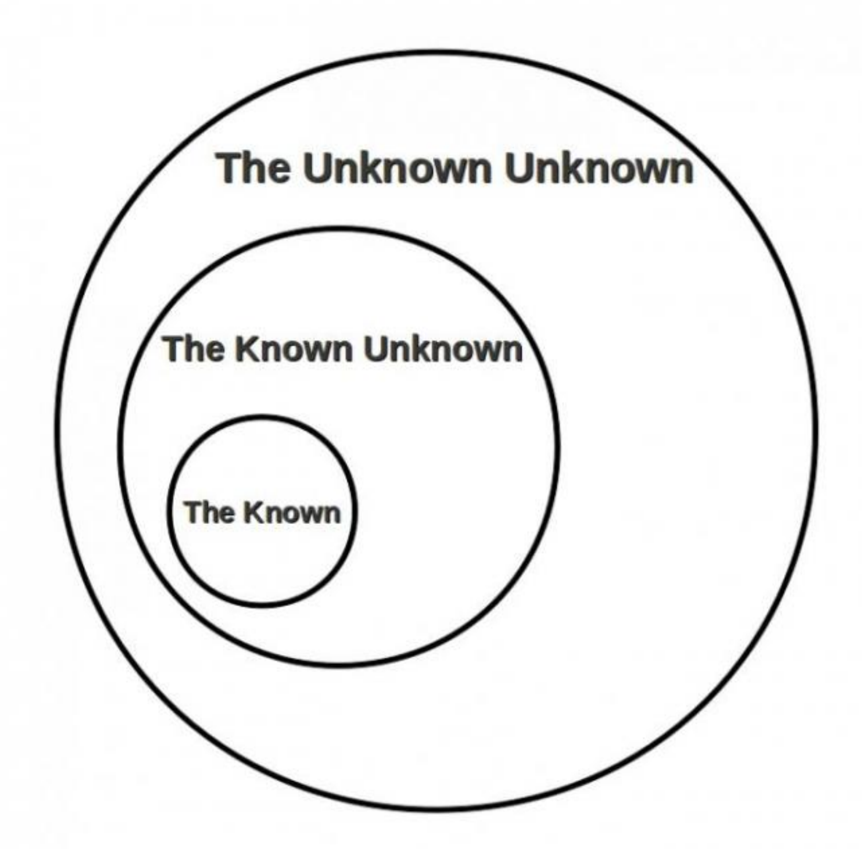
ML in (LHC) Experiments



Deep Thinking
vs
Deep Learning

Deepak Kar
deepak.kar@cern.ch

Session 2:
(ML4)Jets

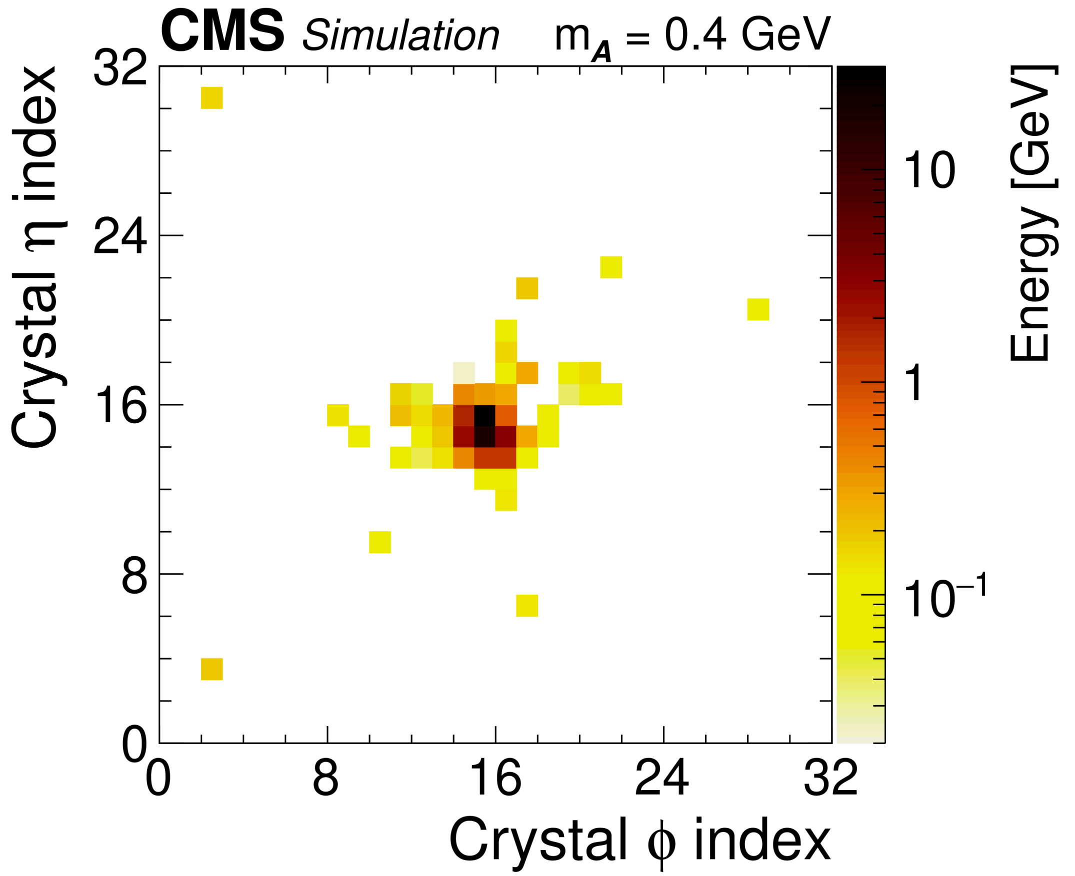
Use of ML in object reconstruction

Example: CMS merged photon reconstruction

Signal: H -> AA ->$\boldsymbol{\gamma\gamma\gamma\gamma}$ (merged)

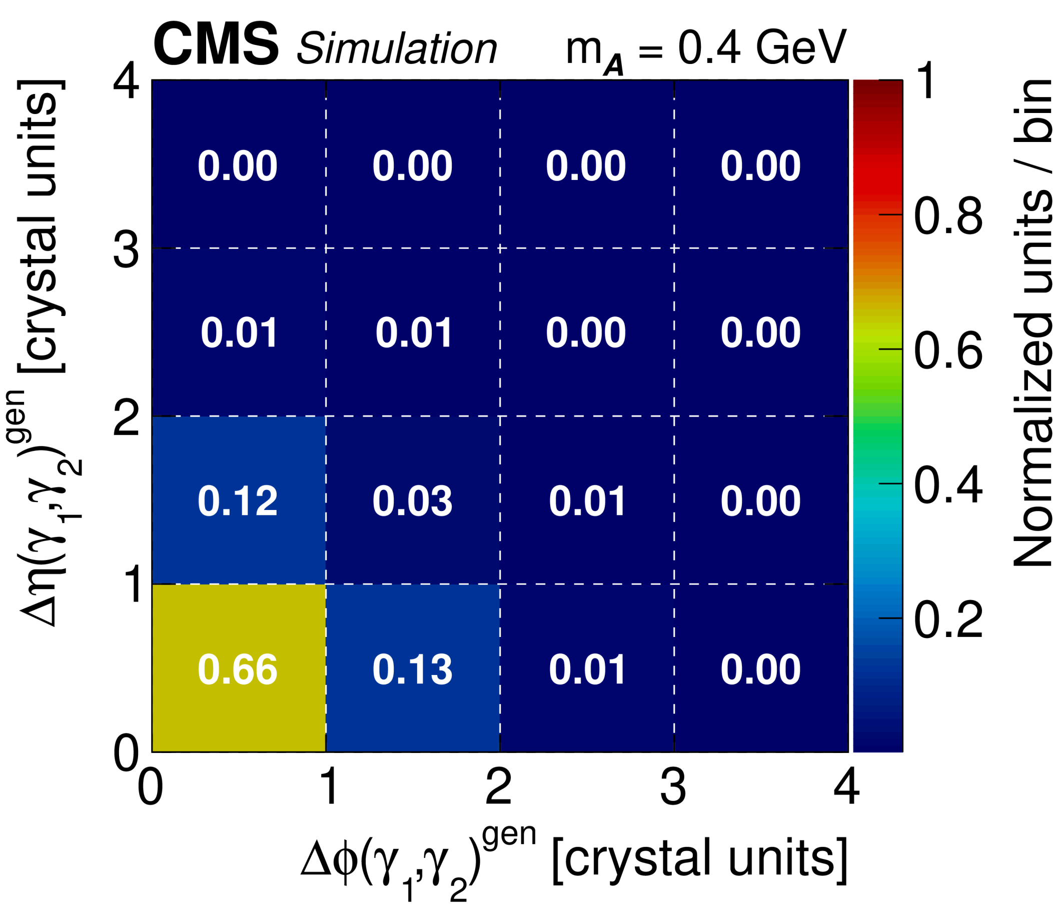
BG: H->$\boldsymbol{\gamma\gamma}$

Use: end to end ML approach, makes use of more granular information



https://cms-results.web.cern.ch/cms-results/public-results/publications/EGM-20-001/

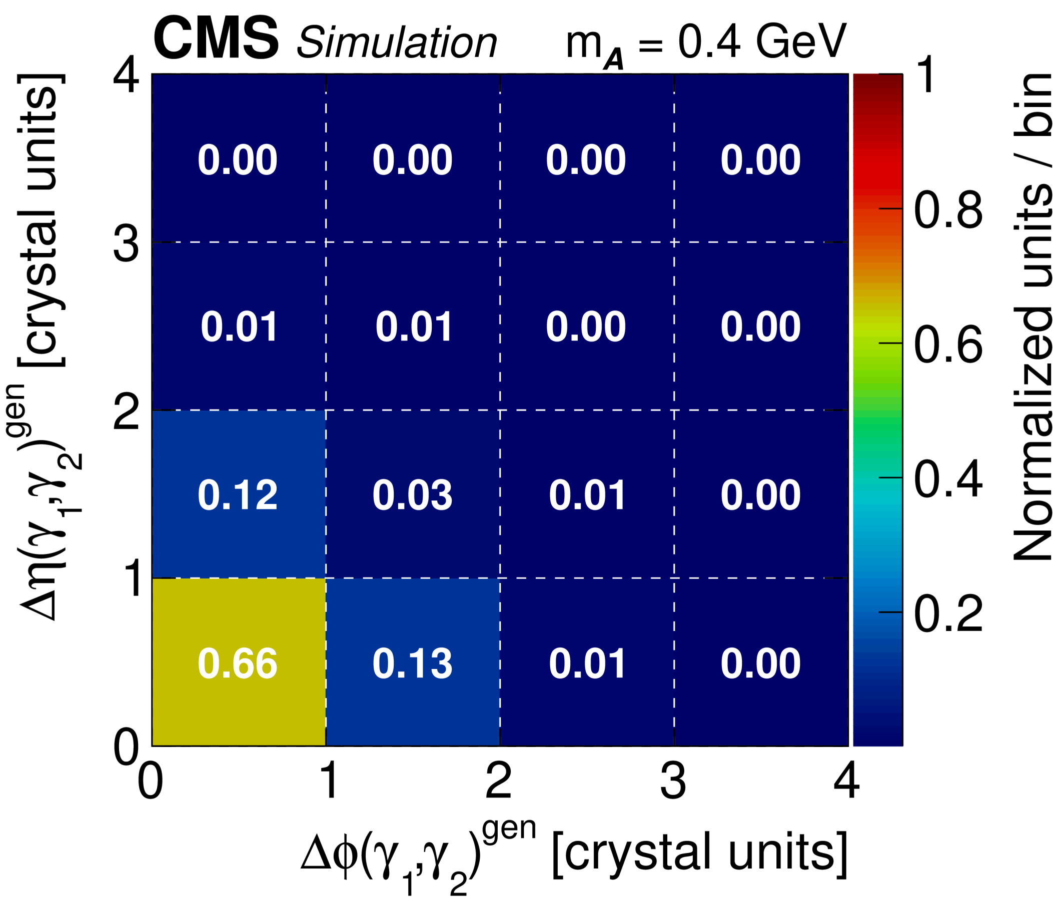
Example: CMS merged photon reconstruction



Number of ECAL crystals in either direction, each pixel in the image exactly corresponds to the energy deposited in a single ECAL crystal.

To improve spatial resolution: split the ECAL images described above into a two-layer image that contains the transverse and longitudinal components of the crystal energy. And include the crystal seed coordinates.

Example: CMS merged photon reconstruction



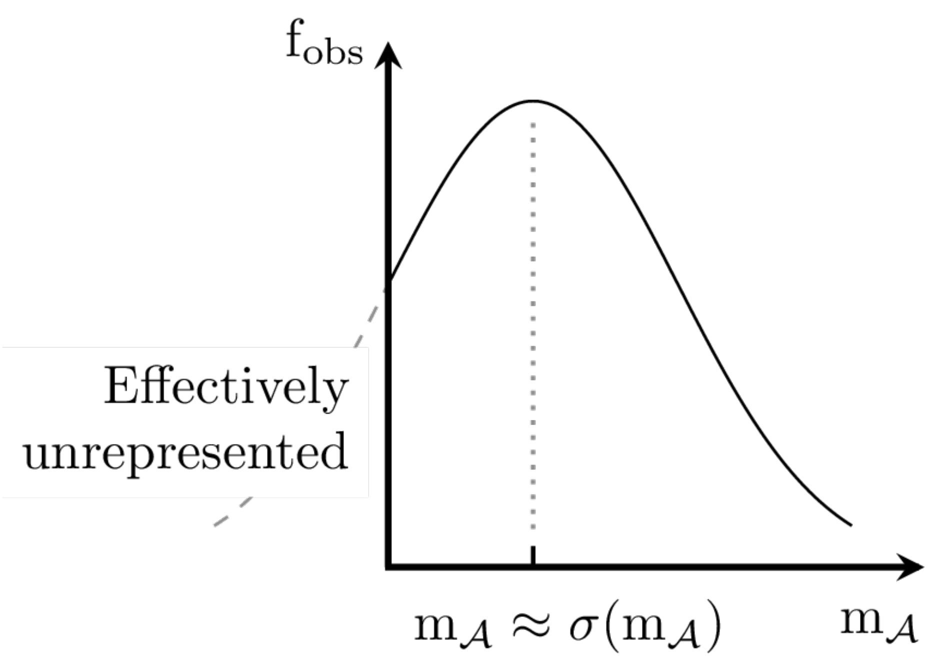
No rotation is performed on the images since electromagnetic showers are not rotationally symmetric. In addition to the η-ϕ symmetry being broken by the CMS magnetic field, general rotations of square pixels are destructive operations that distort the particle shower pattern

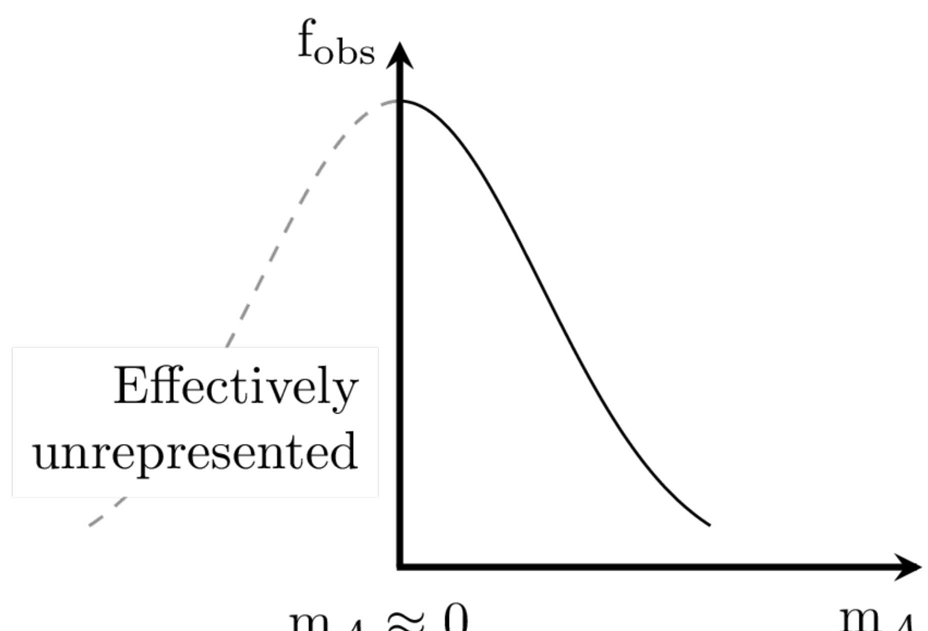
To improve spatial resolution: split the ECAL … ver image that contains the transverse and longitudinal components of the crystal energy. And include the crystal seed coordinates.

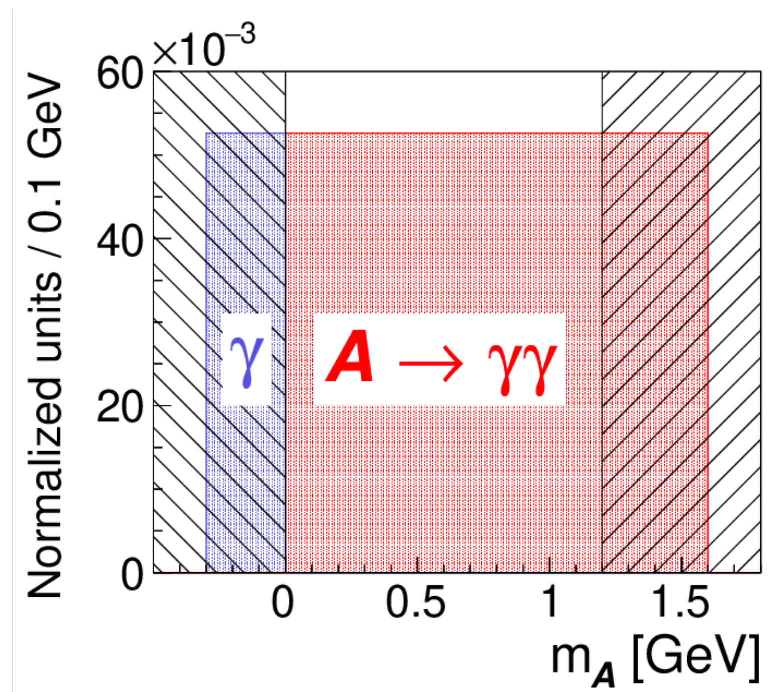
Example: CMS merged photon reconstruction

RESNET CNN, outputs regressed mass in a global maximum pooling layer. Concatenated with the crystal seed coordinates, gives regressed diphoton mass.

BUT…



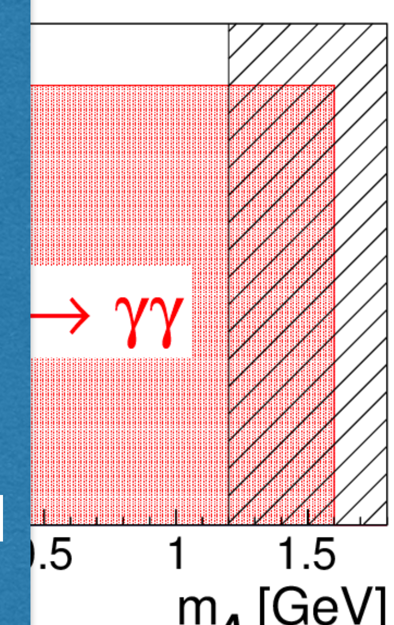




Example: CMS merged photon reconstruction

RESNET CNN, outputs regressed mass in a global maximum pooling layer. Concatenated with the crystal seed coordinates, gives regressed diphoton mass.

BUT…

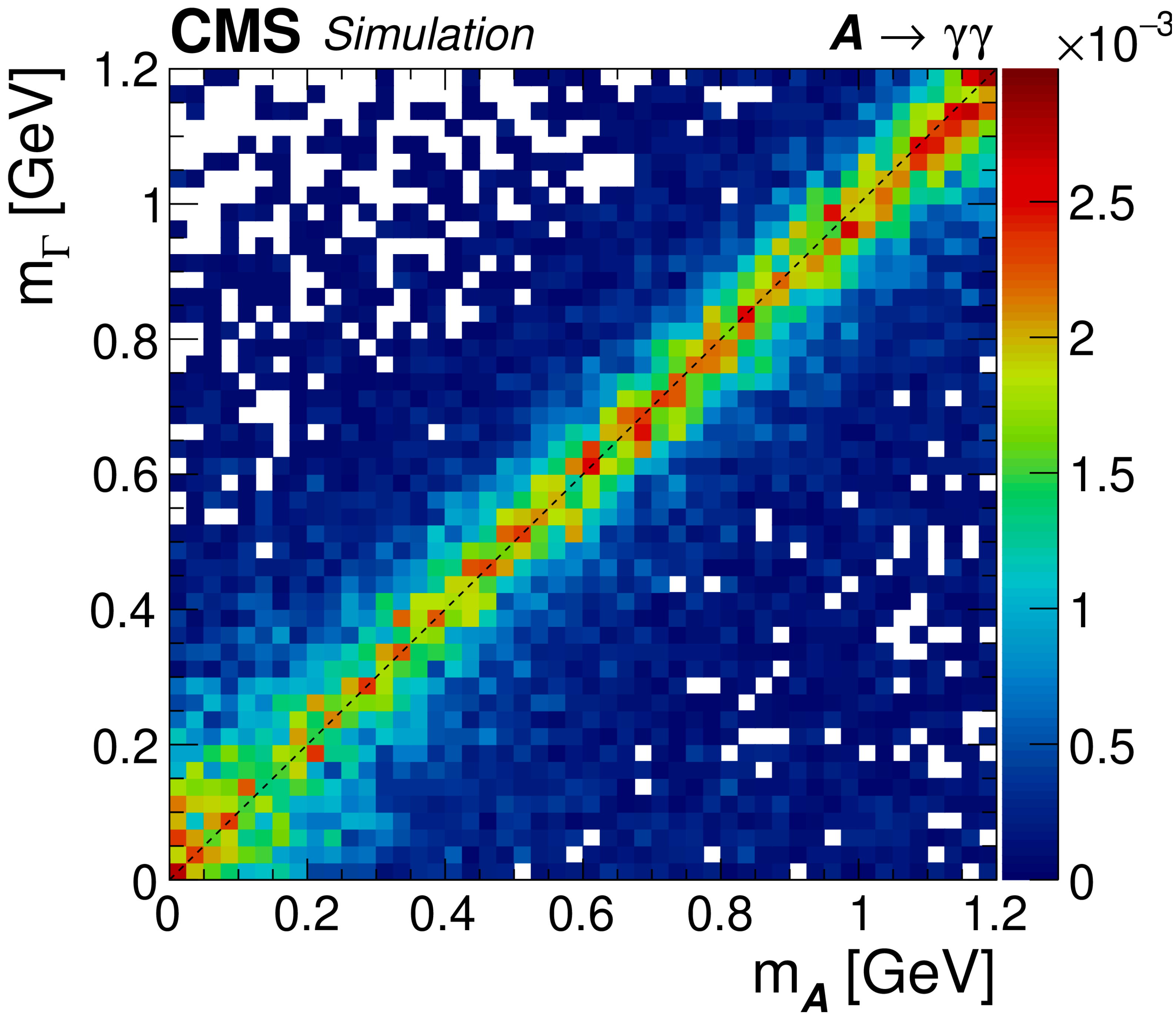


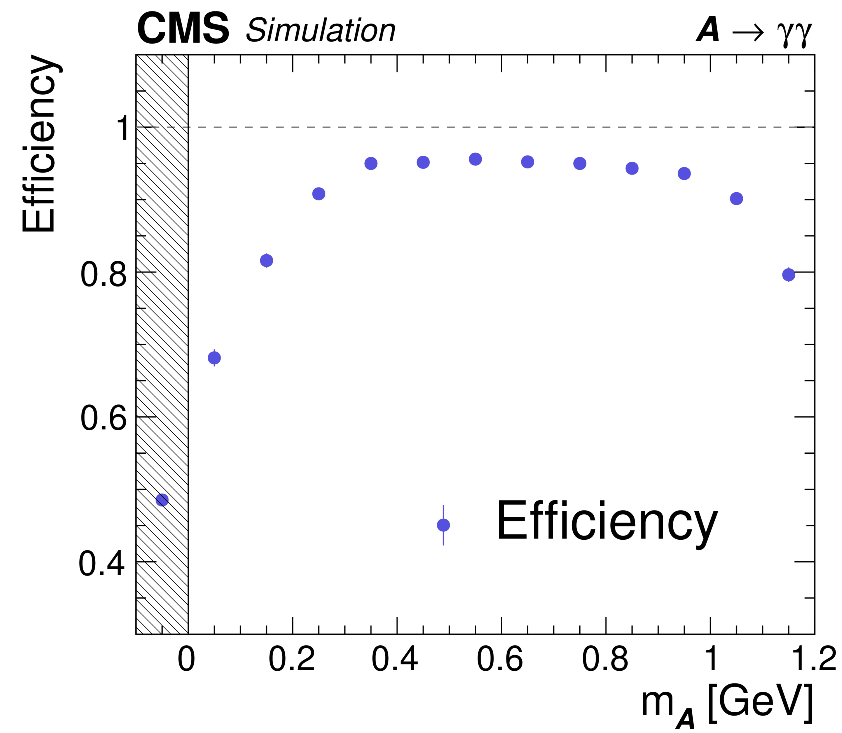
f o

Effectivel
unrepresente

When regressing below the mass resolution, the left tail of the mass distribution becomes underrepresented in the training set. Middle: As mA → 0, only half of the mass distribution is represented. The regressor subsequently defaults to the last full mass distribution at mA ≈ σ(mA). With domain continuation, the generated mass distribution of the original training samples (A → γγ, red region) is augmented with topologically similar samples that are randomly assigned nonphysical masse. This allows the regressor to see a full mass distribution over the entire region of interest

Example: CMS merged photon reconstruction

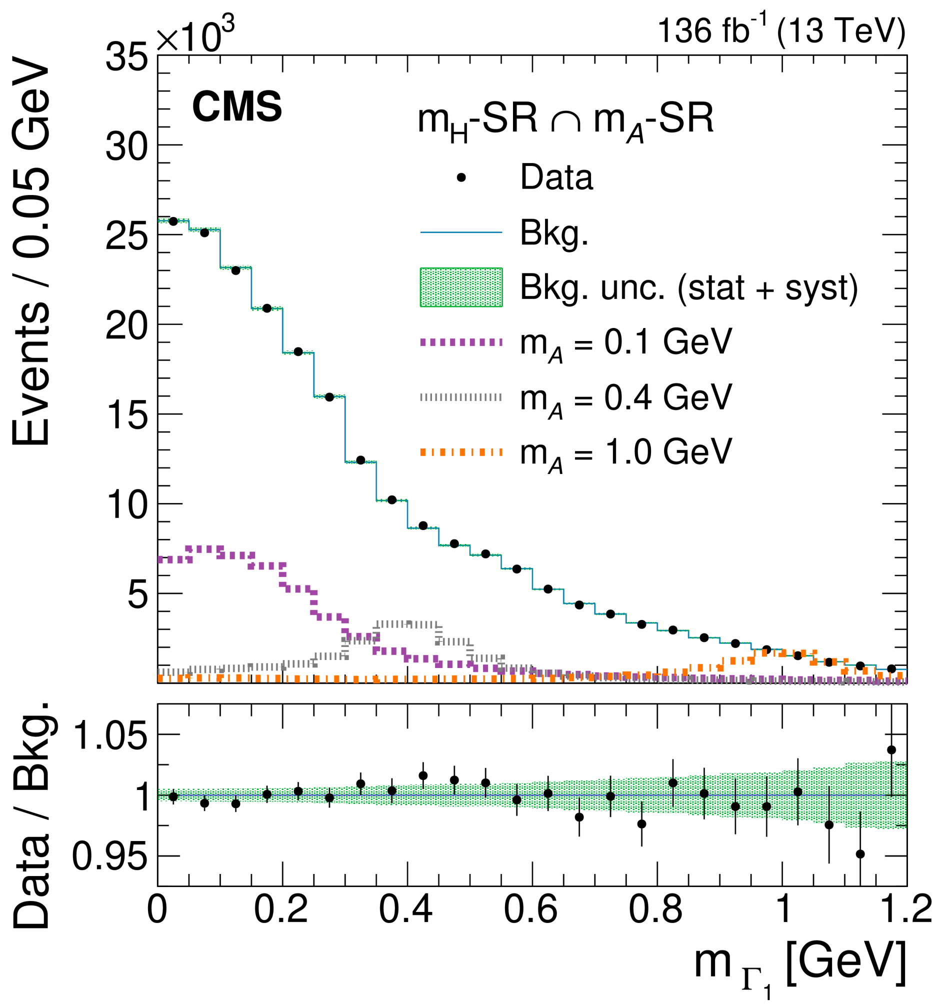


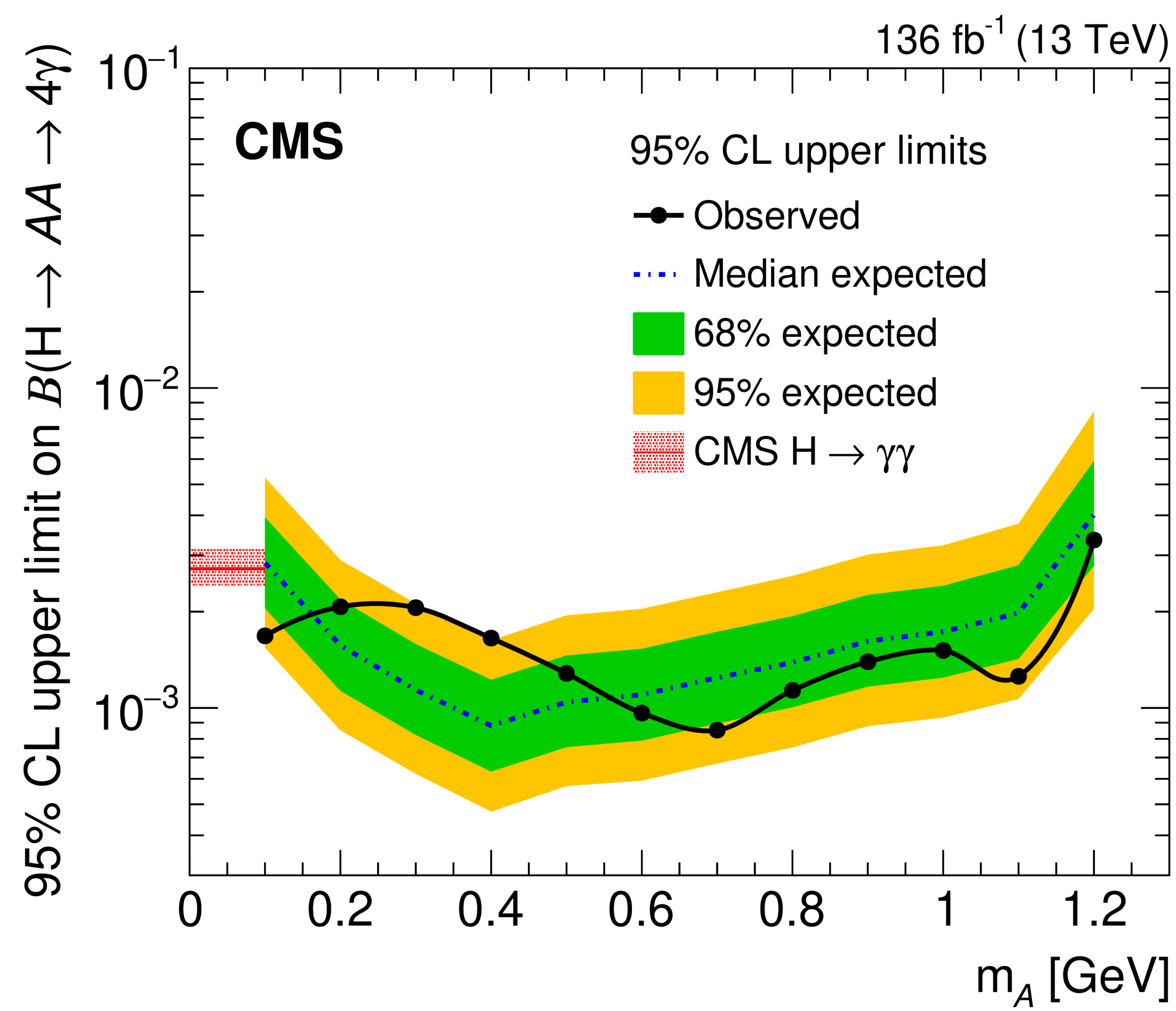


Final result, also validated in data

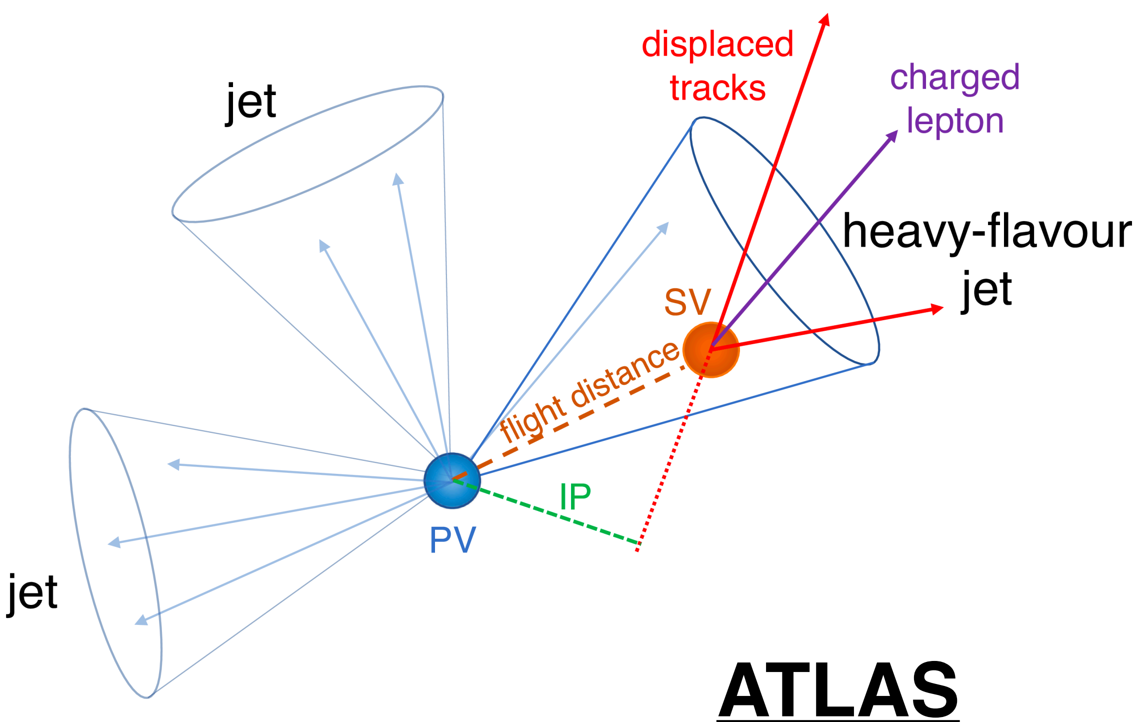
Used in Search:

https://arxiv.org/abs/2209.06197





Example: b/c-jet tagging



CMS

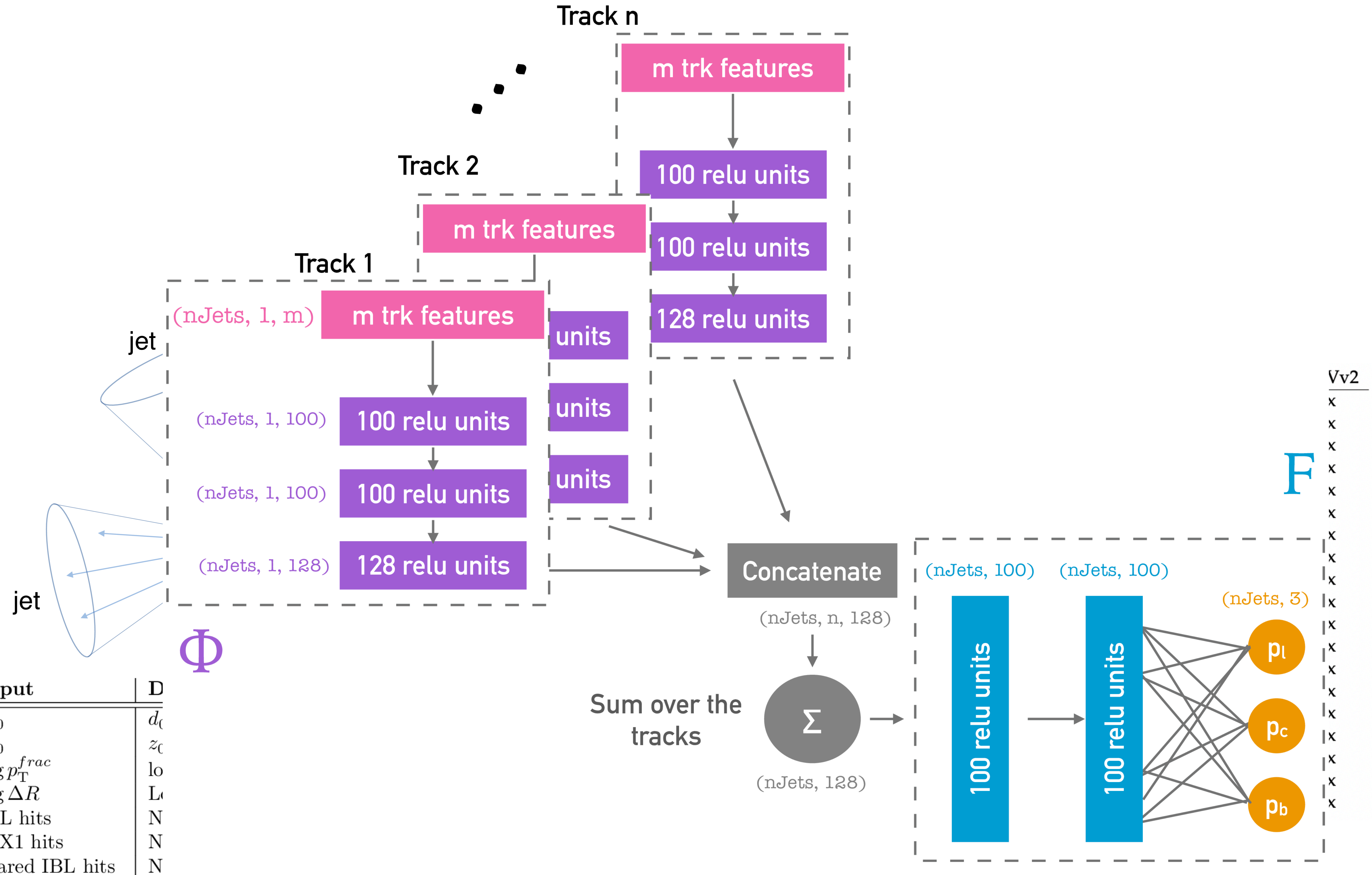
Input variable	Run 1 CSV	CSVv2
SV 2D flight distance significance	x	x
Number of SV	—	x
Track η_{rel}	x	x
Corrected SV mass	x	x
Number of tracks from SV	x	x
SV energy ratio	x	x
ΔR(SV, jet)	—	x
3D IP significance of the first four tracks	x	x
Track $p_{\mathrm{T,rel}}$	—	x
ΔR(track, jet)	—	x
Track $p_{\mathrm{T,rel}}$ ratio	—	x
Track distance	—	x
Track decay length	—	x
Summed tracks E_{T} ratio	—	x
ΔR(summed tracks, jet)	—	x
First track 2D IP significance above c threshold	—	x
Number of selected tracks	—	x
Jet p_{T}		x
Jet η		x

DeepCSV tagger

ATLAS

Input	Description
s_{d0}	d_0/σ_{d0}: Transverse IP significance
s_{z0}	$z_0 \sin\theta/\sigma_{z0\sin\theta}$: Longitudinal IP significance
$\log p_{\mathrm{T}}^{frac}$	$\log p_{\mathrm{T}}^{track}/p_{\mathrm{T}}^{jet}$: Logarithm of fraction of the jet p_{T} carried by the track
$\log \Delta R$	Logarithm of opening angle between the track and the jet axis
IBL hits	Number of hits in the IBL: could be { 0, 1, or 2 }
PIX1 hits	Number of hits in the next-to-innermost pixel layer: could be { 0, 1, or 2 }
shared IBL hits	Number of shared hits in the IBL
split IBL hits	Number of split hits in the IBL
nPixHits	Combined number of hits in the pixel layers
shared pixel hits	Number of shared hits in the pixel layers
split pixel hits	Number of split hits in the pixel layers
nSCTHits	Combined number of hits in the SCT layers
shared SCT hits	Number of shared hits in the SCT layers

Using deep sets, modelling jets as a set of tracks



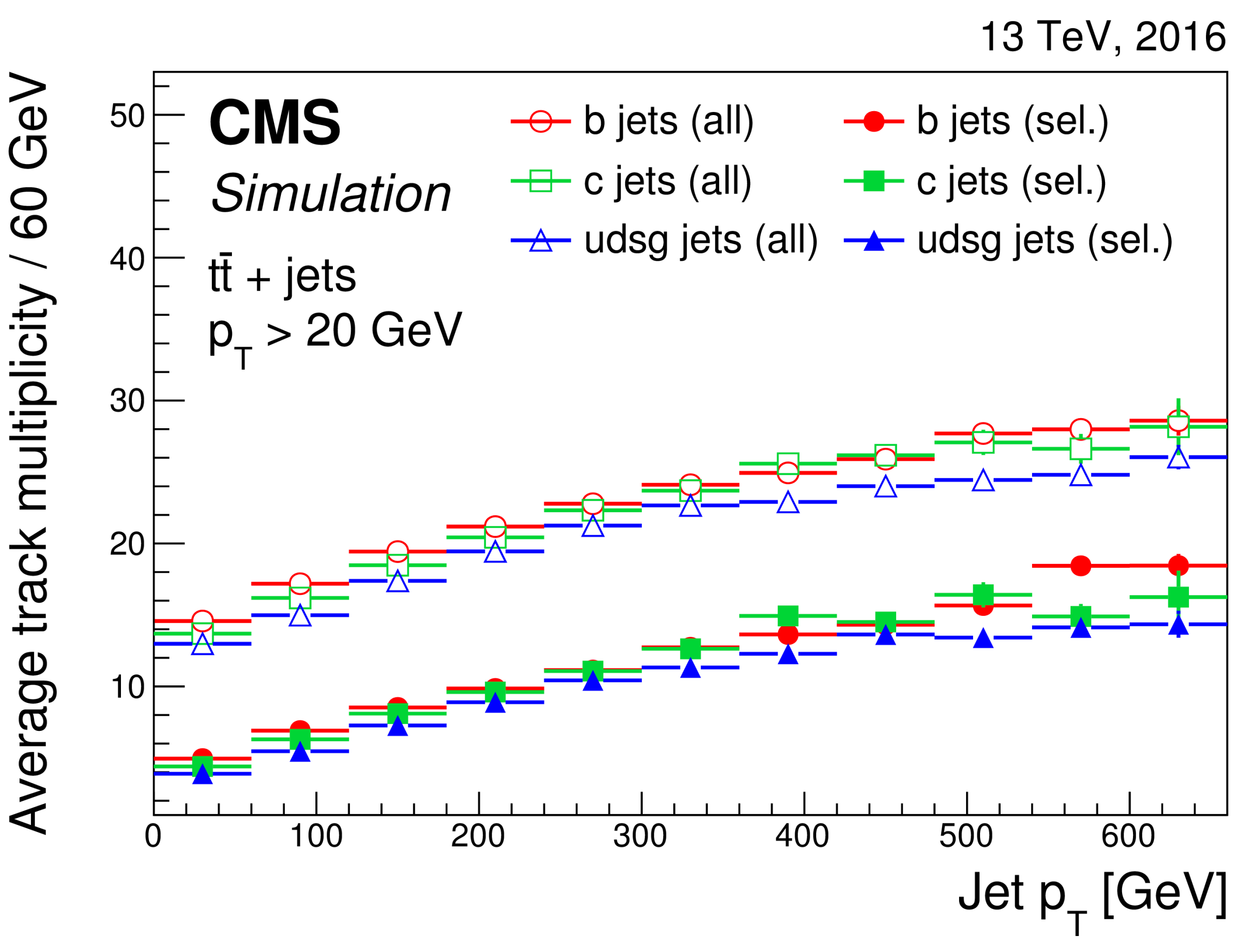
Input	D
s_{d0}	d(
s_{z0}	z(
$\log p_{\mathrm{T}}^{frac}$	lo
$\log \Delta R$	L(
IBL hits	N
PIX1 hits	N
shared IBL hits	N
split IBL hits	Number of split hits in the IBL
nPixHits	Combined number of hits in the pixel layers
shared pixel hits	Number of shared hits in the pixel layers
split pixel hits	Number of split hits in the pixel layers
nSCTHits	Combined number of hits in the SCT layers
shared SCT hits	Number of shared hits in the SCT layers

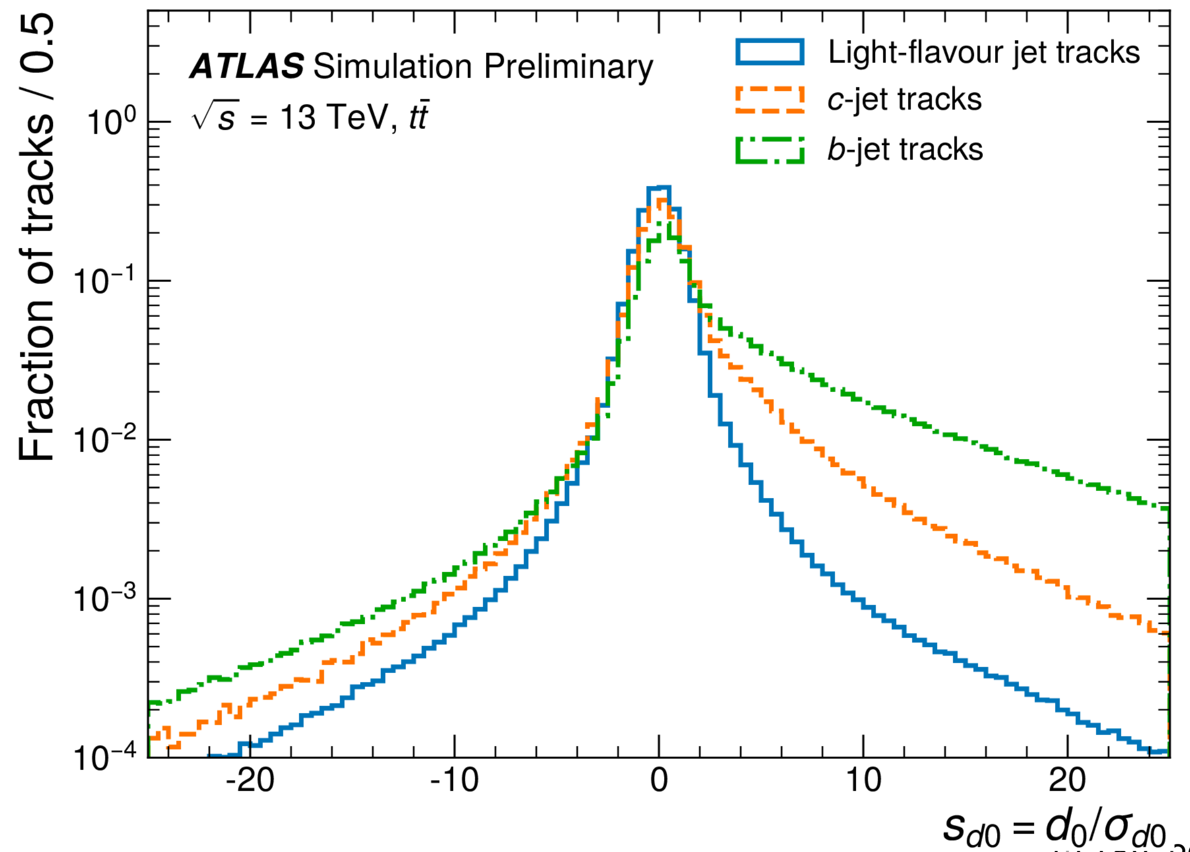
Using deep sets, modelling jets as a set of tracks

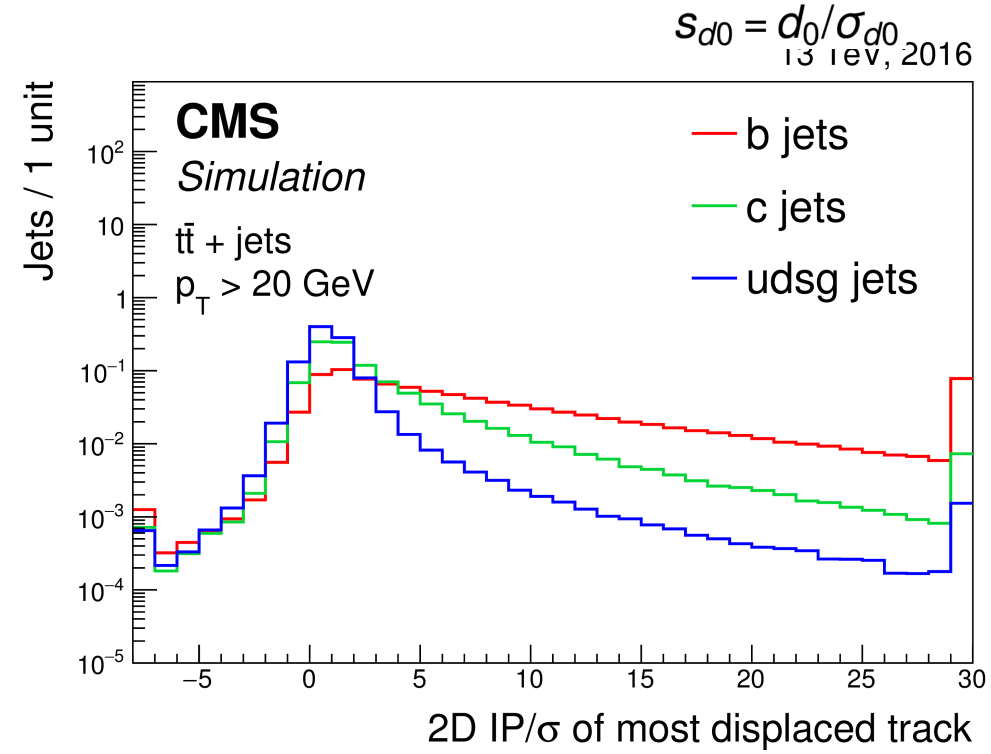
Quiz

Why so much effort on b-tagging?

Example: b/c-jet tagging

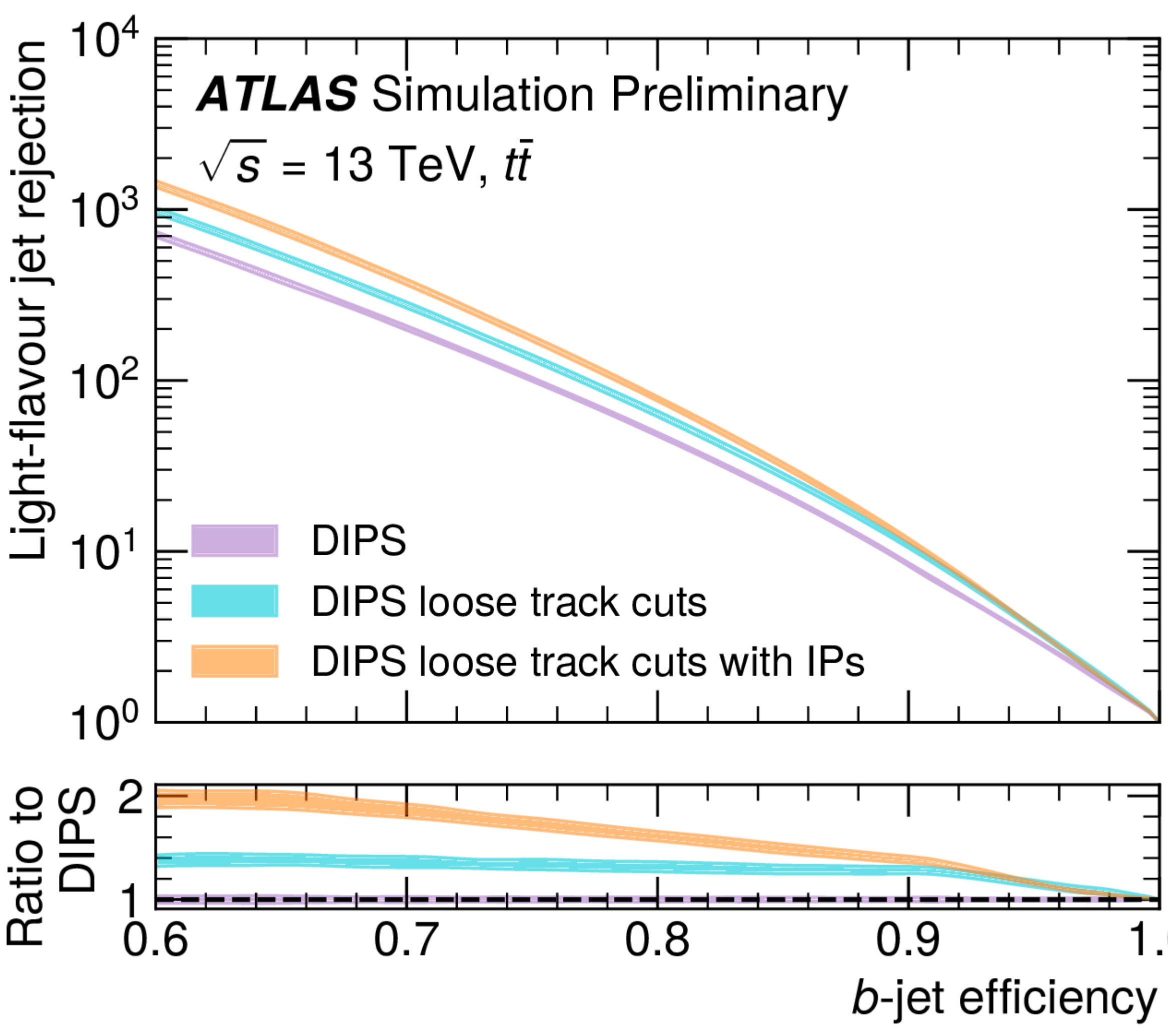




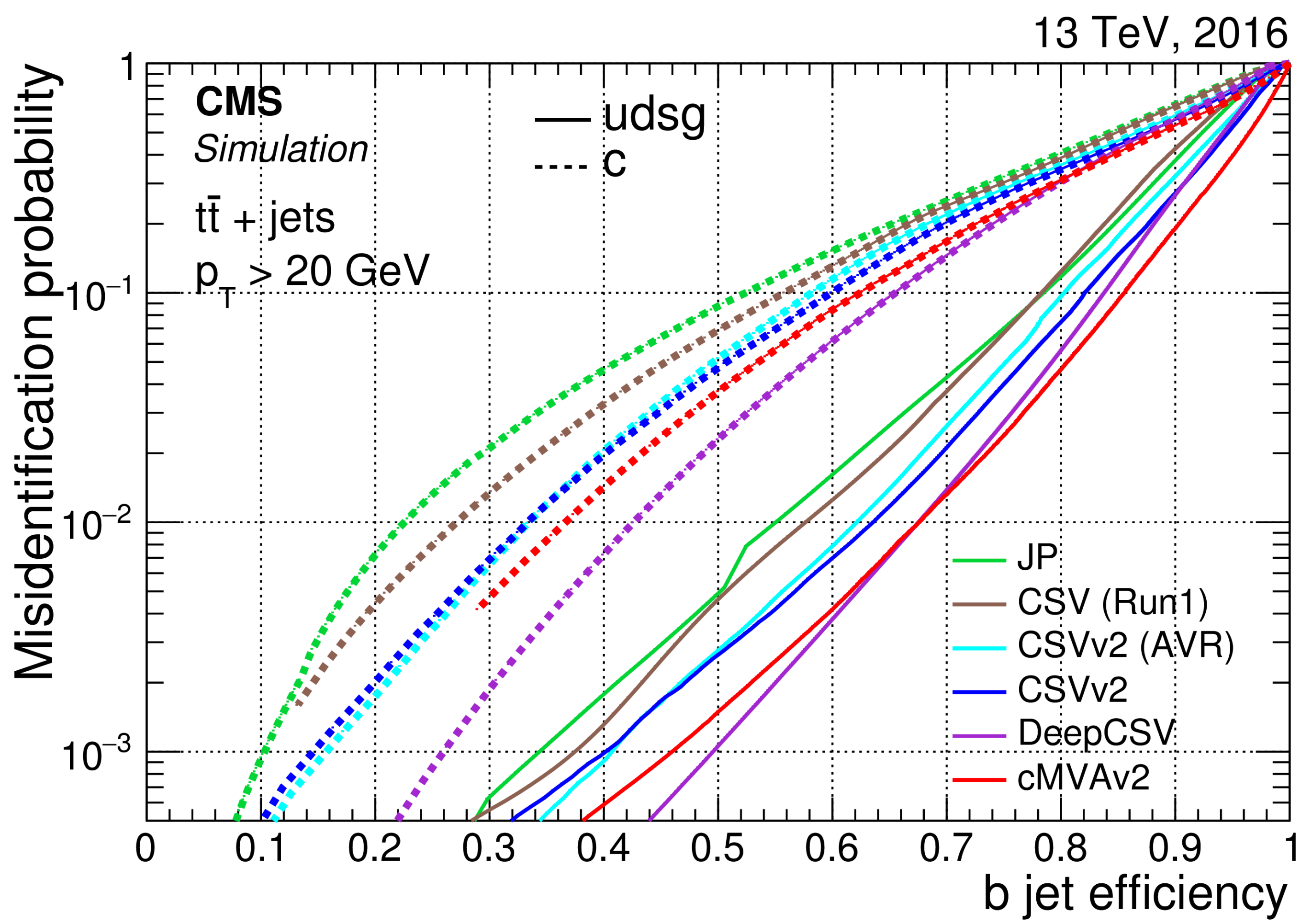


Example: b/c-jet tagging

ATLAS



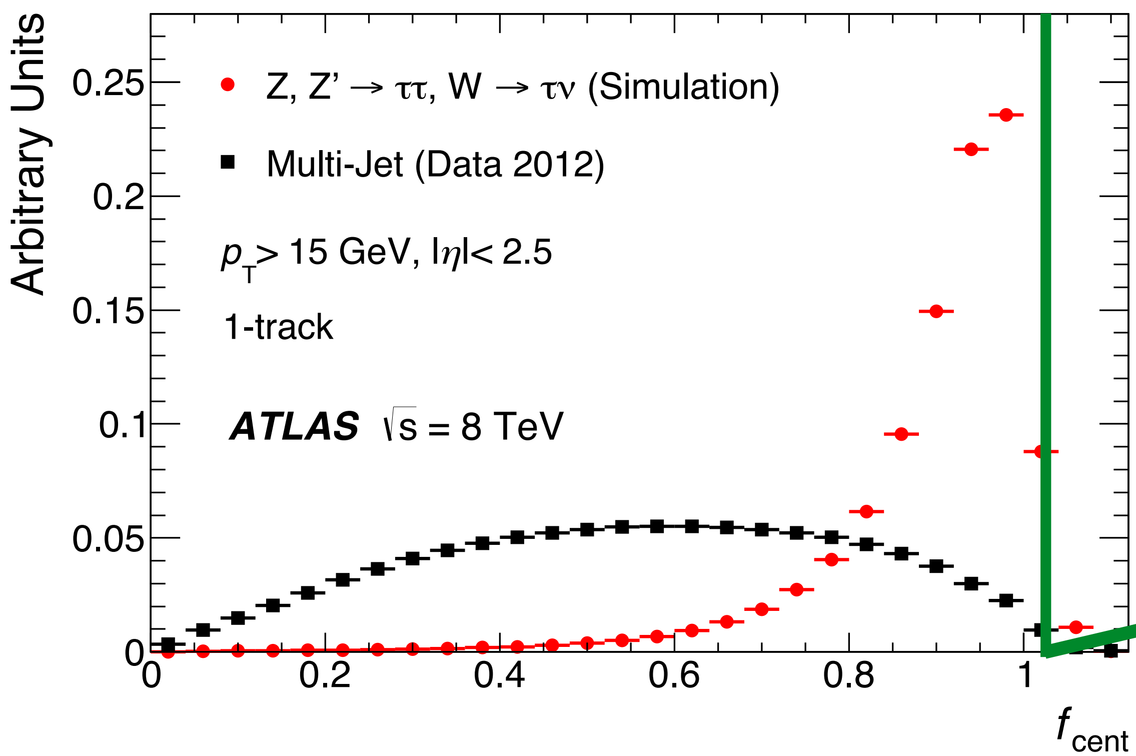
CMS



An aside: ROC Curves

Signal-Background Discrimination/ROC

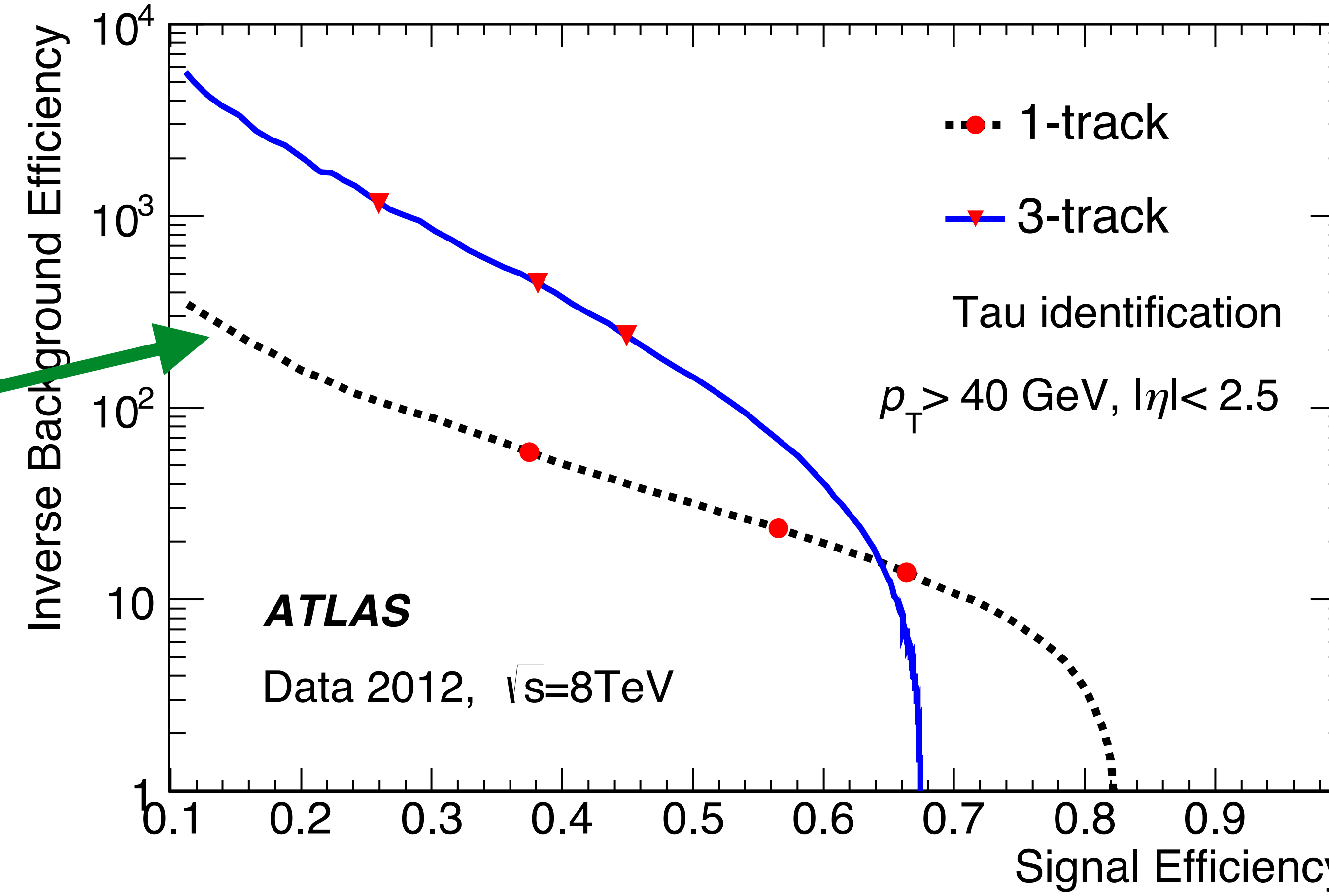
From MC



Scan over the full range

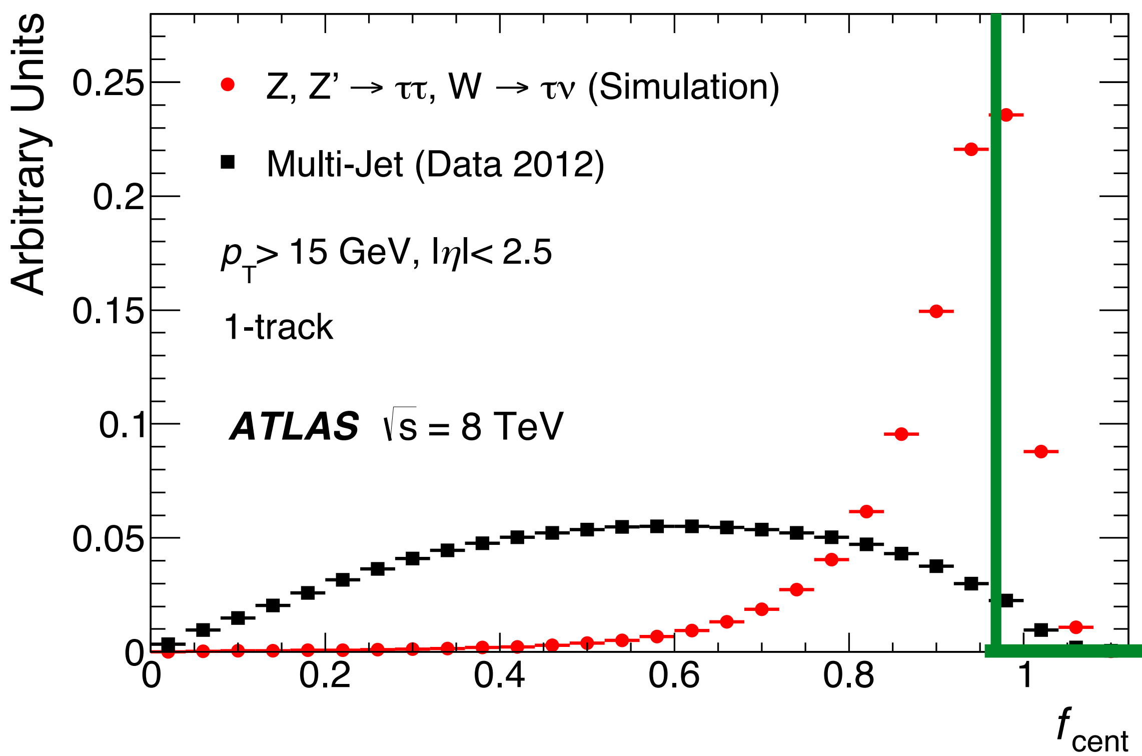
Signal efficiency: fraction of signal events right of the line

Inverse background efficiency: inverse fraction of background events right of the line



Signal-Background Discrimination/ROC

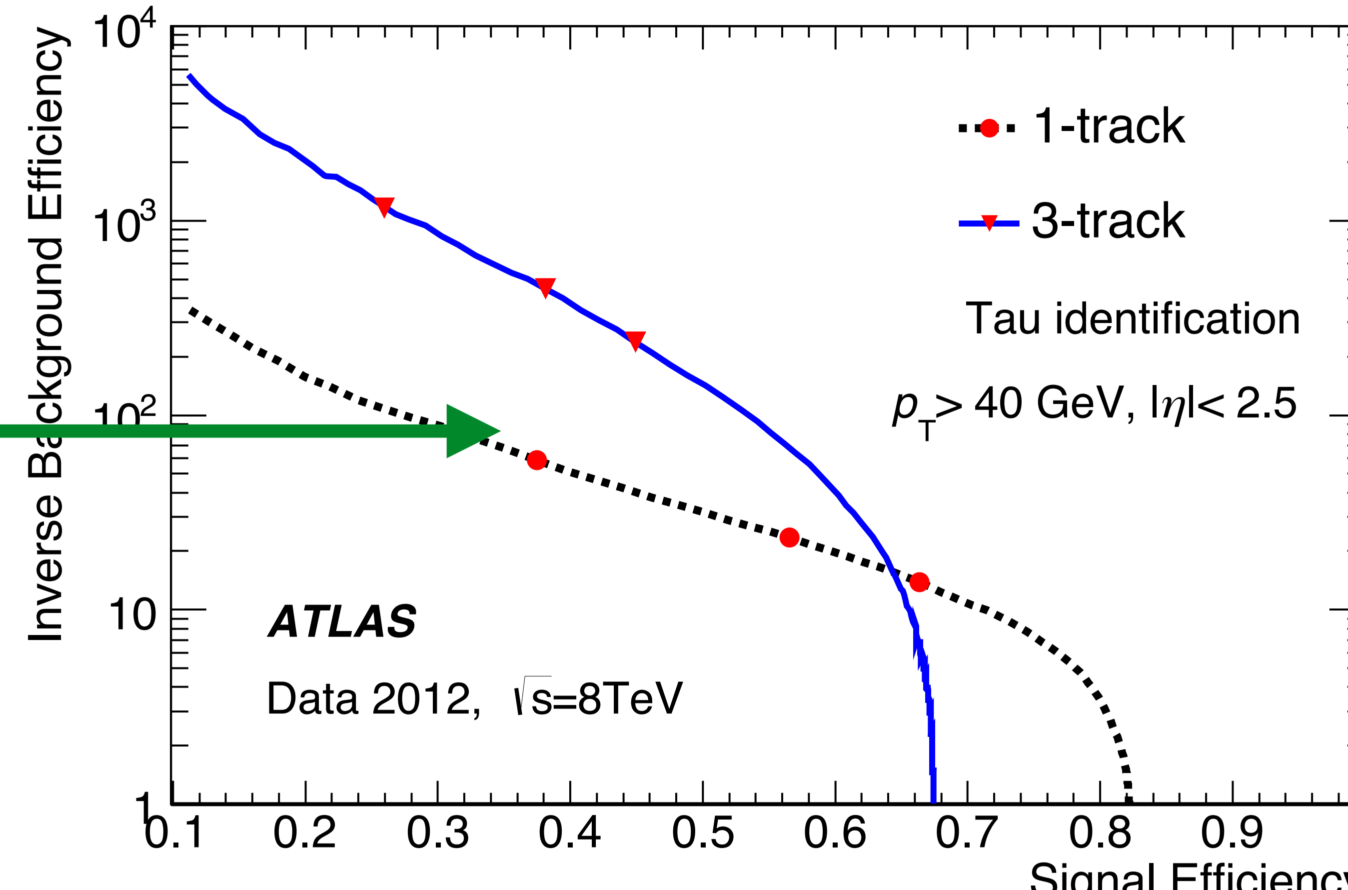
From MC



Scan over the full range

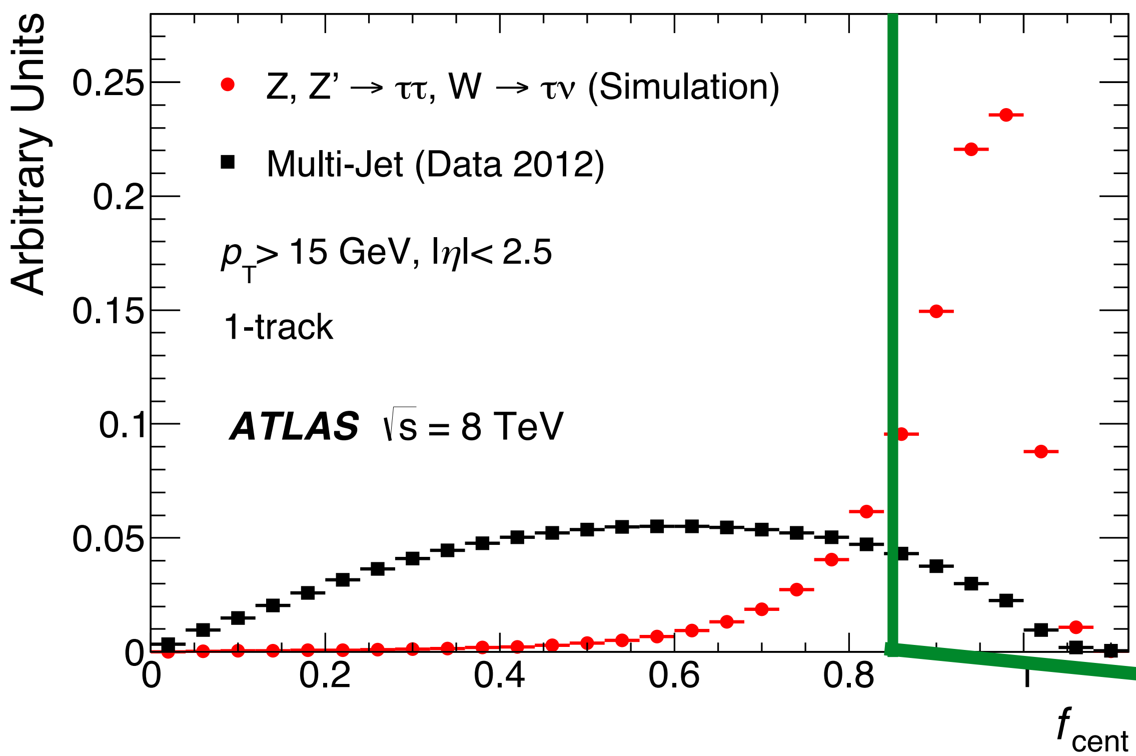
Signal efficiency: fraction of signal events right of the line

Inverse background efficiency: inverse fraction of background events right of the line



Signal-Background Discrimination/ROC

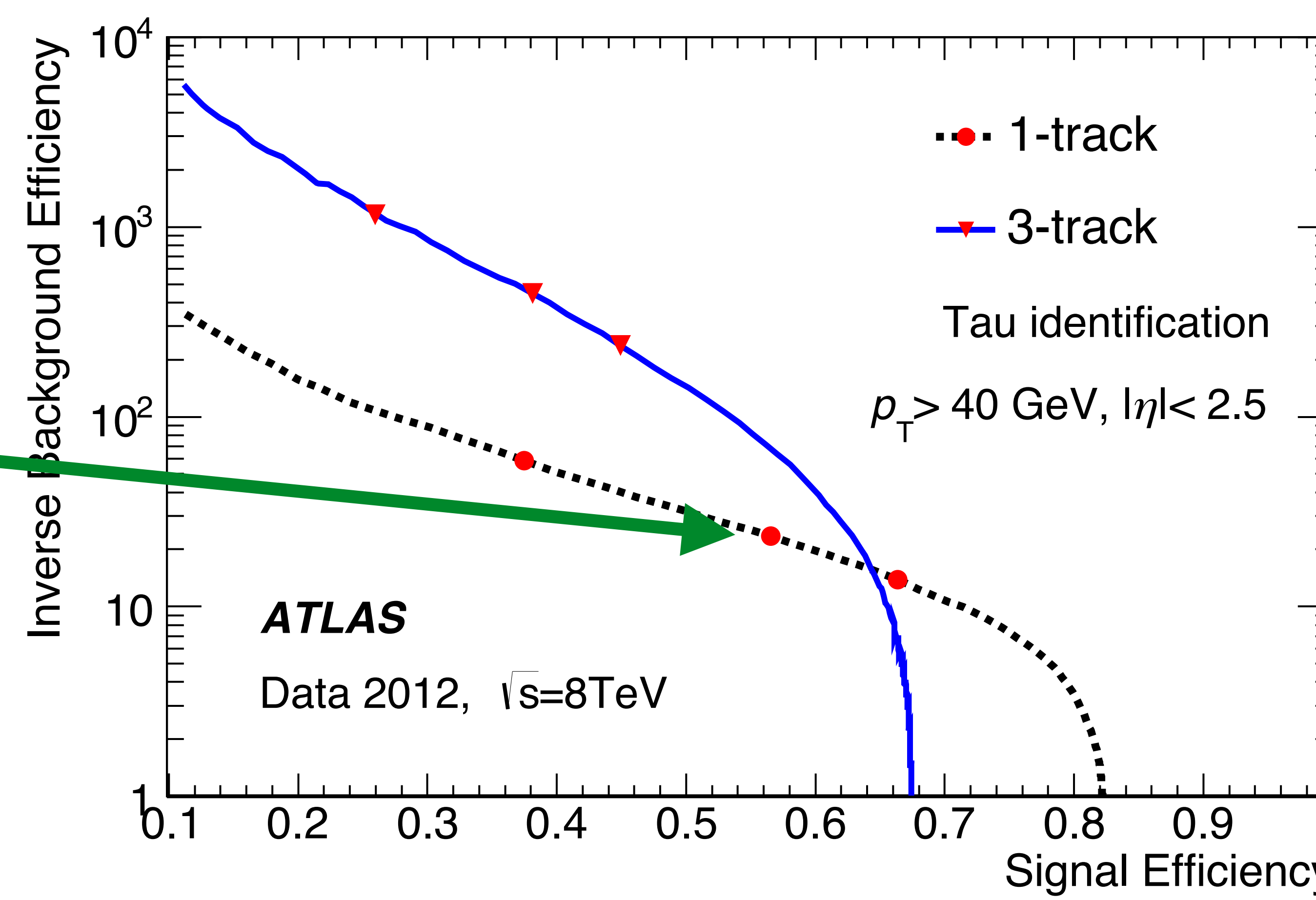
From MC



Scan over the full range

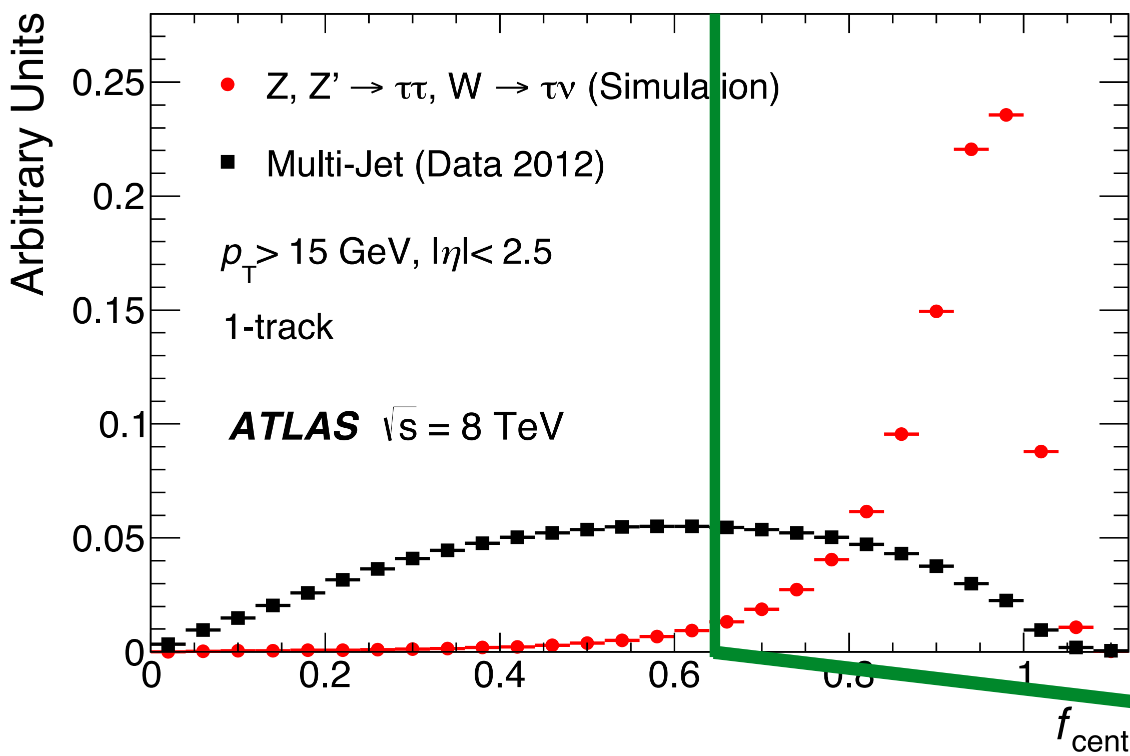
Signal efficiency: fraction of signal events right of the line

Inverse background efficiency: inverse fraction of background events right of the line



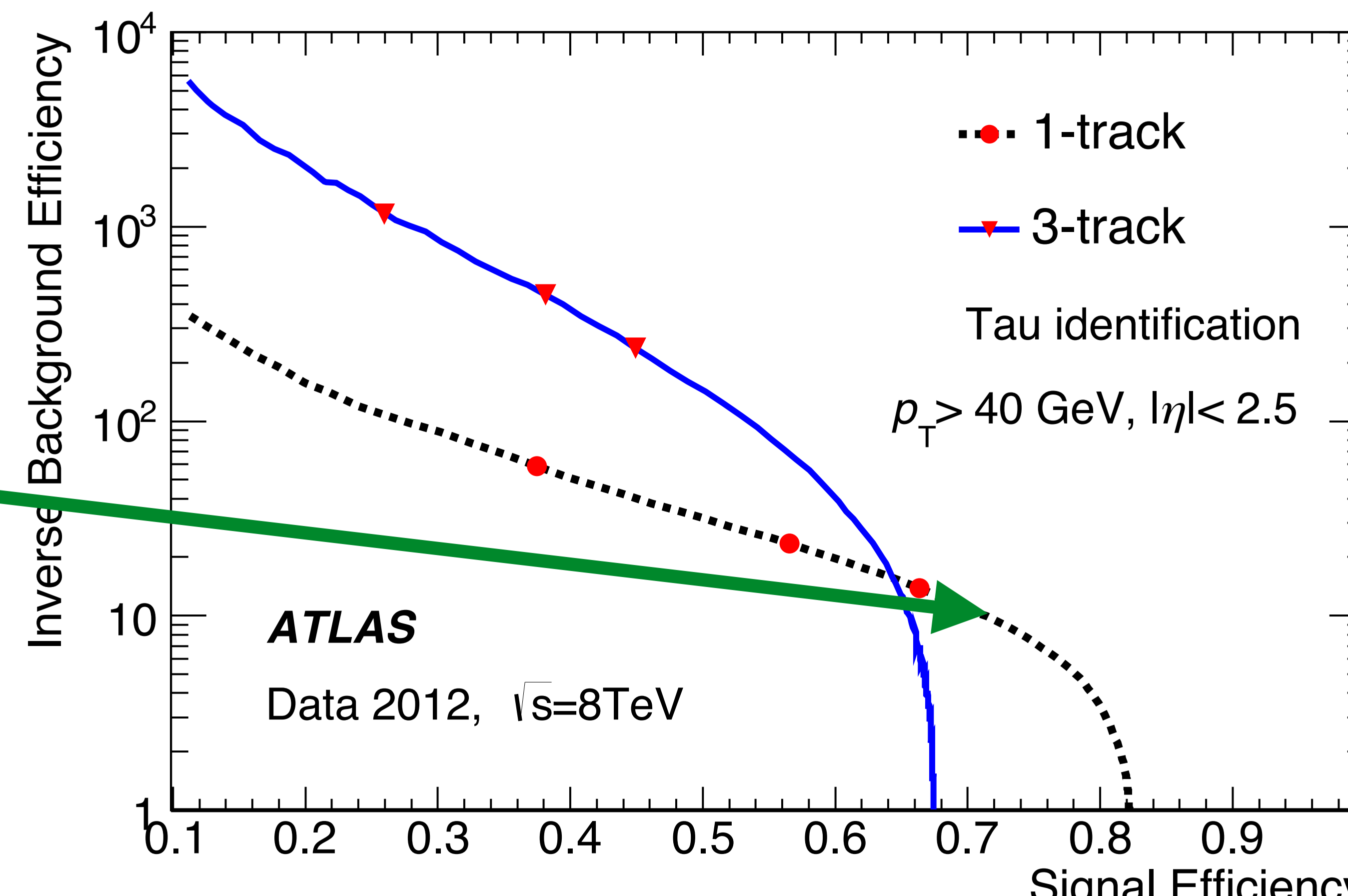
Signal-Background Discrimination/ROC

From MC



Signal efficiency: fraction of signal events right of the line

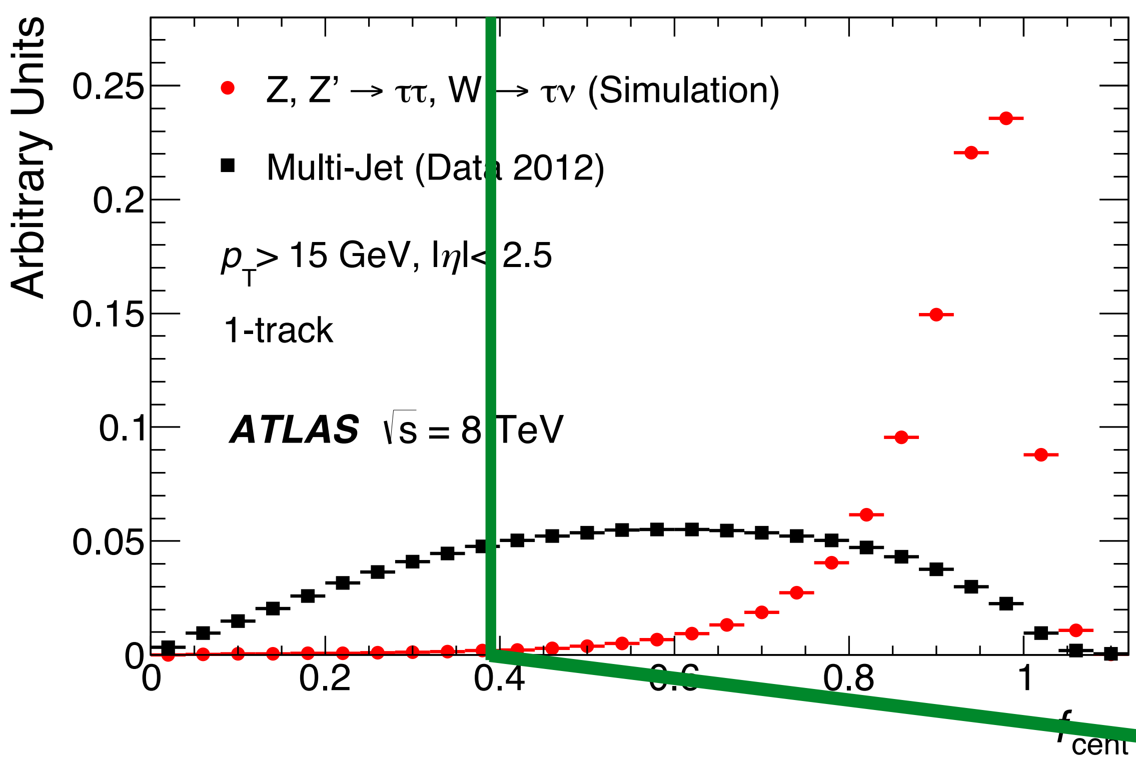
Inverse background efficiency: inverse fraction of background events right of the line



Scan over the full range

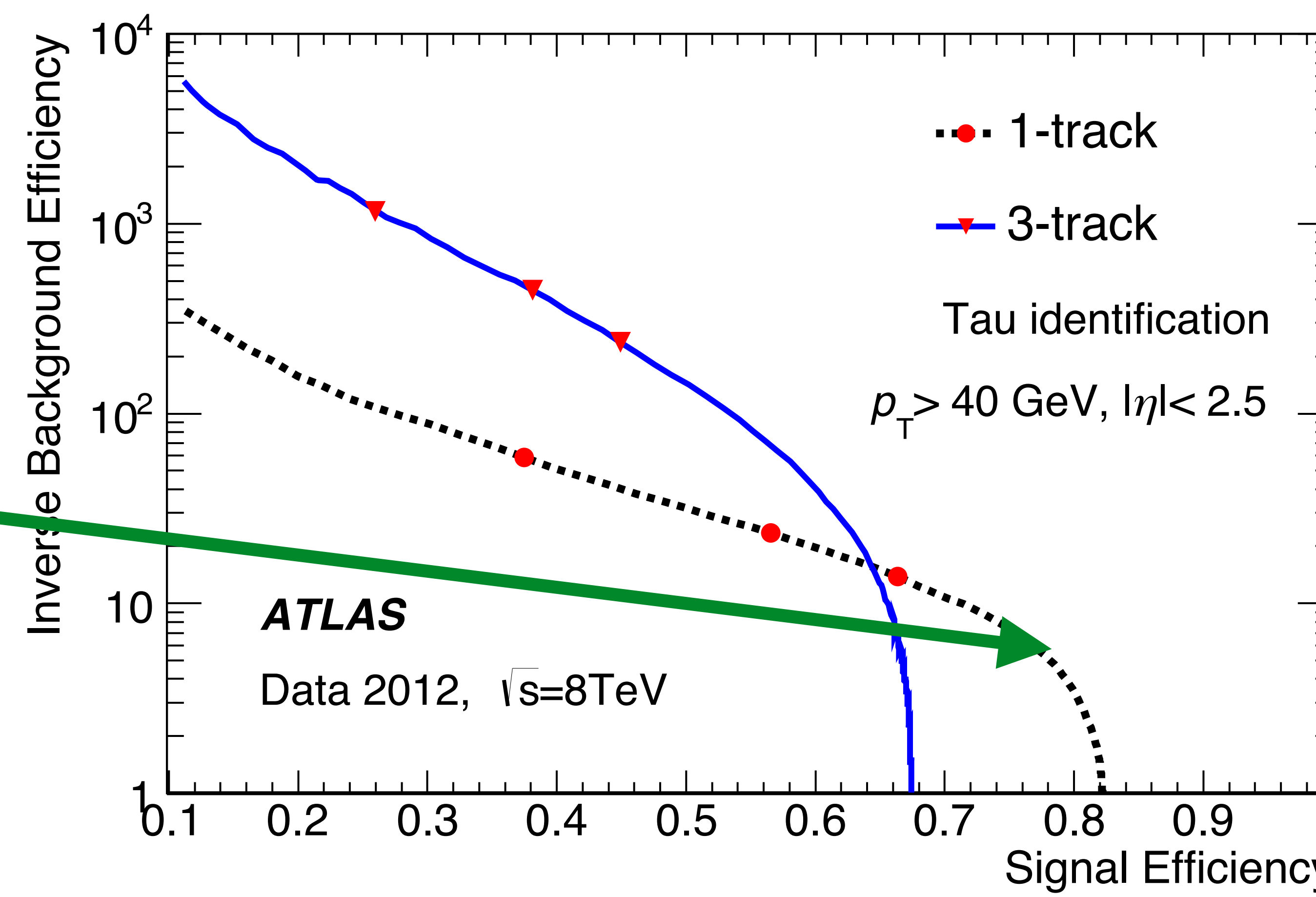
Signal-Background Discrimination/ROC

From MC



Signal efficiency: fraction of signal events right of the line

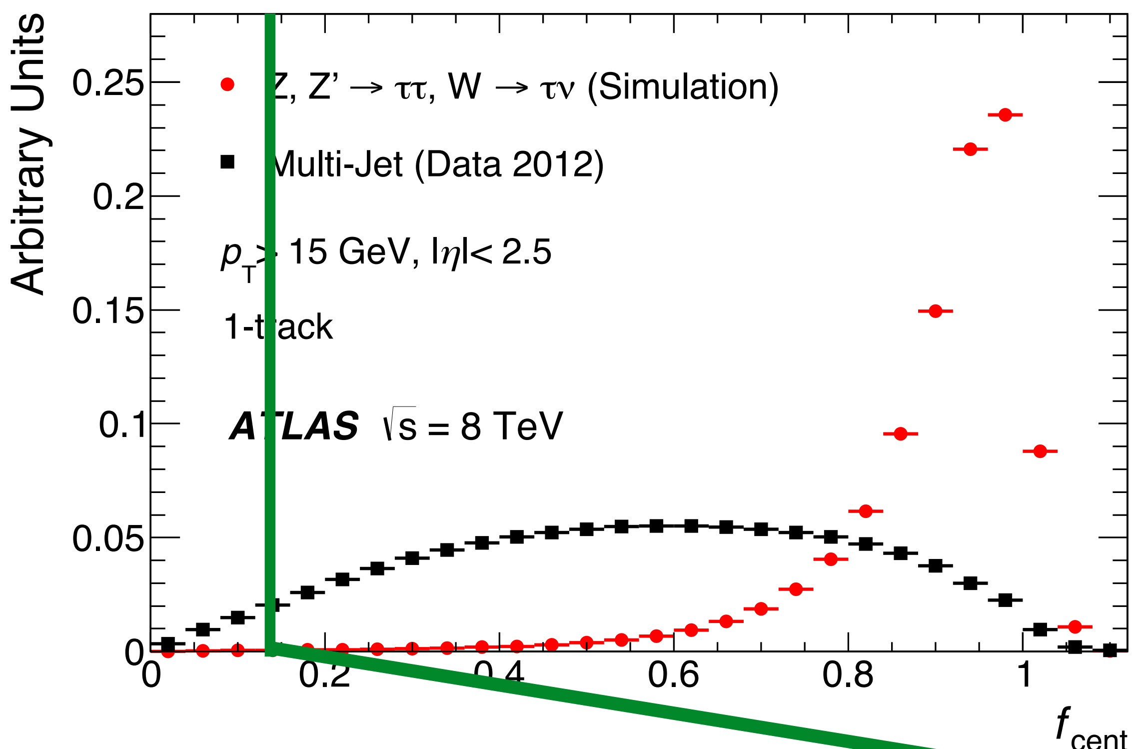
Inverse background efficiency: inverse fraction of background events right of the line



Scan over the full range

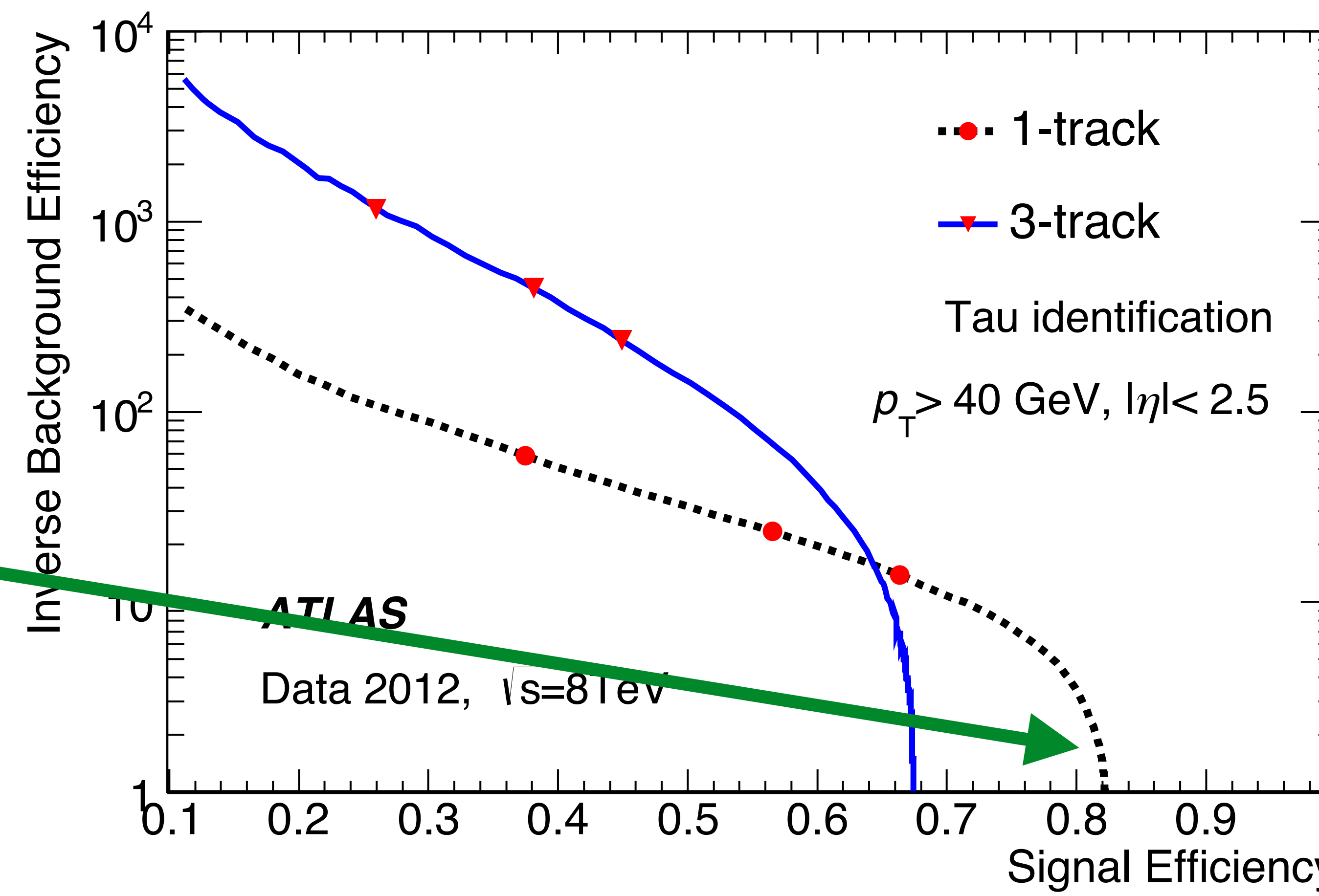
Signal-Background Discrimination/ROC

From MC



Signal efficiency: fraction of signal events right of the line

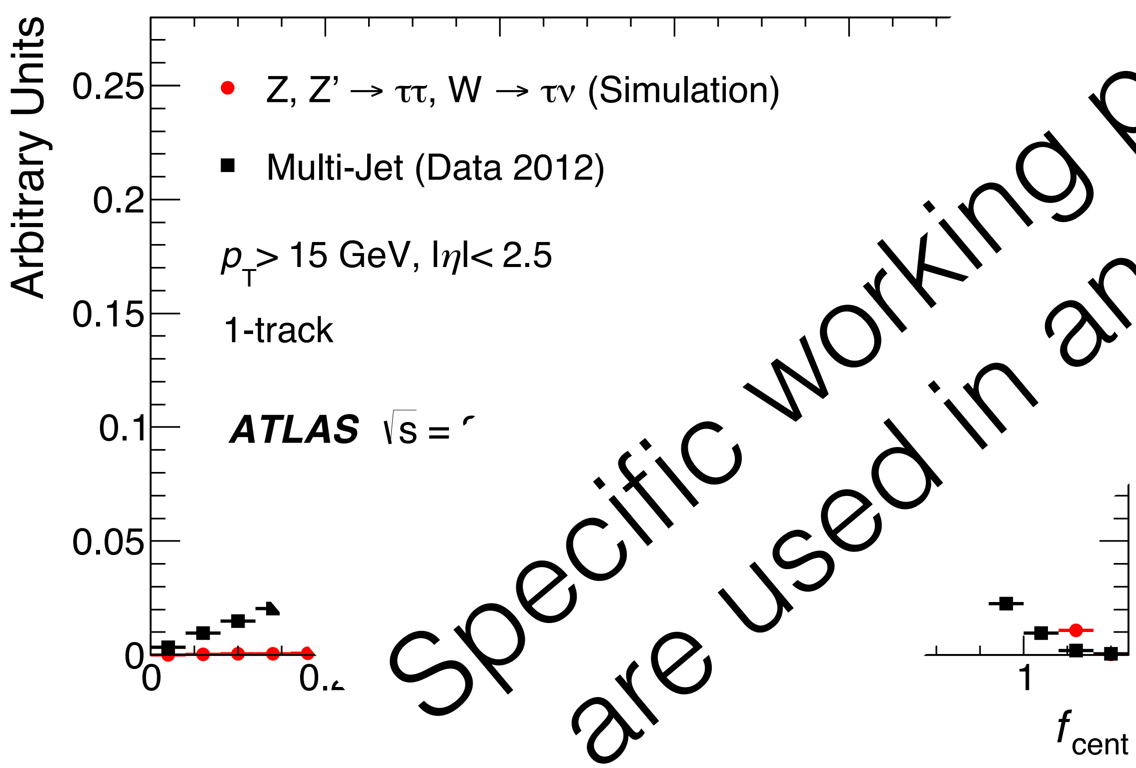
Inverse background efficiency: inverse fraction of background events right of the line



Scan over the full range

Signal-Background Discrimination/ROC

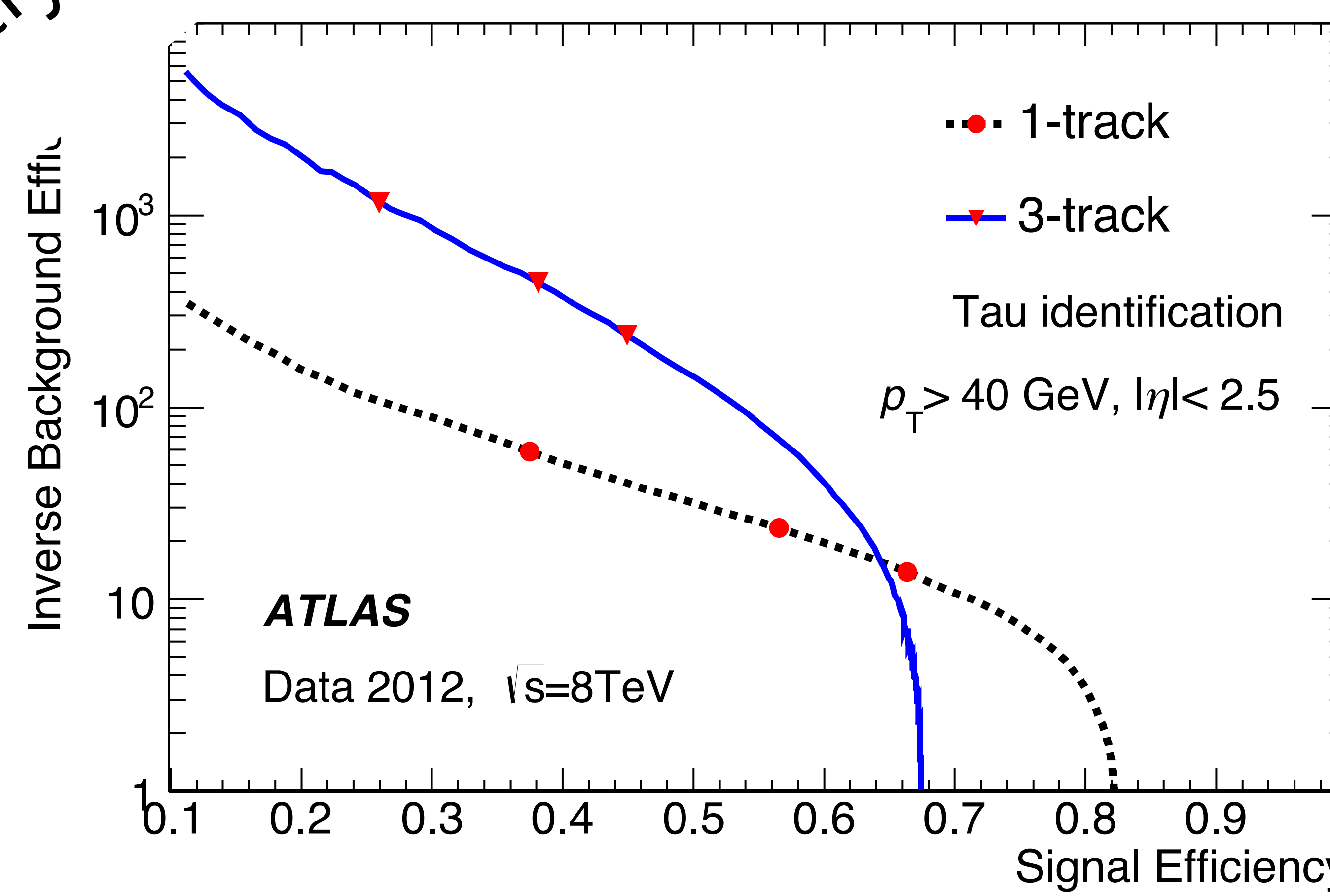
From MC



Specific working points (WP) are used in analysis/data

...iency: fraction of signal events ...right of the line

...e background efficiency: ...e fraction of background events right of the line

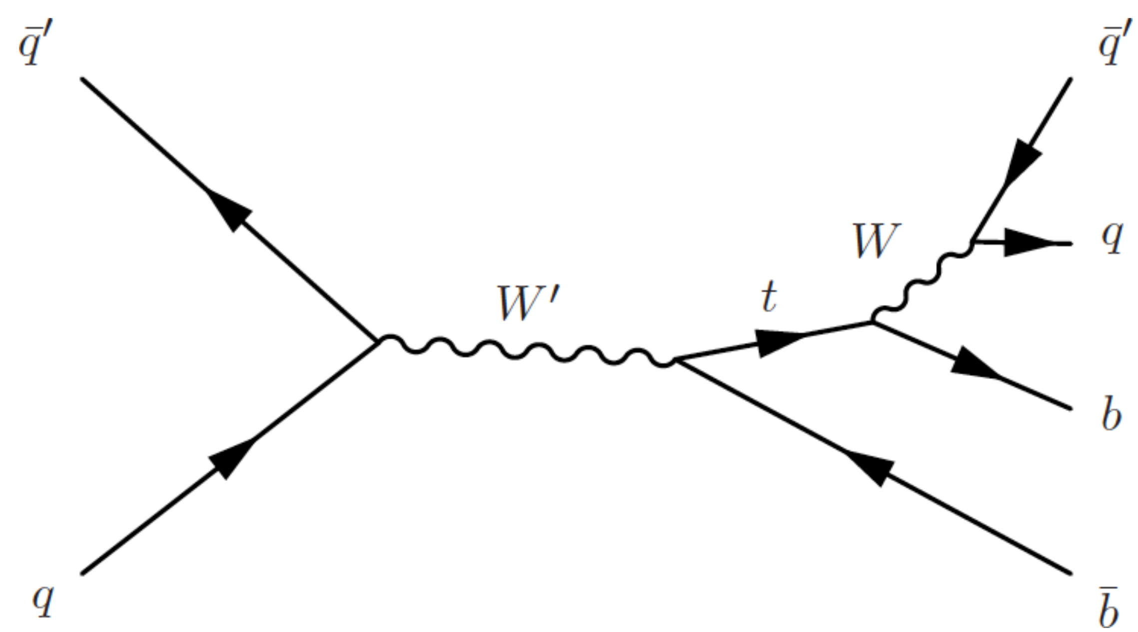


Scan ...er the full range

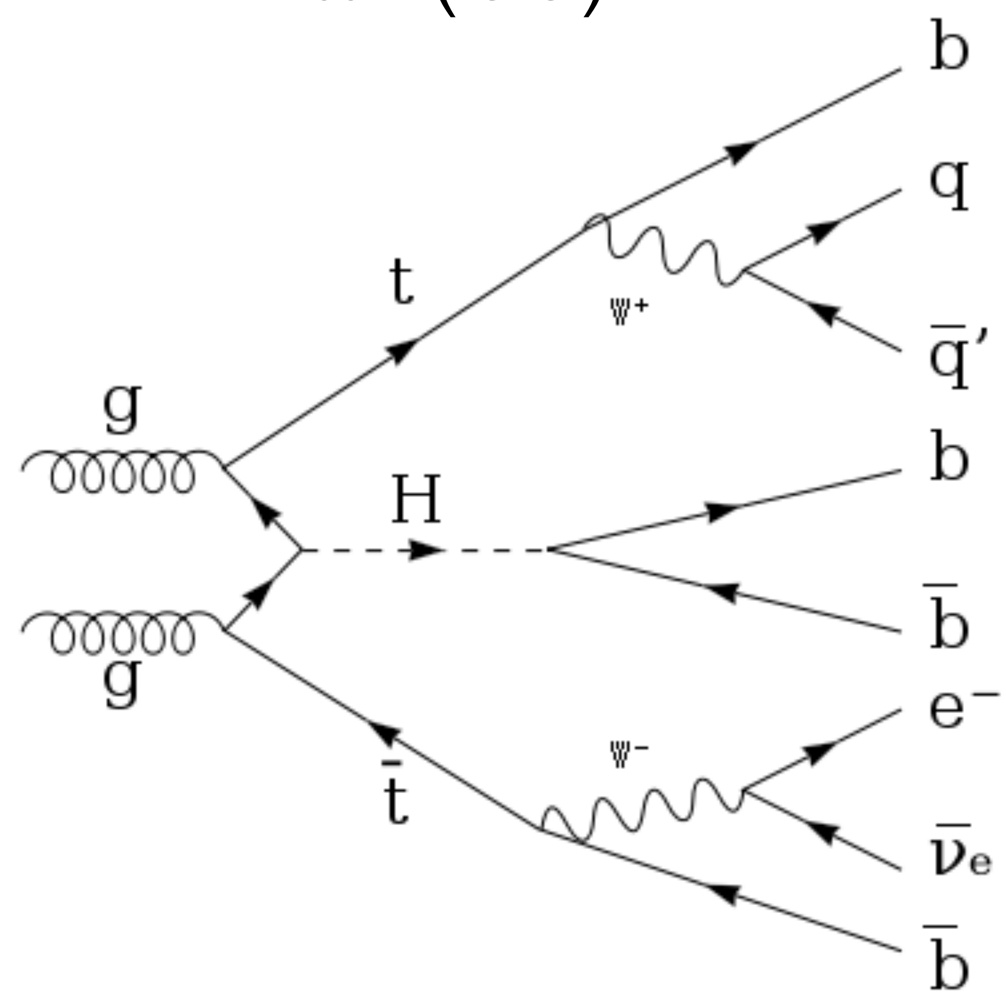
Jet Tagging

Jet Substructure

W prime



ttH(bb)



Motivation

- Large number of jets in final state, a combinatorial nightmare
- Large multijet background

Any way out?

Motivation

- For a two body decay, we can write:

$$p_{T_1} = z p_T$$
$$p_{T_2} = (1-z) p_T,$$

- For a (quasi)collinear splitting:

$$M^2 = (E_1 + E_2)^2 - (\vec{p_{T_1}} + \vec{p_{T_2}}) \times (\vec{p_{T_1}} + \vec{p_{T_2}}) = 2\vec{p_{T_1}}\vec{p_{T_2}}$$

- In terms of the angular separation:

$$M^2 = p_{T_1} p_{T_2} \Delta R^2 = z(z-1) p_T^2 \Delta R^2$$
$$\Delta R^2 = 4M^2 / p_T^2$$

Quiz

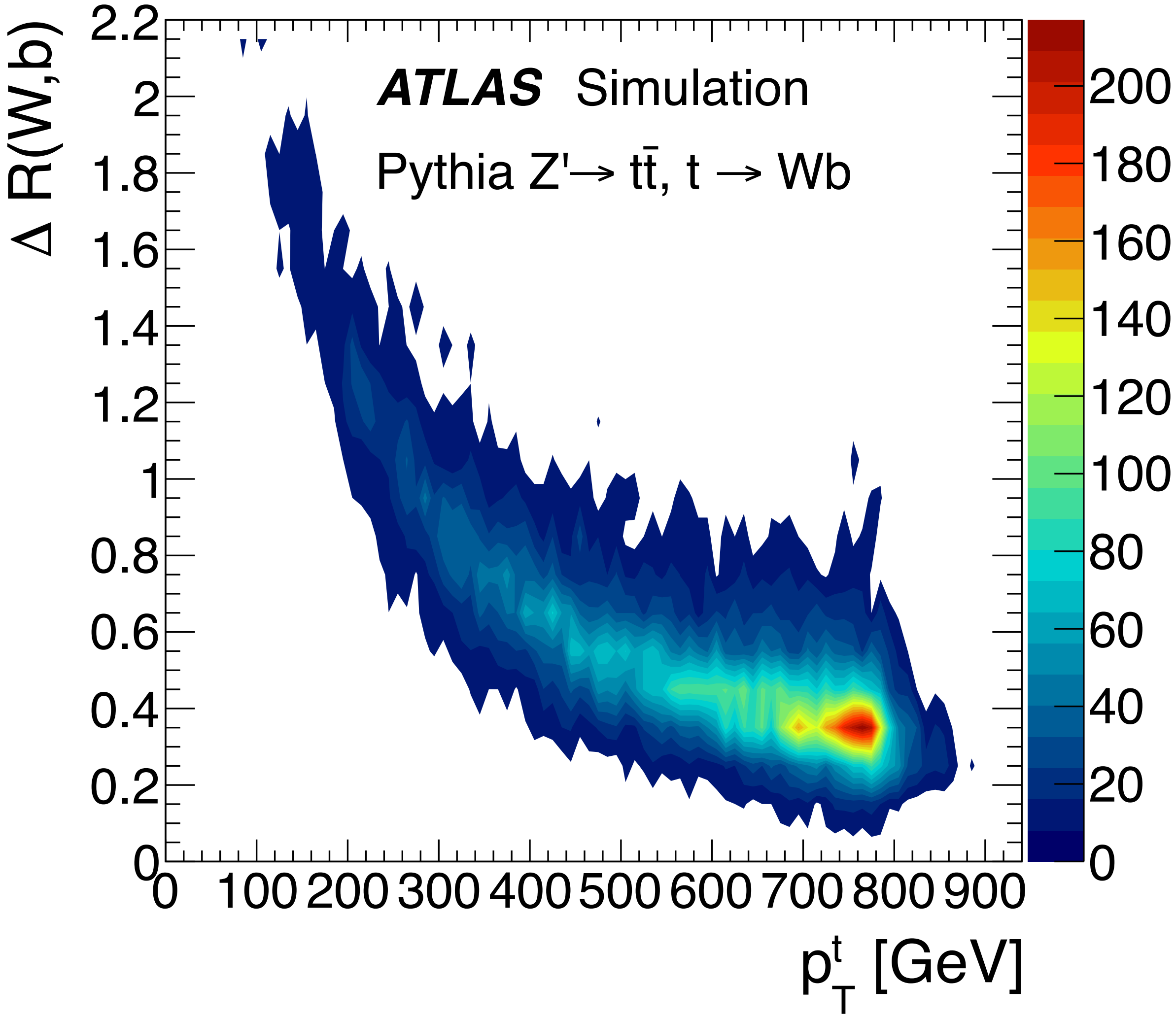
Jet constituents are almost massless, then how can jets have mass?

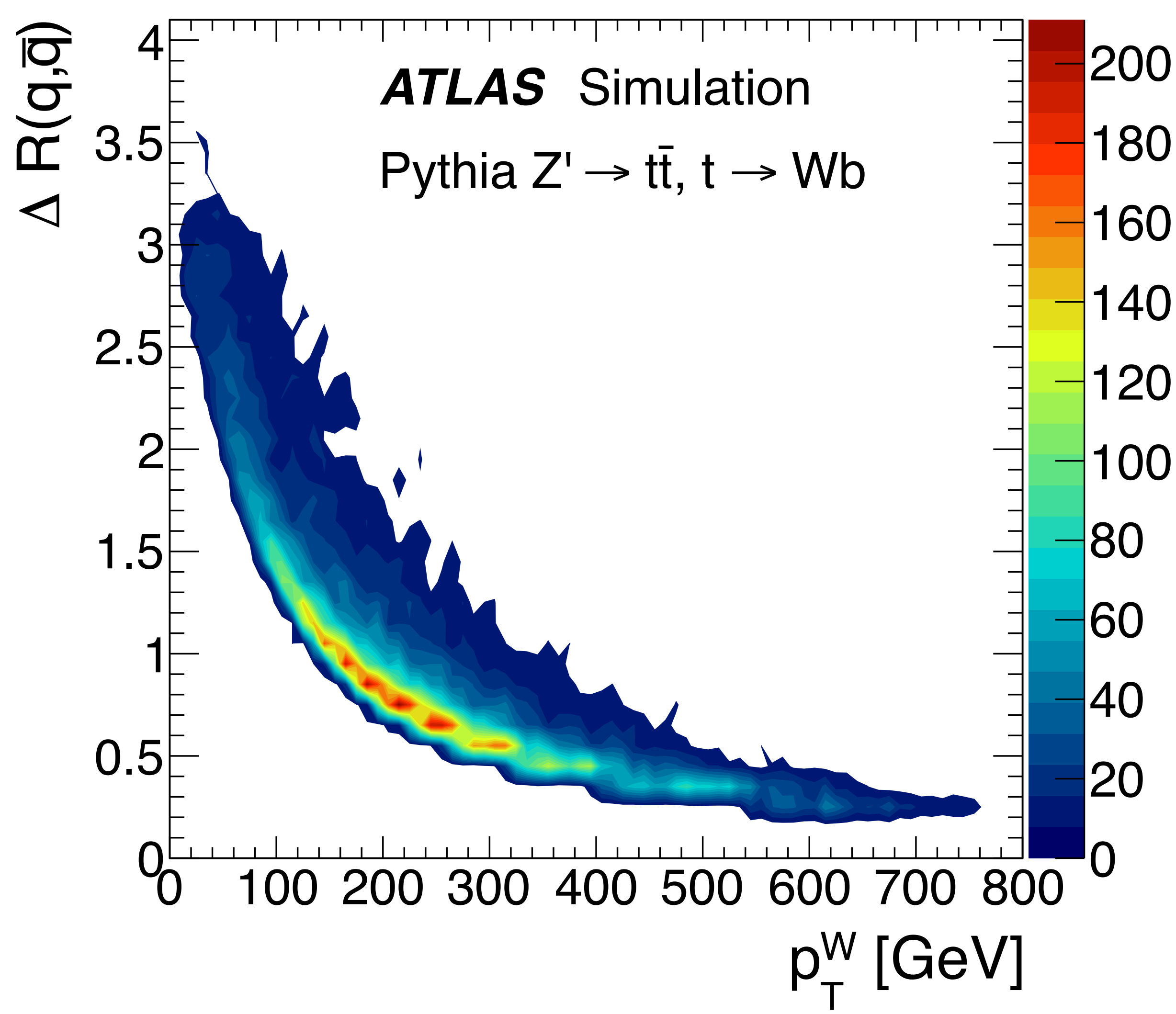
Jet mass

Jet mass: the constituents of a jet usually have zero or very small mass, however the jet itself often ends up having a non-negligible mass, as the jet mass can be calculated as:

$$m = \sqrt{(\sum_{i \in \text{jet}} E_i)^2 - (\sum_{i \in \text{jet}} \vec{p}_i)^2},$$

Motivation





A large radius jet of $R = 2m/p_T$^2 can contain all decay products

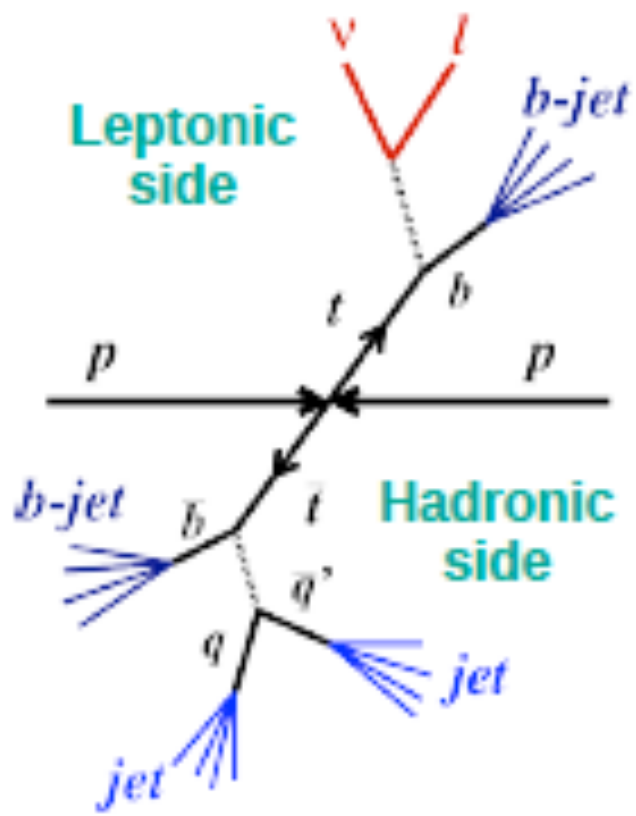
Boost!

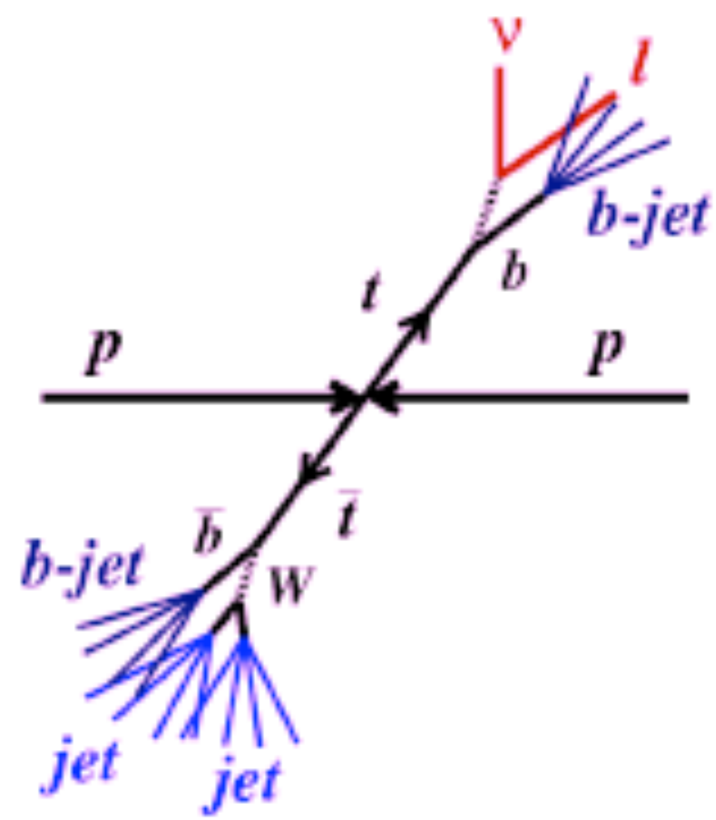
High p_T means transverse Lorentz boost!

More probable with higher LHC energy

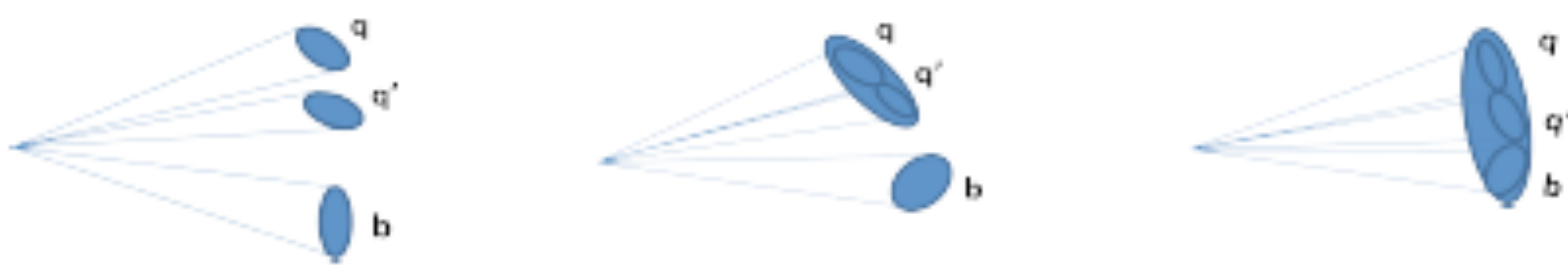
Hadronically decaying top quark, Higgs/W/Z bosons, new heavy particles …

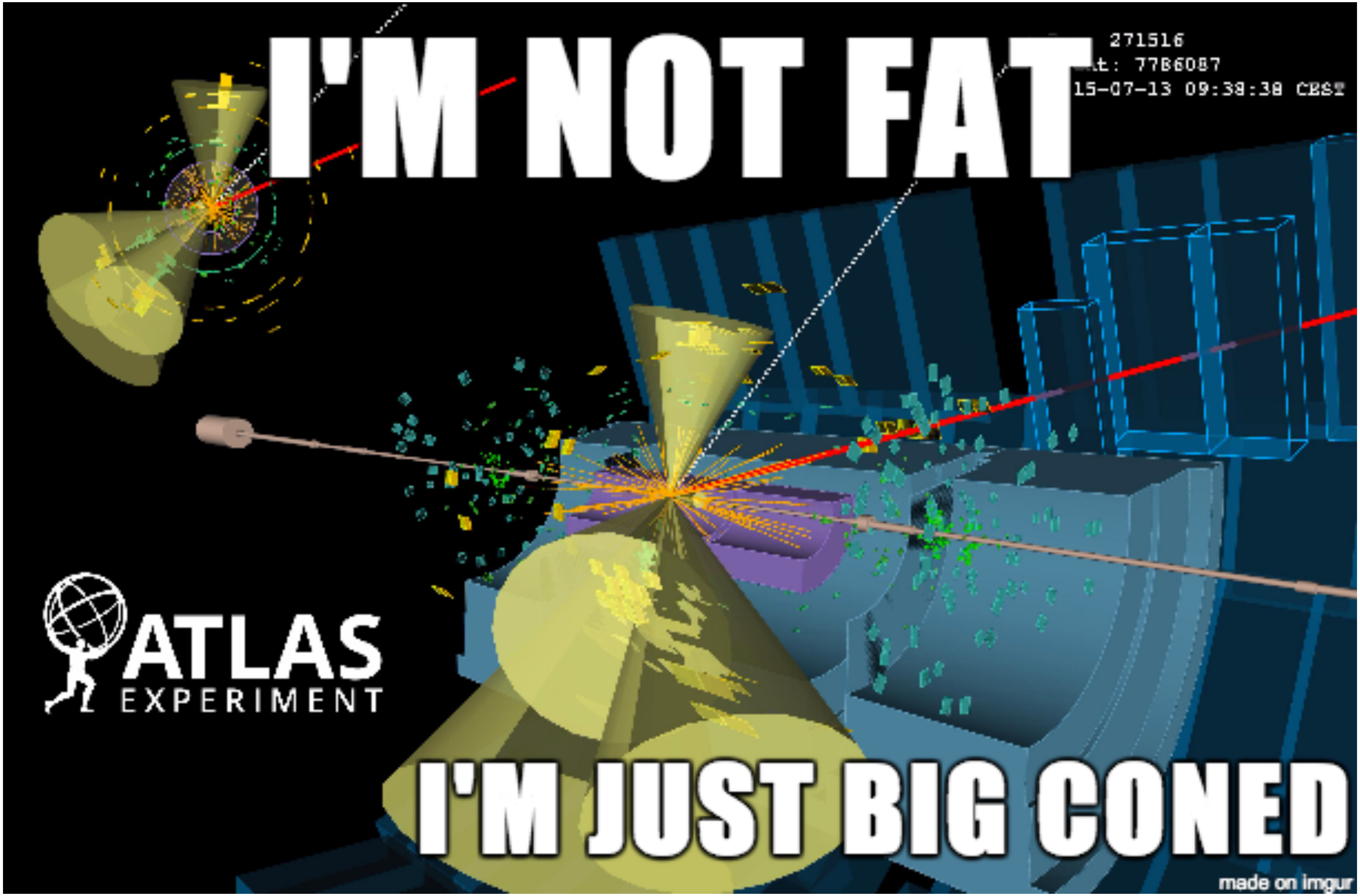
Boost





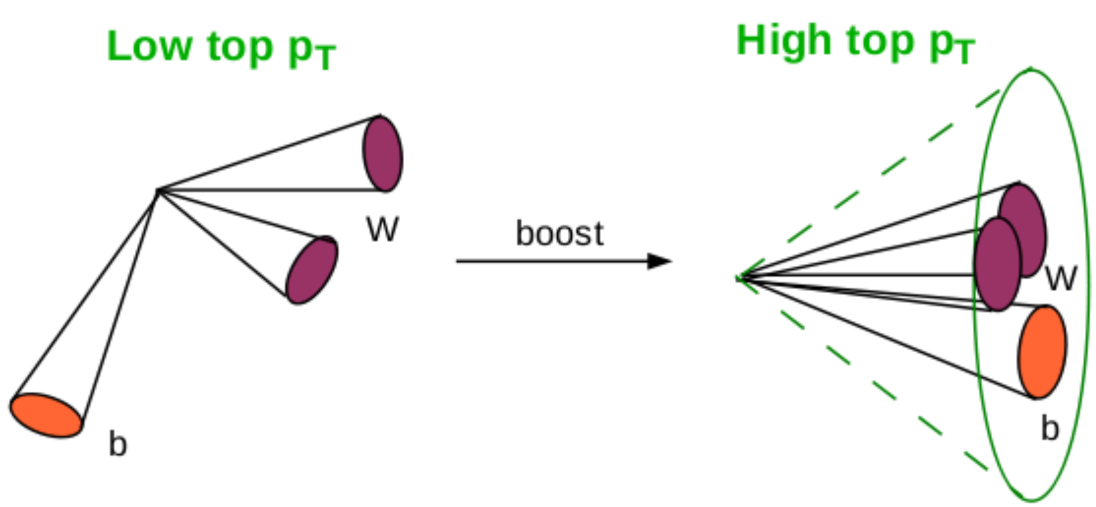
Hadronic side




I'M NOT FAT
271516
15-07-13 09:38:38 CEST
ATLAS
EXPERIMENT
I'M JUST BIG CONED
made on imgur

Signal vs Background

The boosted jet coming from top quark (hadronic) decay should be distinguishable from the boosted jet coming from events with no top quarks.



We want to exploit the "substructure" of the large-radius jet to identify original particles

Background: light quark/gluon/lepton jets

Quiz

Can you think of any possible disadvantage of using large-radius jets?

First Application

- The so-called BDRS paper (Butterworth-Davison-Rubin-Salam, 2008)
- Looked at VH (V=W/Z), with H decaying to bbbar.
- Charged leptons in the final state help to reduce background.
- Conventional methods result in Hbb swamped by multijet background.

BDRS

- Start with fat (C-A 1.2) boosted ($p_T > 200$) b-tagged jet.

- 1. Break the jet j into two subjets by undoing its last stage of clustering. Label the two subjets j_1, j_2 such that $m_{j_1} > m_{j_2}$.

- 2. If there was a significant mass drop (MD), $m_{j_1} < \mu m_j$, and the splitting is not too asymmetric, $y = \frac{\min(p_{tj_1}^2, p_{tj_2}^2)}{m_j^2} \Delta R_{j_1,j_2}^2 > y_{\mathrm{cut}}$, then deem j to be the heavy-particle neighbourhood and exit the loop. Note that $y \simeq \min(p_{tj_1}, p_{tj_2}) / \max(p_{tj_1}, p_{tj_2})$.[1]

3. Otherwise redefine j to be equal to j_1 and go back to step 1.

$$\mu = 0.67 \qquad y_{cut} = 0.09$$

BDRS

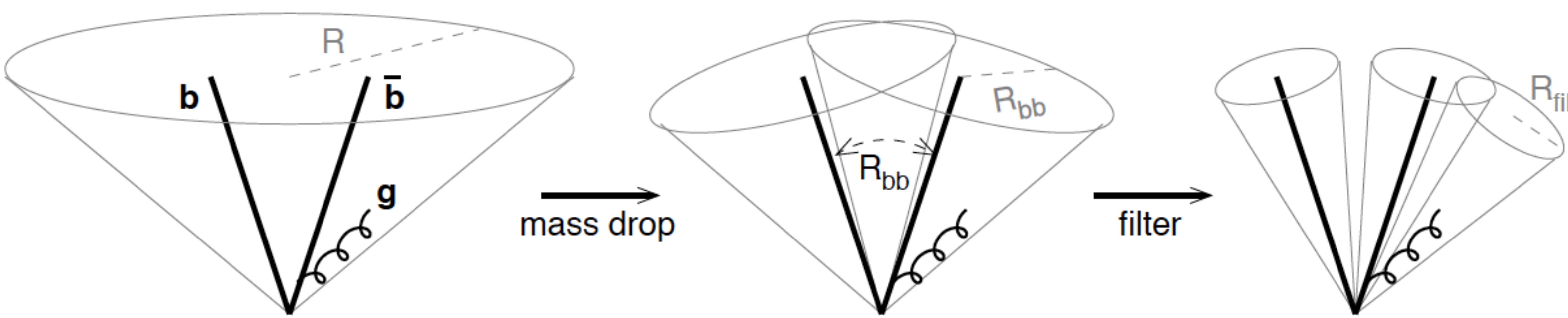
- Start with fat (C-A 1.2) boosted ($p_T > 200$) b-tagged jet.
-
-

$$R_{\text{filt}} < R_{b\bar{b}}$$

$$R_{\text{filt}} = \min(0.3, R_{b\bar{b}}/2)$$

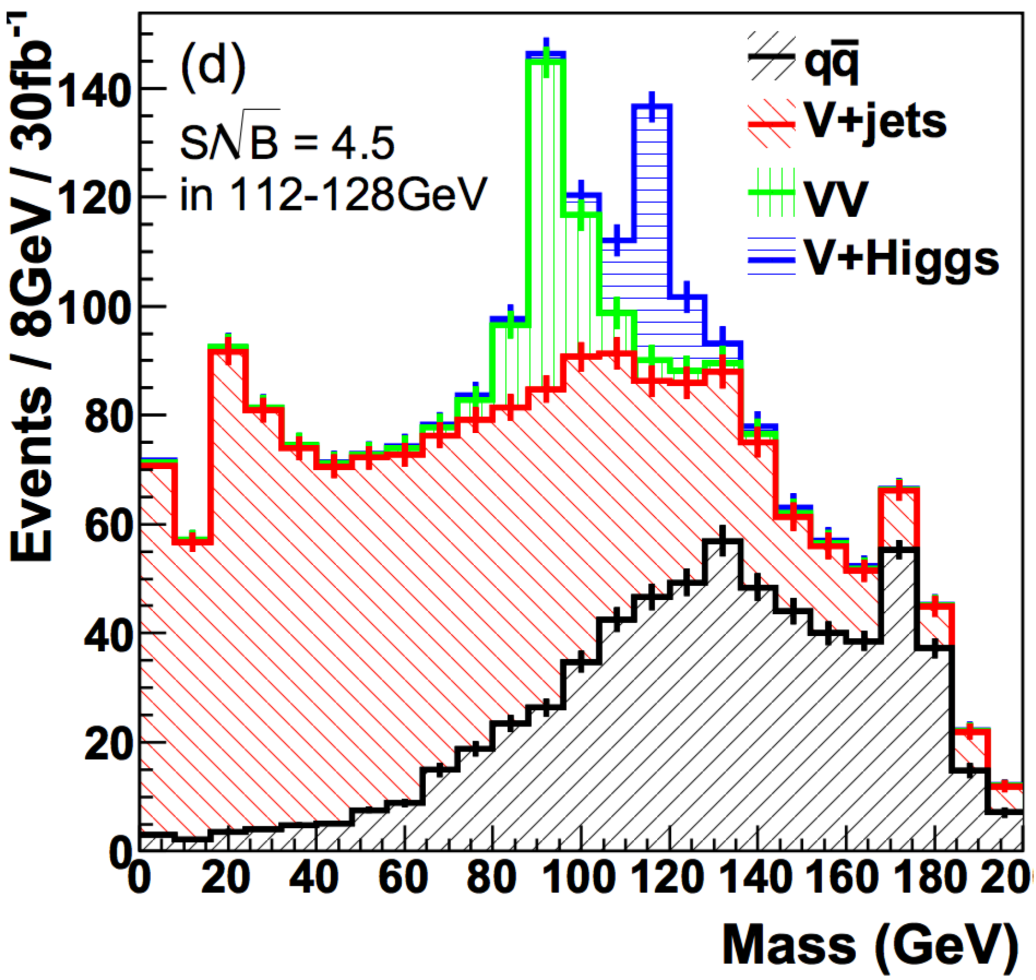
BDRS

- Start with fat (C-A 1.2) boosted ($p_T > 200$) b-tagged jet.
- De-cluster the jet. At each stage, mass drop and symmetric splitting requirement.
- Continue till an interesting splitting has been found.



BDRS

- Then use the three hardest b-tagged subjets to (re)form the large-radius jet.
- The mass of the large-radius jet is Higgs boson mass.
- For a 115 GeV Higgs, 4.5 σ for 30 fb^{-1}.



Lessons

- We need to clean the jet.
- We need observables (like jet mass) to discriminate signal against background.
- BDRS technique did not work in ATLAS/CMS directly.

Jet Grooming

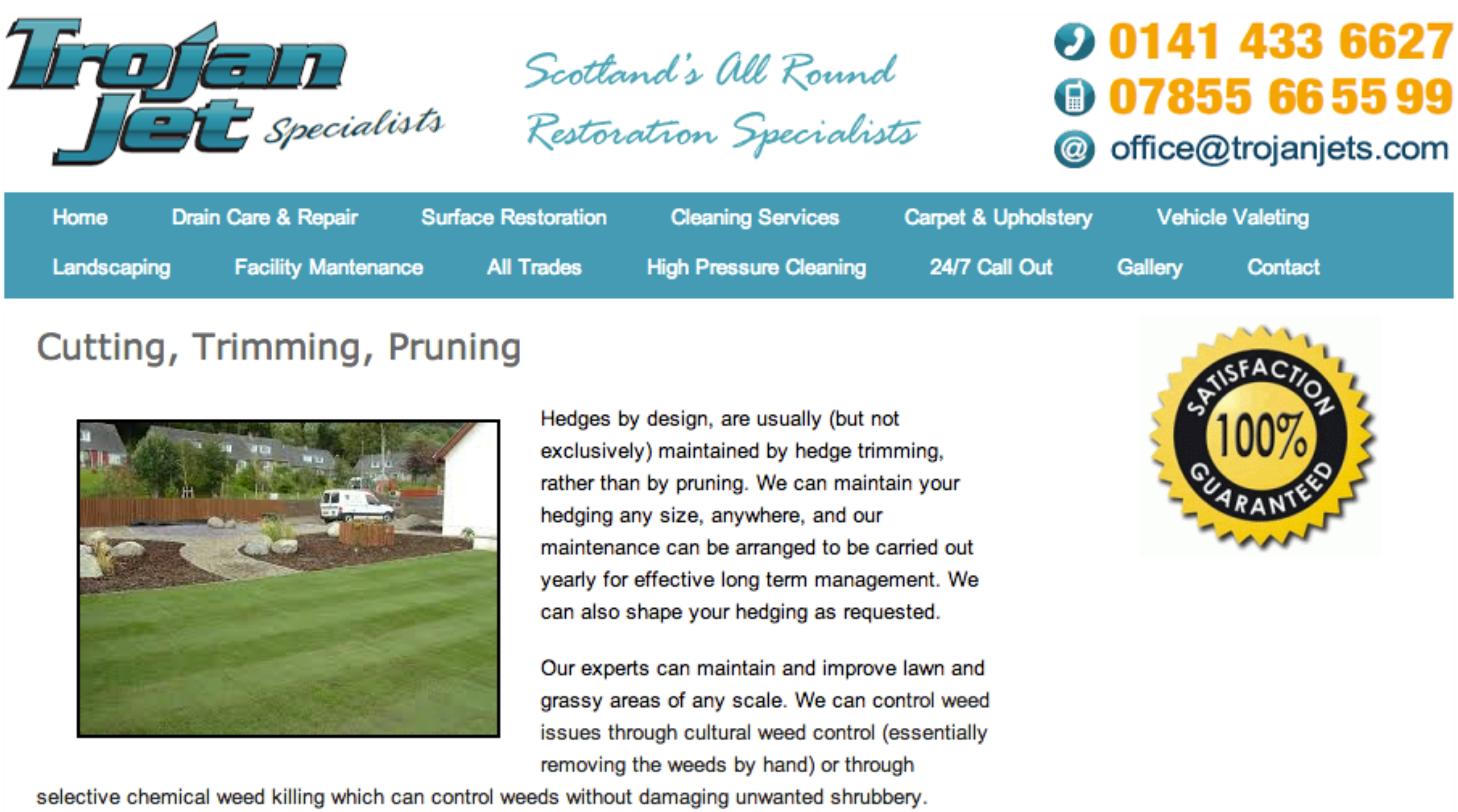
- Jets need to be “groomed”.
- Need observables which would be sensitive to signal-like or background-like nature of these jets.

Why?

The large-radius jets not only include particles coming from the interesting decays, but also from pileup, underlying event ....

Jet Grooming

- Jets need to be "groomed".



Jet Grooming

- Mass drop filtering
- Pruning
- Trimming
- Soft Drop

Soft Drop

Start with a jet j and it is split into last two subjets

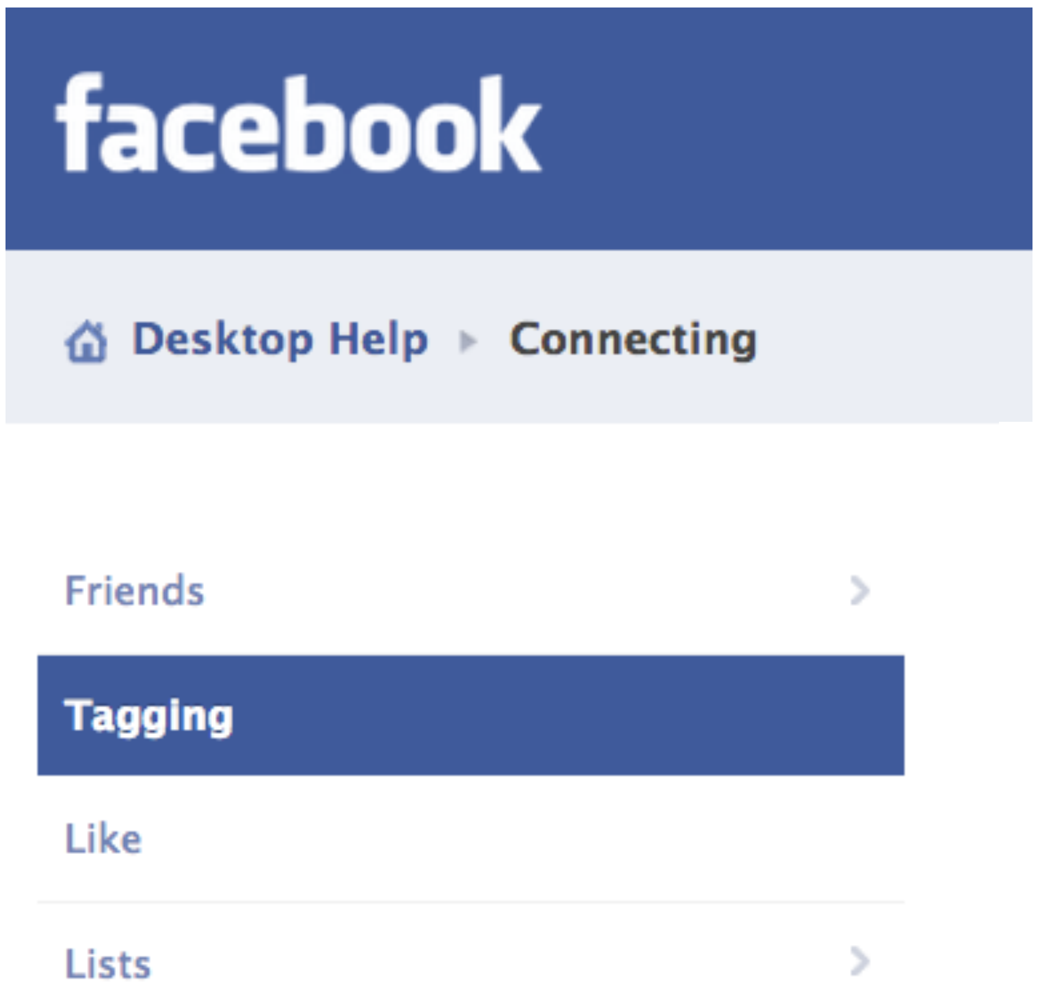
If: $$\frac{min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{cut} \left(\frac{\Delta R_{12}}{R_0}\right)^{\beta}$$

Then j is the final soft drop jet.

Otherwise the higher p_T subjet is taken as j and iterated …

Advantage: can be compared directly to analytic calculations

Tagging boosted objects: observables and taggers



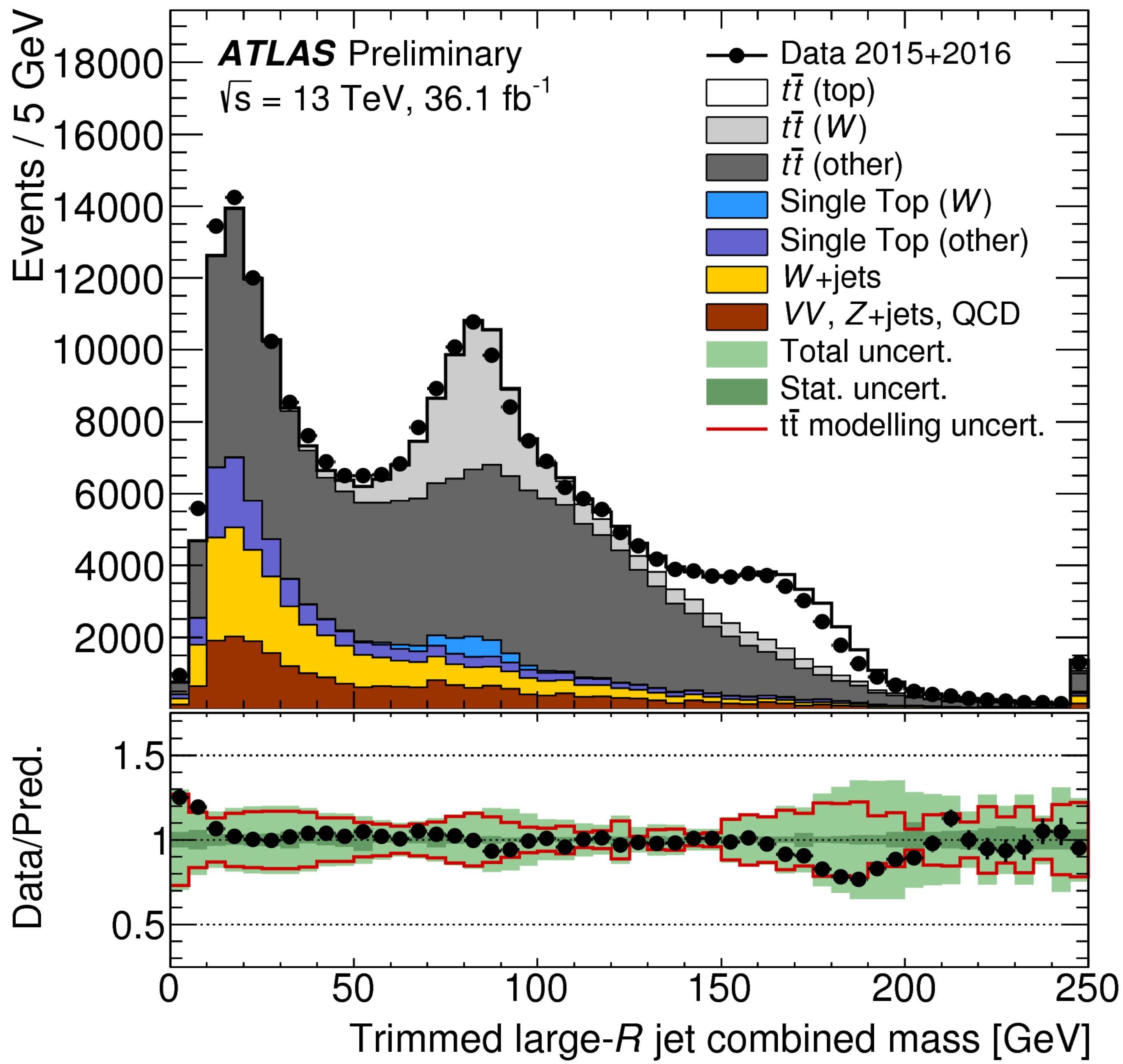
particles!

Tag ~~people~~ in your posts

Add tags to anything you post, including photos and updates. Tags can point to your friends or anyone else on Facebook. Adding a tag creates a link that people can follow to learn more.

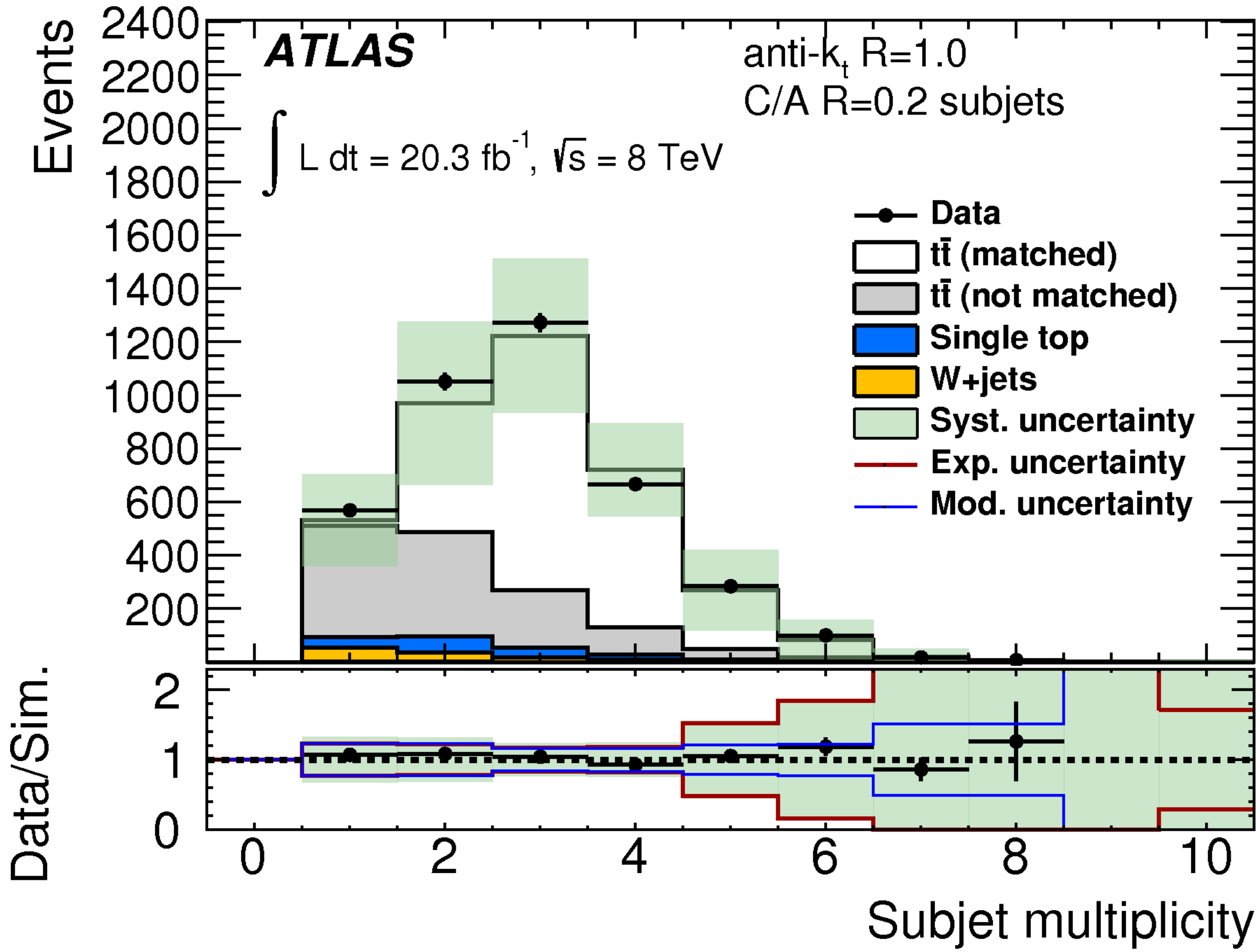
Target is to identify jets resulting from the decay of top quark or Higgs against jets coming from light quark/gluons.

Mass



Mass peaks clearly visible over background!

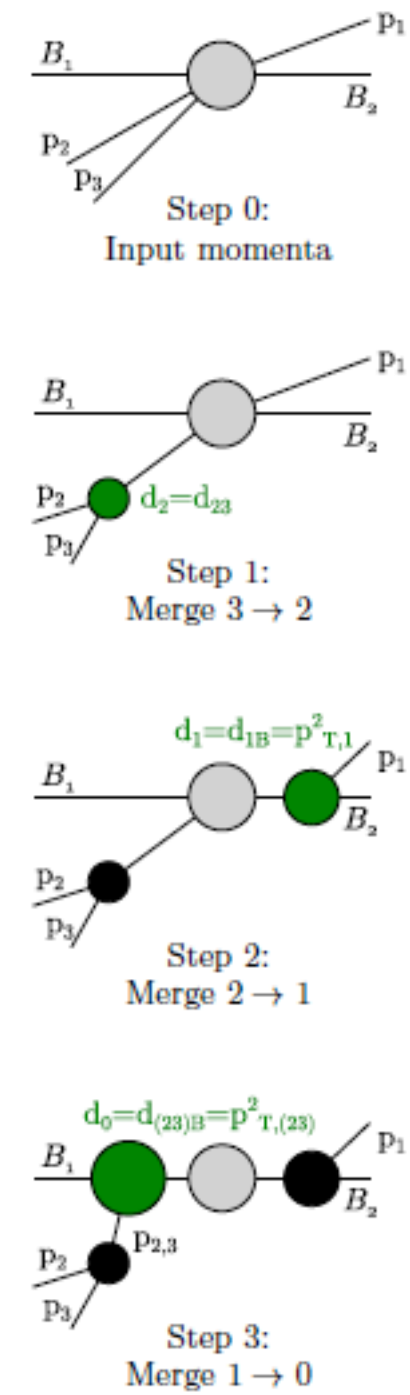
Subjet multiplicity

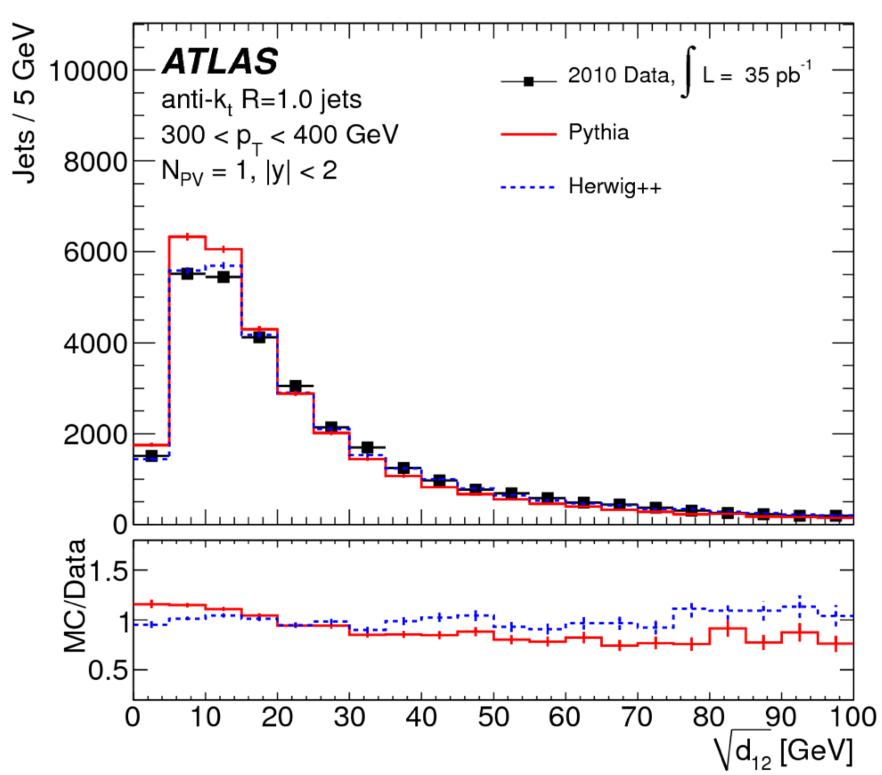


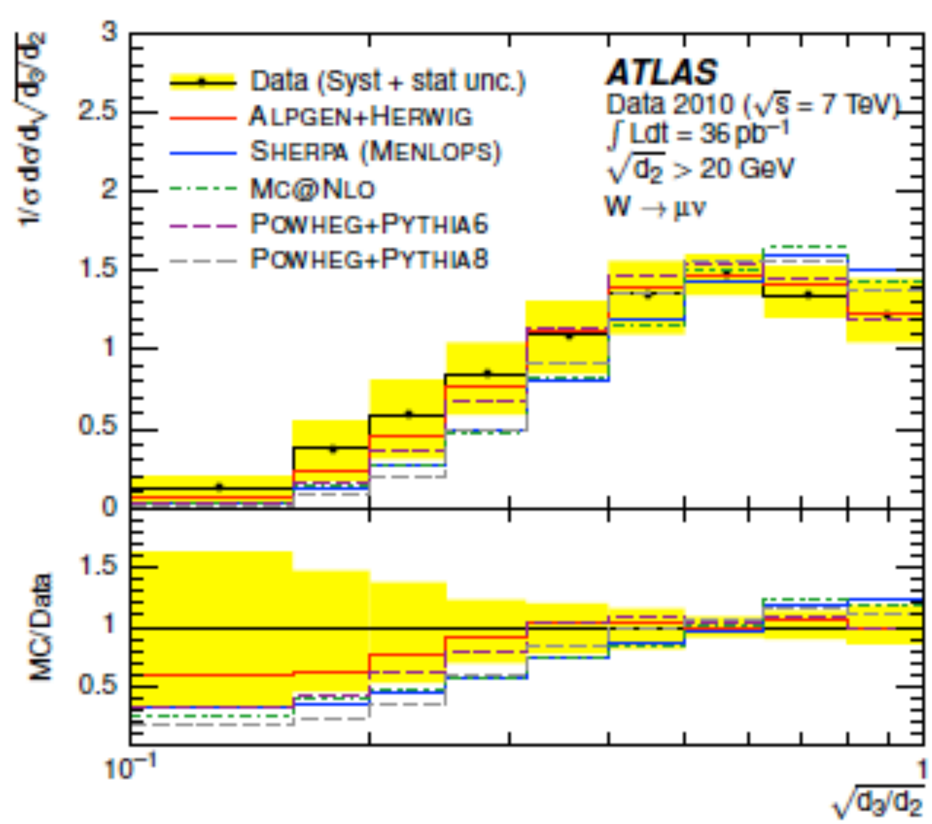
k_t splitting scale

$$\sqrt{d_{ij}} = \min(p_{Ti}, p_{Tj}) \times \Delta R_{ij}$$

When combining two subjets with k_t algorithm:







Symmetric for heavy particle two body decay

N-Subjettiness

Quantify the degree to which jet radiation is aligned along specific subjet axes.

$$\tau_N \equiv \frac{1}{d_0} \sum_{k=1}^{M} \left(p_{\mathrm{T},k} \times \underbrace{\Delta R_{\min,k}}_{\text{distance to nearest subjet}} \right)$$

$d_0 = R \times$ sum of p_T of all constituents

Smaller values: N or less energy deposits

Larger values: more than N energy deposit

$\tau_{N-1} > \tau_N$ for N prong substructure

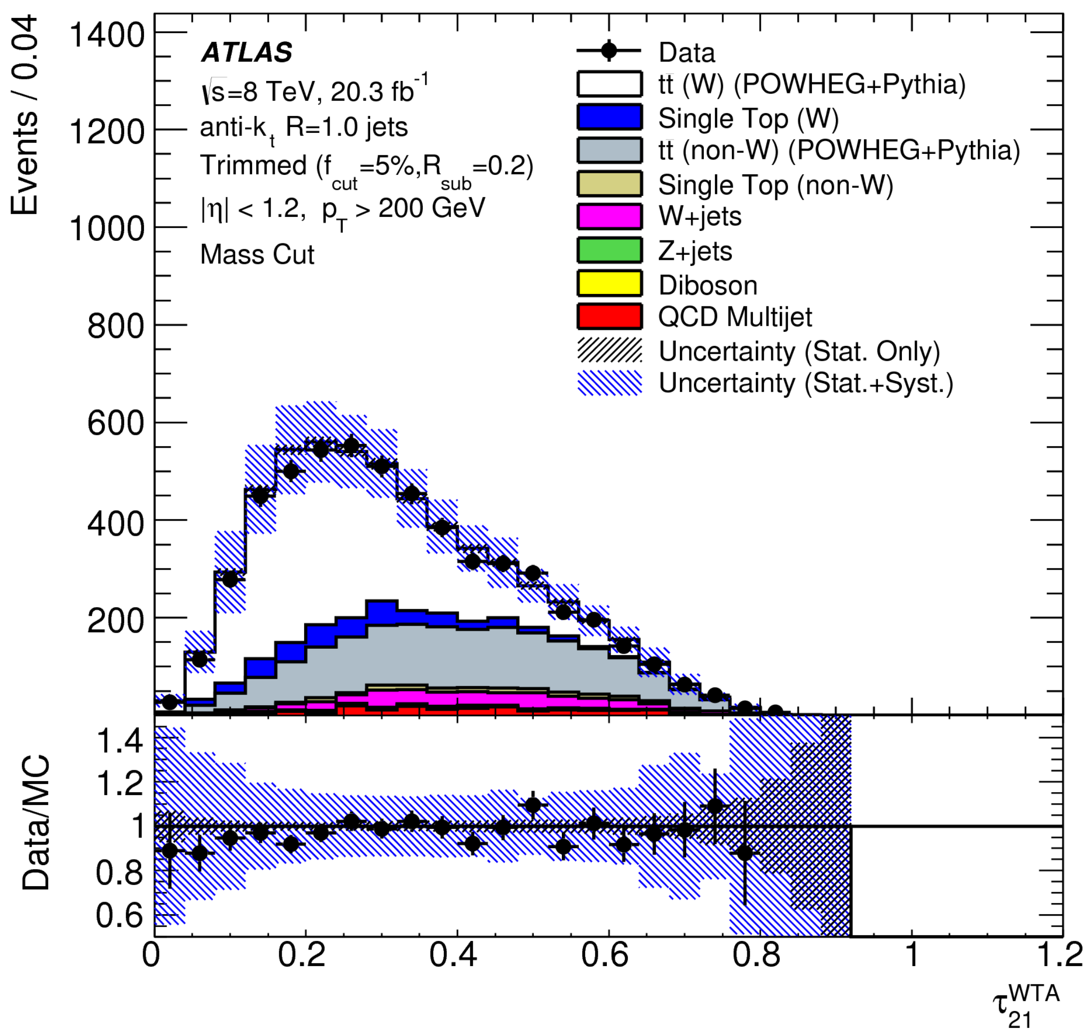
Calculated by k_t clustering the constituents, and requiring exactly N subjets

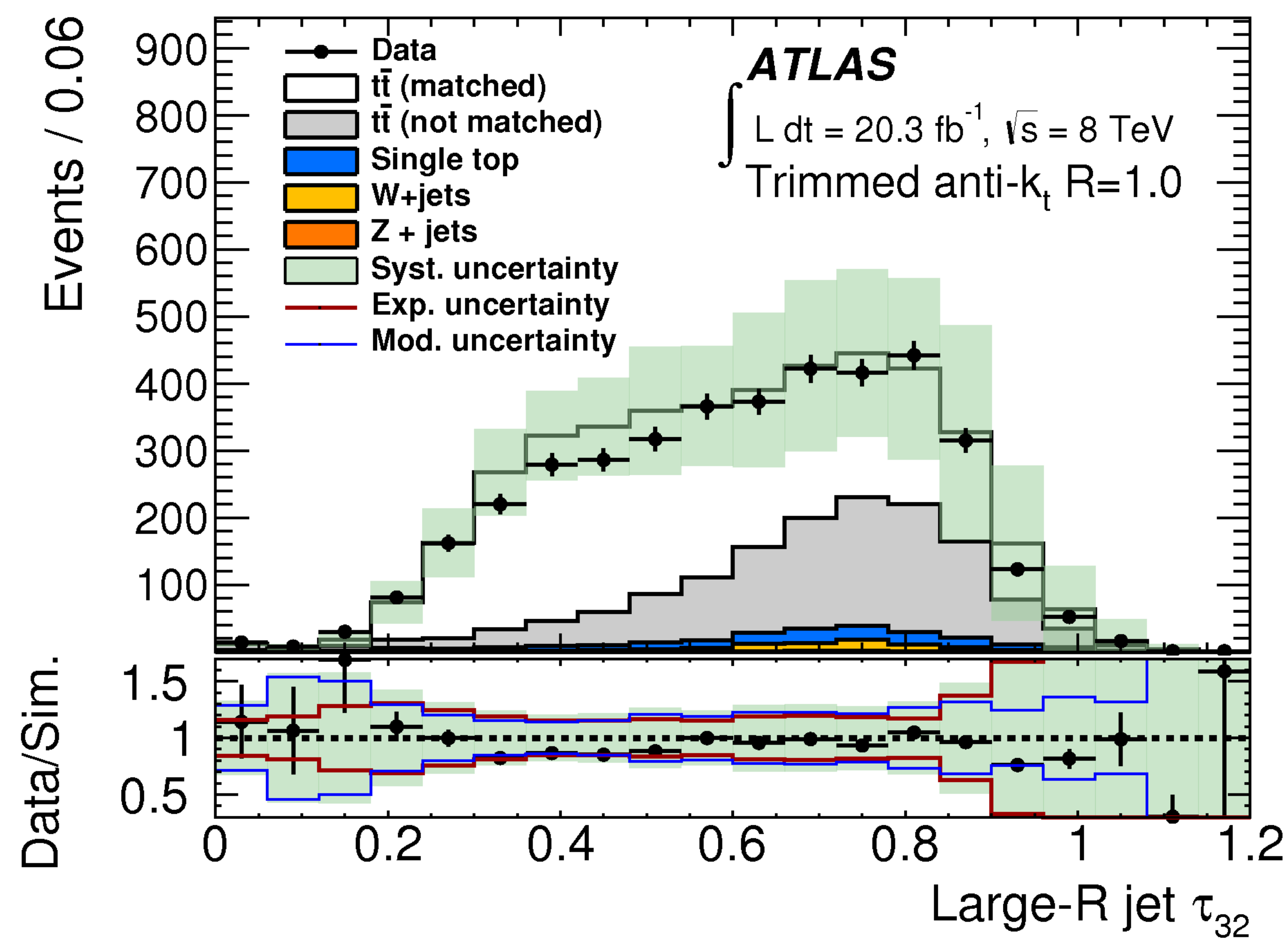
N-Subjettiness

The ratio of τ_N/τ_{N-1} is used as discriminant

More like 2 subjets than 1

More like 3 subjets than 2





W-like ⟷ MJ-like

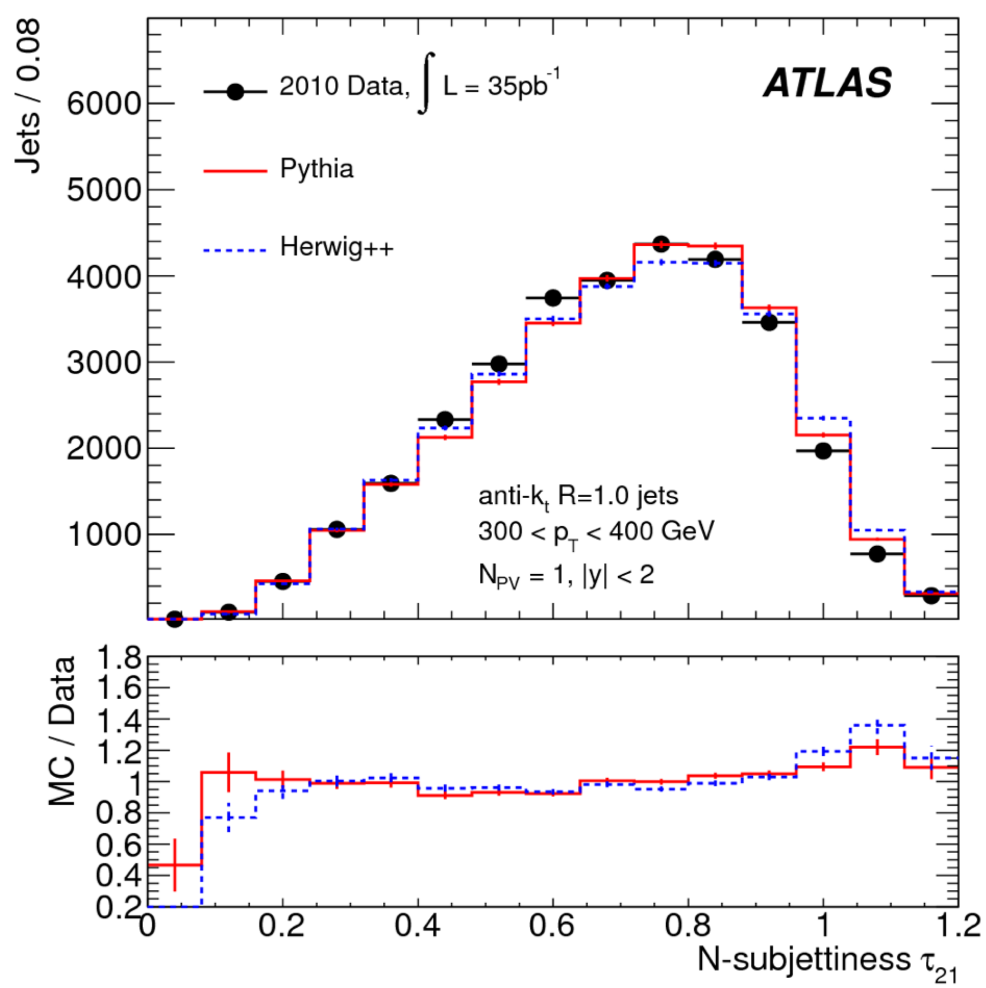
Top-like ⟷ MJ-like

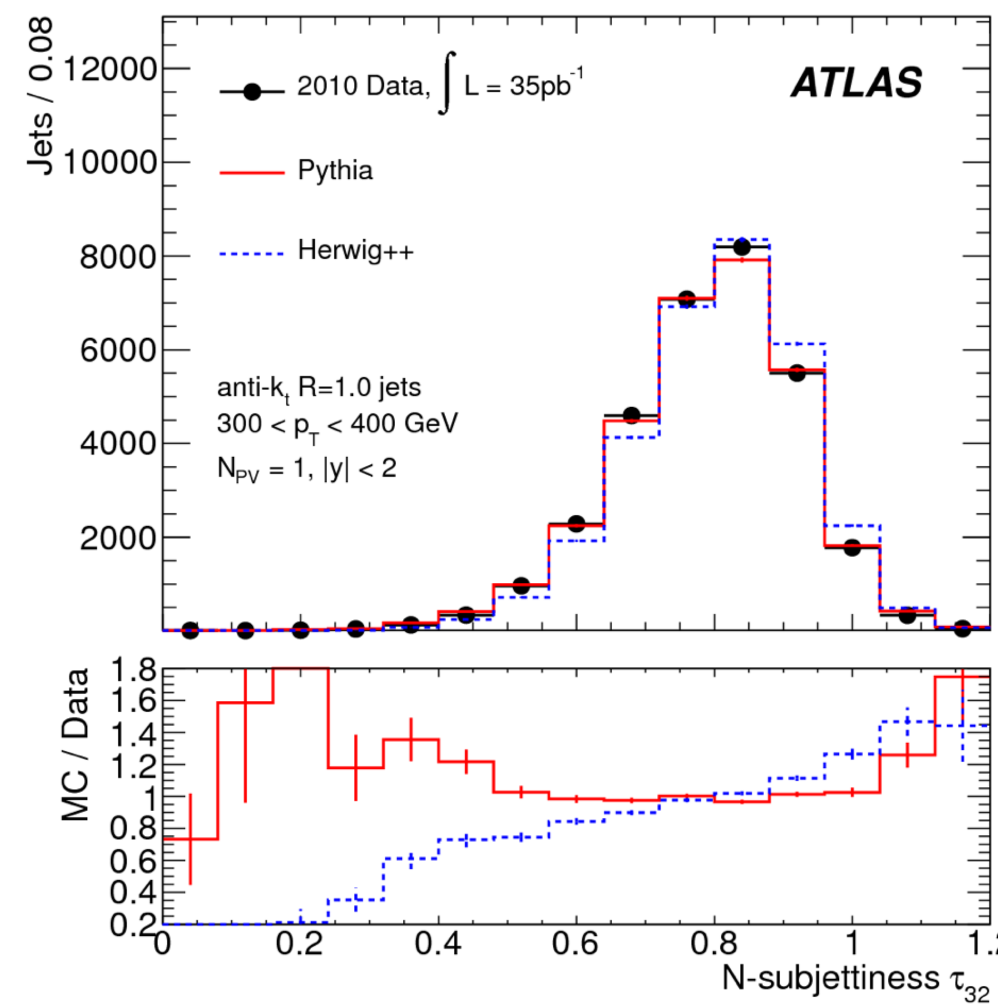
N-Subjettiness

The ratio of τ_N/τ_{N-1} is used as discriminant

More like 2 subjets than 1

More like 3 subjets than 2



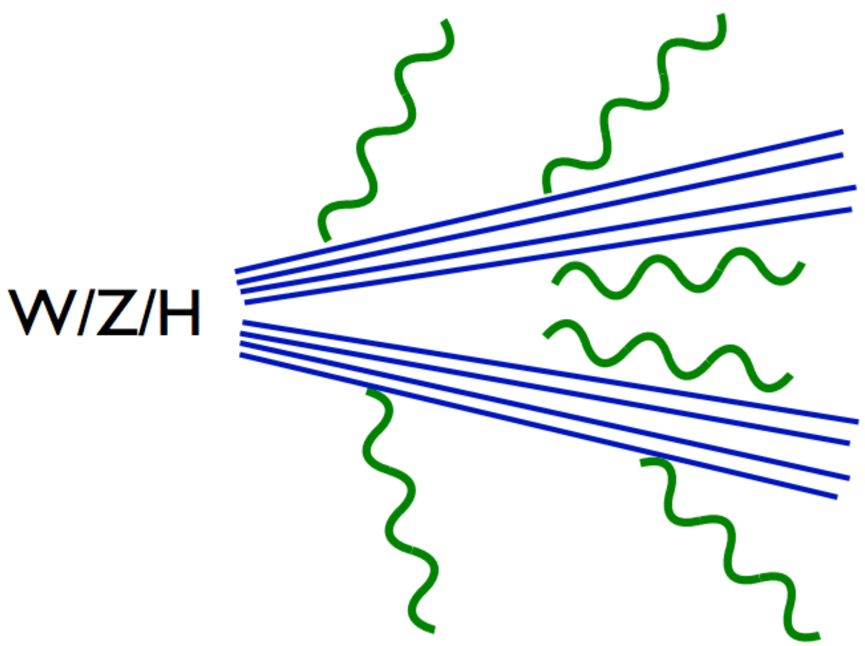


W-like ↔ MJ-like

Top-like ↔ MJ-like

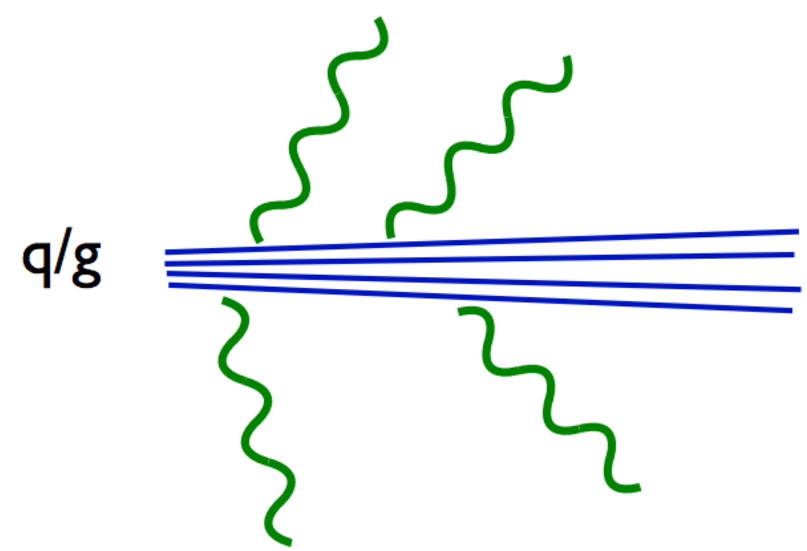
Energy Correlation Functions

Discriminate between:



Signal: Two-prong jet

Characteristic angular size
determined by mass



Background: One-prong jet

No intrinsic angular size

ECF

Over all constituents (beta: angular exponent):

$$\mathrm{ECF}(1,\beta) = \sum_i p_{Ti}$$

$$\mathrm{ECF}(2,\beta) = \sum_{i<j} p_{Ti}\, p_{Tj}\, (R_{ij})^\beta$$ ← [see Banfi, Salam, Zanderighi; Jankowiak, Larkoski]

$$\mathrm{ECF}(3,\beta) = \sum_{i<j<k} p_{Ti}\, p_{Tj}\, p_{Tk}\, (R_{ij} R_{jk} R_{ki})^\beta$$

$$\mathrm{ECF}(N,\beta) = \sum_{\text{sets of } N} (N \text{ energies}) \times \left(\binom{N}{2} \text{ angles} \right)^\beta$$

ECF(N+1) << ECF(N)

for N subjets

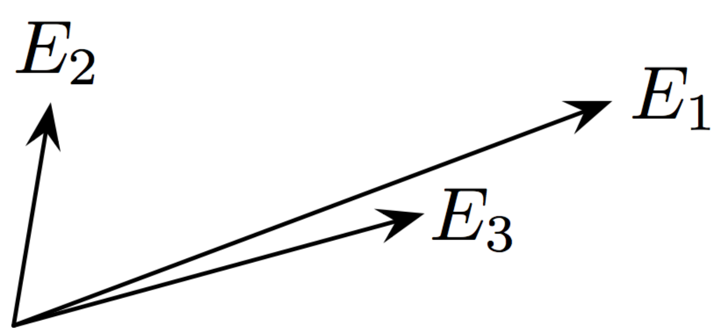
Define (double) ratio = [ECF(N+1)/ECF(N)]/[ECF(N)/ECF(N-1)]

$$C_N^{(\beta)} = \frac{\mathrm{ECF}(N+1,\beta)\, \mathrm{ECF}(N-1,\beta)}{\mathrm{ECF}(N,\beta)^2}$$

Analogous to N-subjettiness ratio

The energy correlation double ratio C_N effectively measures higher-order radiation from leading order (LO) substructure. For a system with N subjets, the LO substructure consists of N hard prongs, so if C_N is small, then the higher-order radiation must be soft or collinear with respect to the LO structure. If C_N is large, then the higher-order radiation is not strongly-ordered with respect to the LO structure, so the system has more than N subjets. Thus, if C_N is small and C_{N-1} is large, then we can say that a system has N subjets.

ECF



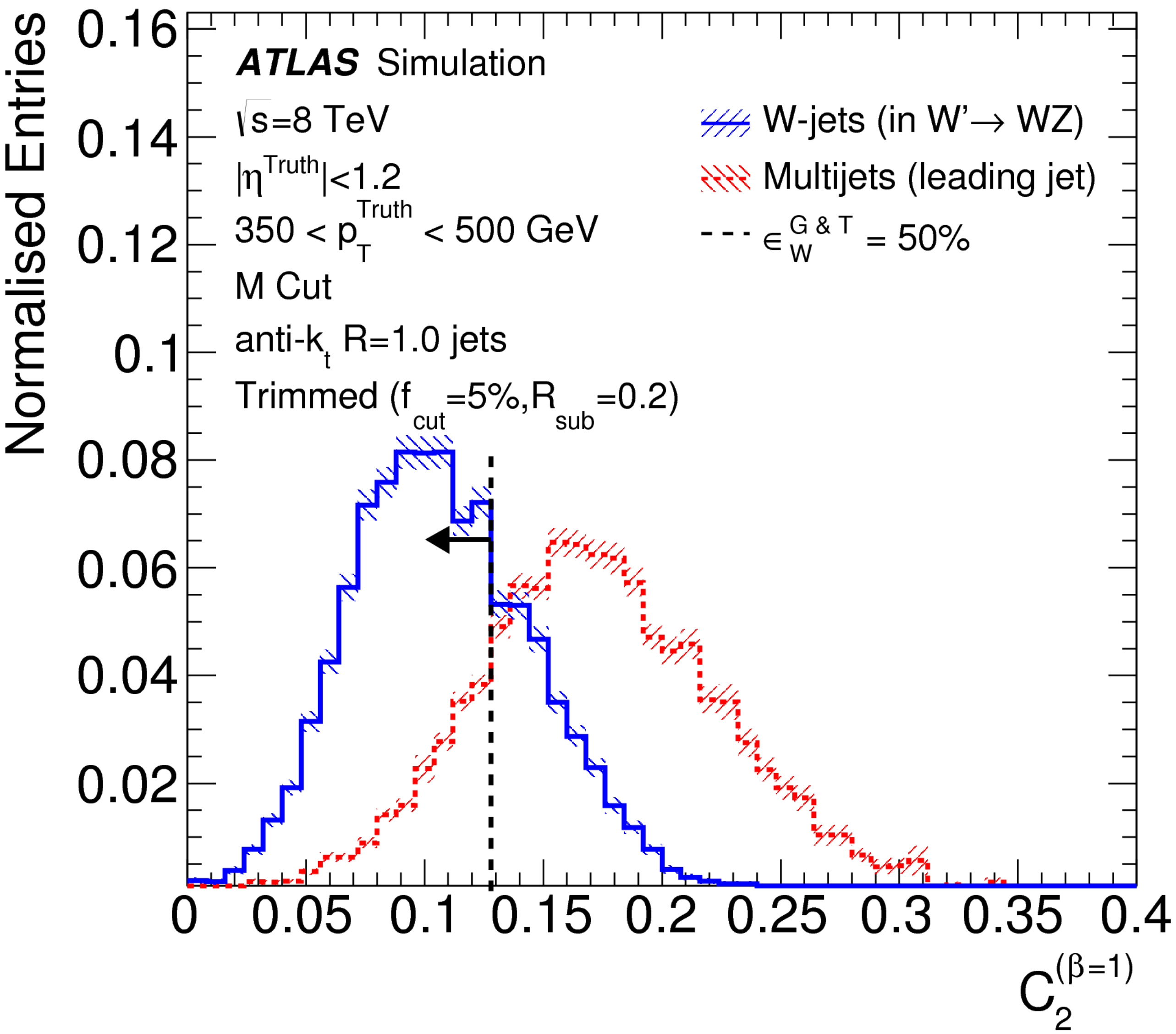
For this multiple soft radiation case, with only 1 **real** subjet

$C_2 > \tau_{21}$

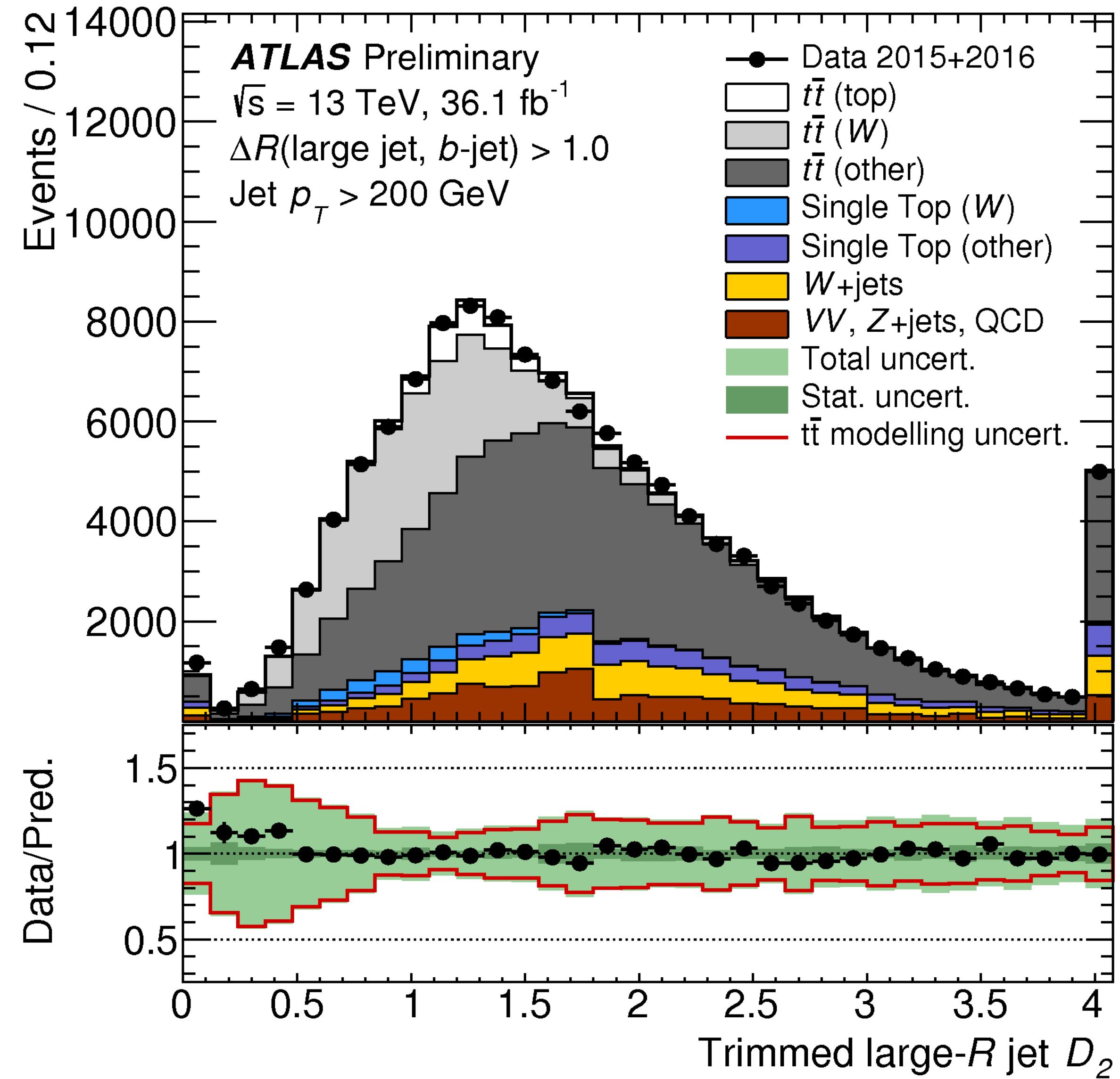
Nsubjettiness will identify this as more 2 subjet-like while ECF will identify more as 1 subjet-like

D-observables are further optimised by exploiting boost-invariance of the difference of one and two prong

ECF results

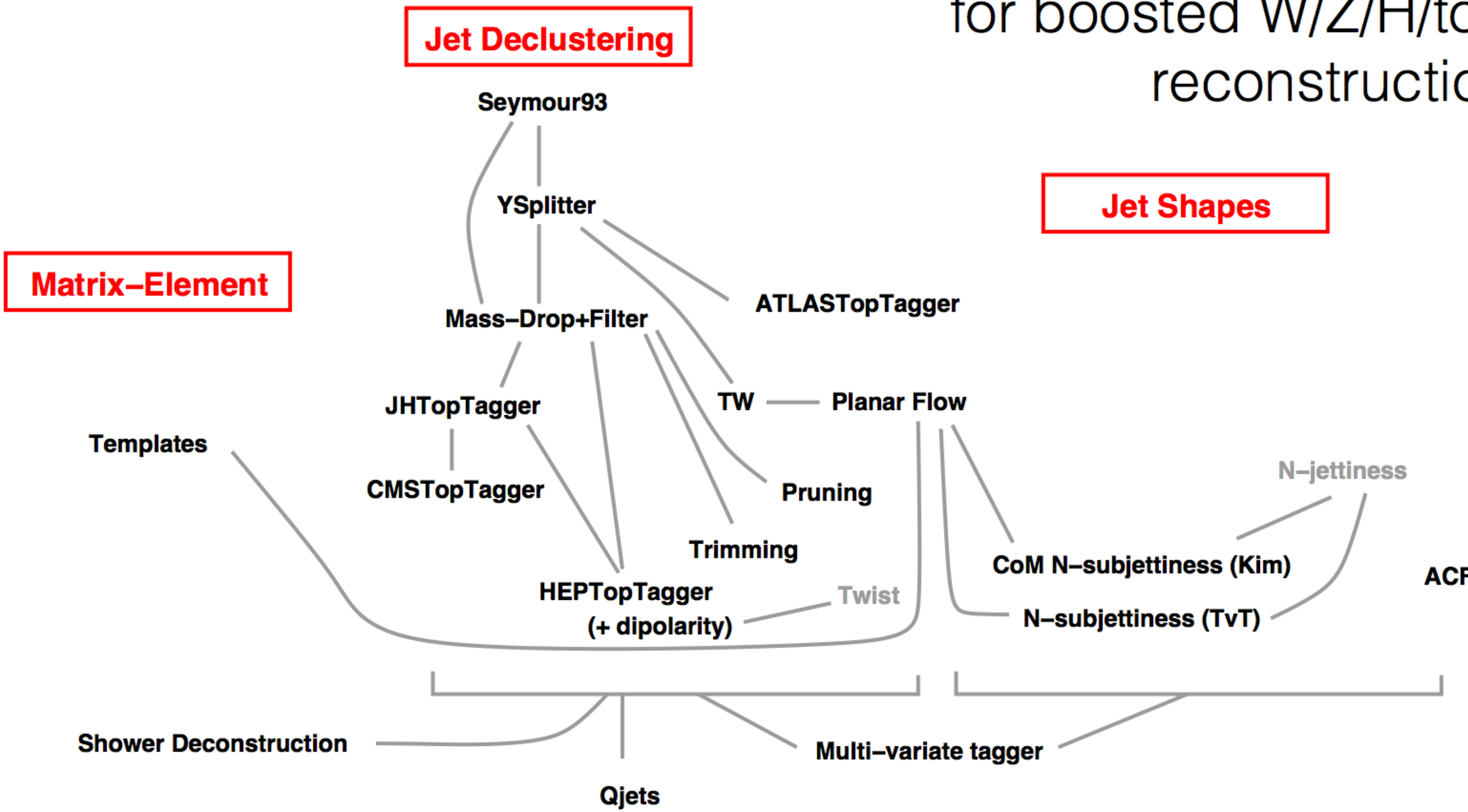


ATLAS Simulation
√s=8 TeV
|η^Truth|<1.2
350 < p_T^Truth < 500 GeV
M Cut
anti-k_t R=1.0 jets
Trimmed (f_cut=5%,R_sub=0.2)
W-jets (in W'→ WZ)
Multijets (leading jet)
∈_W^G & T = 50%
Normalised Entries
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
C_2^(β=1)



ATLAS Preliminary
√s = 13 TeV, 36.1 fb^-1
ΔR(large jet, b-jet) > 1.0
Jet p_T > 200 GeV
Data 2015+2016
tt̄ (top)
tt̄ (W)
tt̄ (other)
Single Top (W)
Single Top (other)
W+jets
VV, Z+jets, QCD
Total uncert.
Stat. uncert.
tt̄ modelling uncert.
Events / 0.12
14000
12000
10000
8000
6000
4000
2000
Data/Pred.
1.5
1
0.5
0
0.5
1
1.5
2
2.5
3
3.5
4
Trimmed large-R jet D_2

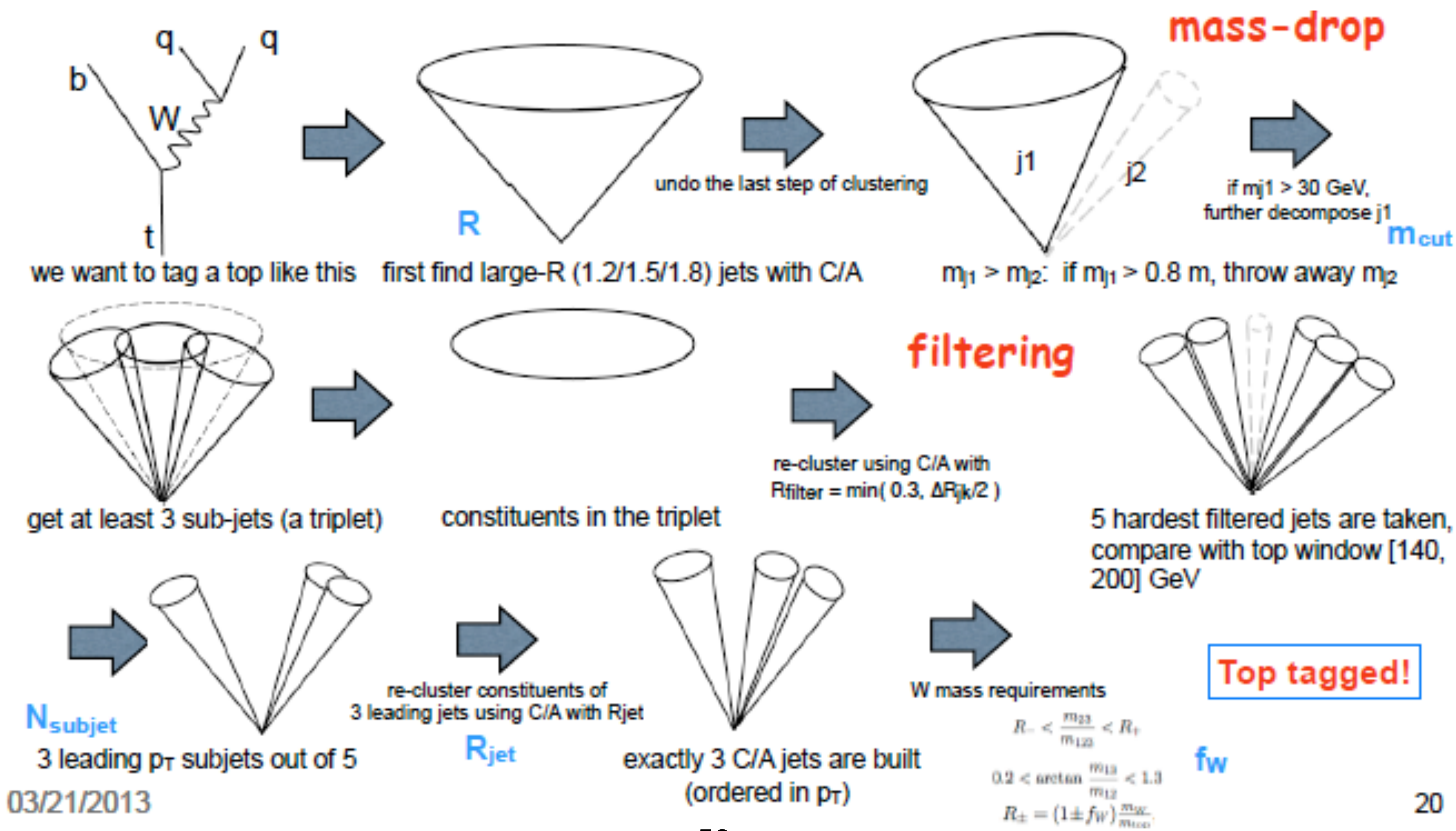
Some of the tools developed for boosted W/Z/H/top reconstruction



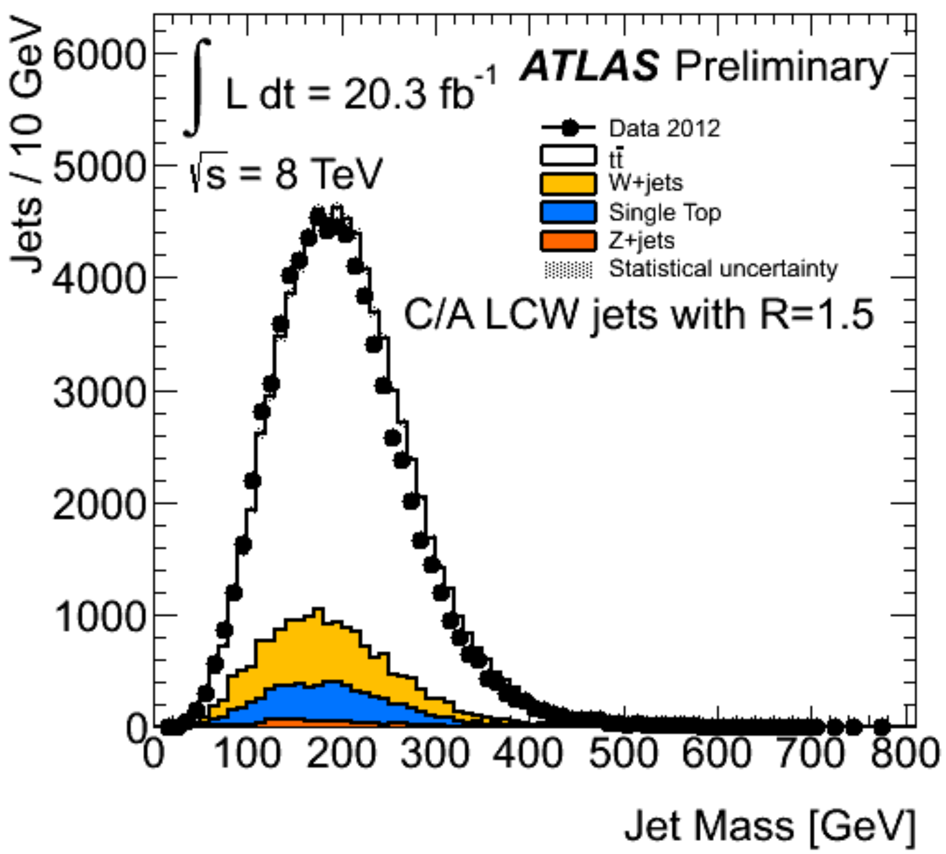
apologies for omitted taggers, arguable links, etc.

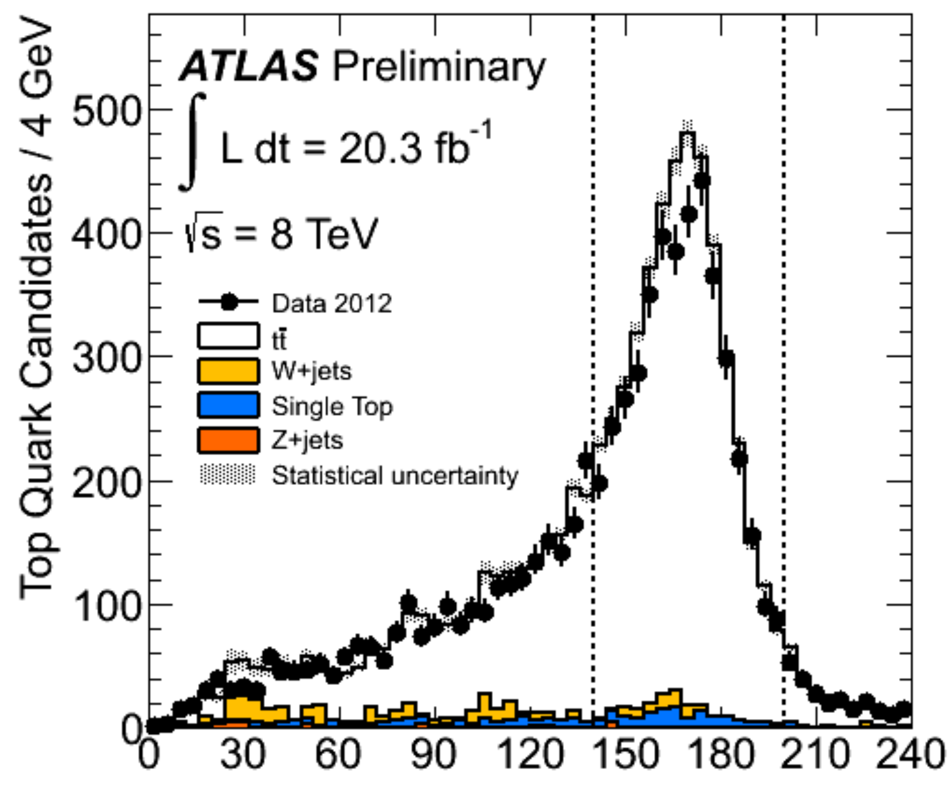
HEPTopTagger

Browsing through all the branches of jet recombination history:

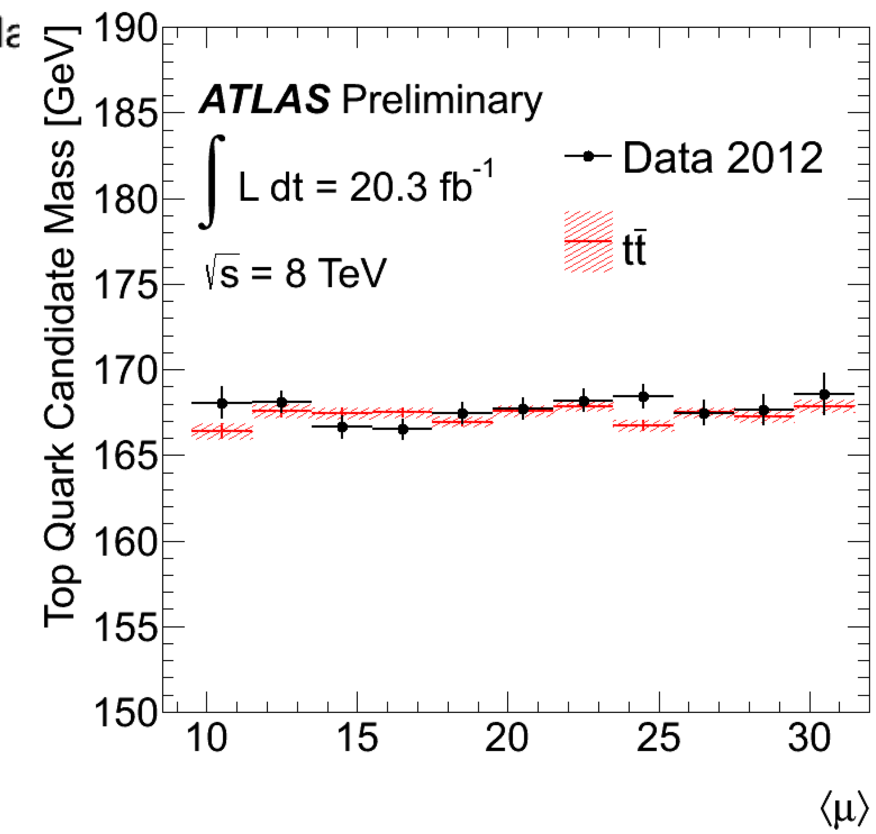


HEPTopTagger results





Before and after tagging by HepTopTagger

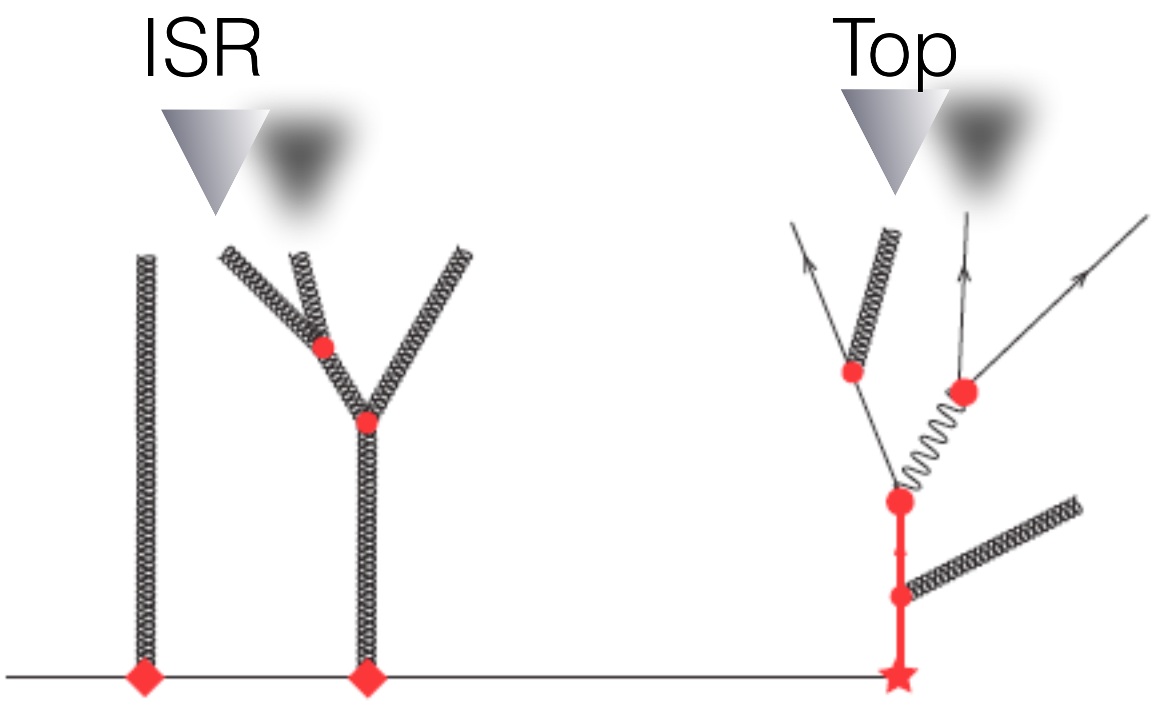


Pileup resilience

Quiz

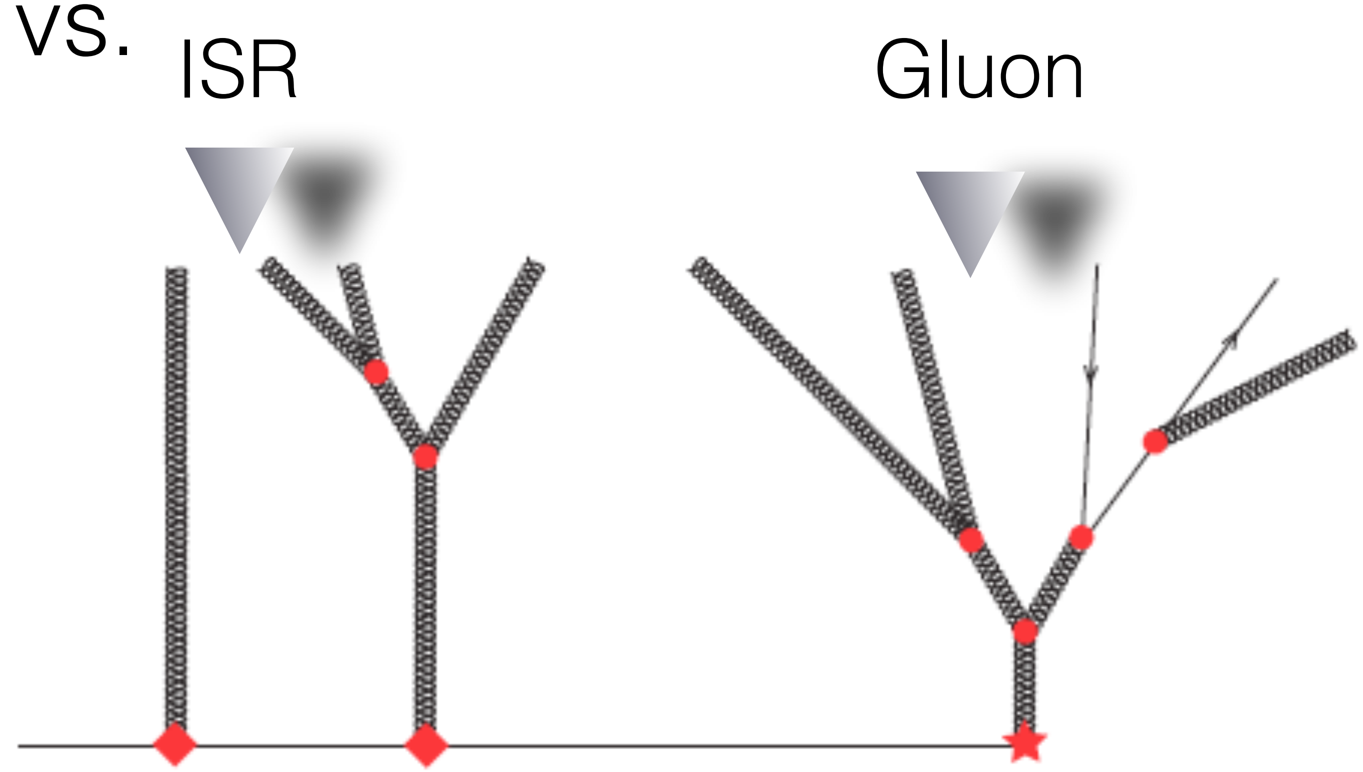
What does HEP in HEPTopTagger stand for?

Shower deconstruction



Top quark jet
shower history

vs.

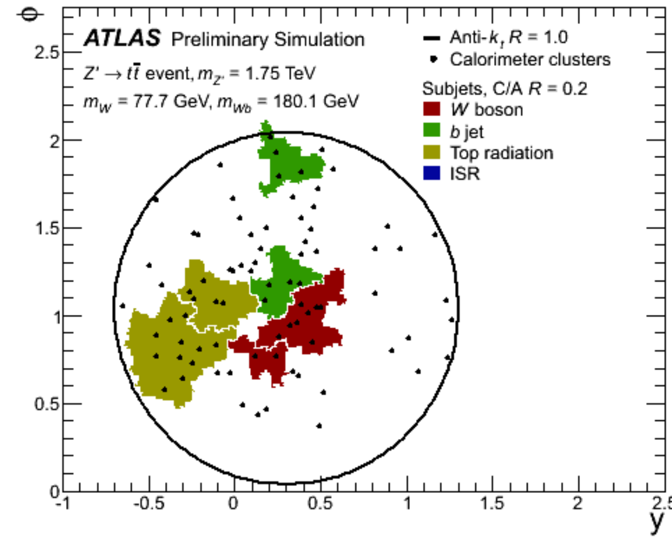


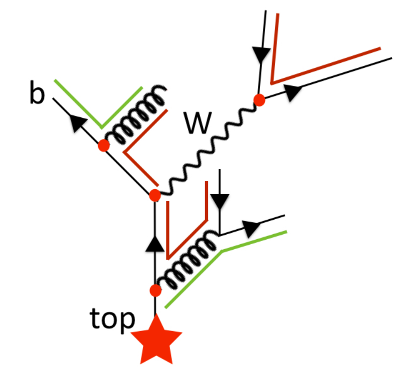
Light quark jet
shower history

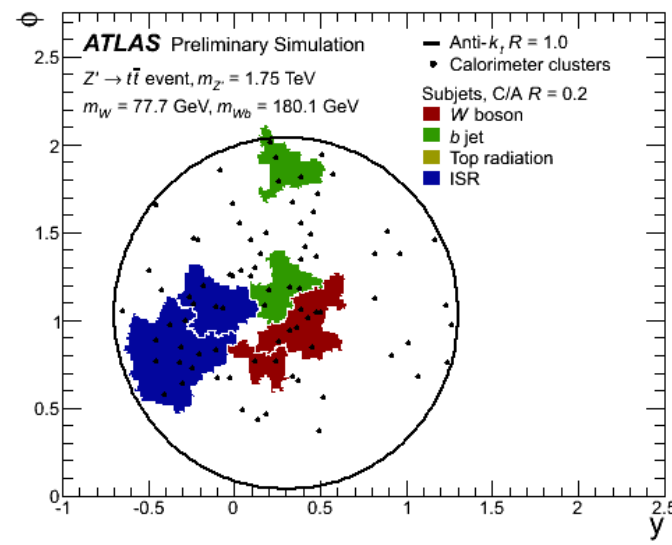
Shower deconstruction

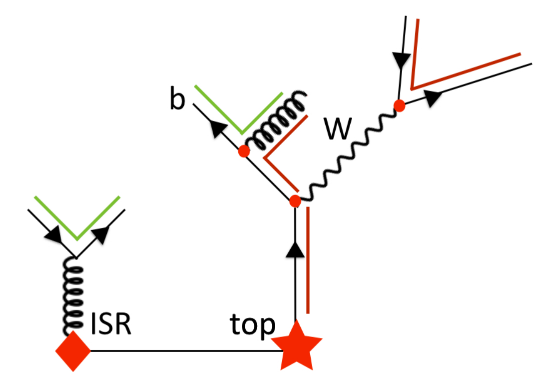
- Decompose the large-radius jet into small radius **subjets.**
- Build all possible shower histories with the subjets.
- Assign probability whether signal-like or background-like.
- A single analytic function:

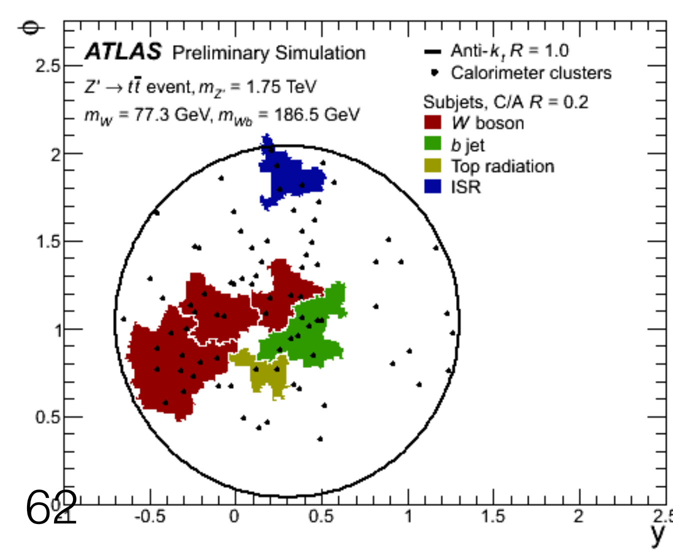
$$\chi(\{p\}_N) = \frac{P(\{p\}_N|\mathrm{S})}{P(\{p\}_N|\mathrm{B})}$$

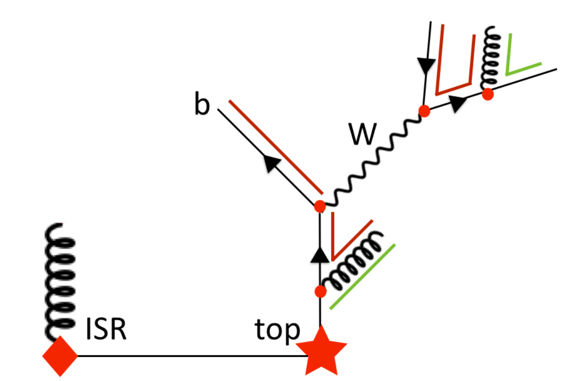




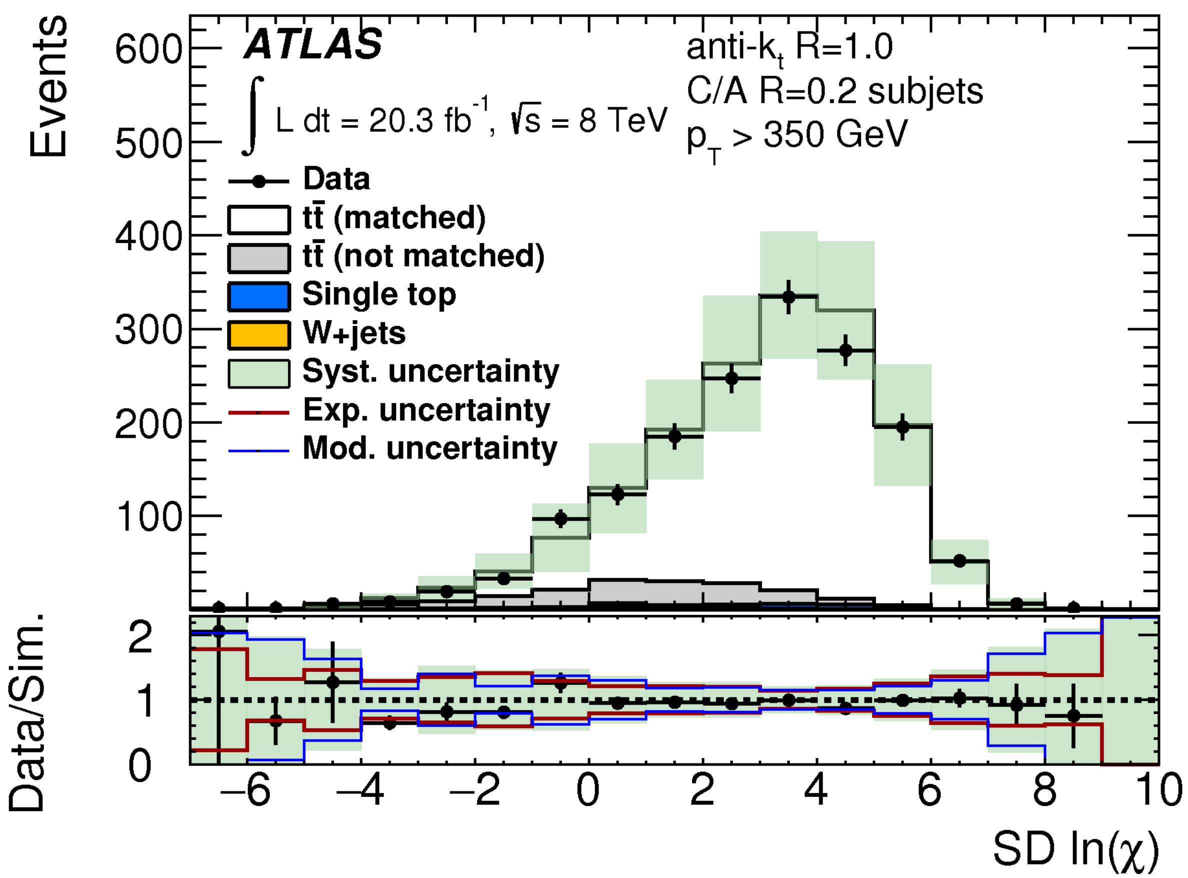




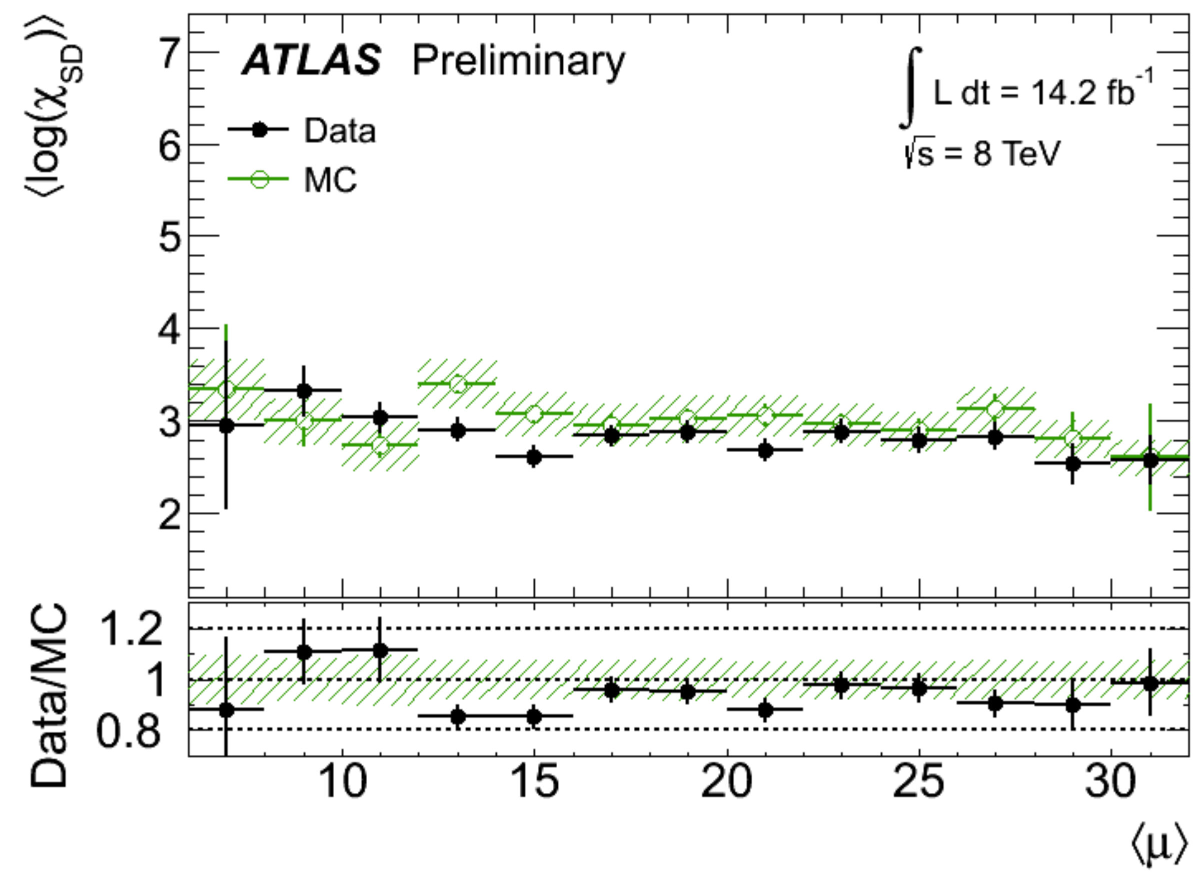




SD with data

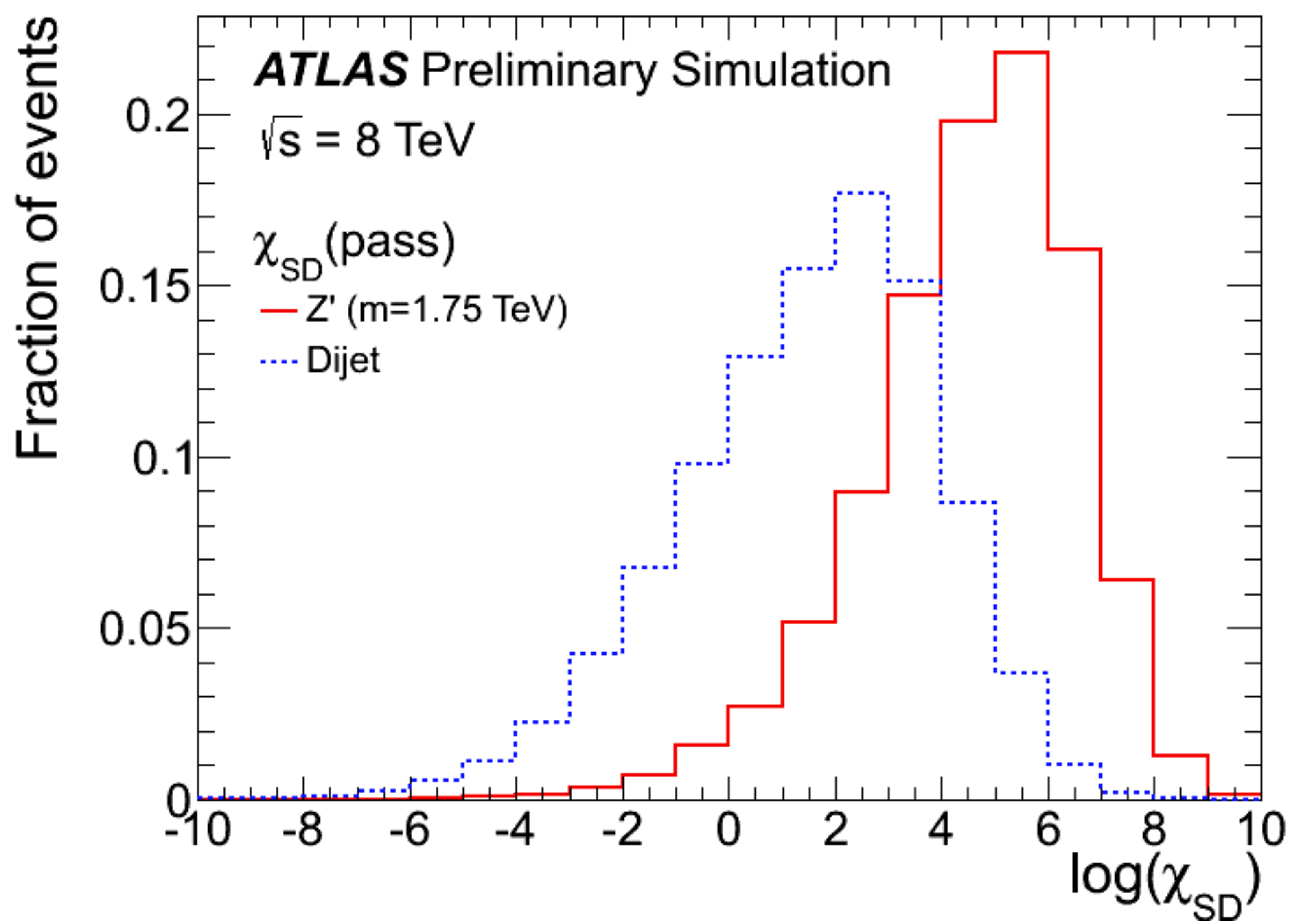


LogChi modelled well by MC

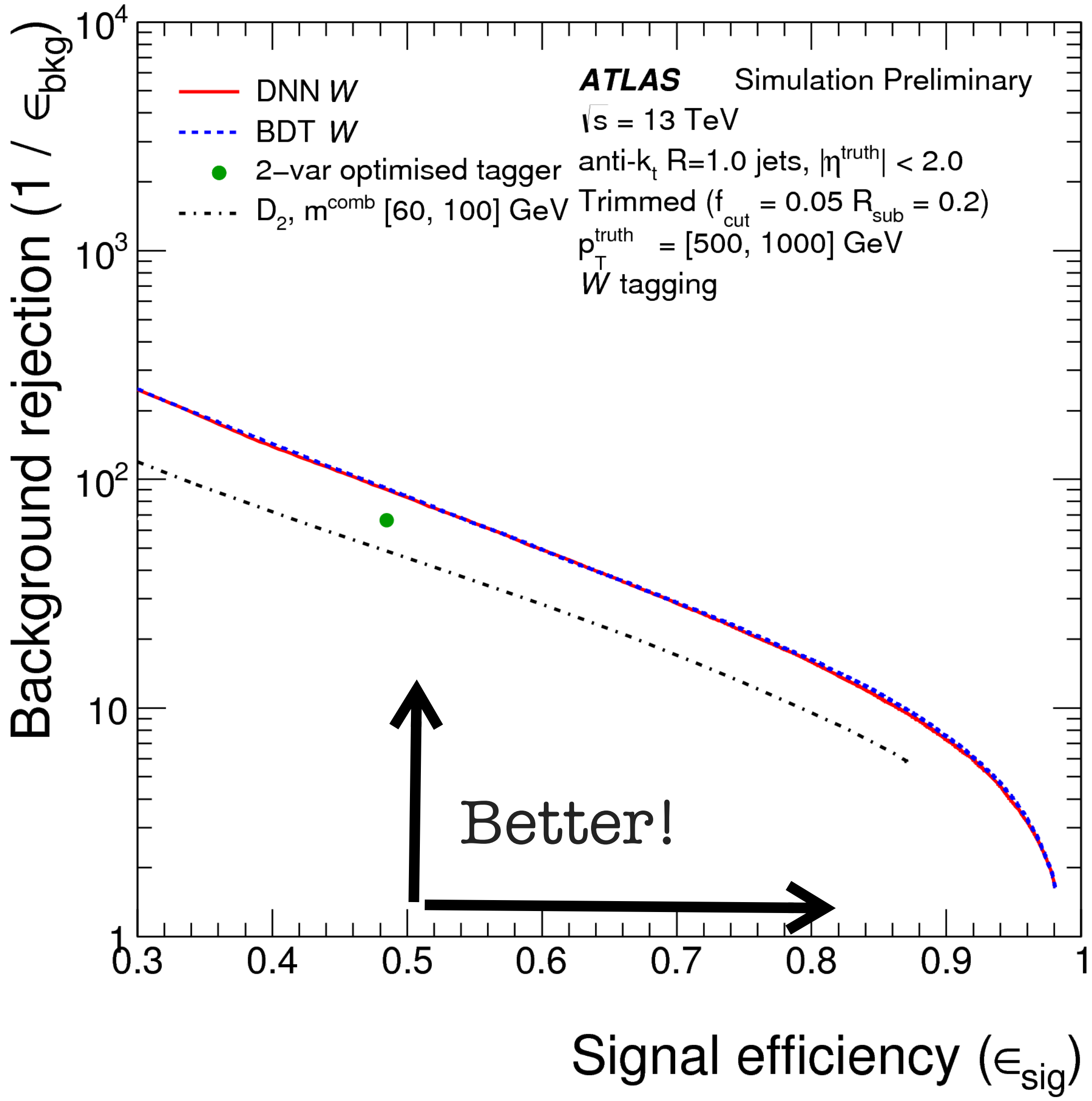


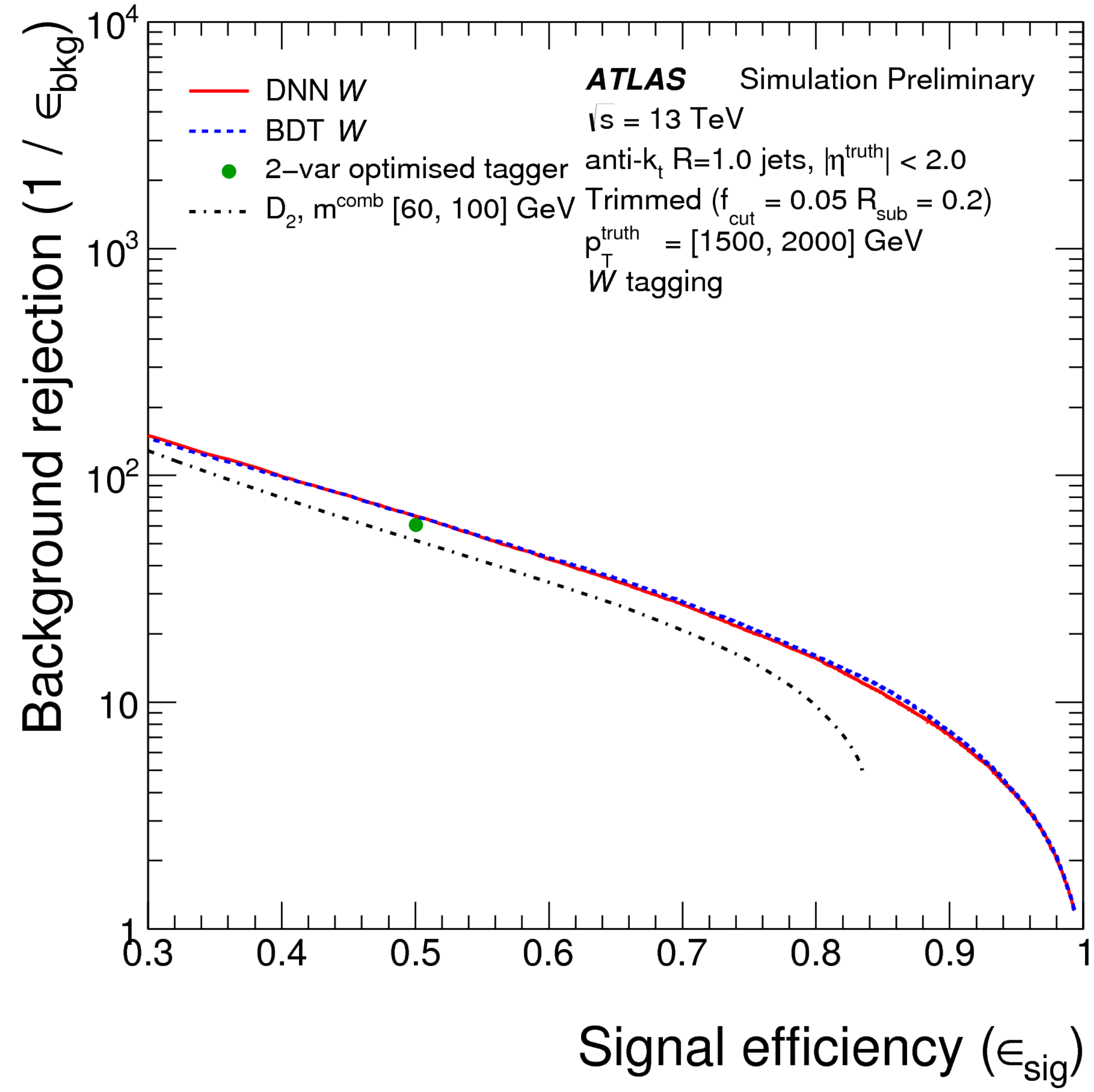
LogChi robust against pileup

Signal over background discrimination

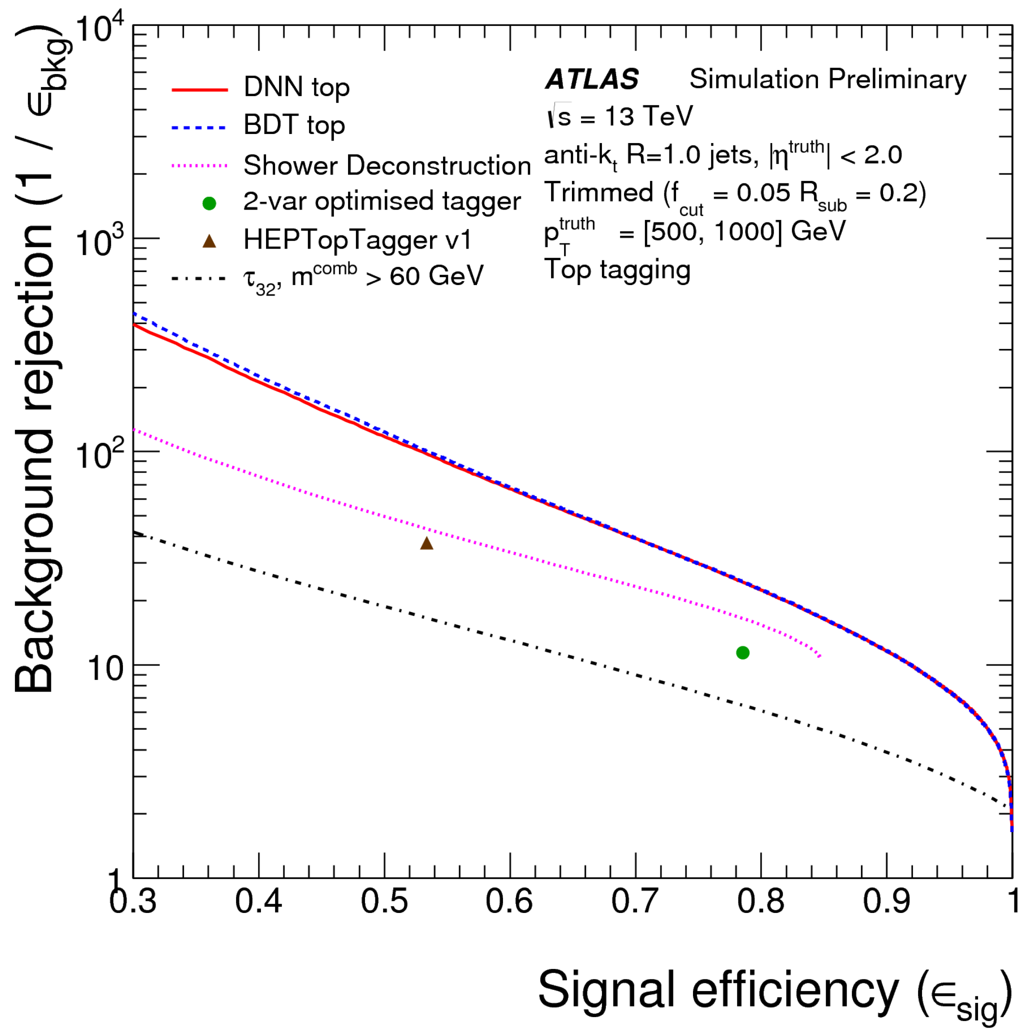


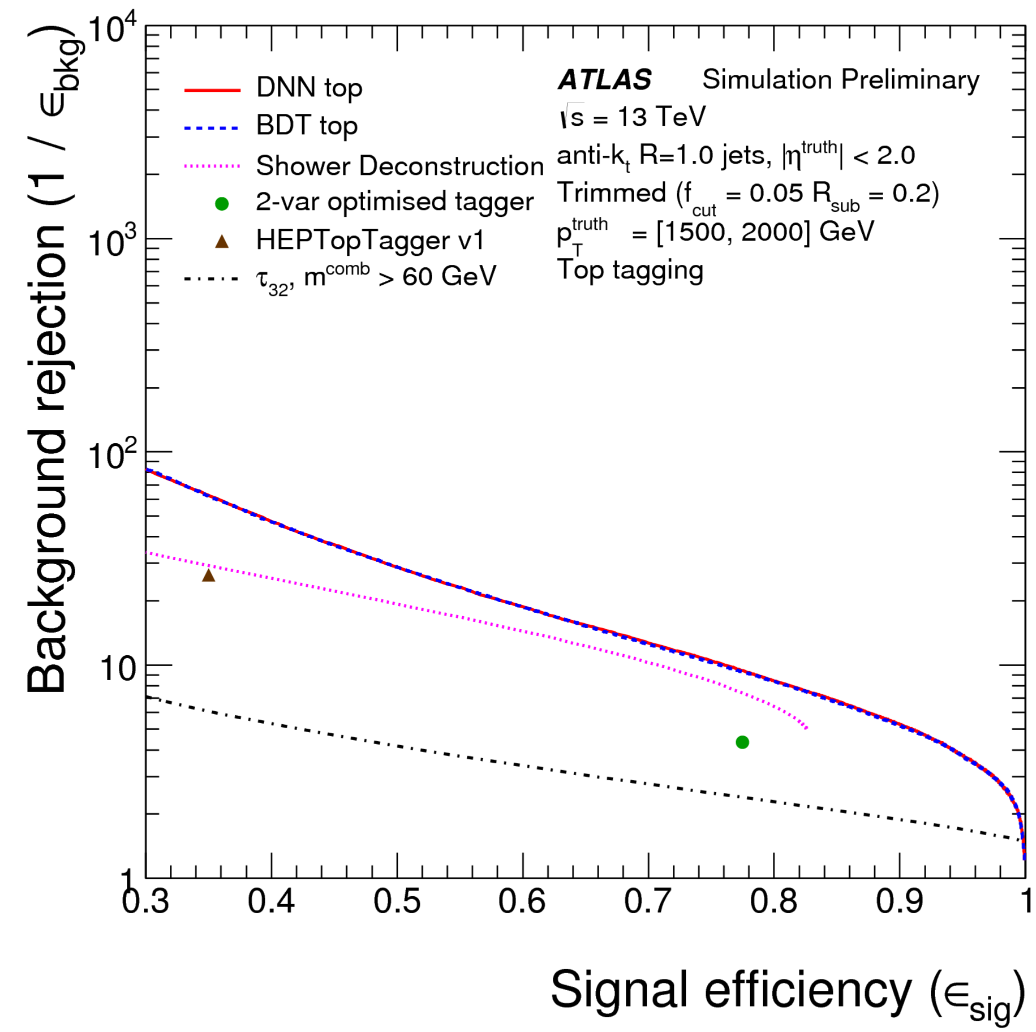
Tagger Comparison (W)





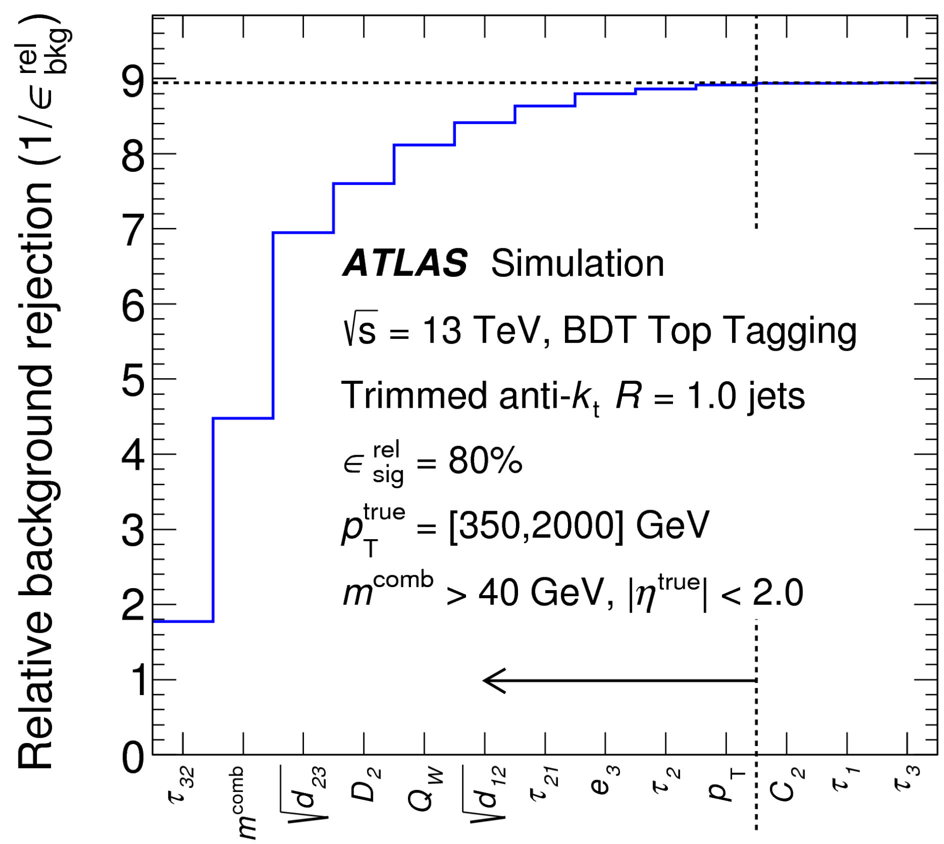
Tagger comparison (top)

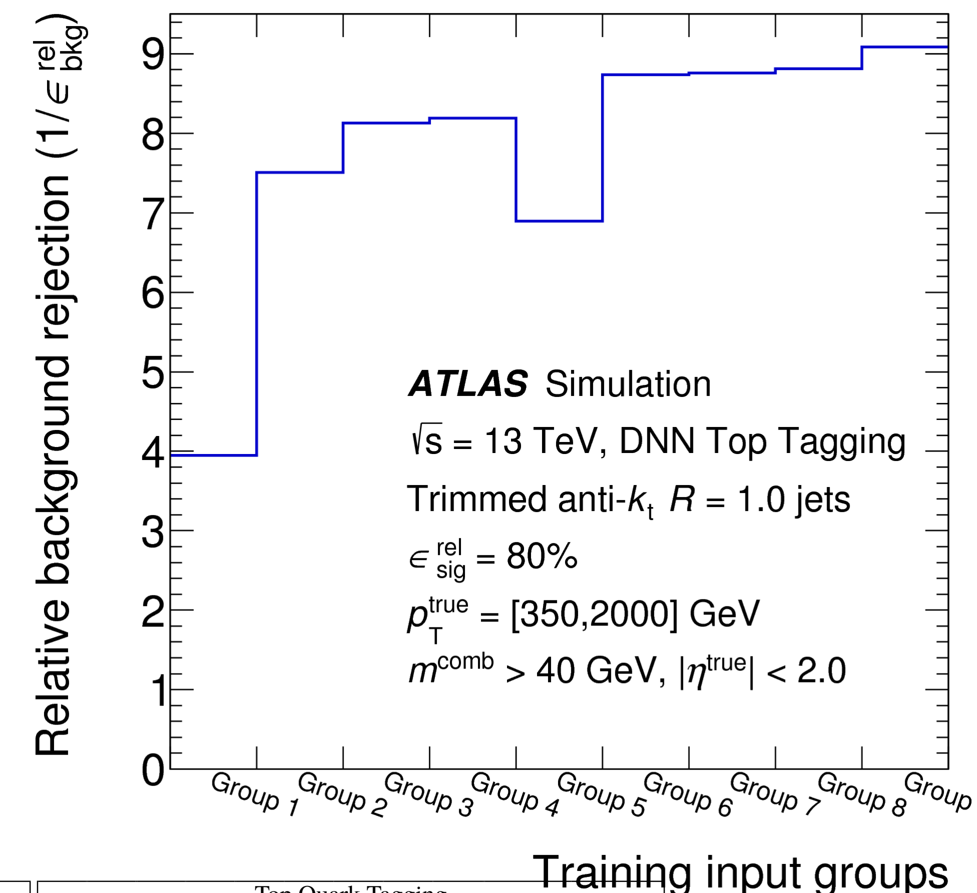




Better top quark finding efficiency with SD at the same rejection of multijets when compared to other taggers

Example: top tagging in ATLAS

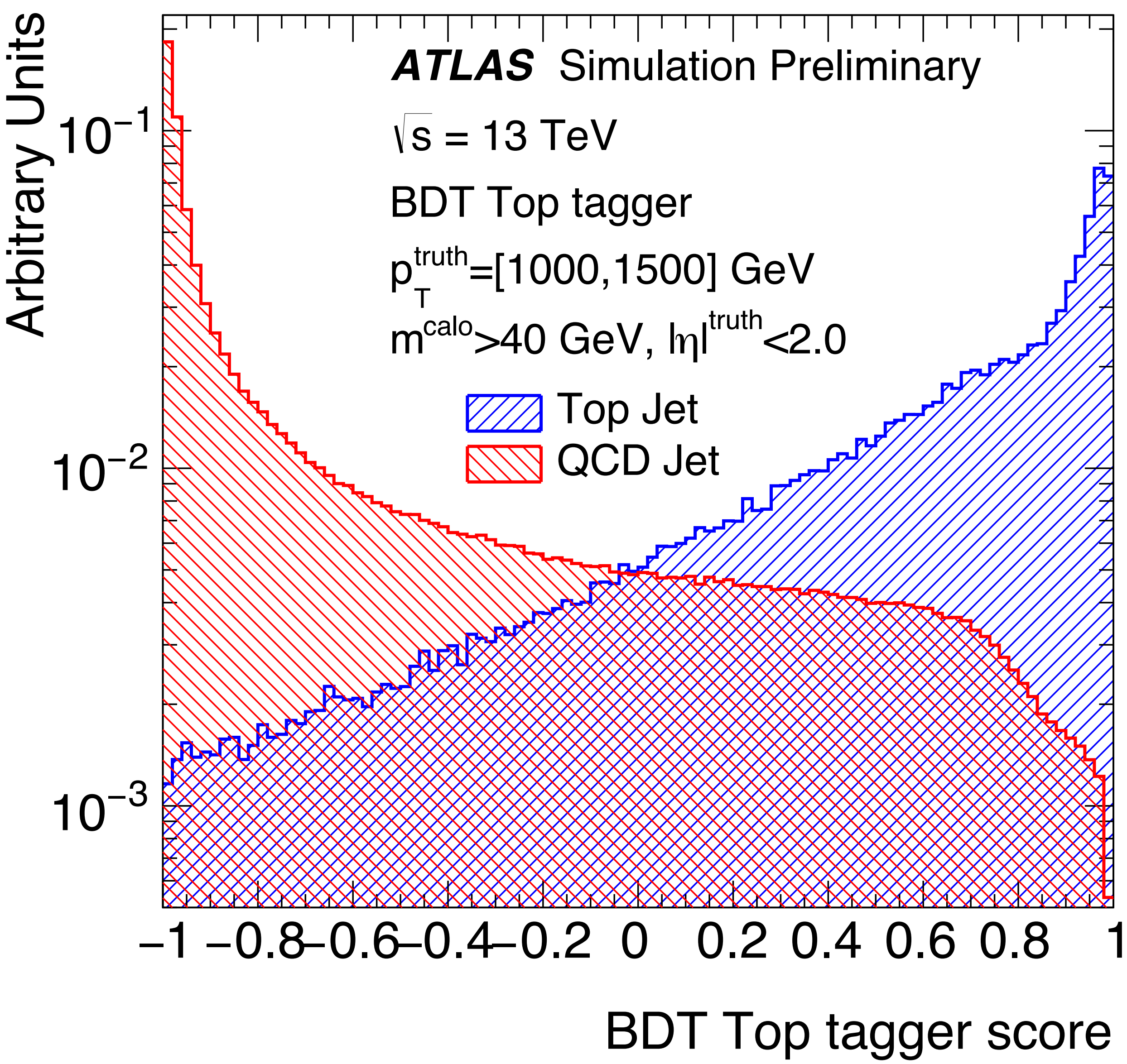


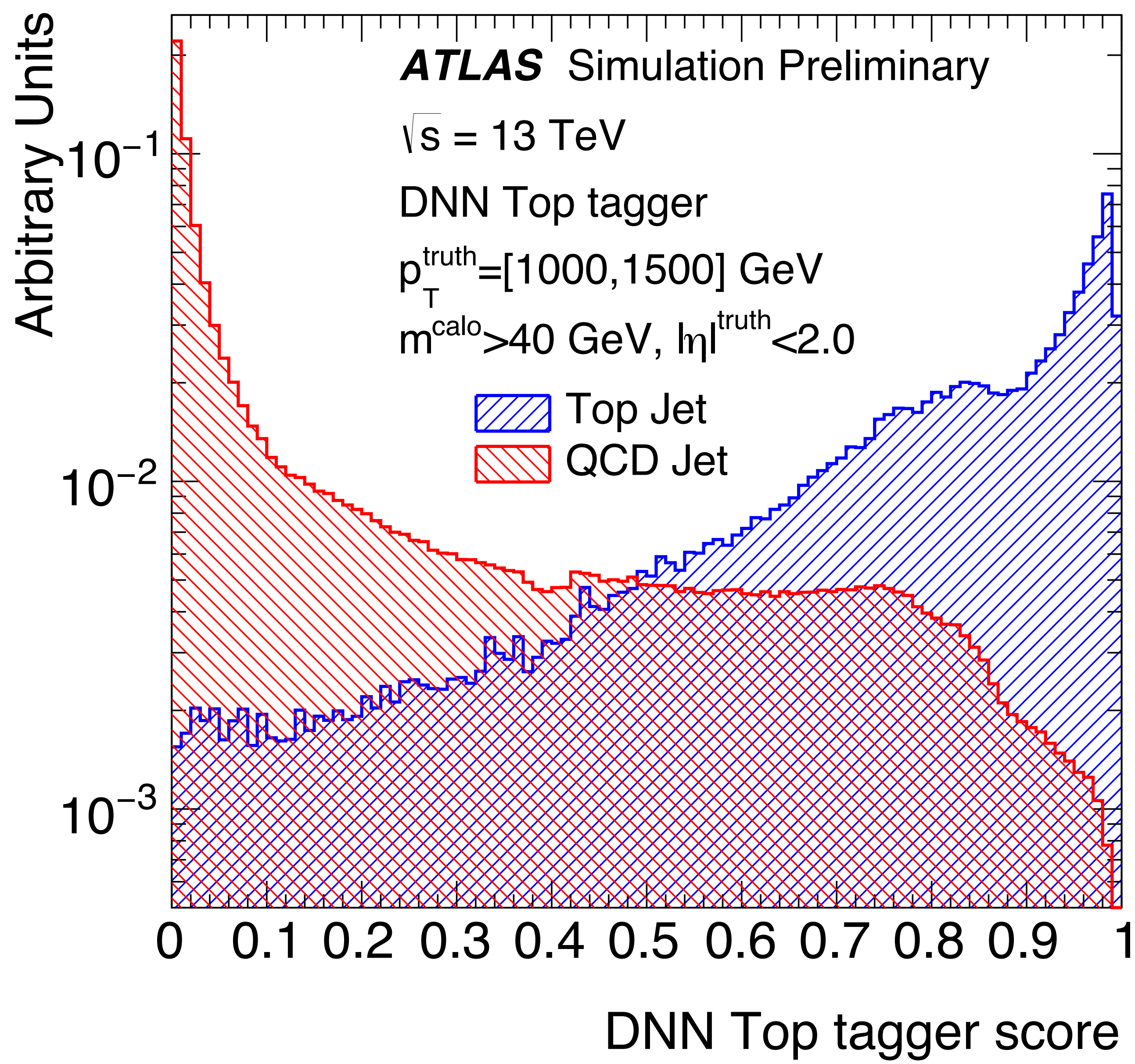


	W Boson Tagging											Top Quark Tagging										
	DNN Test Groups									Chosen Inputs		DNN Test Groups									Chosen Inputs	
Observable	1	2	3	4	5	6	7	8	9	BDT	DNN	1	2	3	4	5	6	7	8	9	BDT	DNN
m^{comb}	○	○		○	○	○	○	○	○	○	○		○	○	○		○	○	○	○	○	○
p_T	○	○			○	○		○	○	○	○			○	○			○	○	○	○	○
e_3	○	○				○			○						○			○		○	○	○
C_2			○	○	○		○	○	○		○	○	○	○		○	○		○	○		○
D_2			○	○	○		○	○	○	○	○	○	○	○		○	○		○	○		○
τ_1	○	○				○			○	○					○			○		○		○
τ_2	○	○				○			○						○			○		○	○	○
τ_3															○			○		○		○
τ_{21}			○	○	○		○	○	○	○	○	○	○	○		○	○		○	○	○	○
τ_{32}												○	○	○		○	○		○	○	○	○
R_2^{FW}			○	○	○	○	○	○	○	○	○											
$\mathcal{P}$			○	○	○	○	○	○	○	○	○											
a_3			○	○	○	○	○	○	○	○	○											
A			○	○	○	○	○	○	○	○	○											
z_{cut}			○	○	○		○	○	○		○											
$\sqrt{d_{12}}$		○				○	○	○	○	○	○					○	○	○	○	○	○	○
$\sqrt{d_{23}}$																○	○	○	○	○	○	○
$KtDR$		○				○	○	○	○	○	○											
Q_w																○	○	○	○	○	○	○

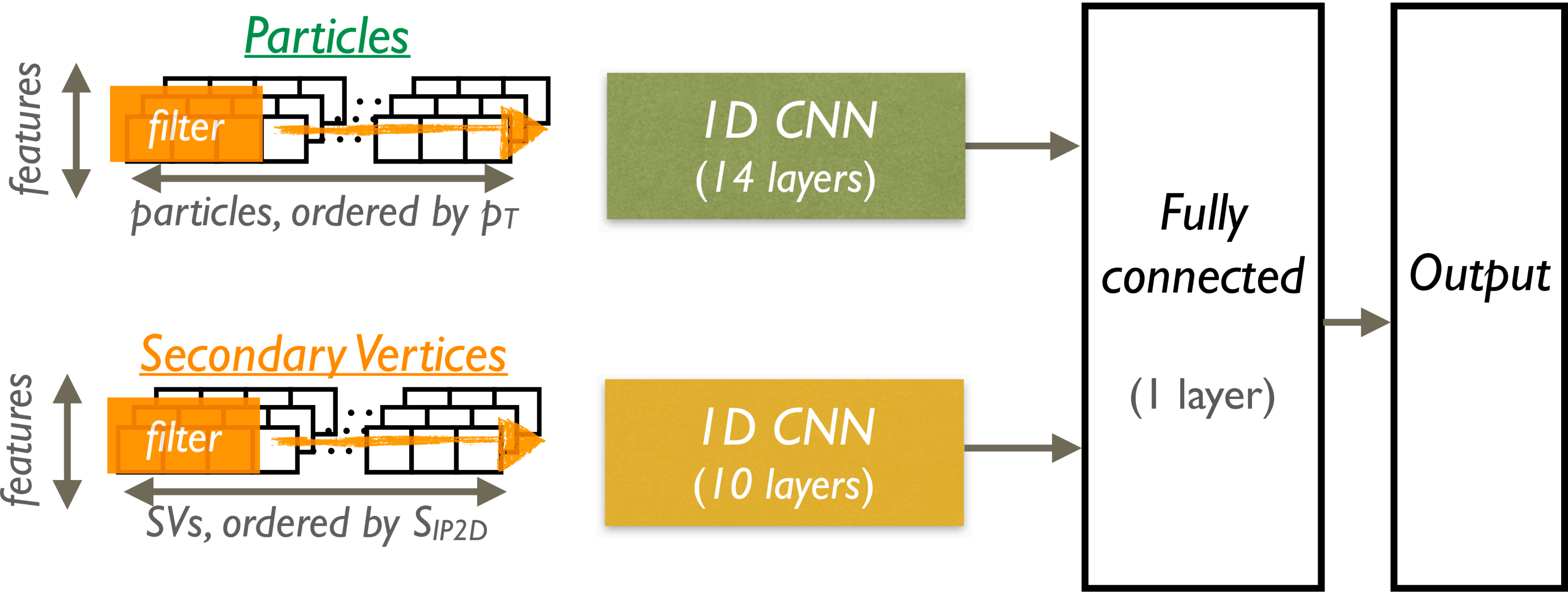
http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/JETM-2018-03/

Example: top tagging in ATLAS





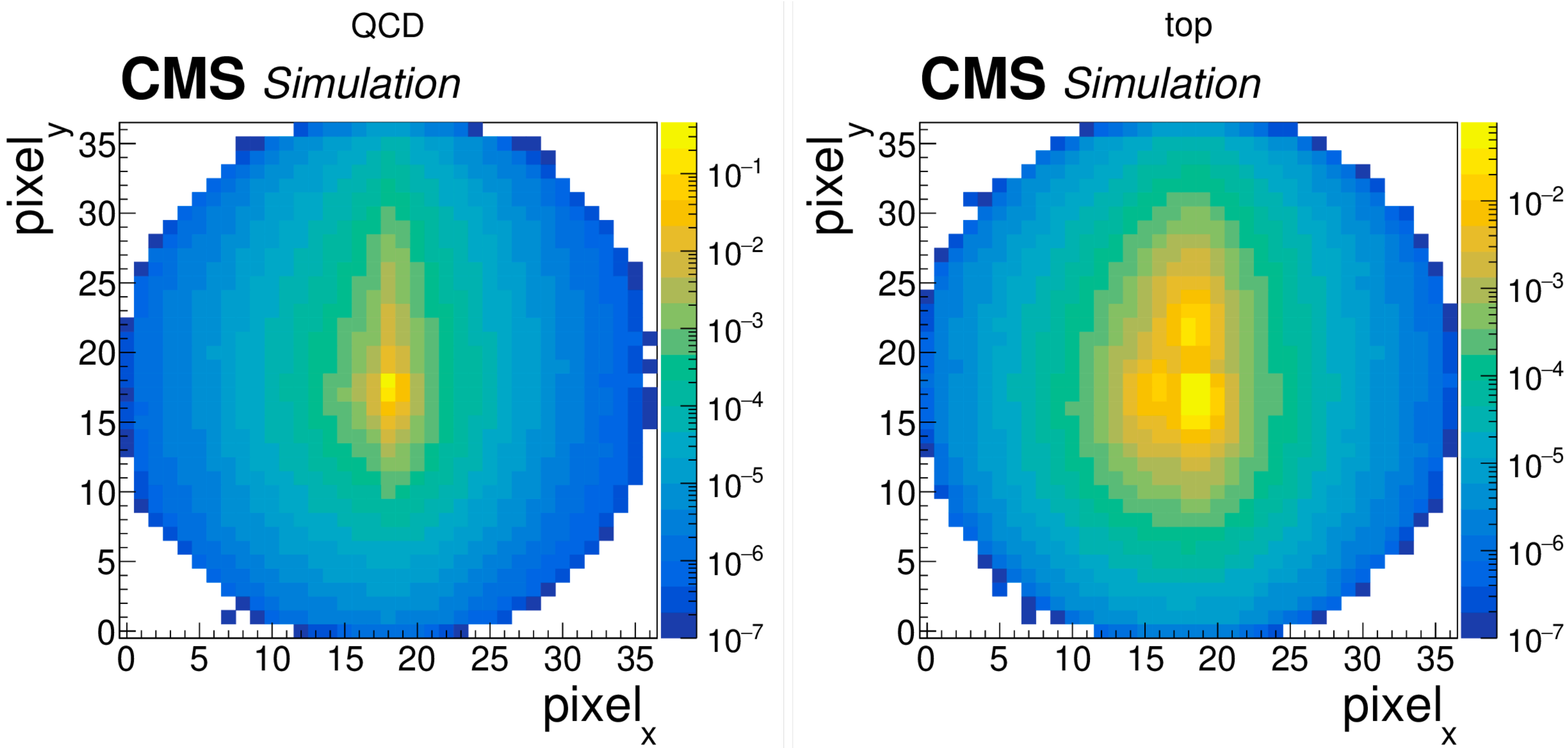
Example: tagging in CMS



DeepAK8: a multiclass classifier for the identification of hadronically decaying particles with five main categories, W/Z/H/t/other. A mass-decorrelated related version as well.

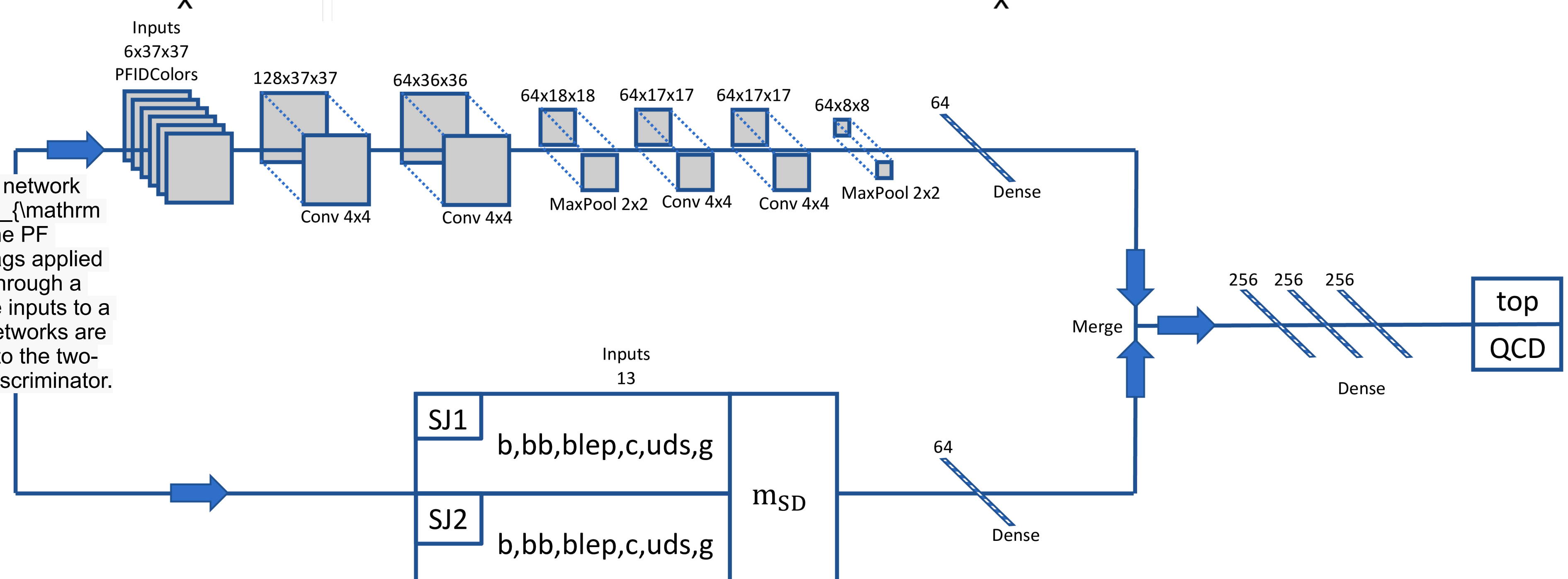
https://cms-results.web.cern.ch/cms-results/public-results/publications/JME-18-002/

Example: top tagging in CMS

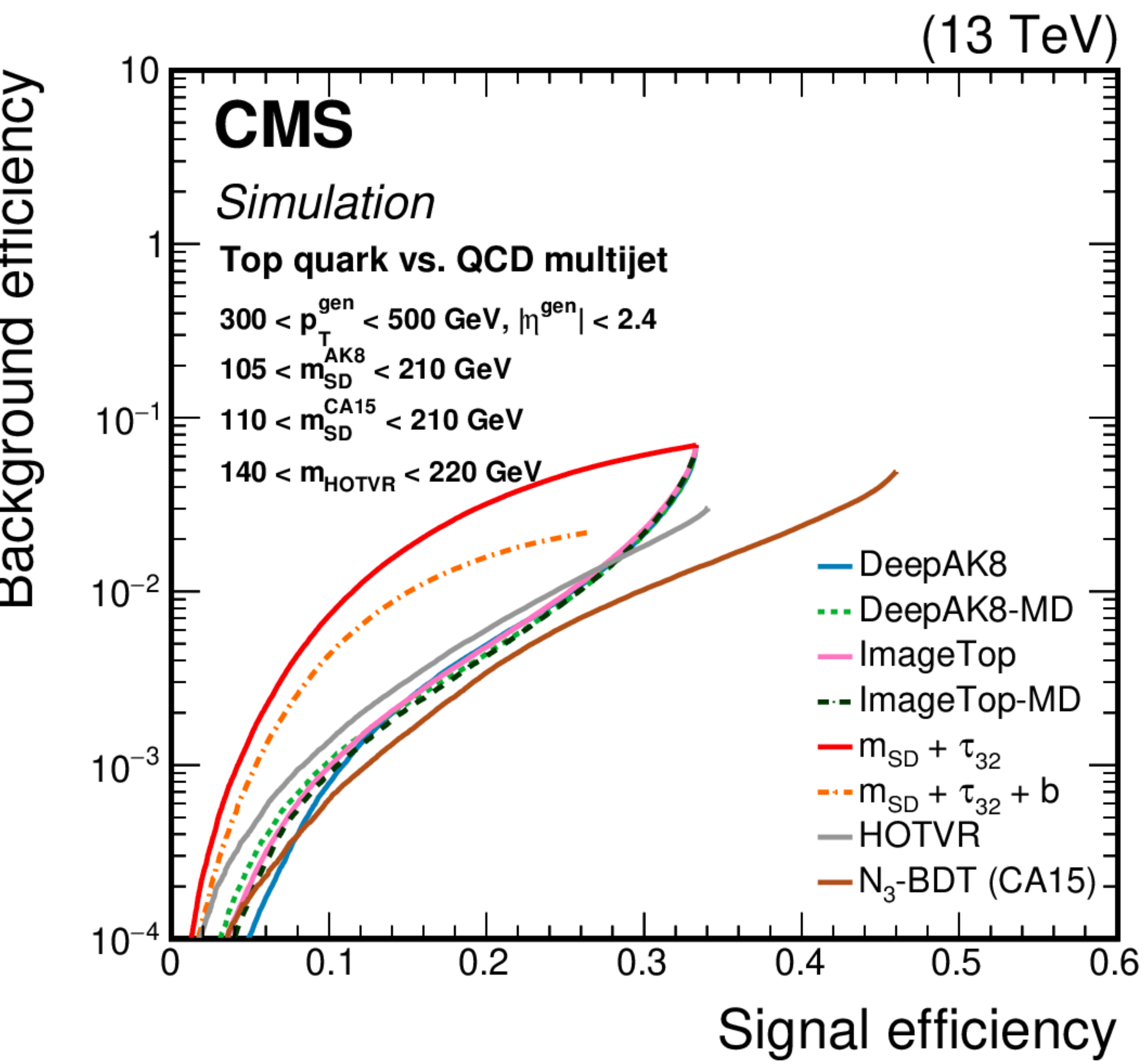


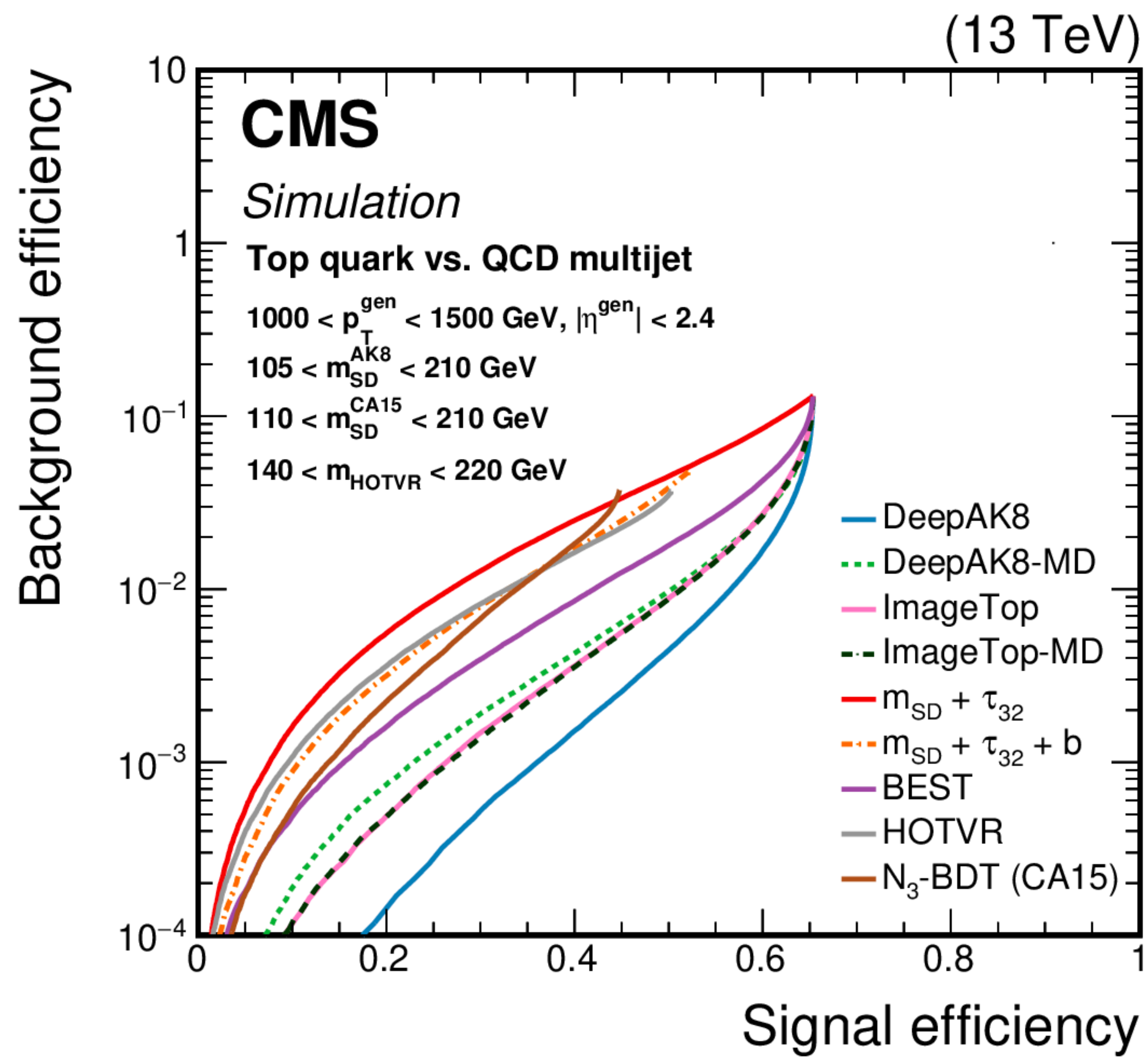
Ensemble of overlaid images after the image post processing

The ImageTop network architecture. The neural network inputs are the 37x37 pixelized PF candidate ${p_{\mathrm{T}}}$ map, which is split into colors based on the PF candidate flavor, and the DeepFlavor subjet b tags applied to both subjets. The pixelized images are sent through a two-dimensional CNN, and the subjet b tags are inputs to a dense layer. After flattening the CNN, the two networks are taken as input to three dense layers and finally to the two-node output, which is used as the top tagging discriminator.



Example: top tagging in CMS



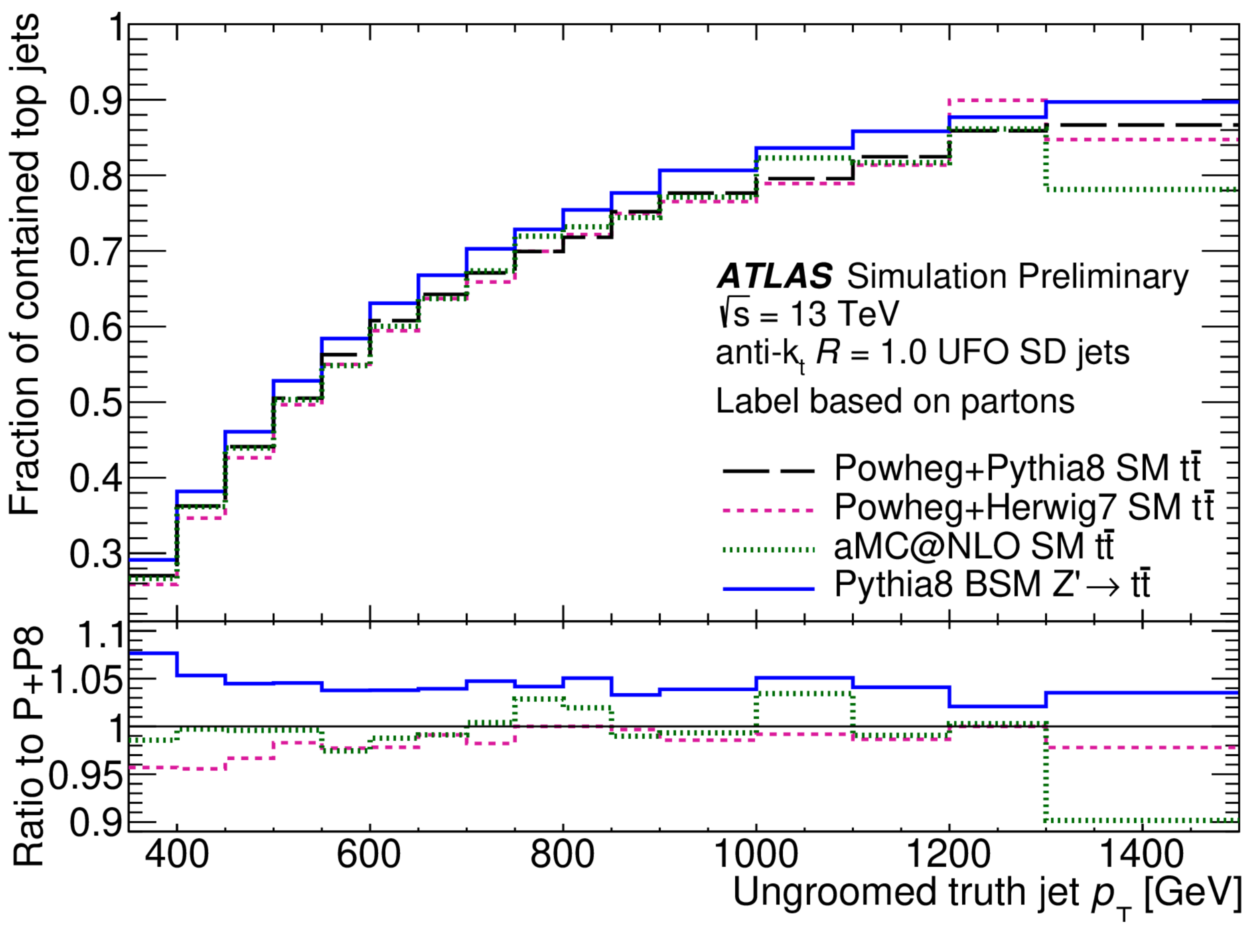


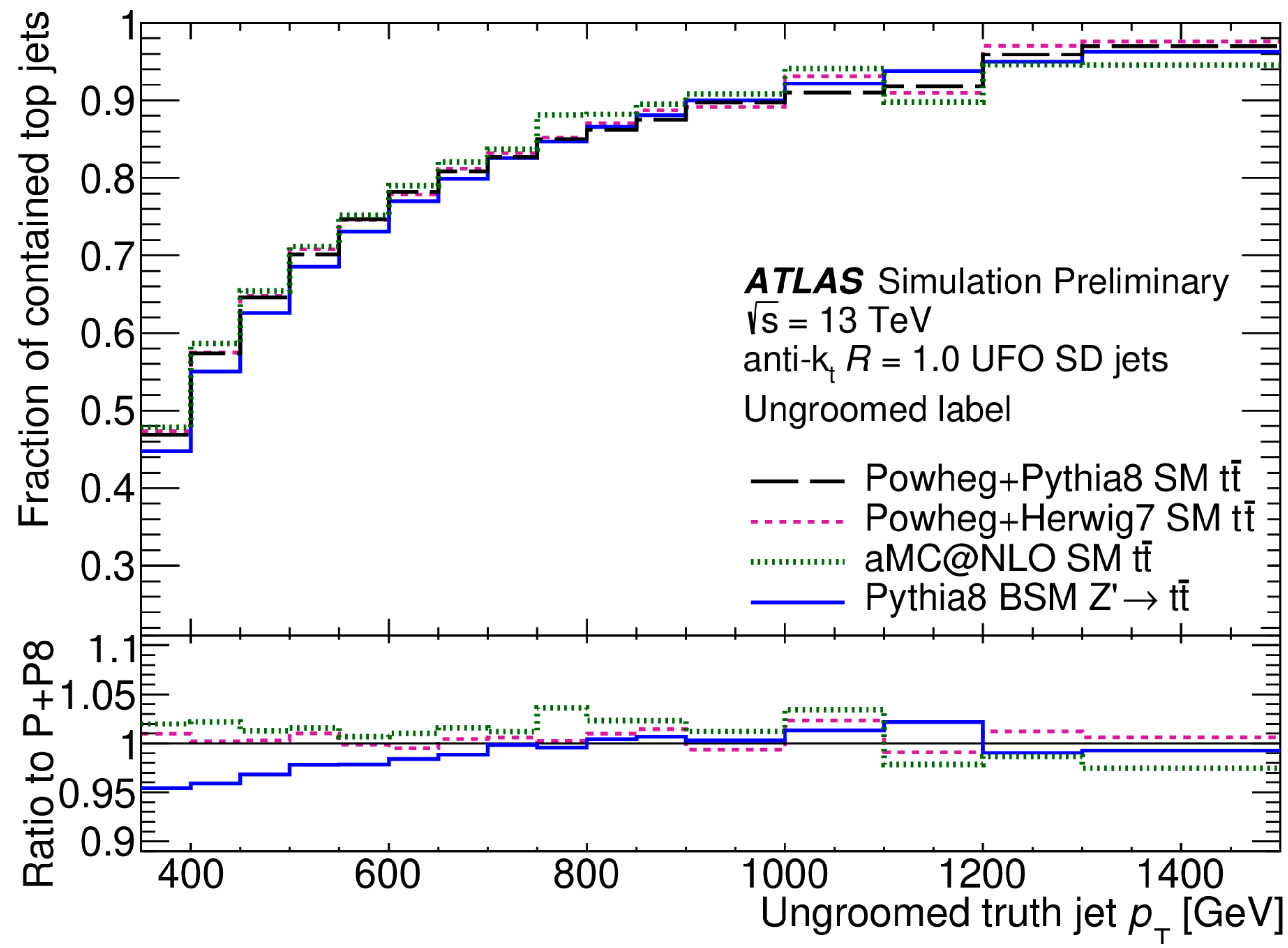
Experimental Considerations in Jet Tagging

Truth Labelling

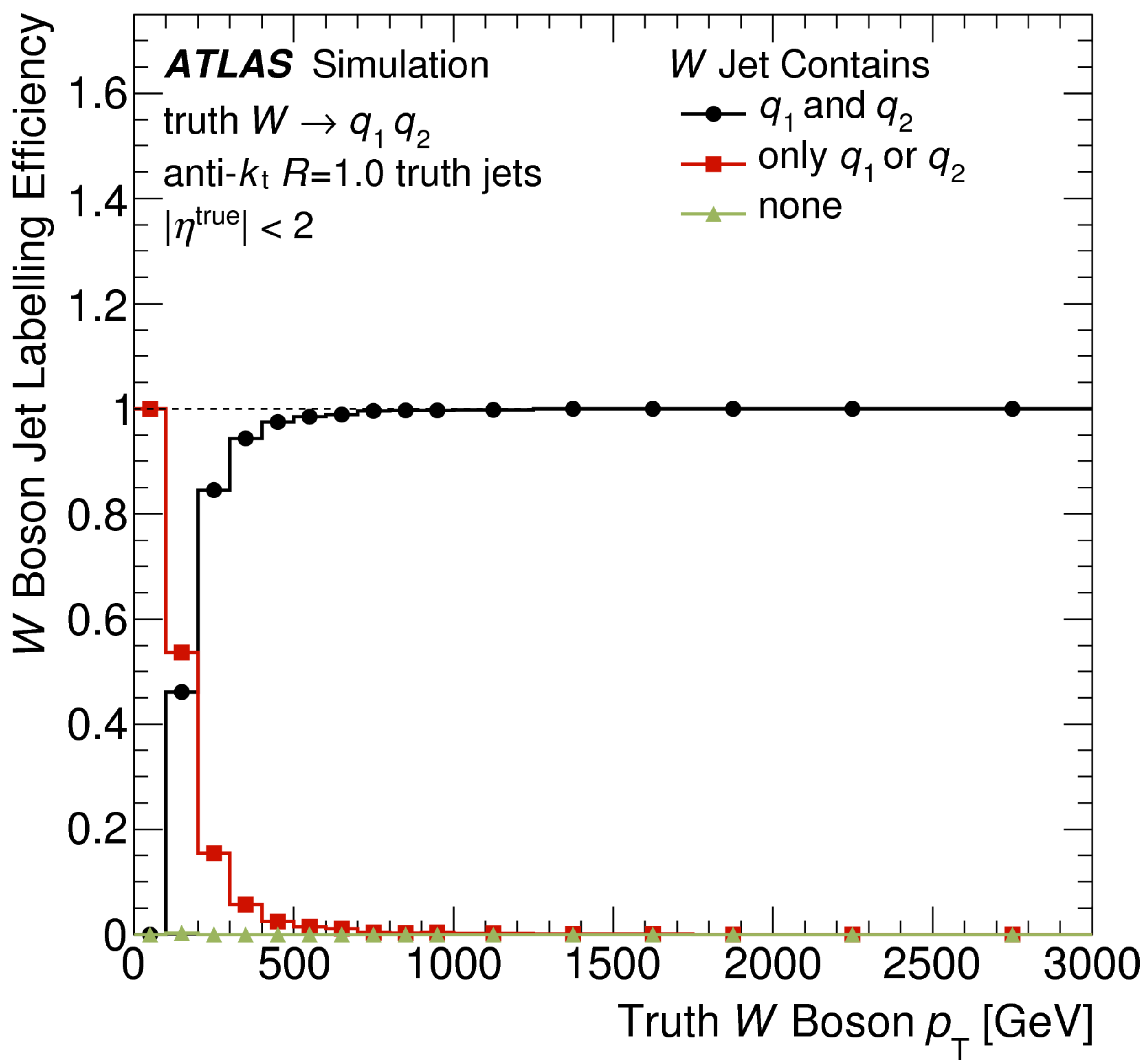
- In order to calculate the efficiency of the taggers, the labelling of the initiating particle is necessary.
- Three step process: match jets to truth jets, truth jets to truth top/W, then partonic decay products of top/W to match jets. Rather generator dependent.
- Update: b-hadron ghost associated, and mass/k_t splitting scale requirments.

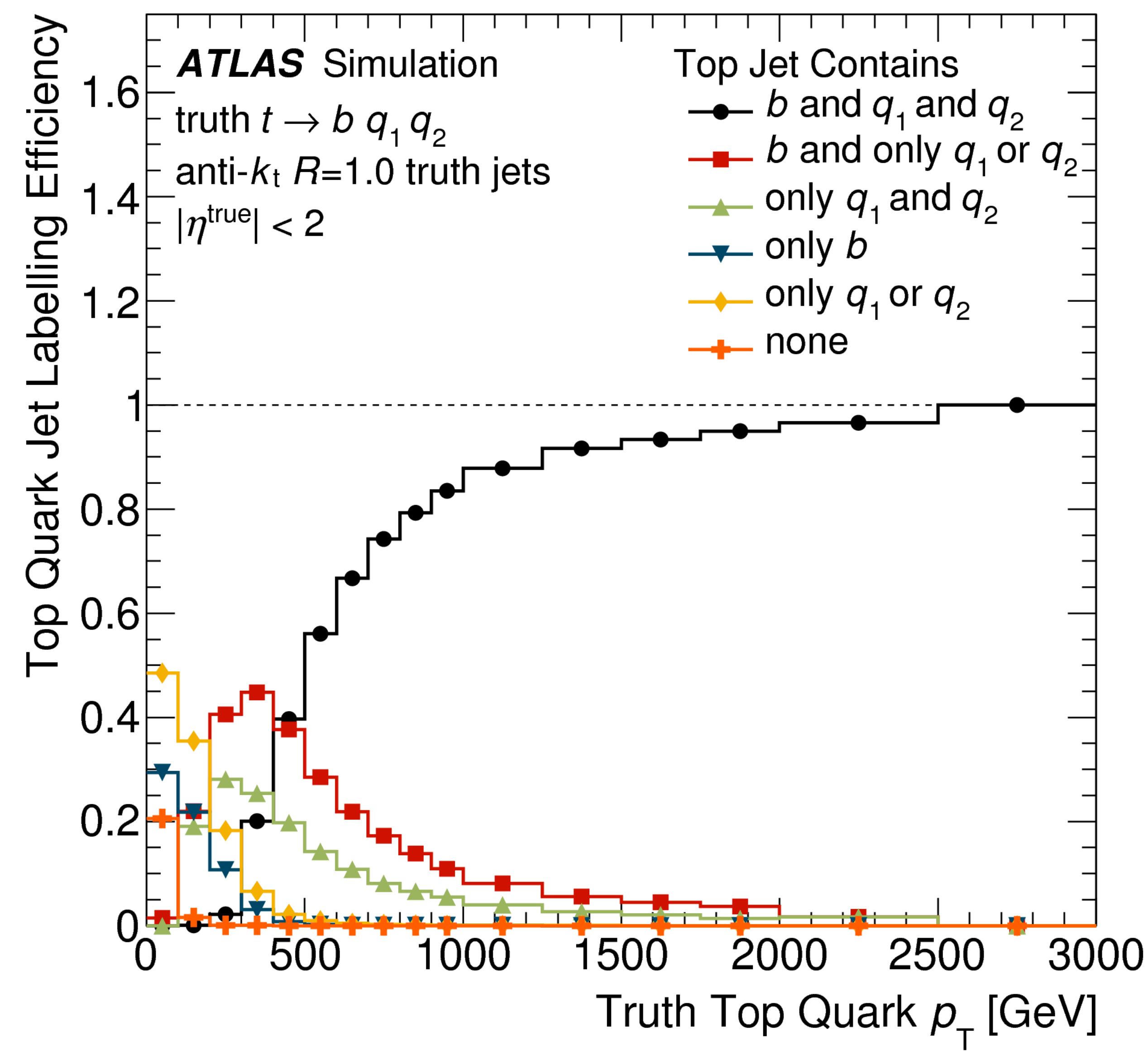
Truth Labelling





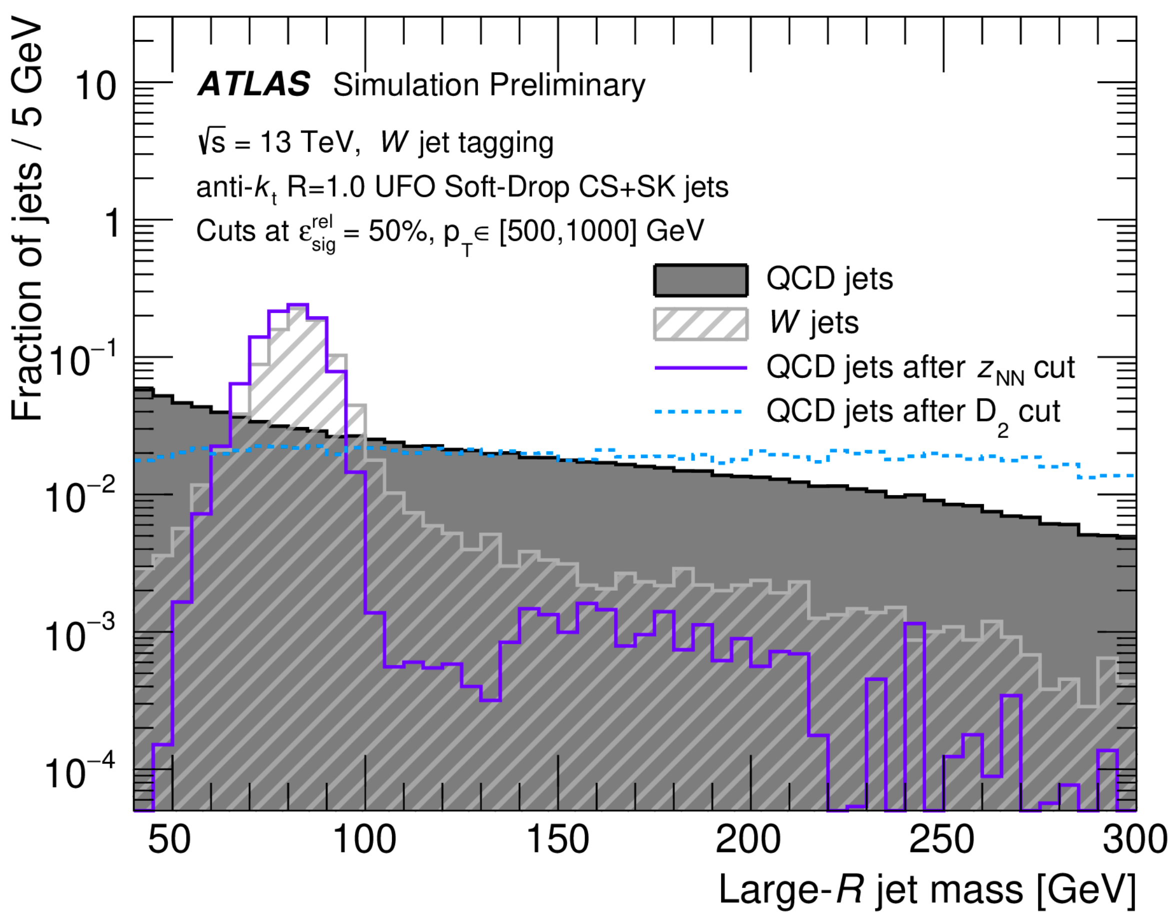
Containment



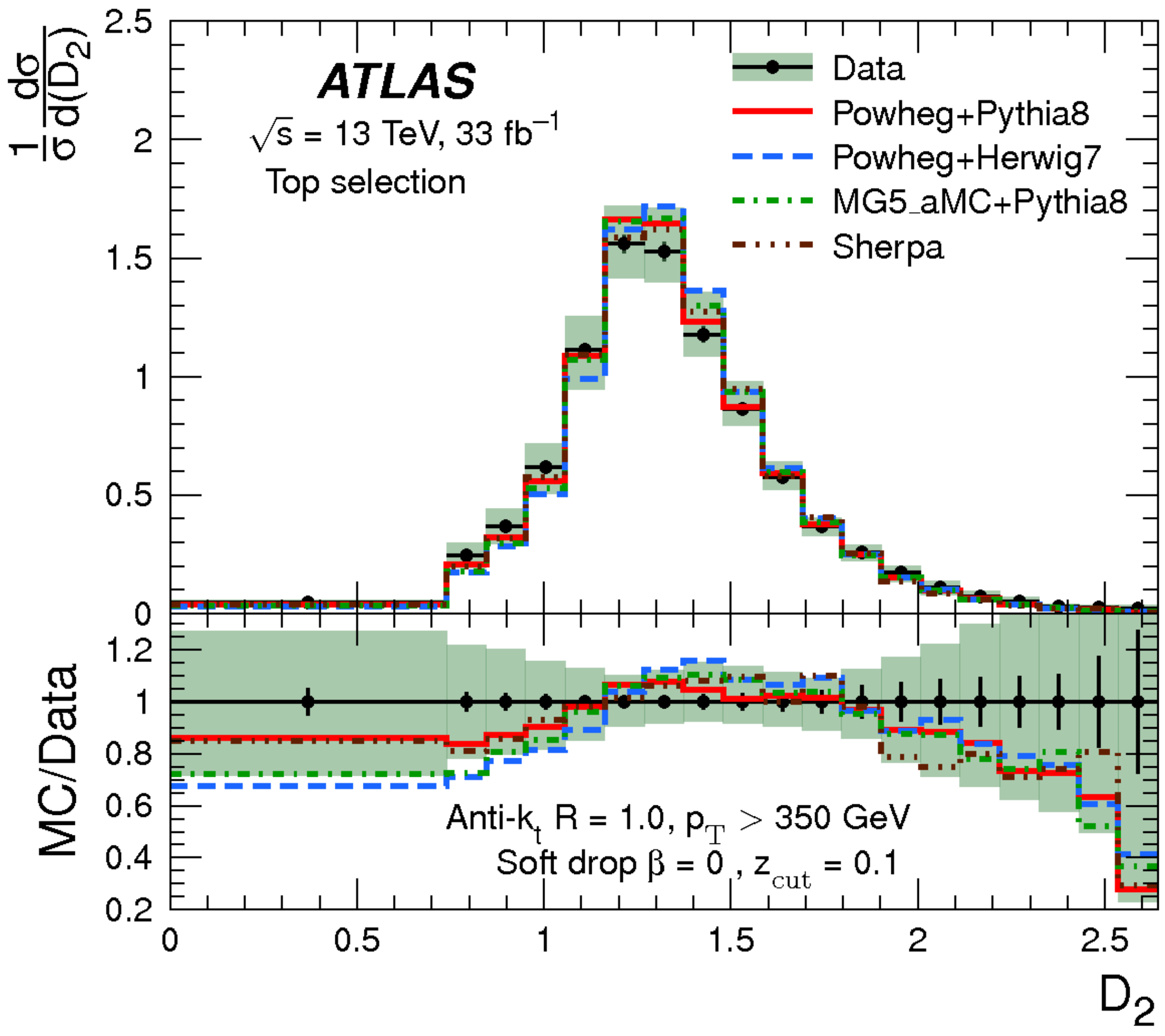


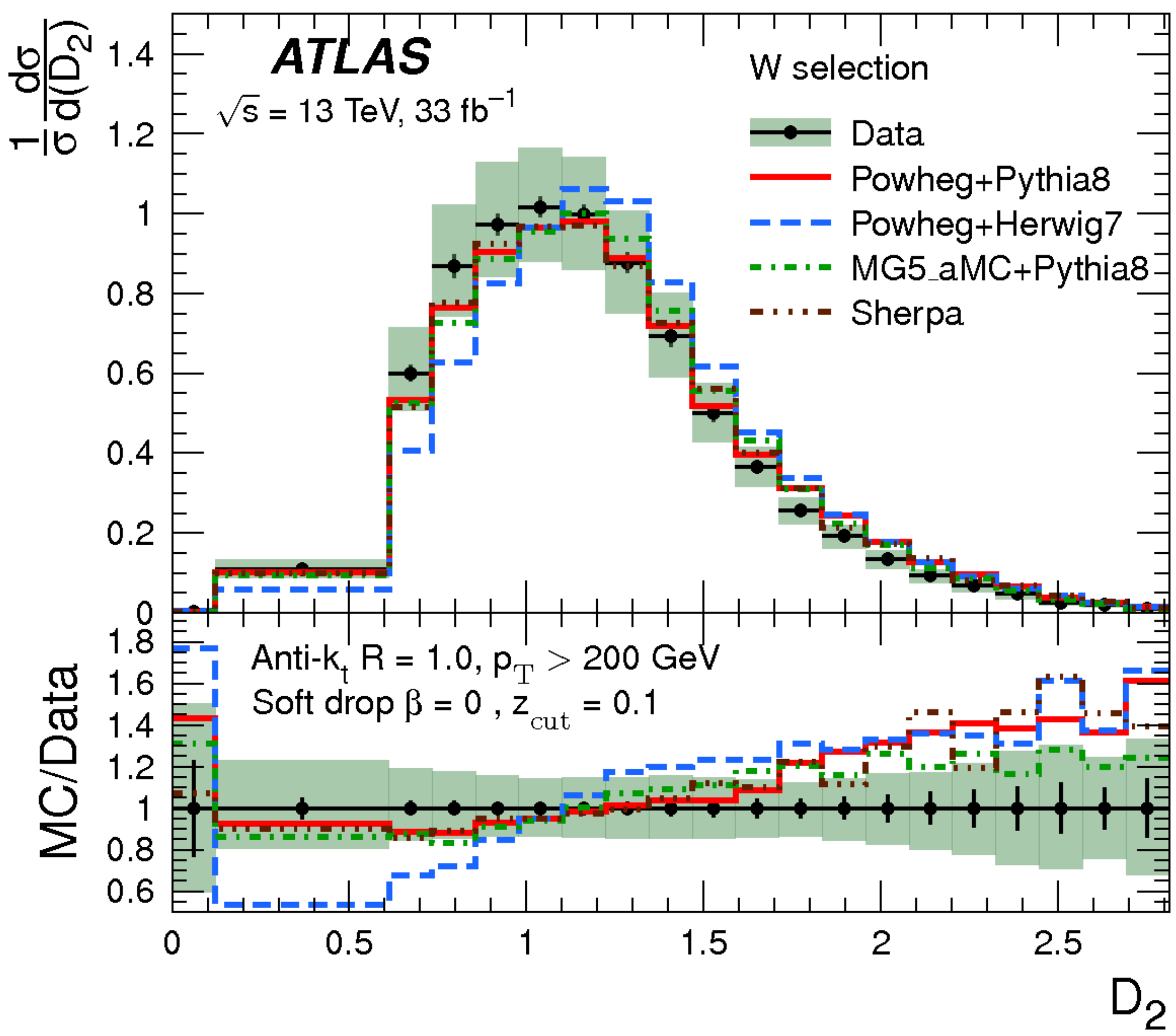
Mass Sculpting

- Jet mass dependence of tagging efficiency
- Difficulty for analyses using sideband for background estimation or bump-hunting

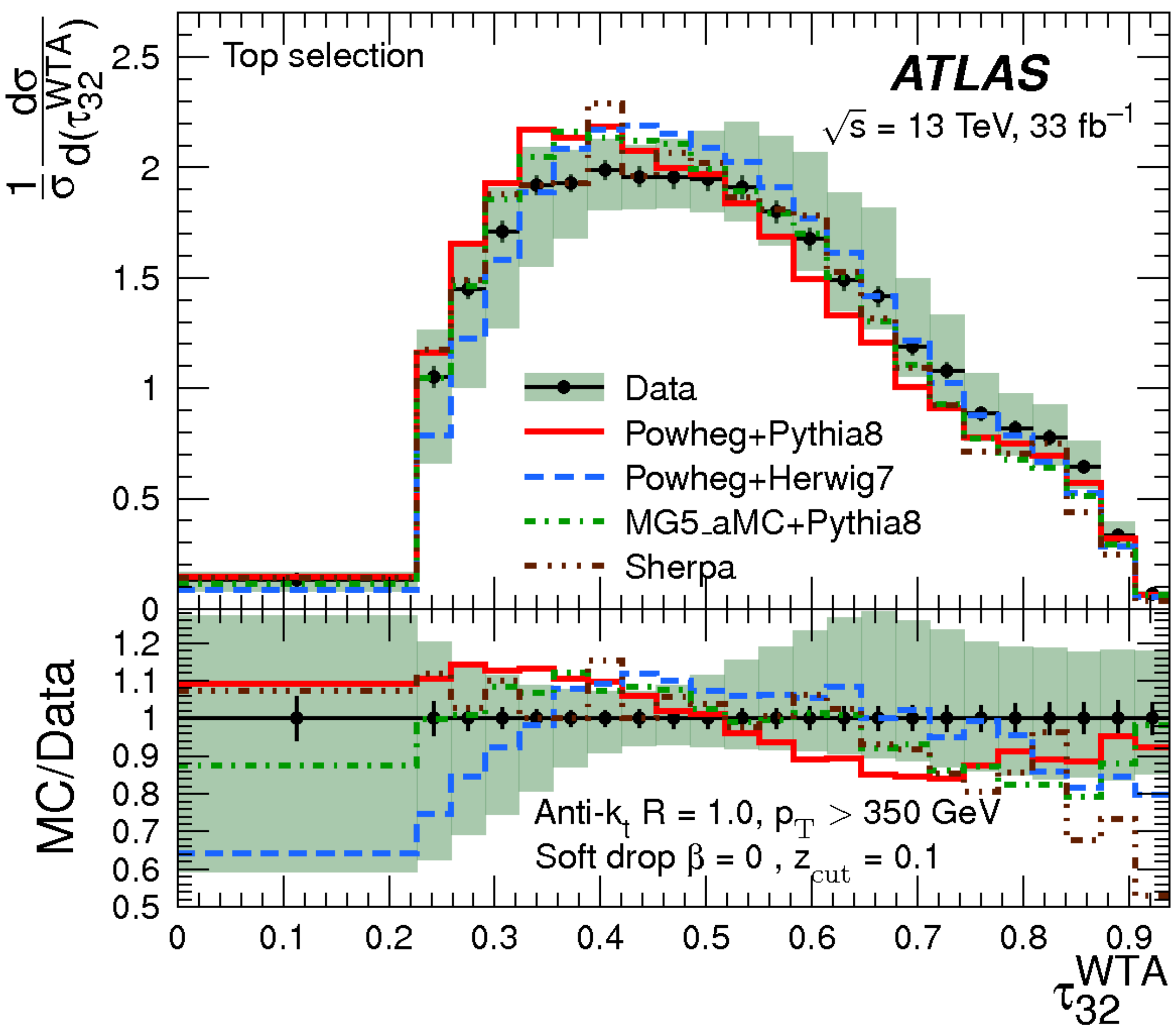


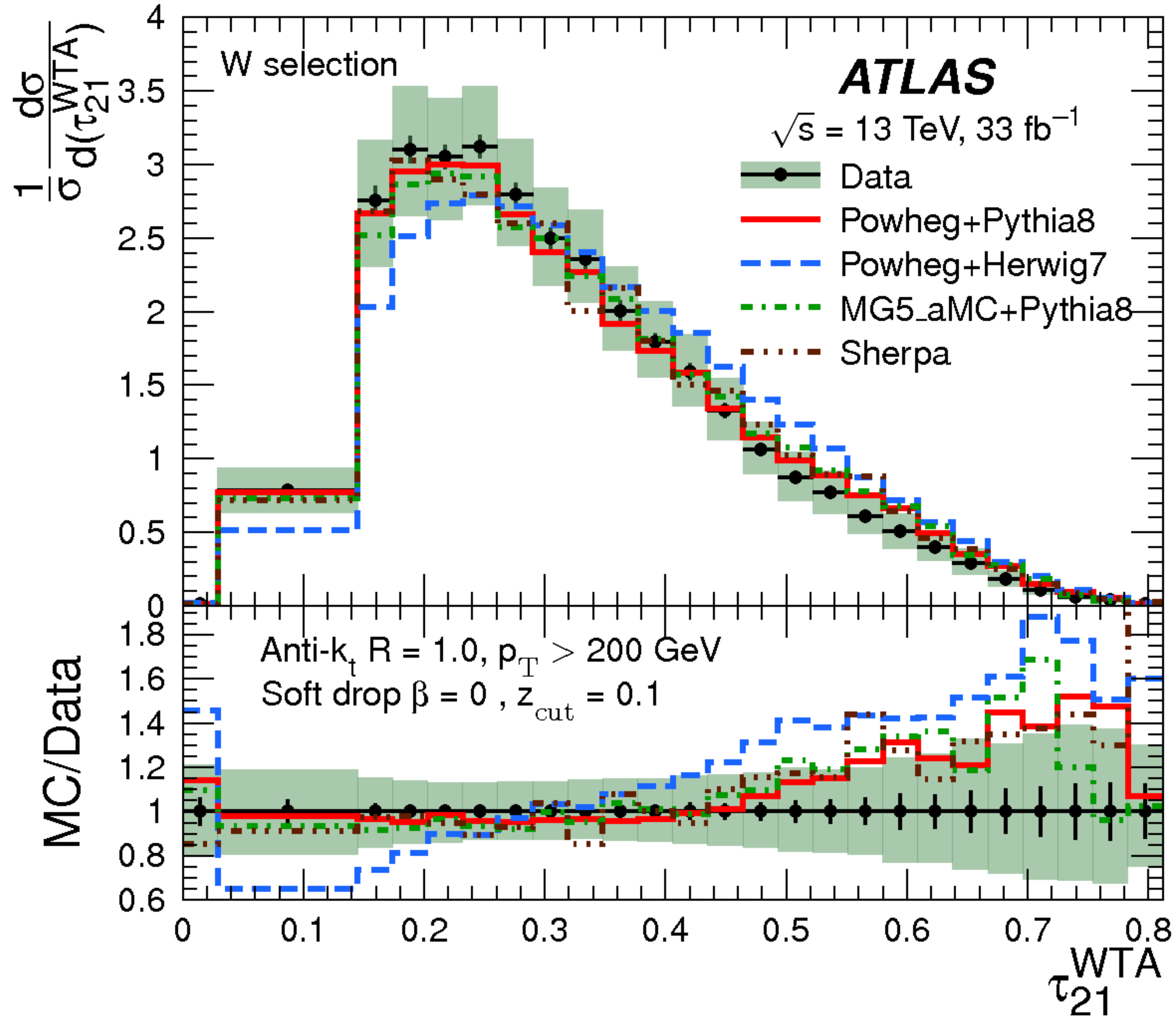
Modelling Dependance





Modelling Dependance





Uncertainties