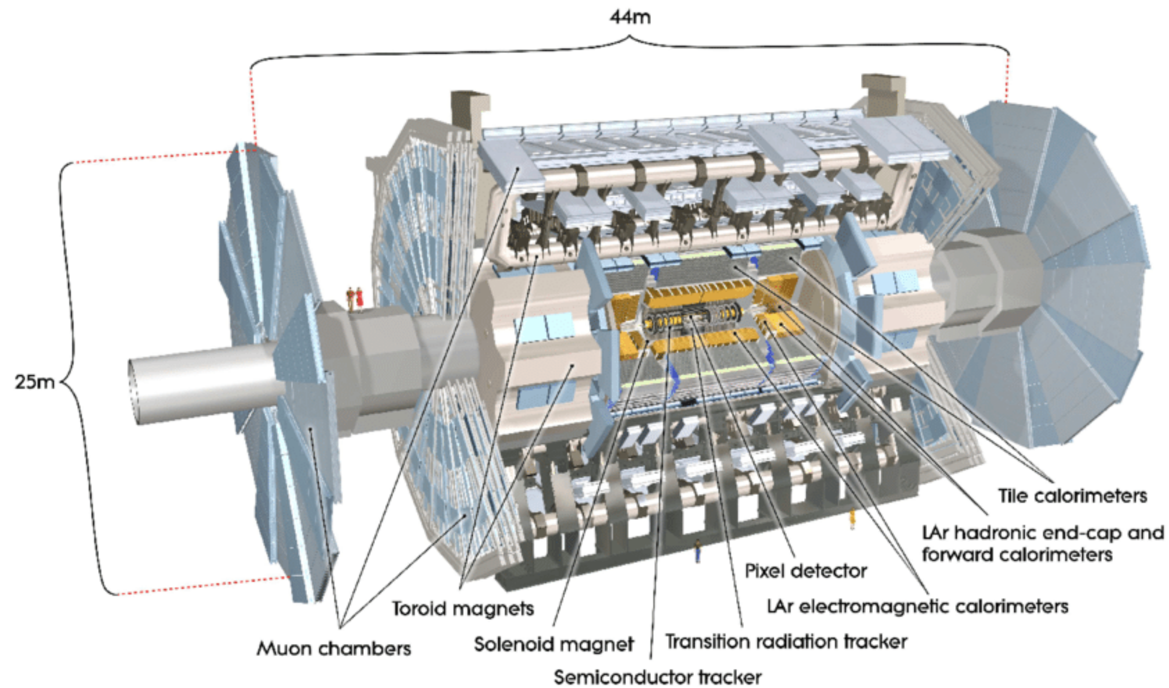
Flavor tagging at the LHC

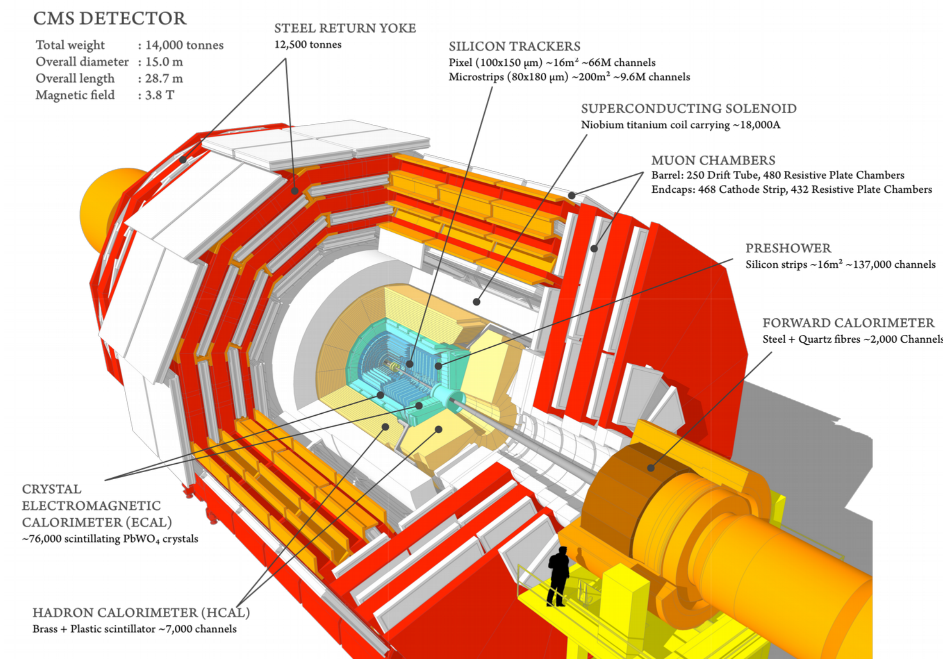
Valerio Dao (CERN)
Silvio Donato (UZH)

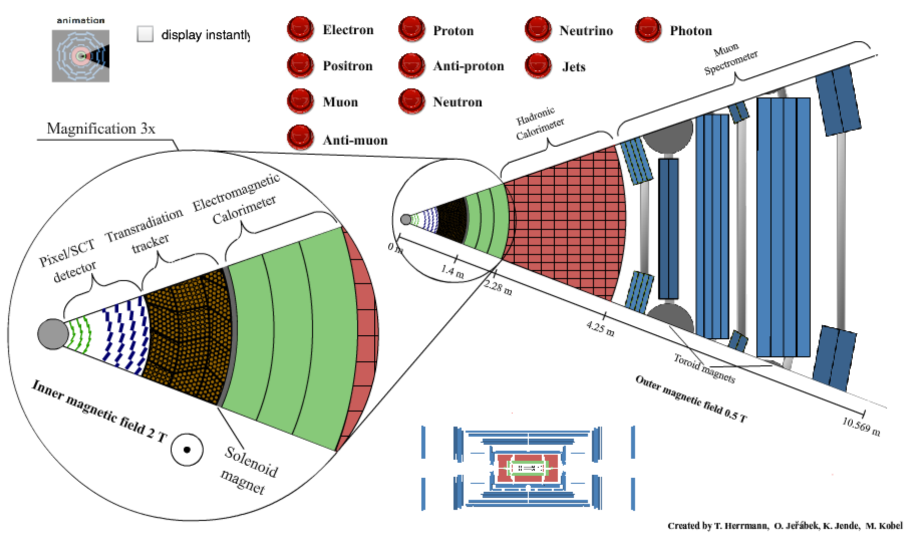
Introduction

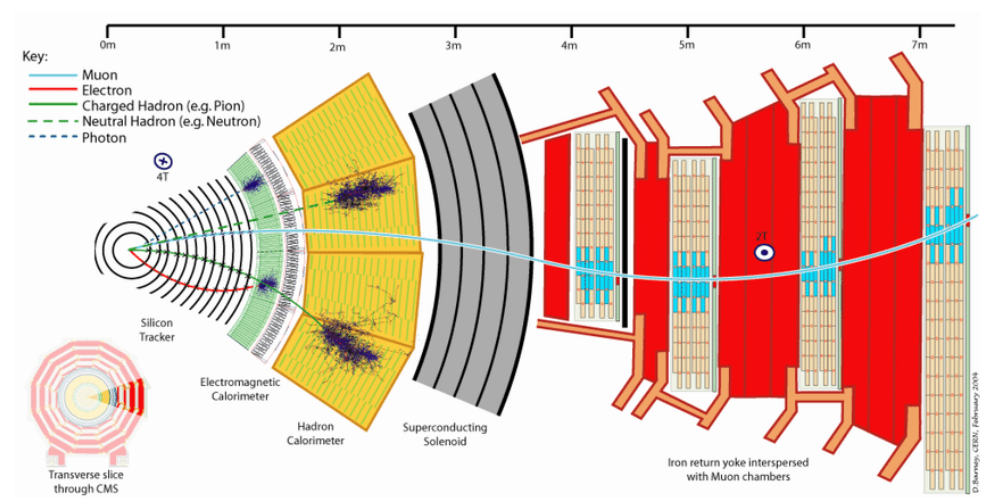
- Flavor tagging is one of the key ingredient of many analysis including ttH.
- **b-tagging** is one of the most discriminating variables of ttH(bb) to reject tt+lights and tt+cc.
- **c-tagger** is currently used in the search for H $\rightarrow$ cc
  - it might be useful to tag jets from W in ttH (BR(W$\rightarrow$cX) ~ 50%).
- **boosted X $\rightarrow$ bb** tagger is currently used for gg $\rightarrow$ boosted H(bb)
  - a natural application would be boosted ttH(bb).
- **quark/gluon discriminator** is currently used to reject QCD in fully hadronic ttH analysis (CMS).

ATLAS and CMS experiments



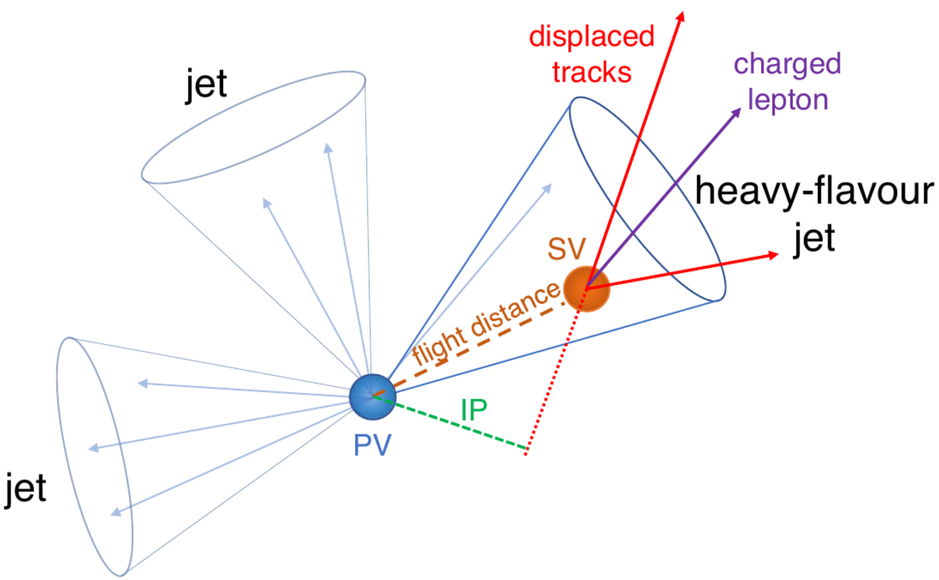






b(c) tagging

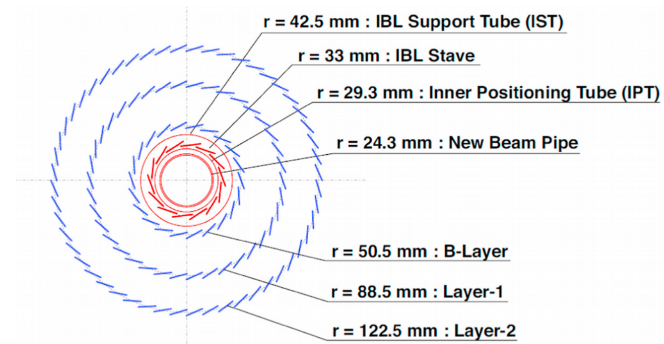
- b(c) tagging is based on B(D) hadron decay features:
  - leptonic decay BR($b \to \mu\ \nu\ X$) ~ 11% + BR($b \to c \to \mu\ \nu\ X$) ~ 10%;
  - sizable lifetime ($c\tau$ ~ 0.45 mm);
  - charged multiplicity (aver. ~ 5.0) and invariant mass (~ 5 GeV).
- Discriminating variables:
  - soft lepton (e,μ);
  - track impact parameter (IP);
  - secondary vertex/vertices (SV).

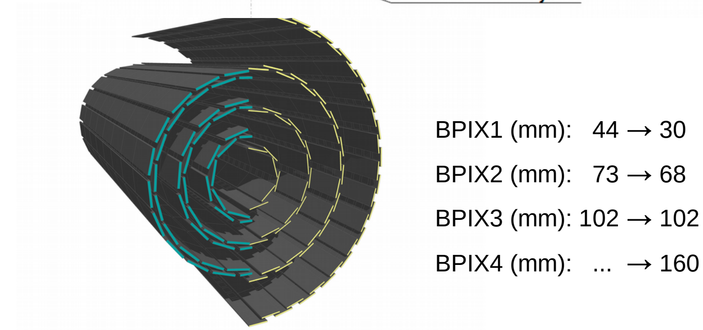


- Fake: long-lived hadron decay (eg. K_S^0 or Λ), material interaction.

b/c tagging

- b/c tagging performance strongly depends on the **pixel detector**.
- **ATLAS** inserted a new pixel layer (Inner Barrel Layer) during LS1.
- **CMS** replaced the whole pixel detector (phase-1 upgrade) in 2016-17.





- **Jet flavor definition**:
  - **ATLAS**: AK4 calo jets p_T>20 GeV, $|\eta|<2.5$, pile-up jet rejection(JVT) (eff. 92% fake 2%). b jet if B hadron in $\Delta R<0.3$.
  - **CMS**: AK4 PF jets p_T>20 GeV, $|\eta|<2.4$. b jet if generator B hadron is into the AK4 jet (ghost association) and $\Delta R<0.25$ from a gen jet p_T>8 GeV.

Low level taggers

- **Impact parameter**:
  - **ATLAS**: log-likelihood ratio (LLR) based on **IP2D** and **IP3D** of jet tracks, Recurrent Neural Network Track-based tagger (**RNNIP**) → output p_b,p_u,p_c and p_τ.
  - **CMS**: JetProbability (**JP**)→ LLR of 3D IP significance. **JBP** → only 4 tracks considered.
- **Secondary vertex**:
  - **ATLAS**: single displaced vertex (**SV1**) starts from two-track vertices, **JetFitter**: cascade vertex algorithm, it includes single prong vertex with jet axis.
  - **CMS**: Adaptive Vertex Fitter (Run-1) → **Inclusive Vertex Fitter** (Run-2): it doesn't use jet direction to find the secondary vertex. Originally developed for g → bb measurement.
- **Soft lepton tagger**:
  - **ATLAS**: Soft Muon Tagger. BDT discriminant (Kinematic + track quality)
  - **CMS**: Soft Muon and Soft Electron tagger. BDT discriminant (Kinematic + elect. ID)

High level taggers

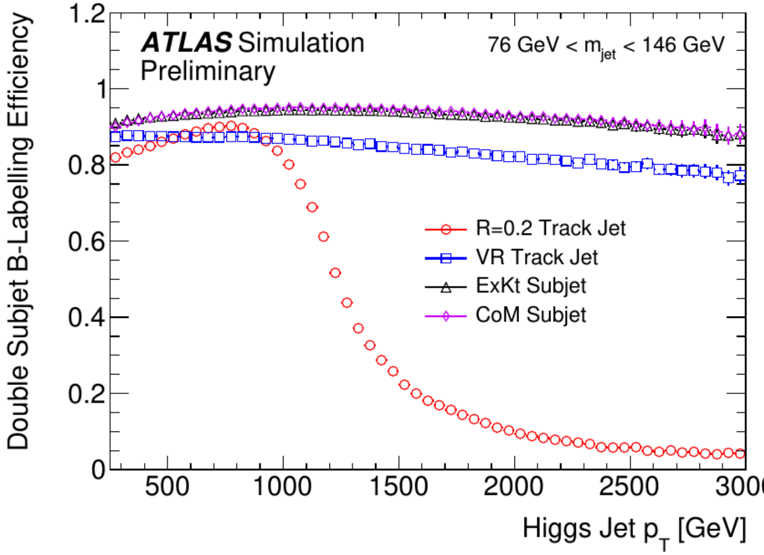
- **ATLAS**:
  - Training sample: hybrid tt + Z'.
  - BDT tagger: kinematic variables (reweighted to avoid correlation) IP2D and IP3D, sec. vertex (SV1, JetFitter) (**MV2**)
    + soft muon tagger (**MV2Mu**)
    + RNNIP (**MV2MuRnn**)
  - BDT **c-tagger**: as MV2 trained against b and light, optimized for c-tagging (JetFitter with 1 vertex, adding track rapidities)
  - Deep Neural Network (**DL1**, **DL1Mu, DL1MuRnn**)
    - Inputs: MV2/MV2Mu/MV2MuRnn inputs + c-tagging variables soft muon tagger replaced by its inputs.
    - Output: p_b,p_u,p_c.
    - Future: Study techniques for systematic and pileup mitigation. RNNIP inputs fully integrated in DL1 training.

High level taggers

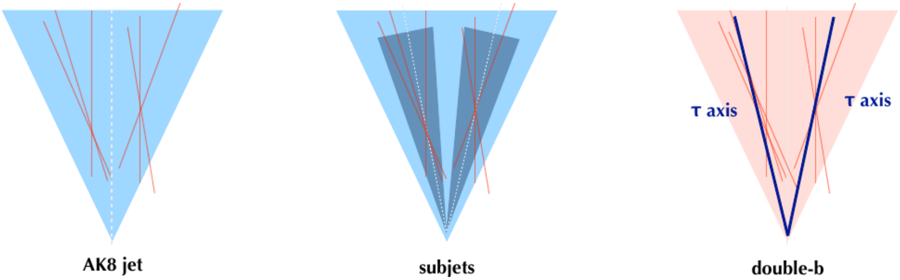
- **CMS**:
  - Training sample: hybrid tt + multijet.
  - **Combined Secondary Vertex** (CSVv2)
    - no intermediate tagger. 18 input variables used (based on kinematic and IP and SV).
  - **DeepCSV**
    - similar input to CSVv2, more layers, more tracks considered, 5 outputs (p_{bb}, p_b, p_c, p_{cc}, p_{light}).
  - **CMVA**v2: CSV + soft muon + soft electron + JP + JPB tagger.
  - **c-tagger**: CvsL and CvsB tagger (very similar to CSV); DeepCSV c-tagger → better performance.
  - Future: **DeepFlavour** uses properties of charged and neutral PF candidates, SV, without any specific preselection (more tracks).

Boosted X→bb taggers

- **ATLAS**: b-tagging on track subjets (ΔR=0.2):
  - several possible tagging (leading subjet pT, highest/lower b-tagging, ...).
  - new subjet definition improves X->bb efficiency (Variable Radius, VR)



- **CMS** AK8 jet, soft drop algorithm:
  - tagging on subjects (as ATLAS);
  - new double b-tagging exploiting correlation between subjects:
    - τ_N algo to define parton direction;
    - track associated minimizing track τ-axis distance;
    - use IP and SV variables, including correlation;
    - z variable to reject g → bb.

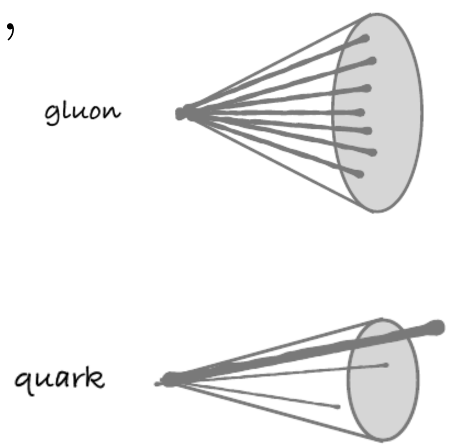


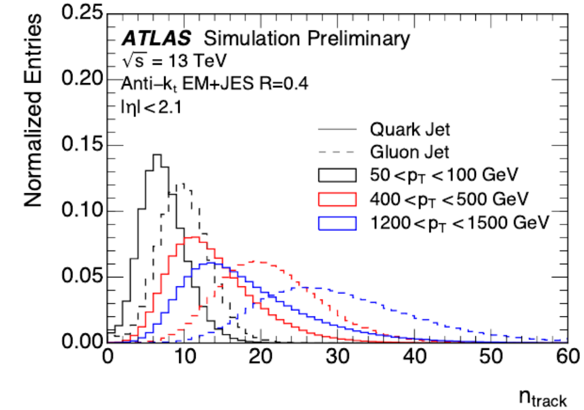
$$\tau_N = \frac{1}{d_0} \sum_k p_T^k \min(\Delta R_{1,k}, \ldots, \Delta R_{N,k}) \quad d_0 = \sum_k p_T^k R_0$$

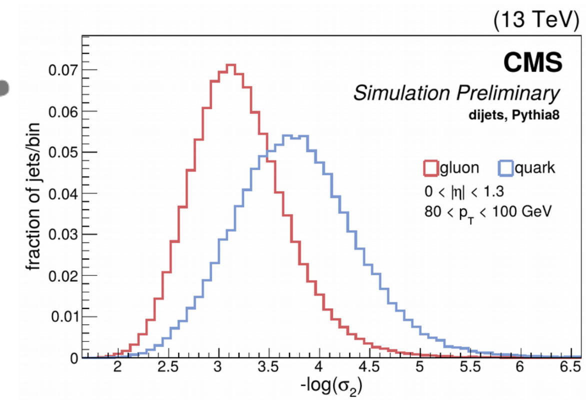
$$z = \Delta R(SV_0, SV_1) \frac{p_T(SV_1)}{m(SV_0, SV_1)}$$

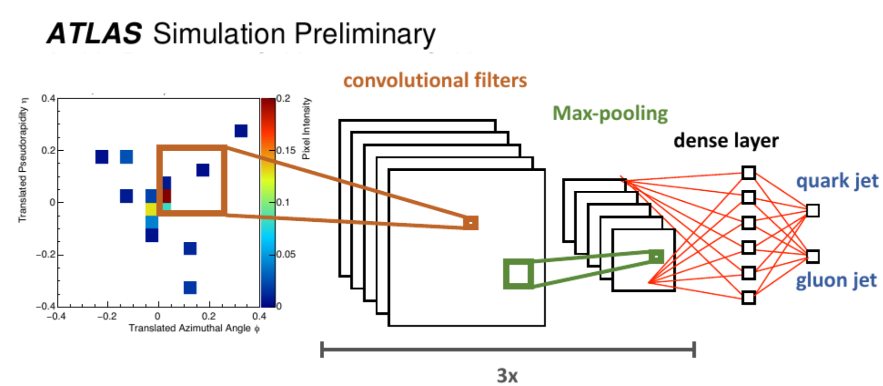
Quark/gluon tagger

- **ATLAS:** likelihood ratio using number of track per jet in (p_T,η) bins,
- **CMS**: likelihood ratio of
  - number of tracks,
  - $p_T D = \frac{\sqrt{\sum_i p_{T,i}^2}}{\sum_i p_{T,i}}$,
  - σ_2 is the ellipse minor axis.
- Future:
  - ATLAS: ECAL towers + tracks images (Convolutional Neural Network tagger);
  - CMS: neutral and charged PF candidates as DNN inputs (DeepJet).



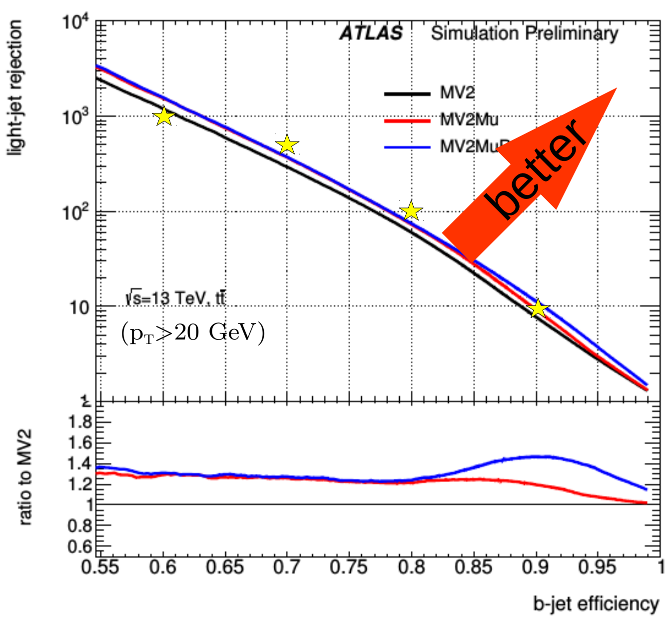


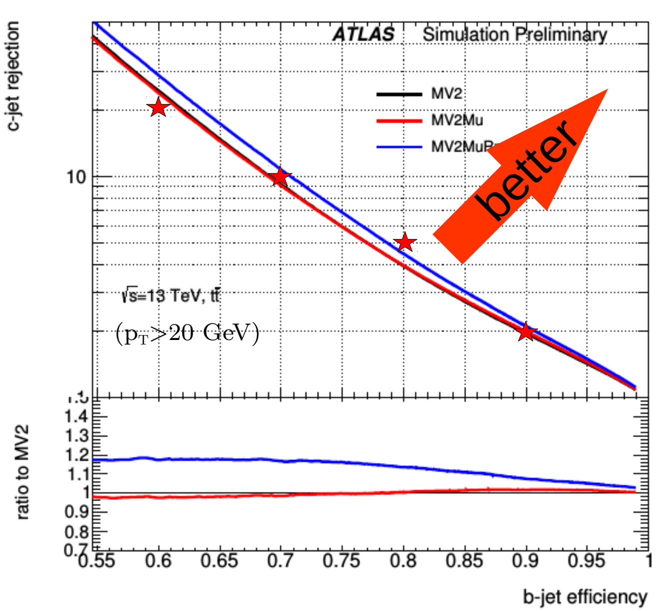


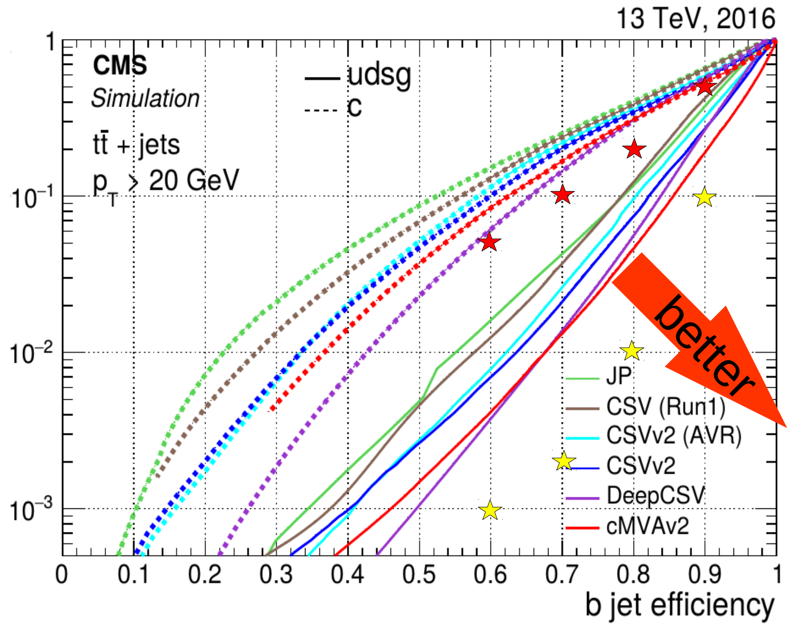


Performance b-tagger

- Plots: b-jet efficiency vs non-b jet rejection (ATLAS) or fake rate (CMS).
- Example b efficiency 80%:
  - light rejection ~ 100; c rejection ~ 5.
- In 2016, ATLAS b-tagging performance was significantly better than CMS.

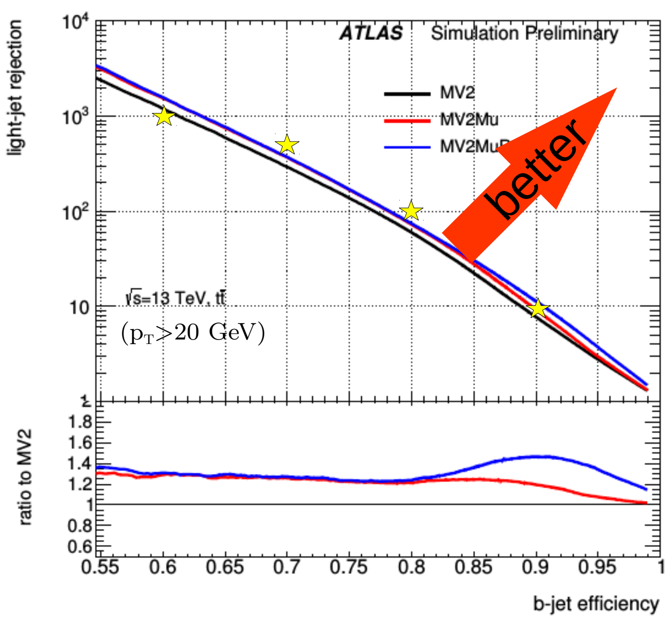


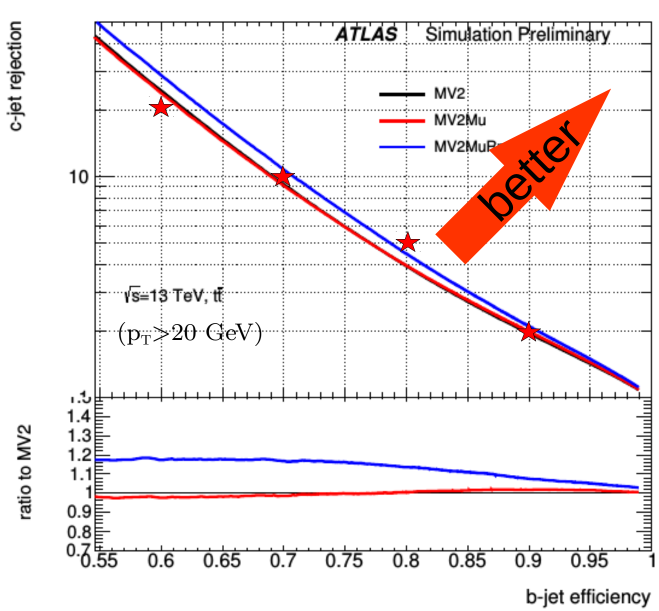


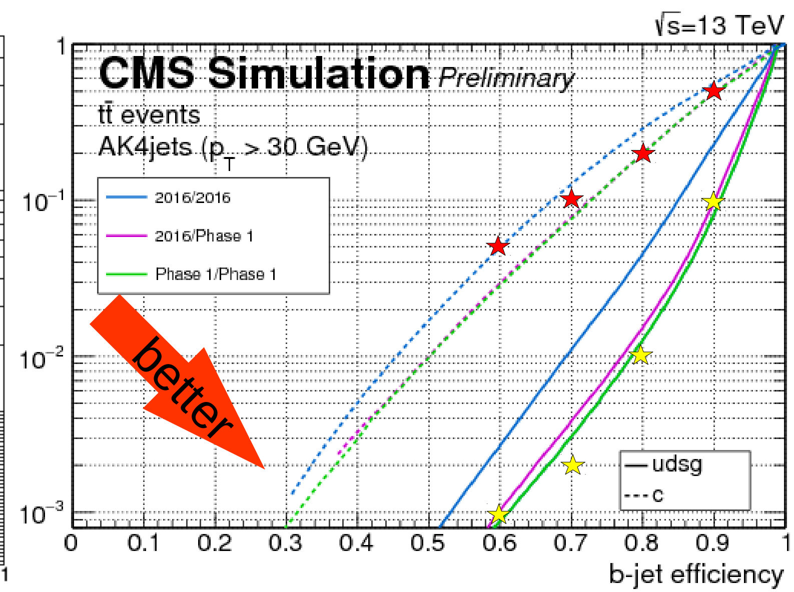


Performance b-tagger

- Plots: b-jet efficiency vs non-b jet rejection (ATLAS) or fake rate (CMS).
- Example b efficiency 80%:
  - light rejection ~ 100; c rejection ~ 5.
- After phase-1 upgrade, the two experiments have a comparable performance similar (in simulation) → warning: plot with different p_T range!

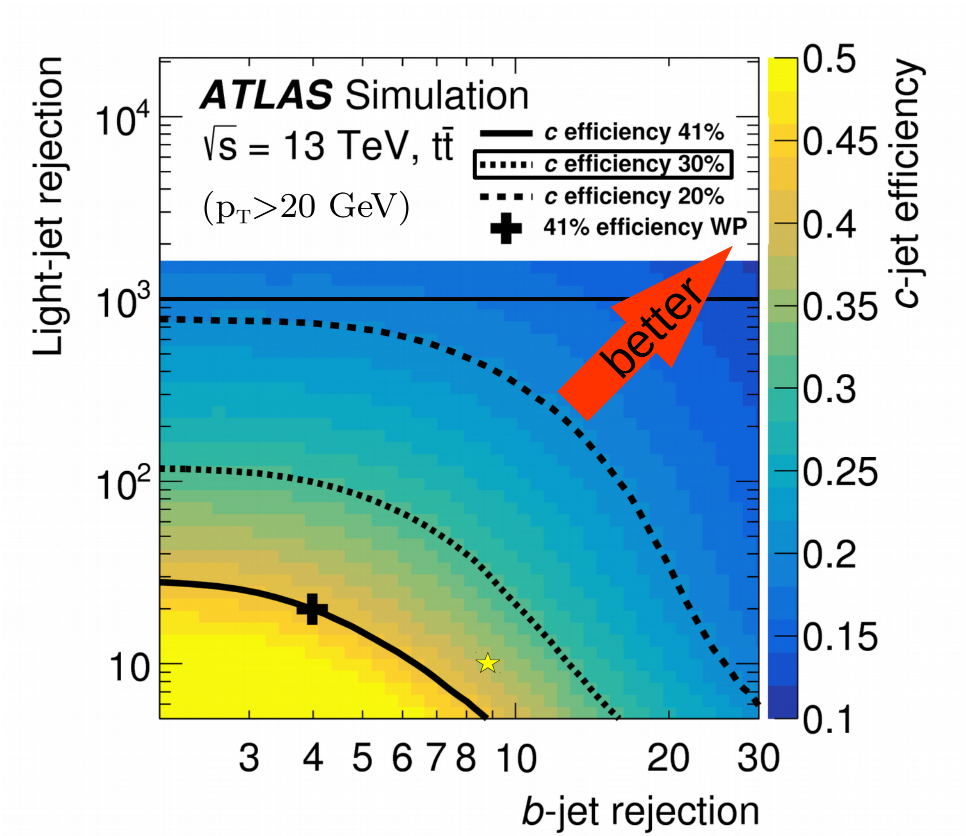


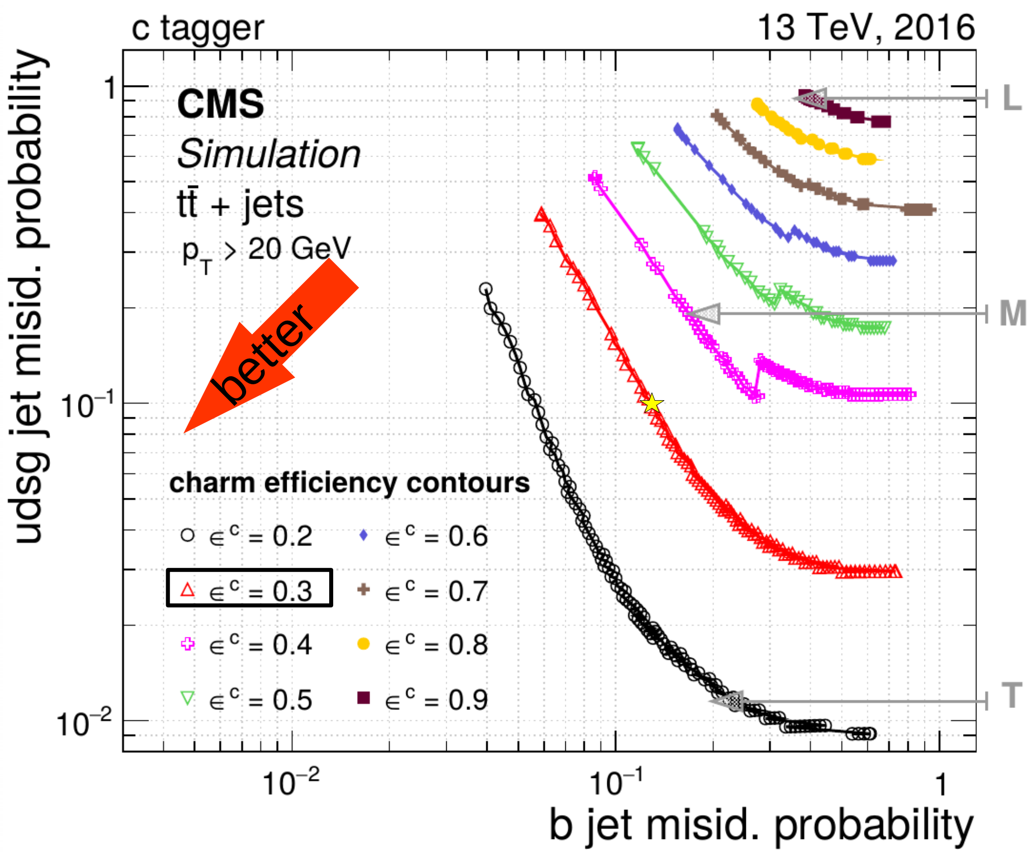




c-tagger

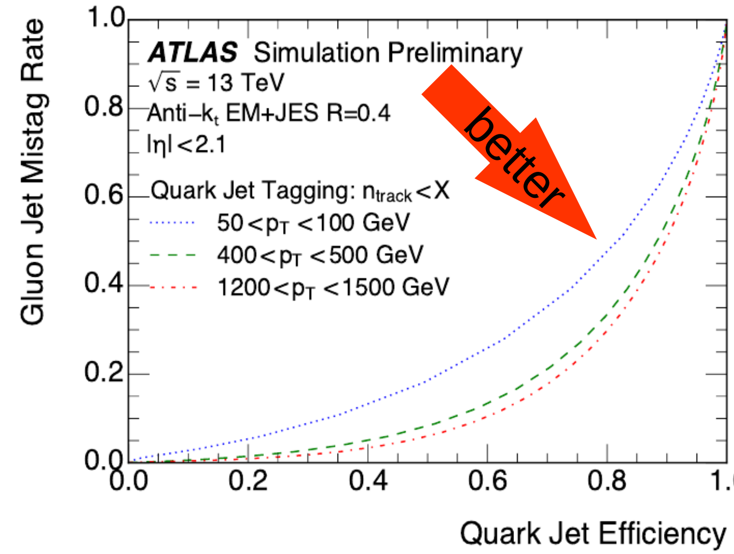
- CMS plot obtained with Phase-0 pixel detector.
- Working point used in $H \rightarrow cc$ analysis (ATLAS)
  - c-efficiency ~41%; light rejection ~20; b rejection ~4.

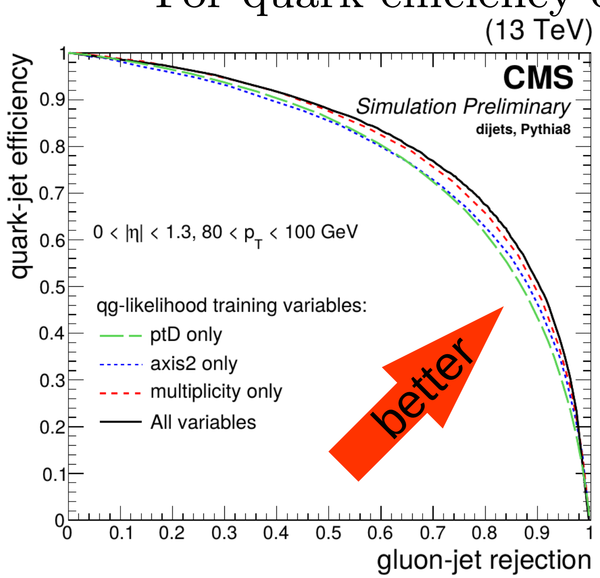


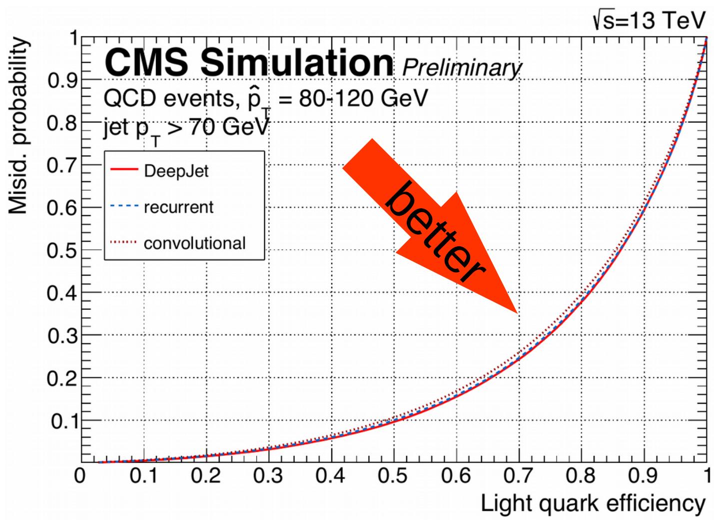


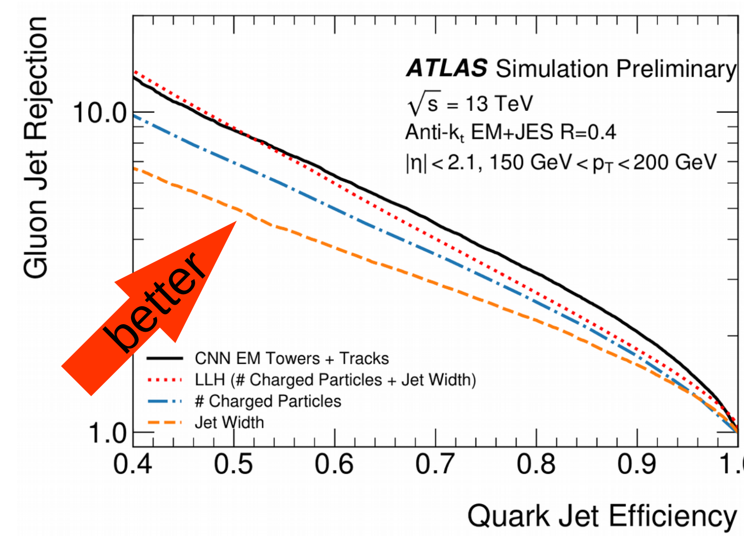
Quark/gluon tagger

- Different jet selection → plots cannot be compared.
- Charged multiplicity is the most discriminating variables
  - CMS q/g discriminator shows a slight improvement including other variables.
- CNN and DeepJet improvement is about ~10-20% in gluon rejection.
  - For quark efficiency of ~70%, gluon fake rate is ~20%.



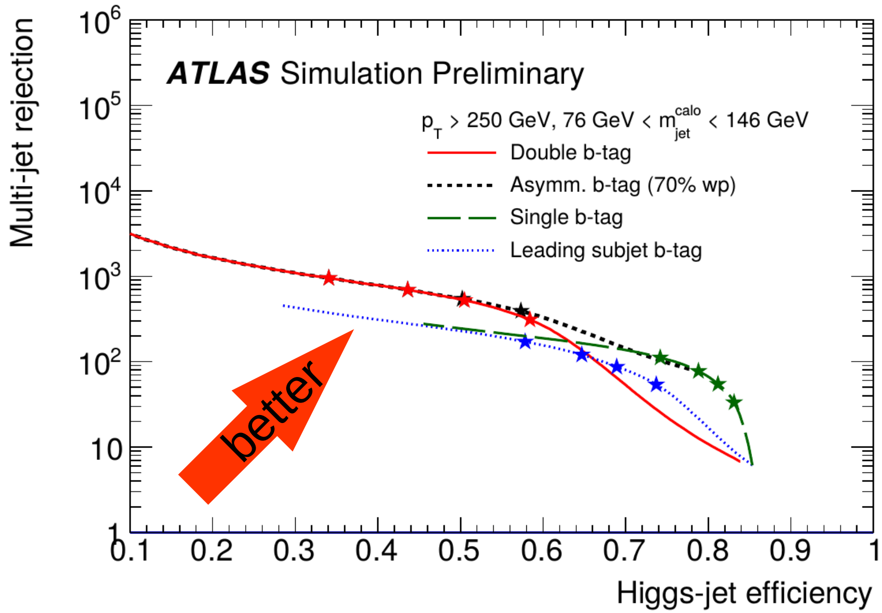


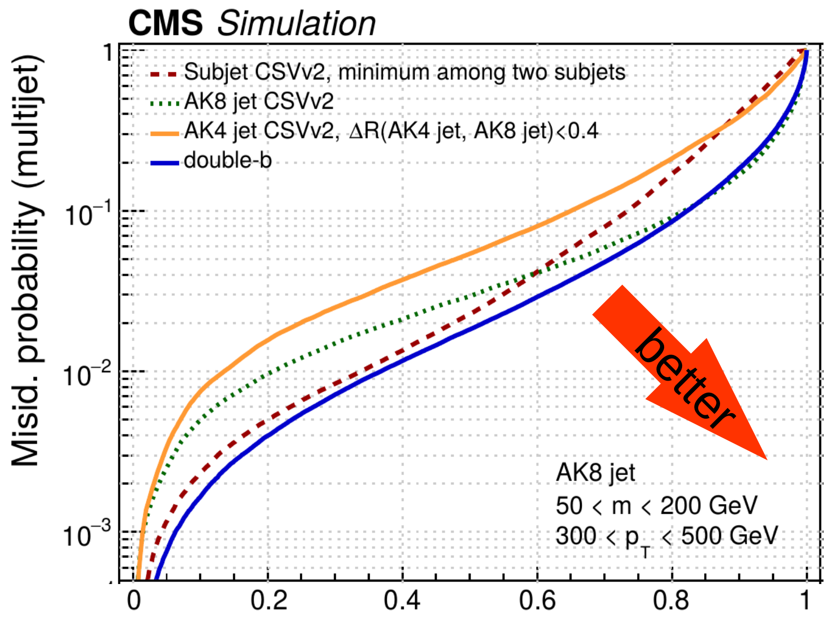




Boosted X → bb

- Double subjet b-tag is better than single subjet b-tag in large multijet rejection region.
- As expected double-b outperforms other taggers, especially in high H → bb efficiency region.
  - multijet rejection of ~500 at H(bb) efficiency of ~50%.
- CMS plot shows only tagging efficiency (plot obtained 2016 pixels)
- Warning: different kinematic cuts → plots cannot be compared.





Reference - CMS

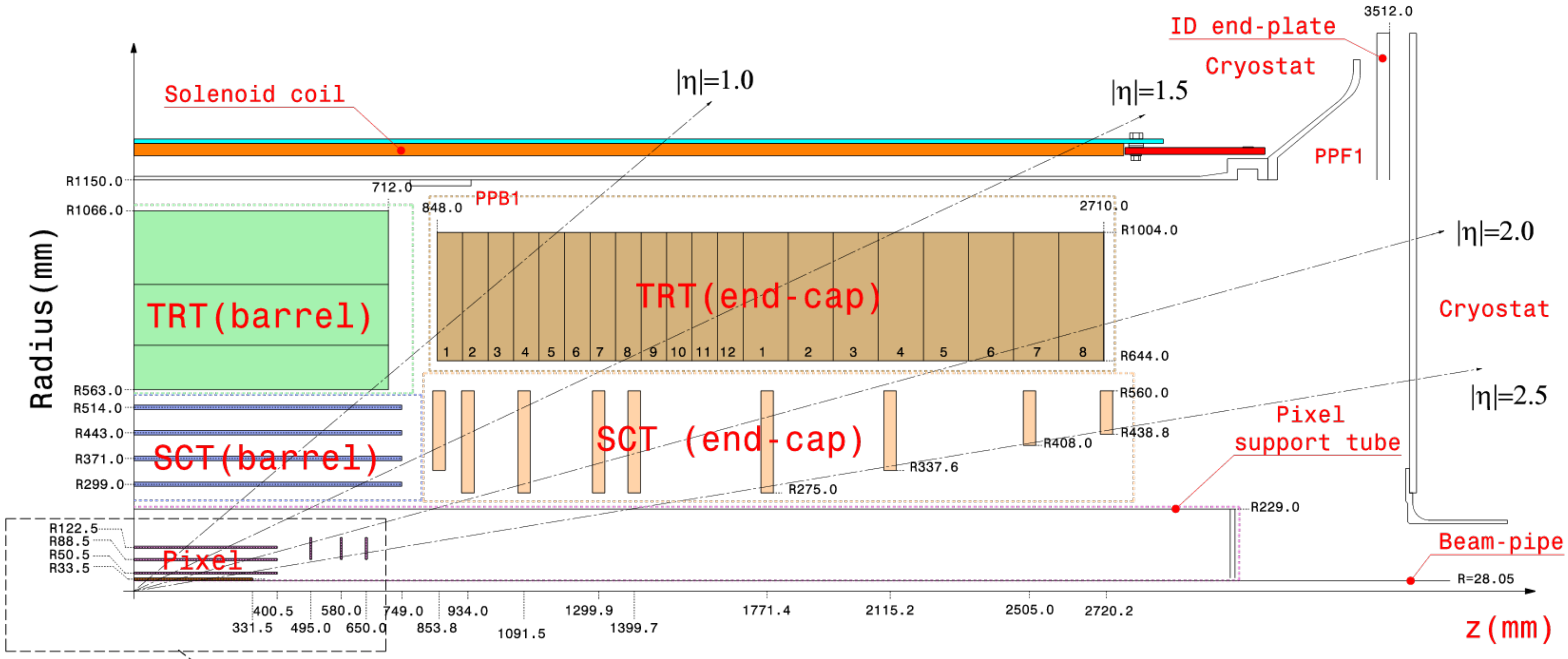
- "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV", JINST 13 (2018) no.05, P05011
- "CMS Phase 1 heavy flavour identification performance and developments", CMS DP-2017/013
- "Heavy flavor identification at CMS with deep neural networks", CMS DP-2017/005
- "Jet algorithms performance in 13 TeV data", CMS-PAS-JME-16-003
- "New developments for Jet Substructure", CMS DP-2017/027

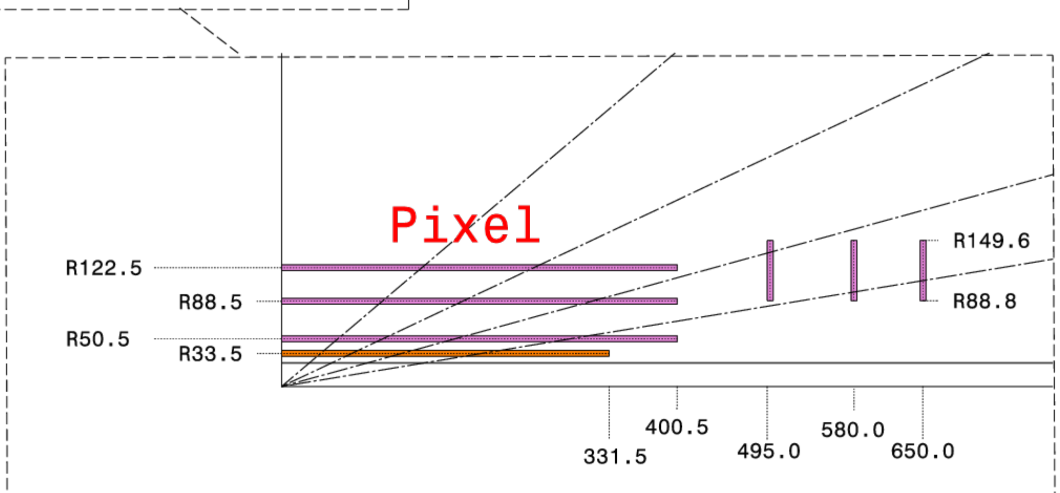
Reference - ATLAS

- "Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run", ATL-PHYS-PUB-2017-013
- "Boosted Higgs ($\rightarrow b\bar{b}$) Boson Identification with the ATLAS Detector at s = 13 TeV", ATLAS-CONF-2016-039
- "Variable Radius, Exclusive-kT , and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS", ATL-PHYS-PUB-2017-010
- "A new tagger for the charge identification of b-jets", ATL-PHYS-PUB-2015-040
- "Quark versus Gluon Jet Tagging Using Charged-Particle Constituent Multiplicity with the ATLAS Detector", ATL-PHYS-PUB-2017-009
- "Quark versus Gluon Jet Tagging Using Jet Images with the ATLAS Detector", ATL-PHYS-PUB-2017-017.

Backup

ATLAS tracker r-z sect.

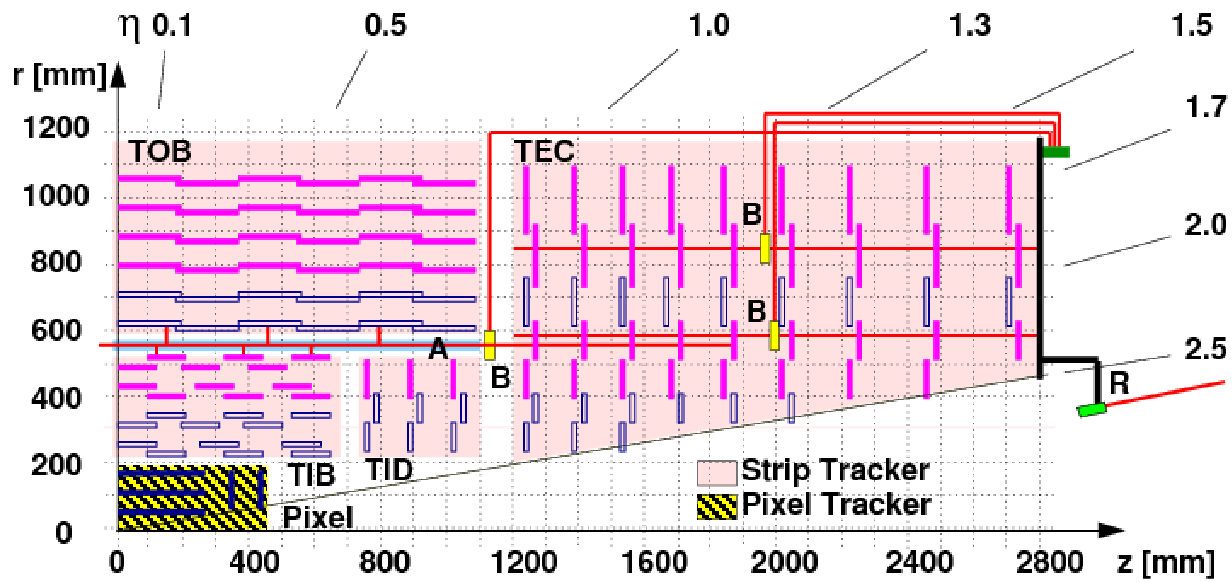


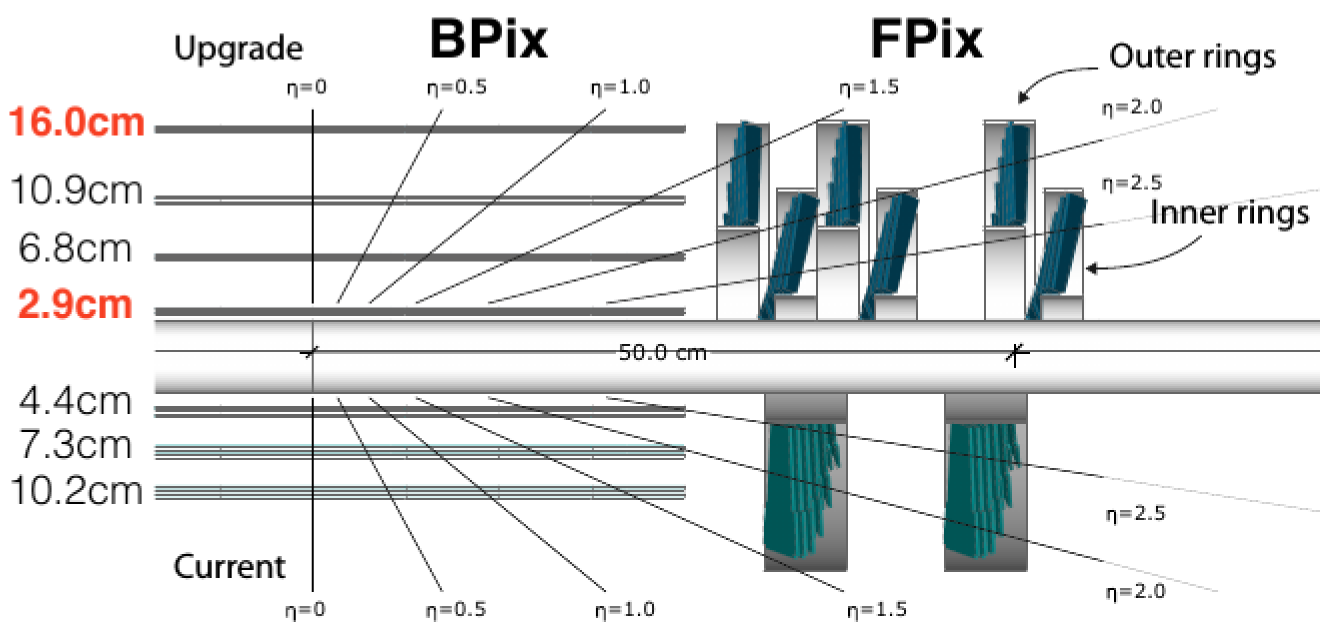


Envelopes

Pixel		31<R<242 (mm)
SCT barrel		255<R<549 (mm)
SCT end-cap		251<R<610 (mm)
TRT barrel		554<R<1082 (mm)
TRT end-cap		617<R<1106 (mm)

CMS tracker r-z section





Working points 2016

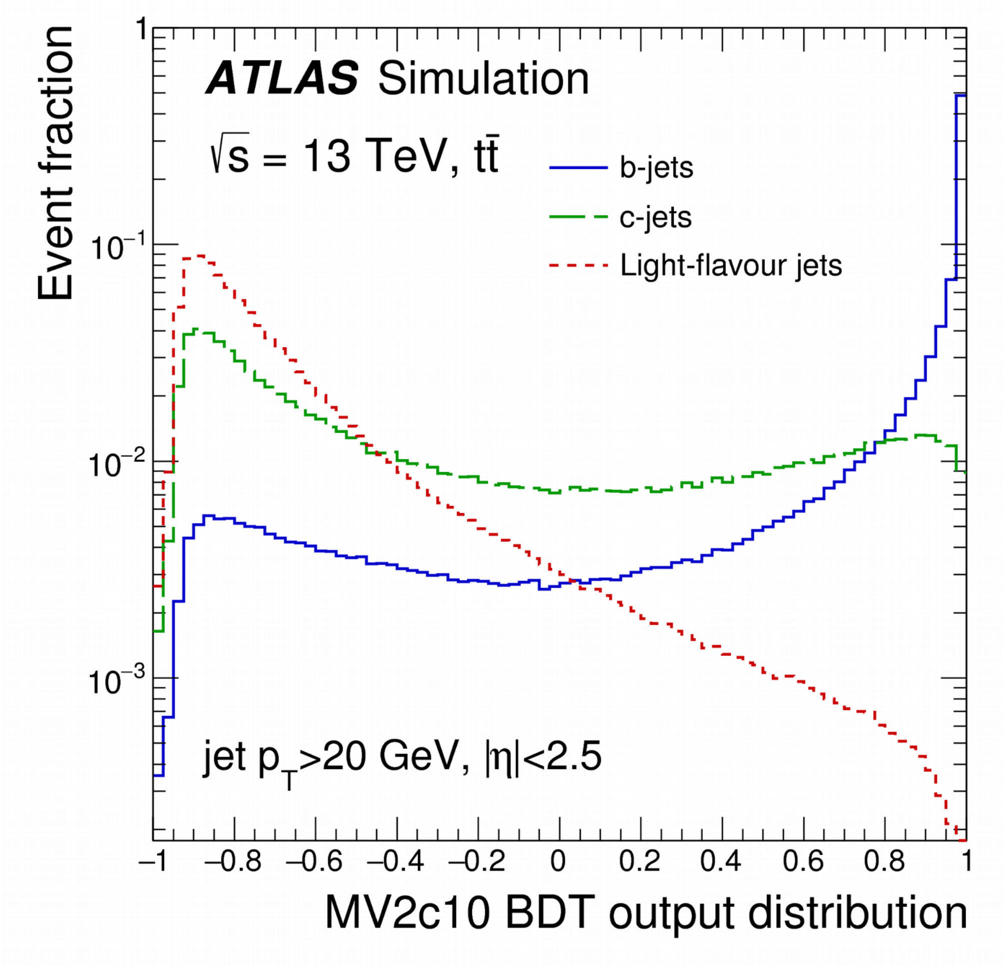
ATLAS (2016)

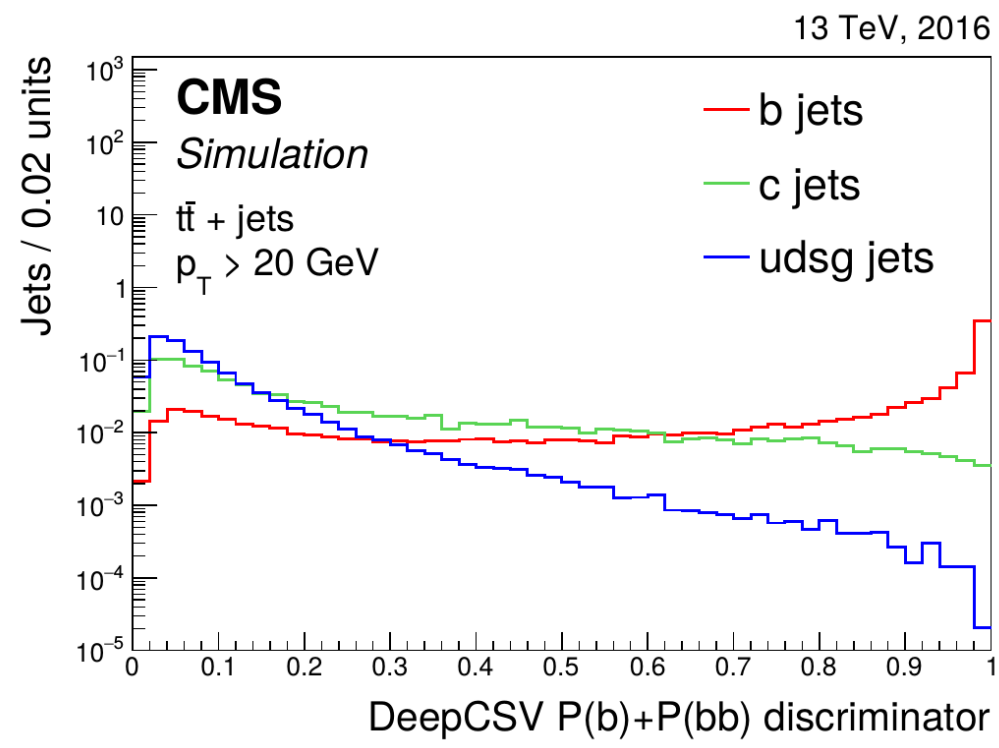
WP	Cut value X	b-jet efficiency (ε_b)	c-jet mistag rate (ε_c)	LF-jet mistag rate ($\varepsilon_{\mathrm{LF}}$)
85%	0.1758	85%	32%	2.9%
77%	0.6459	77%	16%	0.77%
70%	0.8244	70%	8.3%	0.26%
60%	0.9349	60%	2.9 %	0.065%
50%	0.9769	50%	0.94 %	0.017%

CMS (2016)

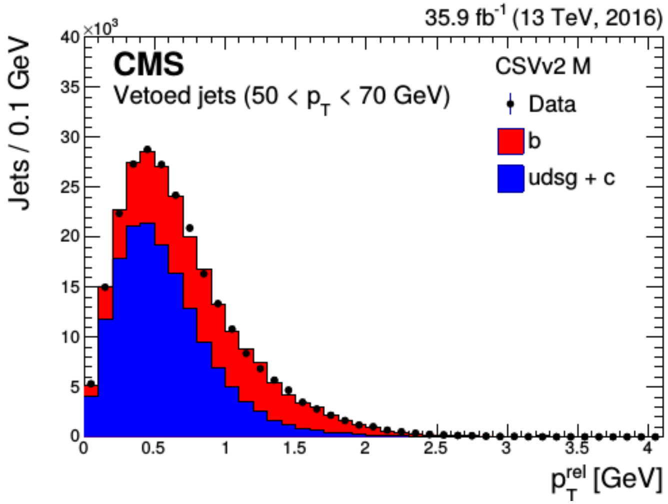
Tagger	Working point	ε_{b} (%)	ε_{c} (%)	$\varepsilon_{\mathrm{udsg}}$ (%)
Deep combined secondary vertex (DeepCSV) $P(\mathrm{b}) + P(\mathrm{bb})$	DeepCSV L	84	41	11
	DeepCSV M	68	12	1.1
	DeepCSV T	50	2.4	0.1

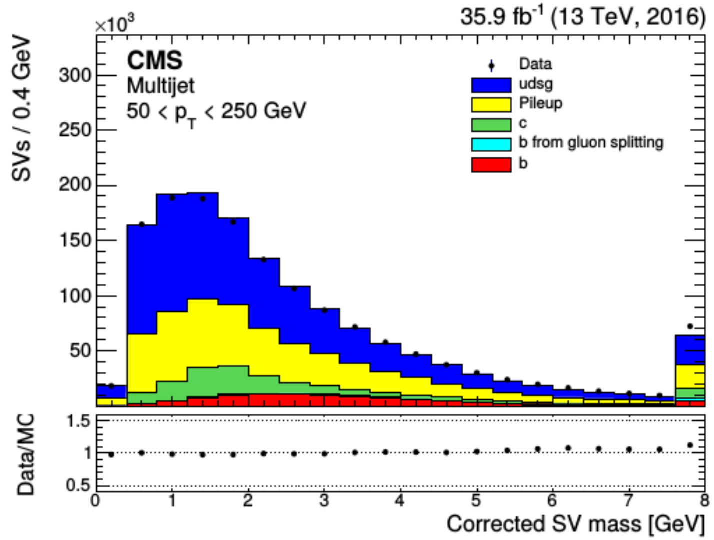
Working points 2016

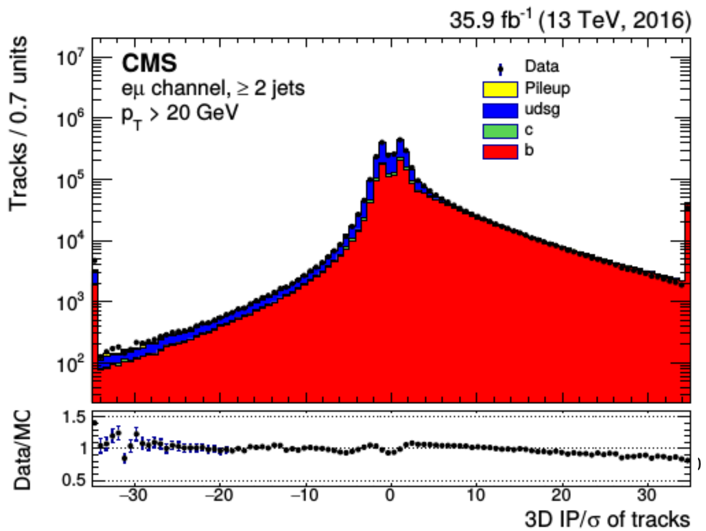


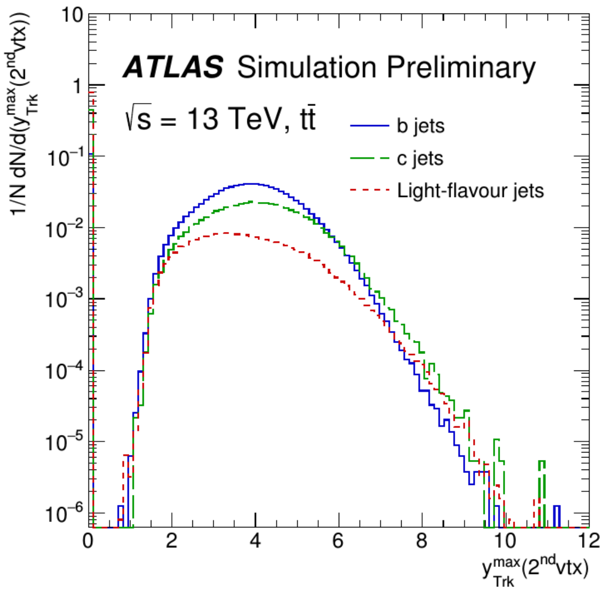


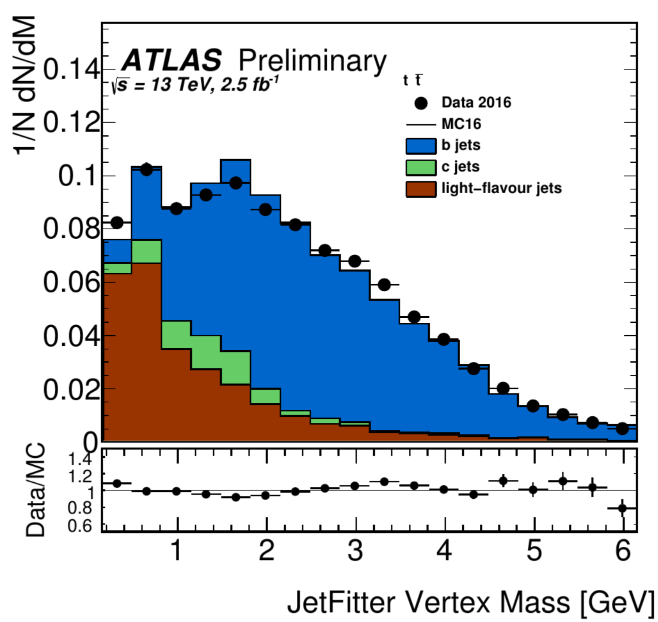
Input variables

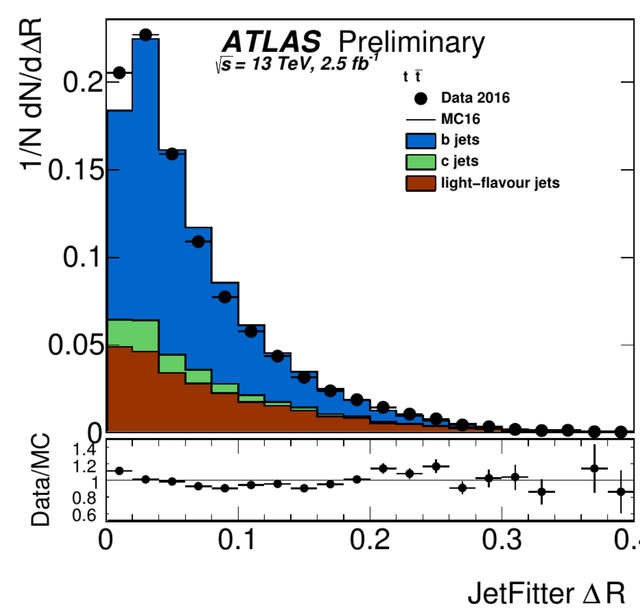






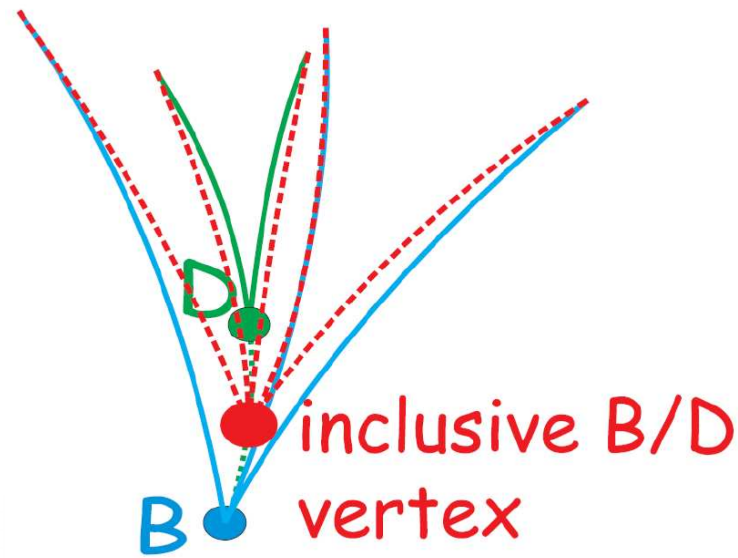




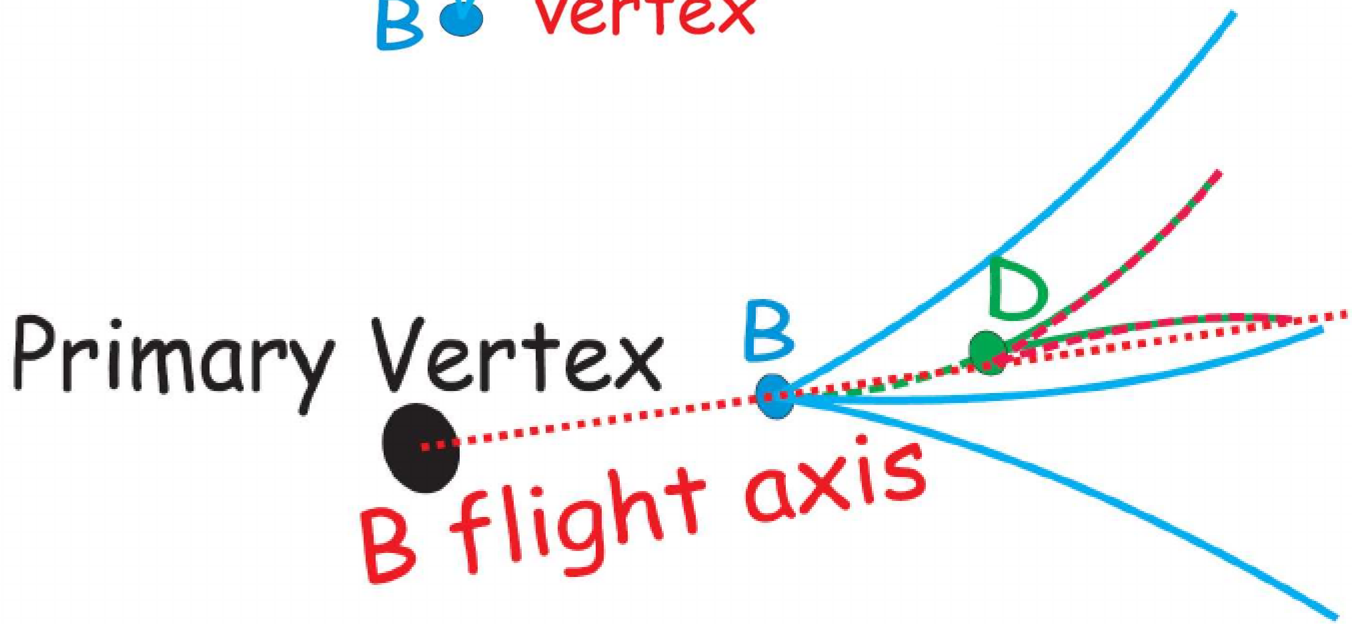


Secondary Vertex Fitter

Default

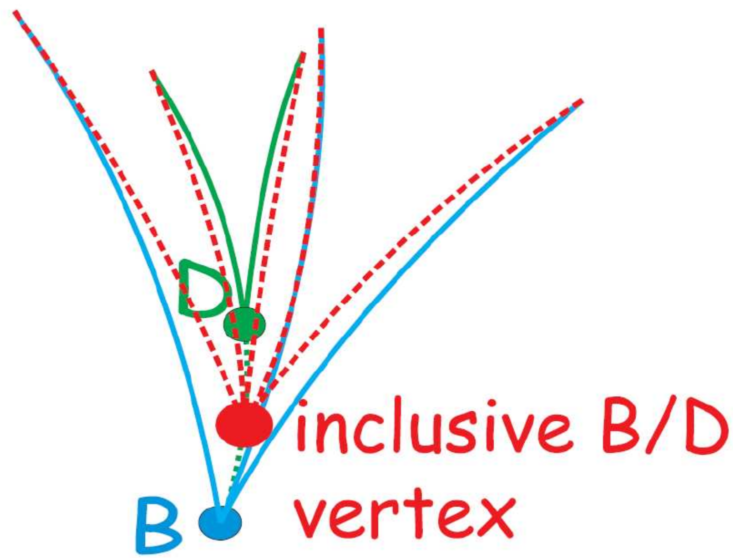


JetFitter

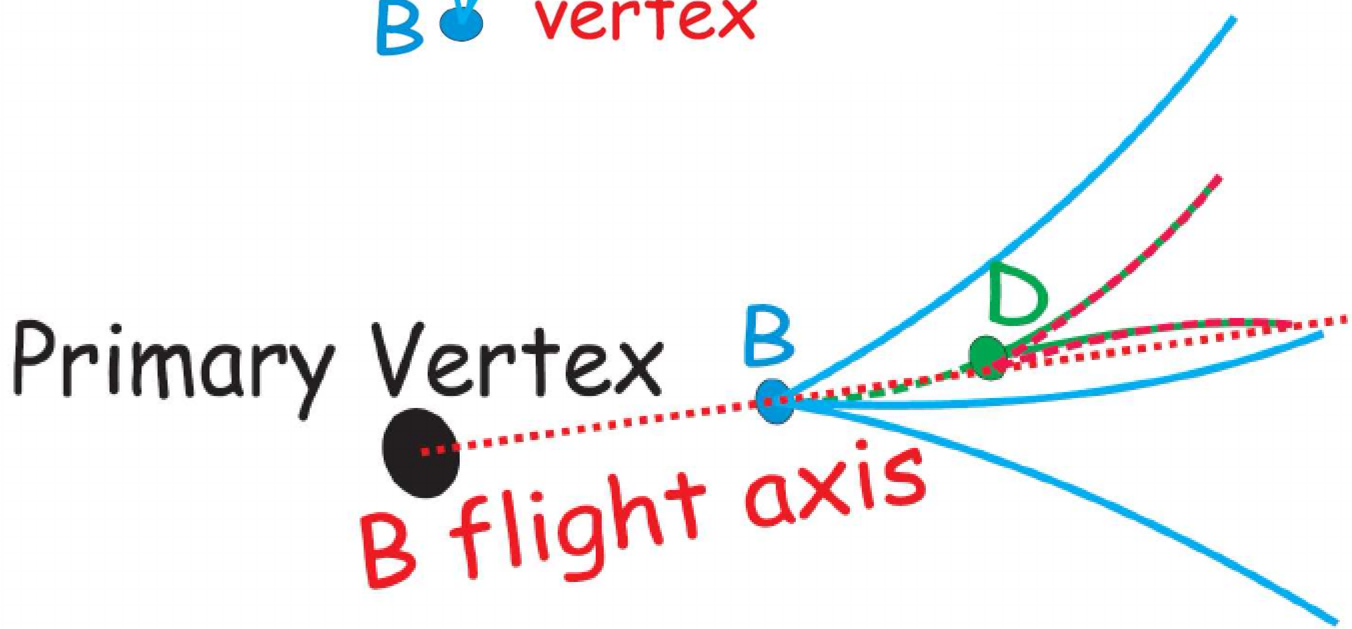


Inclusive Vertex Finder

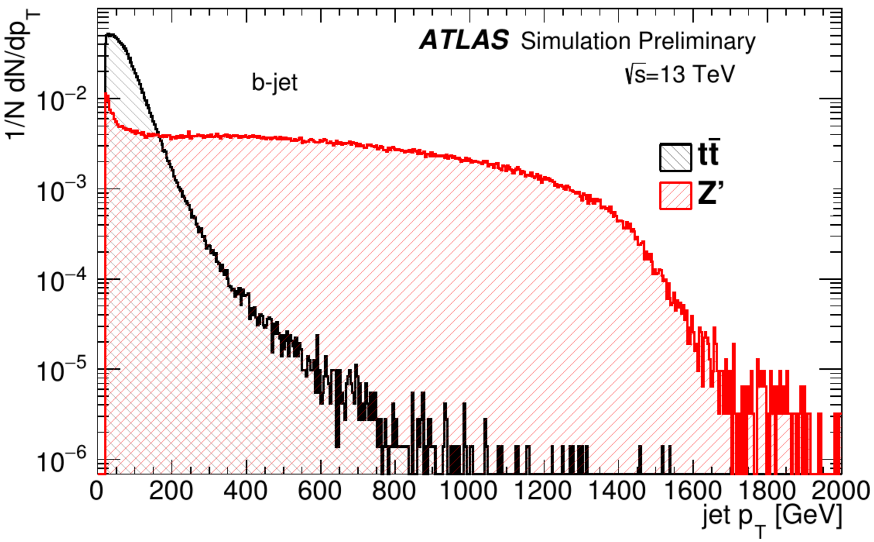
Default

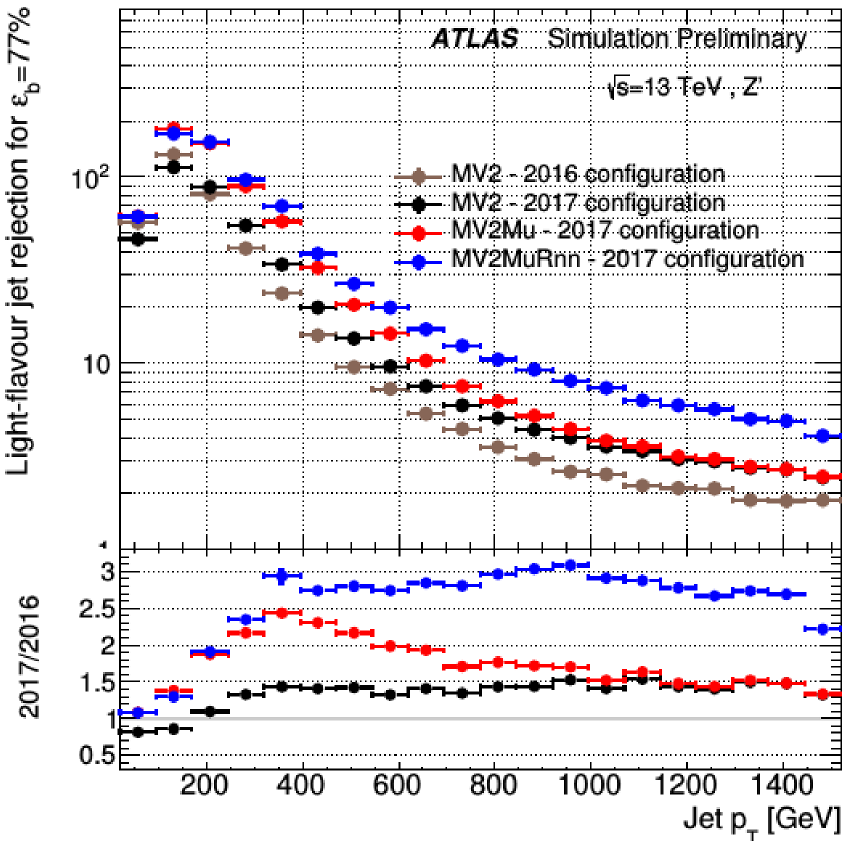


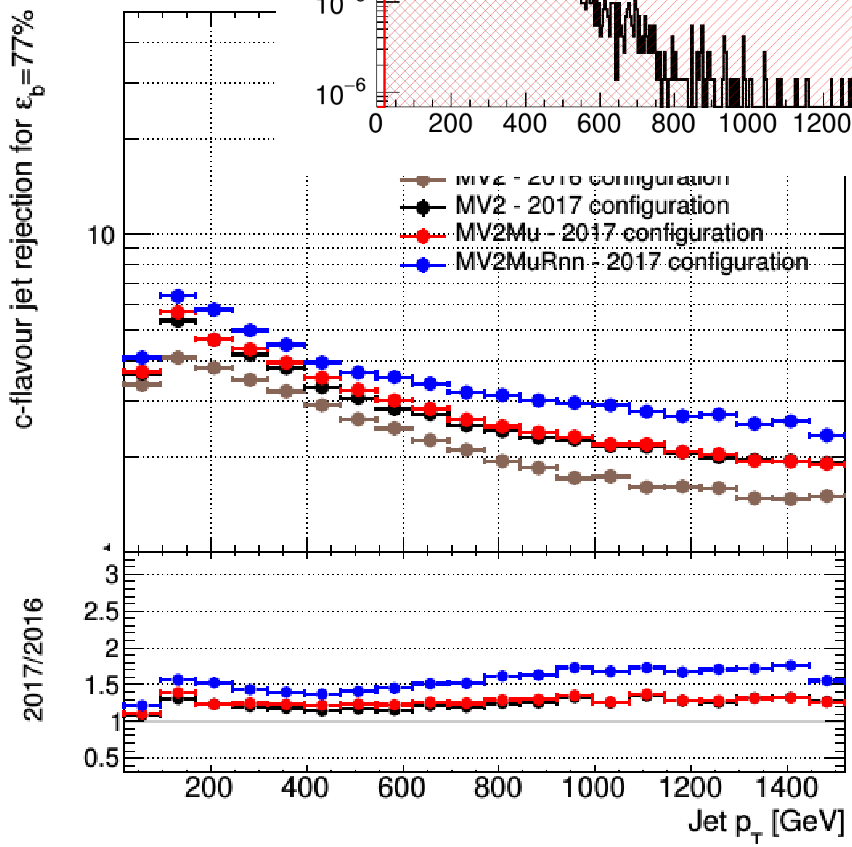
JetFitter



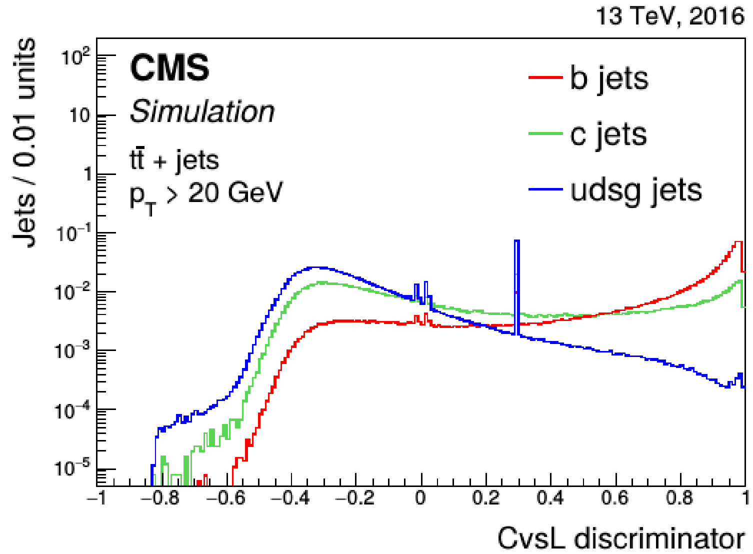
ATLAS 2016 vs 2017

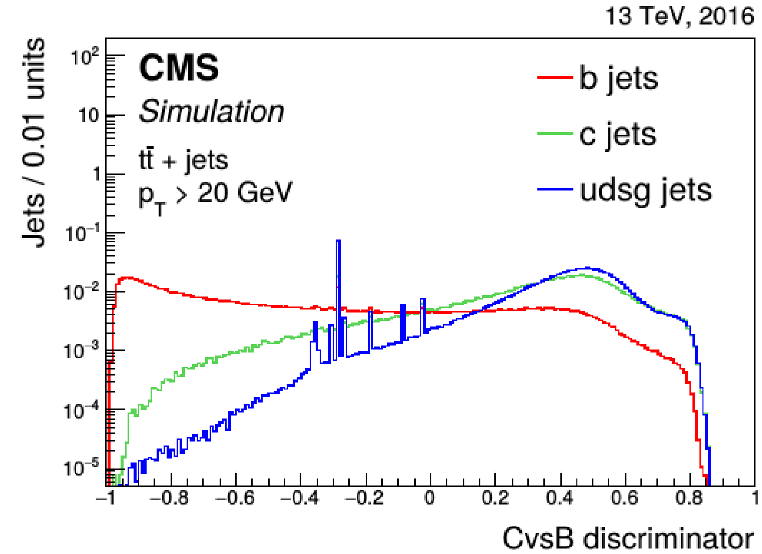


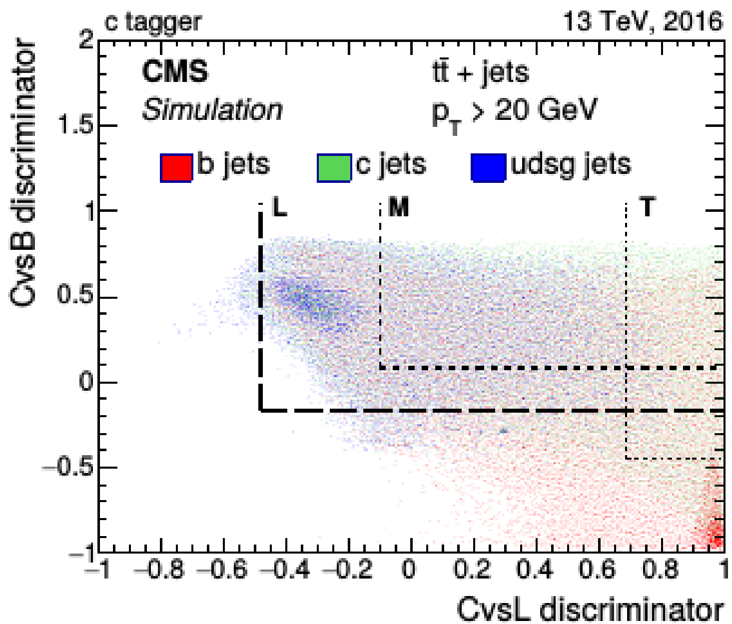


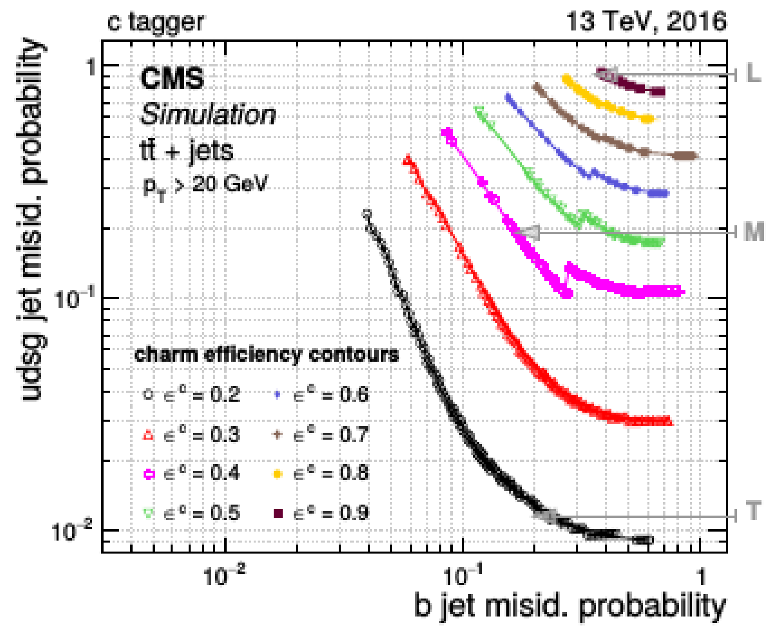


c-tagger

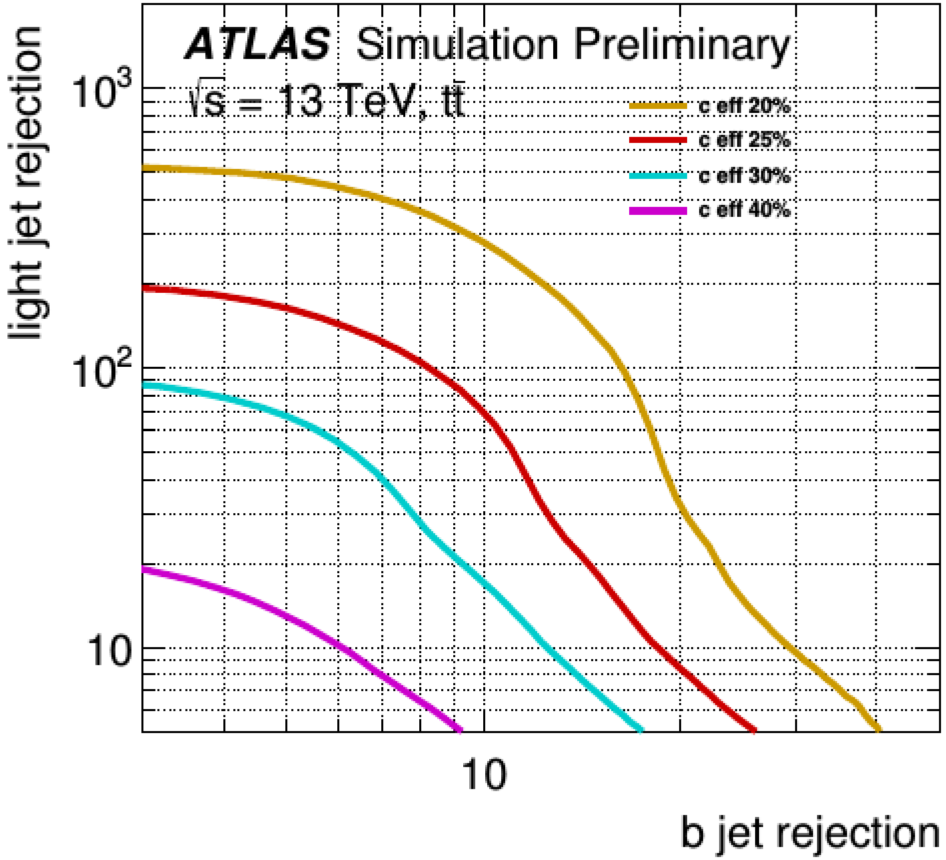

13 TeV, 2016
CMS
Simulation
$t\bar{t}$ + jets
p_T > 20 GeV
b jets
c jets
udsg jets
Jets / 0.01 units
CvsL discriminator


13 TeV, 2016
CMS
Simulation
$t\bar{t}$ + jets
p_T > 20 GeV
b jets
c jets
udsg jets
Jets / 0.01 units
CvsB discriminator

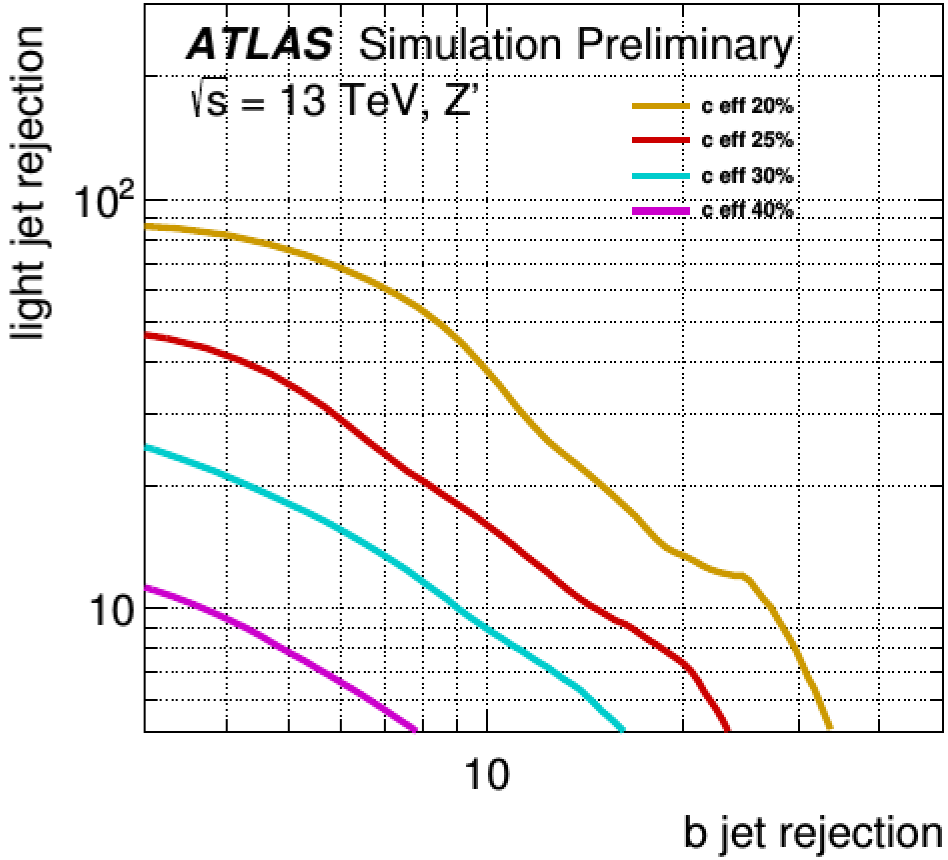

c tagger
13 TeV, 2016
CMS
Simulation
$t\bar{t}$ + jets
p_T > 20 GeV
b jets
c jets
udsg jets
L
M
T
CvsB discriminator
CvsL discriminator


c tagger
13 TeV, 2016
CMS
Simulation
$t\bar{t}$ + jets
p_T > 20 GeV
charm efficiency contours
ϵ^c = 0.2
ϵ^c = 0.3
ϵ^c = 0.4
ϵ^c = 0.5
ϵ^c = 0.6
ϵ^c = 0.7
ϵ^c = 0.8
ϵ^c = 0.9
L
M
T
udsg jet misid. probability
b jet misid. probability

c-tagger



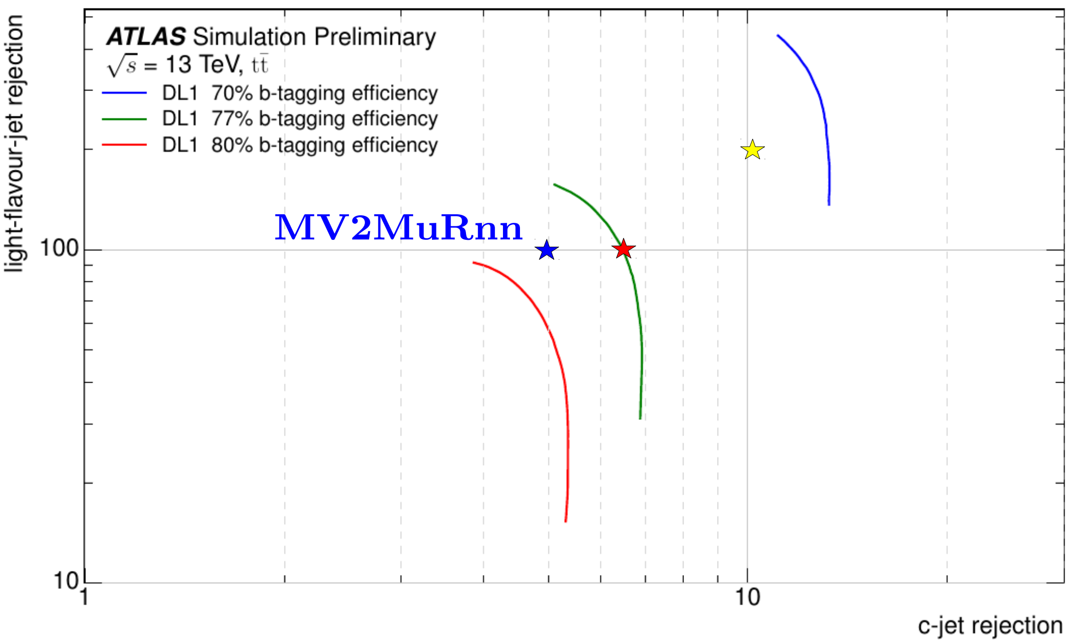
ATLAS Simulation Preliminary
√s = 13 TeV, tt̄
c eff 20%
c eff 25%
c eff 30%
c eff 40%
light jet rejection
b jet rejection
10
10²
10³

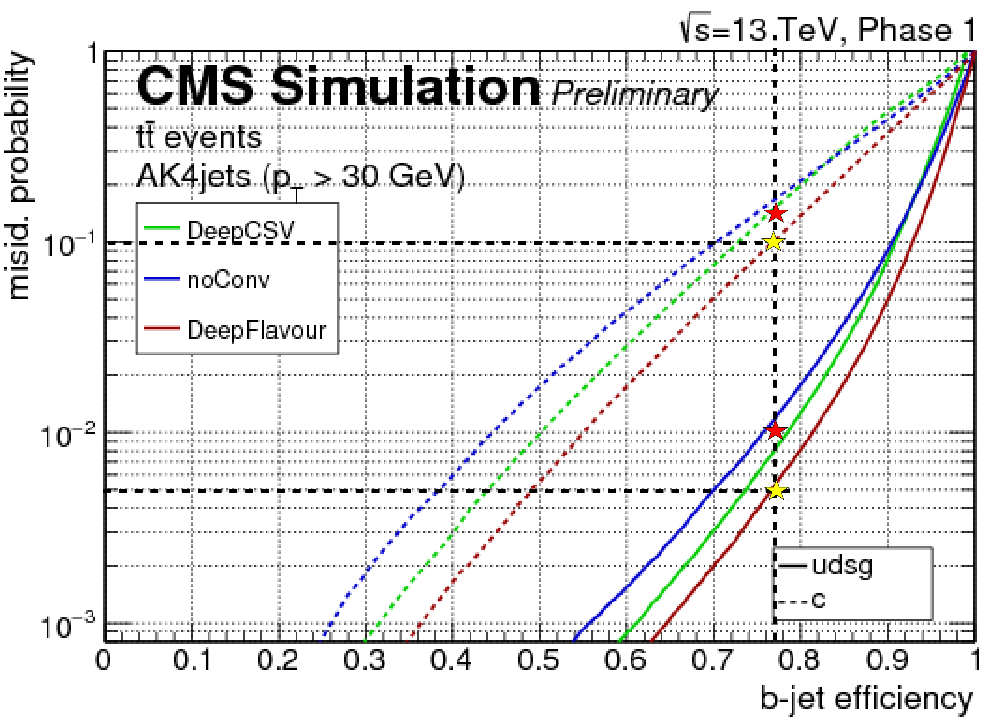


ATLAS Simulation Preliminary
√s = 13 TeV, Z'
c eff 20%
c eff 25%
c eff 30%
c eff 40%
light jet rejection
b jet rejection
10
10²

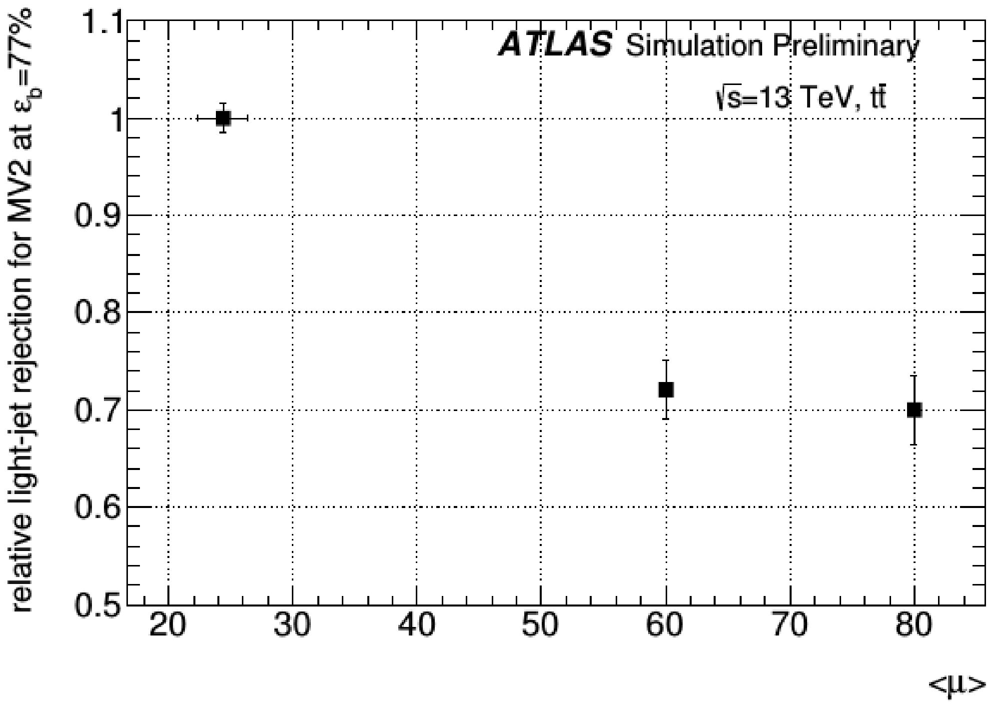
Deep Neural Network

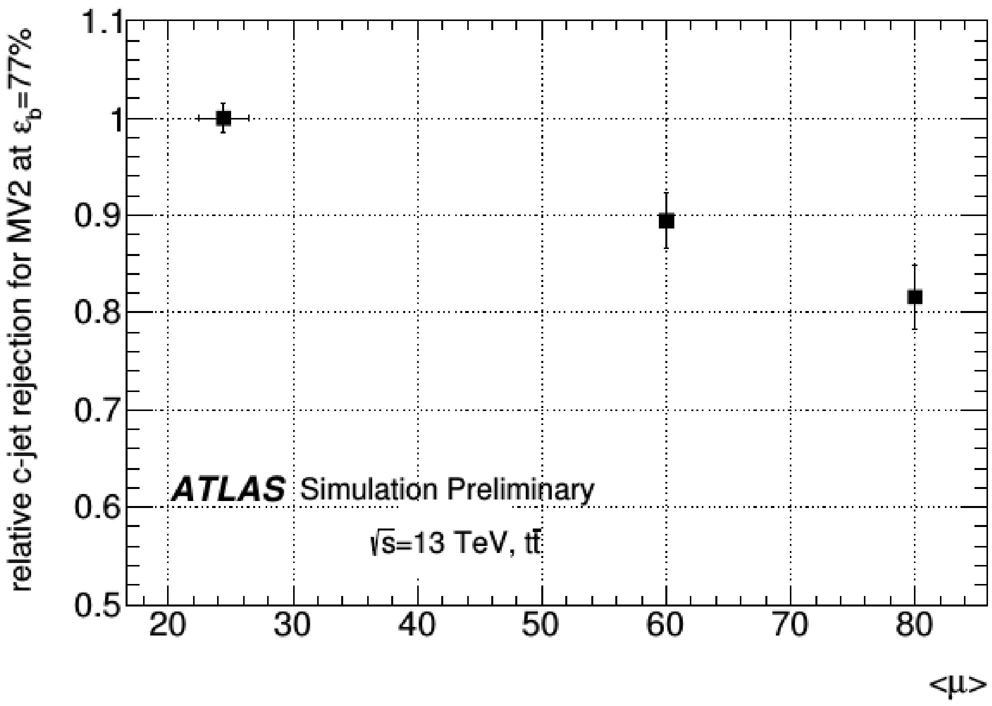
- DeepFlavour is more aggressive than DL1 (more input variables) and hence gives better performance.





b-tag vs PU


ATLAS Simulation Preliminary
√s=13 TeV, tt̄
relative light-jet rejection for MV2 at ε_b=77%
<μ>
0.5
0.6
0.7
0.8
0.9
1
1.1
20
30
40
50
60
70
80


ATLAS Simulation Preliminary
√s=13 TeV, tt̄
relative c-jet rejection for MV2 at ε_b=77%
<μ>
0.5
0.6
0.7
0.8
0.9
1
1.1
20
30
40
50
60
70
80