Bundesministerium
für Bildung
und Forschung

DFG

Measurement of HEPTopTagger performance in the semi-leptonic top decay channel

G. Kasieczka
in consultation with
C. Anders, S. Schätzel, A. Schöning, D. Sosa

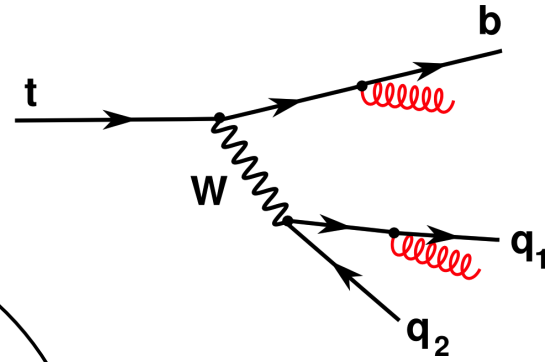
Alliance Top Quark Workshop, 22-23.03.2012

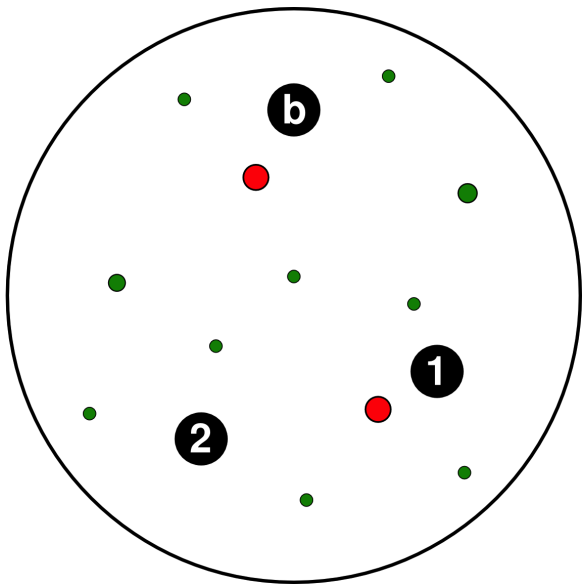
Overview

- Goal: event based identification of boosted tops for New Physics searches
- First: Study HEPTopTagger on full 2011 ATLAS data
  - Compare with MC
  - Stability with pileup
  - Substructure
  - Loose/Tight settings
- Additionally: use MC for
  - Signal/Fake efficiency
  - Mass and p_T scales and resolutions

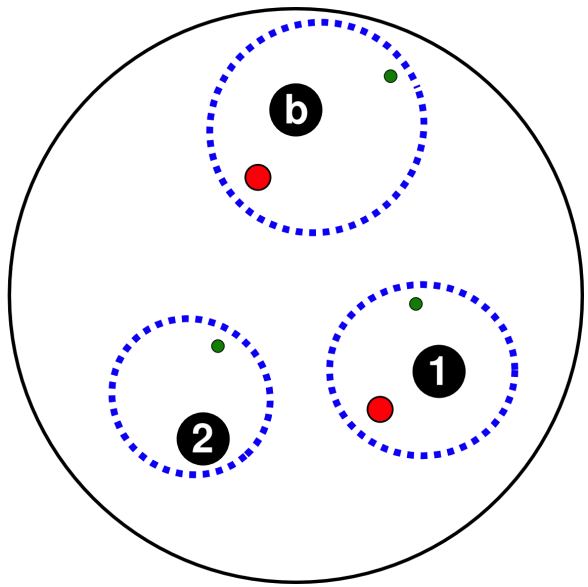
Algorithm

- **HEPTopTagger: Finds and reconstructs hadronically decaying top quark from the substructure of a Cambridge/Aachen (C/A) Fatjet with R=1.5**





fatjet before grooming

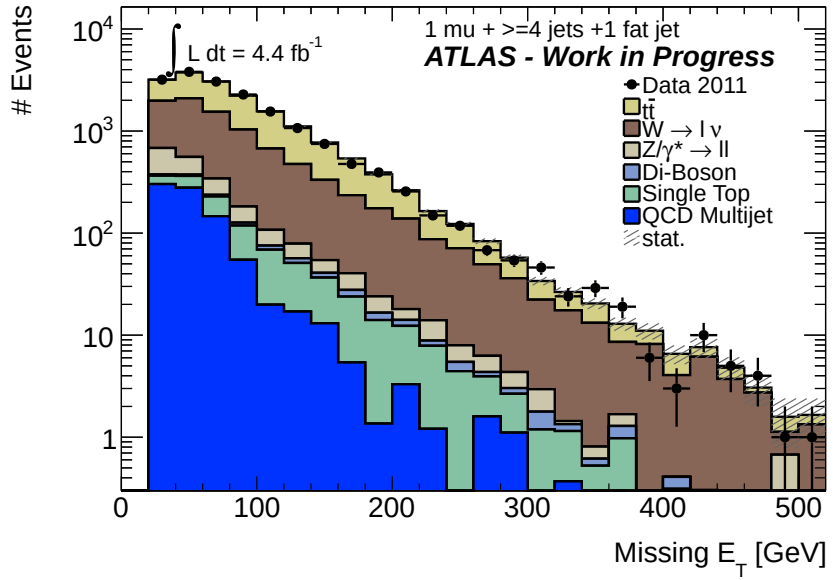


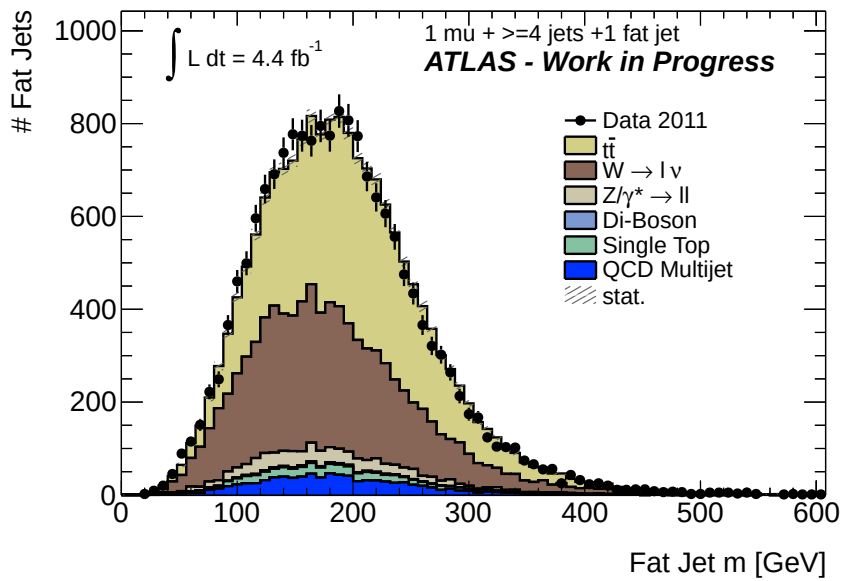
fatjet after grooming

Many thanks to T. Plehn, M. Spannowsky and M. Takeuchi

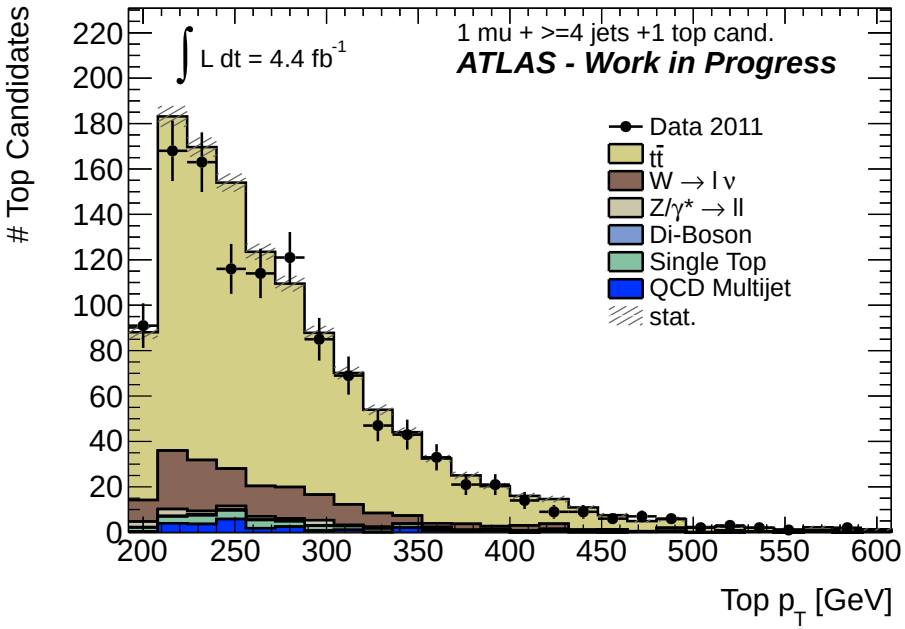
Selection

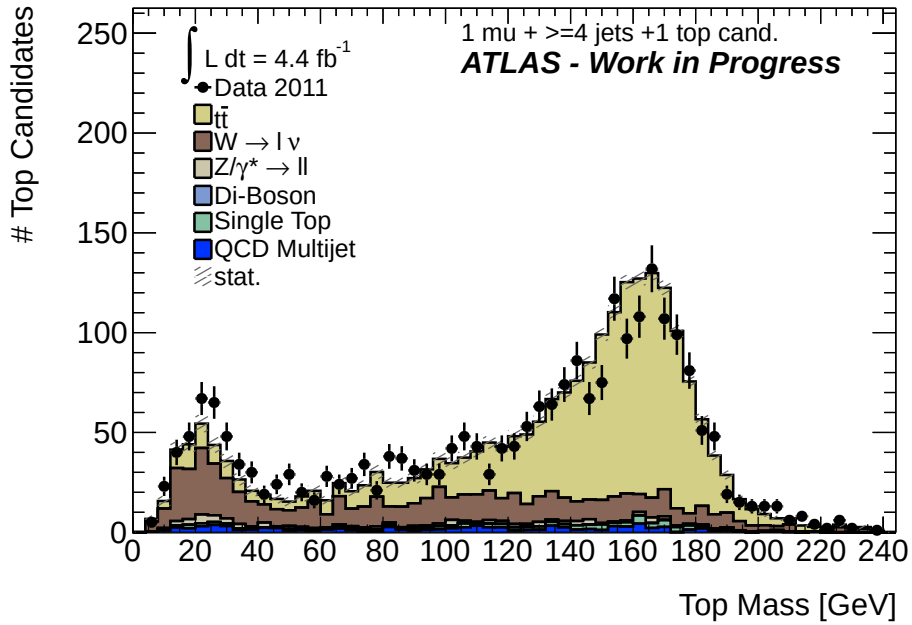
- Use full 2011 ATLAS data set $\sim 4.4\,\mathrm{fb}^{-1}$
- Top-enriched selection:
  - Single Muon Trigger
  - 1 Muon, 0 Electrons
  - $\geq$ 4 Jets (AntiKt, R=0.4, $p_T > 25$ GeV)
  - $E_{T,\,\mathrm{miss.}} > 20$ GeV
  - $E_{T,\,\mathrm{miss.}} + m_T > 60$ GeV
  - 1 Fatjet (C/A, R=1.5, $p_T > 200$ GeV)





Apply the HEPTopTagger



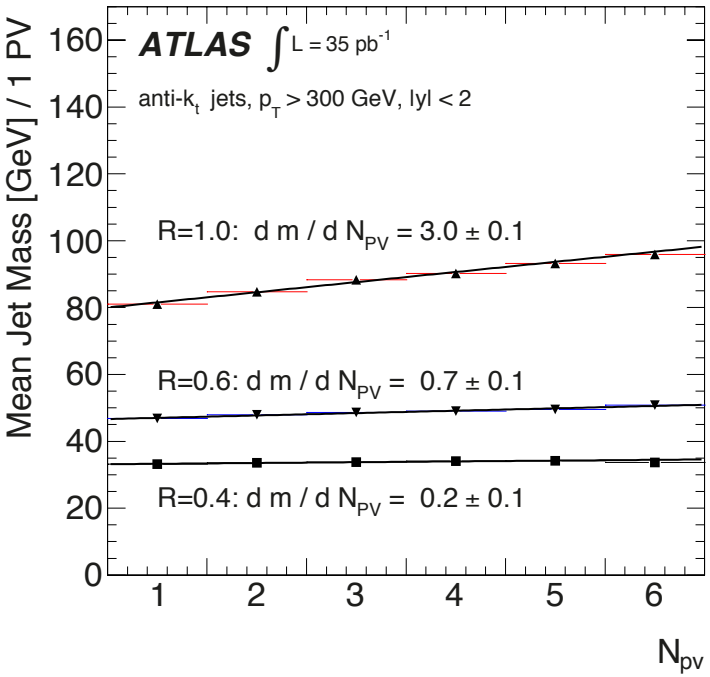


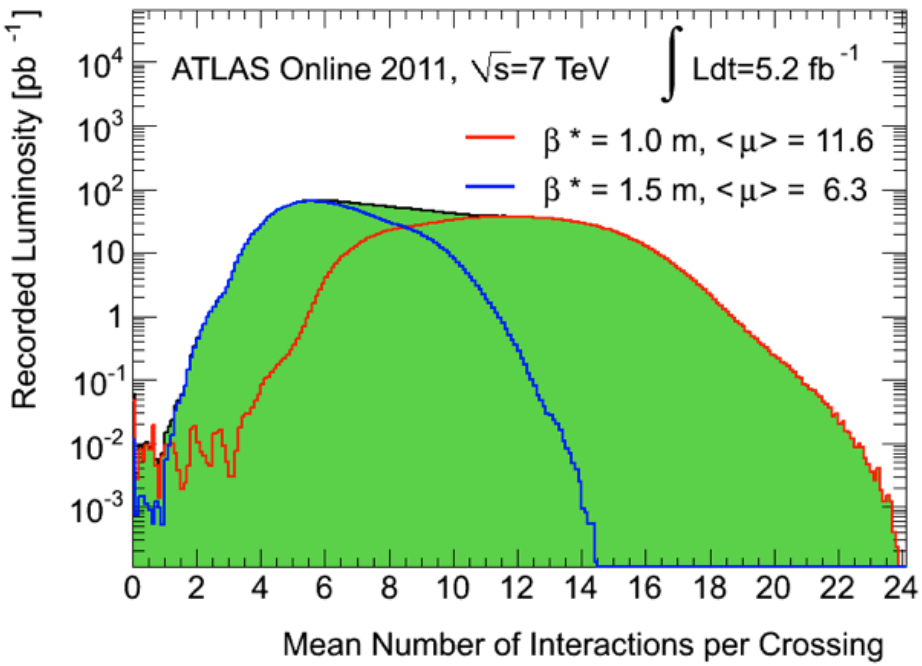
- Well described top mass peak
- Flat backgrounds
- $\Rightarrow$ **HEPTopTagger works on data**

Effect of pileup on fatjets

Ungroomed fatjets are affected by multiple interactions per bunch crossing (*in-time pileup*) and previous events (*out-of-time pileup*).

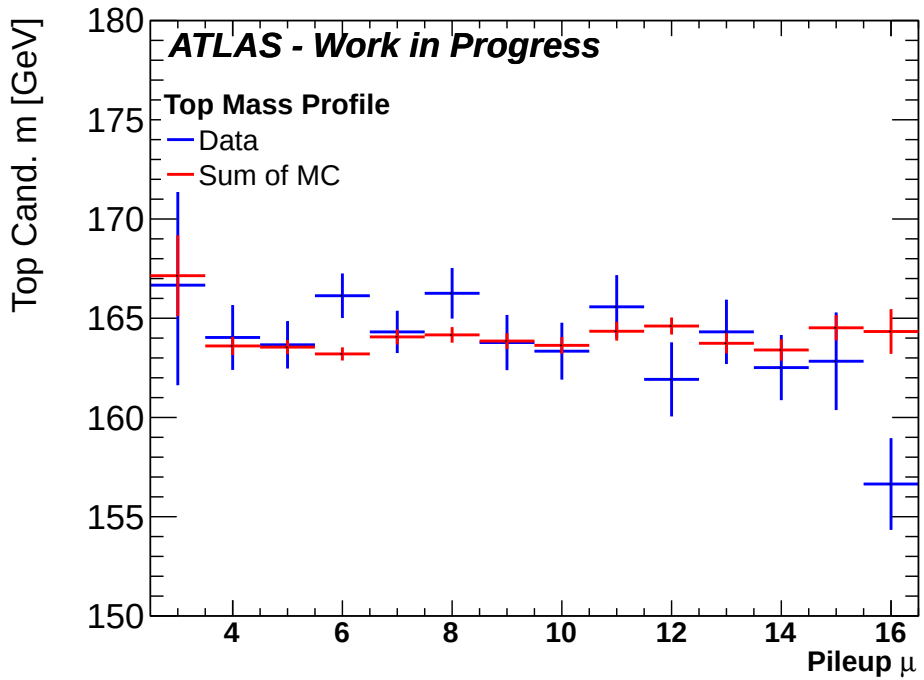
- Average number of interactions ber bunch crossing: μ
- Number of primary vertices: N_{PV}





Grooming removes unwanted constituents

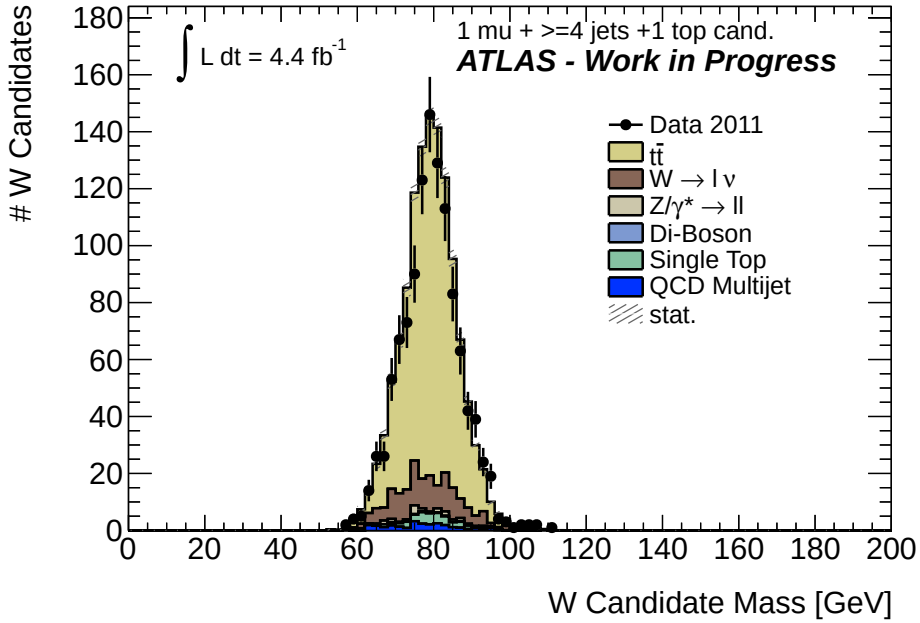
Pileup in the HEPTopTagger

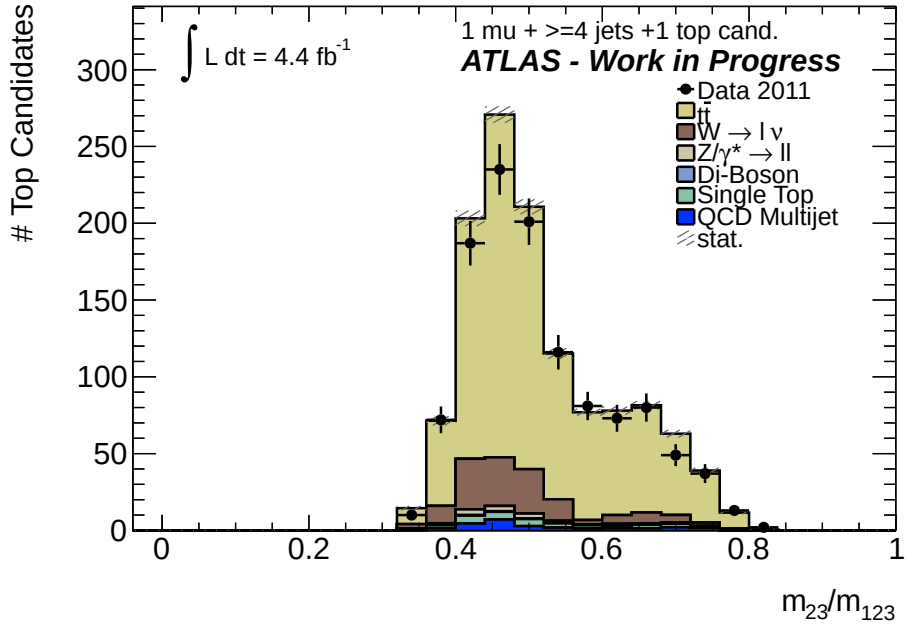


- Look at top candidate mass between 140 and 200 GeV
- Mean reconstructed top mass is stable with pileup
- Fit: $m = a + b \cdot \mu$

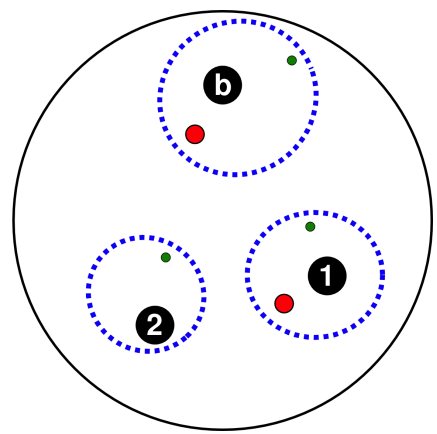
	a	b
data	166.1 ± 2.2 GeV	-0.21 ± 0.23 GeV
$\sum$ MC	165.1 ± 1.9 GeV	-0.08 ± 0.21 GeV

Substructure



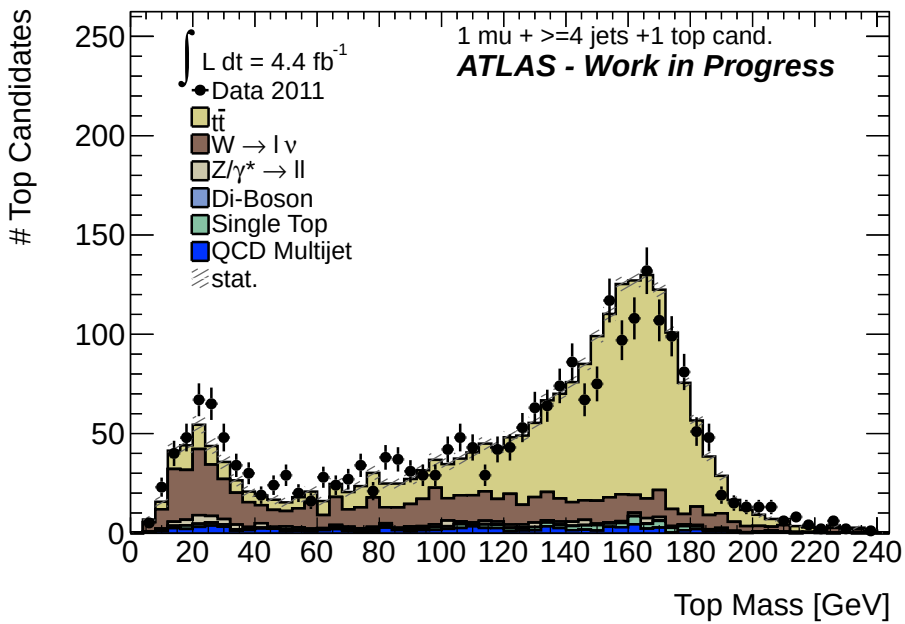


- m_{23}: subjet pair mass (W candidate)
- m_{123}: 3-subjet mass (top candidate)
- Substructure of tagged top candidates modelled well



Optimization

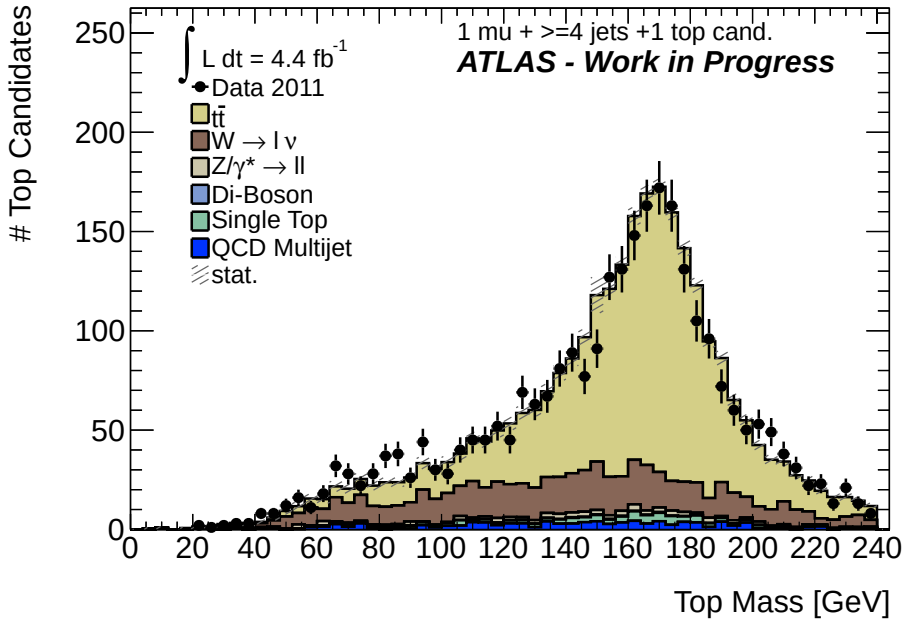
- HEPTopTagger filtering parameters:
  - Decompose substructure with $m > m_{\text{cut}}$
  - Find subjets with max filtering radius $R_{\text{jet}}^{\text{filt}}$
  - Keep $N_{\text{jet}}^{\text{filt}}$ hard subjets
  - m_W requirements (f_W)
  - Tune for efficiency/purity
- MC S($t\bar{t}$)/B(rest) in top mass window $120 - 200$ GeV



Default

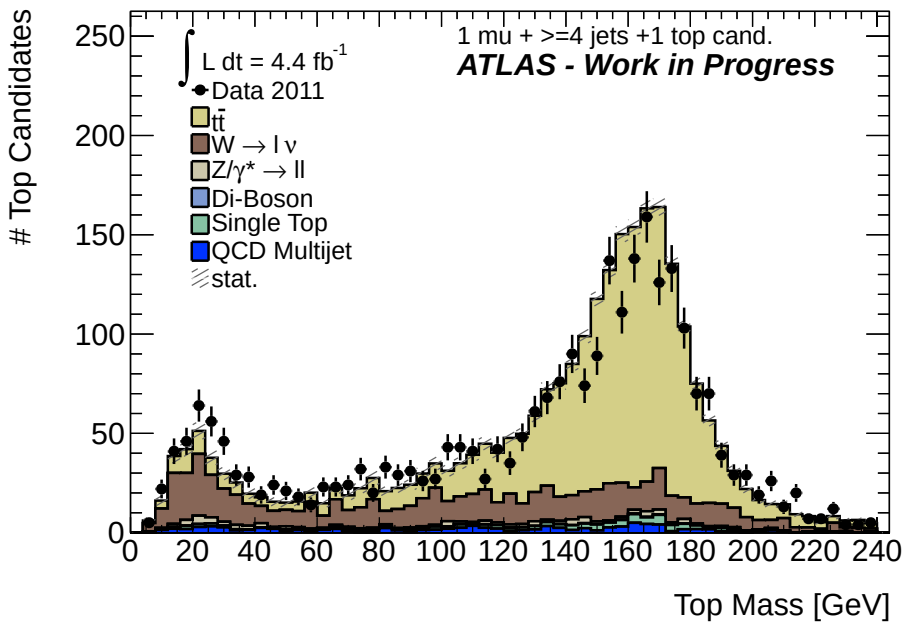
- R=1.5
- $m_{\text{cut}} = 30$ GeV
- $R_{\text{jet}}^{\text{filt}} = 0.3$
- $N_{\text{jet}}^{\text{filt}} = 5$
- $f_W = 15\,\%$
- $\rightarrow S/B = 4.1$

Optimize: Loose/Large Settings



Loose

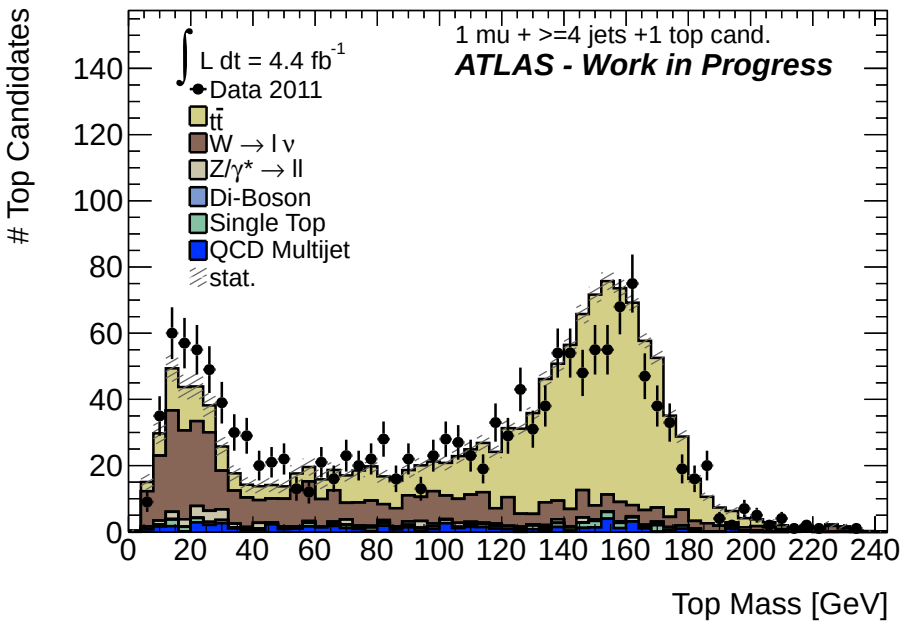
- R=1.5
- $m_{\text{cut}} = 70$ GeV
- $R_{\text{jet}}^{\text{filt}} = 0.5$
- $N_{\text{jet}}^{\text{filt}} = 7$
- $f_W = 20\,\%$
- $\rightarrow S/B = 3.1$



Large

- R=1.8
- $m_{\text{cut}} = 30$ GeV
- $R_{\text{jet}}^{\text{filt}} = 0.3$
- $N_{\text{jet}}^{\text{filt}} = 5$
- $f_W = 15\,\%$
- $\rightarrow S/B = 3.7$

Optimize: Tight Settings



Tight

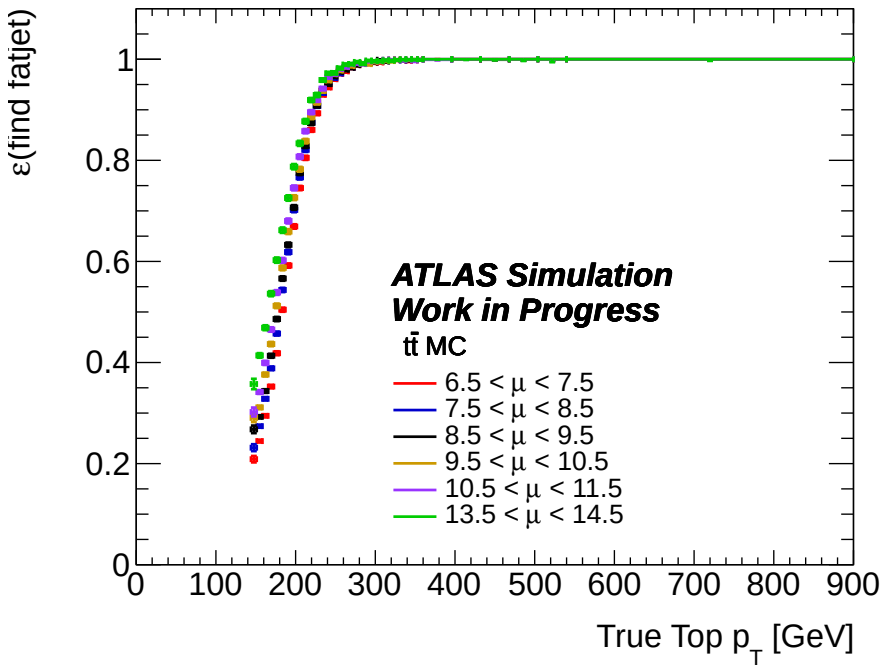
- R=1.5
- $m_{\text{cut}} = 30$ GeV
- $R_{\text{jet}}^{\text{filt}} = 0.2$
- $N_{\text{jet}}^{\text{filt}} = 4$
- $f_W = 10\,\%$
- $\rightarrow S/B = 5.2$

- Trade efficiency against purity via filtering parameters
- Choose ideal working point per analysis
- Loose/Large/Tight settings of the HEPTopTagger are modelled well

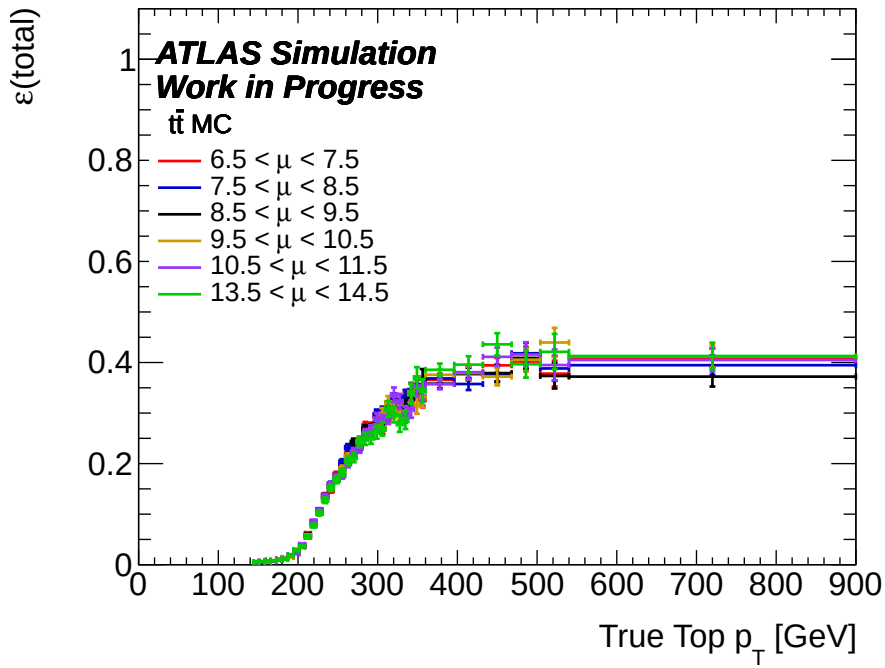
Performance on Monte Carlo

- Use semi-leptonic $t\bar{t}$ sample (MC@NLO) as source of true top events
  - Hadronically decaying top quark with $p_T > 150\,\text{GeV}$ (MC level)
  - No other cuts
- $\varepsilon(\text{total}) = \varepsilon(\text{find fatjet}) \cdot \varepsilon(\text{tag fatjet})$
  - ε(find fatjet): probability to find a fatjet in the event
  - ε(tag fatjet): probability to find a top candidate in a fatjet

Signal Efficiency



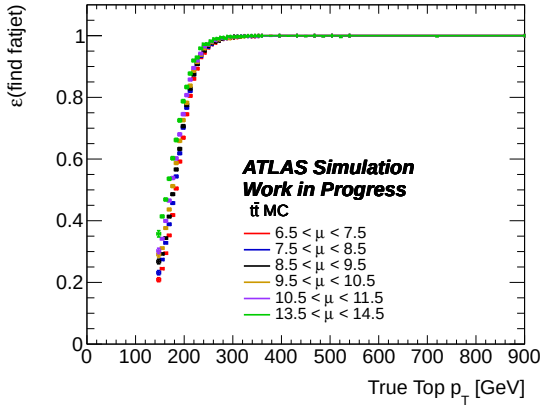
(j) probability to find fatjet in event

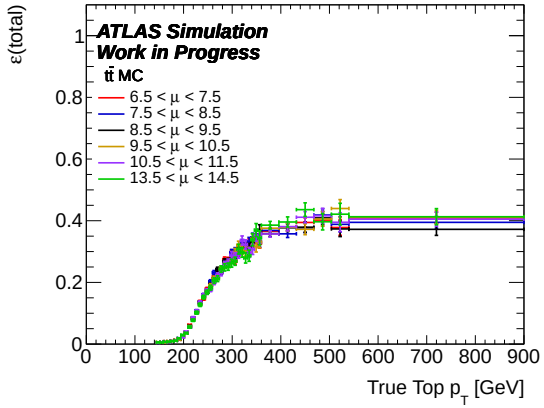


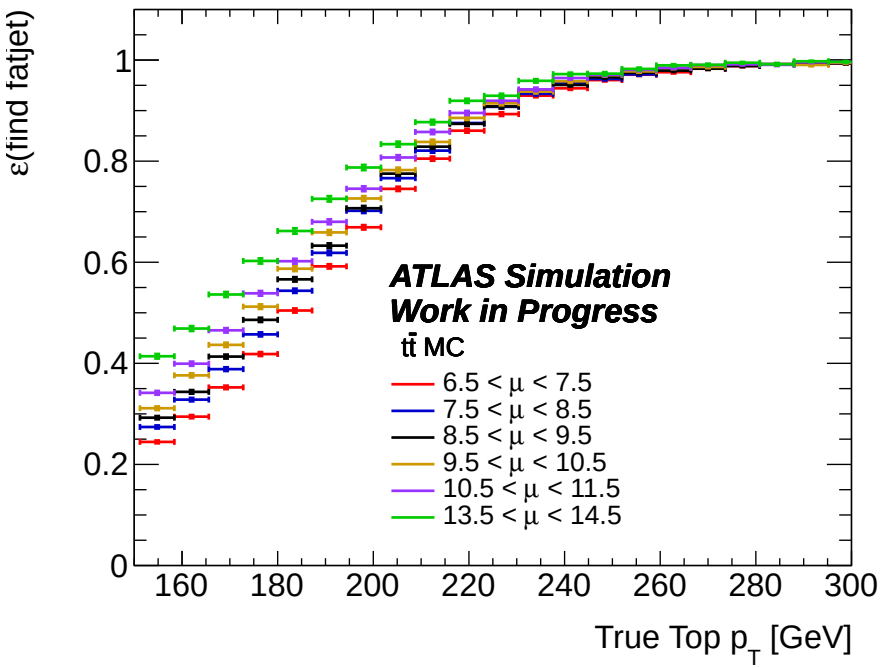
(k) probability to find top cand. in event

- Turn-on determined by ε(fatjet) reaches 100% at 250 GeV
- ε(tot.) plateaus at 40%
- Efficiencies stable with respect to μ, small effect on low-p_T fatjet finding

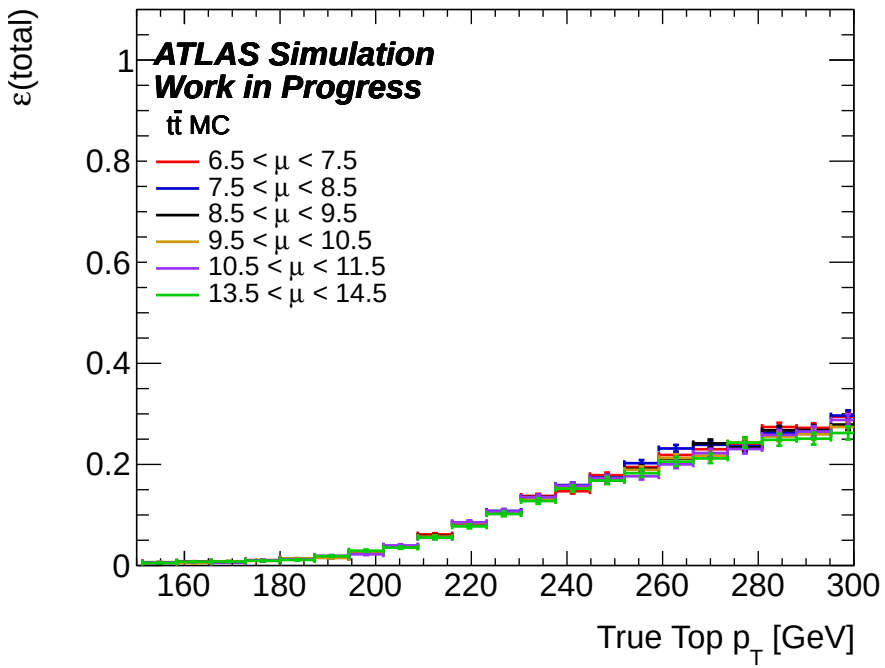
Signal Efficiency





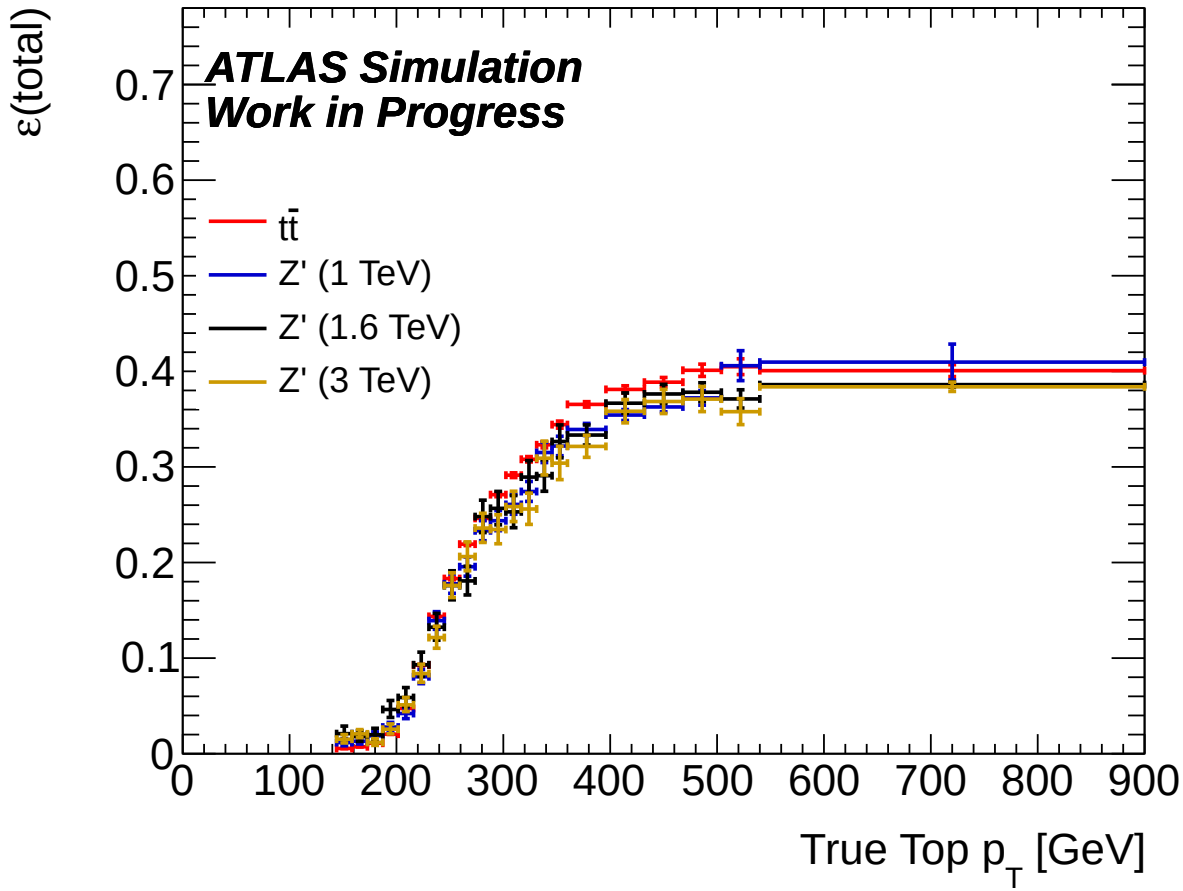


(n) probability to find fatjet in event



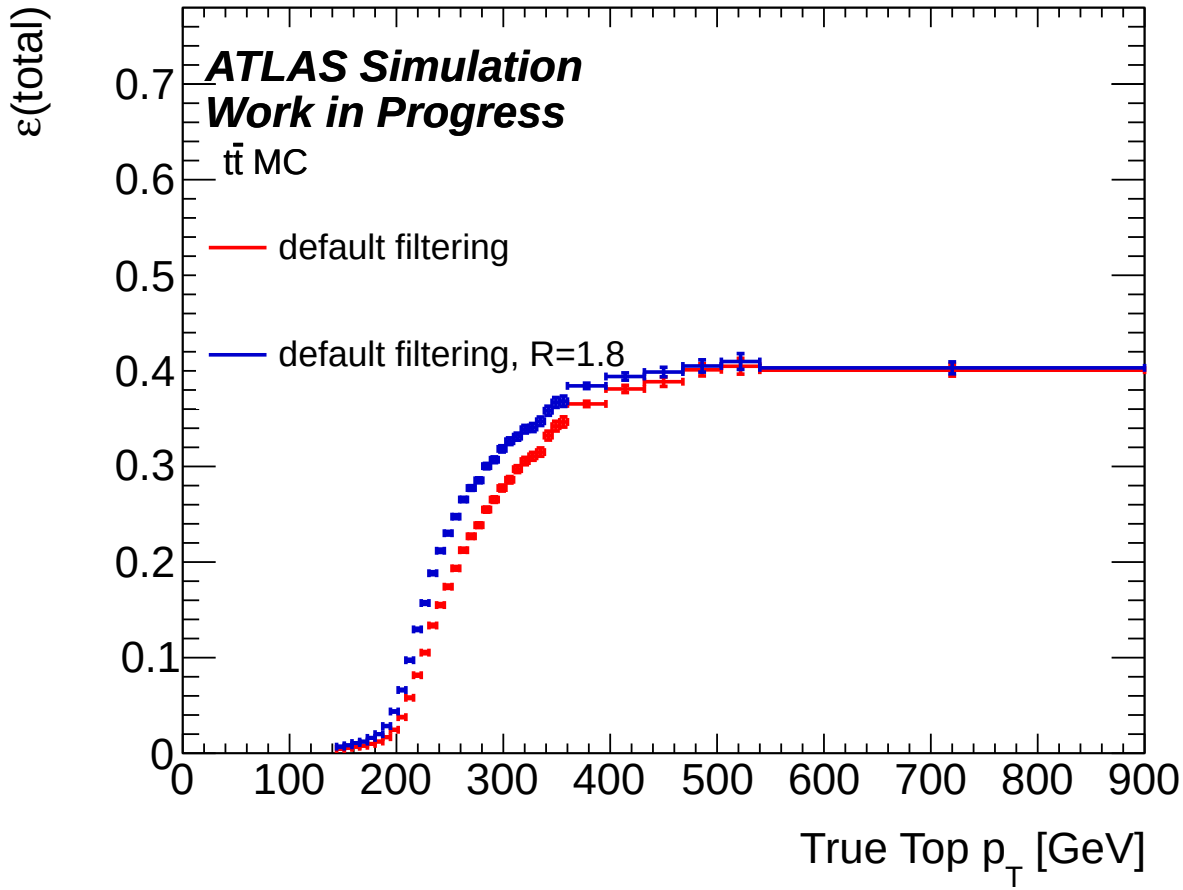
(o) probability to find top cand. in event

Different Samples



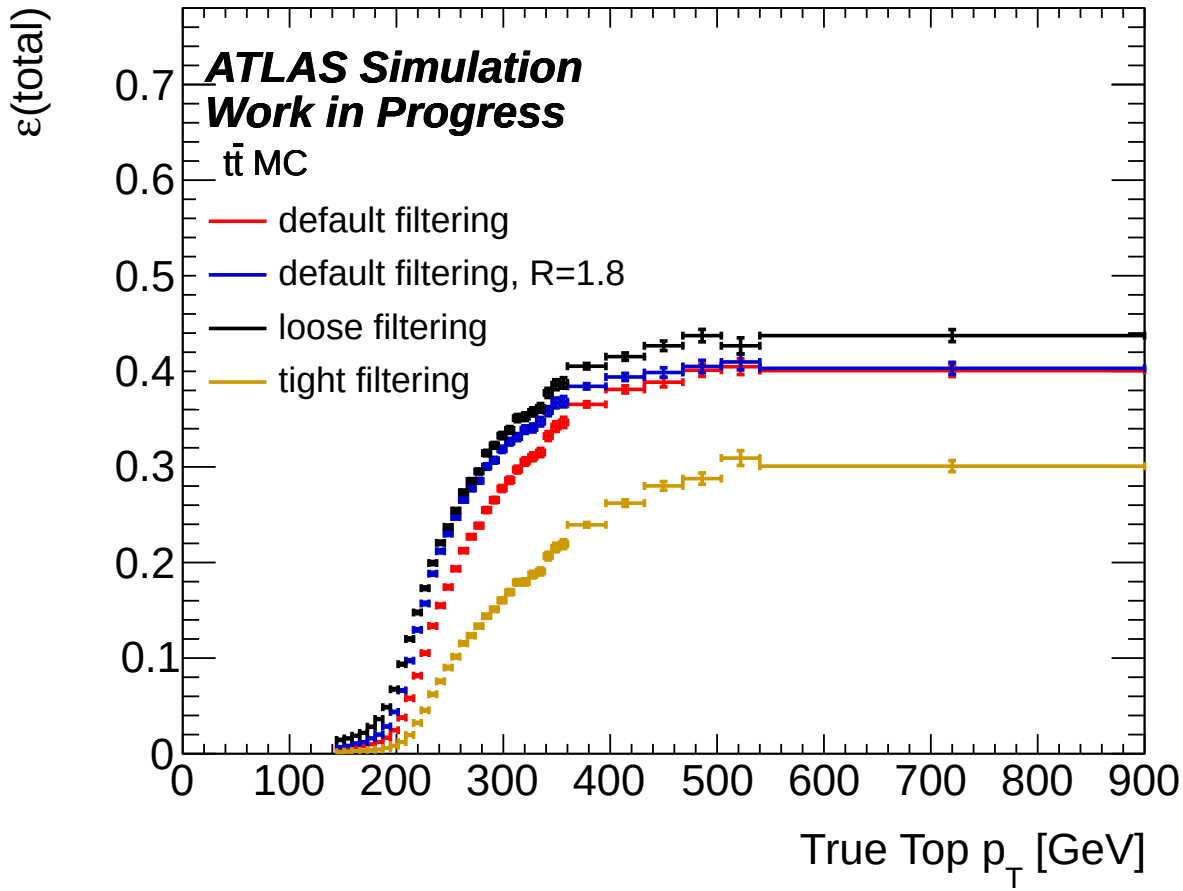
- Signal efficiency does not depend on source of top quarks

Different Taggers



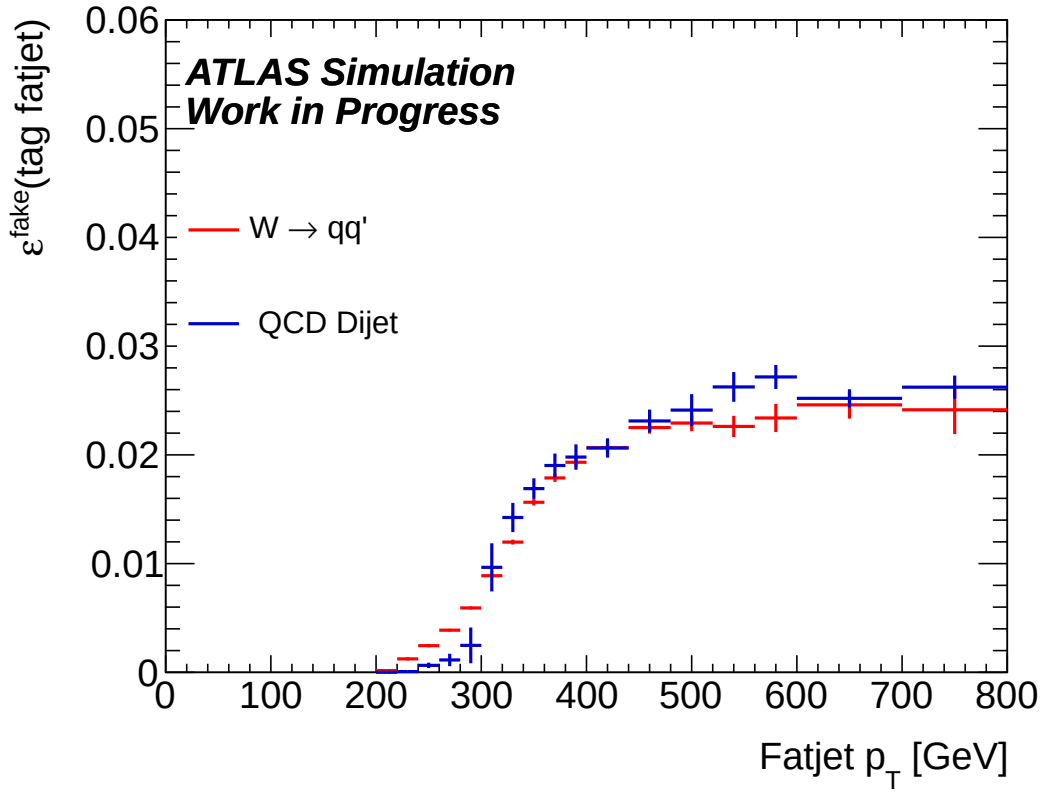
- Larger (R=1.8) fatjets increase efficiency at low p_T
- Converge around 500 GeV
- Low efficiency at $p_T = 200$ GeV due to hard top candidate p_T cut

Different Taggers



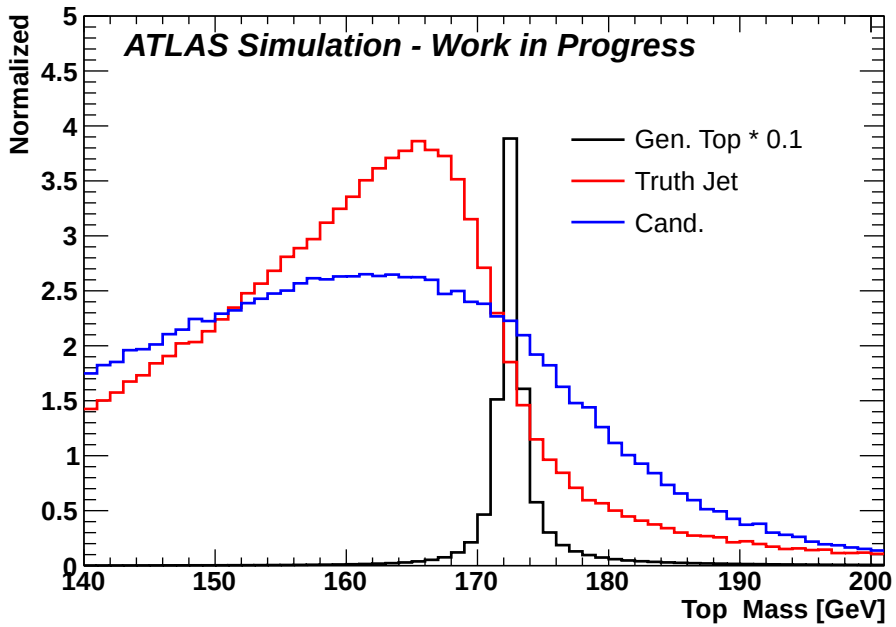
- $\sim 50\,\%$ difference between loose and tight filtering setting

Fake Efficiency in W→ qq' and QCD DiJet MC



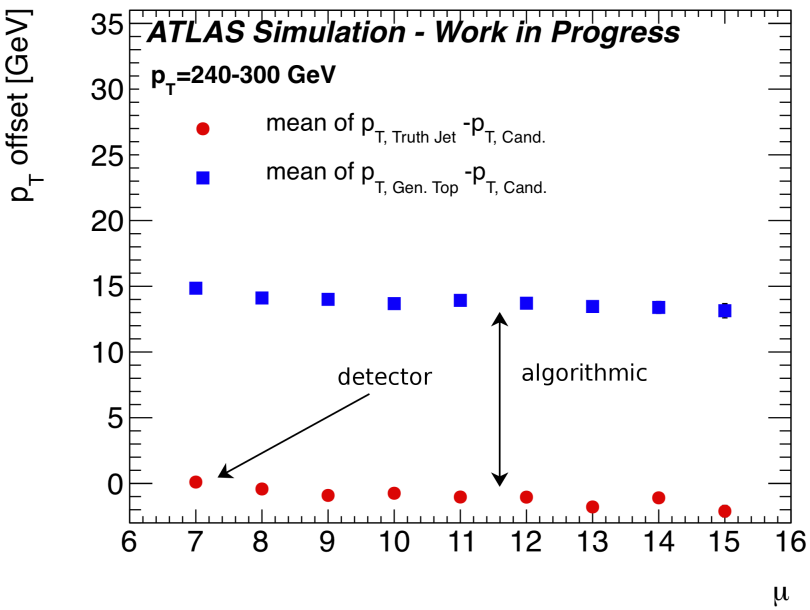
- Inclusive MC preselection
- Fatjets in W→ qq' and QCD DiJet look similar to the HEPTopTagger
- $\varepsilon^{\text{fake}}(\text{tag}) < 3\,\%$
- Total rejection depends on fatjet finding efficiency (analysis dependent)

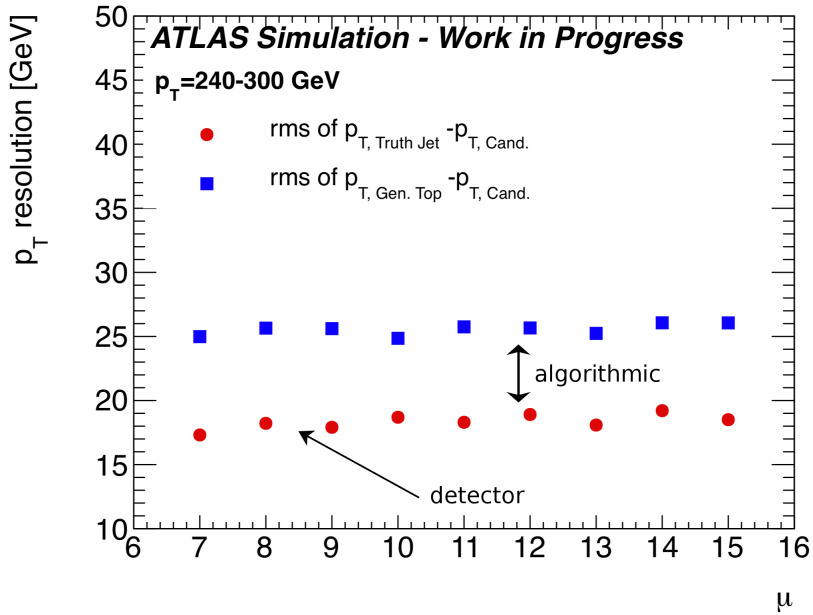
Disentangle algorithmic effects from detector effects



- "Truth Jet": HEPTopTagger on generated hadrons
- "Cand.": HEPTopTagger on calorimeter clusters

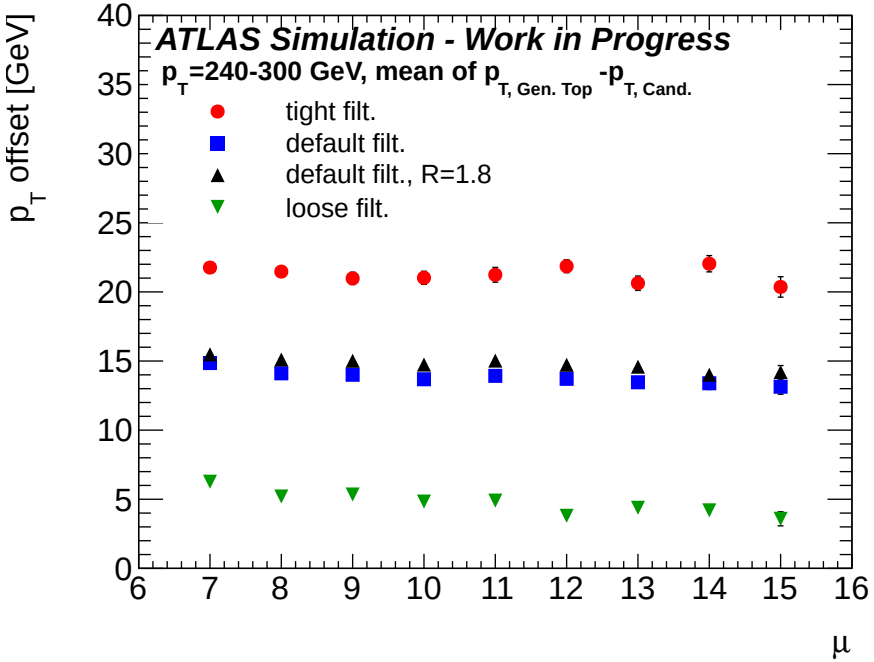
Reconstruction of top p_T

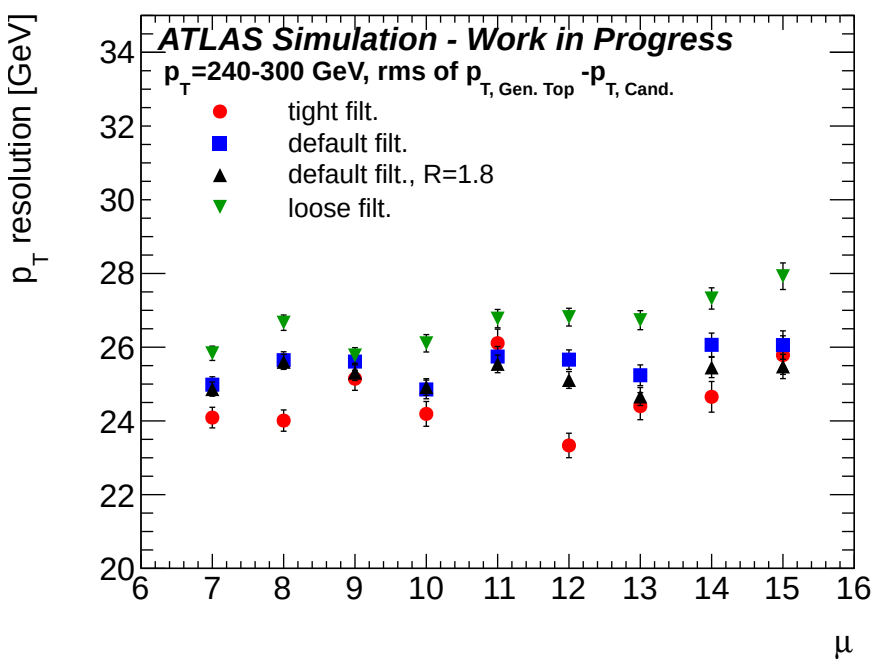




- p_T offset: 15 GeV (dominated by algorithm)
- p_T resolution: 25 GeV (18 GeV from detector)

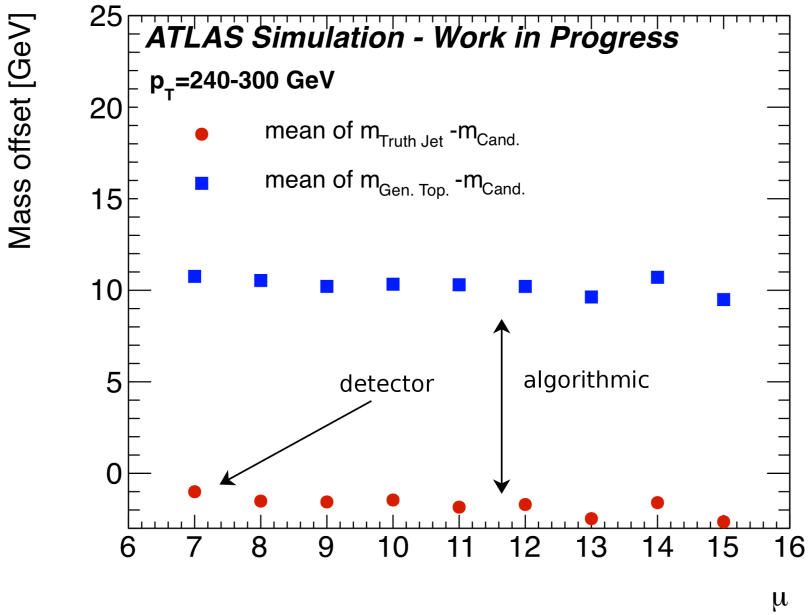
Effect of different filtering settings on p_T reconstruction

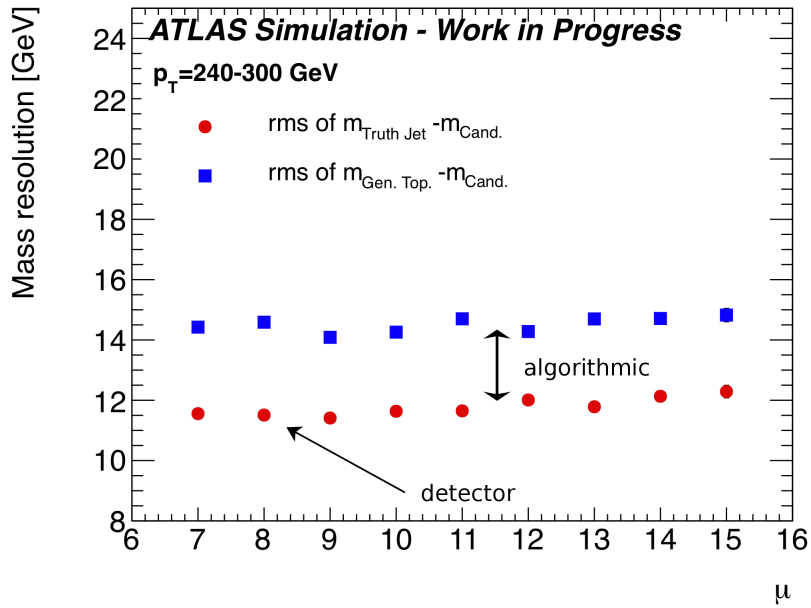




- Looser filtering decreases total p_T offset but worsens resolution
- Looser filtering increases effect of pileup

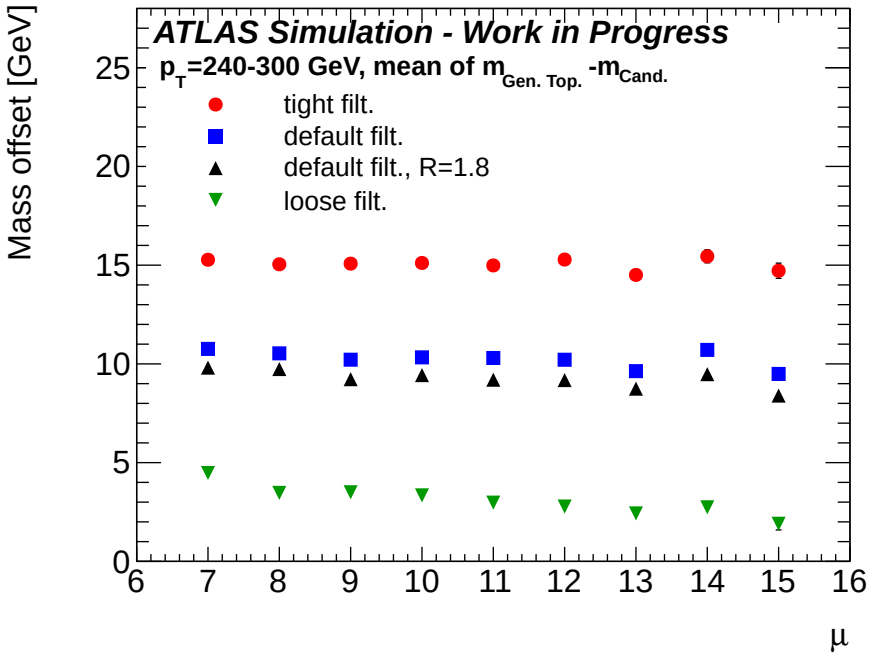
Reconstruction of top mass

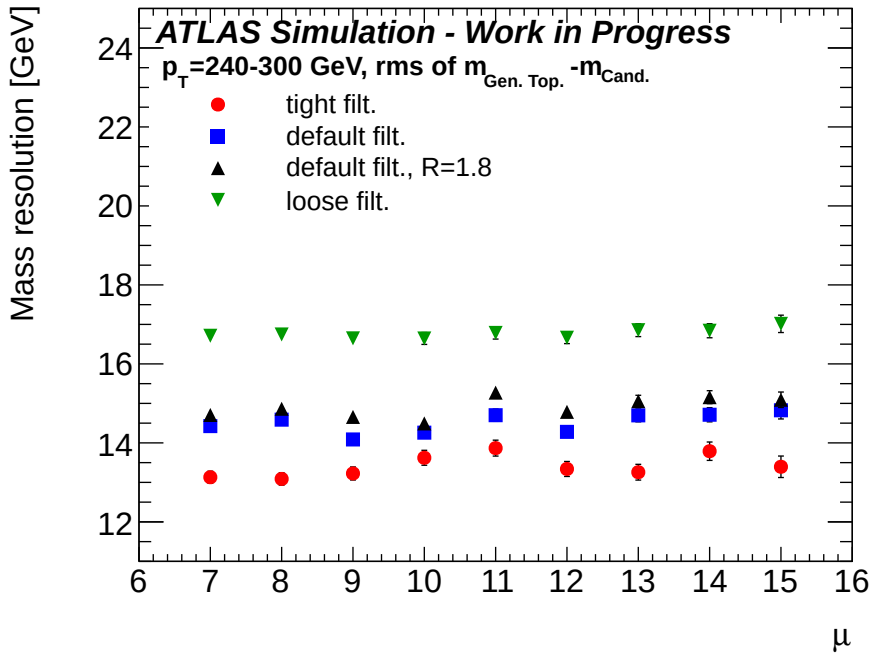




- Mass offset: 11 GeV (dominated by algorithm)
- Mass resolution: $\sim$ 15 GeV ($\sim$ 12 GeV from detector)
- Stable against pileup

Effect of different filtering settings on mass reconstruction

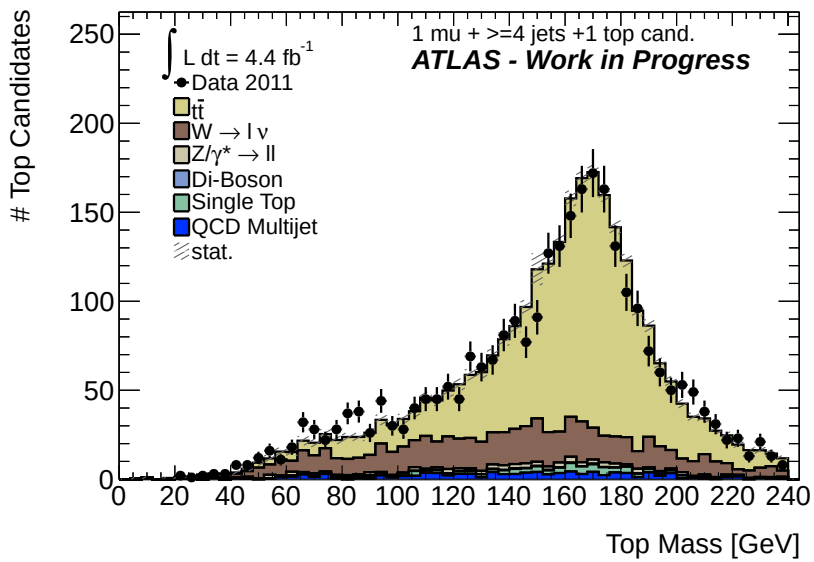




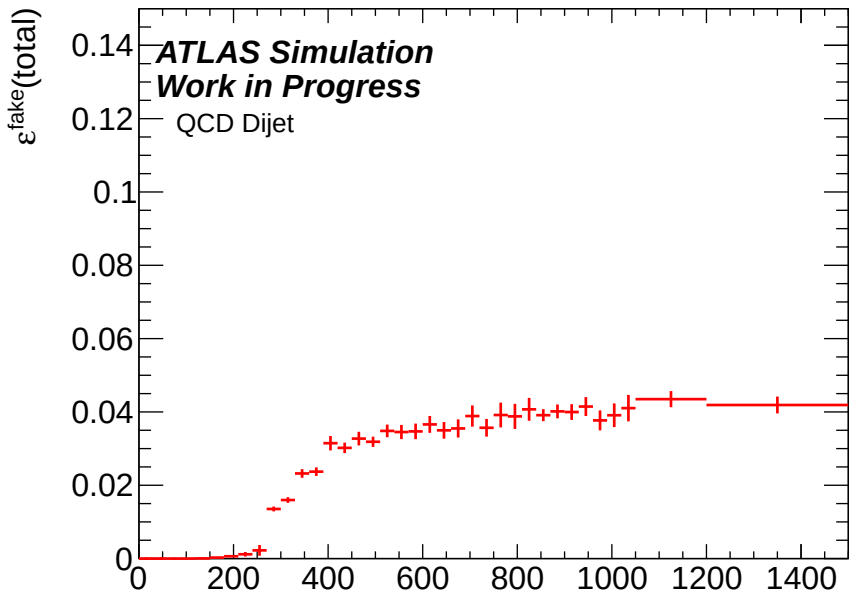
- Looser filtering decreases total mass offset but worsens resolution
- Looser filtering increases effect of pileup

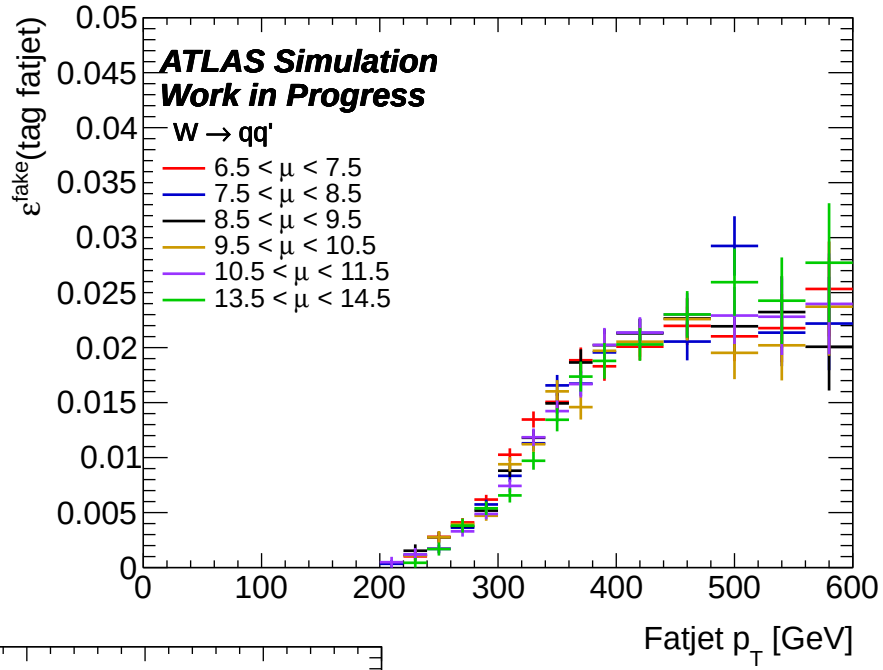
Summary

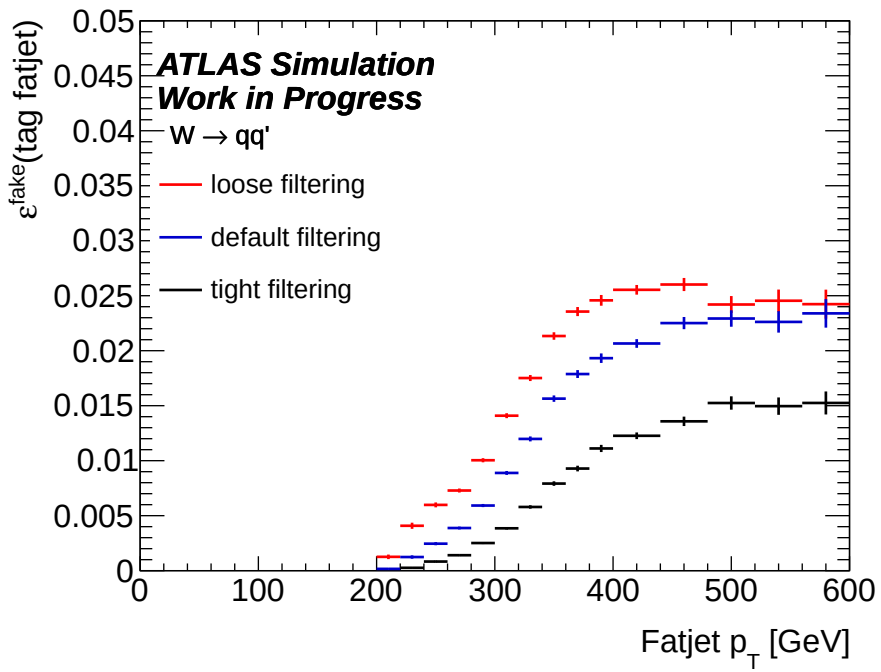
- Commissioned HEPTopTagger for use in ATLAS
- Fatjet kinematics and substructure well described by simulation
- Top tagging efficiency $= 40\,\%(20\,\%)$ for $p_T = 400\,\text{GeV}(200\,\text{GeV})$
- Reconstruction at low p_T :
  - top mass: shift $= -4\,\text{GeV}$, resolution $= 17\,\text{GeV}$
  - top p_T: shift $= -4\,\text{GeV}$, resolution $= 26\,\text{GeV}$
- Reconstruction robust against pile-up



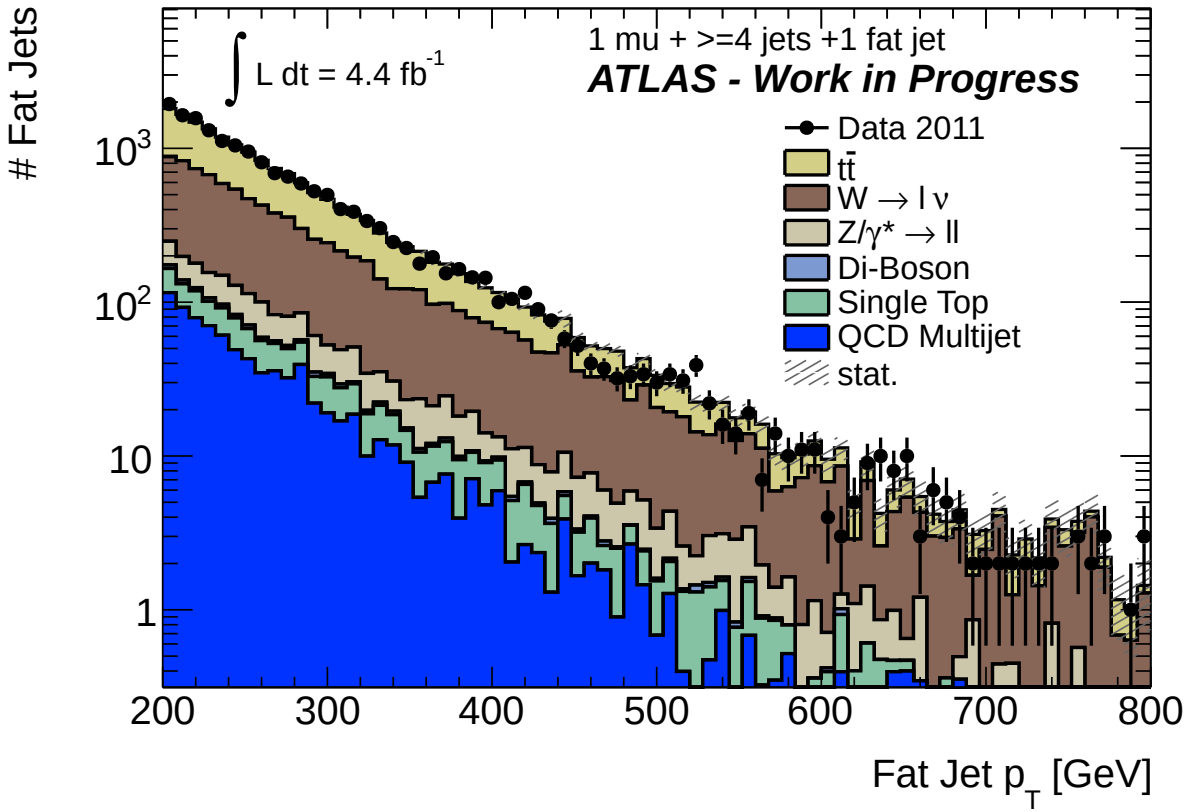
Fake Efficiency - Backup







Fatjet p_T before tagging



Mass Scale/Resolution

ATLAS Simulation - Work in progress

mean of $m_{\text{Truthjet}} - m_{\text{Cand.}}$

	offset	slope
$p_{\text{T, Gen.}} = 150 - 240$ GeV	-0.45 $\pm$ 0.28	-0.09 $\pm$ 0.03
$p_{\text{T, Gen.}} = 240 - 300$ GeV	0.01 $\pm$ 0.27	-0.14 $\pm$ 0.03
$p_{\text{T, Gen.}} = 300 - 900$ GeV	0.68 $\pm$ 0.27	-0.17 $\pm$ 0.03

mean of $m_{\text{Gen}} - m_{\text{Cand.}}$

	offset	slope
$p_{\text{T, Gen.}} = 150 - 240$ GeV	11.30 $\pm$ 0.26	-0.02 $\pm$ 0.02
$p_{\text{T, Gen.}} = 240 - 300$ GeV	11.62 $\pm$ 0.30	-0.09 $\pm$ 0.03
$p_{\text{T, Gen.}} = 300 - 900$ GeV	12.23 $\pm$ 0.34	-0.10 $\pm$ 0.03

rms of $m_{\text{Truthjet}} - m_{\text{Cand.}}$

	offset	slope
$p_{\text{T, Gen.}} = 150 - 240$ GeV	10.78 $\pm$ 0.20	0.13 $\pm$ 0.02
$p_{\text{T, Gen.}} = 240 - 300$ GeV	10.94 $\pm$ 0.19	0.09 $\pm$ 0.02
$p_{\text{T, Gen.}} = 300 - 900$ GeV	11.11 $\pm$ 0.19	0.07 $\pm$ 0.02

rms of $m_{\text{Gen}} - m_{\text{Cand.}}$

	offset	slope
$p_{\text{T, Gen.}} = 150 - 240$ GeV	14.37 $\pm$ 0.18	0.01 $\pm$ 0.02
$p_{\text{T, Gen.}} = 240 - 300$ GeV	14.42 $\pm$ 0.21	0.02 $\pm$ 0.02
$p_{\text{T, Gen.}} = 300 - 900$ GeV	15.99 $\pm$ 0.24	0.03 $\pm$ 0.02