b-Tagging in ATLAS and CMS

Luca Scodellaro
Instituto de Física de Cantabria - CSIC
On behalf of the ATLAS and CMS Collaborations

Fifth Large Hadron Collider Physics Conference
May 15th-20th, 2017
Shanghai (China)

Outline

- b-tagging algorithms in ATLAS and CMS
  - Optimization of algorithms for the 2016 data taking Run
  - New developments
- Measurements of b-tagging performance in data:
  - Commissioning of b-tagging observables
  - Data-MC scale factors measurements
- b-tagging in events with boosted topologies
- Preparation for the future runs:
  - Data taking in 2017, with the new pixel Phase 1 in CMS
  - Upgrades for High-Luminosity LHC

b-Tagging in ATLAS and CMS

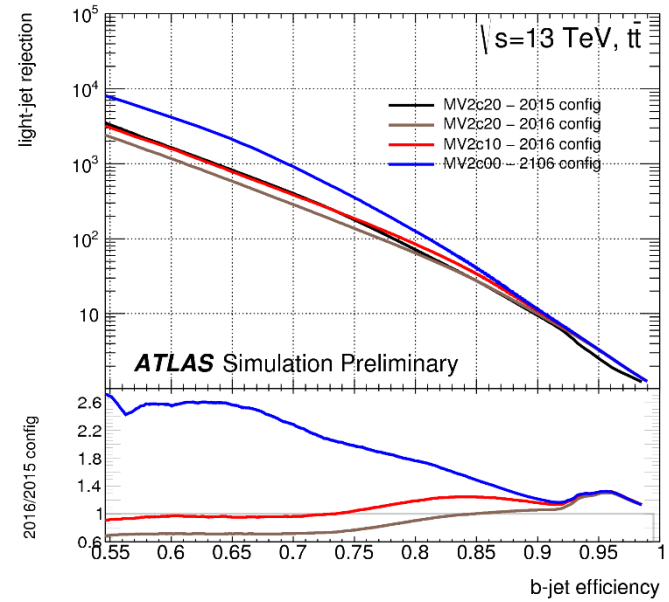
Properties	ATLAS	CMS
Tracks with large impact parameters (IP)	IP2D, IP3D, JetProb TrackCounting	TCHP, TCHE, JetProb JetBProb
Secondary vertices (SV)	SV0, SV1, SV	SSVHP, SSVHE, IVF
Multiple vertices from B->C->X decay chains	JetFitter	
Soft leptons from semi-leptonic B decays	SMT, p_TRel	Soft Lepton (SL) Taggers
Multivariate combinations	MV1c, MV2c00, MV2c10, MV2c20	CSV, CSVv2, cMVAv2, DeepCSV

- The algorithms provide a discriminant value for each jet
- Operating points defined as thresholds on the discriminant:
  - ATLAS: guaranteeing efficiencies for b-jets of 60, 70, 77 or 85%
  - CMS: reducing mis-identification rates on light jets to 0.1, 1 or 10%

RunII Algorithms in ATLAS

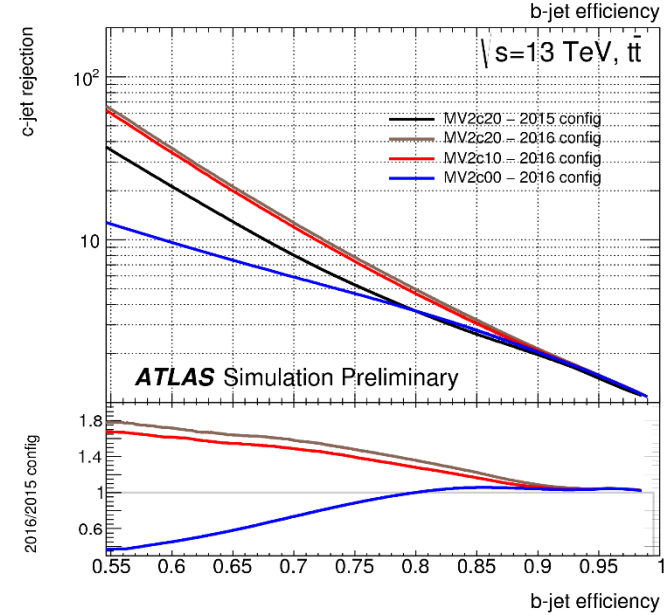
- Detector, tracking and algorithm improvements from RunI:
  - Light (c) jet rejection improved by a factor 4 (1.5-2) for a b-tag efficiency of 70%

ATL-PHYS-PUB-2015-022



- Further optimization for 2016 LHC Run:
  - Track selection criteria and secondary vertex cleaning in IP2D, IP3D and SV
  - MVA training procedure improved
  - MV2c10: similar light jet rejection, but 40% improvement in c-jet rejection for a b-tag efficiency of 70%

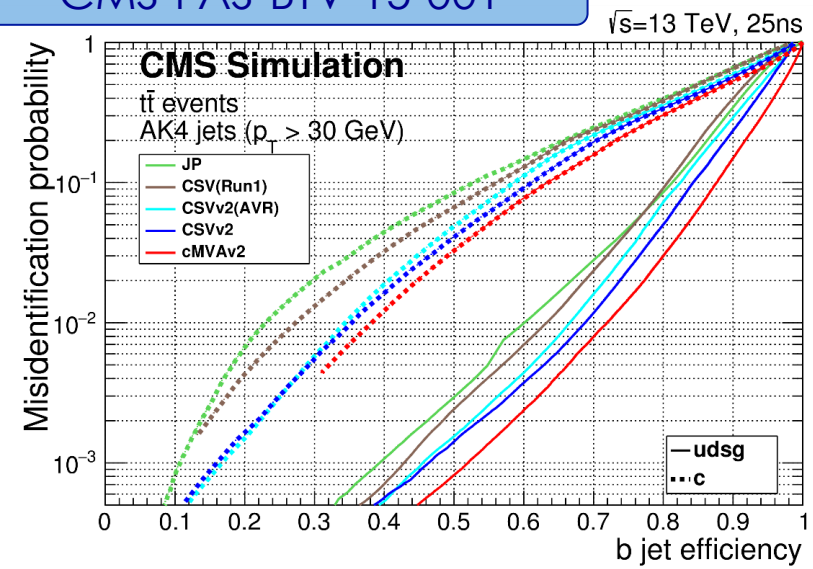
ATL-PHYS-PUB-2016-012

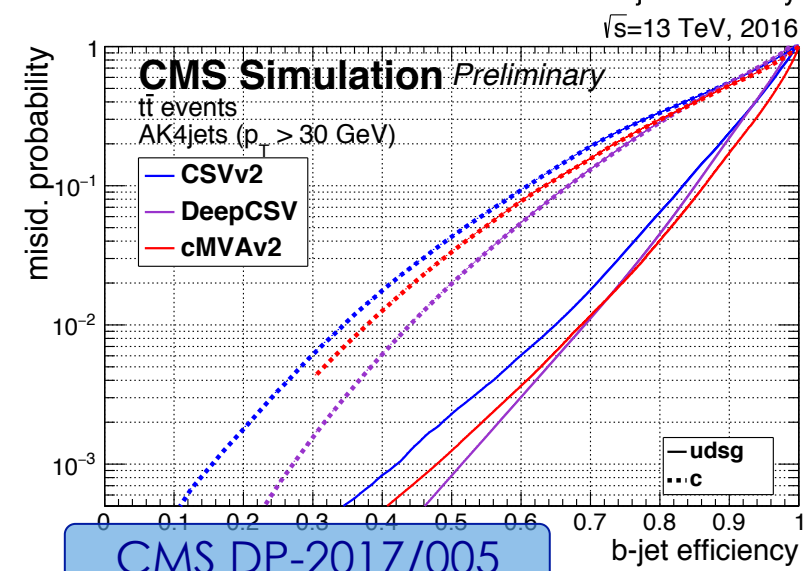


RunII Algorithms in CMS

- The new CSVv2 is an evolution of the RunI CSV algorithm:
  - Neural network instead of likelihood ratio allows to combine more variables
  - Uses Inclusive Vertex Finder
- The cMVAv2 tagger combines CSVv2, JP and soft lepton information
- DeepCSV use more sophisticated machine learning classifiers:
  - See next slide

CMS-PAS-BTV-15-001





CMS DP-2017/005

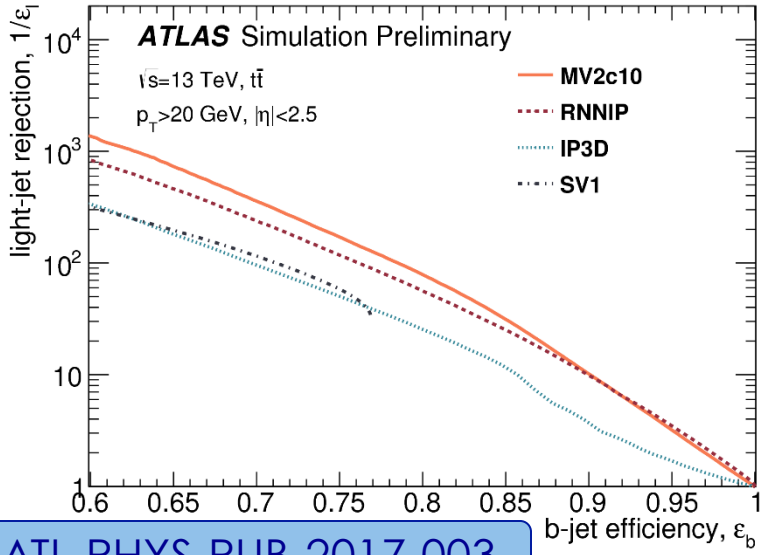
New Developments

- Use of more sophisticated neural network classes allows to better exploit the information available for b-tagging:
  - Can combine a large number of input features
  - Can handle more low-level information
  - Allows for multi-classification, providing an output probability for each jet flavor hypothesis

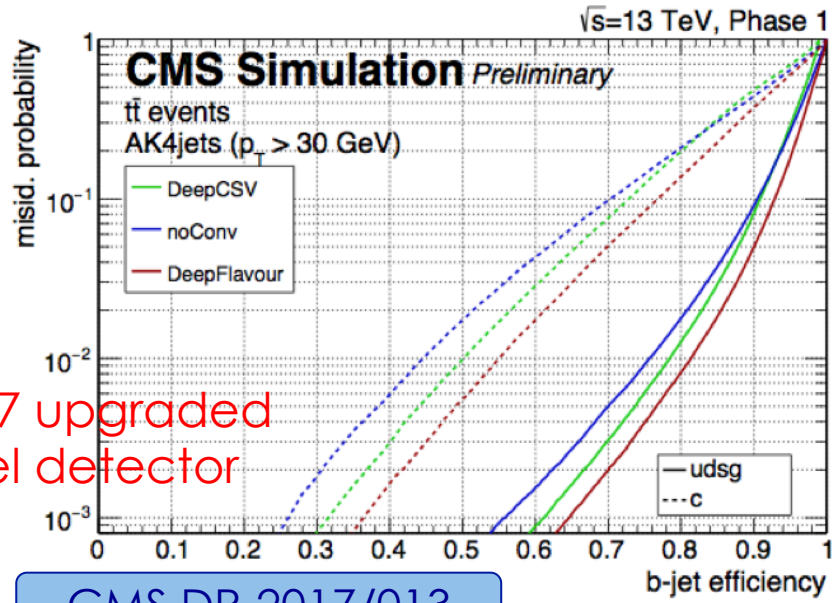
- ATLAS and CMS already developed algorithms based on:
  - Recurrent Neural Networks (ATLAS):
    - Networks with directed cycles
    - Can process arbitrary sequence of inputs
  - Deep Neural Networks (CMS):
    - Feedforward networks with many hidden layers

New Developments

- ATLAS RNNIP tagger:
  - Built using a sequence of track-by-track variables as input to a Recurrent NN
  - Outperforming IP3D: 2.5 times its light-jet rejection for a b-tag efficiency of 70%



ATL-PHYS-PUB-2017-003



2017 upgraded pixel detector

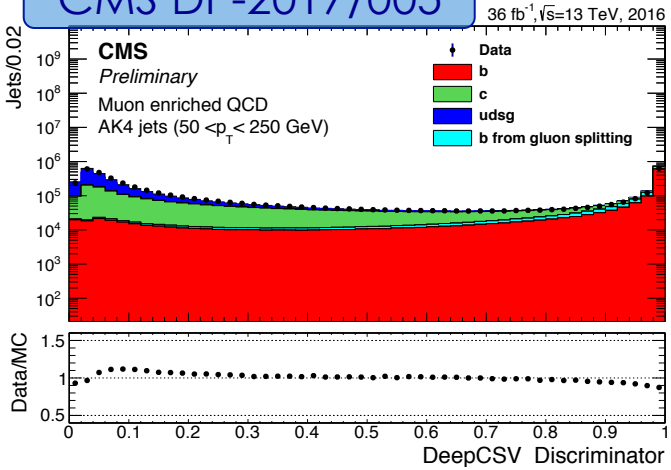
CMS DP-2017/013

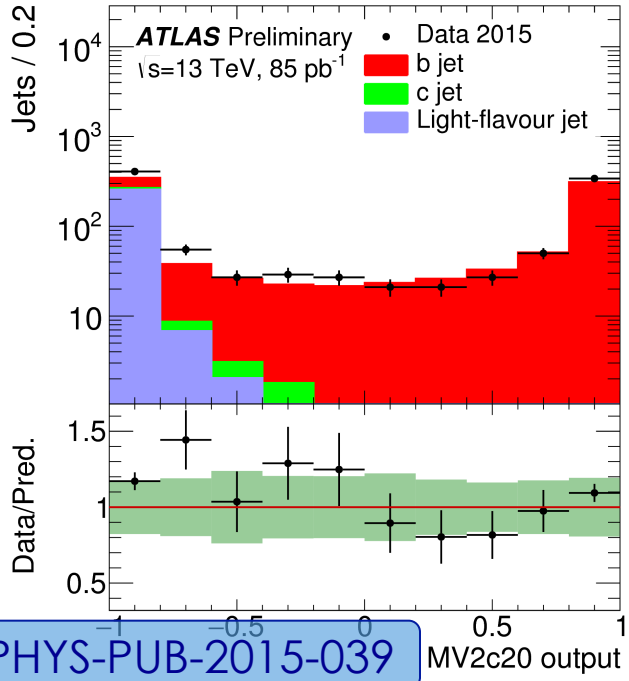
- CMS DeepFlavour tagger:
  - Using properties of particle-flow candidates, and SV in the jets
  - 4% absolute improvement in b-tag efficiency for a mistag rate of 0.1% against DeepCSV

Performance In Data

- Since simulations do not describe perfectly b-tagging observables, performance is studied in data
- Samples with different jet flavor composition are exploited to commission the algorithms:
  - Inclusive jets from QCD processes
  - Jets from QCD with an embedded soft muon
  - Top pair production events
- To correct b-tagging efficiencies in physics analysis, data-to-MC scale factors are computed for each operating point through data driven techniques

CMS DP-2017/005

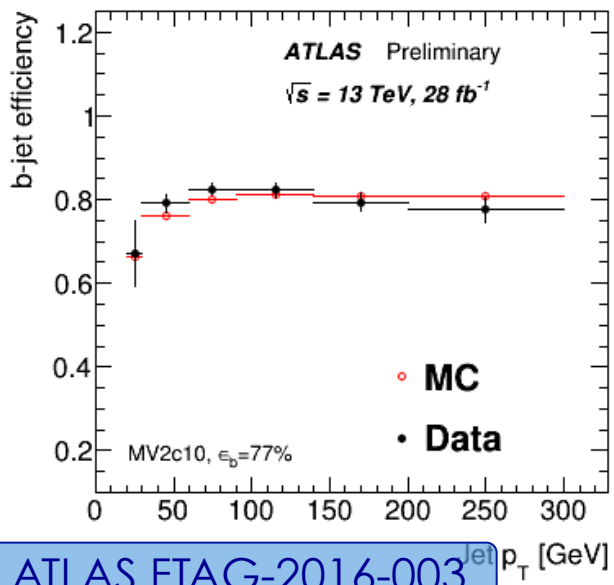




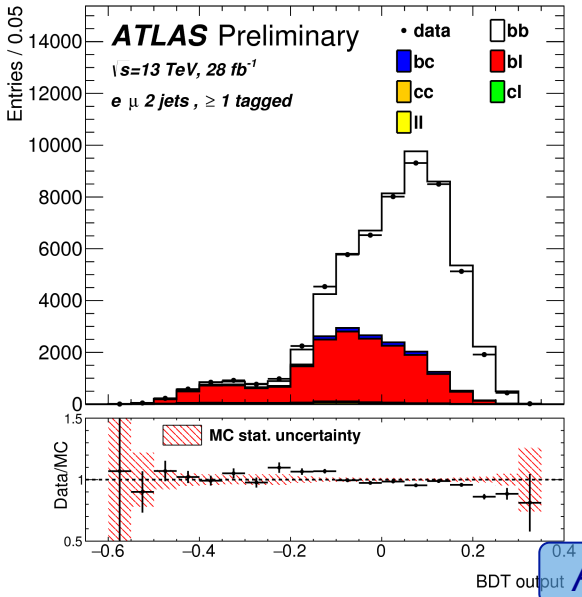
ATL-PHYS-PUB-2015-039

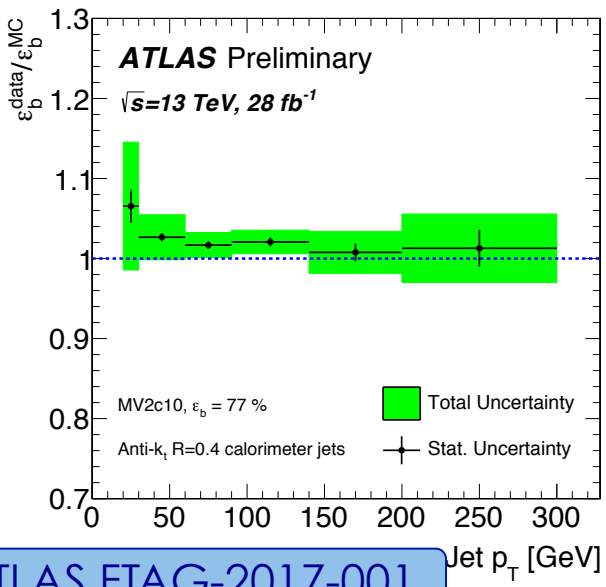
Scale Factors in ATLAS

- In RunII, scale factors measurements in ATLAS exploit the highly b-enriched jet sample from dilepton ttbar events
- Combinatorial likelihood PDF methods:
  - Efficiencies extracted from a likelihood fit to discriminant PDFs
- Tag-and-probe method:
  - Multivariate kinematic discriminator to suppress non-b jet probes



ATLAS FTAG-2016-003

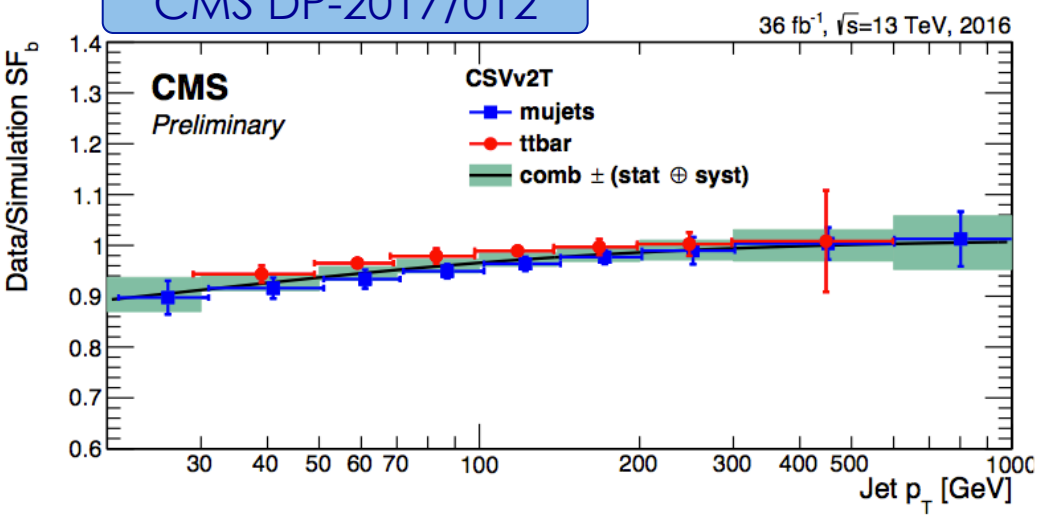


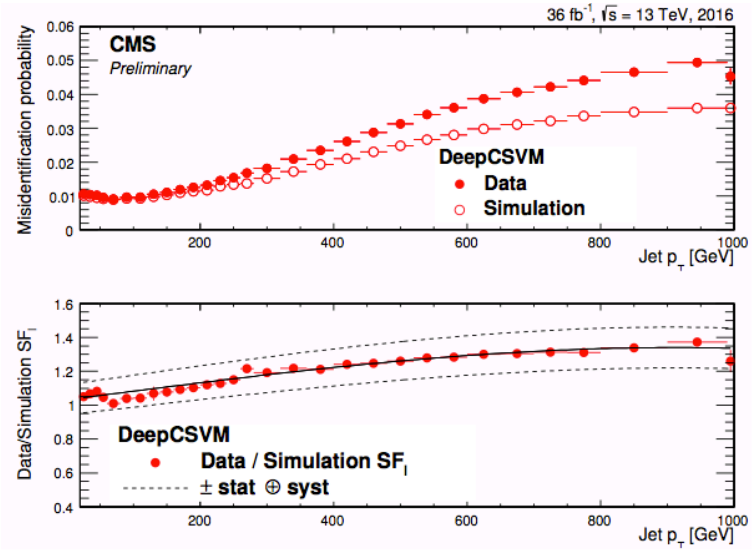


ATLAS FTAG-2017-001

Scale Factors in CMS

CMS DP-2017/012

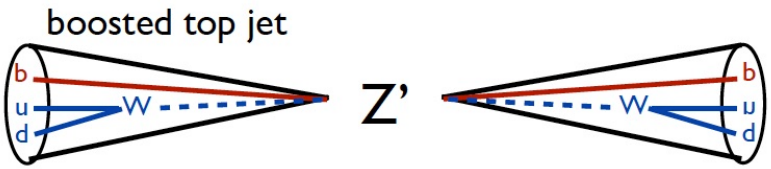




- Scale factor measurements in CMS exploits various methods:
  - QCD muon-eriched based: PtRel, System8, Lifetime Tagger
  - ttbar based: kinematic fits in dilepton and lepton+jets channels
  - Single measurements are combined to reach the best precision
- Discriminant reshaping factors are also determined in ttbar events for analyses using its whole shape in signal extraction
- SFs for light-jet mistag rate measured in inclusive jet samples

b-Tagging in Boosted Topologies

- In high energy collisions, particles decaying to b quarks can be produced with large momentum (boosted topology):
  - B decay products can overlap with particles from other jets
  - Important in many BSM searches



- Both experiments developed specific b-tagging algorithms for boosted topologies, using slightly different approaches

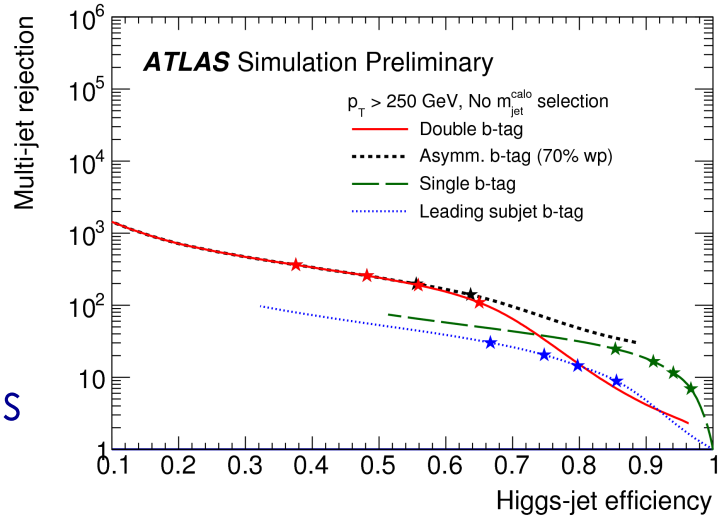
- ATLAS:
  - Calorimeter large-R (R=1.) jets
  - Trimming algorithm to discard softer components of the jet
  - b-tagging applied on ghost-associated track jets (R=0.2)

- CMS:
  - Particle flow AK8 jets (fatjets)
  - Soft drop declustering to resolve jet substructure
  - b-tagging applied on tracks in fatjets and/or subjets

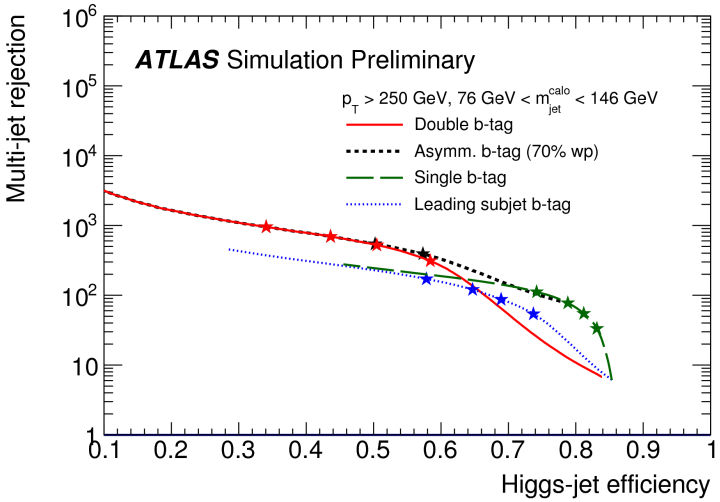
Boosted b-Tagging in ATLAS

ATLAS-CONF-2016-039

- Dedicated boosted Higgs boson tagger developed for RunII
- Various b-tagging options on the associated track jets compared:
  - In general, requiring two track jets satisfying (asymmetric) b-tagging requirements is more performant
  - Single track jet b-tagging becomes competitive at high efficiencies

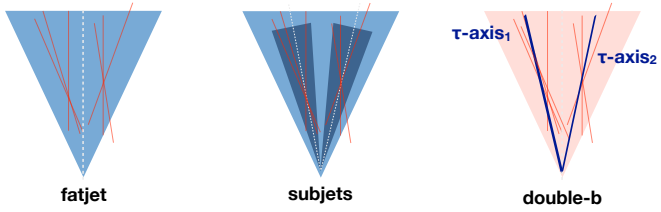


- Additional requirements applied to enhance the performances:
  - Large-R jet mass window cut around the Higgs boson mass
  - Cut on large-R jet substructure variables (e.g $D_2^{(\beta=1)}$)

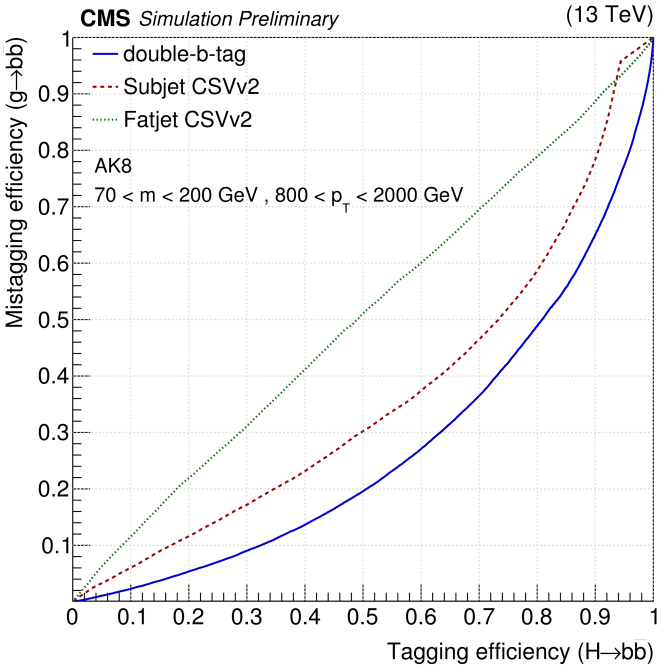


Boosted b-Tagging in CMS

- Two approaches followed since RunI:
  - Apply b-tagging algorithms on all the tracks in the fatjets
  - Identify subjets and apply algorithms only on their tracks (still baseline for boosted top quarks)
- In RunII, new dedicated algorithm to tag boosted decays to b pairs:
  - Build observables from n-subjettines axes to exploit the correlations between the b flight directions
  - Double-b tagger outperforming fatjet and subjet b-tagging for H->bb against multijet and g->bb backgrounds



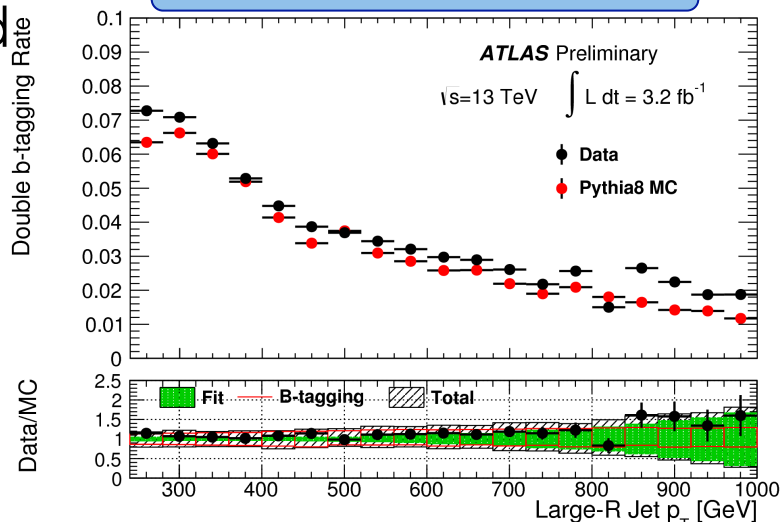
CMS-PAS-BTV-15-002

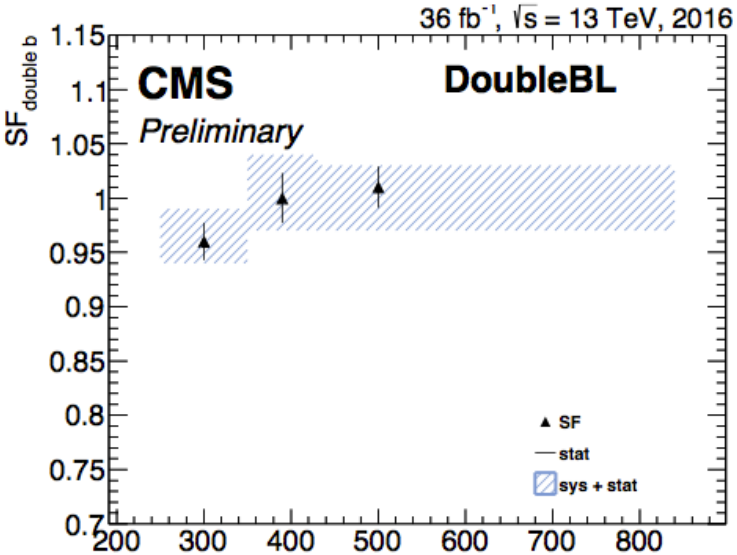


Boosted b-Tagging Validation

- ATLAS: performance of the boosted H->bb tagger validated on candidate g->bb large-R jets:
  - Two track-jets associated, with at least one containing a soft muon
  - Observed rate of double b-tagged large-R jets agree reasonably well with simulations
- CMS: measuring specific scale factors by applying standard techniques (LT method) to AK8 jets with two soft muon-tagged subjets

ATLAS-CONF-2016-039



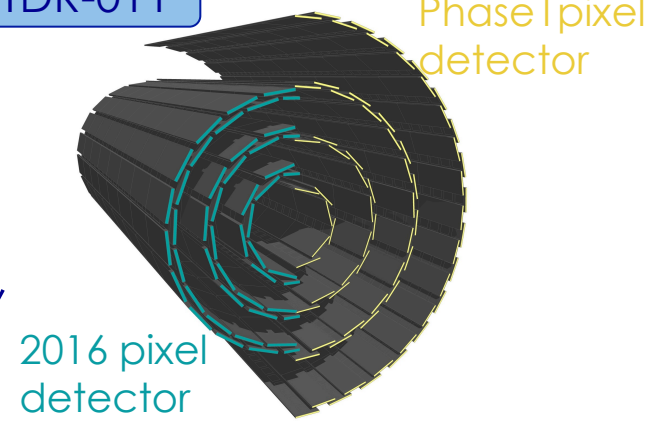


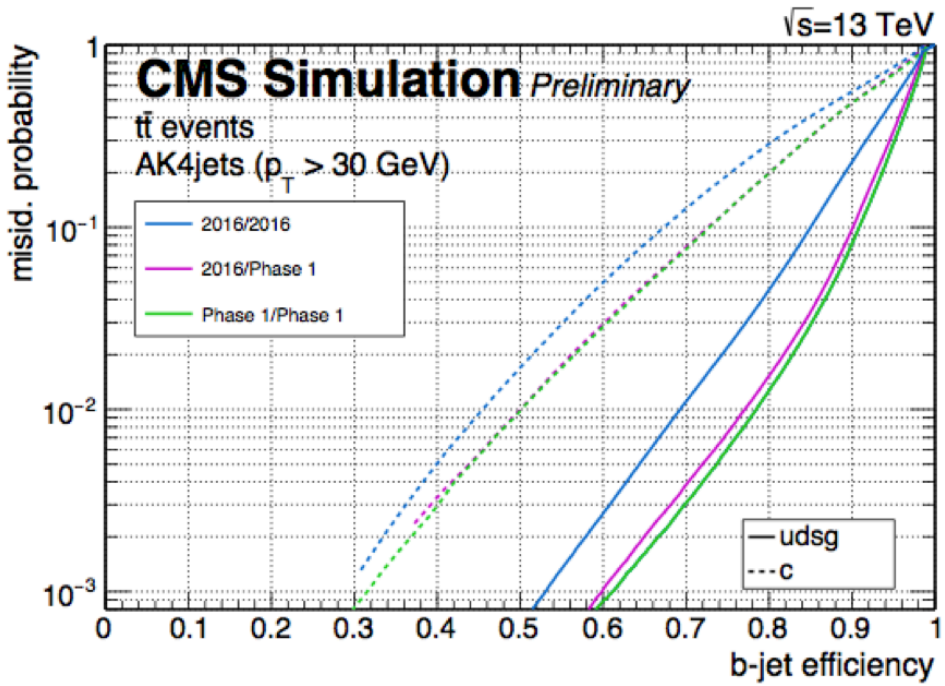
CMS DP-2017/012

Phase 1 Upgrade in CMS

CMS-TDR-011

- The CMS Phase 1 upgrade includes a new pixel detector with an additional layer, closer to the beam spot:
  - Expected improved resolution in tracking, especially on impact parameter



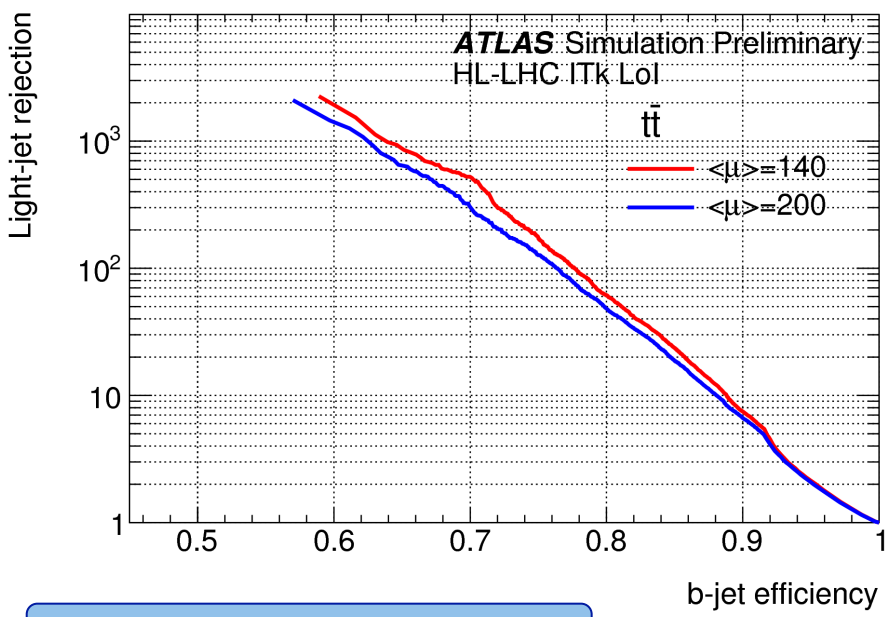


- Comparison of DeepCSV performance with 2016 detector, Phase 1 detector and 2016 training, and with Phase 1 detector and new dedicated training

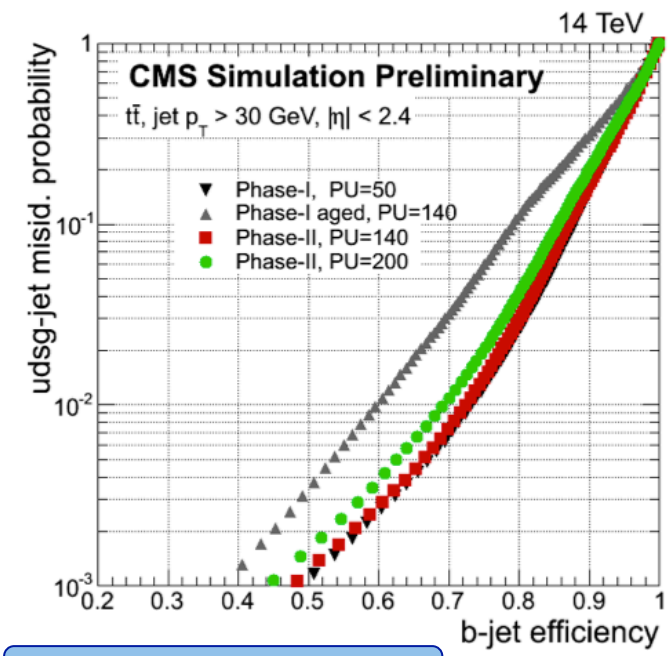
CMS DP-2017/013

High Luminosity Upgrades

- Major upgrades of ATLAS and CMS detectors planned to operate during the High Lumiosity (HL) LHC phase
  - Trackers will be replaced with new detector with higher granularity, radiation robustness and extended coverage
- First studies show that the b-tagging algorithms can operate in the complex high PU environment expected during HL-LHC



ATL-PHYS-PUB-2016-026



CMS DP-2016/065

Conclusions

- b-Tagging is a fundamental tools in most physics analyses
- Both ATLAS and CMS reached a significant improvement on their algorithms in RunII, and new promising ideas for further developments are being explored
- Not only algorithms, but also the measurements of their performance on data had benefited from new ideas (and of increased sample statistics) in 2016
- Techniques are being extended to cover more specific topologies becoming ever more important with the increase of the LHC collisions center-of-mass energy
- More challenge ahead: already working to maintain b-tagging a successful tool in the next decade of data taking

Backup Material

Identification of c-Jets

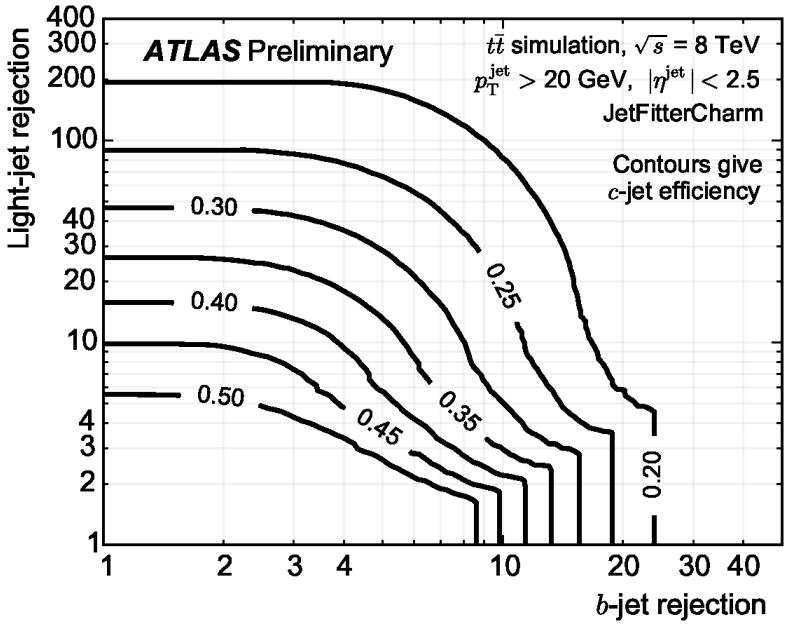
ATL-PHYS-PUB-2015-001

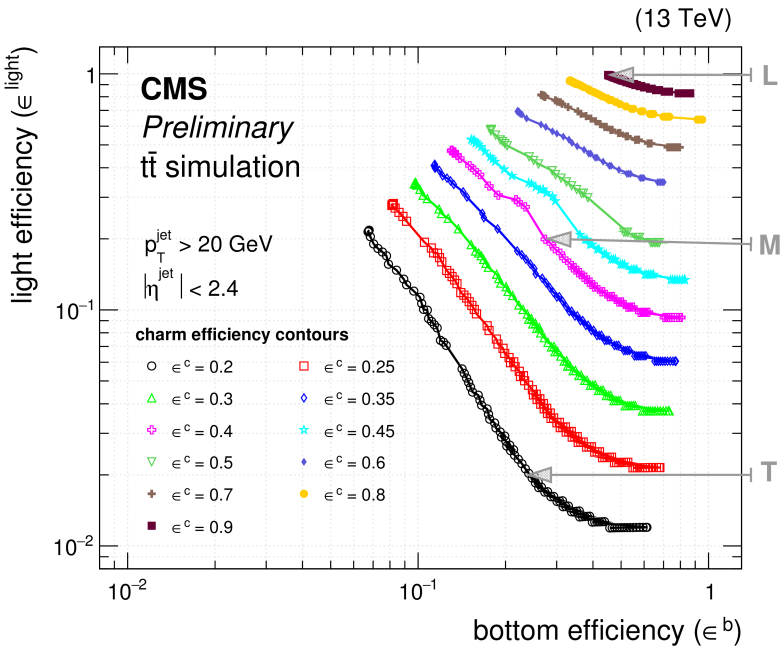
- ATLAS RunI JetFitterCharm uses a design similar to MV1:
  - NN combination of a modified version of the IP3D, SV1 and JetFitter taggers

CMS-PAS-BTV-16-001

- The CMS c-jet tagging algorithm uses a design similar to CSVv2:
  - Adding soft lepton information to add more observables and more jet categories

- Operating points defined by 2D cuts on c-jets vs light-jets and c-jets vs b-jets discriminants:

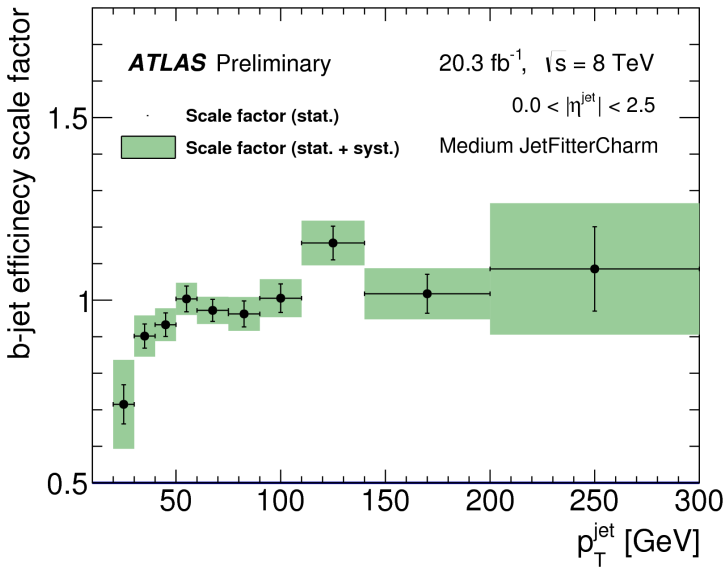


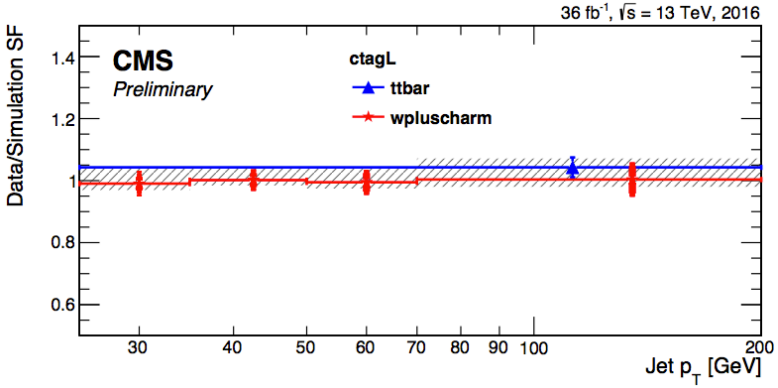


Scale Factors for c-Taggers

- In ATLAS, efficiencies for c-jet identification are measured in a multijet event samples where jets contain D* mesons

ATL-PHYS-PUB-2015-001



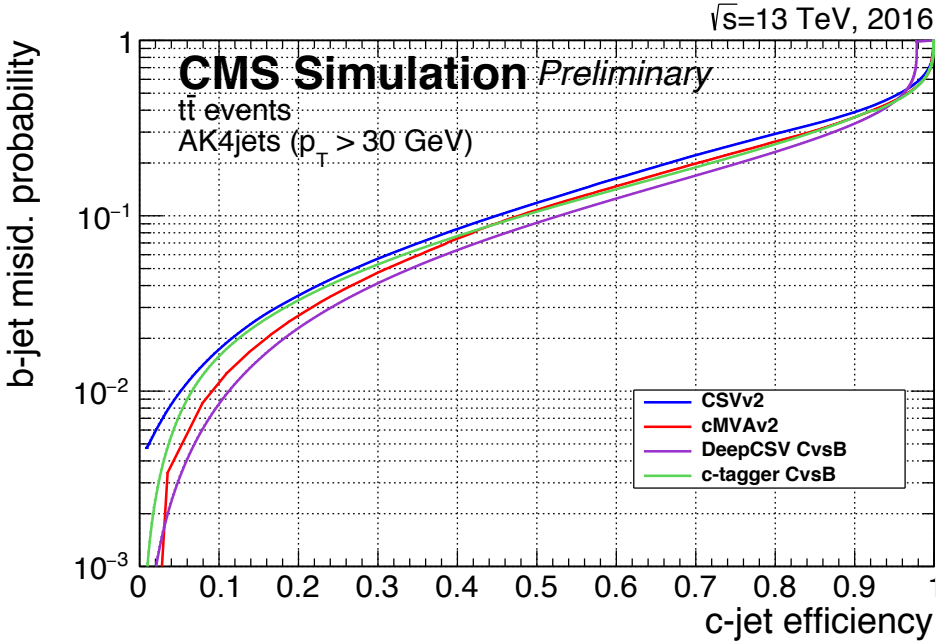


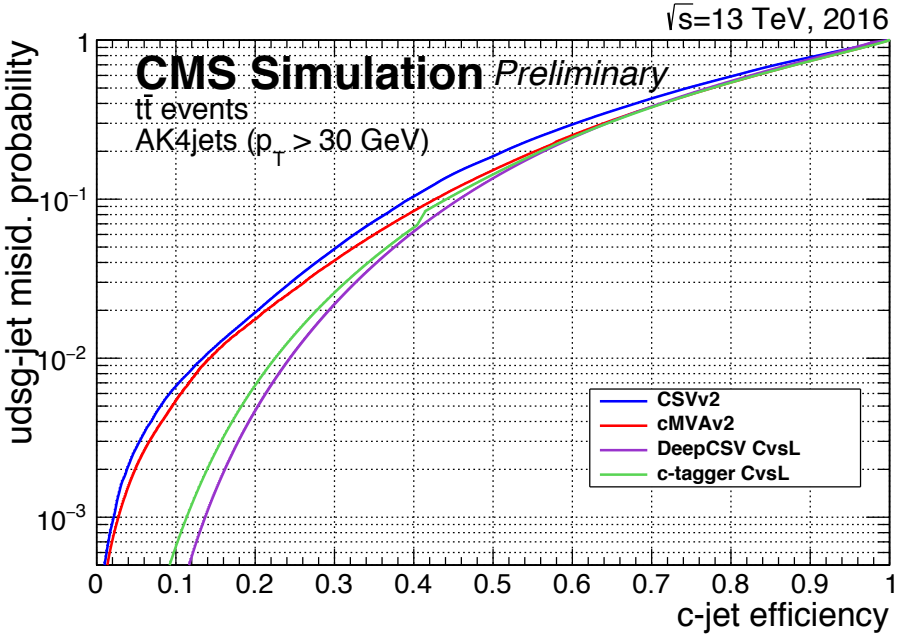
CMS DP-2017/012

- CMS: performance measured in two c-jet enriched samples:
  - Wc events, selected by requiring a soft muon in the c-jet
  - ttbar events in lepton+jets final states

c-Tagging with Multiclassifiers

- DeepCSV is already outperforming the dedicated c-tagger





CMS DP-2017/005