b-tagging: current and future performance

L. Mastrolorenzo[1]
on behalf of the ATLAS and CMS Collaboration

[1] RWTH Aachen University

Double Higgs Production at Colliders Workshop 2018
Fermilab, USA, 4-8 September 2018

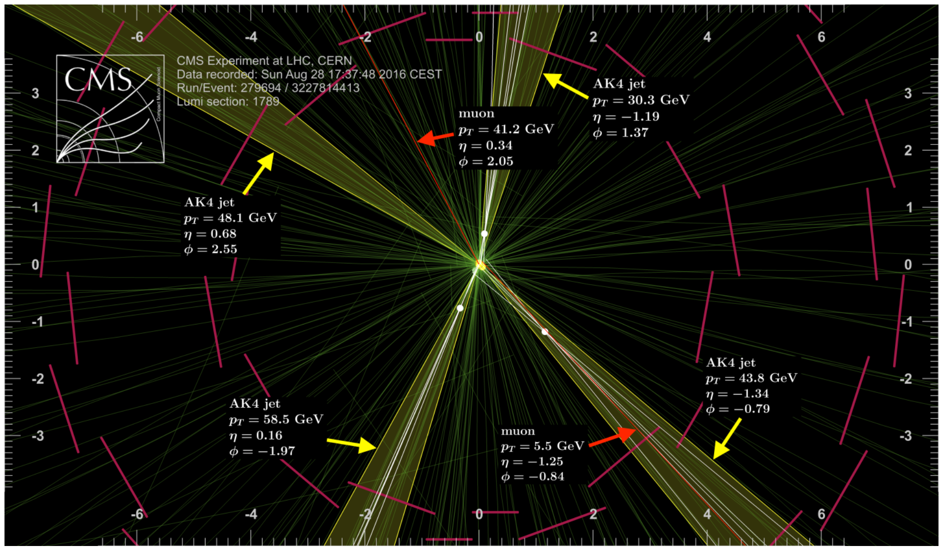
Outlook

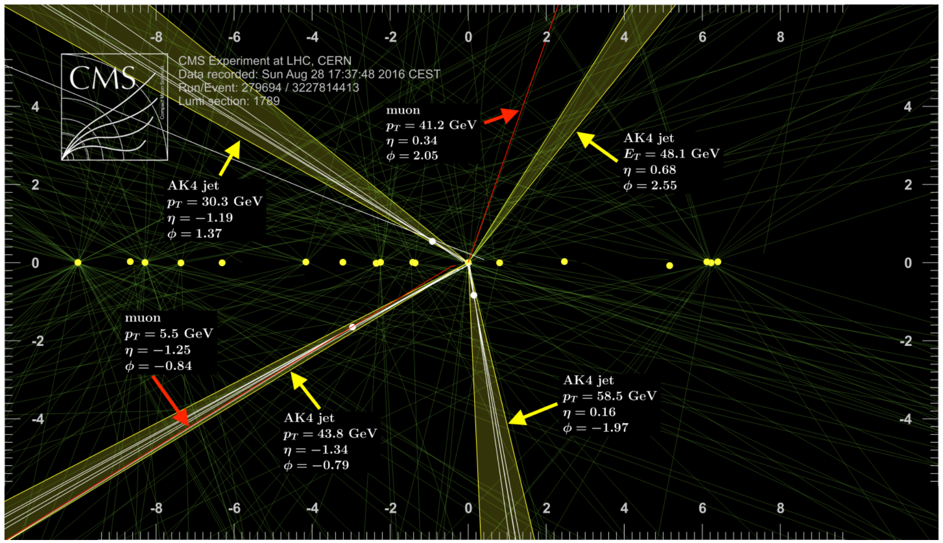
In this talk:

- Introduction: Physics case for heavy-flavour tagging
- Overview of b-tagging technique in ATLAS and CMS
- Focus on cutting-edge taggers in ATLAS and CMS (focussing on Deep Neural Network based algorithm)
- Performance in data and data/MC scale factor measurements
- Future prospects and Conclusions

- **Why heavy flavour tagging?**
  - The ATLAS and CMS scientific program at LHC strongly relies on the most accurate identification of the particles reconstructed by the detectors
  - Partons originating from hard interactions, after showering and hadronization, give rise to jets observed in the detector
  - Heavy-flavour jet identification exploits the properties of the hadrons originated in the jet to discriminate heavy flavour (b-,c-) initiated jets from those arising from light partons

CMS DP-2017/032

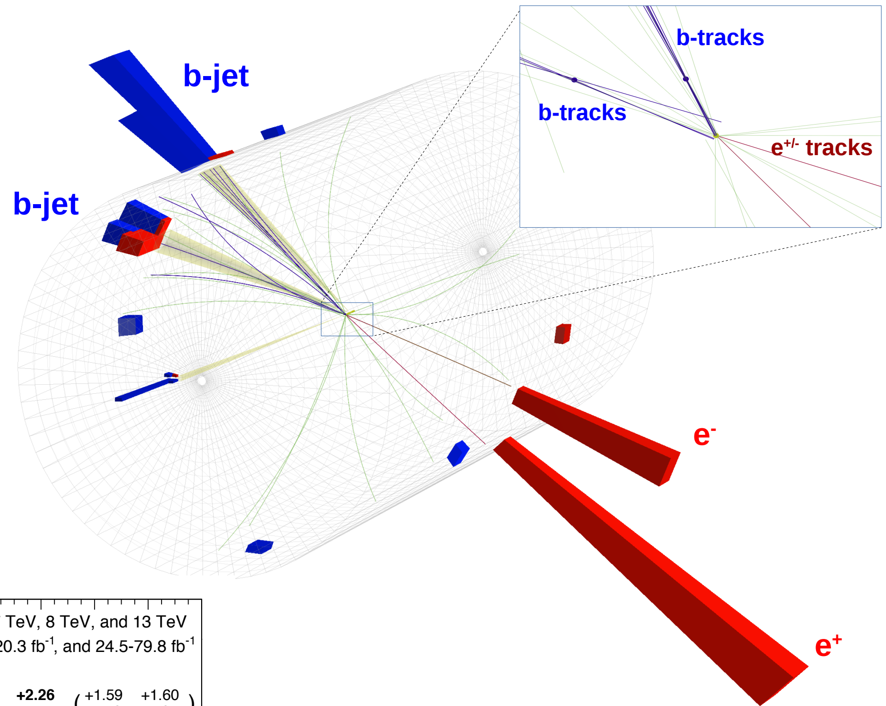




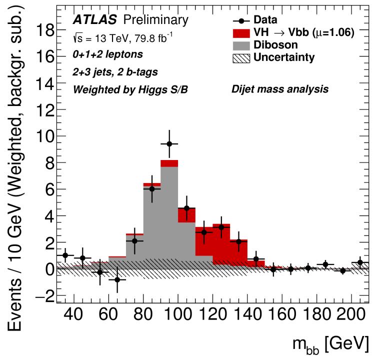
Introduction

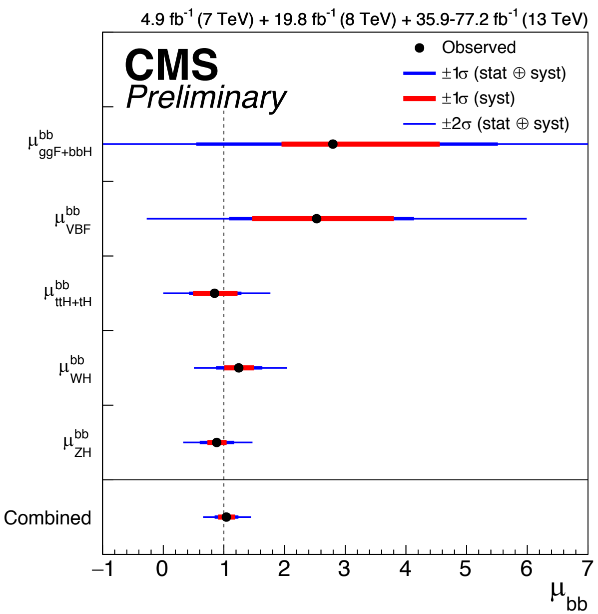
Why b-tagging?

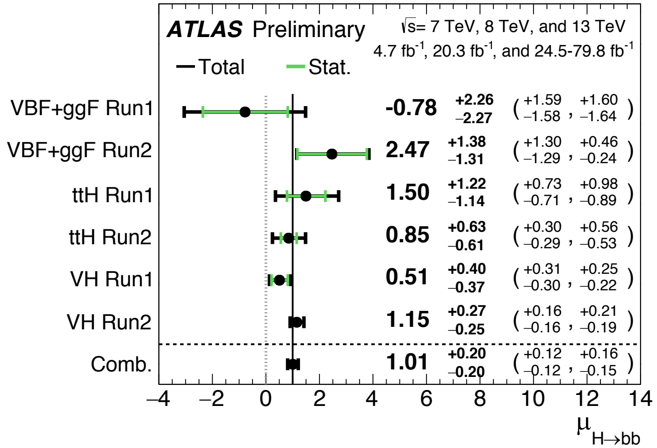
- B-tagging is an essential tool to be exploited to study physics processes with b-jets in their final state:
  - **SM Higgs sectors (H->bb, HH->bbbb,…)**
  - **Top physics (t->Wb)**
  - **BSM searches (X -> bY)**
  - **Also used as veto for many backgrounds (H->WW)**



→Lead to discovery of H→bb!



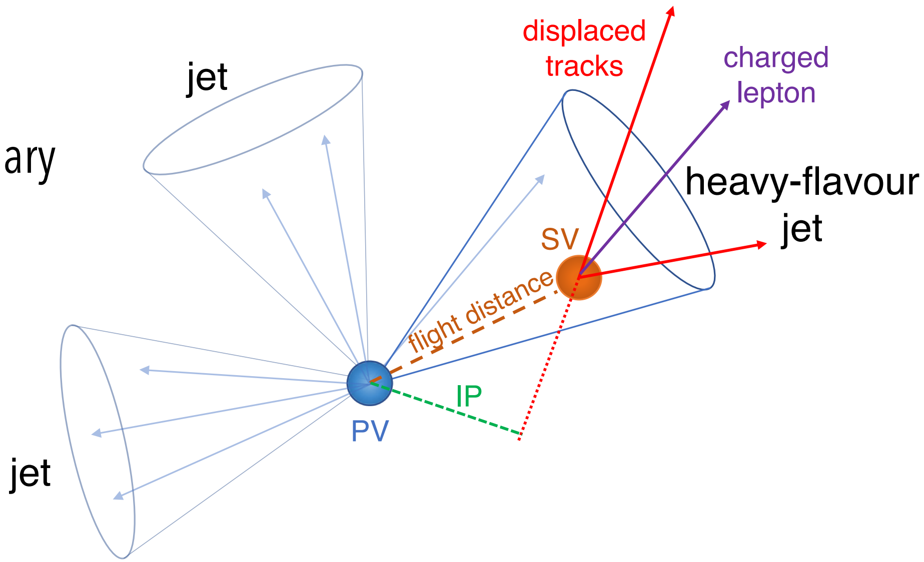




H->bb (cmb)	Exp.	Obs.	μ
ATLAS	5.5	5.4	1.01 ± 0.20
CMS	5.6	5.5	1.04 ± 0.20

▪ Overview on b-tagging

- b-jet tagging rely on b-hadron properties
  - Displaced vertex (secondary vertex) from primary vertex due to its long life (~1.5ps)
  - Large B-hadron mass
  - Large impact parameters (d0)
  - Semi-leptonic e/μ decay of B-hadron (~40% total B hadron decays)



- Different optimized WPs in term of b-efficiency vs mistag rate
- b-jet efficiency and purity is an important metric to assess tagger performance

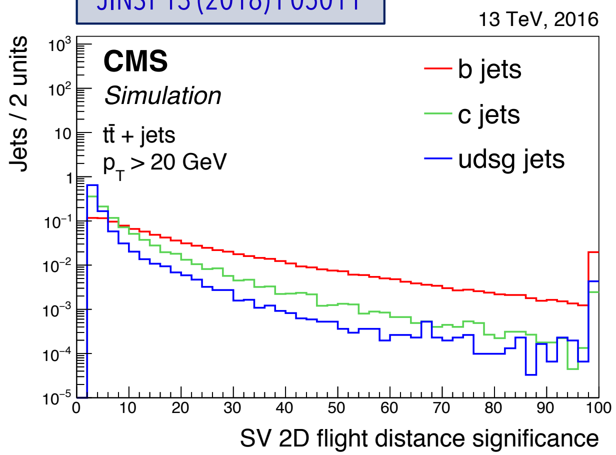
A variety of b tagging algorithms has been developed by ATLAS and CMS. After a short overview
➔ focus on latest state-of-the-art techniques.

Tracking and Vertexing in ATLAS and CMS

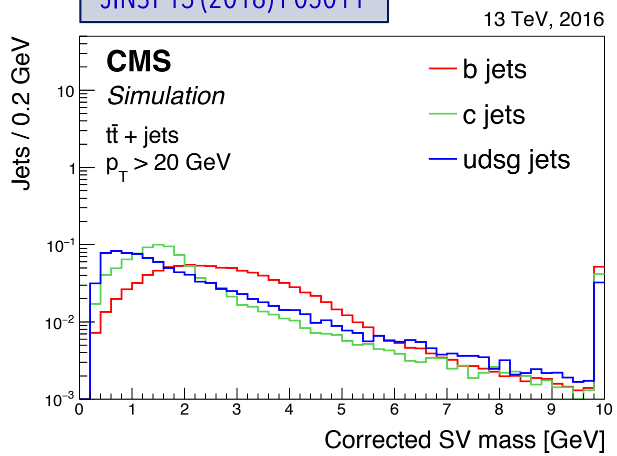
Track reconstruction

- Both in ATLAS and CMS, iterative algorithms have been developed:
- ➔Inside-out: Silicon seed → extrapolate to TRT (primary tracking), Outside-in:TRT seed → extrapolate to silicon
- ➔ Iterative combinatorial track finding (Seed gen+track-fitting+track-selection)

JINST 13 (2018) P05011



JINST 13 (2018) P05011



Secondary Vertex reconstruction (CMS)

- Two algorithms have been designed: **Adaptive Vertex Reconstruction (AVR)** used during Run-1 and **Inclusive Vertex Finding (IVF)** developed for Run-2:
- ➔ AVR: Iteratively fit tracks clustered within a jet (dR<0.3) after passing basic selection
- ➔ IVR: Starts from all reco tracks with pt>0.8 GeV: track clustering + Secondary vertex fitting + track arbitration & cleaning + Secondary vertex refitting

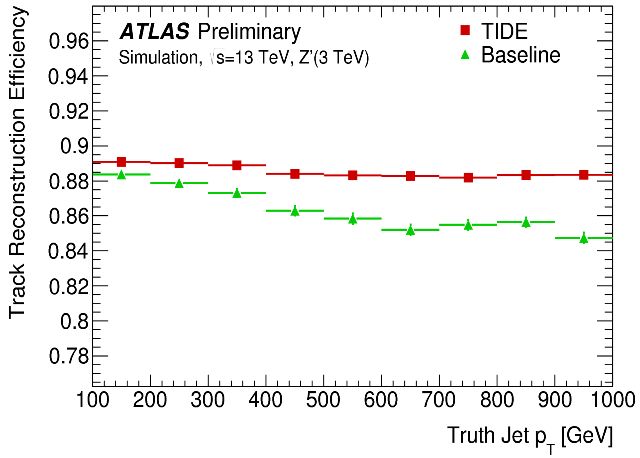
b-tagging at LHC

Tracking and Vertexing in ATLAS and CMS

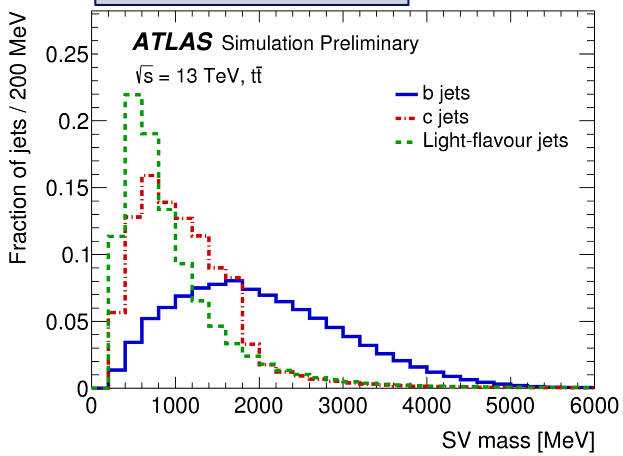
Track reconstruction

- Both in ATLAS and CMS, two method have been developed:
- → Inside-out tracking: Silicon seed → extrapolate to TRT (primary tracking)
- → Outside-in tracking: TRT seed → extrapolate to silicon (conversion recovering)

ATL-PHYS-PUB-2015-006



ATL-PHYS-PUB-2017-011



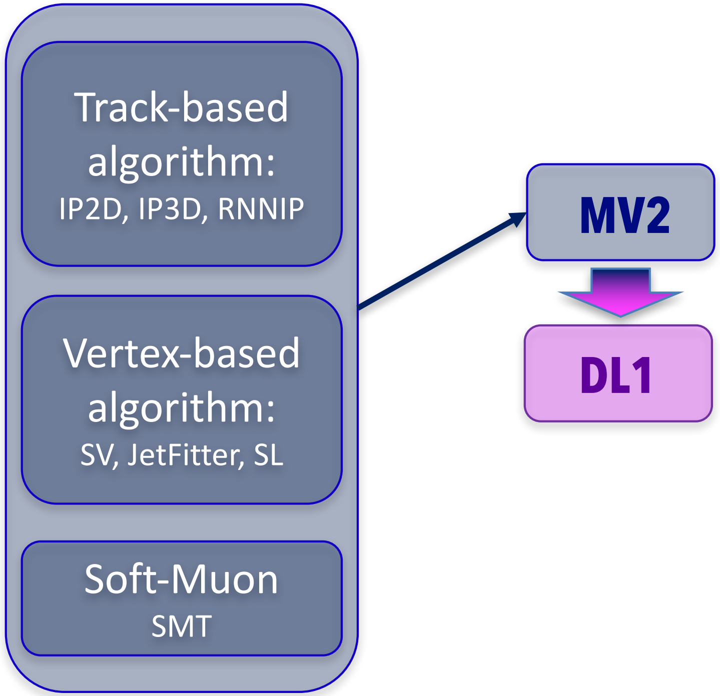
Secondary Vertex reconstruction (ATLAS)

- **Single Secondary Vertex Finder (SSVF):**
- → SVF: consider tracks inside a jet cone f(pT), jet axis direction and PV position
- → SSVF: based on SVF, tighter track selection + 3D-IP of a track with respect to the PV
- → Iteratively reconstruct a single secondary vertex per jet (b and c too close, or highest track multiplicity)

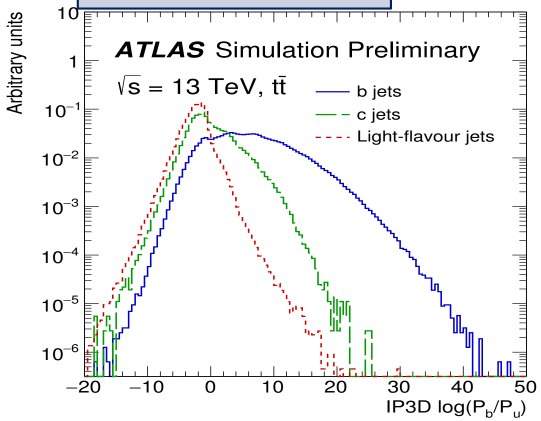
b-tagging strategies in ATLAS

▪ Overview on b-tagging algorithm in ATLAS

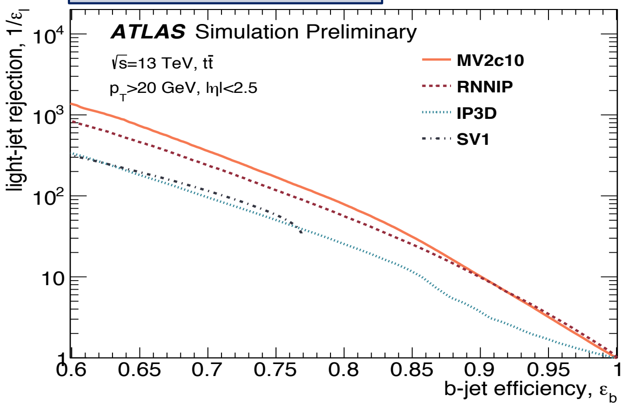
- ➢ IP2D, IP3D
  - lifetime-based tagging algorithms
- ➢ RNNIP
  - Recurrent NN tagger: better exploits correlations between tracks associated do jets
- ➢ SV, JetFitter
  - SV: reconstruction of inclusive displaced SV within the jet. JetFitter: Reconstruct the decay chain exploiting decay topology
- ➢ MVA discriminators (MV2* and MV2RNN, details next slides)
  - MVA based discriminator family exploiting jet kinematics, IPs, SVs and JetFitter information. MV2* are based on BDT
- ➢ DL1 (more details in the next slides)
  - DNN Multiclassifier: same inputs as MV2



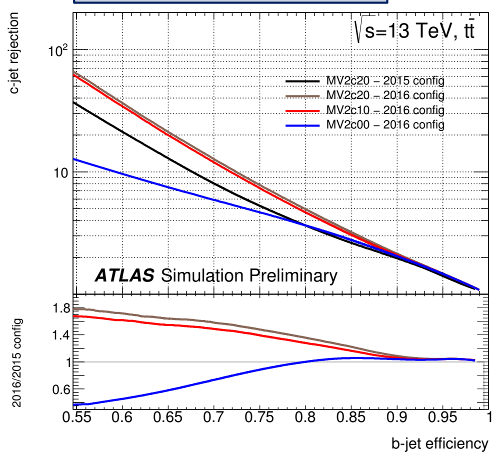
ATL-PHYS-PUB-2016-012



ATL-PHYS-PUB-2017-003



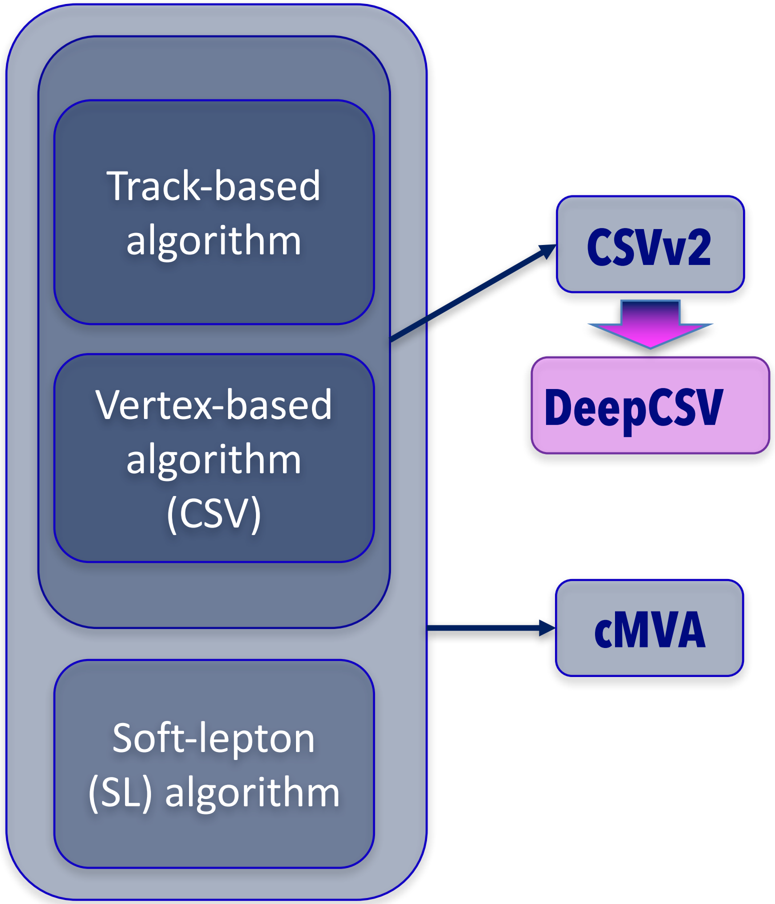
ATL-PHYS-PUB-2016-012



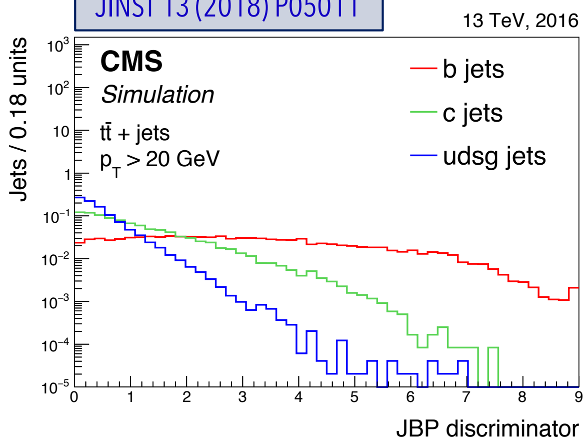
b-tagging strategy in CMS

Overview on b-tagging algorithm in CMS

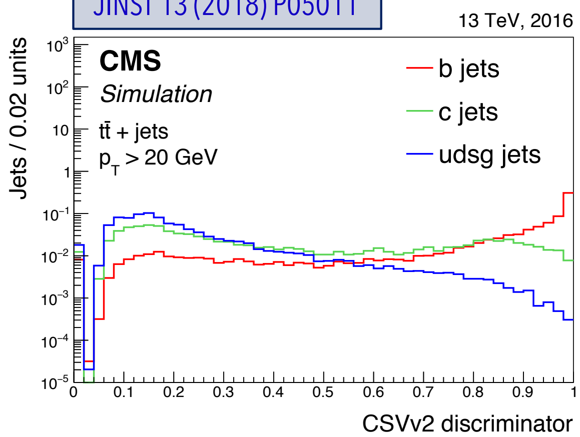
- JP and JBP
  - Likelihood based on the track properties (displacement). Returns p(b-jet)
- CSV and CSVv2
  - Combine displaced tracks with secondary vertices in BDTs (CSV) and in multilayer perceptrons (CSVv2)
- DeepCSV (more details in the next slides)
  - DNN Multiclassifier: same inputs as CSVv2 with a simple extension to use more charged particle tracks
- DeepFlavour (more details in the next slides)
  - DNN Multiclassifier



JINST 13 (2018) P05011



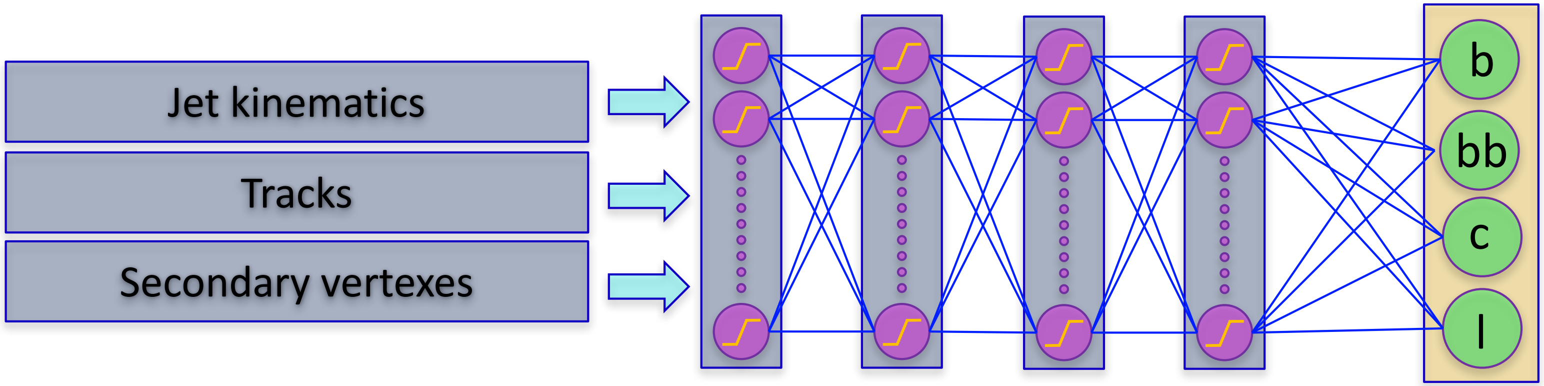
JINST 13 (2018) P05011



- cMVA: combined multivariate analysis (cMVA) tagger, combines the discriminator values of low-level tagger

Cutting edge technology - CMS

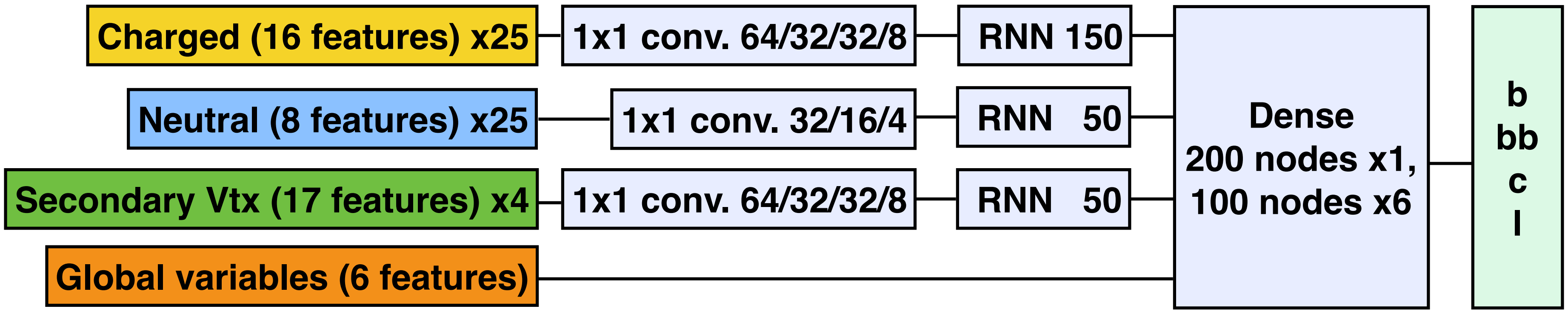
- **DeepCSV: DNN architecture**
  - Input variables go through 4 fully connected layers, each layer has 100 nodes
  - ReLu activation function used in each of the hidden nodes
  - Output layer → softmax activation function → multiclassification



- **DeepCSV: Training**
  - Training performed with Keras DL-library interfaced with Tensorflow
  - Jets with pT in [20,1000] GeV and flavour ratio fixed to **2 : 1 : 4** for **b : c : light**
  - Mixture of tt and multijets events →reduce dependence on heavy-flavour quarks production process

Cutting edge technology - CMS

- **DeepFlavour: DNN architecture**
  - Inputs: 16 properties of up to 25 charged and 6 properties up to 25 neutral PF-jet constituents, as well as 17 properties from up to 4 secondary vertices associated with the jet.
  - Separate 1x1 convolutional layers are trained: 4 hidden layers with 64,32,32, and 8 filters for charged candidates and vertices and 3 hidden layers with 32,16, and 4 filters for neutral particles
  - Output is then separately fed into 3 (LSTMs) with 150, 50 and 50 output nodes and then merged with global jet properties and fed through one dense layer with 200 nodes, followed by 7 hidden dense layers with 100 nodes



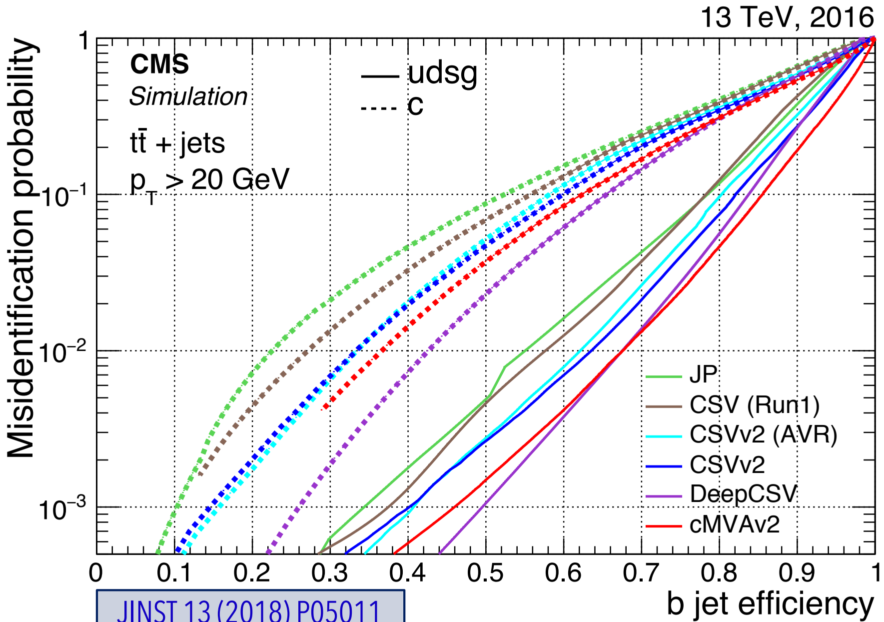
- **DeepFlavour: Training**
  - Trained on 40M jets from top-quark pair and QCD events
  - Almost full set of jet constituents without strong preselection

Performance in Simulation - CMS

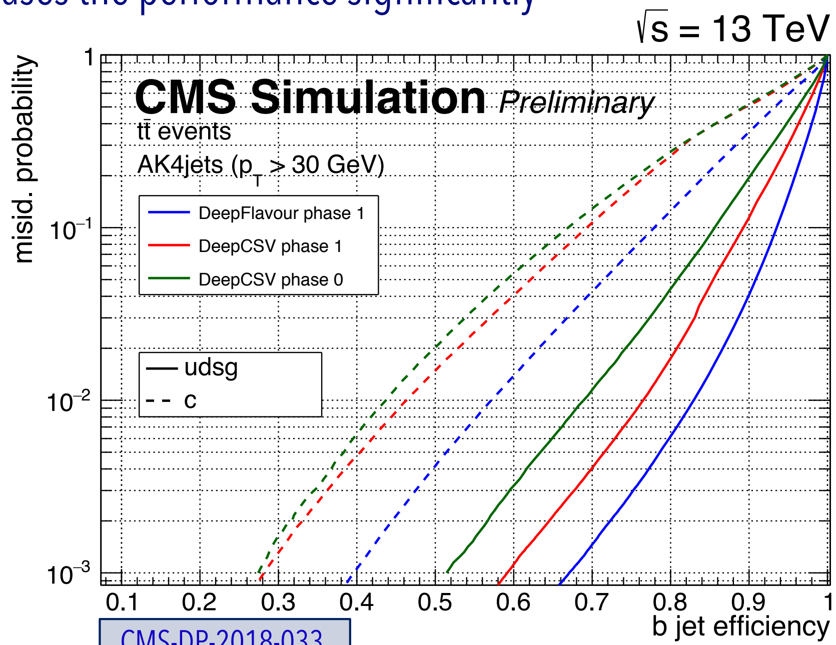
- **DeepCSV**
  - Performance evaluated in simulated tt events, considering AK4 Jets with pT>20GeV
  - DeepCSV performance are compared against those of other commissioned taggers in CMS
  - DeepCSV WPs are defined as values of the discriminator cut for which the light mistag-rate is 10%, 1%, and 0.1%,

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

- **DeepFlavour**
  - Performance evaluated in simulated tt events (plot), considering AK4 Jets with pT>30GeV
  - Simply adding more information can even degrade performance
  - Adding convolutional layers (exploiting structures) increases the performance significantly



JINST 13 (2018) P05011



CMS-DP-2018-033

Performance in Simulation - CMS

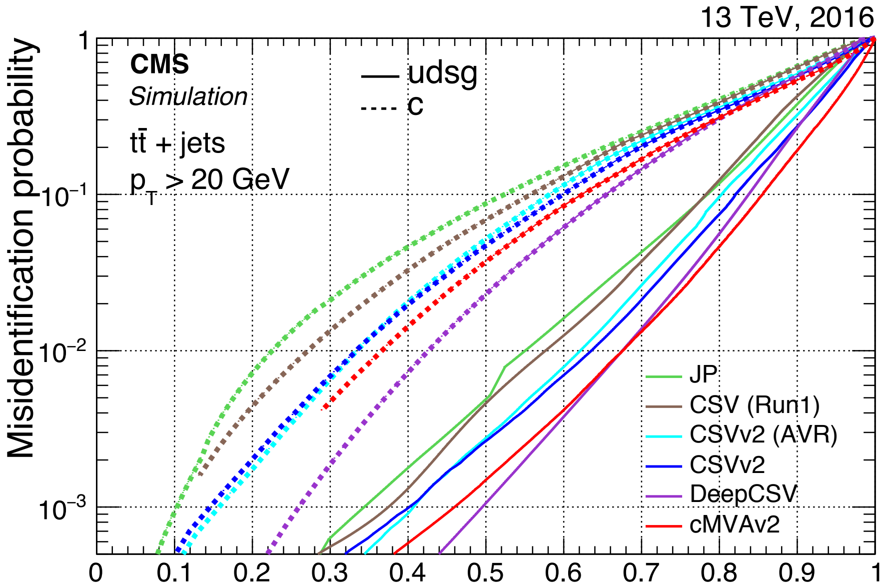
DeepCSV

- Performance evaluated in simulated tt events, considering AK4 Jets with pT>20GeV
- DeepCSV performance are compared against those of other commissioned taggers in CMS
- DeepCSV WPs are defined as values of the discriminator cut for which the light mistag-rate is 10%, 1%, and 0.1%,

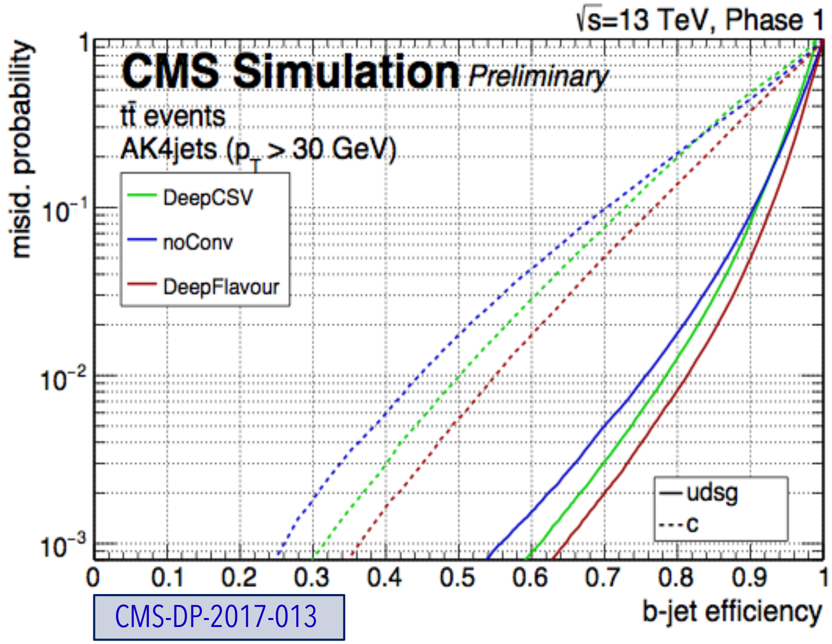
Tagger	Working point	ε_b (%)	ε_c (%)	ε_{udsg} (%)
	DeepCSV L	84	41	11
Deep combined secondary vertex	DeepCSV M	68	12	1.1
(DeepCSV) $P(b) + P(bb)$	DeepCSV T	50	2.4	0.1

DeepFlavour

- Performance evaluated in simulated tt events (plot), considering AK4 Jets with pT>30GeV
- Simply adding more information can even degrade performance
- Adding convolutional layers (exploiting structures) increases the performance significantly



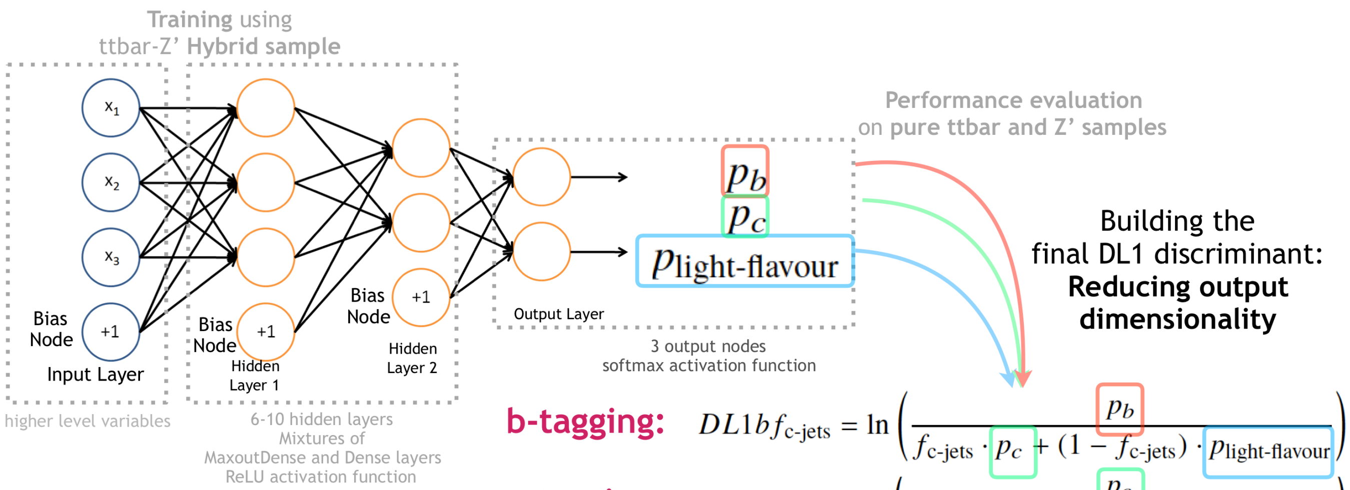
JINST 13 (2018) P05011



CMS-DP-2017-013

- **MV2 tagger**
  - High level tagger that combine inputs from IP2D, IP3D, RNNIP, JetFitter and SL taggers
  - Based on the training of a BDT algorithm, targetting b-jet as signal and c-(light-)jets as background
  - Different versions of MV2 have been developed by ATLAS, relying on improvement in low-level taggers
  - c-jet fraction in the training is set to 7% and that of light-flavoured jets to 93%.

- **DL1 tagger**
  - Based on artificial DNN, trained with keras and Theano backend using ADAM optimizer
  - Multiclassifier: returns the probability for a jet to be a b-, c- or light-flavour jet
  - It makes use of the same input features used in MV2 + full set of SMT and JetFitter c-taggiing variables



b-tagging: $DL1b f_{\text{c-jets}} = \ln\left(\frac{p_b}{f_{\text{c-jets}} \cdot p_c + (1 - f_{\text{c-jets}}) \cdot p_{\text{light-flavour}}}\right)$

c-tagging: $DL1c f_{\text{b-jets}} = \ln\left(\frac{p_c}{f_{\text{b-jets}} \cdot p_b + (1 - f_{\text{b-jets}}) \cdot p_{\text{light-flavour}}}\right)$

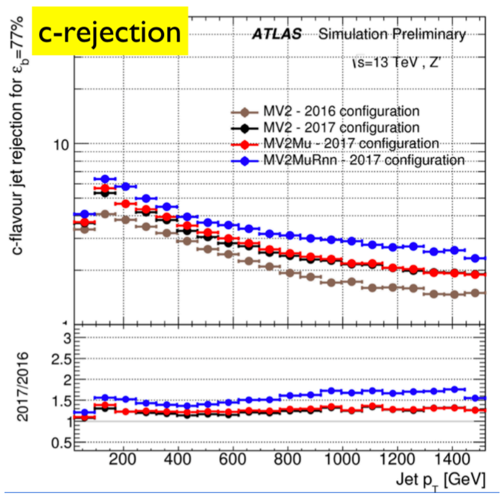
From M. Lanfermann

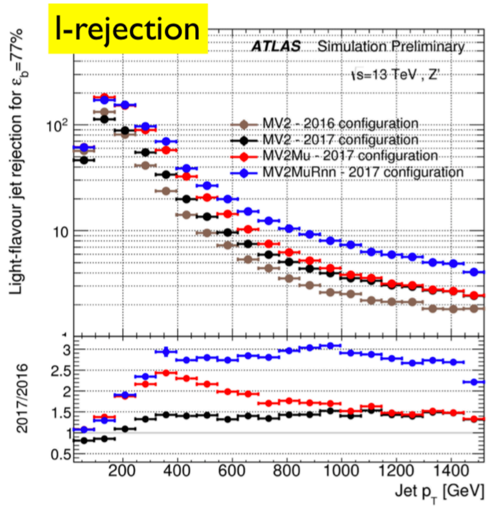
—> **Increased Flexibility:**
+ Background weighing tuneable after training
+ Same training usable for b- and c-tagging

NN config file size ~1MB

Cutting edge technology - ATLAS

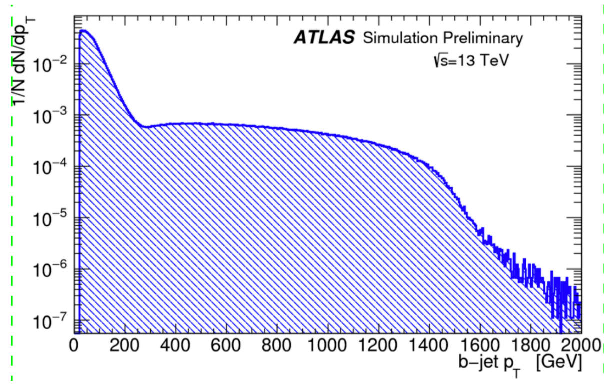
- **Three variants of the MV2 (and DL1) algorithms have been deployed**
  - standard impact parameter and secondary vertex-based inputs + kinematics of the jet (MV2/DL1)
  - standard inputs + soft muon tagger (MV2Mu/DL1Mu)
  - standard inputs + soft muon tagger+ RNNIP (MV2MuRnn/ DL1MuRnn)





Improvements on the full pt spectrum (SMT at low-medium pt, RNN at high pt)

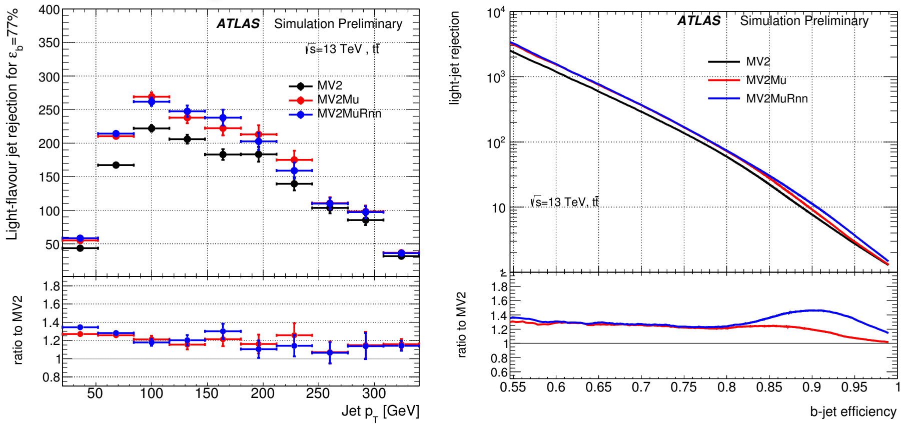
- **Trainings:**
  - Hybrid training procedure
    - ➔ **significant impact on high-pt performance**
  - low pt ttbar + high pt Z' sample composition
  - similar performance at low pt



Performance in Simulation - ATLAS

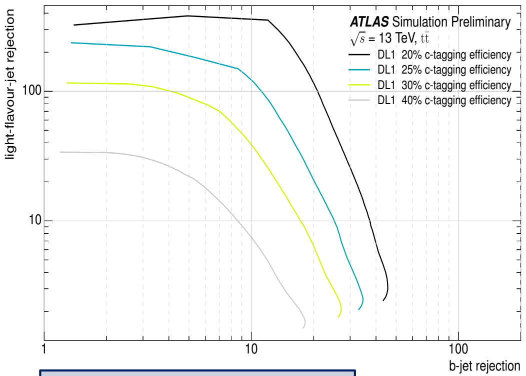
- **MV2**
  - Light-jet rejection as a function of *b*-jet efficiency for MV2 (black), MV2Mu (red), MV2MuRnn (blue)
  - Light-flavour rejection as a function of the jet pT for MV2 (black), MV2Mu (red), MV2MuRnn (blue).
  - The algorithm evaluation is performed on *tt* events for a flat *b*-jet eciency of 77% for each *p*T bin.
  - The ratio calculated for each MV2 variant vs MV2.

BDT Cut Value	*b*-jet Efficiency [%]	*c*-jet Rejection	Light-jet Rejection	τ Rejection
0.9349	60	34	1538	184
0.8244	70	12	381	55
0.6459	77	6	134	22
0.1758	85	3.1	33	8.2



ATL-PHYS-PUB-2017-013

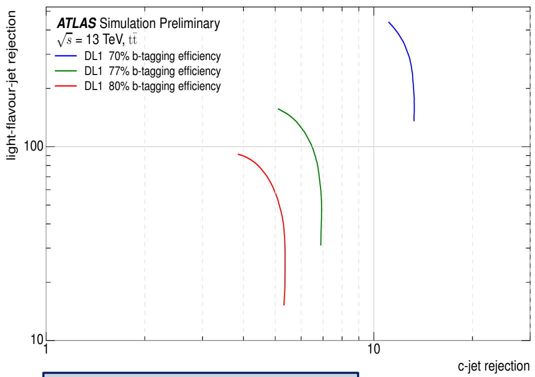
- **DL1**
  - light-flavour vs *b*-jet rejection (a) and *c*-jet rejection (b) for a set of working points corresponding to different values of the *c*-jet and *b*-jet efficiency. The evaluation is performed on $t\bar{t}$ events



ATL-PHYS-PUB-2017-013

B-TAGGING *slightly more inputs*	light-jet rejection	c-jet rejection
77% b-tagging efficiency	~same	(5.8->6.3) 9%

From M. Lanfermann



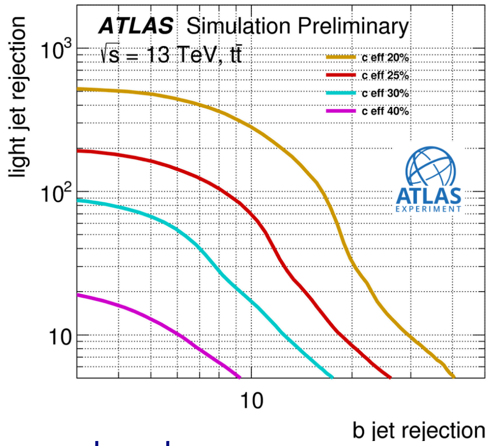
ATL-PHYS-PUB-2017-013

C-TAGGING *same inputs*	light-jet rejection	b-jet rejection
40% c-tagging efficiency	(16—>27) +68%	same
20% c-tagging efficiency	(53—>104) 195%	same

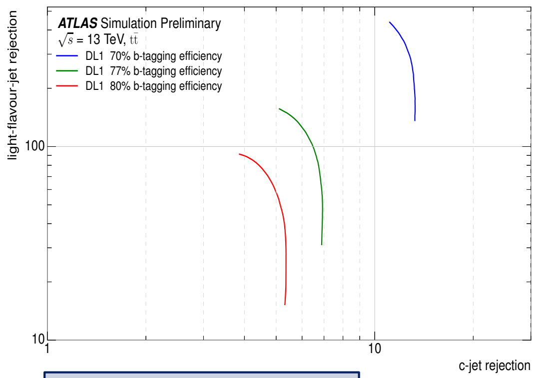
From M. Lanfermann

ATLAS

- Topology of the displaced vertex reconstructed by the JetFitter algorithm in addition to b- tagging inputs used in a dedicated BDT for c- tagging
- MV2c100 (b discrimination), MV2cl100 (light-flavour discrimination)



- DL1 taggers show improved performance



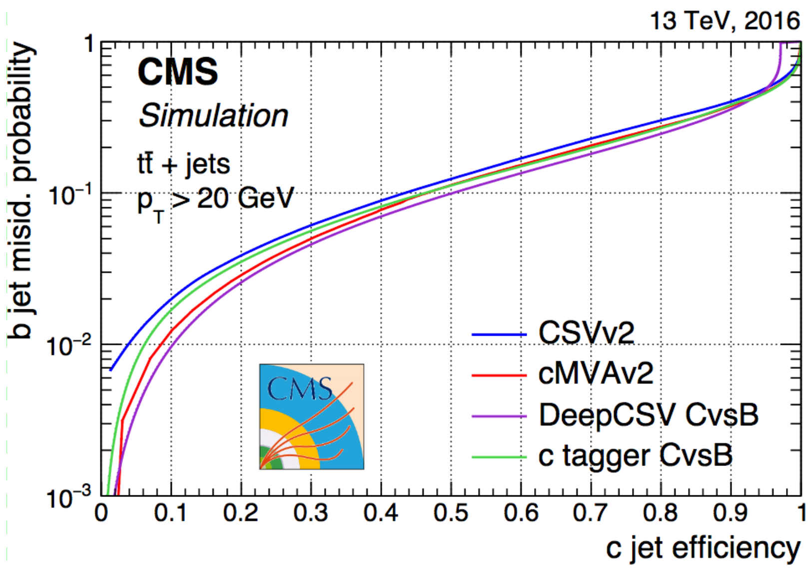
ATL-PHYS-PUB-2017-013

C-TAGGING *same inputs*	light-jet rejection	b-jet rejection
40% c-tagging efficiency	(16—>27) +68%	same
20% c-tagging efficiency	(53—>104) 195%	same

From M. Lanfermann

CMS

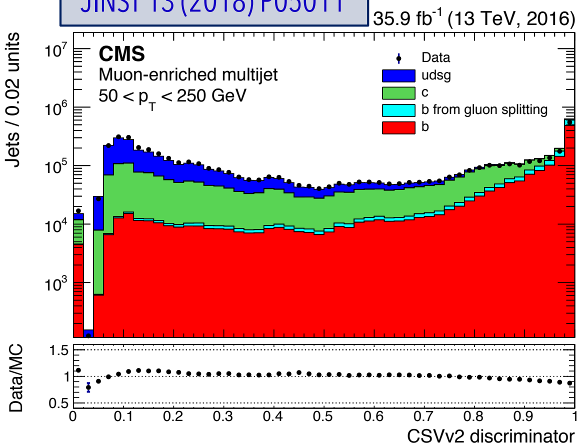
- c-tagging identification based on CSVv2 (now baseline is DeepCSV)
- similar inputs as in b-tagging + additional kinematics of the soft-lepton taggers
- discrimination exploited for c- vs light-flavour and c- vs b-jets (using Gradient Boosting Classifier)
- focus on DNN-based c-tagging response, i.e. DeepCSV - outperforming dedicated CSVv2 algorithm



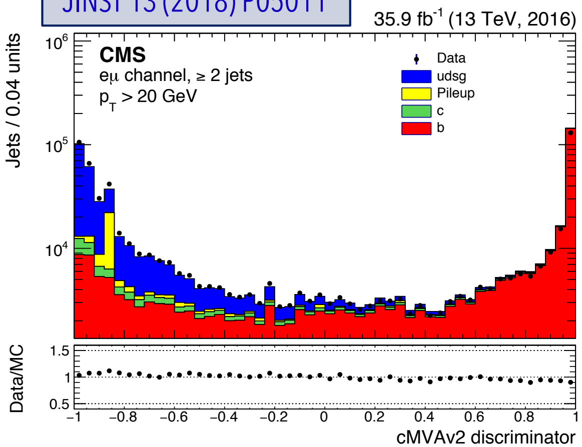
Performance in Data - CMS

- **Efficiency for b-jets (Measurement in ttbar events, 4 methods)**
  - Kin: di-leptonic channel, based on template fit
  - TnP: semileptonic channel. Tag: CSV medium requirement to either b-jets. Probe: the other
  - TagCount: di-leptonic channel, based on counting events with two b-tagged jets in selected sample
  - Iterative-Fit: di-leptonic channel, based on calibration of the full b-tagging shape
  - Systematics uncertainties are small and mostly dominated by ttbar and HF modeling

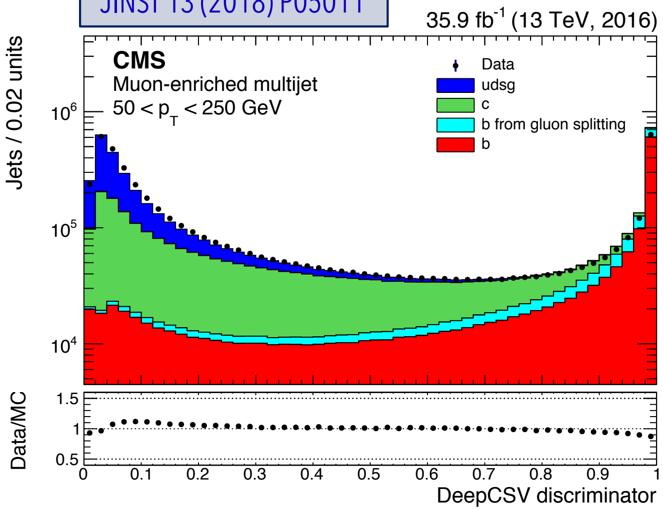
JINST 13 (2018) P05011



JINST 13 (2018) P05011



JINST 13 (2018) P05011



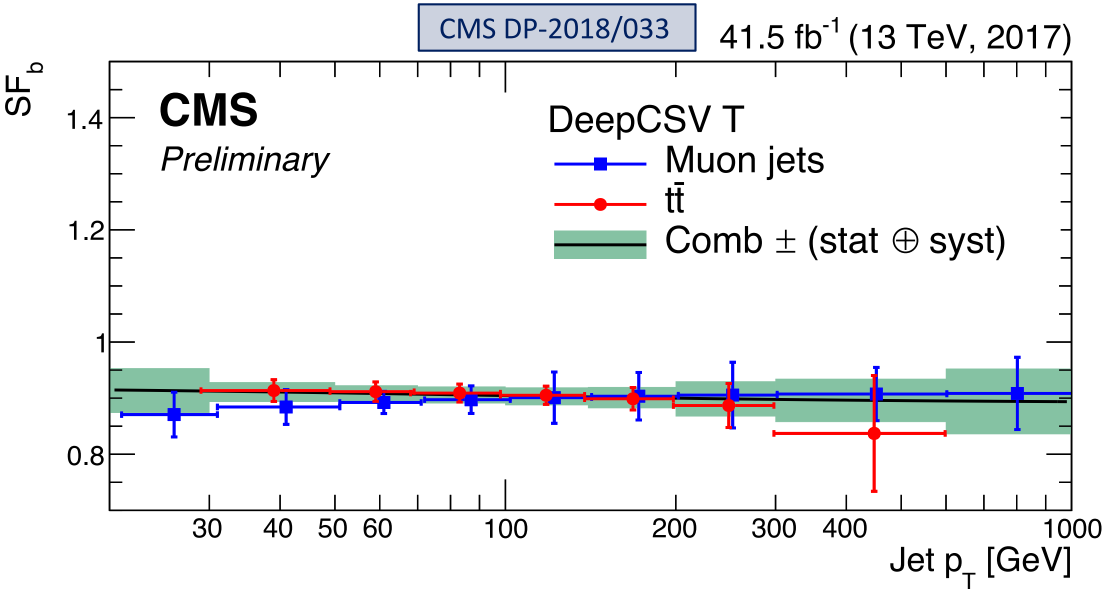
- **Efficiency for b-jets (Mu-enriched sample)**
  - 2 AK4 jets with 50<pT<250 GeV, with at least one containing a muon with pT>5GeV
  - Events enriched in high-flavour jets

Determination of data/MC scale factors

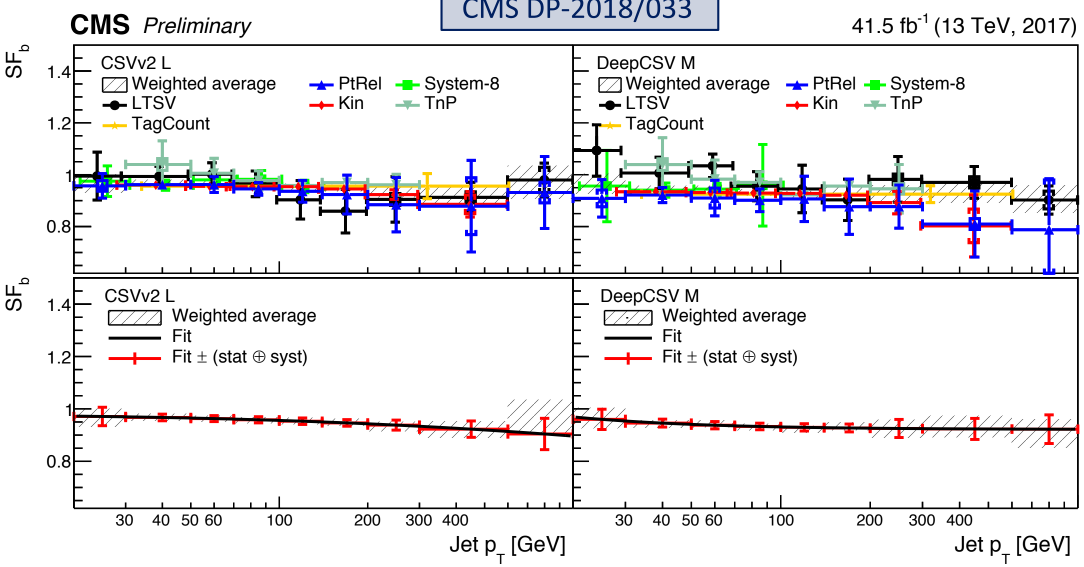
CSVv2 tight WP

- Measured b-tagging efficiency in multijet events with a muon, based on the combination of the results from different measurements, obtained using the PtRel, the LT and the System8 methods.
- Comparison of the b-jet SF for the tight working point in mu-enriched multijets and in tt events + combination
- including in the systematics an additional 1% uncertainty to cover any residual sample dependence
- The SFs found via the different methods agree within the experimental uncertainty

CMS DP-2018/033



CMS DP-2018/033



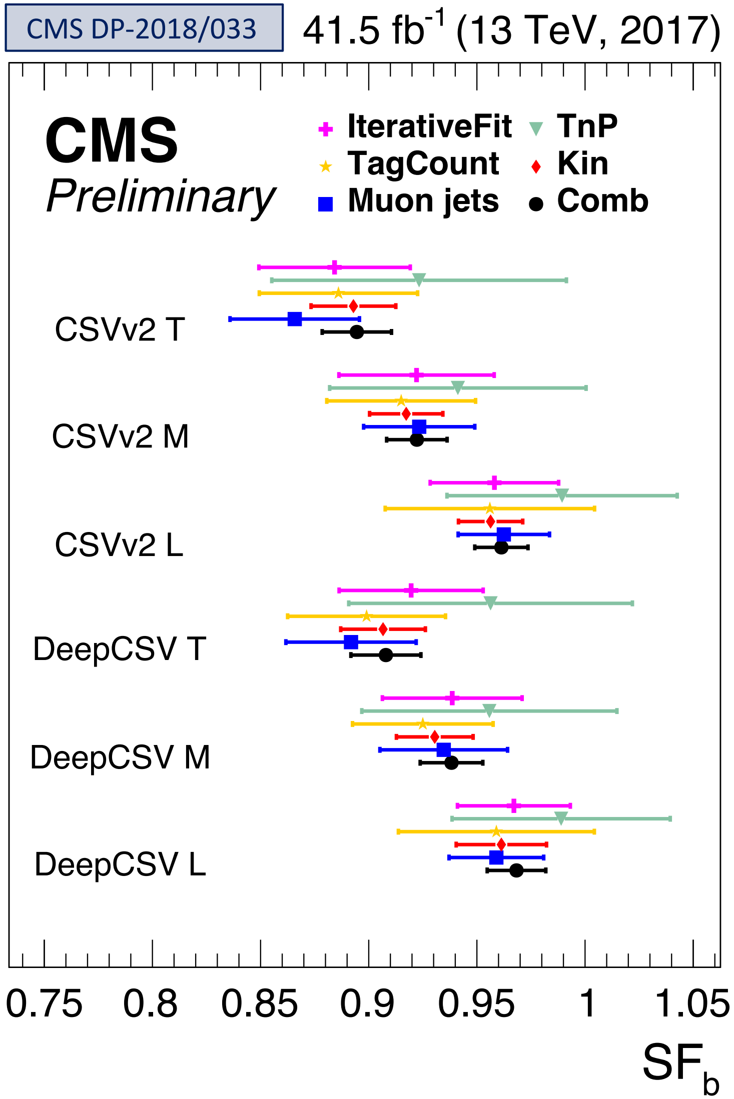
CSVv2 sv DeepCSV

- SF for b jets as a function of the jet pT for the loose CSVv2 (left) and the medium DeepCSV (right) WPs
- SFs for tagging b as a function of the jet pT are shown measured with the various methods
- Statistical uncertainty (inner bar) + statistical and systematic uncertainty (outer bar)

Determination of data/MC scale factors

Summary of SF derivation

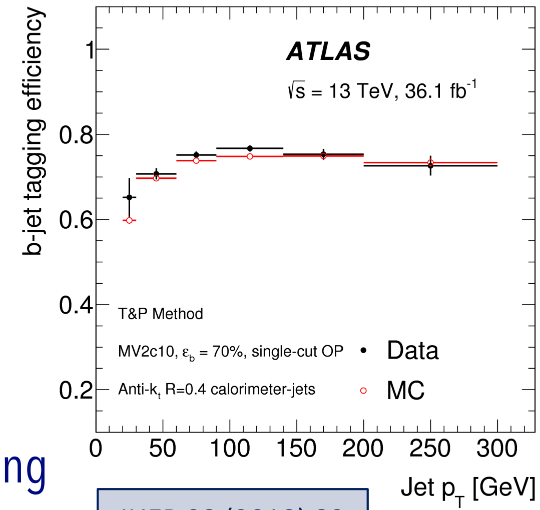
- Comparison of the b-jet data-to-simulation scale factors derived with various methods and their combination.
- The values of the scale factors are averaged over the spectrum of b-jets from ttbar events.
- The scale factors measured with the different methods agree within their uncertainties.
- The combination includes the results from **all** the measurements with the exception of the IterativeFit and the TagCount methods.



Performance in Data - ATLAS

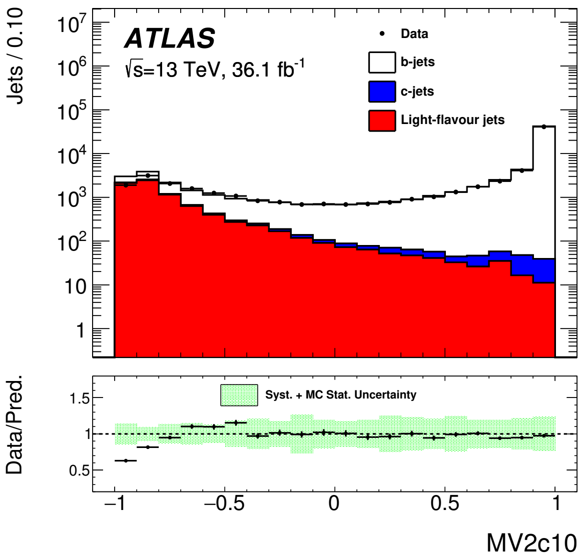
▪ Efficiency for b-jets (High purity di-leptonic ttbar sample)

- Combinatorial Likelihood approach
- TnP: 90% pure sample in ttbar events. Jet is considered a probe if the other jet is b-tagged at the 85% efficiency OP
- Events are selected using a dedicated BDT not using the b-tagging information (used to select events with two b-jets)
- Calibration for MV2c10 and DL1 taggers
- Systematics are small and mostly dominated by ttbar and HF modeling

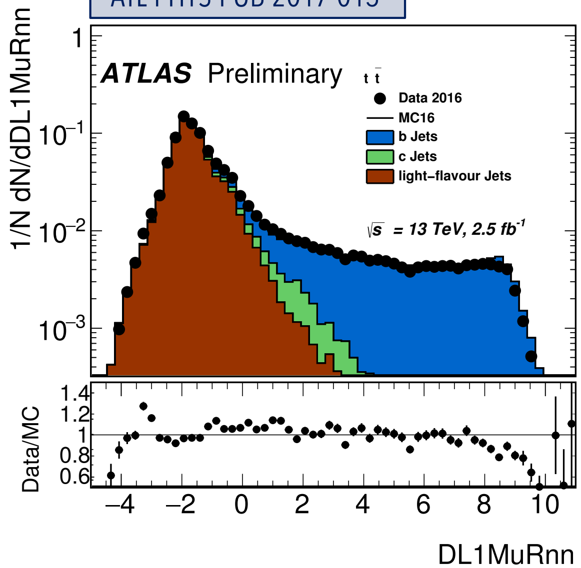


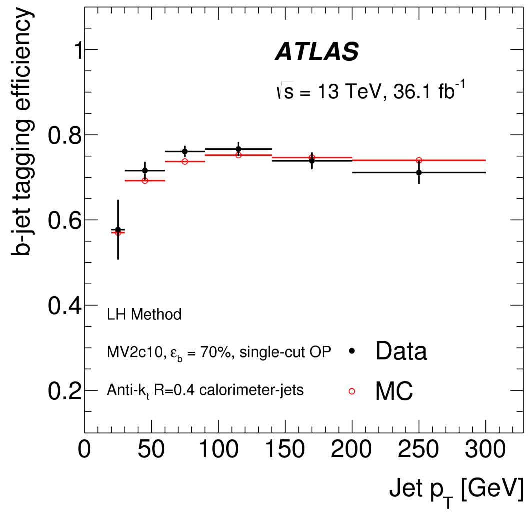
JHEP 08 (2018) 89

JHEP 08 (2018) 89



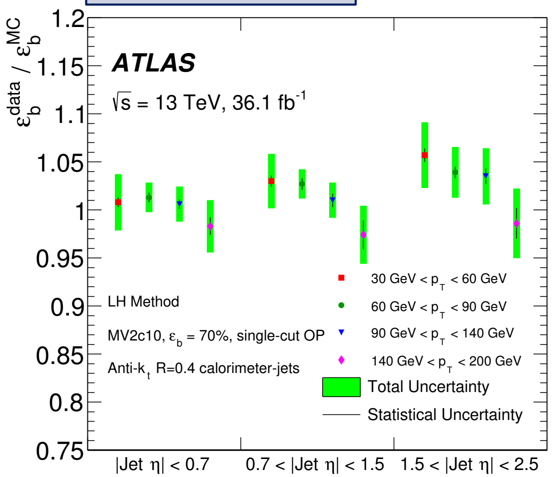
ATL-PHYS-PUB-2017-013



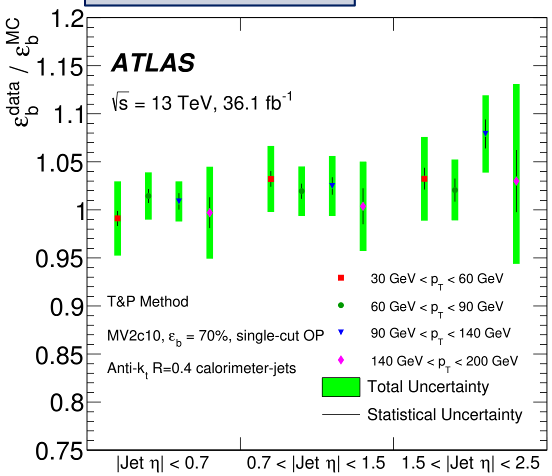


- **70% eff OP**
  - Data-to-simulation scale factors, corresponding to the 70% b-jet tagging efficiency single-cut OP
  - Scale factors have been derived in different pseudorapidity bins
  - LH method (left) is compared to the T&P one (right)
  - The SFs found via the two different methods agree within the experimental uncertainty
  - Comparison between data and simulation SF as function of Jet-pT (before/after smoothing) for LH and T&P methods

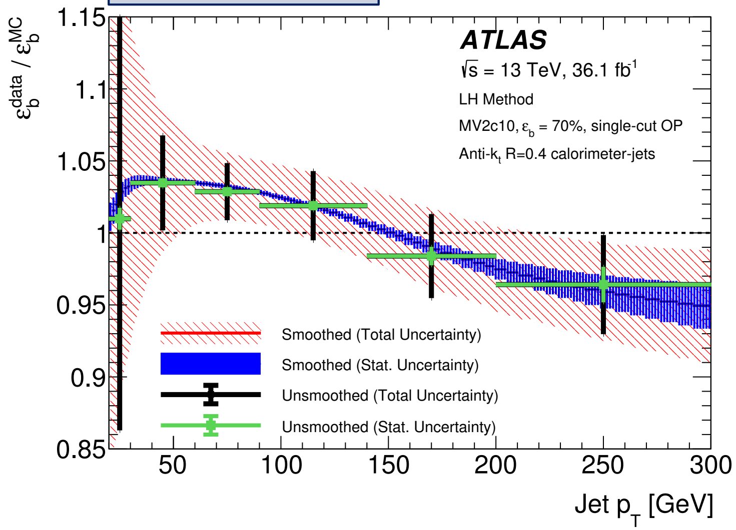
JHEP 08 (2018) 89



JHEP 08 (2018) 89



JHEP 08 (2018) 89

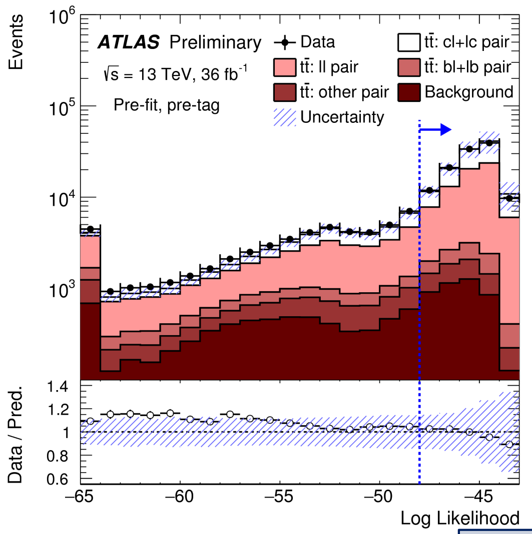


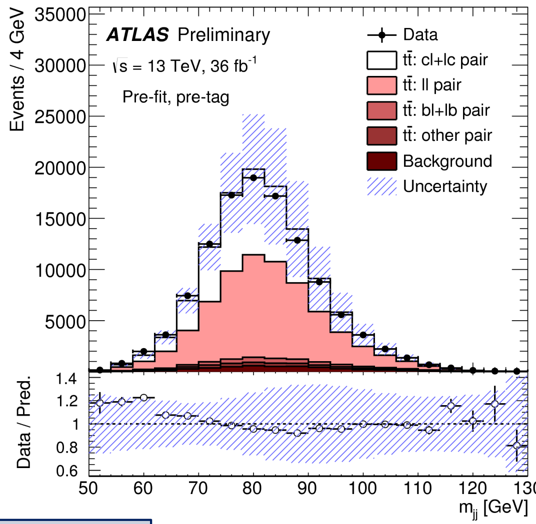
Performance in Data – ATLAS & CMS

Efficiency for c-jets

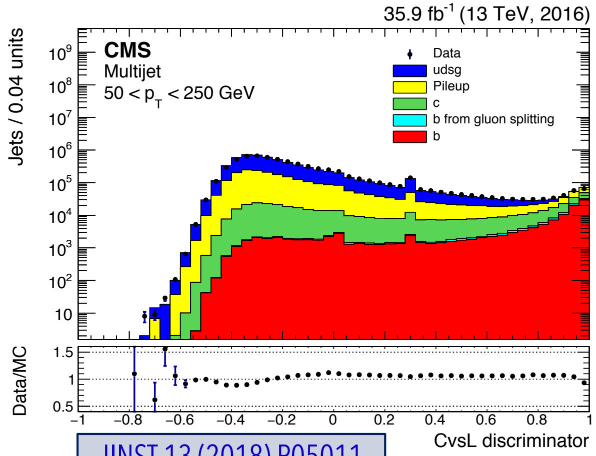
- ATLAS and CMS use two c-enriched topologies, ttbar events in single lepton final state (W→lν, W→cs) and W+c selected by searching for a soft muon (W→μν) in the c-jets
- cut-and-count analysis in W+c (ATLAS) and fit to discriminant to extract c-tagging efficiency in ttbar (ATLAS and CMS)
- uncertainties dominated by ttbar modeling

  (10-20% depending on pT)

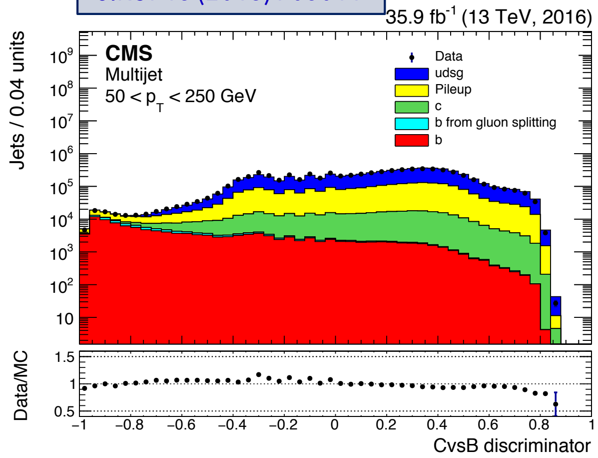




ATLAS-CONF-2018-001



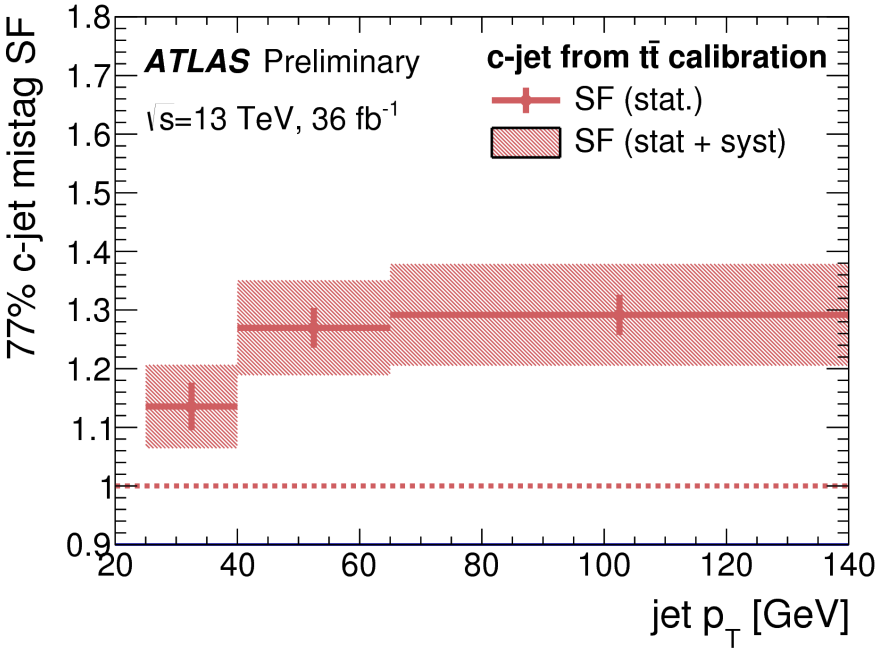
JINST 13 (2018) P05011

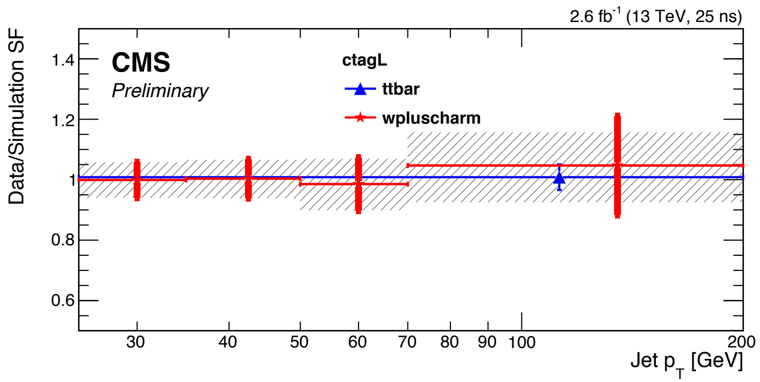


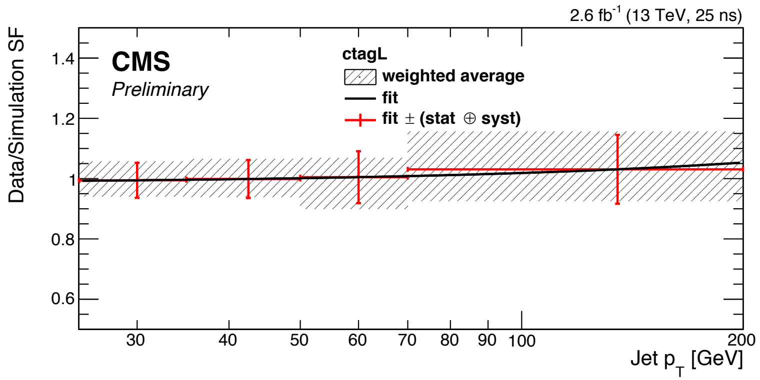
Determination of data/MC scale factors

c-tagging SF measurement – ATLAS & CMS

- ATLAS and CMS use two c-enriched topologies, ttbar events in single lepton final state (W→lν, W→cs) and W+c selected by searching for a soft muon (W→μν) in the c-jets
- cut-and-count analysis in W+c (ATLAS) and fit to discriminant to extract c-tagging efficiency in ttbar (ATLAS and CMS)
- uncertainties dominated by ttbar modeling (10-20% depending on pT)



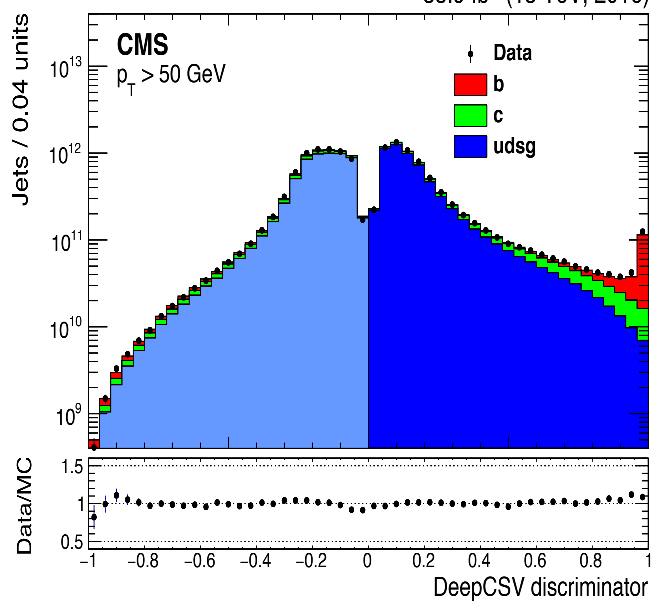




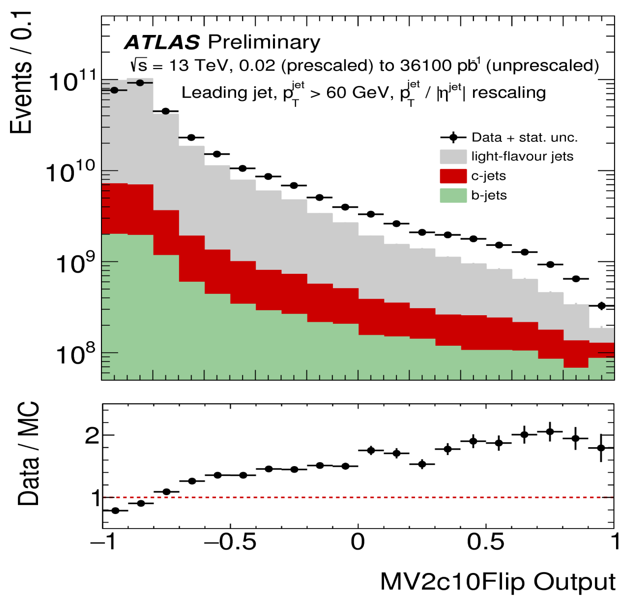
Performance in Data – ATLAS & CMS

- **Light-jets misitag rate measurement**
  - **ATLAS & CMS:**
    - flipped-taggers to calibrate fake-jets generated from track- resolution effect (larger uncert.)
    - flipped taggers exhibit similar mistag rate for light-jets and much smaller discrimination power for b/c-tagging
      → light- enriched sample posttag
  - **ATLAS:**
    - Using data sample enriched in light flavour (exploiting dedicated algorithm to reduce the b--jet efficiency)
    - Bottom-Up approach: simulated underlying track variable are required to match distribution in data
      → b-tagging algorithms are re-evaluated
      → access change in light-jet mistag rate

JINST 13 (2018) P05011



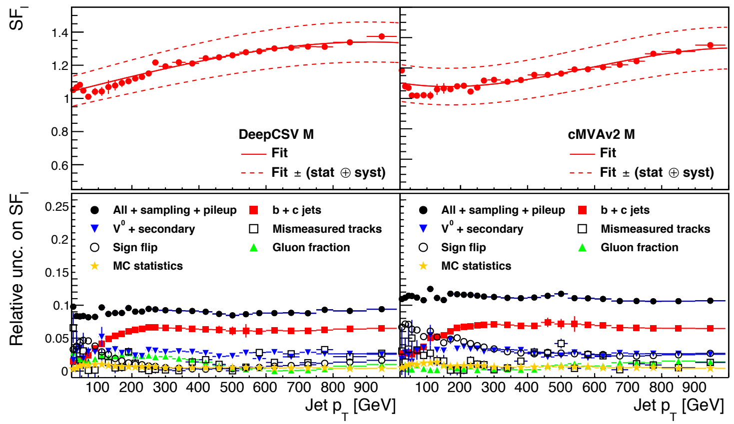
ATLAS-CONF-2018-006



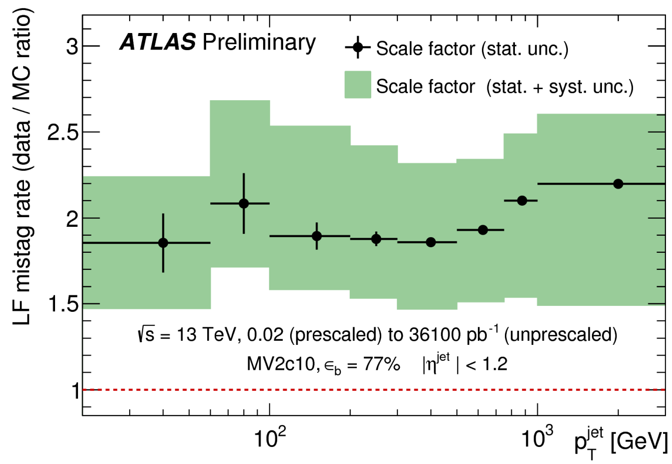
Performance in Data – ATLAS & CMS

Light-jets misitag rate measurement

- **ATLAS & CMS:**
  - flipped-taggers to calibrate fake-jets generated from track- resolution effect (larger uncert.)
  - flipped taggers exhibit similar mistag rate for light-jets and much smaller discrimination power for b/c-tagging
    → light- enriched sample posttag



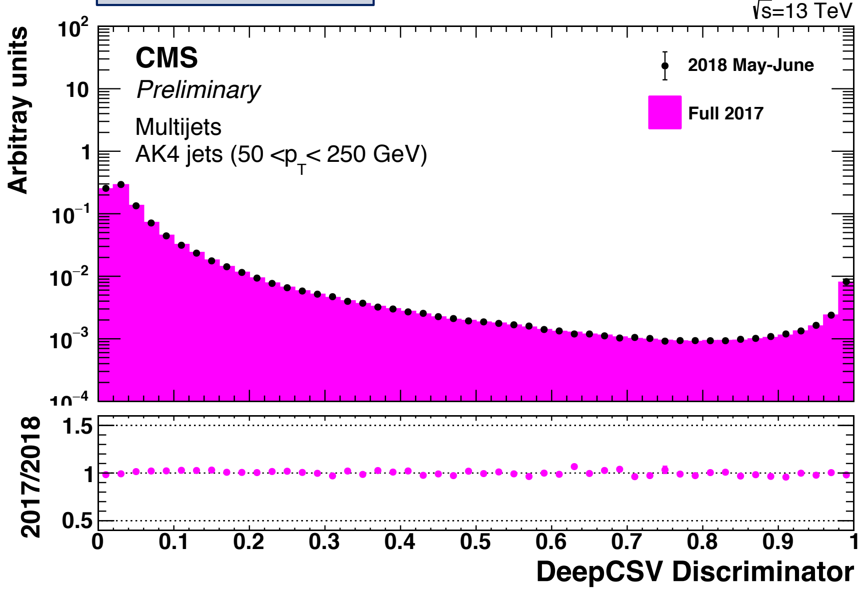
- **ATLAS:**
  - Using data sample enriched in light flavour (exploiting dedicated algorithm to reduce the b--jet efficiency)
  - Bottom-Up approach: simulated underlying track variable are required to match distribution in data
    → b-tagging algorithms are re-evaluated
    → access change in light-jet mistag rate



Commissioning with 2018 Data

- **CMS (DeepCSV)**
  - Inclusive multijet event sample.
  - Online selection: AK4 jets with a pT>40 GeV.
  - Offline selection: pT in [50, 250] GeV and a $|\eta|<2.4$
  - Best knowledge of the 2018 detector condition
  - Full 2017 data set used for comparison
  - For comparison purposes, the integrals of all are normalized to unity.

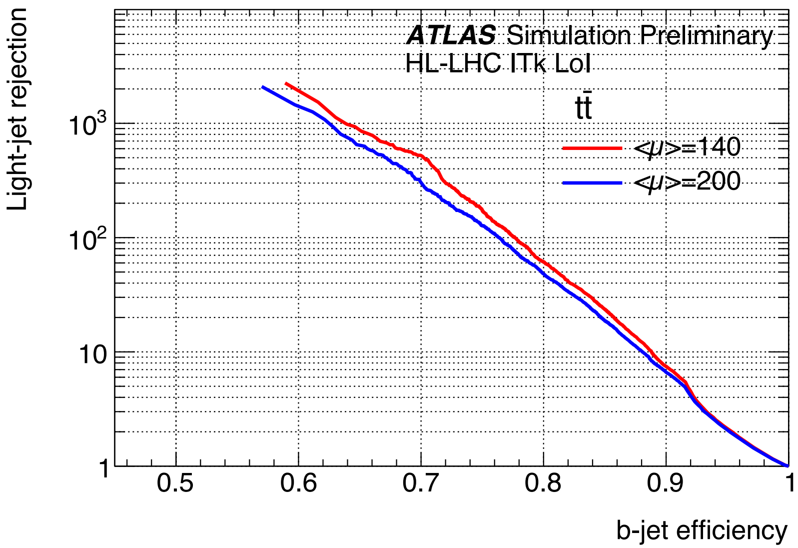
CMS DP-2018/033



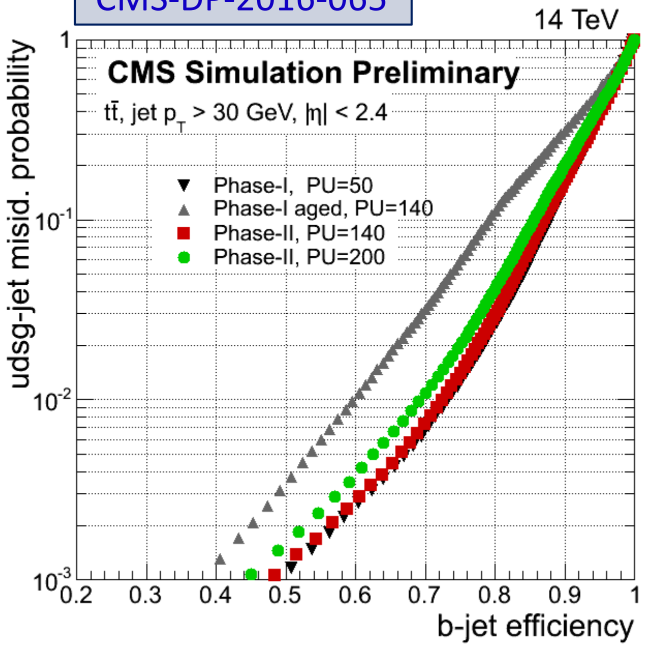
Future prospect – HL-LHC

- Major upgrade of ATLAS and CMS detectors planned to operate during the High Luminosity (HL) LHC phase
  - Trackers will be replaced with new detectors 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



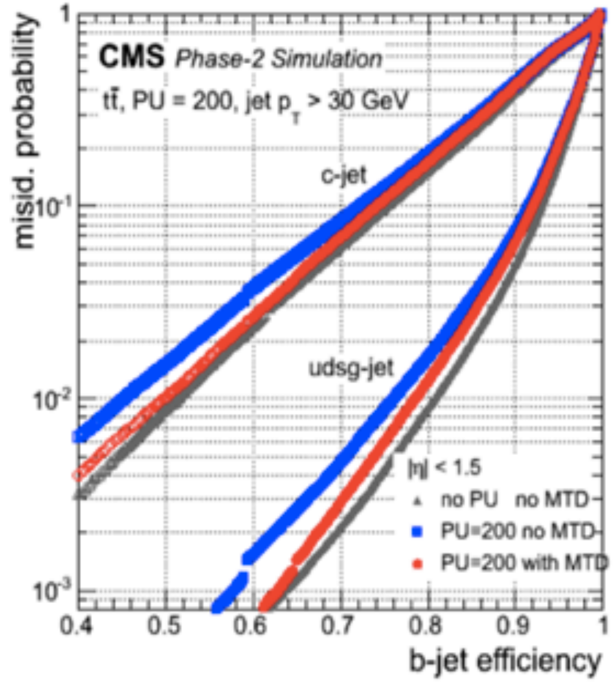
CMS-DP-2016-065



Future prospect – HL-LHC

Considerations about HL-LHC

- ***SM HH discovery is challenging*** but analysis improvements thus far are fa
- We will have a **new tracker detector at HL-LHC…**
  - *10% improvement in signal acceptance* for H($b\bar{b}$)H($b\bar{b}$) from extended t
  - *10-15% increase for the VBF* process
  - b-tagging performance will benefit from a more granular detector
- We will have a **timing detector at HL-LHC…**
  - *~ 30% improvement in light-jet discrimination* by removing spurious tra reconstructing
  - *~20% increase in effective integrated luminosity for HH*
- **Better background discrimination** from selection optimization with the larg



Mip Timing Detector

Channel	Signal increase (%)
HH→ $bb\gamma\gamma$	22
HH→ $b\bar{b}b\bar{b}$	18

From C. Vernieri

Conclusions

- b-Tagging is a fundamental tools in most physics analyses
- Both ATLAS and CMS reached a significant improvement on their algorithms in Run-II, in particular through the development of Deep Neural Network based algorithms

Properties	CMS	ATLAS
Tracks with large IP	IP2D,IP3D, JP,	JP,JBP
Secondary verices (SV)	SV,SV0,SV1,RNNIP	IVF
Multi-vtx – decay chain	JetFitter	-
Soft-leptons from B-decay	SMT	Soft Lepton Tagger (SL)
MVA combinations	MV1c10, DL1 (+RNN,+SMT)	CSV, cMVA, DeepCSV DeepFlavour

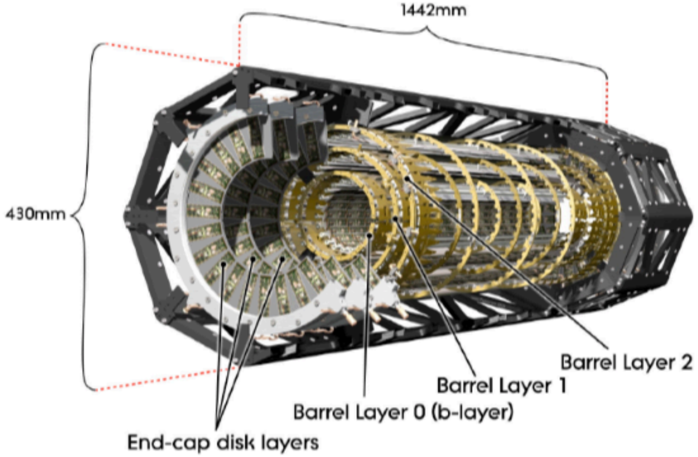
- Not only algorithms, but also the measurements of their performance on data had benefited from new ideas (and of increased sample statistics) in 2016
- More challenge ahead: already working to maintain b-tagging a successful tool in the next decade of data taking

Back-Up

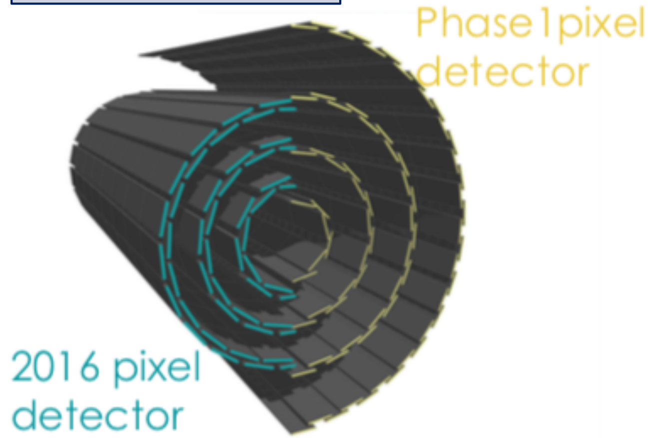
Detector overview

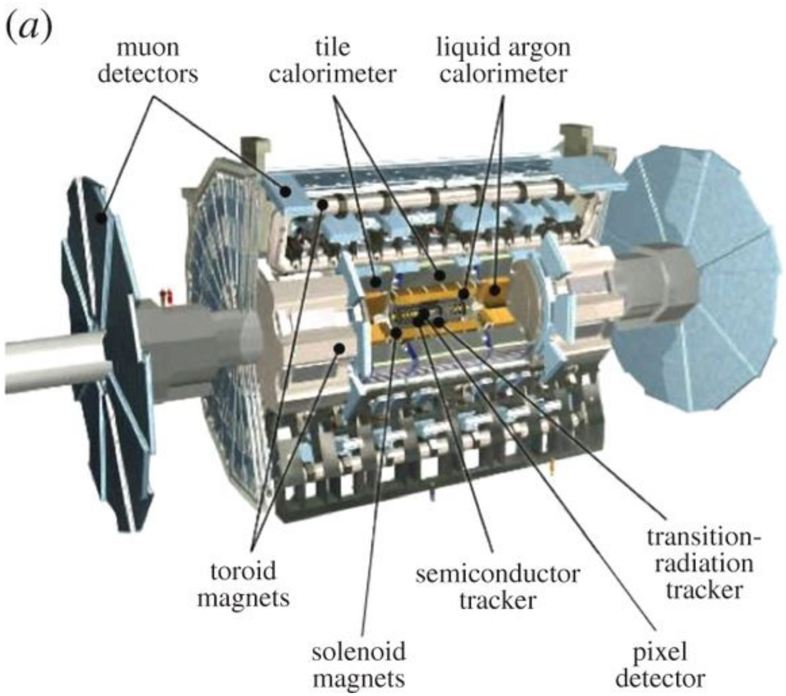
ATLAS

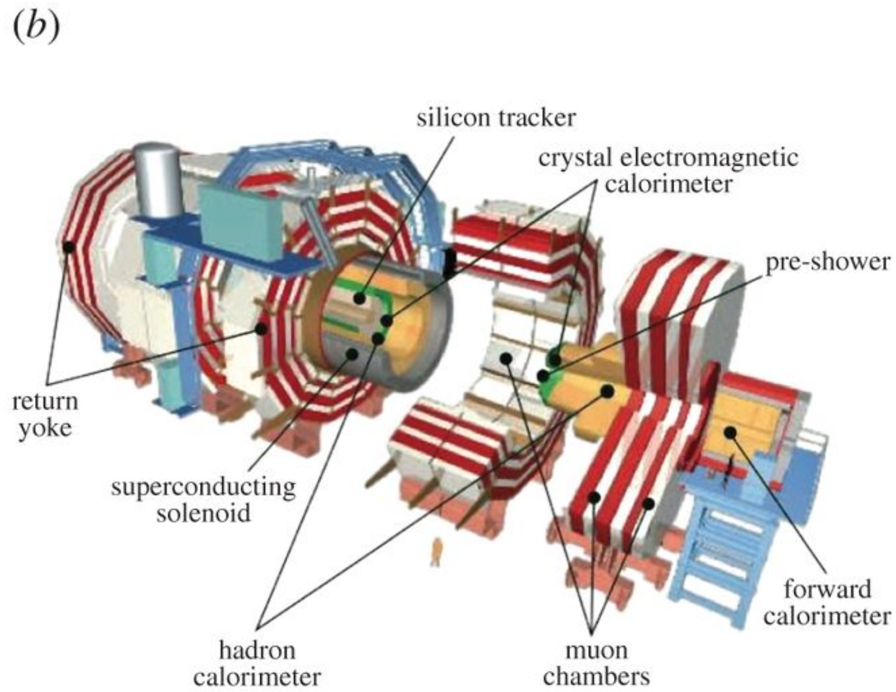
Inner detector & IBL



CMS-TDR-011

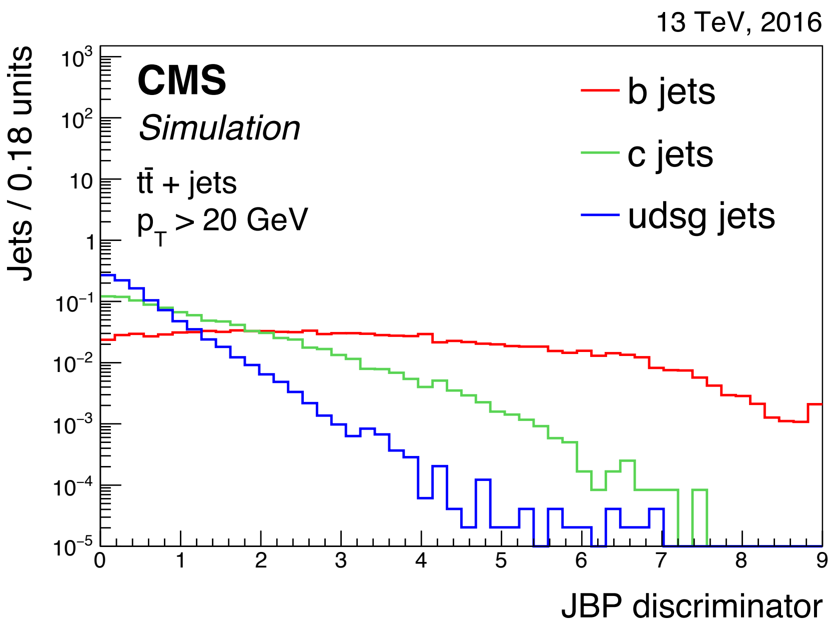


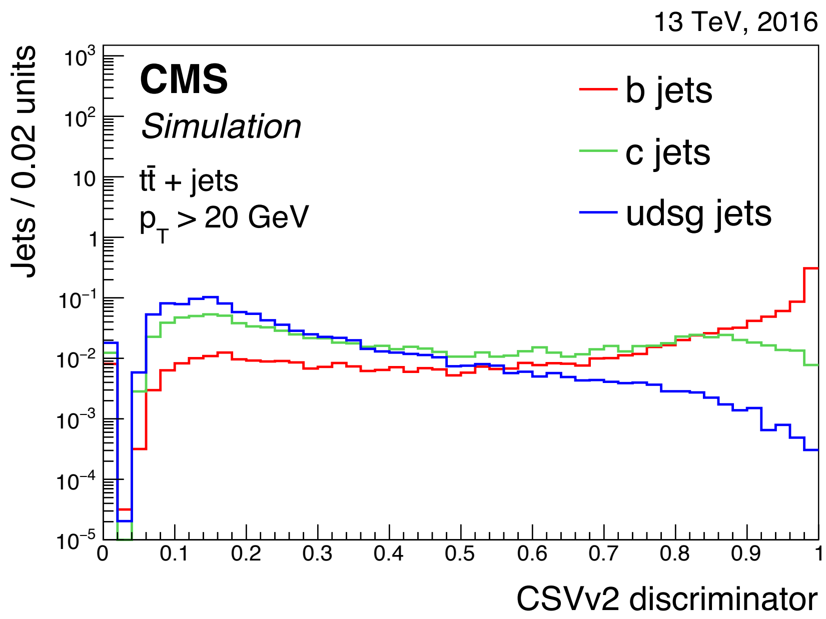




Overview on b-tagging algorithm in CMS

- JP and JBP
  - Likelihood based on P(for jet to be originated by primary vertex)
- CSV and CSVv2
  - Combine displaced tracks with secondary vertices in multilayer perceptrons
- DeepCSV (more details in the next slides)
  - DNN Multiclassifier: same inputs as CSVv2 with a simple extension to use more charged particle tracks
- DeepFlavour (more details in the next slides)





Physics case for H→cc

Importance of searches for H→cc

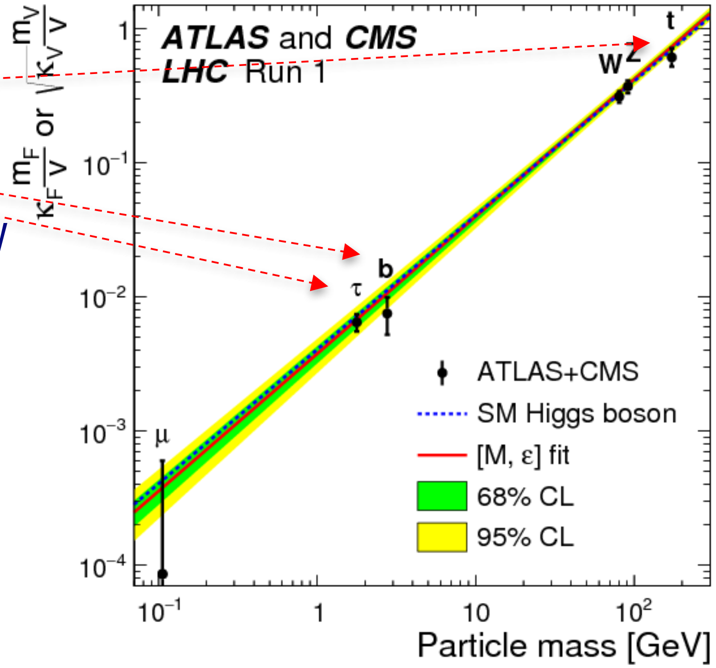
- From LHC Run-I ➔ strong indication that the Yukawa couplings scale with fermion masses

$$L_Y = f_l \overline{\chi}_L \phi l_R + f_u \overline{q}_L \tilde{\phi} u_R + f_d \overline{q}_L \phi d_R + \text{h.c.}$$

$$\phi = \begin{pmatrix} 0 \\ v+h \end{pmatrix} \rightarrow \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix}$$

$$L_Y = \frac{v f_l}{\sqrt{2}} \left(\overline{l}_L l_R + \overline{l}_R l_L \right) + \frac{v f_u}{\sqrt{2}} \left(\overline{u}_L u_R + \overline{u}_R u_L \right) + \frac{v f_l}{\sqrt{2}} \left(\overline{d}_L d_R + \overline{d}_R d_L \right)$$

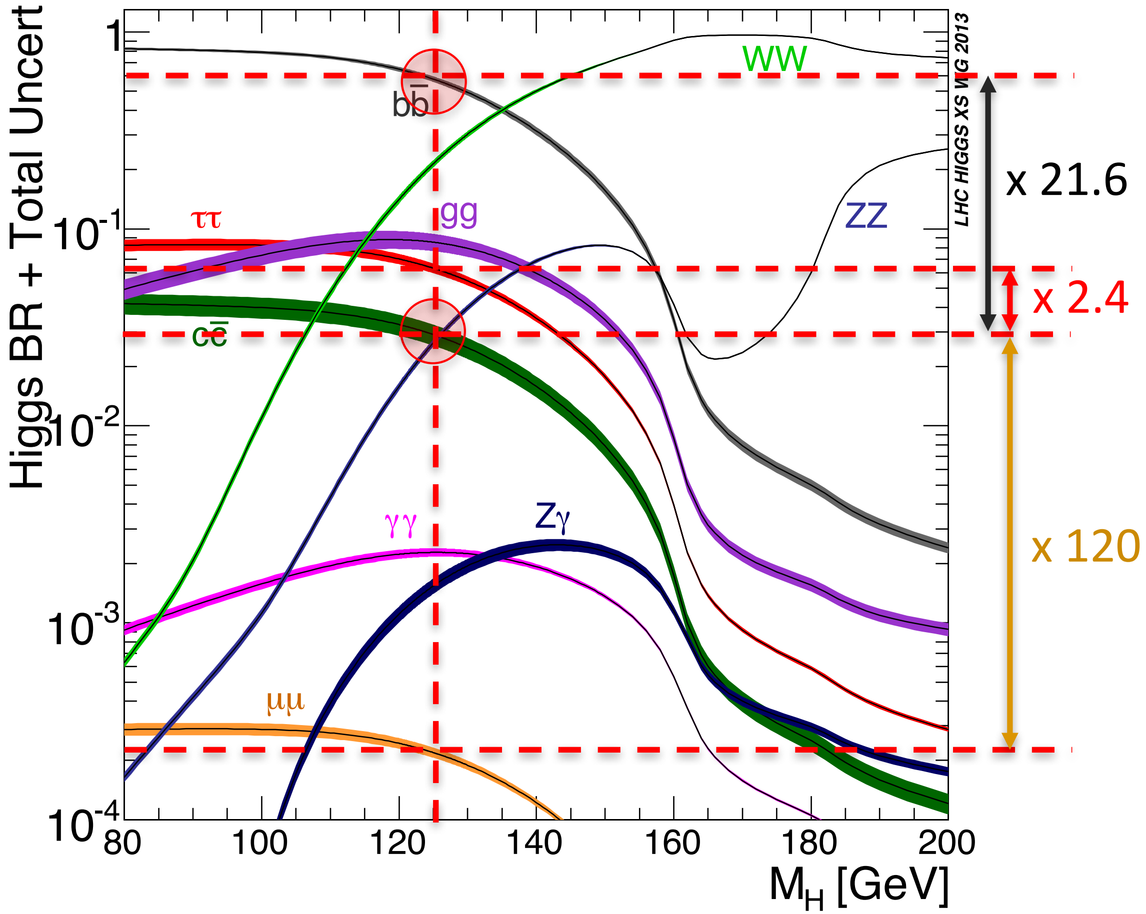
$$f_i = \frac{m_i}{v}\sqrt{2}$$



- While Higgs couplings to the vector bosons are precisely established, those relative to the fermions are affected by larger uncertainties
- Direct observation of H→ττ first and ttH more recently, has been proven.
- July 2018: H→bb decay has been observed by ATLAS and CMS

Physics case for H→cc

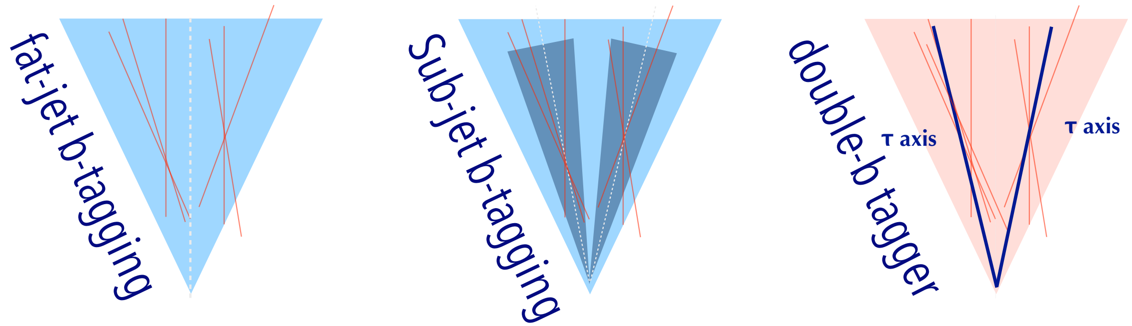
- **H→cc to test the Standard Model**
  - Only H→cc could provide some prospects to constrain the Higgs coupling to the second generation fermions for the end of HL-LHC



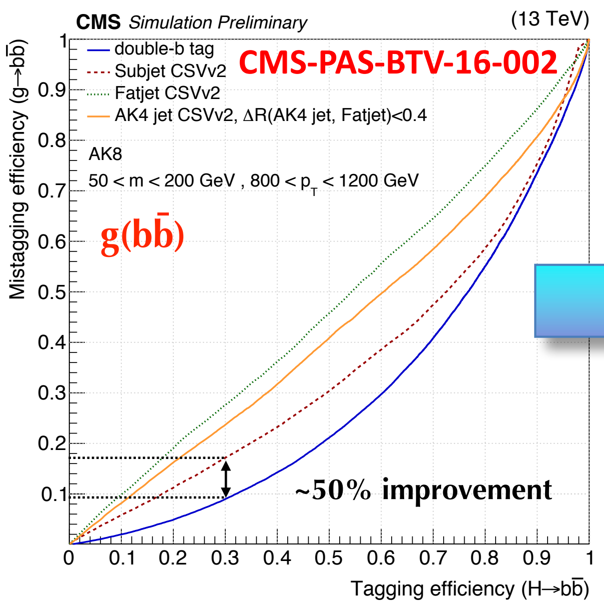
- For Higgs mass of ~125 GeV, the expected BR(H→cc) is 2.67 %
- BR(H→μμ) = 0.02 %
  BR(H→ττ) = 6.37 %
  BR(H→bb) = 57.7 %
- H → cc decays constitute the largest part of the SM prediction for fermionic H decays for which we have no experimental evidence

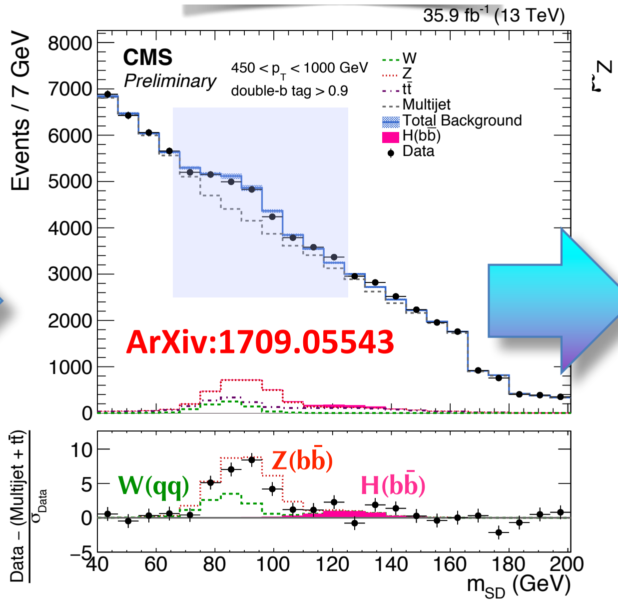
Boosted topologies taggers

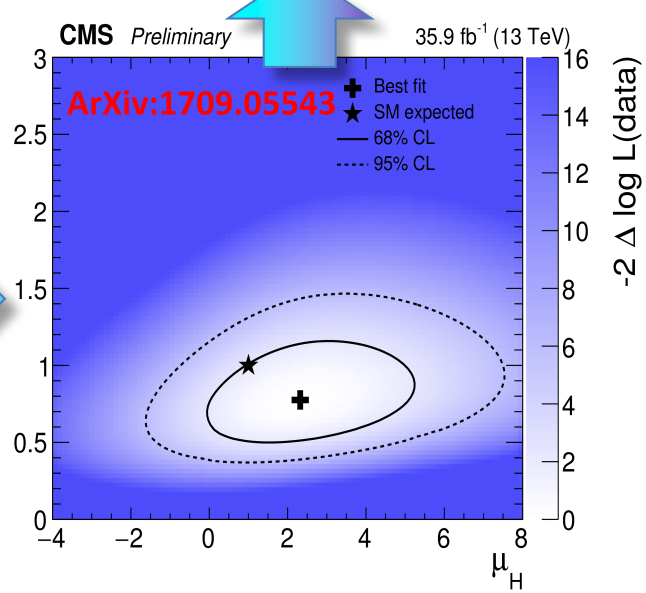
- **Double-b:**
  - Development of a discriminator based on the exploitation of the **"fat-jet" sub-structure** and **correlation** between two sub-jets coming from a **boosted resonance** (impressive results already with the double-b)



$$S^{Z}_{exp.}(S^{Z}_{obs.}) = 5.8\sigma\ (5.1\sigma)$$
$$S^{H}_{exp.}(S^{H}_{obs.}) = 0.7\sigma\ (1.5\sigma)$$



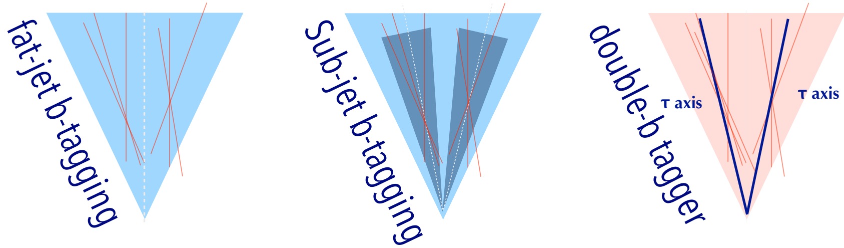




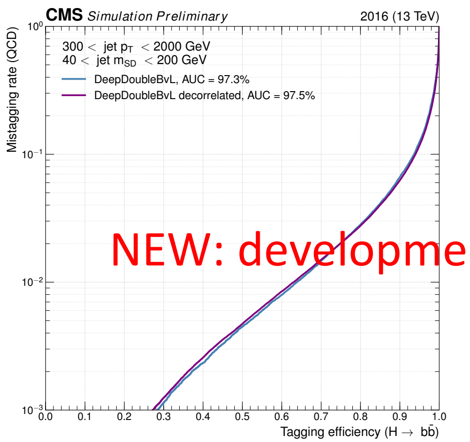
Boosted topologies taggers

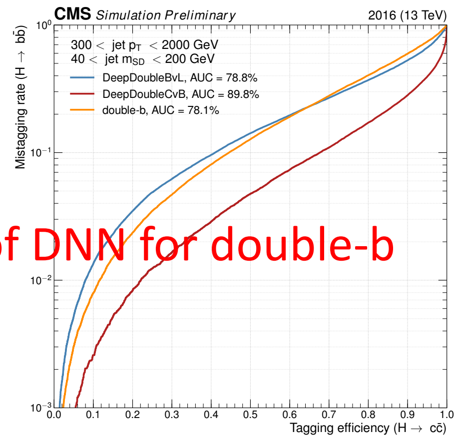
▪ Double-b taggers:

- Development of a discriminator based on the exploitation of the **"fat-jet " sub-structure** and **correlation** between two sub-jets coming from a **boosted resonance** (impressive results already with the double-b)

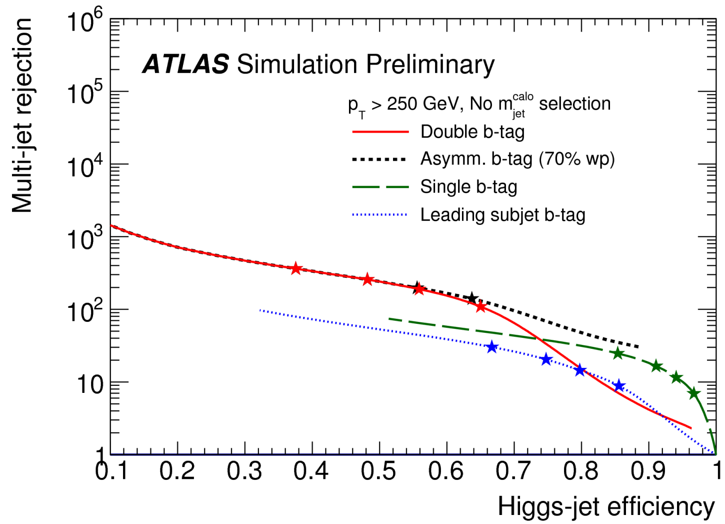


- CMS
  - Particle flow AK8 jets (fatjets)
  - Soft drop declustering to resolve jet substructure
  - b-tagging applied on tracks in fatjets and/or subjets
- ATLAS
  - Calorimeter large-R (R=1.) jets
  - Trimming algorithm to discardsofter components of the jet
  - b-tagging applied on ghost-associated track jets (R=0.2)





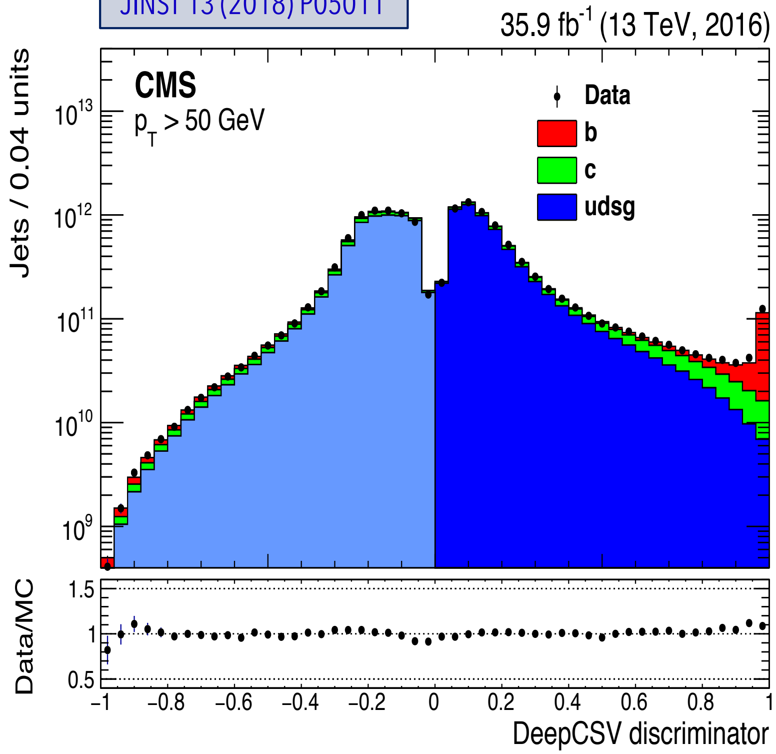
NEW: development of DNN for double-b



Performance in Data - CMS

- Distributions of the DeepCSV (left) and the cMVAv2 (right) discriminators for jets in an inclusive multijet sample.
- Output of the negative DeepCSV tagger is shown with a negative sign. Normalization vs data

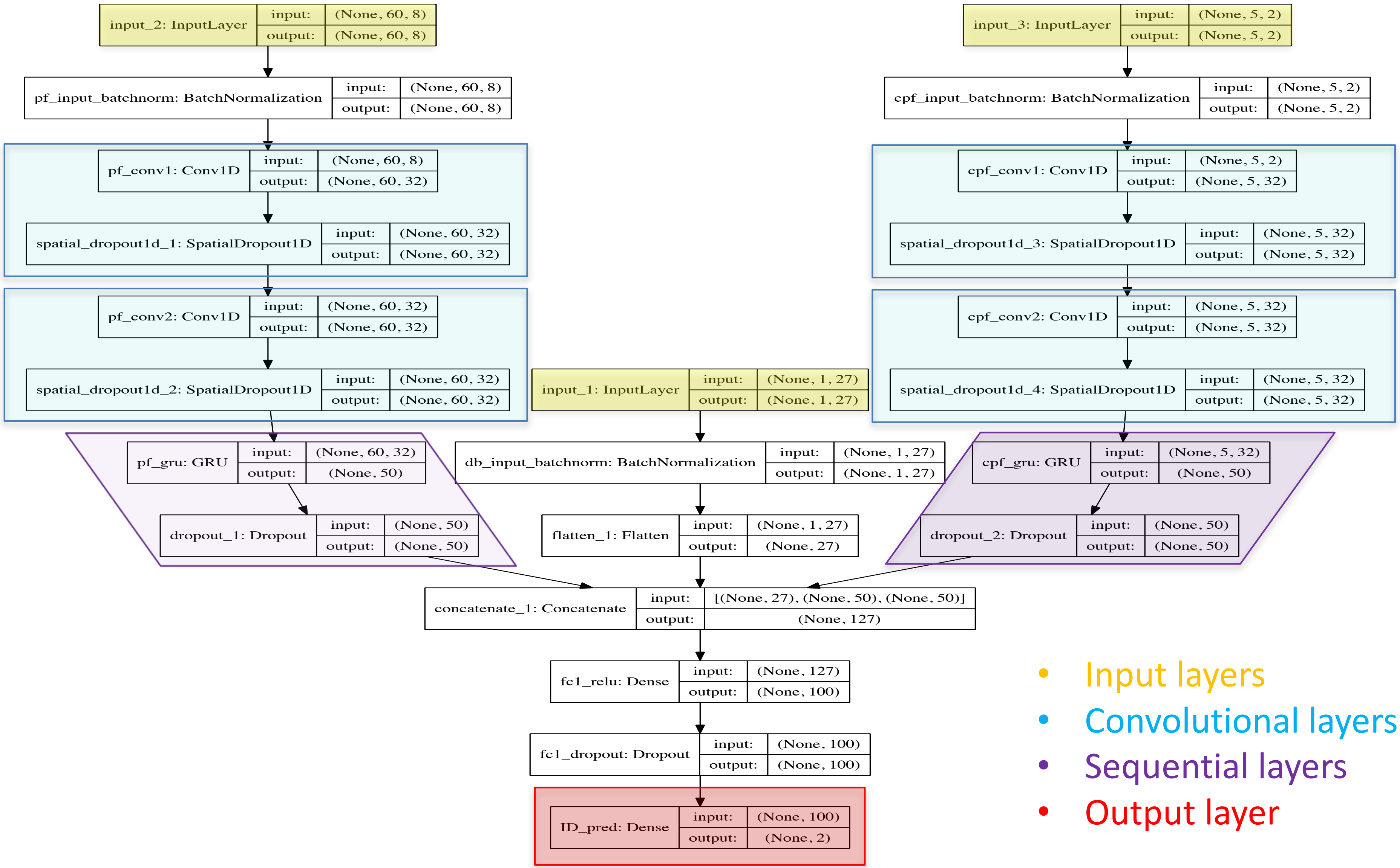
JINST 13 (2018) P05011



Main techniques to perform measurement in data:

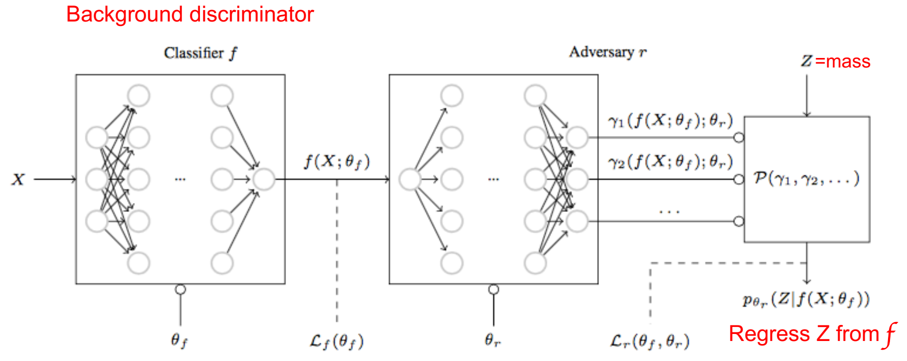
- **mu+jets:** Measured b-tagging efficiency in multijet events with a muon, based on the combination of the results from different measurements, obtained using the PtRel, the LT and the System8 methods.
- **Kin:** Method for the measurement of the b-tagging efficiency in ttbar events in the dileptonic channel, based on a template fit to an MVA discriminator combining kinematic variables.
- **TnP:** Method for the measurement of the b-tagging efficiency in ttbar events in the semileptonic channel. The b-tagging efficiency is measured with a tag and probe method (TnP). As a tagging requirement, the tagger requirement is applied to either the b-jet on the hadronic or leptonic side, while the b-jet from the other side is used as probe.
- **TagCount:** Method for the measurement of the b-tagging efficiency in ttbar events in the dileptonic channel. The b-tagging efficiency is obtained by counting the number of events with two b-tagged jets in the selected sample of events.
- **IterativeFit:** Method for the measurement of the b-tagging efficiency in ttbar events in the dileptonic channel. This method is based on the calibration of the full b-tagging discriminator shape.

▪ Double-c: Neural Network Architecture



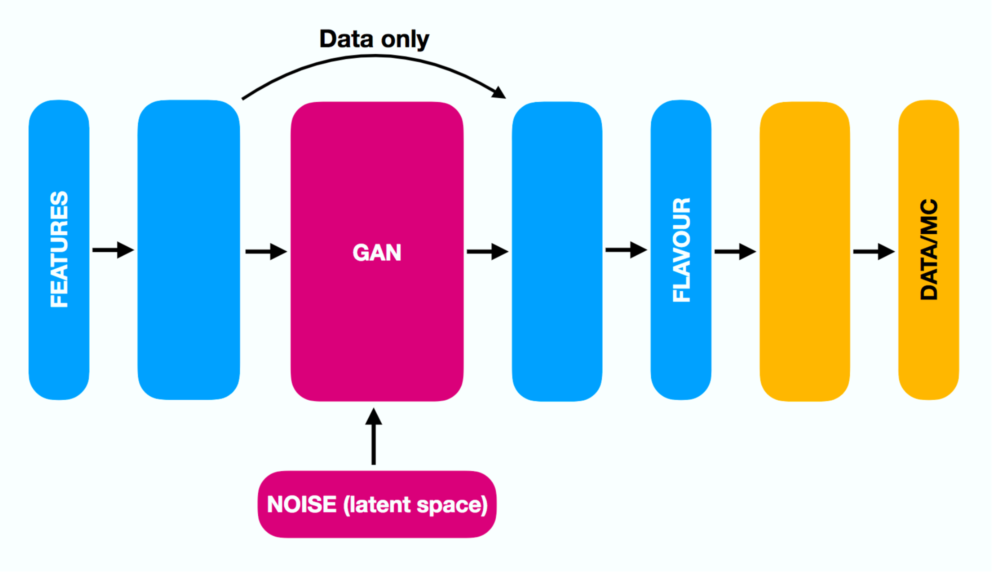
- Input layers
- Convolutional layers
- Sequential layers
- Output layer

- **Mass sculpting**
  - Use classifier to increase signal purity, but want to avoid artificial bump in background
  - Many features depend on mass
  - Enforce independence of classifier on mass



$$\hat{\theta}_f, \hat{\theta}_r = \arg\min_{\theta_f} \max_{\theta_r} \mathcal{L}_f(\theta_f) - \mathcal{L}_r(\theta_f, \theta_r)$$

- **Data/MC scale factors**
  - SF: $f : \mathbb{R}^n \to \mathbb{R}$
  - Basically what a DNN does!
  - Future possible extension of the reweighting method

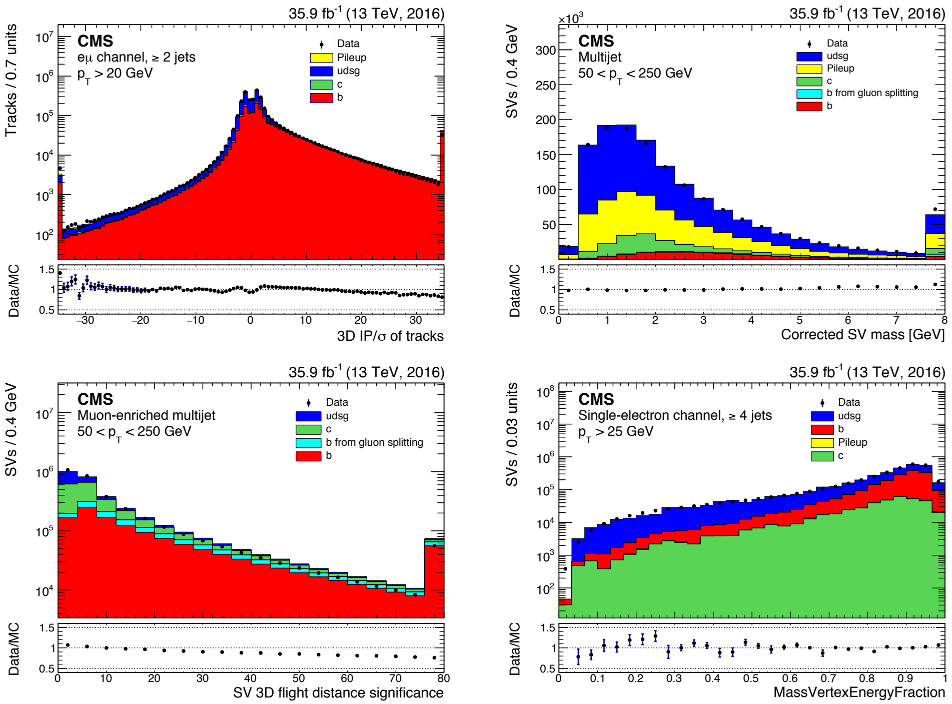


Performance in Data - CMS

▪ DeepCSV commissioning: event topologies:

- **Inclusive Multijet sample**: Online: 1 AK4 hets with pt>40 GeV. Offline selection: 50<pT<250. Sample dominated by light jets with relatively large contribution from pile-up jets
- **Muon enriched jet sample**: Online: 2 AK4 hets with pt>40 GeV, with at least one containing a muon with pT>5GeV. Offline selection: 50<pT<250 + at least a muon. Sample enriched by heavy-flavour jets
- **Di-Lepton tt**: Online: isolated electron and a muon. Offline: at least two AK4 with pT>20GeV + electron and muon with pT>25 GeV. Sample dominated by b-jets from the top quark decays
- **Single-Lepton tt**: Online: One isolated lepton. Offline: at least two AK4 with pT>20GeV + electron and muon with pT>25 GeV. Sample dominated by b-jets from the top quark decays with higher fraction in c-jets

CMS DP-2018/033 ***All the MC distribution are normalised to data!***



CMS DP-2018/033