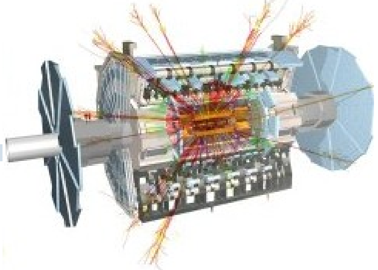
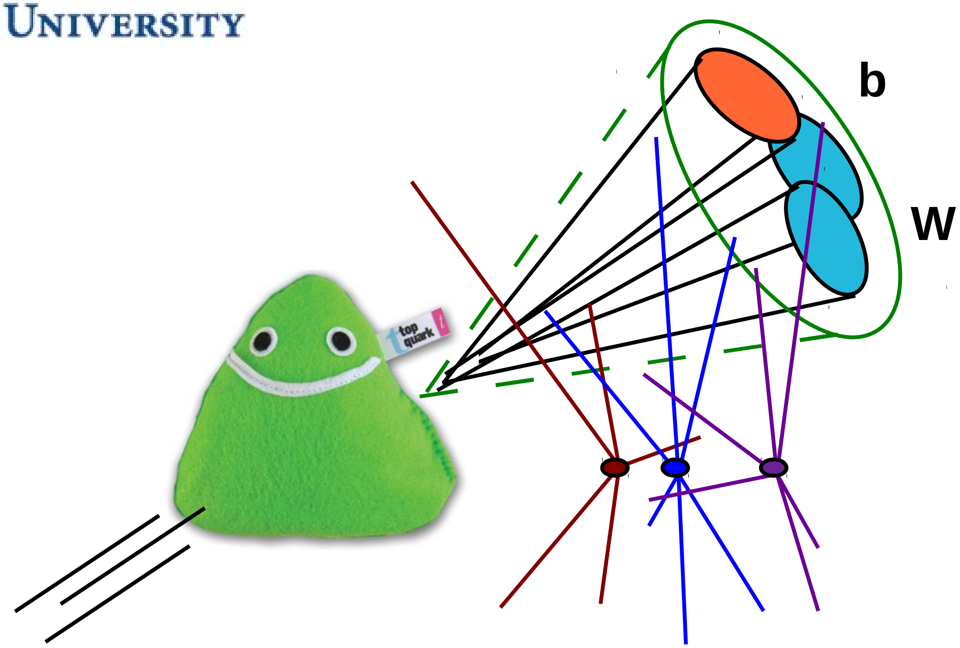




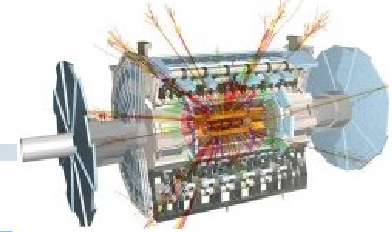
Jet Substructure and Boosted top-tagging at ATLAS

Emily Thompson – *Columbia University*
on behalf of the ATLAS Collaboration

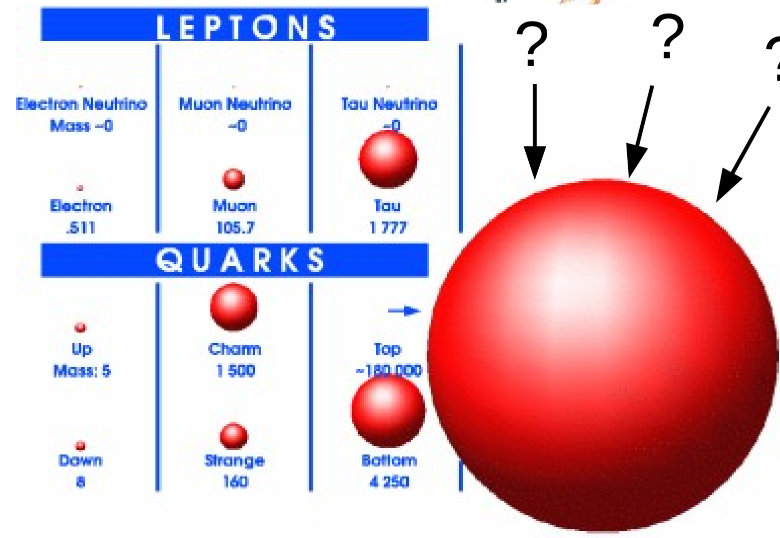
Workshop on LHC Physics

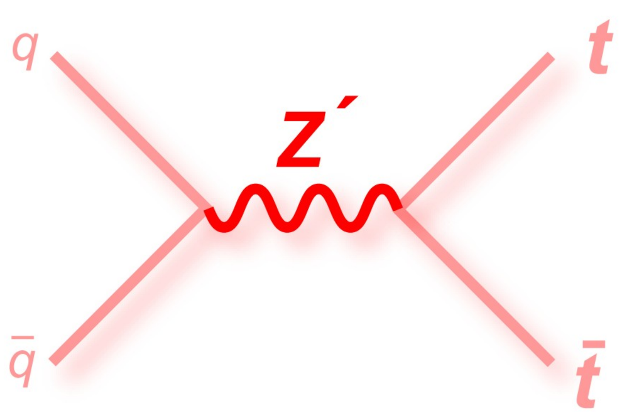
- Outline:
  - Motivation for looking for boosted tops
  - Jet substructure
  - Current results with boosted tops
  - The future of top-tagging and jet grooming in ATLAS

Introduction – High-pt top quarks



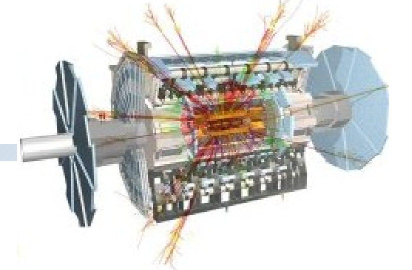
- The LHC is a top-producing machine
  - Pair production cross section @7(8) TeV LHC: **~20 (30) times** Tevatron
- Top quark properties: highly coupled to electroweak symmetry breaking mechanism....and new physics??
- New physics (such as heavy gauge bosons) can produce tops with relatively high transverse momentum.... “boosted”





- Many physics analyses in ATLAS are beginning to use boosted objects signatures to search for new phenomena in high energy collisions provided by the LHC (Higgs, Exotics, SUSY)
- Need to develop tools in order to identify top quarks over the full momentum range

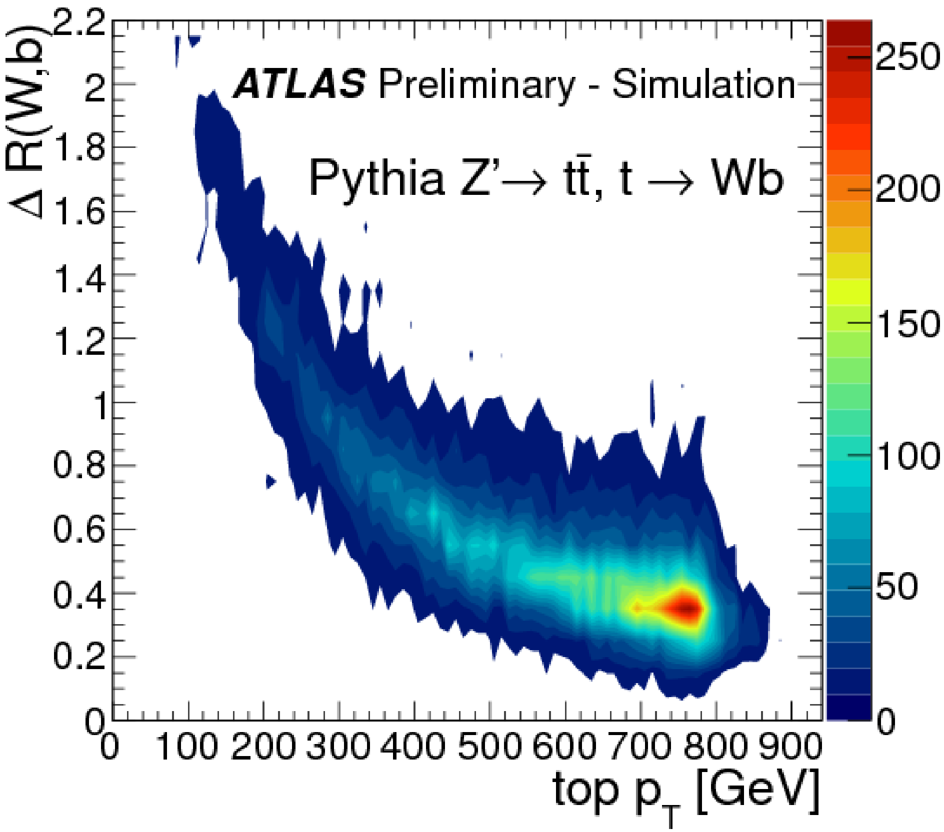
Entering the boosted regime

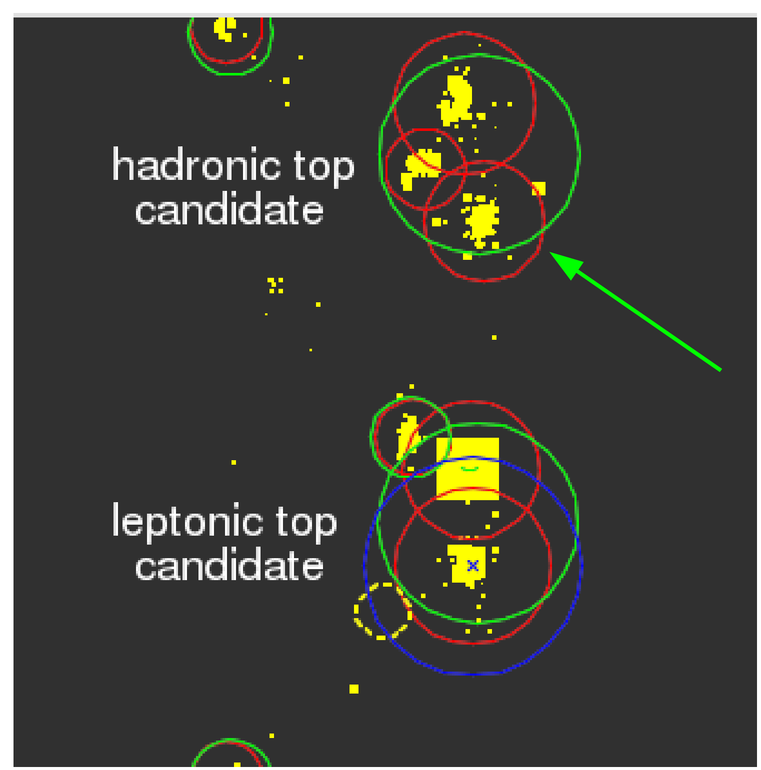


- At high-p_T, hadronic decay products ($t \rightarrow Wb \rightarrow qq'\, b$) collimate into a single large-radius jet
- Rule of thumb: opening angle of decay products of a boosted object has a $1/p_T$ dependence

$$dR \sim \frac{2m}{p_T}$$

→ eg: top with $p_T > 350$ GeV or so will have decay products within a separation $dR \sim 1$



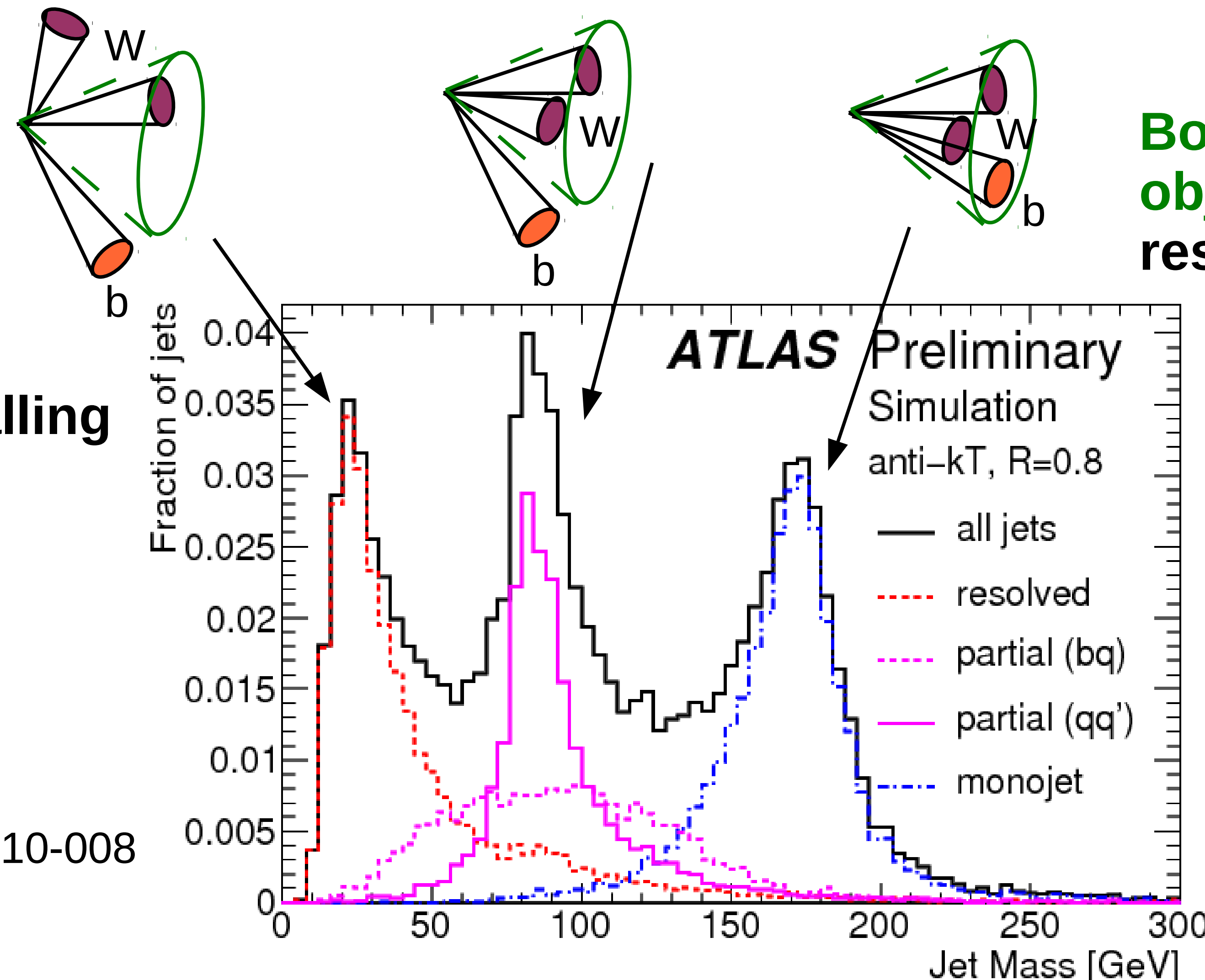


R = 1.0
m_j = 197 GeV
E_T = 356 GeV

Jet substructure

- Utilize information we have inside the jet to pick out boosted object
- Best substructure variable: **Jet Mass** (aka: mass of the 4-vector of the jet)

QCD:
steeply falling spectrum



Boosted object: resonance

ATL-PHYS-PUB-2010-008
(early simulation of
$g_{kk} \rightarrow t\bar{t}$ events
at various masses, sqrt(s)=10 TeV)

Jet substructure

- global shape of the jet
- how the clusters are distributed relative to the jet axis
- how the jet was created

- Other information stored in a jet:
- eg: **kt-splitting scales:**

$$\sqrt{d_{ij}} = \min(p_{\mathrm{T}i}, p_{\mathrm{T}j}) \times \Delta R_{ij}$$

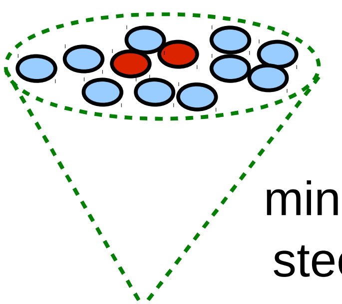
(i, j = last constituents before final reclustering)

- Jet constituents reclustered with k_t algorithm

  → highest p_T clustered last

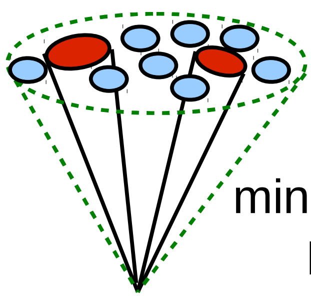
- Hard scale tends to be less sensitive to pileup

QCD



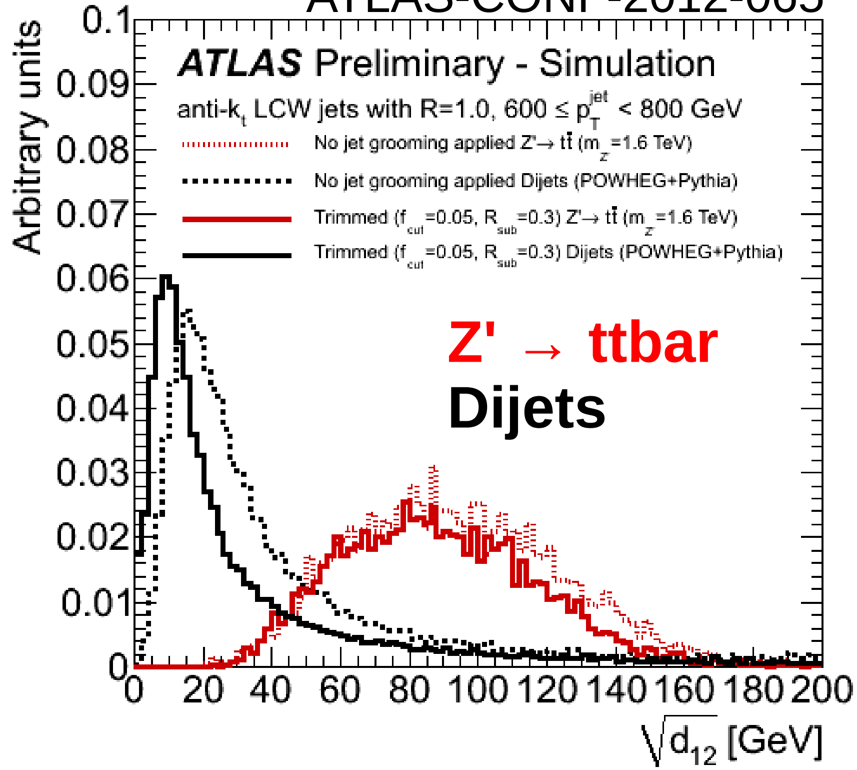
min p_T x dR = steeply falling spectrum

Boosted W boson



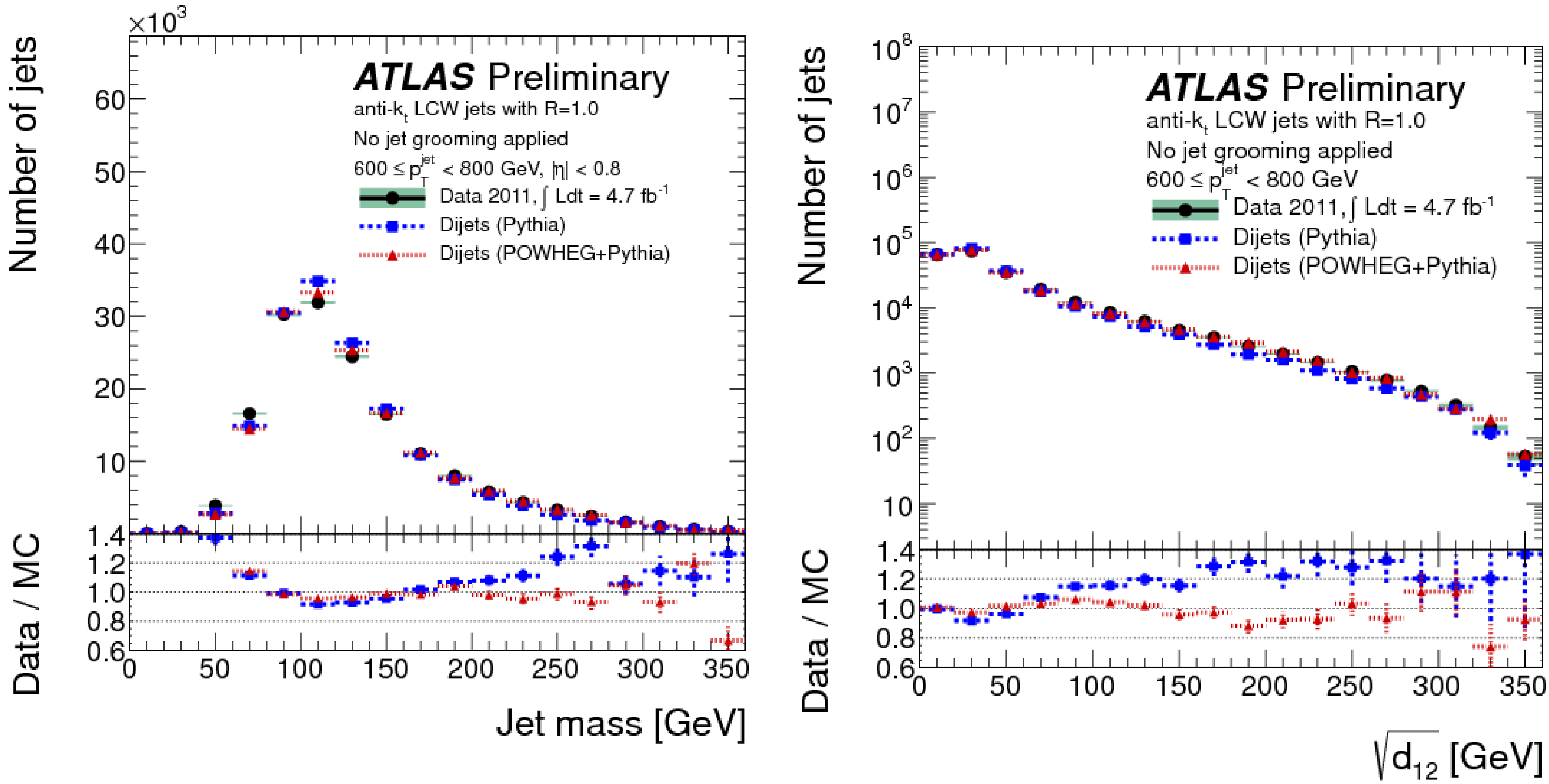
min p_T x dR = M_{jet} / 2

ATLAS-CONF-2012-065



Validation in data

- Mass and underlying structure much better described with the addition of NLO in shower calculations (PoWHEG+Pythia)

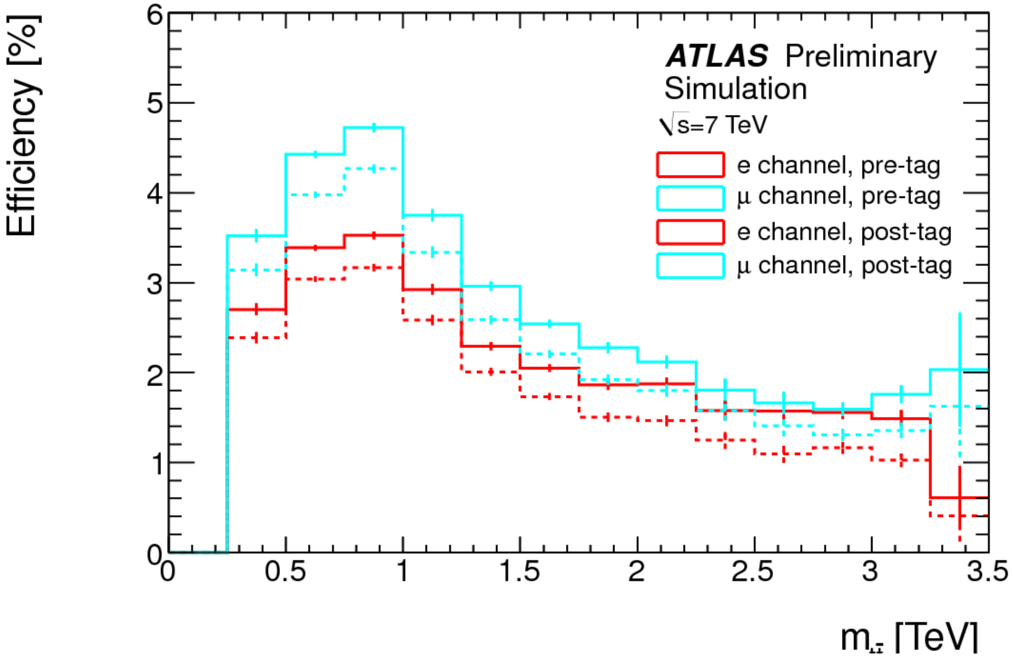


- Data driven jet mass scale uncertainty ~5% above jet p_T>350 GeV

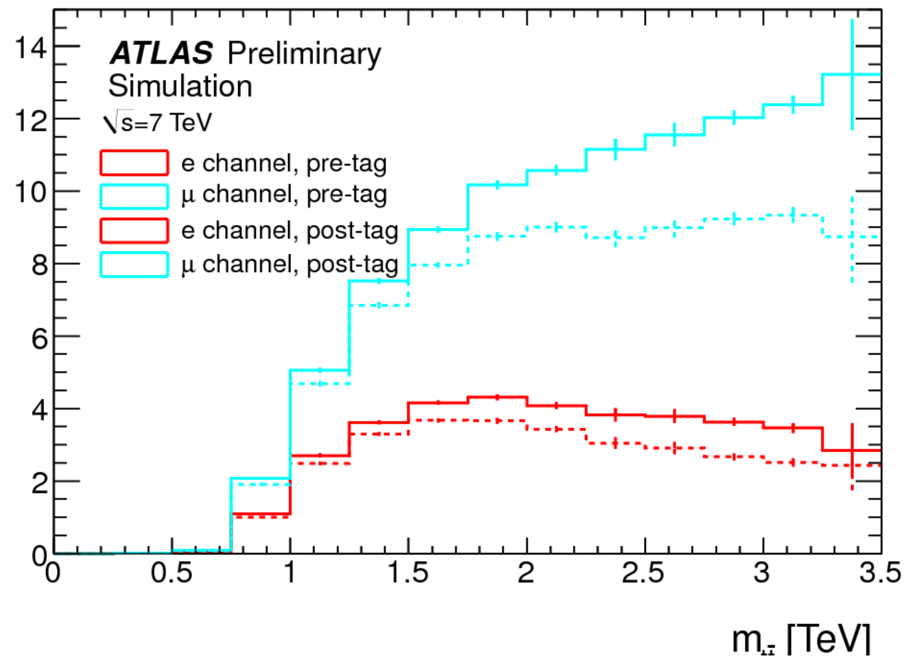
Recent ATLAS Results

- Two ttbar resonance searches using jet substructure in ATLAS:
  - semi-leptonic (ttbar→WbWb → qq'b lνb) (ATLAS-CONF-2012-136)
  - fully hadronic (ttbar→WbWb → qq'b qq'b) (arXiv:1211.2202, submitted to JHEP)
- Benchmark signals: Z' (heavy gauge boson), g_{KK} (RS KK gluons)
- Adding the boosted selection largely increases new physics sensitivity as Z' mass increases

"resolved"



"boosted"

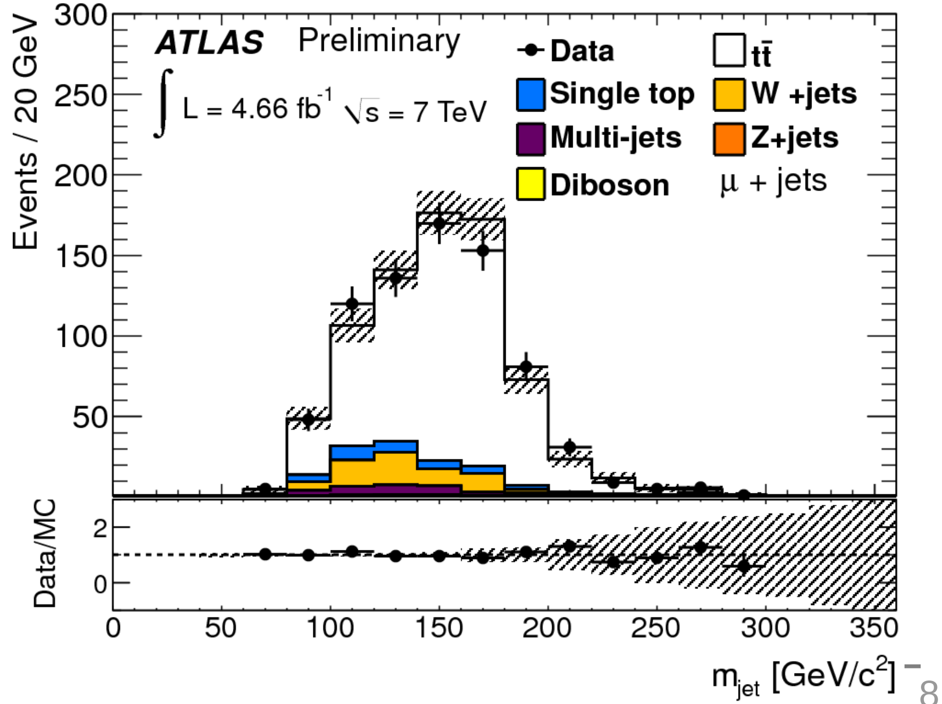


Semi-leptonic ttbar resonances

- Leptonic top selection:
  - "Mini-isolation" $I^{\ell}_{\text{mini}} = \sum_{\text{tracks}} p_T^{\text{track}}$

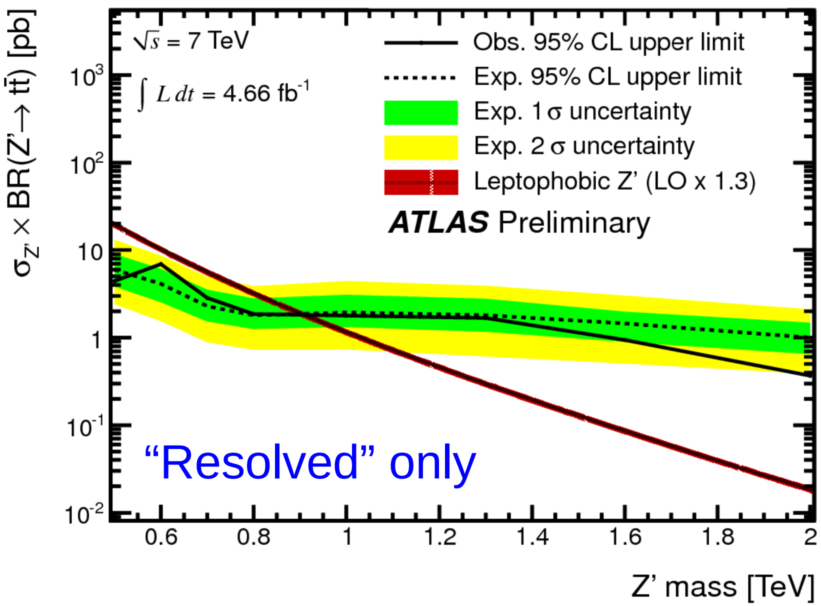
    (for all tracks with p_T>1 GeV and dR(lepton, track) < 10/p_T(lepton))
  - Select leptons with $I^{\ell}_{\text{mini}}/p^{\ell}_T < 0.05$
  - Event: at least one b-tagged (R=0.4) jet, missing energy requirement
- Resolved hadronic top selection:
  - Event level single-lepton trigger (mu or e)
  - 3 anti-k_t R=0.4 jets, p_T > 25 GeV
- Boosted hadronic selection:
  - Large-R jet trigger (fully efficient for R=1.0 jets with p_T>350 GeV)
  - 1 anti-kt R=1.0 jet with p_T > 350 GeV, M > 100 GeV, sqrt(d_{12}) > 40 GeV

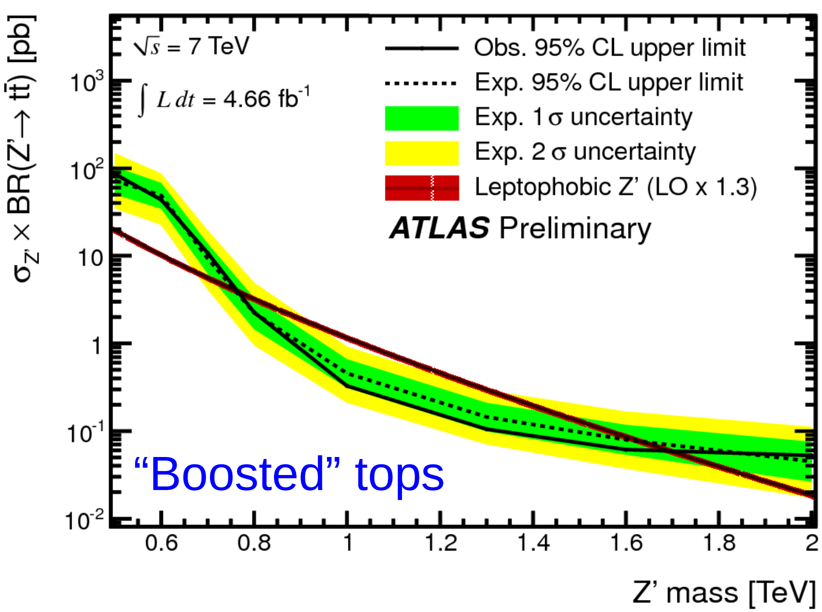
ATLAS-CONF-2012-136



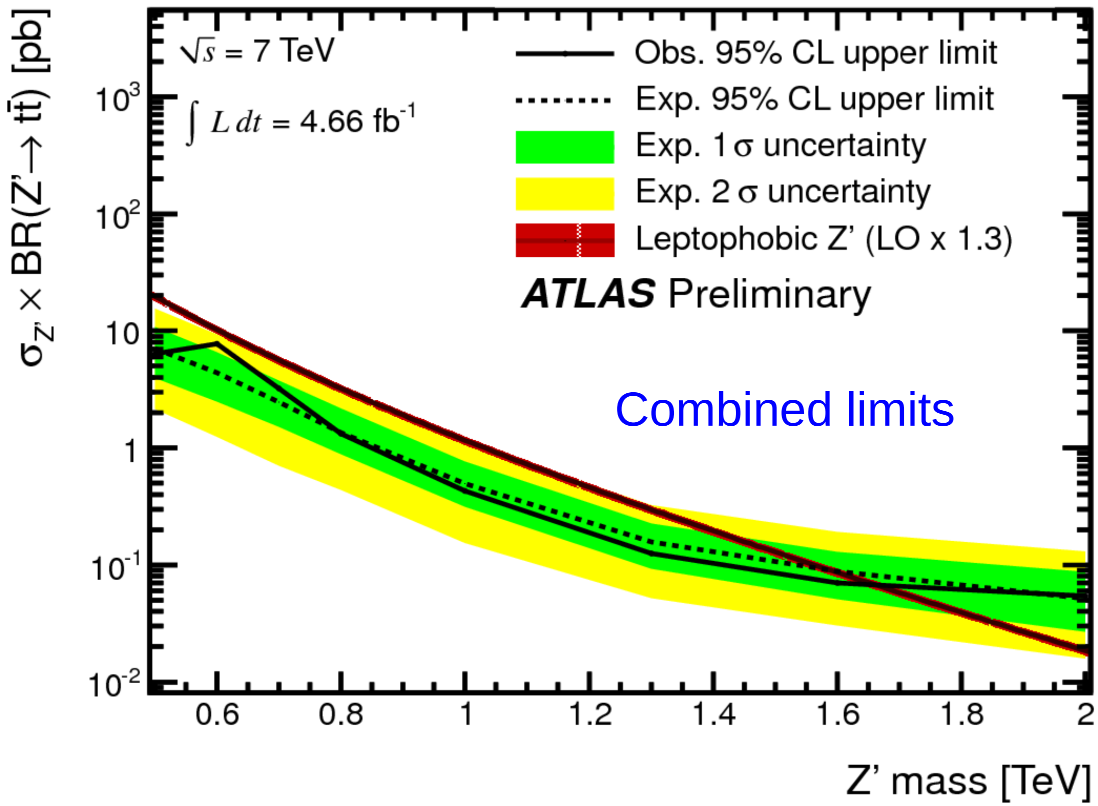
Semi-leptonic ttbar resonances

- With added boosted selection, Z' limit extended by ~800 GeV!





ATLAS-CONF-2012-136

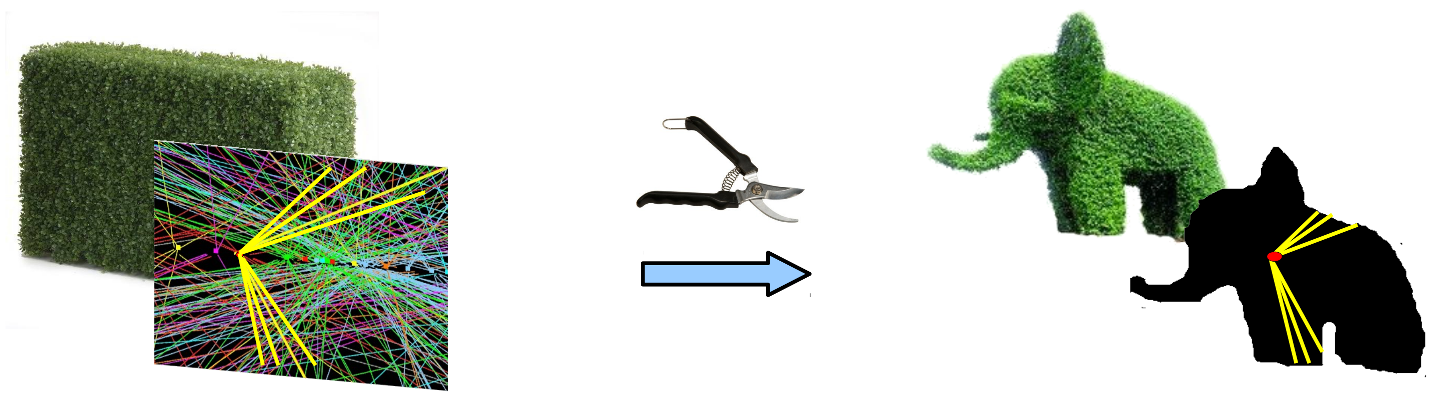


Z' exclusion: $0.5 < m_Z < 1.7$ TeV

g_{kk} exclusion: $0.7 < m_{gkk} < 1.9$ TeV

Turning up the luminosity...

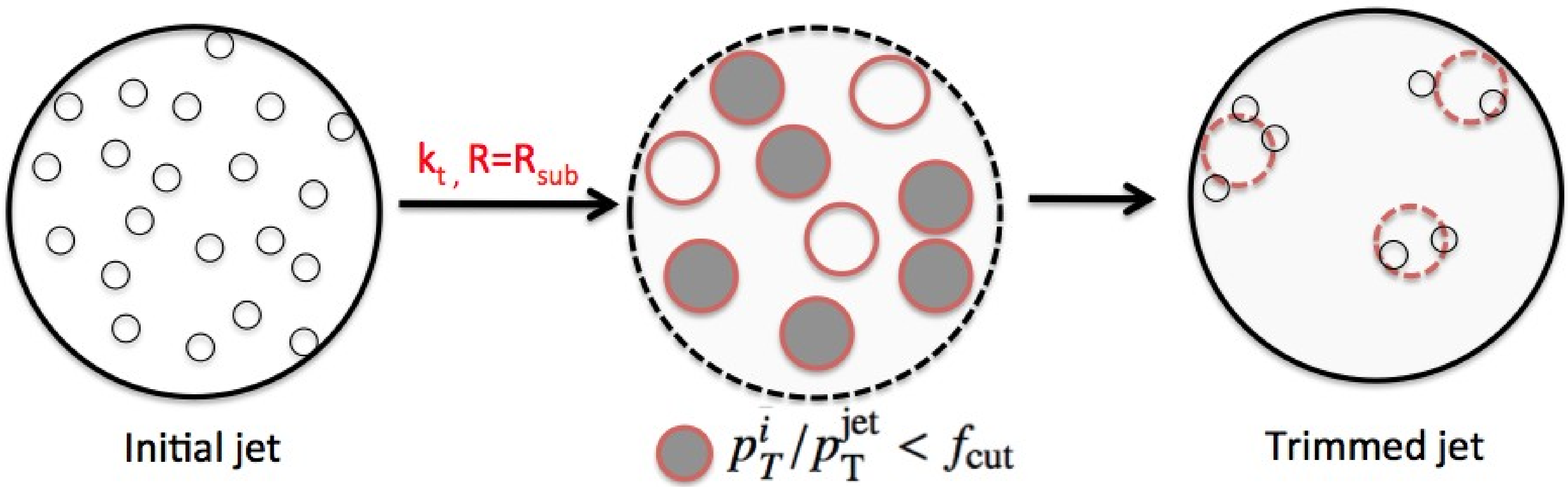
- Still a lot of unused information hidden in the substructure of a jet
- Problem: pileup and the underlying event can contaminate the jet, making it look like it has more structure than it does.
- **Jet grooming:** seeks to get rid of softer components in a jet from UE or pileup and leave constituents from the hard scatter behind
  - Better mass resolution expected after grooming
  - Mass less dependent on pileup
- Strategy: start with a large-R jet created using a standard algorithm (C/A, anti-k_t), then make choices whether or not to keep constituents (or even the whole jet) after grooming



Trimming

arxiv:0912.1342
(D. Krohn, J. Thaler, L. Wang)

- Contamination from pileup/UE is much softer than hard-scatter partons
- Uses k_t algorithm to create subjets of size R_{sub} from the constituents of the large-R jet
- Any subjets with too low pt compared to whole jet (f_{cut}) are removed

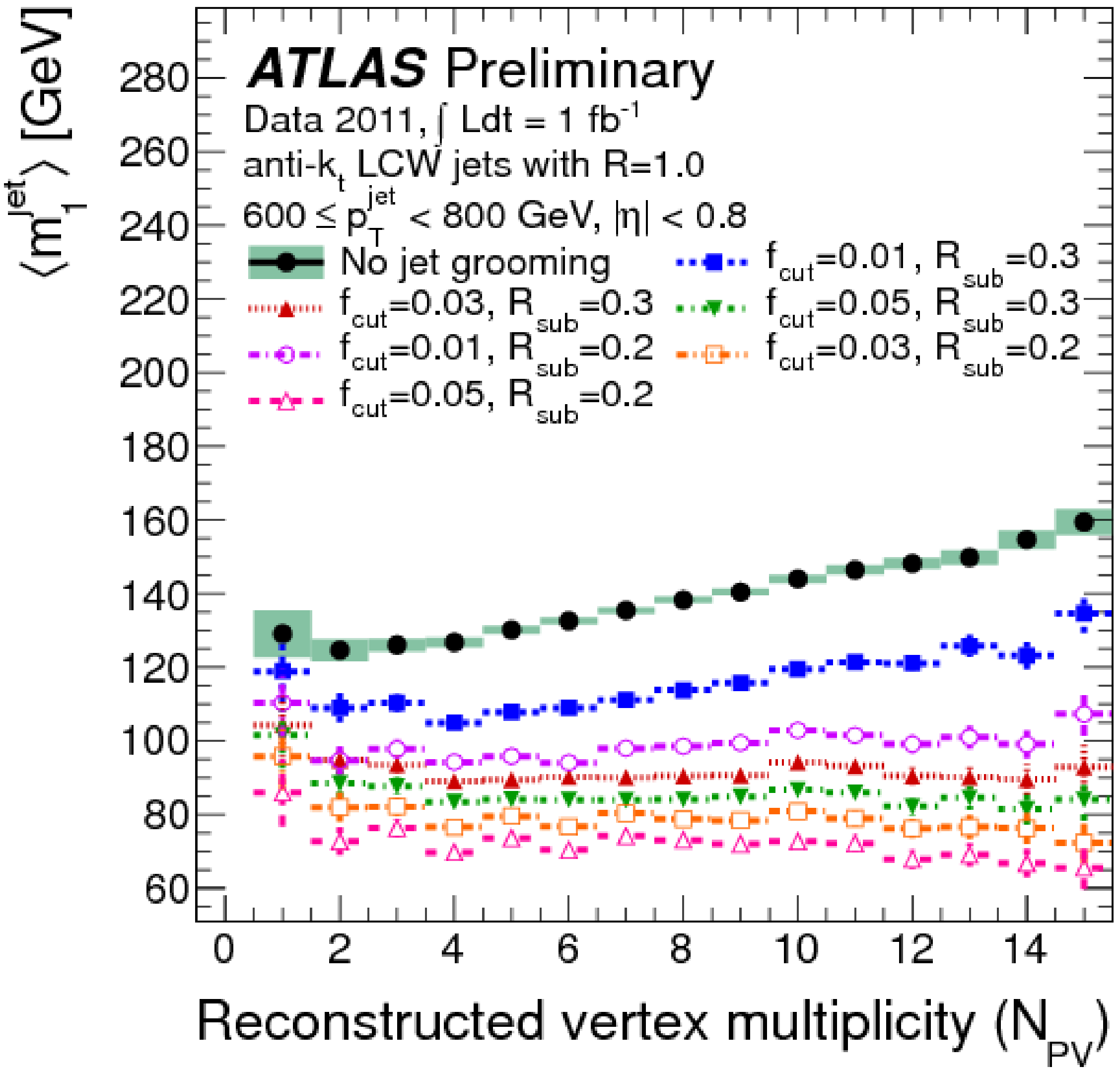


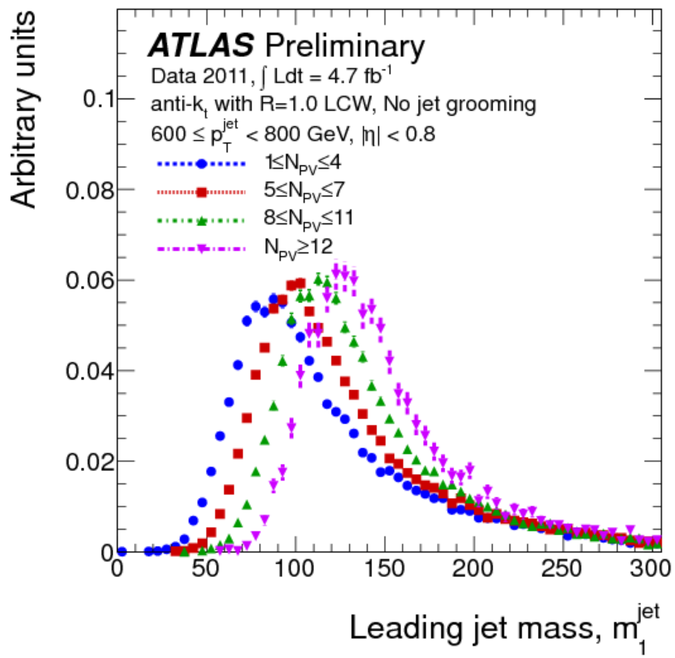
- Larger f_{cut} → tighter cut (bigger p_T fraction subjets are cut out)
- Smaller R_{sub} → tighter cut (smaller R carries less pt of the whole jet)

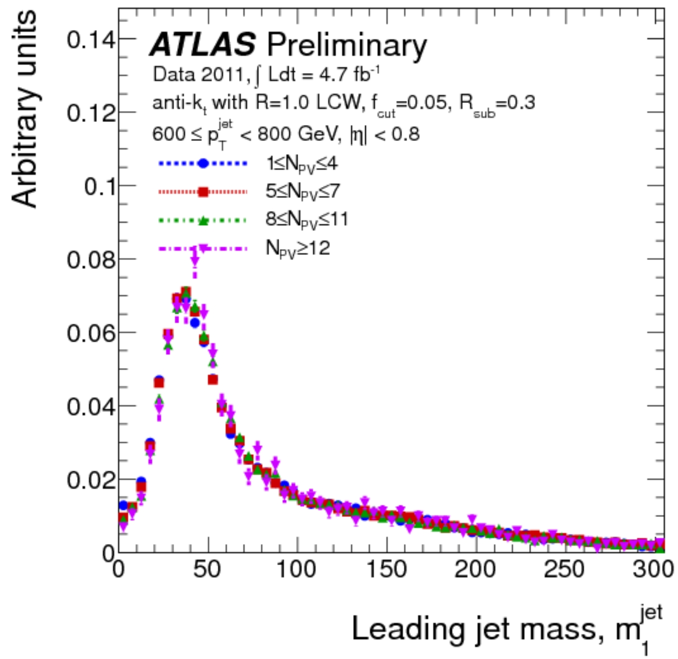
Impact of trimming in data

ATLAS-CONF-2012-066

- Trimming parameters were optimized in ATLAS



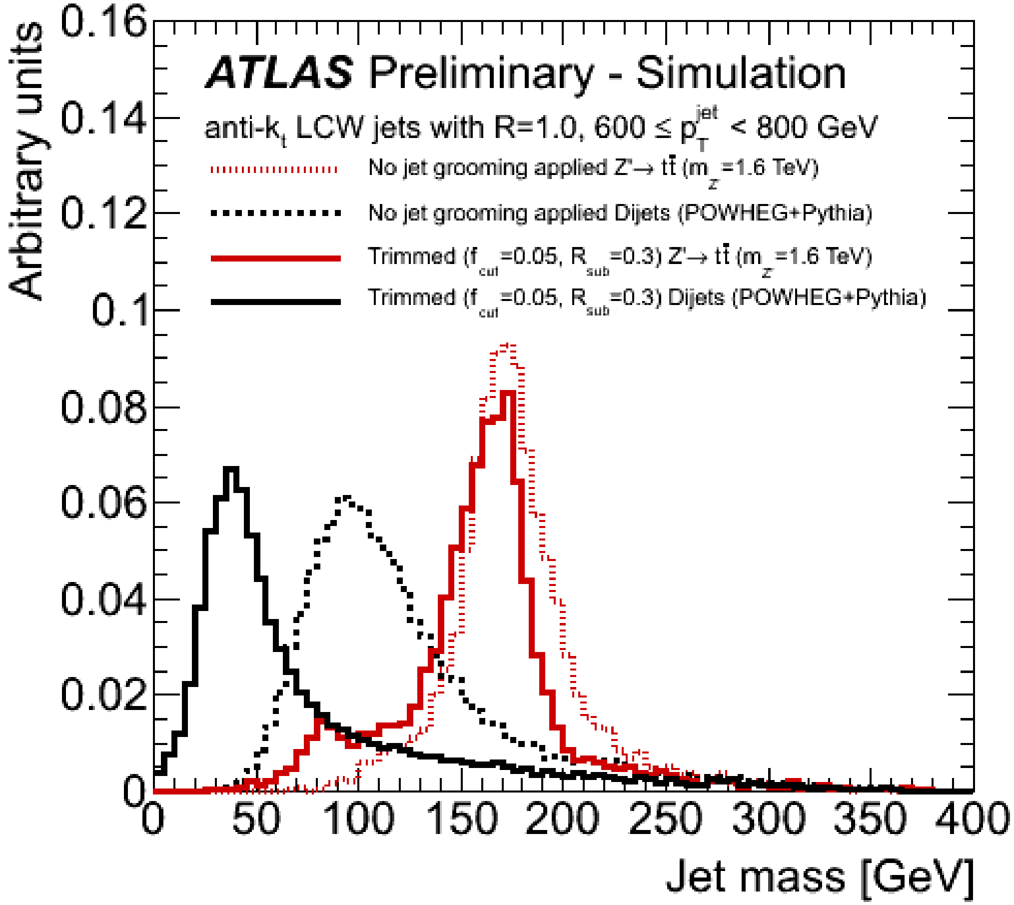




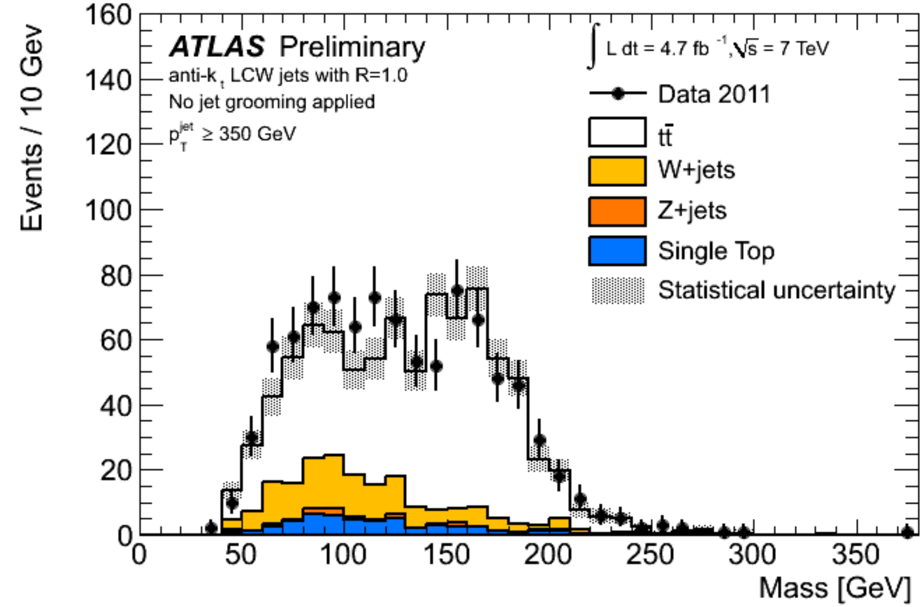
Trimming on the signal

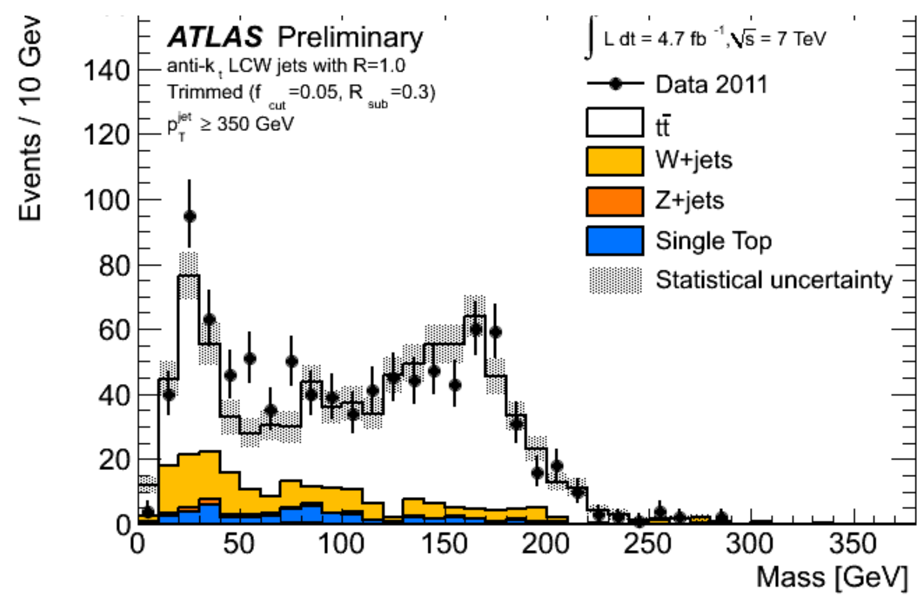
ATLAS-CONF-2012-065

- Mass signal peak remains relatively unaffected
- Background (non-substructure) jets are systematically shifted lower in mass



leading-pT anti-k_t R=1.0 jet mass in ttbar-enriched selection

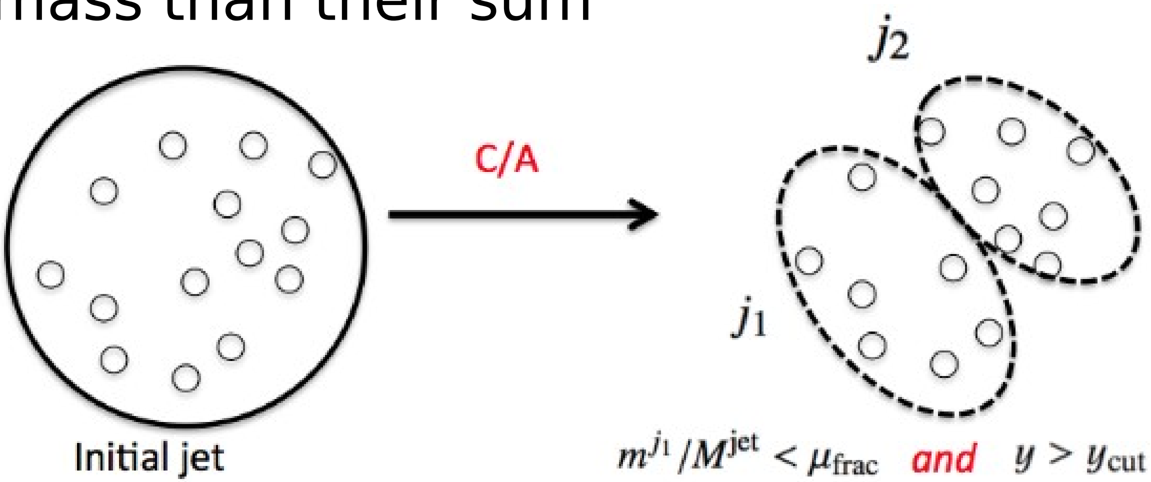




HEPTopTagger

arXiv:1006.2833
Plehn et al

- Example of how substructure/grooming techniques may be used to optimize the selection of hadronicly-decaying tops
- Utilizes a recursive "mass drop"/filtering approach
  - Was optimized for H→bb search using C/A jets   arxiv.org:0802.2470 (BDRS)
- Start with large-R (R=1.5, 1.8) Cambridge/Aachen jets, p_T>200 GeV
- Undo the last stage of C/A clustering to create two subjets
- Identify relatively symmetric subjets, each with significantly smaller mass than their sum

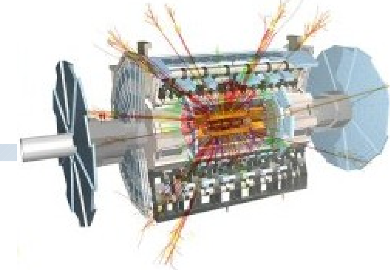


$$\frac{\min[(p_T^{j_1})^2, (p_T^{j_2})^2]}{(M^{\text{jet}})^2} \times \Delta R^2_{j_1, j_2} > y_{\text{cut}}$$

...continue undoing last steps of C/A until all subjets have m<50 GeV

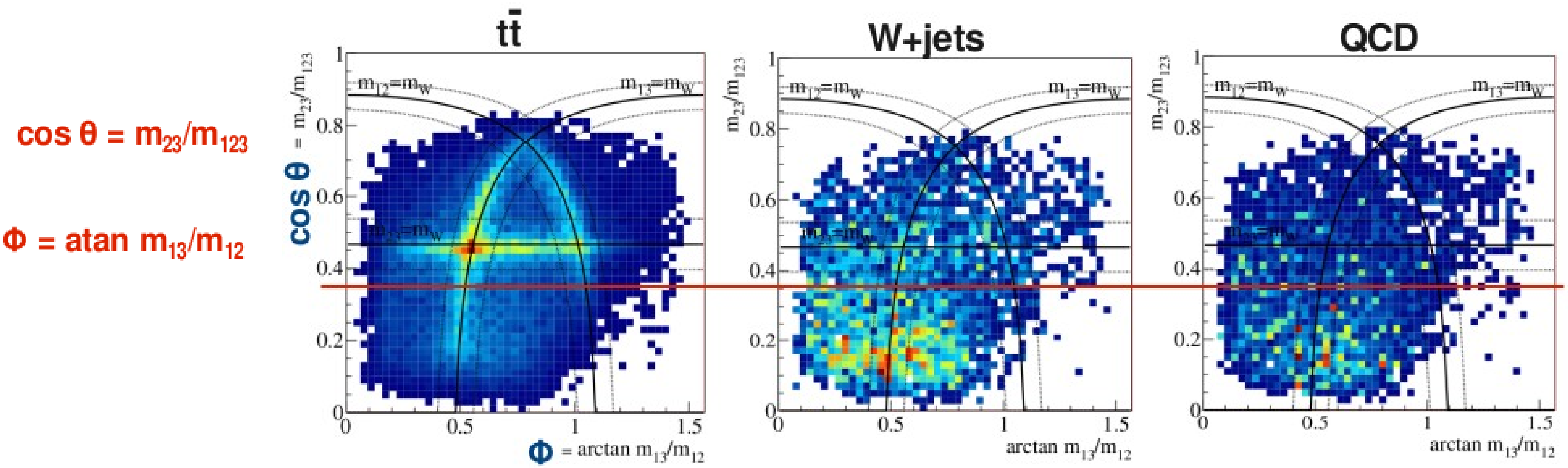
HEPTopTagger

arXiv:1006.2833
Plehn et al



- Test triplets of subjets for compatibility with top quark
    - Form (tiny) C/A jets out of topoclusters from particular subjet triplet and “filter” to remove pileup/underlying event contamination
    - Re-cluster again exclusively into exactly three jets
- Reconstruct hadronic top from combinations of final subjets and identify top quark via subjet mass ratios (subjets are calibrated)

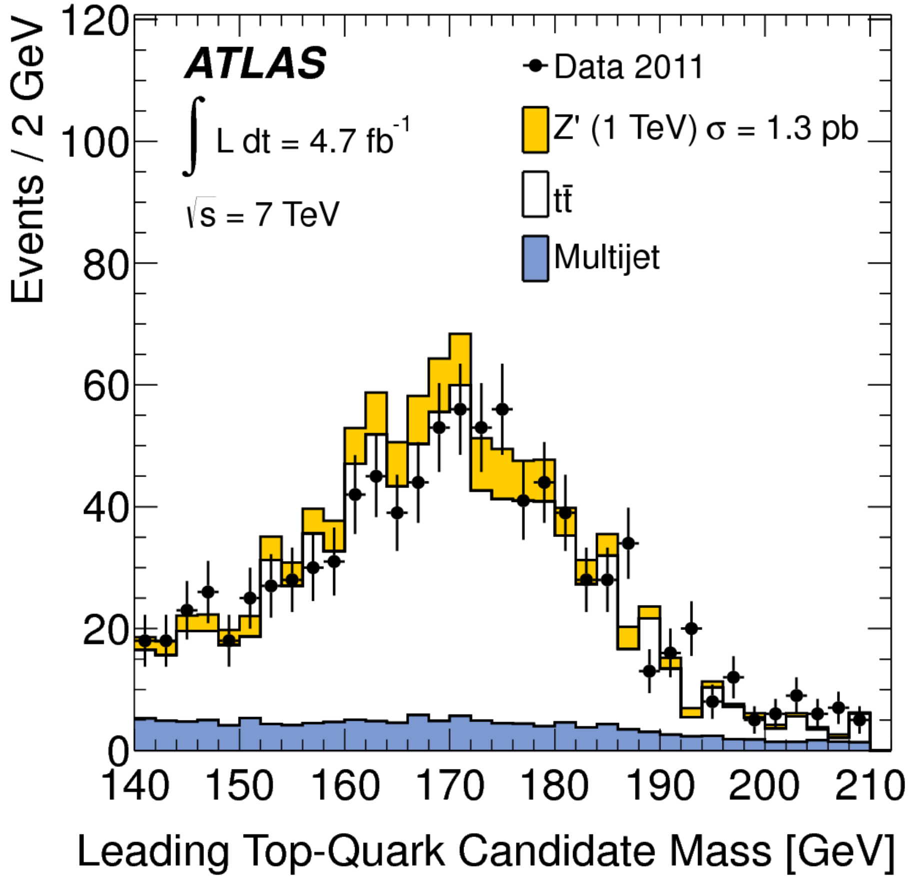
$$\text{for } m_i^2 \approx 0: \quad m_{123}^2 \approx m_{12}^2 + m_{13}^2 + m_{23}^2$$

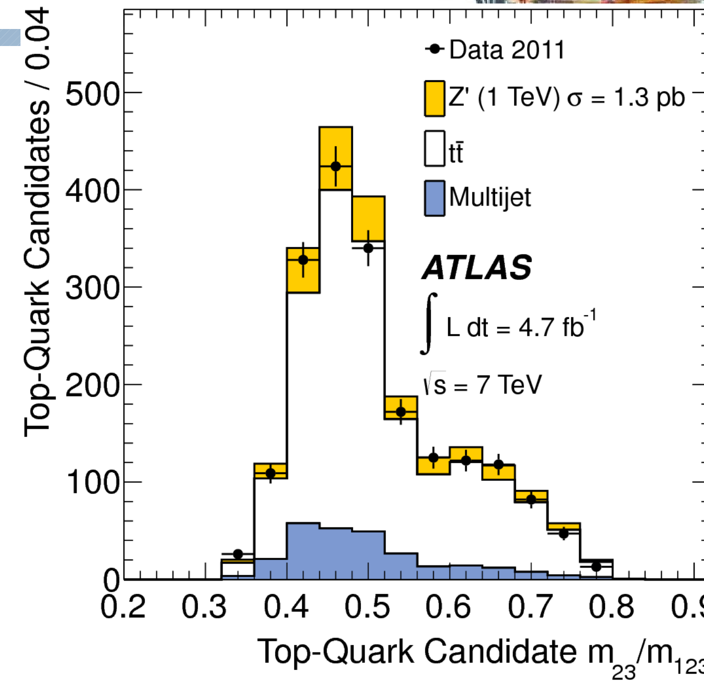


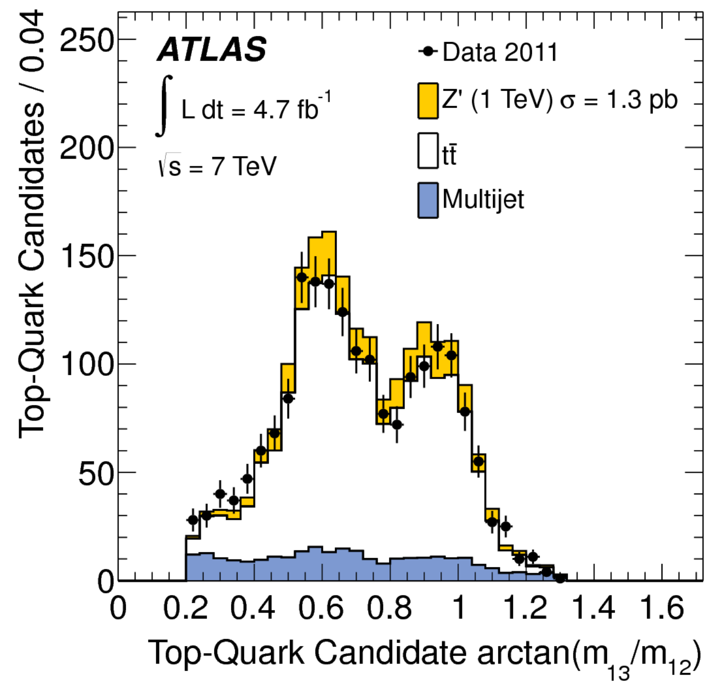
HEPTopTagger

arXiv:1211.2202, submitted to JHEP

- Fully hadronic ttbar resonance search
- Dominant dijet background





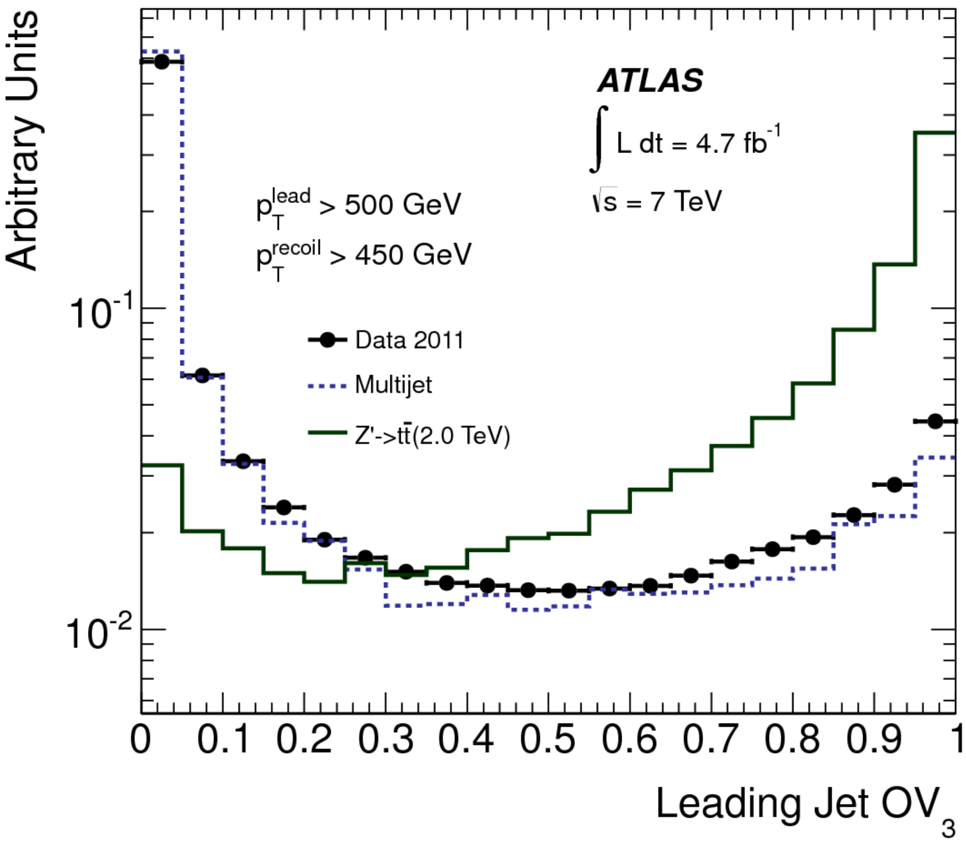


- 953 events observed in signal region (expected 770^{+220}_{-180} ttbar events, 130±70 dijet events)

Top Templating

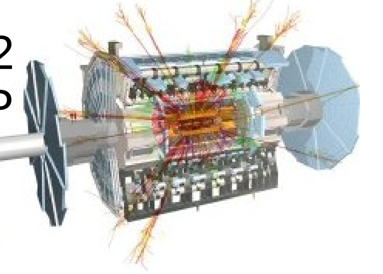
arXiv:1211.2202
submitted to JHEP

- Complimentary approach to HEPTopTagger for fully-boosted fully-hadronic ttbar searches
- Energy flow inside a jet should match the 4-momenta of top quark decay products at the parton-level
- Look at all parton templates at a given p_T, and find the template that maximizes this overlap (OV_3)
- In boosted ttbar selection, $OV_3 > 0.7$ is required on both leading and recoil jet
- Narrower cone-size used (R=1.0), optimized for higher-p_T jets (pT > 450 GeV)
- 123 observed in signal region (expected 59^{+27}_{-26} ttbar events, 53±6 dijet events)



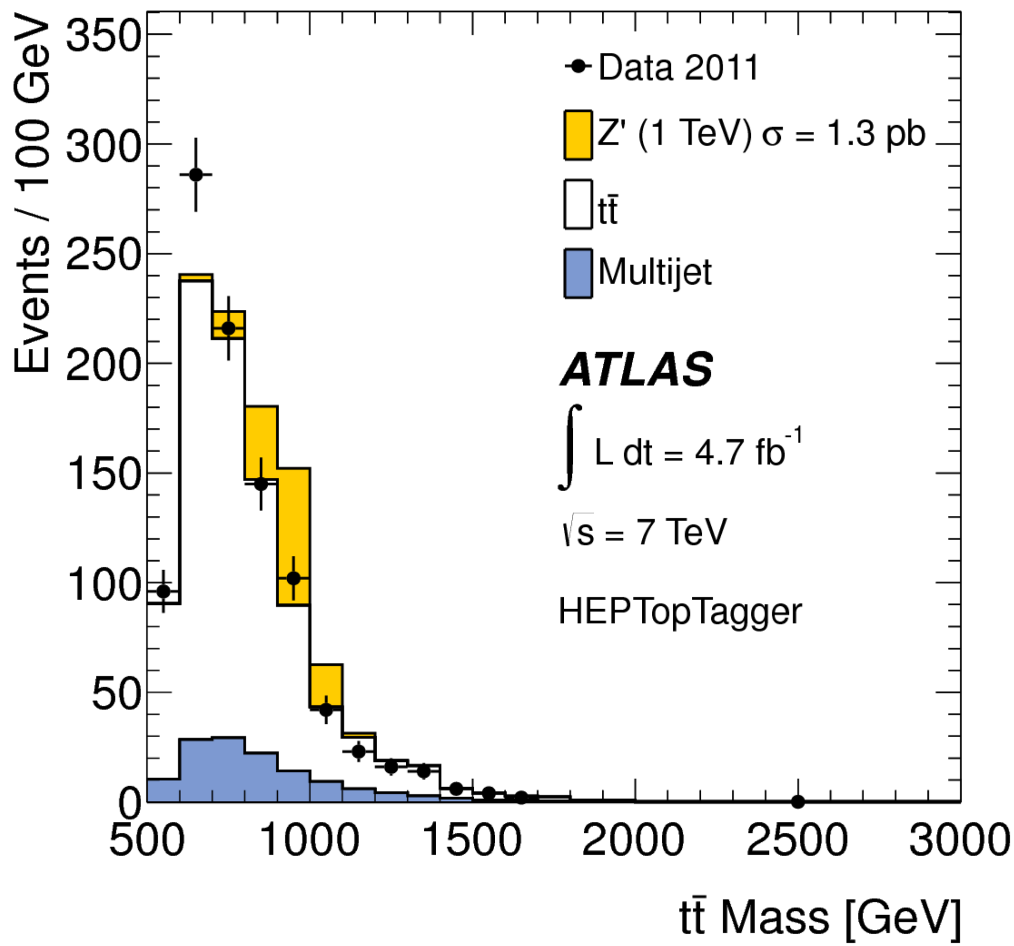
Fully hadronic ttbar

arXiv:1211.2202
submitted to JHEP



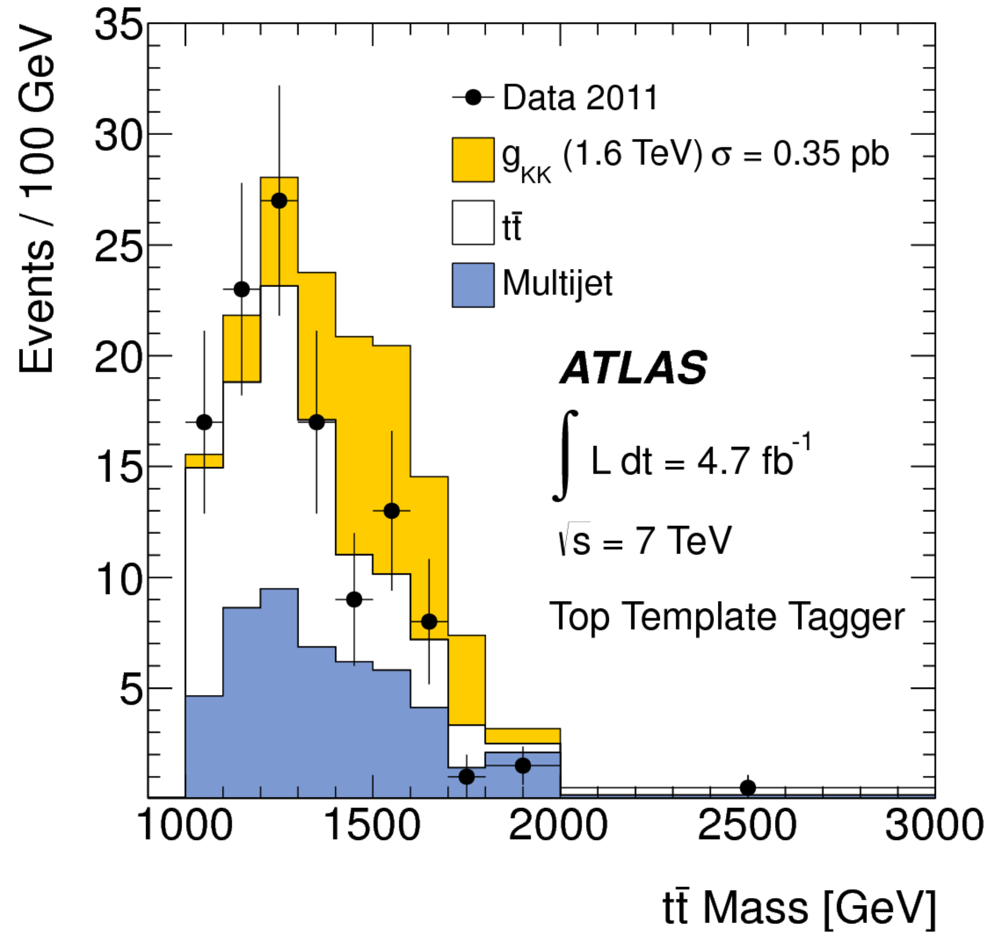
- Results

HEPTopTagger



M(Z') = 1 TeV shown

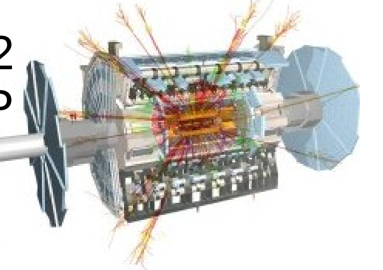
Top Templating



$M(g_{kk})$ = 1.6 TeV shown

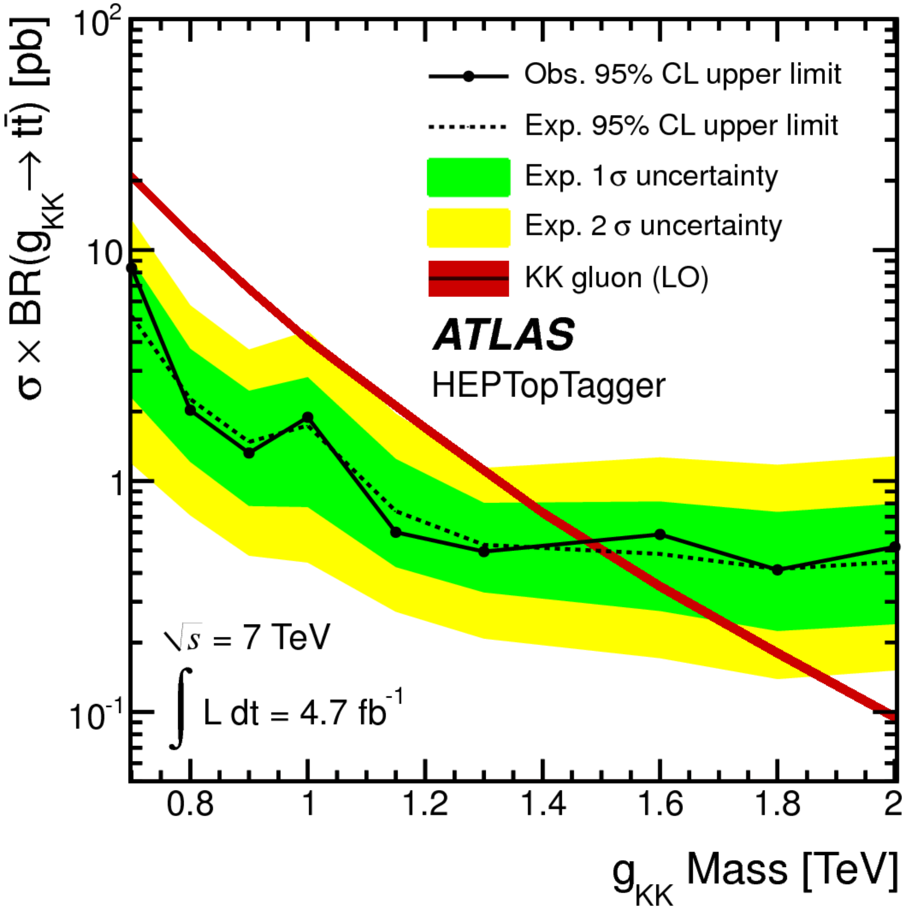
Fully hadronic ttbar

arXiv:1211.2202
submitted to JHEP



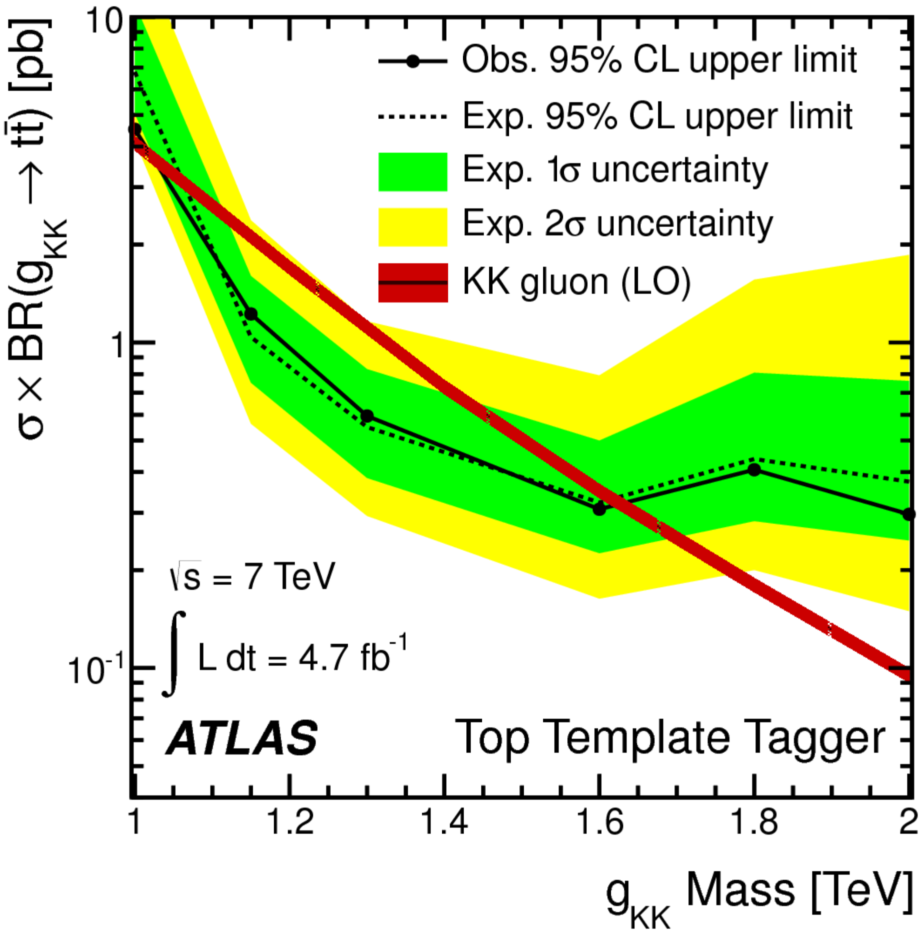
- Results

HEPTopTagger



Z' exclusion: $0.7 < m_Z < 1.32$ TeV
g_{kk} exclusion: $0.7 < m_{gkk} < 1.48$ TeV

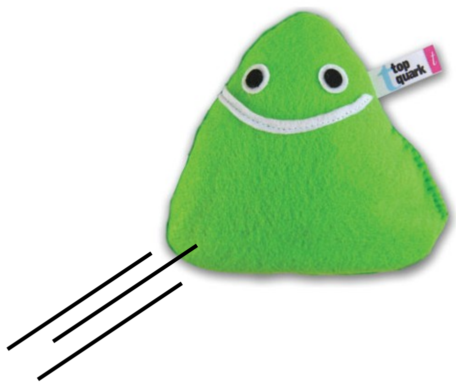
Top Templating



g_{kk} exclusion: $1.02 < m_{gkk} < 1.63$ TeV

Conclusions and outlook

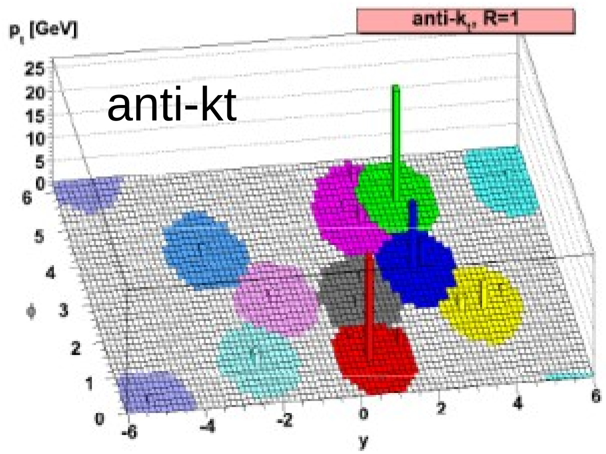
- Jet substructure can be a powerful tool to discriminate between a dominant QCD background and heavy particle reconstruction
- Boosted selection → increased signal sensitivity
- Up and coming jet reconstruction: grooming
  - Reduced sensitivity to underlying event and pileup
  - Mass signal peak remains relatively unaffected (and resolution improves!), but background non-substructure jets are systematically shifted lower in mass
  - Lots of performance/optimization studies on jet grooming and substructure in ATLAS already, now being implemented in analyses
- Ready for 8 TeV luminosity conditions and beyond! Stay tuned!

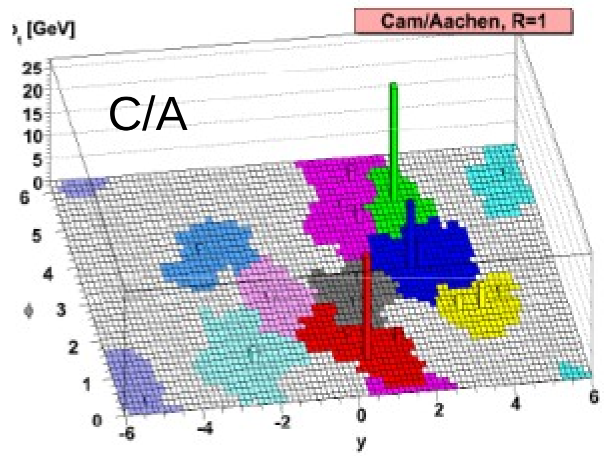


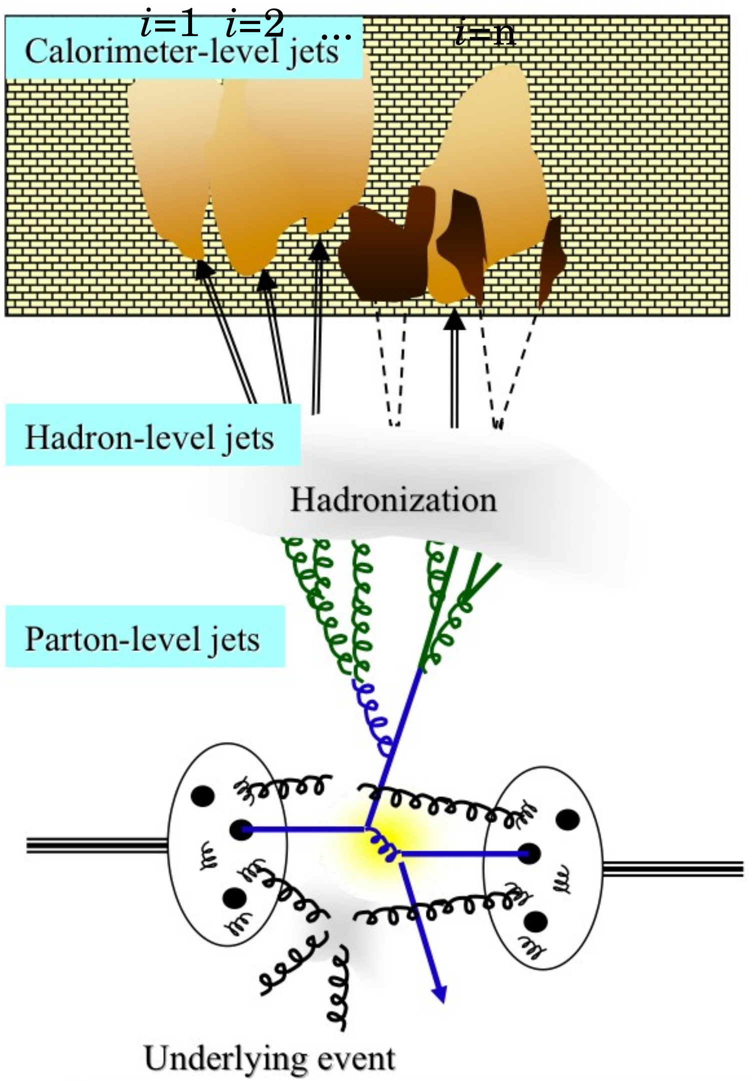
backup

Jets in ATLAS

- Jet Reconstruction: iterative recombination algorithms combine pairs of protojets
  - $d_{ij} = \min(k_{t,i}^{2p}, k_{t,j}^{2p})\ (\Delta\phi_{ij}^2 + \Delta\eta_{ij}^2)$ (while $d_{ij} < R$)
  - kt: p = 1 (smallest pT clusters first)
  - anti-kt: p = -1 (largest pT clusters first)
  - Cambridge-Aachen: p = 0 (angle-ordered)

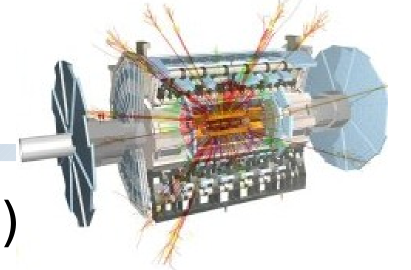






- Inputs to jet reconstruction:
  - “Truth jet”: Stable truth particles (lifetime > 10 ps)
  - “Reconstructed jet”: 3D topological clusters of calorimeter cells
- Calibration on jet mass and energy in simulation for large-R jets (R=1.0, 1.2)

Jet mass scale uncertainty

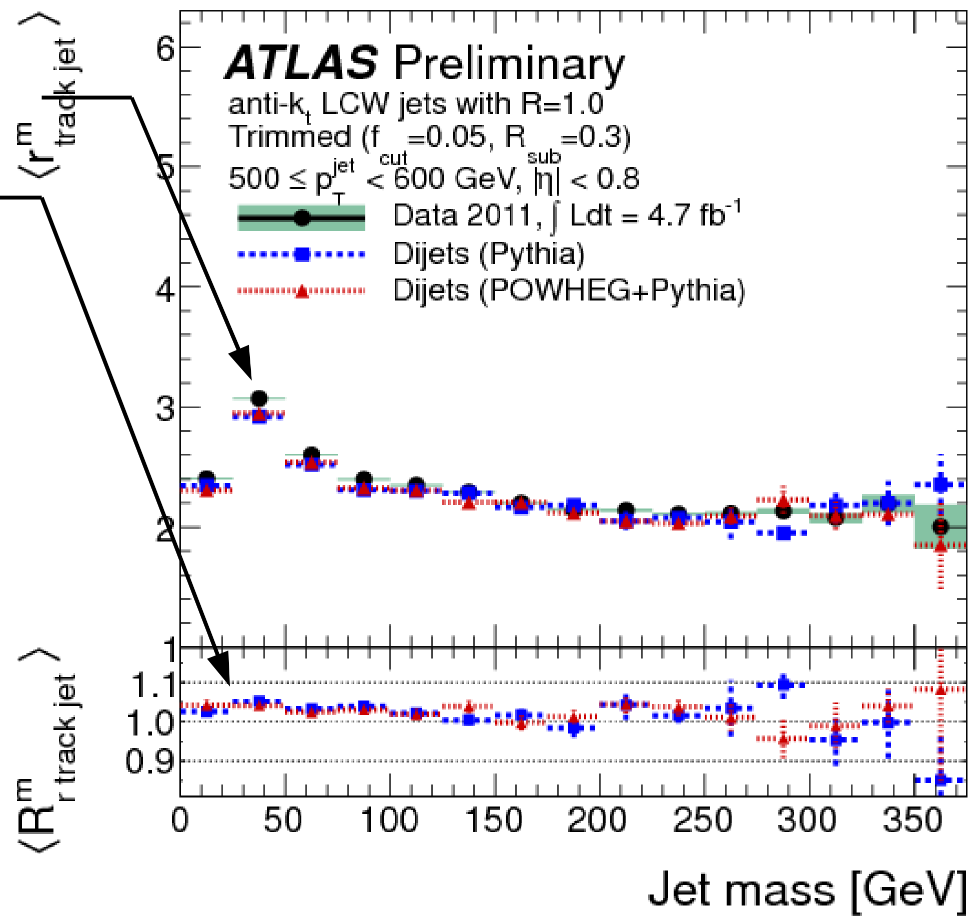


- Build jets out of charged tracks (minimal pileup dependence)
- **Double ratio method** for relative calorimeter-based jet mass scale uncertainty (track jet measurement largely uncorrelated)

  $r^{m}_{\text{track jet}}$ = calo jet / track jet

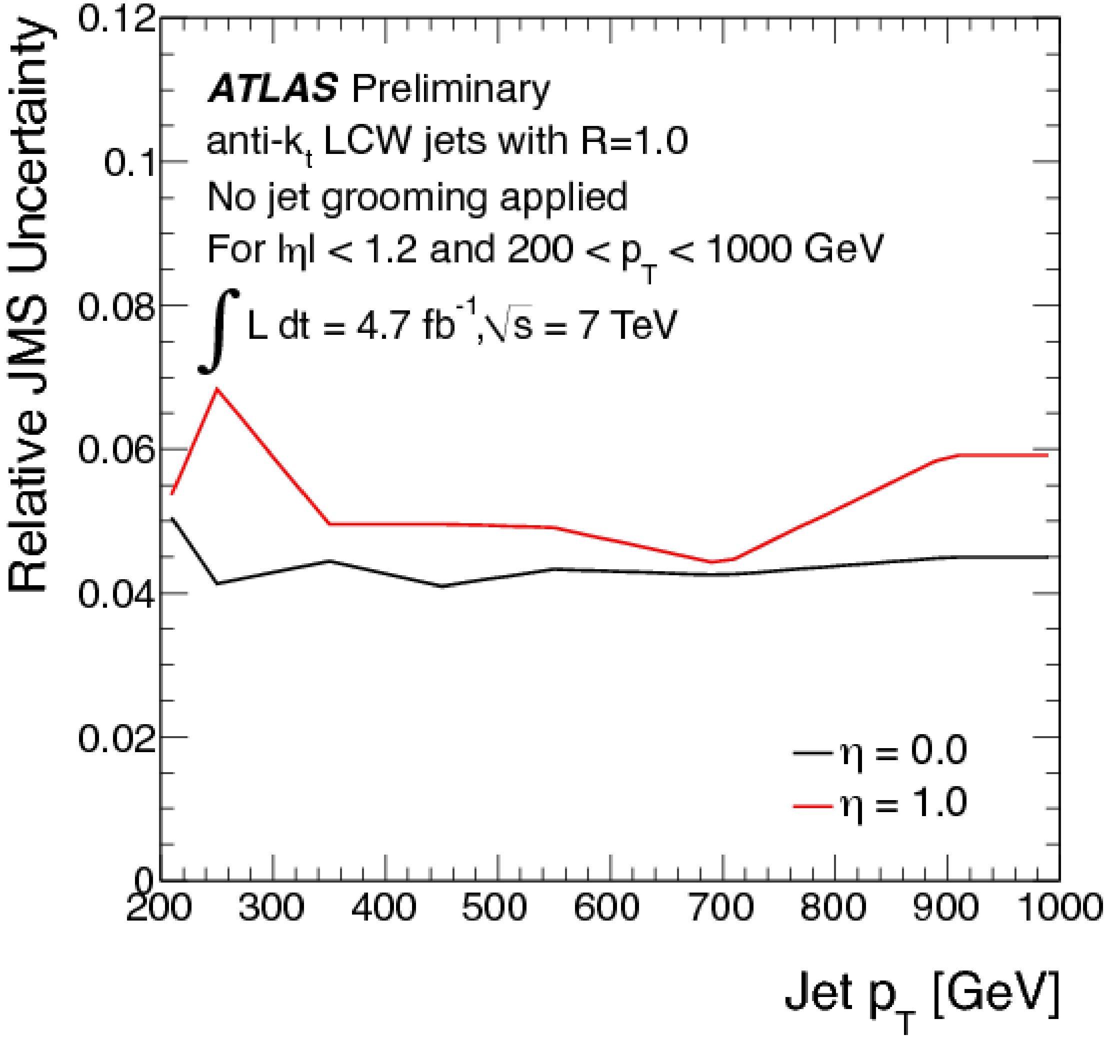
- Take mean of $r^{m}_{\text{track jet}} = R_{\text{data}}$ or R_{MC}
- Compute $R_{\text{data}}/R_{\text{MC}}$
- Systematic is deviation from 1, taking the statistical uncertainty of the calo-to-track jet ratio into account
  - weighted average (w_bin is the stat uncertainty):

$$\langle \delta_{\text{MC}} \rangle = \frac{\sum_{\text{bins}} w_{\text{bin}} \left(R^{m}_{r\text{track}-\text{jet}} - 1 \right)}{\sum_{\text{bins}} w_{\text{bin}}}$$

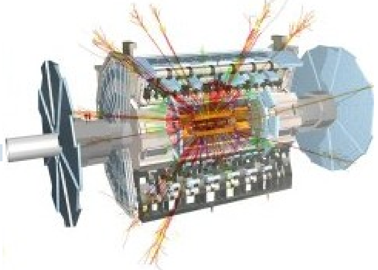


Jet mass scale uncertainty

- JMS is fairly stable above jet pT > 350 GeV



Fully hadronic ttbar background



- Top Templating analysis QCD background estimation
  - b: b tagged
  - t: top tagged with OV3 > 0.7

P = signal region

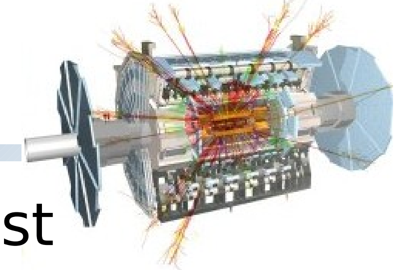
Recoil Jet \ Leading Jet	no-tag	t	b	t + b
t + b	J	K	L	**P**
b	B	D	H	N
t	E	F	G	M
no-tag	A	C	I	O

$$K' = N_J \times \frac{N_F}{N_E}$$

$$M' = N_F \times \frac{N_O}{N_C}$$

$$P' = K' \times \frac{M'}{N_F} = \frac{N_J \times N_O \times N_F}{N_E \times N_C}$$

N-subjettiness



- "To what degree a jet can be regarded as having at least N subjets" (Thaler and Van Tilburg, arxiv:1011.2268)
- Use k_t to recluster jet constituents into exactly N subjets
- τ_N: pt weighted distance of cluster to closest axis.

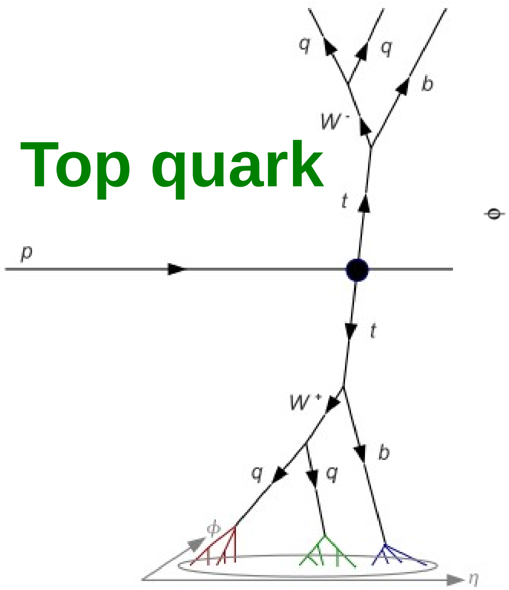
  $\tau_N \rightarrow 0$ means that the jet is well described as having N subjets

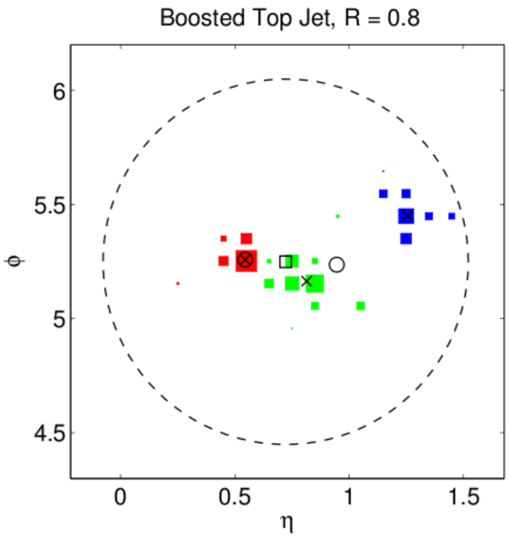
  $\tau_N \rightarrow$ large = clusters are distributed randomly

$$\tau_N = \frac{\sum_k p_{T,k} \min\{\Delta R_{1,k}, R_{2,k}, \cdots, R_{N,k}\}}{\sum_k p_{T,k} R_0}$$

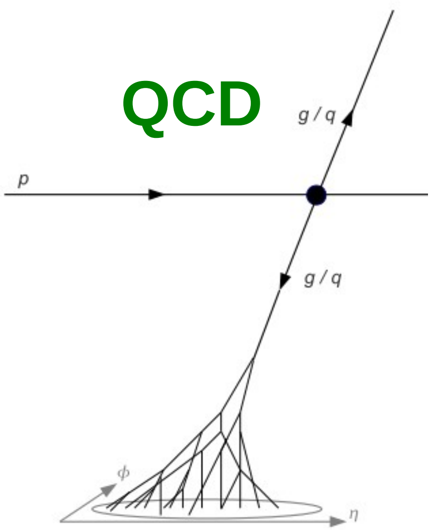
(for k constituent particles)

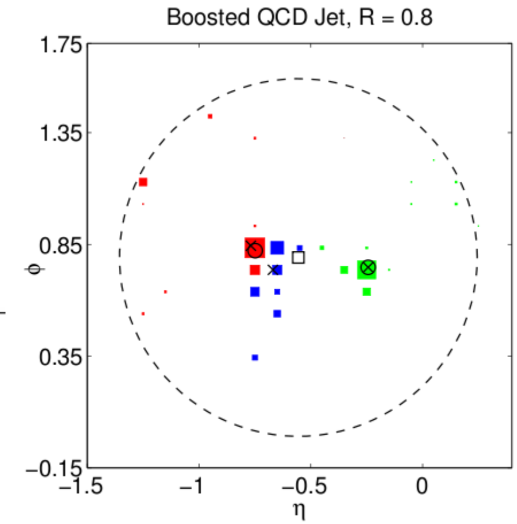
Top quark



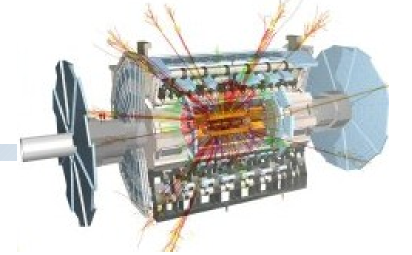


QCD

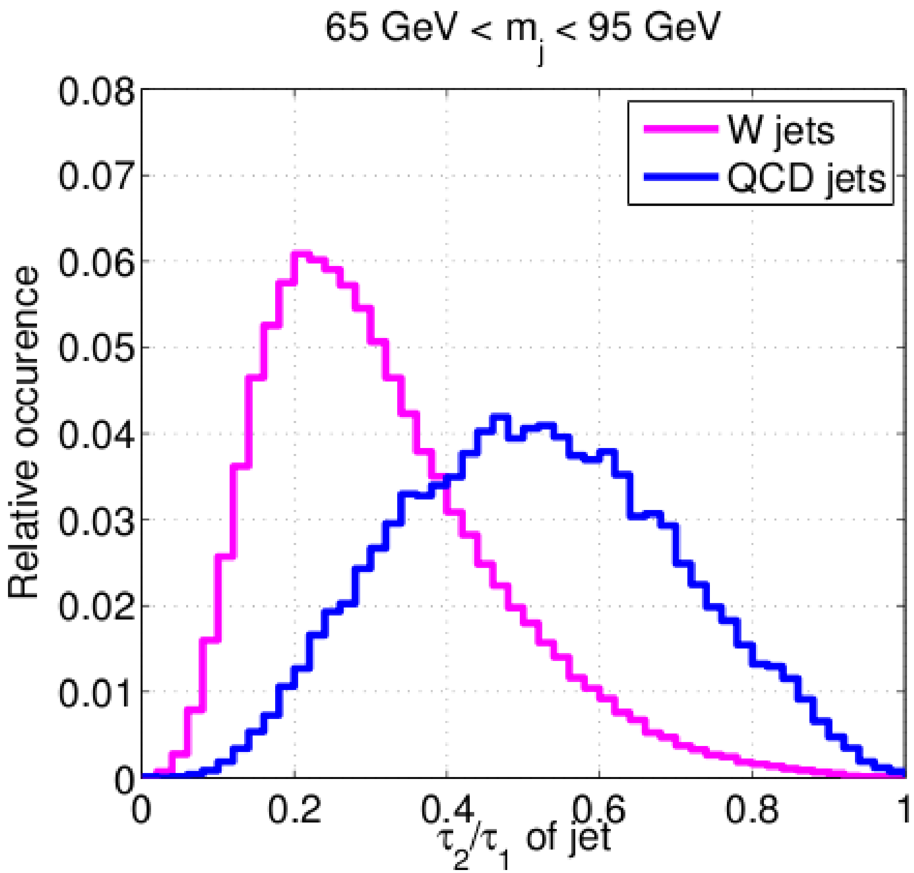


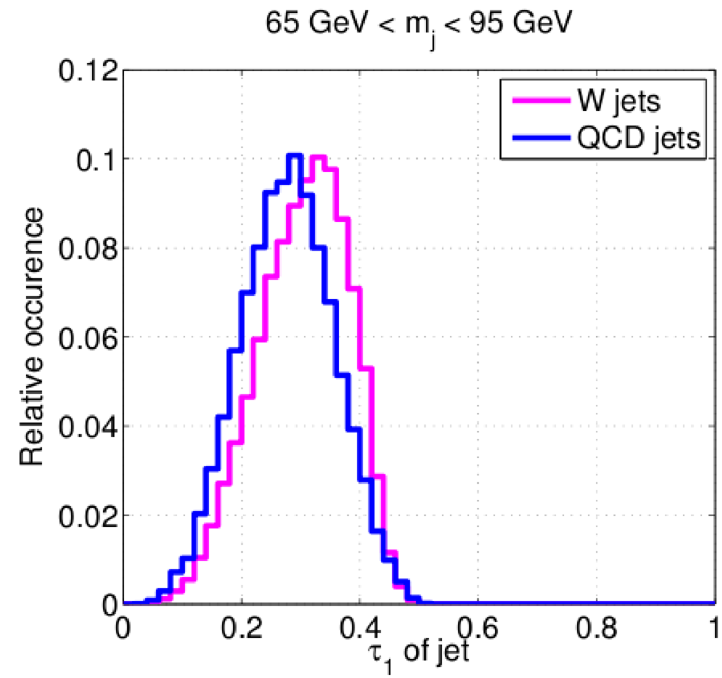


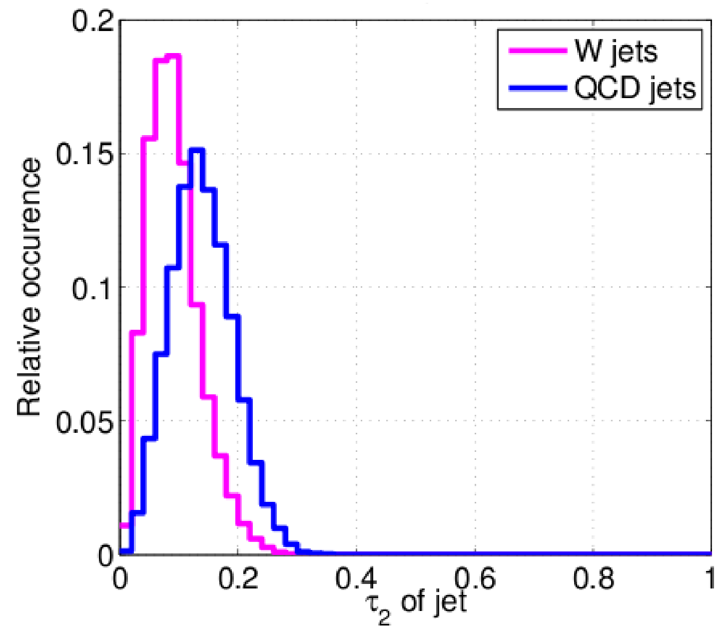
N-subjettiness



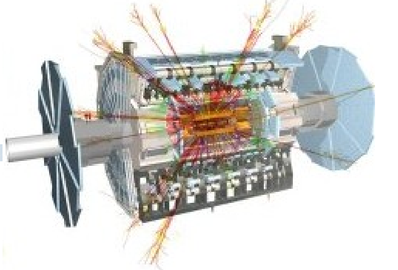
- QCD jets with large values of τ_1 generally also have large values of τ_2
- Good discriminating variable for 2-pronged decay: τ_2/τ_1 (or τ_3/τ_2 for top decay)







Impact of trimming on top taggers



- Nsubjettiness as a top tagger
- Trimming reduces QCD mis-tag rate (in the top mass window)

800 < pT < 1000 GeV (100 < M < 250 GeV)