Hadronic top tagging using ASF

Lene Bryngemark, Torsten Åkesson

with warm gratitude to Andrew Larkoski for sharing his code

Boston Jets Workshop, January 22, 2014

LUND UNIVERSITY

Intro

1. I will show you some work I've done on top tagging using jet substructure
2. The jet substructure is the Angular Structure Function, ASF

This is work in progress! But since it's a workshop I'd like to share it with you at this point.

What is ASF?

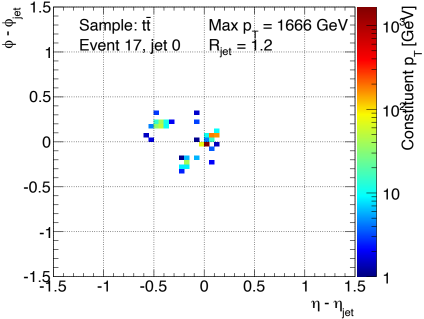
The Angular Structure Function is defined as [1]:

$$\Delta G(R) \equiv R \frac{\sum_{i \neq j} p_{Ti} p_{Tj} \Delta R_{ij}^2 \delta_{dR}(R - \Delta R_{ij})}{\sum_{i \neq j} p_{Ti} p_{Tj} \Delta R_{ij}^2 \Theta(R - \Delta R_{ij})} \quad (1)$$

where the i, j indicate jet constituents, and ΔR_{ij} is their relative distance. δ_{dR} is a narrow Gaussian approximating a δ function.

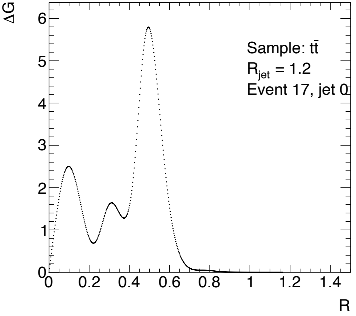
The ASF maps out **distances between the interesting features** of a jet.

jet



$\Leftrightarrow$

ASF



Idea/some concepts

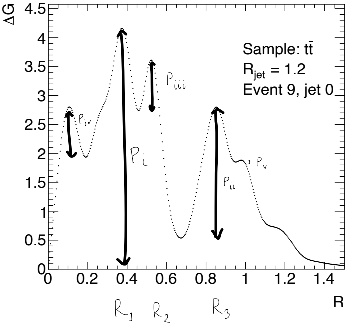
Most of what I'll show is based on the general behaviour of this plot:

The prominence of a peak is

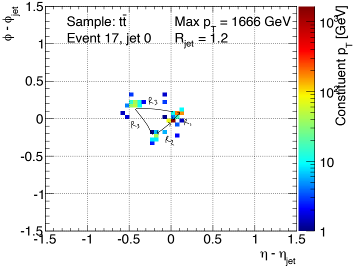
- the minimum descent from the top needed to reach a higher peak
- simply the height, in case of the highest peak

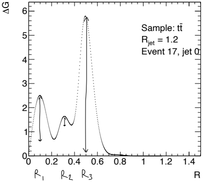
For my top tagging, this is done:

- For each jet, order the peaks by falling prominence: $P_i, P_{ii}, P_{iii}, \ldots$
- Order the 3 most prominent peaks in increasing R. "R_1" is the position of the first of these prominent peaks. It has prominence "P_1".



R_1 is the smallest prominent angular scale of the jet.

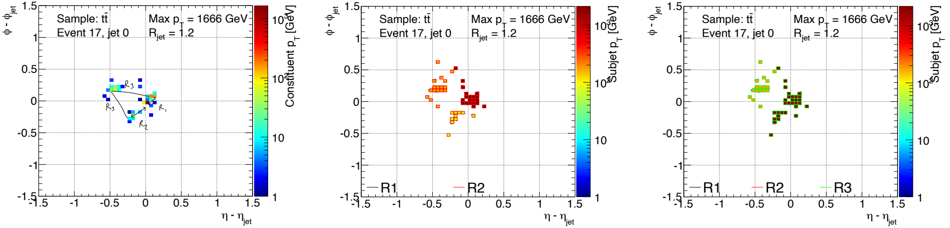




Now we have found the important distances in the jet – use this for reclustering!

Use k_t, $R_{kt} = R_i/2 + 0.1$.

Opening angles are boost dependent – here measured jet by jet.



Original jet... ... reclustered with R_2... ... and R_3

Reclustering $\rightarrow$ subjets.
Subjet four-vectors $\rightarrow$ subjet masses,
combined to form di-subjets, etc.

There is no requirement on the number of subjets, as this is boost dependent.

For R_2, $M_{02} = 175.3$ GeV, for R_3, $m_0 = 175.3$ GeV

Evaluating the usefulness for top tagging

Input: Pythia8 simulated events producing

- $t\bar{t}$ or
- light quarks and gluons (q, g)
    - qq
    - gg
    - qg

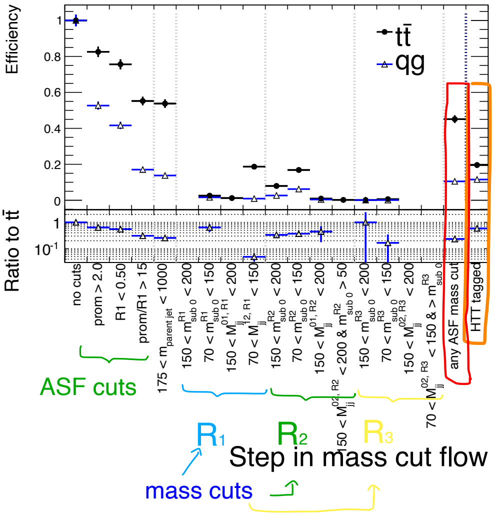
$\sim$flat $\hat{p}_{\mathrm{T}}$ in 1 - 2.5 TeV range

Take out the two leading Anti-k_{t}, $R = 1.2$ jets

Performance: cuts on (di-)subjet masses

Cut flow, comparing $t\bar{t}$ and q, g efficiency. Not optimised!

Pythia 8, generated with $1000 < \hat{p}_T \lesssim 2500$ GeV.



Last two bins:

- OR of my mass cuts,
- HTT [2] tagging efficiency.

  (ATLAS default settings)

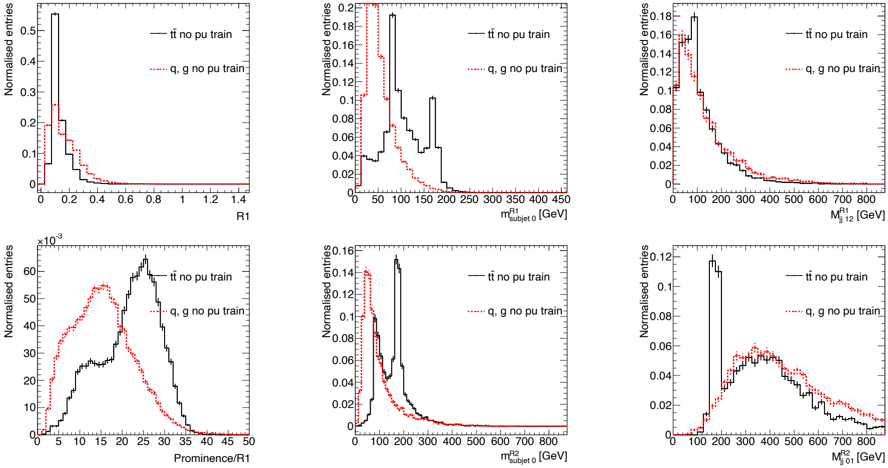
Promising.

Many variables, some correlations – what would happen if we used a multivariate technique?

BDT: input

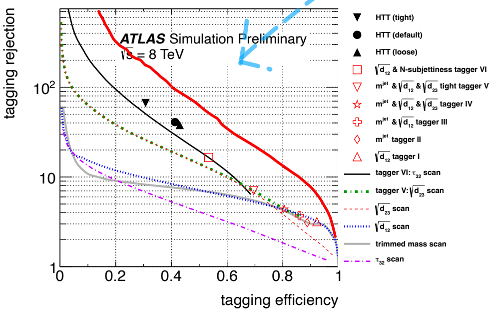
Boosted Decision Tree, from the TMVA package [3]. Very little tweaking.

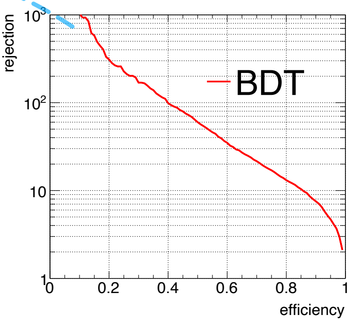
Some example training distributions: compare $t\bar{t}$ and q, g:



Performance: ROC curve

ATLAS-CONF-2013-084





ATLAS efficiency/rejection plot

jet p_T from **500 to 1000** GeV

Full ATLAS machinery

My efficiency/rejection plot

jet p_T from **1500 to 2500** GeV

Pythia truth

BUT:

this is too good to be true

– my efficiency/rejection plot was done on truth, without pile-up.

Technical reasons – takes more computational time to produce a large enough pile-up training sample than I had before this talk.

A detector simulation is hard to do yourself.

How would pile-up might affect mass? Angular scales?

– can do a simple pile-up overlay!

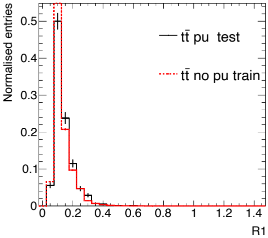
(the cut flow is done with pile-up...)

Including pile-up

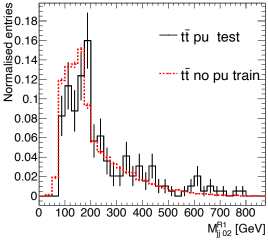
Deal with pile-up: trim the jets before the ASF is calculated.

I have run smaller files with pile-up overlay (Pythia min-bias, Poisson $\langle\mu\rangle = 20$) and **trimming** ($k_t, R_{kt} = 0.3$, p_{T} ratio 0.03).

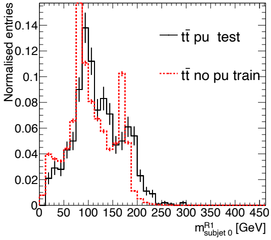
Some variables are stable; compare **test** (pile-up) and **train** (no pile-up):



R_1



$M_{02}^{R_1}$ – a di-subjet mass

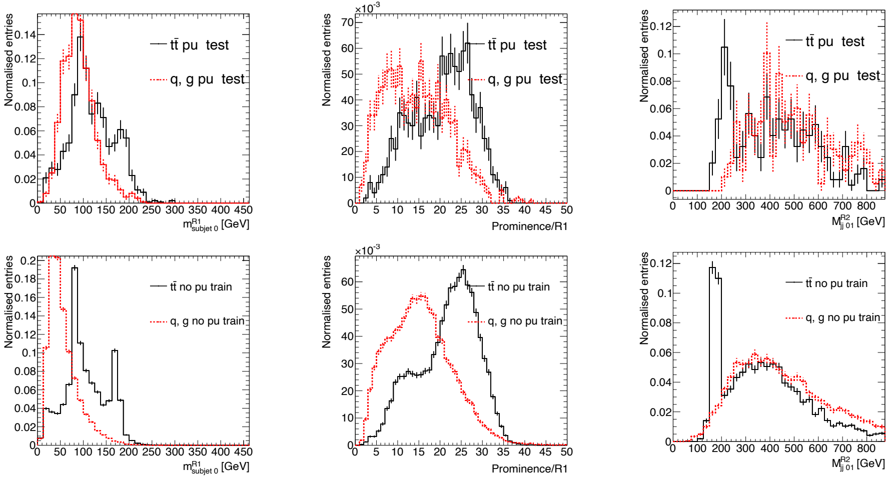


$m_0^{R_1}$ – leading R_1 subjet mass

Pile-up shifts, that's right, single jet mass.

What is the expected performance in pile-up?

Does pile-up destroy the signal/background separation?



Mass shifts ⇒ trained cuts less valid. (But this could be mended...)

Performance in pile-up?

A preliminary test, for this workshop:

- Use the BDT trained on a large sample without pile-up
- test it on a (small) sample with pile-up.

Clearly not optimal! Poor person's (temporary) solution.

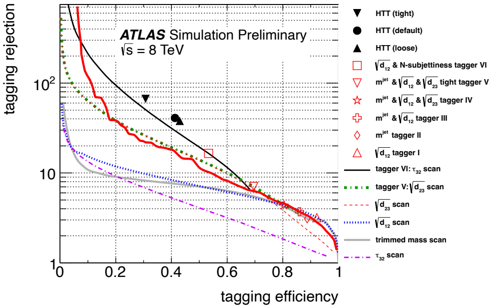
Performance in pile-up?

A preliminary test, for this workshop:

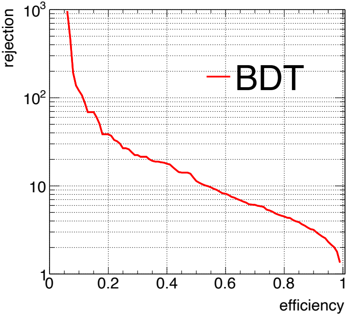
- Use the BDT trained on a large sample without pile-up
- test it on a (small) sample with pile-up.

Clearly not optimal! Poor person's (temporary) solution.

ATLAS-CONF-2013-084



ATLAS efficiency/rejection plot
jet p_T from **500 to 1000** GeV



My efficiency/rejection plot
jet p_T from **1500 to 2500** GeV

Further considerations

There are a few things to think about, for example:

- Calorimeter granularity – lower limit on R_i
- Pile-up correction! A jet area based $\rho \cdot A^{\mu}$ mass correction
- Training and testing on events with pile-up
- p_{T} range? How does it compare to e.g. HTT for other p_{T}? HTT has been shown to be quite stable in pile-up...
- Optimisation of parameters – initial R parameter, cuts, reclustering algorithm, trimming parameters, ASF parameters...

Conclusions

- I have showed you one way to make use of the jet-by-jet information on opening angles in a jet
- This I use for hadronic top tagging at high p_{T}
- Some initial studies show promising results
- ... also with respect to pile-up stability...
  - first coarse check gives results comparable to those published for other taggers at high efficiency
- ... but a dedicated pile-up treatment would help.

References

- Martin Jankowiak and Andrew J. Larkoski, *Jet Substructure Without Trees*, SLAC–PUB–14433, arXiv:1104.1646.
- T. Plehn, G. P. Salam and M. Spannowsky, *Fat Jets for a Light Higgs*, Phys. Rev. Lett. **104** (2010) 111801 [arXiv:0910.5472 [hep-ph]].
- A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, *"TMVA: Toolkit for Multivariate Data Analysis"*, PoS A CAT 040 (2007) [physics/0703039].

Backup

Simulations

- Input: stable particles ($c\tau < 1$ m) in Pythia
  - no muons or neutrinos
  - signal: fully hadronic $t\bar{t}$ decays
  - background: qq, qg, gg
  - generated with flat distribution in $\hat{p}_{\mathrm{T}}$, 1 – 2.5 TeV
- Pile-up:
  - Pythia min-bias, AMBT2B tune with CTEQ6L1 PDFs
  - Added to hard scatter event according to number of pile-up events $\mu \in$ Poisson($\langle\mu\rangle$), $\langle\mu\rangle = 20$

Jets

- Jets:
  - anti-k_t, $R = 1.2$
  - $p_T^{jet} > 100$ GeV in pile-up environment, 20 GeV otherwise (makes no difference since $\hat{p}_T$ is large and ...)
  - Leading or subleading
- Trimming:
  - recluster with k_t, $R_{kt} = 0.3$
  - rejecting subjets with p_T fraction < 0.03
- If trimming is done, this is done *before* the ASF machinery
- ASF reclustering:
  - recluster with k_t, $R_{kt} = \frac{R_i}{2} + 0.1$
- Histograms as well as training/testing:
  - Require that the jet is leading or subleading

More ASF details

We have

$$\Delta G(R) \equiv R \frac{\sum_{i \neq j} p_{Ti} p_{Tj} \Delta R_{ij}^2 \delta_{dR}(R - \Delta R_{ij})}{\sum_{i \neq j} p_{Ti} p_{Tj} \Delta R_{ij}^2 \Theta(R - \Delta R_{ij})}$$

- i, j indicate jet constituents, and
- ΔR_{ij} is their relative distance.
- δ_{dR} is a narrow Gaussian approximating a δ function.
  - $\delta_{dR}(x) = \frac{\exp(-x^2/2dR^2)}{\sqrt{2\pi}dR}$
  - Its width $dR = 0.06$
- $\Theta(x)$ is the Heaviside step function.
  - In practice, replaced by the error function as the δ function is replaced by a Gaussian, with the same width

MVA settings

One of the most performant out of the methods available in the TMVA package, without much tuning, was a Boosted Decision Tree (BDT) with Adaptive Boost. Parameters are given below.

parameter	value
N_{trees}	850
$N^{min}_{leafsinnode}$	150
Maximum depth of forest	4
N_{cuts}	40
β_{boost}	0.75
Separation type	Gini index

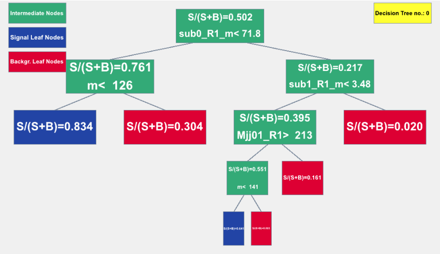
MVA variables

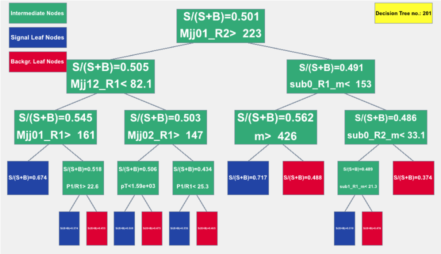
Variables, ordered by importance ranking in BDT method. No large differences though.

rank	parameter	value
1	m_{sub0}^{R1}	$1.042 \cdot 10^{-1}$
2	m	$9.547 \cdot 10^{-2}$
3	$P_1\,/\,R_1$	$9.495 \cdot 10^{-2}$
4	p_{T}	$9.062 \cdot 10^{-2}$
5	R_1	$8.465 \cdot 10^{-2}$
6	P_1	$7.862 \cdot 10^{-2}$
7	$Mjj01_{R1}$	$7.403 \cdot 10^{-2}$
8	m_{sub1}^{R1}	$6.549 \cdot 10^{-2}$
9	$Mjj02_{R1}$	$6.432 \cdot 10^{-2}$
10	$Mjj12_{R1}$	$6.238 \cdot 10^{-2}$
11	m_{sub0}^{R2}	$6.065 \cdot 10^{-2}$
12	$Mjj01_{R2}$	$4.601 \cdot 10^{-2}$
13	$Mjj02_{R2}$	$4.303 \cdot 10^{-2}$
14	m_{sub0}^{R3}	$0.559 \cdot 10^{-2}$

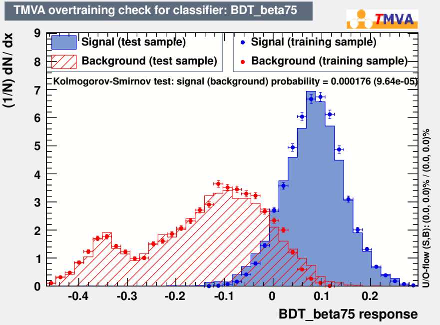
Example BDTs

$\leftarrow$: no, yes : $\rightarrow$



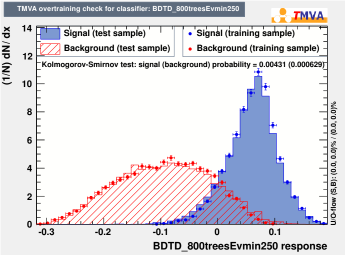


Overtrain check

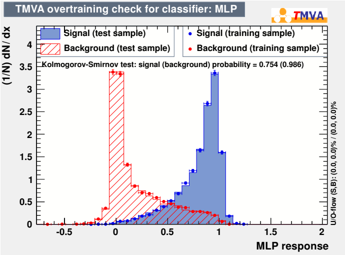


"Standard" BDT

All of these show similar performance.

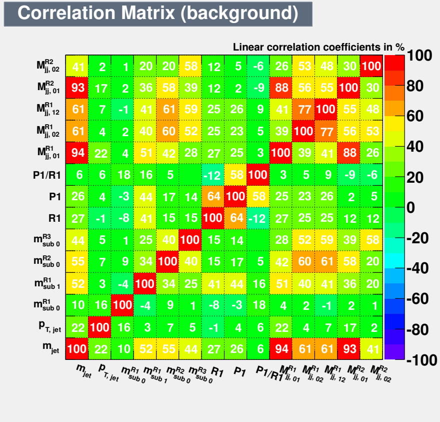


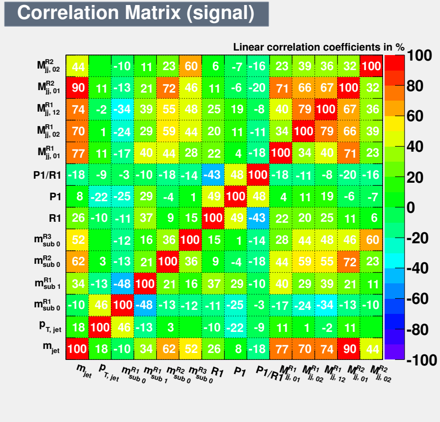
Using decorrelated variables



Neural network

Correlations





For some variables, very different correlations between signal and background

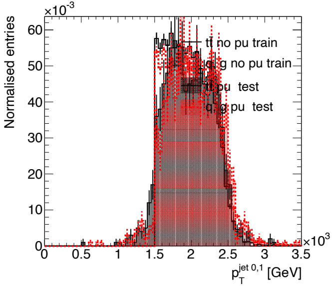
Distributions

Training and testing samples overlaid

$t\bar{t}$ and q, g

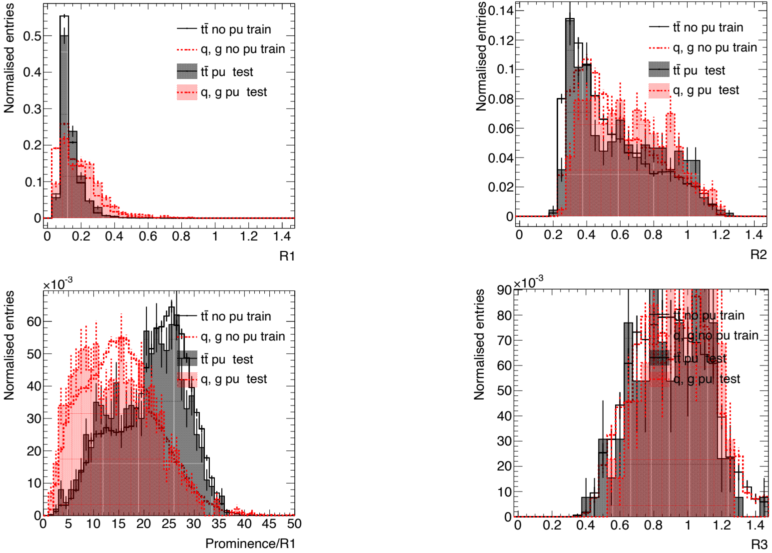
$p_{\mathrm{T}}^{jet} > 1500$ GeV

jet p_T

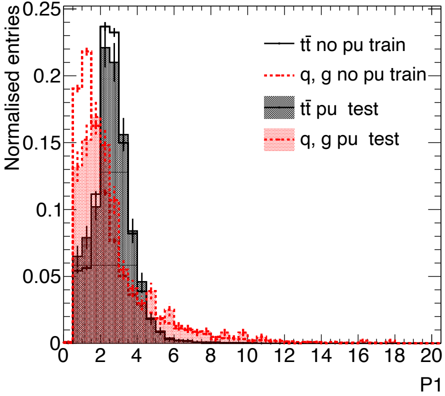


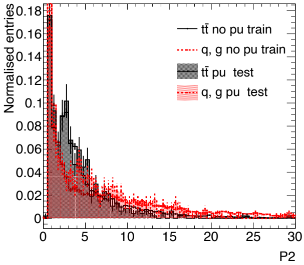
In pile-up run: passing initial cuts (p_T, index)
Training sample: 22769/25435 sig/bkg
Testing sample: 948/964 sig/bkg

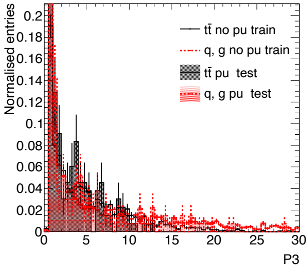
R_i and P_1 / R_1


Normalised entries
tt̄ no pu train
q, g no pu train
tt̄ pu test
q, g pu test
R1
R2
×10^-3
Prominence/R1
R3

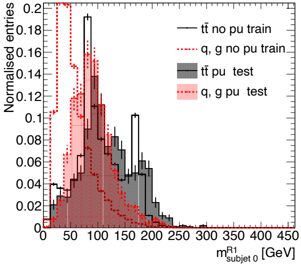
P_i

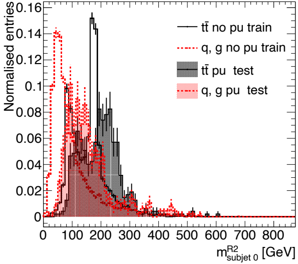

Normalised entries
0.25
0.2
0.15
0.1
0.05
0
0 2 4 6 8 10 12 14 16 18 20
P1
tt̄ no pu train
q, g no pu train
tt̄ pu test
q, g pu test

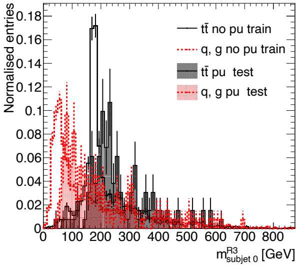

Normalised entries
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0 5 10 15 20 25 30
P2
tt̄ no pu train
q, g no pu train
tt̄ pu test
q, g pu test

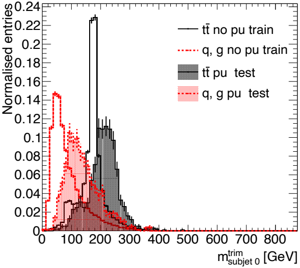

Normalised entries
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0 5 10 15 20 25 30
P3
tt̄ no pu train
q, g no pu train
tt̄ pu test
q, g pu test

Leading ASF R_i subjet mass + leading trim subjet mass

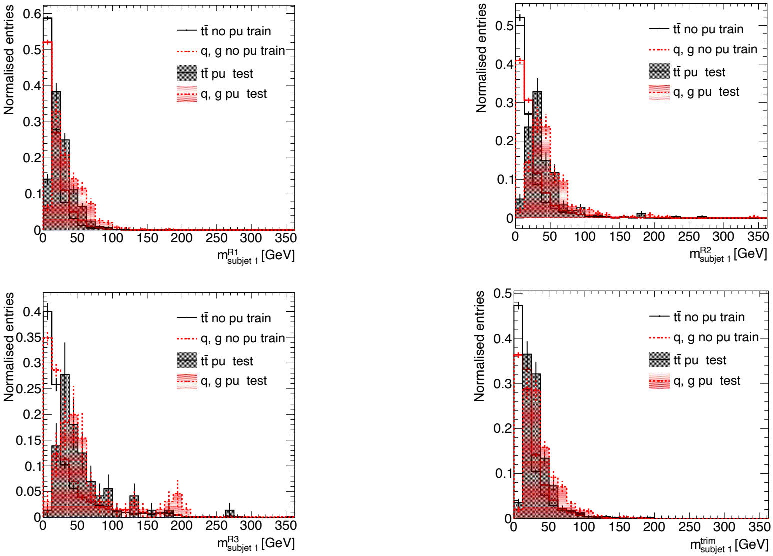




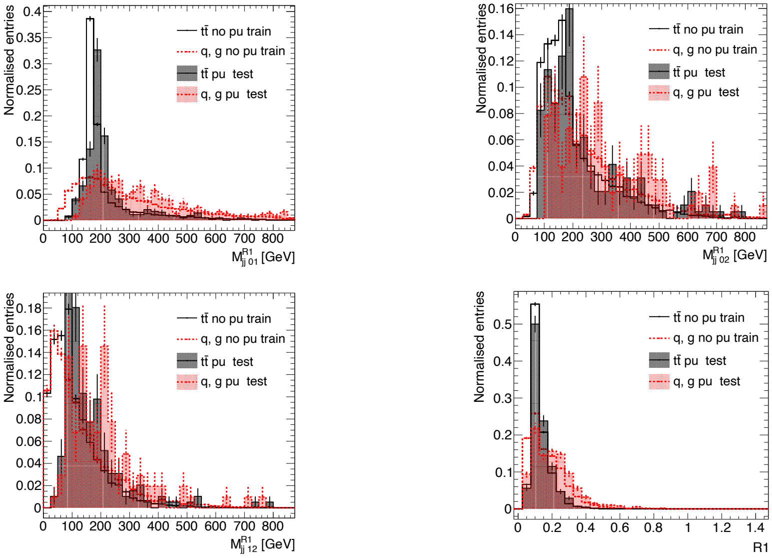




Subleading ASF R_i subjet mass



ASF R_1 di-subjet mass



ASF R_2 di-subjet mass