Getting to the “bottom” of precision measurements and/or new physics

Petra Van Mulders

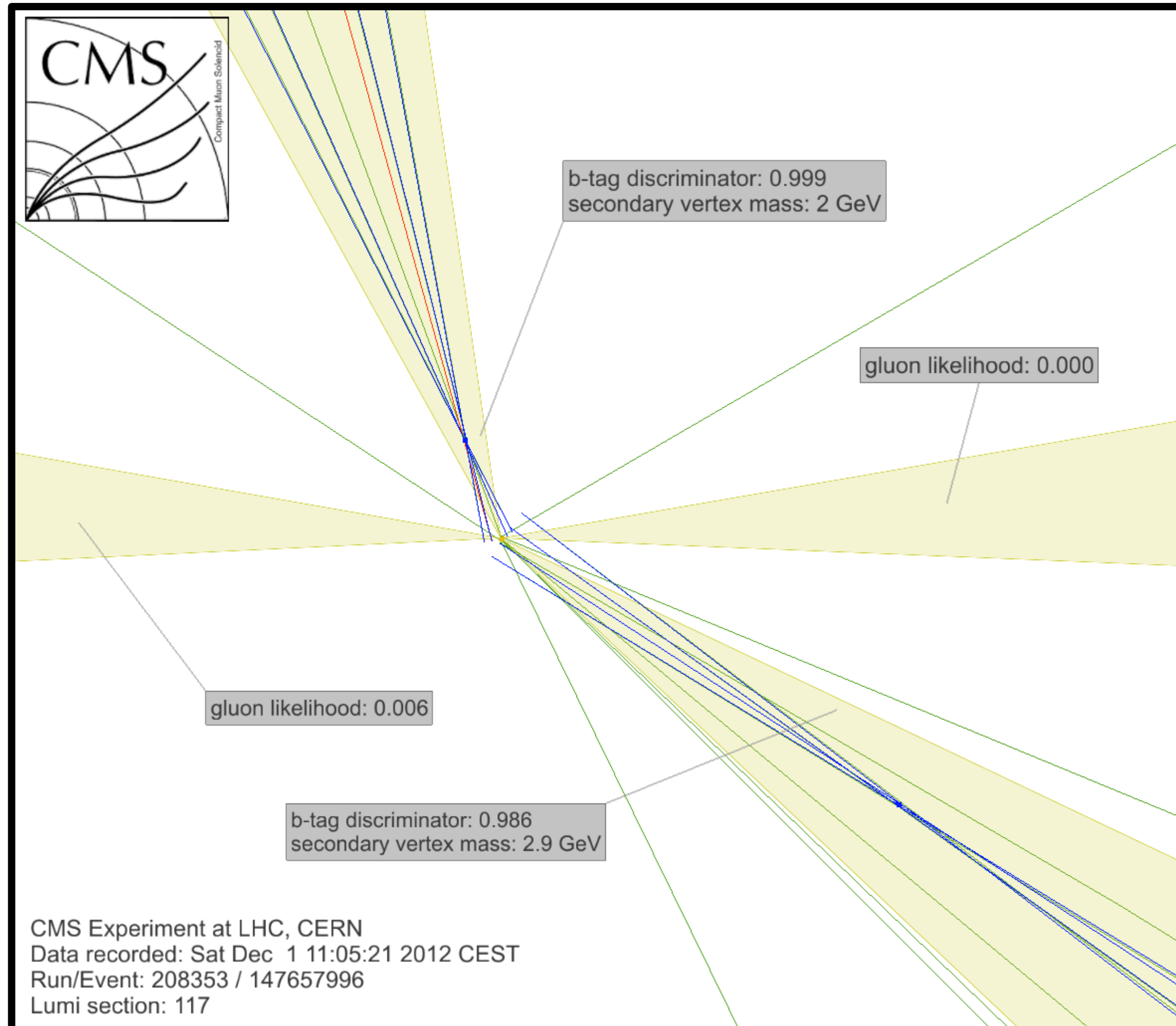
Vrije Universiteit Brussel

iihe BRUXELLES BRUSSEL

fwo Research Foundation Flanders
Opening new horizons

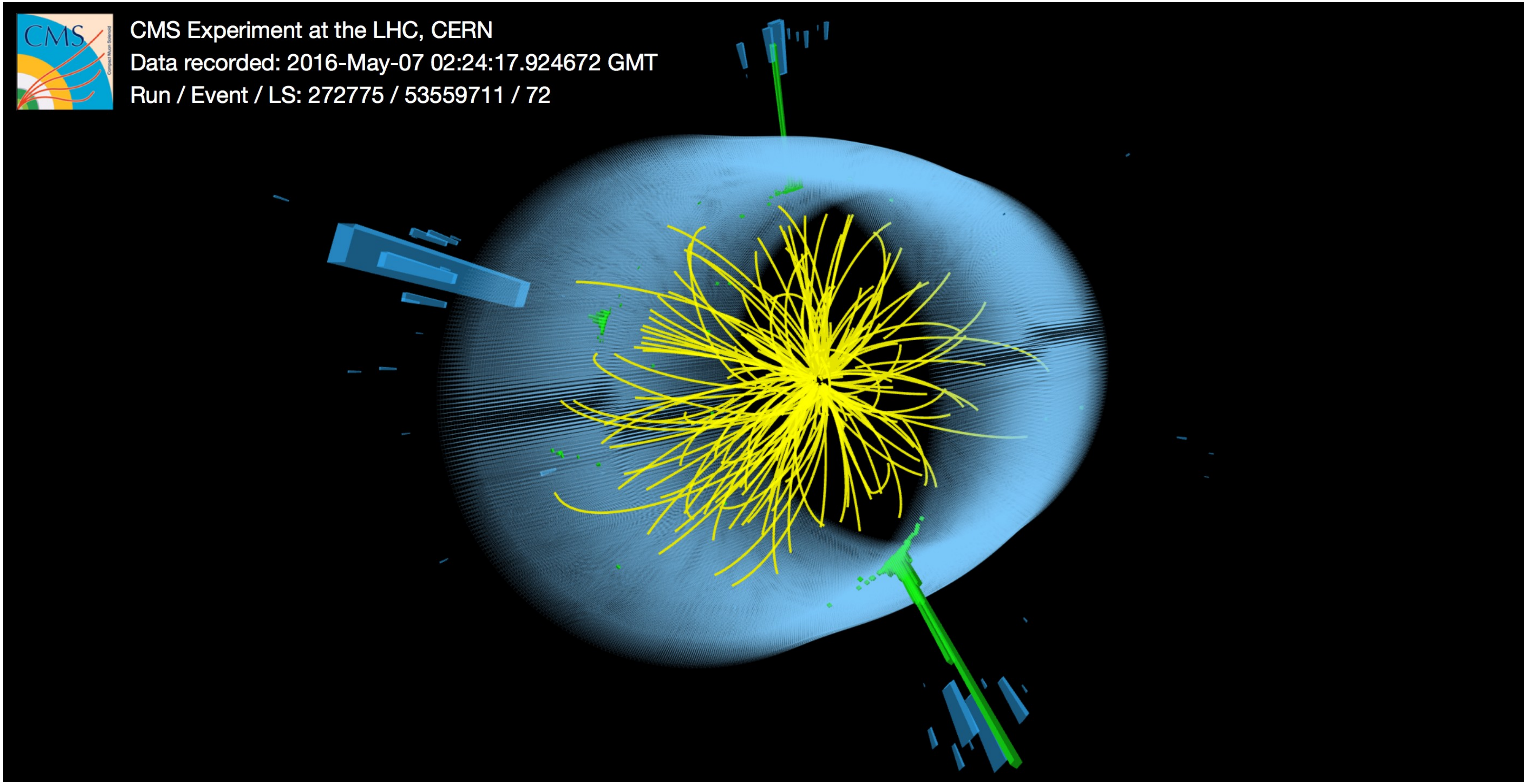
Fundamental Interactions .be

Belgian Science Policy Office belspo



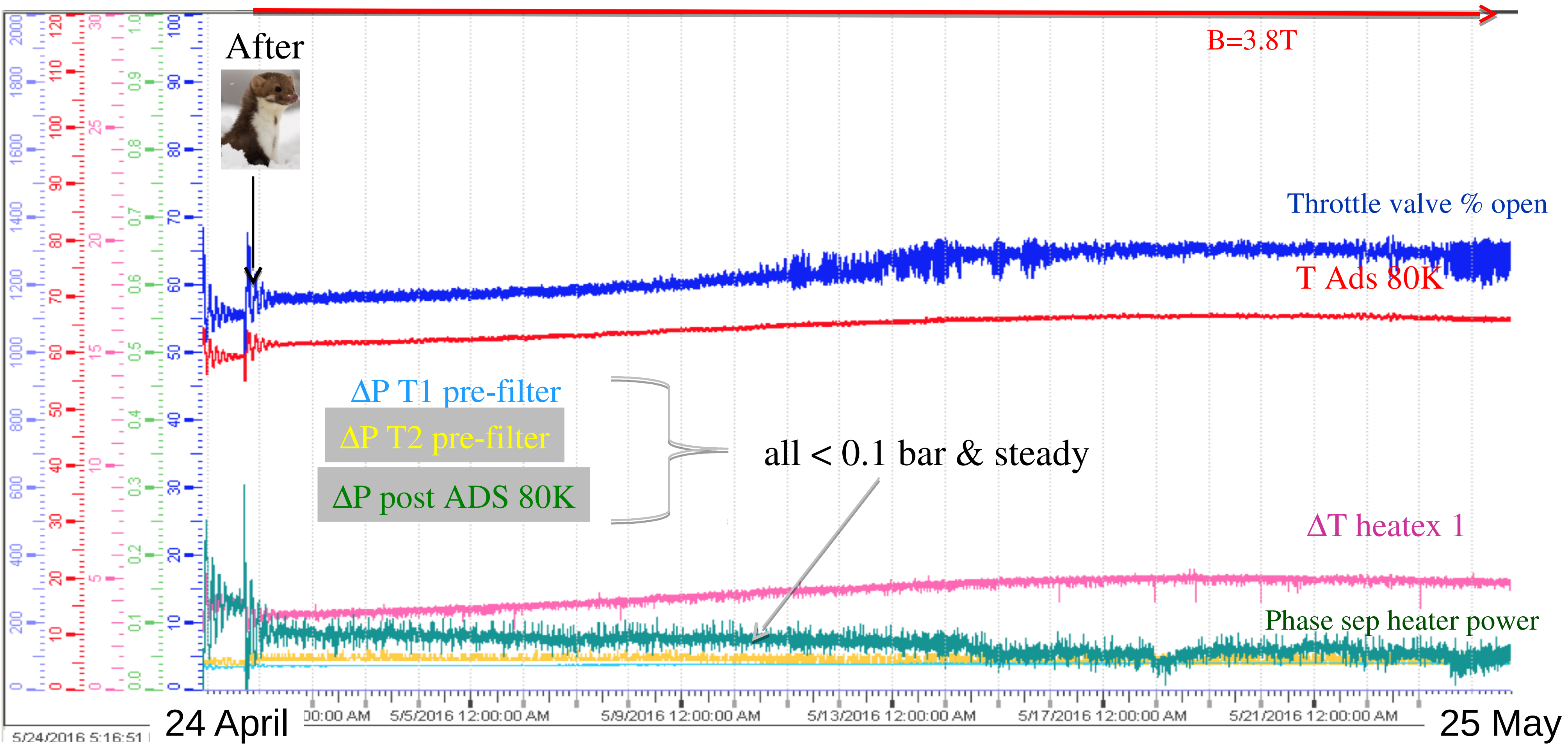
source: HIG-13-011

Collisions at 13 TeV



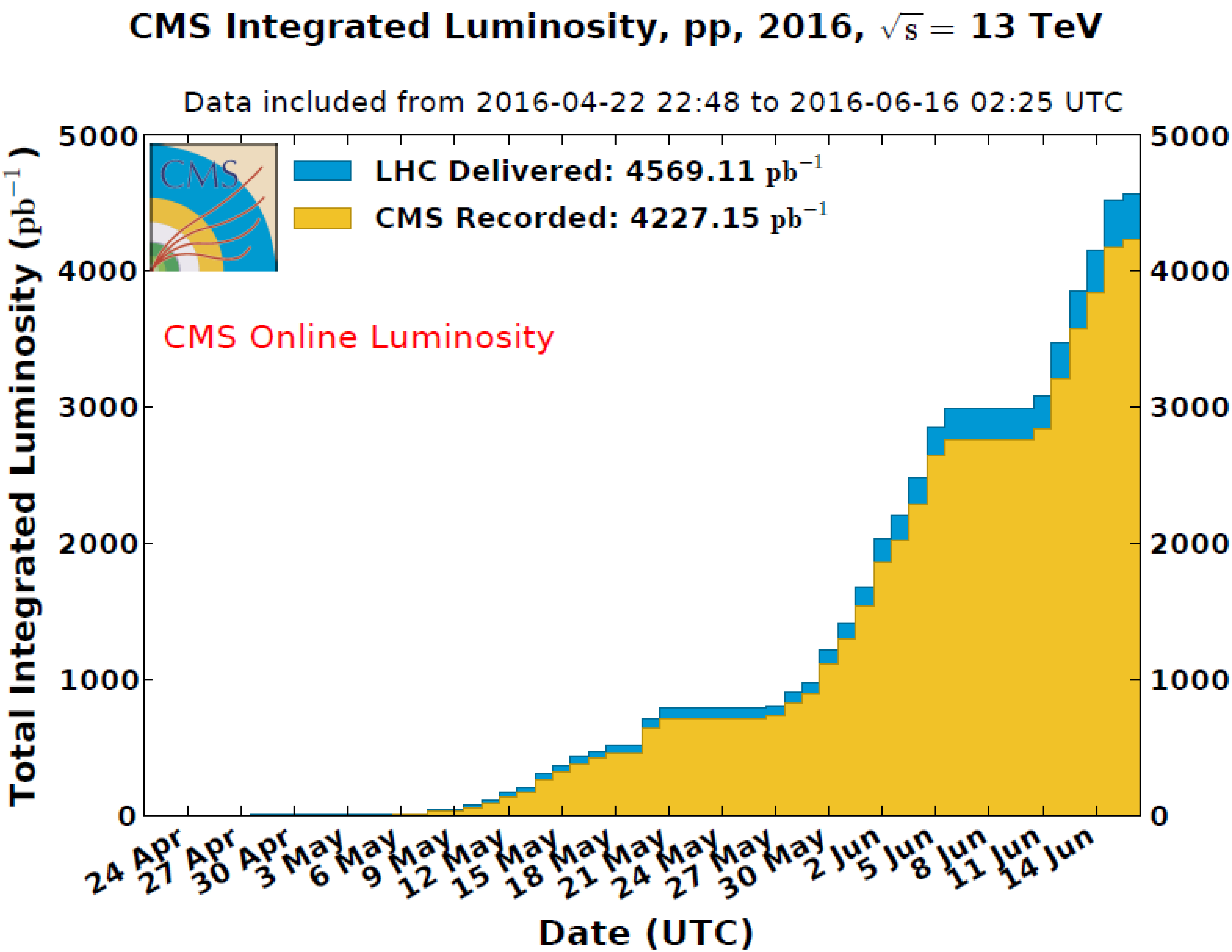
CMS magnet cyrogenic system

Reached 3.8T on April 28, operational parameters stable in time



A warm thanks to colleagues from CERN-TE dept, technical support from other CERN depts, CERN-EN,EP, CERN Management, CMS Magnet team and integration office, contractors (particularly Altead, ZEC service), CMS members for support and advice

CMS is efficiently taking data



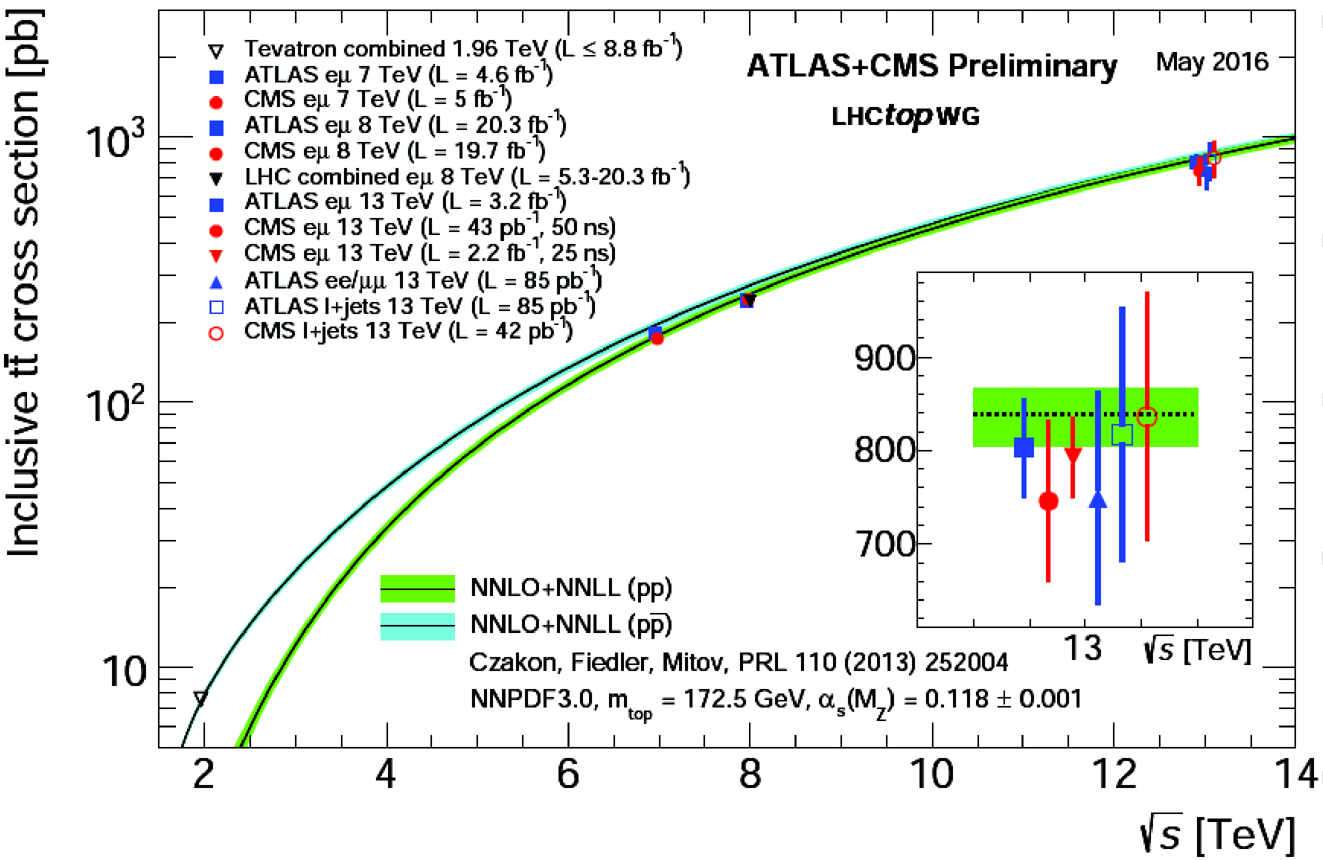
Special thanks to the LHC team!!!

Content

- A few physics highlights

- b jet identification

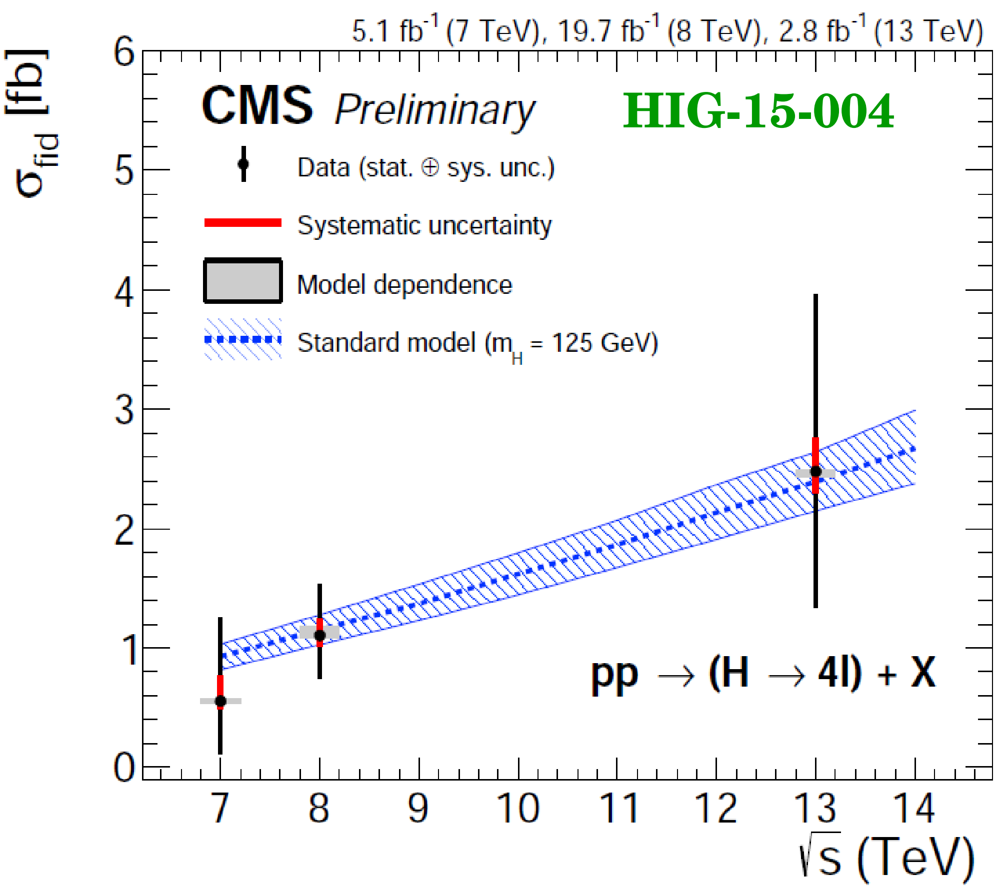
Top quark production



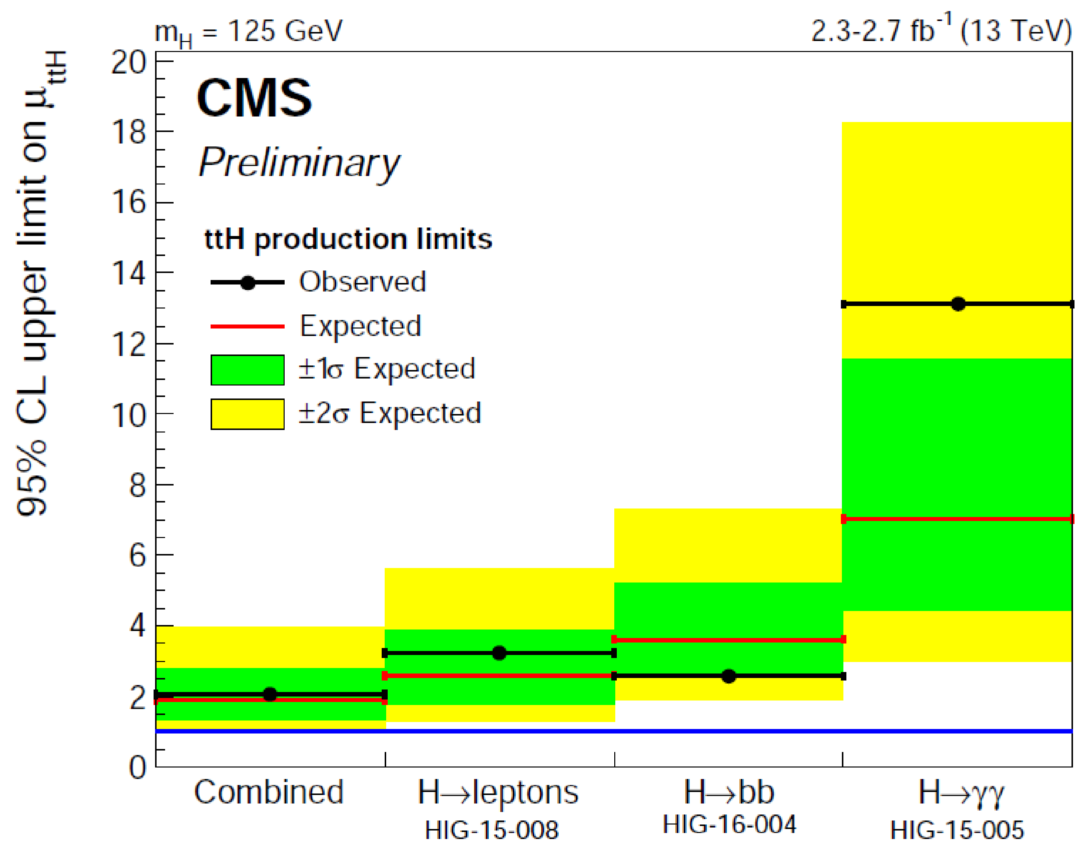
Differential cross section measurements performed in all decay channels for Run 1 and Run 2

- Cross section at 13 TeV measured with a precision of 5.5% **TOP-16-005**
- Cross section measured at 5 TeV **TOP-16-015**
- Single top t-channel: 15% uncertainty **TOP-16-003**
- Top quark mass measured with a precision of 0.3% (Run 1) **TOP-14-022**
- W boson polarization → most precise measurements of helicity fractions to date (Run 1) **TOP-13-008**

Higgs physics at 13 TeV



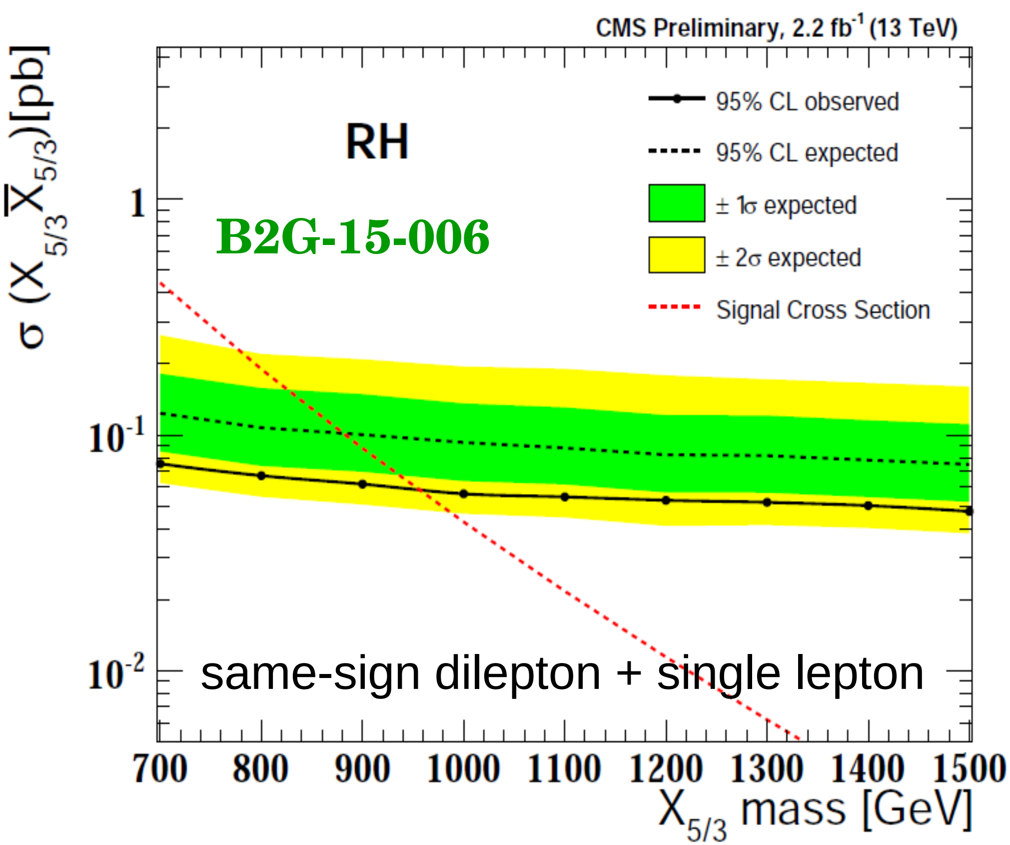
Large jump in sensitivity due to larger cross section at 13 TeV



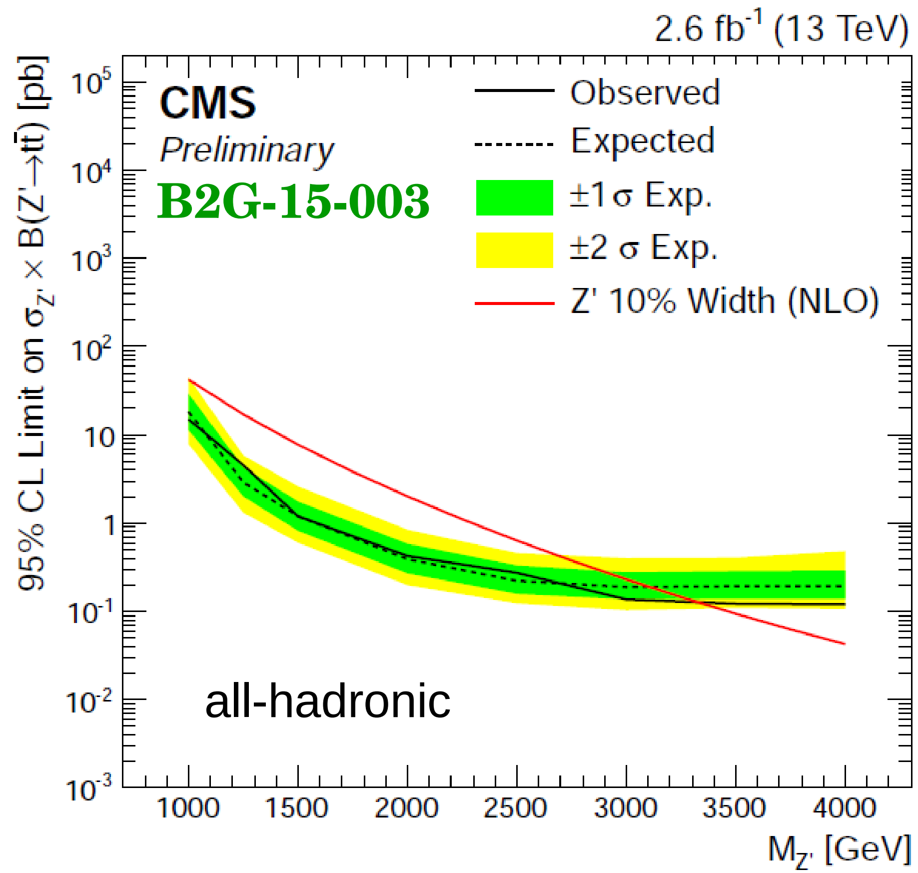
μ_{ttH}: $0.15^{+0.95}_{-0.81}$ **TTHCombMoriond2016**

(Run 1 μ_{ttH}: $2.9^{+1.0}_{-0.9}$) **HIG-15-002**

Searching new “exotic” particles

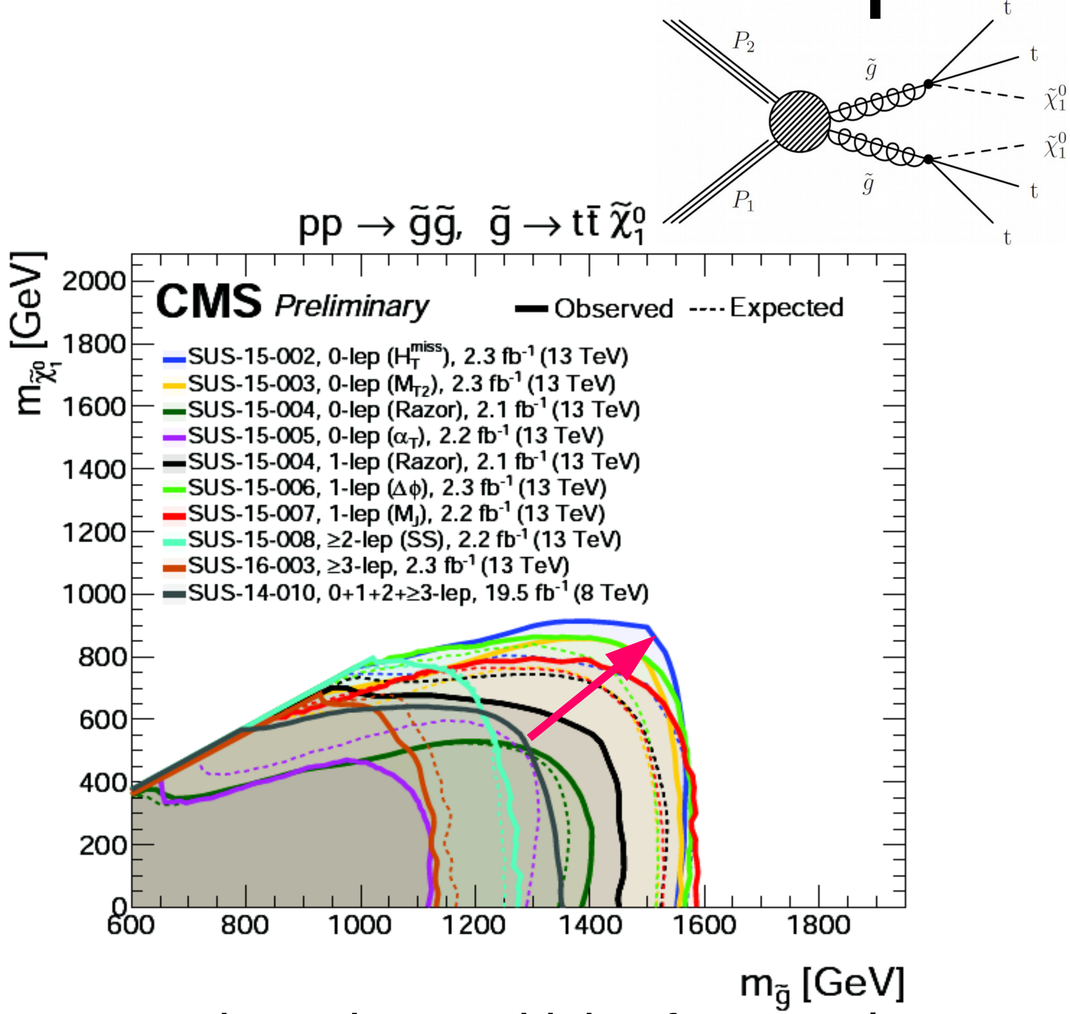


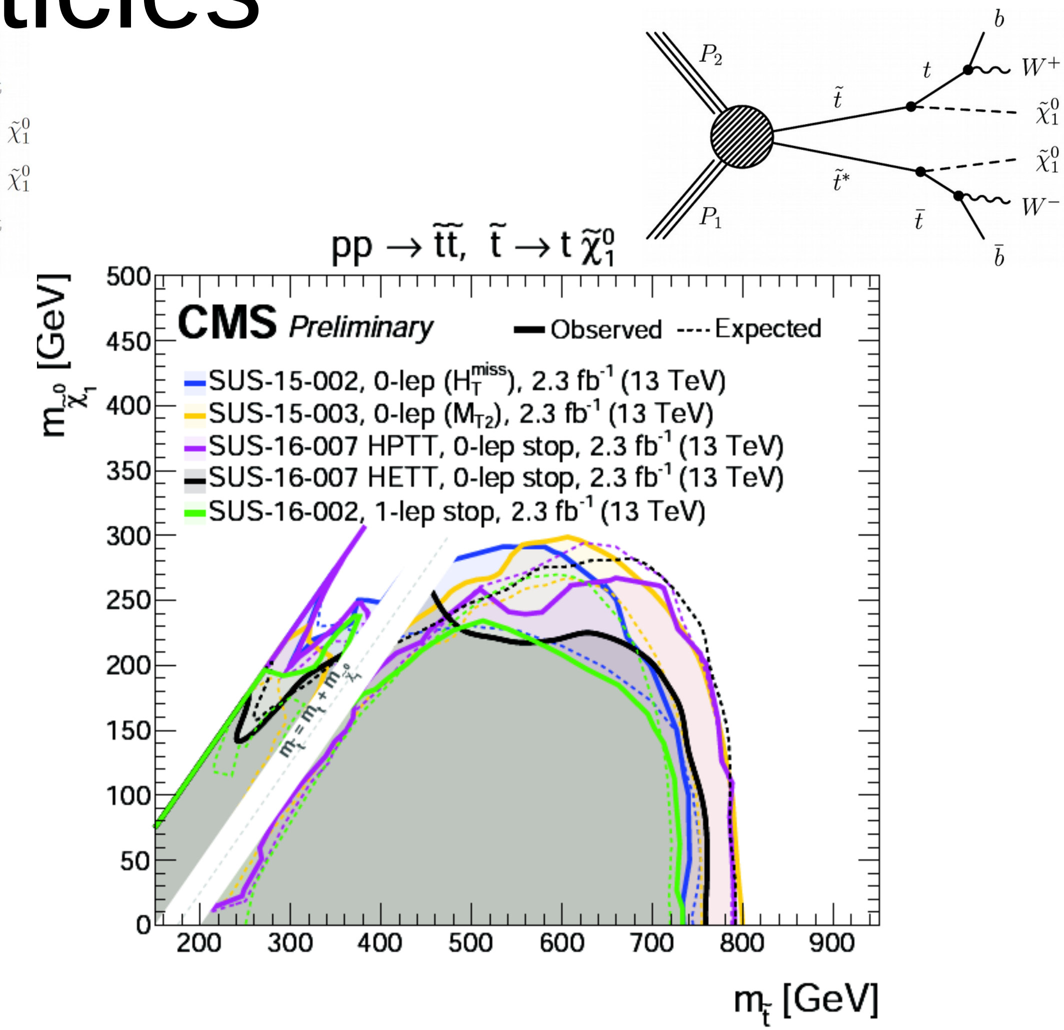
Existence of top quark partner $X_{5/3}$ is excluded with masses below 0.96TeV (Run 1 limit: 0.8TeV)



Existence of Z'→ tt (10% width) excluded with masses between 1 and 3.3 TeV (Run 1 limit, all decay channels combined: 2.9TeV)

Searches for sparticles





Huge jump in sensitivity, for massless neutralino → limit on the gluino mass at least 200 GeV higher

The “stealth” region is probed by precision measurements of top quark pair production

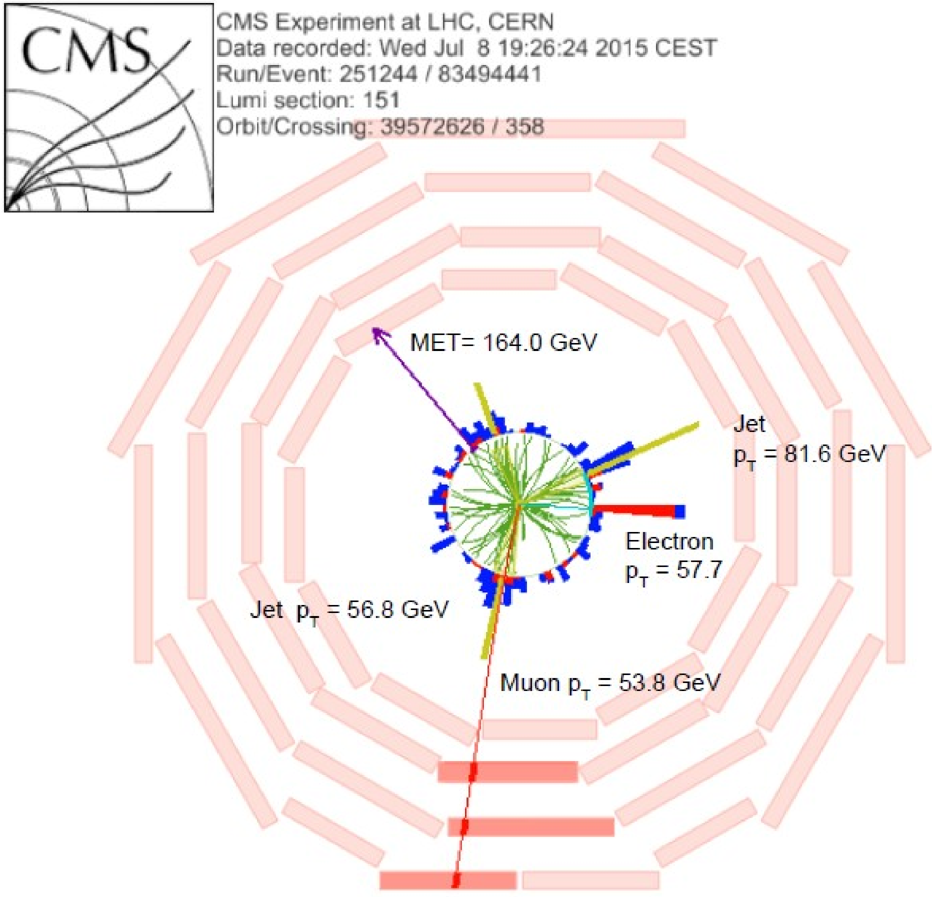
Content

- A few physics highlights
  - Accurate reconstruction algorithms are vital for the sensitivity of physics analyses
  - All of the highlights shown before rely on b jet identification

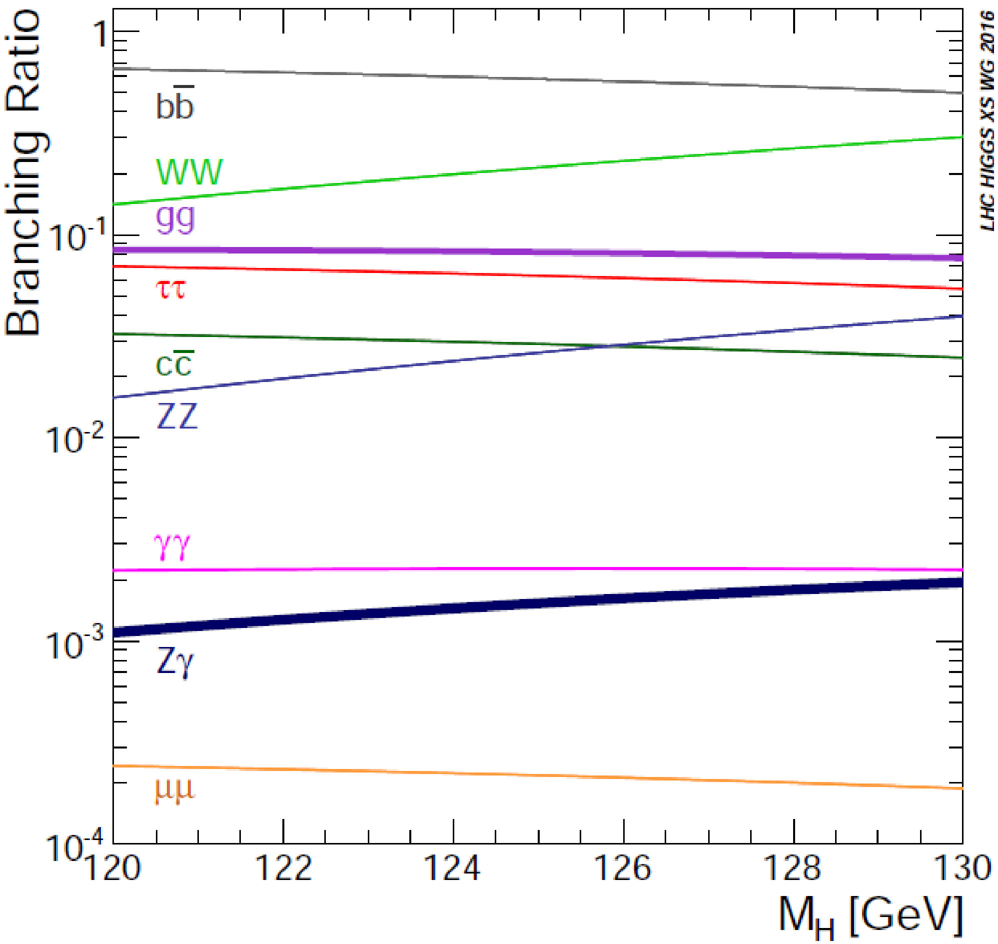
- <u>b jet identification</u>

Many measurements and searches rely on accurate b jet identification

- Top quark: BR(t → b W) ~ 100%



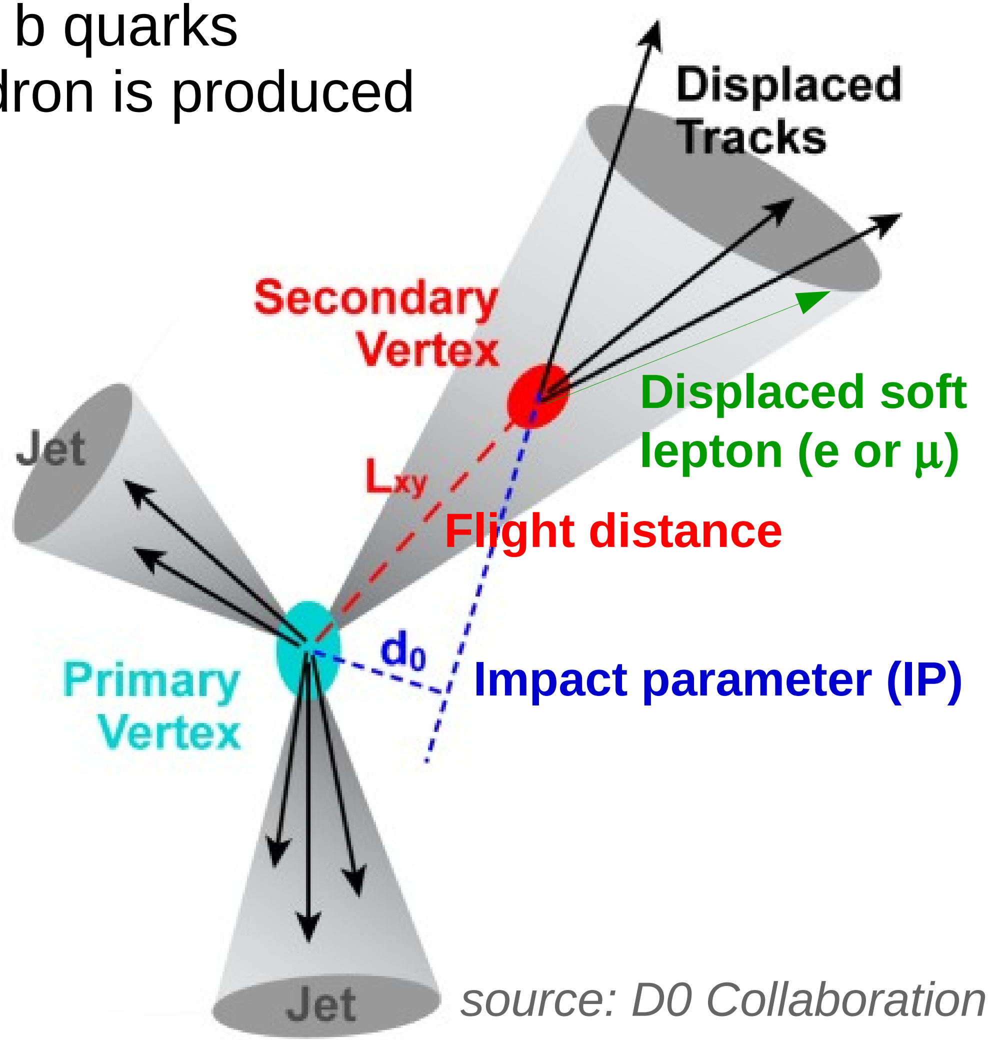
- H-boson: BR(H → bb) ~ 60%



Searches for new physics with third generation quarks or H-bosons in the final state rely on b jet identification

Particle trajectories are most important input for b jet identification

- b jets arise from the hadronization of b quarks
  and during the hadronization a B hadron is produced
- B hadron properties:
  - Relatively large mass [5-6 GeV]
  - Long lifetime [$c\tau$ ~450 μm]
    decay displaced by ~5 mm
    (E=70 GeV)
  - Semi-leptonic decays
    BR (B→ μ or e) ~ 40%

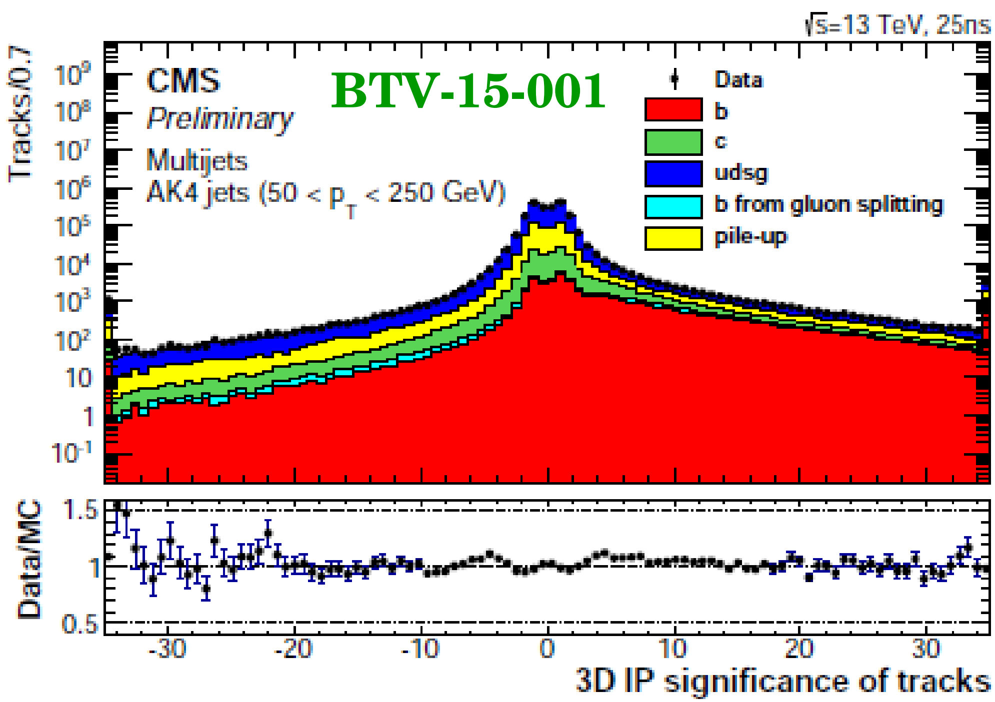


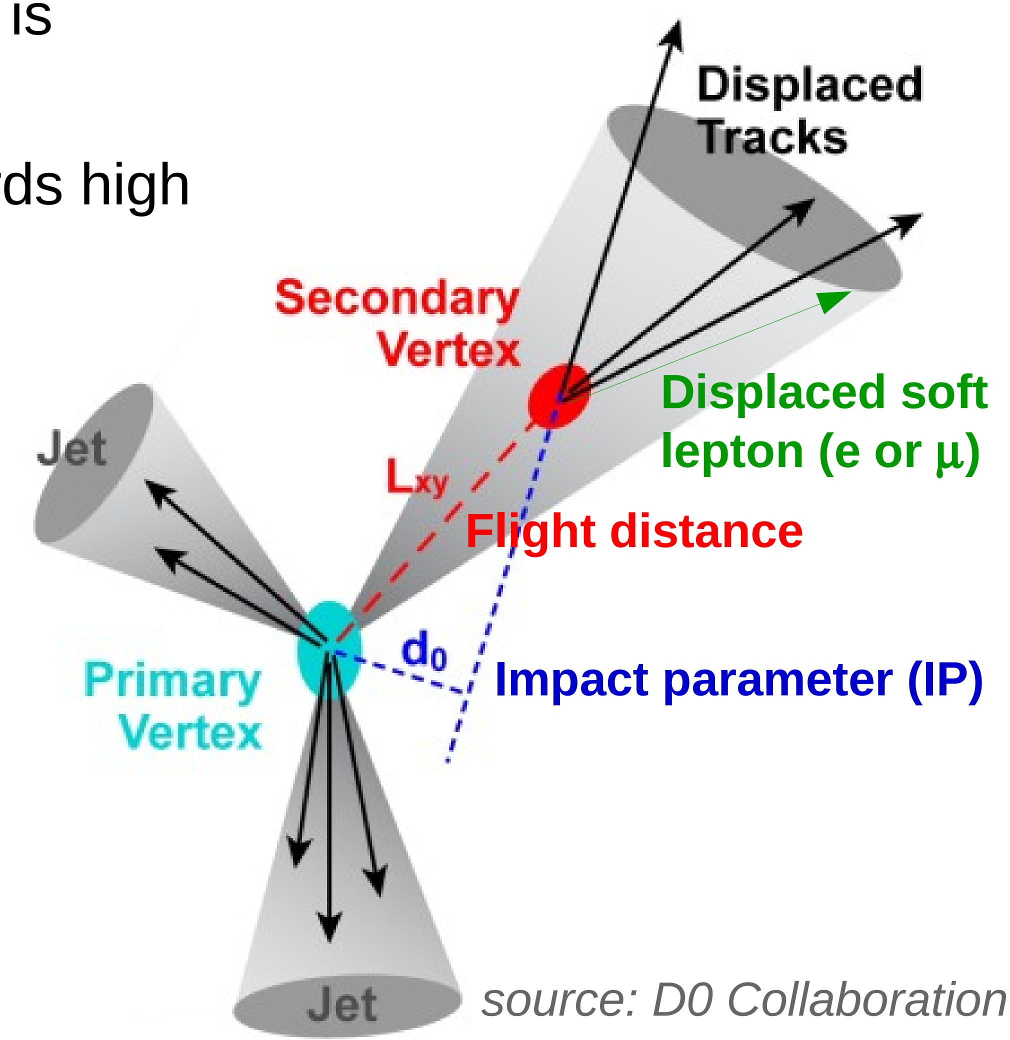
source: D0 Collaboration

Accurate reconstruction of particle trajectories is essential!

The track impact parameter significance discriminates

- For light jets the IP value distribution is symmetric around 0
- For b (and c) jets there is a tail towards high positive IP values



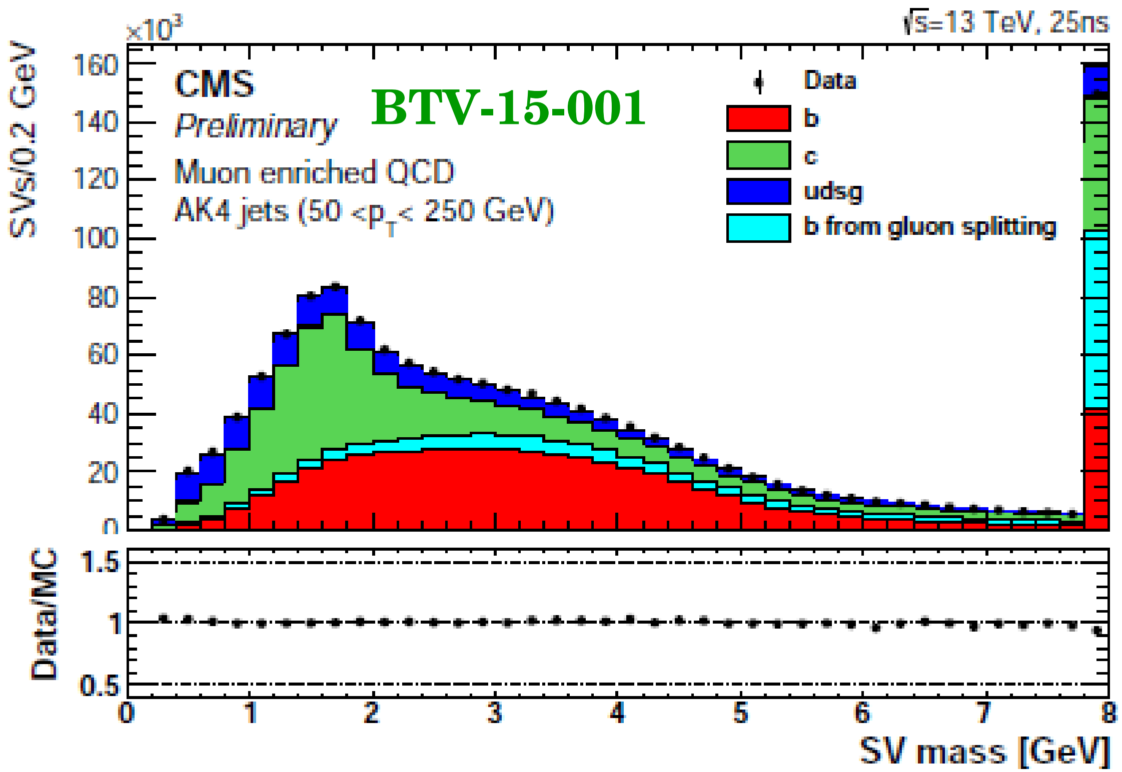


source: D0 Collaboration

IP significance = IP value / IP uncertainty

We can also exploit the information from secondary vertices

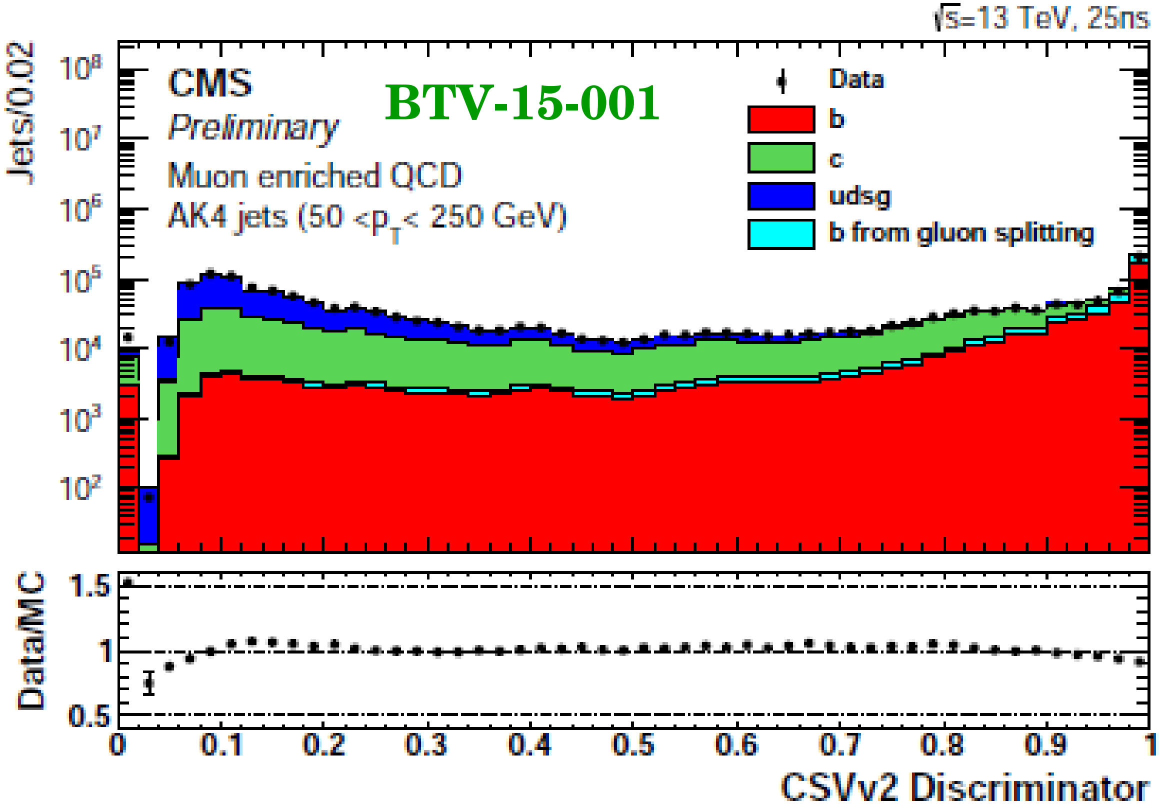
- Optimization for Run 2 → 'inclusive' secondary vertex (SV) finding
  - Going beyond tracks associated to the jet
  - Cluster tracks nearby in space and fit the secondary vertex position



- SV finding efficiency increases with:
  - 10% for b jets
  - 15% for c jets
  - 8% for d,u,s,g jets
- Exploit the properties of the secondary vertex for b-tagging

The CombinedSecondaryVertex (CSV) discriminator

- Several variables are combined with a neural network
- Combines information of displaced tracks as well as fitted secondary vertex

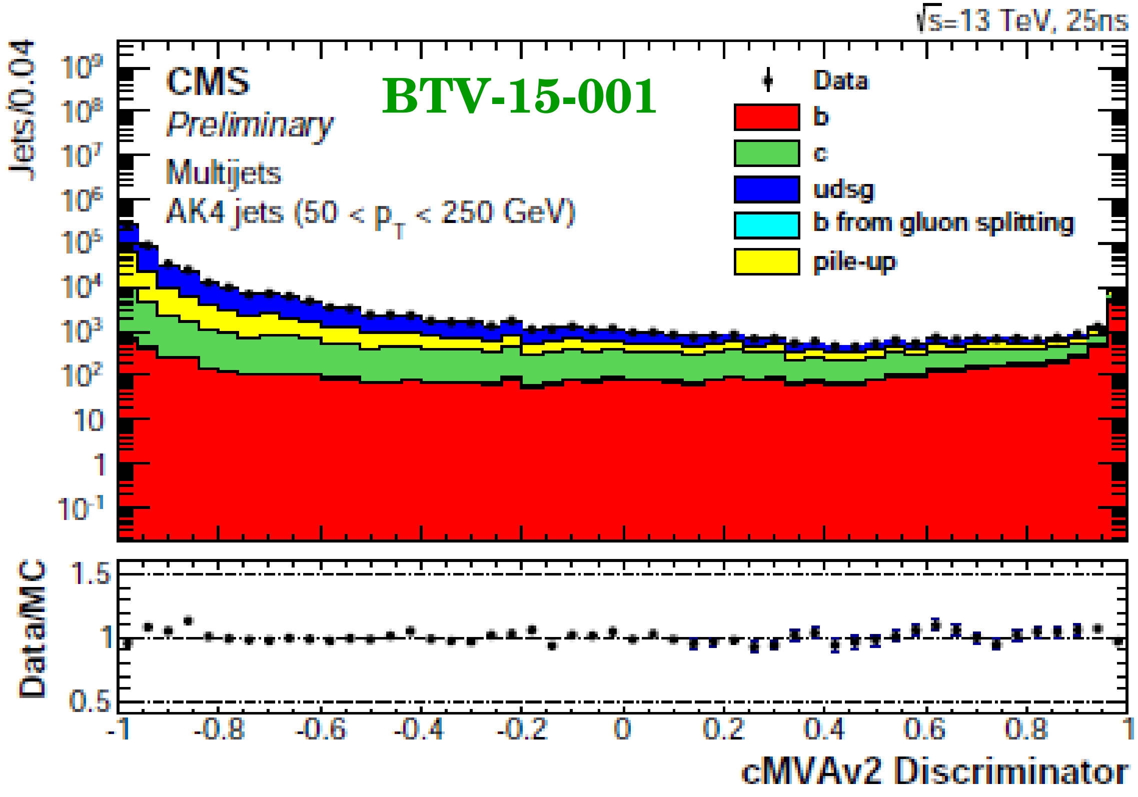


- Training performed in vertex categories
- Combining 22 variables
- Flat jet p_T and η spectrum
- Two trainings:
  - b jets against c jets
  - b jets against d,u,s,g jets
- Linear combination of two discriminators

We can also combine multiple discriminators into a single one

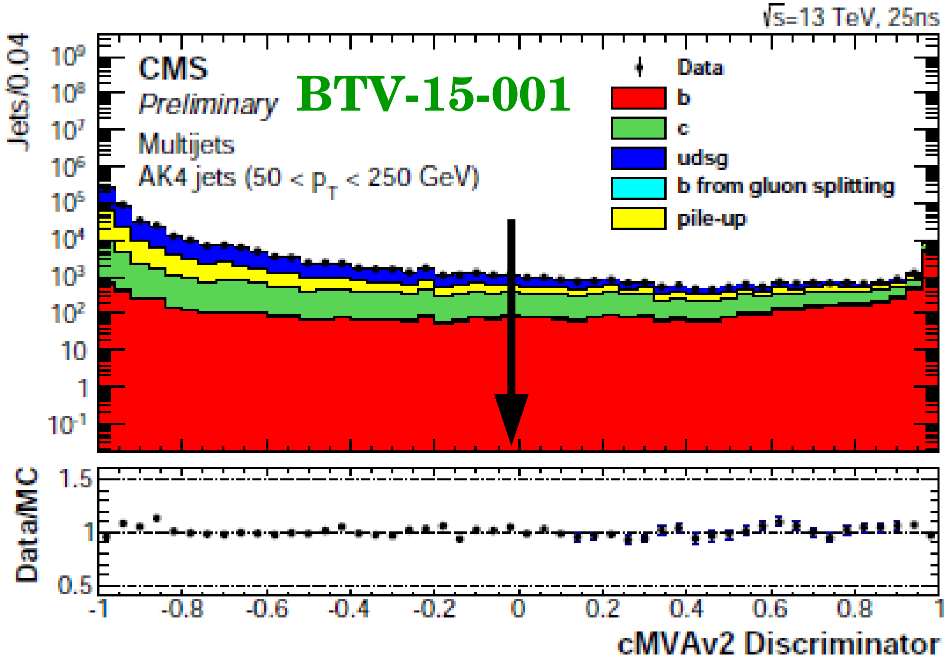
- Combined MVA (cMVA) = combination of various discriminators:
  - Discriminators exploiting the presence of muons or electrons in the jet
  - Two types of JetProbability discriminator
  - Two types of CSV

- Combining 6 discriminators with a Boosted Decision Tree



The performance of the algorithms is quantified and compared

- For each cut on a discriminator we can determine:
  - the b-tagging efficiency, ε_b
  - the probability to misidentify a non-b jet as a b jet, $\varepsilon_{non\text{-}b}$
- Perform a scan over all discriminator thresholds
- For each threshold, plot ε_b vs $\varepsilon_{non\text{-}b}$ → Receiver Operating Characteristic curve (ROC curve)



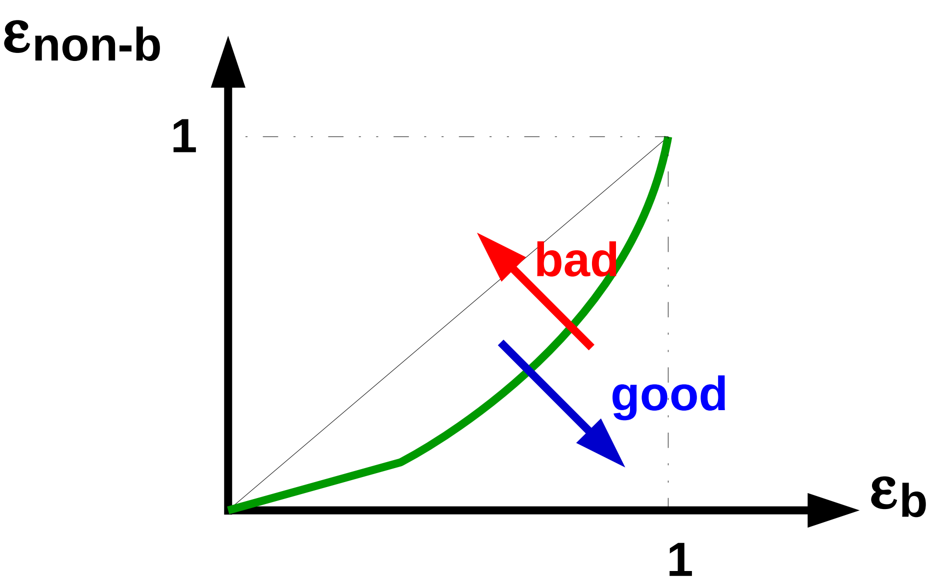
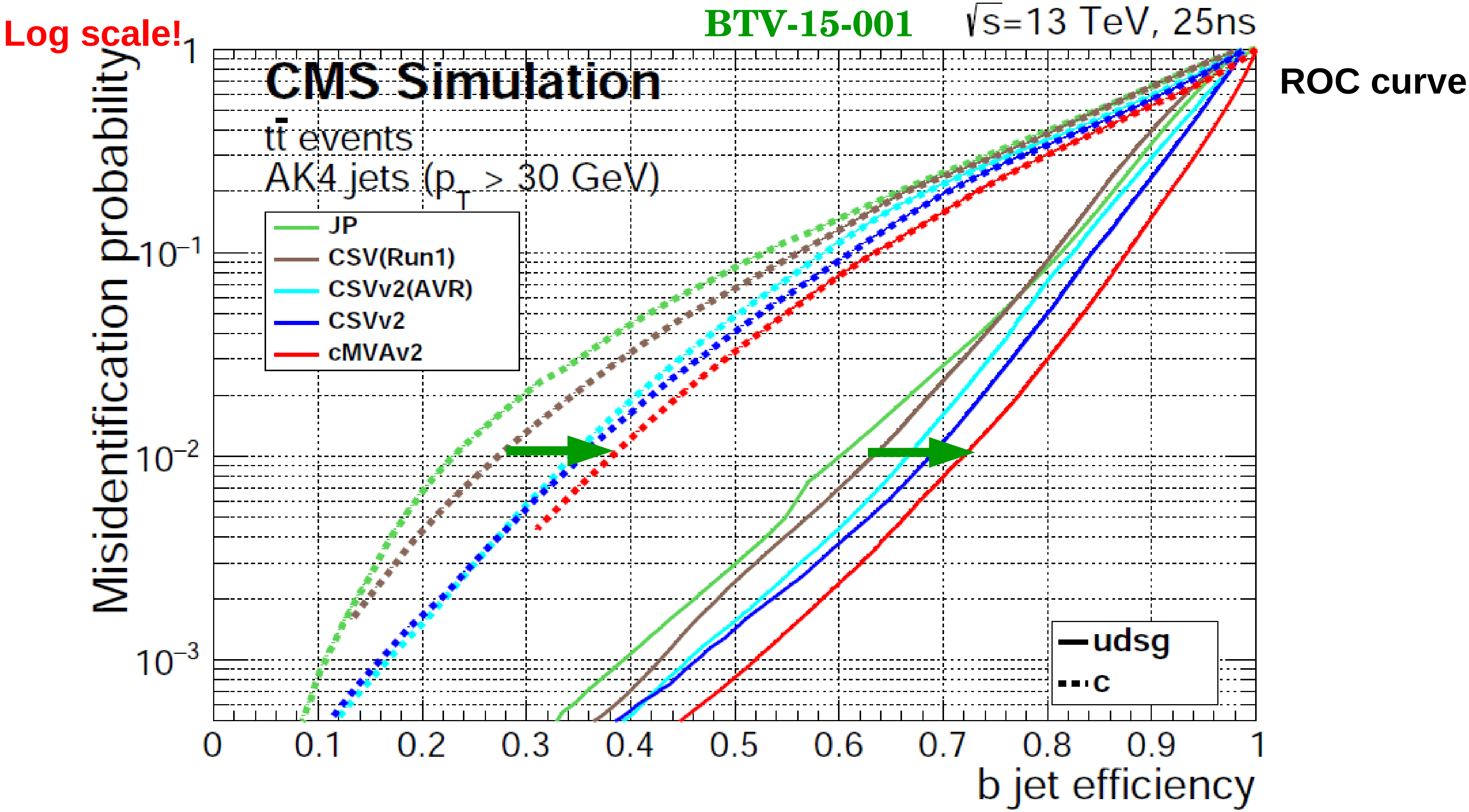


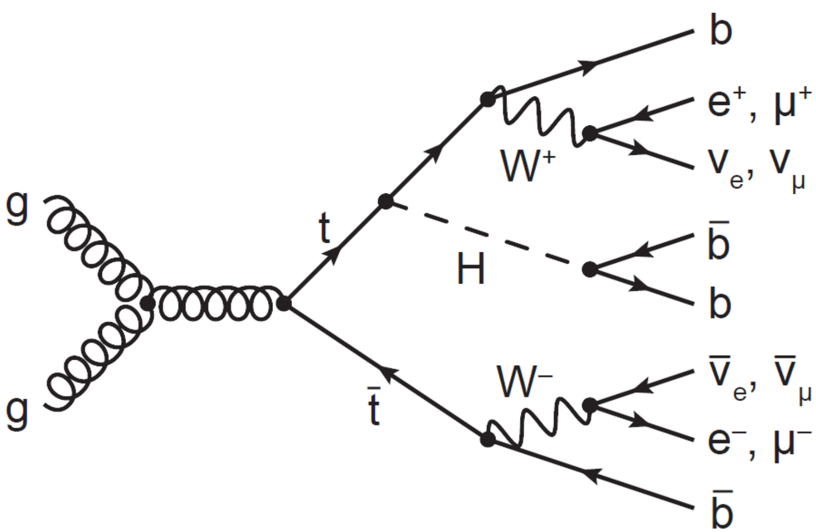
Illustration ROC curve

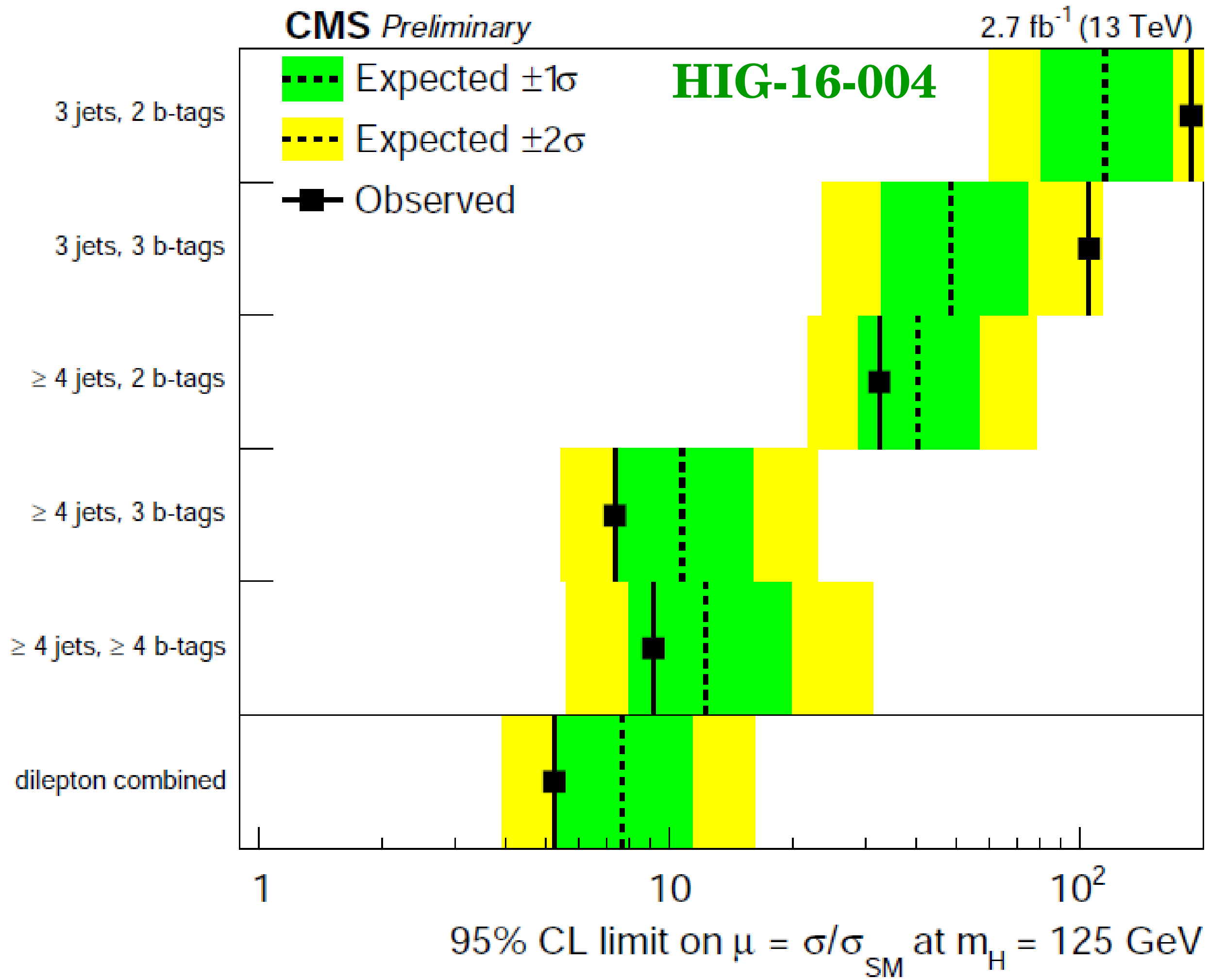
A clear improvement with respect to the performance in LHC Run 1!



For the same misidentification probability (1% for u,d,s,g jets): a relative improvement of >10% in b jet identification efficiency

Impact of improved b jet identification on the observation of the ttH process





- Number of jets identified as b is crucial ingredient!
- Categories with higher number of b jets are more sensitive

In general with four b-tagged jets, a 10% higher b-tagging efficiency corresponds to 60% more ttH signal

Summary & Outlook

- About 10% better performing b jet identification algorithms
- We have also developed tools for b-tagging in boosted topologies
  → double-b tagger public material expected to be released next week
- Charm identification algorithm has been developed
  → public material expected to be released in July
- Paper to be submitted this Fall with the latest greatest status for identification of b, c and boosted b jets

JINST
PUBLISHED BY IOP PUBLISHING FOR SISSA MEDIALAB
RECEIVED: November 19, 2012
REVISED: February 11, 2013
ACCEPTED: February 21, 2013
PUBLISHED: April 15, 2013
LHC REFERENCE VOLUME

Identification of b-quark jets with the CMS experiment

The CMS collaboration

E-mail: cms-publication-committee-chair@cern.ch

ABSTRACT: At the Large Hadron Collider, the identification of jets originating from b quarks is important for searches for new physics and for measurements of standard model processes. A variety of algorithms has been developed by CMS to select b-quark jets based on variables such as the impact parameters of charged-particle tracks, the properties of reconstructed decay vertices, and the presence or absence of a lepton, or combinations thereof. The performance of these algorithms has been measured using data from proton-proton collisions at the LHC and compared with expectations based on simulation. The data used in this study were recorded in 2011 at $\sqrt{s} = 7$ TeV for a total integrated luminosity of 5.0 fb^{-1}. The efficiency for tagging b-quark jets has been measured in events from multijet and t-quark pair production. CMS has achieved a b-jet tagging efficiency of 85% for a light-parton misidentification probability of 10% in multijet events. For analyses requiring higher purity, a misidentification probability of only 1.5% has been achieved, for a 70% b-jet tagging efficiency.

KEYWORDS: Large detector-systems performance; Pattern recognition, cluster finding, calibration and fitting methods; Performance of High Energy Physics Detectors

ARXIV EPRINT: 1211.4462

2013 JINST 8 P04013

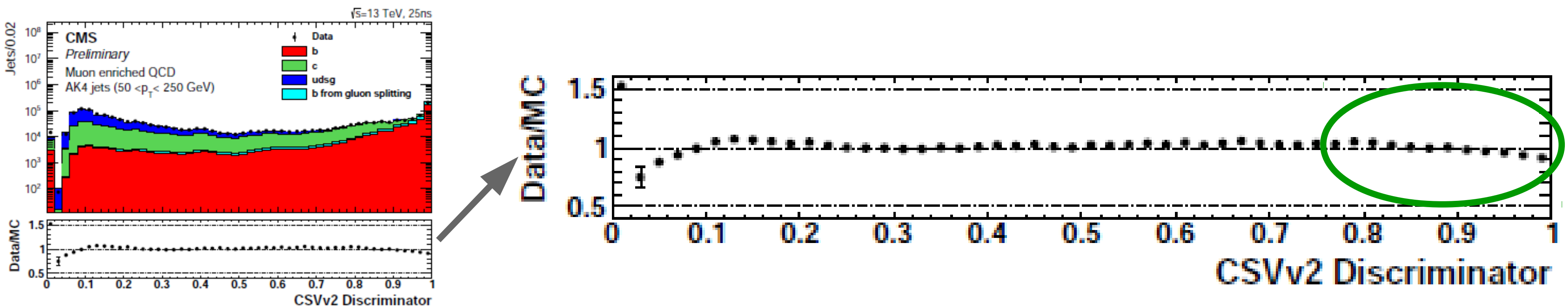
Season 2 expected soon

Lots of new developments are needed in the near and far future

Additional material

- Efficiency measurements from data (scale factors)
  - Examples of various methods
- b tagging in boosted topologies
  - Scale factors measurements
  - Double-b tagger
- Phase 1 and HL-LHC
- Systematic uncertainties

Data and simulation does not agree perfectly → scale factors are needed



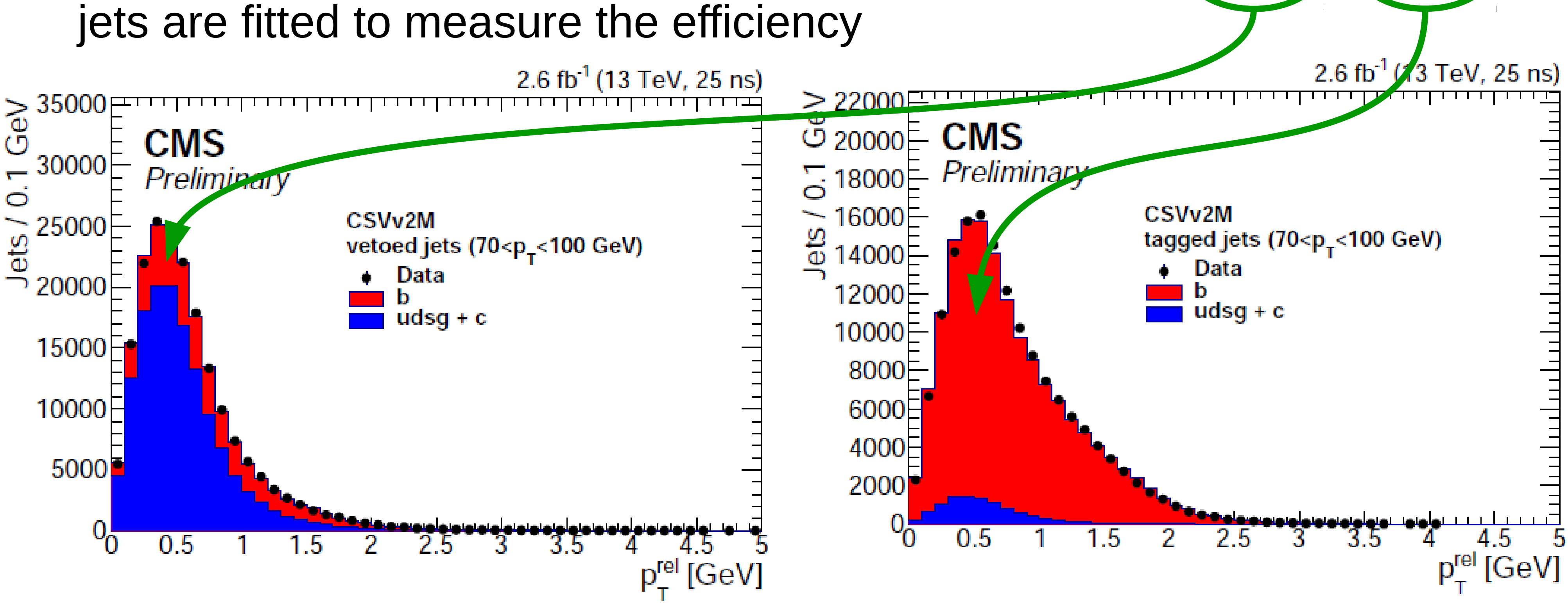
- Scale factor SF = $\varepsilon_X^{data} / \varepsilon_X^{MC}$
  - Depends on jet flavour X
  - Measured for 3 discriminator thresholds corresponding to a misidentification probability of 10% (loose), 1% (medium) and 0.1% (tight)

- Various methods to measure efficiency from data, general strategy
  - SF_b → select data events enriched in b jets
  - SF_l → select data events enriched in light jets

PtRel method fits the distribution of muon p_T relative to the jet axis

- Select multijet events with at least 1 jet containing a muon (muon-jet); request 2nd jet in the event away from the first one passing a b-tag cut
- Muon-jet is either b-tagged or b-vetoed
- Distributions of muon p_T for tagged and vetoed jets are fitted to measure the efficiency

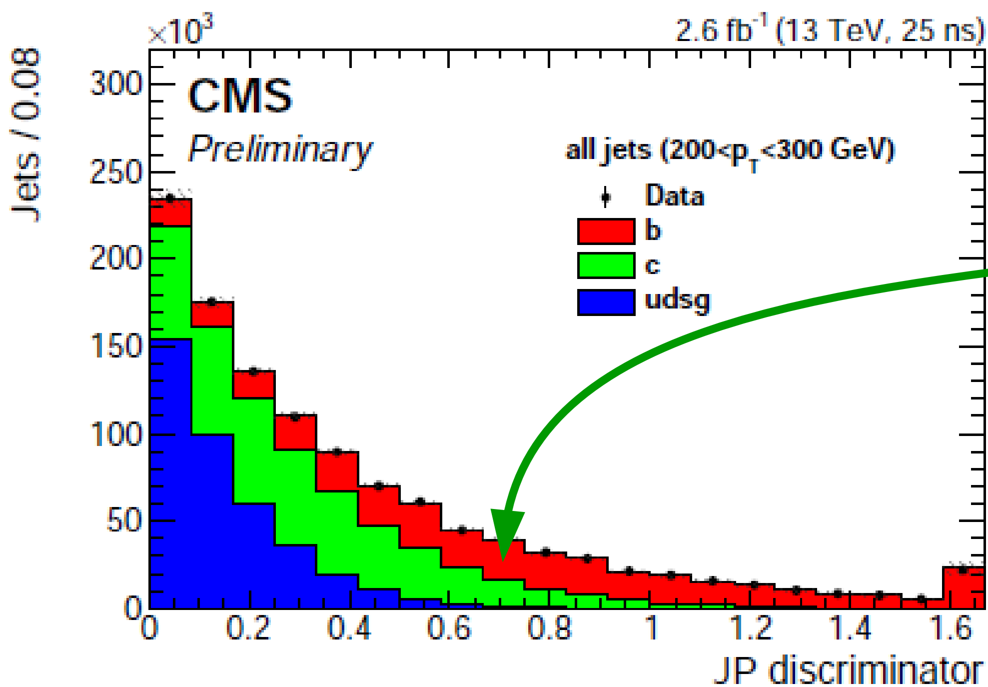
$$\epsilon_b = \frac{N_b^{tagged}}{N_b^{vetoed} + N_b^{tagged}}$$

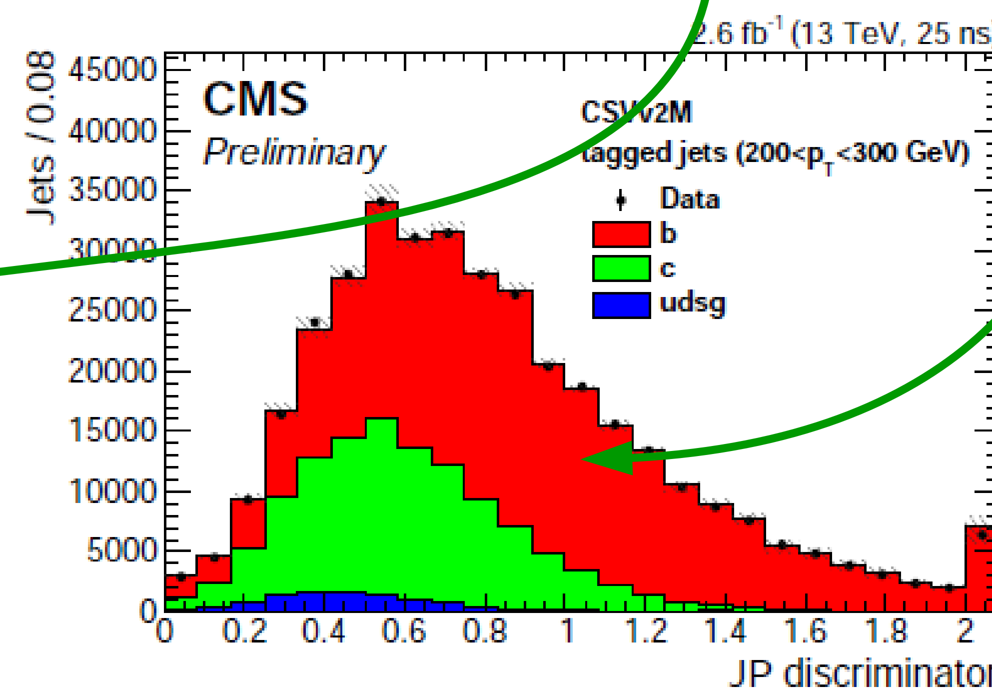


LifeTime (LT) method uses similar concept, fitting JP distribution

- Select multijet events with at least 1 jet containing a muon (muon-jet); request 2nd jet in the event away from the first one
- We consider all muon-jets and tagged muon-jets
- Distributions of the muon-jet JetProbability are fitted to measure the efficiency:

$$\epsilon_b = \frac{N_b^{tagged}}{(N_b^{vetoed} + N_b^{tagged})}$$





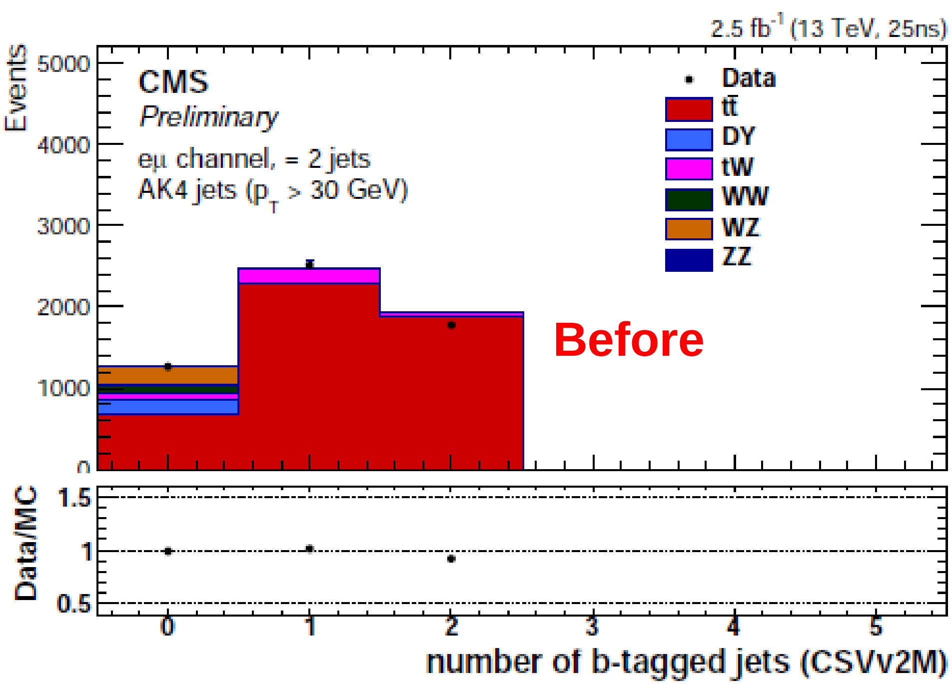
A simple and robust method, relying only on fractions of events

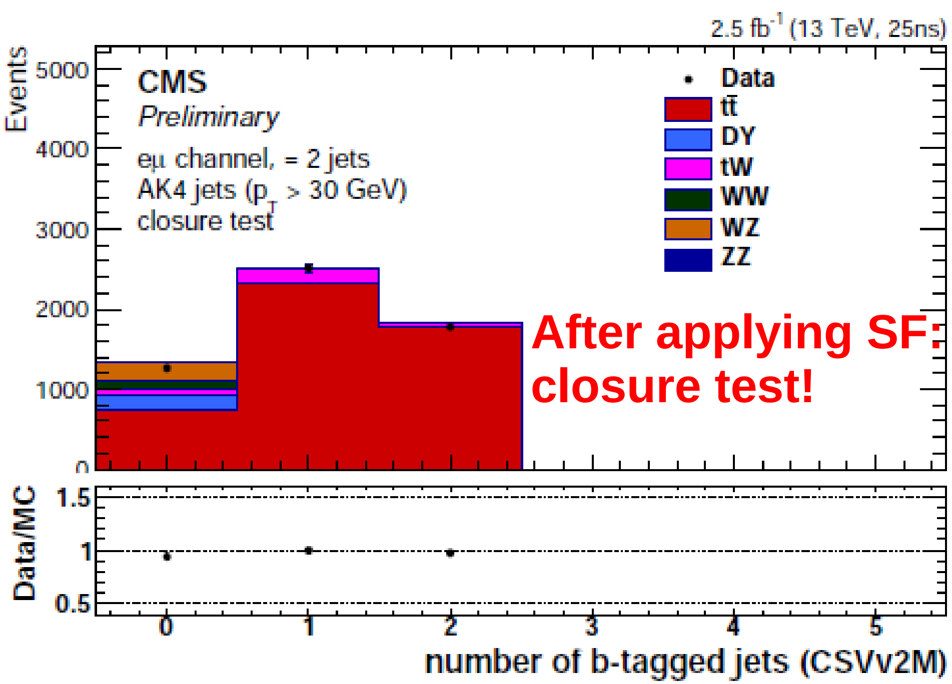
- Select top-quark pair events with 2 jets, 1 electron and 1 muon
- Efficiency obtained from the fraction of events with 2 b-tagged jets in data as:

$$\varepsilon_b = \sqrt{\frac{F_{2btag} - F_{2btag}^{non-bjet}}{f_{2b}}}$$

MC fraction of events with 2 true b jets

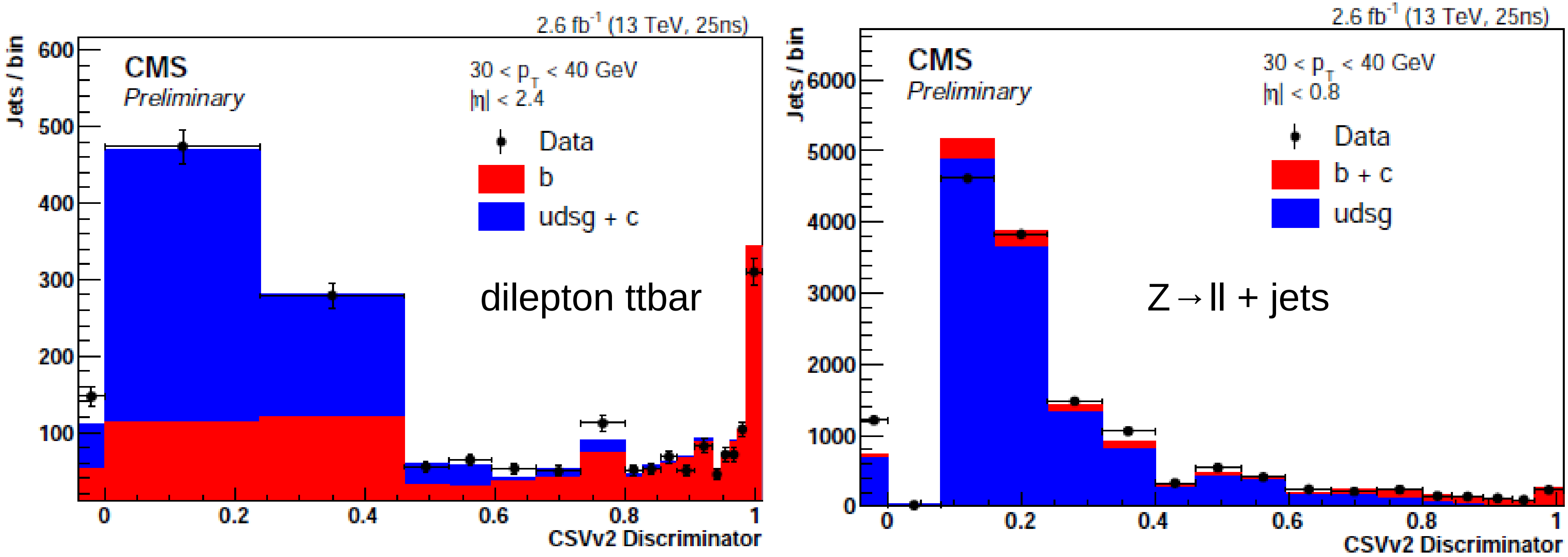
MC fraction of events with 2 tagged jets with at least 1 not a true b jet





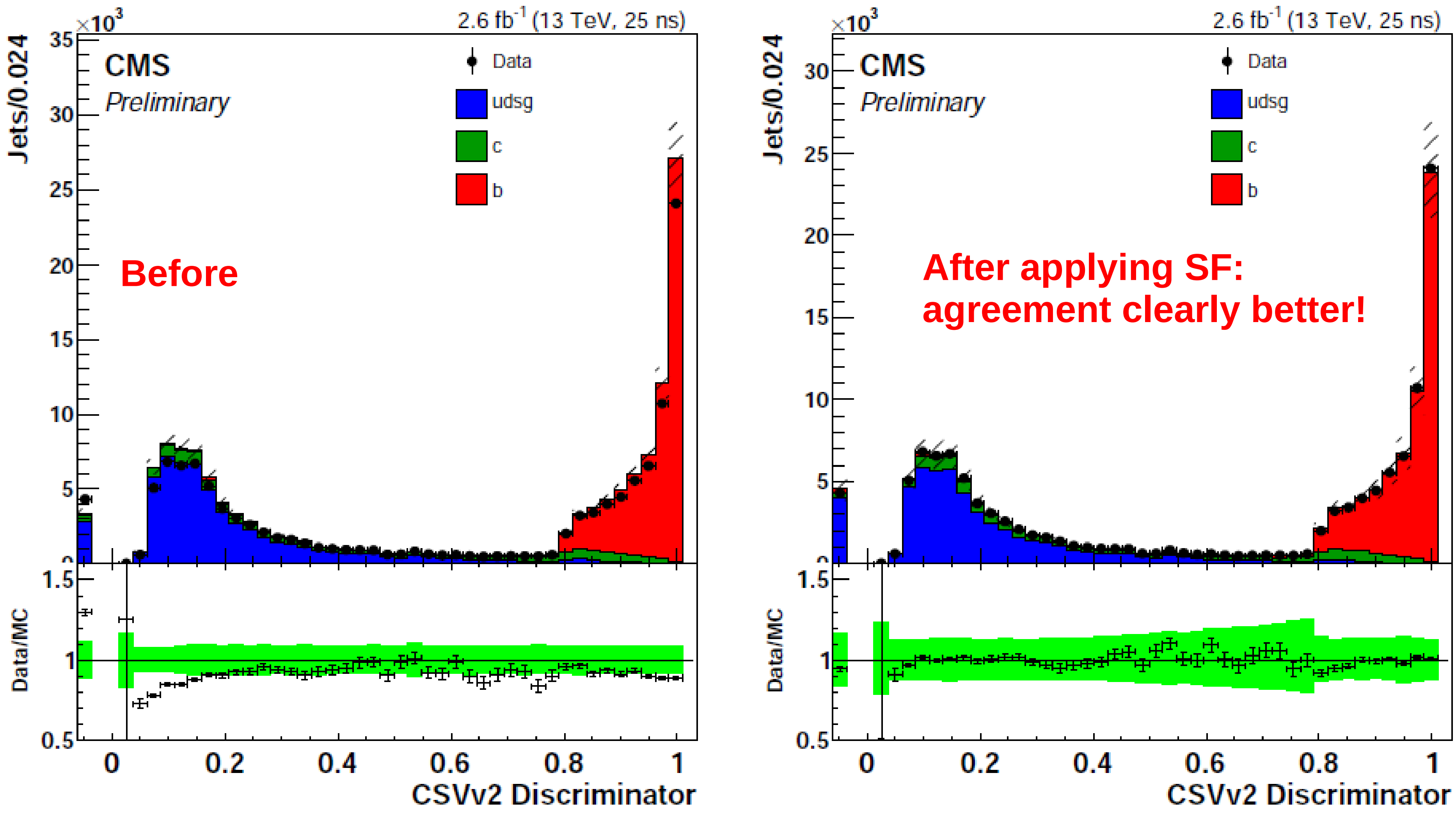
More ambitious technique aims to correct the discriminator distribution

- Tag & probe method, iterative procedure
- SF_b measured in dilepton ttbar events after subtracting non-b background → iterate until convergence
- SF_l measured in Z → ll + jets events after subtracting b and c background → iterate until convergence

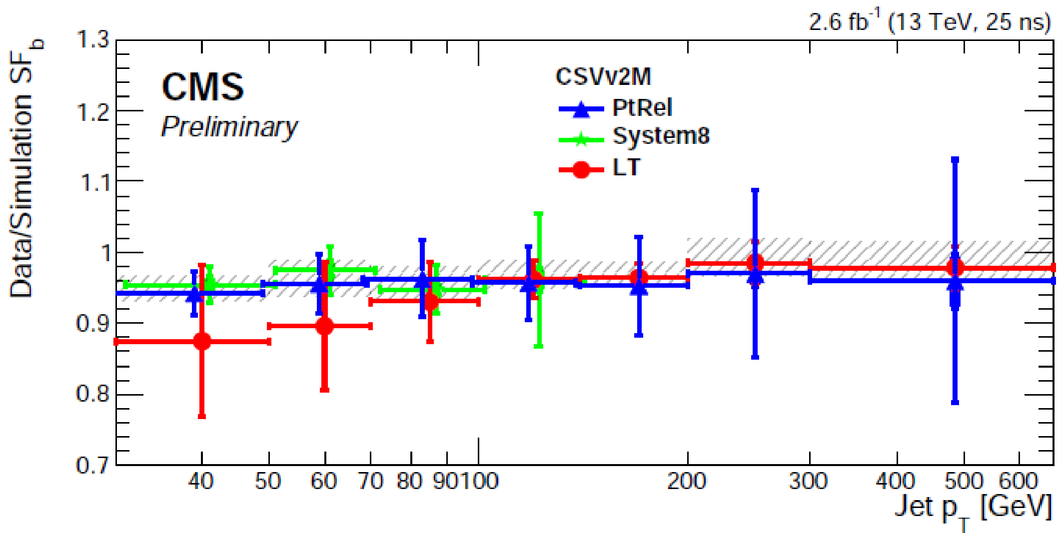


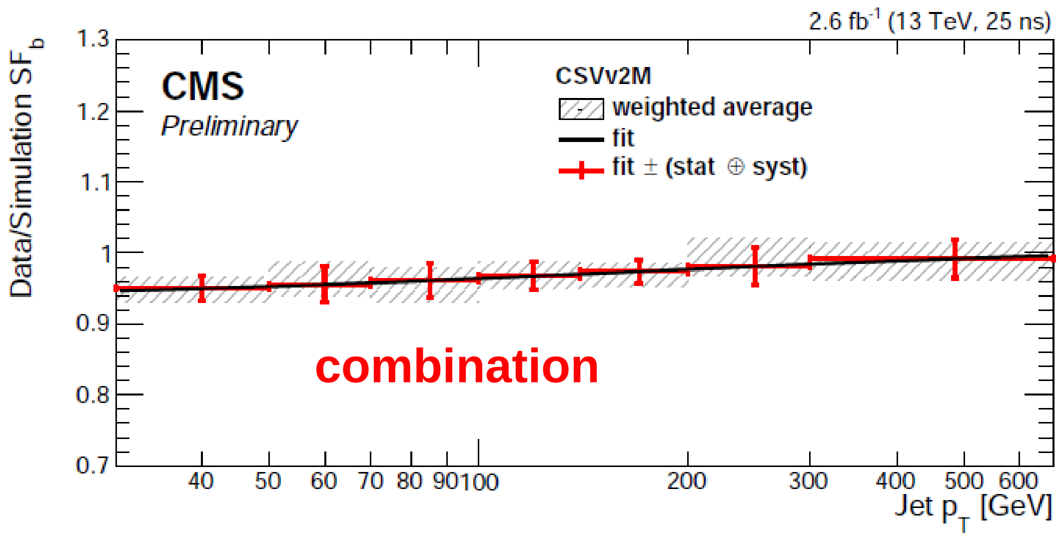
The measured scale factors are stable after 3 iterations

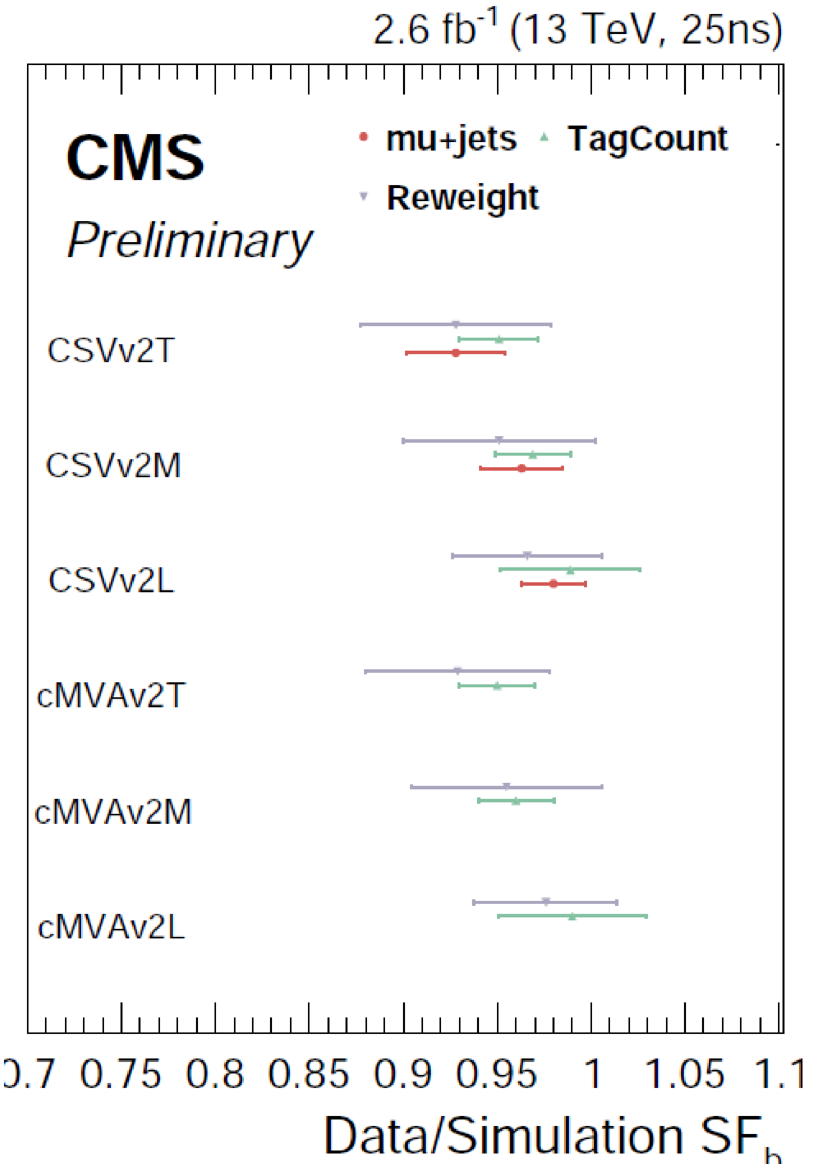
- Closure test of this method performed in ttbar events with 1 lepton, exactly 4 jets of which 2 are b-tagged



The various measurements agree; a combination is performed







The precision on the scale factor for b jets is 2 to 4%

Negative tag method is used to measure the mistag scale factors

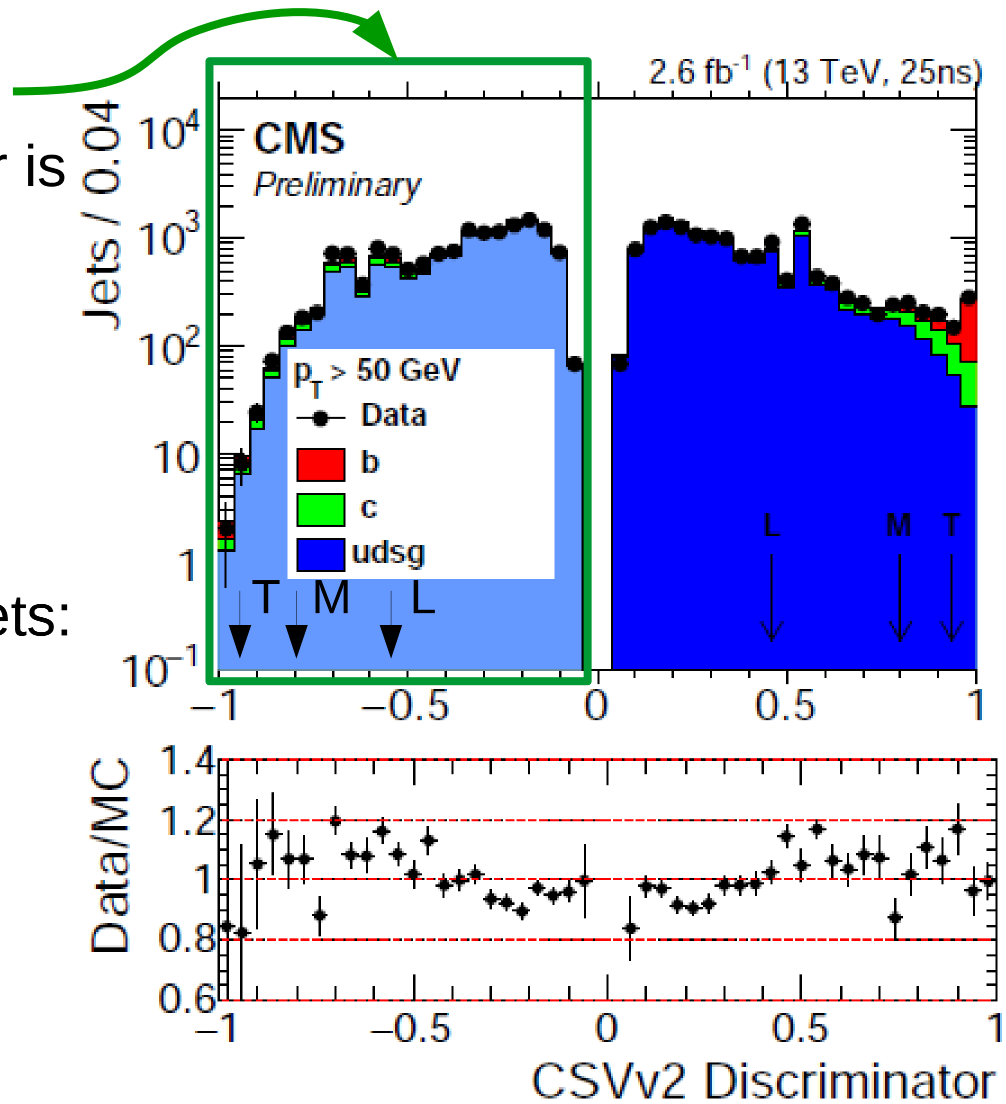
- Using only tracks with negative IP and negative decay length for the secondary vertex a negative tagger is defined
- Similarly for the positive tagger
- To first order, should be symmetric for light-flavour jets

  → derive ε_l from negative-tagged jets:

$$\varepsilon_l^{data} = \varepsilon^{neg,data} R_l$$

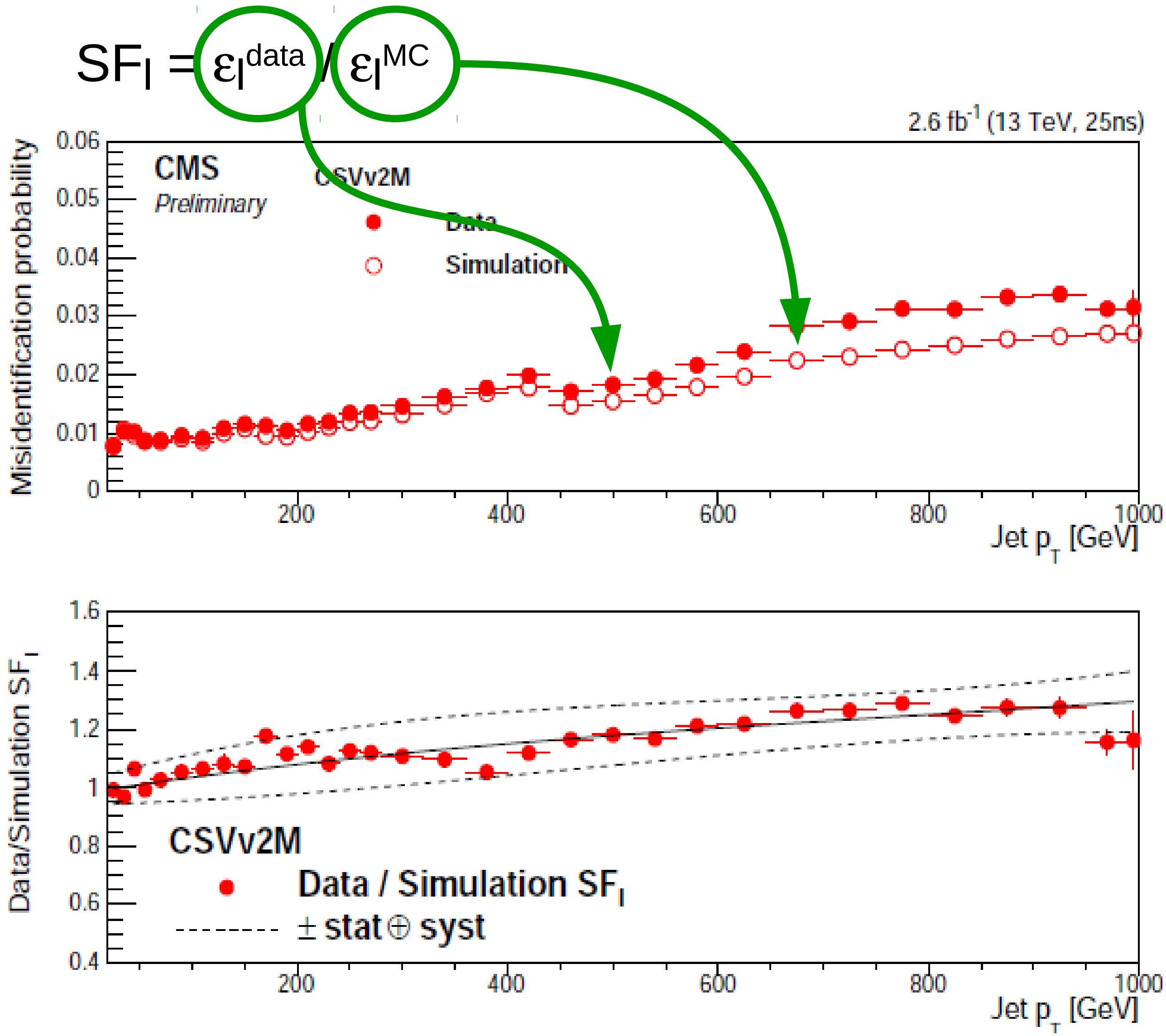
$$R_l = \varepsilon_l^{MC} / \varepsilon^{neg,MC}$$

- R_l corrects for asymmetry and contamination of non-l jets



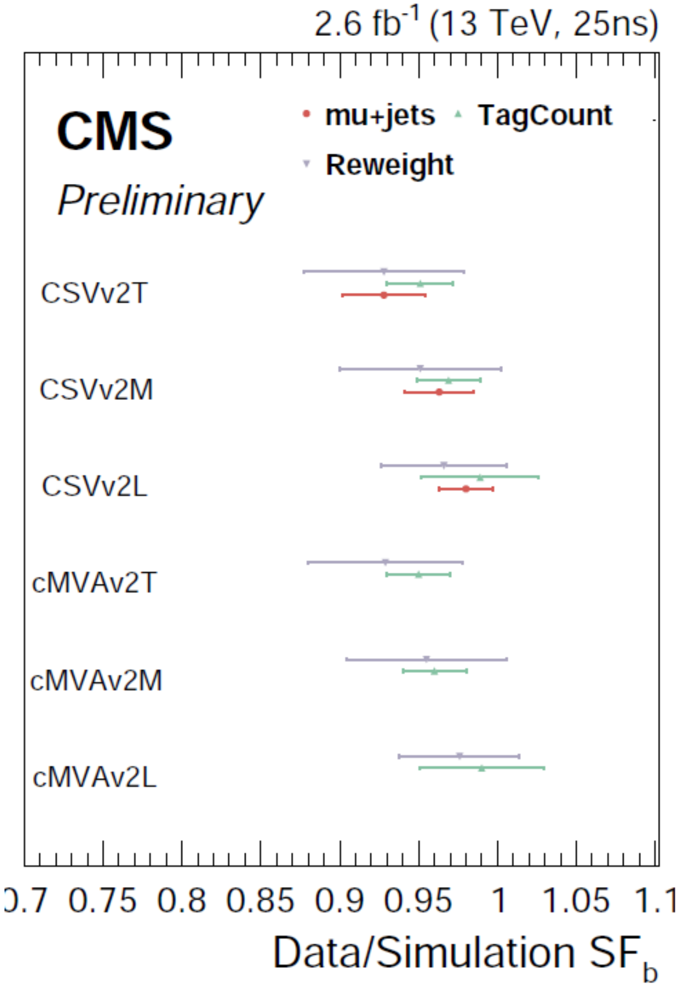
The mistag scale factor is typically > 1, i.e. more mistags in data

$SF_l = \varepsilon_l^{data} / \varepsilon_l^{MC}$



The precision on the mistag scale factor is 5 to 10%

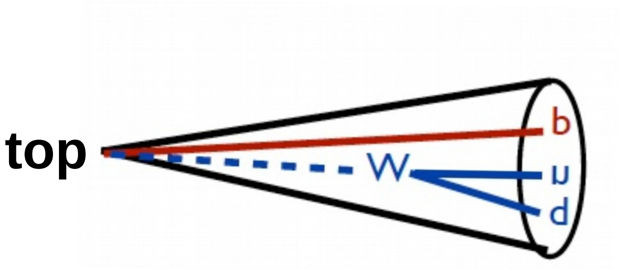
To apply b-tagging in physics analyses, scale factors are essential to correct for data/MC differences

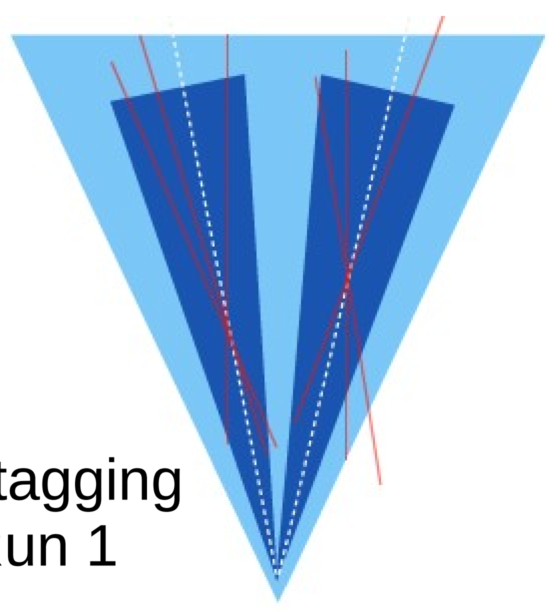


- Scale factor SF = $\varepsilon_X^{data} / \varepsilon_X^{MC}$ are derived with various methods
- Uncertainties of 2 – 4% achieved for SF_b and 5 – 10% for SF_l
- Dominated by systematic uncertainties
  → precision measurements require careful assessment of these uncertainties

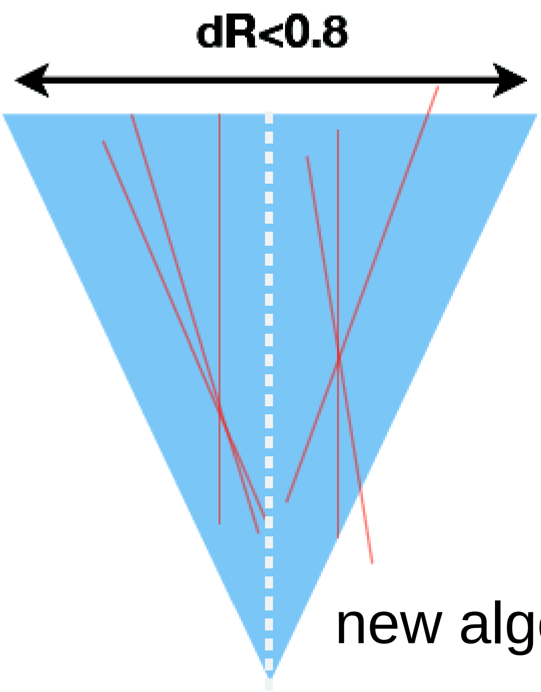
b jet identification in boosted topologies was already used in Run 1

- The produced particles may be boosted due to high energy
- Decay products are clustered in a single 'fat' jet
- b-tagging for boosted objects, two possibilities:





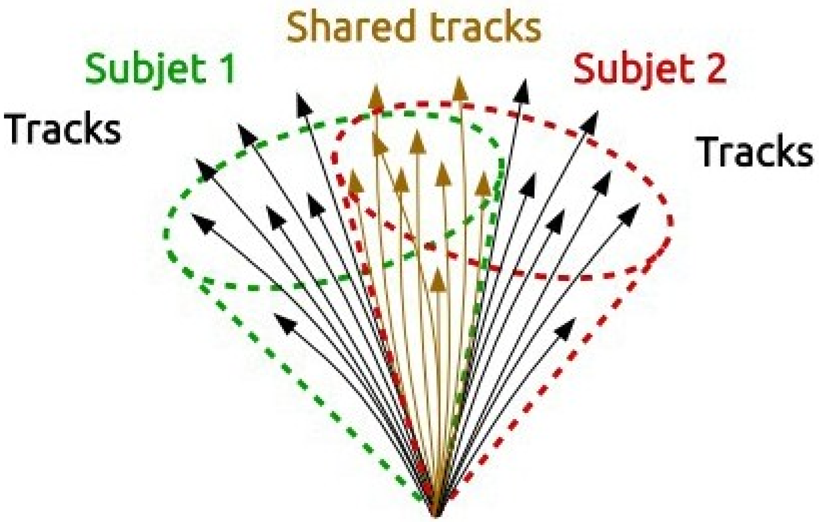
Subjet b-tagging used in Run 1

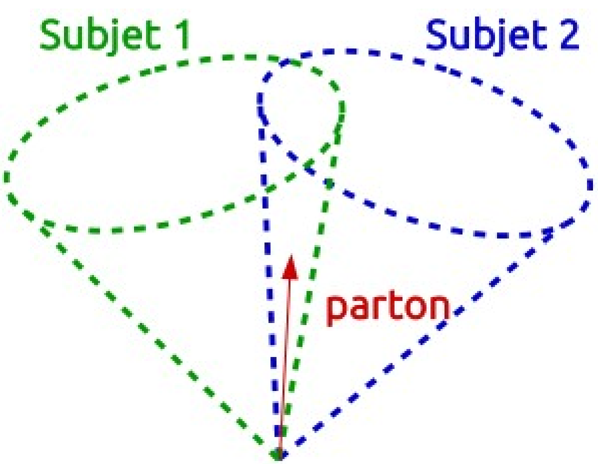


Fat jet b-tagging new algorithm developed!

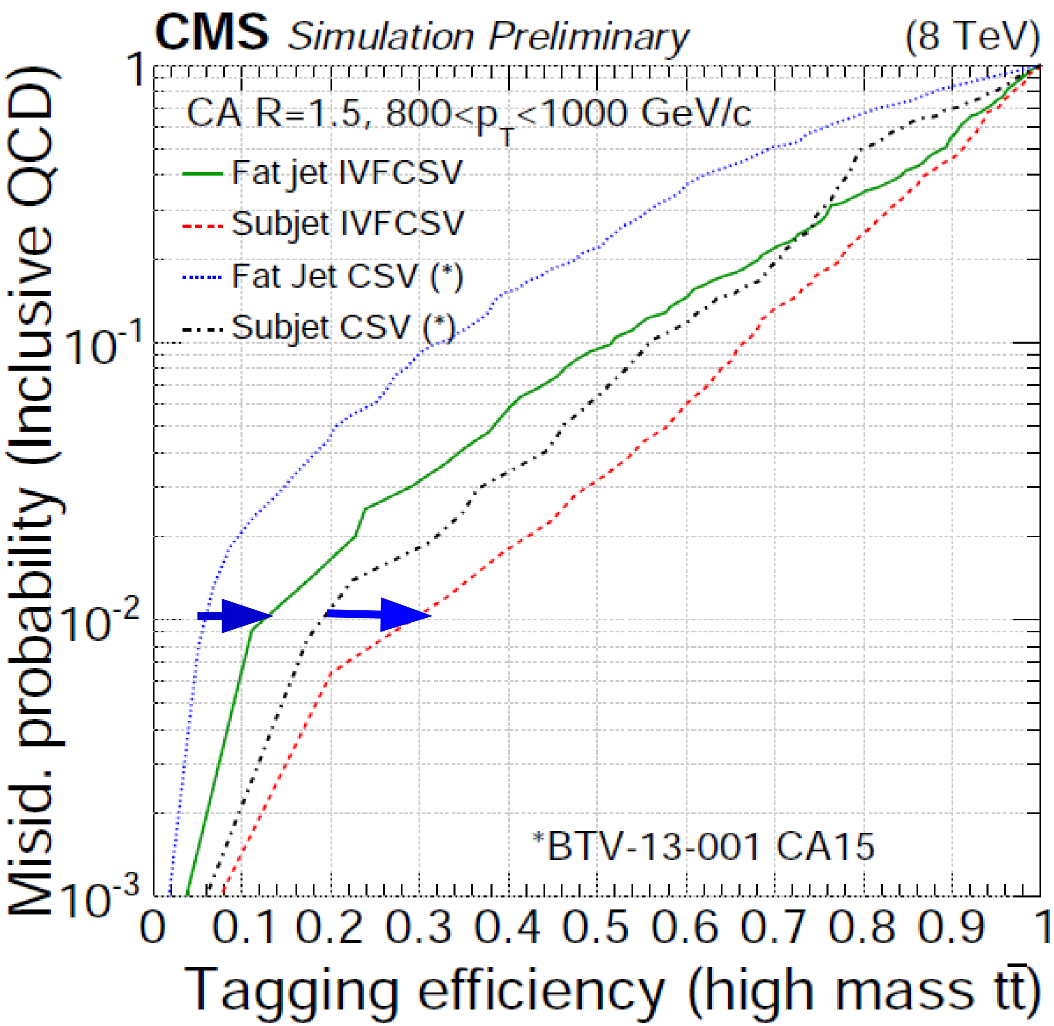
b-tagging in boosted topologies was optimized for Run 2

- Tracks used for b-tagging after subjet reconstruction
  - Run 1: tracks in fixed cone-size → track can be shared resulting in ambiguities
  - Run 2: PF particles clustered in the subjet
- Jet-flavour assignment
  - Run 1: ΔR matching with the parton
  - Run 2: virtual clustering of b and c hadrons in the subjet
- Benefiting also from improvements for regular b jet identification





Optimized b-tagging algorithms perform better in boosted topologies

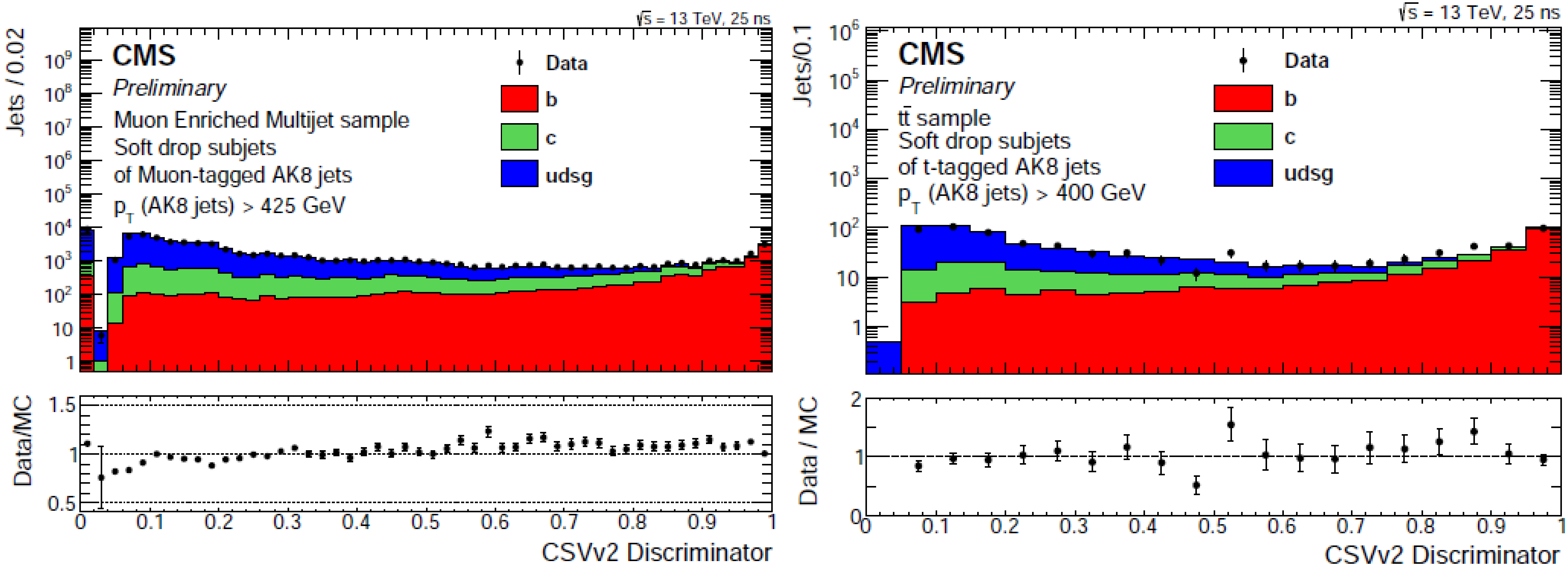


- Boosted hadronic top-quark jets in high-mass ttbar events
  - Hadronic top quark identified by matching fat jet ($\Delta R<1.5$) with generator $t \rightarrow bqq$ decay
  - HEPTopTagger using CA15 jet clustering algorithm
- Subjet b-tagging works better than fatjet b-tagging for this topology

For the same misidentification probability (1% for u,d,s,g jets): a relative improvement of ~50% in b jet identification efficiency

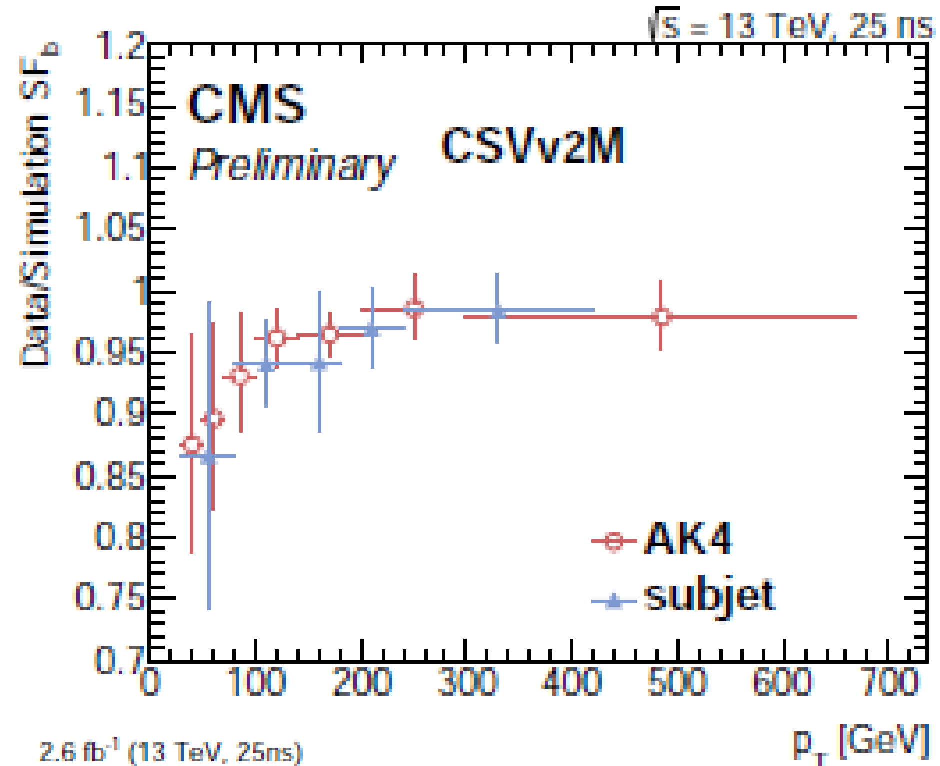
The discriminator for subjets in boosted topologies looks good

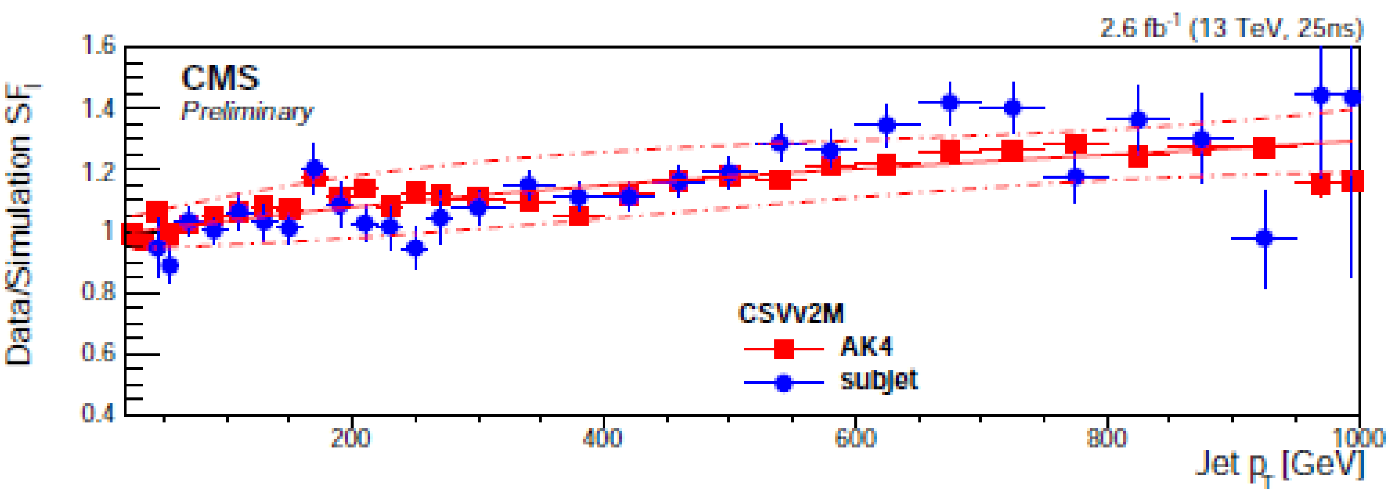
- The agreement between data and MC of the input variables is checked using muon-enriched multijet ($g \to bb$) events and boosted hadronic top-quark jets



The scale factors for the regular jets agree with those for subjets

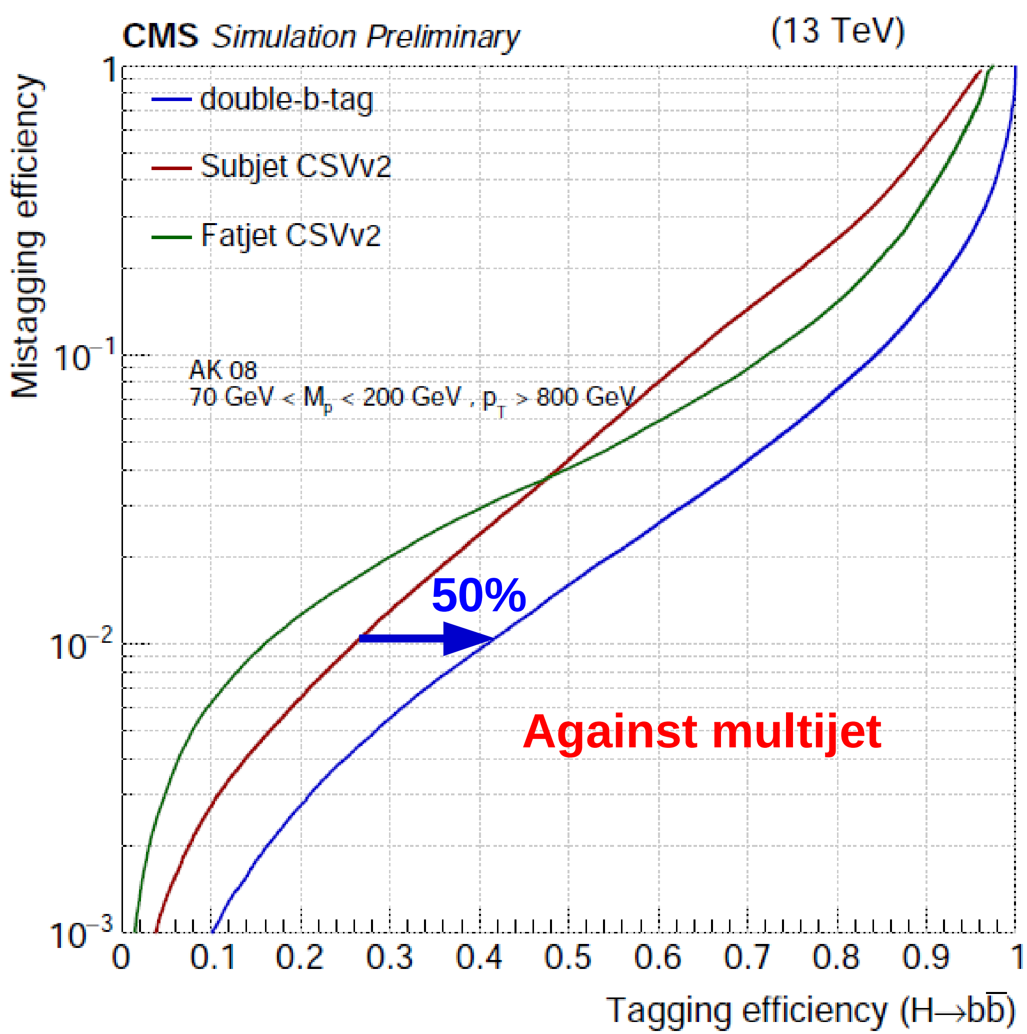
- SF_b measured with LifeTime method
- SF_l measured with negative tag method
- Good agreement with SF for regular jets



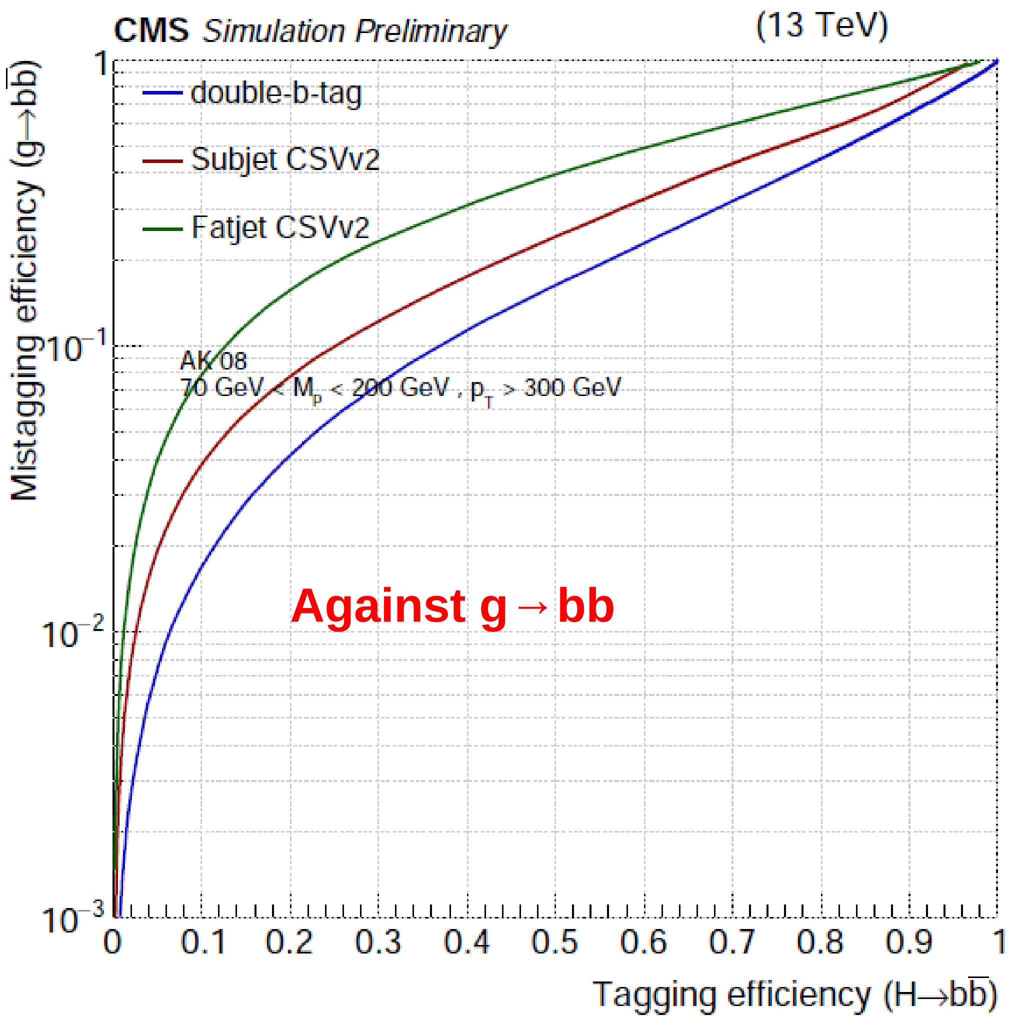


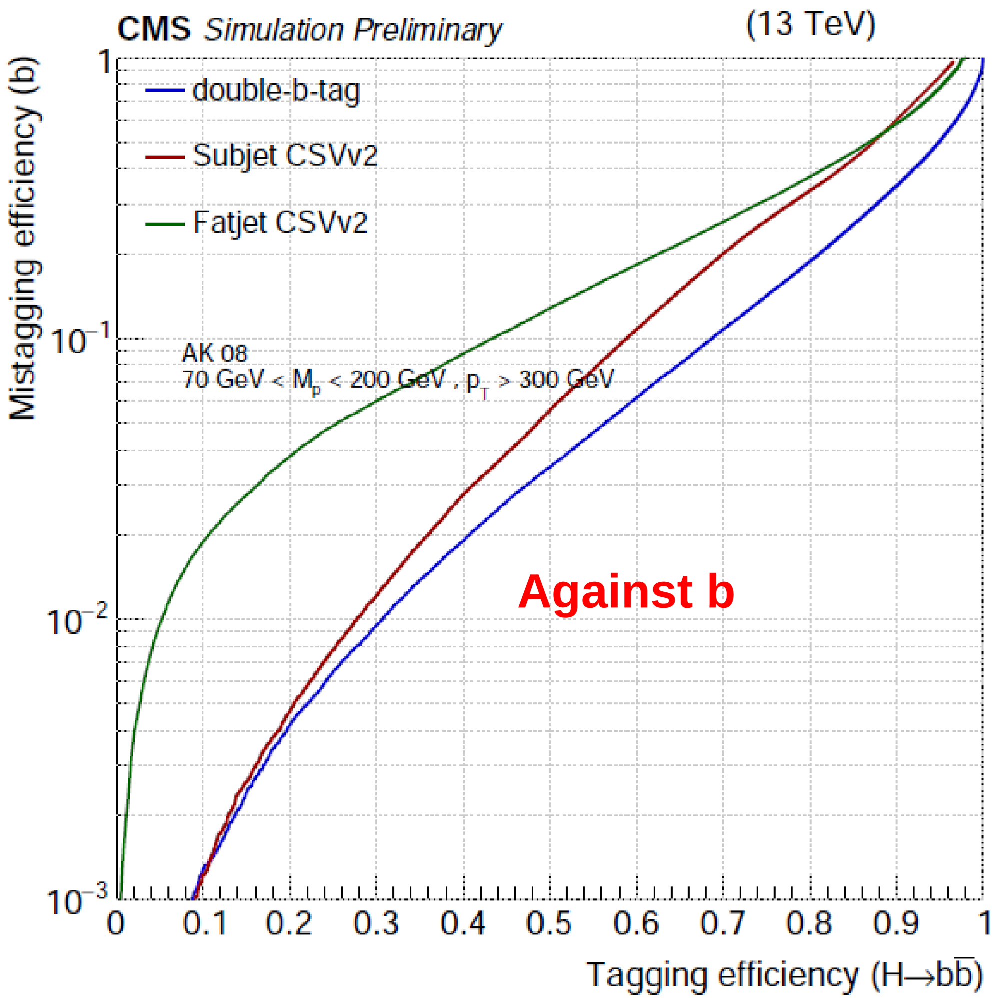
The double-b tagger is more efficient to identify boosted X → bb objects

- Double-b tagger aims to identify fat jets containing two b hadrons
- Boosted Decision Tree:
  - Signal: spin-0 radion → HH → 4b combining various mass points
  - Background: inclusive multijet
- Input variables based on track information, secondary vertices, leptons and minimum subjet CSVv2 value
- Performance compared with CSVv2 for subjet and fat jet b-tagging



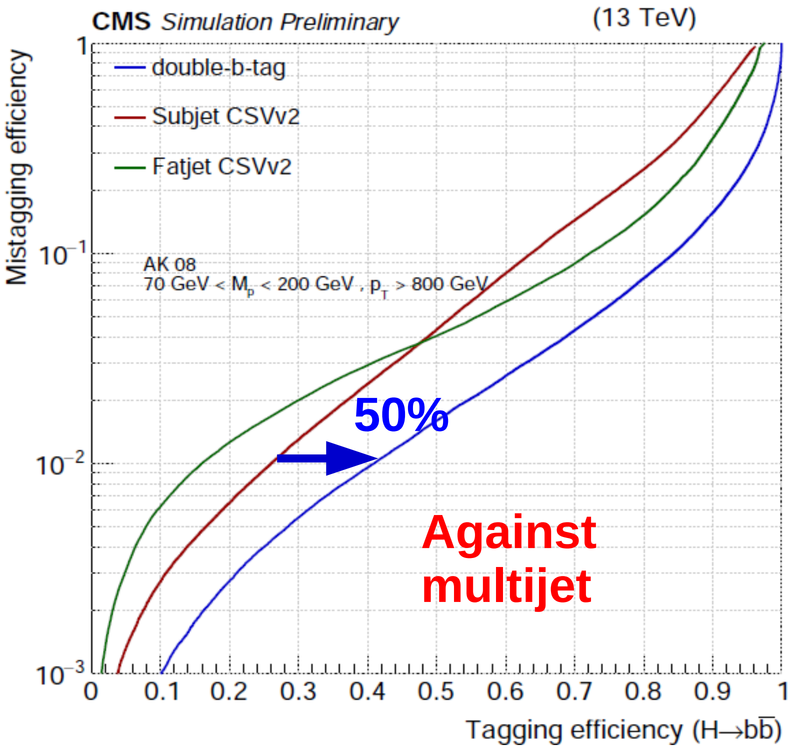
The double-b tagger is more efficient to identify boosted X → bb objects





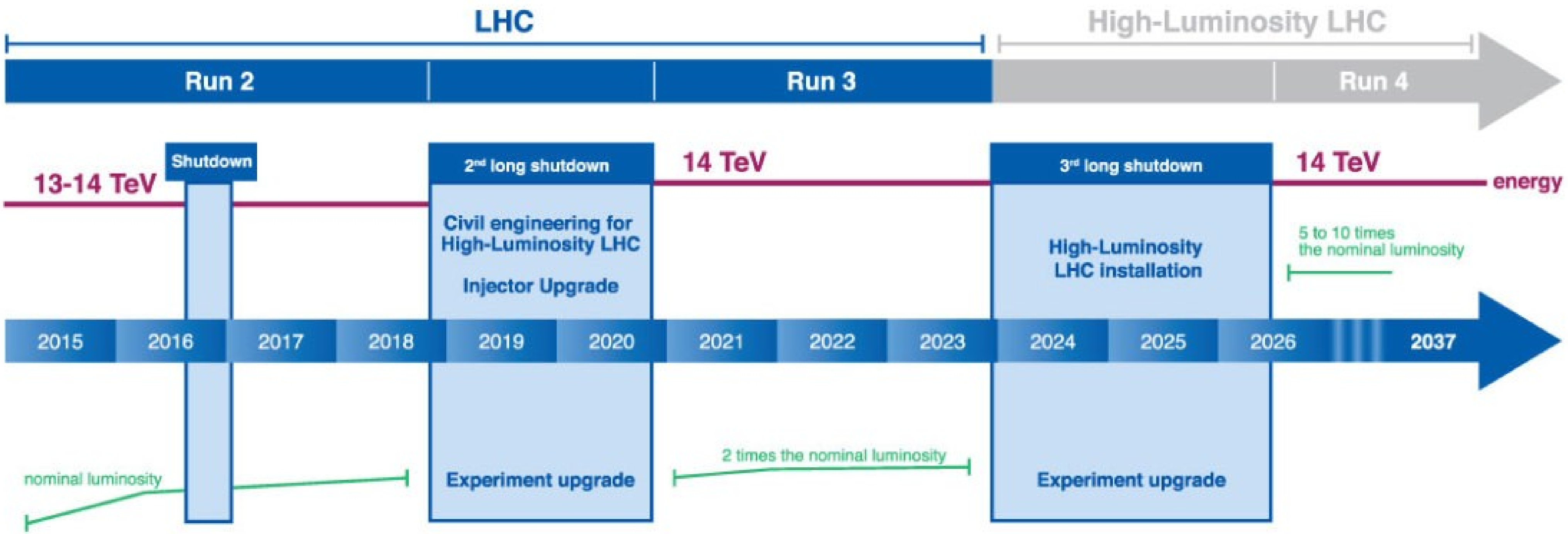
A new version of the double-b tagger will appear soon (BTV-15-002), with scale factors measured using g → bb jets

Dedicated tools are developed for the special case of b-tagging in boosted event topologies



- With increasing centre of mass energy, objects containing b jets may be boosted → different strategies developed
- Scale factors derived for subjet b-tagging
  → consistent with scale factors for 'standard' b-tagging
- Double b-tagger is 50% more efficient compared to subjet b-tagging

Upgrades to the CMS detector planned in the near and far future



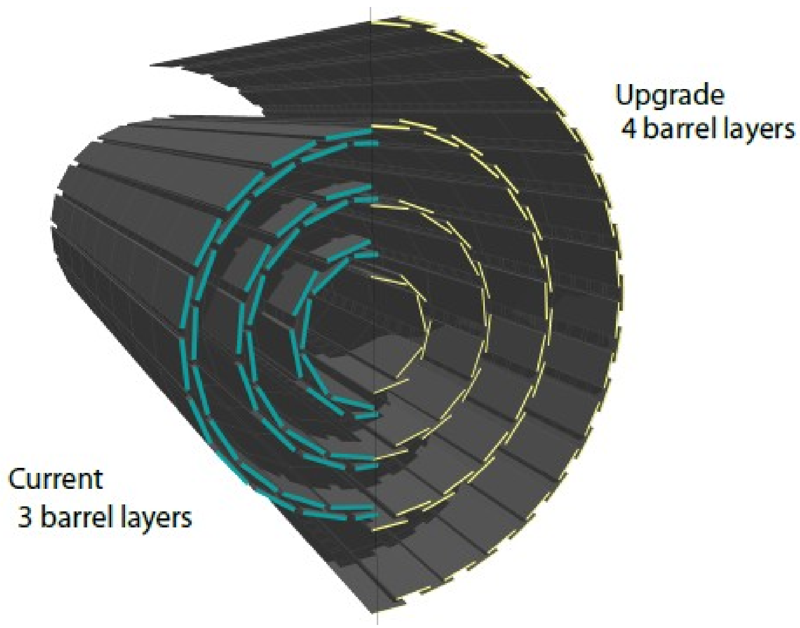
- Run 2: just started
- In 2017: new pixel tracker
- 100/fb

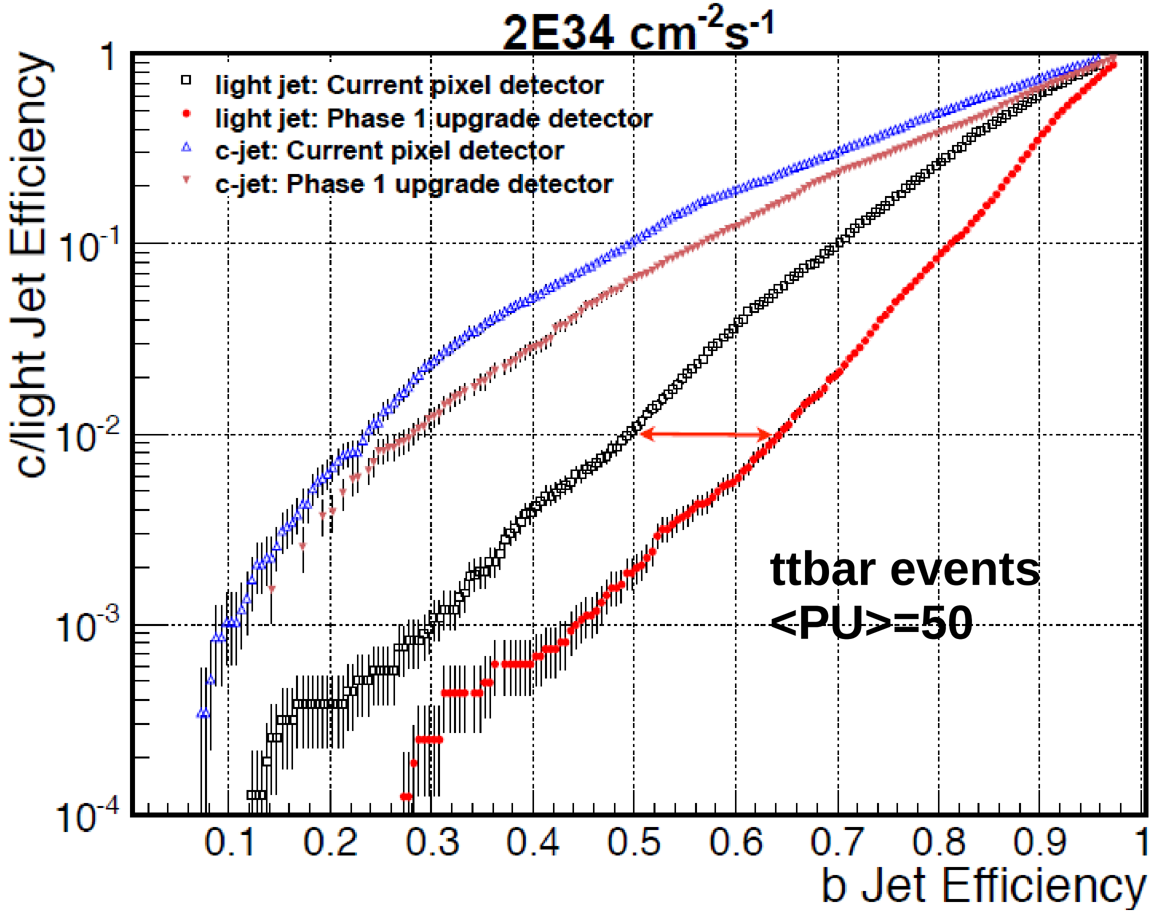
- Run 3: 2020
- 300/fb

- HL-LHC: 2026
- New tracker
- 3000/fb

Upgraded pixel tracker with 4 layers will result in increased performance

- From 2017 onwards:
  - 4 layers in the pixel tracker
  - First layer closer to the beam pipe

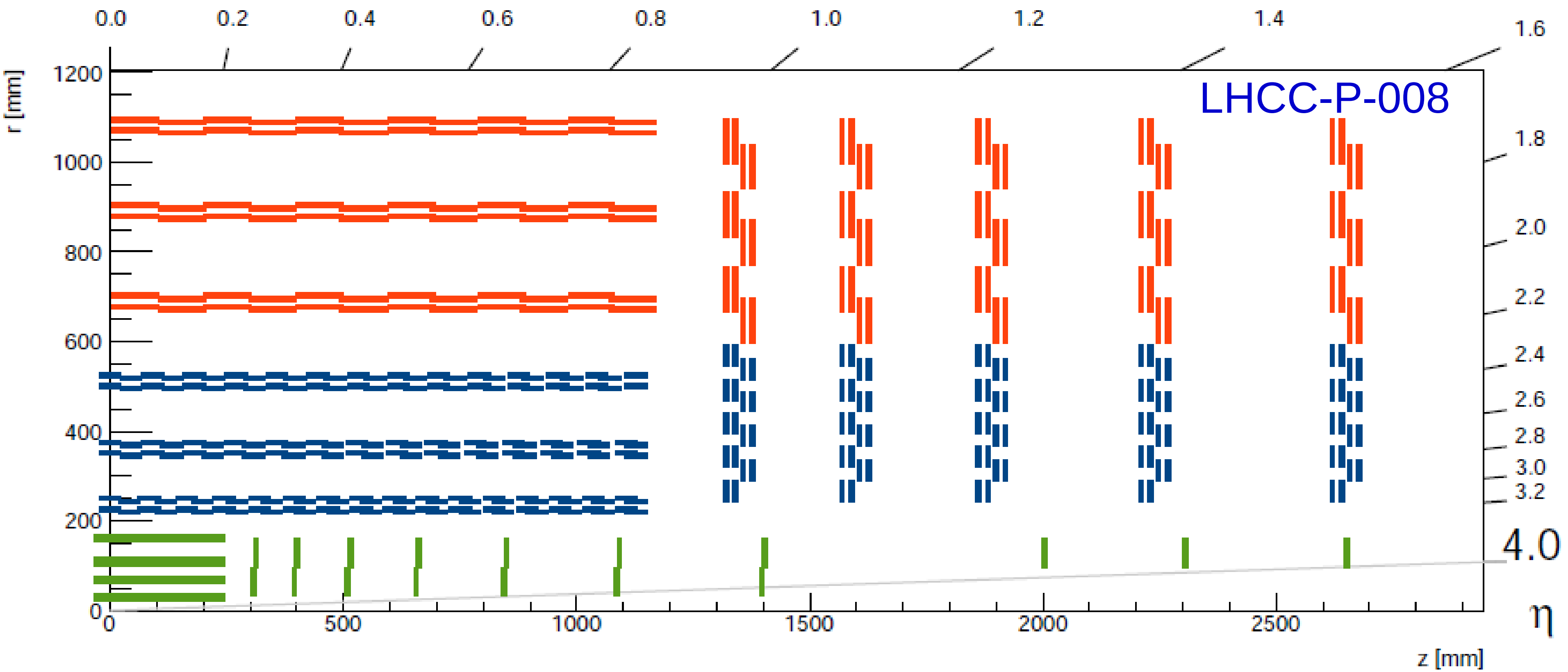




Plots from CMS-TDR-011 → old CSV tagger, but clear improvement!

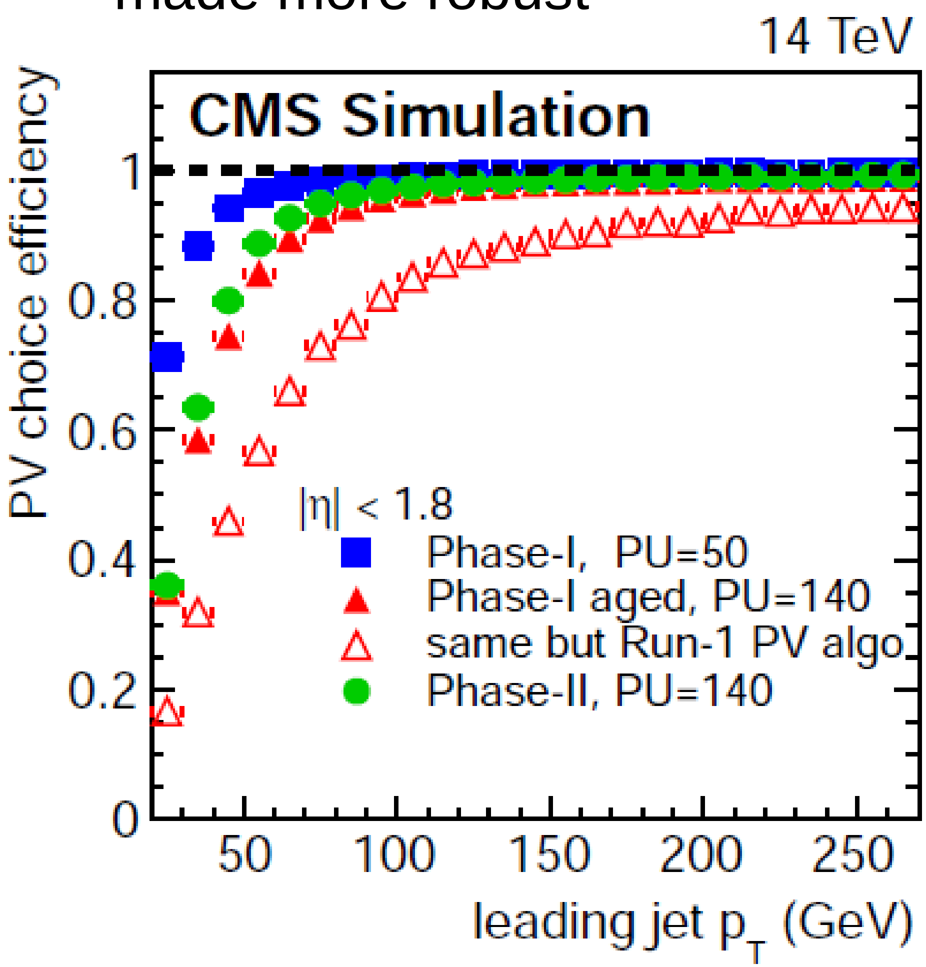
The conditions at the HL-LHC will be extreme → special tracker device

- Tracker will be used already for L1 (=hardware) trigger
- Extension of the tracker towards higher pseudorapidity
  → precision studies of e.g. H production through vector boson fusion

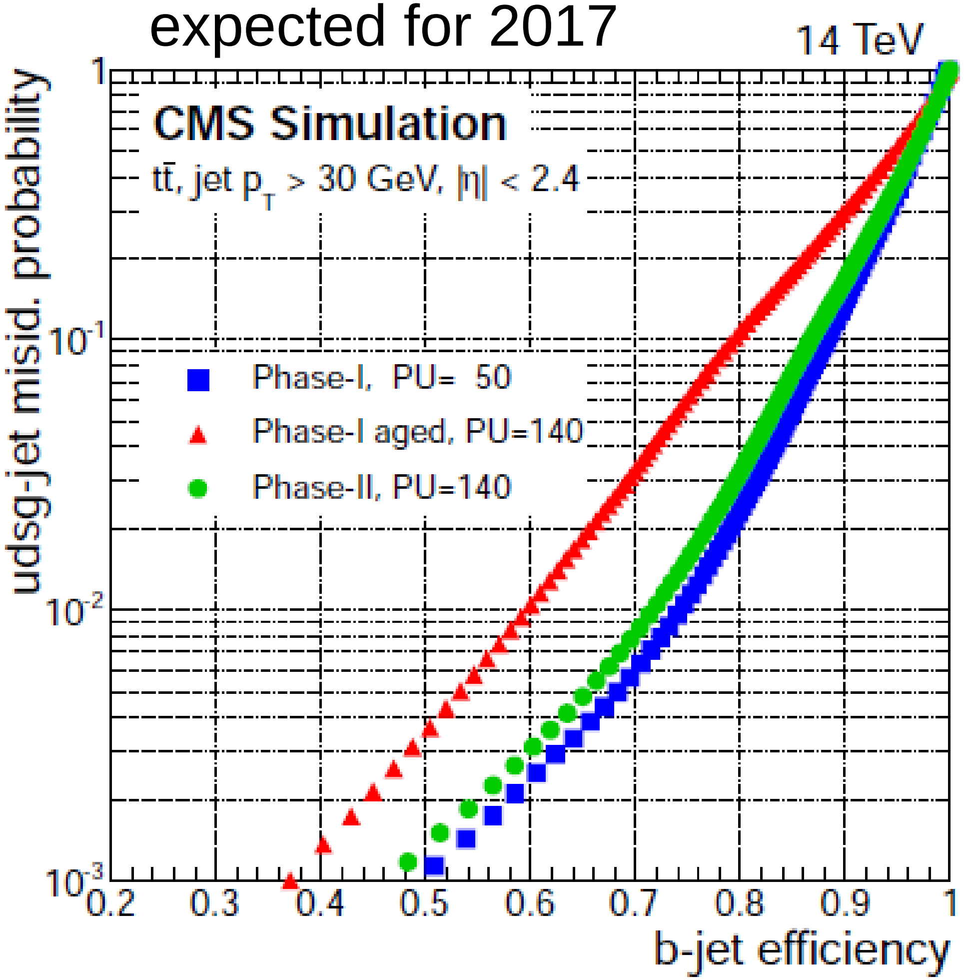


Several optimizations for primary vertex finding and b-tagging

- Primary vertex finding has been made more robust



- b jet identification close to what is expected for 2017



Plots from LHCC-P-008 → old CSV tagger, but clear improvement!

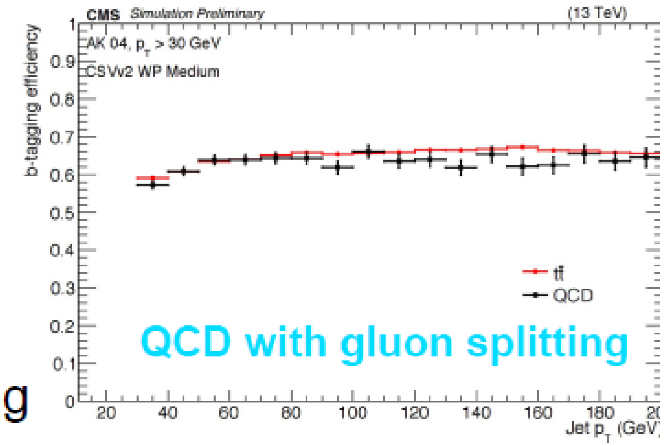
Precision measurements require reduced systematic uncertainties

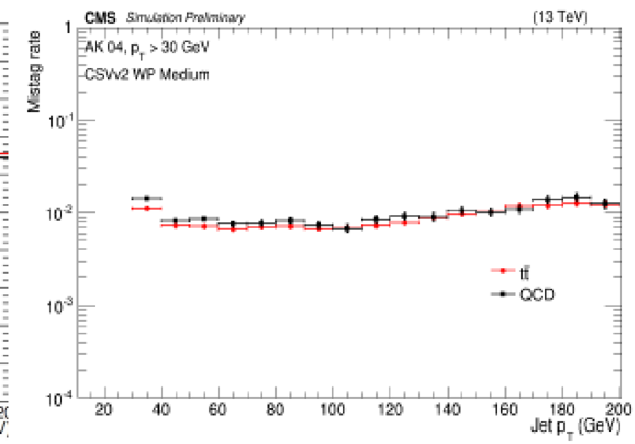
- In Run 2 statistical uncertainty will become negligible, hence we need to reduce the systematic uncertainty of the scale factor measurements to improve the precision measurements relying on b jets
- Systematic uncertainties:
  - Gluon splitting
  - b/c quark fragmentation
  - Muon p_T
  - Away-jet requirement
  - Ratio of c over light jets
  - Selection on PtRel
  - Pile up, JES, ...
  - Difference between muon-jets and inclusive jets
  - Generator uncertainties: PDF, parton showers, ISR and FSR, underlying event, B decay, …
- Starting to think about ways to reduce those

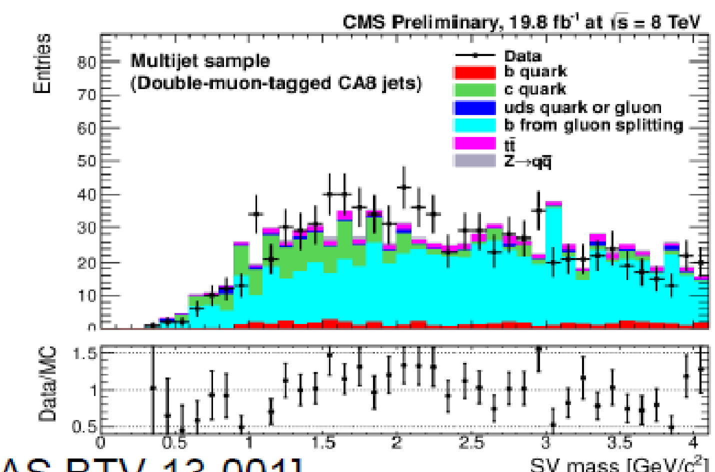
b/c prod.	low pT: 0.1% - 0.3%, high pT: 0.5% - 1.3%
mu pT	low pT: 0.1% - 1.1%, high pT: 0.1 - 0.9%
c/l ratio	<0.1% - 0.2%
b-frag	0.2% - 0.8%
PS	0.3% - 0.6%
IFSR	0.3% - 0.6%

Different efficiency for b jets in QCD and ttbar not fully understood

- **Gluon splitting**:
  - **large impact** on the QCD performance
  - removing gluon splitting component **ttbar** and **QCD** performances diverge, still being fully understood:
    - **different content of the jet pT**: larger HF hadron contribution in ttbar events
    - **more gluons around b (c) jets** in QCD?
    - inputs welcome...
  - experience in getting gluon splitting enriched regions in data (e.g. boosted topologies studies)

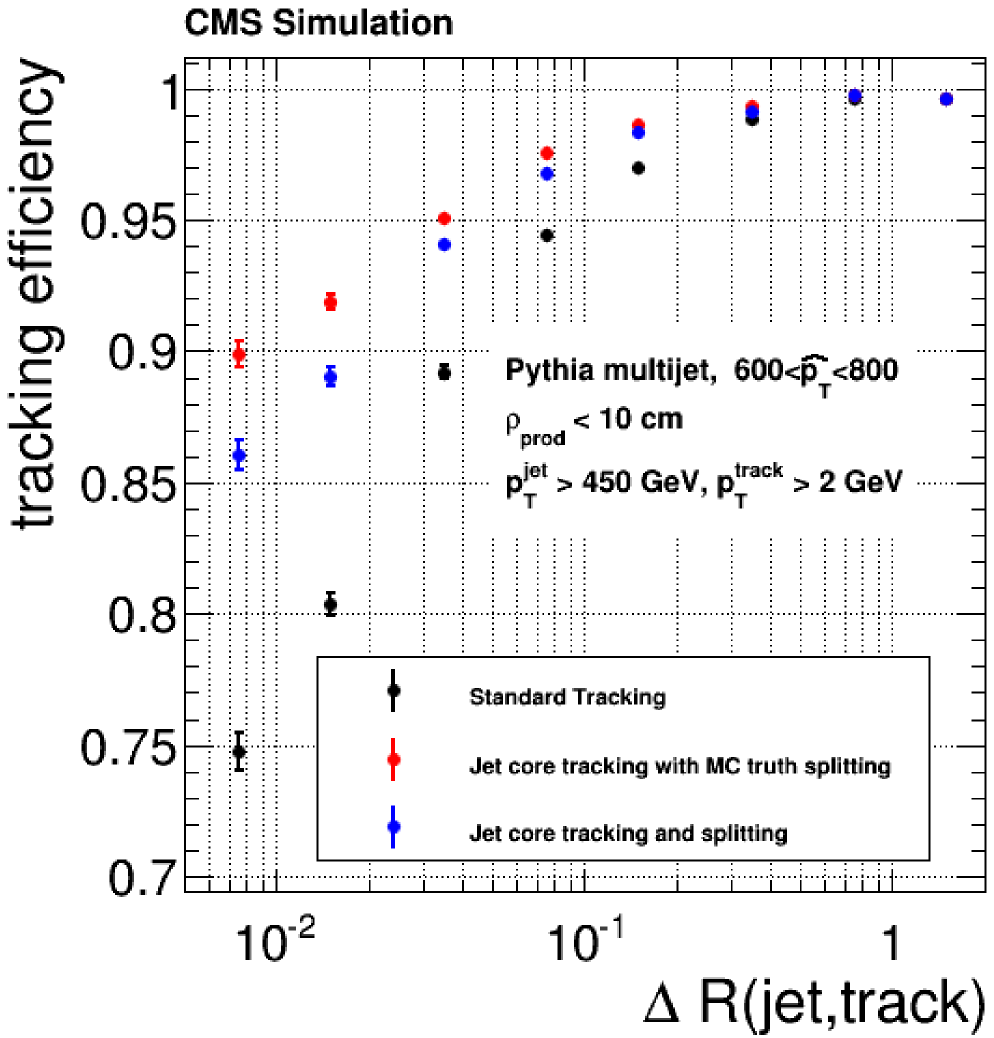






[CMS PAS BTV-13-001]

Tracking efficiency in high pt jets



The track reconstruction efficiency as function of the ΔR between the track and the vertex is computed for different scenarios: standard tracking, jet core tracking with a cluster splitting based on MC truth, and jet core tracking with cluster splitting.