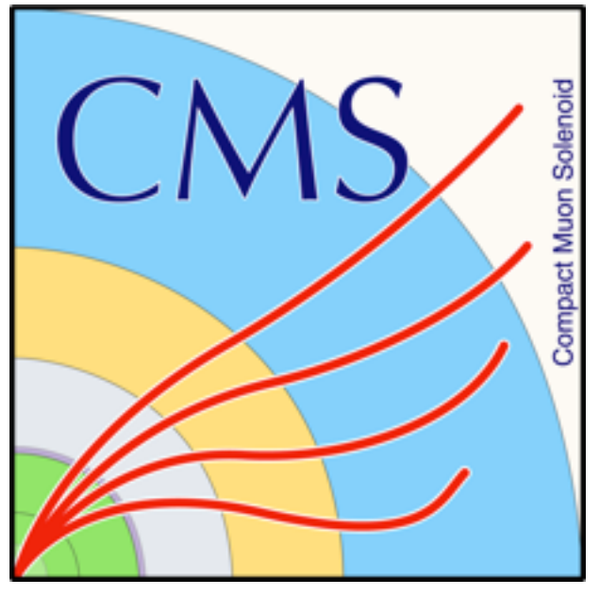
W/Z/Top tagging for Run 2 in CMS

Boost 2015, Chicago
Gregor Kasieczka for the CMS Collaboration

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich



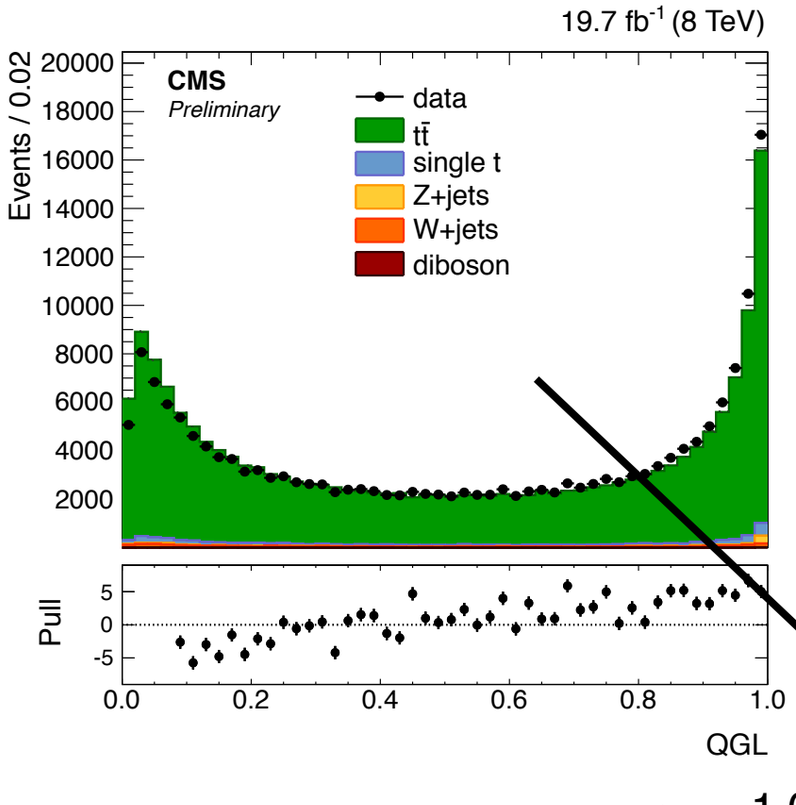
W/Z tagging: resolved

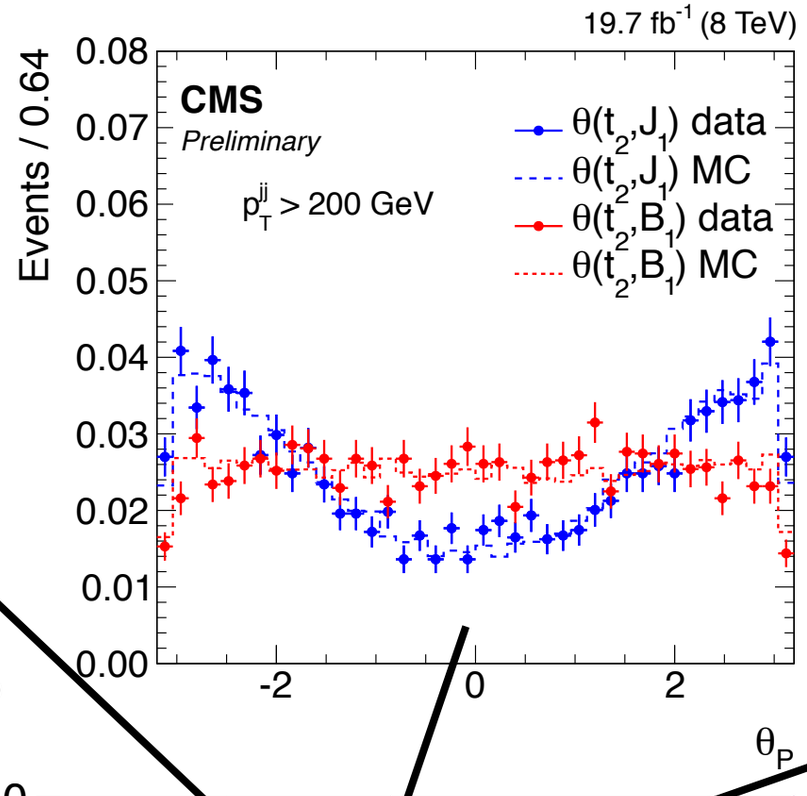
CMS PAS JME-14-002

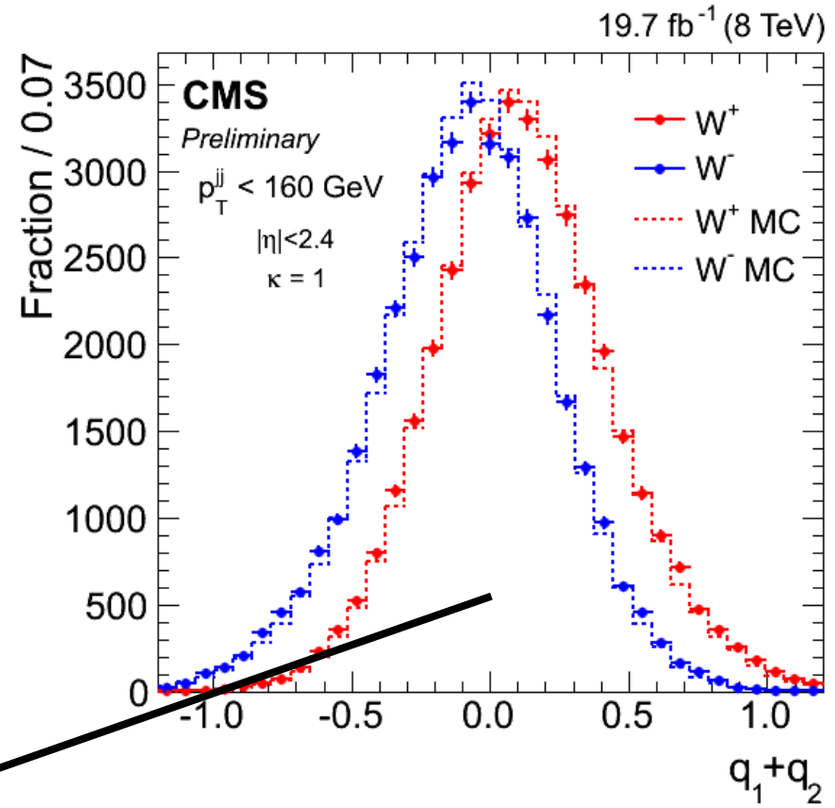
Basic

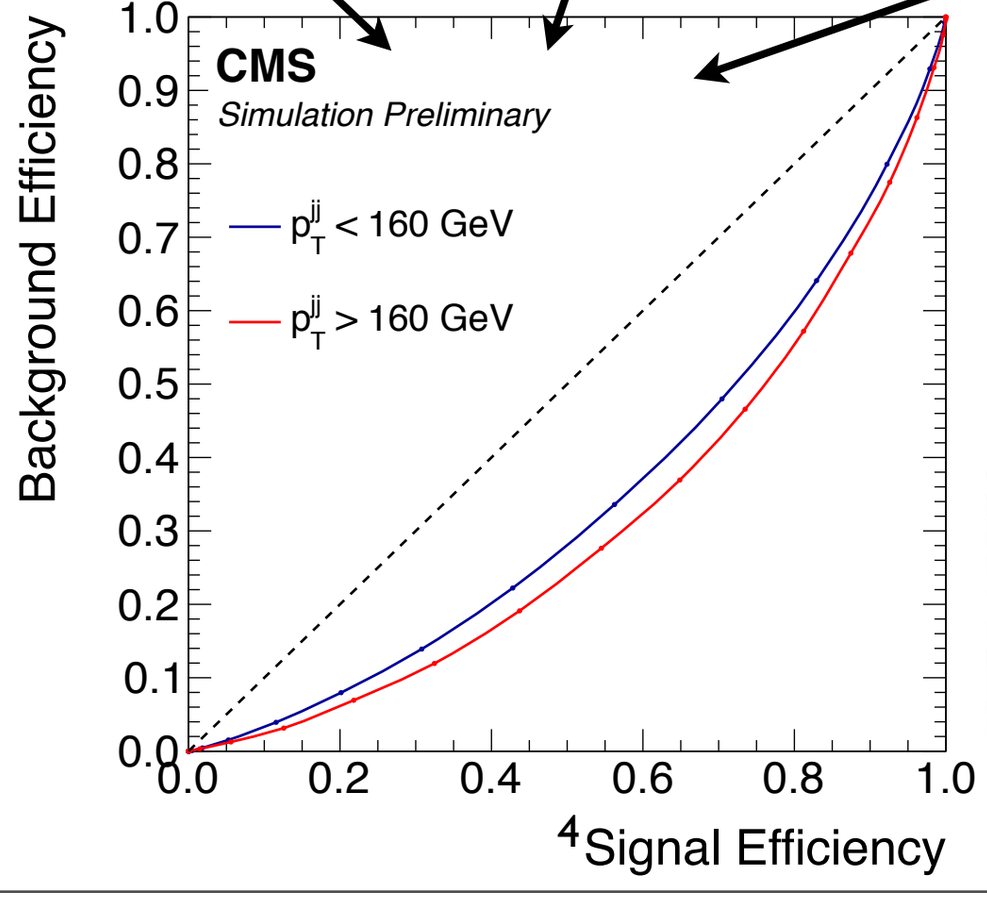
- Dataset: 19.7 fb-1 at 8 TeV
- Selection:
  - Single muon trigger
  - =1 isolated muon
  - >= 4 AK (R=0.5) jets, >=2 *b*-tagged
  - Consider pairs of un-tagged jets as W candidate if 40 < m < 130 GeV
- Observables:
  - Quark Gluon likelihood (QGL) [CMS-PAS-JME-13-002]
    - ptD: $\frac{\sqrt{\sum p_T^2}}{\sum p_T}$
    - PF candidate multiplicity in jet
    - σ_2: Minor axis in η-phi plane of PF candidates
  - Jet charge [CMS-PAS-JME-13-006] $q = \sum_i q_i \cdot \left(\frac{(p_T)_i}{(p_T)_{\text{jet}}} \right)^\kappa$
  - Jet Pull Angle [1001.5027] $\vec{t} = \sum_i \frac{(p_T)_i \, |r_i|}{(p_T)_{\text{jet}}} \vec{r}_i,$

Data vs MC









Good individual data/ MC agreement combine into MVA

W/Z tagging: boosted

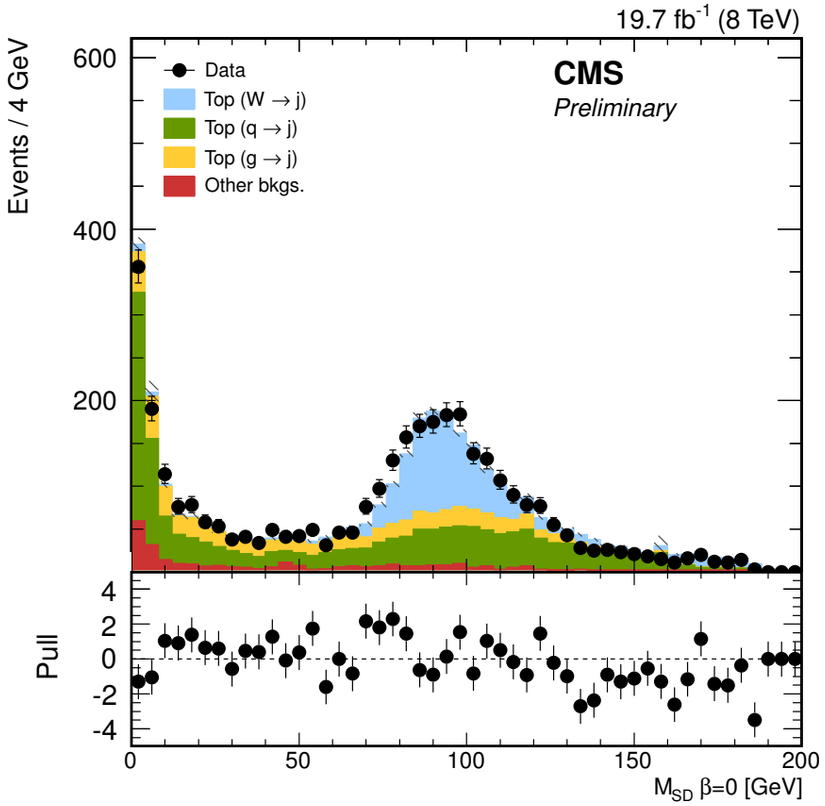
CMS PAS JME-14-002

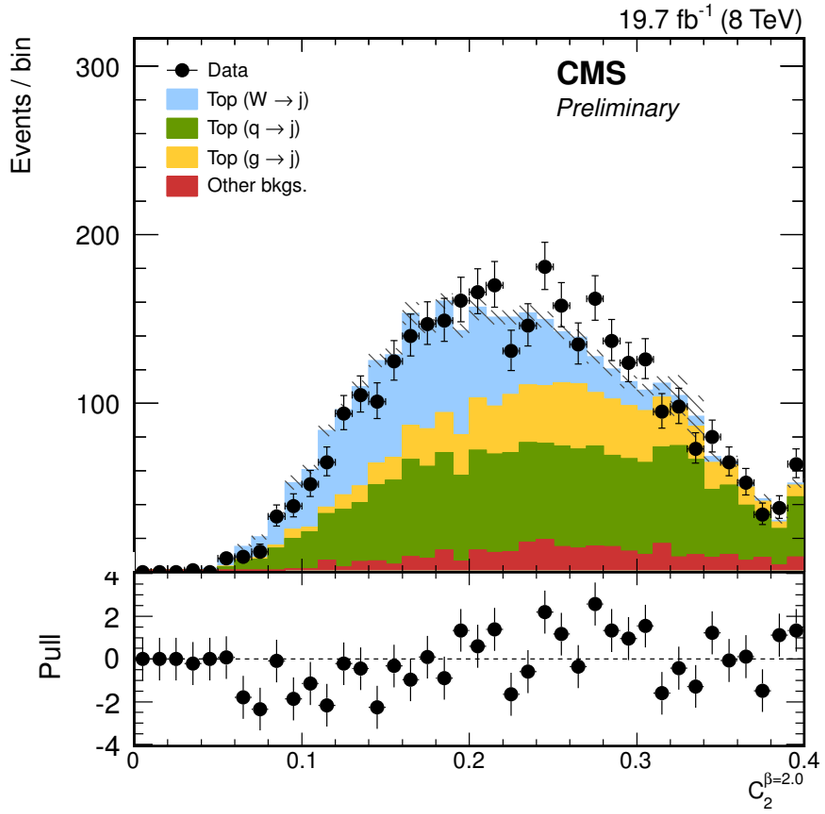
Basics

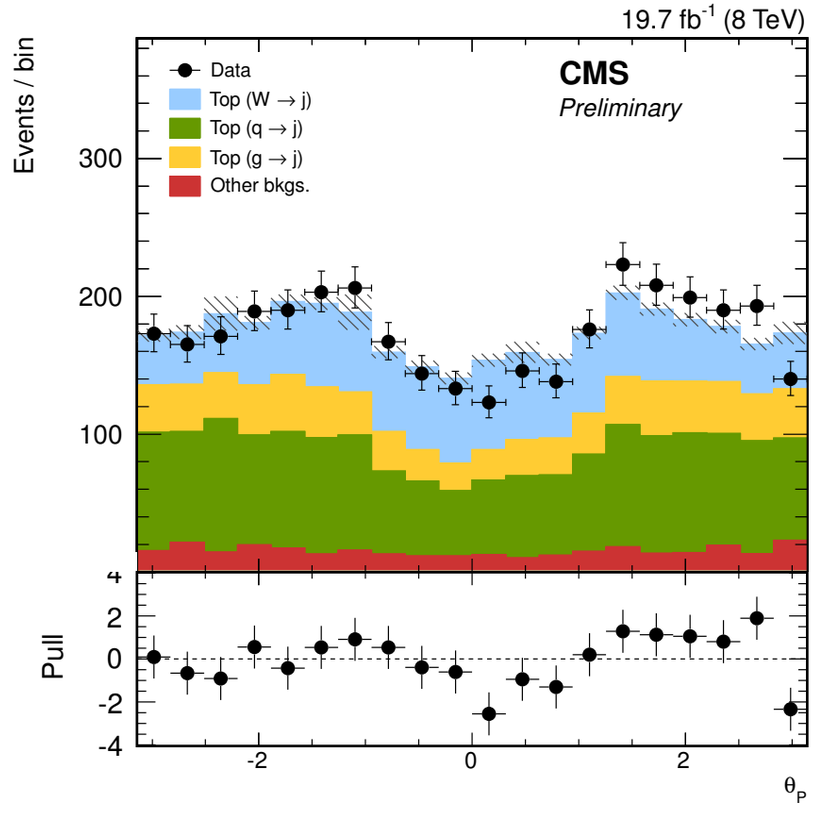
Algorithm	Parameters	Reference
Filtering	R=0.2, 3 jets	[0802.2470]
Trimming	R=0.1, f=3%	[0912.1342]
Pruning	z=10%, rf=0.5	[0912.0033]
Soft Drop	z=10%, beta={-1,0,1,2}	[1402.2657]
QGL	(see resolved)	[CMS-PAS-JME-13-002]
Subjet QGL	QGL for pruned subjets	-
Energy Correlation Functions	C_2, beta={0, 0.2, 0.5, 1, 2}	[1305.0007]
tau2/tau1	kT subjets	[1011.2268]
Q-jet volatility	alpha=0.1	[1201.1914]
Jet Pull Angle	(see resolved)	[1001.5027]
Jet Pull Magnitude	magnitude of pruned subjet pull-vector	-

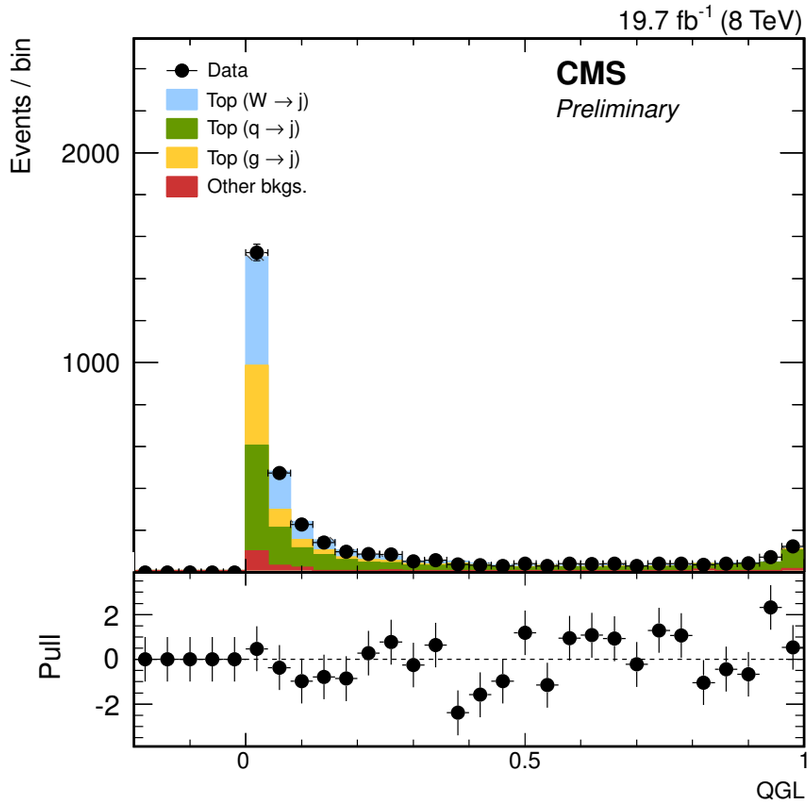
- Selection:
  - ttbar: single muon, two b-tagged AK5 jets, 1 AK8 (pT>250 GeV) jet (no overlap with b-jets)
  - Z+jets: 2 OS muons, with 65 < m 105 GeV

Control Distributions



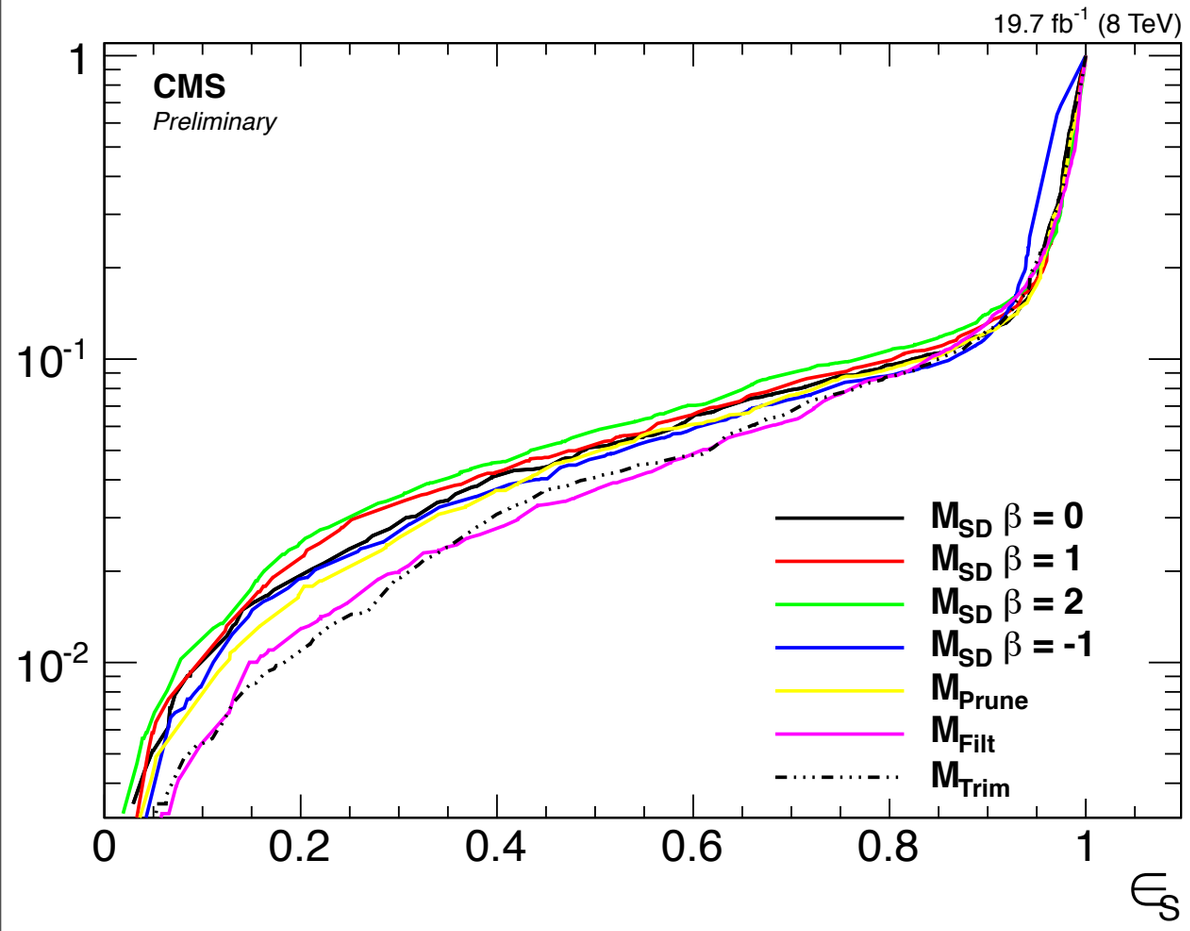


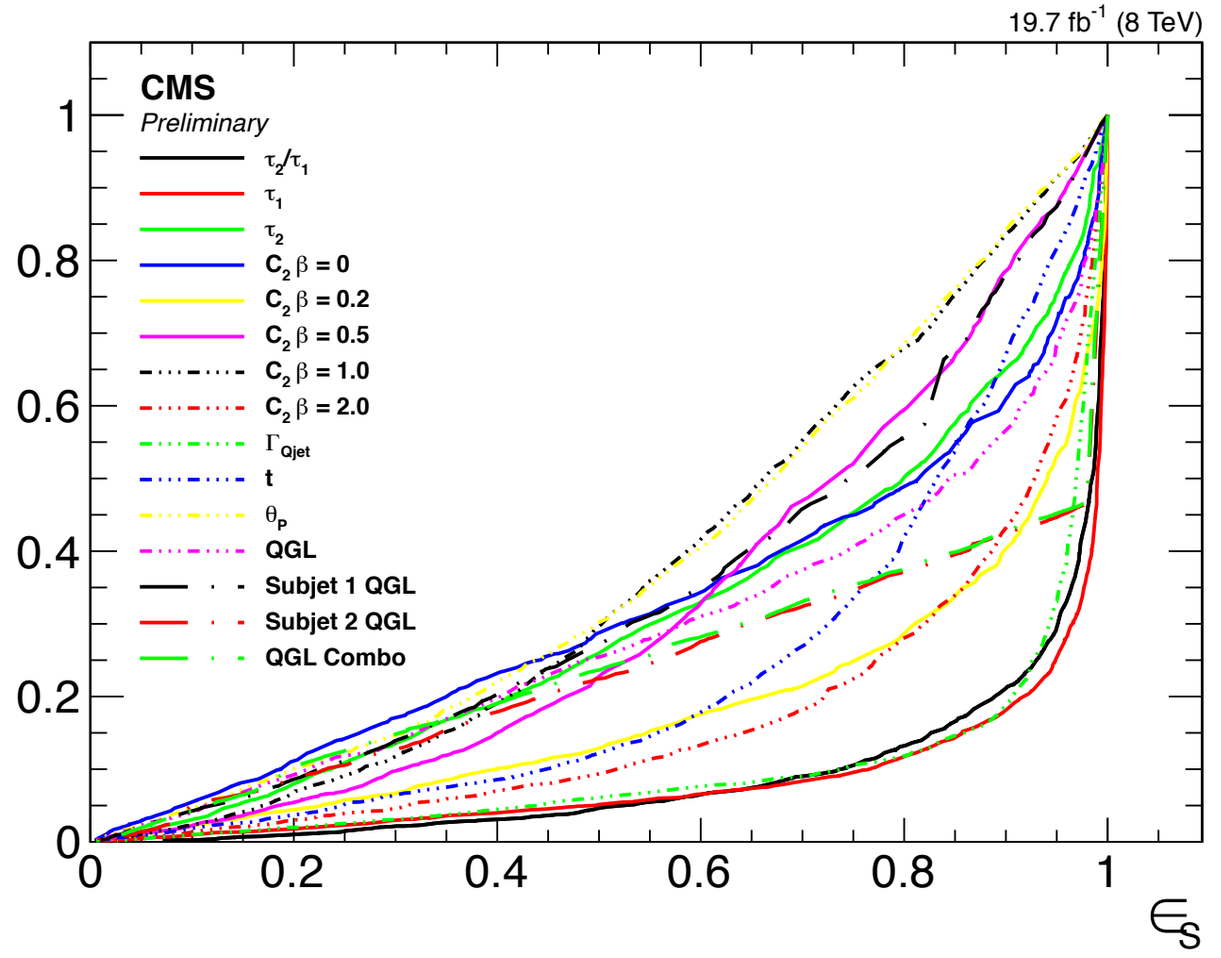




Data well modeled by simulation
(some discrepancy for ECF)
QGL for merged W more gluon than quark like

ROCs

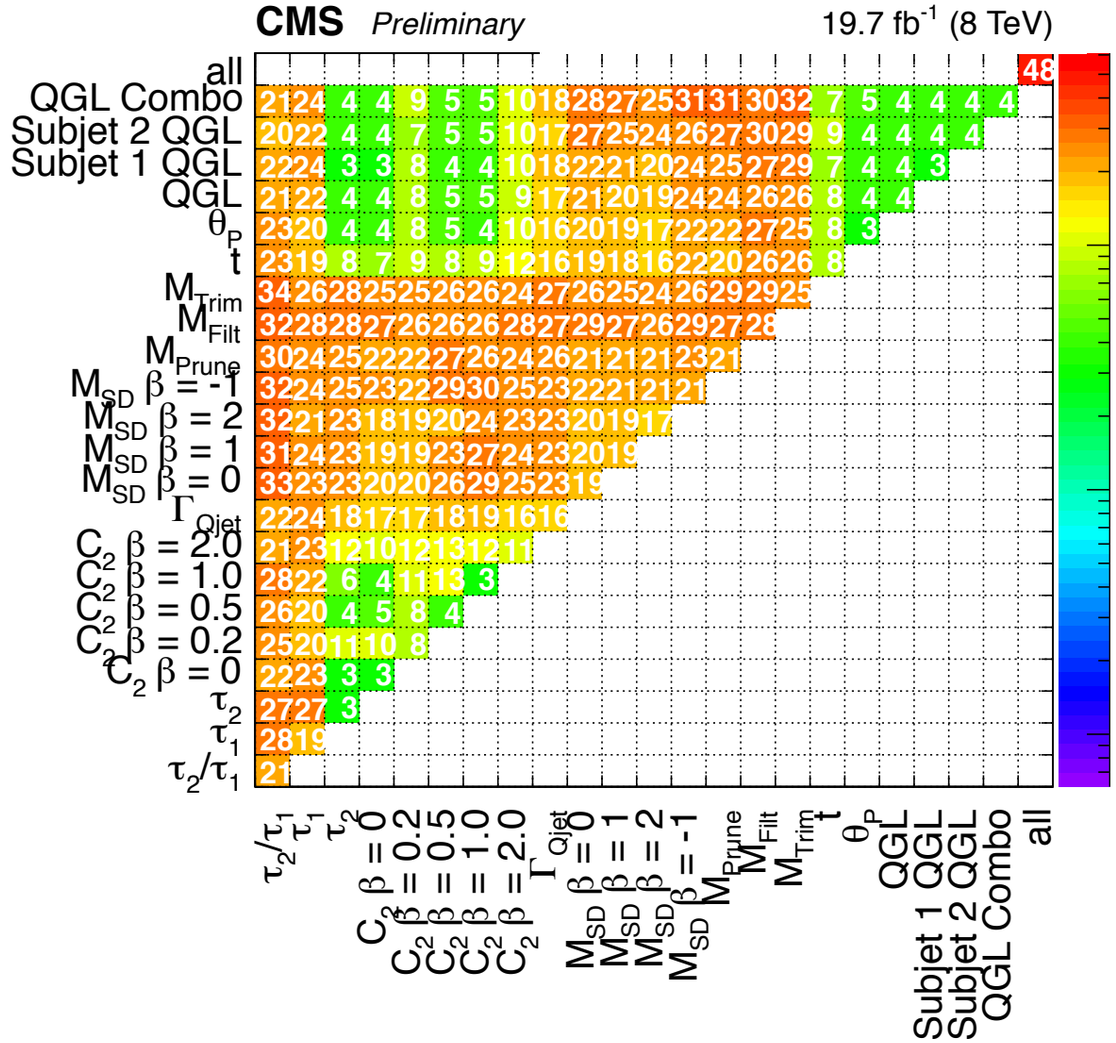




Single variable ROC curves:
Signal: MC
Background: Z+jets data

Trimmed and Filtered mass perform well
Shape variables tau2/tau1 and Q-vol do good

Pairs



Train BDTs
Z score: 1/background efficiency
for a signal efficiency of 50%

Signal: simulation
Background: Z+jets data

Rank	Z	Pairs of variables	
1.	38.5 ± 1.6	M_{Prune}	τ_2/τ_1
2.	37.9 ± 1.6	M_{Filt}	τ_2/τ_1
3.	37.8 ± 1.6	M_{Trim}	τ_2/τ_1
4.	37.7 ± 1.6	M_{Trim}	QGL Combo
5.	37.2 ± 1.6	M_{Prune}	QGL Combo
6.	36.7 ± 1.5	$M_{\text{SD}}\ \beta = -1$	τ_2/τ_1
7.	36.3 ± 1.5	$M_{\text{SD}}\ \beta = 0$	τ_2/τ_1
8.	35.8 ± 1.5	$M_{\text{SD}}\ \beta = 2$	τ_2/τ_1
9.	35.3 ± 1.4	$M_{\text{SD}}\ \beta = 1$	τ_2/τ_1
10.	35.0 ± 1.4	$M_{\text{SD}}\ \beta = -1$	QGL Combo

Run 2 V tagging baseline:
- AntiKt, R=0.8
-m pruned/softdrop (z=0.1, beta=0)
- N-subjettines tau2/tau1

Top Tagging

Overview

13 TeV MC

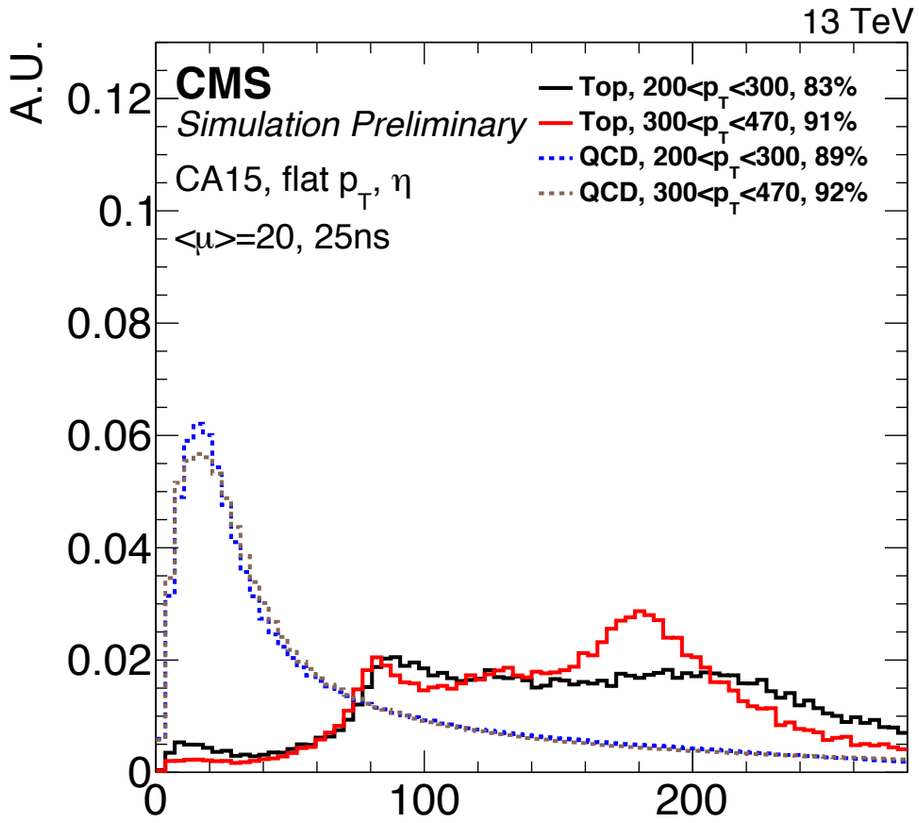
pT*	|eta| cut	QCD	Z'	Fatjet	Top size**	Match R
200-300	2.4	170-300	750	CA/1.5	0.8	1.2
300-470	2.4	300-470	1000	CA/1.5	0.8	1.2
470-600	2.1	470-600	2000	AK/0.8	0.6	0.6
600-800	2.1	600-800	2000	AK/0.8	0.6	0.6
800-1000	1.5	800-1000	2000	AK/0.8	0.6	0.6
1000-1400	1.3	1000-1400	3000	AK/0.8	0.6	0.6

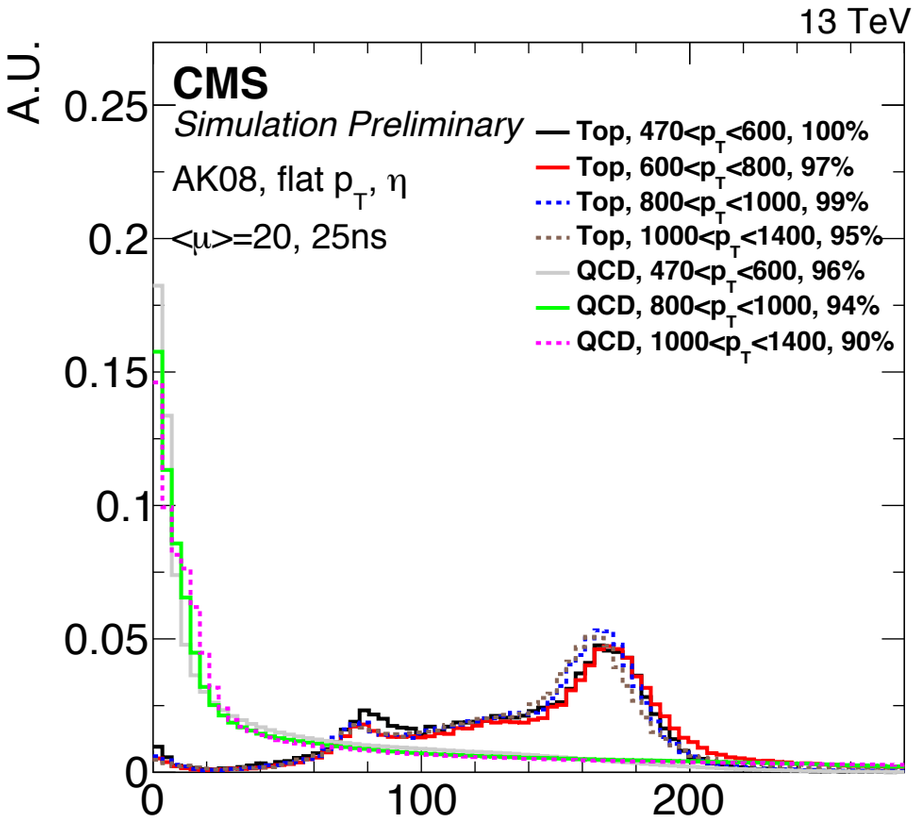
*pT: of matched parton from hard process
**Top size: maximal truth level deltaR of quark from decay and top-axis

Only reconstructed jets matched to partons from the hard process are considered.

- Groomed masses (Filtering, Trimming, Pruning, **Softdrop**)
- N-Subjettiness
- Q-jet volatility
- Unclustering-based taggers
  - CMS Top Tagger [CMS-PAS-JME-009-01]
  - HEPTopTagger V2 [1503.05921]
- Shower Deconstruction [1211.3140]

Softdrop Mass

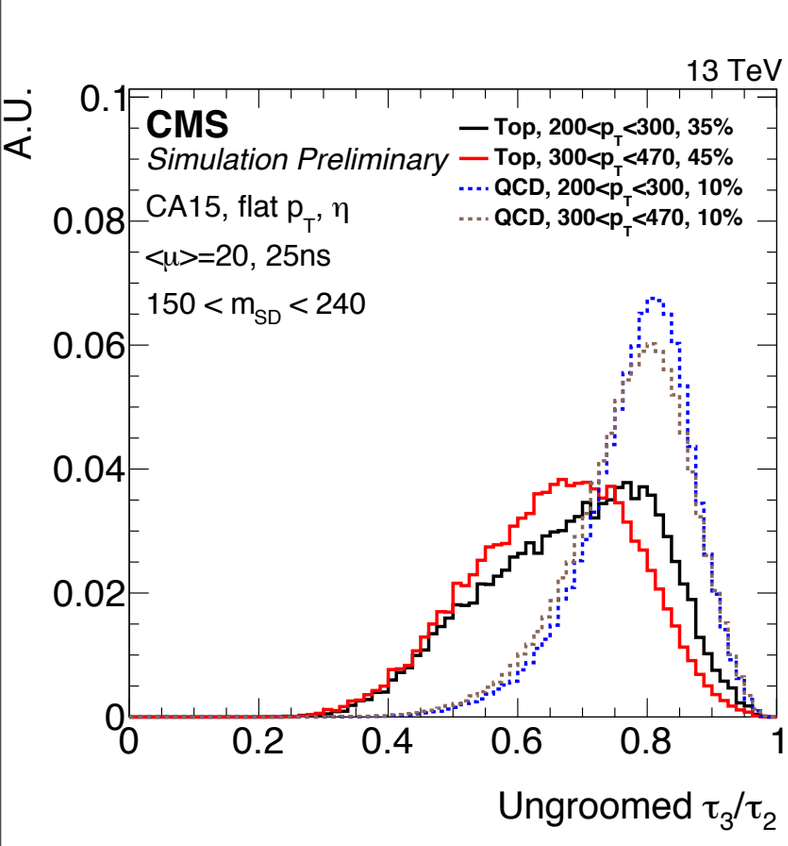


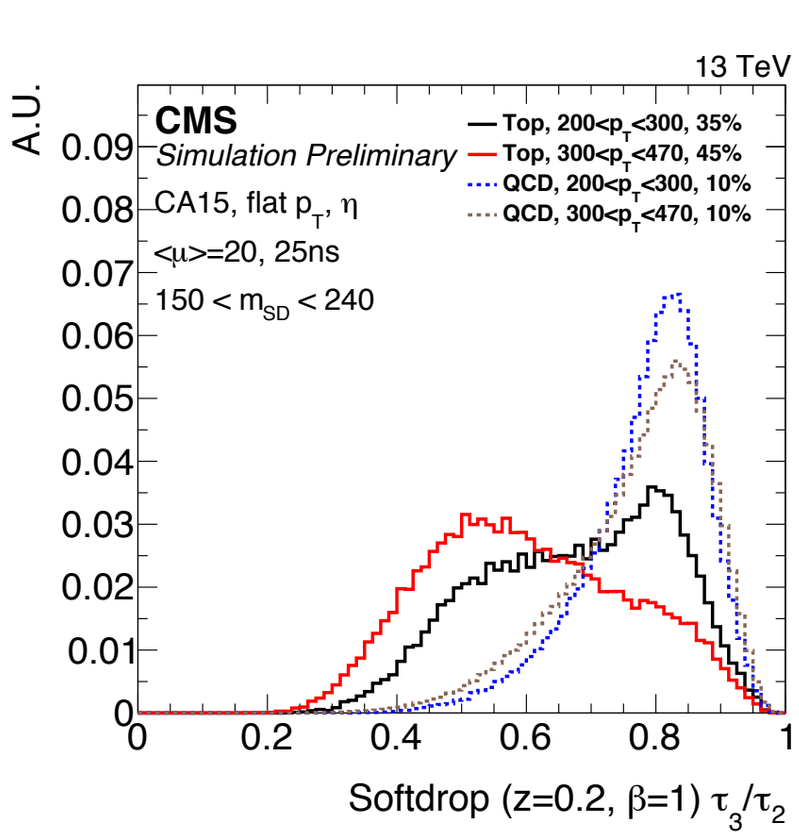


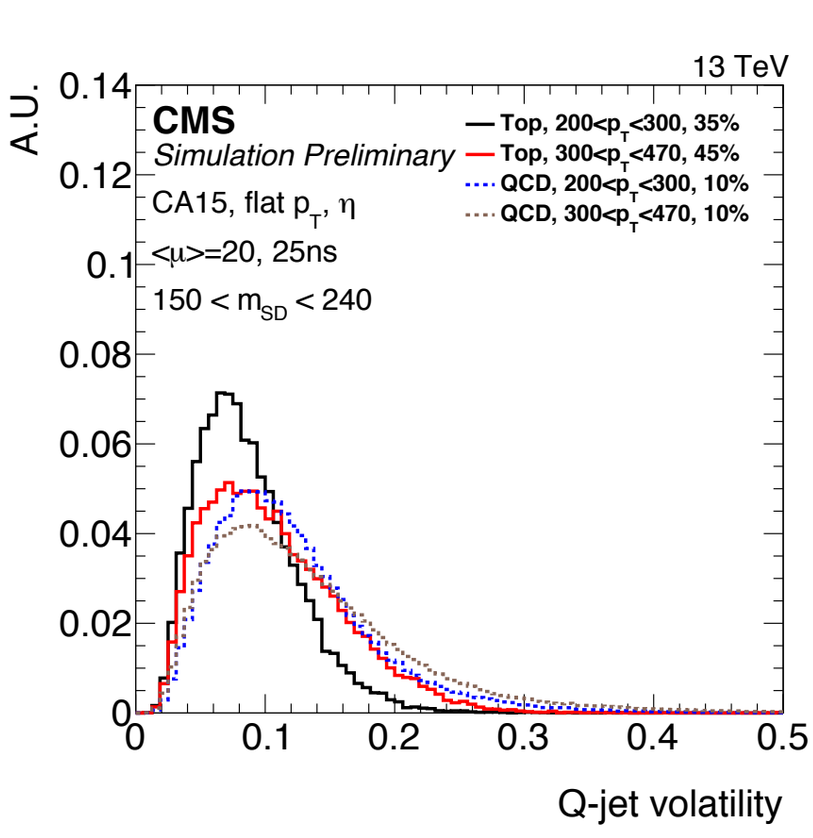
Using W-tagging default softdrop for high pT
At low pT a larger beta allows slightly higher total signal efficiency

Shape Variables

After mass-cut:
At low pT groomed n-subjettiness has clear advantage
Discrimination power of q-jet volatility strongly reduced

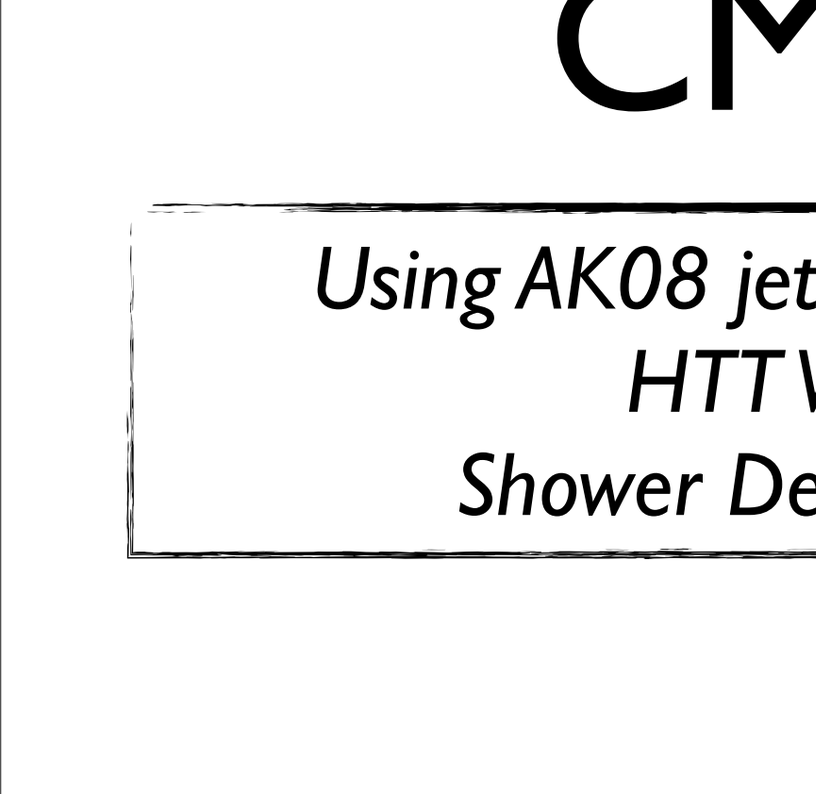


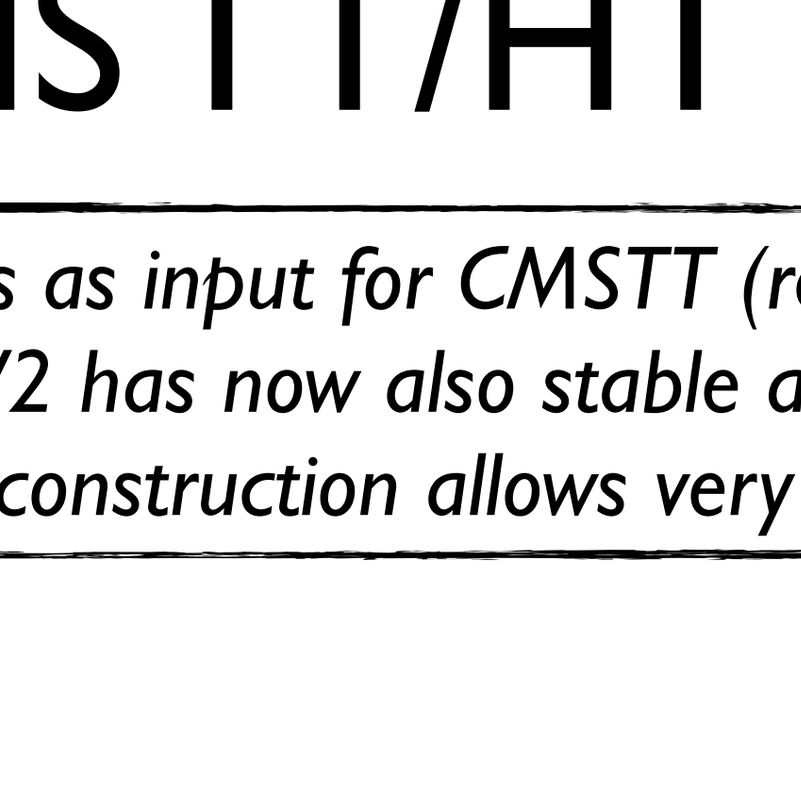


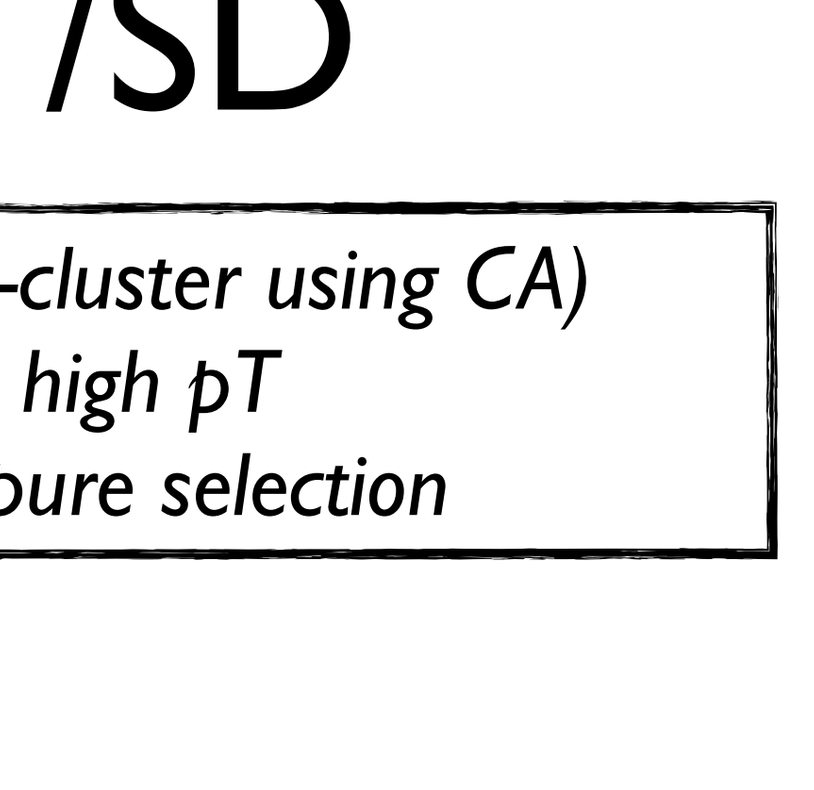


CMS TT/HTT/SD

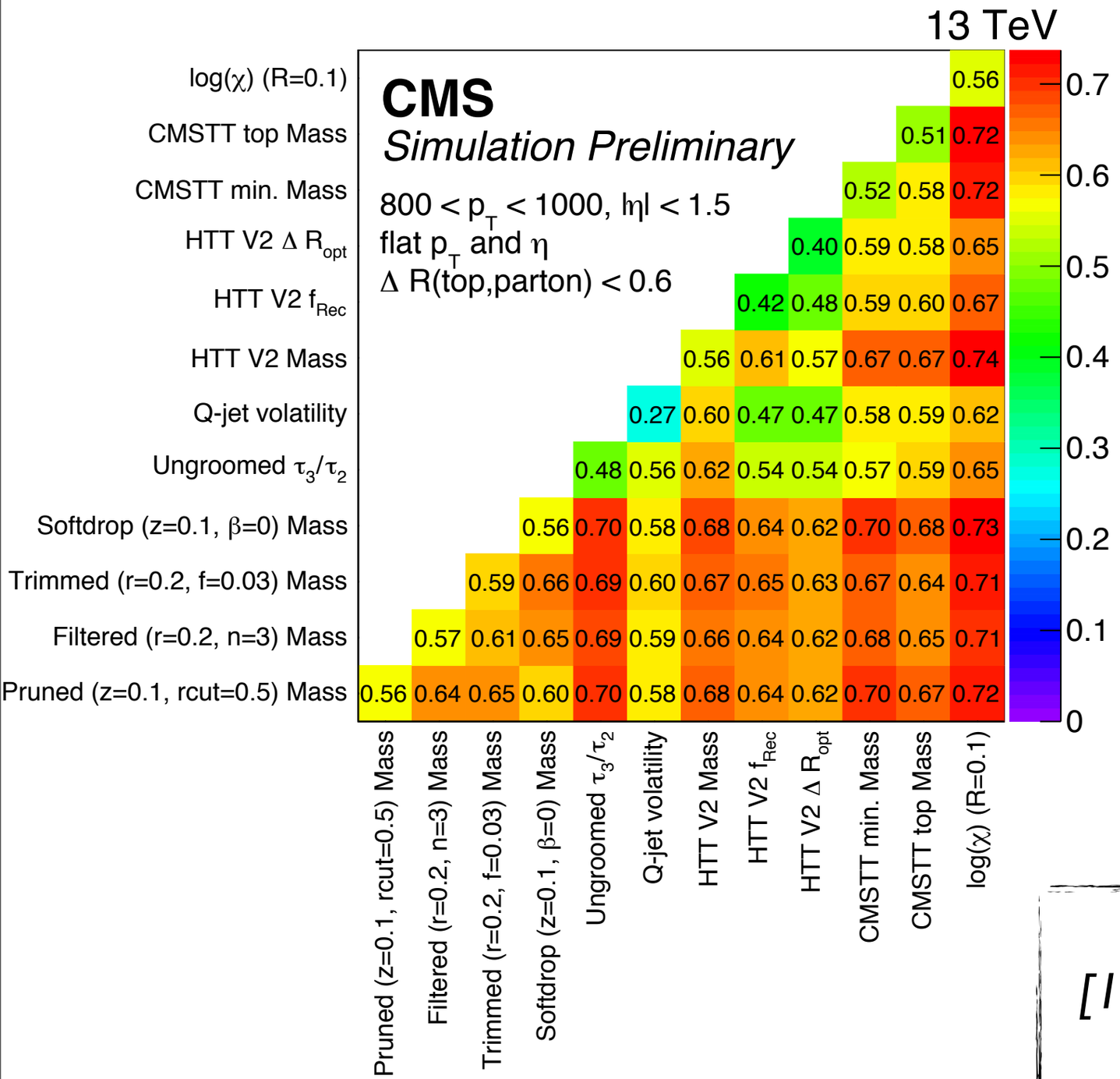
Using AK08 jets as input for CMSTT (re-cluster using CA)
HTTV2 has now also stable at high pT
Shower Deconstruction allows very pure selection

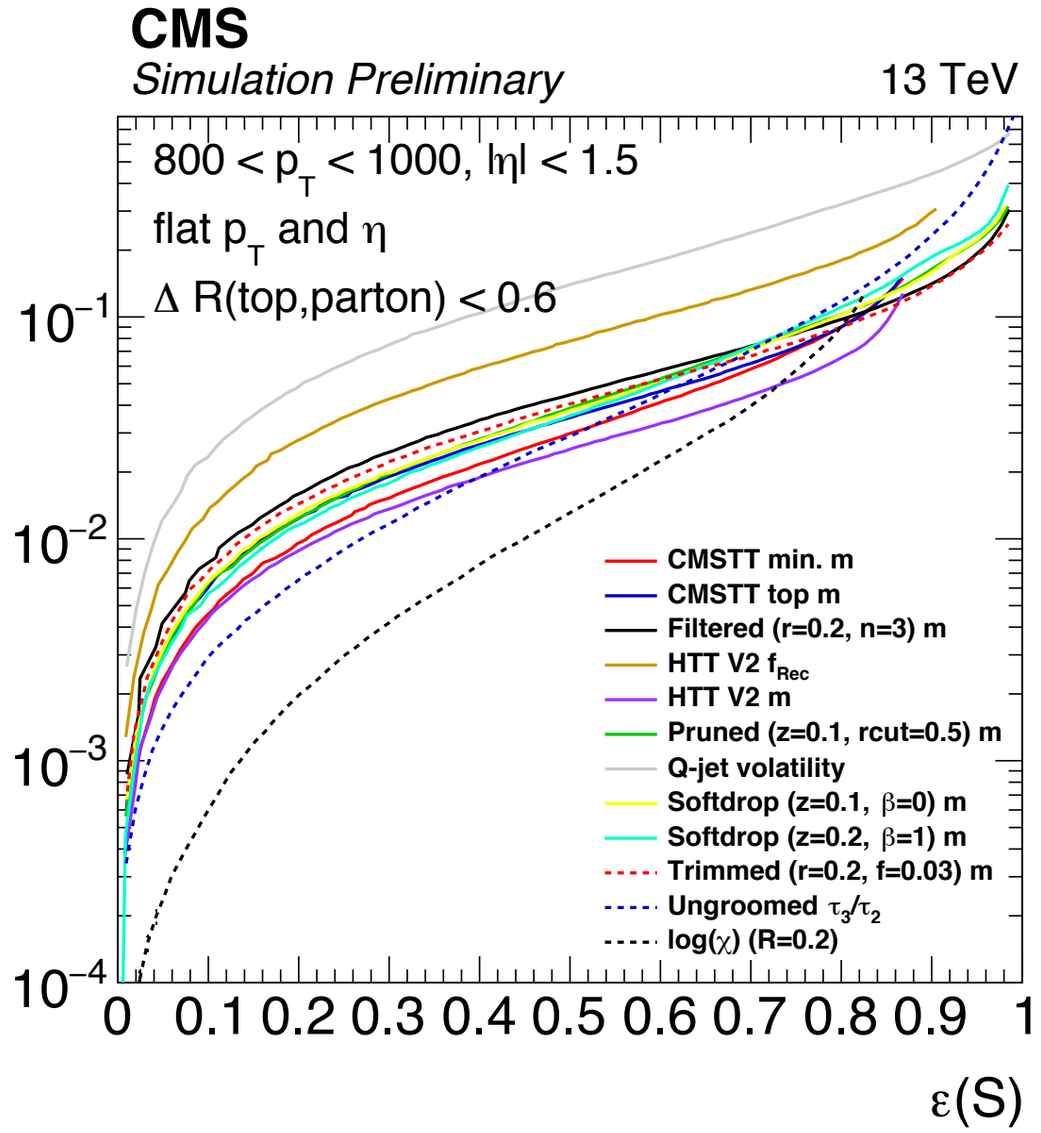






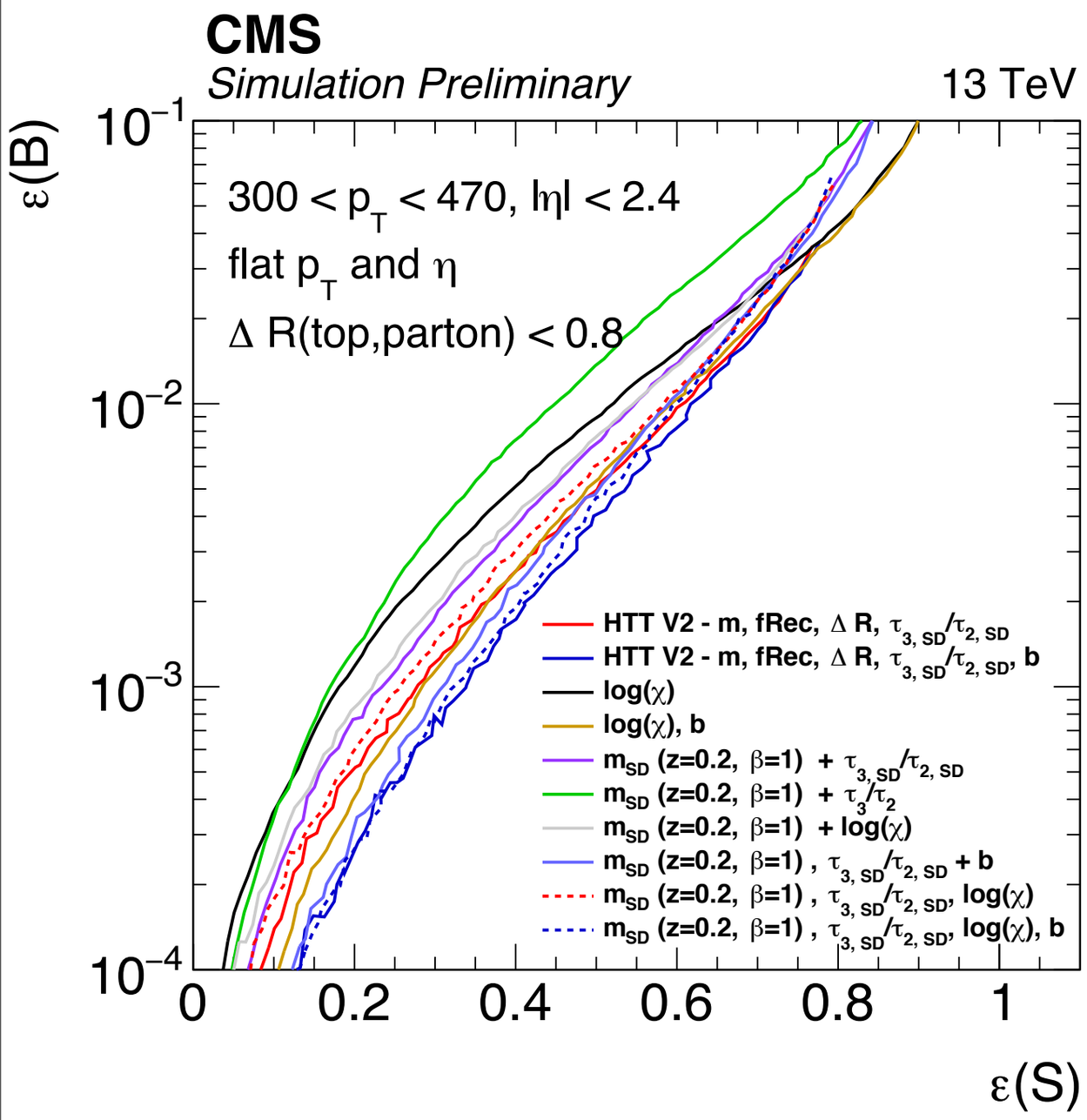
MI vs ROC

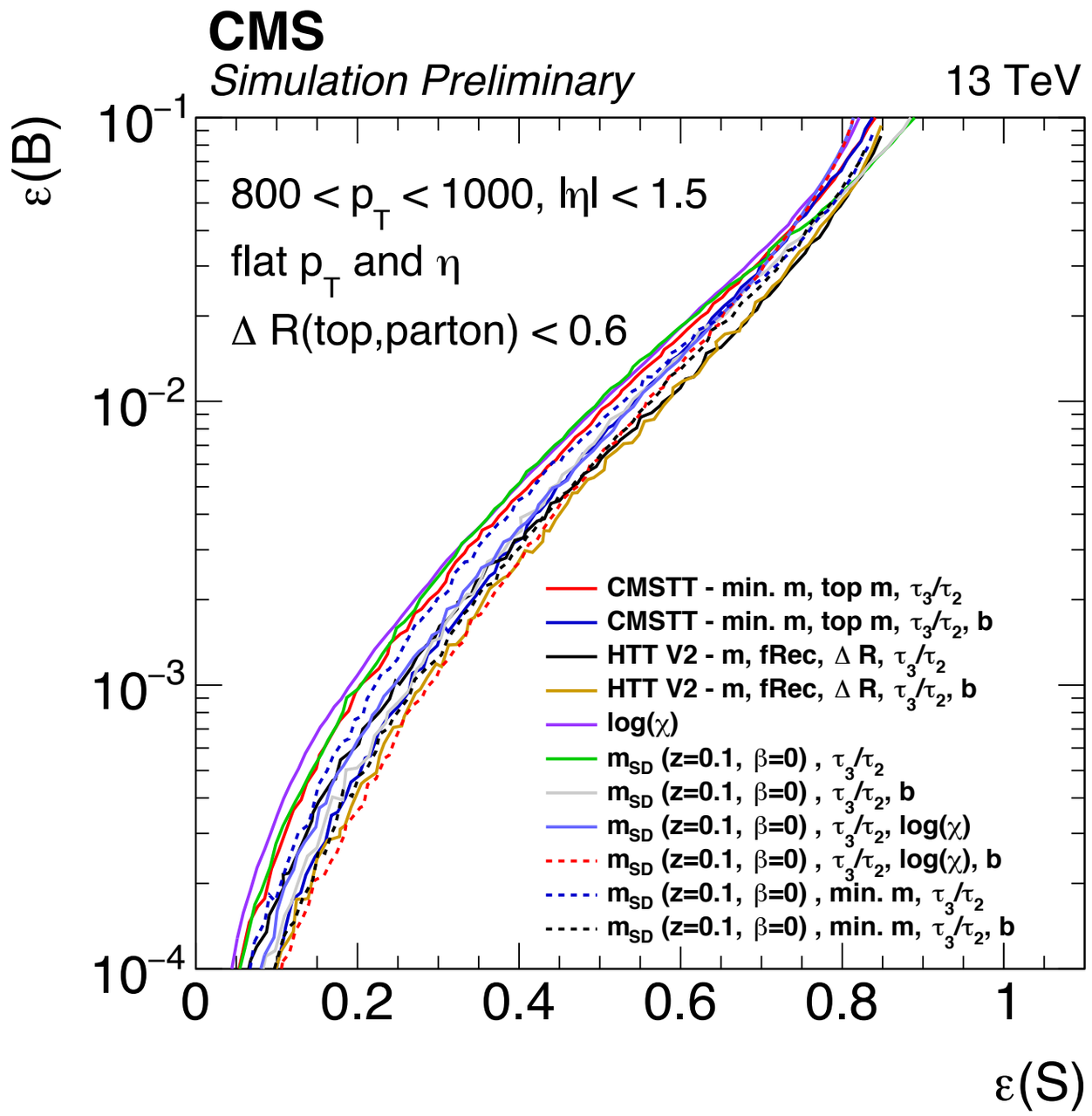




Comparison of Mutual Information [1408.3122] and single variable ROC curves
Shower Deco. single most discriminating variable
MI also influenced by maximal efficiency

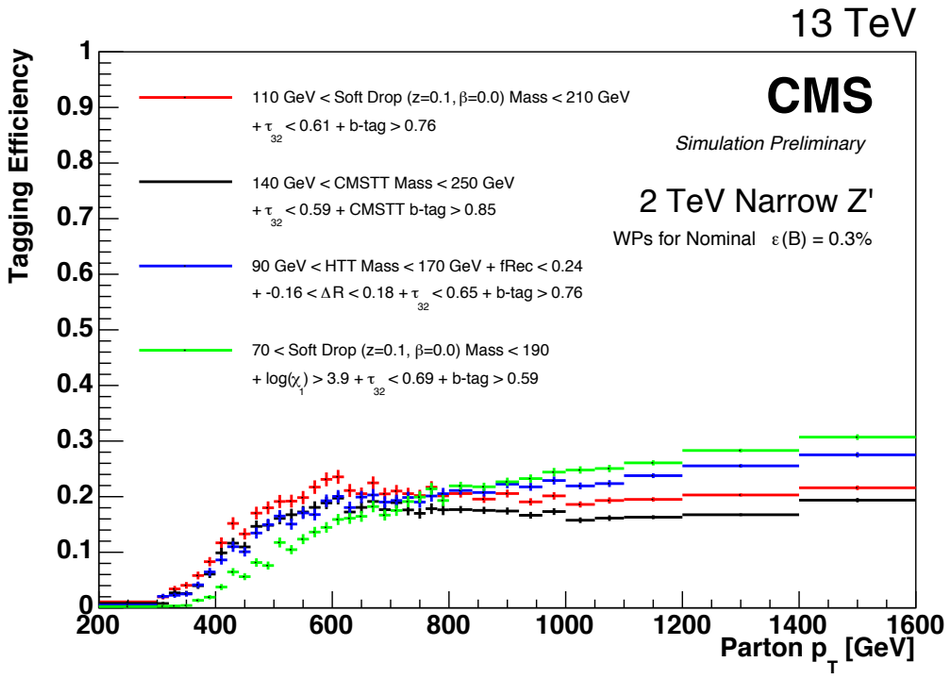
Performance in Simulation

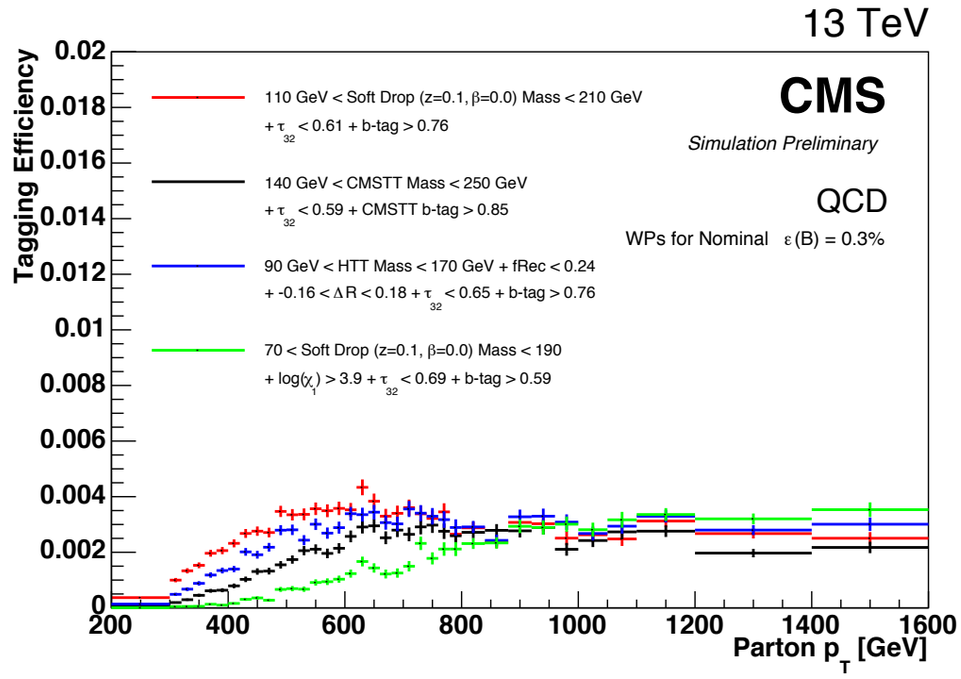


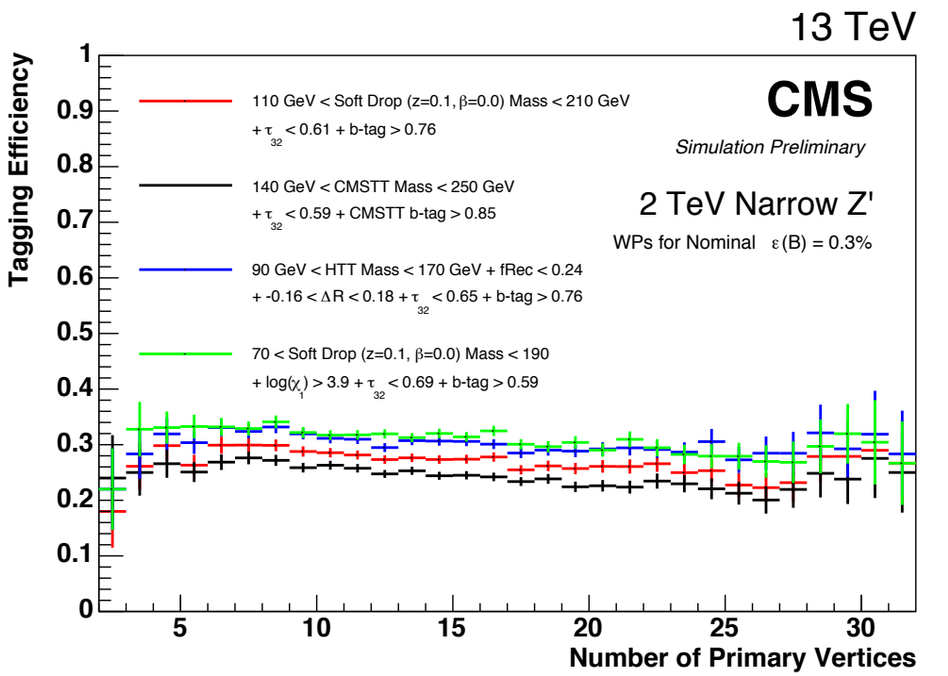


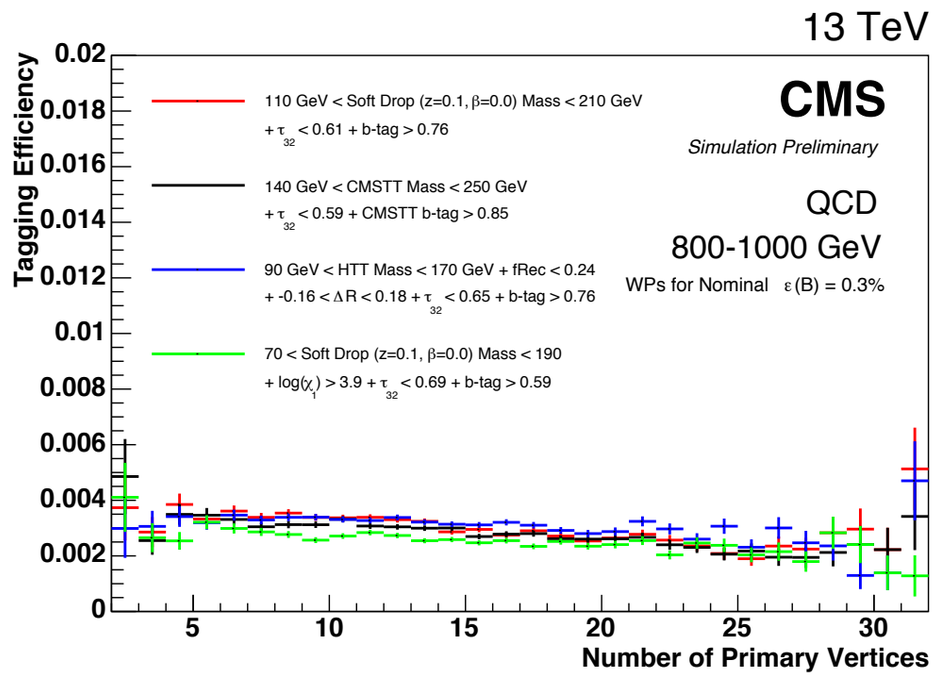
Combining variables (cut based) yields similar performance frontier for different sets of variables

Efficiency



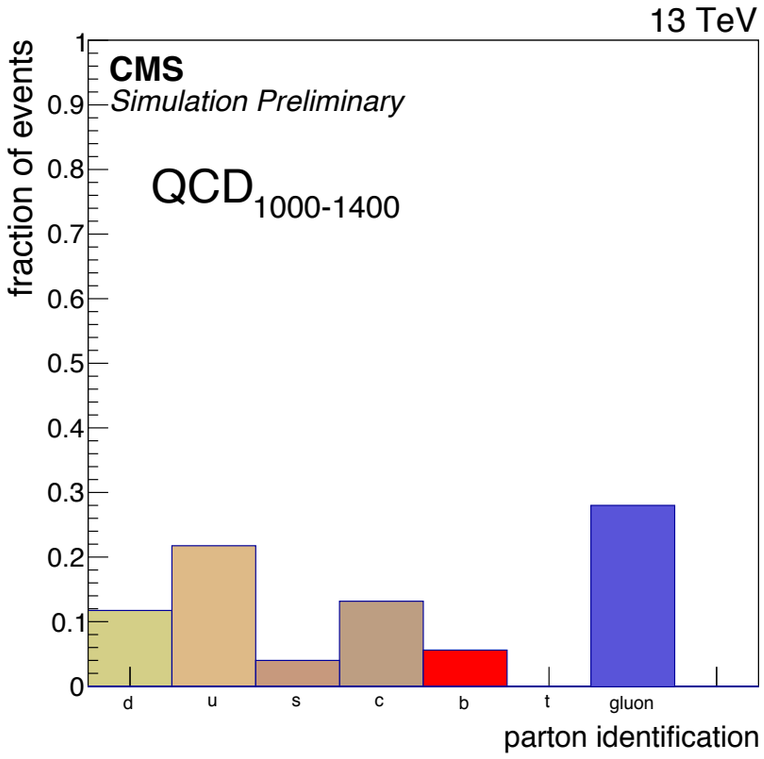


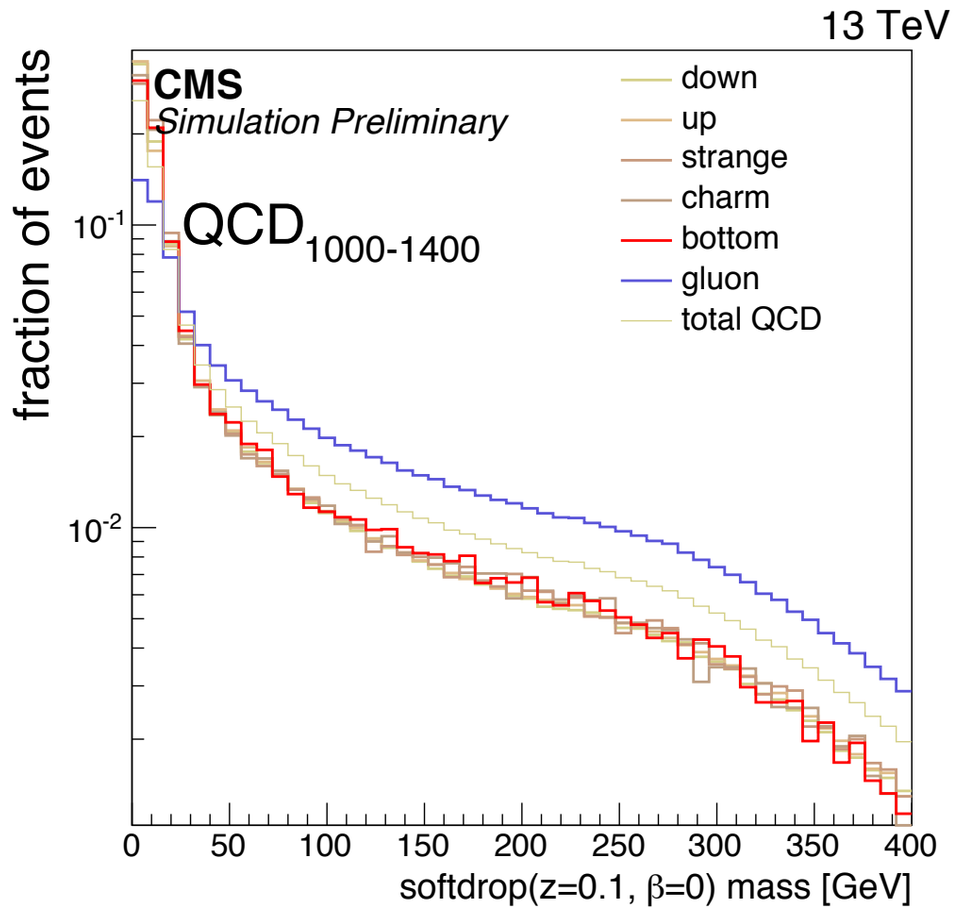


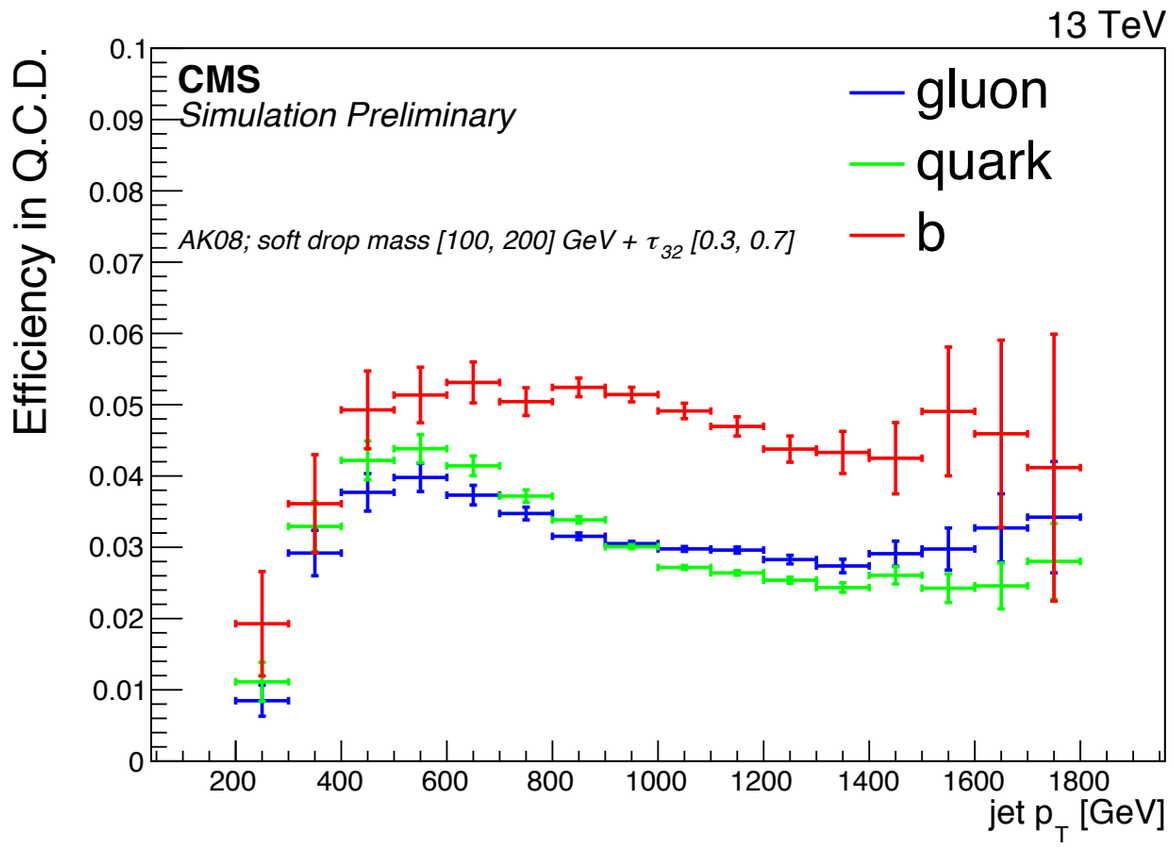


Combined taggers offer similar stability at high pT and against pile up

Background Source



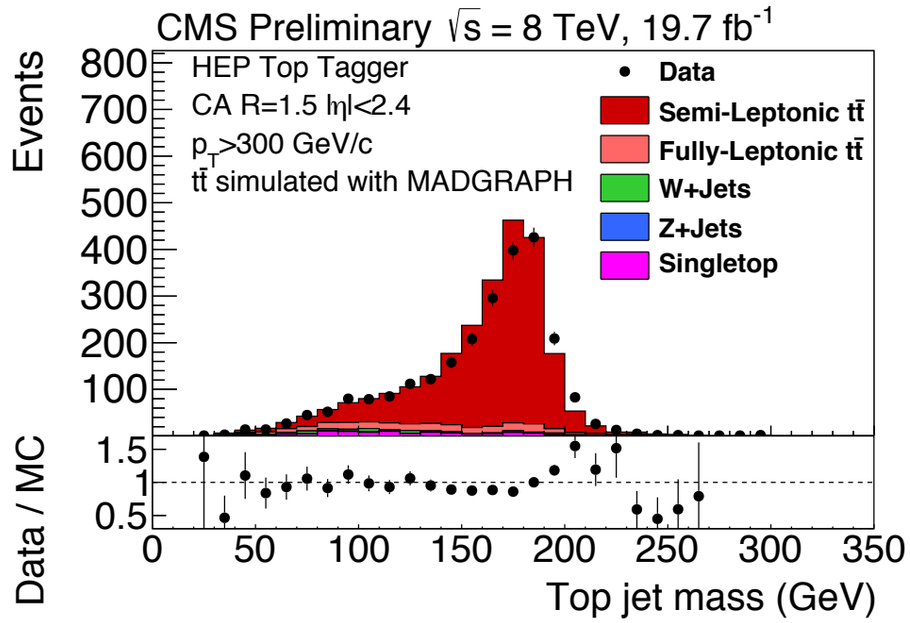


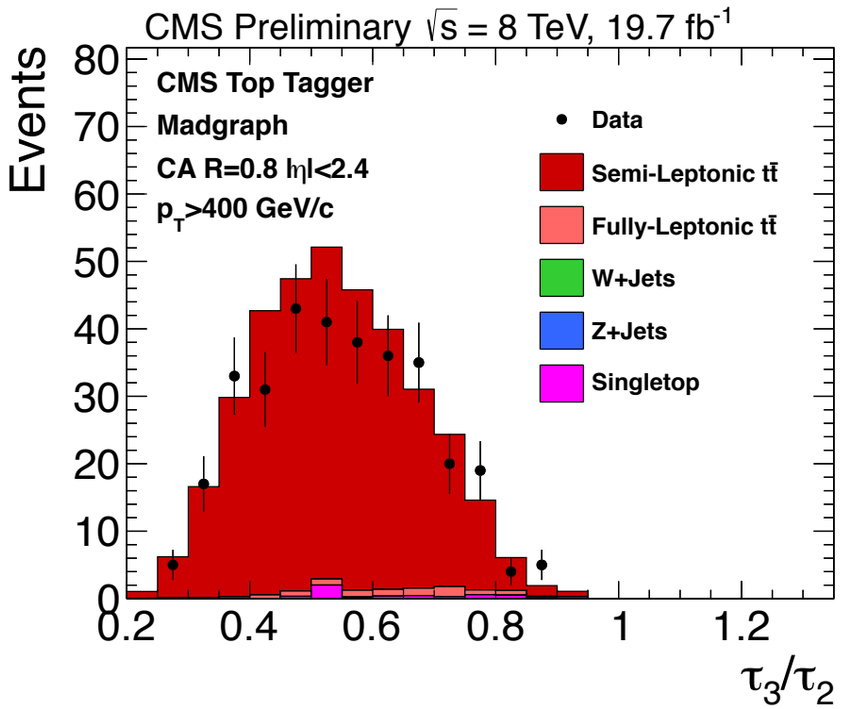


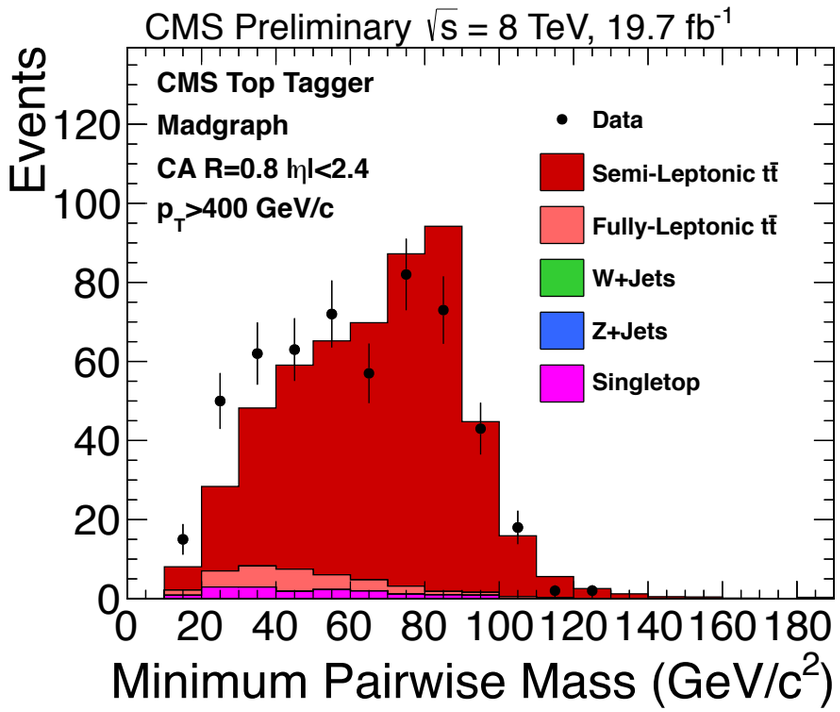
Investigating how mistag-rate depends on parton species. light quark and gluon are similar, higher for b-quarks

Data/MC

- Dataset: 19.7 fb-1 at 8 TeV
- Selection:
  - =1 isolated muon
  - >= 1 AK (R=0.5) jets, *b*-tagged
  - Consider the highest pT jet in the hemisphere opposite of the muon as top candidate
    - CMS TopTagger:
      C/A, R=0.8
      pT > 400 GeV
    - HEPTopTagger:
      C/A, R=1.5
      pT > 200 GeV







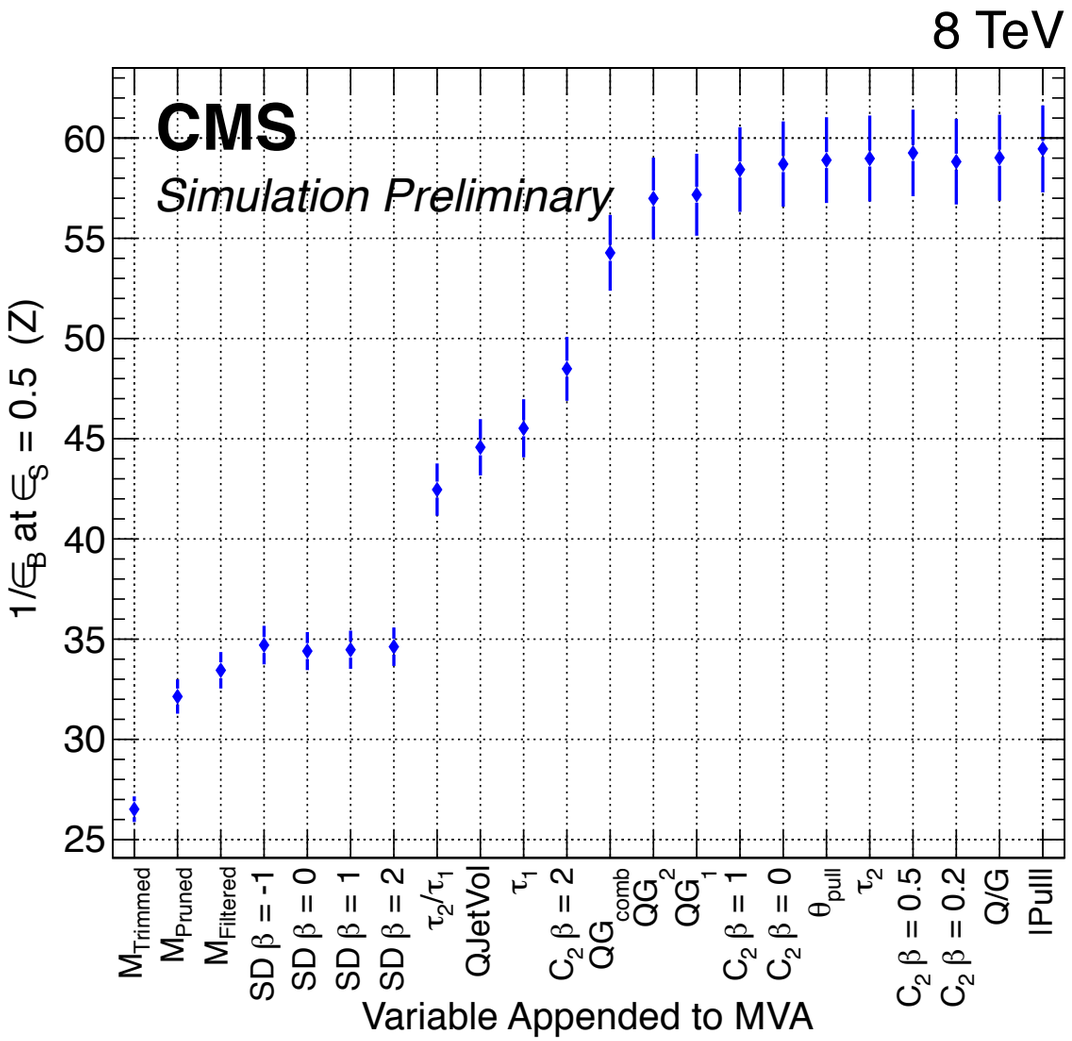
Conclusions

- Softdrop mass adapted both for W, Z and top-tagging
- Improve V tagging performance further by combining masses, shape variables and QGL information
- Different approaches lead to similar efficiency *ceiling* for top tagging - also studying correlations/mutual information and stability
- *Ready for 13 TeV and high pile-up!*

Thank You!

Backup

BDT



	Z-score
best pair	39
best tripe	43
full BDT	60