b tagging at CMS

Caroline Collard (IPHC Strasbourg)

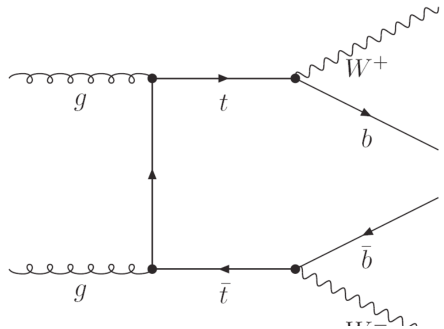
GDR Terascale, Nantes

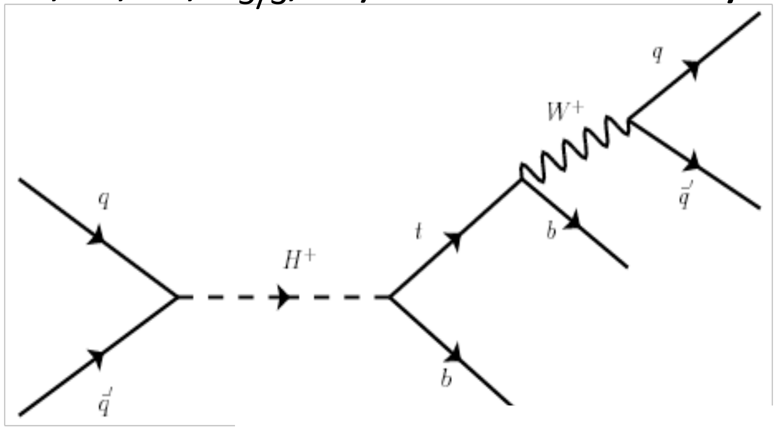
May 23-25, 2016

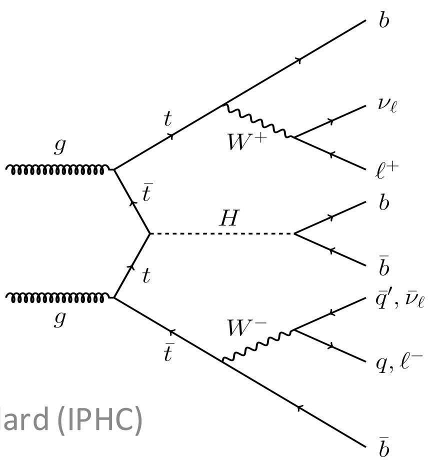
Motivation

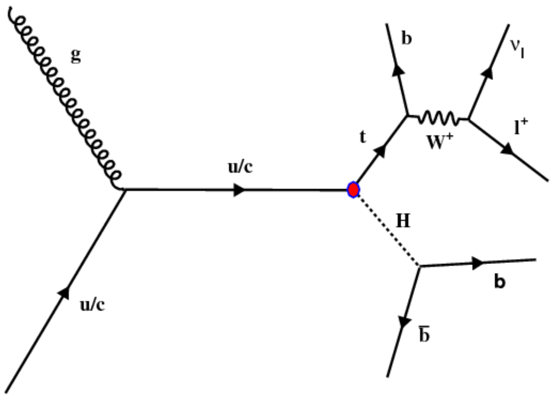
Identification of jets originating from b quarks (= b tagging) is important for many SM analyses & BSM searches:

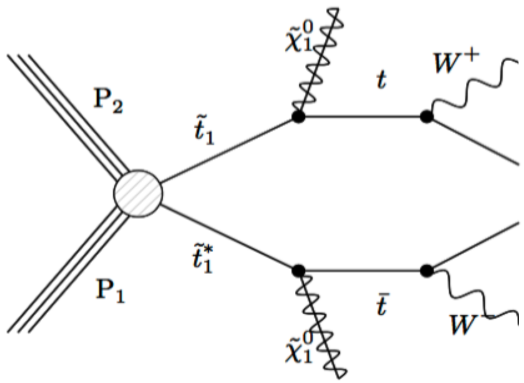
- Used in Top, SM (bb, V+bb, V+cc) and Higgs (H->bb) studies, and in 3rd generation in SUSY and BSM searches (W', Z', T', b', $T_{5/3}$, ...) + in other analyses with veto against top background.











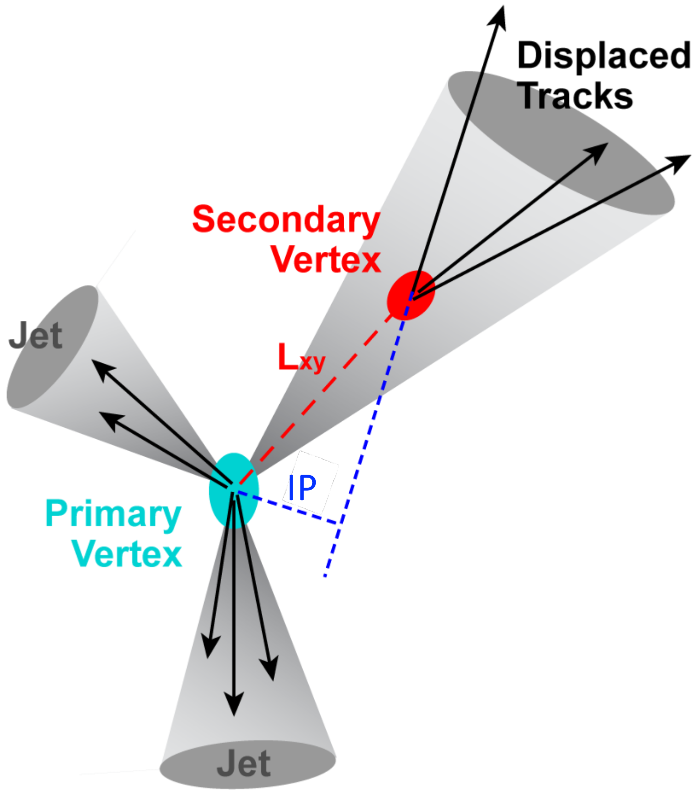
Outline of the talk

- Strategy for b tagging → Definition of b taggers
- Performances of these b taggers
- Performance measurements in data
- Special case for the boosted topologies
- What is next?
- Conclusions

This talk is highly inspired by recent presentations given by members of the CMS b tagging team.

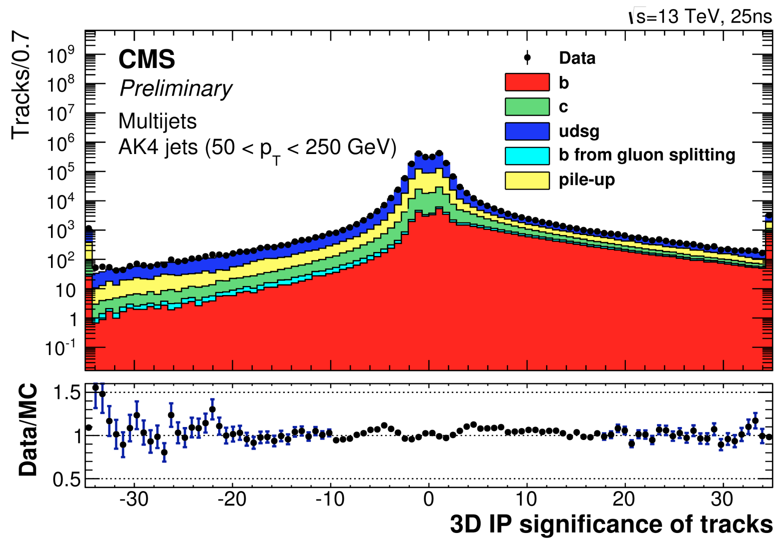
Basics on b tagging

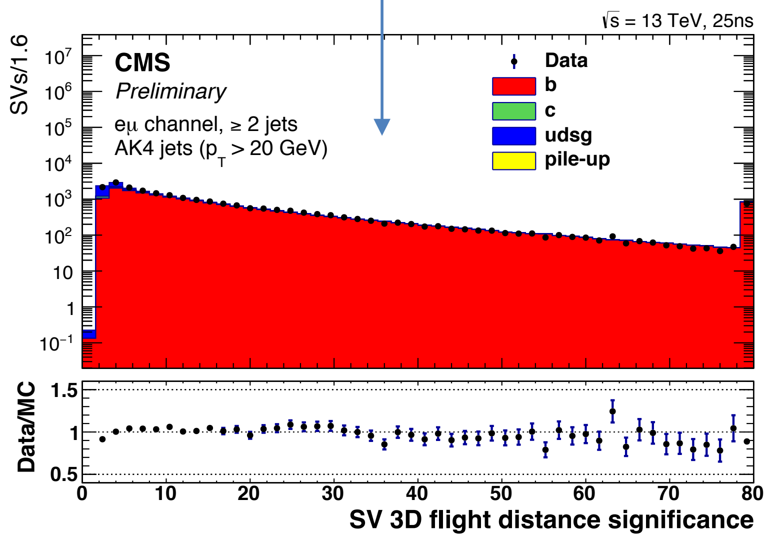
- b tagging = tagging of b jets, which are jets arising from the process of hadronization of b quarks
- Use B-hadron properties to identify b jets:
  - Relatively large mass [5-6 GeV]
  - Long lifetime [**cτ** ≈ 450 μm]
    E = 70 GeV gives **βγcτ** ≈ 5 mm
  - Daughter particle multiplicity
    ≈ five charged tracks per decay
  - Possible presence of semileptonic decays
    b→μνX [Br ≈ 11%], b→c→μνX [Br ≈ 10%]
  - Tertiary vertex
    (B-meson decay to a charmed hadron),
    cτ ≈ 120-310 μm

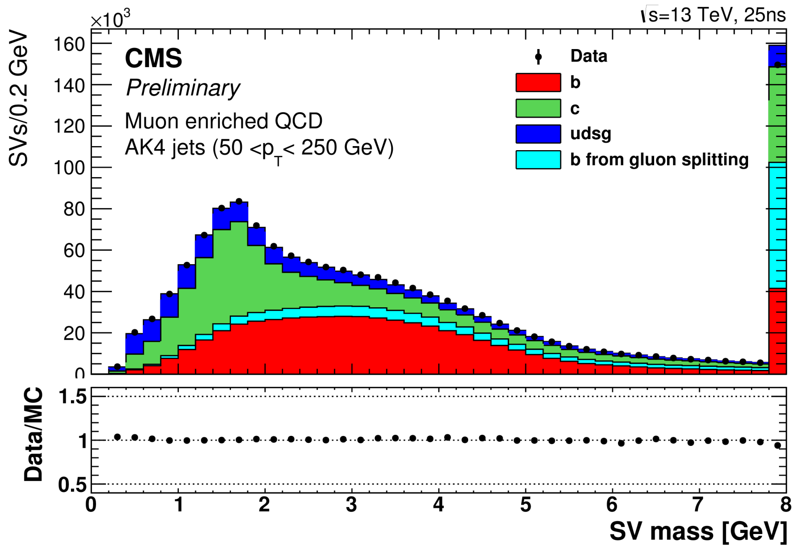


Some properties of b jets

- Information on the displaced tracks and secondary vertices,
- Check Data vs MC in different topologies :
  - Multijets (QCD),
  - jets with a soft muon (μ-enriched QCD),
  - eμ + at least 2 jets (dilepton ttbar)





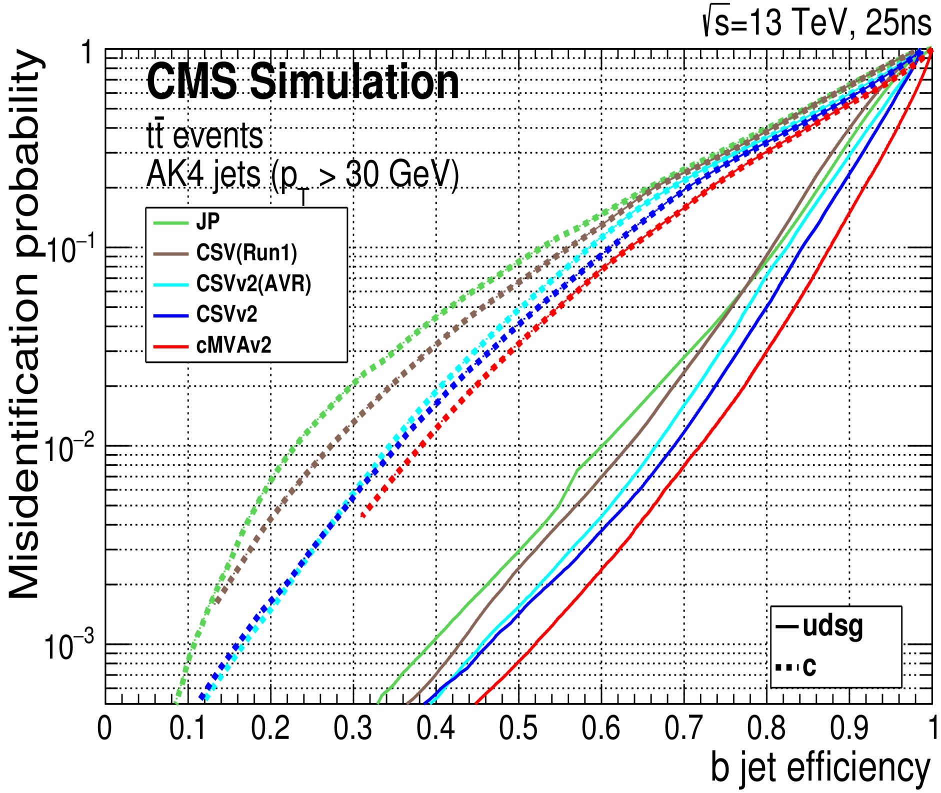


ref: BTV-15-001

Taggers for b tagging

- **CSV (Combined Secondary Vertex)** flagship tagger for Run 1, exploiting displaced tracks and AVR secondary vertices
- **CSVv2** improved version of CSV for Run2: **neural network** instead of a Likelihood Ratio, additional variables, improved track selection, use of **IVF** secondary vertices
- **JP (Jet Probability)**: Likelihood to estimate the probability of jet tracks to come from the primary vertex, mostly used for performance measurements, calibrated separately in data and MC

CSV(Run1) → CSVv2 (Run2)



- **cMVAv2** (combined MVA v2): new algorithm developed in Run 2, combining in a boosted decision tree (BDT) the discriminators from other algorithms: **JP**, **CSVv2(IVF)** and **CSVv2(AVR),** Soft Muon (**SM**) and Soft Electron (**SE**) taggers

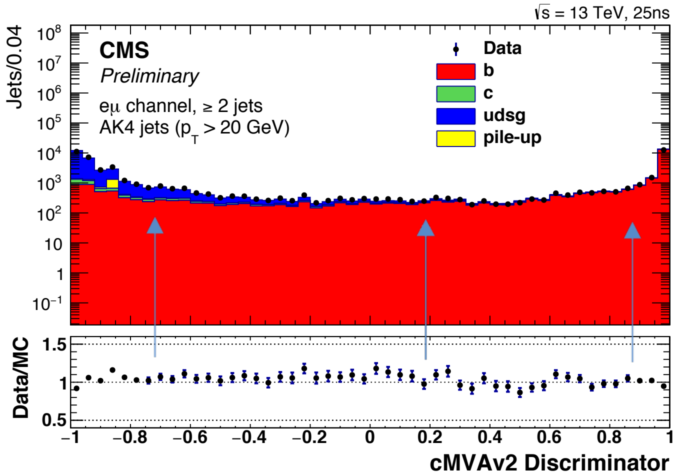
ref: BTV-15-001

b tagging efficiencies in MC

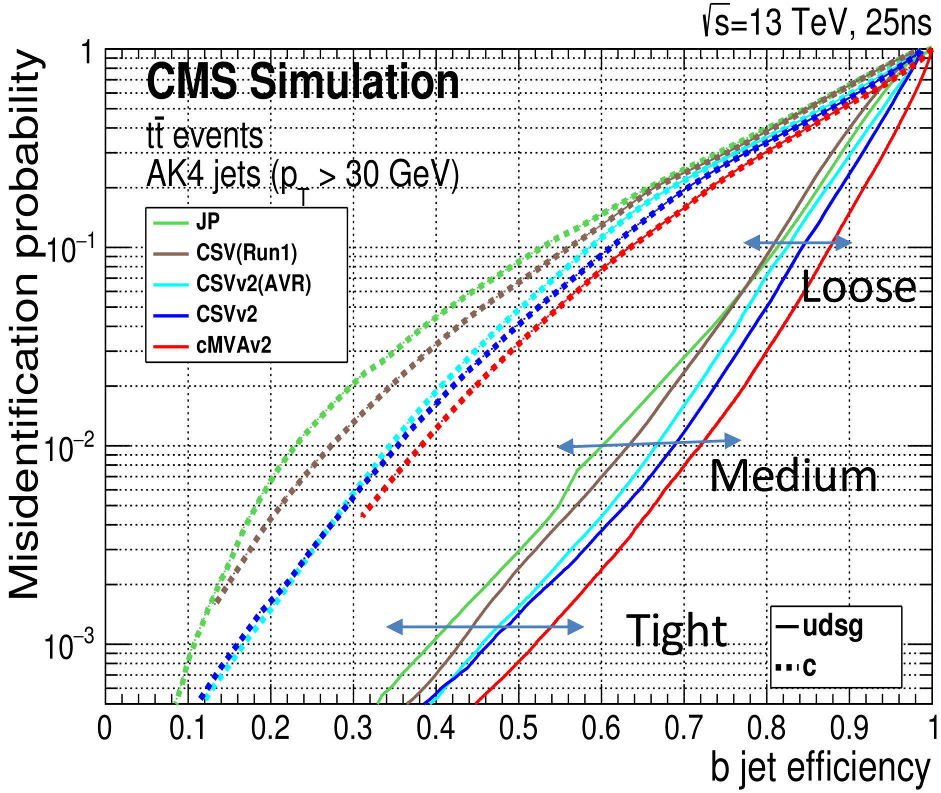
Definition of 3 working points:

Loose, Medium & Tight, in order to have a mistag rate of 10%, 1% and 0.1% respectively.

Tagger	operating point	discriminator value	ϵ_b (%)
JetProbability (JP)	JPL	0.245	≈ 82
	JPM	0.515	≈ 62
	JPT	0.760	≈ 42
Combined Secondary Vertex (CSVv2)	CSVv2L	0.460	≈ 83
	CSVv2M	0.800	≈ 69
	CSVv2T	0.935	≈ 49
Combined MVA (cMVAv2)	cMVAv2L	-0.715	≈ 88
	cMVAv2M	0.185	≈ 72
	cMVAv2T	0.875	≈ 53



CSV(Run1) → CSVv2 (Run2)

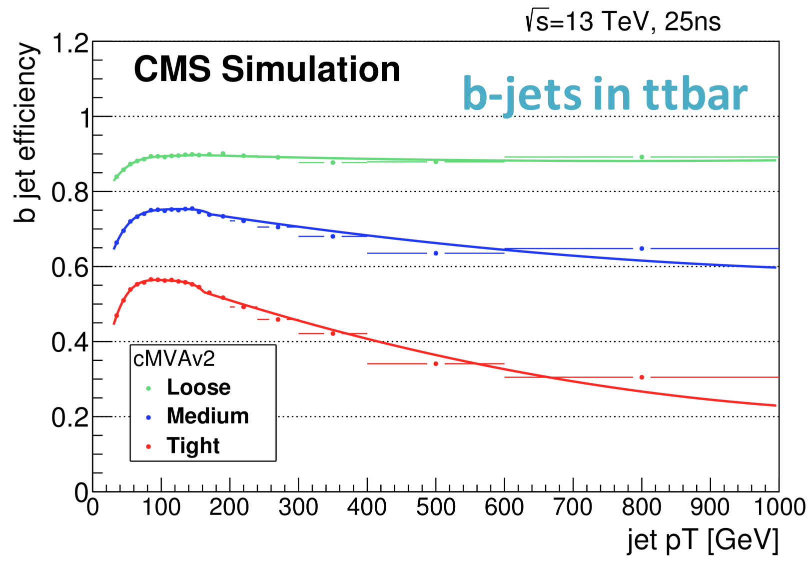


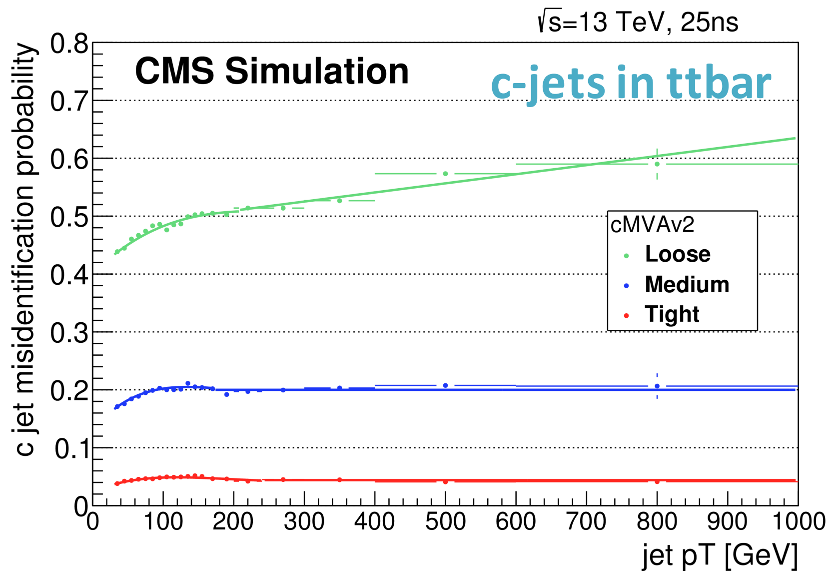
ref: BTV-15-001

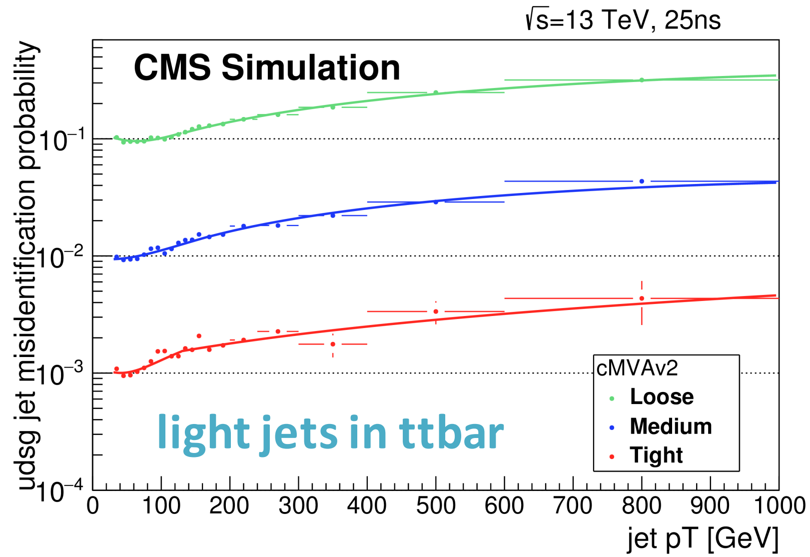
b tagging efficiencies in MC

Definition of 3 working points:
Loose, Medium & Tight, in order to have a mistag rate of 10%, 1% and 0.1% respectively.

Efficiencies as a function of pT for cMVAv2 for b, c and light jets separately.







ref: BTV-15-001

Performance Measurements

Need to correct the MC efficiencies to account for possible data/MC discrepancies in the b tagging performances:

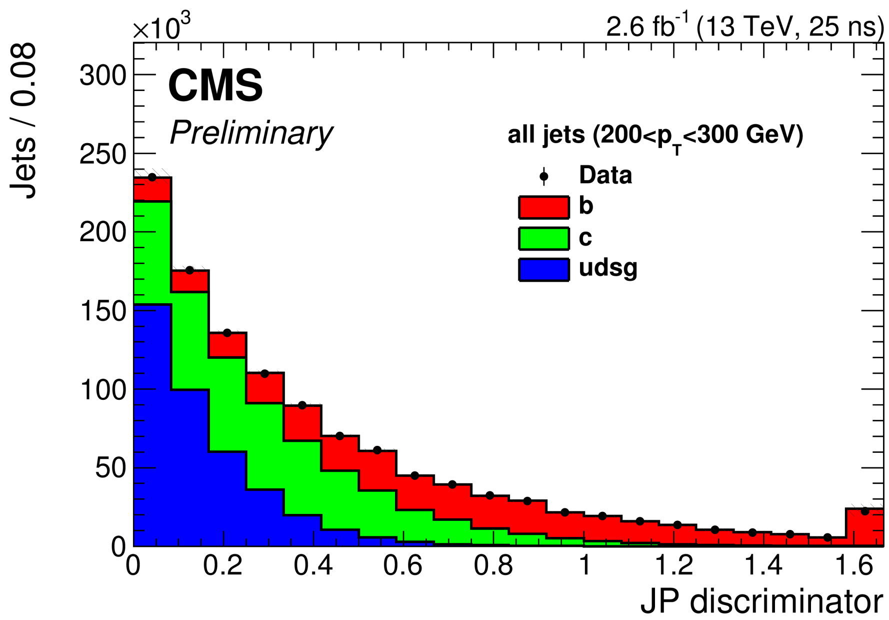
1) Scale factors ($\varepsilon^{Data}/\varepsilon^{MC}$) to correct for a given WP
   - Measurement of the b tagging efficiency, based on samples enriched in b jets:
     - jets with a soft muon coming from a semileptonic decay of a B hadron:
       - PtRel method,
       - Lifetime Tagger method,
       - System8 method

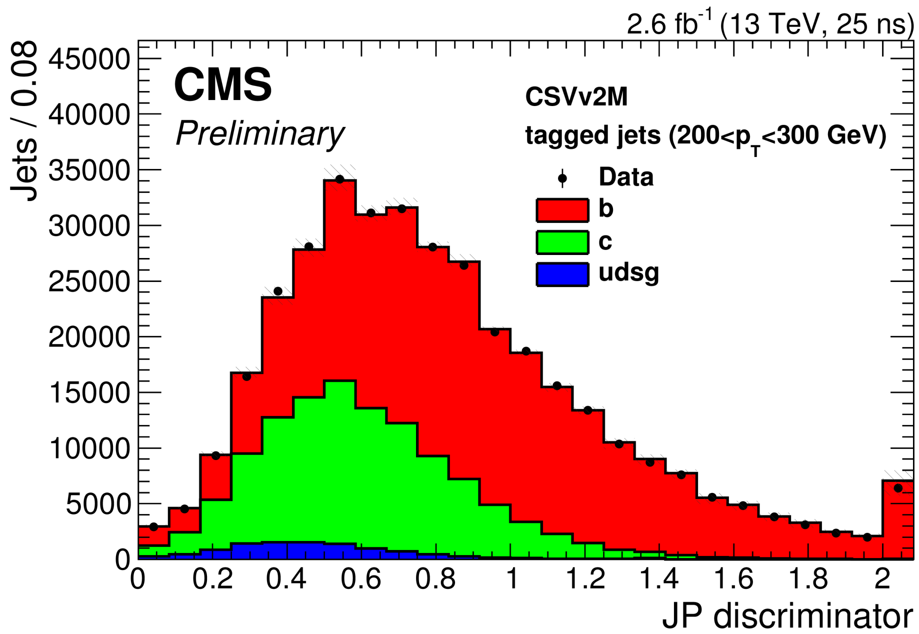
       } combined
     - ttbar dilepton sample: Tag Counting method
   - Measurement of the misidentification probability for light jets: performed on inclusive QCD sample with the negative tag method
   - Measurement of the misidentification probability for c jets: work ongoing
2) Correction factors for reshaping the whole discriminator distribution, for analysis exploiting the shape (e.g. in MVA): Reweighting method which provides SF for both b jets (based on ttbar dilepton events) and light jets (based DY dilepton events).

1^{st} example: the Lifetime Tagger method

Template fit method based on the Jet Probability (JP) discriminant:

- Use jets containing a soft muon, to enrich the b contribution
- Templates from MC
- Fits are done before and after b tagging requirement to measure the efficiency
- $\varepsilon_b^{tag} = N_{b\text{-}jet}^{tagged} / N_{b\text{-}jet}^{total}$





ref: BTV-15-001

Combination of the QCD-based SF

Combination of all the methods with the BLUE method

Treatment of systematics:

- Common (PU, gluon splitting, P_T^{μ}) or for 2 of them (away-jet tagger)
  → 100% correlated or anti-corr
- Other specific to 1 method: uncorr

The event overlap has been taken into account in the combination.

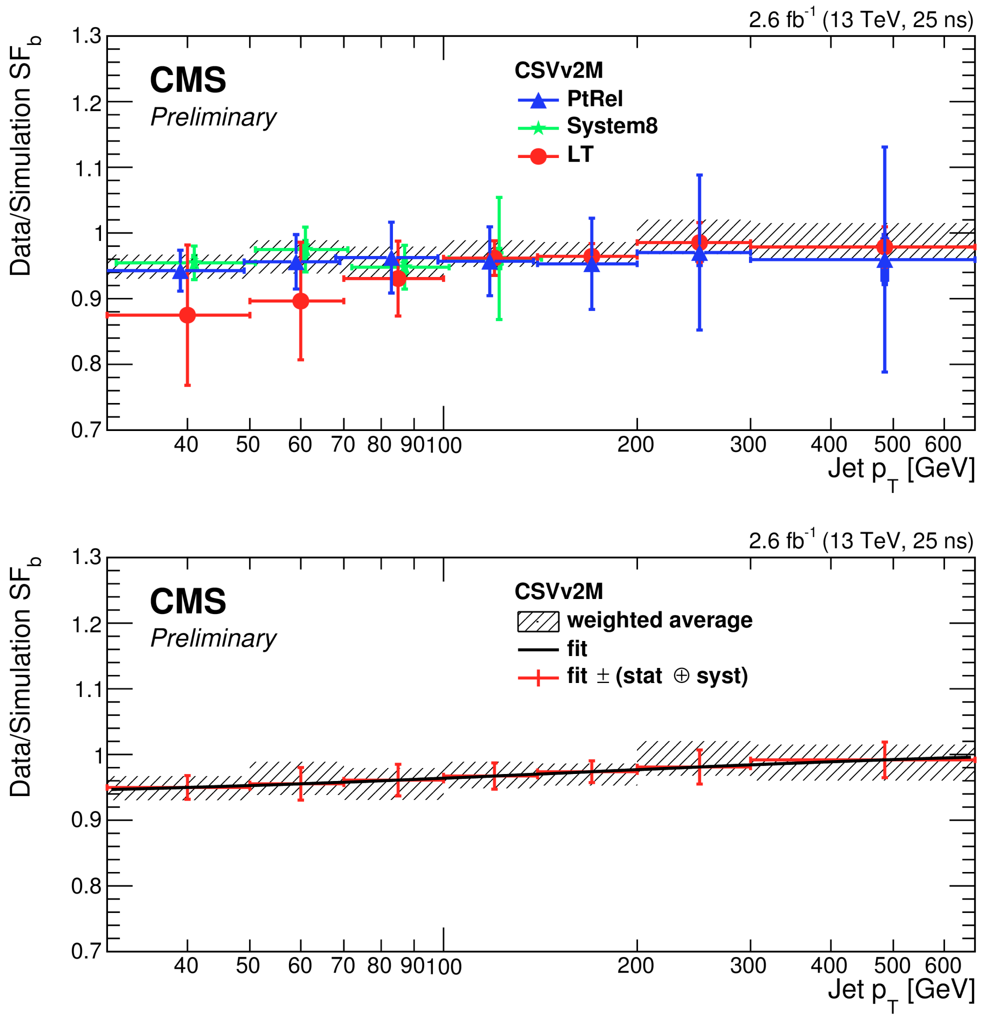
Precision: σ(stat) ~15-30% σ(tot)

To quantify the relative $\sigma[SF_b]$:

For 70 < p_T<100 GeV : 1.7% (L) → 3% (T)

For 300 < p_T<670 GeV: 4% (L) → 5% (T)

QCD-based SF combination



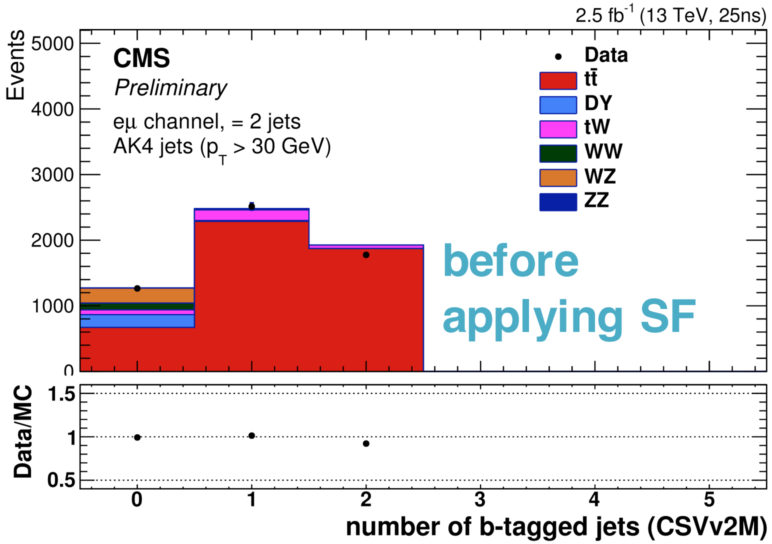
ref: BTV-15-001

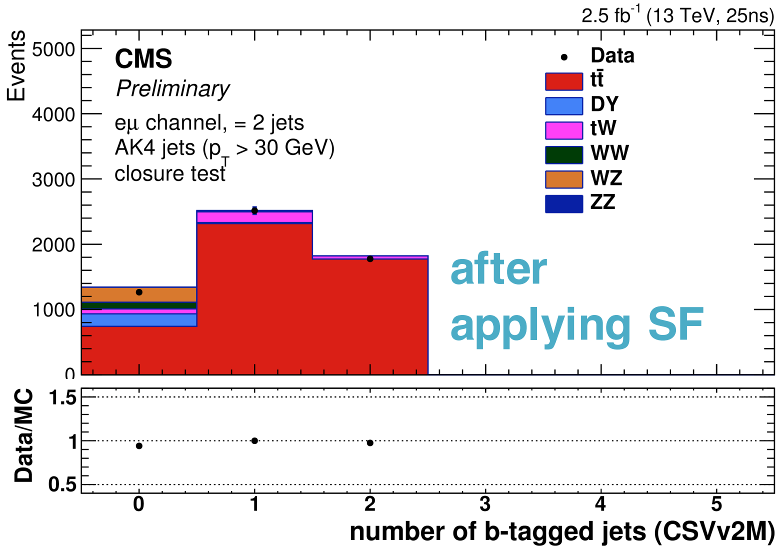
2nd example: the TagCounting method

Count fraction of events with $N_{btag} = 2$ in a sample with two jets:

- Use dilepton ttbar eμ events → high b jet purity
- Based on fractions → event yield systematics cancel out, but sensitive to modeling uncertainties (fragmentation and normalization scales)
- No fit performed, calculate b tagging efficiency as:

$$\varepsilon_b = \sqrt{\frac{F_{2tag} - F_{non2b}^{truth}}{f_{2b}}}$$



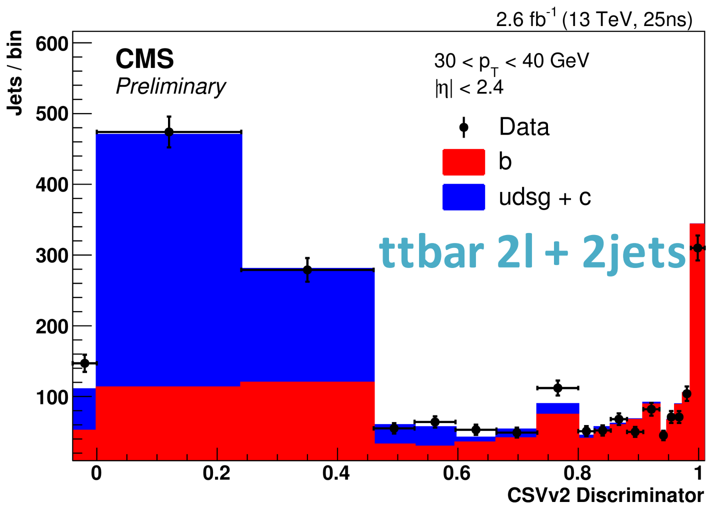


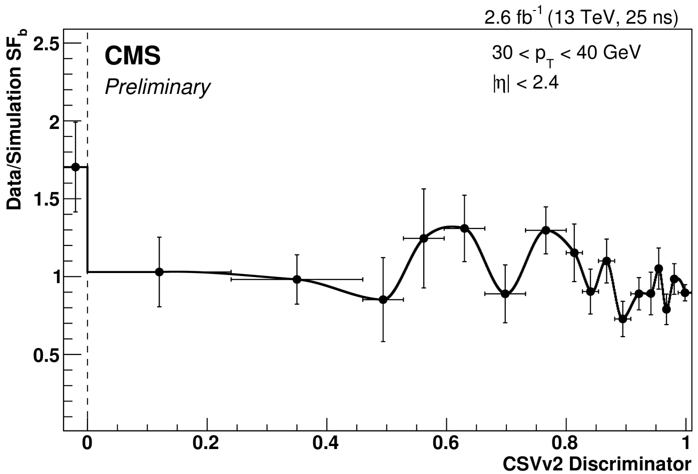
ref: BTV-15-001

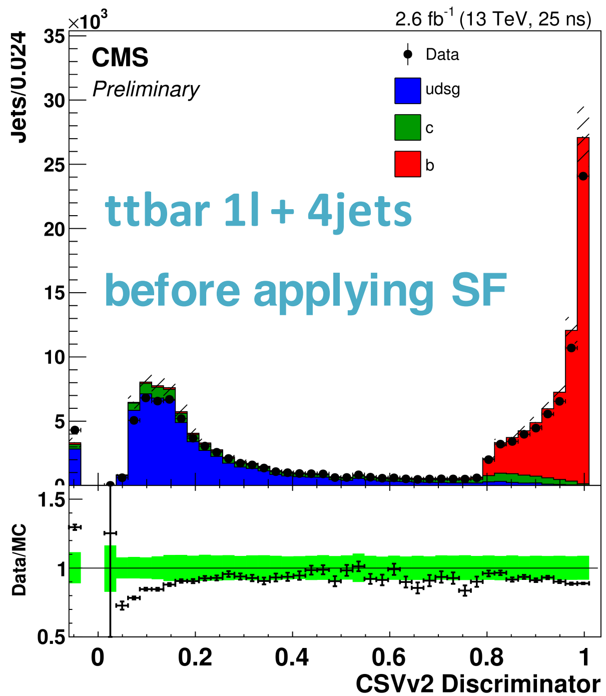
3rd example : the Reweighting method

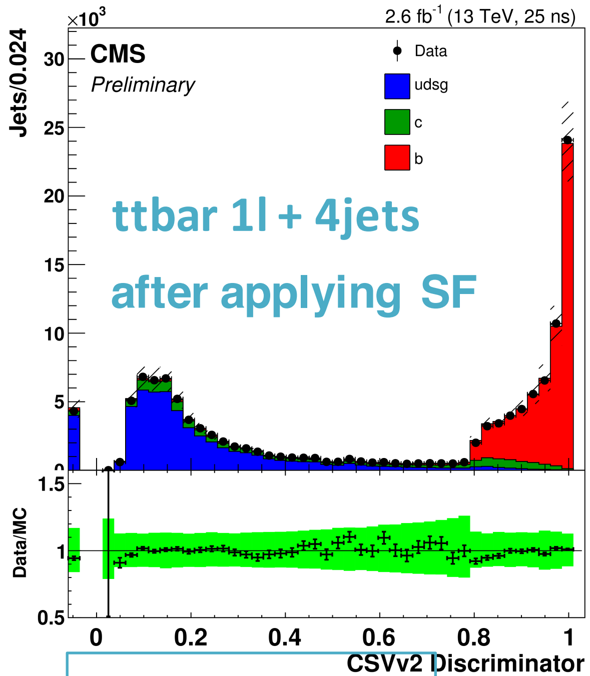
Tag&Probe method to extract shape SF in ttbar 2l for b-jets
and in DY 2l for light jets.

Closure test done on ttbar 1l events.









ref: BTV-15-001

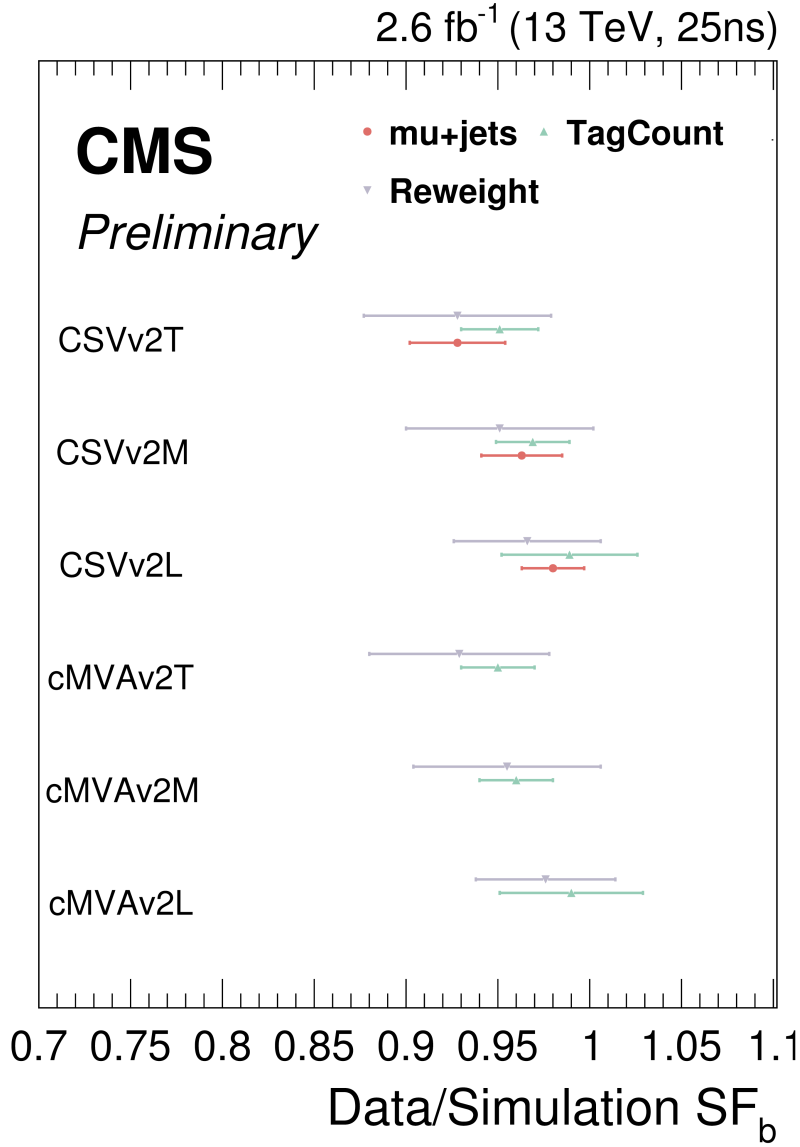
Comparison of the QCD-based and ttbar-based SF

Consistent results from different techniques and different samples

Here compared:

- combined results from **muon-enriched QCD**, averaged over the pT spectrum of b-jets from ttbar
- **TagCount** method results (ttbar)
- average scale factors obtained applying the **reweighting** method on ttbar events

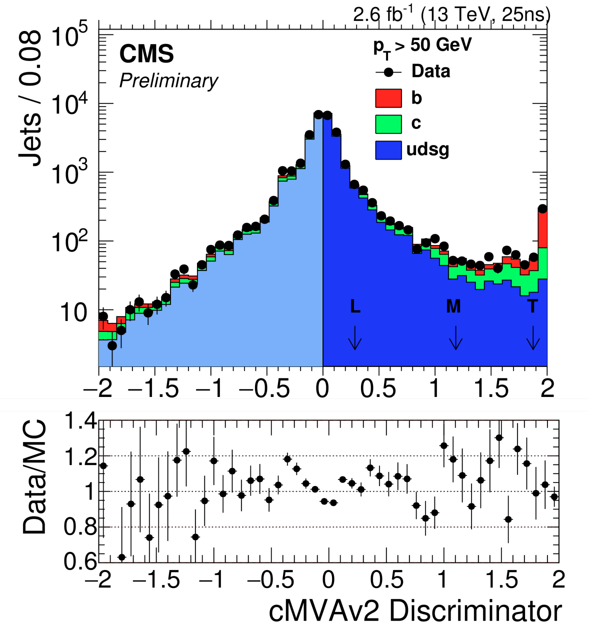
Note: No cMVAv2 results for the mu+jets combination because of a possible bias as cMVAv2 uses the the soft lepton info as input.



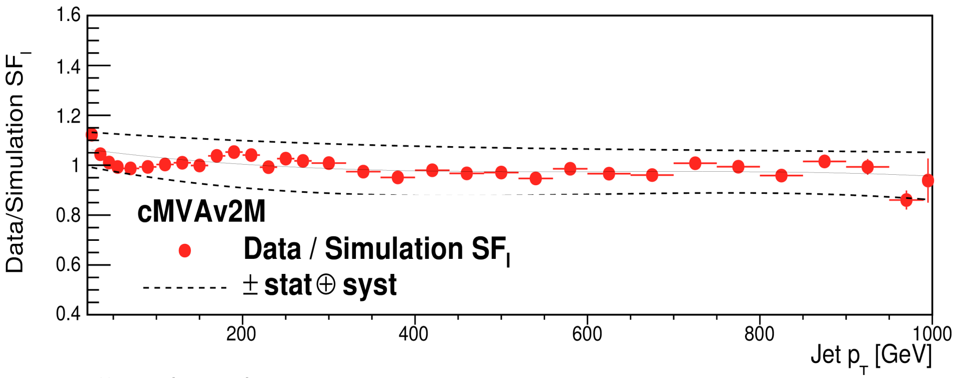
ref: BTV-15-001

4th example: the negative tag method

Method to measure the mistag rate (= the efficiency to tag a light jet as a b jet) on multijets events, based on negative and positive taggers.



- Negative (or positive) tagger = similar to the default algo but using only tracks with IP<0 (or >0) or SV decay lengths <0 (or >0).
- For light jets, negative & positive taggers are expected to be symmetric (as the sign of the IP or decay length is mostly due to resolution effects in track reco)
- Efficiency from negative taggers, corrected for b/c jet contamination and long-lived particles.
- Correction factors in pT and η bins.



Precision: almost fully dominated by systematic effects.

To quantify the relative $\sigma[SF_{light}]$:

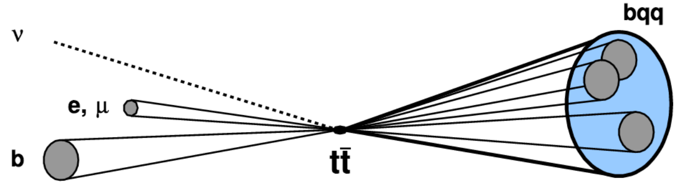
For 80 < p_T< 320 GeV : 5% (L) → 20% (T)

ref: BTV-15-001

Boosted b tagging

Special consideration for the case of b quarks arising from highly Lorentz-boosted particles (boosted top or boosted Higgs).

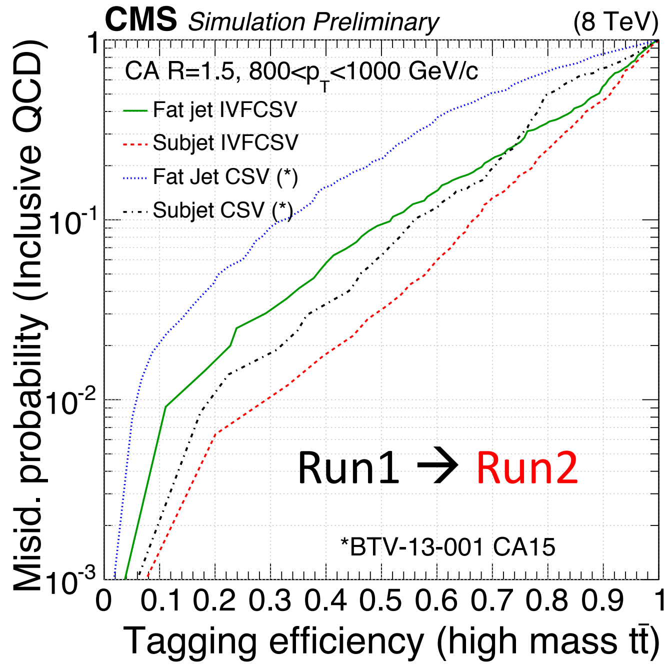
Consequence of the boost of the parent particle: collimated decay products, merged into a single ”fat” (large R) jet.



Developed at Run1: b tagging for fat jets (using all jet tracks) and subjets (based on subjet tracks).

Subjet b tagging outperforms fatjet b tagging in most of the cases.

Improvement for Run2: jet-track association, jet flavour definition, and benefits from improvements to the standard CSVv2.

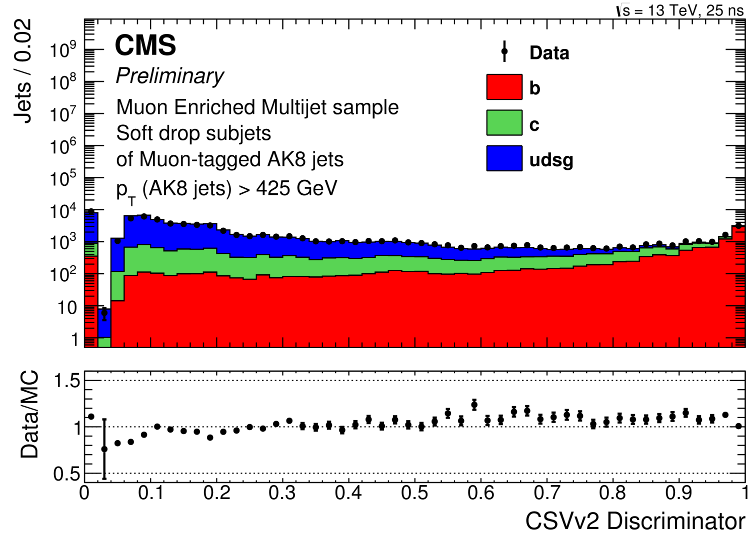


ref: DP-2014-031

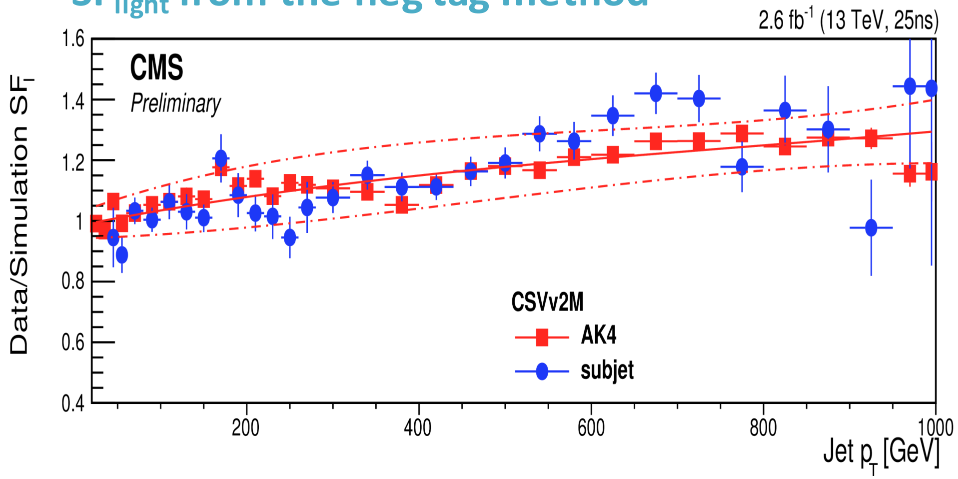
Performance measurements in boosted topologies

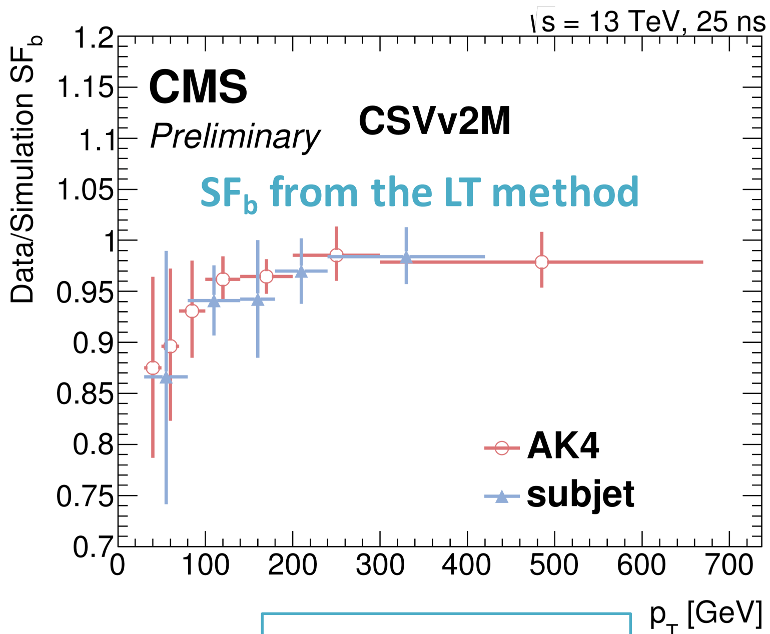
Performance measurements performed on AK4 subjets reconstructed within the AK8 fat jets.

- CSVv2 algo with the same Loose & Medium WP
- Same methods as for AK4 jets used here.
- Good agreement between the results of the 2 jet sizes.



SF_{light} from the neg tag method





ref: BTV-15-001

Ongoing developments on boosted topologies

New strategy for the boost H→bb topology: design a double b tagger

Specifications: do better than subjet or fatjet b tagging, be stable against p_T & independent from particle mass

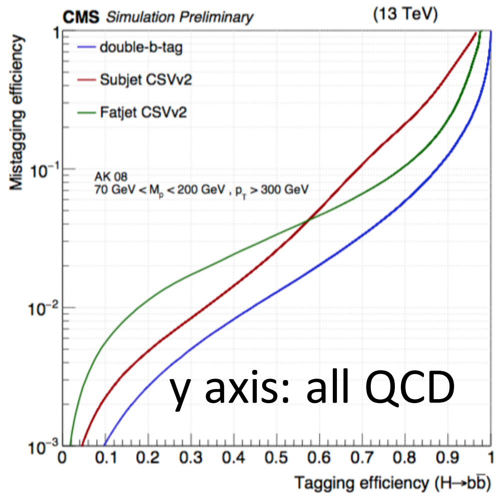
BDT training on G*→ HH→ 4b against QCD using

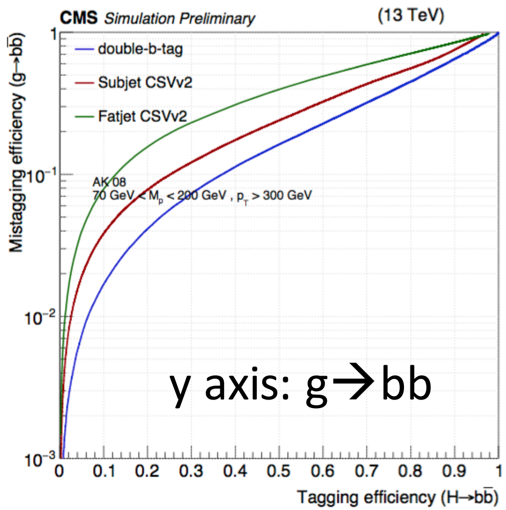
- track info,
- secondary vertex info,
- the minimum CSVv2 subjet score,
- and if two SVs found:

$Z = \Delta R(SV_1,SV_2)^* z$ with $z = p_{T\,1}/\mathrm{mass}(SV_1 + SV_2)$

Overall outperforms subjet and fatjet b tagging

A new version of this tagger is available within CMS
→ PAS BTV-15-002 in preparation





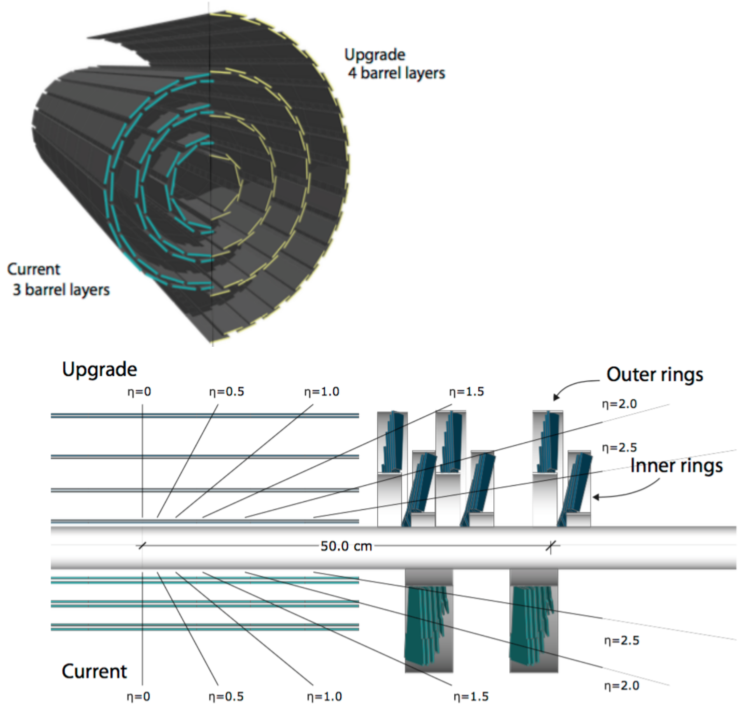
ref: DP-2015-038

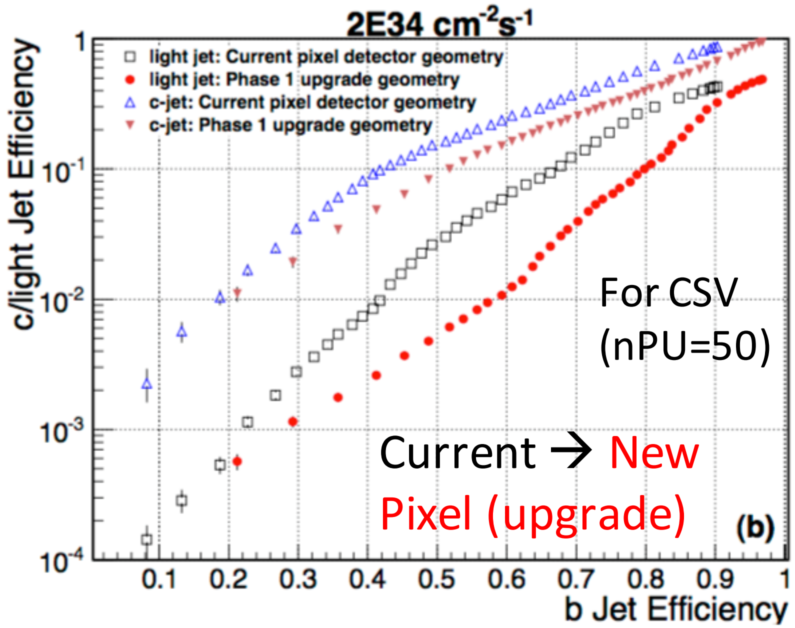
What is next?

End of 2016 : installation of the new pixel detector of CMS (Phase1).

Significant improvement in b tagging due to extra layers, finer granularity, decrease in the amount of material:

- For an efficiency(bjet) = 60%, a factor of 6 for the light reduction is expected.
- For a mistag of 1%, a relative 40% improvement in b-tagging efficiency.





ref: CMS-TDR-011,
CERN-LHCC-2011-006

To conclude

Overview of b tagging in CMS: it is working well at 13 TeV

- in standard jet configurations
- in boosted topologies
- [not mentioned in the talk, but working well too
  - at the trigger level
  - in events from heavy-ion collisions]

Additional developments ongoing to improve b-tagging in AK4 & AK8 jets, as well as in view of the new pixel detector.

New public results coming soon: double b tagger (PAS BTV-15-002) and c tagger (PAS BTV-16-001).

Parametrization for cMVAv2

Jet flavour	operating point	jet p_T range	function
b	Loose	$30 \leq p_T < 150\,\text{GeV}$	$0.707 + 5.6 \cdot 10^{-3} \cdot p_T - 6.27 \cdot 10^{-5} \cdot p_T^2 + 3.10 \cdot 10^{-7} \cdot p_T^3 - 5.63 \cdot 10^{-10} \cdot p_T^4$
		$150 \leq p_T$	$0.906 - 6.39 \cdot 10^{-5} \cdot p_T + 4.11 \cdot 10^{-8} \cdot p_T^2$
	Medium	$30 \leq p_T < 175\,\text{GeV}$	$0.421 + 0.0107 \cdot p_T - 1.314 \cdot 10^{-4} \cdot p_T^2 + 7.268 \cdot 10^{-7} \cdot p_T^3 - 1.523 \cdot 10^{-9} \cdot p_T^4$
		$175 \leq p_T$	$0.79 - 3.17 \cdot 10^{-4} \cdot p_T + 1.24 \cdot 10^{-7} \cdot p_T^2$
	Tight	$30 \leq p_T < 160\,\text{GeV}$	$0.127 + 0.01578 \cdot p_T - 2.126 \cdot 10^{-4} \cdot p_T^2 + 1.273 \cdot 10^{-6} \cdot p_T^3 - 2.88 \cdot 10^{-9} \cdot p_T^4$
		$160 \leq p_T$	$0.634 - 6.74 \cdot 10^{-4} \cdot p_T + 2.69 \cdot 10^{-7} \cdot p_T^2$
c	Loose	$30 \leq p_T < 205\,\text{GeV}$	$0.40 + 1.23 \cdot 10^{-3} \cdot p_T - 4.60 \cdot 10^{-6} \cdot p_T^2 + 5.71 \cdot 10^{-9} \cdot p_T^3$
		$205 \leq p_T$	$0.478 + 1.573 \cdot 10^{-4} \cdot p_T$
	Medium	$30 \leq p_T < 170\,\text{GeV}$	$0.13 + 1.48 \cdot 10^{-3} \cdot p_T - 1.00 \cdot 10^{-5} \cdot p_T^2 + 2.65 \cdot 10^{-8} \cdot p_T^3 - 2.36 \cdot 10^{-11} \cdot p_T^4$
		$170 \leq p_T$	0.20
	Tight	$30 \leq p_T < 240\,\text{GeV}$	$0.024 + 5.27 \cdot 10^{-4} \cdot p_T - 3.72 \cdot 10^{-6} \cdot p_T^2 + 9.87 \cdot 10^{-9} \cdot p_T^3 - 8.83 \cdot 10^{-12} \cdot p_T^4$
		$240 \leq p_T$	0.044
light	Loose	$30 < p_T < 130\,\text{GeV}$	$0.124 - 1.0 \cdot 10^{-3} \cdot p_T + 1.06 \cdot 10^{-5} \cdot p_T^2 - 3.18 \cdot 10^{-8} \cdot p_T^3 + 3.13 \cdot 10^{-11} \cdot p_T^4$
		$130 \leq p_T$	$0.055 + 4.53 \cdot 10^{-4} \cdot p_T - 1.6 \cdot 10^{-7} \cdot p_T^2$
	Medium	$30 \leq p_T < 170\,\text{GeV}$	$9.59 \cdot 10^{-3} - 1.96 \cdot 10^{-5} \cdot p_T + 4.53 \cdot 10^{-7} \cdot p_T^2 - 1.08 \cdot 10^{-9} \cdot p_T^3 + 7.62 \cdot 10^{-13} \cdot p_T^4$
		$170 \leq p_T$	$5.07 \cdot 10^{-3} + 6.02 \cdot 10^{-5} \cdot p_T - 2.3 \cdot 10^{-8} \cdot p_T^2$
	Tight	$30 \leq p_T < 130\,\text{GeV}$	$1.24 \cdot 10^{-3} - 1.27 \cdot 10^{-5} \cdot p_T + 1.98 \cdot 10^{-7} \cdot p_T^2 - 7.46 \cdot 10^{-10} \cdot p_T^3 + 8.35 \cdot 10^{-13} \cdot p_T^4$
		$130 \leq p_T$	$1.08 \cdot 10^{-3} + 3.54 \cdot 10^{-6} \cdot p_T$

ref: BTV-15-001

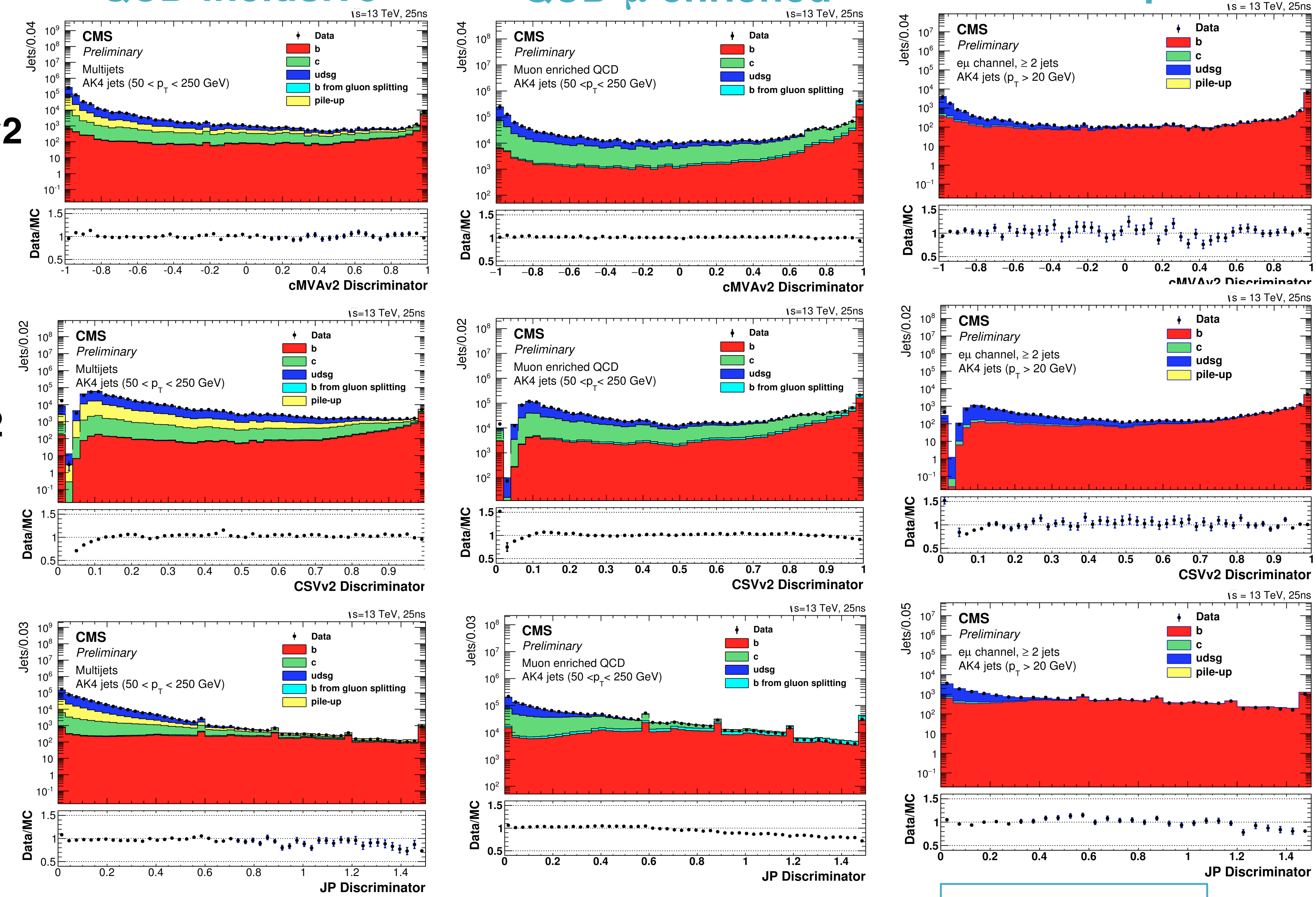
Commissioning for ak4 jets

QCD inclusive | QCD μ-enriched | ttbar 2leptons

cMVAv2

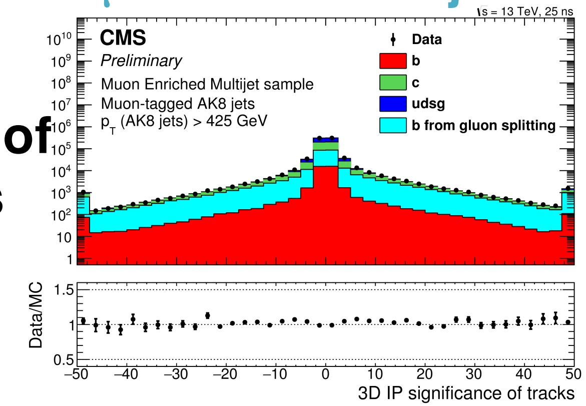
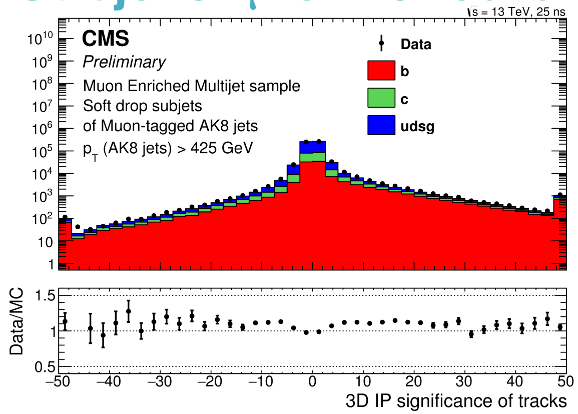
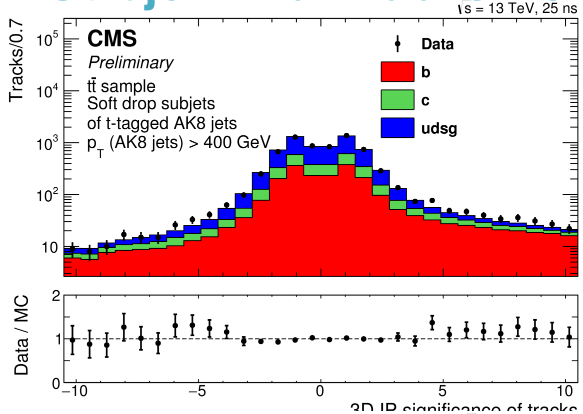
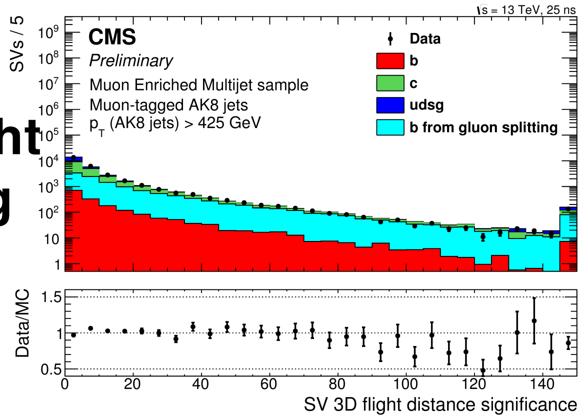
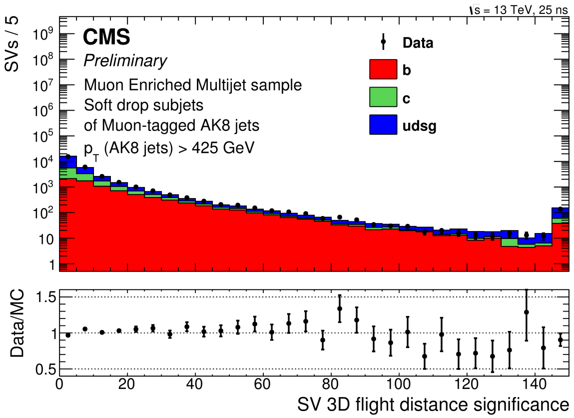
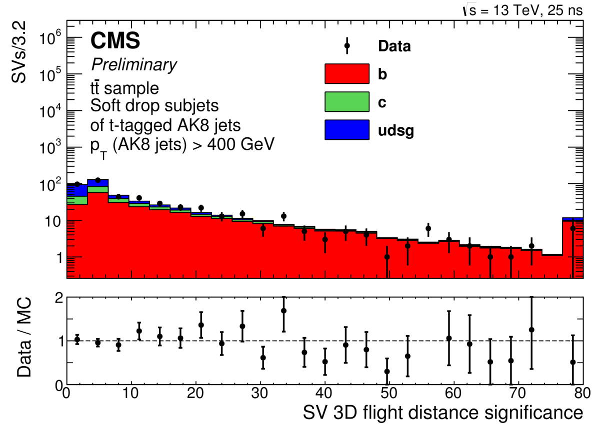
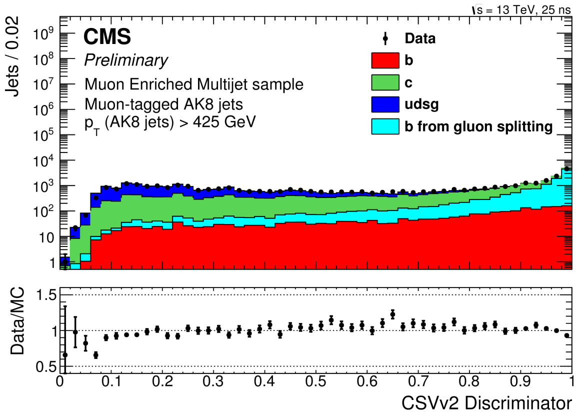
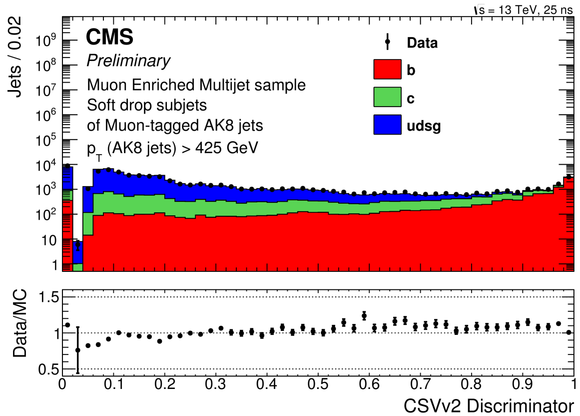
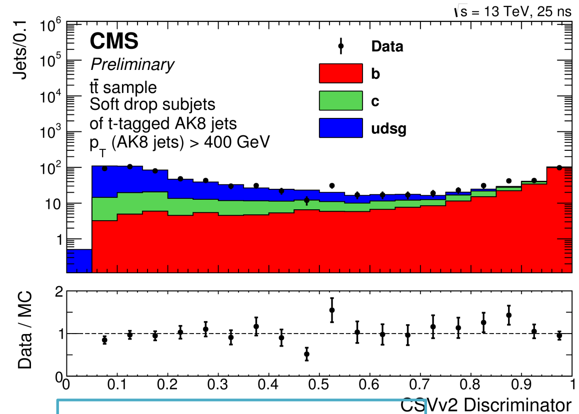
CSVv2

JP



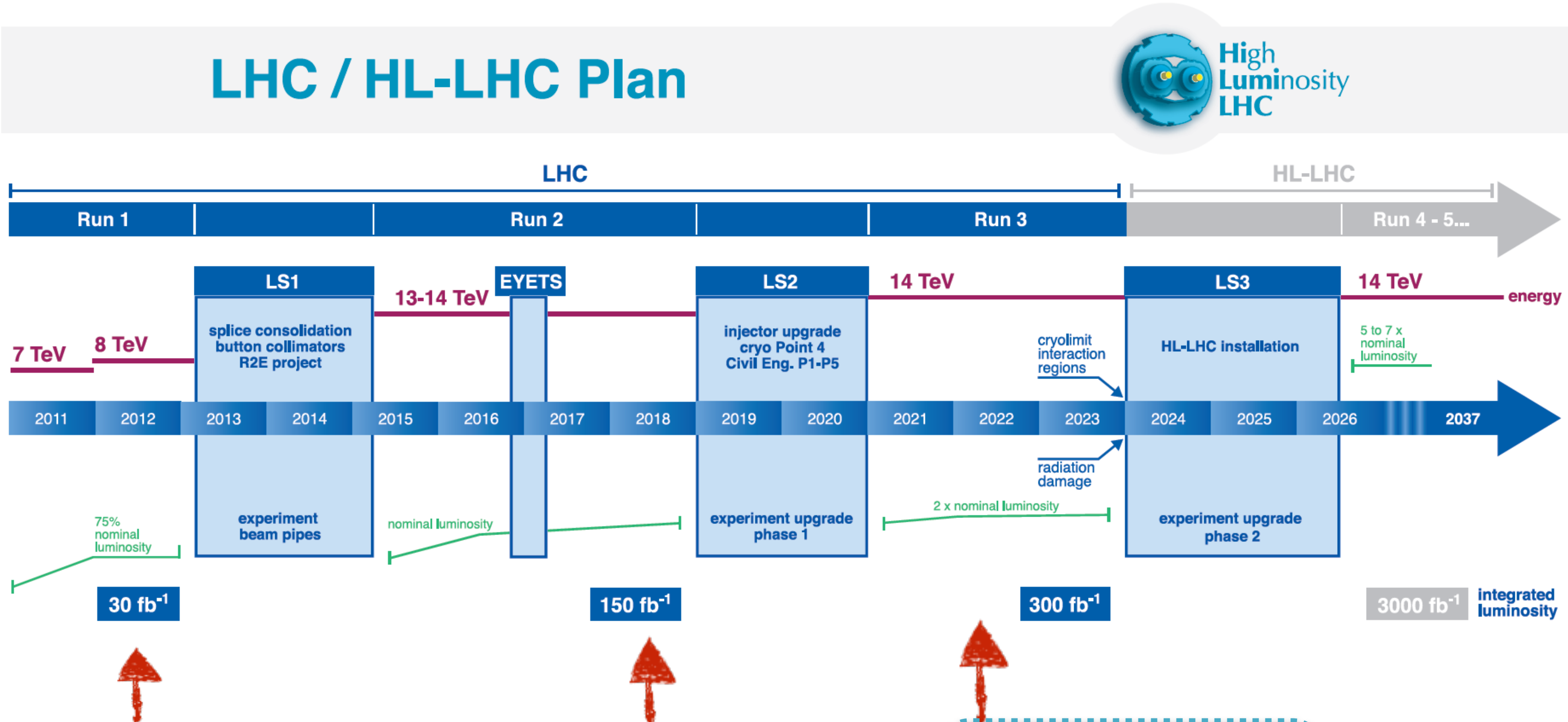
ref: BTV-15-001

Commissioning for ak8 jets

	μ-enriched fat jet	Subjet of μ-enriched fat	Subjet in fat hadronic t
IP sig of tracks			
SV flight dist sig			
CSVv2			

ref: BTV-15-001

LHC planning



L = 0.75×10^{34}
50ns bunch
pileup ≈ 25

L = 1.5×10^{34}
25ns bunch
pileup ≈ 20-50

L = $1.7\text{-}2.2 \times 10^{34}$
25ns bunch
pileup ≈ 60

ref: http://hilumilhc.web.cern.ch