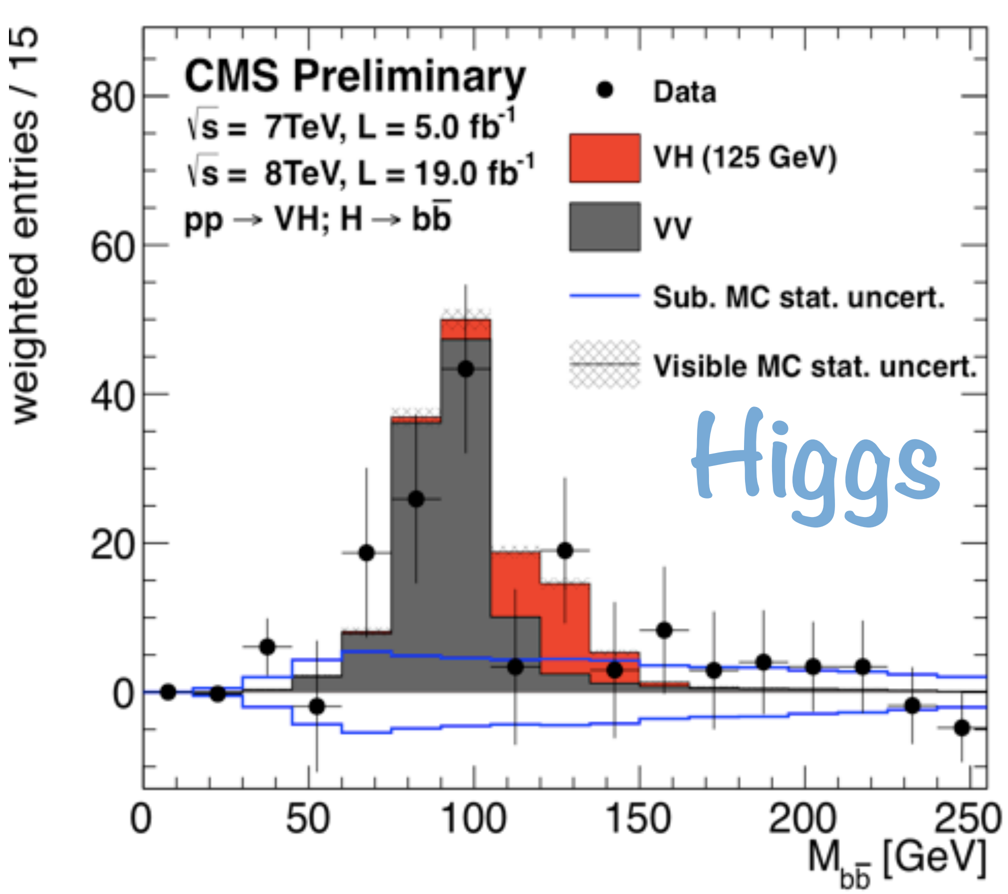
CMS Results on b-tagging in boosted topologies

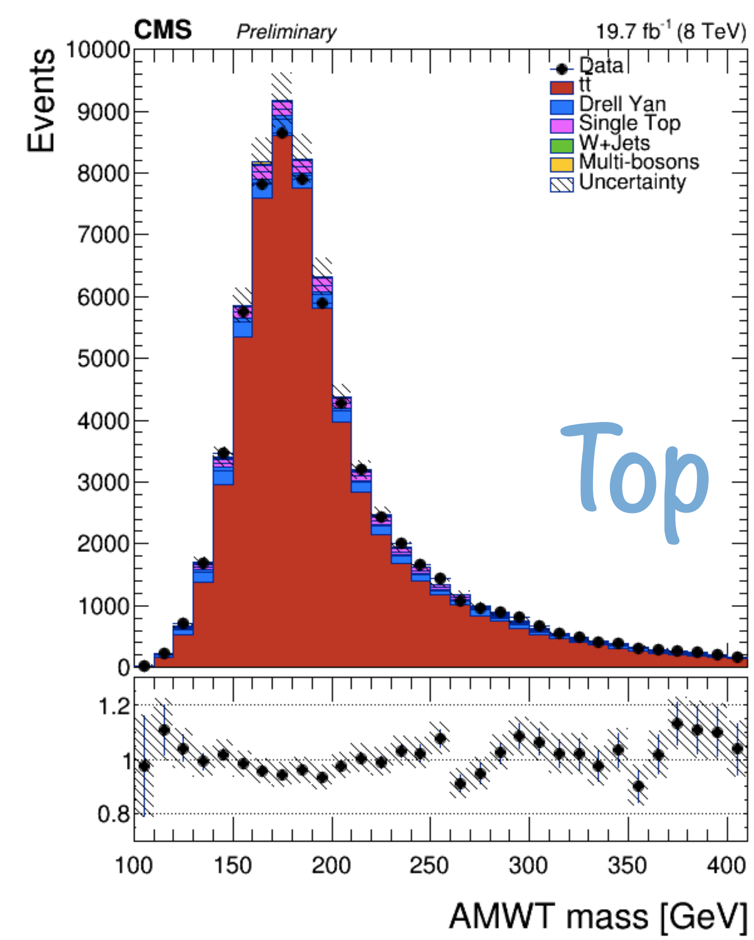
Caterina Vernieri
on behalf of the CMS Collaboration
FNAL

BOOST2015, The University of Chicago
10-14 August

b-tagging

- Jets from **b-quark** hadronization are present in many physics processes
  - decay of **Top** quarks ($t \to Wb$)
  - the **Higgs** boson ($h \to b\bar{b}$)
  - important also for new final state predicted by **BSM** models
- b-jet efficiency and purity is very important
  - crucial to reduce the otherwise overwhelming background





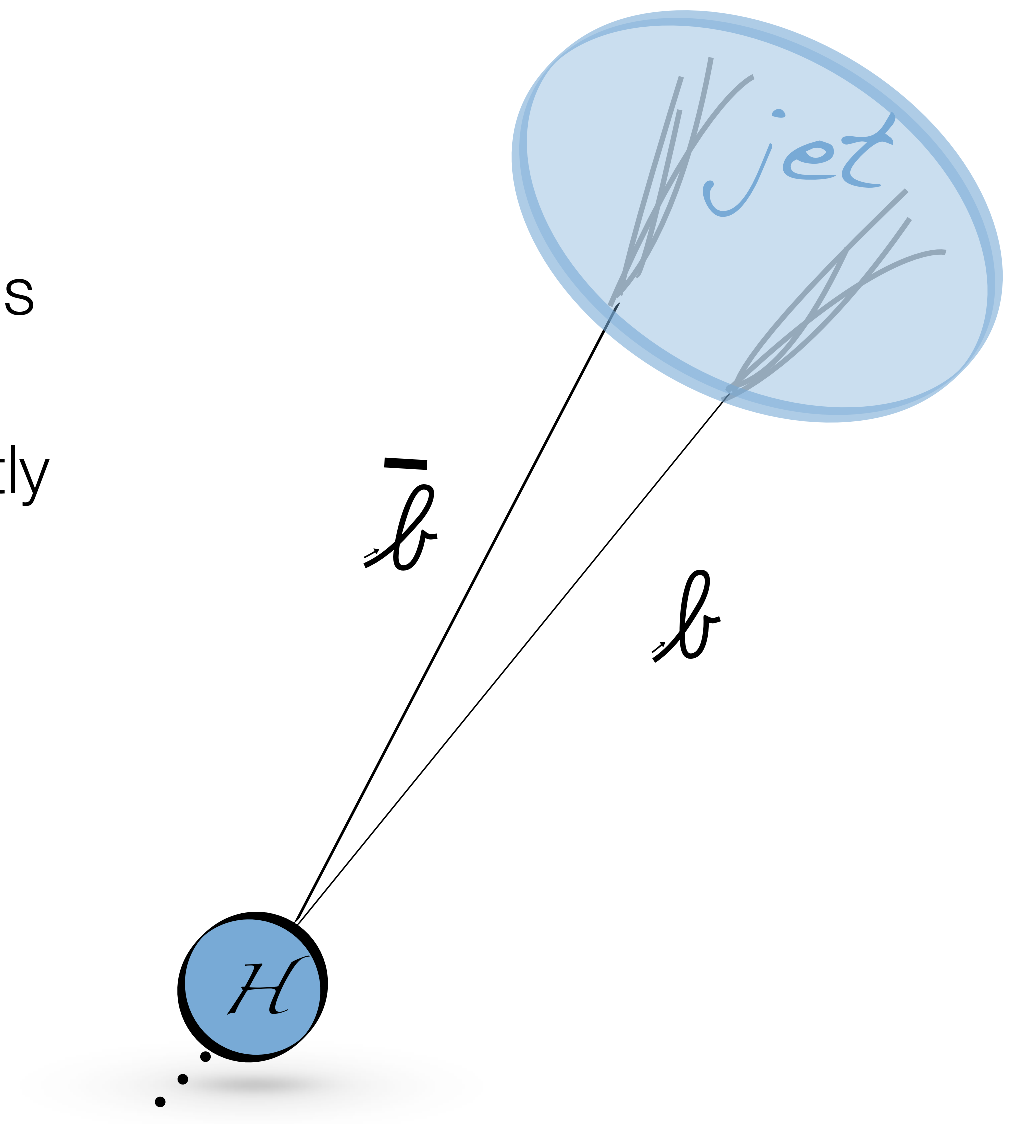
double-b-tagging

Boosted h($b\bar{b}$) is expected to be:

- a **single "fat" jet**

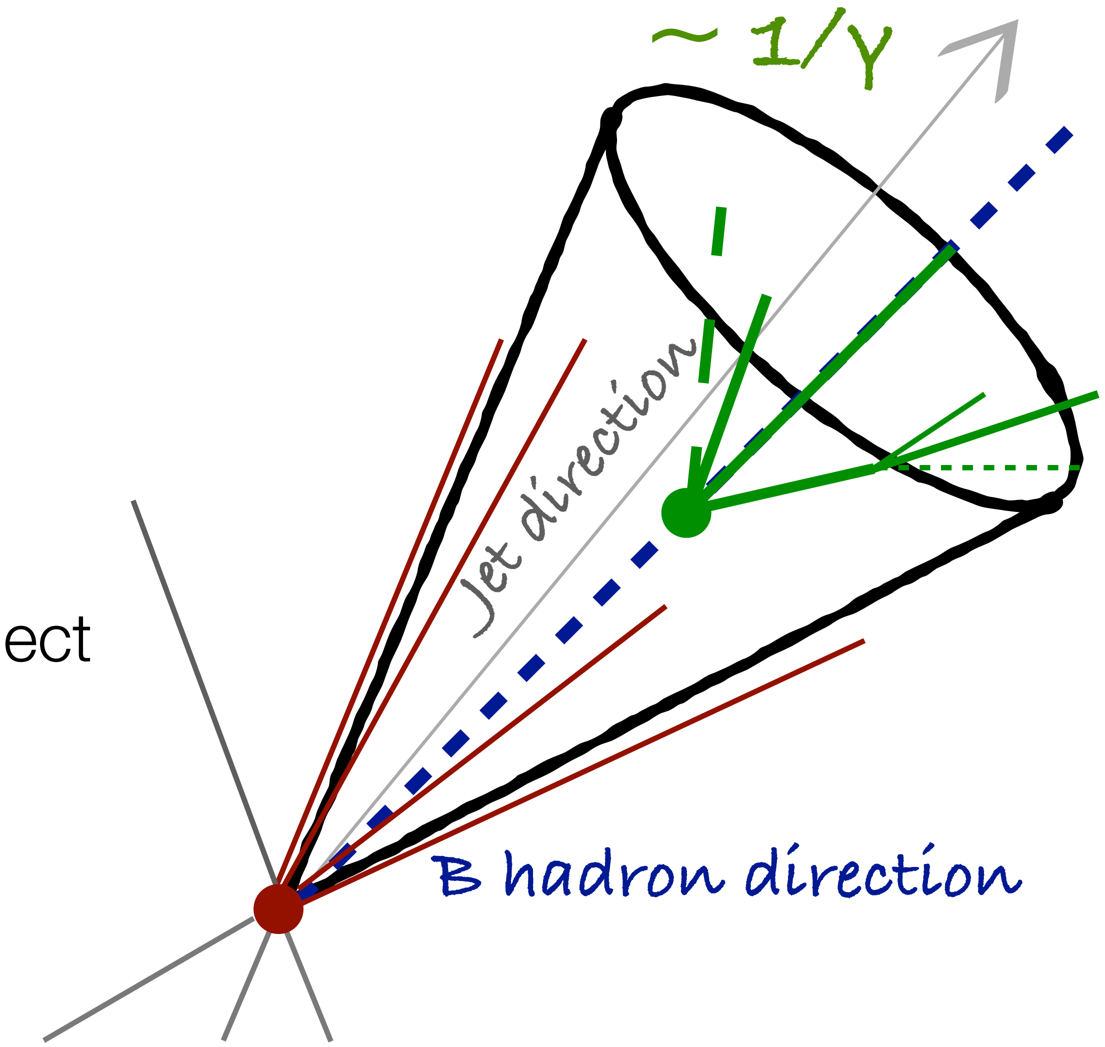
Signal can be identified by exploiting:

- **b-tagging** to reconstruct the two B hadrons from the b and $\bar{b}$ within the same fat jet
  - We identify secondary vertex independently of jet direction and reconstructs the B-hadron decay chains
  - displaced tracks linked to charged constituents of particle flow jets
- the composite nature of the jet using **substructure**
  - **uncorrelated** to b-tagging, can be exploited separately



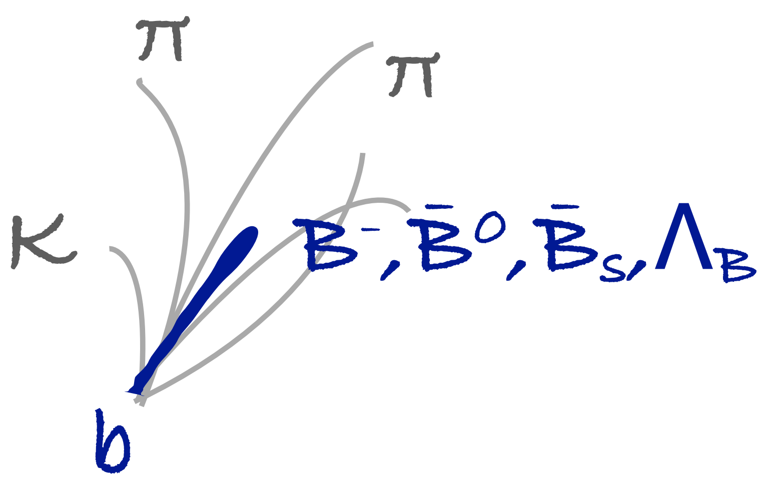
“b-tagging” task

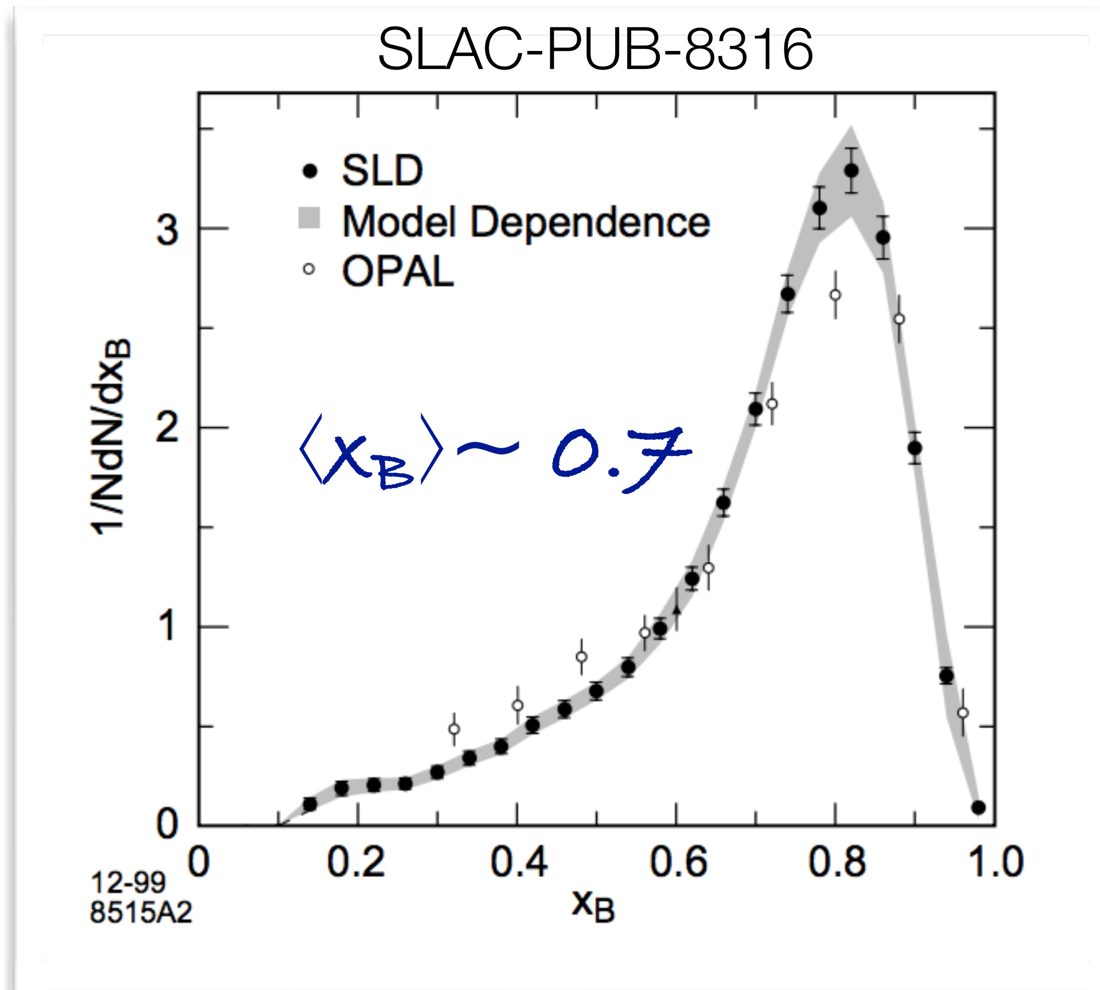
- b hadronizes in “jet” of particles
  ➞ need to associate tracks to jet
- B-hadron decay tracks are in a forward cone
  ➞ B flight ~ jet direction
  - looks for tracks in the jet cone
- A dedicated selection to maximize purity and reject fakes
- Inclusive Vertex Finder (IVF) to reconstruct secondary vertices
  - Independent of the jet direction



b-tagging, Properties

b quarks hadronize





- Measurable lifetime

  $c\tau \sim 500\ \mu m \rightarrow \beta\gamma c\tau \sim 5mm$ @ 50 GeV

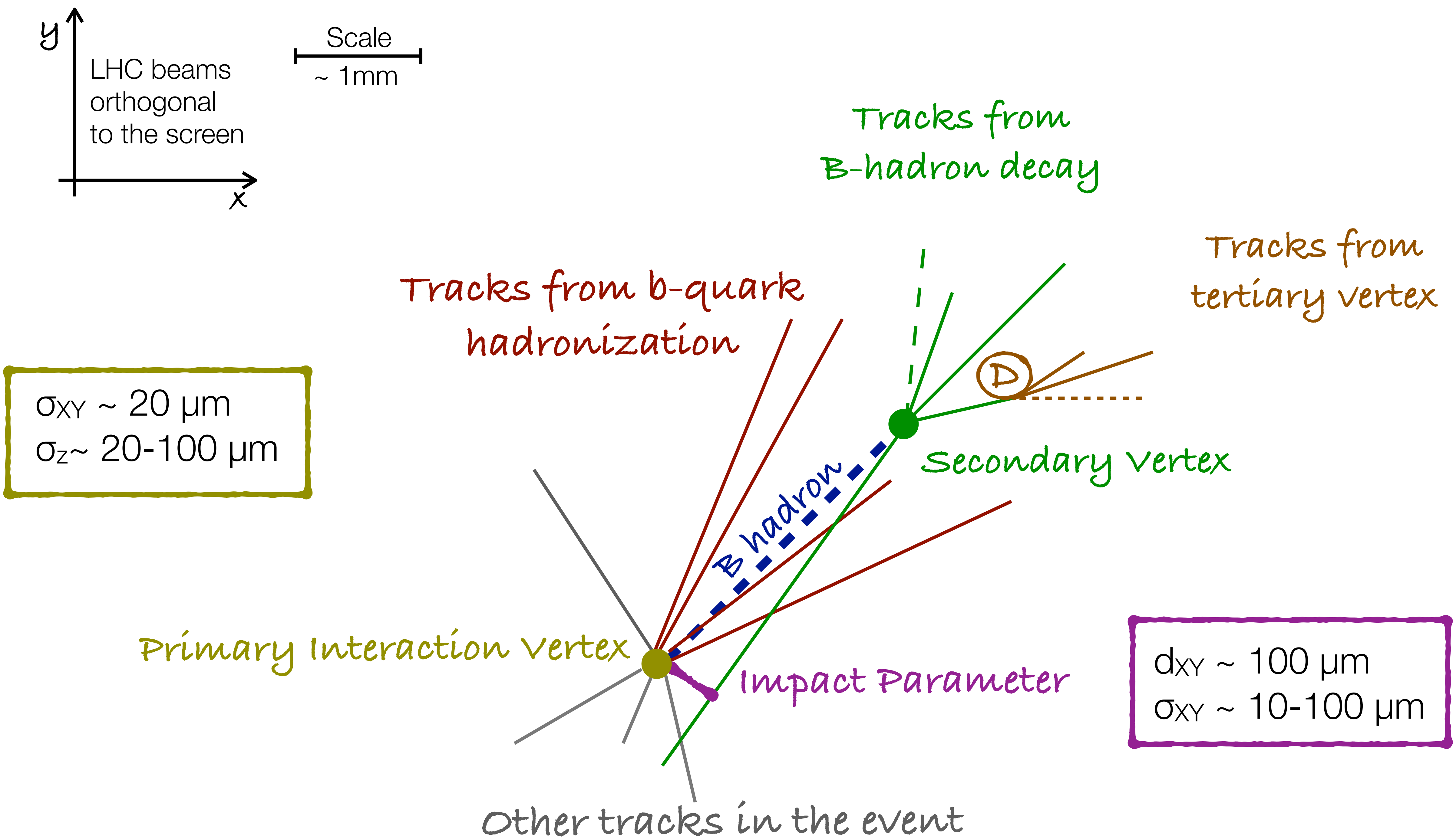
- High momentum transferred to the B-hadron
- Large mass (~5 GeV)
- The weak b-decay often produces leptons

  BR: $B \rightarrow l + \nu + X$ ~ 25%

  $B \rightarrow D \rightarrow l + \nu + X'$ ~ 20% tertiary vertex

  where l = e or μ

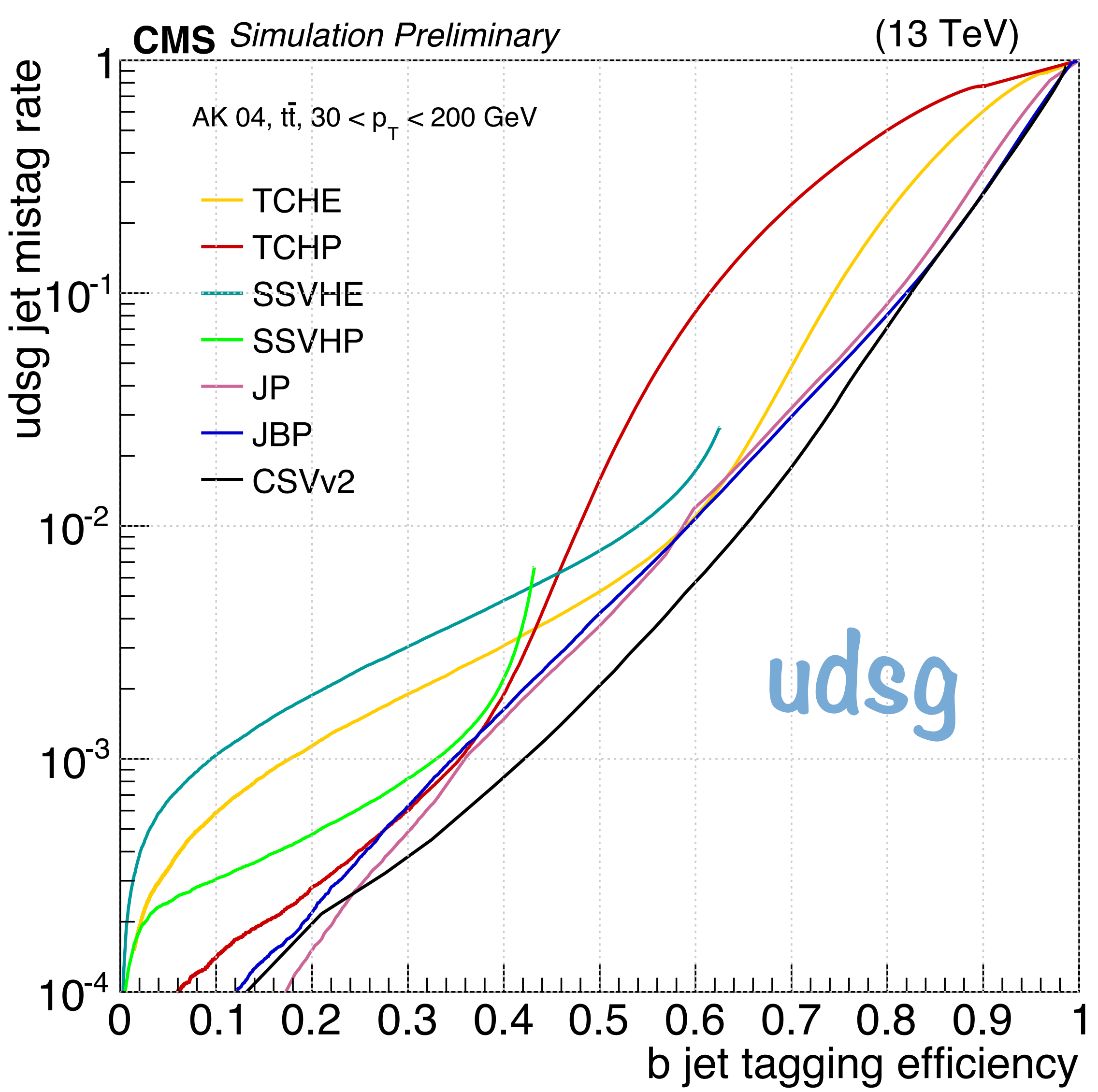
b-tagging, **Observables**

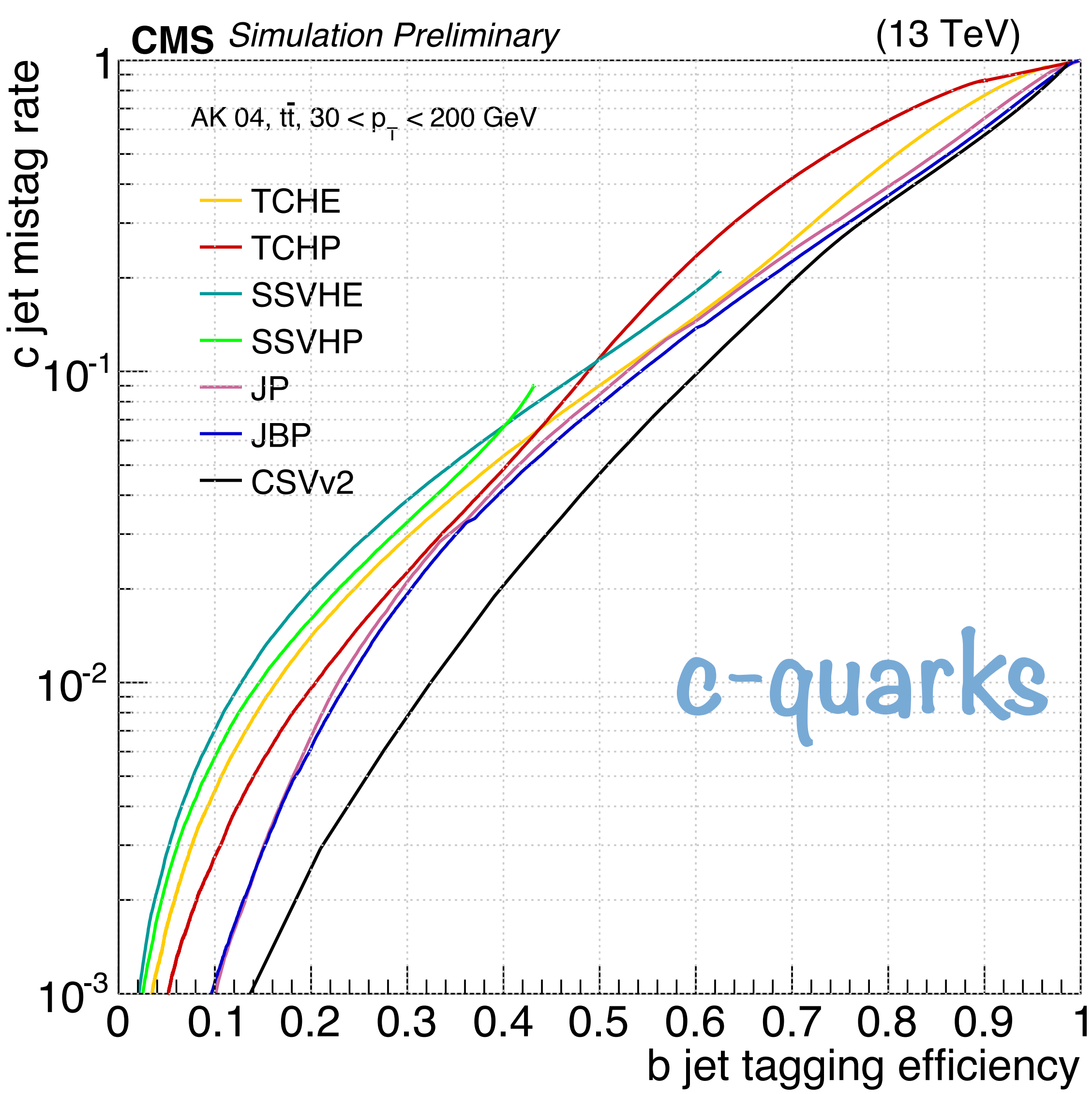


b-tagging, algorithms at CMS

b-tagging algorithms use the information from:

- **impact parameter** significance of charged-particle tracks
- the presence and properties of reconstructed **decay vertices**
  - flight distance, mass, energy ratio, # charged tracks at SV





CSVv2 Performance

- The Combined Secondary Vertex through multivariate technique combines

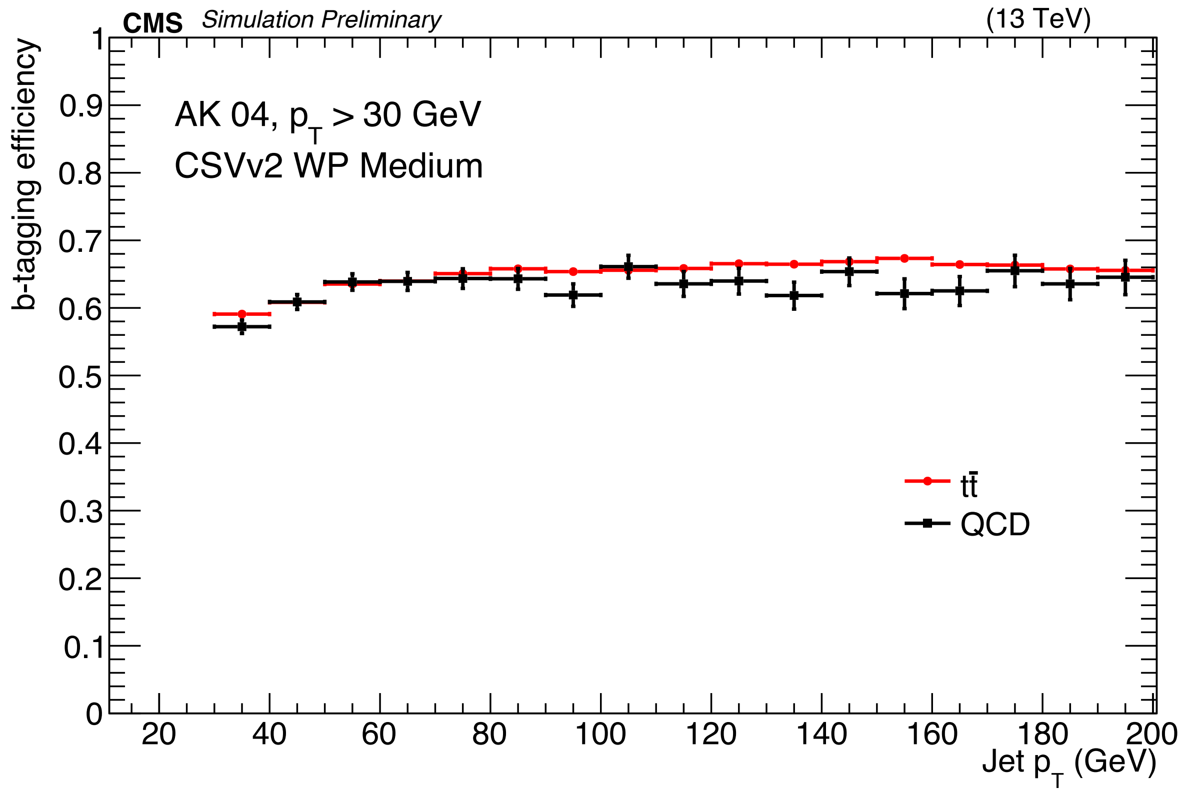
  Track information

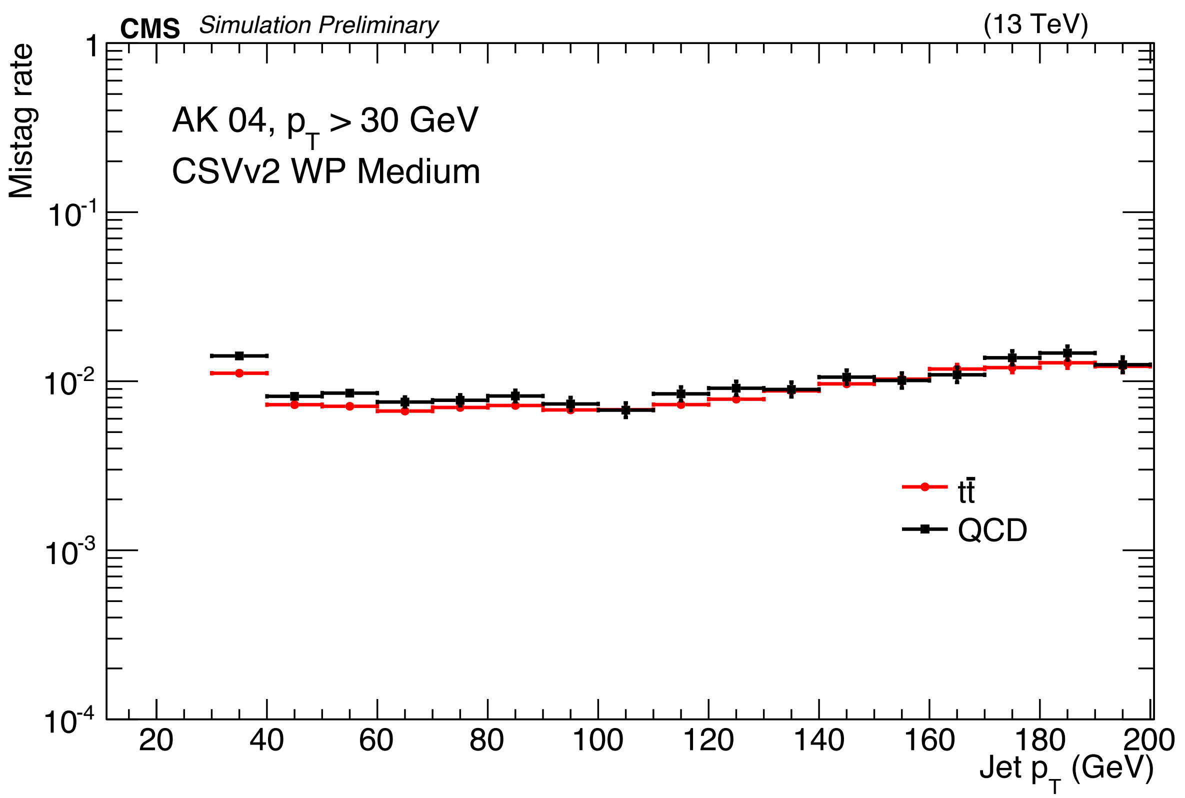
  - 3D IP significance of the most displaced tracks

  Vertex information

- For fixed working points (WP), performance are:

	b	c	light [%]
Loose	85	32	10
Medium	70	10	1
Tight	43	2	0.1





$h(b\bar{b})$ as tool for discovery

LHC Run I legacy

H

$J^P = 0^+$

m_H = 125.03 ± 0.26 (stat.) ± 0.14 (syst.) GeV

production and decay rates are consistent with a SM Higgs boson

CMS reported **2.6σ** evidence in the **$b\bar{b}$ final state** consistent with SM h(125 GeV)

- high b-tag efficiency plays a critical role
- helps to reduce high QCD background contribution

SM predicts $h(b\bar{b})$ to dominate total width

- Higgs, as a new powerful **tool to search for new physics**
- large number of events vs. $Z(b\bar{b})$: BR=15% vs 58%

h($b\bar{b}$) and BSM

New physics might preferentially couple to EWK sector

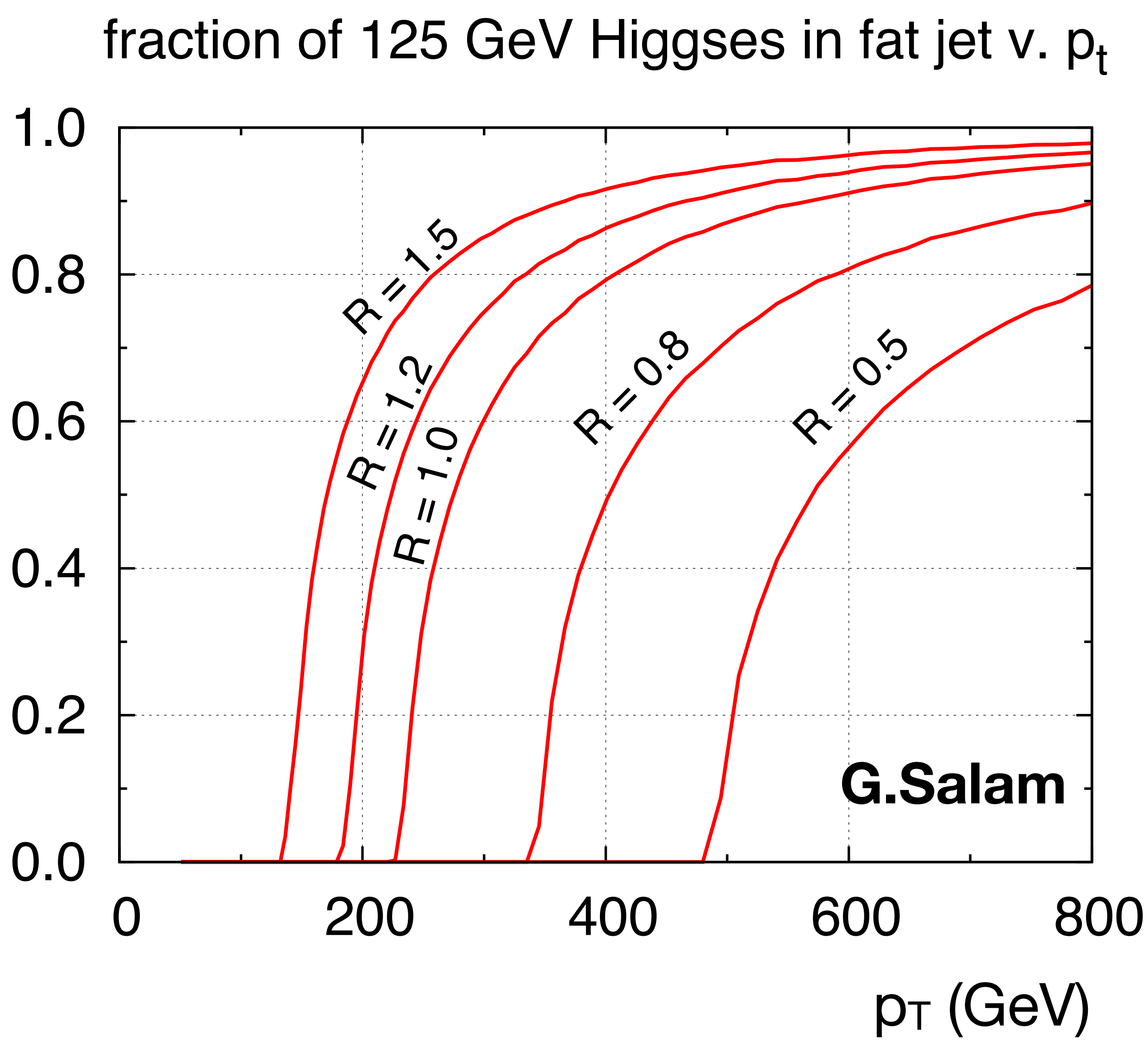
If BSM particles are heavy then **boosted h boson as final state**

- b-tagging helps to reduce QCD background
- many many searches can benefit from this.

Vector-Like Quarks	$b'\bar{b}' \to b\bar{b}$ + $h(b\bar{b})h(b\bar{b})$ + X $t'\bar{t}' \to t\bar{t}$ + $h(b\bar{b})h(b\bar{b})$ + X	CMS-B2G-14-001 CMS, JHEP 06 (2015) 080
Warped Extra Dimensions 2HDM,(N)MSSM	$X \to hh \to 4b$, $b\bar{b}\gamma\gamma$, $b\bar{b}\tau\tau$	CMS,arXiv:1503.04114 CMS-HIG-13-032/14-034
SUSY	$X \to hh/V$ + LSP	CMS Phys.Rev.D90 (2014)092007

boosted h($b\bar{b}$)

h($b\bar{b}$) from decay of heavy objects is expected to be produced with high p_T



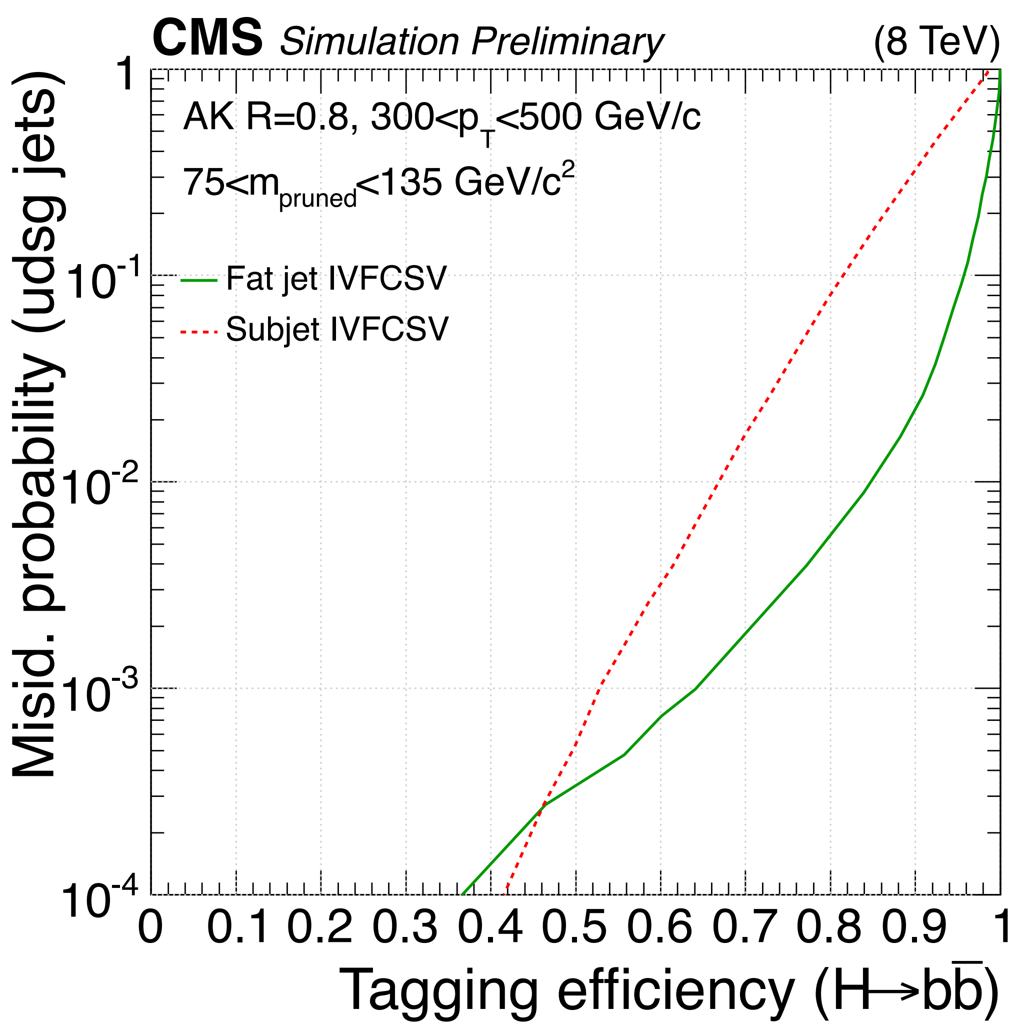
$dR(b\bar{b}) \sim 2m/p_T$

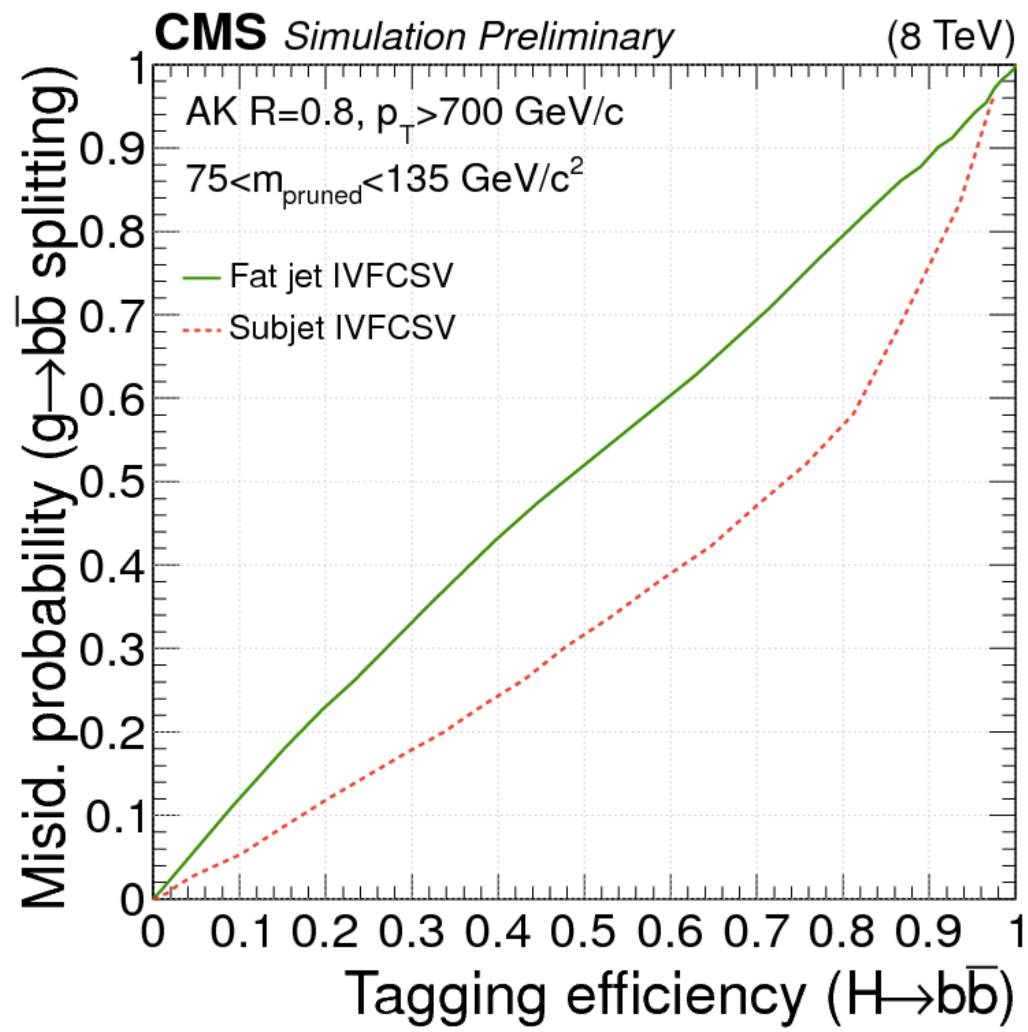
The boosted h($b\bar{b}$) signal is expected to be a **single "fat" jet**

Fully contained in a jet of radius :

- R =1.5 for H p_T ~ 200 GeV
- R = 0.8 for H p_T ~ 500 GeV

Run I, fat-jet vs. sub-jet approach





fat-jet b tagging based on the standard b-tagging algorithm which is not designed for two b quarks

sub-jet b tagging

fat-jet b-tagging works but substructure does better against $g(b\bar{b})$

At very high p_T jets from $h(b\bar{b})$ get too close

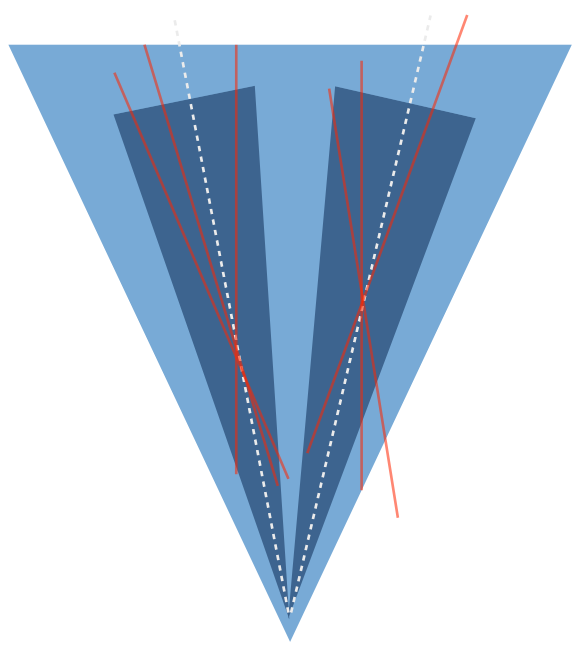
less discrimination with respect to $g(b\bar{b})$

remove dependency on sub-jets definition

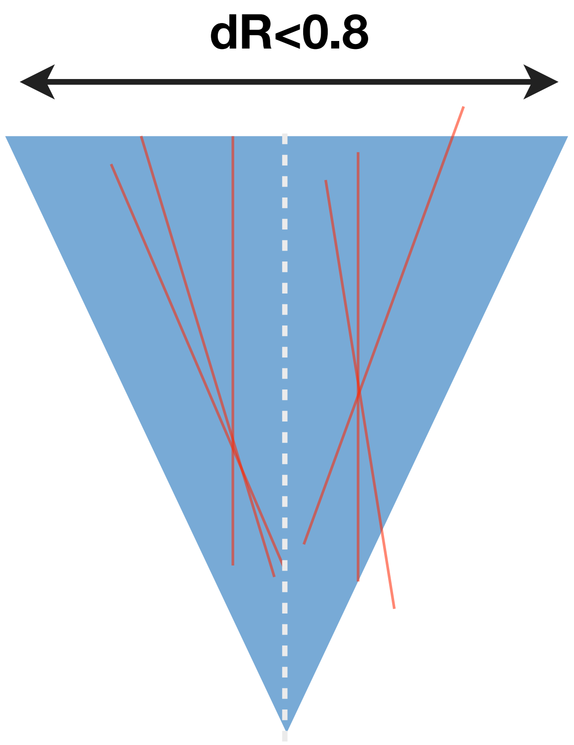
Run II, a new strategy

- Dedicated b-tagging strategy to identify boosted h($b\bar{b}$)
- We target $b\bar{b}$ signal from **any resonance**, not just Higgs.
- This approach is
  - stable against $\mathbf{p_T}$
  - **mass** independent
    - easier to validate
    - it can be applied to Z to $b\bar{b}$ as well as any BSM particles decaying to $b\bar{b}$ and having an electroweak-scale mass
- **flexible cone sizes** to better combine with top tagger and different kinematic regime
  - 0.8 for high and 1.5 for low boosted regime (same as top tagging)
- preliminary results, work is ongoing…

sub-jets vs. double-b tagging



sub-jets



double-b

Defines sub-jets
b-tagging observables for each sub-jet
explicit jet track association

substructure observables
Secondary Vertex within the fat-jet cone
observables from SV and tracks collections for the fat-jet

double-b discriminant

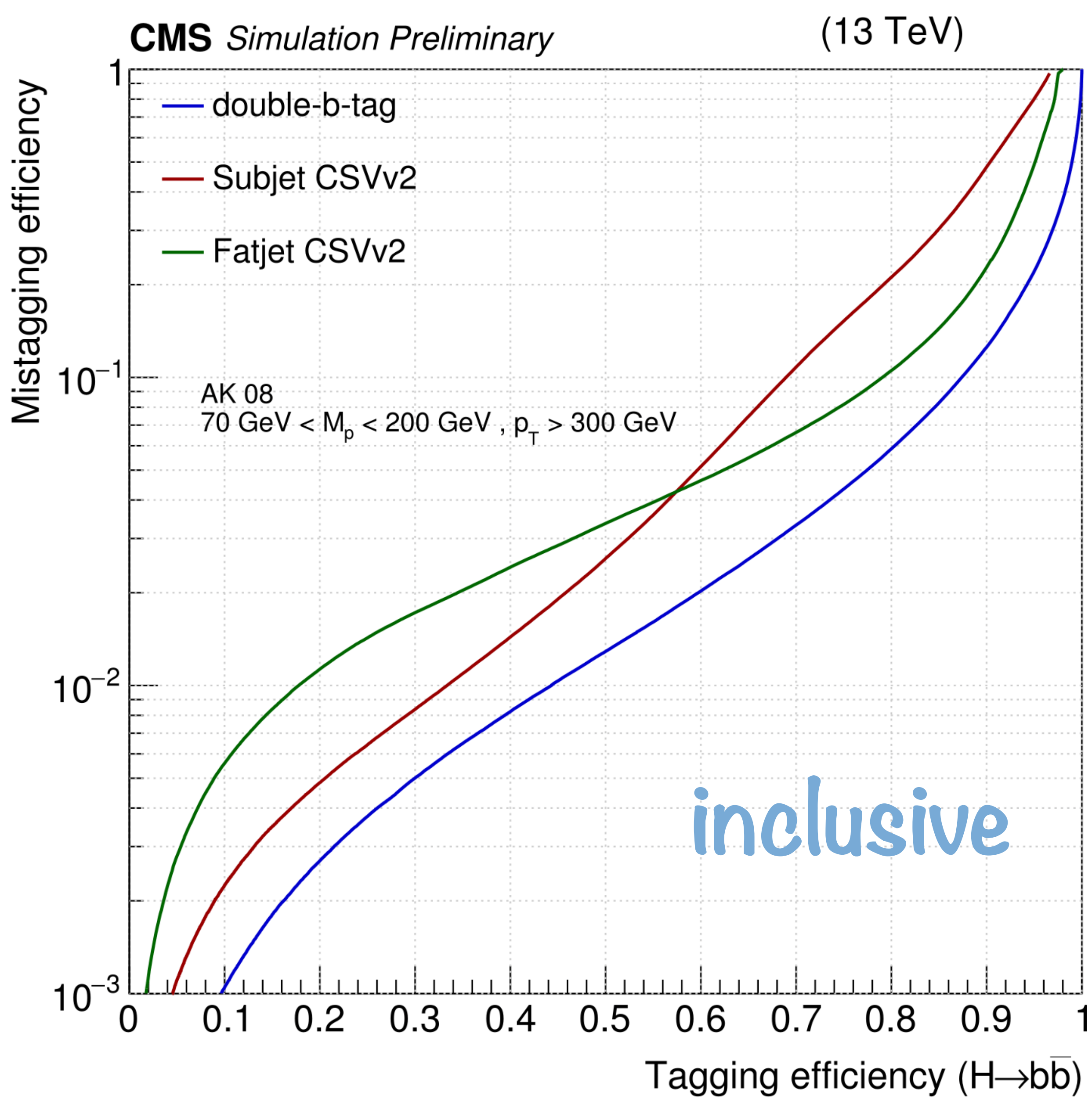
- A Boost Decision Tree in TMVA is used for the training
  - QCD as **background**
- Inputs exploit same information as CSVv2 related to

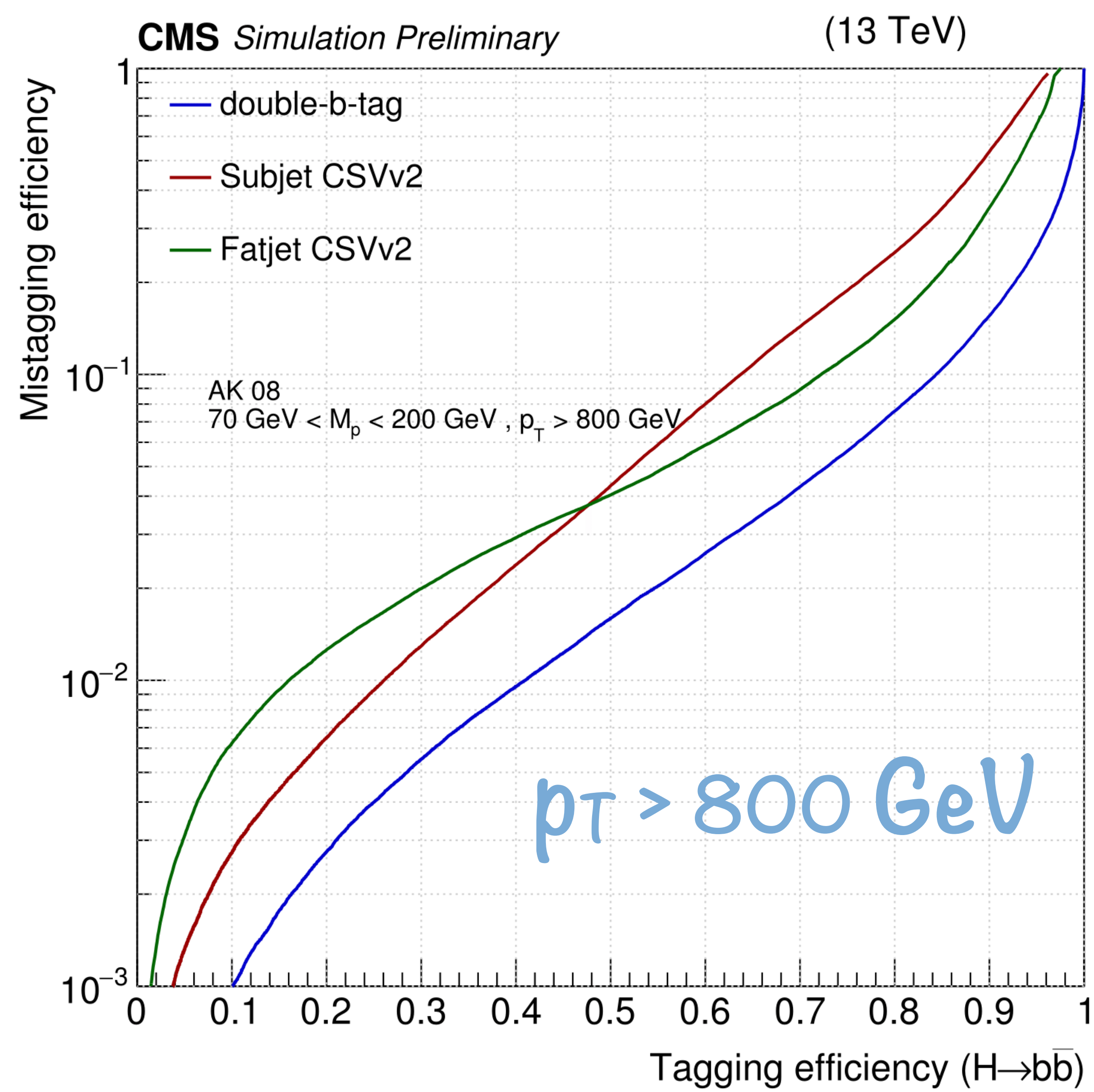
  Tracking

  Secondary Vertex (only one is considered)
- Also additional information from **Soft Lepton** is used

  the presence of a lepton in the jet and its p_T relative to the jet
- The minimum score of the two sub-jet CSVv2 b-tag value

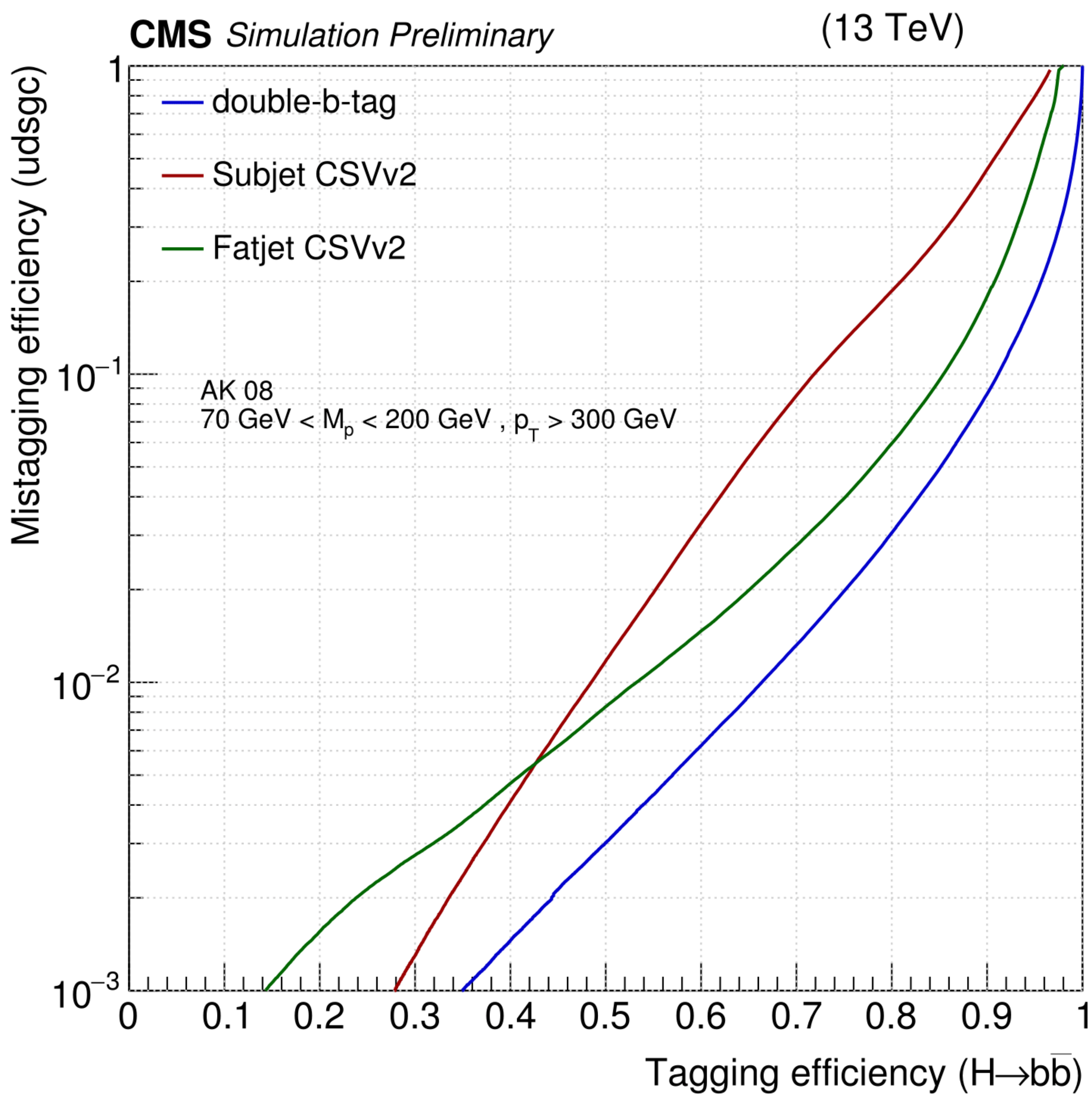
h(b$\bar{b}$) vs. Inclusive QCD





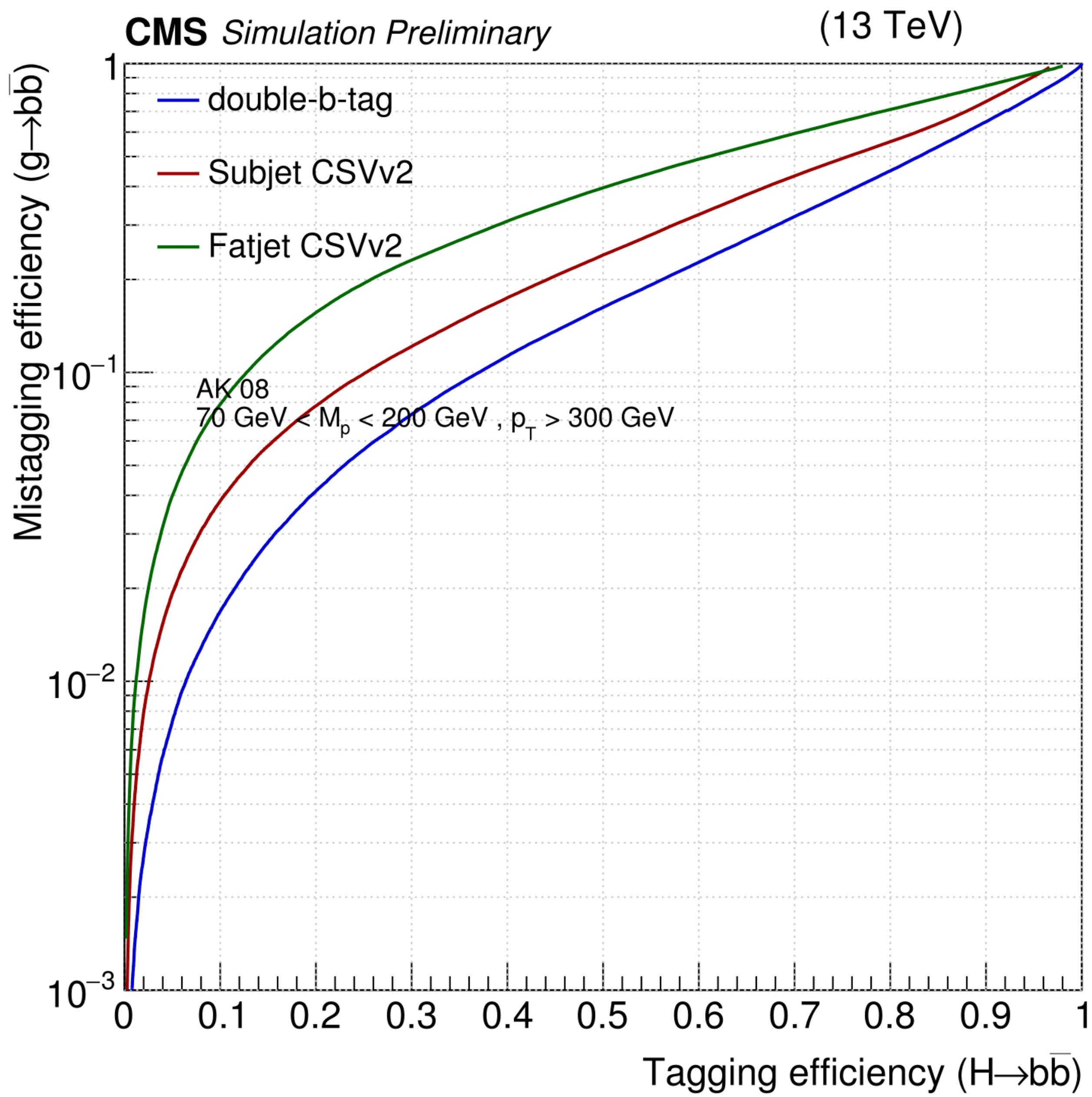
- double b-tag performs better than both sub-jet and fat-jet b-tagging in the entire p_T range considered

h($b\bar{b}$) jets vs.udsgc jets



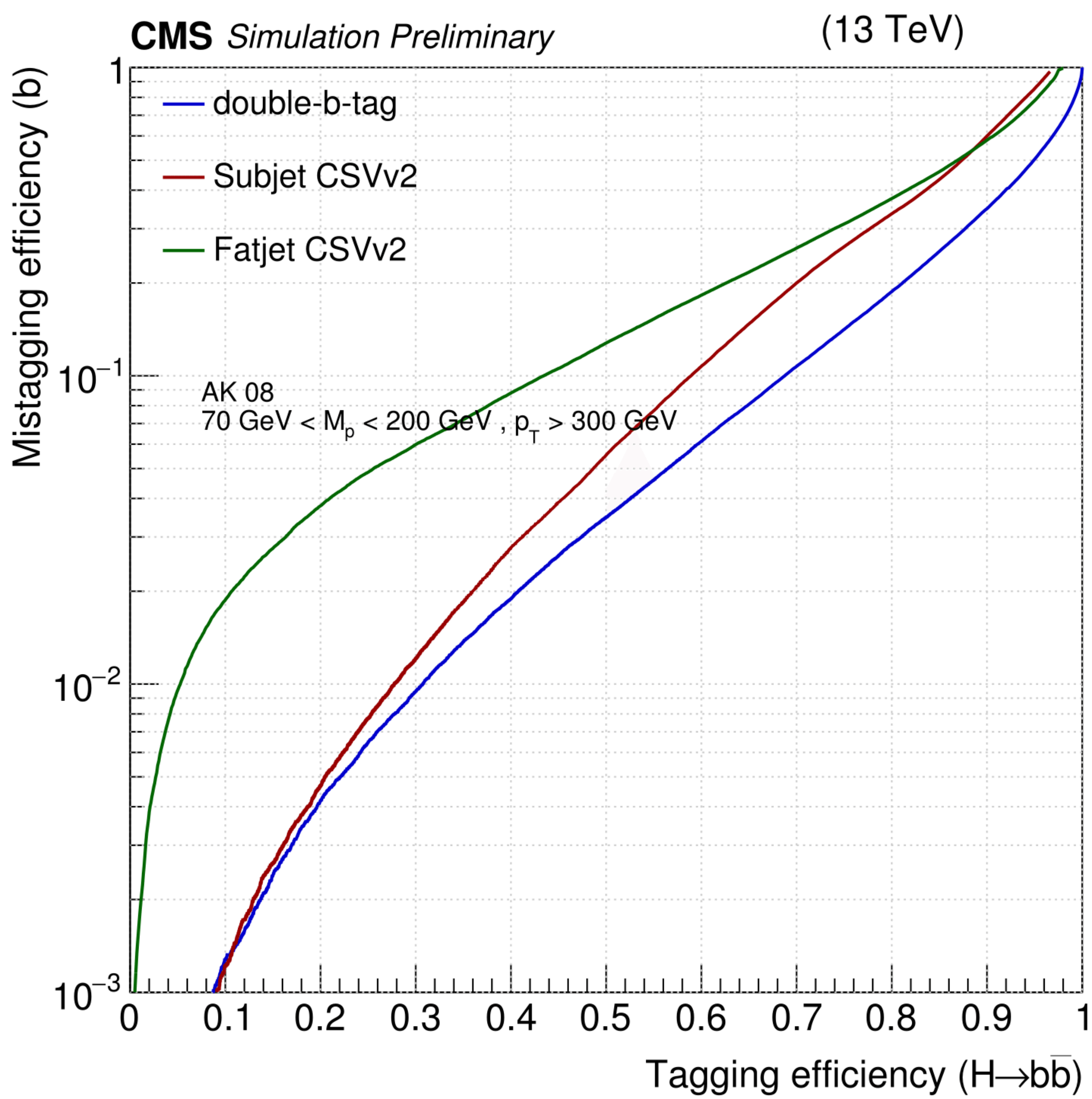
- double b-tag performs better than both sub-jet and fat-jet b-tagging in the entire p_T range considered

h(b$\bar{b}$) jets vs. g(b$\bar{b}$) jets



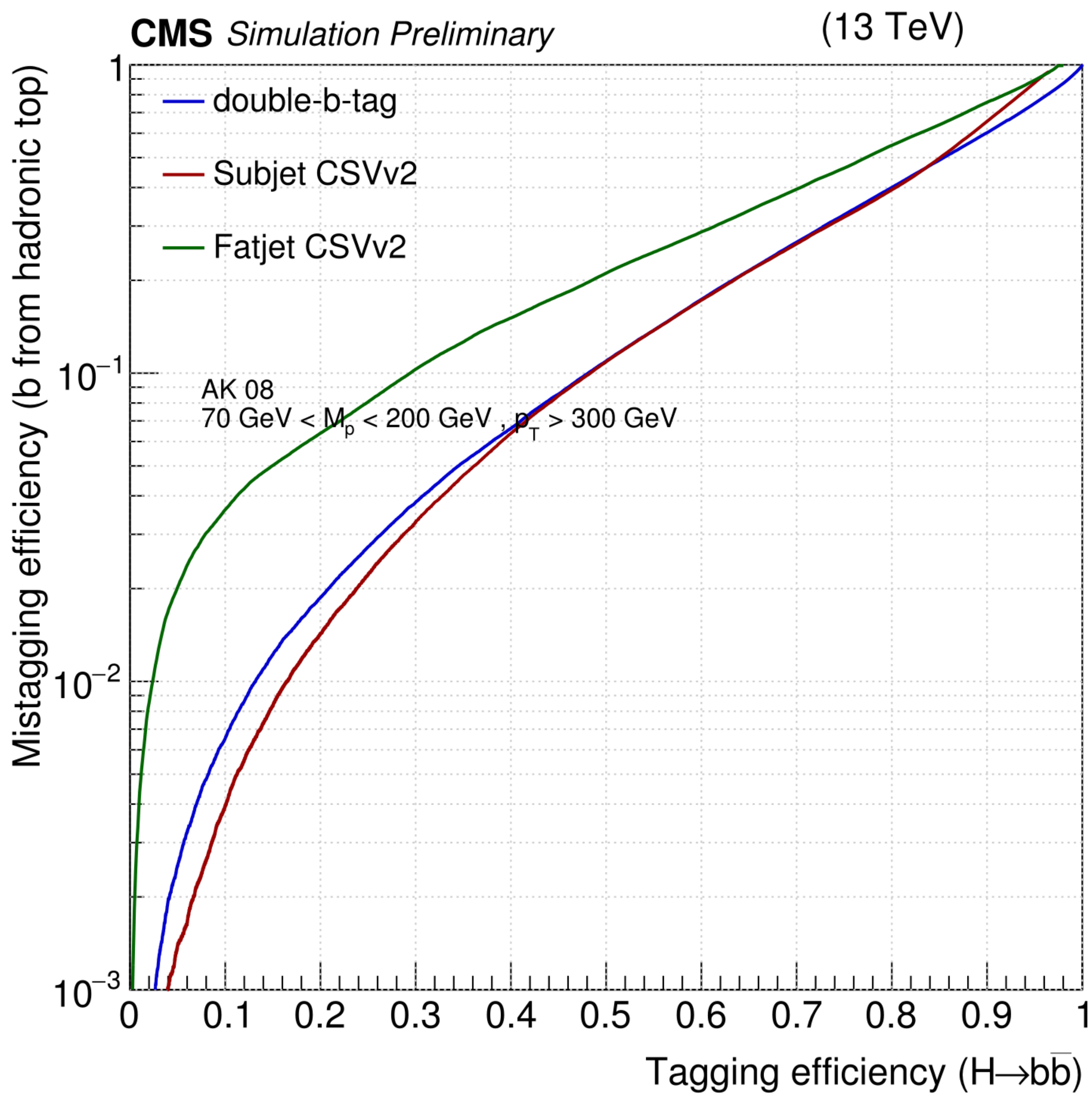
- double b-tag performs better than both sub-jet and fat-jet b-tagging

h($b\bar{b}$) jets vs. b-jets



- double b-tag performs better than both sub-jet and fat-jet b-tagging

h($b\bar{b}$) against single b from top background



first look at top background performance

- double b-tag and sub-jet b-tag perform better than fat-jet b-tagging

Conclusions

We presented a first dedicated strategy towards double-b-tagging

We currently focus on H to $b\bar{b}$ signal and we compare to sub-jet and fat-jet b-tagging

- it can be used for any other resonance decaying to $b\bar{b}$ as well

We show the performance that can be gained with a dedicated approach over the past results

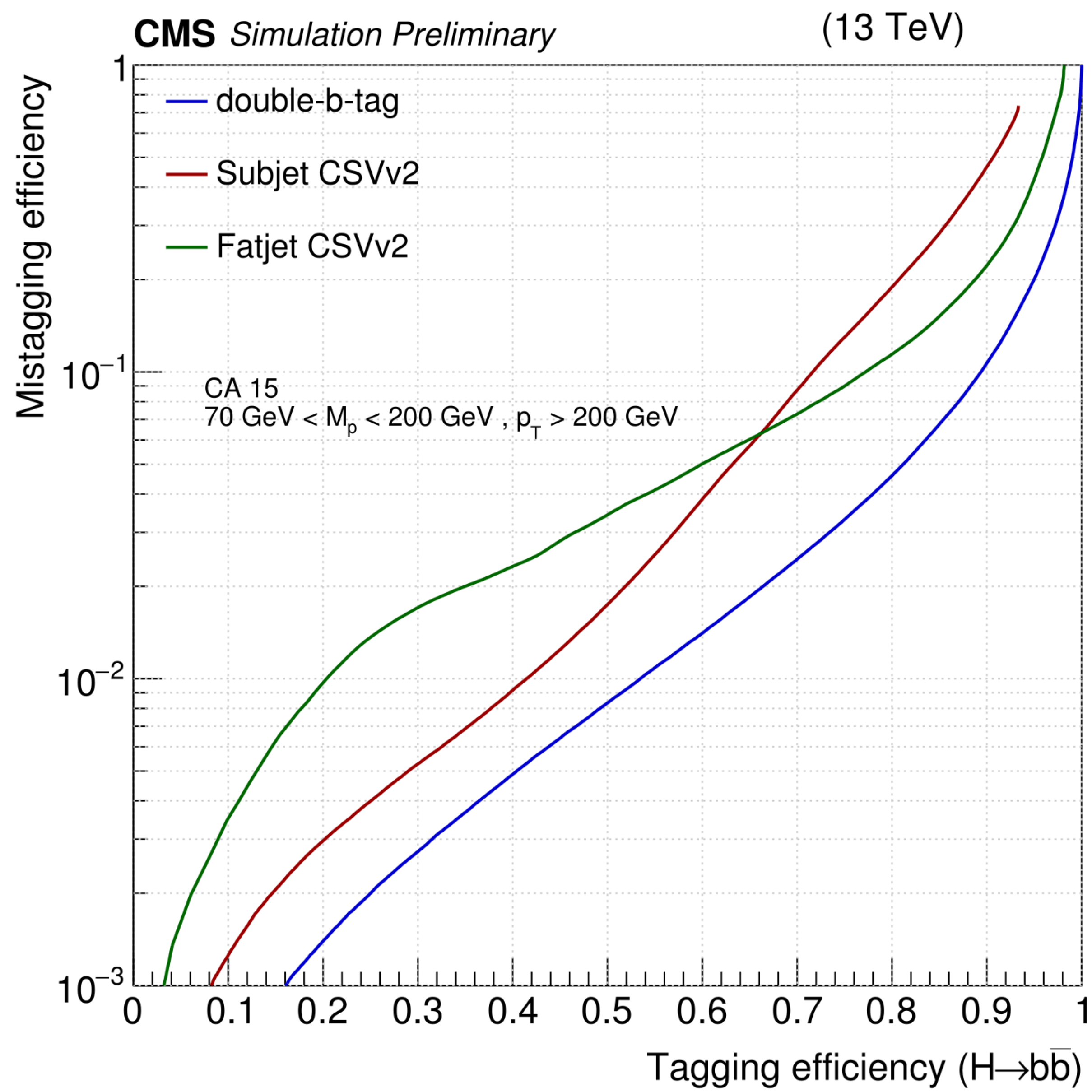
Studies are still on going

Just started, but …

- already good performance with respect to the previous approach
- performance are stable against different backgrounds and jet p_T

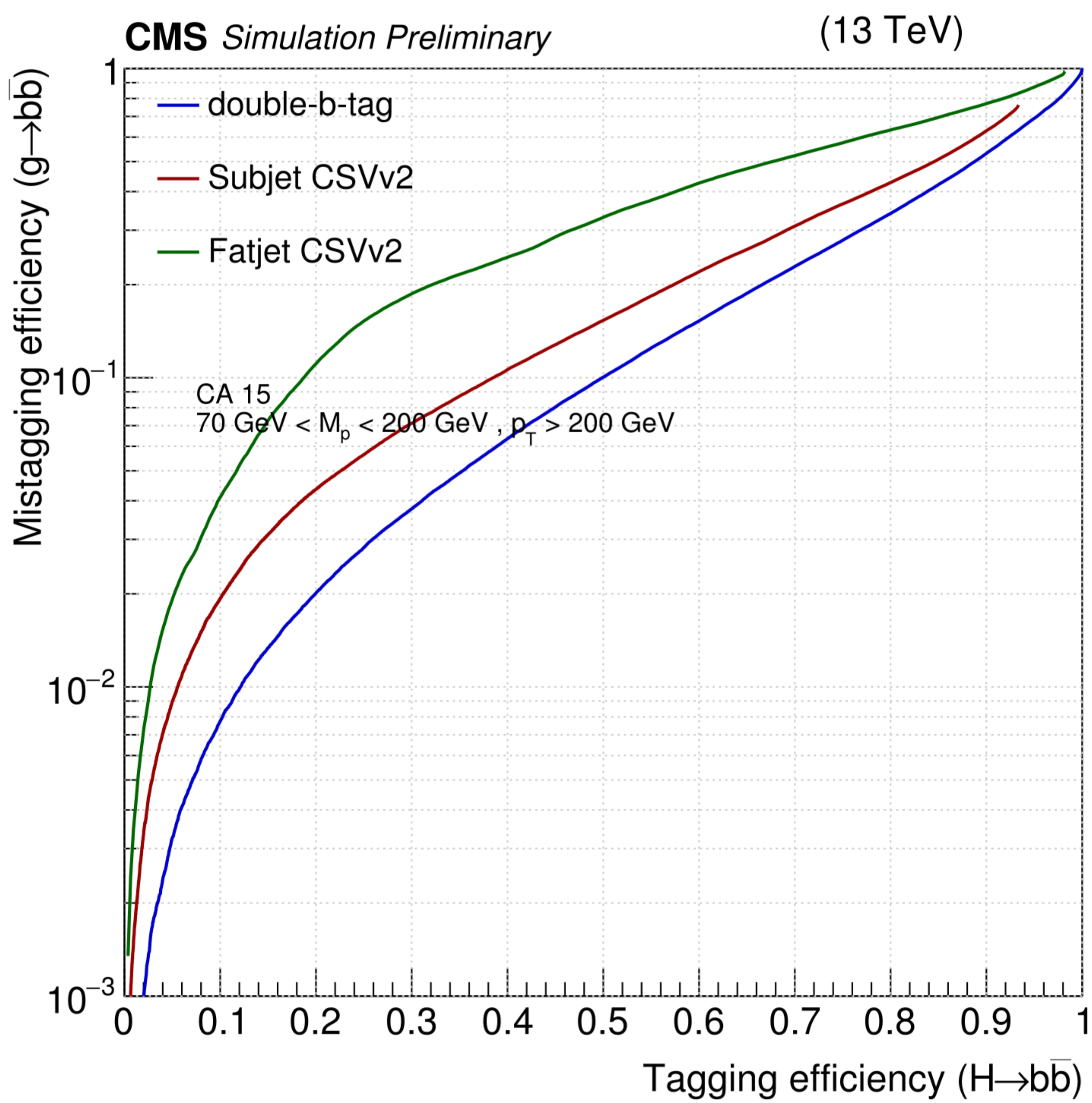
“ Additional Material ”

h(b$\bar{b}$) vs. Inclusive QCD



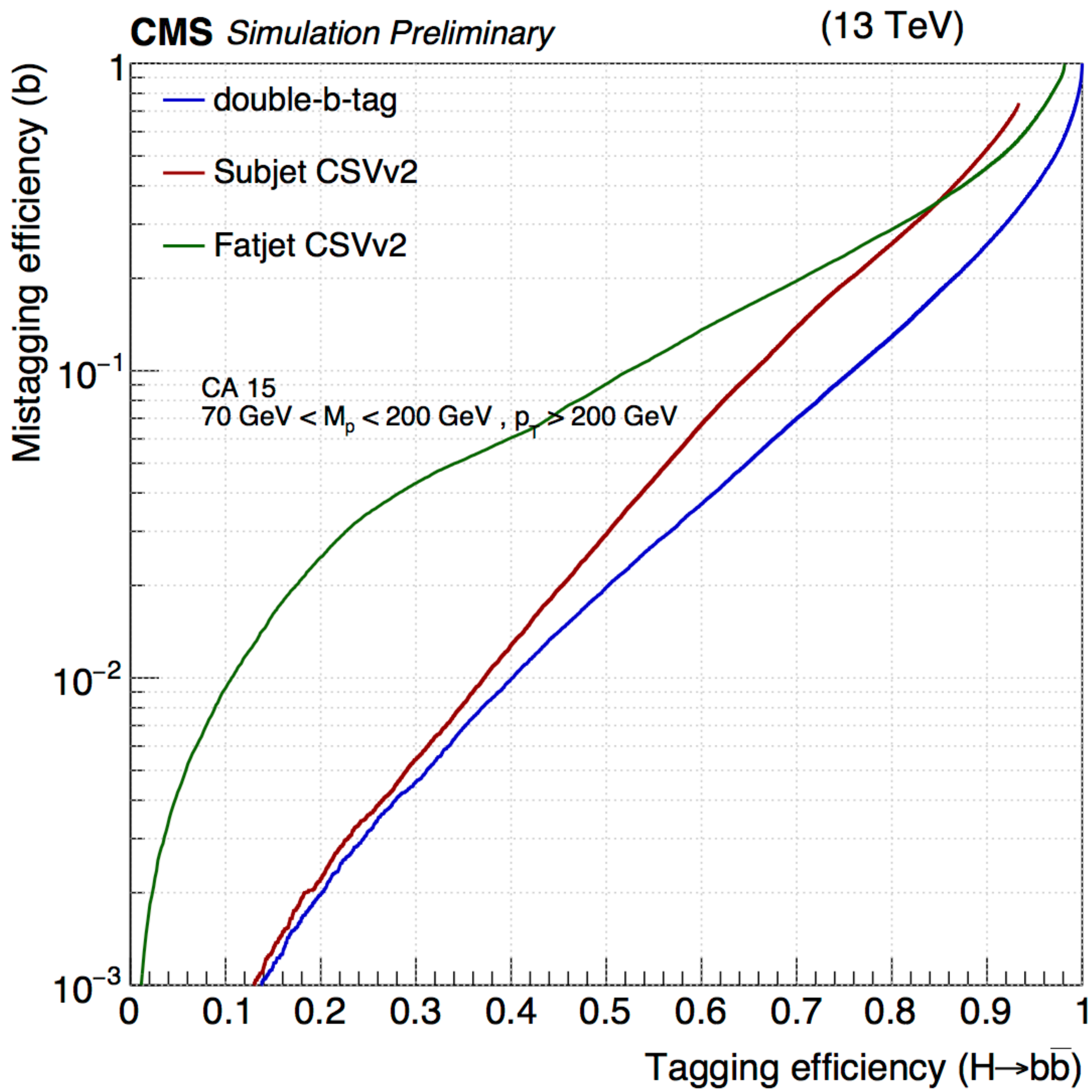
- double b-tag performs better than both sub-jet and fat-jet b-tagging in the entire p_T range considered

H($b\bar{b}$) jets vs. g($b\bar{b}$) jets



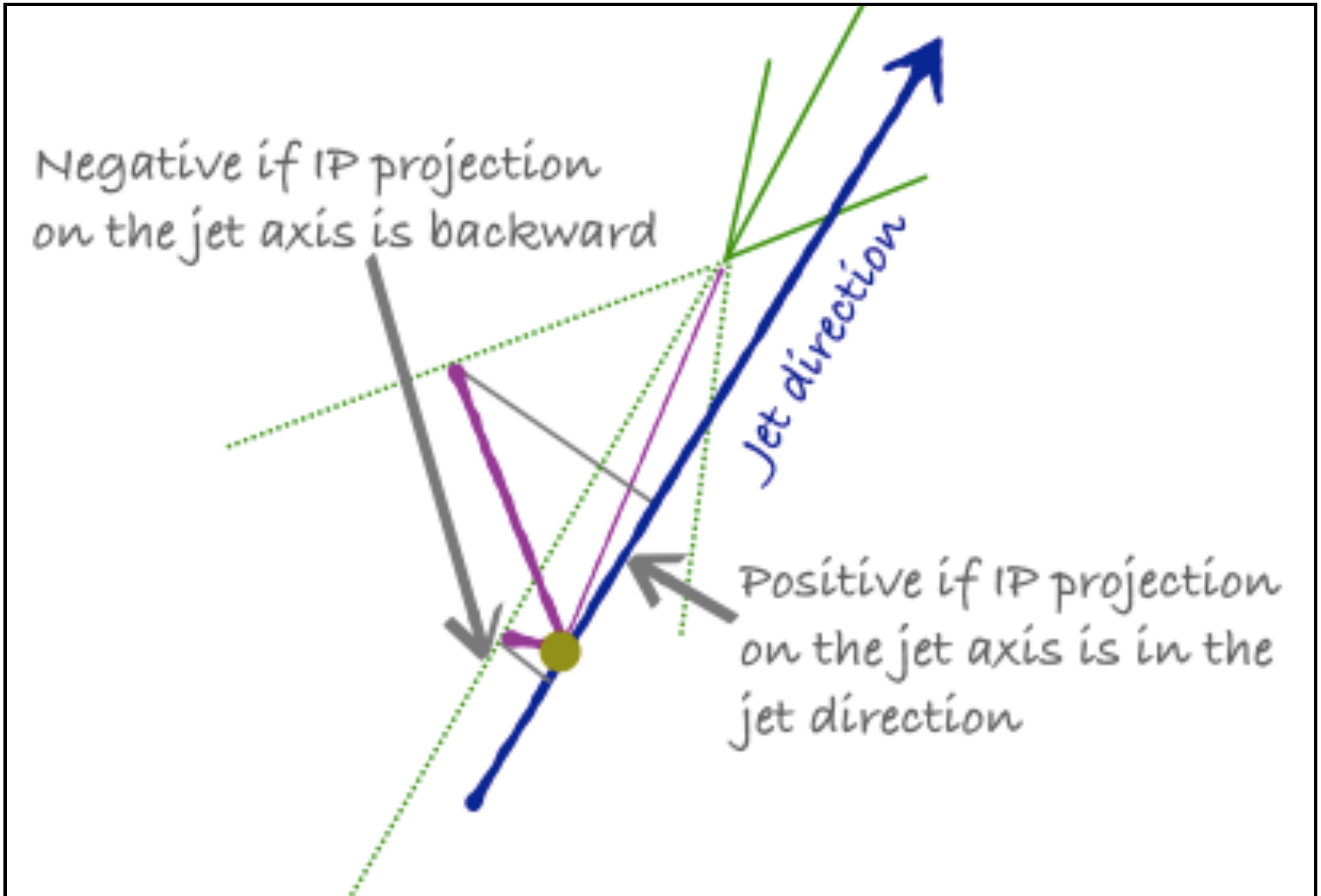
- double b-tag performs better than both sub-jet and fat-jet b-tagging

H($b\bar{b}$) jets vs. b-jets



- double b-tag and sub-jet b-tag performs better than fat-jet b-tagging

Impact parameter



Input Variables

Tracking

* number of tracks in the jet
* 3D signed IP significances
* IP significance of the first track that raises the invariant mass above the charm threshold of 1.5 GeV
* the pseudo-rapidity of the tracks at the vertex with respect to the jet axis

Secondary Vertex identified through the IVF algorithm. All the SV associated to the AK08/CA15 jet are considered and ordered according to their p_T.

* number of tracks at the vertex;
* number of secondary vertices
* ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet;
* 2D flight distance significance;
* The mass of a SV is computed as the invariant mass of tracks associated to the SV
* ΔR between the secondary vertex flight direction and the jet axis

Soft Lepton All the SL associated to the AK08/CA15 jet are considered and ordered according to their $p_{T,rel}$ relative to the jet. The leading one is considered.

* number of SL .
* p_T rel of the leading SL
* p_T ratio of the leading SL and the jet p_T

SubJetCSVIVF, the minimum score of the two subjet CSVv2IVF b-tag

Inclusive Vertex Finder Algorithm

- IVF does not depend on jet direction
  - Make a list of seeding tracks from high IP tracks
  - Check every other track against the seed and associate them in a cluster
- All tracks in the event (or in the jet if one want to speed up the algorithm) are checked for compatibility with a given seed
  - The track should be close enough to the seed
  - The point of closest approach between the tracks should be far enough from the PV compared to the distance between the two tracks
- The main compatibility criterion is the minimum distance between the two tracks, compared to the distance of their 2- track-vertex to the primary vertex
  - track density decreases with distance from PV, so the chances to have a track near the seed decreases with with r
- For each cluster try a vertex fitting