The Boost13 Working Group Report

Boost14

University College London, 18-22nd August 2014

Ben Cooper (UCL)

On behalf of the Boost13 working group

The Boost Reports Series

Boost2010 Report

Eur. Phys. J. C (2011) 71:1661
DOI 10.1140/epjc/s10052-011-1661-y

THE EUROPEAN PHYSICAL JOURNAL C

Special Article - Tools for Experiment and Theory

Boosted objects: a probe of beyond the standard model physics*

Boost2011 Report

IOP PUBLISHING — JOURNAL OF PHYSICS G: NUCLEAR AND PARTICLE PHYSICS

J. Phys. G: Nucl. Part. Phys. **39** (2012) 063001 (44pp) doi:10.1088/0954-3899/39/6/063001

TOPICAL REVIEW

Jet substructure at the Tevatron and LHC: new results, new tools, new benchmarks*

Boost2012 Report

Eur. Phys. J. C (2014) 74:2792
DOI 10.1140/epjc/s10052-014-2792-8

THE EUROPEAN PHYSICAL JOURNAL C

Special Article - Tools for Experiment and Theory

Boosted objects and jet substructure at the LHC. Report of BOOST2012, held at IFIC Valencia, 23rd–27th of July 2012

The Boost Reports Series

- And now…Boost2013

Towards an Understanding of the Correlations in Jet Substructure

Report of BOOST2013, hosted by the University of Arizona, 12^{th}-16^{th} of August 2013.

D. Adams[1], A. Arce[2], L. Asquith[3], M. Backovic[4], T. Barillari[5], P. Berta[6], D. Bertolini[2], A. Buckley[8], J. Butterworth[9], R. C. Camacho Toro[10], J. Caudron[9], Y.-T. Chien[11], J. Cogan[12], B. Cooper[9], D. Curtin[17], C. Debenedetti[18], J. Dolen[9], M. Eklund[22], S. El Hedri[22], S. D. Ellis[22], T. Embry[22], D. Ferencek[23], J. Ferrando[24], S. Fleischmann[16], M. Freytsis[25], M. Giulini[21], Z. Han[27], D. Hare[4], P. Harris[4], A. Hinzmann[4], R. Hoing[4], A. Hornig[22], M. Jankowiak[4], K. Johns[28], G. Kasieczka[23], T. Knight[24], G. Kasieczka[29], R. Kogler[30], W. Lampl[4], A. J. Larkoski[4], C. Lee[31], R. Leone[31], P. Loch[31], D. Lopez Mateos[27], H. K. Lou[27], M. Low[27], P. Maksimovic[32], I. Marchesini[32], S. Marzani[32], L. Masetti[33], R. McCarthy[32], S. Menke[32], D. W. Miller[35], K. Mishra[36], B. Nachman[32], P. Nef[4], F. T. O'Grady[24], A. Ovcharova[23], A. Picazio[37], C. Pollard[38], B. Potter Landua[29], C. Potter[29], S. Rappoccio[39], J. Rutherfoord[40], G. P. Salam[10,11], J. Schabinger[23], A. Schwartzman[4], M. D. Schwartz[27], B. Shuve[43], P. Sinervo[44], D. Soper[45], D. E. Sosa Corral[45], M. Spannowsky[32], E. Strauss[34], M. Swiatlowski[4], J. Thaler[34], C. Thomas[34], E. Thompson[1], N. V. Tran[36], J. Tseng[36], E. Usai[36], L. Valery[36], J. Veatch[23], M. Vos[23], W. Waalewijn[4], and C. Young[47]

Arriving at the Scope of this Report

Topic	Volunteers
Systematic comparisons of MC generators to all available unfolded experimental data (using Rivet)	Andy Buckley Marat Freytsis
Analytical calculations vs MC vs data? New measurements?	James Ferrando
Systematic study of taggers/observables (correlations etc)	Nhan Viet Tran Andreas Hinzmann Pekka Sinervo Gregor Kasieczka Jesse Thaler Emanuele Usai

Topic	Volunteers
Explore performance of observables/taggers and MC generator comparisons at much higher boosts (1-2 TeV)	Brain Shuve Marcel Vos David Lopez (H->bb) Ben Nachman Andy Buckley Sebastian Fleischmann Lucia Masetti
Comment on Snowmass	James Dolen
Prospects for analytical calculations at high pT	Simone Marzani

- After Boost13 discussion session we had a number of topics…
- …but final scope determined largely by the interests of those people with time to work on the report, and what can realistically be done with the limited manpower/computing resources available.

2014 Report Overview

- A systematic exploration of the correlations/overlap/complementarity between different groomed jet mass definitions and substructure variables in the context of:
  - W tagging
  - Top tagging (including HTT and John Hopkins tagging algorithms)
  - q/g discrimination

- Exploration of correlations done largely through examining the ROC curves for BDT combinations of the groomed masses/variables.
  - If two variables are strongly correlated the BDT combination ROC will not improve on the single variable ROC curves. Variables which do not share a lot of information will improve in combination.

- Correlations and performance are explored as a function of anti-k_T jet radius and jet p_T, going beyond p_T > 1 TeV.

2014 Report Overview

- We are not trying to make quantitative statements about which groomed mass + variable combination makes the best tagger
  - No pile-up.
  - No detector simulation/emulation.
  - No rigorous comparison between generators.

- But statements on the correlations, and how these evolve with different p_T and R, should not be too dependent on these factors.

W-tagging Studies

All studies by Nhan Viet Tran

Fastjet 3.03 jet framework used

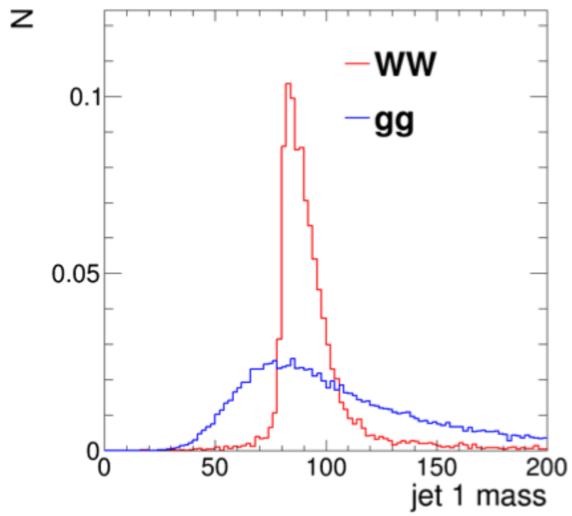
W-Tagging MC Samples

- All samples at $\sqrt{s}$ = 8 TeV using the CTEQ6L1 PDF.
- QCD background samples:
  - Madgraph5 + Pythia8 (Tune 4C).
  - Only pp→gg samples used (will check results with qq).
  - Generated in exclusive parton p_T bins, with additional cut on leading (ungroomed) jet.
- Resonant (scalar) X→WW→qqqq signal samples:
  - JHU Generator + Pythia8 (Tune 4C).
  - Generated in exclusive in W p_T bins, with additional cut on leading (ungroomed) jet.

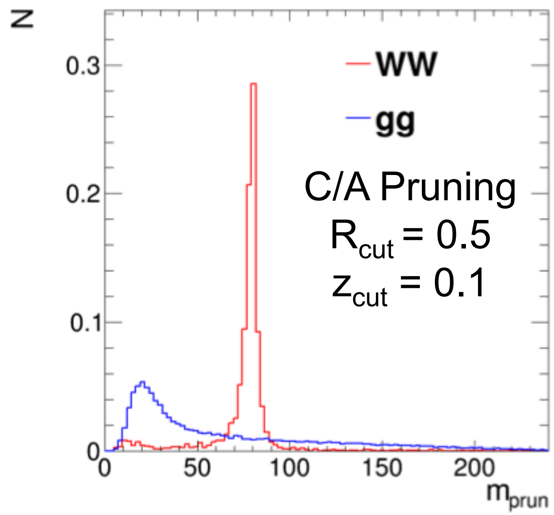
- Exclusive p_T bins generated:
  - 300-400 GeV, 500-600 GeV, 1.0-1.1 TeV.
- Range of anti-K_T jet radii used R=0.4, 0.8, 1.2.

Single Variables: Mass

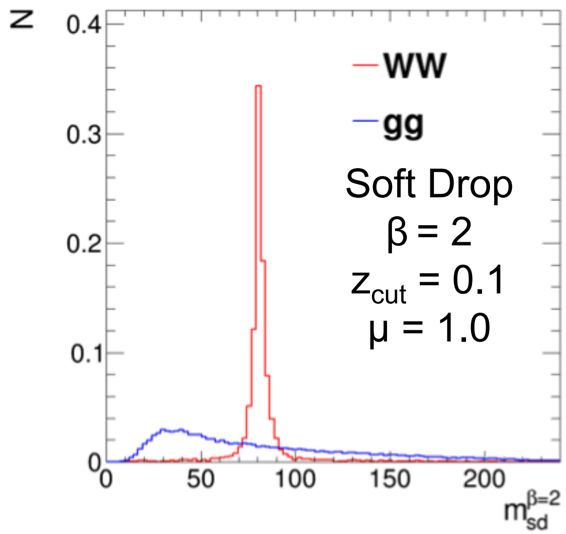
R=0.8, p_T 500-600 GeV



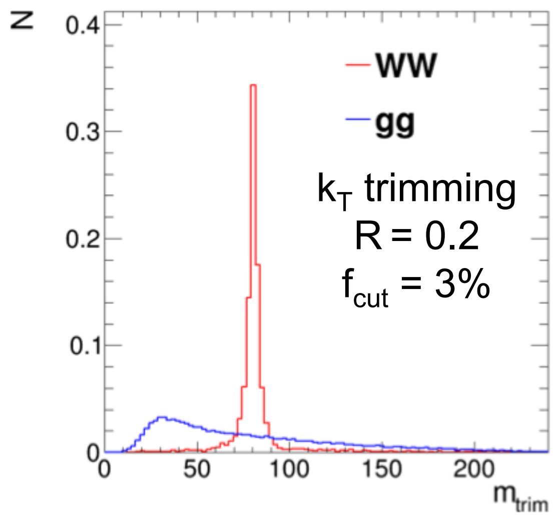
(a) Ungroomed mass

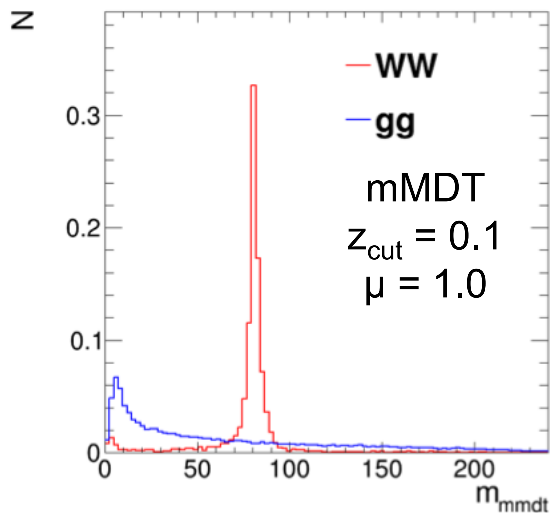


(b) Pruned mass



(e) Soft-drop $\beta = 2$ mass

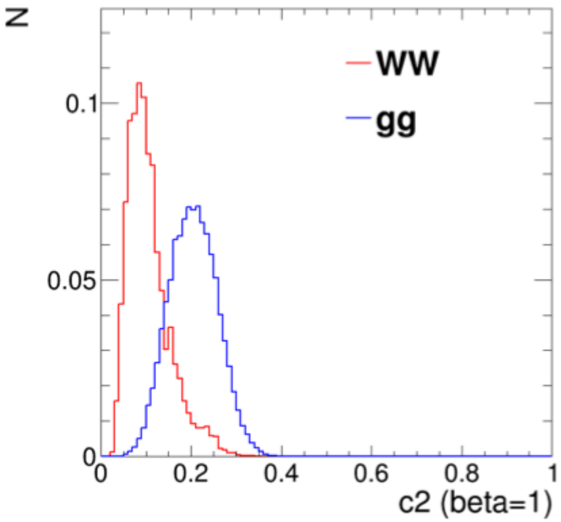




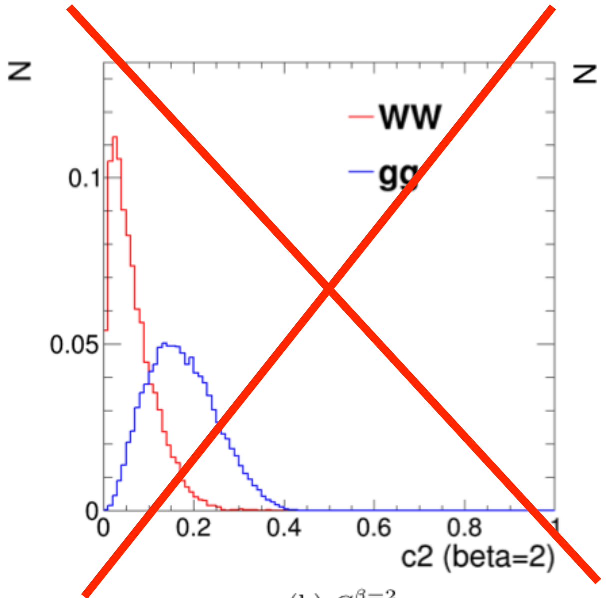
- Fixed grooming parameters used.
- No optimization of grooming parameters.

Single Variables: Substructure

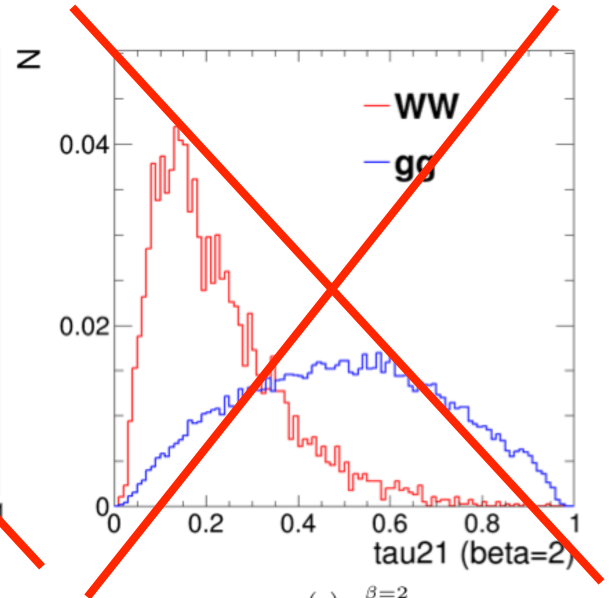
R=0.8, p_T 500-600 GeV



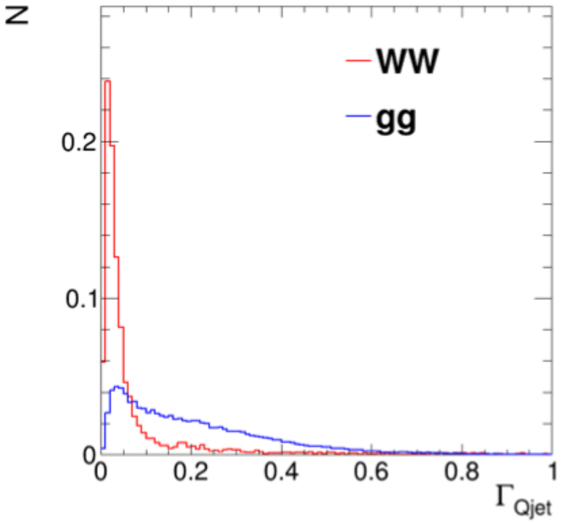
(a) $C_2^{\beta=1}$

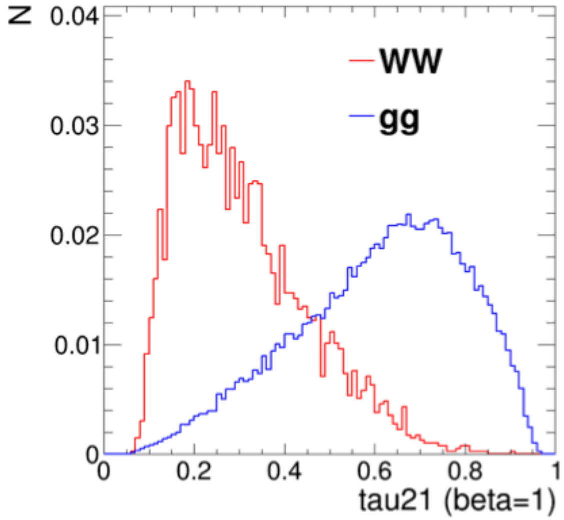


(b) $C_2^{\beta=2}$



(e) $\tau_{21}^{\beta=2}$



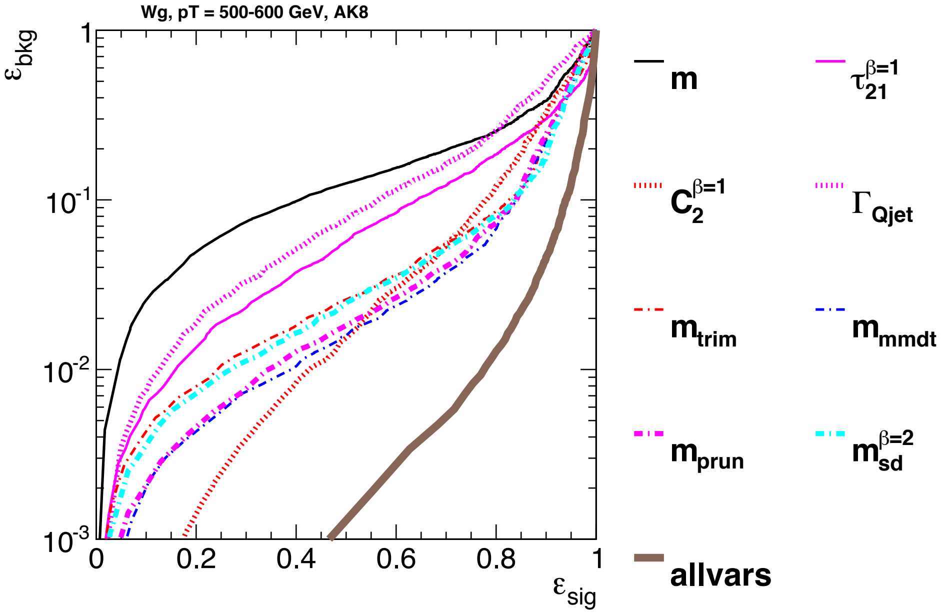


Found that the β=2 variants of these discriminants were not as performant as β=1.

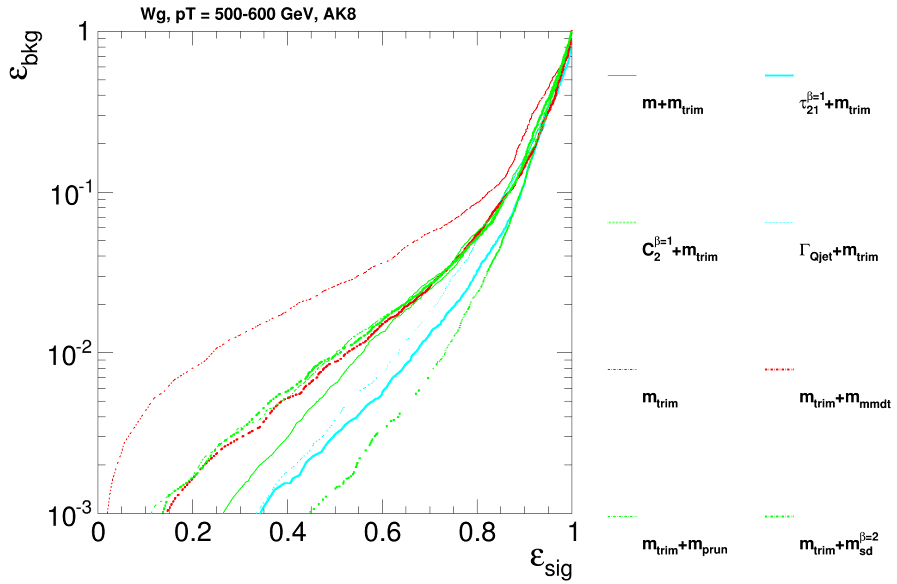
ROC Curves

R=0.8, p_T 500-600 GeV

Individual variables

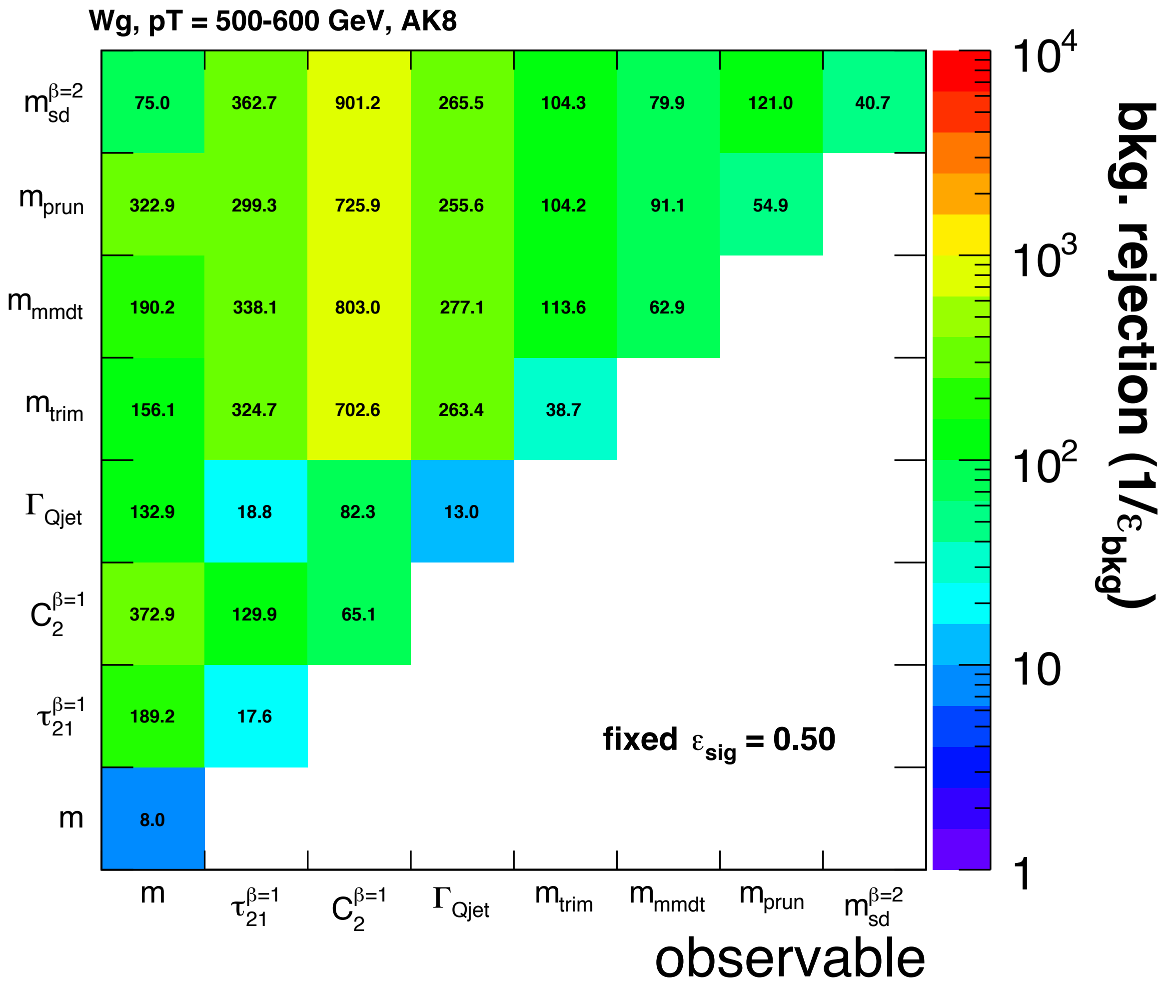


m_{trim} + X combinations



- Clearly much to be gained from combining the variables…
- …but combinations produced too many ROC curves to digest!

Combined Variable Performance

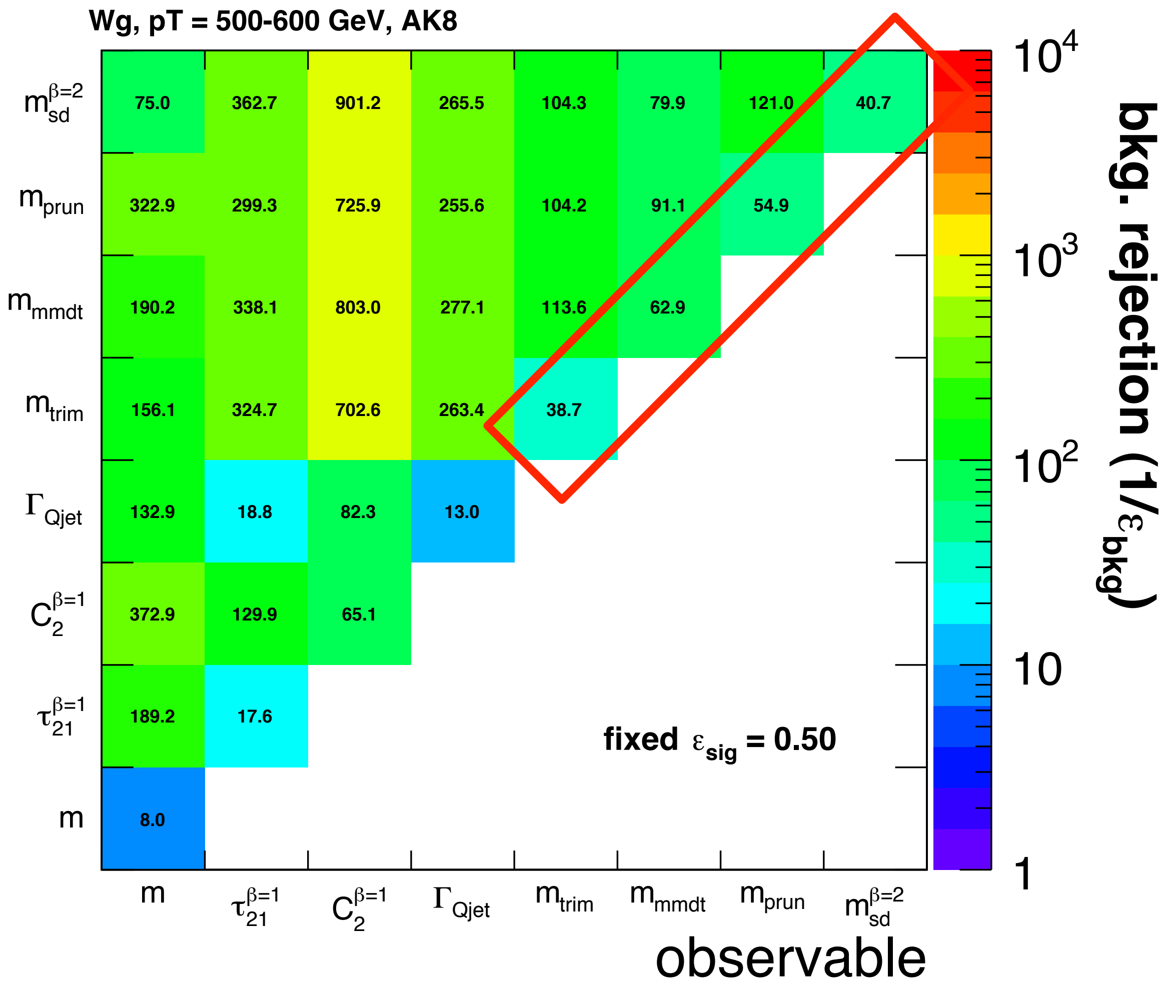


R=0.8,
p_T 500-600 GeV

- Background rejection at signal efficiency of 50% for each combination of variable (and for single variables along the diagonal).

Combined Variable Performance

Individually groomed masses are more powerful discriminants

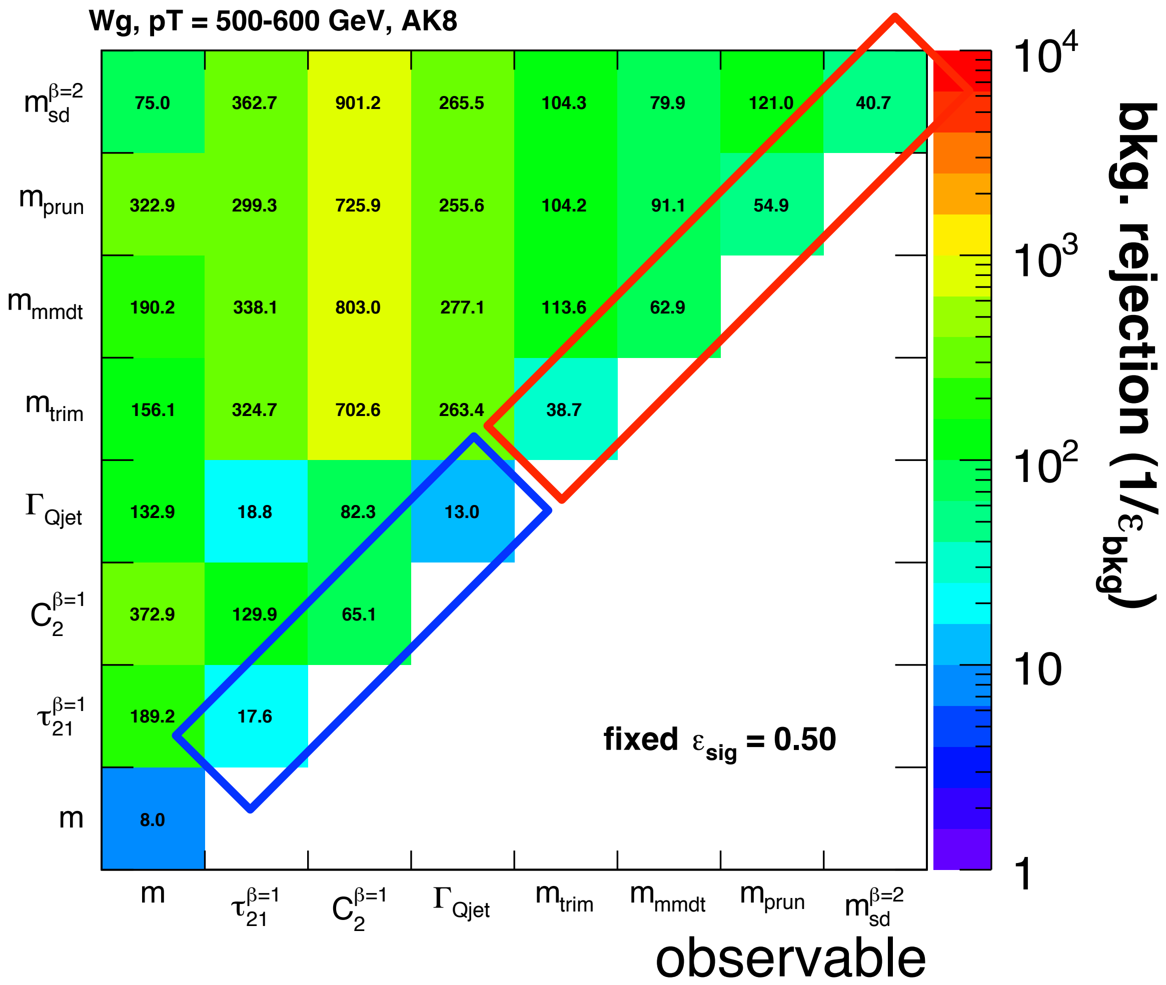


R=0.8, p_T 500-600 GeV

Combined Variable Performance

Individually groomed masses are more powerful discriminants

than the substructure shape variables considered



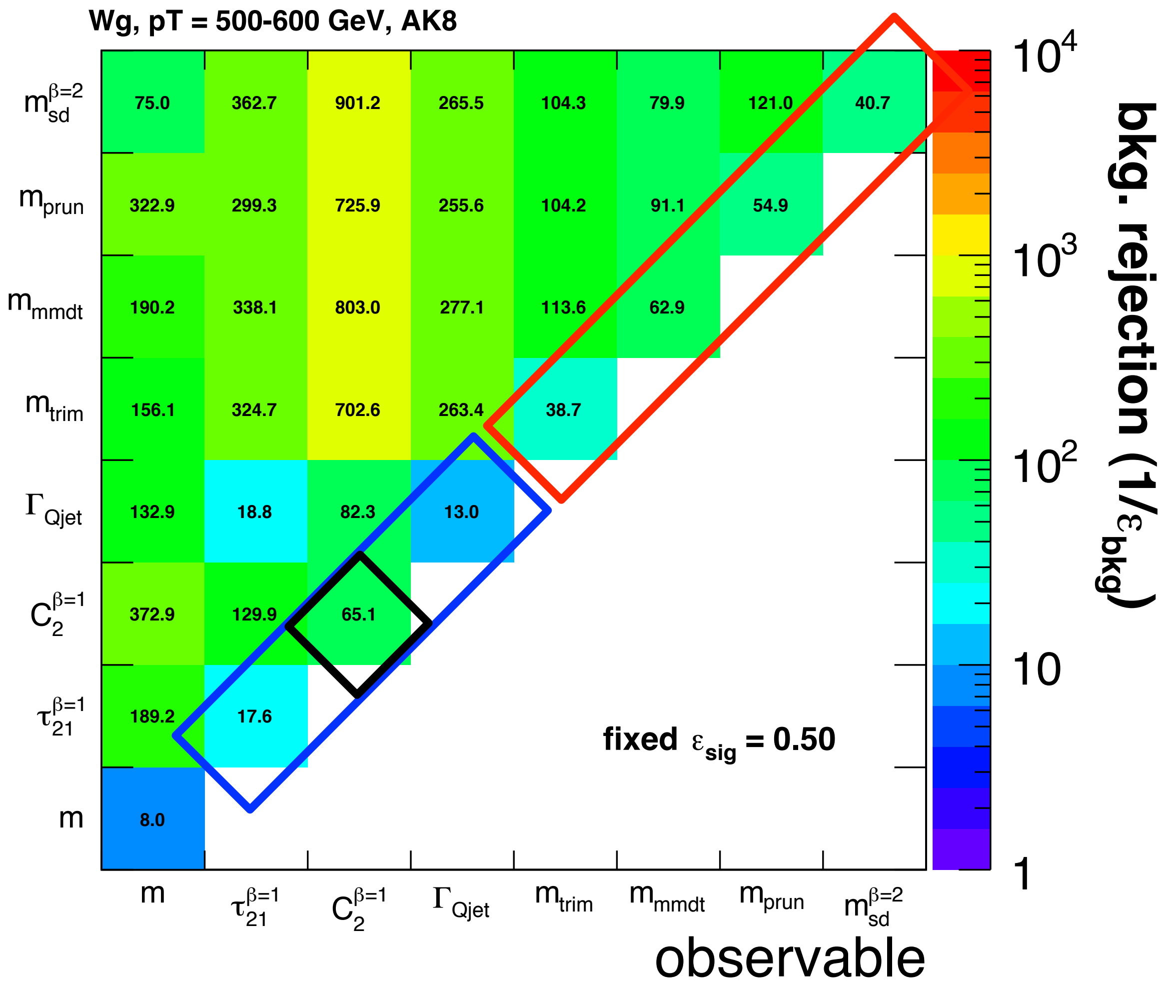
R=0.8, p_T 500-600 GeV

Combined Variable Performance

Individually groomed masses are more powerful discriminants

than the substructure shape variables considered

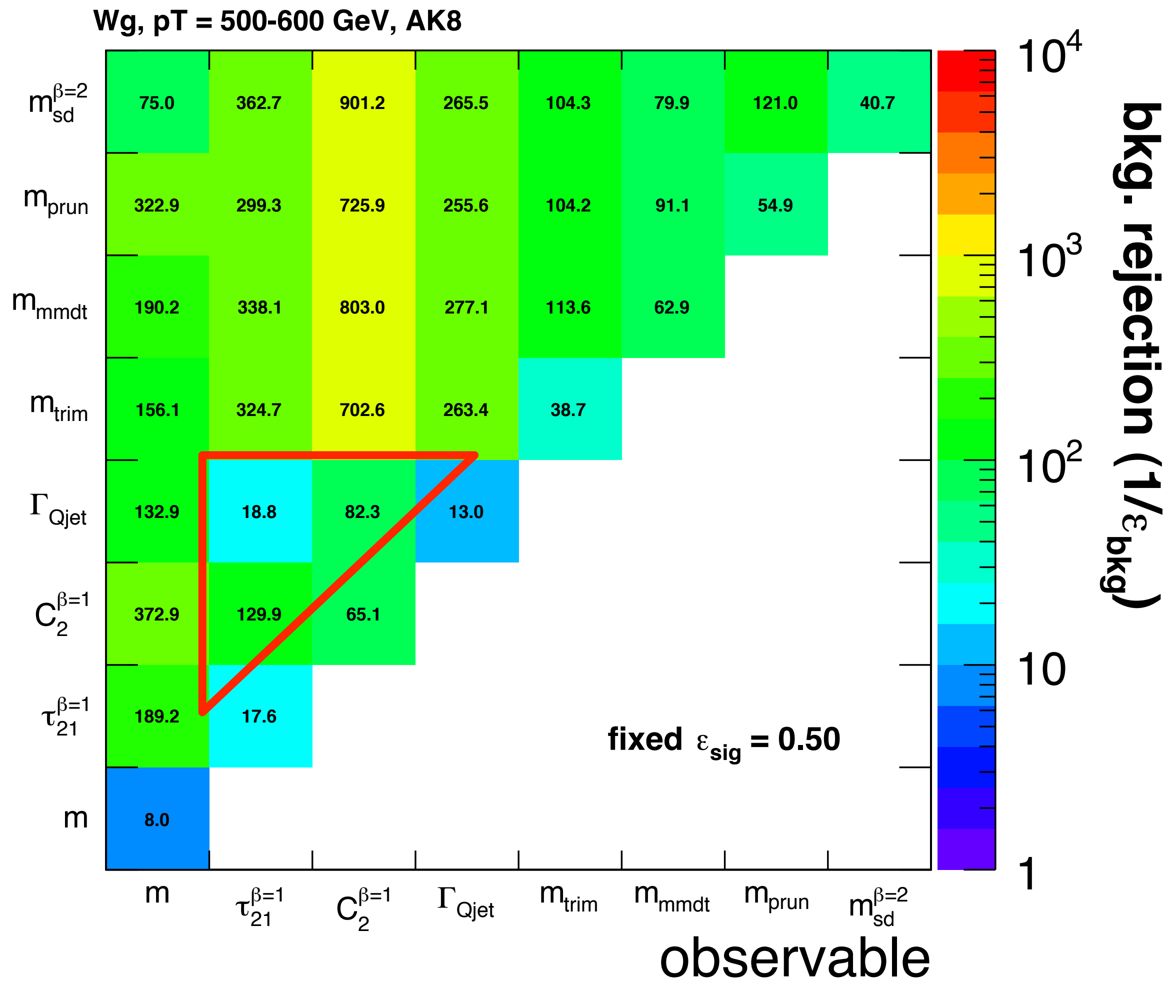
with the exception of C_2 (in this R bin)



R=0.8, p_T 500-600 GeV

Combined Variable Performance

Combinations of substructure variables are not as powerful



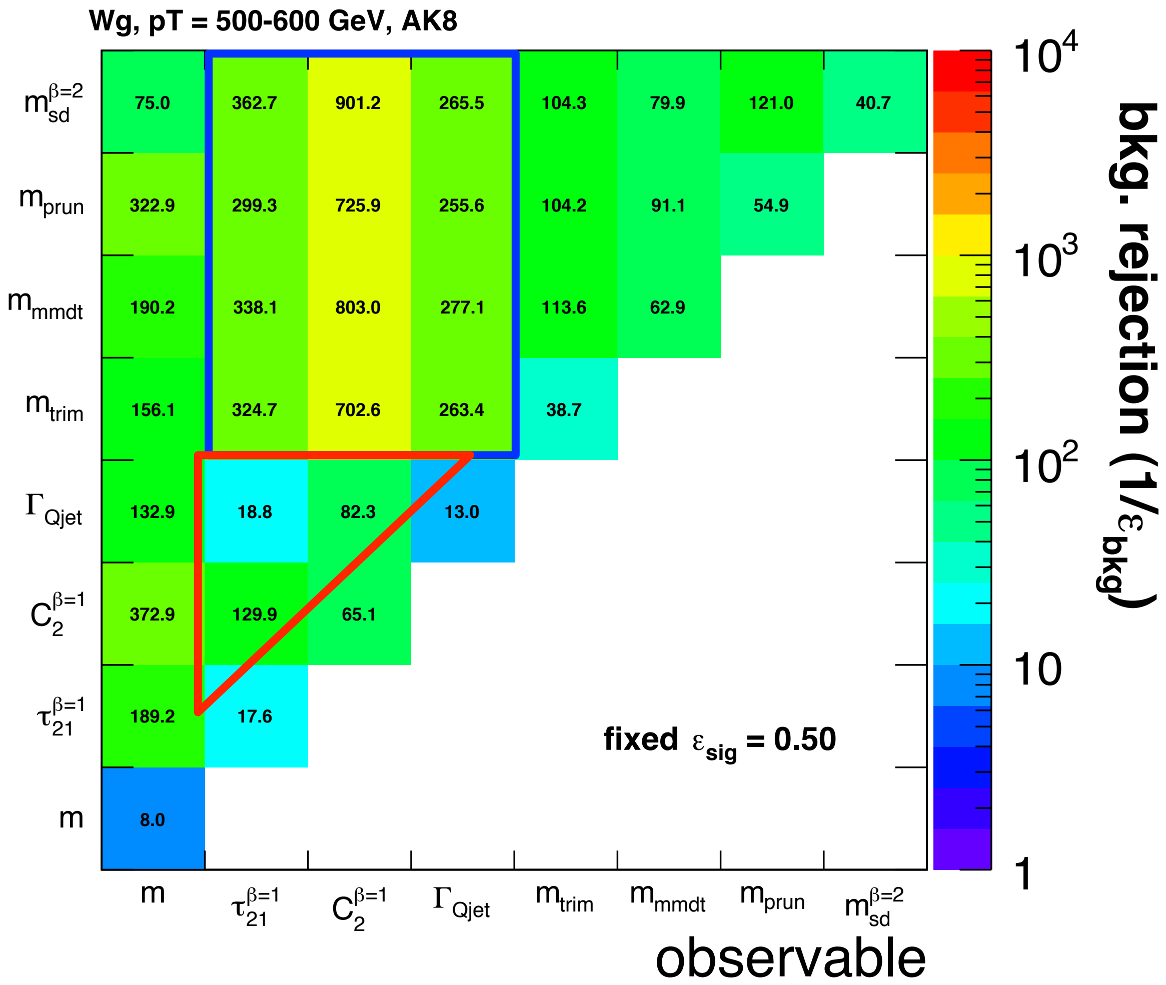
	m	$\tau_{21}^{\beta=1}$	$C_2^{\beta=1}$	Γ_{Qjet}	m_{trim}	m_{mmdt}	m_{prun}	$m_{sd}^{\beta=2}$
$m_{sd}^{\beta=2}$	75.0	362.7	901.2	265.5	104.3	79.9	121.0	40.7
m_{prun}	322.9	299.3	725.9	255.6	104.2	91.1	54.9	
m_{mmdt}	190.2	338.1	803.0	277.1	113.6	62.9		
m_{trim}	156.1	324.7	702.6	263.4	38.7			
Γ_{Qjet}	132.9	18.8	82.3	13.0				
$C_2^{\beta=1}$	372.9	129.9	65.1					
$\tau_{21}^{\beta=1}$	189.2	17.6						
m	8.0							

R=0.8, p_T 500-600 GeV

Combined Variable Performance

Combinations of substructure variables are not as powerful

as combinations of groomed mass and substructure



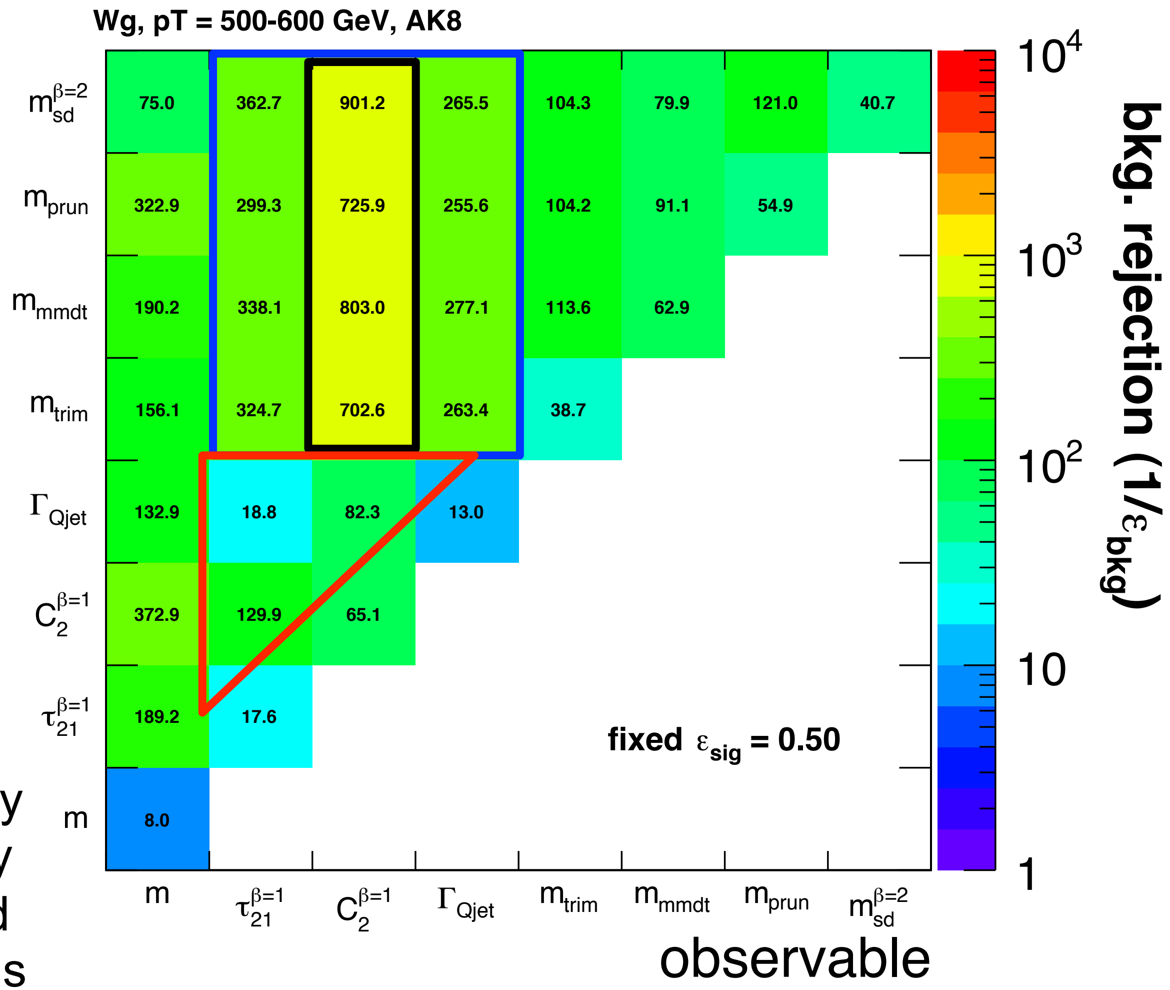
R=0.8, p_T 500-600 GeV

Combined Variable Performance

Combinations of substructure variables are not as powerful

as combinations of groomed mass and substructure

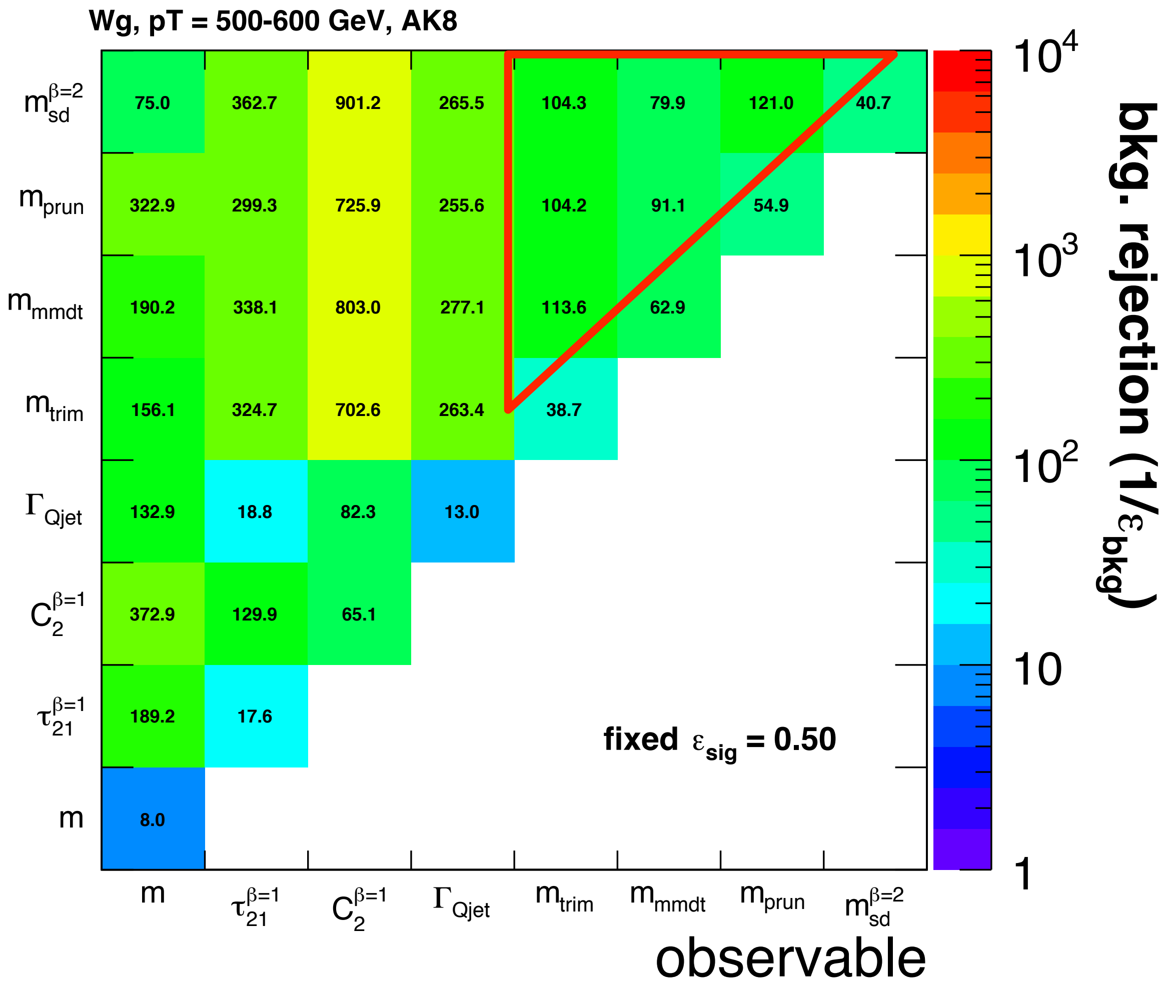
C_2 is particularly complementary to the groomed masses (but this is R dependent!)



R=0.8, p_T 500-600 GeV

Combined Variable Performance

There is complementary information between the different groomed masses

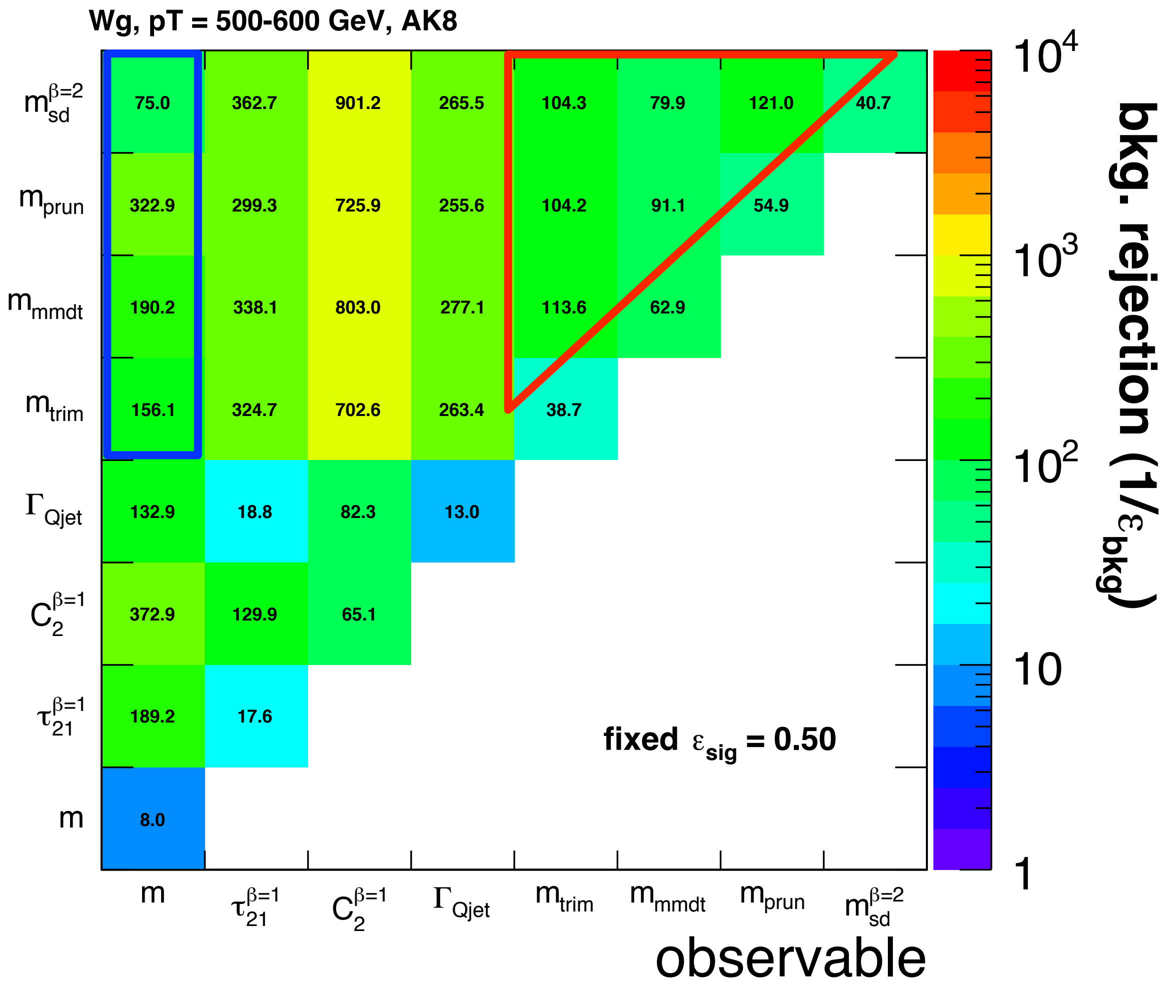


R=0.8, p_T 500-600 GeV

Combined Variable Performance

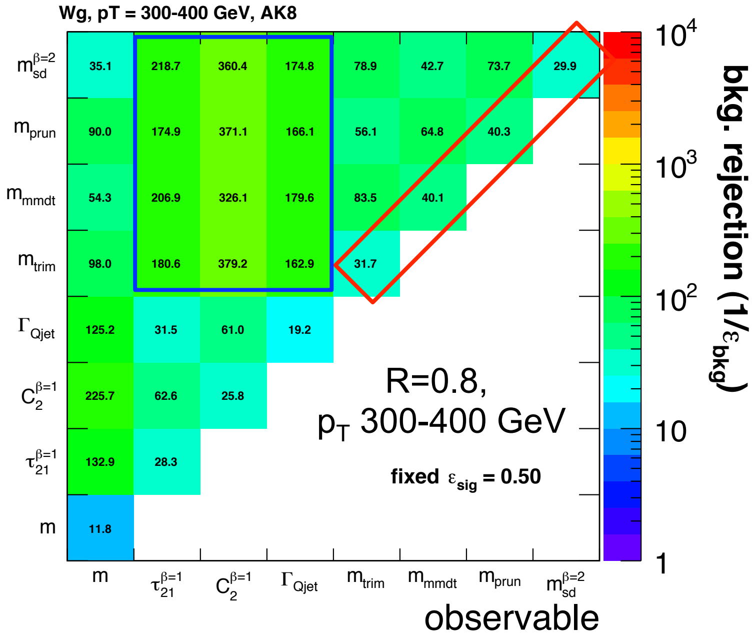
There is complementary information between the different groomed masses

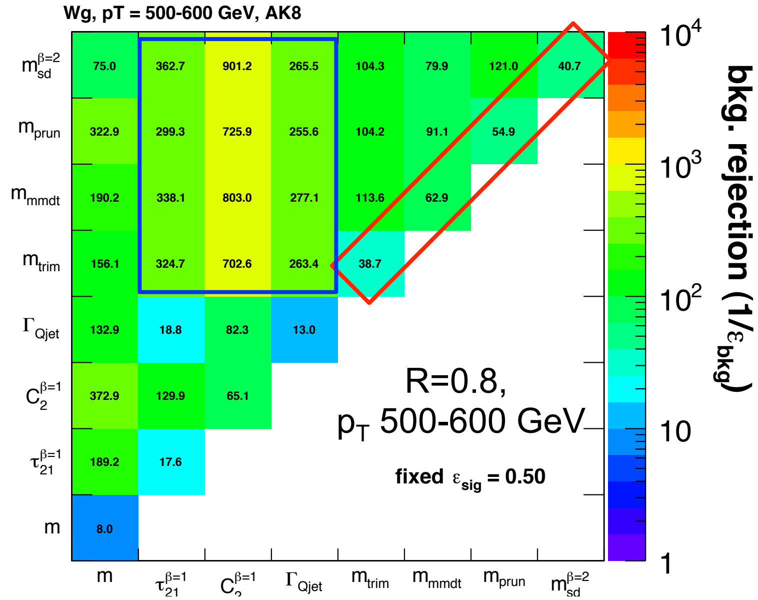
and between the groomed masses and ungroomed mass. Grooming cuts out some useful information!

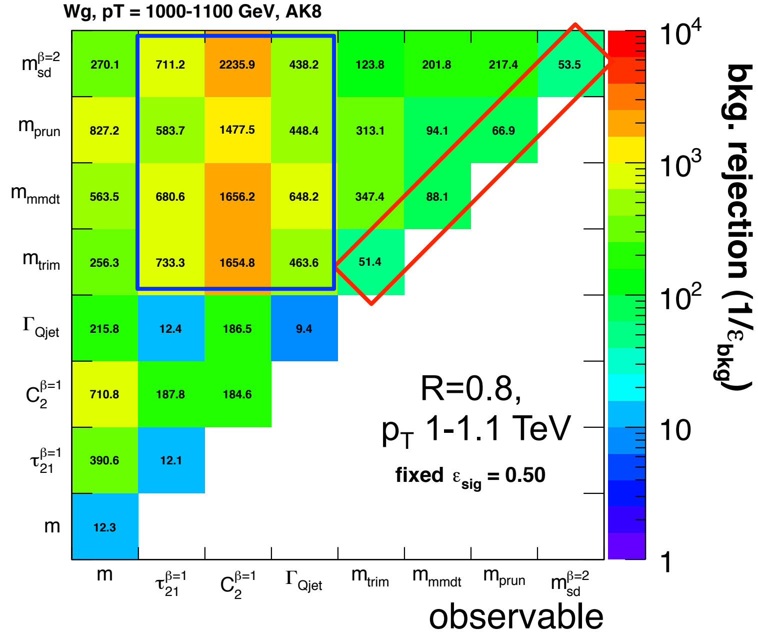


R=0.8, p_T 500-600 GeV

Dependence on p_T

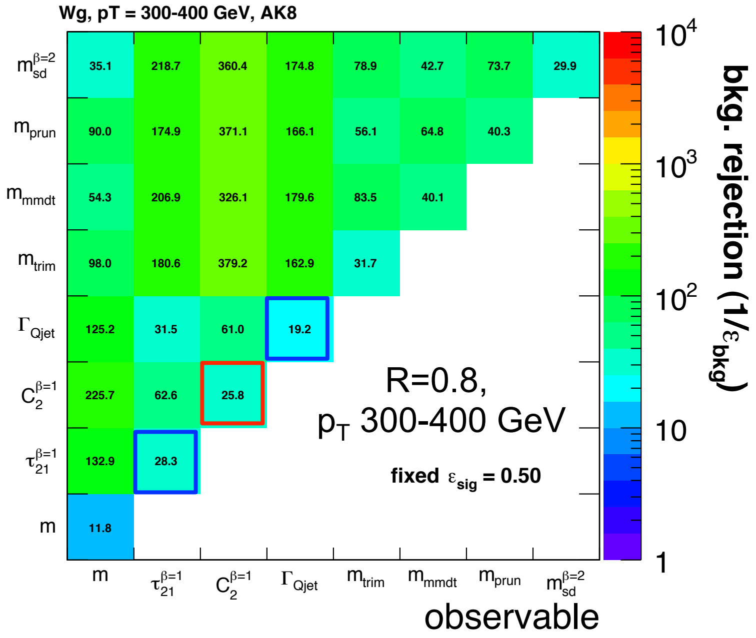


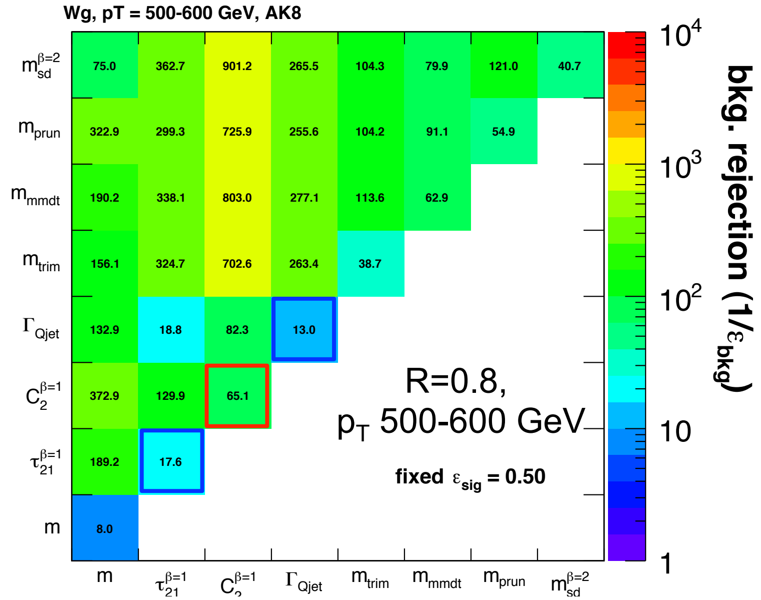


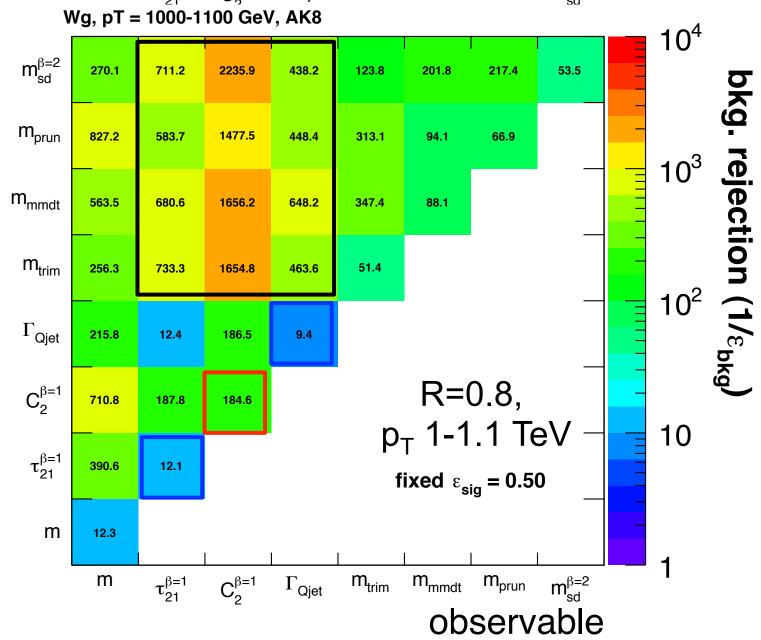


- As p_T increases the power of the groomed masses stays relatively constant (up to x 2)
- But the power of mass+substructure combination increases dramatically.
  - Addition of substructure information increasingly important to get best tagging at higher p_T.

Dependence on p_T

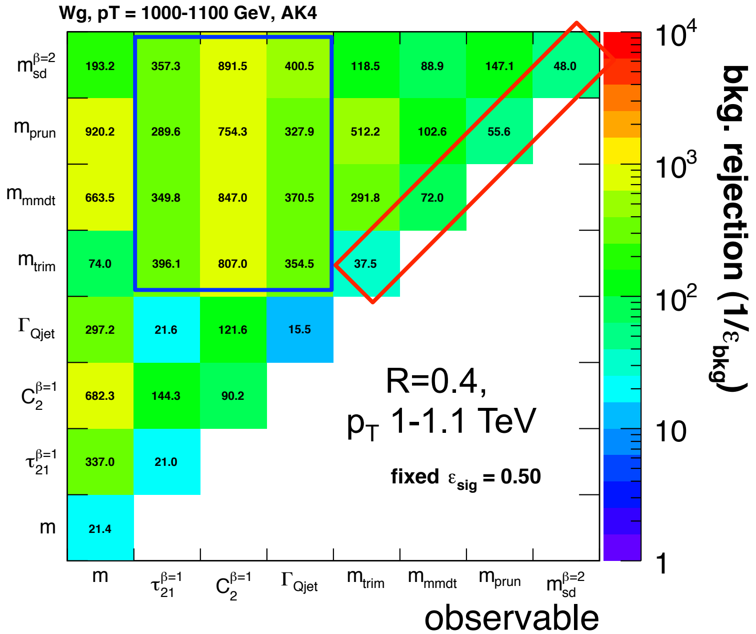


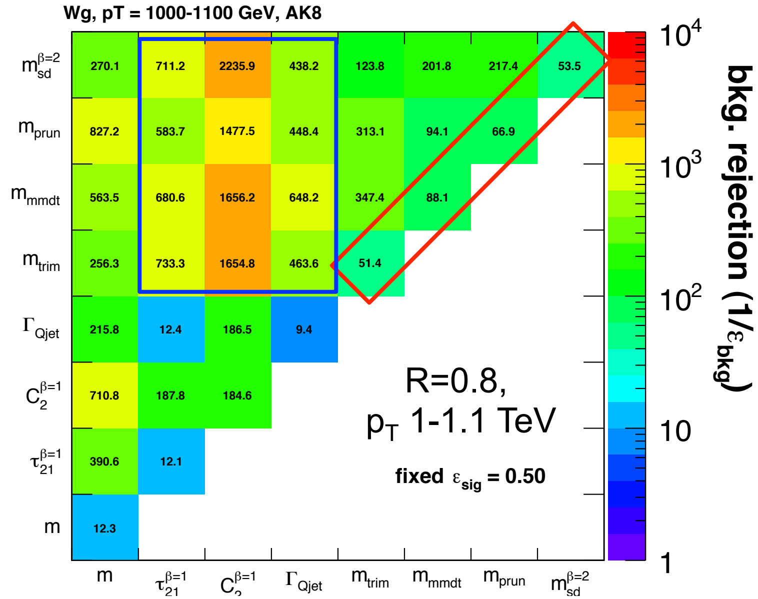


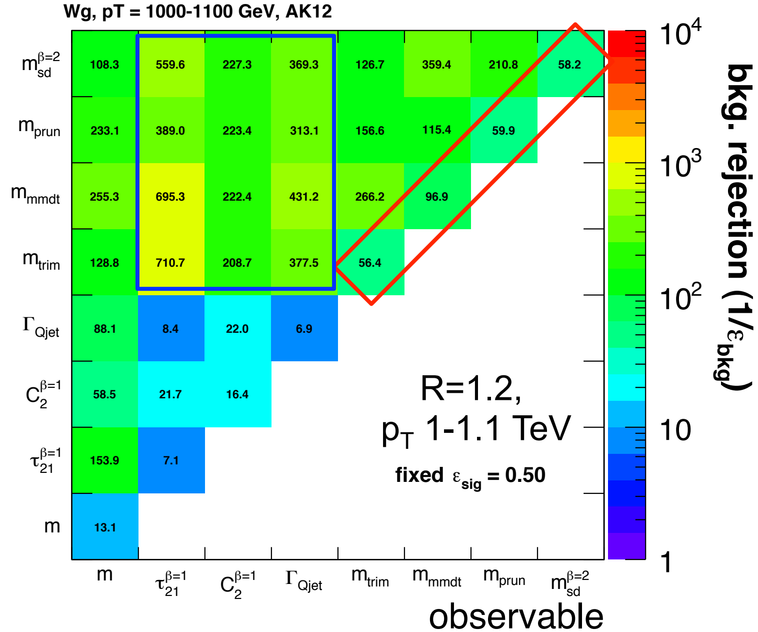


- Individual power of C2 increases dramatically with increased p_T.
- But individual power of τ_{21} and Γ_{Qjet} gets worse.
- Interesting differences in the performance of mass+shape taggers, especially evident at higher p_T.

Dependence on R







- Again, individual power of groomed masses stays relatively constant.
- But dramatic changes in the power of substructure variables as R changes.
  - At R=0.4 and 0.8 C_2 is by far the most powerful…
  - …but this all changes for R=1.2. τ_{21} becomes most powerful in combination.
- Power of groomed mass + shape taggers varies substantially with jet radius.

Main Conclusions from W Tagging

- Individually, groomed masses are more powerful discriminants than the substructure variables examined:
  - Exception to this is $C_2^{\beta=1}$, as powerful as groomed masses for R=0.4, R=0.8.
  - Groomed mass power does not vary too much with p_T or R.

- Taggers should be built from a combination of groomed mass + substructure variable:
  - Great improvements in rejection power, especially at high p_T.
  - Performance of combined taggers improves with p_T….
  - …but varies substantially with jet radius R.
  - Most performant substructure variable depends on R
    - For R=0.8 it is C_2
    - For R=1.2 it is τ_{21}
  - Different substructure variables prefer to be used in combination with different groomers e.g. C_2 best with $m_{sd}^{\beta=2}$

Top tagging Studies

All studies by Brian Shuve

Fastjet 3.03 jet framework used

Top Tagging MC Samples

- All samples at $\sqrt{s}$ = 14 TeV.

- Three different exclusive p_T bins:
  - 600-700 GeV, 1-1.1 TeV and 1.5-1.6 TeV.

- QCD background samples:
  - Sherpa 2.0.0, 2→2+2 generation (both qq and gg).
  - p_T cut on leading parton-level jet p_T.

- All-hadronic ttbar samples:
  - Sherpa 2.0.0
  - p_T cut on top/anti-top p_T.

Top Tagging Options Explored

- Study the performance of the following Top tagging strategies:
  - HepTopTagger (HEP)
  - John Hopkins Tagger (JH)
  - Trimming & Pruning

- First compare the “mass” performance of these taggers.
  - Across different p_T bins, with different jet radii (R=0.4,0.8,1.2).

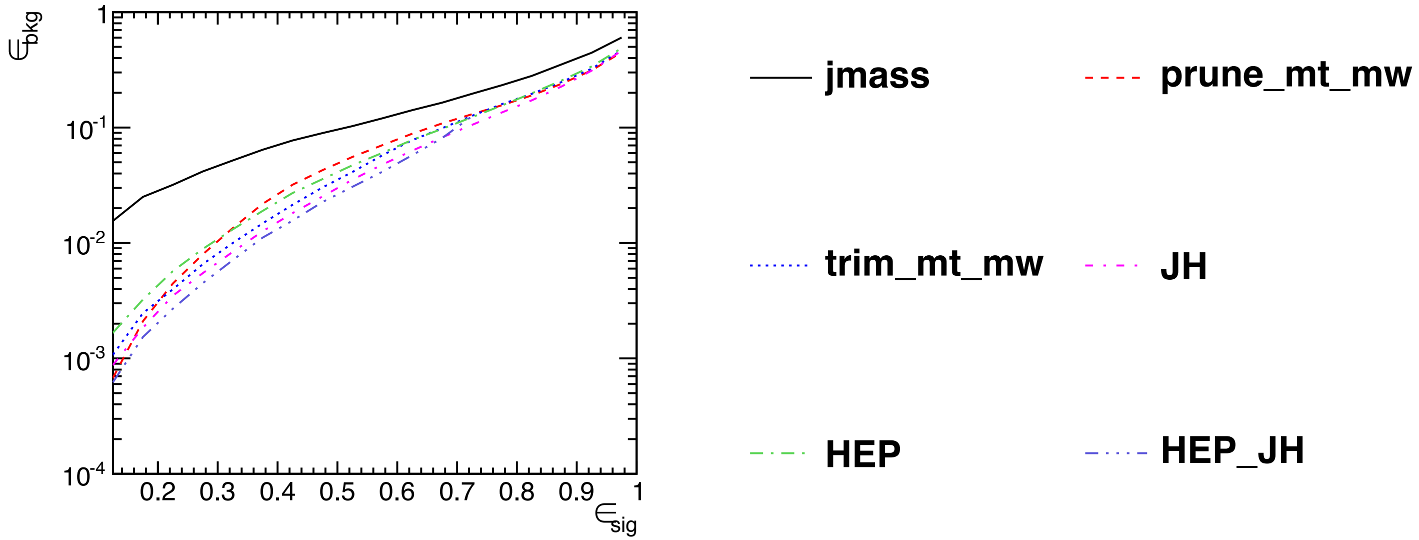
- Then investigate adding shape information:
  - n-subjettiness (τ_{21} and τ_{32})
  - ECF ratios (C_2 and C_3)
  - Qjet volatility

“Mass” Tagging

- **HEP and JH taggers:**
  - Output (when tagging requirements pass) a Top mass (m_{top}) and W mass (m_W) hypothesis, as well as a helicity angle.
  - We study the performance when m_{top}, m_W and helicity angle are combined in a BDT…(similar to Boost2011 report).

- **Trimming/Pruning** “mass” top tagging works as follows:
  - m_{top} = the full groomed jet mass.
  - m_W = lowest mass pair (if >=3 subjets) or highest mass subjet (if ==2 subjets).
  - Combine m_{top} and m_W in a BDT discriminant.

Continuous Optimisation

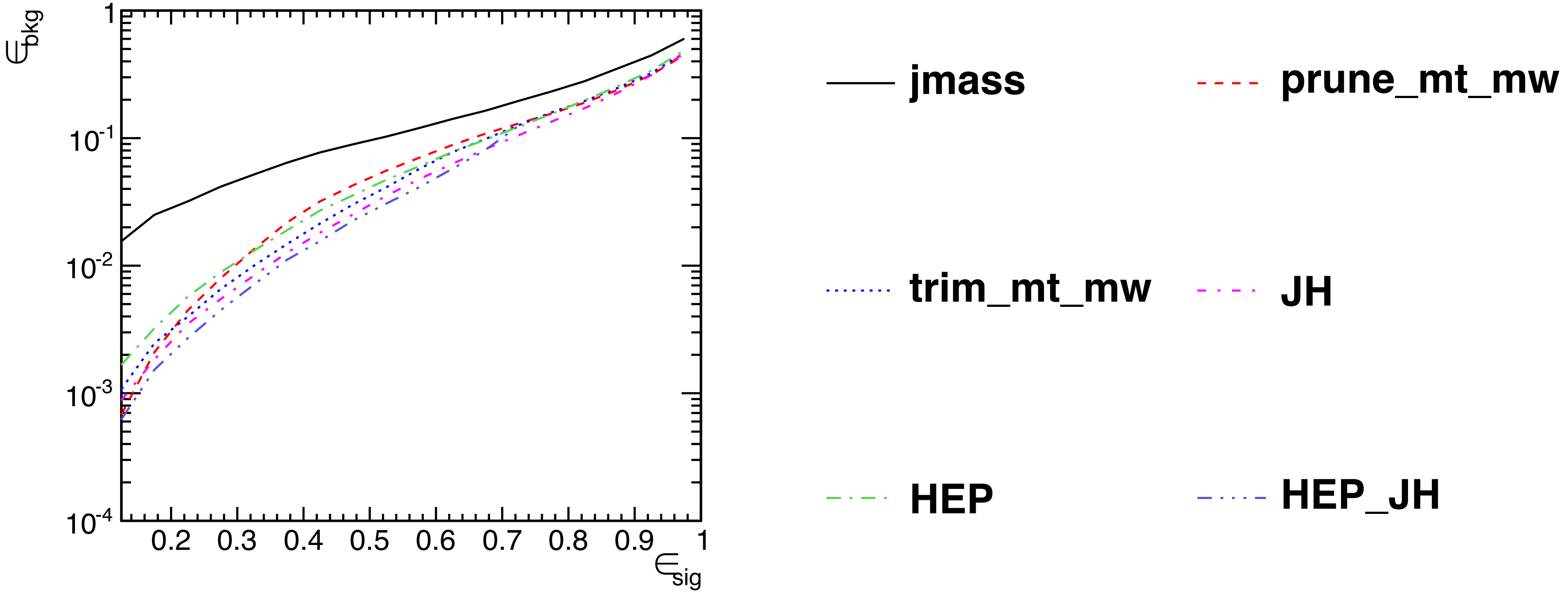


- For each point on each ROC curve the tagger "inputs" are scanned over to give the optimal background rejection at that efficiency.

HEPTopTagger: $m \in [30, 100]$ GeV, $\mu \in [0.5, 1]$
JH Tagger: $\delta_p \in [0.02, 0.15]$, $\delta_R \in [0.07, 0.2]$
Trimming: $f_{\mathrm{cut}} \in [0.02, 0.14]$, $R_{\mathrm{trim}} \in [0.1, 0.5]$
Pruning: $z_{\mathrm{cut}} \in [0.02, 0.14]$, $D_{\mathrm{cut}} \in [0.1, 0.6]$

Tagger Performance

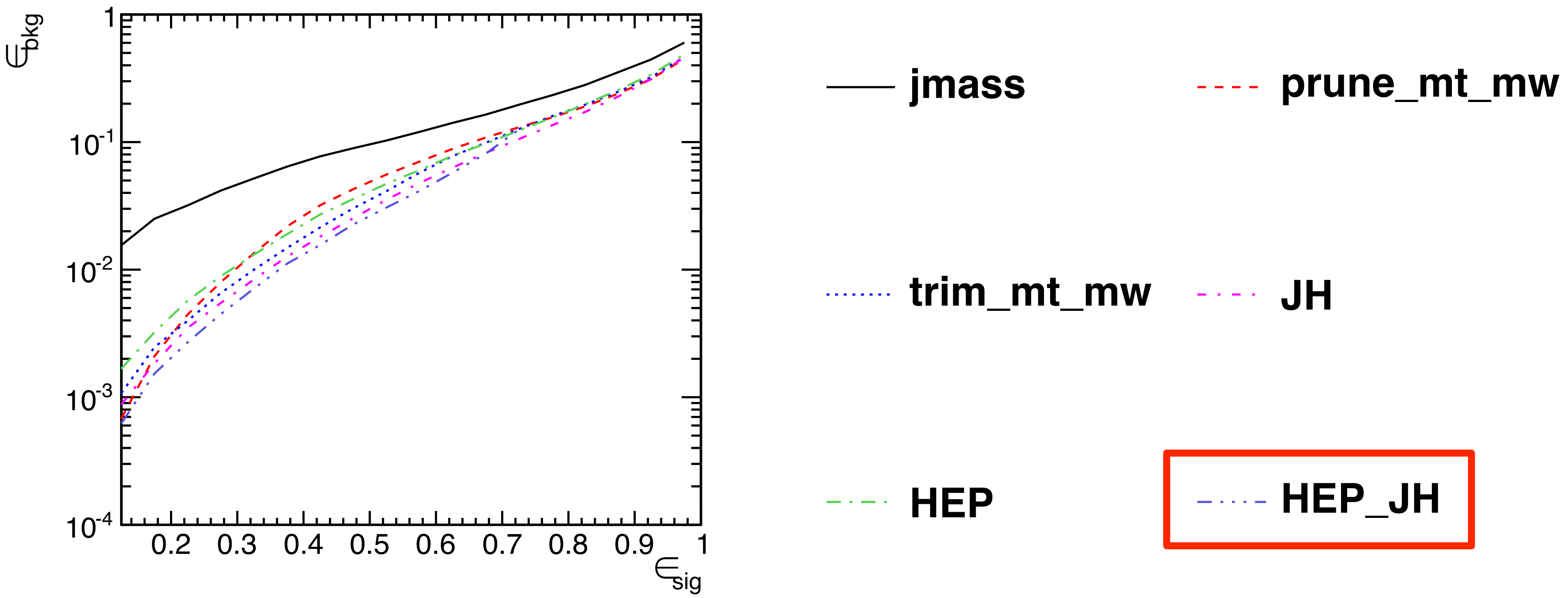
- All approaches are similar in performance.
- JH appears to perform slightly better than HEP.
- Trimming slightly better than pruning.



p_T 1000-1100 GeV bin, R=0.8

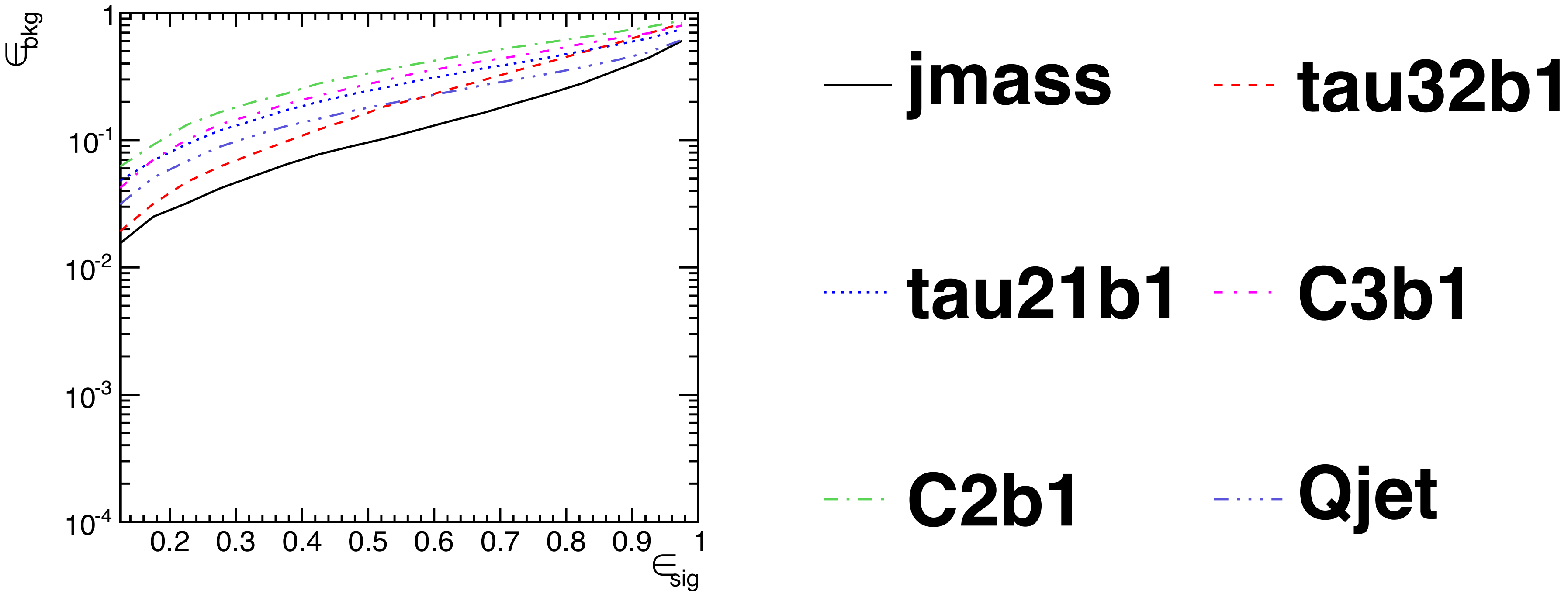
Correlations Between Taggers

- The JH tagger can be improved by BDT combination of the JH outputs with the HEP tagger outputs.
- There is complementary information in the outputs of these taggers!



p_T 1000-1100 GeV bin, R=0.8

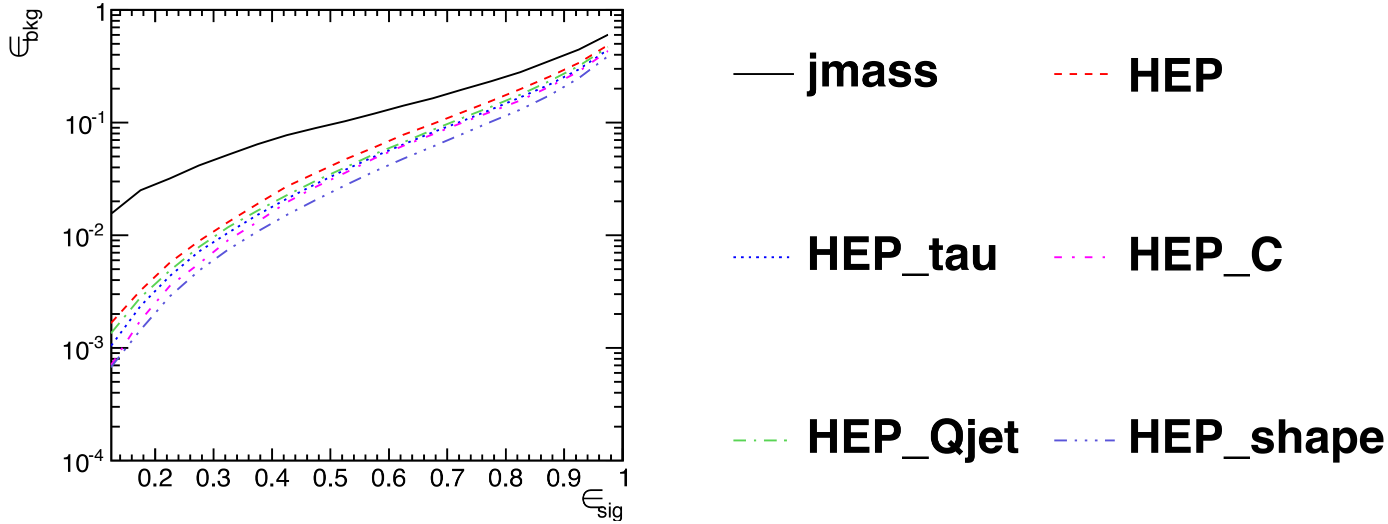
Shape-Only Performance



p_T 1000-1100 GeV bin,
R=0.8

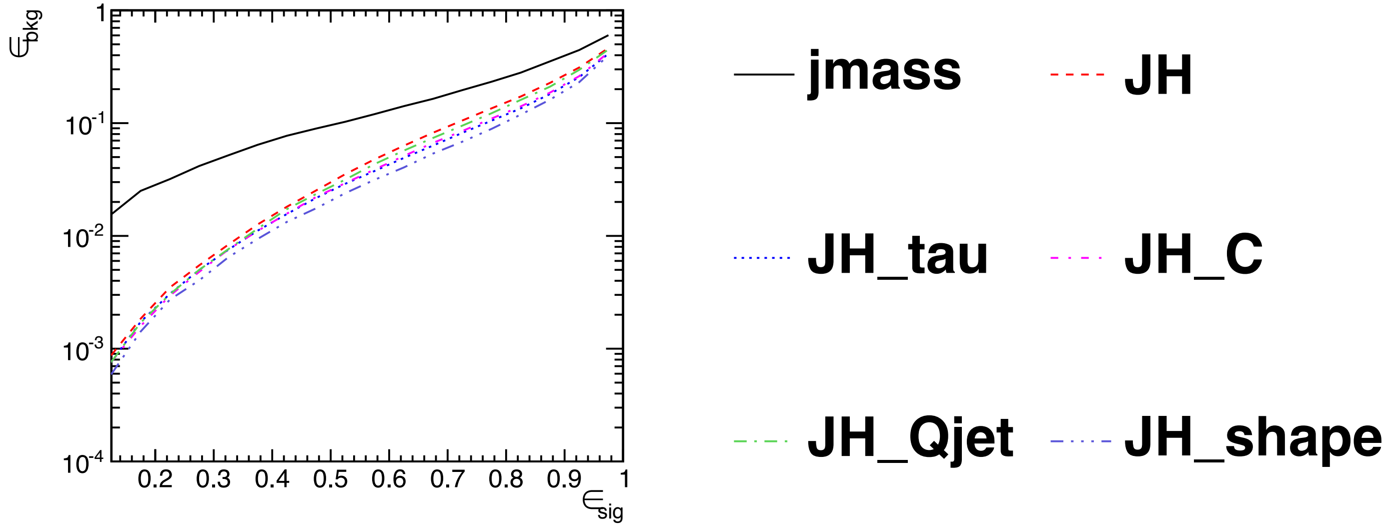
- Jet shape/substructure variables not as powerful as the ungroomed mass.

Adding Shape Information



p_T 1000-1100 GeV bin, R=0.8

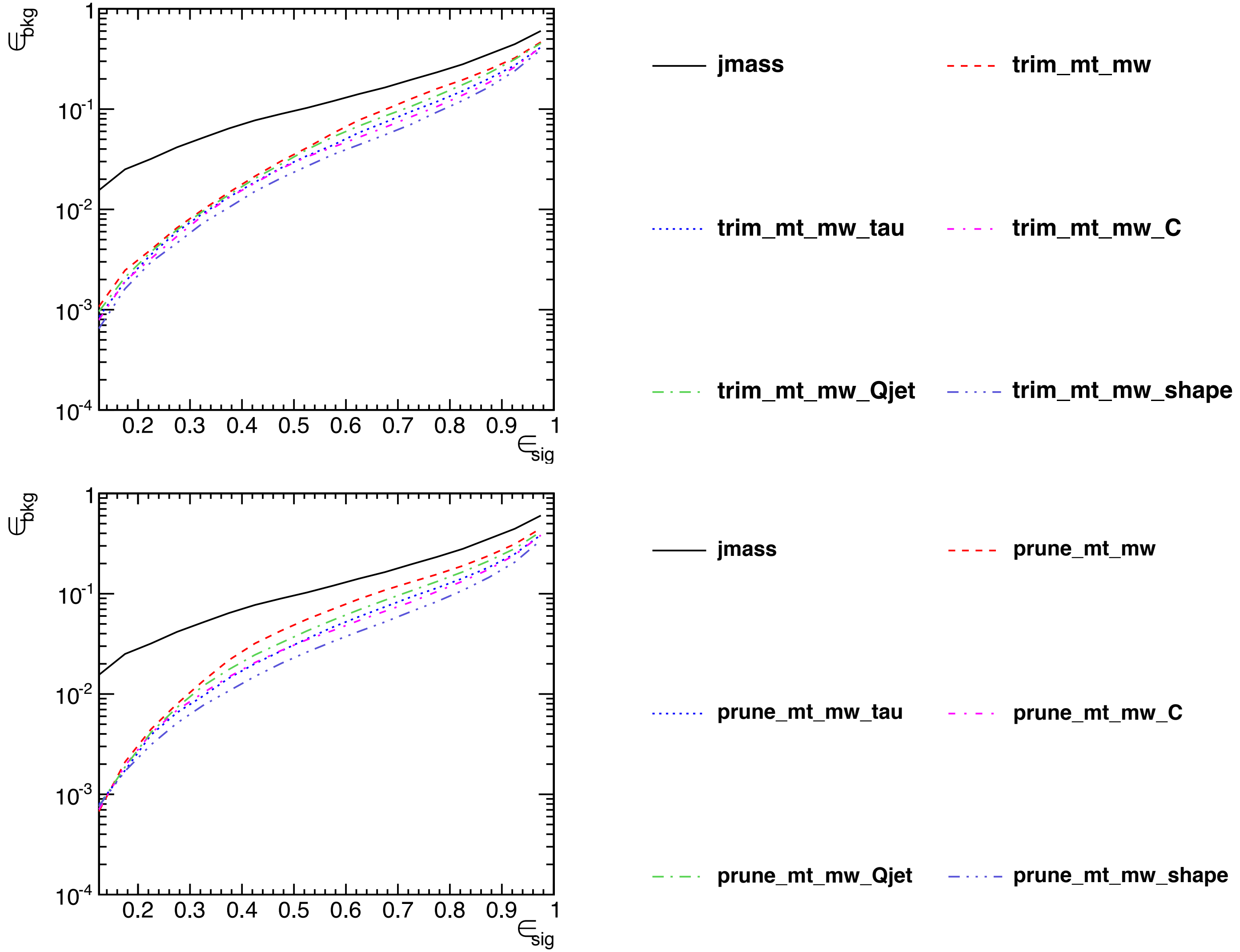
HEP tagger shows bigger improvement when adding shape. No one shape variable dominant.



JH tagger shows less improvement.

- Use BDT from m_{top}, $m_{W,}$ helicity and single (or all) shape variable(s).
- Both HEP and JH are complimented by additional shape info…

Adding Shape Information



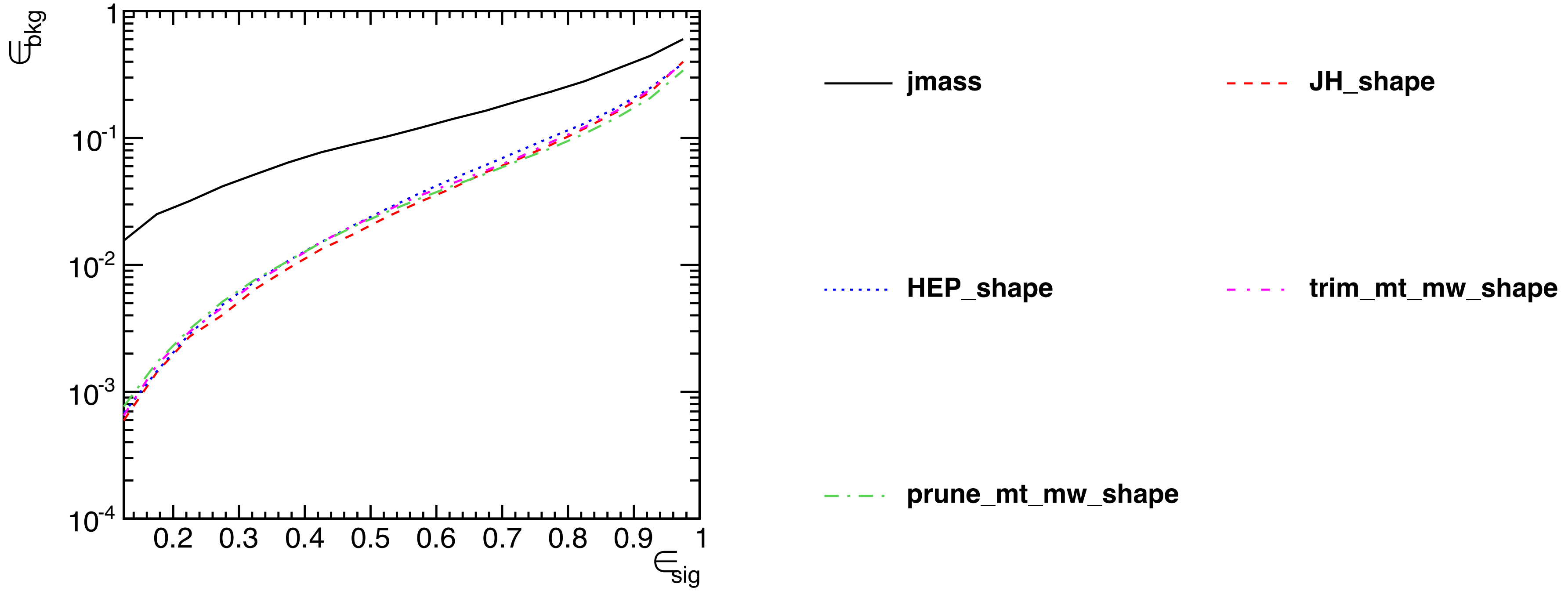
p_T 1000-1100 GeV bin, R=0.8

C_2/C_3 and Tau21/Tau23 seem to be the most complementary shape variables.

- Use BDT from m_{top}, m_W and single (or all) shape variable(s).
- Trimming and pruning complimented by shapes also.

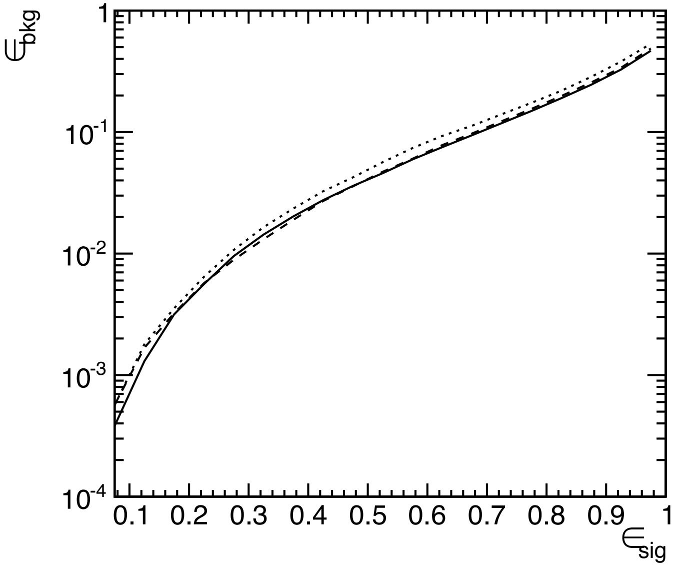
Adding Shape Information

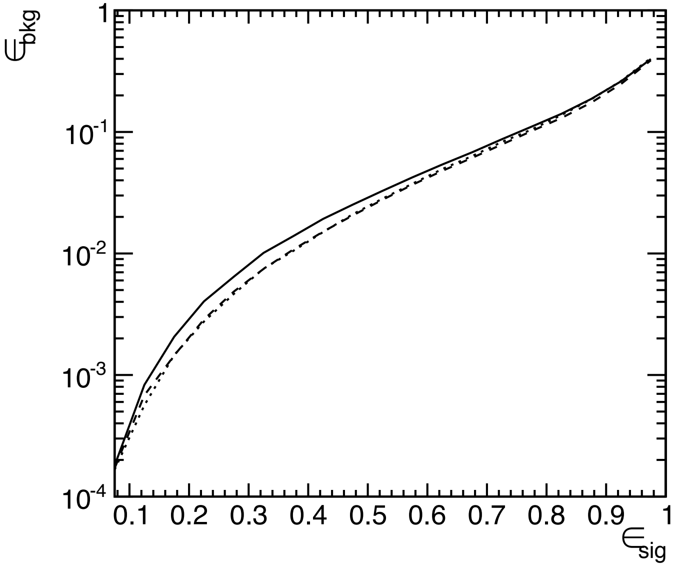
p_T 1000-1100 GeV bin, R=0.8



- Performance of the various strategies very close after adding all shape information (at least all explored here).

HEP p_T dependence (R=0.8)





——— **HEP.pT.600**

- - - - - **HEP.pT.1000**

·········· **HEP.pT.1500**

——— **HEP_shape.pT.600**

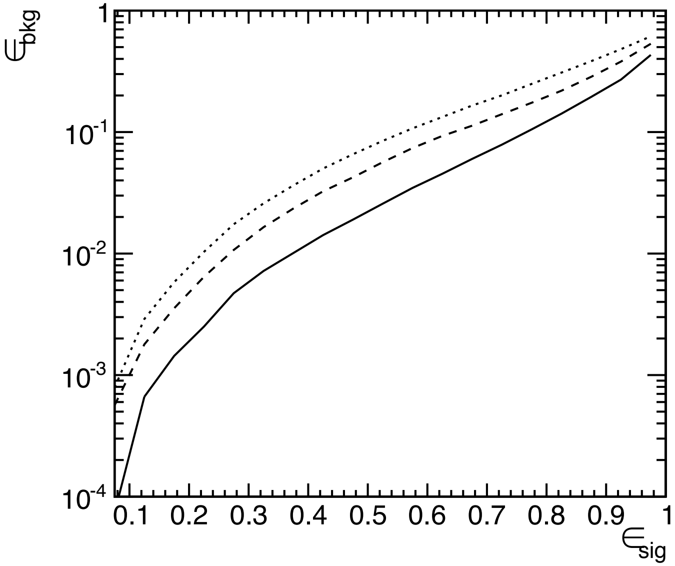
- - - - - **HEP_shape.pT.1000**

·········· **HEP_shape.pT.1500**

Compare performance of taggers in different p_T bins (each p_T bin individually optimised).

In all taggers, optimal performance stays fairly constant with p_T.

HEP R dependence (p_T = 1.5 TeV)

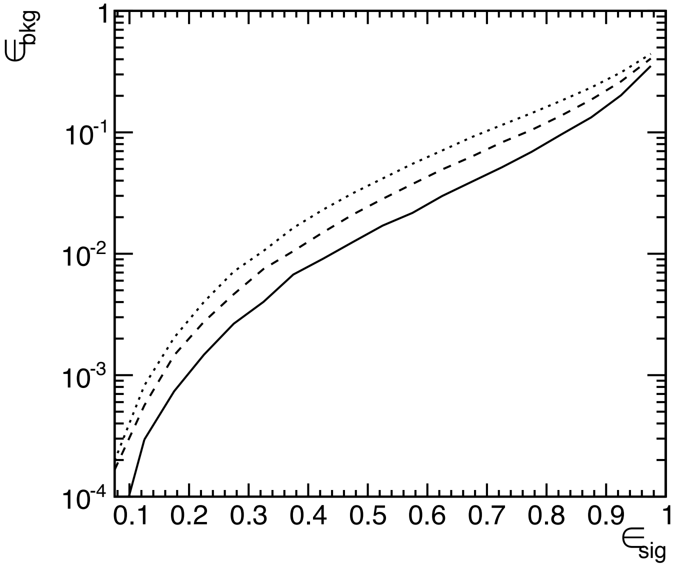


HEP.R.0.4

HEP.R.0.8

HEP.R.1.2

Compare tagger performance as a function of jet radius in the most boosted bin.



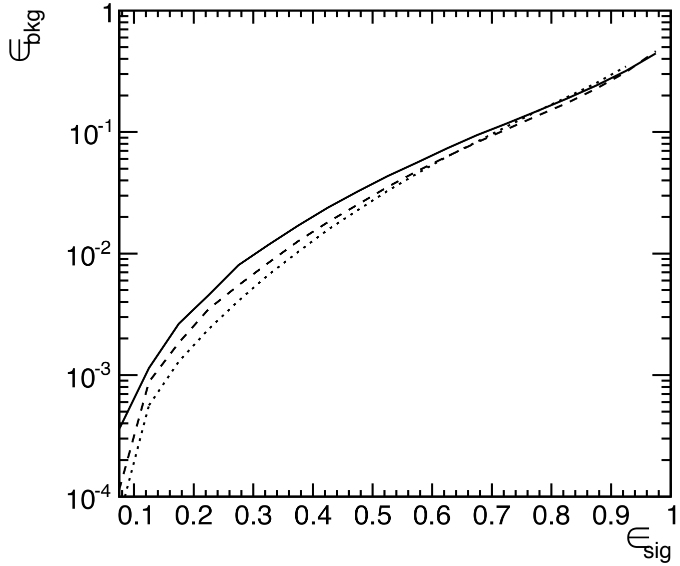
HEP_shape.R.0.4

HEP_shape.R.0.8

HEP_shape.R.1.2

Taggers prefer smaller jet radius (as also observed in W-tagging)

Optimization Transfer Studies

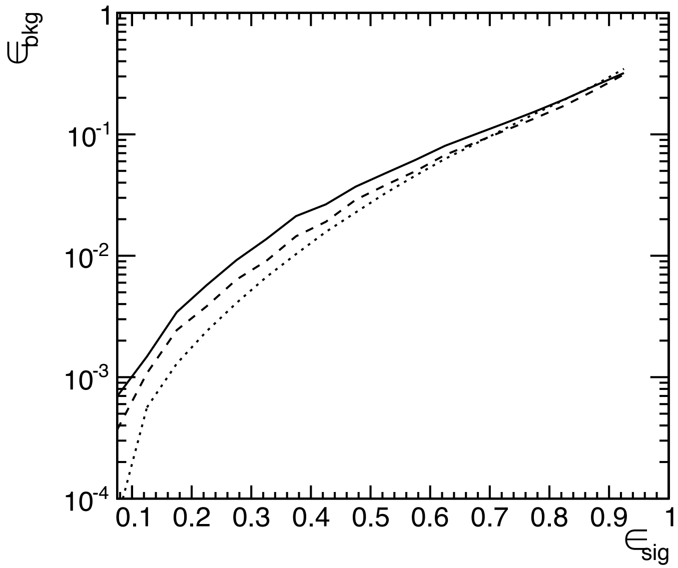


JH.pT.600 (solid line)

JH.pT.1000 (dashed line)

JH.pT.1500 (dotted line)

Individually optimised in each p_T bin



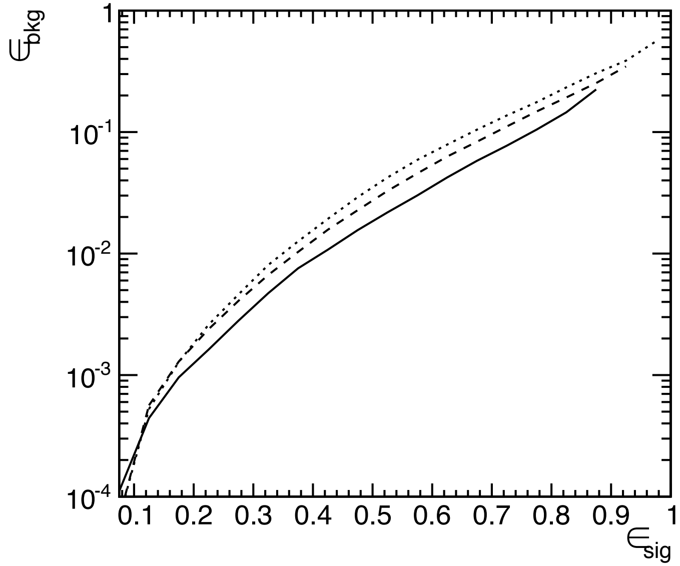
JH.pT.1500 (dotted line)

JH.pT.1000 (dashed line)

JH.pT.600 (solid line)

Only optimised in the p_T 1.5 TeV bin. Performance is very similar!

Optimization Transfer Studies

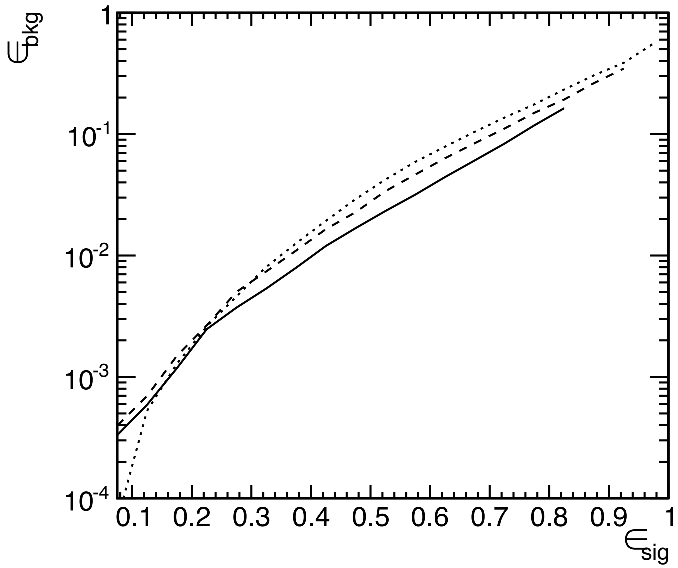


JH.R.0.4

JH.R.0.8

JH.R.1.2

Individually optimised for each jet radius



JH.R.1.2

JH.R.0.8

JH.R.0.4

Only optimised for R=1.2. Performance is very similar!

Main Conclusions Top Tagging

- When optimising over all inputs HEP, JH, trimming and pruning "mass" taggers can produce similar Top tagging performance.
  - Performance does not vary strongly with p_T, but does vary strongly with R (lower R preferred).
  - There is complementary information between these taggers.

- All of the "mass-based" taggers can be improved using additional shape information.
  - Performance of taggers becomes very close when shape added.
  - C2/C3 and tau21/tau32 are most complementary to trimming/ pruning.

Conclusions in common with W-tagging study

Summary

Report Status

Towards an Understanding of the Correlations in Jet Substructure

Report of BOOST2013, hosted by the University of Arizona, 12^{th}-16^{th} of August 2013.

D. Adams[1], A. Arce[2], L. Asquith[3], M. Backovic[4], T. Barillari[5], P. Berta[6], D. Bertolini[2], A. Buckley[8], J. Butterworth[9], R. C. Camacho Toro[10], J. Caudron[9], Y.-T. Chien[11], J. Cogan[12], B. Cooper[9], D. Curtin[17], C. Debenedetti[18], J. Dolen[9], M. Eklund[22], S. El Hedri[22], S. D. Ellis[22], T. Embry[22], D. Ferencek[23], J. Ferrando[24], S. Fleischmann[16], M. Freytsis[25], M. Giulini[21], Z. Han[27], D. Hare[4], P. Harris[4], A. Hinzmann[4], R. Hoing[4], A. Hornig[22], M. Jankowiak[4], K. Johns[28], G. Kasieczka[23], T. Knight[24], G. Kasieczka[29], R. Kogler[30], W. Lampl[4], A. J. Larkoski[4], C. Lee[31], R. Leone[31], P. Loch[31], D. Lopez Mateos[27], H. K. Lou[27], M. Low[27], P. Maksimovic[32], I. Marchesini[32], S. Marzani[32], L. Masetti[33], R. McCarthy[32], S. Menke[32], D. W. Miller[35], K. Mishra[36], B. Nachman[32], P. Nef[4], F. T. O'Grady[24], A. Ovcharova[23], A. Picazio[37], C. Pollard[38], B. Potter Landua[29], C. Potter[29], S. Rappoccio[39], J. Rutherfoord[40], G. P. Salam[10,11], J. Schabinger[23], A. Schwartzman[4], M. D. Schwartz[27], B. Shuve[43], P. Sinervo[44], D. Soper[45], D. E. Sosa Corral[45], M. Spannowsky[32], E. Strauss[34], M. Swiatlowski[4], J. Thaler[34], C. Thomas[34], E. Thompson[1], N. V. Tran[36], J. Tseng[36], E. Usai[36], L. Valery[36], J. Veatch[23], M. Vos[23], W. Waalewijn[4], and C. Young[47]

- A lot of material already in the report.
  - Current draft is 58 pages, with 62 figures!
- More work needed to distill the plots and fill in the text.
- Nothing shown here on q/g tagging – plots produced but not digested.
- Aiming to wrap this up by Autumn/Fall (meaning journal submission).

Summary

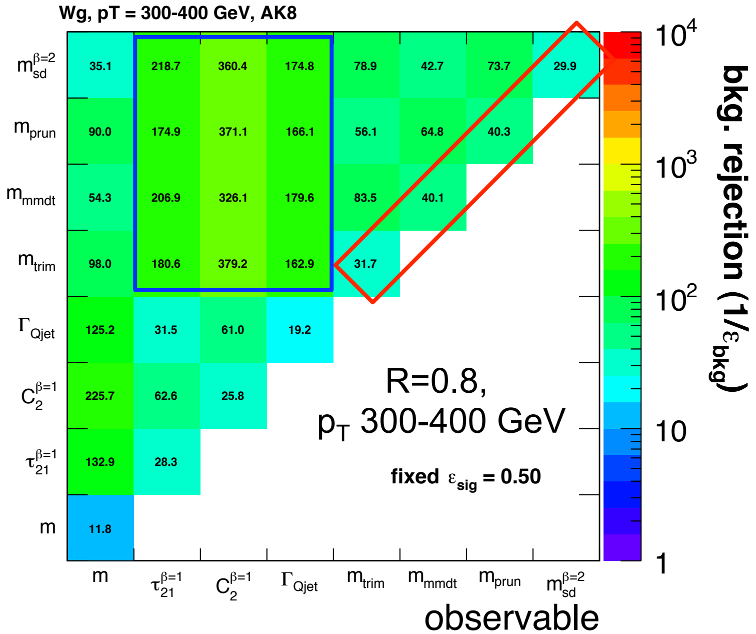
- The Boost13 Report systematically examines the complementarity and overlap in different tagging approaches, and how this changes as a function of p_T and jet radius, for W, Top and q/g tagging.

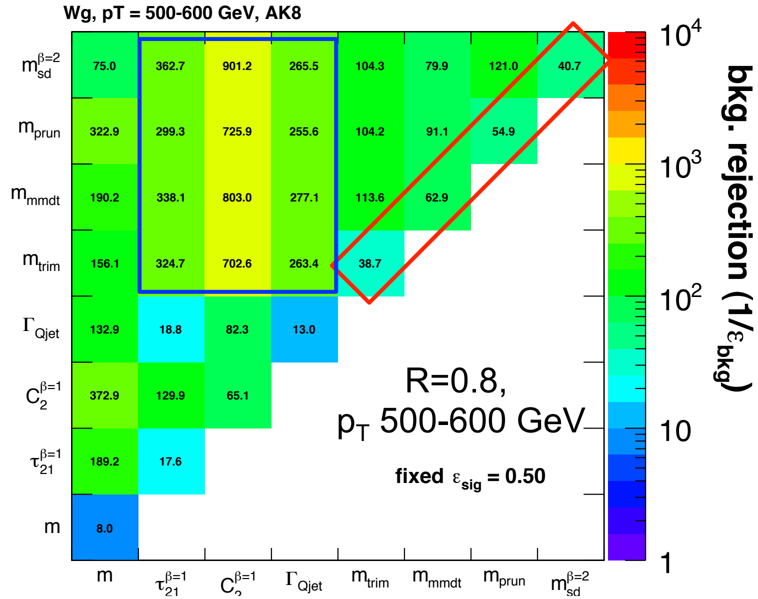
- Hopefully it can be instructive for both phenomenological and experimental communities.
  - What is there left to exploit?
  - How can we build a better tagger?

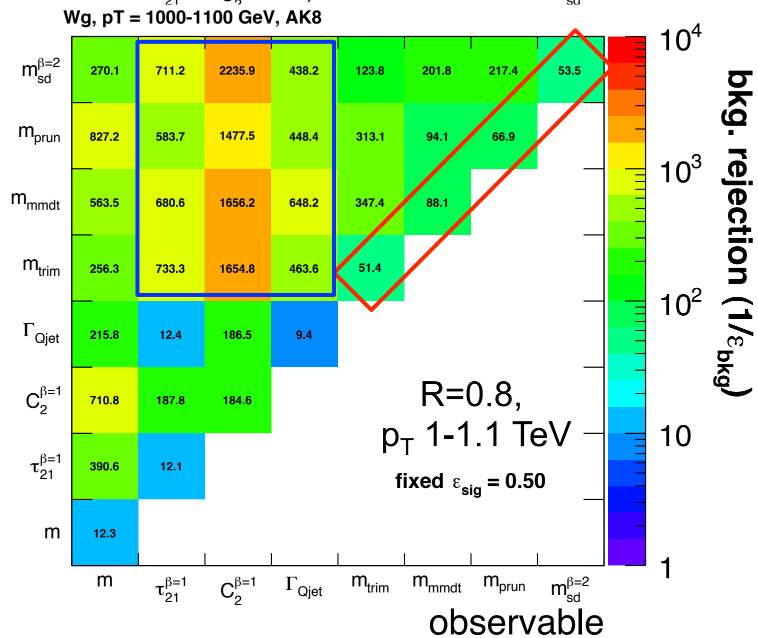
- Expect a complete first draft in next 4 weeks.

Backups

Dependence on p_T (R=0.8)

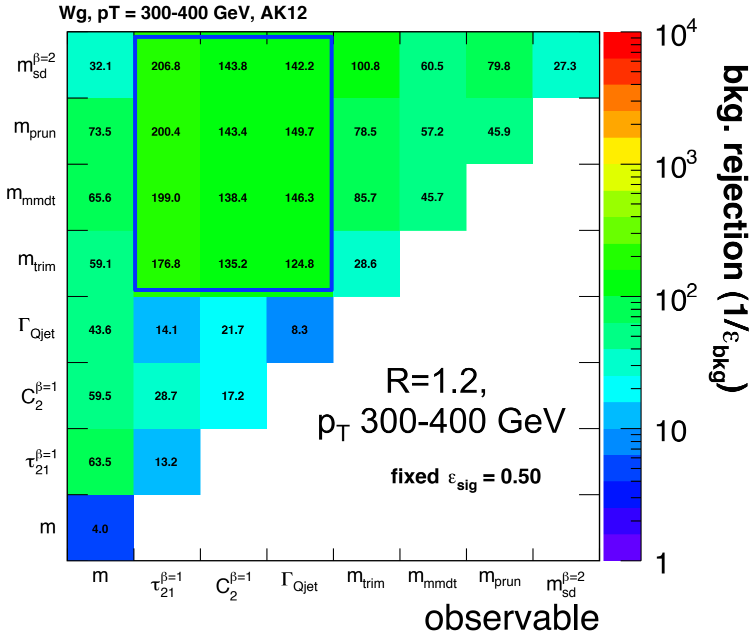


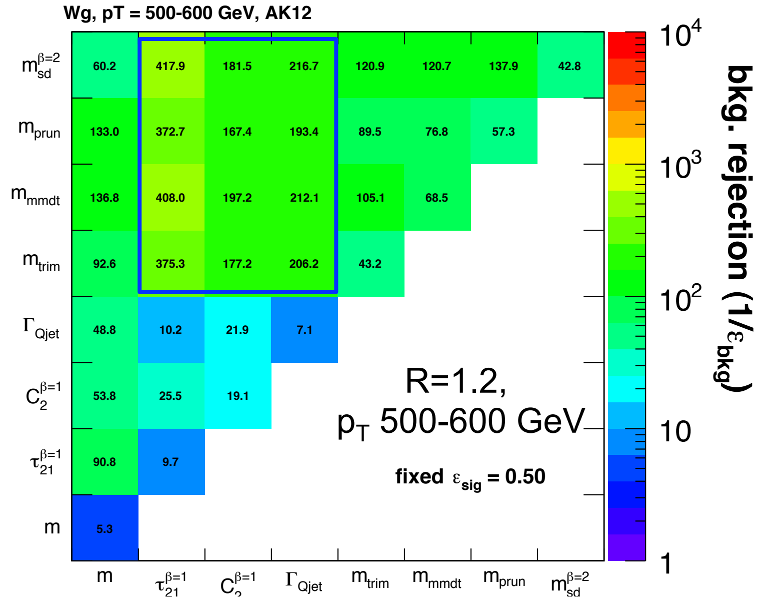




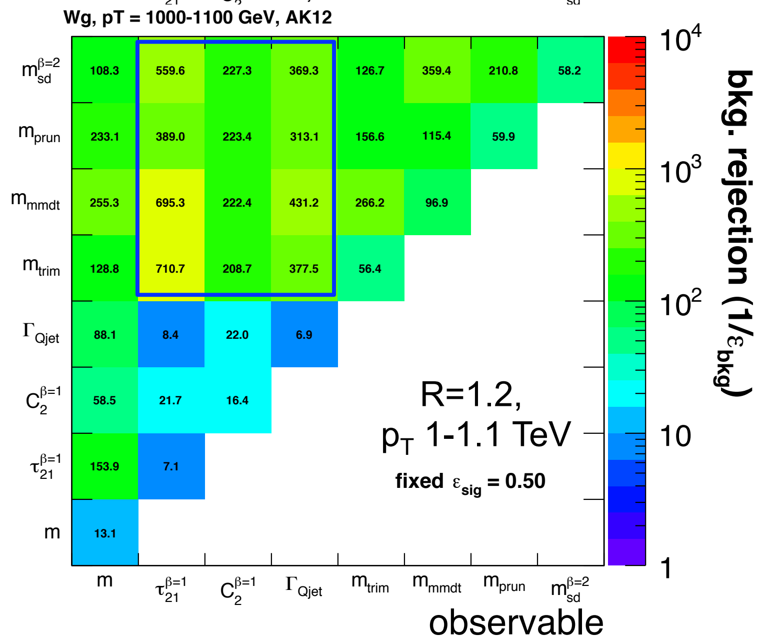
- As p_T increases the power of the groomed masses stays relatively constant (up to x 2)
- But the power of mass+substructure combination increases dramatically.
  - Addition of substructure information increasingly important to get best tagging at higher p_T.

Dependence on p_T (R=1.2)

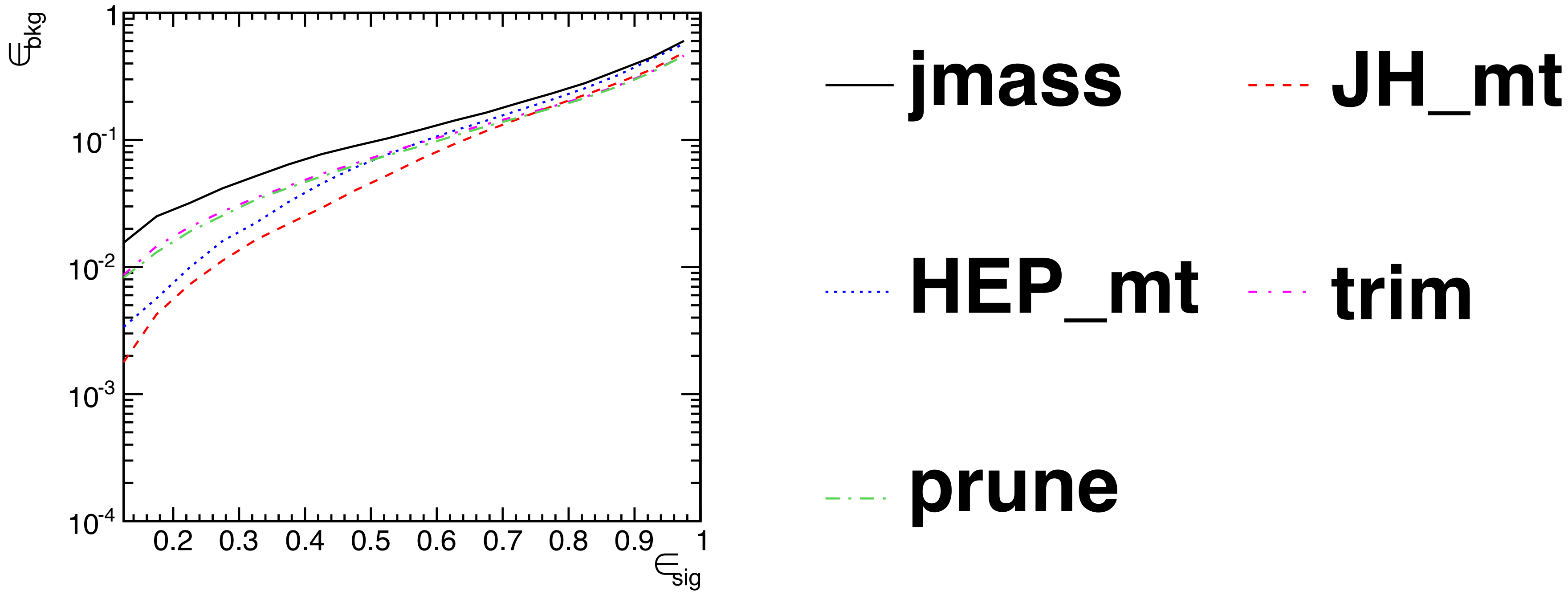




- C2 loses power relative to the other variables as p_T increases.
- Tau21 is the most powerful variable in combination at all p_T when using R=1.2.



Single Variable m_{top} Performance



p_T 1000-1100 GeV bin,
R=0.8

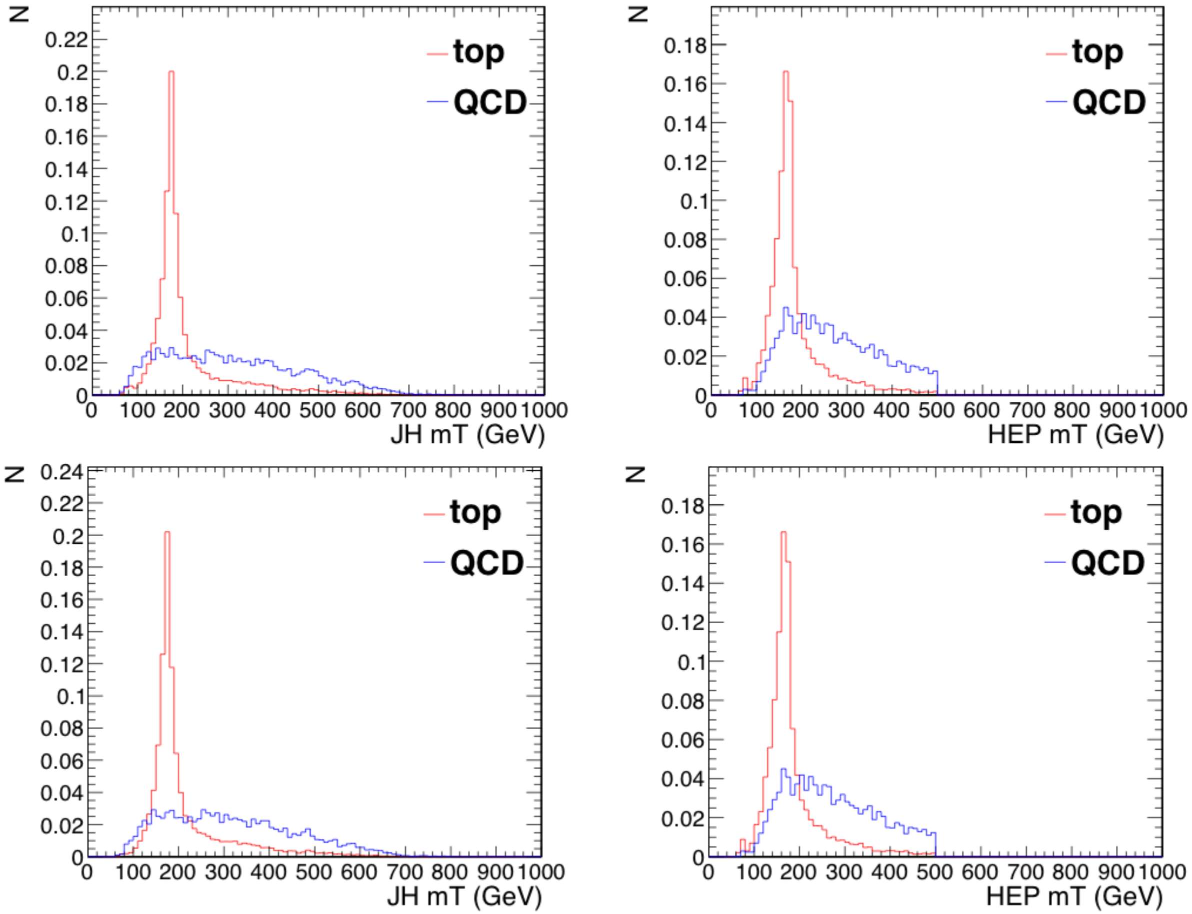
- Just using the reconstructed top mass for discrimination.
- Trimming and pruning perform very comparably.
- JH outperforms HEP.

Single Variable m_{top} Performance

p_T 1500-1600 GeV bin, R=0.8

Top: A particular optimised point on the ROC, where full 3 variable (m_{top}, m_W, helicity) BDT is used in discrimination

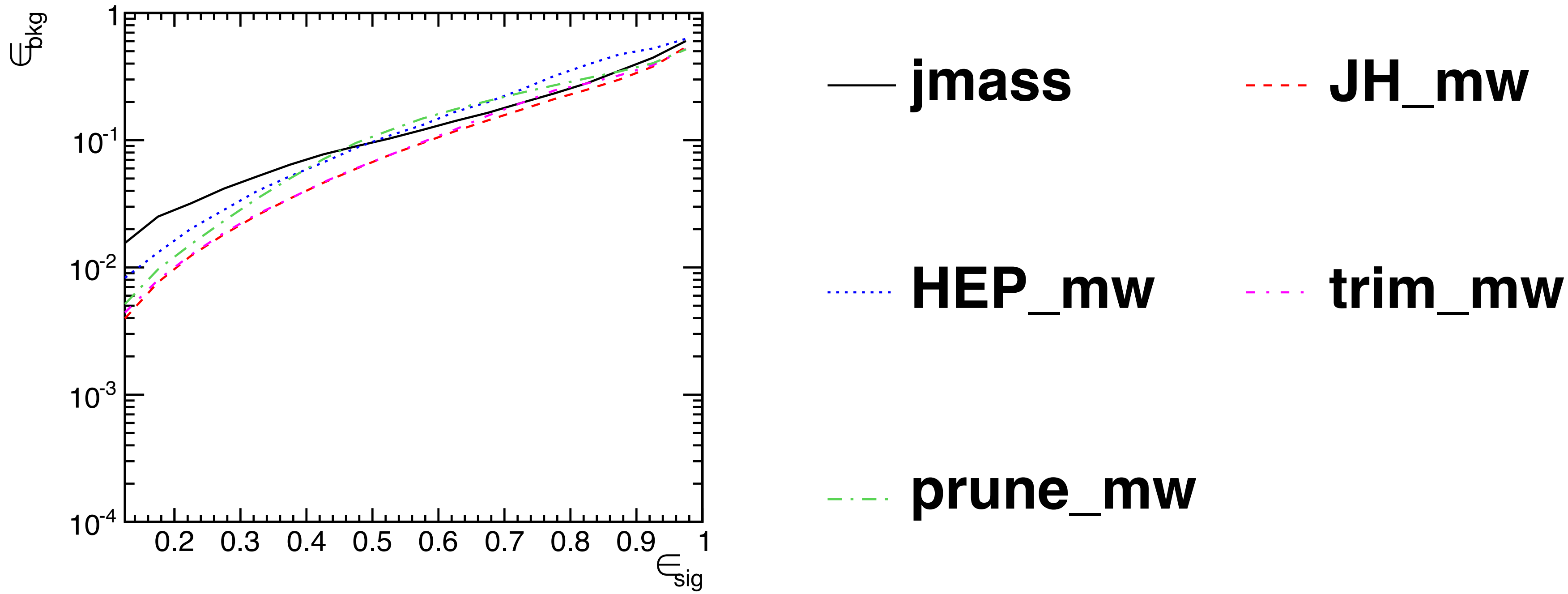
Bottom: A particular optimised point on the ROC, where m_{top} only is used in discrimination



(c) Johns Hopkins Tagger, $R = 0.8$ (d) HEPTopTagger, $R = 0.8$

- The HEP sculpts the QCD mass distribution to look more like Top.
  - Due to selection of subjet “triplet” closest to Top mass.
- Therefore you don’t get such a good Top mass discrimination.

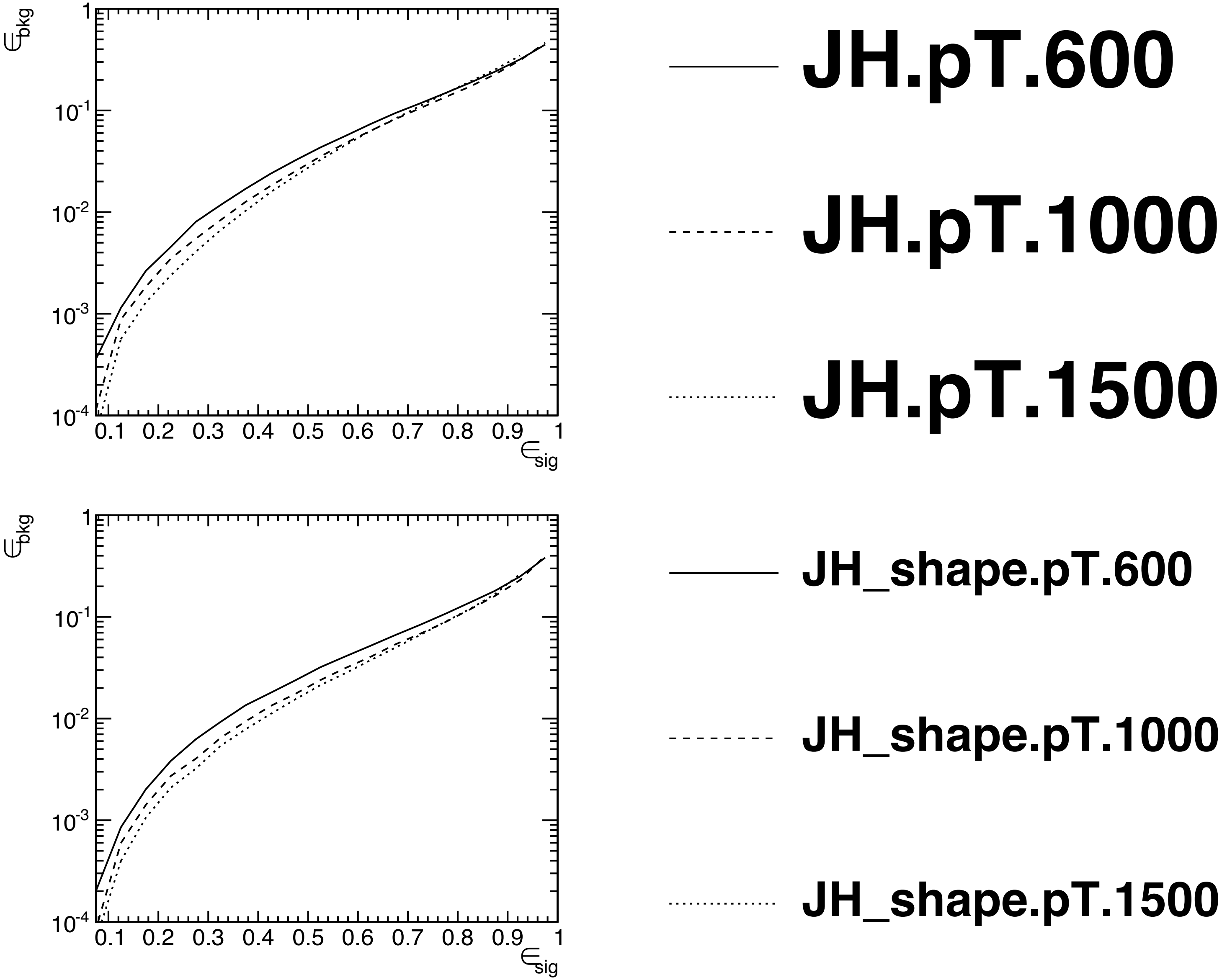
Single Variable Performance



p_T 1000-1100 GeV bin,
R=0.8

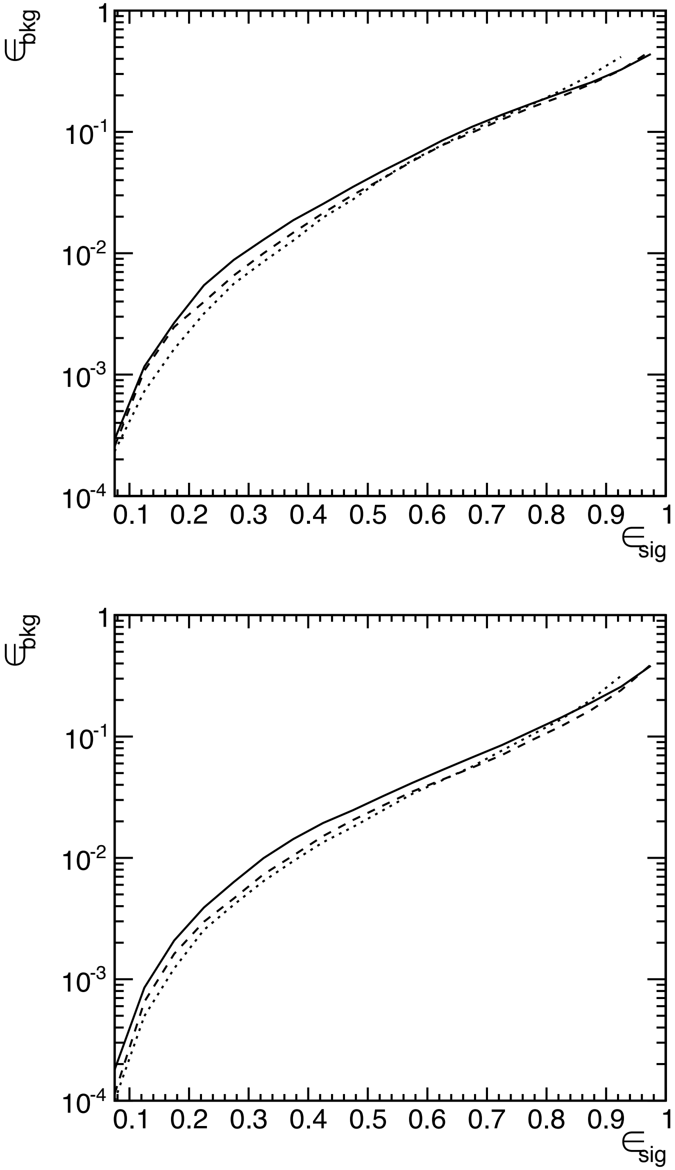
- Just using the reconstructed W mass for discrimination (except “jmass” curve, which always uses ungroomed full jet mass)
- Trimming is better than pruning here (very close to JH).
- JH again outperforms HEP.

JH p_T dependence (R=0.8)



JH performance improves very slightly with p_T

Trimming p_T dependence (R=0.8)



trim_mt_mw.pT.600

trim_mt_mw.pT.1000

trim_mt_mw.pT.1500

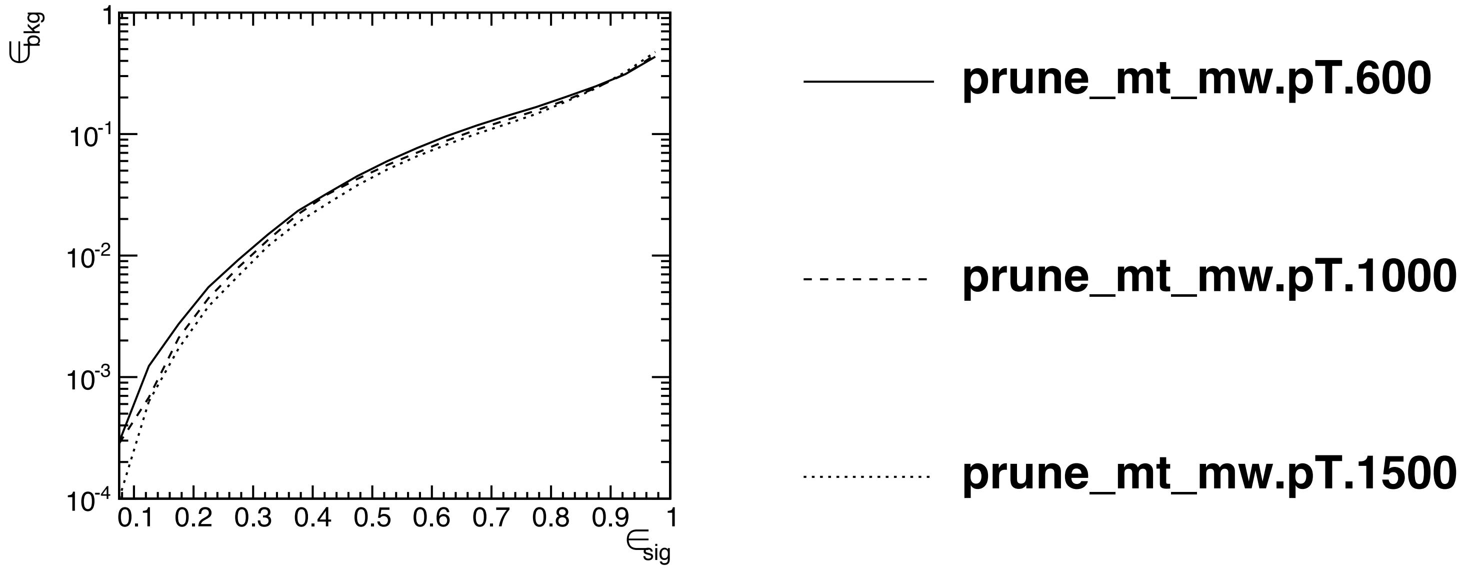
trim_mt_mw_shape.pT.600

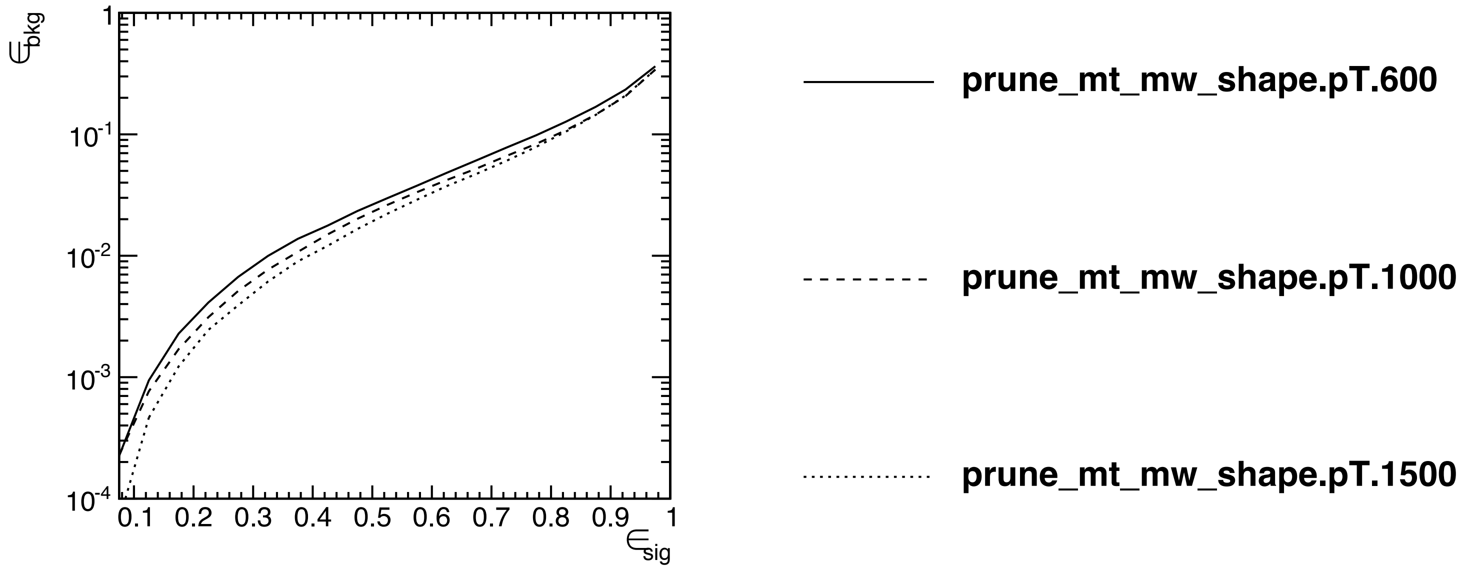
trim_mt_mw_shape.pT.1000

trim_mt_mw_shape.pT.1500

Trimming performance improves slightly with p_T

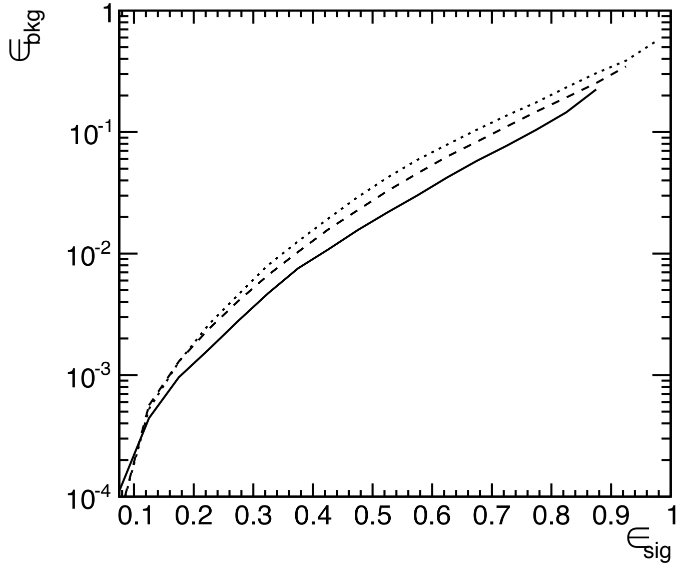
Pruning p_T dependence (R=0.8)





Pruning performance improves slightly with p_T

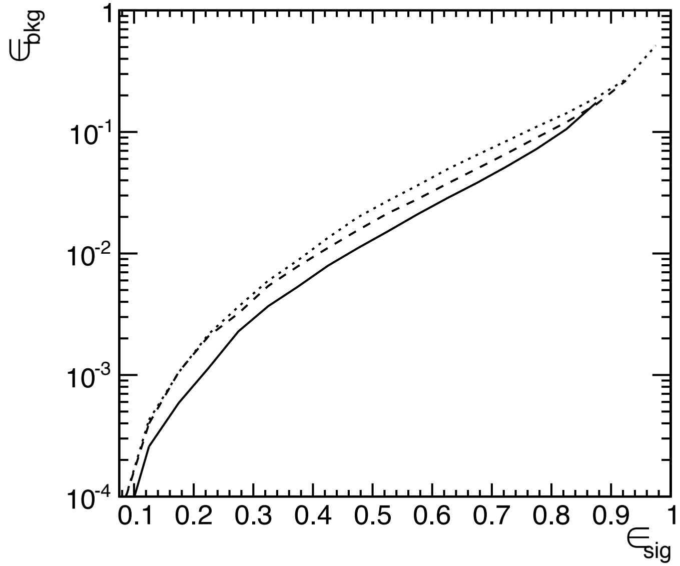
JH R dependence (p_T = 1.5 TeV)



JH.R.0.4

JH.R.0.8

JH.R.1.2

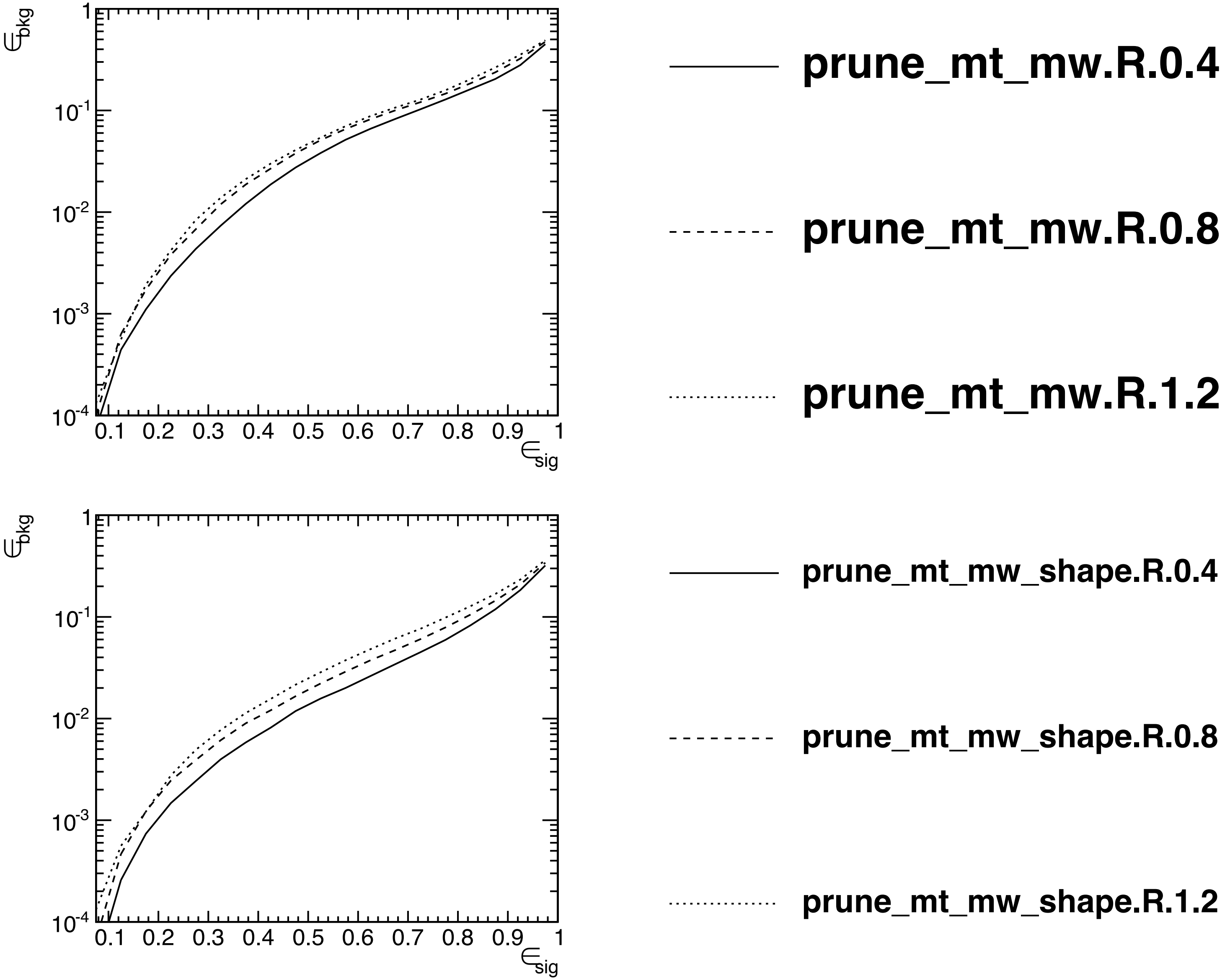


JH_shape.R.0.4

JH_shape.R.0.8

JH_shape.R.1.2

Prune R dependence (1.5 TeV)



Trim R dependence (1.5 TeV)