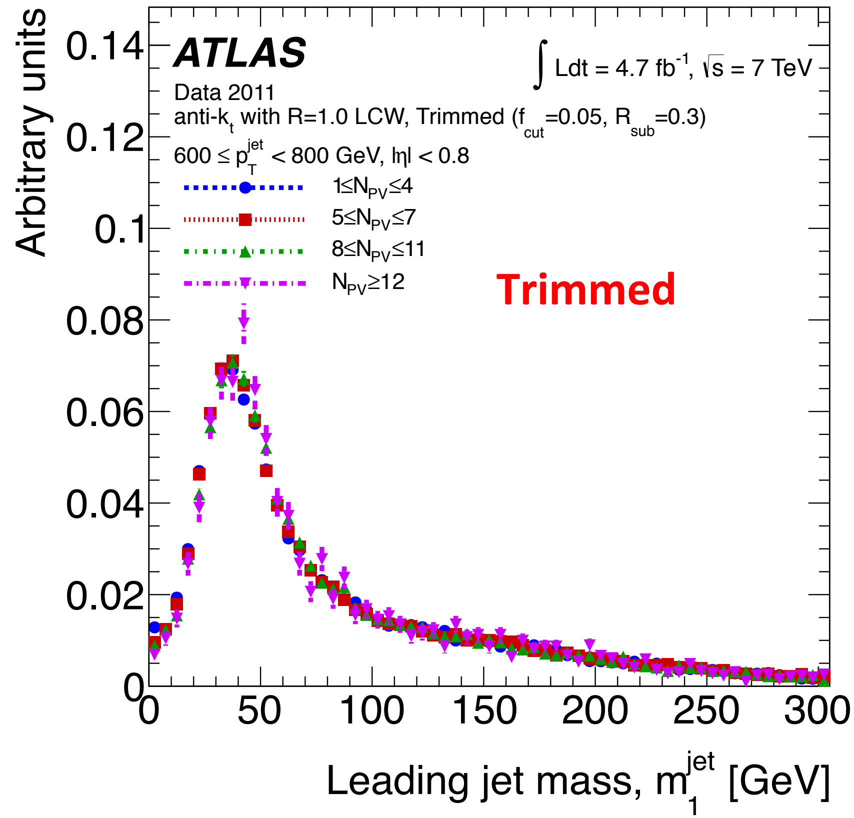
Left: a similar plot for mass-drop filtered C/A jets with R=1.2 .

Effect of pile-up and grooming on full mass spectrum

R1:fig. 20a



R1:fig 20b



Beyond providing a pile-up independent average jet mass, the optimal grooming configurations render the full jet mass spectrum insensitive to high luminosity. The figures show, for 4 different ranges of NPV, the effect of pile-up on ungroomed jets (left) and trimmed jets (right) .

Boosted Top Pt Spectrum (2012 data)

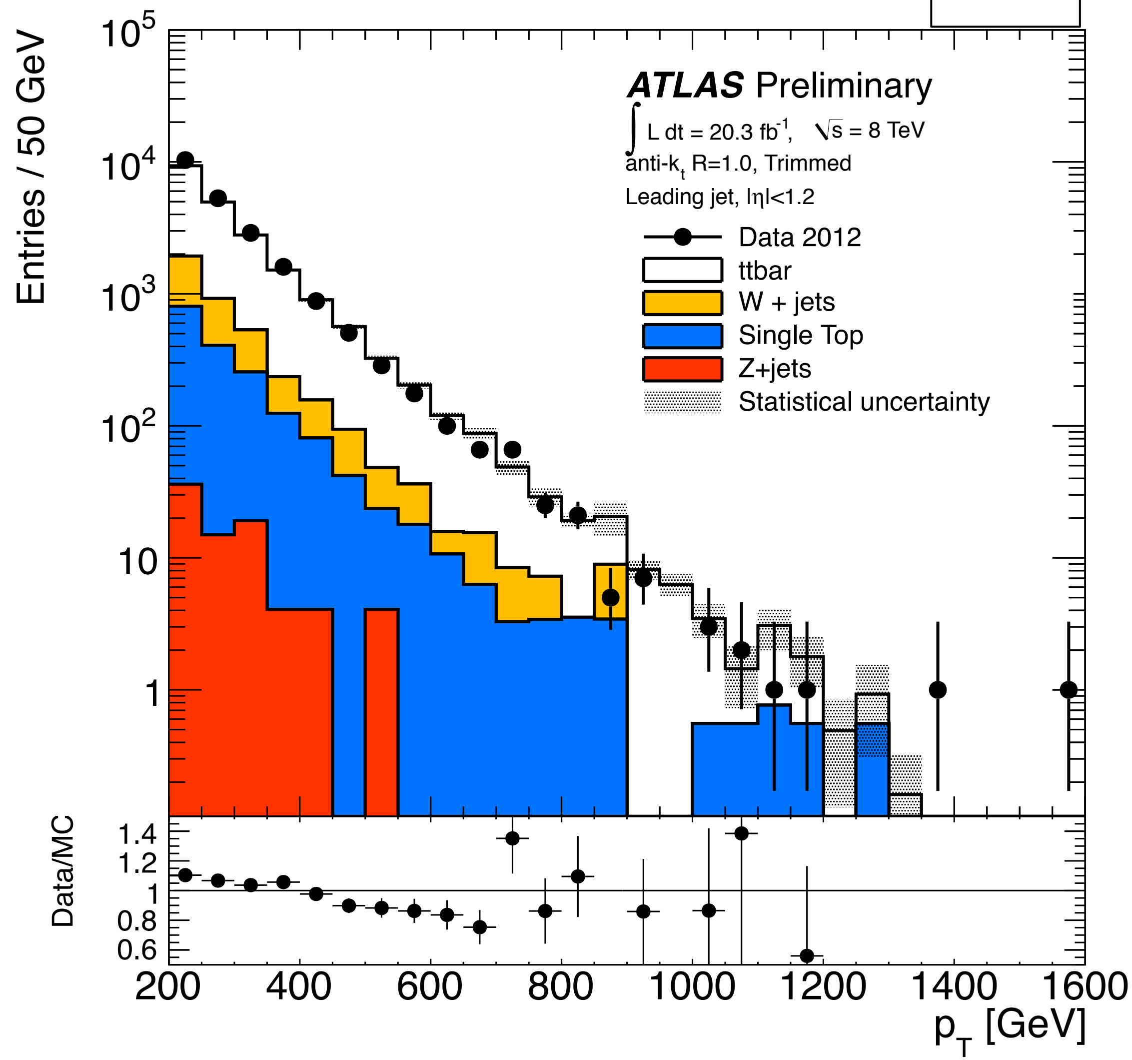
R3: fig1

Highly boosted top events have provided one of the strongest motivations, and one of the best test-beds, for substructure analysis.

Over 5000 t-tbar events with p_T>200 GeV were collected during 2012.

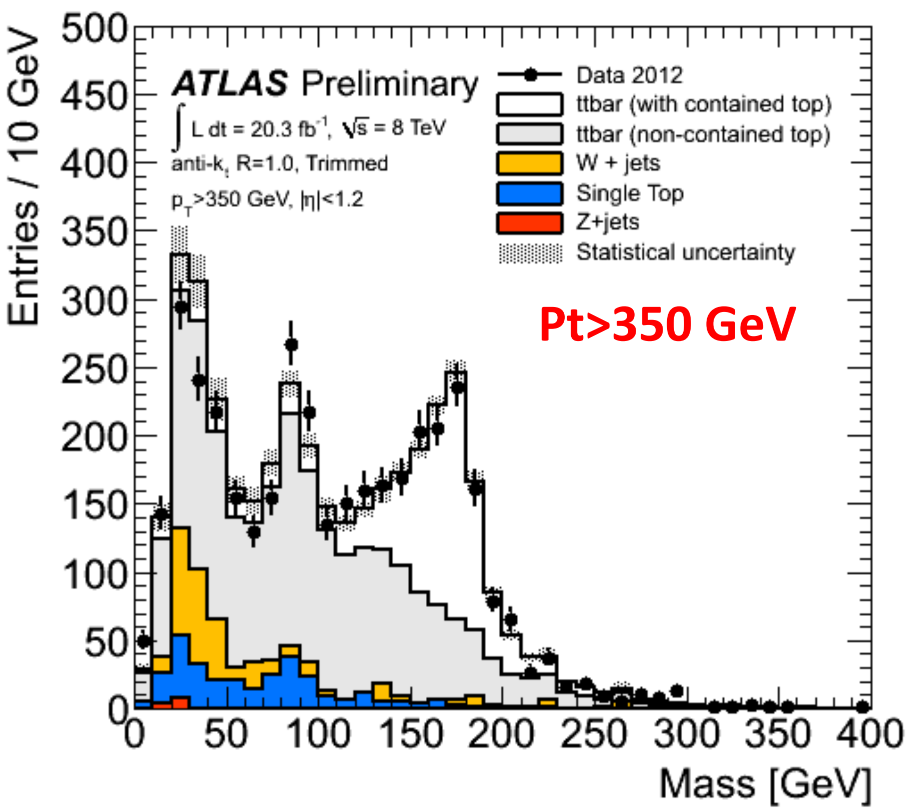
Right: The pt spectrum of t-tbar events with one top decaying semileptonically and one decaying hadronically.

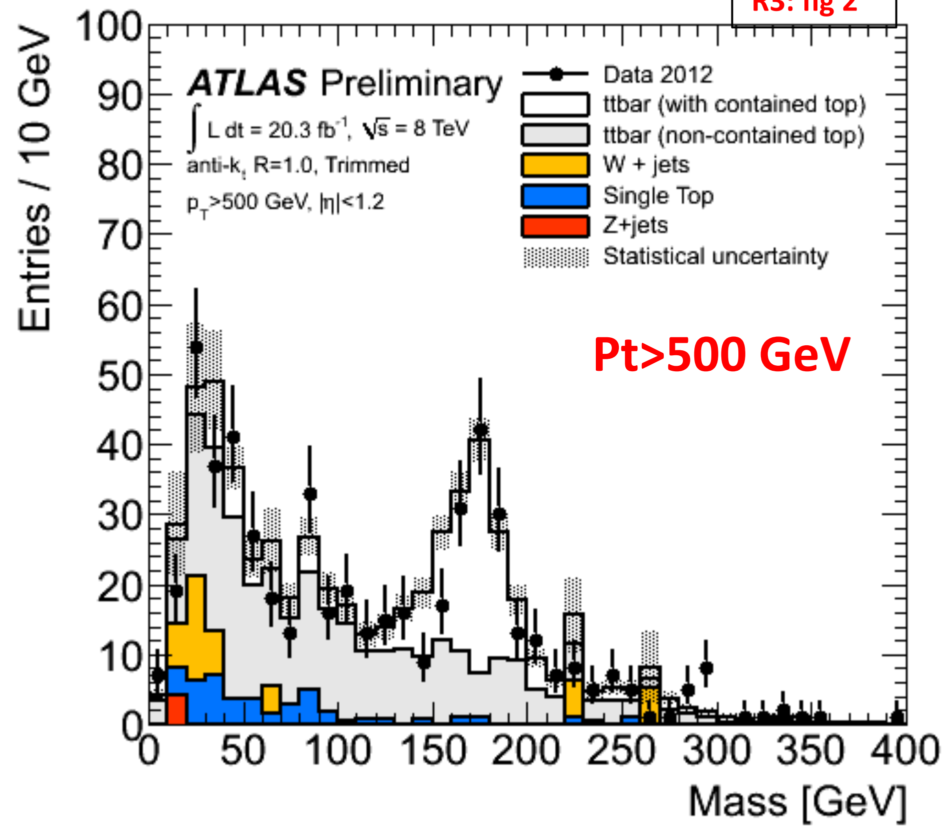
Note small slope in ratio. No correction has been applied to the data.



Mass spectrum of reconstructed trimmed top (2012 data compared to total MC)

R3: fig 2



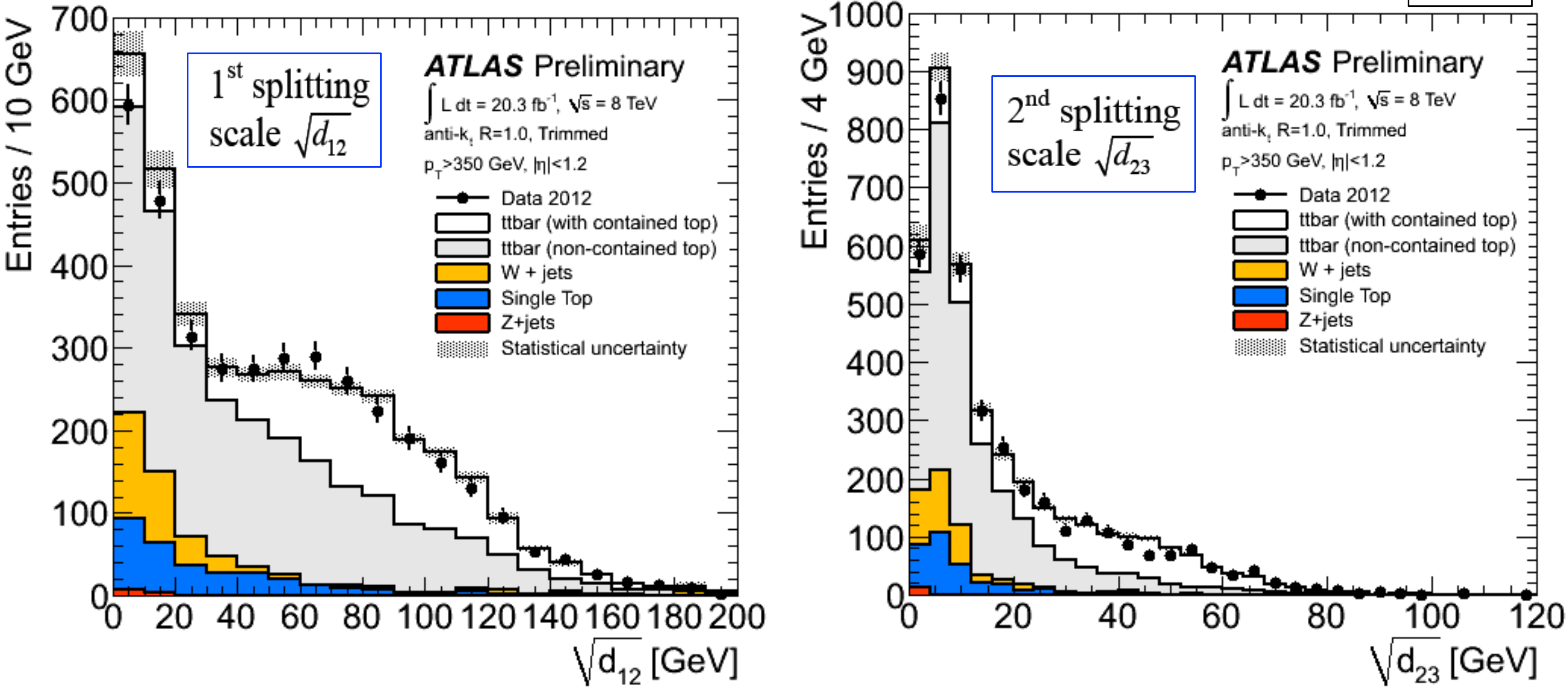


Jet mass for leading p_T anti-kt trimmed jets with R=1.0, $|\eta|<1.2$, and two jet p_T thresholds (before final cuts). Here, "contained" events are those having a hadronically-decaying top quark with all daughter quarks (at the truth level) contained within the large-R jet. "Non-contained" events have one or more of the truth quarks outside the superjet.

The shaded band represents the bin-by-bin uncertainty in simulation.

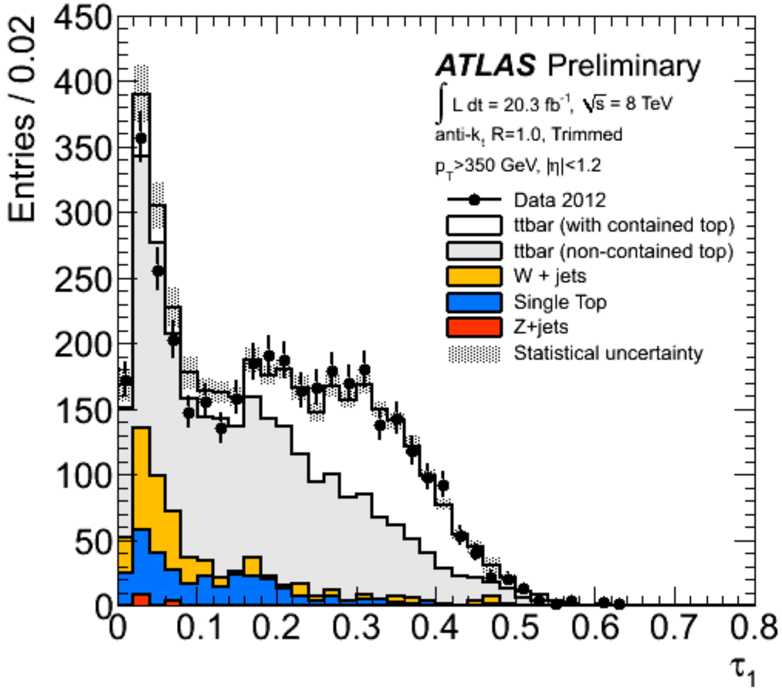
Kt splitting scales from hadronically decaying tops

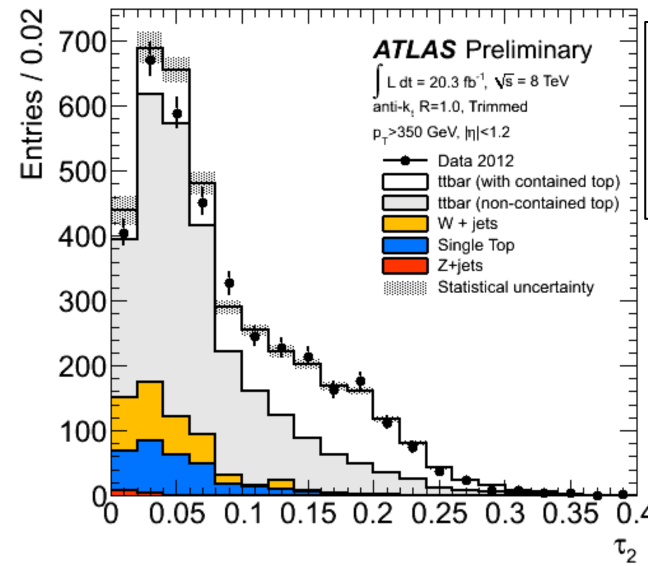
R3: fig 4



Jet splitting scales for leading anti-kt trimmed jets with R=1.0, $p_T > 350$ GeV, and η<1.2 for events selected after a b-tag requirement. Left: 1st splitting scale $\sqrt{d_{12}}$; right: 2nd splitting scale $\sqrt{d_{23}}$. (Recall $\sqrt{d_{12}} \approx M_P/2$).

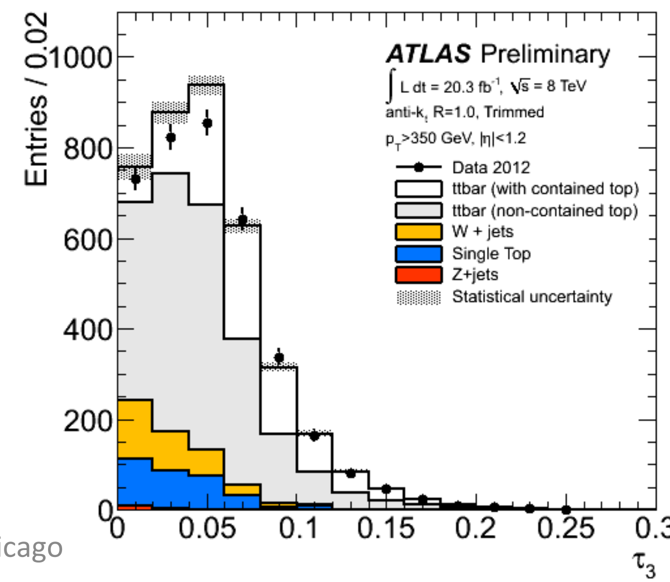
N-subjettiness parameters from top decay





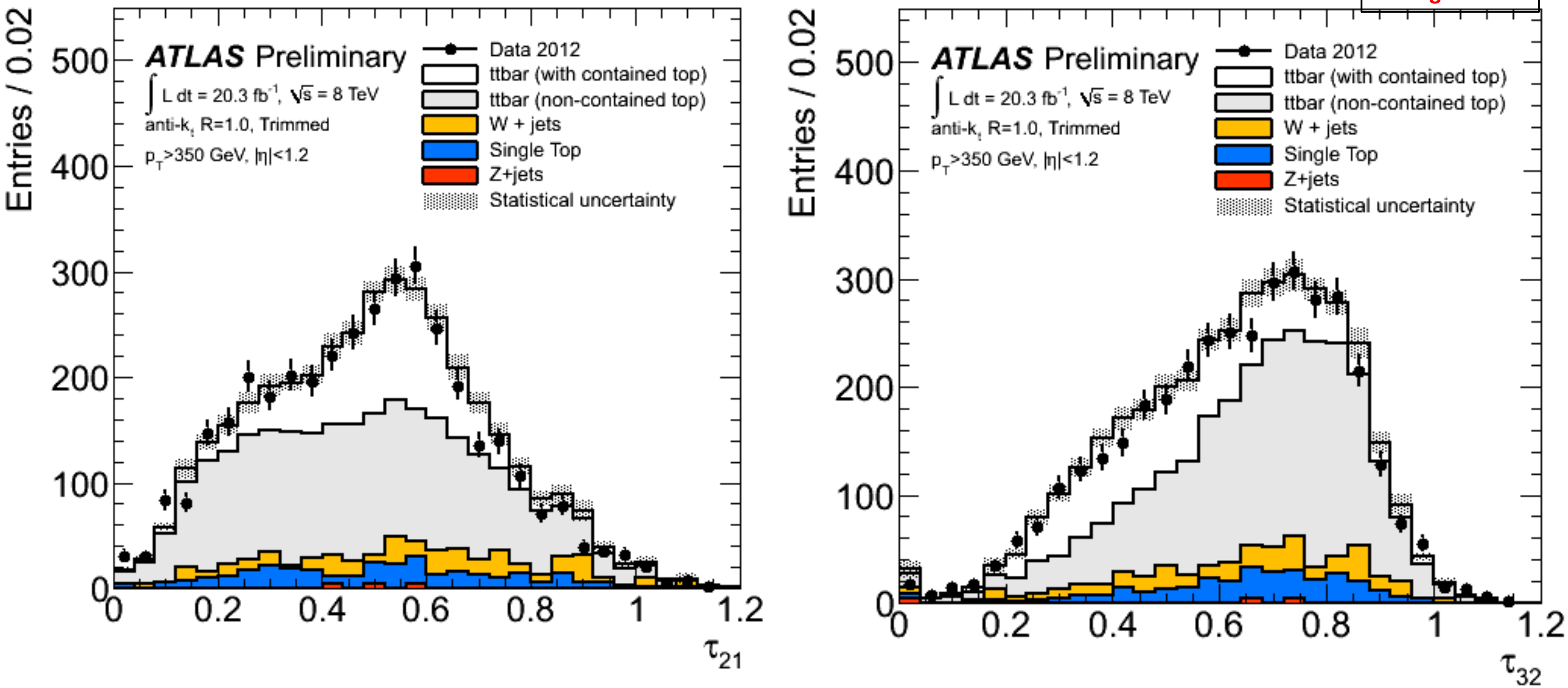
R3: fig 5abc

N-subjettiness parameters (τ_1, τ_2, τ_3) for leading p_T anti-kt trimmed jets with R=1.0, p_T>350 GeV and |η|<1.2 for events selected after a b-tagging requirement.



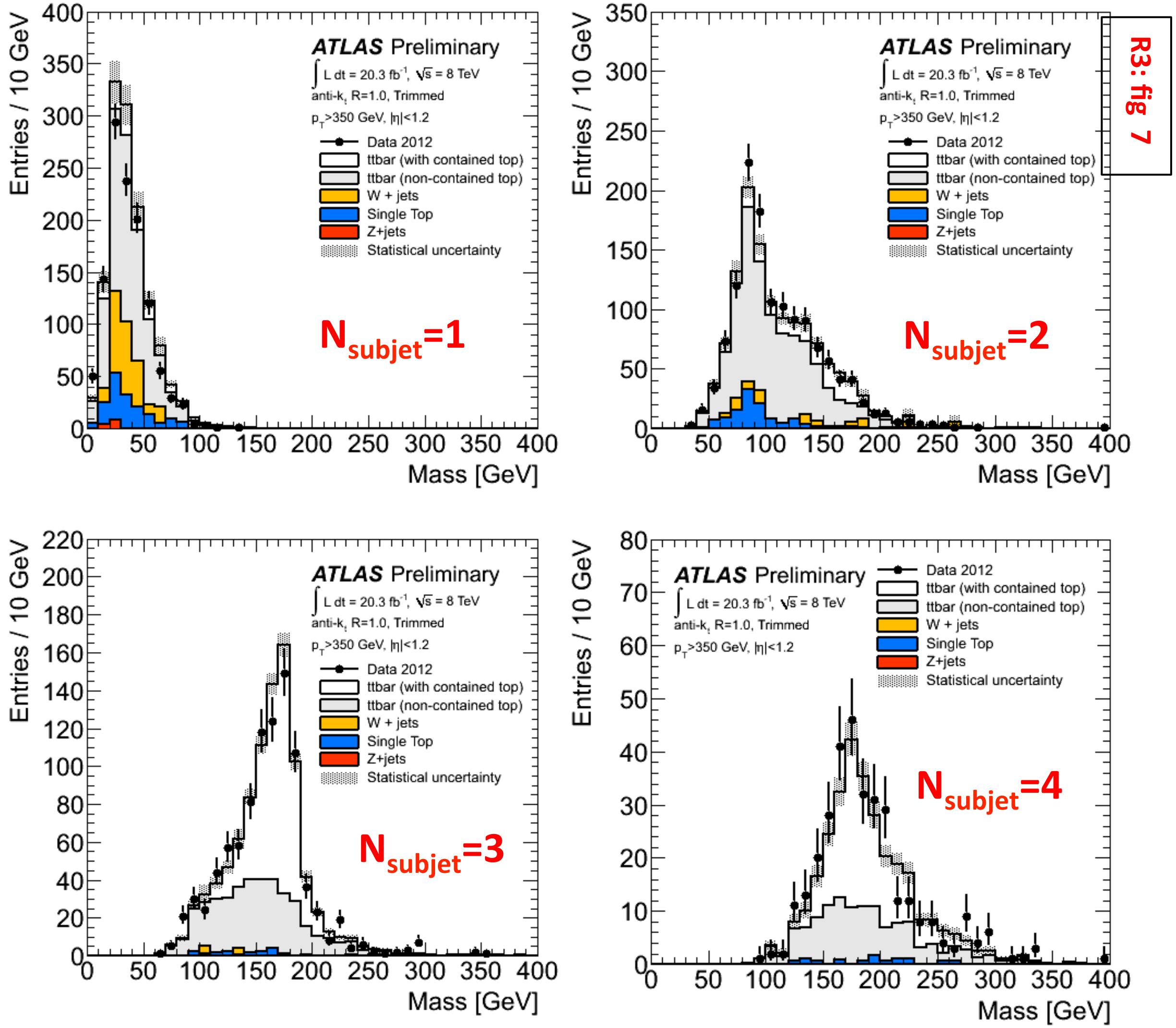
Ratios τ_{21} and τ_{32} from top decay

R3: fig 5de



N-subjettiness ratios τ_{21} and τ_{32} for leading p_T anti-kt trimmed jets with R=1.0 and $|\eta|$ <1.2 for events selected after b-tagging requirement.

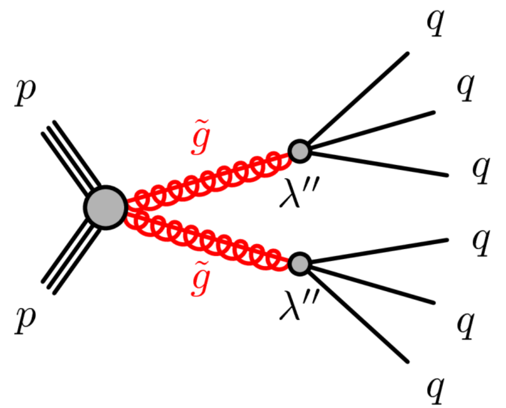
Top Mass spectrum as a function of N_{subjet}



Jet mass for leading trimmed jets with R=1.0, p_T>350, η<1.2, with N_{subjet}=1, 2, 3, and 4 subjets after trimming.

“Search for pair production of massive particles decaying into 3 quarks with the ATLAS detector in sqrt(s)=7 TeV pp collisions at the LHC”

Submitted to JHEP (http://arXiv.org/abs/1210.4813)



Search for R-Parity Violating (RPV) decays of massive pair-produced gluinos,

$$\tilde{g}\tilde{g} \to (\tilde{q}q) + (\tilde{q}q) \to (qqq) + (qqq) \quad \text{with} \quad m_{\tilde{q}} >> m_{\tilde{g}}$$

giving 6 high-pt jets in the final state. Most gluino searches require substantial missing p_T , and would not be sensitive to this final state.

The search was carried out in two analysis channels, which cover approximately orthogonal regions of phase space:

- a resolved-jet channel which required the 6 jets to be well separated, and
- a boosted analysis channel which searched for 2 high-pt jets with a radius of R=1.0, and used substructure analysis to resolve the 3 hard components of these large-R jets.

I will only describe the second of these channels, which was the first time that substructure methods have been used in SUSY searches at the LHC.

Substructure parameters for data, signal and background

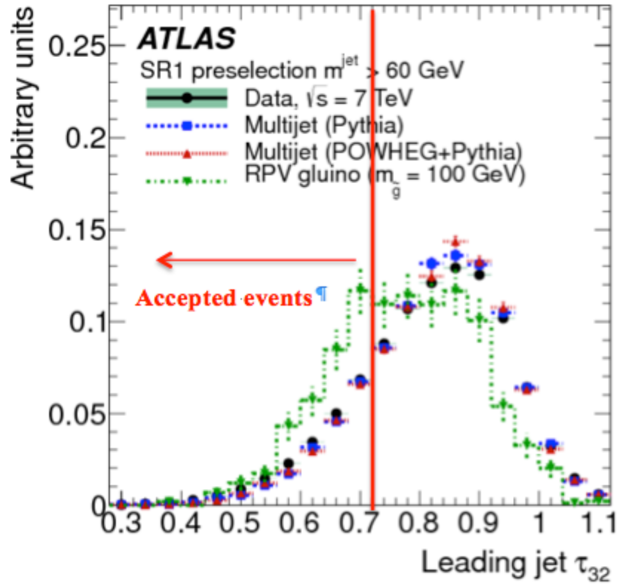
R4: fig 3

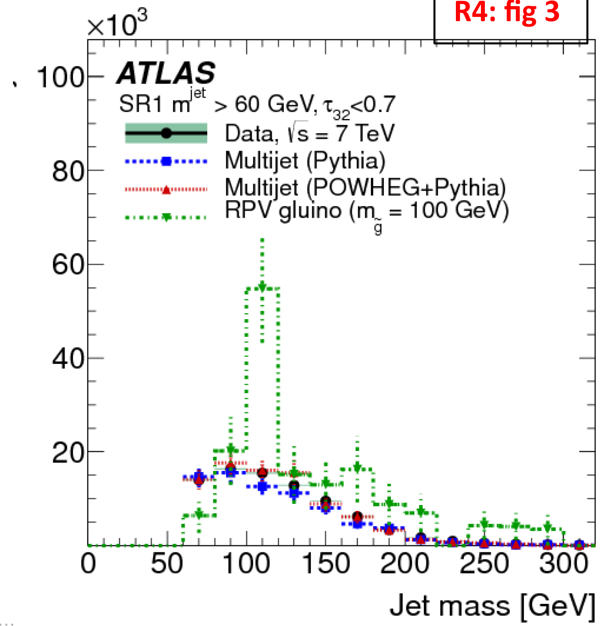
Left: The parameter τ_{32} for data, SM background, and MC signal. A small value (≤ 0.7) indicates 3 hard constituents; a value near 1 indicates 2 is a better representation.

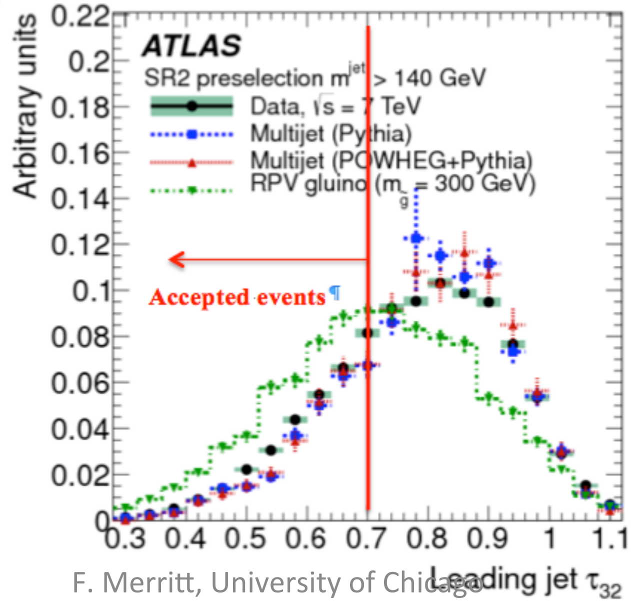
Require $\tau_{32} < 0.7$, increasing the S/B ratio by about a factor of 3.5

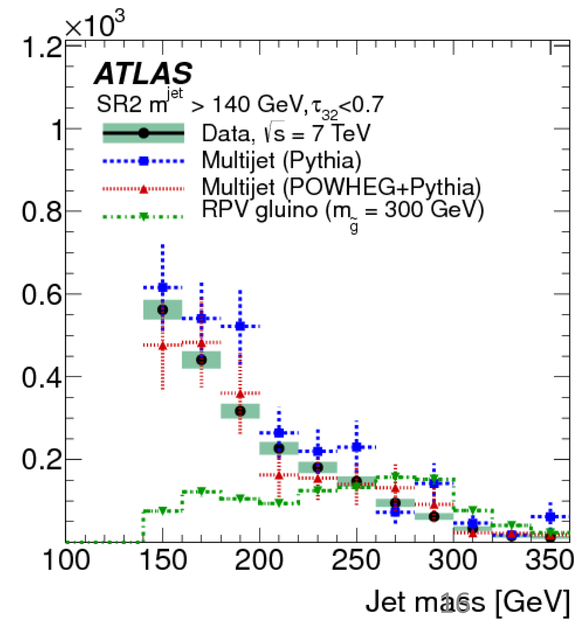
Right: The mass spectrum after requiring $\tau_{32} < 0.7$, showing a clear signal in the region of low gluino mass.

In all of these plots, note the close agreement between data and multijet background, in both shape and normalization.





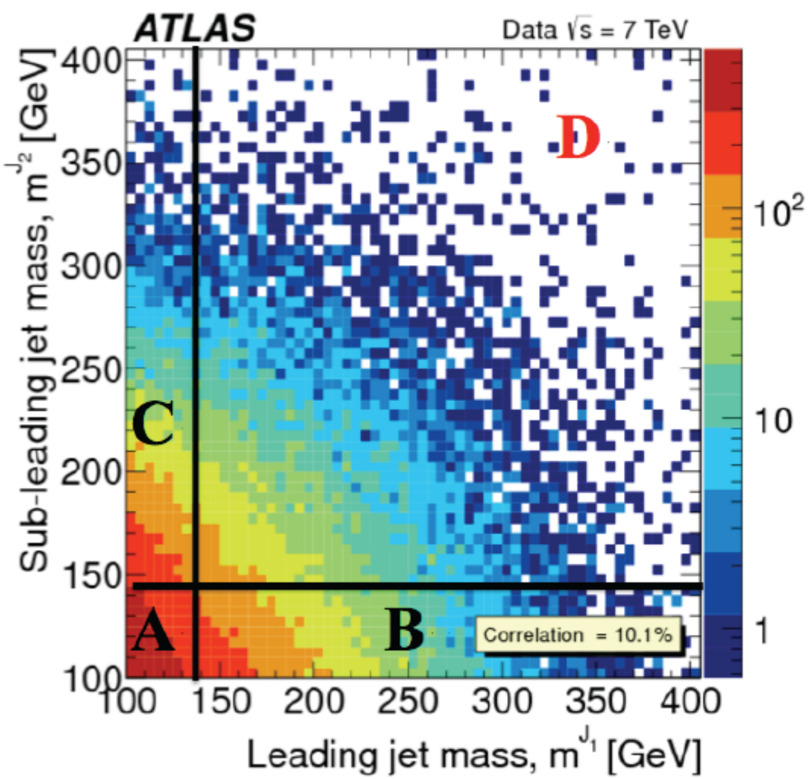


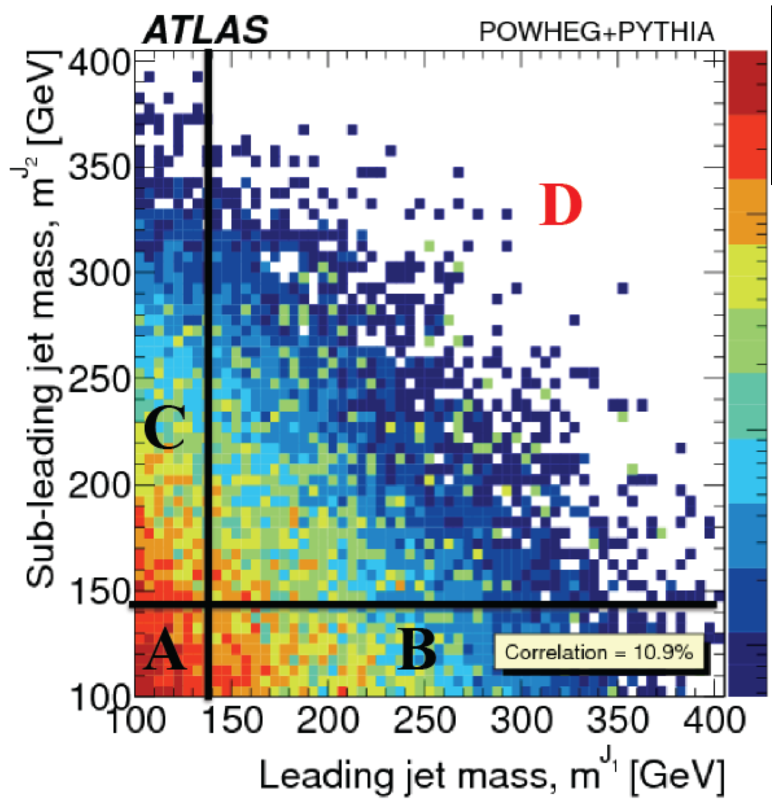


Correlation plots of J_1 and J_2 masses for data and SM background

Signal region: M_{J1}, $M_{J2} > M_{thresh}$ (M_{thresh}=140 GeV below).

Determine bakground in signal region D from data in control regions A, B, C using MC as a template (ABC-D method). Systematic error ≈ 20%).





R4: fig 4cd

Boosted analysis limits as a function of gluino mass

Two selection channels were used:

a) p_T^{Jet}>200 GeV, M_{thresh}=60 GeV

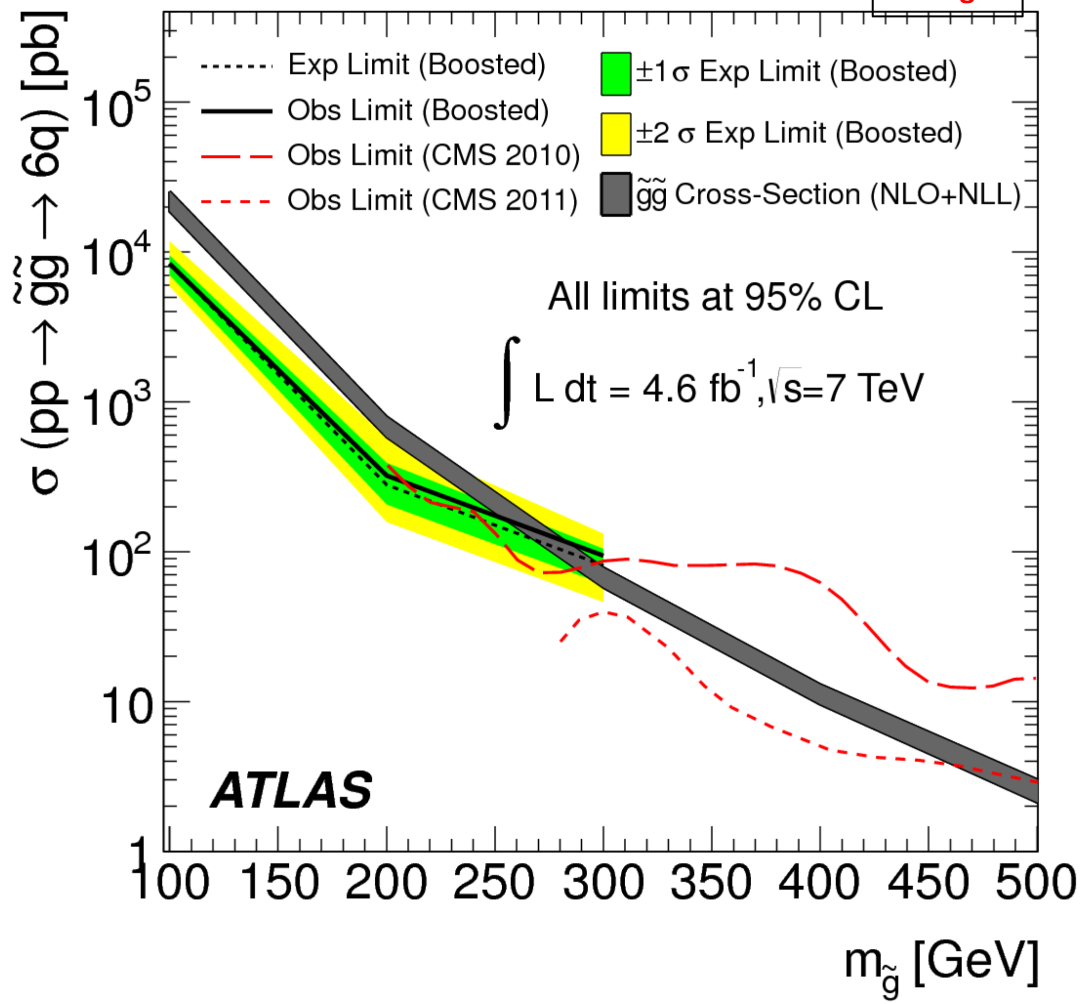
b) p_T^{Jet}>350 GeV, M_{thresh}=140 GeV

excluding gluon masses below 255 GeV as shown in the figure.

(When the limits from resolved events are included, the excluded range is extended up to 666 GeV.)

Note that the boosted analysis deals with a purely hadronic state; also, it represents the first application of substructure analysis to SUSY searches at the LHC.

R4: fig 6



"Search for t-tbar resonances in lepton plus jets events with ATLAS using 14 fb^{-1} of proton-proton collisions at $\sqrt{s}$=8 TeV"

ATLAS-CONF-2013-052

A search is made for a new particle (Z') decaying into two high-p_T tops, with one top decaying leptonically (e or μ) and one decaying hadronically:

$$Z' \to t\bar{t} \to bW + bW \to bl\nu + bqq$$

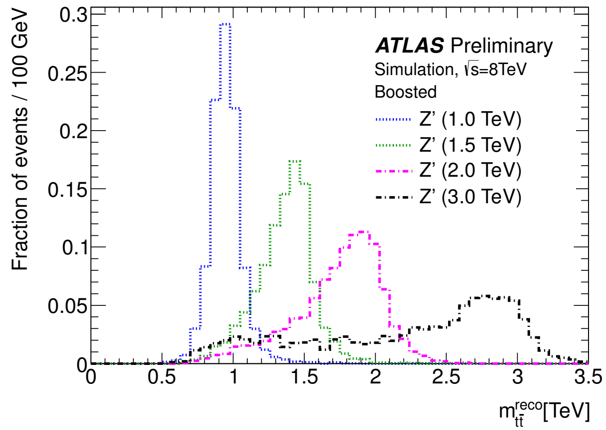
The search is carried out in two selection channels, one using "resolved" anti-kt jets with R=0.4, and one also requiring a large-R "boosted" anti-kt jet with R=1.0 and $p_T > 300$ GeV to capture the hadronic top decay products. I will only describe the boosted channel.

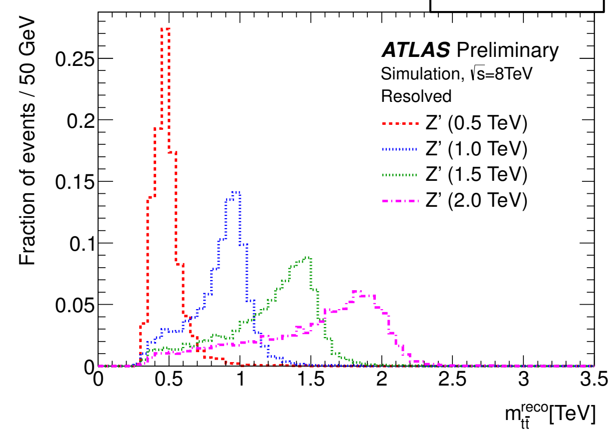
- Large-R jets are trimmed:
  - Constituents (topoclusters) are reclustered using kt-algorithm with $R_{sub} = 0.3$
  - The iterative recombination procedure is "unwound", and at each step subjets with $p_T^{subjet} < 0.05\, p_T^{jet}$ are removed, resulting in 1-6 final subjets.
  - The properties of the trimmed jet are then recalculated. Both small-radius and large-radius jets have their final η and p_T adjusted with energy- and η-dependent MC correction factors (similar to JES corrections for R=0.4 jets).
- Small-radius jets must have $p_T > 25$ GeV and $|\eta| < 2.5$
- Large-radius jets must have $p_T > 300$ GeV and $|\eta| < 2.0$
- There must be at least one small-R jet and one large-R jet in the event.
- The large-R jet must have $m_{jet} > 100$ GeV.
- The first kt spliiting scale must have $\sqrt{d_{12}} > 40$ GeV.
- There must be at least one small-radius jet that is b-tagged.

Efficiency of both channels and both selections

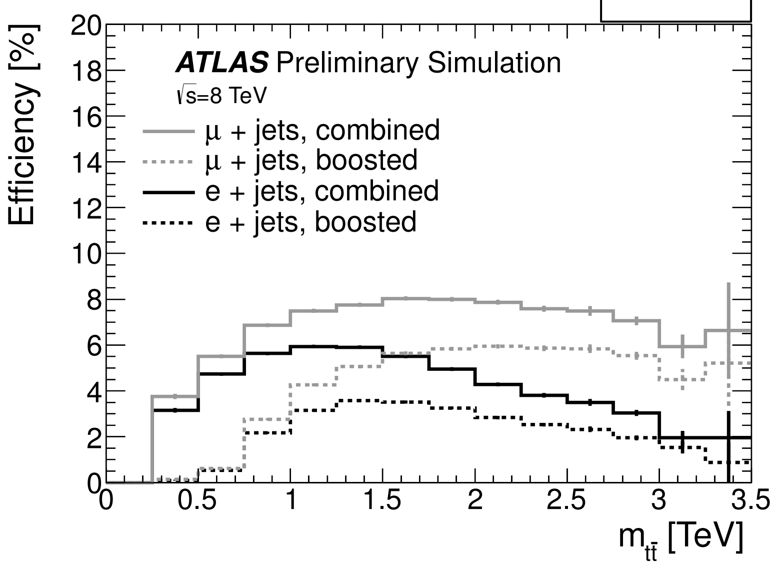
The reconstructed Z-prime mass, reconstructed in both the resolved and boosted selections.

R6: fig 2ac





R6: fig 1



The selection efficiency as a function of true m_{tt} for Z' → tt events. The solid lines show the total efficiency (resolved+boosted) and the dashed lines show boosted alone.

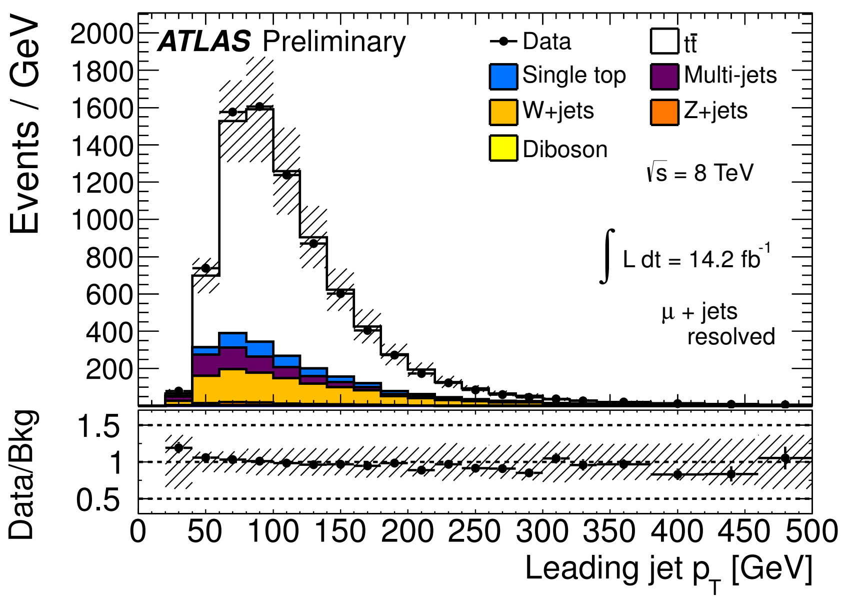
Reconstructed pt for resolved and boosted selections (muon channel only)

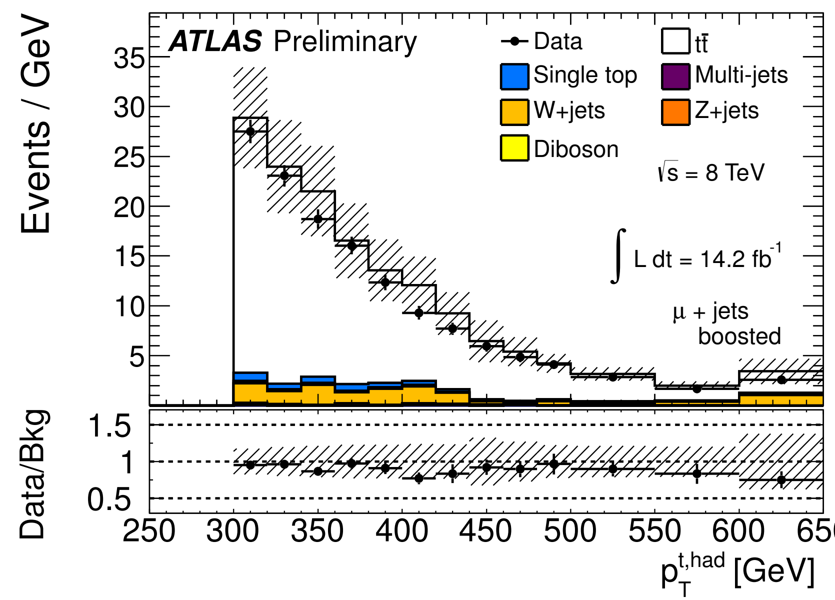
Reconstructed pts are shown for data and MC (muon channel only) for resolved and boosted channels.

Left: Leading (small radius) jet pt of the hadronic system.

Right: transverse momentum of the large-radius jet after the boosted selection.

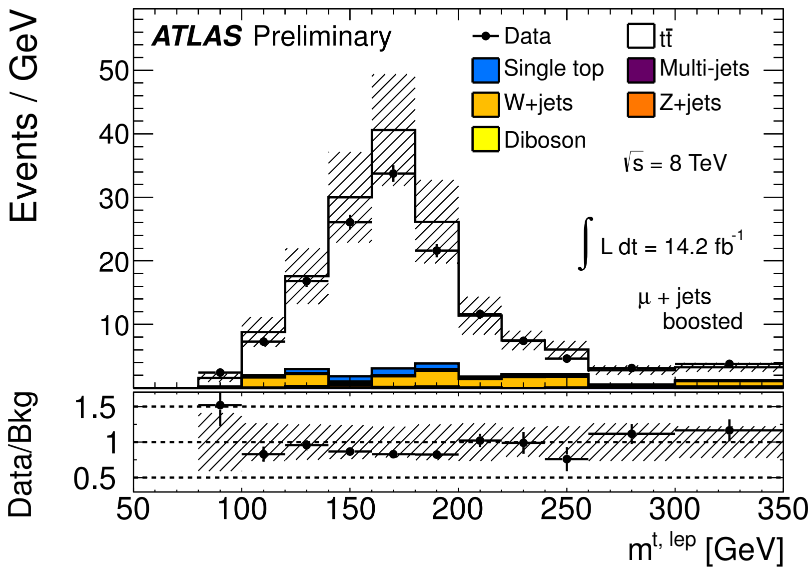
R6: fig 3b, 4b

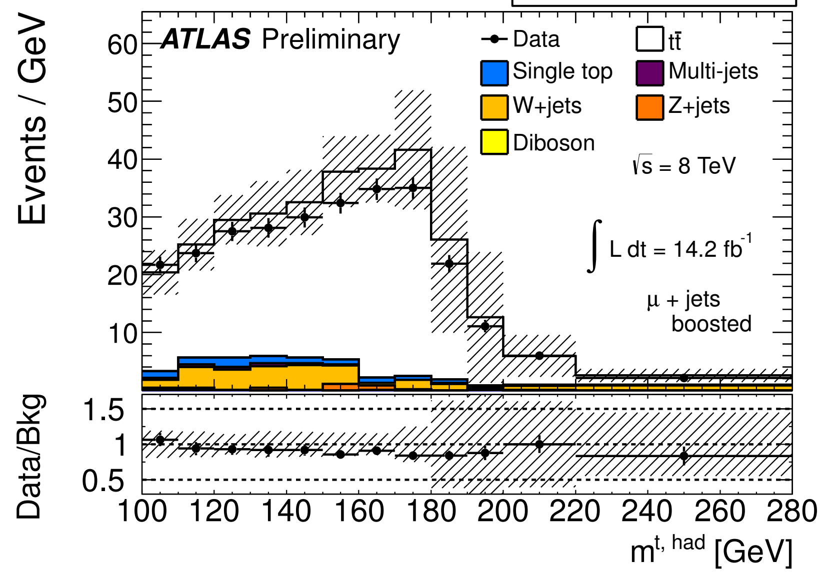




Reconstructed masses and kt scale

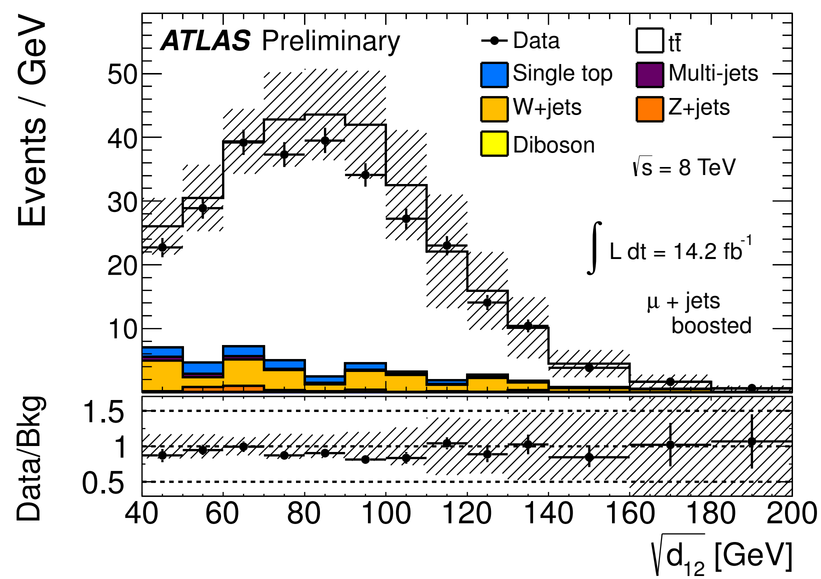
R6: fig 5b, 6b, 7b





Top: Reconstructed mass of boosted tt from (left) the lepton and ME, and (right) the large-R hadronic system.

Right: The substructure variable $\sqrt{d_{12}}$ ($\approx M_P/2$).



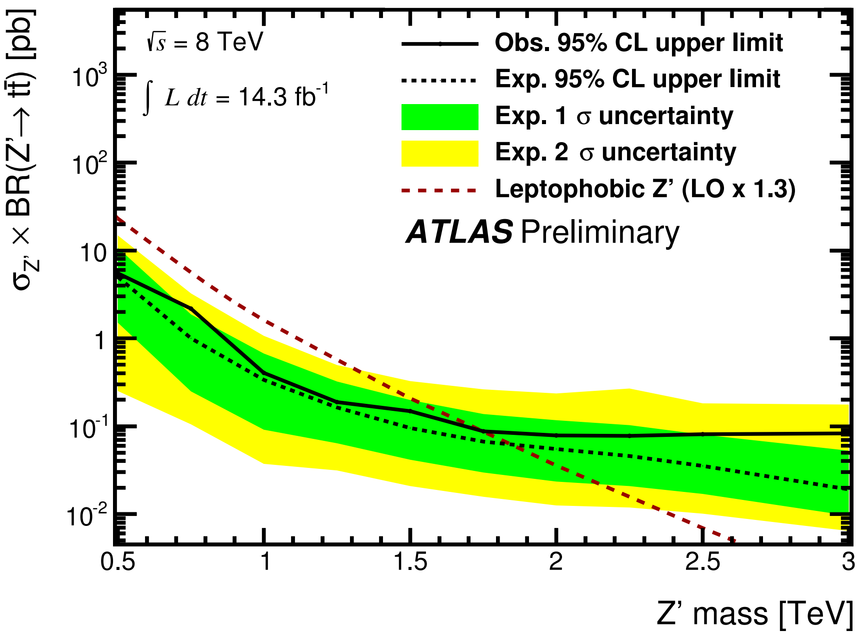
Limits from 2012 data (August 2013)

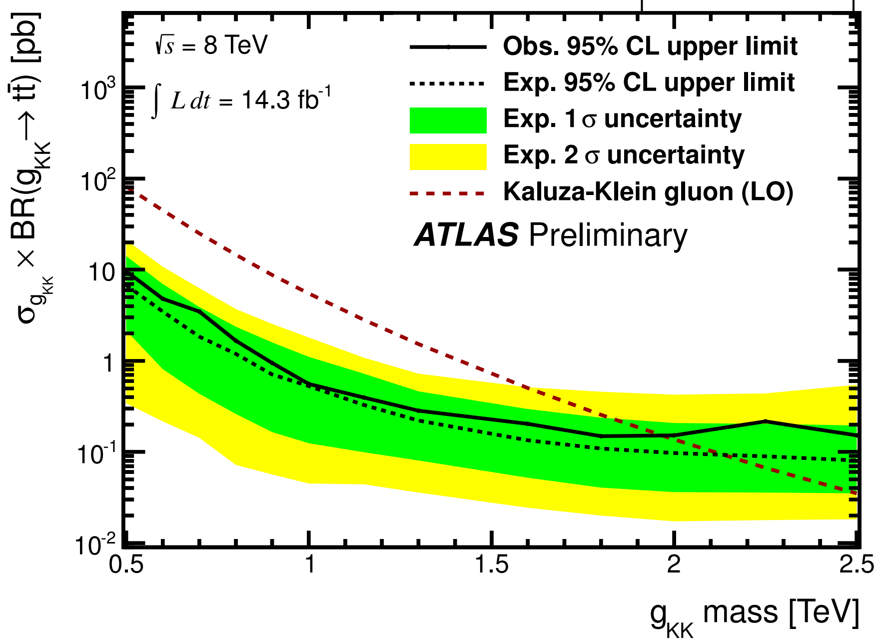
The plots below show the limits from combining the resolved and boosted analyses for the Z' (narrow resonance) and thebroader KK gluino (15.3% width).

Both analyses are important, but the boosted one gives higher efficiency for all masses above 1.4 TeV. It is essential for extending the reach of this (and other searches) into the highest p_T regions.

(Note: Analysis is still underway, and dominant error is calculation of tt tails).

R36 fig 10ab





Summary

- Very extensive work in comparing substructure parameters to data, giving very good agreement between data and MC.
- Significant improvement in resilience to pileup.
- New ideas and ongoing work to improve grooming.
- Significant extensions to search sensitivity in top resonance searches and RPV SUSY searches.

- It is clear that substructure analysis will be increasing important as energy and pileup are both raised in 2015, and a number of searches are extending their high-pt sensitivity by incorporating substructure analysis

Primary Sources

Performance and validation studies

1. Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s}=7$ TeV using the ATLAS detector; submitted to JHEP, 20-June-2013.http://arXiv.org/abs/1306.4945

2. Performance of pile-up subtraction for jet shapes in pp collisions at $\sqrt{s}=8$ TeV. ATLAS-CONF-2013-085, 11-August-2013. ATLAS-CONF-2013-085

3. Performance of boosted top quark identification in 2012 ATLAS data. ATLAS-CONF-2013-084; 11-August-2013. ATLAS-CONF-2013-084

Searches for new physics

4. Search for pair production of massive particles decaying into three quarks with the ATLAS detector in $\sqrt{s}=7$ TeV pp collisions at the LHC, submitted to JHEP. http://arXiv.org/abs/1210.4813

5. A search for t-tbar resonances in the lepton plus jets final state with ATLAS using 4.7 fb^{-1} of pp collisions at $\sqrt{s}=7$ TeV. Phys. Rev D 88, 012004, submitted 12-May-2013. http://arXiv.org/abs/1305.2756v2

6. A search for t-tbar resonances in lepton plus jets events with ATLAS using 14 fb^{-1} of proton-proton collisions at $\sqrt{s}=8$ TeV. ATLAS-CONF-2013-052.

Note: Figures are labeled as [Rn: figX], which means Figure X of the n^{th} reference of this list.

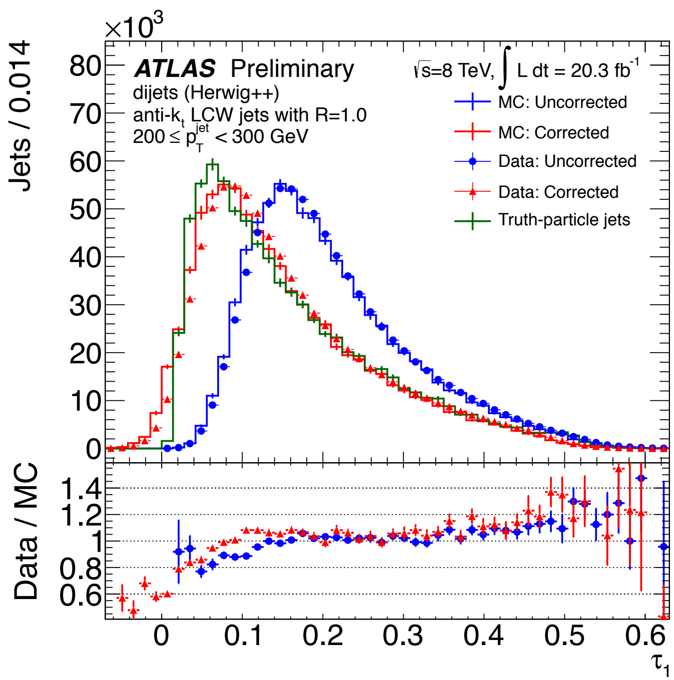
Backup slides

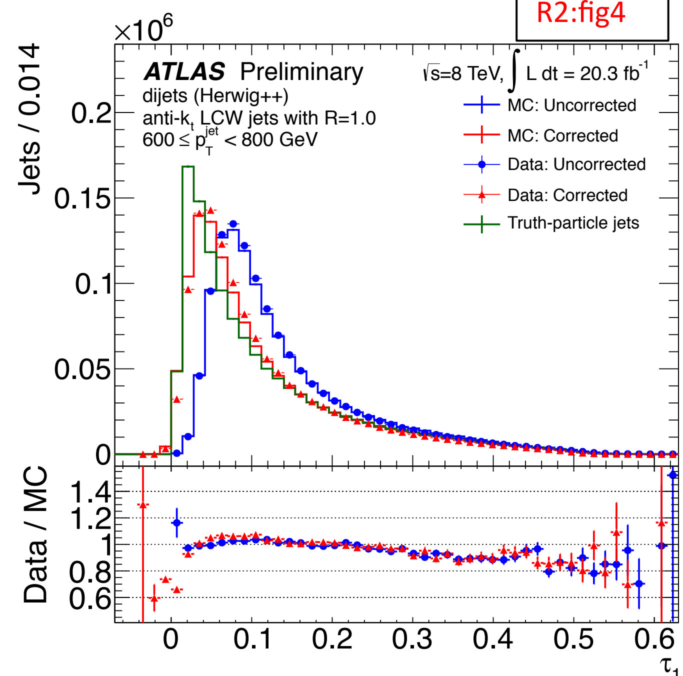
New shape-dependent Pileup subtraction: Reco vs truth (1/2)

In the following figures, τ_N variables calculated from “truth” final state particles (green) are compared to the corresponding variables calculated from the reconstructed MC events as described previously, (blue), and also after corrections based on pile-up and the shape subtraction procedure (red) described in R04. In general, the agreement between the corrected distribution and the truth distribution is striking, as is the improvement from uncorrected to corrected, especially at high pt.

For the plots below and on the next 2 pages, the right-hand plots show the low pt region (200-300 GeV) and the left-hand plots show high pt (600-800 GeV).

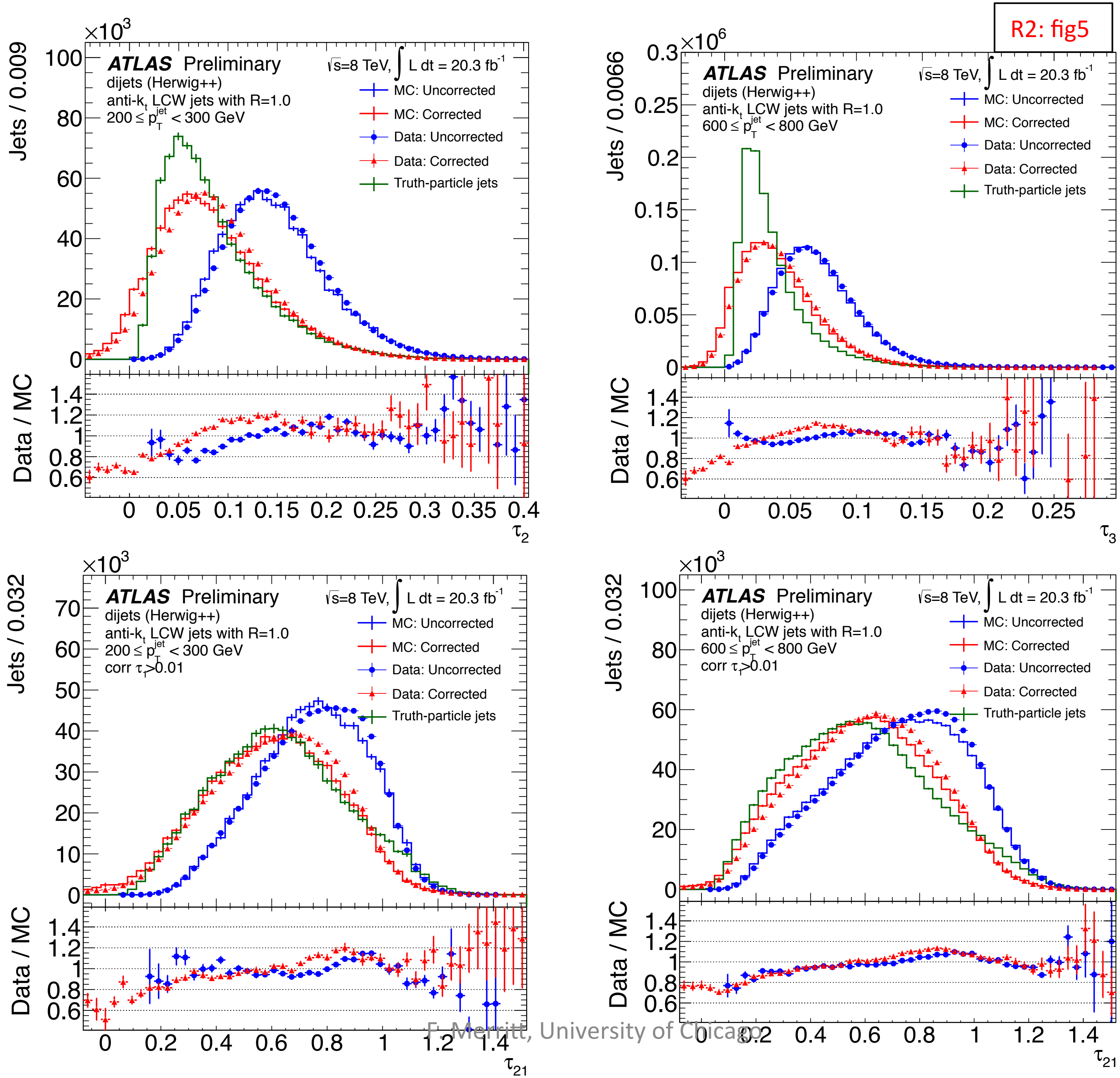
R2:fig4





Shape-dependent pileup subtraction: Reco vs truth (2/2)

R2: fig5



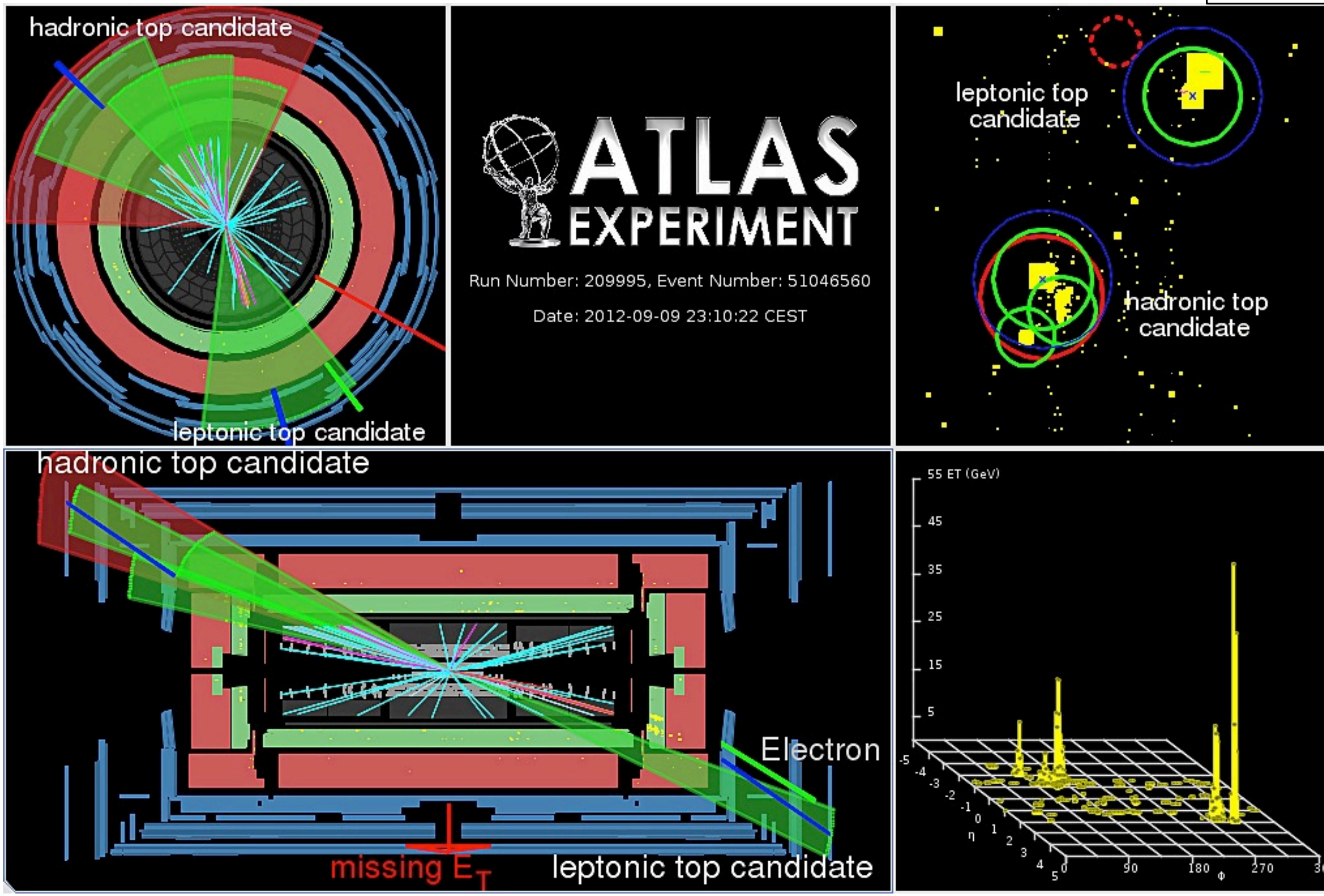
Event display of a boosted t-tbar candidate

R6: fig13b

Upper left: The r-φ Jets are represented by cones, red for R=1.0 (upper part) and green for R≤0.4. Note that the red cone captures, but does not distinguish, the 3 components of the top decay. The electron is shown as a green track, and the missing pt as a red one.

Lower left: The same, but viewed from the side.

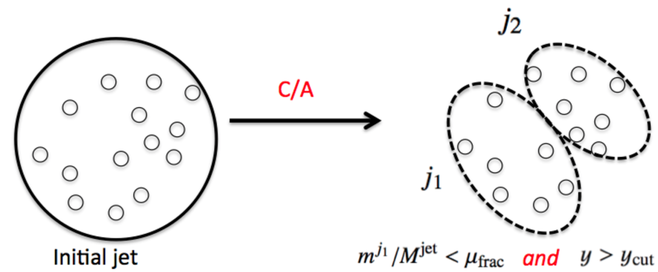
Lower right: A lego plot of the raw calorimeter energy in the η-Φ plane. The leptonic candidate is in the right front part.



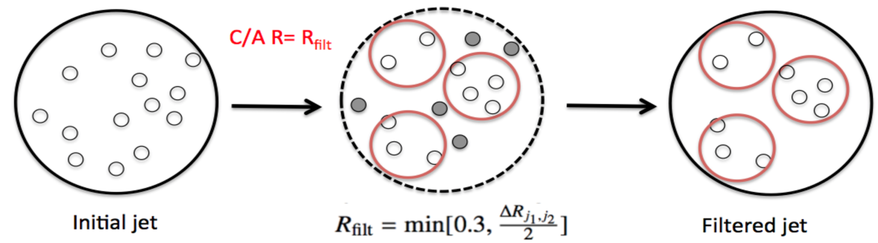
Upper right: This is again the η-Φ plane, now showing the jets as circles. The red circle is the R=1.0 jet of the hadronically decaying top, and the green jets are the R≤0.4 jets. The two light blue outer concentric circles indicate that the jet is b-tagged. The dotter red circle near the center top indicates missing p_T.

Jet Substructure techniques

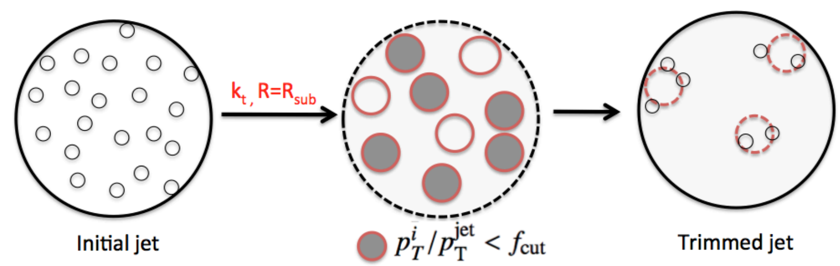
Mass-drop and symmetry: Iteratively undo last stage of kt clustering until largest subjet mass is less than $\max(m_{j1},m_{j2})<\mu\bullet M_{j1+j2}$. Also require $[\min(p_T^{j1}, p_T^{j1})xR_{12}]^2>y_{cut}\approx0.09$. Discard remaining constituents.



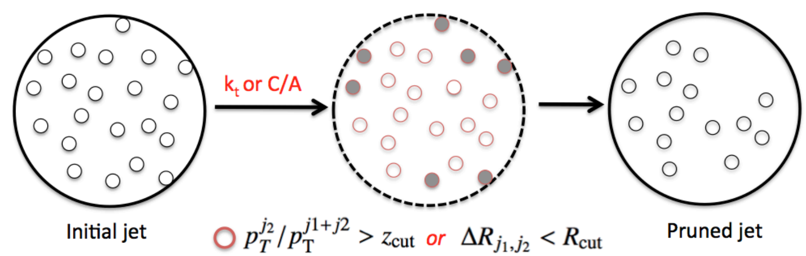
Filtering: Recluster constituents j1, j2 with C/A, $R=\min[0.3, \Delta R_{j1j2}]$. Drop all constituents outside the 3 hardest.



Trimming: Recluster constituents with kt to create subjets of size R_{sub}. Remove any with $p_{Ti}<p_T^{jet}\bullet f_{cut}$ with $f_{cut}\approx0.03$.

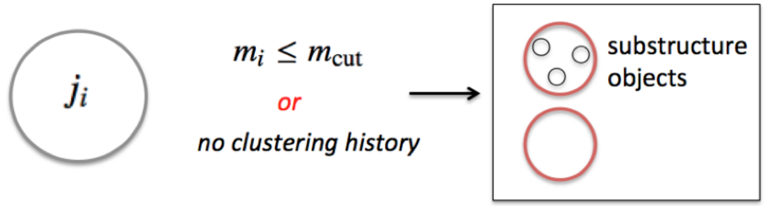


Pruning: At each reclustering step, with $p_T^{j2}<p_T^{j1}$, require either $p_T^{j2}/p_T^{j1+j2}>z_{cut}$ or $\Delta R_{j1,j2}<R_{cut}x(2m^{jet}/p_T^{jet}$. If not, drop j_2 and continue.

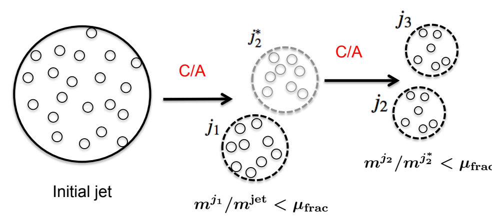


The HEPTopTagger

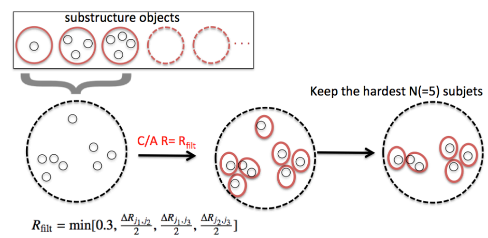
(a) Every object encountered in declustering is considered a "substructure object" if it is of sufficiently low mass or has no clustering history.



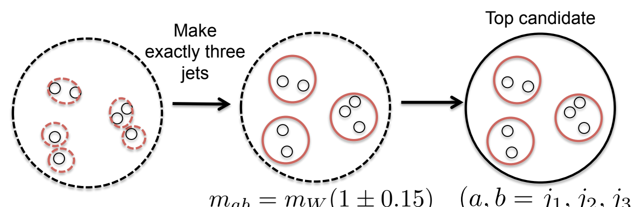
(b) The mass-drop criterion is applied iteratively, following the highest subject-mass line throughout the clustering history, resulting in N_i substructure objects.



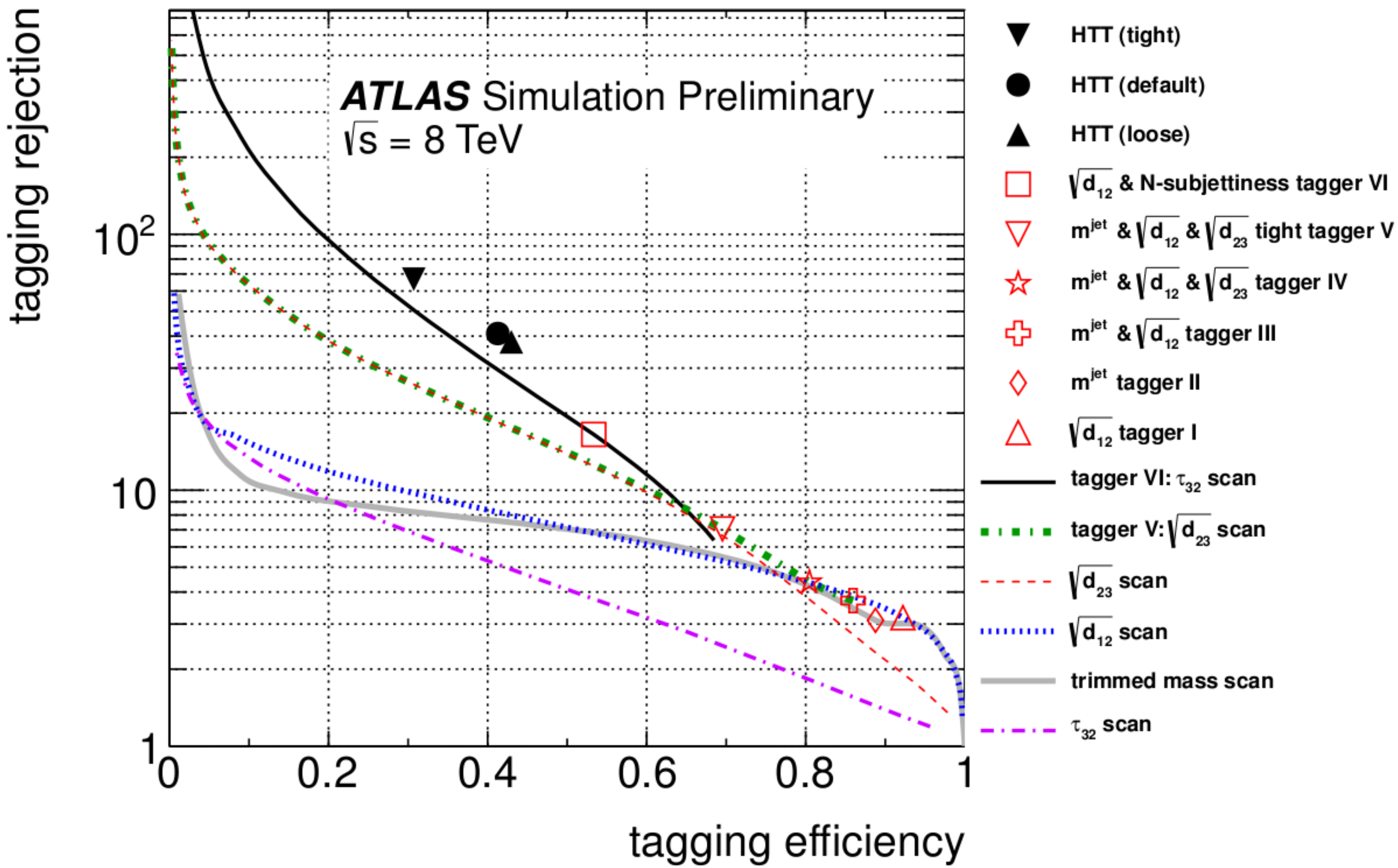
(c) For every triple-wise combination of the substructure objects found in (b), recluster the constituents into subjects and select the N_{subjet} leading-p_T subjets, with $3 \leq N_{subjet} \leq N_i$ (here, $N_{subjet}=5$).



(d) Recluster the constituents of the N_{subjet} subjets into exactly 3 subjets to make the top candidate for this triple-wise combination of substructure objects.

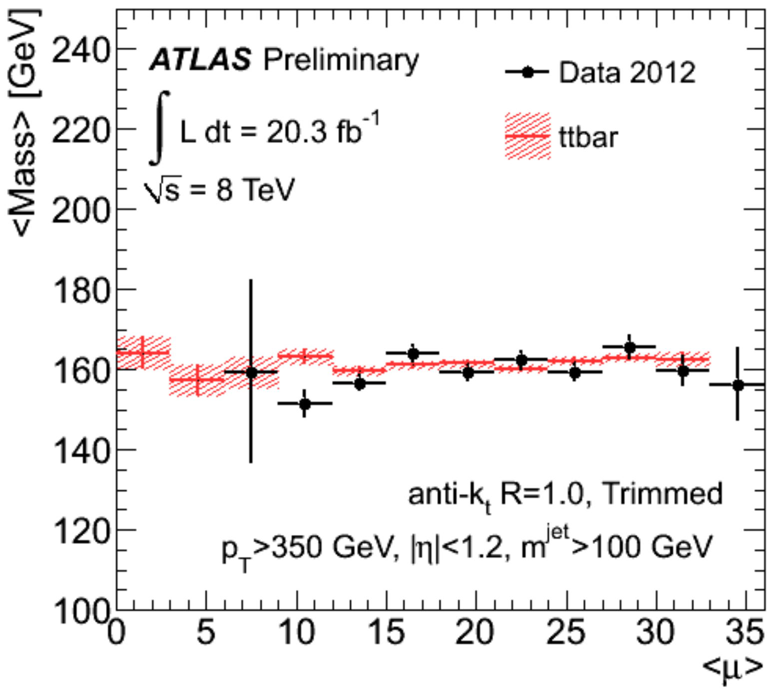


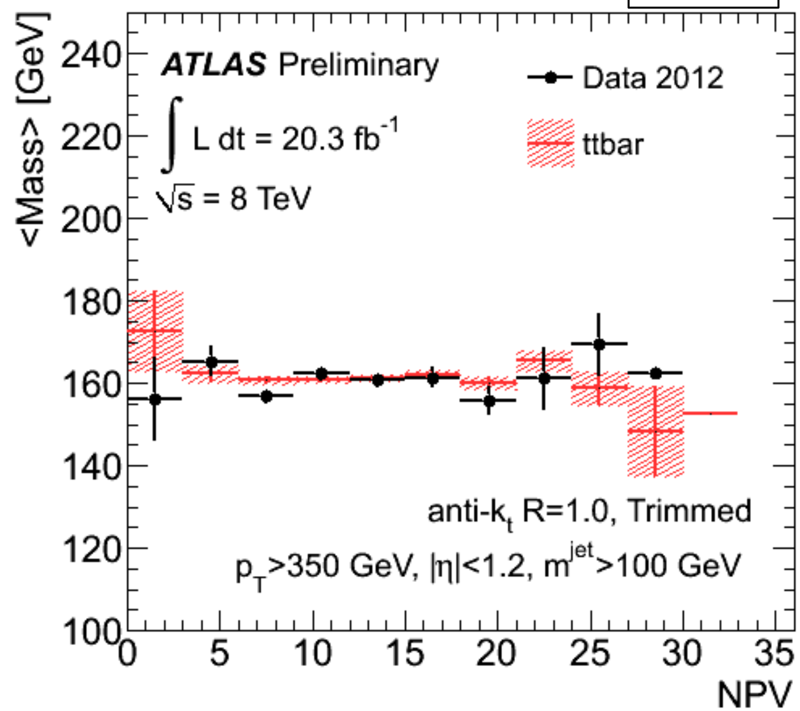
Top Taggers: eff vs rejection



Top mass and pilupe (2012 data)

R3: fig 3





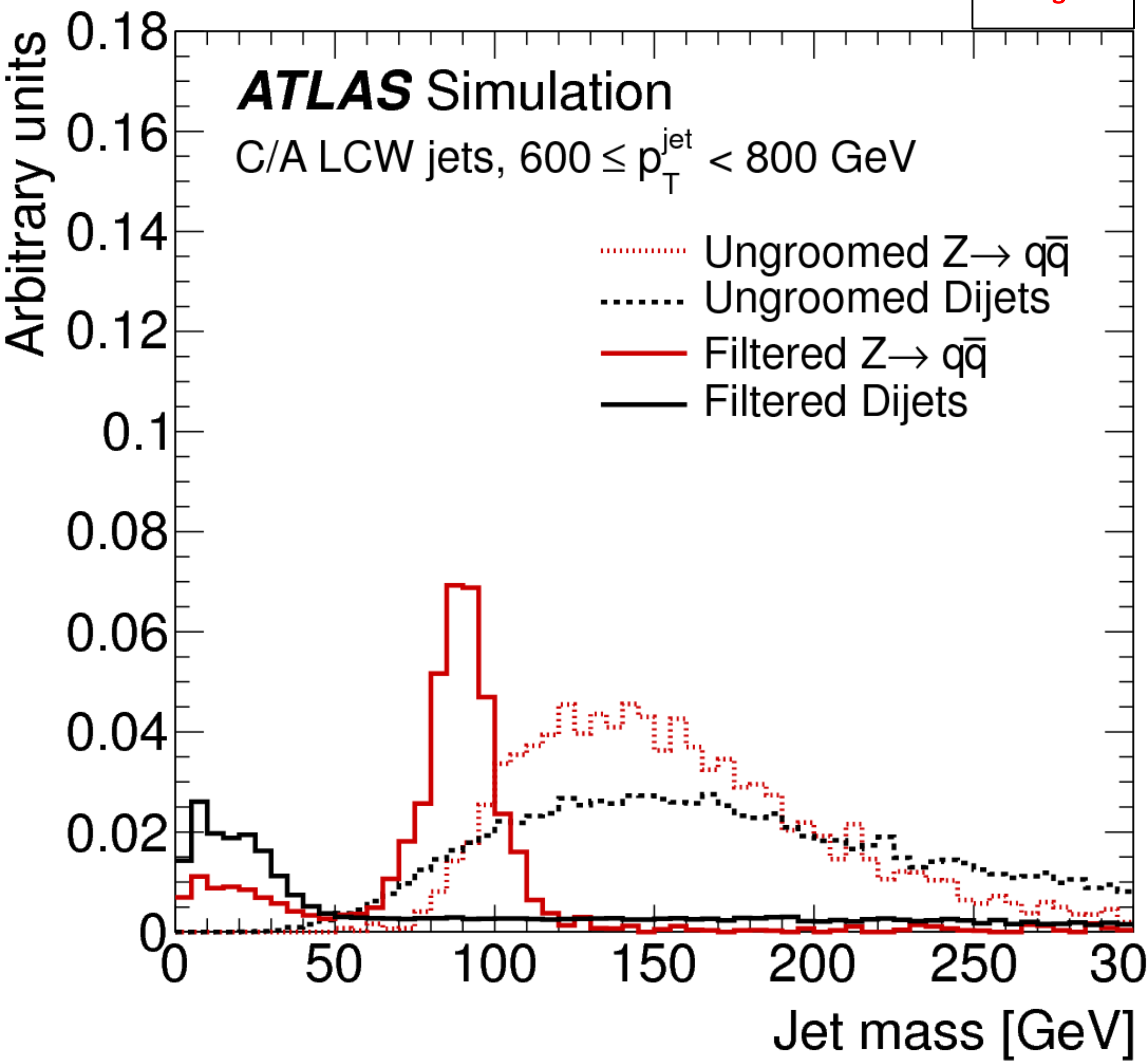
Average mass for leading p_T anti-kt trimmed jets with R=1.0, p_T >350 GeV, |η|<1.2, and m^{jet}>100 after the b-tagging requirement. Left: versus average number of collisions per bunch crossing. Right: versus the number of primary vertices in the event. Shaded band represents bin-by-bin statistical uncertainty in simulation.

The effect of grooming on reconstructed jet masses

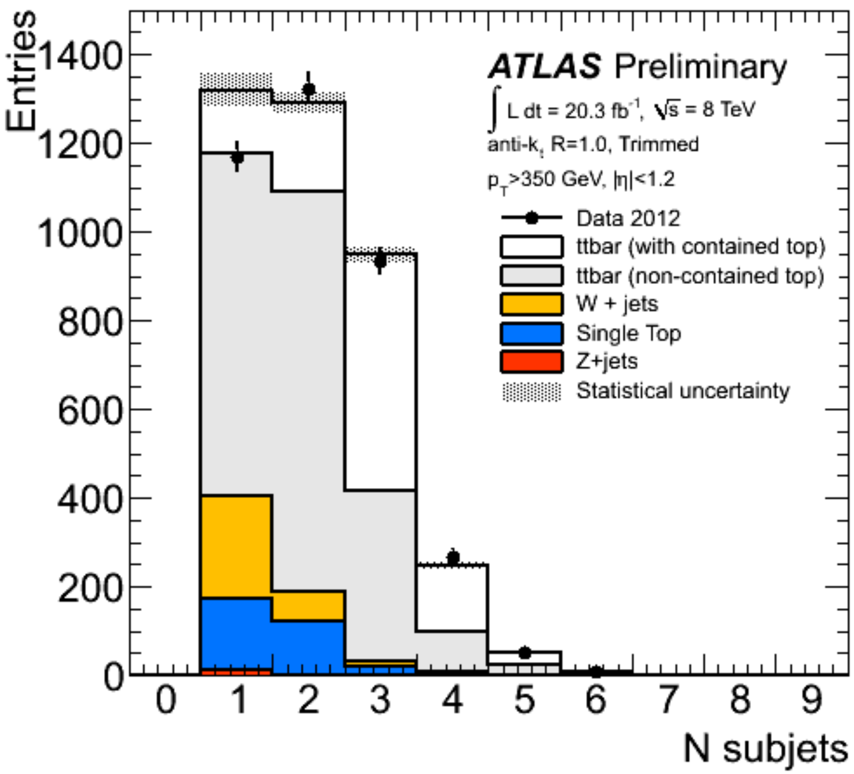
Large-R jets are essential for capturing the decay products from a highly boosted particle, but they also pick up a lot of UE and pile-up background that greatly distorts the mass. Grooming is essential to recover the true mass.

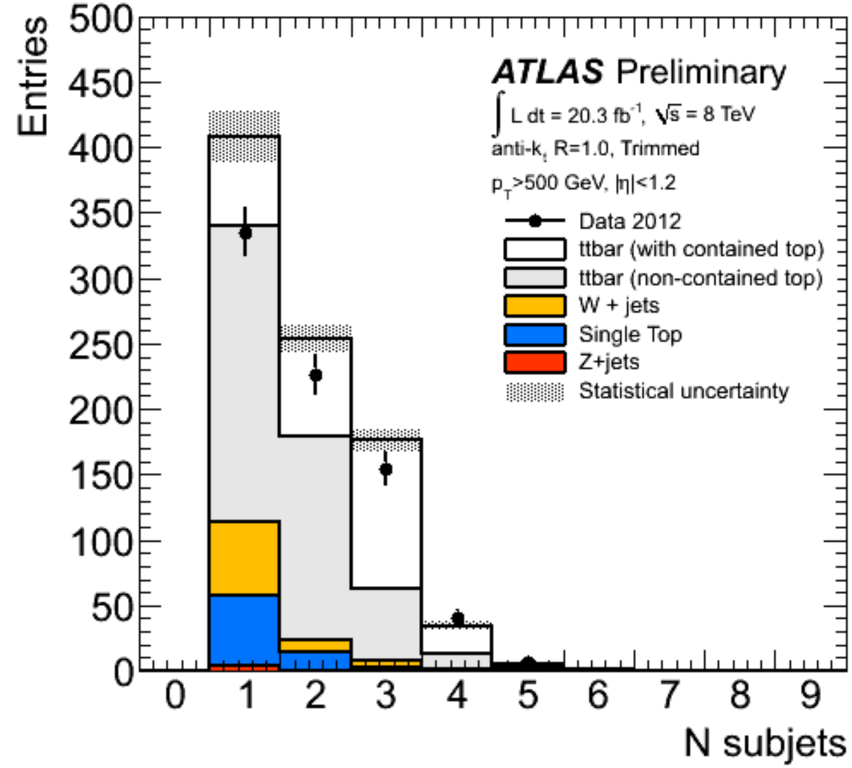
The figure compares the masses before and after grooming for a Z^0 and a light quark jet (from a dijet event). In both cases, the particles have a pt of 600-800 GeV. In both cases, the large jets are C/A jets with R=1.2.

R1: fig29b



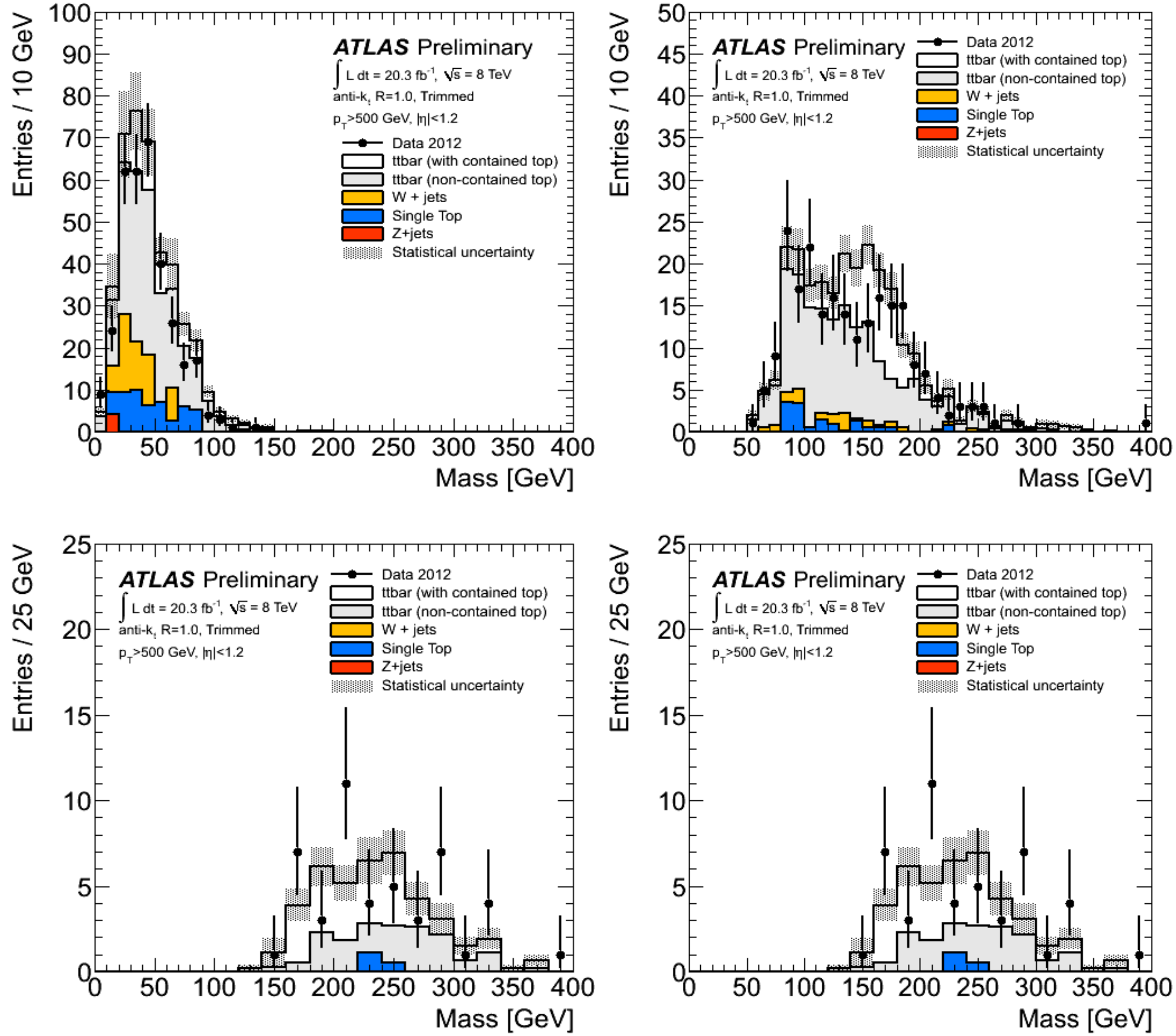
Boosted Top Validation (2012 data) 2/N





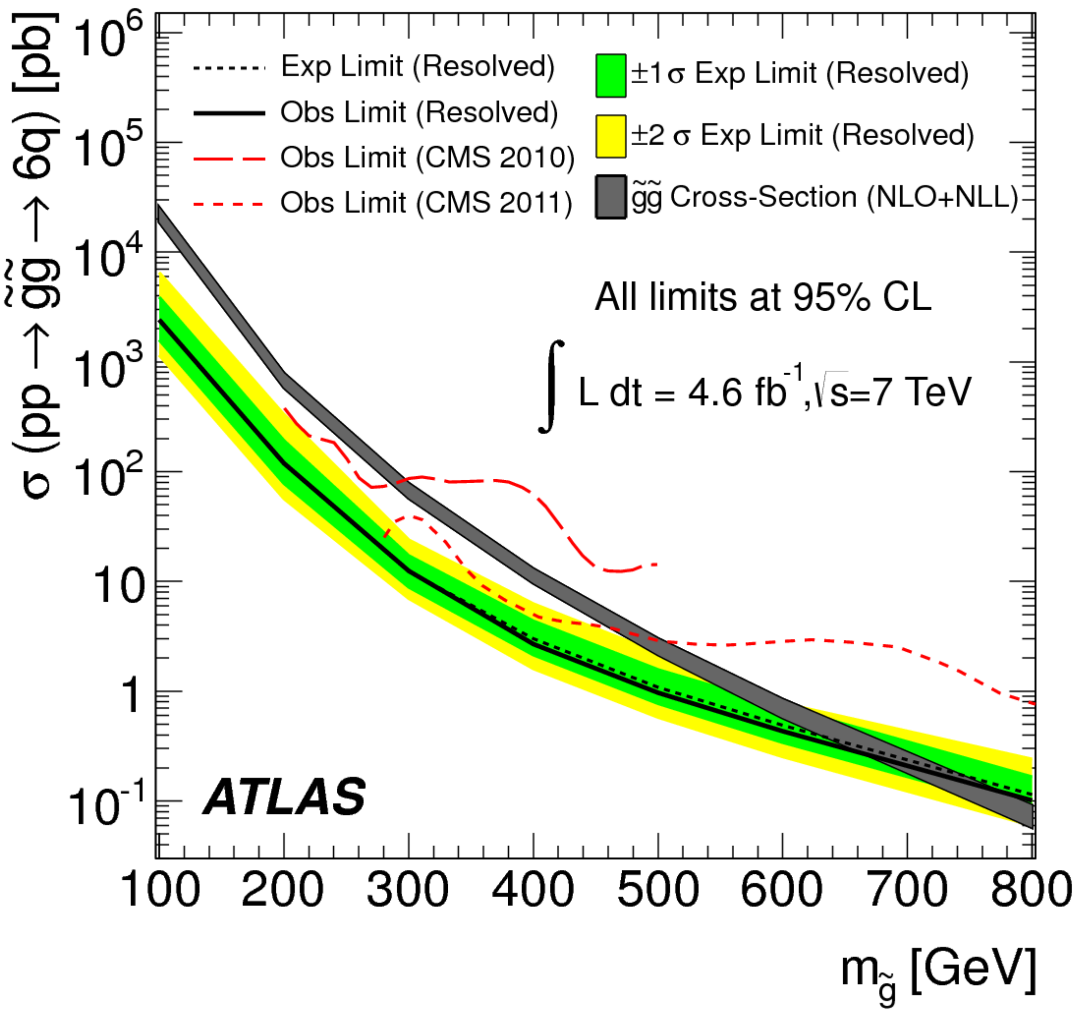
Number of kt subjets after trimming on leading-p_T anti-kt subjets with R=1.0 and |η|<1.2. The shaded bands represent bin-by-bin statistical MC uncertainty and do not include systematics.

Boosted Top Validation (2012 data) 2/Nbackup



Jet mass for leading trimmed jets with R=1.0, p_T>500, η<1.2

Resolved limit (gluino search)



“A search for t-tbar resonances in the lepton plus jets events final state with ATLAS using 4.7 fb^{-1} of pp collisions at √s=7 TeV”

Phys. Rev. D 88, 012004

This recently published search, using the 7 TeV data of 2011, set the final 2011 ATLAS limits shown on the right for (top) narrow Z’ resonances, and (bottom) Kaluza-Klein gluons.

Since both the method and results are similar to the ongoing analysis of 2012 data at 8 TeV, I will only discuss details of the latter.