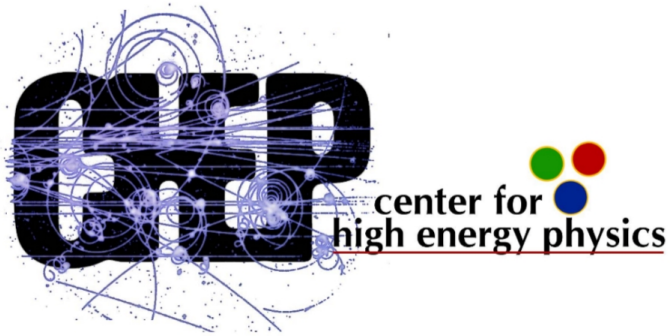
Heavy flavour jet identification at CMS

Jyothsna Rani Komaragiri
Centre for High Energy Physics (CHEP),
Indian Institute of Science (IISc), Bengaluru

for CMS collaboration

Jets@LHC : Discussion Meeting
21st – 28th Jan 2017
@
ICTS, Bengaluru

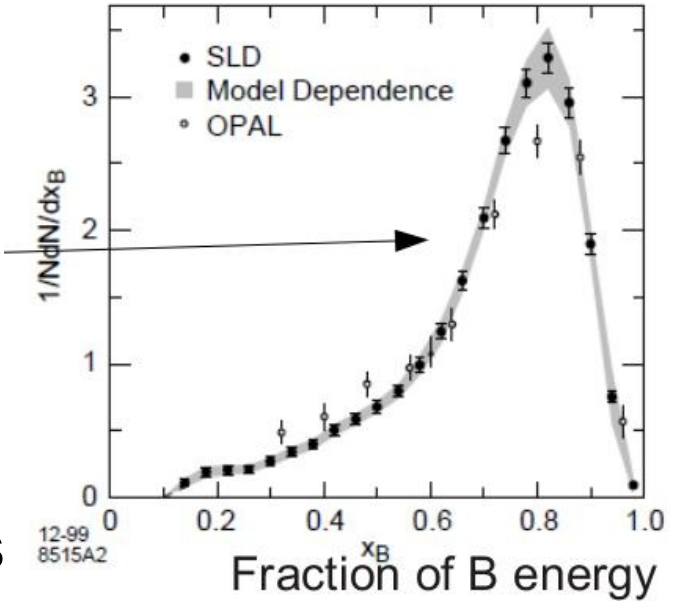


Outline

- **Introduction**
- **Heavy flavour tagging**
  - **b – tagging algorithms**
  - **c – tagging algorithms**
  - **Double b – tagger**
- **Performance in simulations**
- **Performance measurements in data**
- **Summary**

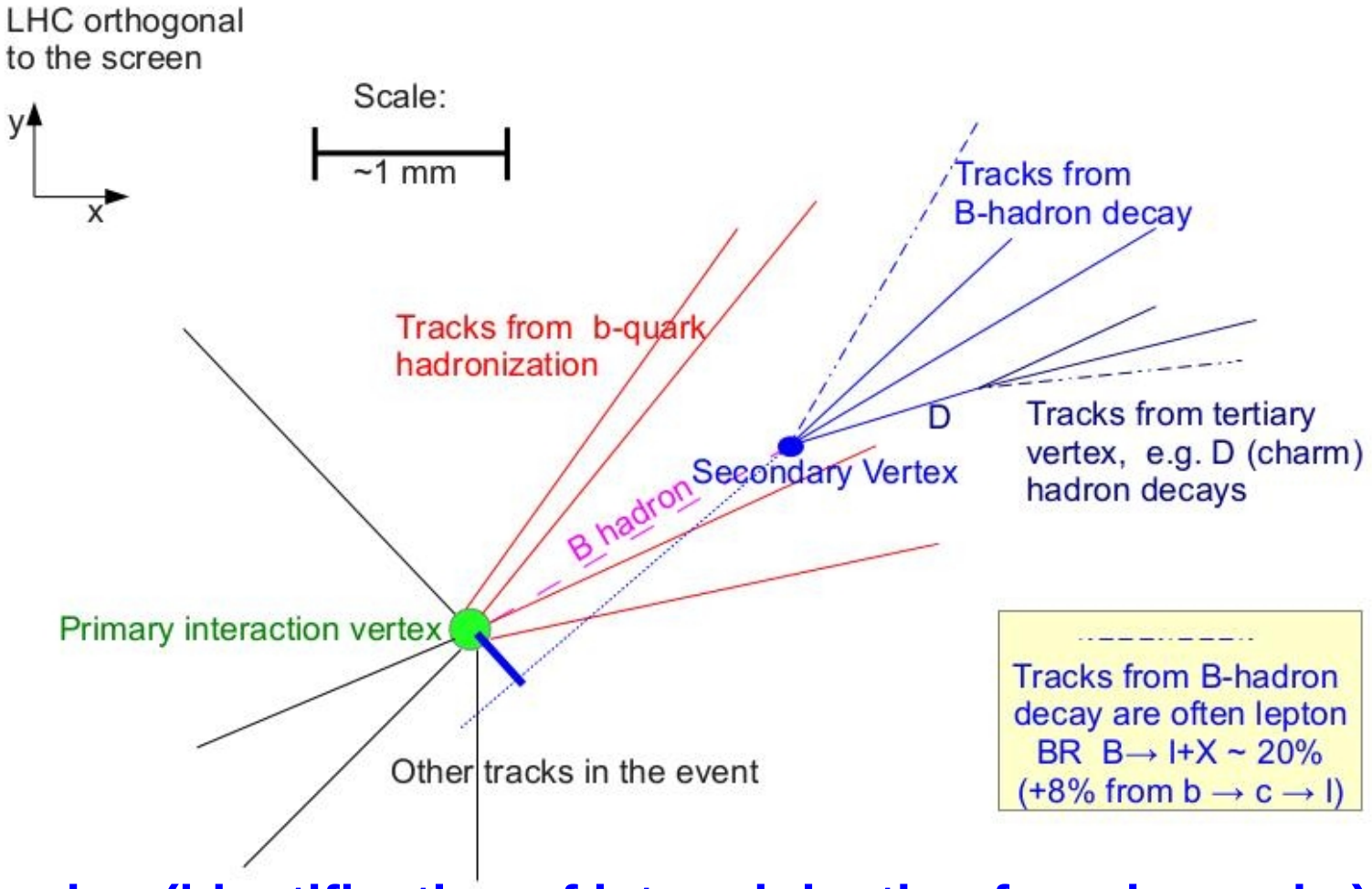
Introduction

- **b-quarks hadronize into B-hadrons (B^+,B^- ,B^0 ,B_s, LambdaB)**
  - **1 B-hadron and several hadronization particles**
- **b-quarks / B-hadrons have sizeable lifetime**
  - **$c\tau$ ~ 500um → $\beta\gamma c\tau$ ~ 5mm @ 50 GeV**
- **The b quark, and then the B-hadrons, have a large mass (~5 GeV)**
- **A large fraction of the original b-quark momentum is carried by the B-hadron**



- **The weak b-decay often produces leptons**

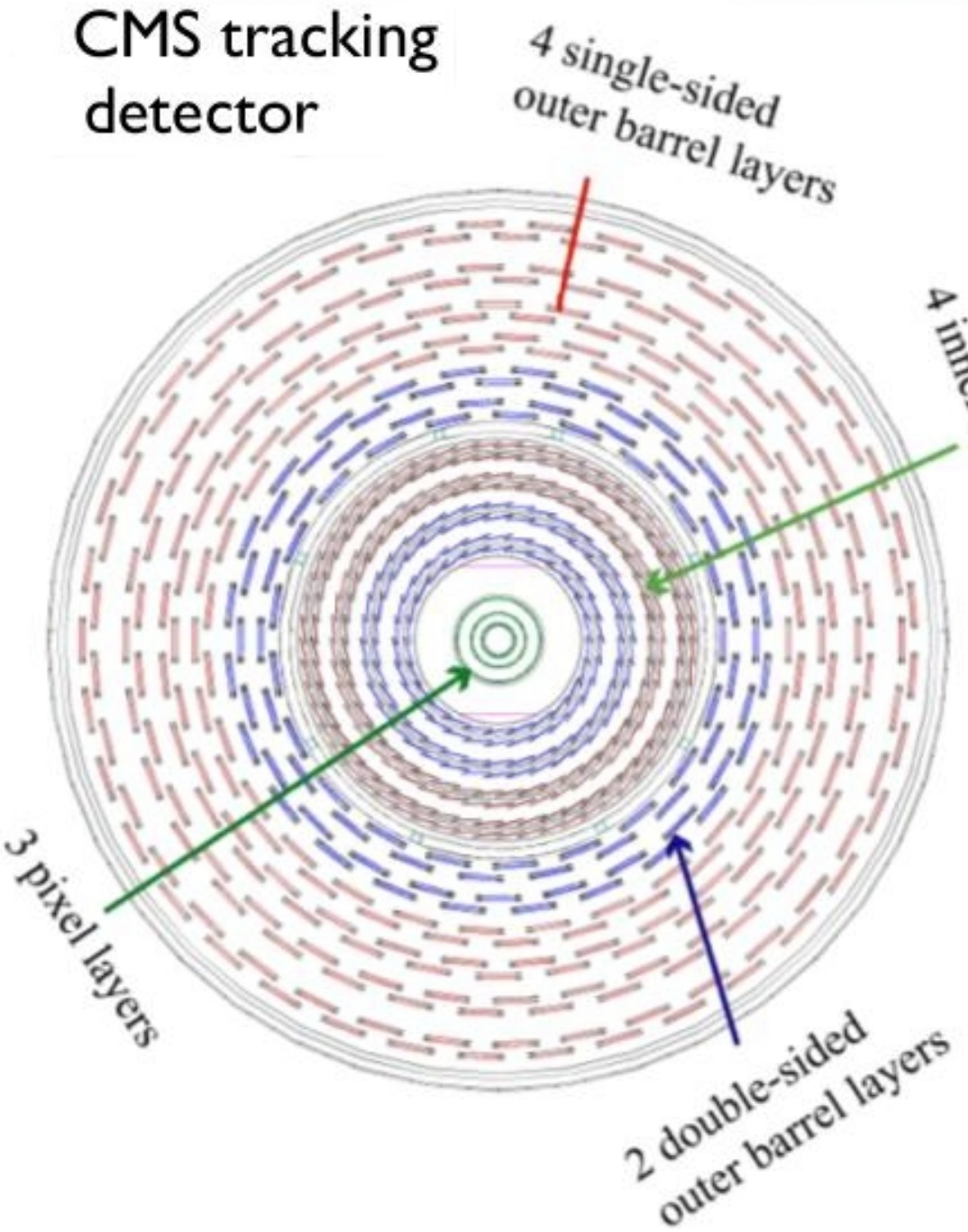
Identification of b-jets



b-tagging (identification of jets originating from b-quarks):
look for displaced tracks and vertices (secondary) within jets

Many SM and BSM events contain b jets:
b-tagging can be used to separate signal from background.

How to look for tracks & vertices?

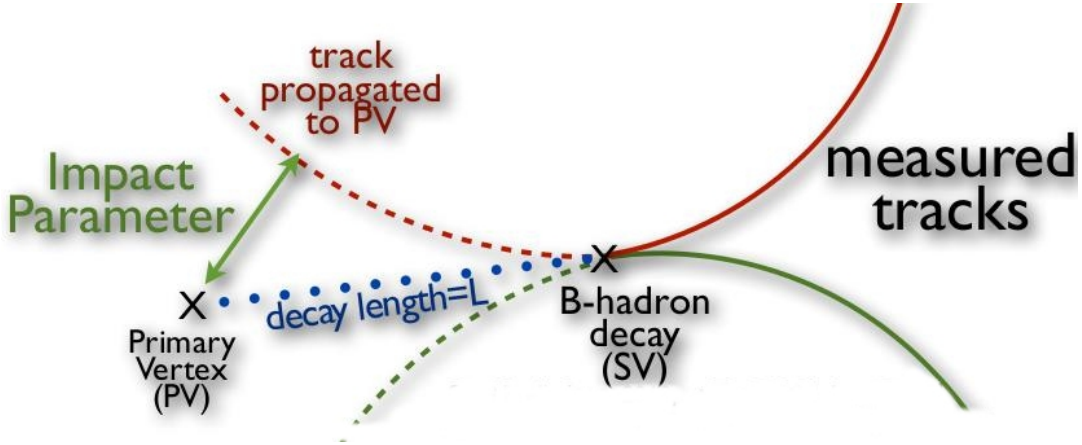


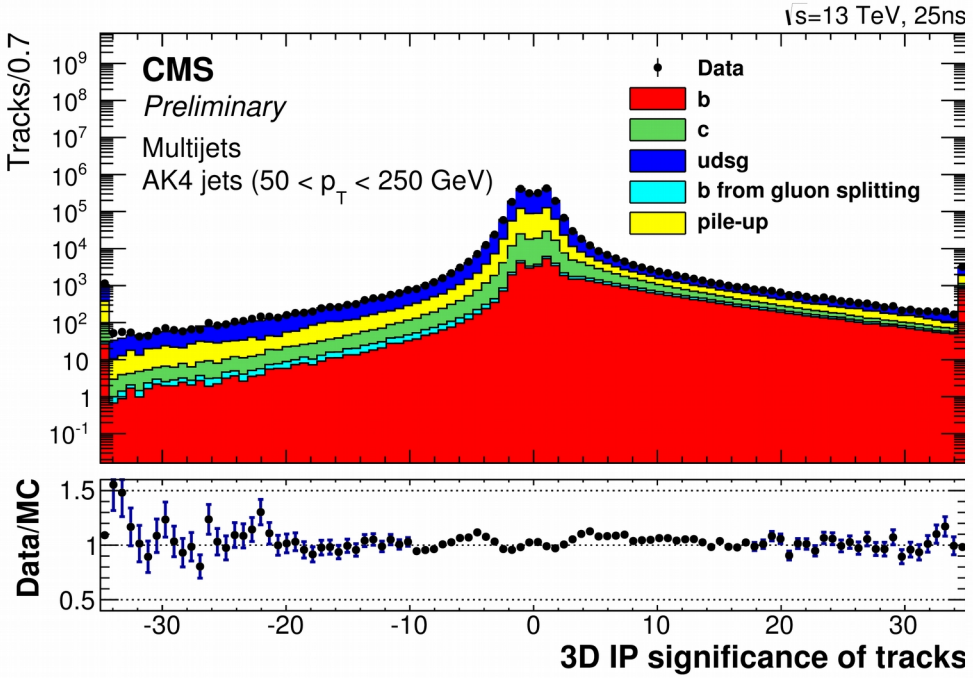
- 10 barrel, 9+3 endcap layers
- radius 1.1 m, length 5.8m
- 200 m^2 active silicon
- acceptance up to $|\eta|<2.5$
- pixel detector central importance: most B hadrons decay within few cm

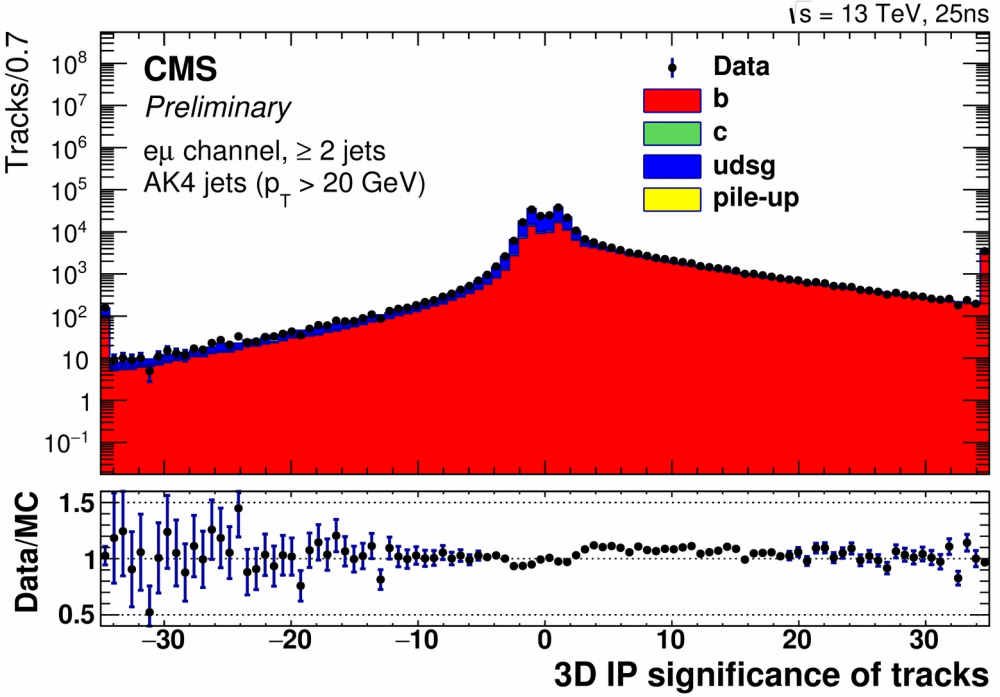
- Tracking combined with information from other detectors (calorimeters) in an optimal way: particle flow (PF) reconstruction
- Tracks denoted as charged PF candidates

Impact parameter of tracks

- **Impact Parameter (IP) defined as distance between track and PV at their point of closest approach**
- **IP has typical size of few 100 μm. Resolution on IP good but mis-measured tracks can cause large uncertainty: 10-100 μm.**
- **Use Track IP significance: S = IP / σ_IP**



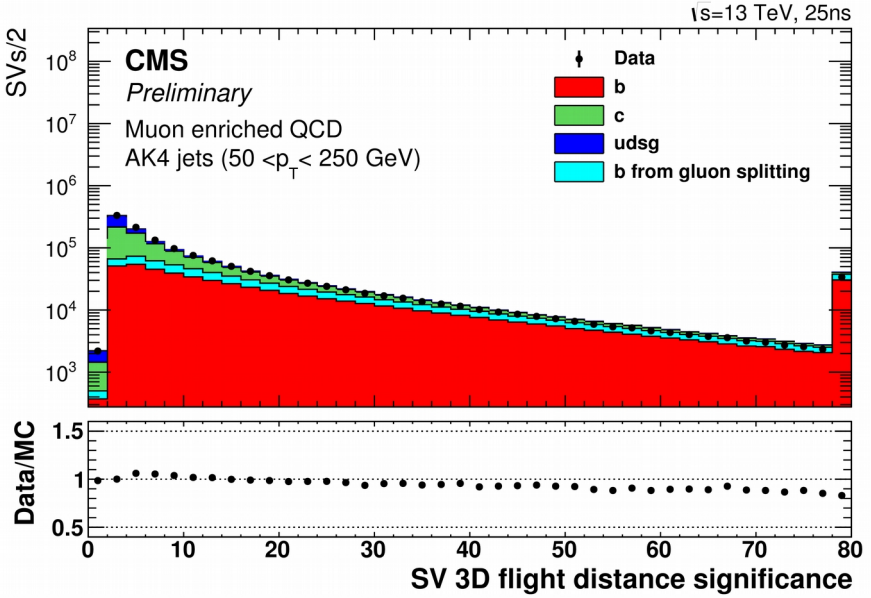


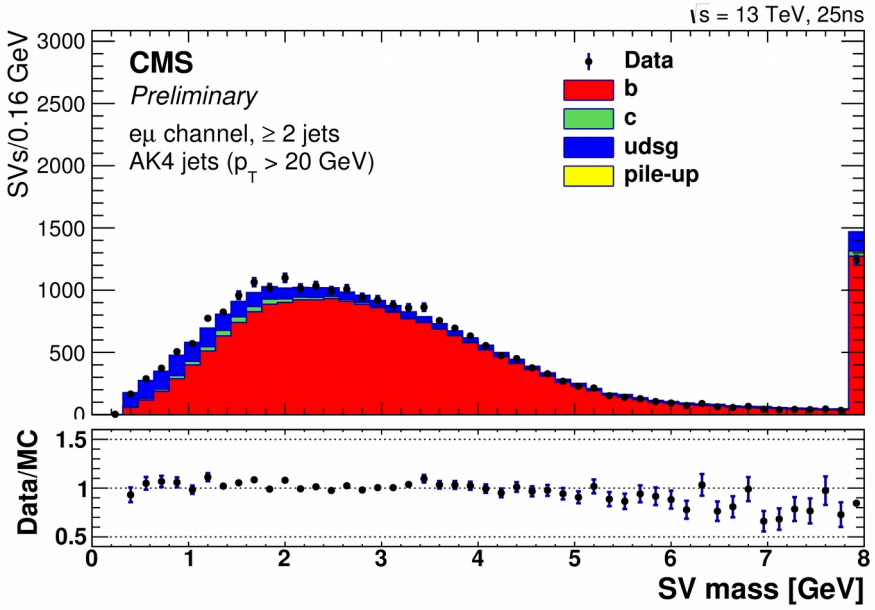


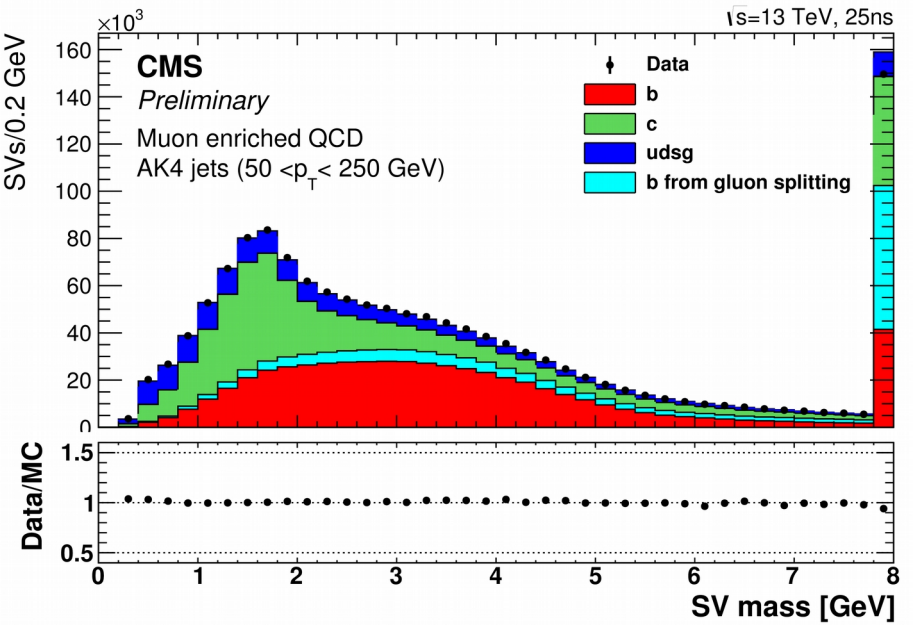
Secondary vertex properties

Some properties that can be used:

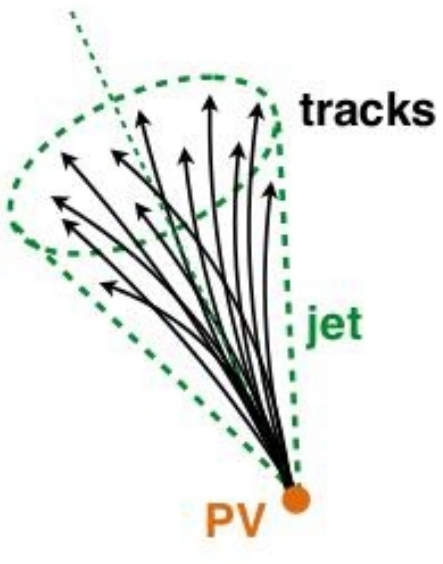
- **Flight distance (significance)**
- **Vertex mass**
- **Number of tracks @ SV**
- **Ratio of energy @SV / jet energy**
- **…..**



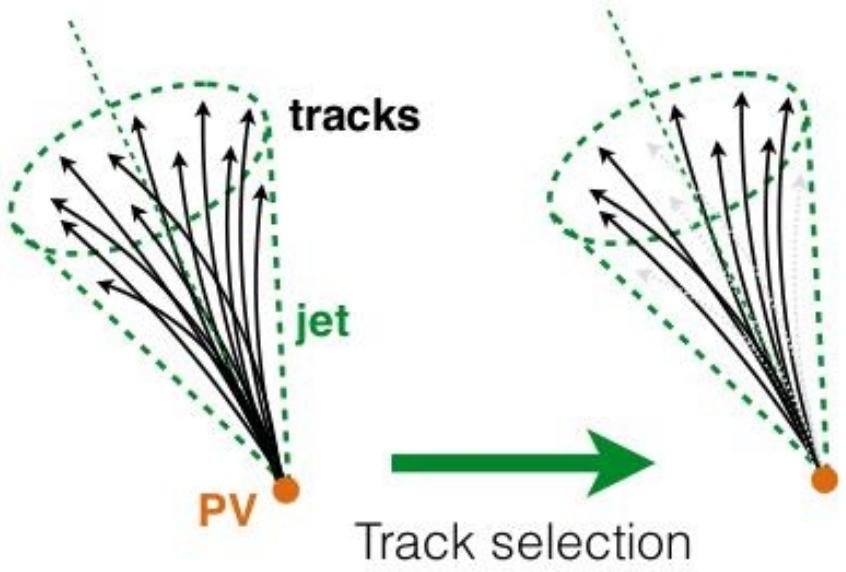




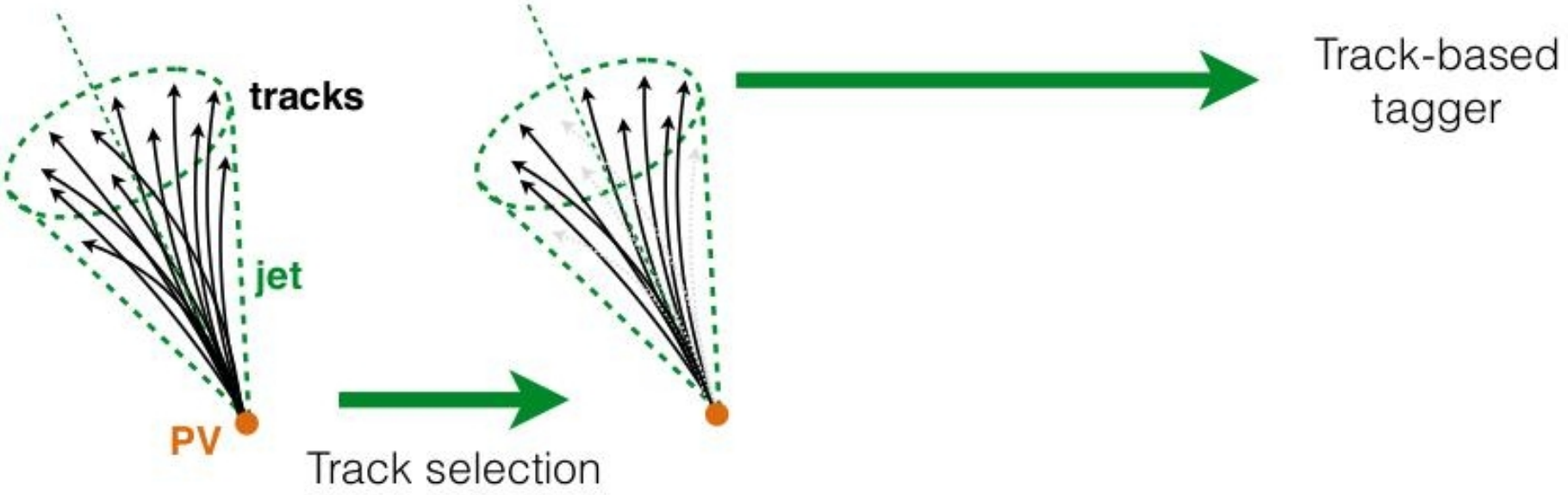
Heavy flavour tagging :: Workflow



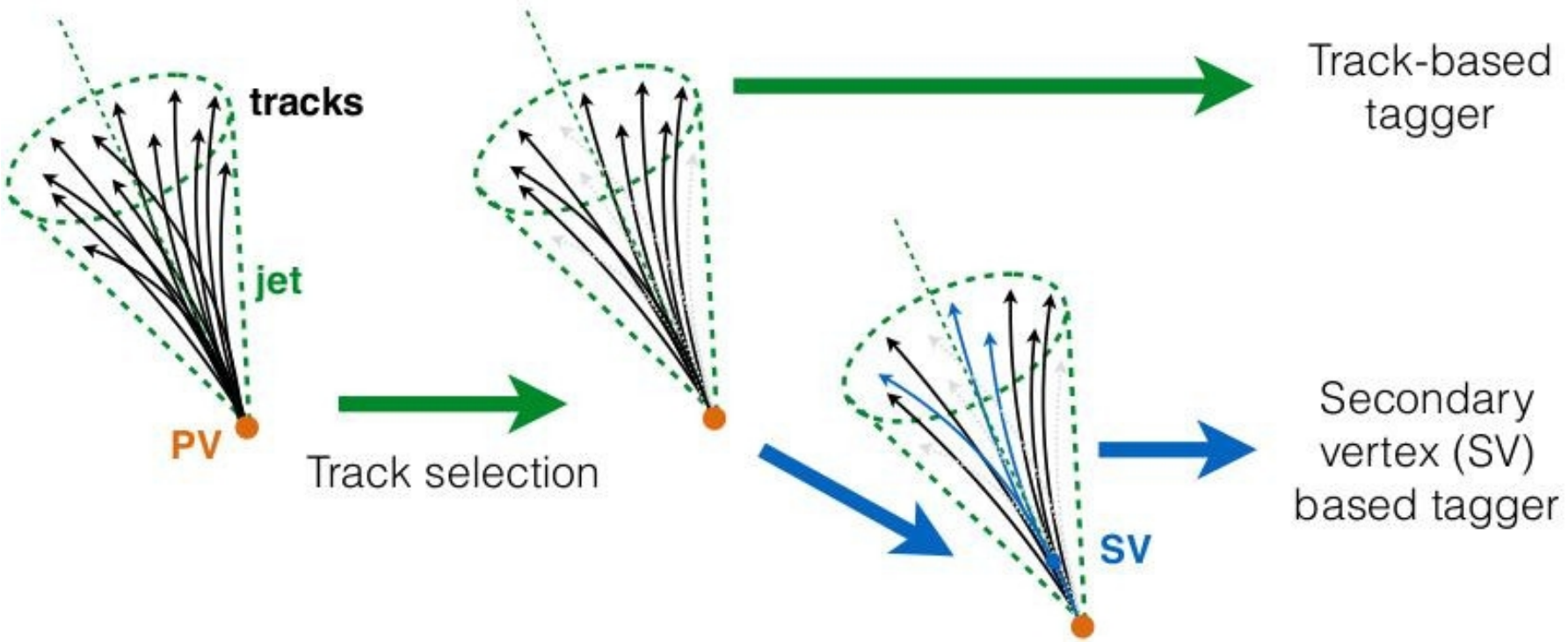
Heavy flavour tagging :: Workflow



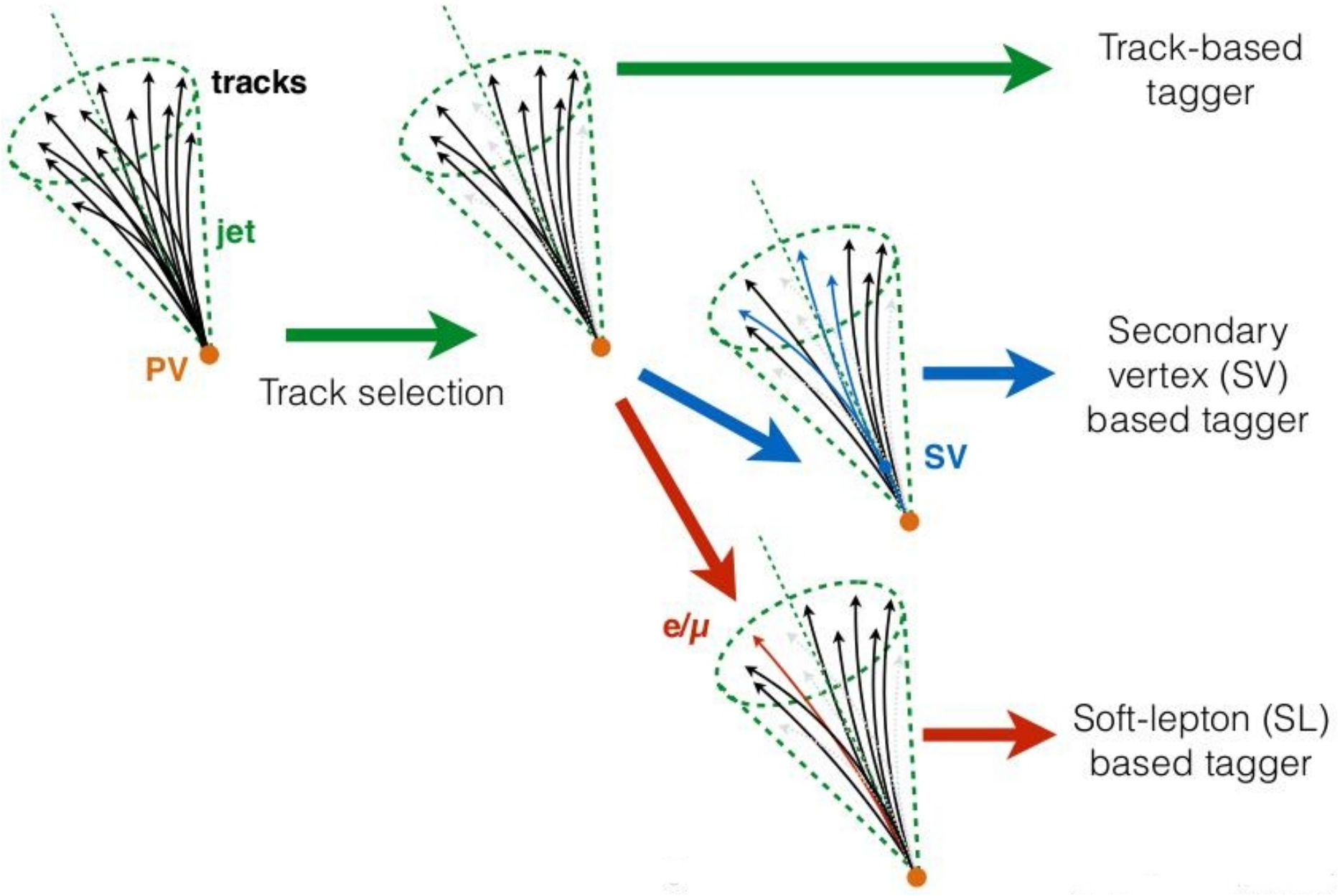
Heavy flavour tagging :: Workflow



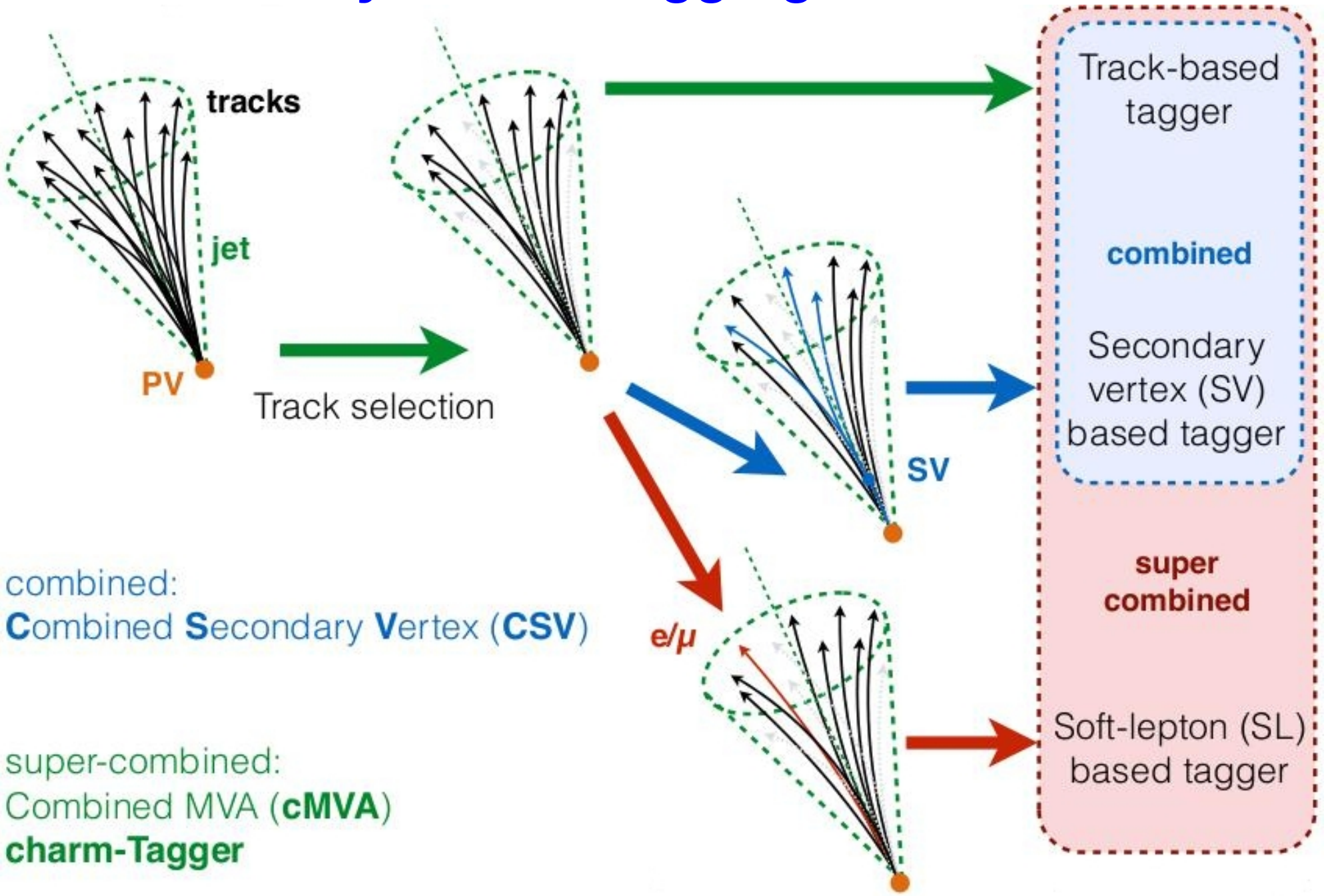
Heavy flavour tagging :: Workflow



Heavy flavour tagging :: Workflow



Heavy flavour tagging :: Workflow



combined:
Combined **S**econdary **V**ertex (**CSV**)

super-combined:
Combined MVA (**cMVA**)
charm-Tagger

b-tagging algorithms

We call b-tagging algorithms OR discriminator computers some pieces of code that:

- **Given the track IP information**
- **and/or given the vertices information**
- **and/or given the lepton information**

- **produces a single float that has the only property of being higher for B and lower for UDSG/C**

For example, more variables are exploited in Combined secondary vertex algorithms

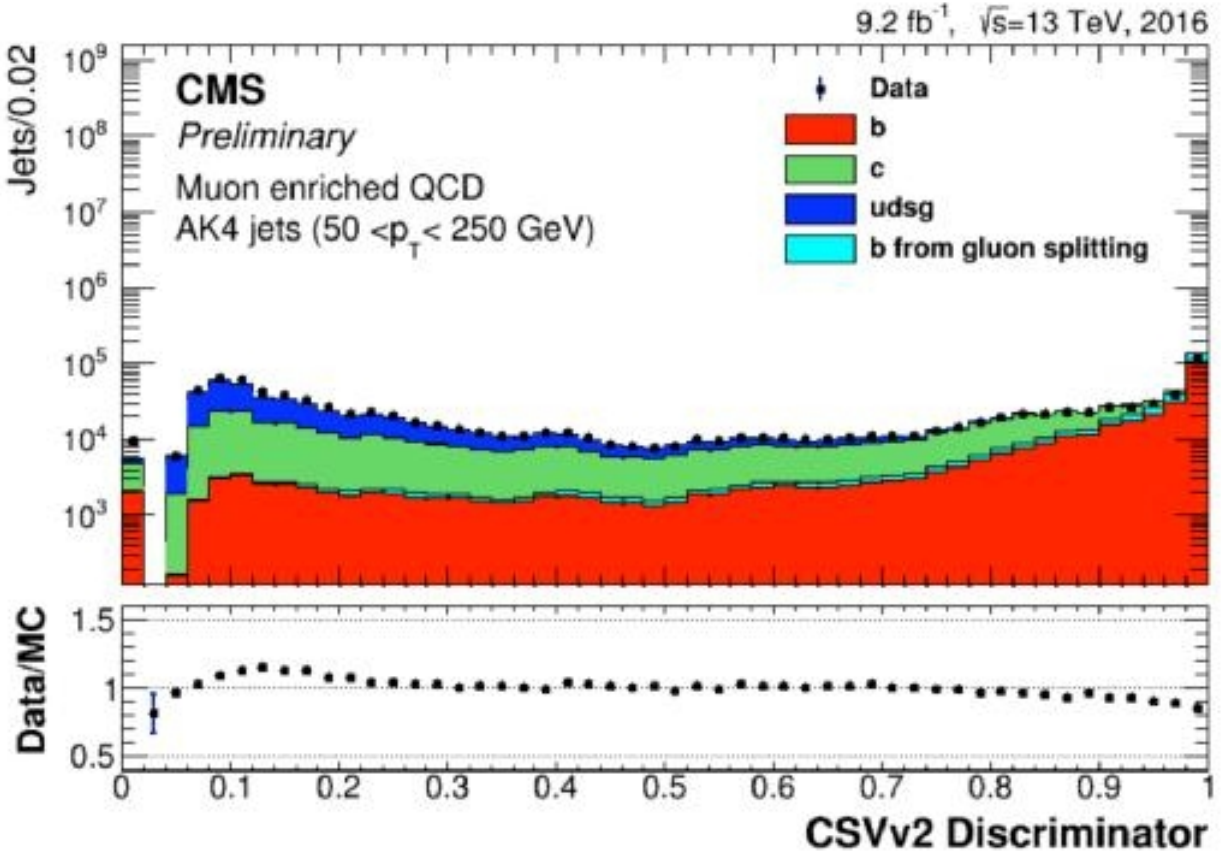
- **Use vertex variables when vertex is available**
- **Make "pseudo vertex" from displace tracks if possible**
- **Use single track variables otherwise**

Multivariate techniques (MVA) can be used to combine all the information

- **Likelihood, Neural Network (NN), Boosted Decision Trees (BDT)**

combined MVA (cMVA) uses Neural Networks

Combined Secondary Vertex b-tagger



Define three operating points based on mistag Probabilities

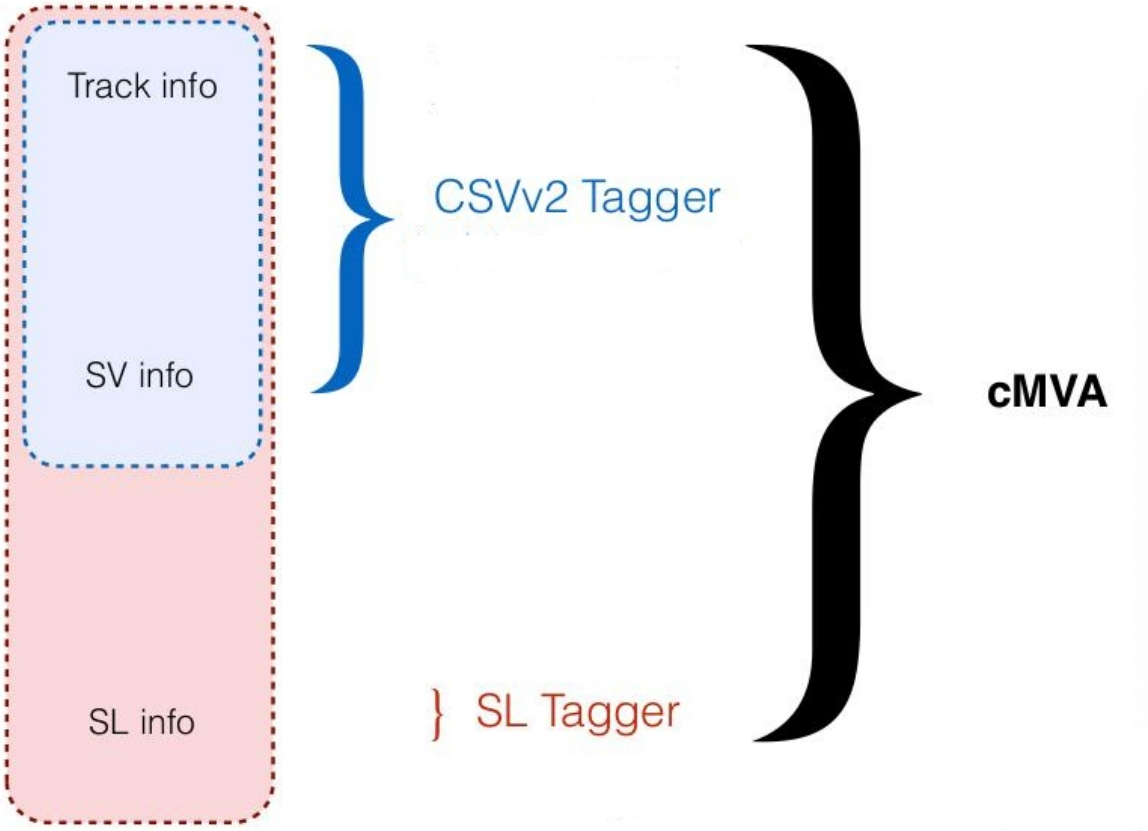
Tight: 0.1%,
Medium: 1%
Loose: 10%

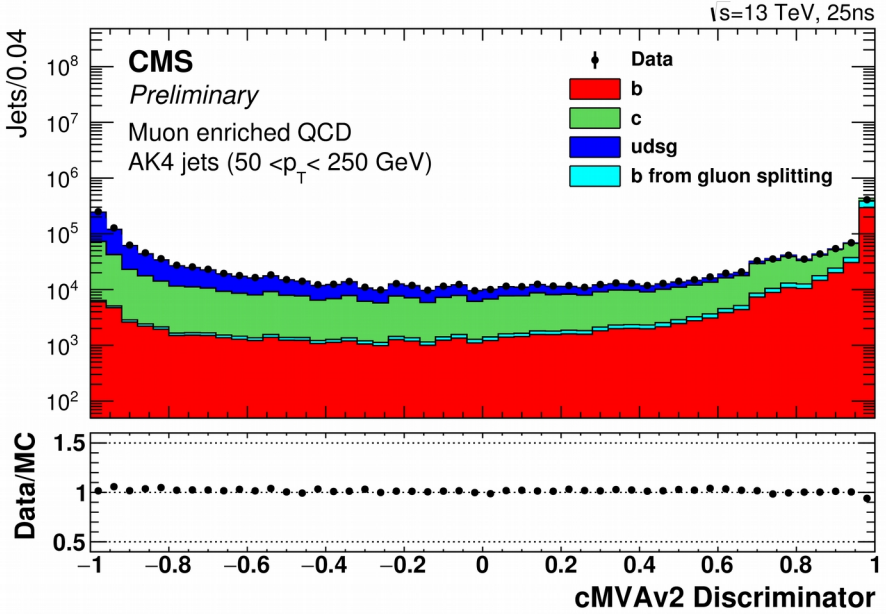
CSVv2>0.935 : Tight
CSVv2>0.80 : Medium
CSVv2> 0.46 : Loose

CSV for AK4 jets in muon enriched QCD

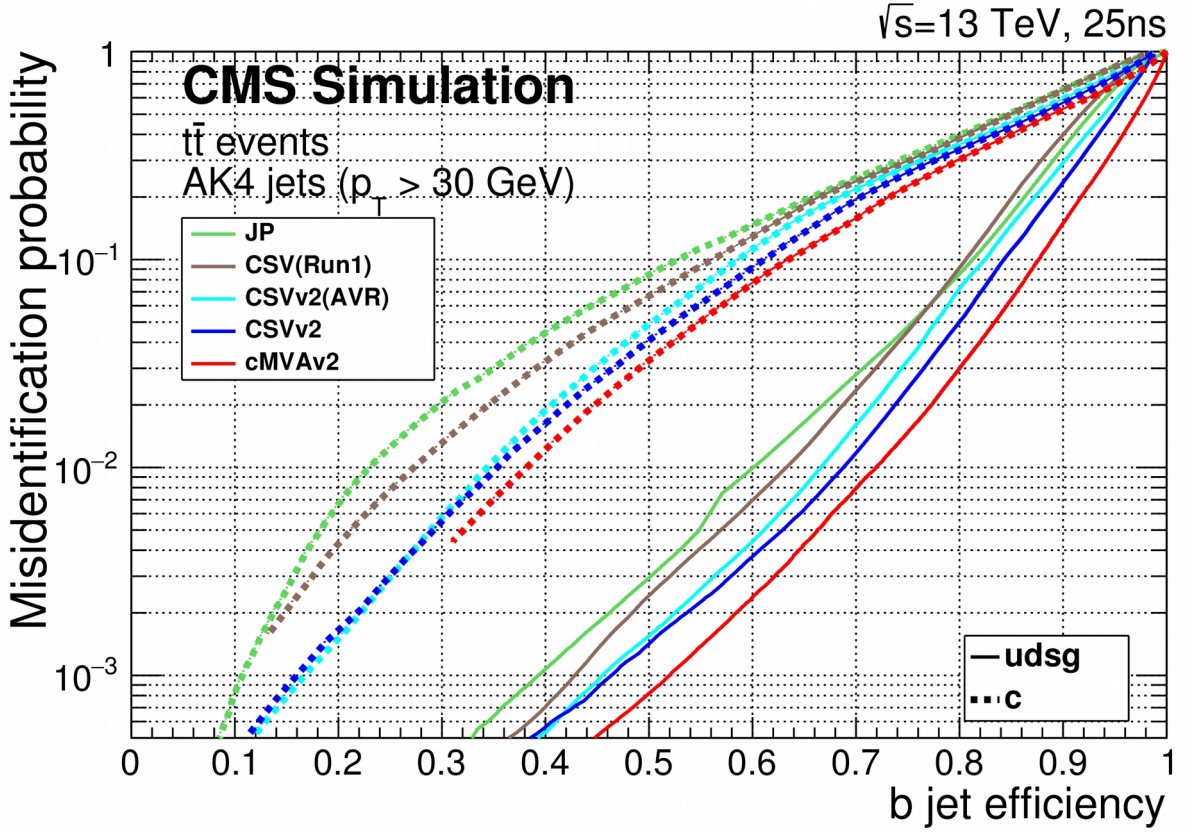
CMS-DP-2016-042

combined MVA b-tagger :: cMVA





Performance of b-taggers in simulation



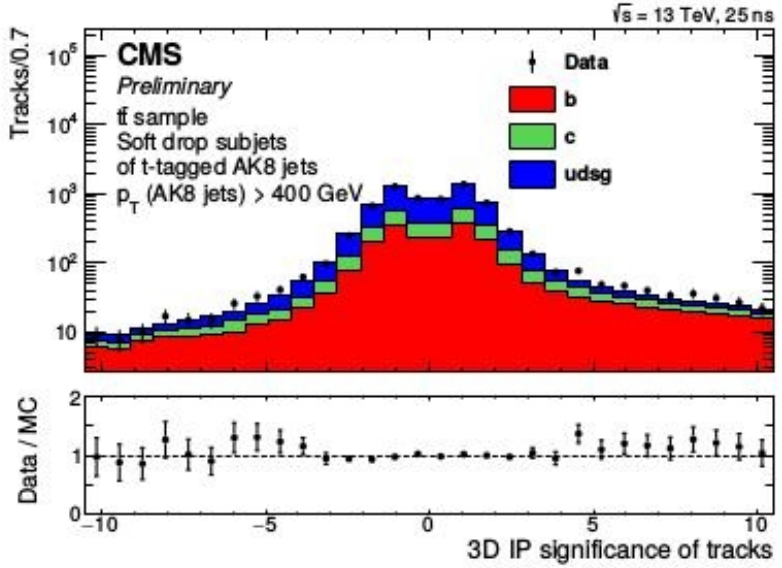
A variety of algorithms to select b-quark jets based on the impact parameters of the charged-particle tracks, reconstructed decay vertices, and the presence of a lepton, or combinations.

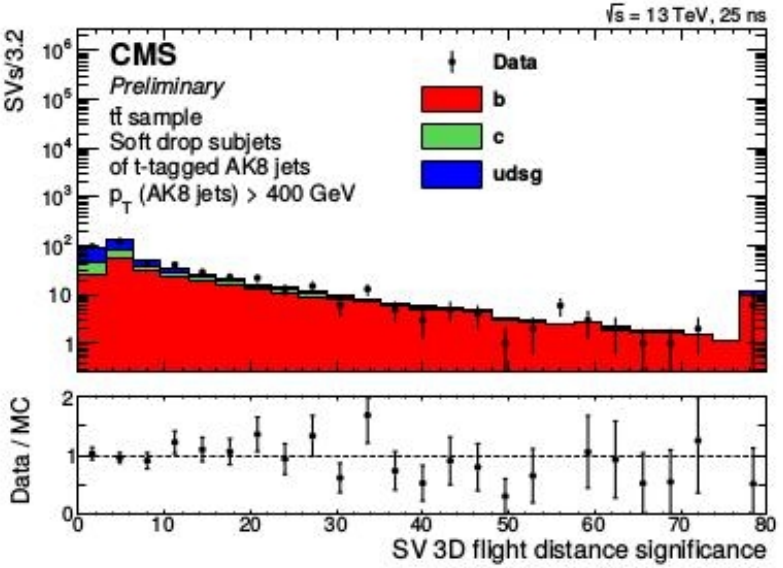
CMS-PAS-BTV-15-001

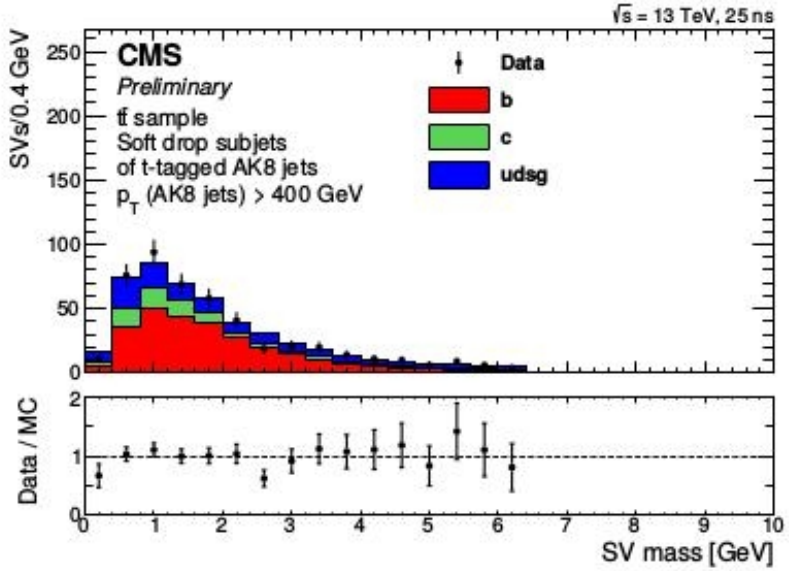
- **Combined Secondary Vertex, CSV (tracks + SV)**
- **CombinedMVA, cMVA (tracks+SV+soft lepton)**

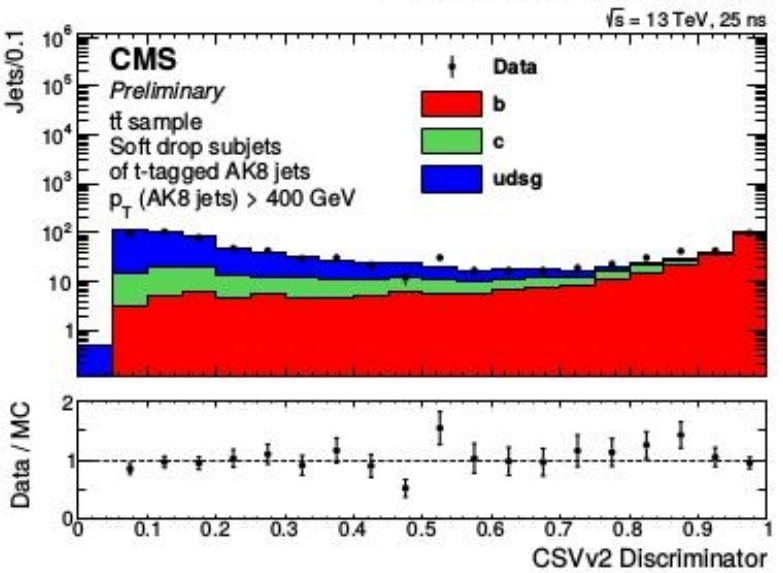
- **Subjet CSV for boosted jets**
- **Two new algorithms: double b-tagger (boosted) & charm tagger**

Subjet CSV for boosted jets





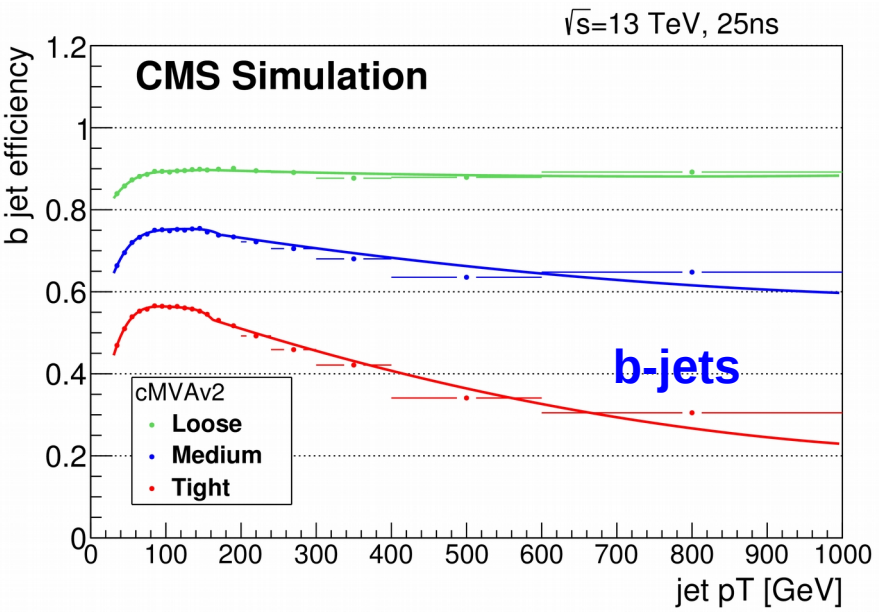


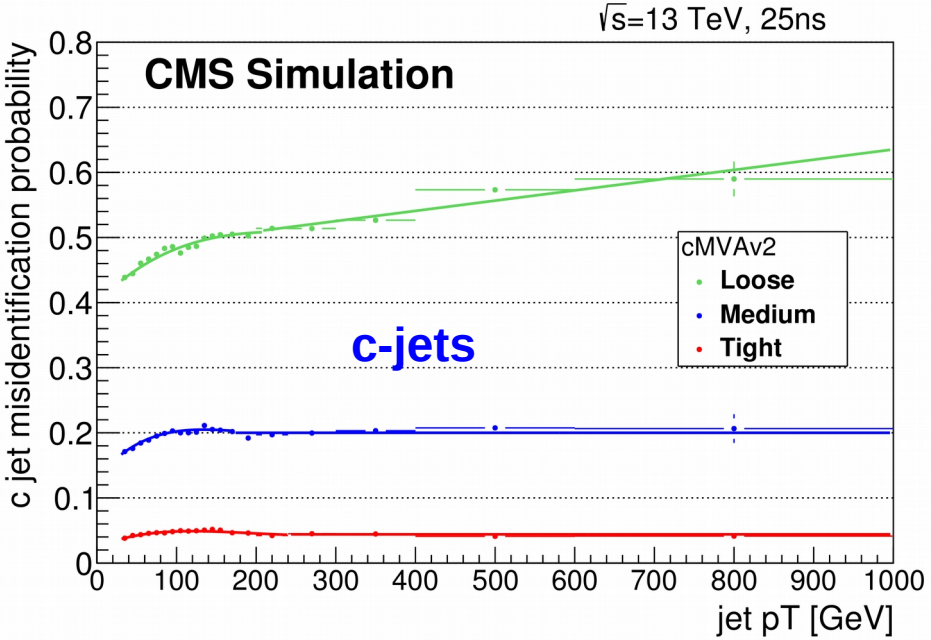


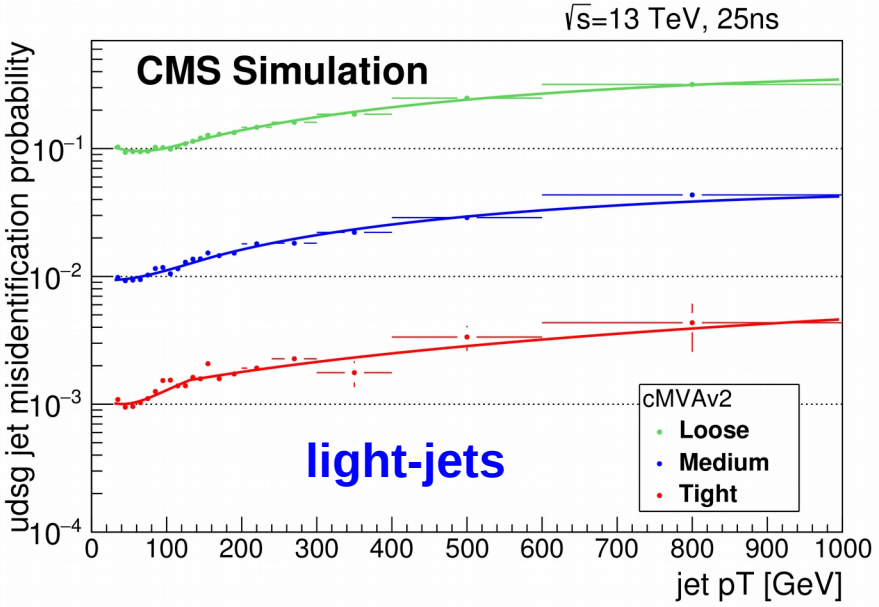
Parameterization of b-tagging efficiencies

- **b-tagging efficiencies provided by collaboration are often useful for theorists and phenomenologists looking for accurate detector simulation (DELPHES etc.)**

- **CMS has provided b-tagging efficiencies in parametric form in CMS-BTV-15-001**

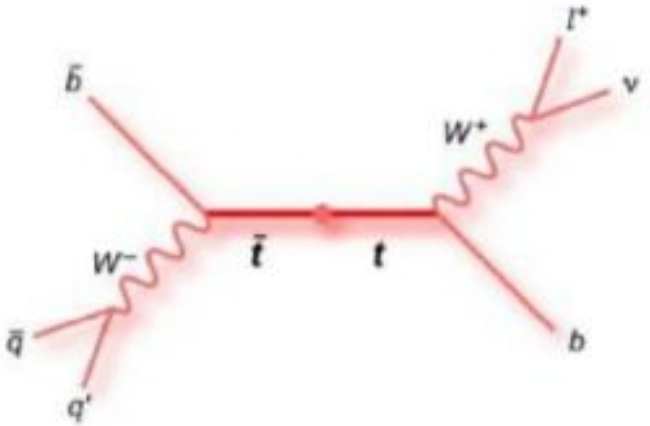


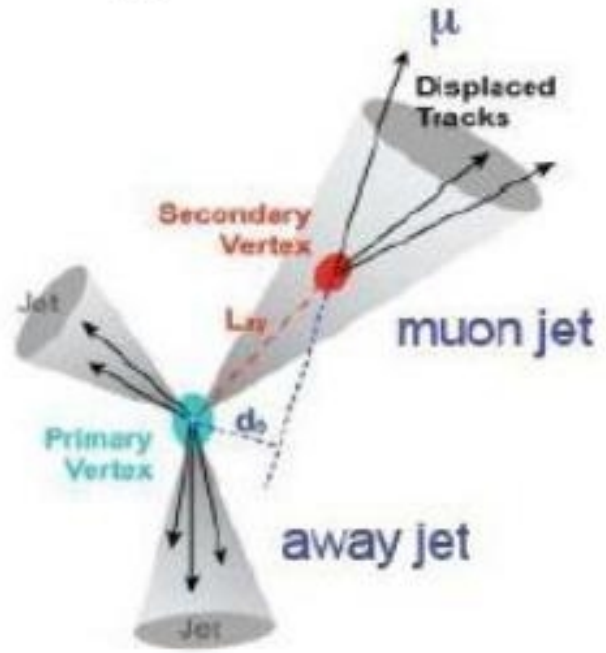




Performance on Data

- **Semi-leptonic ttbar events** are an excellent place to measure b-tag efficiencies:
  - top quarks decay to Wb with BR close to 100%;
  - isolated leptons from W decays allow background reduction.

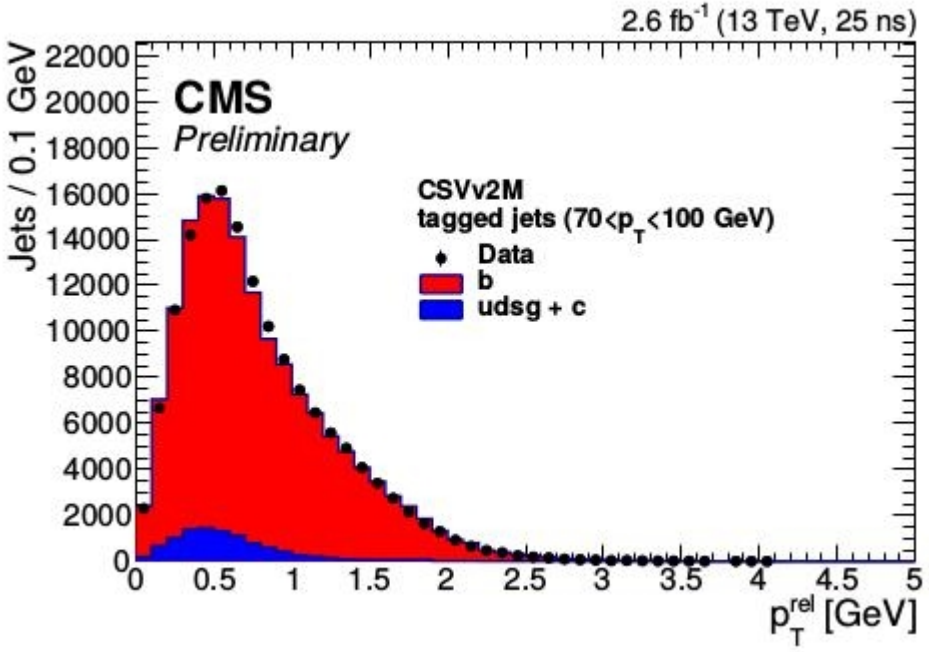


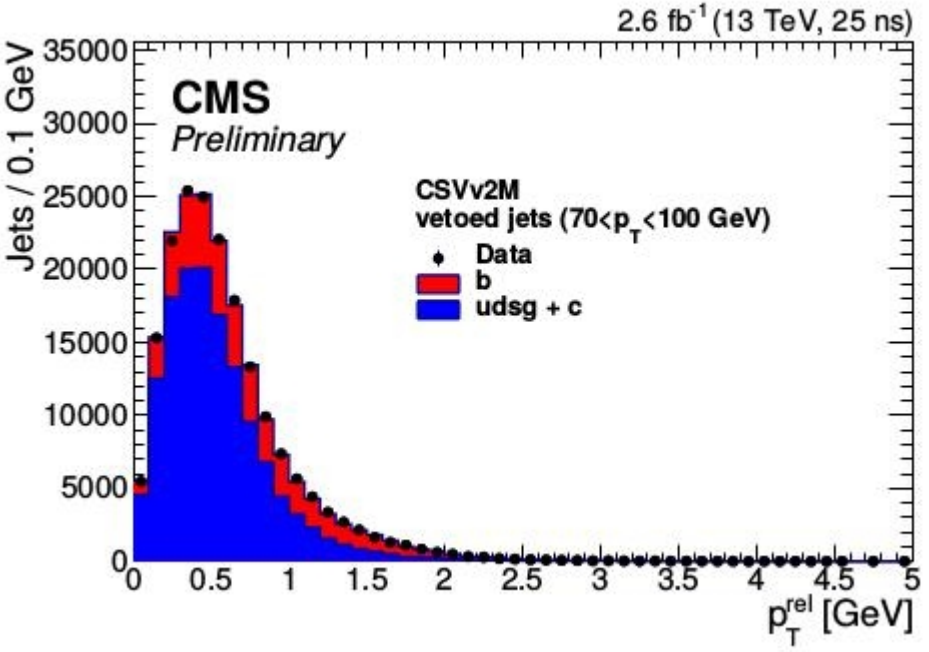


- **Jets containing soft muon** are an alternative:
  - high semileptonic BR of B-hadrons, give b-enriched sample;
  - requiring additionally tagged jet (away jet), to further enrich the sample in bb̄.

pT^{Rel} method : Efficiency measurement

- **Split sample with jets containing muons into two sets:**
  b-tagged (left) and not b-tagged (right)



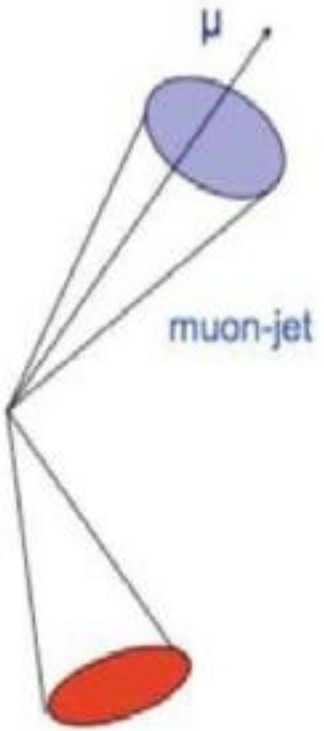


- **Fit the blue (light jets) and red (b-jet) shapes to match the data, extract the fractions f_b^{tag} and f_b^{untag}**

$$\varepsilon_{\mathrm{b}}^{\mathrm{tag}} = \frac{f_{\mathrm{b}}^{\mathrm{tag}} \cdot N_{\mathrm{data}}^{\mathrm{tag}}}{f_{\mathrm{b}}^{\mathrm{tag}} \cdot N_{\mathrm{data}}^{\mathrm{tag}} + f_{\mathrm{b}}^{\mathrm{untag}} \cdot N_{\mathrm{data}}^{\mathrm{untag}}}$$

System8 method : Efficiency measurement

- Apply two weakly correlated taggers to jet with muon:
  - p_T^{rel} > 0.8 GeV cut on the muon (**tag**): p_T of the muon relative to the jet direction. Due to the b-mass, its decay products have a few GeV momentum in its rest frame.
  - tagger whose efficiency being measured (**probe**).



- Two samples: tagged/untagged **away jet**.
- System of **8 equations**, with 8 unknowns.
- **MC gives only the correlation** between two tags on the jet containing the soft-muon.

System8 method : Efficiency measurement

$$n = n_b + n_{cl} \qquad \text{muon-in-jet + away-jet}$$

$$p = p_b + p_{cl} \qquad \text{muon-in-jet + tagged-away-jet}$$

$$n^{\text{tag}} = \varepsilon_b^{\text{tag}} n_b + \varepsilon_{cl}^{\text{tag}} n_{cl}$$

$$p^{\text{tag}} = \beta\, \varepsilon_b^{\text{tag}} p_b + \alpha\, \varepsilon_{cl}^{\text{tag}} p_{cl} \qquad \text{apply "probe" tagger}$$

$$n^{\text{pTrel}} = \varepsilon_b^{\text{pTrel}} n_b + \varepsilon_{cl}^{\text{pTrel}} n_{cl}$$

$$p^{\text{pTrel}} = \delta\, \varepsilon_b^{\text{pTrel}} p_b + \gamma\, \varepsilon_{cl}^{\text{pTrel}} p_{cl} \qquad \text{apply "tag" tagger}$$

$$n^{\text{tag,pTrel}} = \kappa_b\, \varepsilon_b^{\text{tag}} \varepsilon_b^{\text{pTrel}} n_b + \kappa_{cl}\, \varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{pTrel}} n_{cl}$$

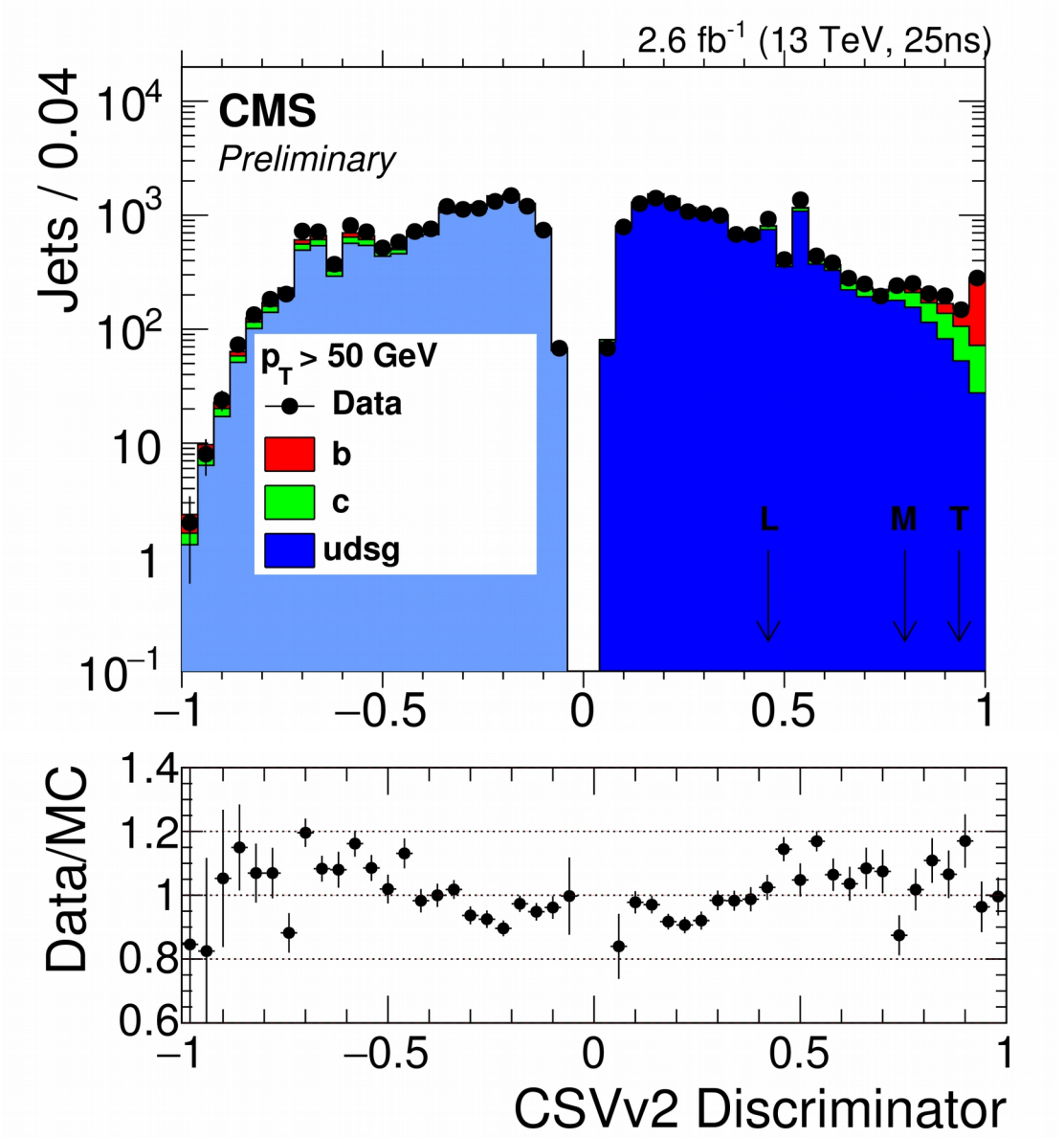
$$p^{\text{tag,pTrel}} = \kappa_b\, \beta\, \delta\, \varepsilon_b^{\text{tag}} \varepsilon_b^{\text{pTrel}} p_b + \kappa_{cl}\, \alpha\, \gamma\, \varepsilon_{cl}^{\text{tag}} \varepsilon_{cl}^{\text{pTrel}} p_{cl} \qquad \text{apply "tag" and "probe" taggers}$$

LHS: observables, GREEK: correlation factors from simulation

Measurement of mistag rate

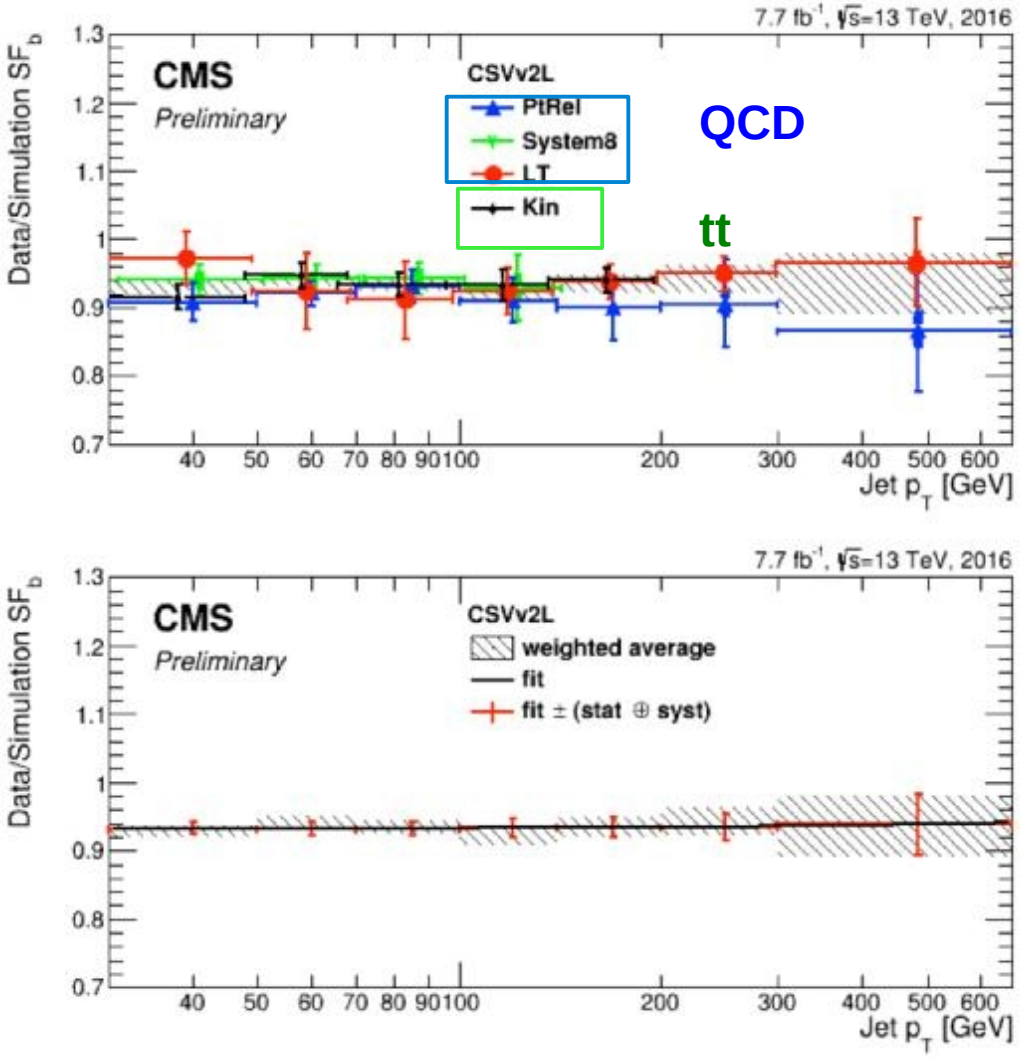
sign creates asymmetric discriminator distributions

➡ negative side can be used to predict positive side for light flavour jets

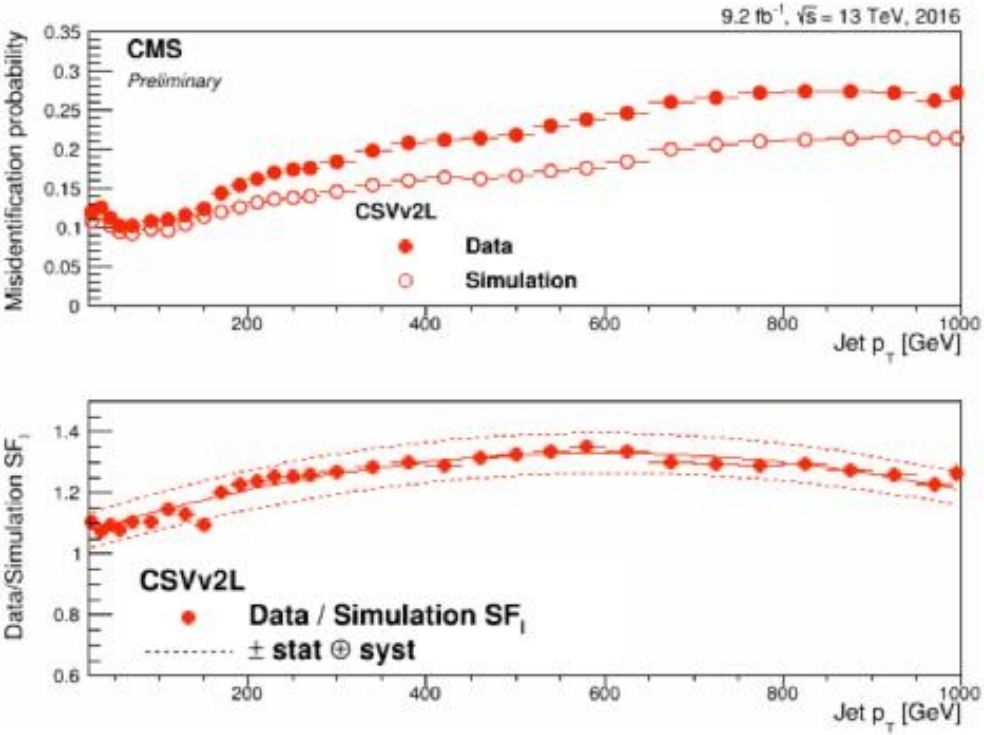


Performance on Data

b-jet Scale Factor



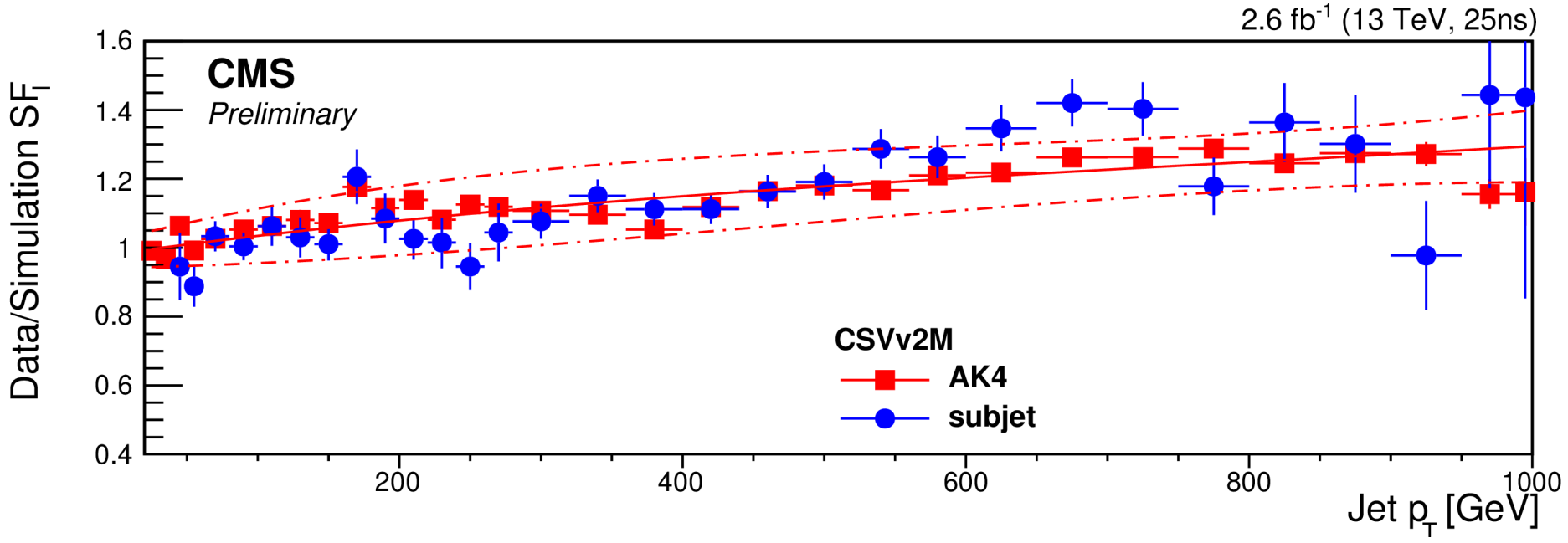
light-jet Scale Factor



CMS-DP-2016-042

Performance on Data

- **Comparison for the CSVv2M scale factor of the misidentification probability obtained with the negative tag method for AK4 jets and softdrop subjets of AK8 jets as a function of the (sub)jet tranvserse momentum.**
- **Within the uncertainty, the scale factor measurements are compatible.**



c-tagger

Track info

SV info

SL info

Fast
Reliable

c-Tagger { **light jets** } **CvsL**

c jets

{ **b jets** } **CvsB**

We need separation against two kind of backgrounds!

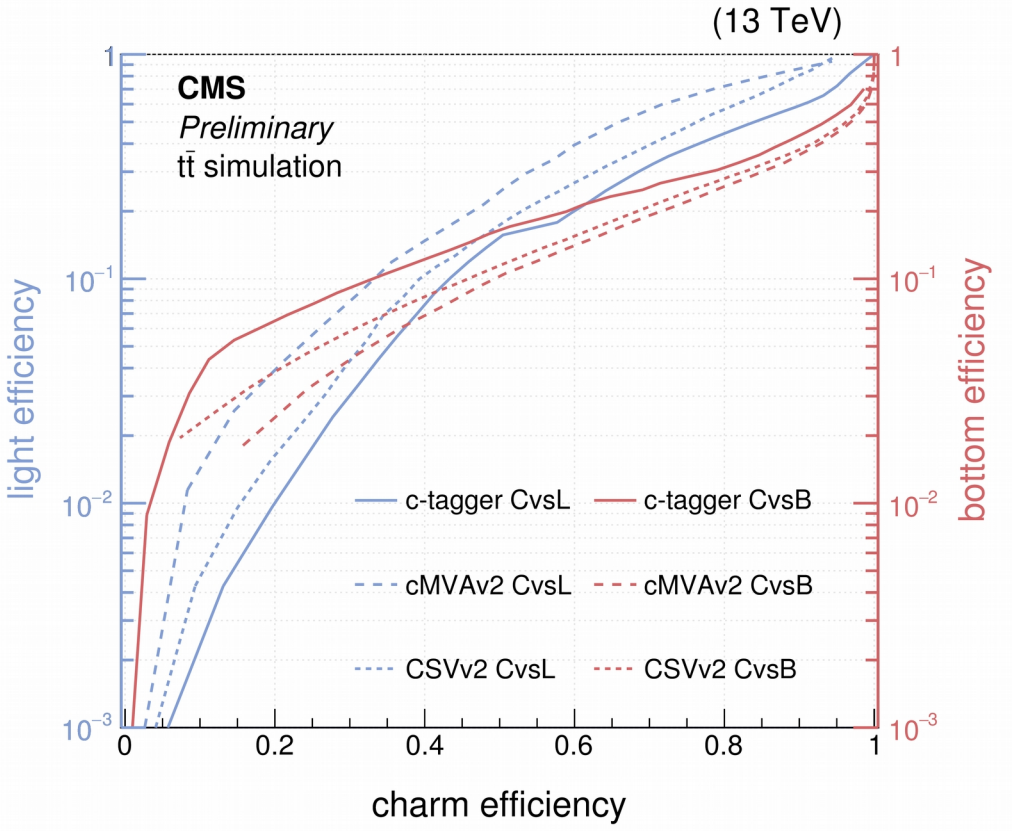
Solution: use two BDTs

CMS-PAS-BTV-16-001

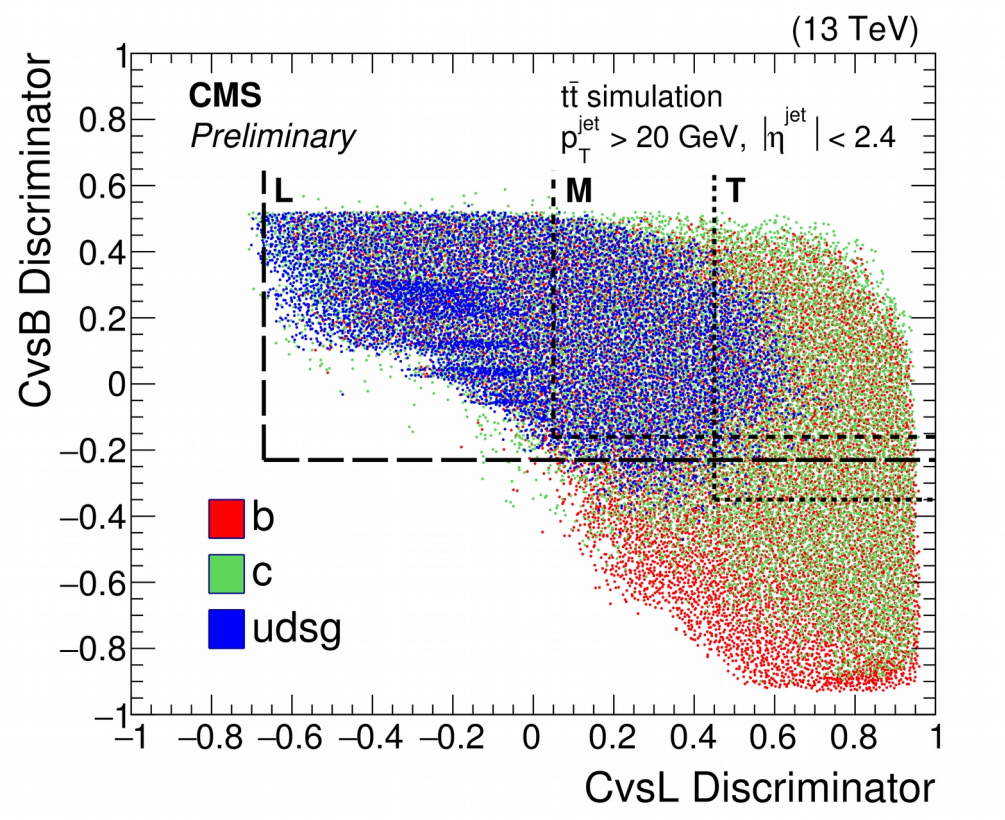
Useful for FCNC t→ Z+c, $H^{\pm}$ → c+s, stop → c+MET etc

c-tagger

Performance of the CvsL (blue full line and axis) and CvsB (red full line and axis) trainings, validated on the tt¯ sample.



Two dimensional scatter overlay of the BDT discriminators for b (red), c (green), and light jets (blue).



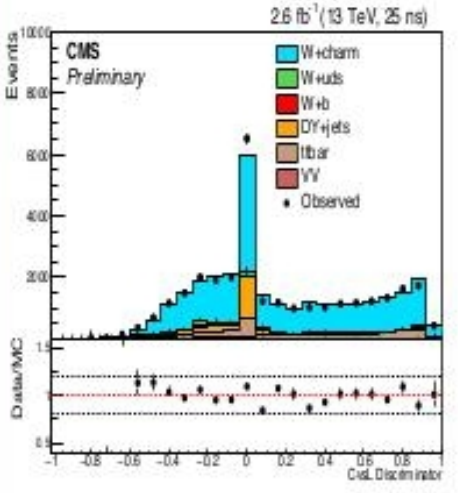
CMS-PAS-BTV-16-001

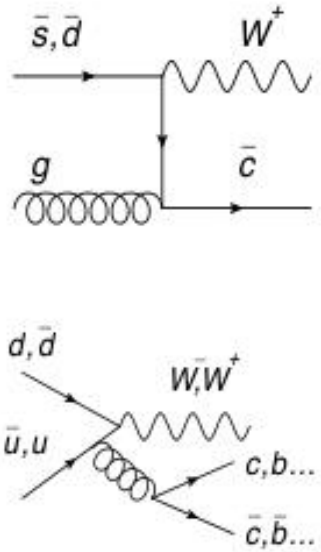
Performance on data :: c-tagger

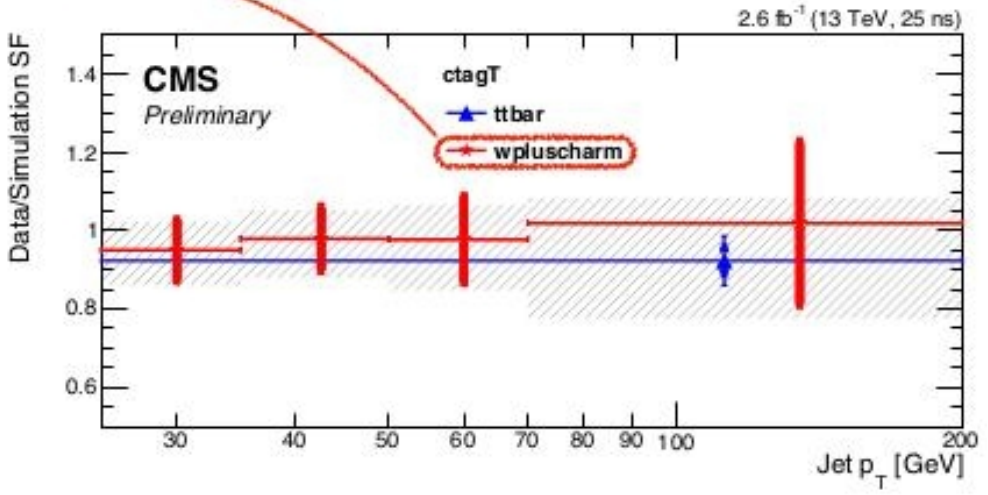
W+c events require a μ in the jet.

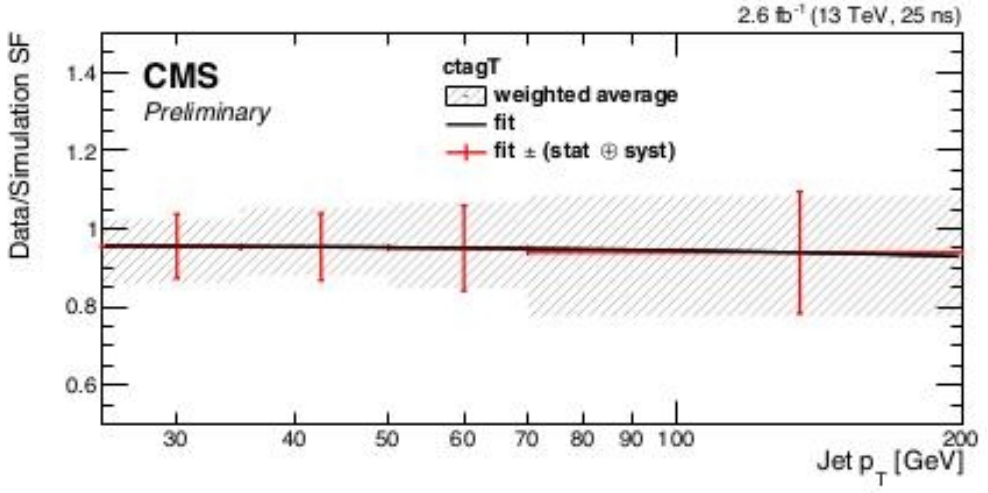
Background removed by OS - SS

High purity sample. Possible p_T(jet) binning









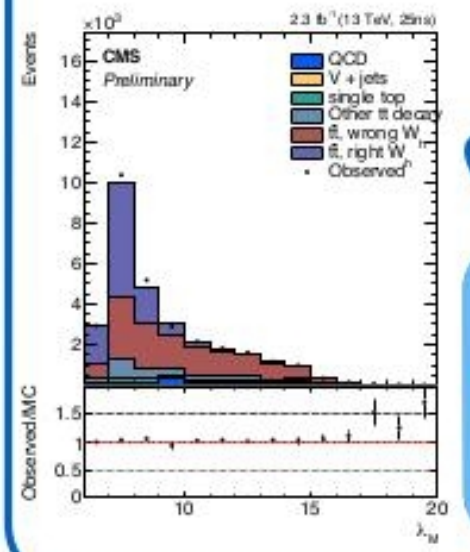
CMS-PAS-BTV-16-001

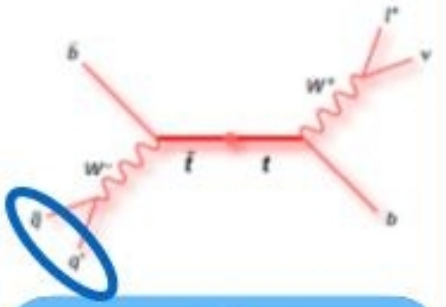
Performance on data :: c-tagger

Semi-leptonic tt events contain 25% of c-jets in the W decay.

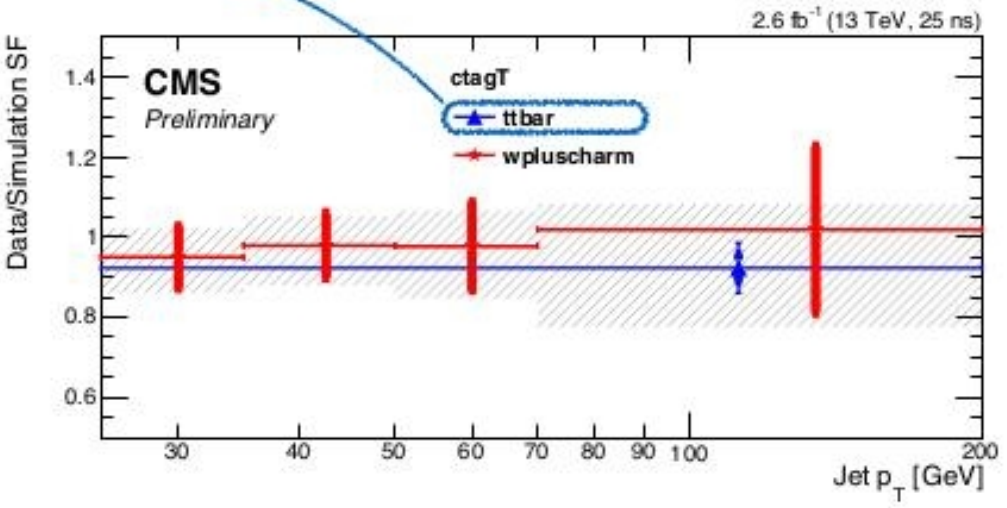
Two jet samples (leading, subleading) enriched/depleted with c-jets by double weak decay

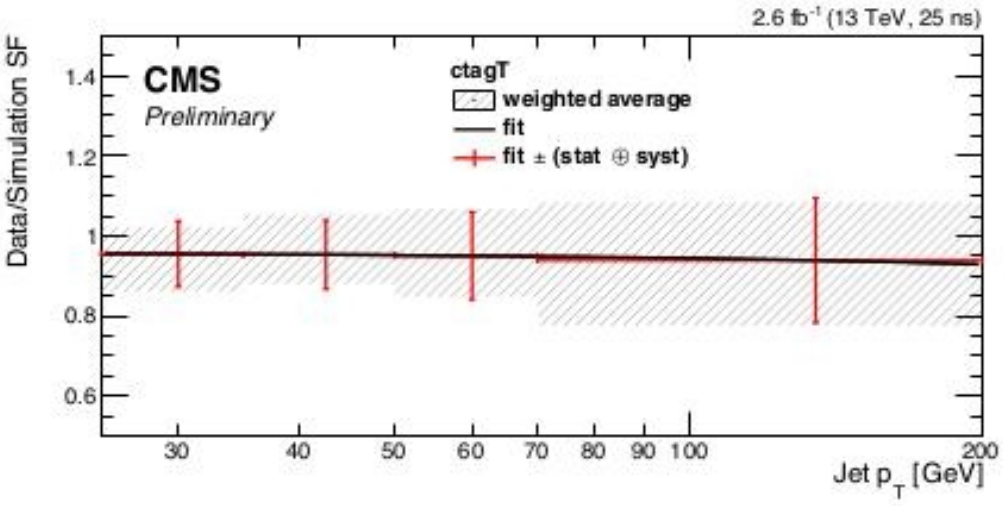
Simultaneous template fit in four categories to extract the SF value





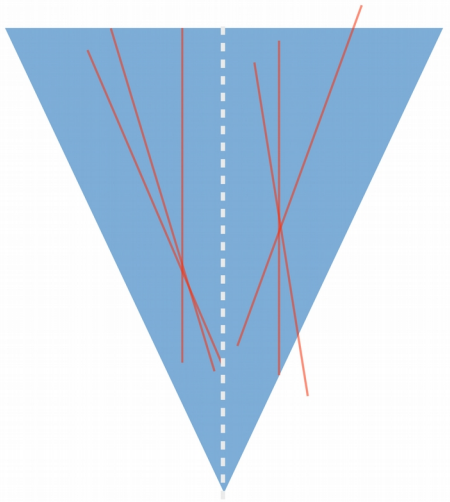
No c-tagged jet	Leading c-tagged jet
Subleading c-tagged jet	Both c-tagged jet



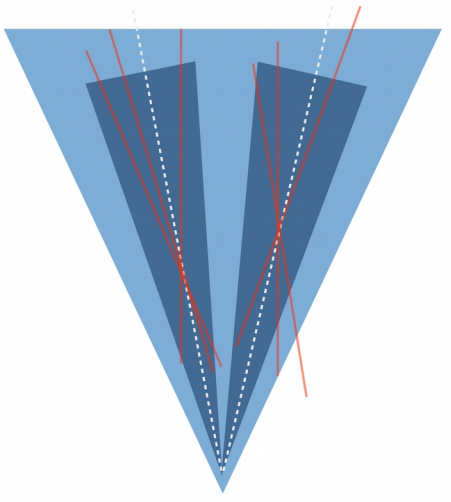


CMS-PAS-BTV-16-001

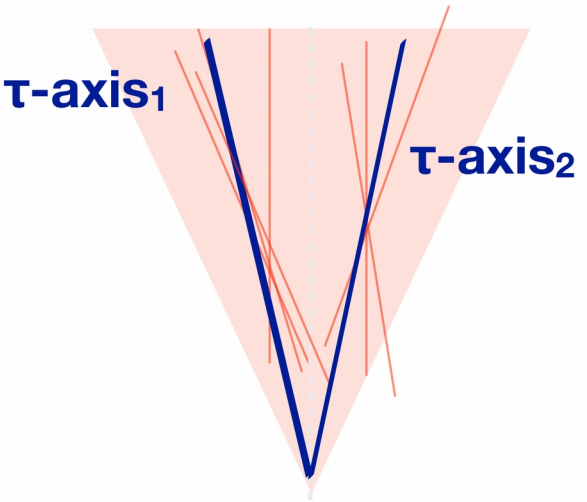
Boosted b-tagging



fatjet



subjets



double-b

FatJet: CSV w/o retraining. Custom (relaxed) track and SV association directly on anti-k T 0.8

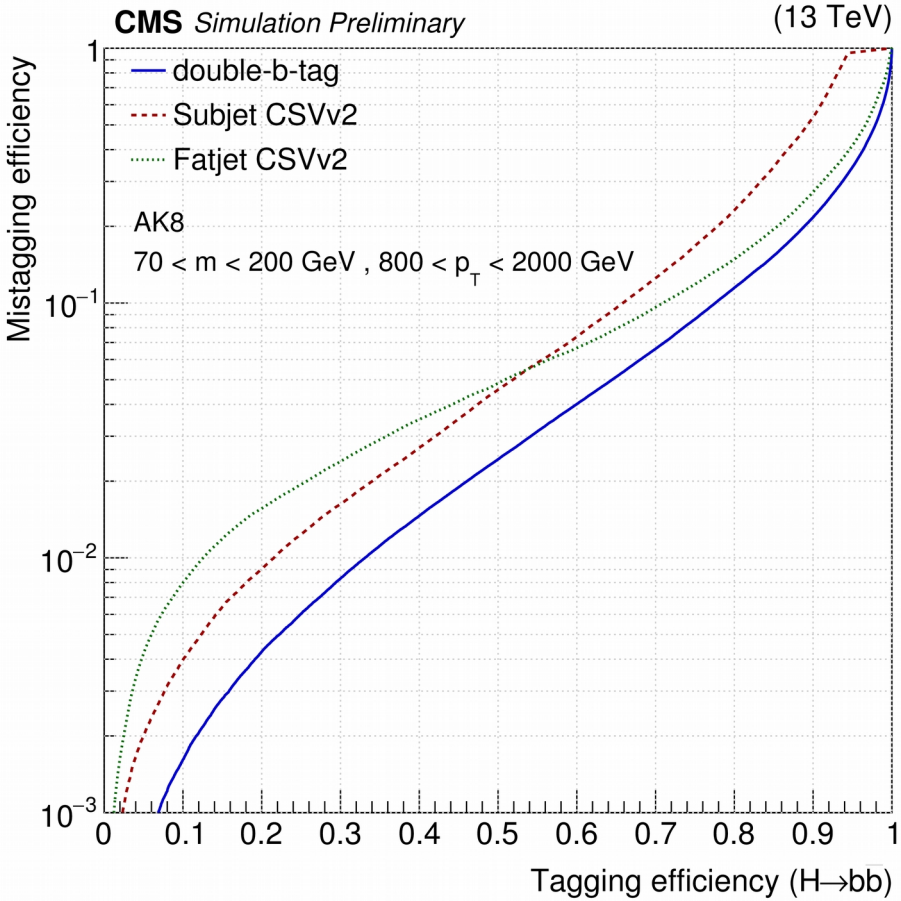
Sub-jet: CSV w/o retraining applied to sub-jets (soft drop, pruned, etc...). Used for boosted top

Double b: dedicated training targeting boosted resonances X→bb

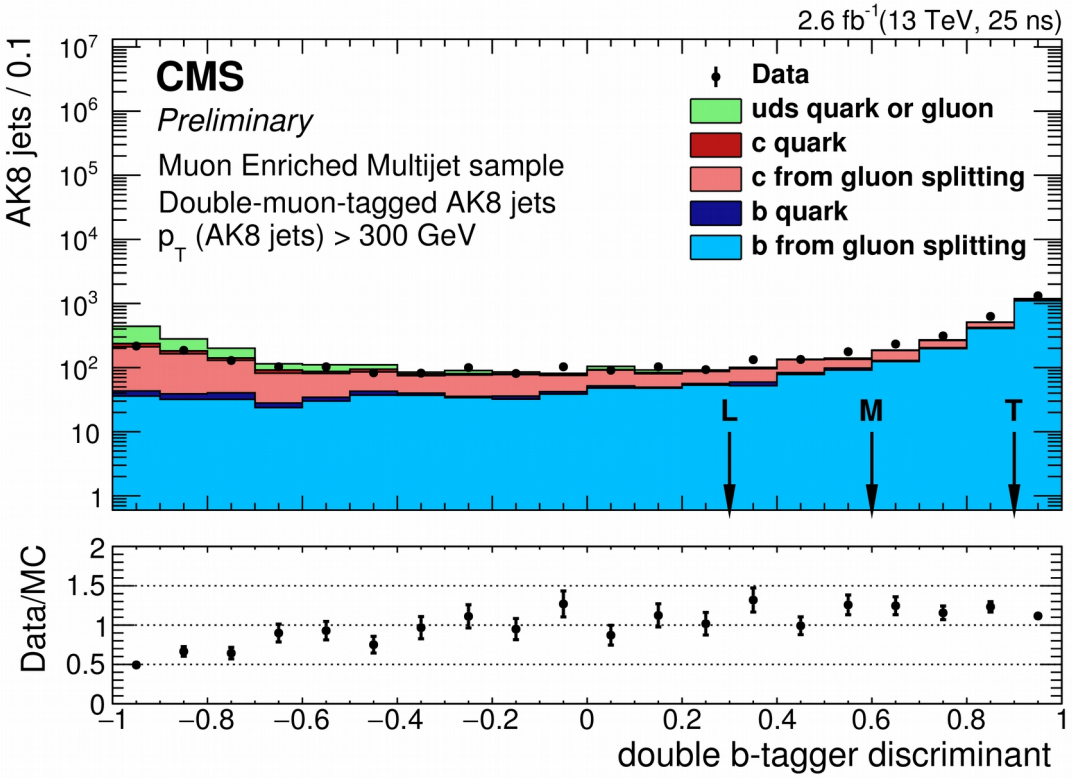
CMS-PAS-BTV-15-002

Double b-tagging

Performance of the double-b tagger, the minimum CSV among the two subjets b tag scores, and fat jet b tag which exploits CSV algorithm.



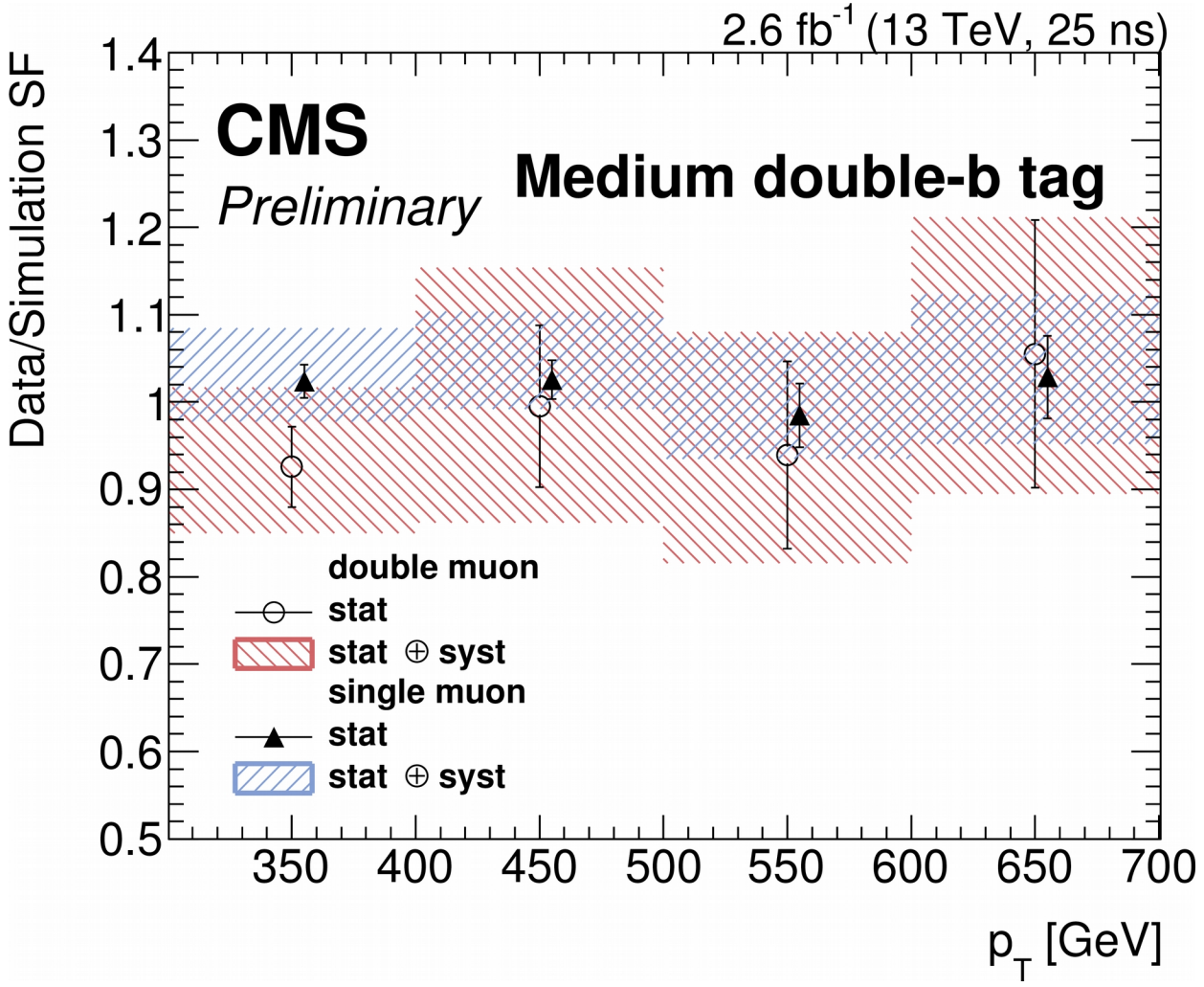
Double-b tagger discriminant distribution in data and simulated samples for the double-muon tagged jets selection



CMS-PAS-BTV-15-002

Double b-tagging : Performance

Data/MC efficiency ratio (SF) for medium double-b tagger requirement obtained with single and double-muon tagged selections.

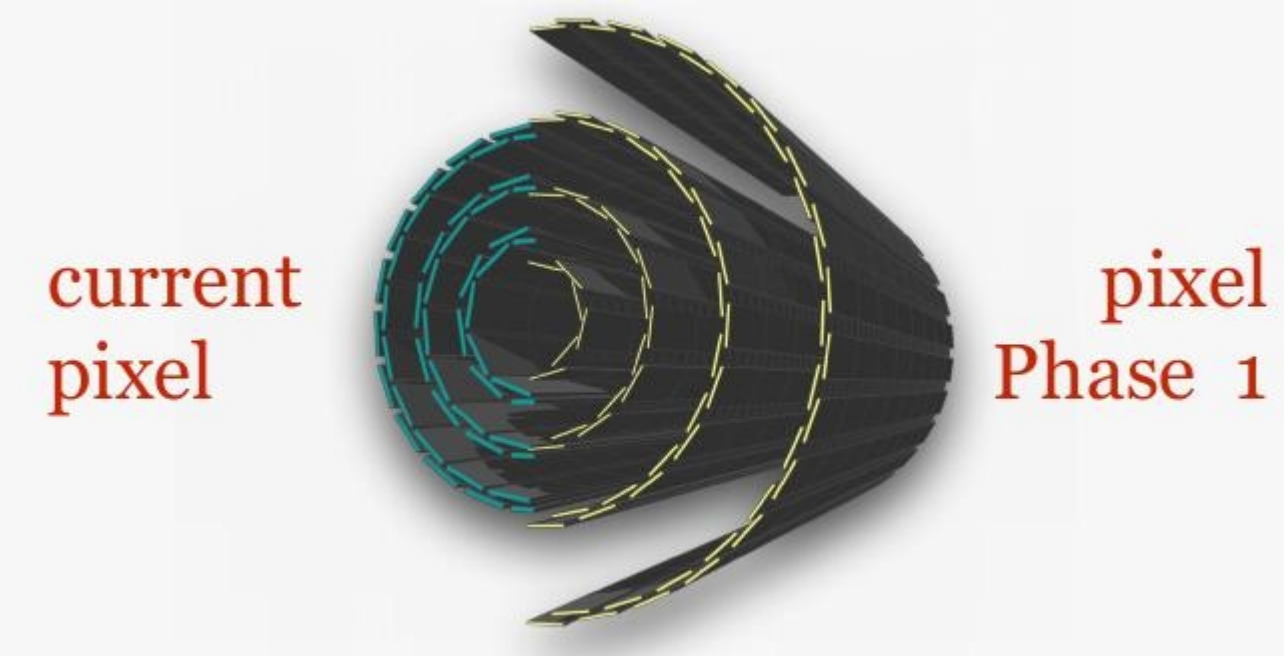


Efficiency using high pT jets enriched in bb from gluon splitting and requiring jet to be
→ matched to at least two muons (double muon)
→ matched to at least one muon (single muon)

Misidentification rate is based on the single lepton tt final state, where one top quark decays leptonically and the other hadronically

CMS-PAS-BTV-15-002

Upgrade studies

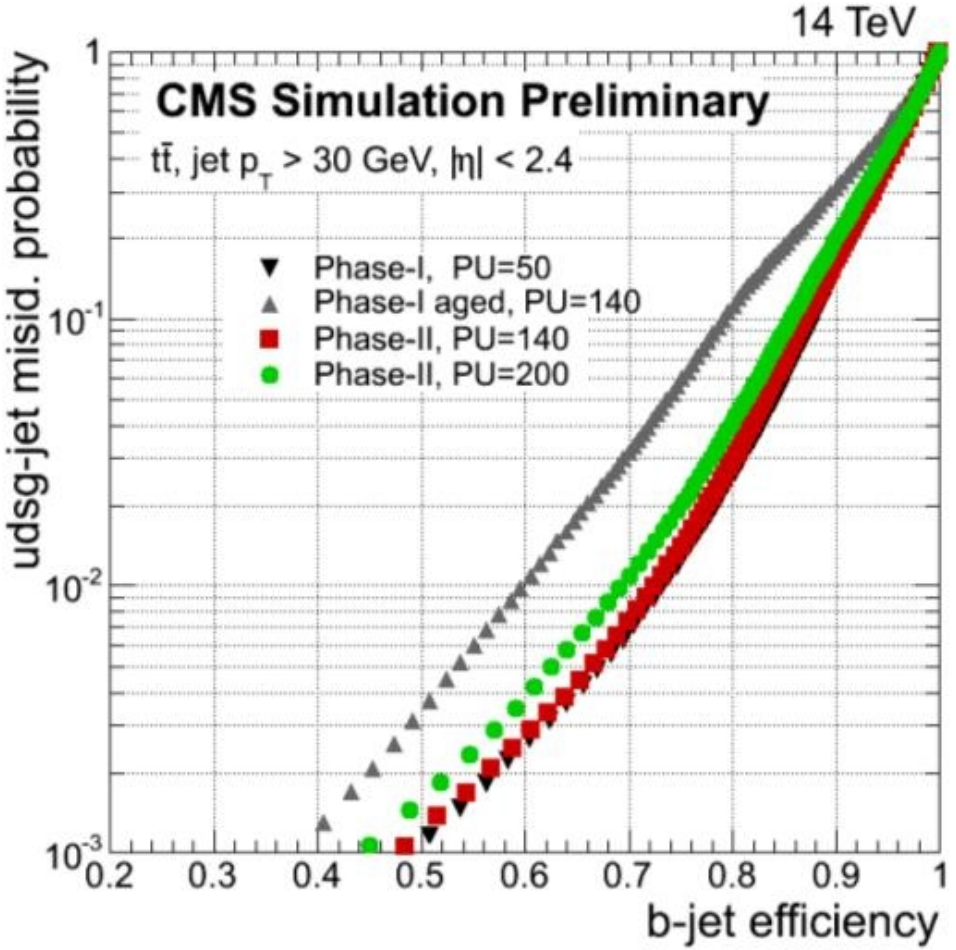


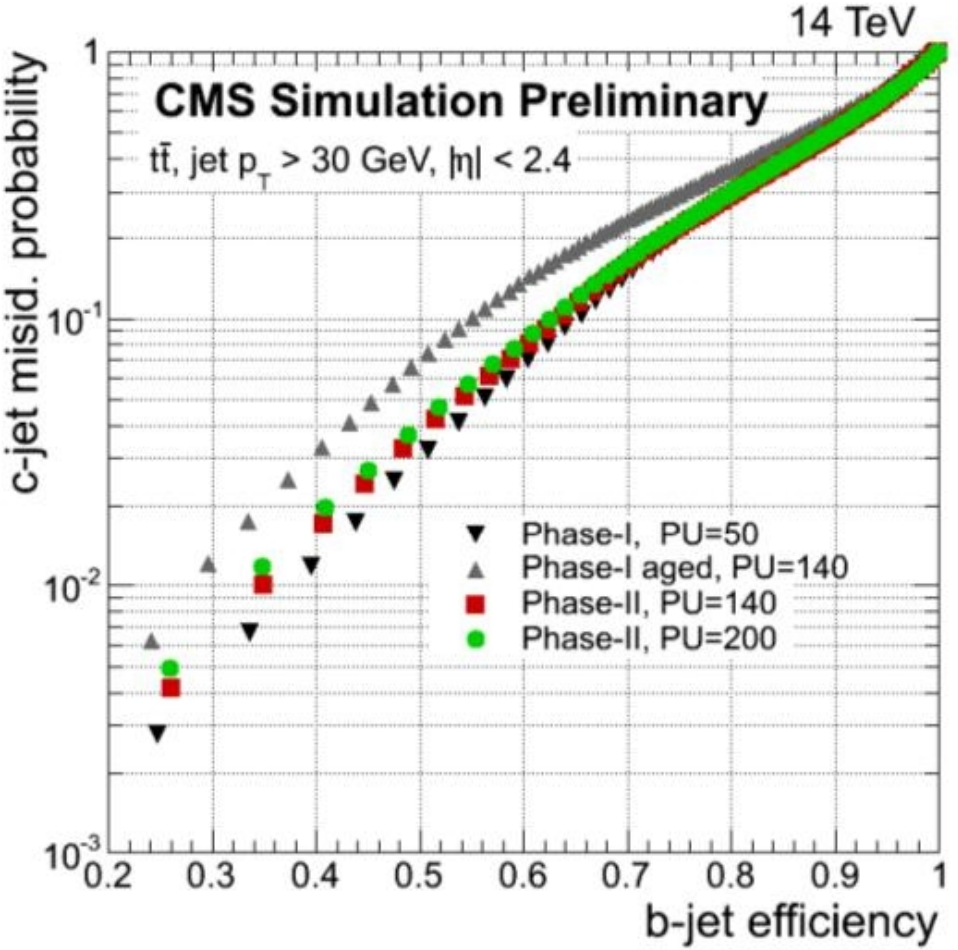
- **Phase 1 Pixel Upgrade:**
  - **additional layer, though less material**
  - **Innermost layer closer to beamline;**
- **Improvement with new detector out of the box, further improvements expected with dedicated optimizations.**

Upgrade studies

The b-tagging performance for the Phase-I aged detector scenario at high pileup is significantly degraded compared to Phase-I without aging at medium pileup.

The performance in Phase-II largely compensates that performance loss, despite the higher number of pileup collisions.





b-tagging & Vertexing: Datasets & Triggers

- **Standard AK4 PF jets:**
  - **Inclusive QCD:**
    - **JetHT PD: Inclusive PF jet triggers (co-owned by SMP)**
    - **Commissioning of b-tagging observables & for Mistag rate measurement**
  - **Muon-enriched QCD: Jets containing soft muons**
    - **BTagMu PD: DiJetMu Calibration triggers by BTV**
    - **Commissioning of b-tagging observables & b-tagging efficiency measurement**
- **Boosted AK8 PF jets:**
  - **Muon-enriched QCD: Jets containing soft muons**
    - **BTagMu PD: Dedicated AK8 SingleJet with soft muon path is added from 2016D era**
    - **AK8 topologies also based on AK4 triggers: subjet b tagging and double b tagging**
    - **Planning to add more dedicated paths for subjet b-tagging in 2017 (BTagMu_AK8Jet*_Mu5, BTagMu_AK8DiJet*_Mu5)**
    - **For Double b-tagger planning to add double muon inside AK8 jet trigger in 2017 (BTagMu_AK8Jet*_DiMu5)**
    - **Commissioning of b-tagging observables & b-tagging efficiency measurement**
- **We also use ttbar signatures for commissioning and performance measurements**

Summary

- **Heavy flavour tagging in CMS is presented.**
- **Performance of these algorithms using data is also looked at.**
- **Many improvements updated/adapted to RunII conditions are on going. (eg. selection of inputs for object reconstruction)**
- **Online integration: b-tagging at trigger level mirrors quite closely what is used offline**
- **Exploring the use of Deep NN**
- **Improvements in algorithms for Phase-1 upgrade studies are on going.**

Thanks

BACKUP

b-tagging & Vertexing: Datasets & Triggers

- **Standard AK4 PF jets:**
  - **Inclusive QCD:**
    - **JetHT PD: Inclusive PF jet triggers (co-owned by SMP)**
    - **Commissioning of b-tagging observables & for Mistag rate measurement**
  - **Muon-enriched QCD: Jets containing soft muons**
    - **BTagMu PD: DiJetMu Calibration triggers by BTV**
    - **Commissioning of b-tagging observables & b-tagging efficiency measurement**
- **Boosted AK8 PF jets:**
  - **Muon-enriched QCD: Jets containing soft muons**
    - **BTagMu PD: Dedicated AK8 SingleJet with soft muon path is added from 2016D era**
    - **AK8 topologies also based on AK4 triggers: subjet b tagging and double b tagging**
    - **Planning to add more dedicated paths for subjet b-tagging in 2017 (BTagMu_AK8Jet*_Mu5, BTagMu_AK8DiJet*_Mu5)**
    - **For Double b-tagger planning to add double muon inside AK8 jet trigger in 2017 (BTagMu_AK8Jet*_DiMu5)**
    - **Commissioning of b-tagging observables & b-tagging efficiency measurement**
- **We also use ttbar signatures for commissioning and performance measurements**

Performance on Data

two main methods:

events with muons in jets

- use characteristic properties of the muon from B decays to measure (fit) the b-content
- relative momentum to the jet axis (p_t^{rel}) is large because of the B hadron mass
- possible to use IP of muon as well (although correlated with b-taggers)

events with top quarks

- select sample with a high fraction of top quark events
- the branching fractions of top quarks are well known
  → the flavour composition of the sample is known as well