# Flavour tagging at the LHC experiments

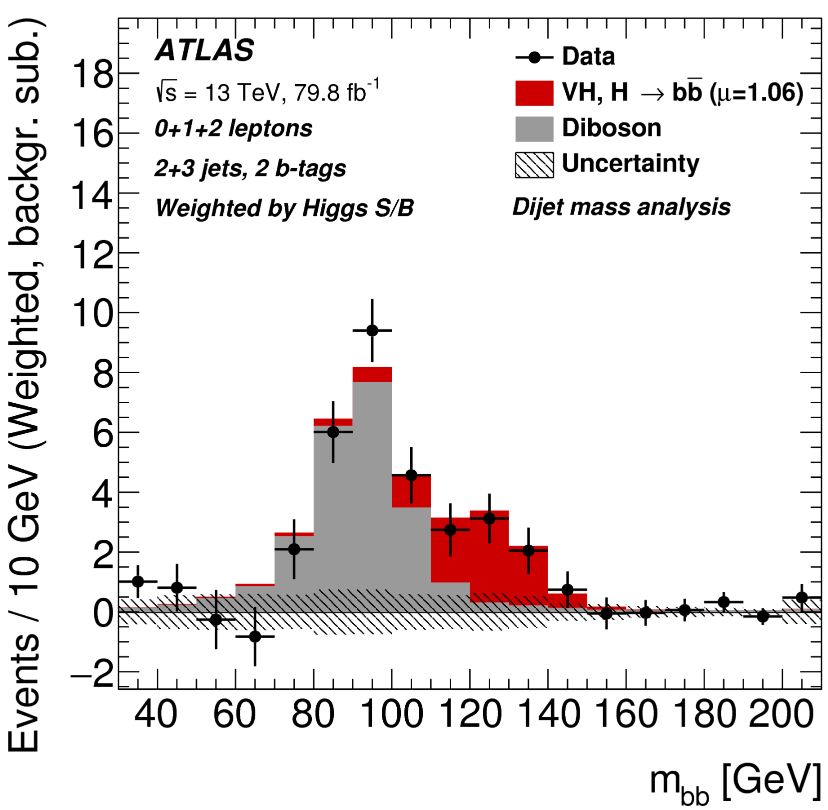
Riccardo Manzoni - ETH Zürich
on behalf of ATLAS, CMS and LHCb Collaborations

20/5/2019 LHCP, Puebla, Mexico

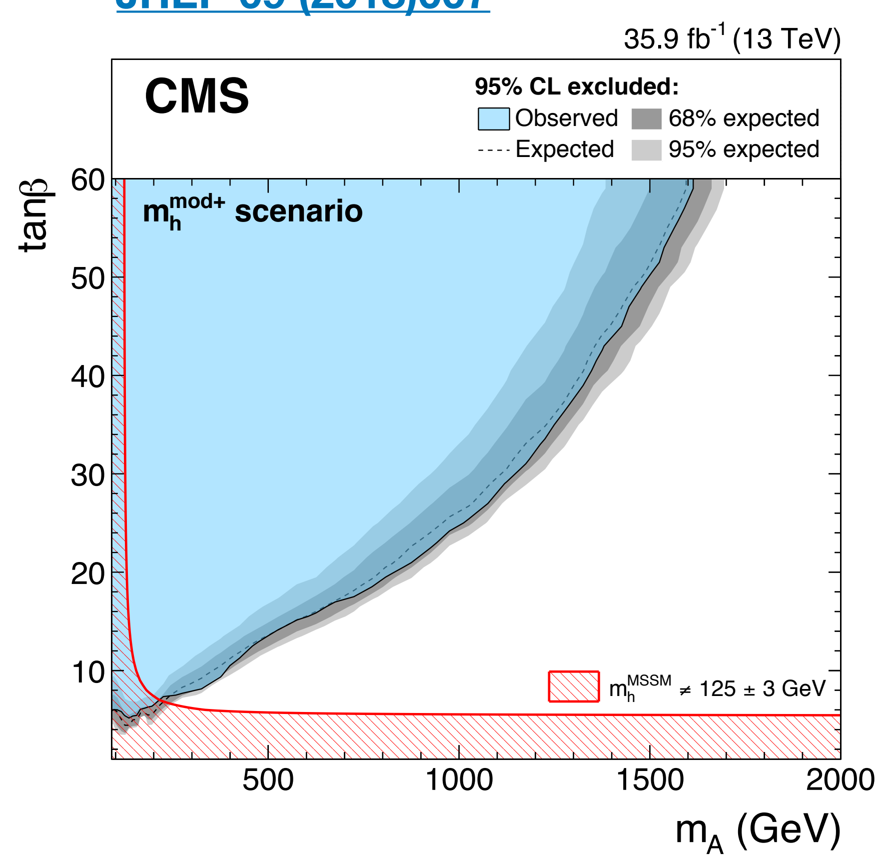
# Flavour tagging in physics analyses

- **Higgs physics:** H→cc, H→bb, ttH, H→ττ, HH
- **top quark** BR(t→bW) ~100%
- **BSM:** new particles coupling strongly with t / b / τ

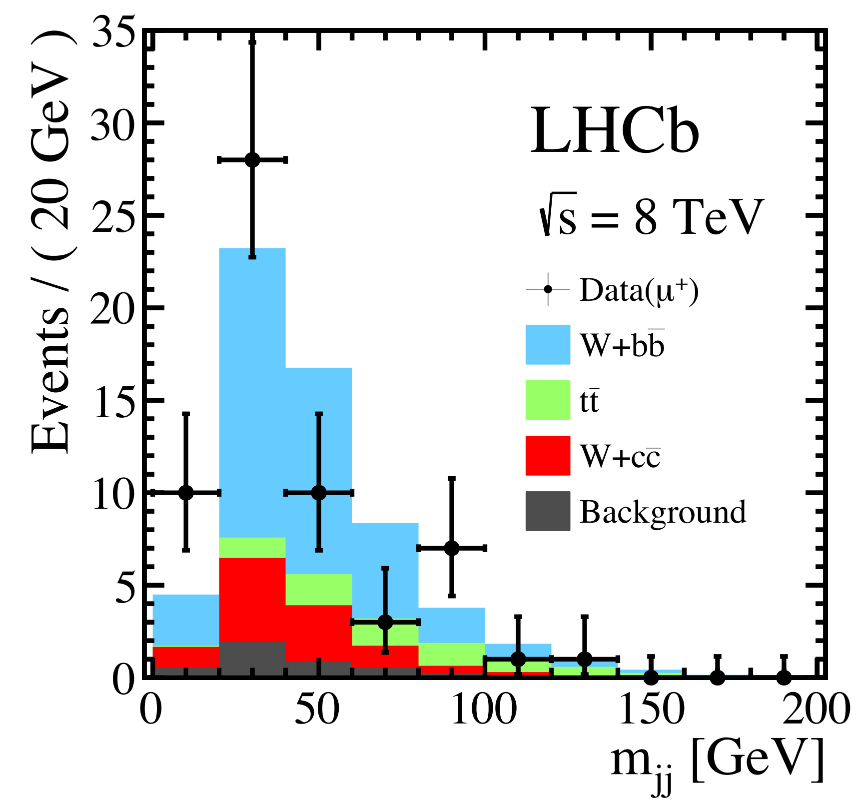
Phys. Lett. B 786 (2018) 59




JHEP 09 (2018)007

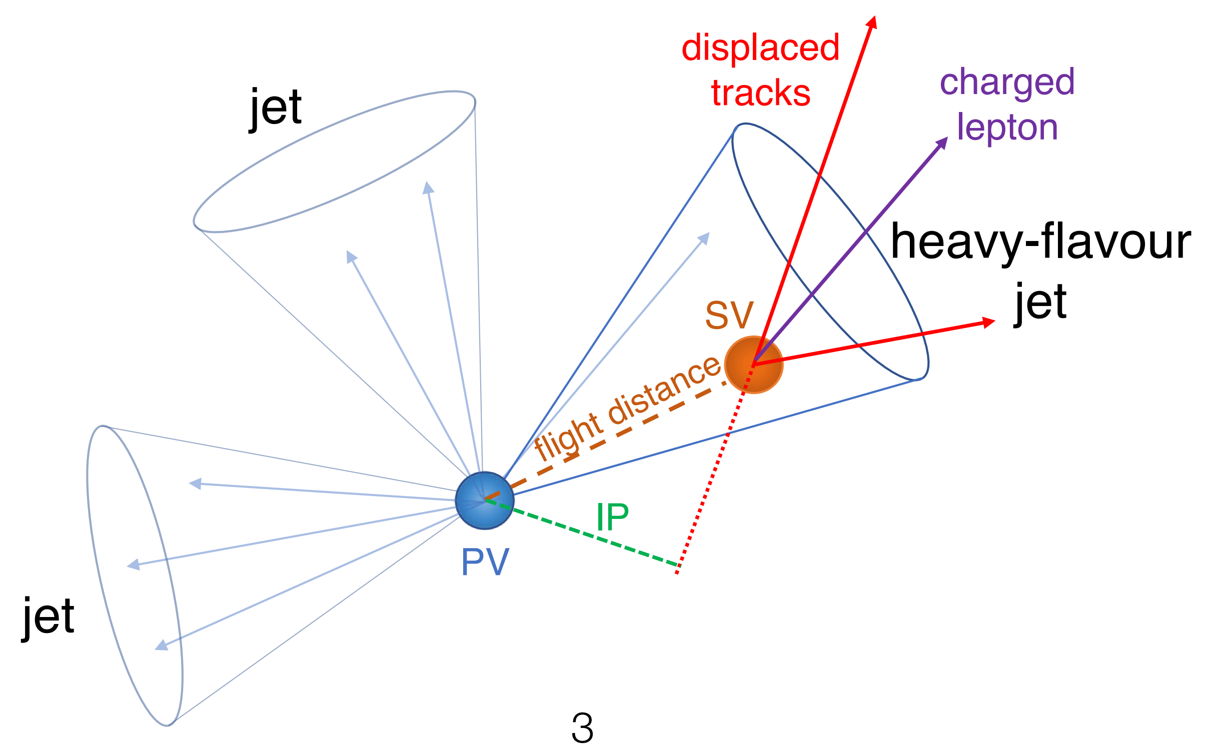



PHYS. LETT. B767 (2017) 110



# Heavy flavour jets

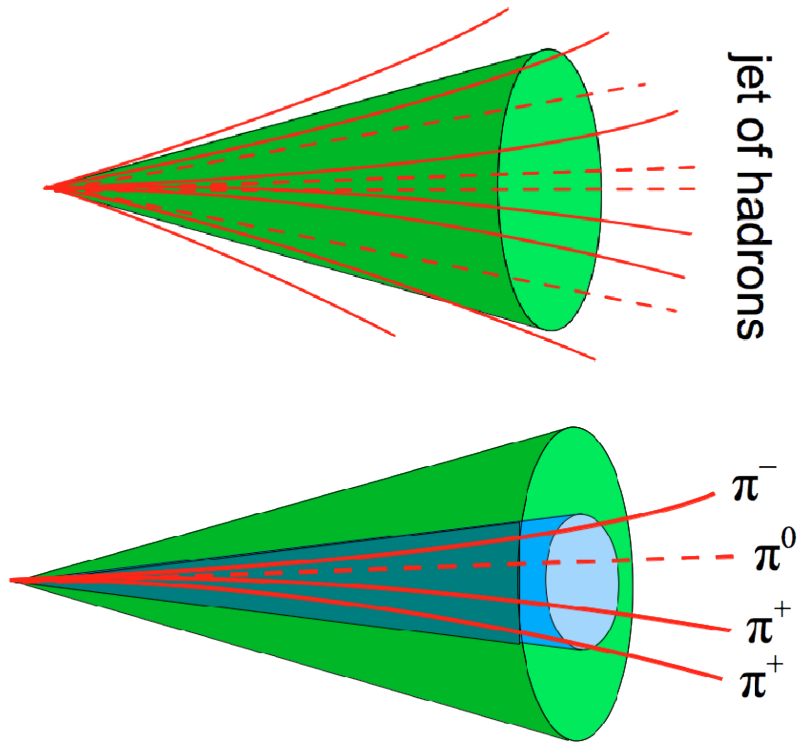
- **quarks hadronise and fragment**
- **b- and c-quarks vs gluon and uds-jets/hadrons**
  - larger mass and larger fraction of initial quark momentum carried by the corresponding hadron
  - longer lifetime → displaced decays for b/c hadron
  - 20 (10) % of decays to leptons for b (c) hadrons



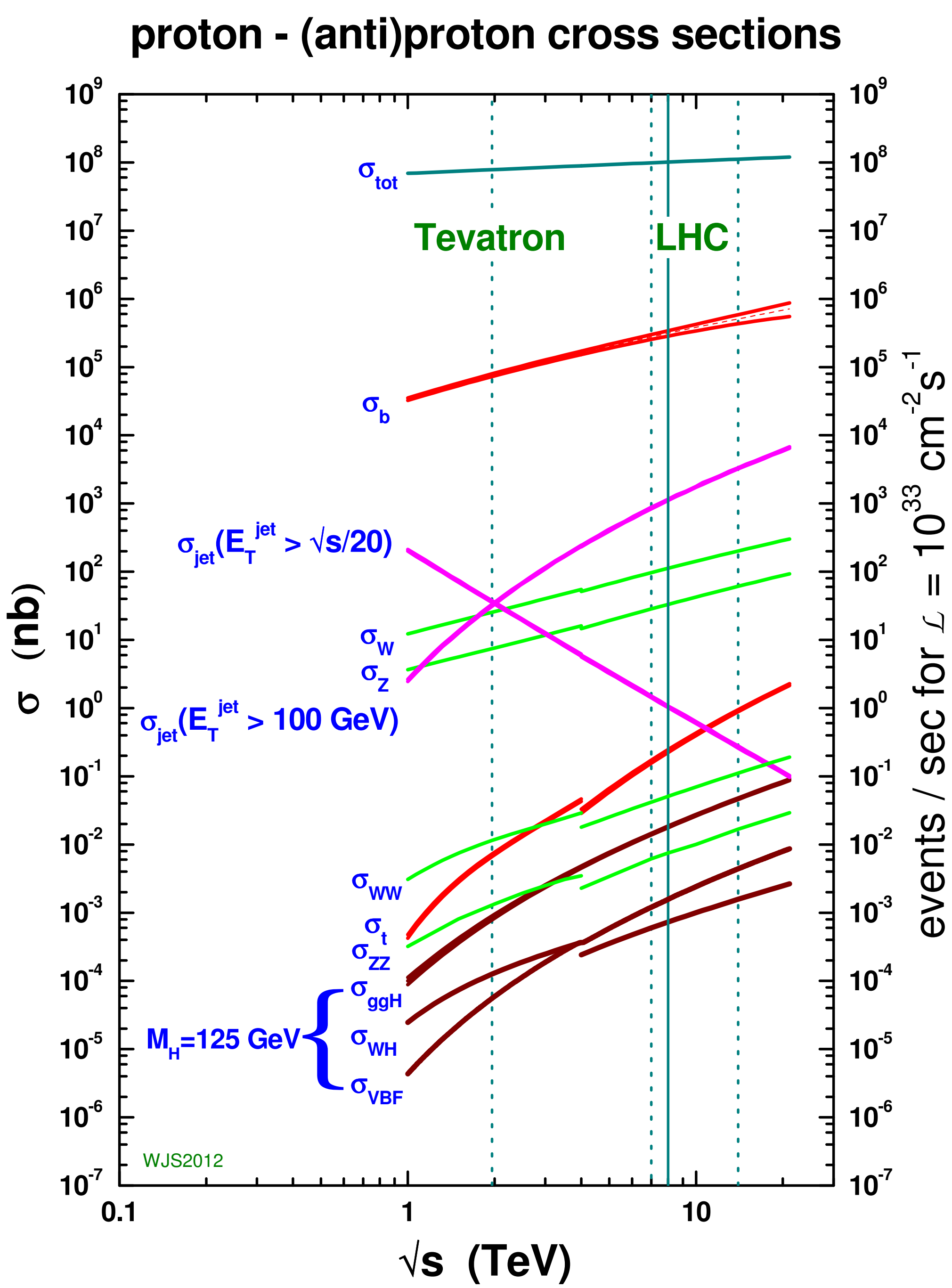
# Tau leptons

- **τ is the only lepton heavy enough to decay into hadrons (BR = 65%)**
  - $m_\tau$ = 1.78 GeV, lifetime $2.91 \cdot 10^{-13}$ s, $c\tau$ = 90 μm
- **τ vs jets**
  - lower multiplicity: mostly 1 or 3 charged plus 0 to $2\pi^0$
  - lifetime longer than for light hadrons (but shorted than b- an c-hadrons)
  - τ non-coloured → isolated
- **electrons and muon can 'fake' $\tau_h$ too**

| Decay mode | Meson resonance | $\mathcal{B}$ [%] |
|---|---|---|
| $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ | | 17.8 |
| $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$ | | 17.4 |
| $\tau^- \to h^- \nu_\tau$ | | 11.5 |
| $\tau^- \to h^- \pi^0 \nu_\tau$ | $\rho(770)$ | 26.0 |
| $\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$ | $a_1(1260)$ | 10.8 |
| $\tau^- \to h^- h^+ h^- \nu_\tau$ | $a_1(1260)$ | 9.8 |
| $\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$ | | 4.8 |
| Other modes with hadrons | | 1.8 |
| All modes containing hadrons | | 64.8 |



# The challenge




- **overwhelming production of light-jets**

# The approach

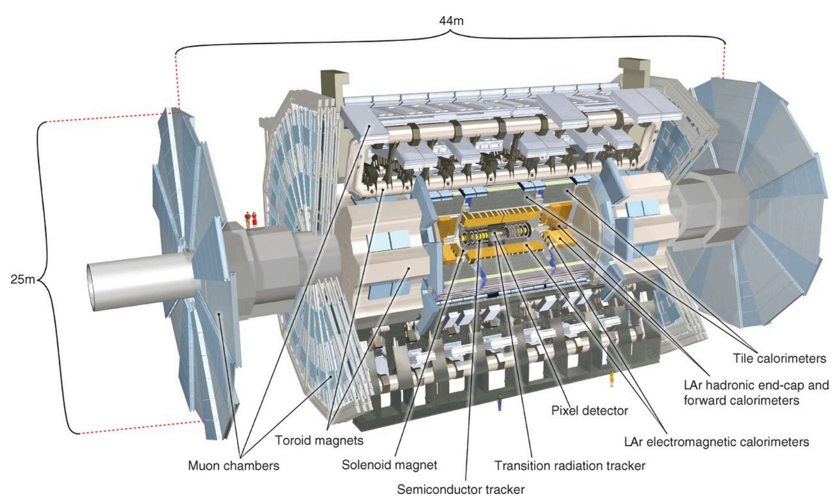

- Machine Learning
- **… and domain knowledge & ingenuity**

# Outline

- **Jet flavour tagging in ATLAS, CMS and LHCb**
  - algorithms, calibrations and performance
- **$\tau_h$ reconstruction in ATLAS and CMS**
  - LHCb look for exclusive $\tau$ decays

# Jet flavour tagging

# ATLAS algorithms




**inner detector tracks**

**EM topological clusters**
**Variable Radius jets**

**$p_T$ dependent jet-trk association**

**Low level algorithms**
- 2D and 3D impact parameter significance
- secondary and tertiary vertices
- JetFitter
- sof-lepton

**High level taggers**
combining info from low level algos
- MV2 (BDT)
- DL1 (DNN)

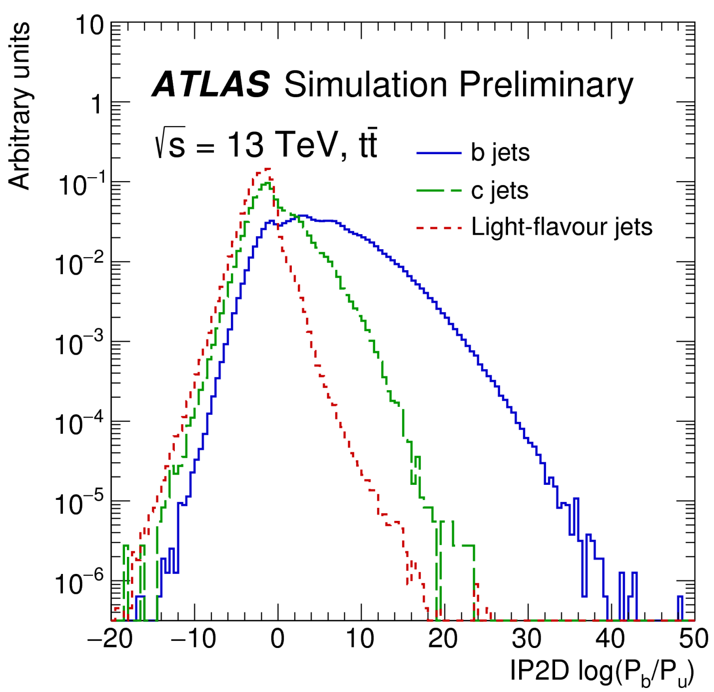
**identify b-jet vs rest (light/charm)**
**separate tagger for charm**

# Low-level algorithms
# IP taggers

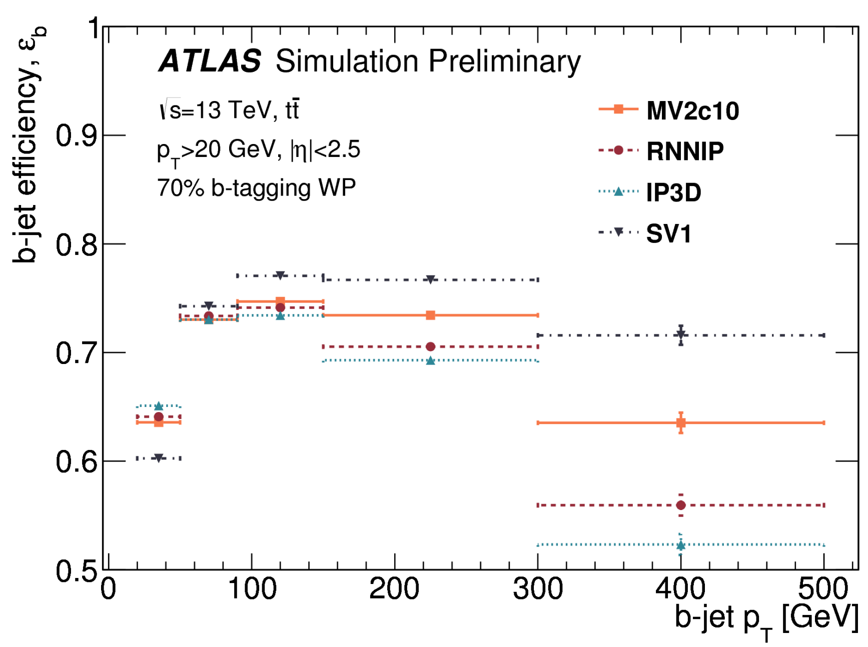
- **IPTag: 2D and 3D track impact parameter significance likelihood**
  $d_0/\sigma(d_0)$ and $z_0 \cdot \sin\theta/\sigma(z_0 \cdot \sin\theta)$

- **RDNN:** Recurrent Deep Neural Network

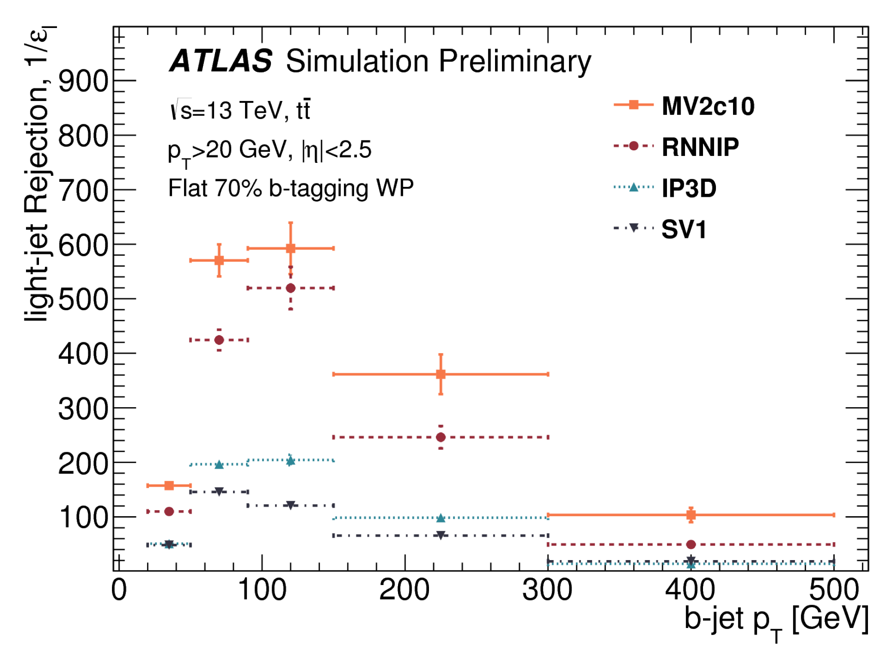
- using same tracks/inputs:
  $p_T$>1 GeV, $|d_0|$< 1mm, $|z_0 \sin\theta|$< 1.5mm, 14 track quality categories




**2D IPTag likelihood**




**b-jet efficiency**




**light-jet rejection**

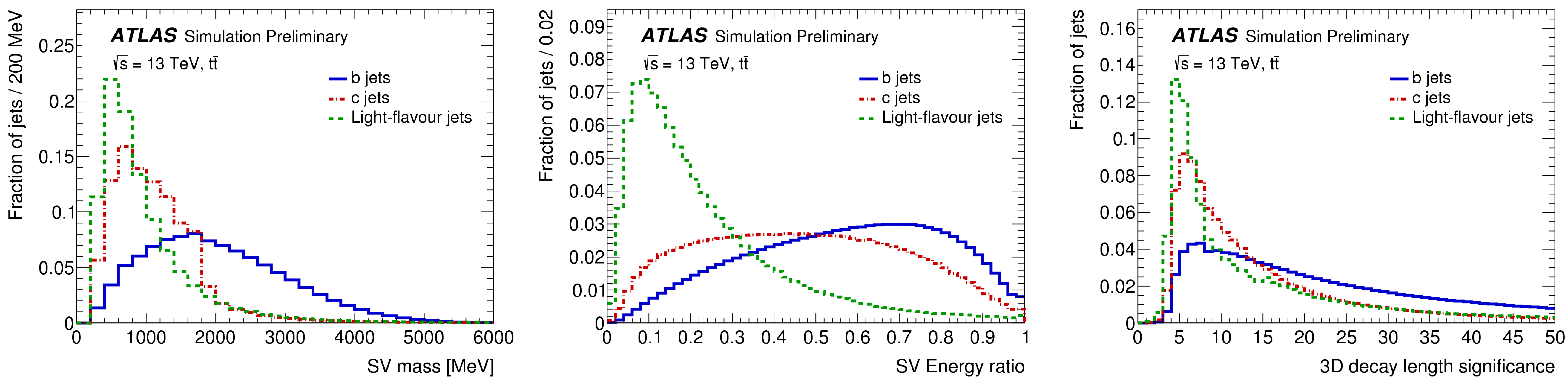
ATL-PHYS-PUB-2016-012



ATL-PHYS-PUB-2017-003

# Low-level algorithms Secondary Vertex (SV1)

- **single-secondary-vertex-finding (SSVF) algorithm to identify jets with SV consistent with a b-hadron decay**
  - form 2-track vertices using all tracks in a jet and iteratively merge until *one* secondary vertex (SV) remains
  - **$\chi^2$, SV mass, SV energy fraction, decay length** information used later in high-level taggers (after removing $K_s$, $\Lambda_0$, $\gamma$ conversions)

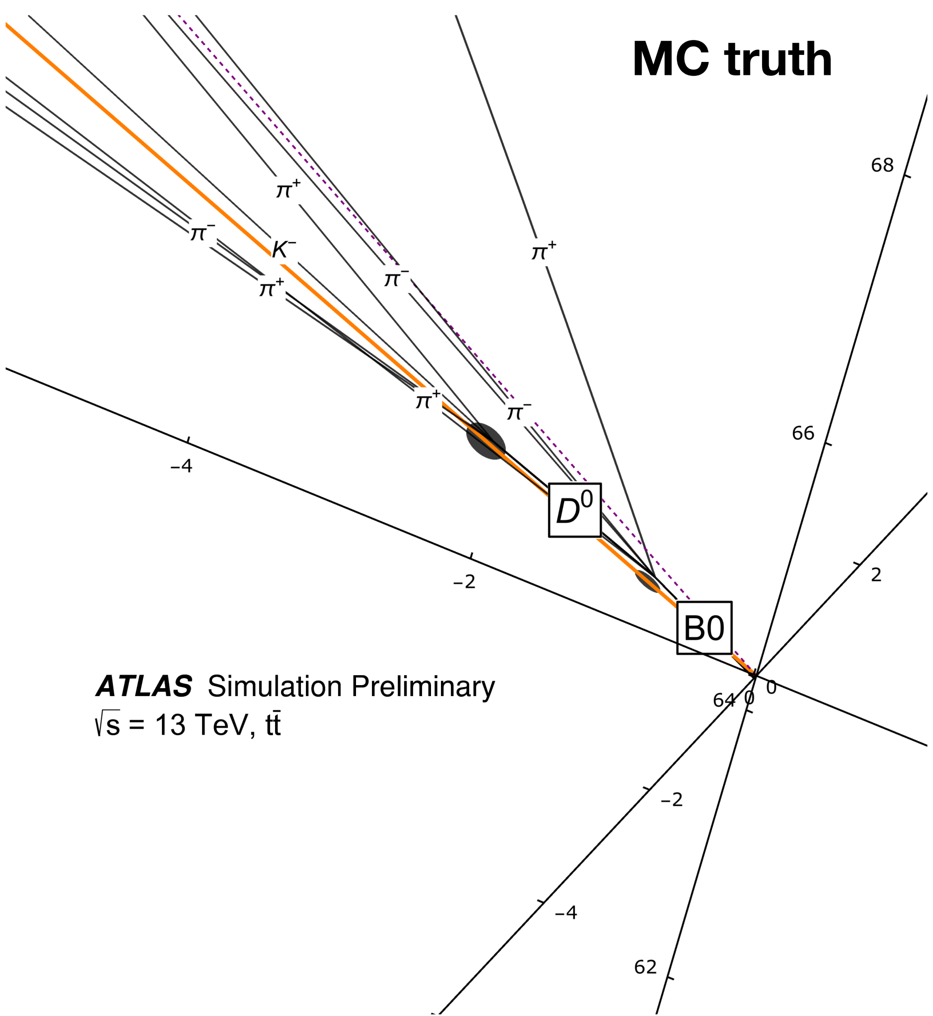


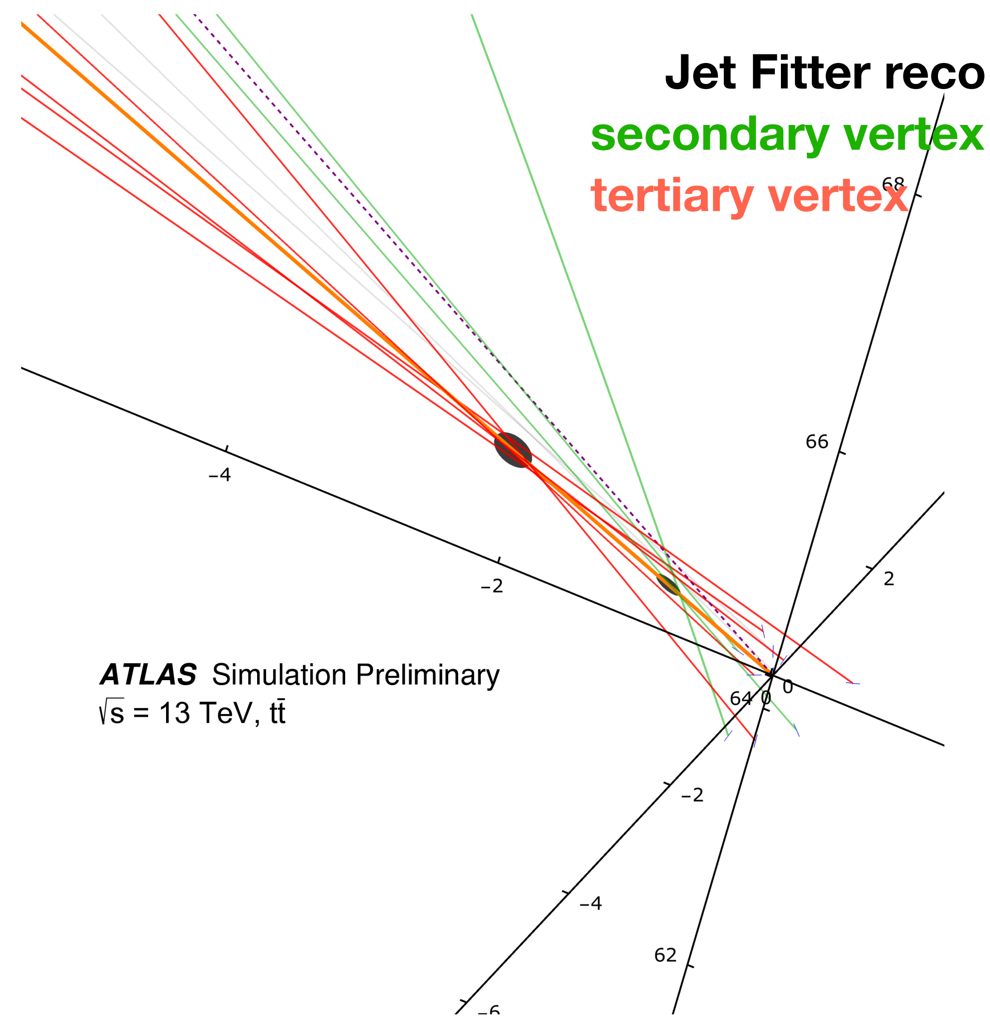

ATL-PHYS-PUB-2017-011

# Low-level algorithms
# Jet Fitter

ATL-PHYS-PUB-2018-025

- **fit cascade vertices separately**
    - secondary/tertiary vertex $\chi^2$, mass, vertex energy fraction, decay length used later in high level algorithms



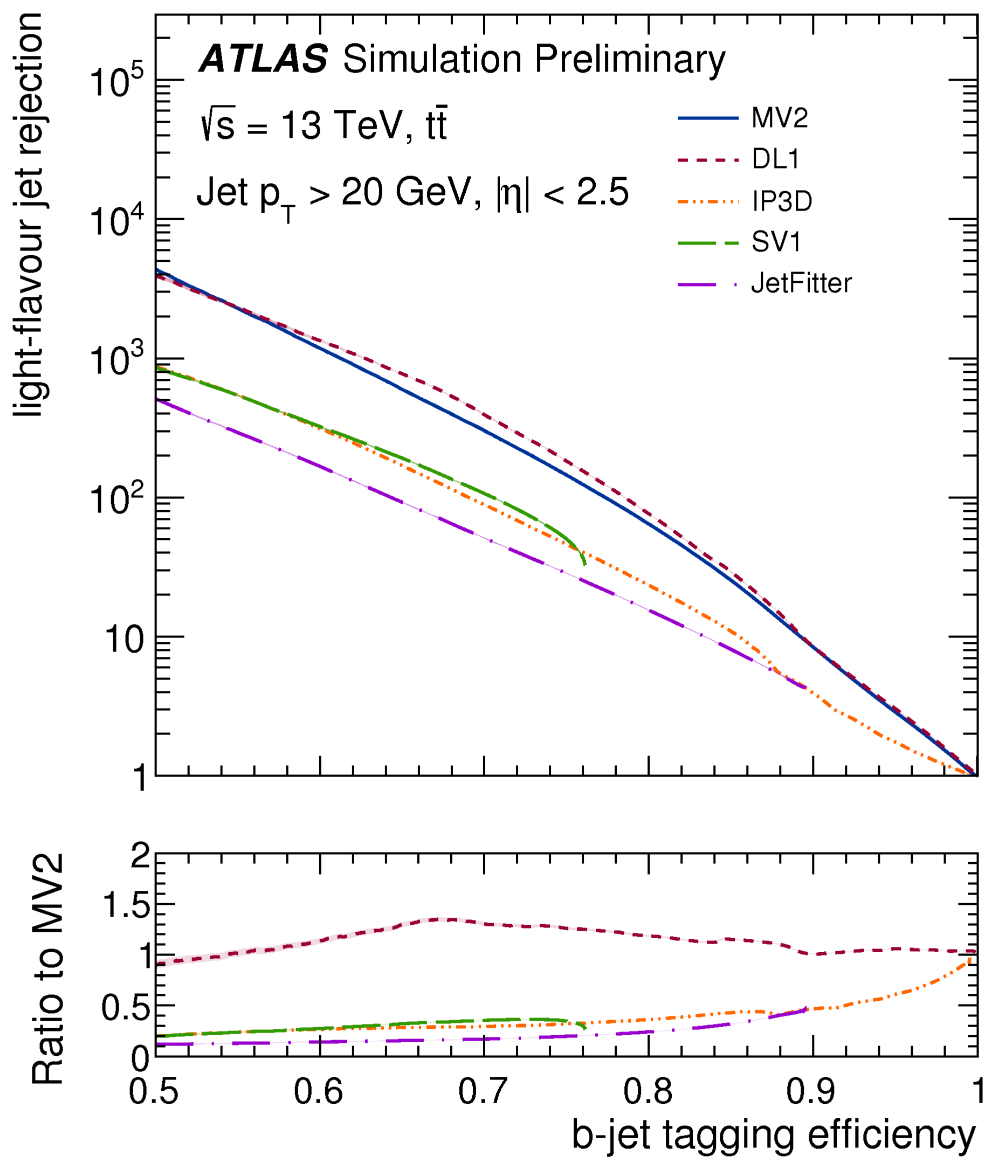



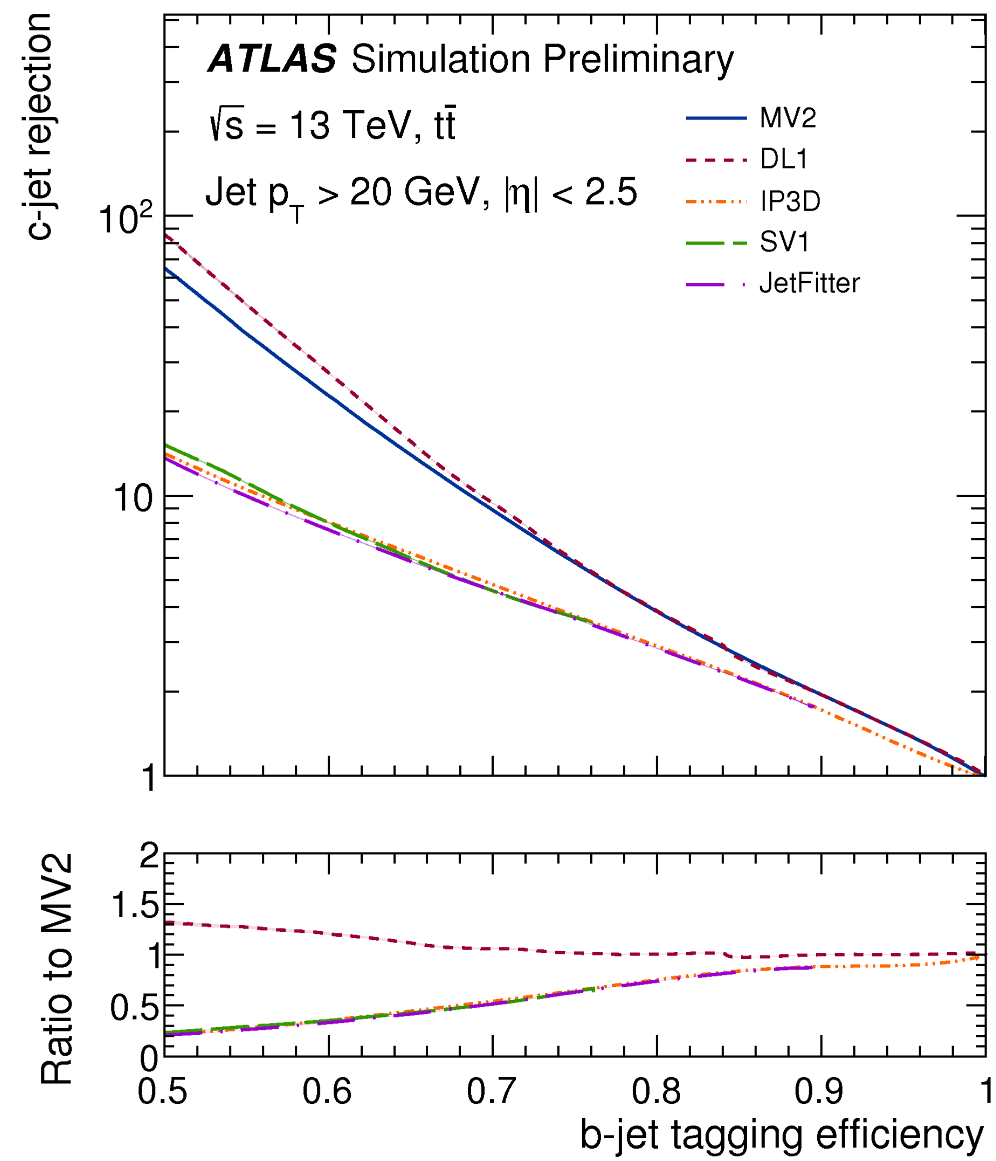
ATLAS EXPERIMENT

FTAG-2019-002

# High-level taggers MV2 and DL1

- **machine learning based discriminators: MV2 (BDT), DL1 (NN)**
  - feed on low-level tagger inputs shown in previous slides



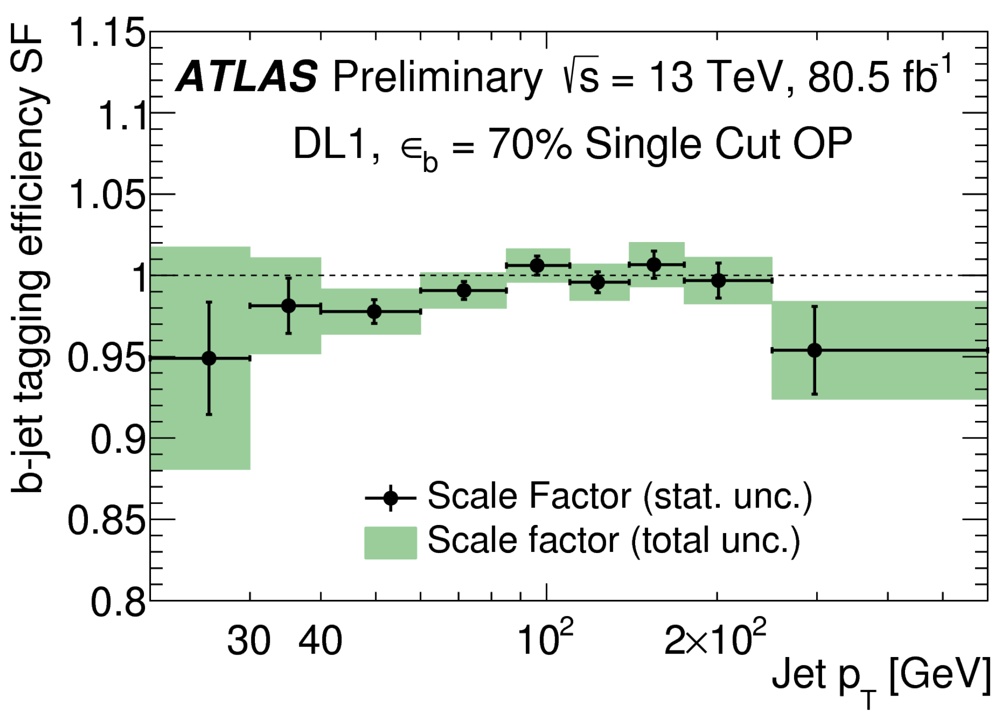



# Calibration in data

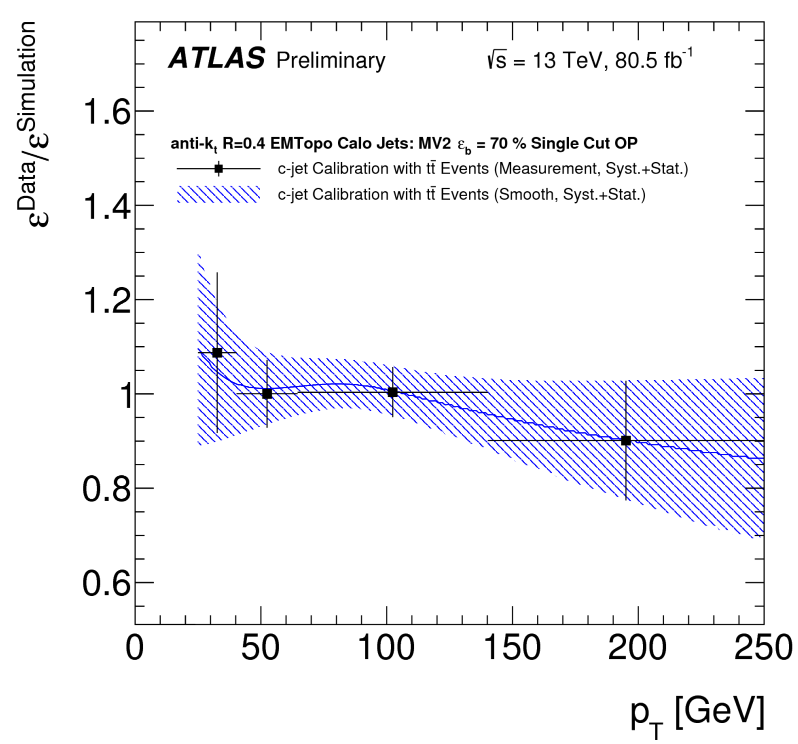
- **fits to control regions enriched in b- (tt→eμ + bb), c- (tt→ℓ cq + bb) and light-jets (Z+jets)**

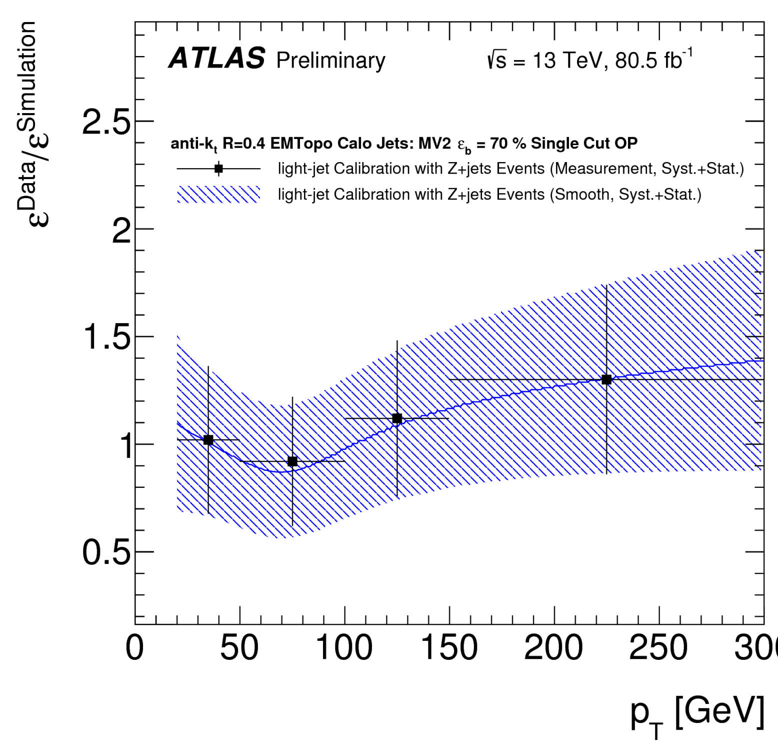
- improved uncertainties for light-jet SF

FTAG-2019-002

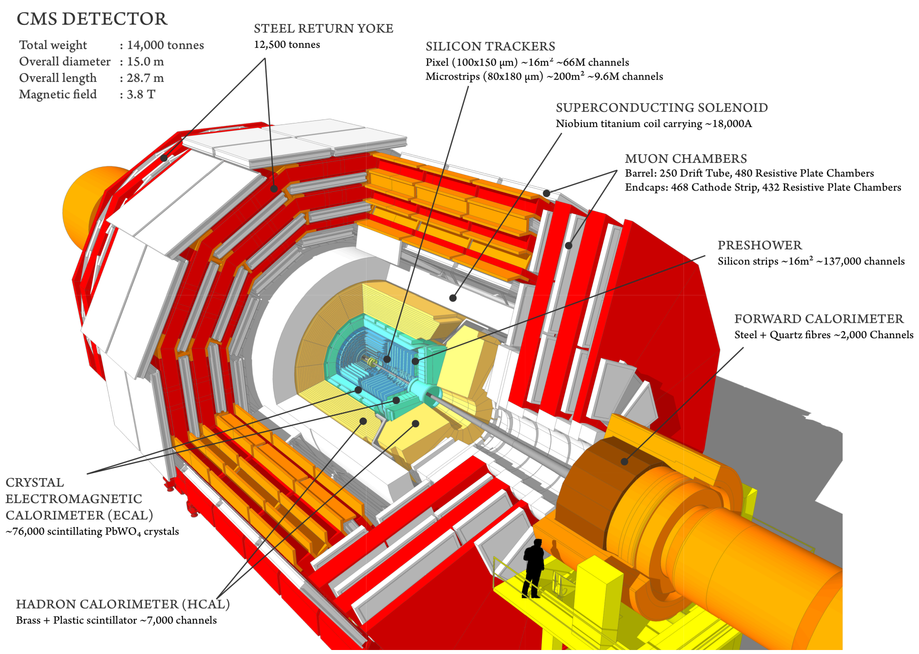


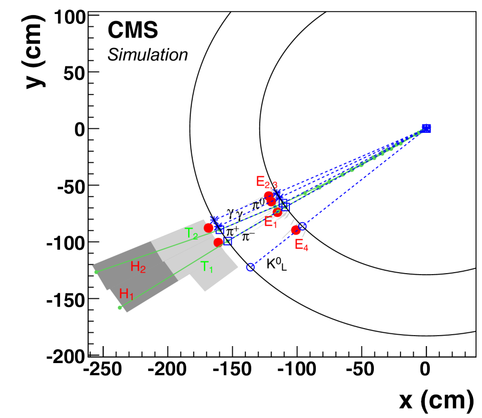

FTAG-2019-003

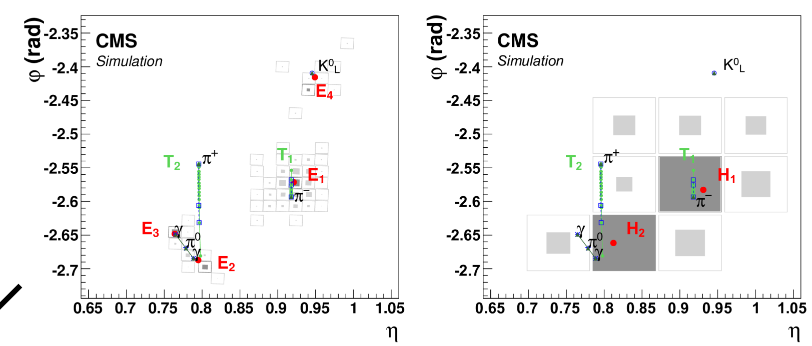


# CMS algorithms



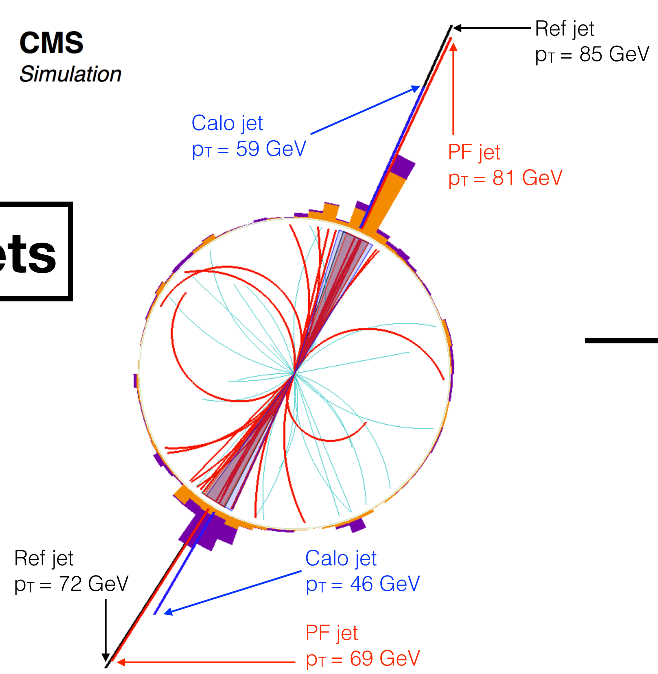




**Reconstruction & Particle Flow**
combine complementary detector info to create stable particles: e, μ, γ, charged/neutral hadrons

*JINST* 12 (2017) P10003

**PF Jets**

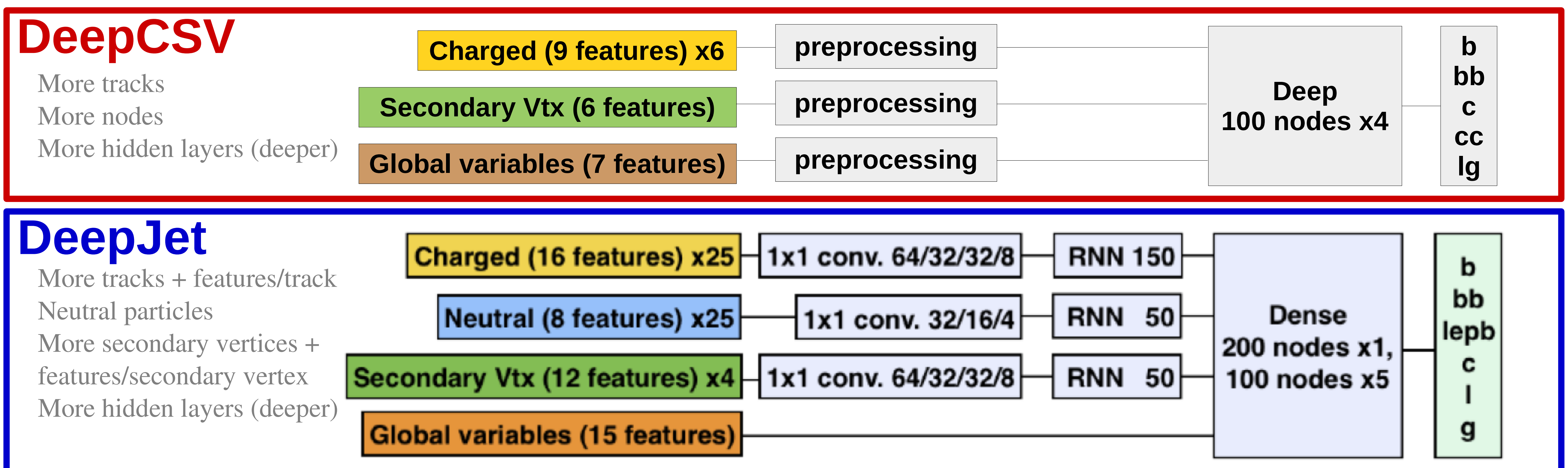



**high-level multi class taggers DeepCSV and DeepJet**

# DeepCSV & DeepJet design

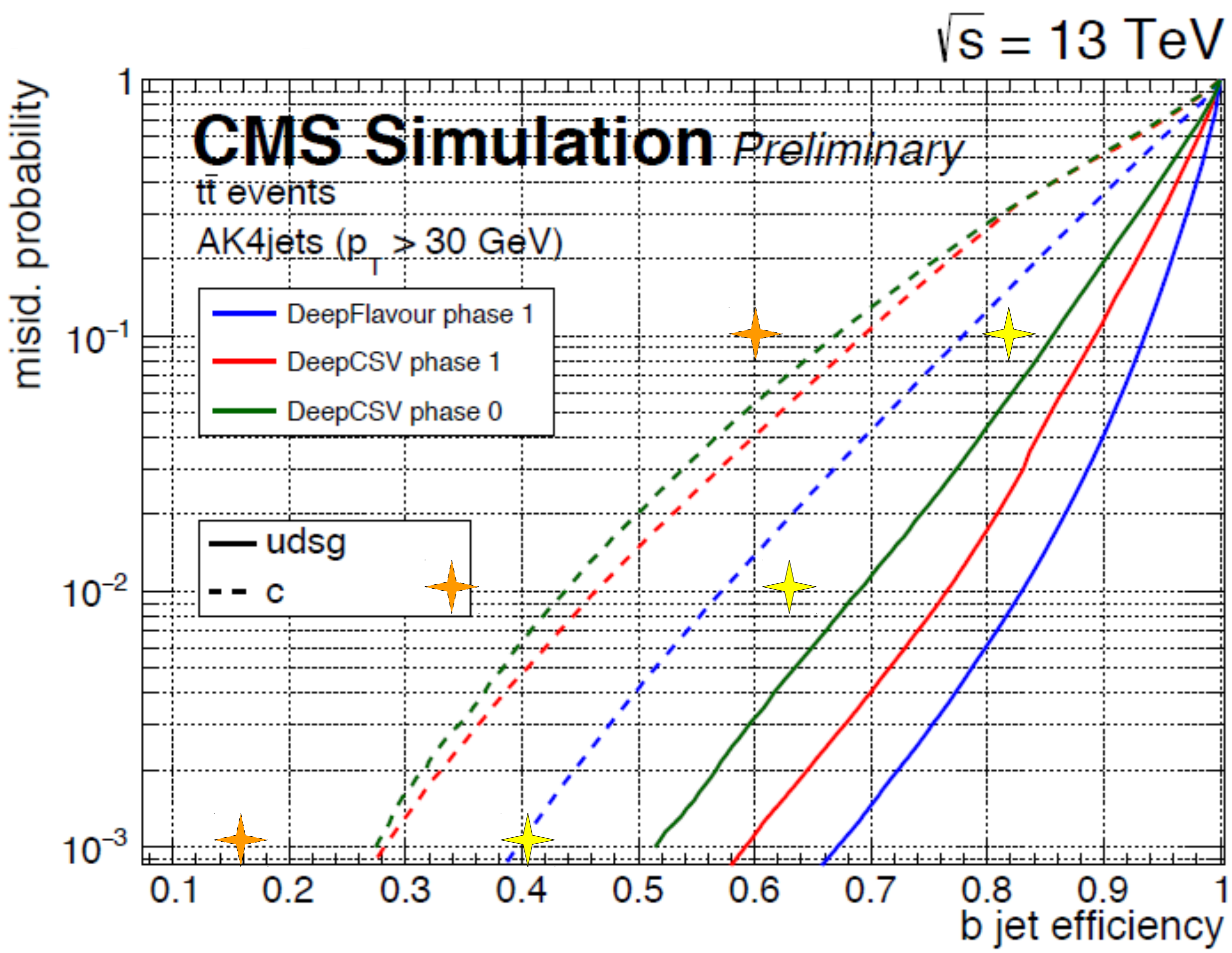
- **two flagship algorithms DeepJet and DeepCSV (Combined Secondary Vertex)**
    - DeepJet most performant, DeepCSV less affected by possible miss-modelings (see next slide)

- **both based directly on low level inputs** (different from ATLAS' approach)



# DeepCSV & DeepJet performance

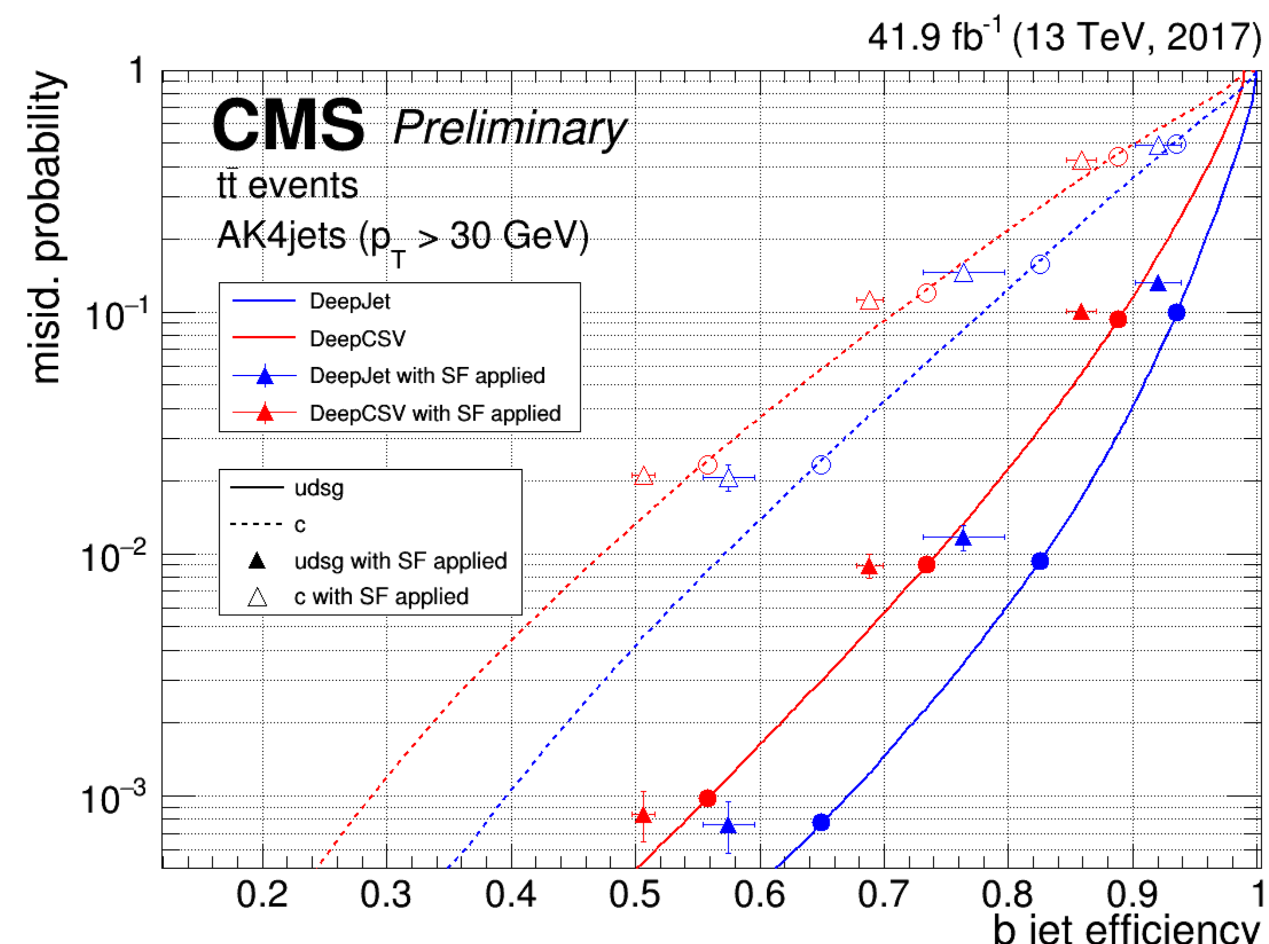
- **large performance gain compared to Run1 algos**
- data/MC discrepancies result in small performance loss




CSVv2 phase 0 – udsg misid. probability

CSVv2 phase 0 – c misid. probability

Approximation based on [JINST 13 (2018) P05011]



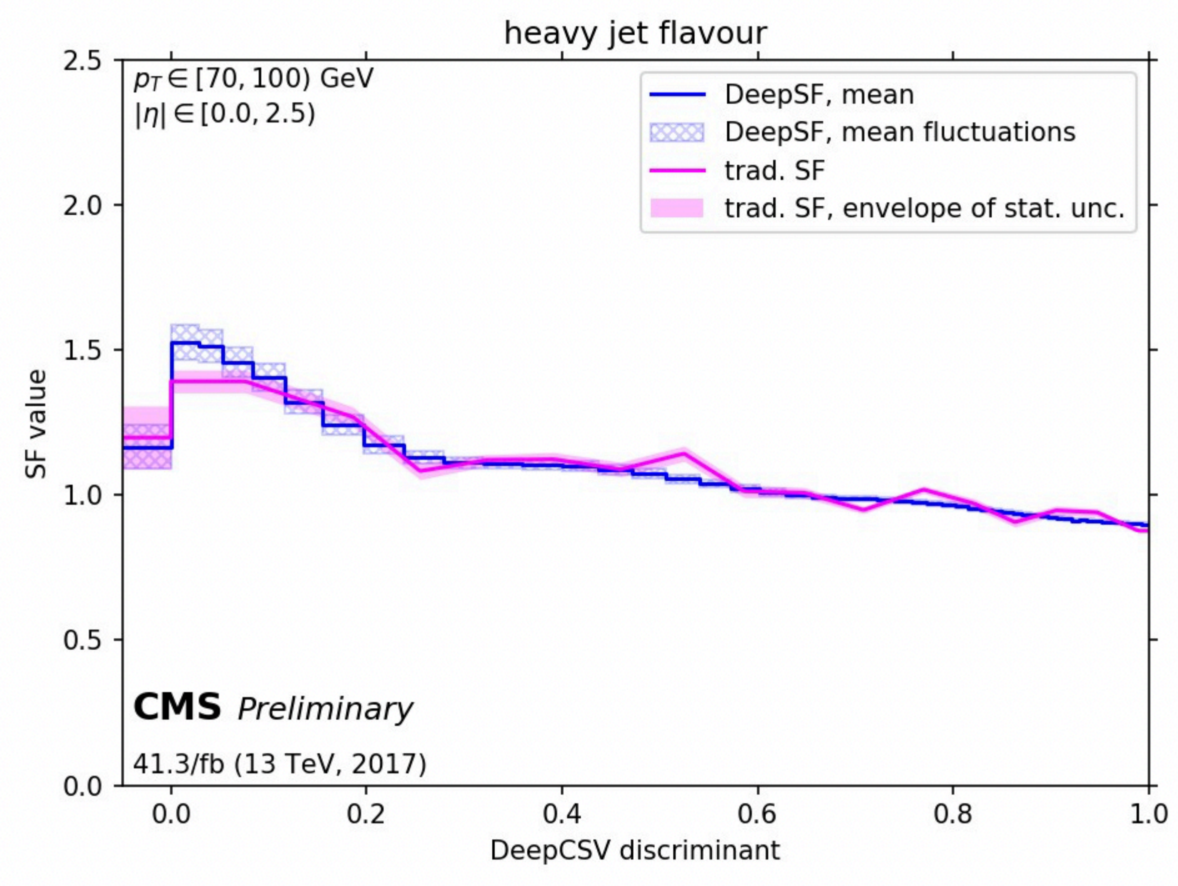

CMS-DP-2018-058

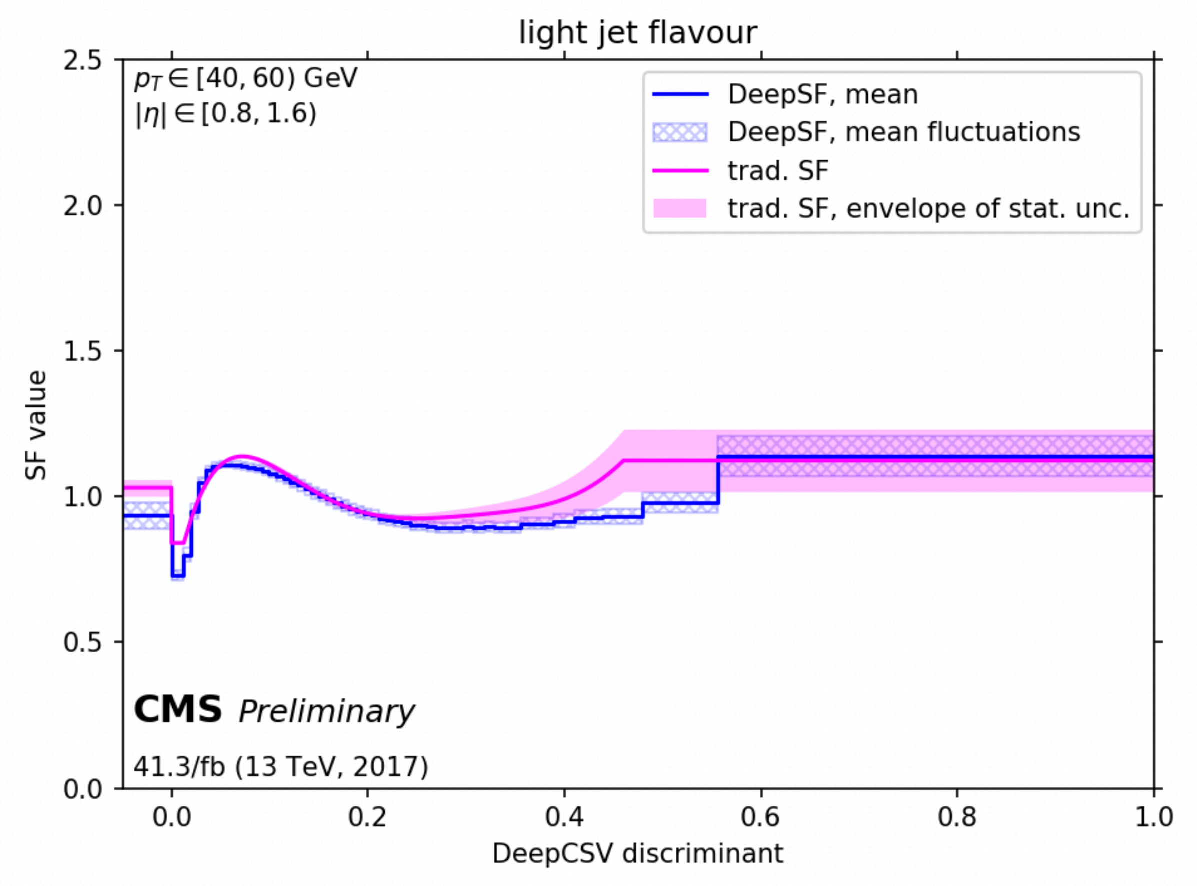
# Calibration in data

- **traditional approach** simultaneous fits (polynomials, splines) to control regions enriched in b-, c- and light-jets (tt, W+c, QCD…)

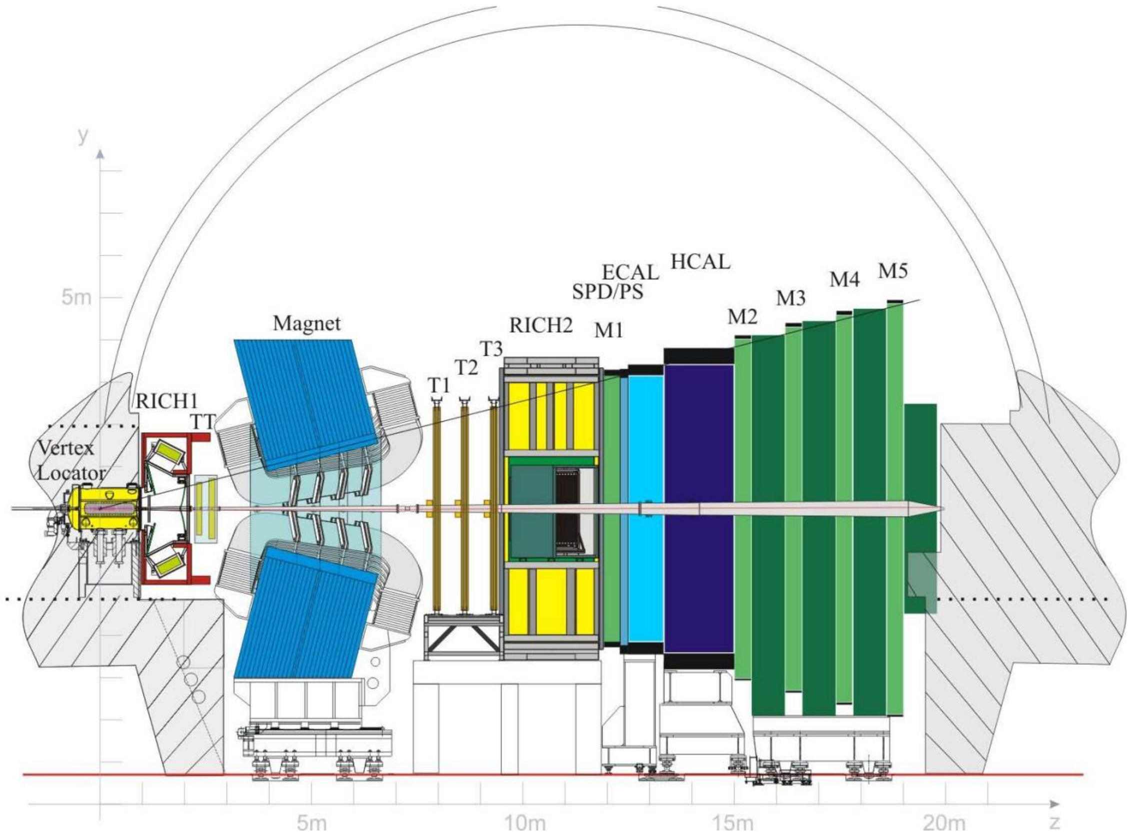
- **novel approach** adversarial neural network
  - smoother SF and better uncertainty

CMS-DP-2019-003



# LHCb SV tagger




- tracks $p_T$>0.5
- not belonging to IP $\chi^2$>16
- no PID, $m_\pi$ assigned to all tracks
- no ΔR<0.5 requirement
  - aimed at low $p_T$ jets, tracks outside the done

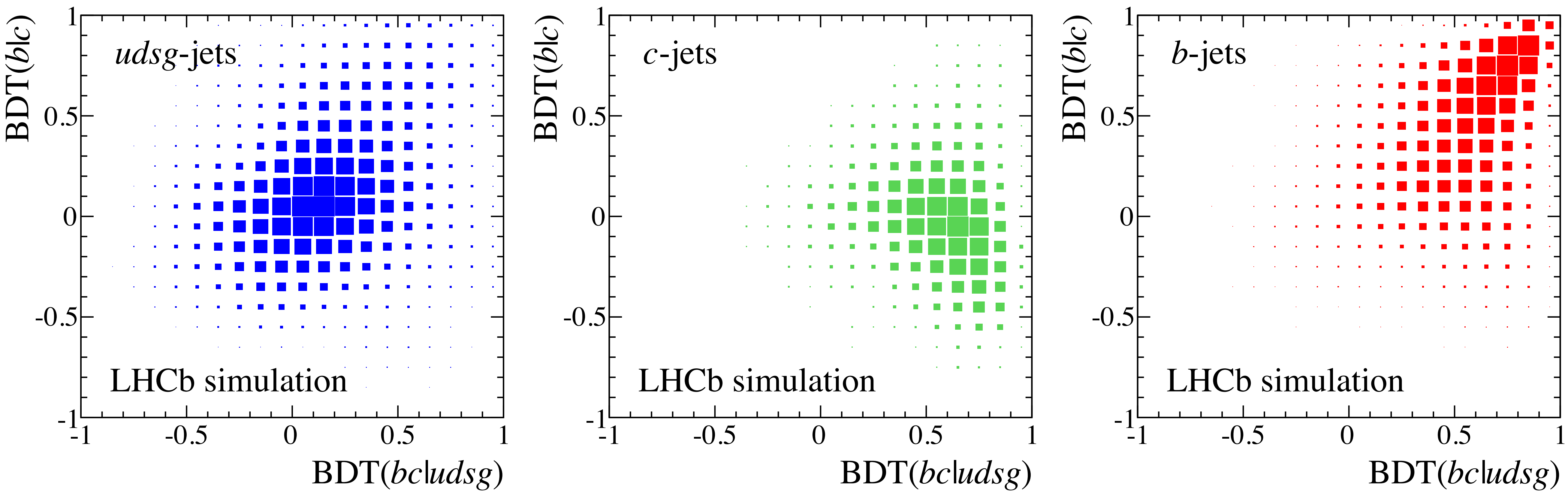
- build all 2-track SV
- require 0.4 GeV < $M_{SV}$ < M($B^0$)
- at this stage, the same track can belong to more than one SV

- linking: merge SV that share tracks and are in ΔR<0.5 form the jet axis
- $M_{corr} = (M^2 + p^2\sin^2\theta)^{1/2} + p\cdot\sin2\theta$ minimum mass compatible with the direction of flight of the SV

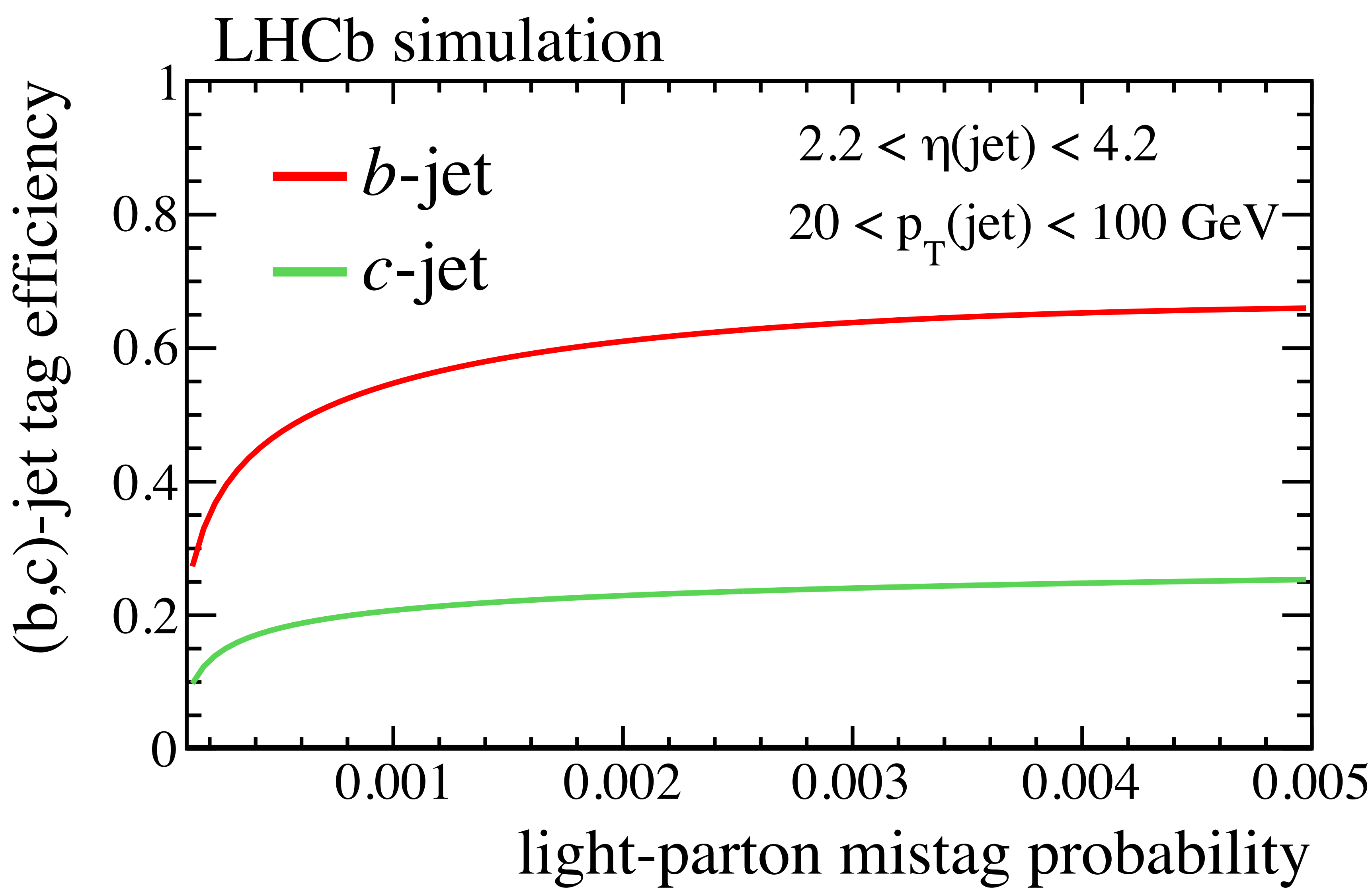
# SV tagger

JINST 10 P06013

- properties of the identified SV are passed to two BDTs
- BDT(bc|udsg) to separate heavy from light jets and BDT(b|c) to separate b- from c-jets



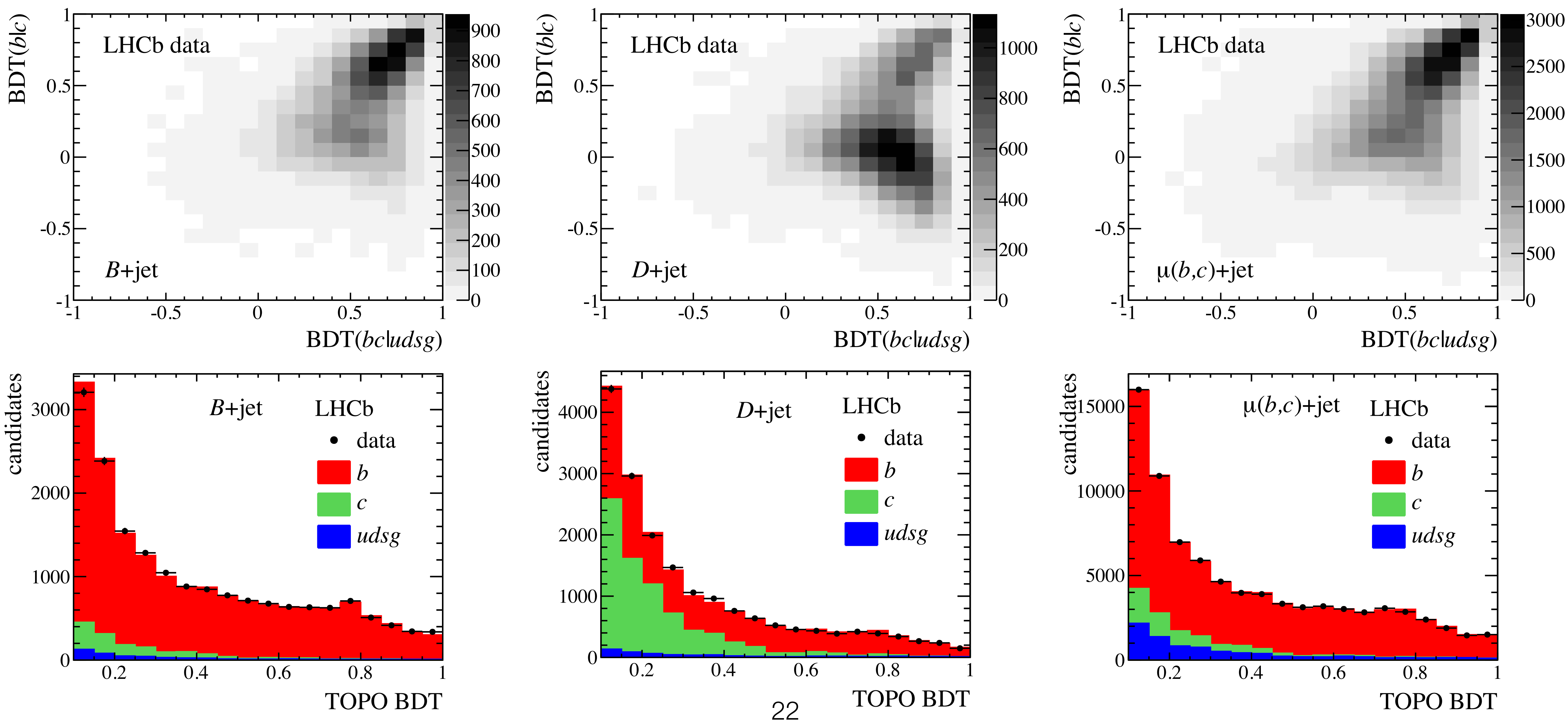
# SV tagger performance



# Calibration in data

JINST 10 P06013

- using B, D and μ + jets data



# Tau reconstruction

# Tau reconstruction and identification

Eur. Phys. J. C75 (2015) 303

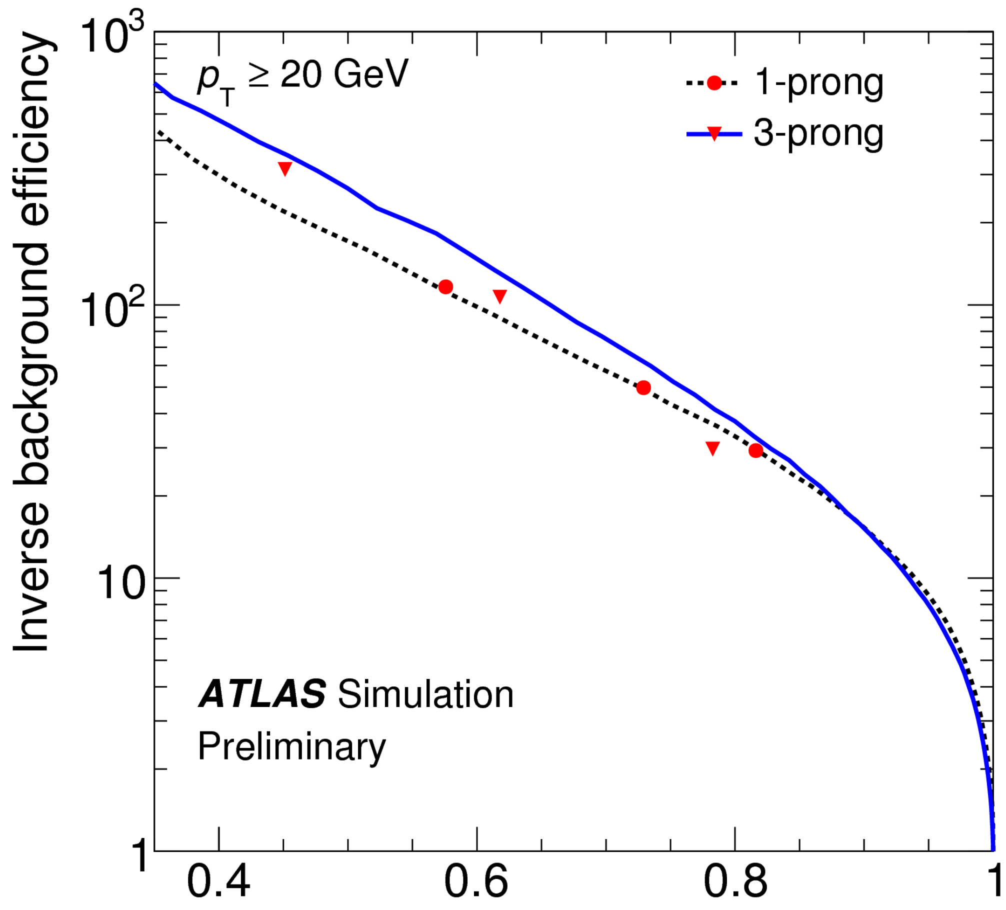
- **seeded from AK4 jets from calorimetric *TopoClusters***
  - tracks are assigned to *core* and *isolation* regions if $\Delta R(\text{track, jet-axis}) < 0.2$ and $0.2 < \Delta R < 0.4$ respectively
  - track-vertex compatibility requirement

- **$\pi^0$ reconstruction**
  - first looks for $\pi^0$ candidates in the core region using BDT
  - foreach candidate, a $\pi^0$-likeness score is assigned. Onlythe higher scoring $\pi^0$ are considered

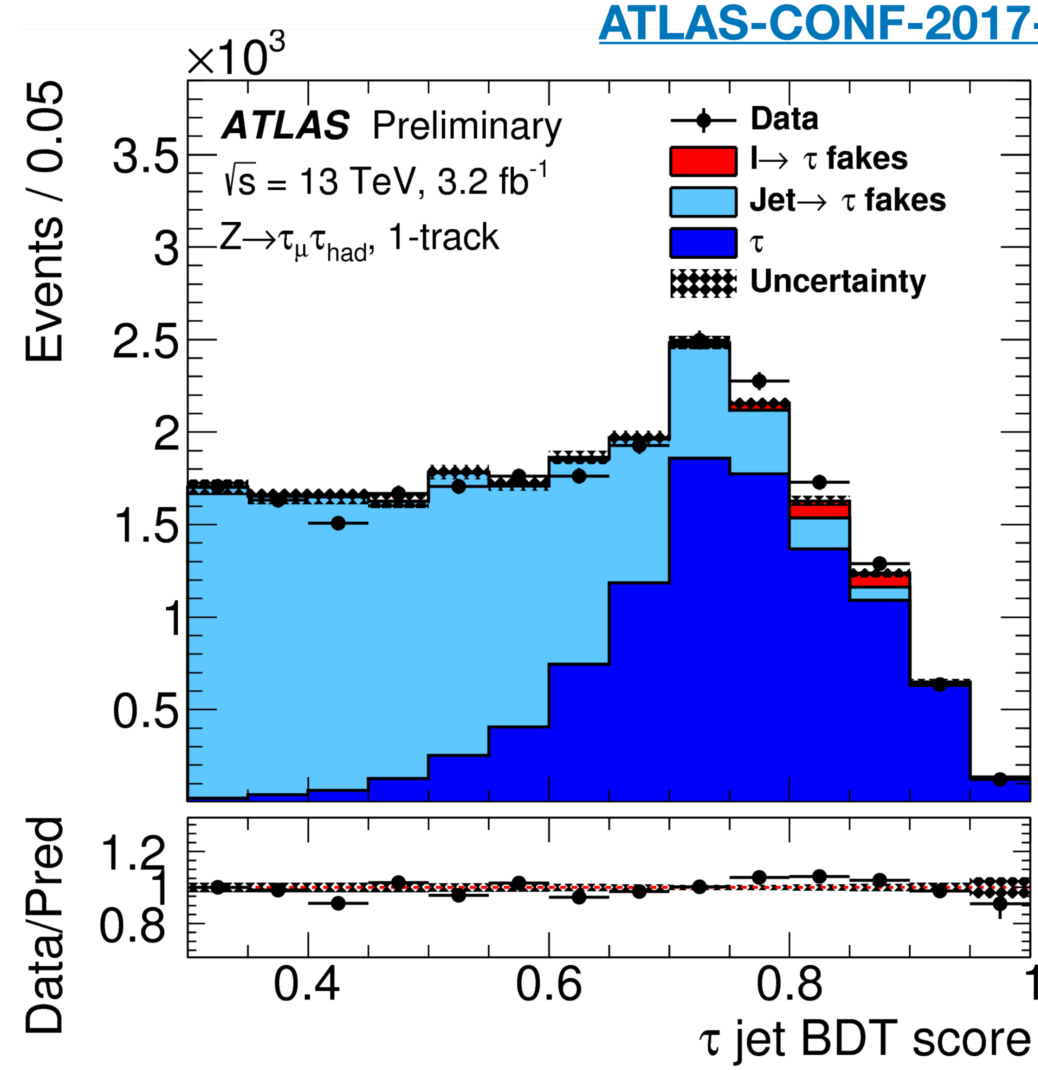
# Tau identification

- **BDT discriminator against jet fakes, trained for 1/3 tracks.**
  BDT- and cut-based discriminators against e and μ respectively
- efficiency, fake-rates and tau energy scale measured in data using mainly $Z \to \tau\tau$ and $Z \to \ell\ell$ candles

ATLAS-CONF-2017-029



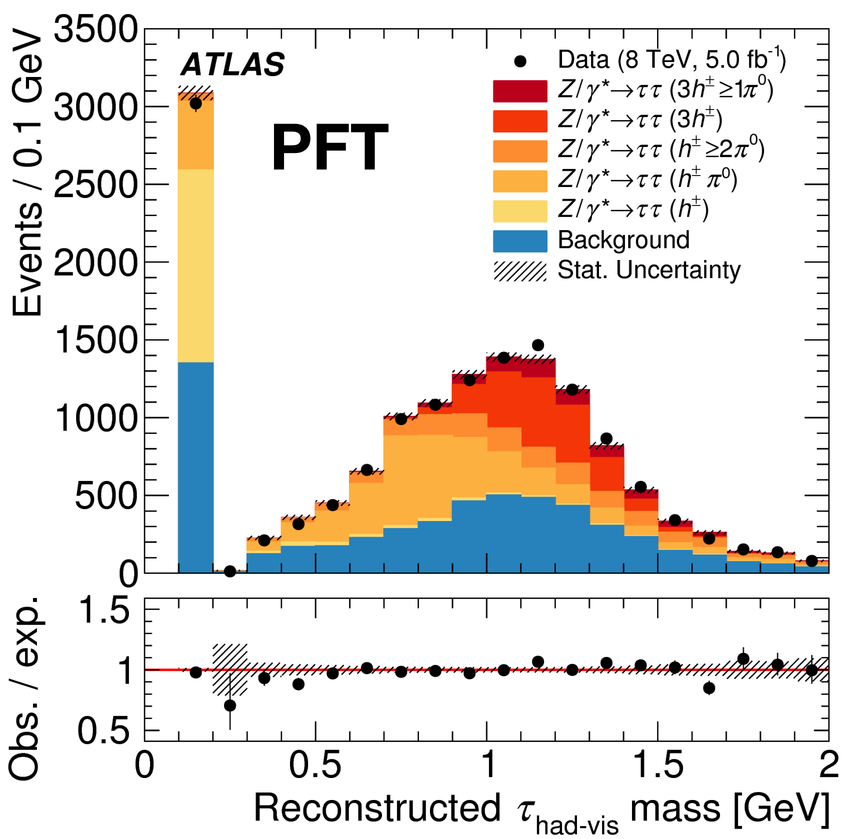

ATL-PHYS-PUB-2015-045

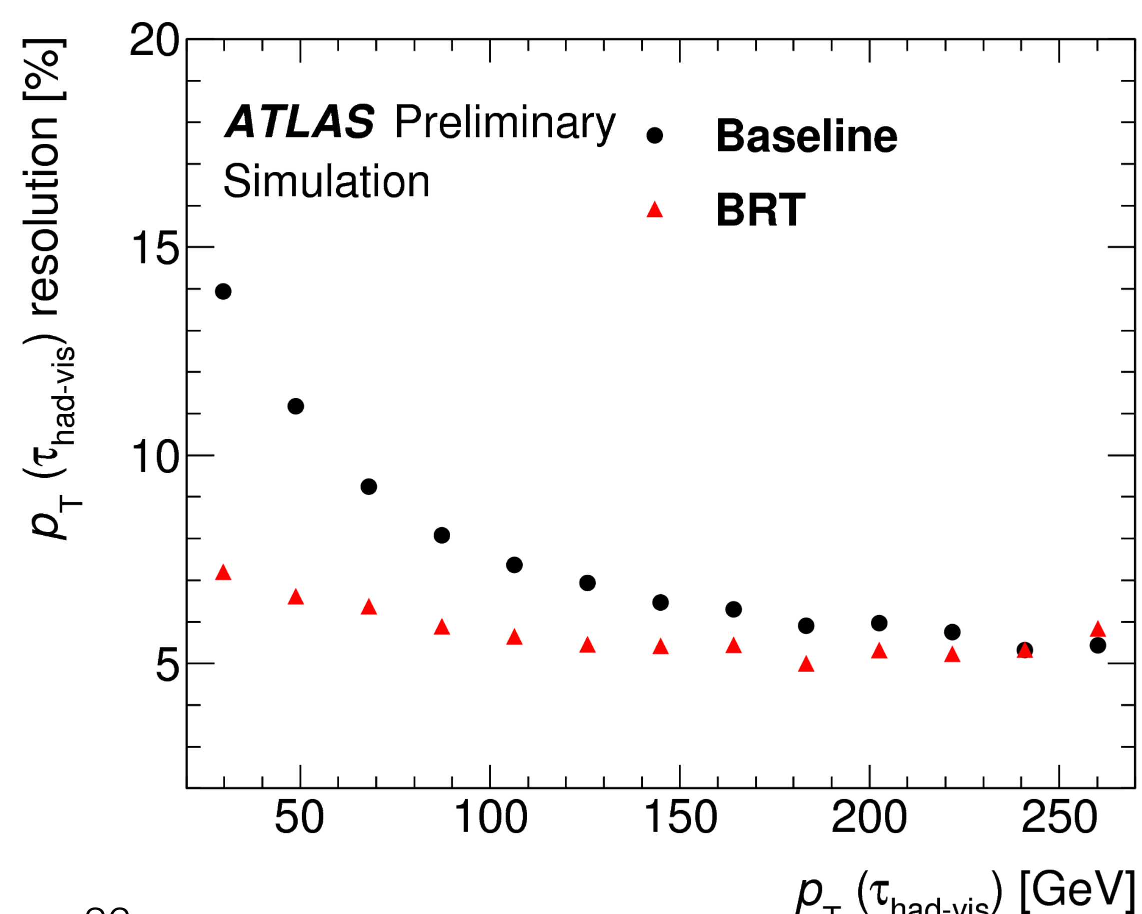


# τ energy calibration

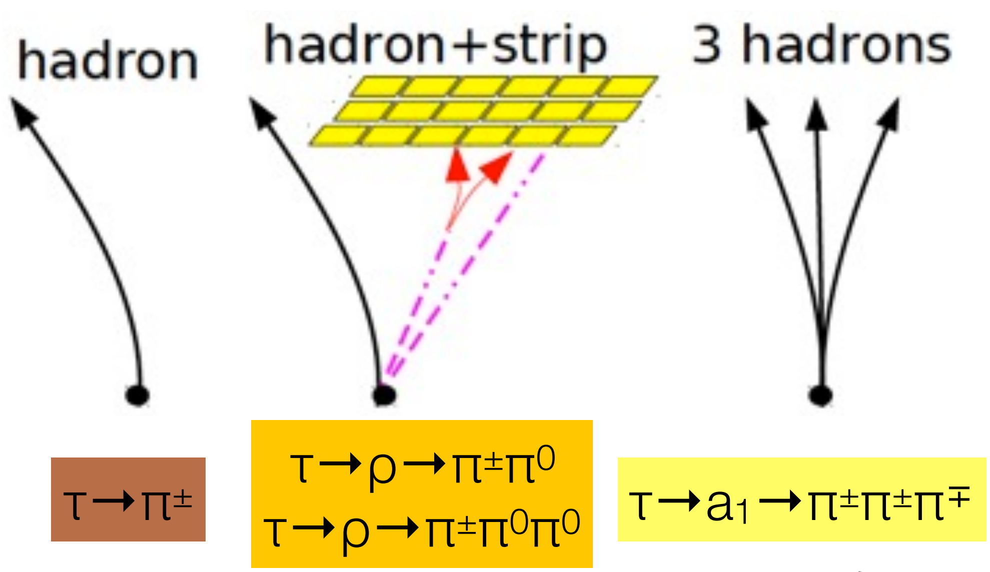
[Eur. Phys. J C 76 (2016) 295]

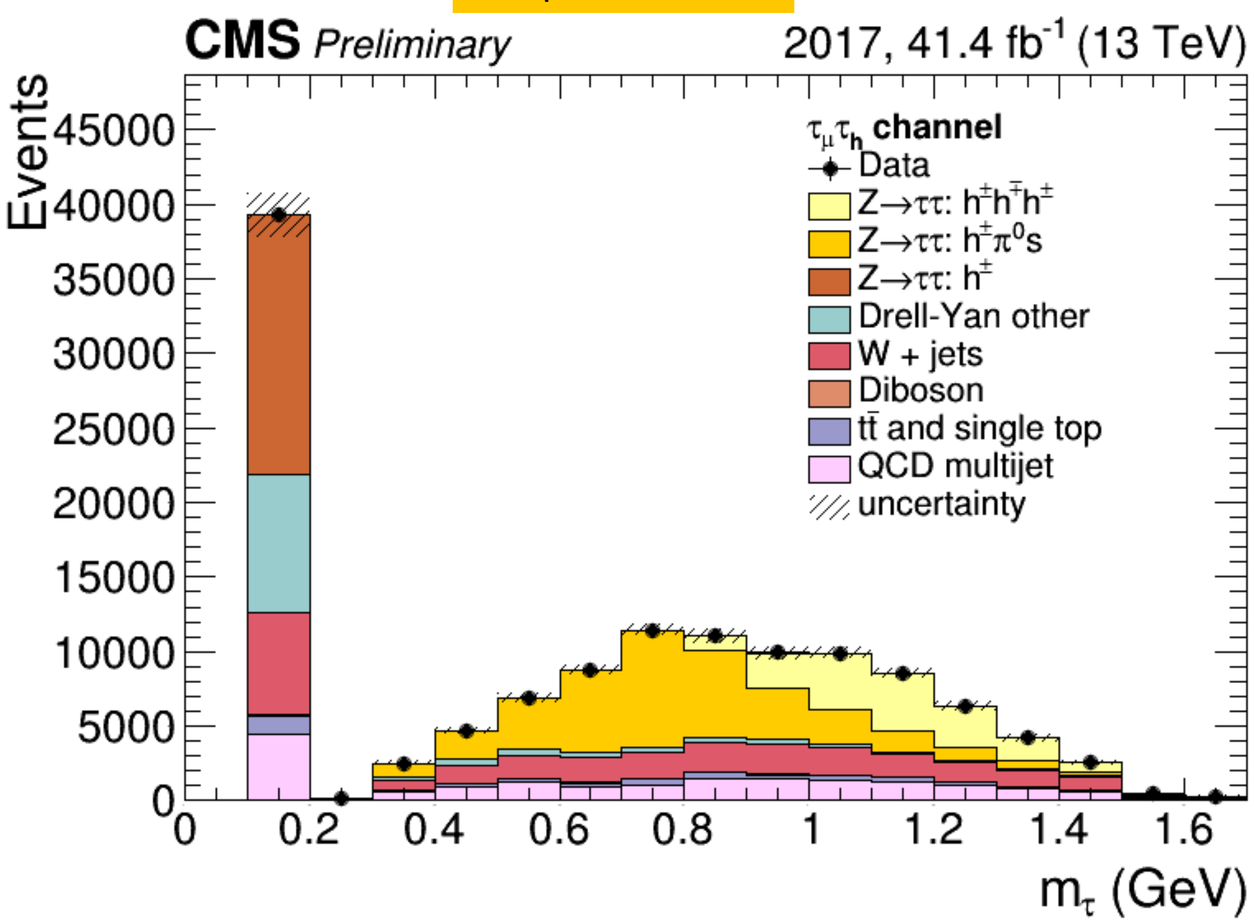
- **baseline calo-based:** removes PU contribution, good at high $p_T$, degrades at lower $p_T$
- **Boosted Regression Tree:** improves at low $p_T$ including information from Particle Flow Taus leveraging on track info



# Tau reconstruction: Hadron Plus Strip



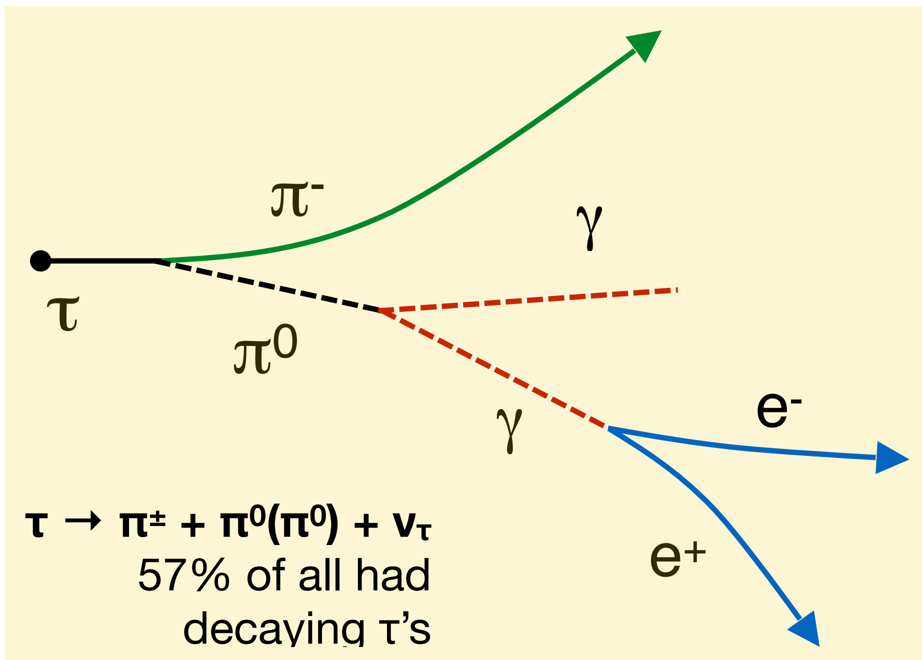



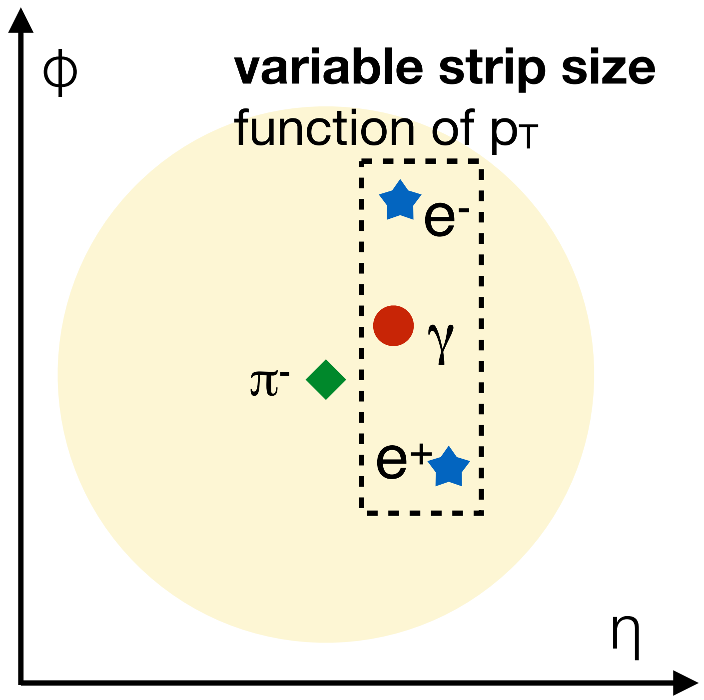

JINST 13 (2018) P10005

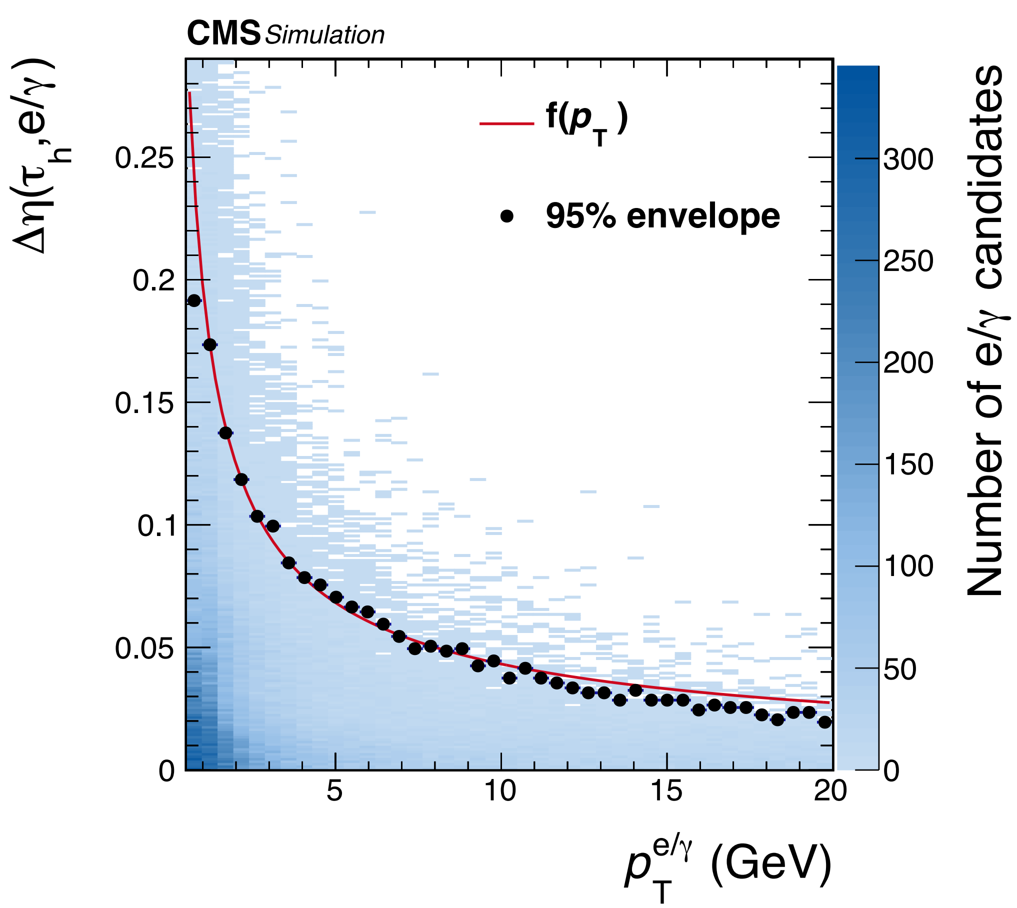
- **Particle Flow inputs**
  jets and their charged and neutral constituents
- **identifies $\tau_h$ decay mode**
  - exploits the ρ(770) and $a_1$(1220) intermediate resonances through mass window requirements

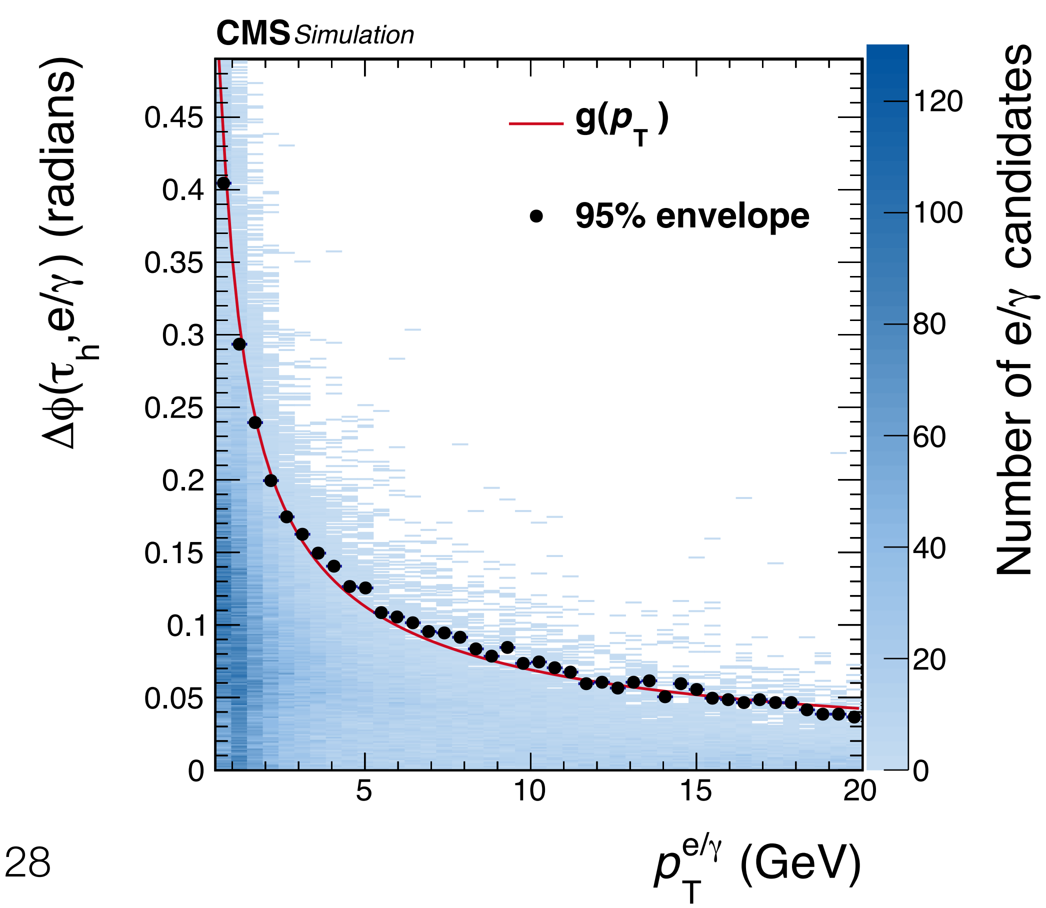
# Dynamic strip ($\pi^0$)

JINST 13 (2018) P10005



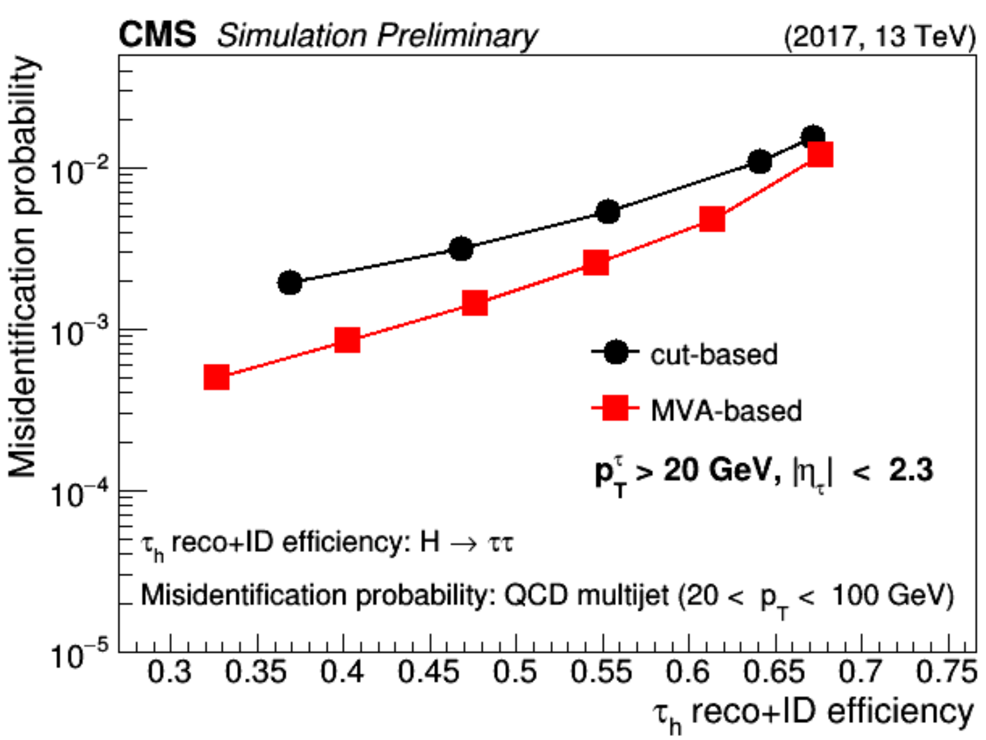



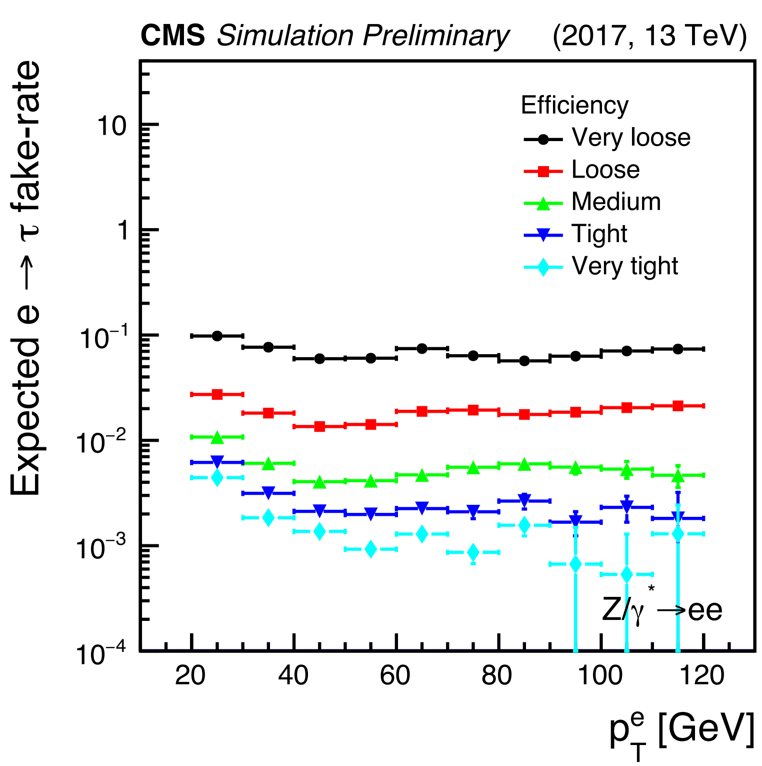
# Tau identification

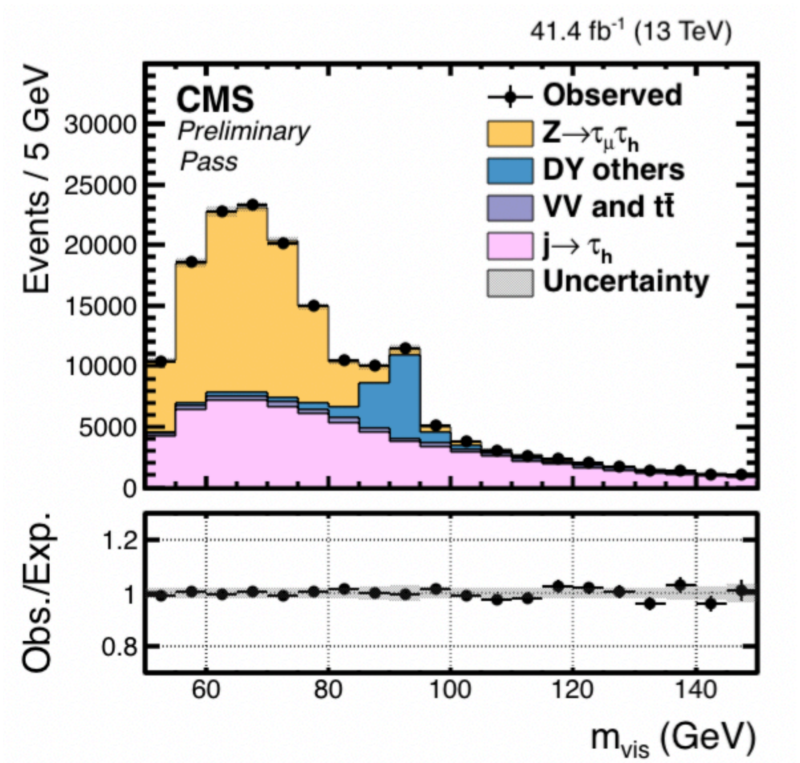
[CMS-DPS-2018-026]

- **BDT discriminators to suppress jets and electrons faking taus,** cut-based anti-muon discriminator

- efficiency, fake-rates and tau energy scale measured in data using mainly Z→ττ and Z→ℓℓ candles



# Conclusions

- **b-, c- and τ tagging crucial for large part of physics searches and measurements at LHC**
- **widespread usage of modern machine learning tools**
  - main driver of performance improvement wrt Run 1, together with detector upgrades
  - calibrations and data/MC modelling critical → discrepancies tend to hit performance
- **ATLAS and CMS obtain remarkably similar performances while often adopting different approaches, LHCb explores complementary phase space**