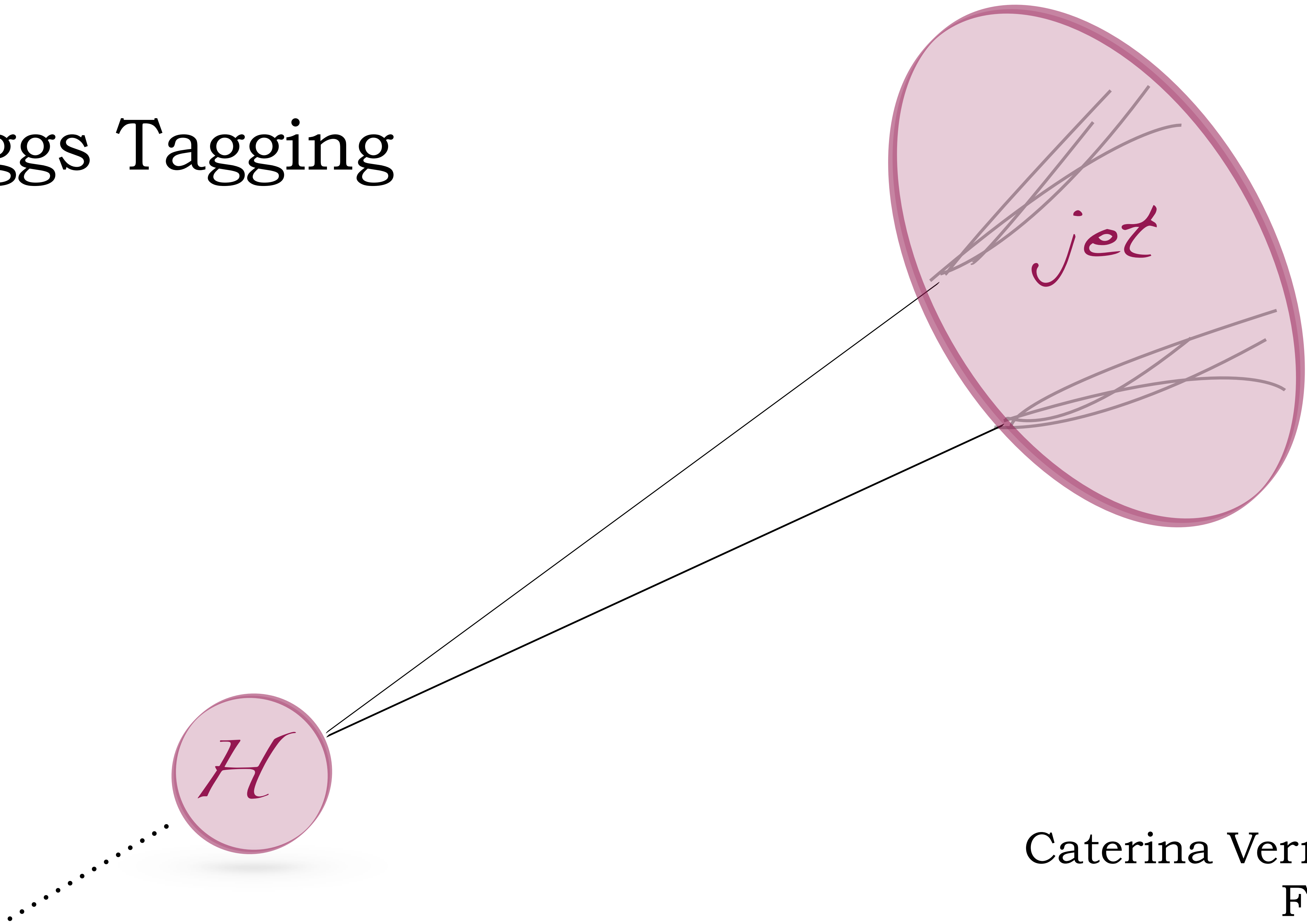
Higgs Tagging



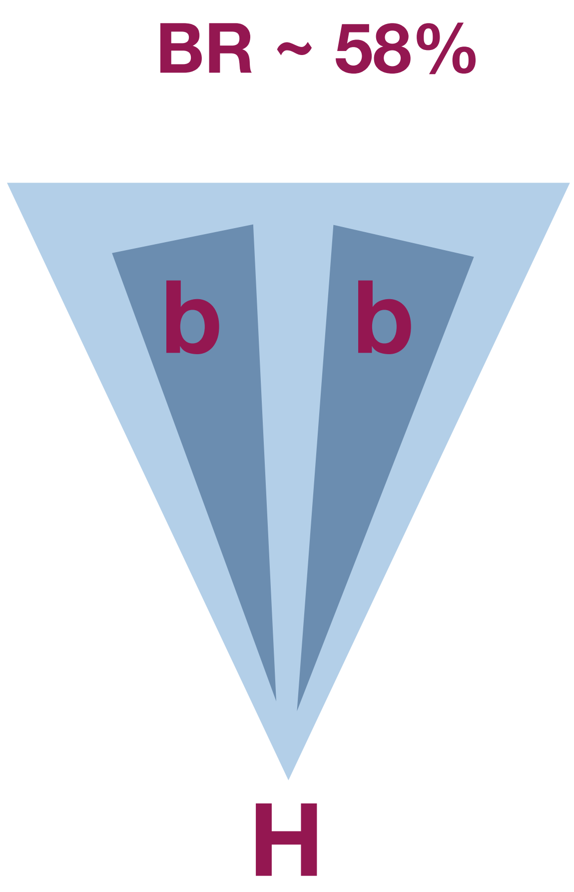
Caterina Vernieri
FNAL

Jet Substructure "Planning for the future" workshop @ LPC
November 30 2016

Higgs as tool for discovery

- In many **BSM scenarios** the H could be used as tool for discovery new states:
  - hh, Vh resonance
  - t’→ th, b’→bh searches
  - χ0→ hG
- SM decay **h→ $b\bar{b}$** provides the **largest branching ratio**
  - Boosted topology allows a few handles to suppress dominant multi-jet background
  - **Boosted h→ $b\bar{b}$** tagger becomes an essential tool for new physics search
    - In principle the same approach should work for Z to $b\bar{b}$ and X to $b\bar{b}$
- But also **h→WW** and **h→ττ** are being exploited
  - They add sensitivity to searches for new physics states
  - Dedicated tagging techniques available

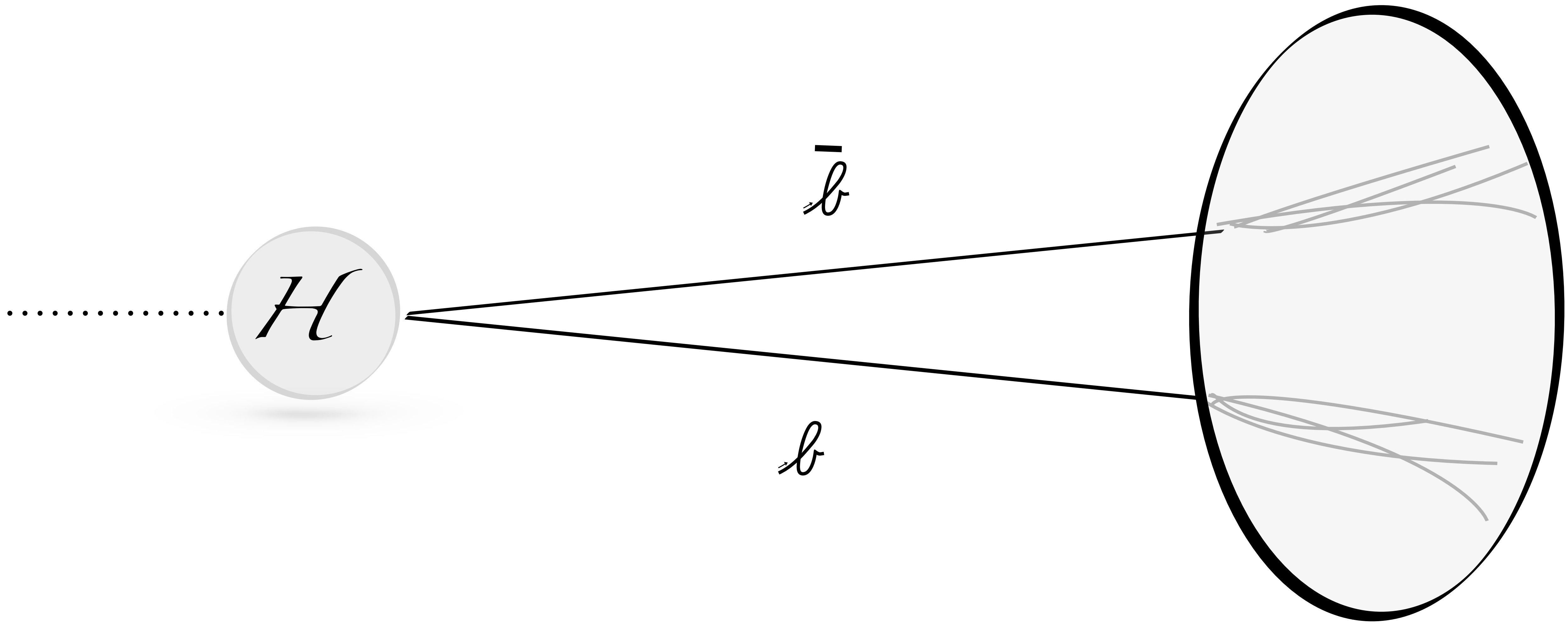
$h(b\bar{b})$ tagging



The boosted $h(b\bar{b})$ signal can be identified by exploiting:

- large-R jet **mass**
- the composite nature of the jet using **substructure**
- **b-tagging** to reconstruct the two B hadrons from the b and $\bar{b}$ within the same fat jet

I.MASS

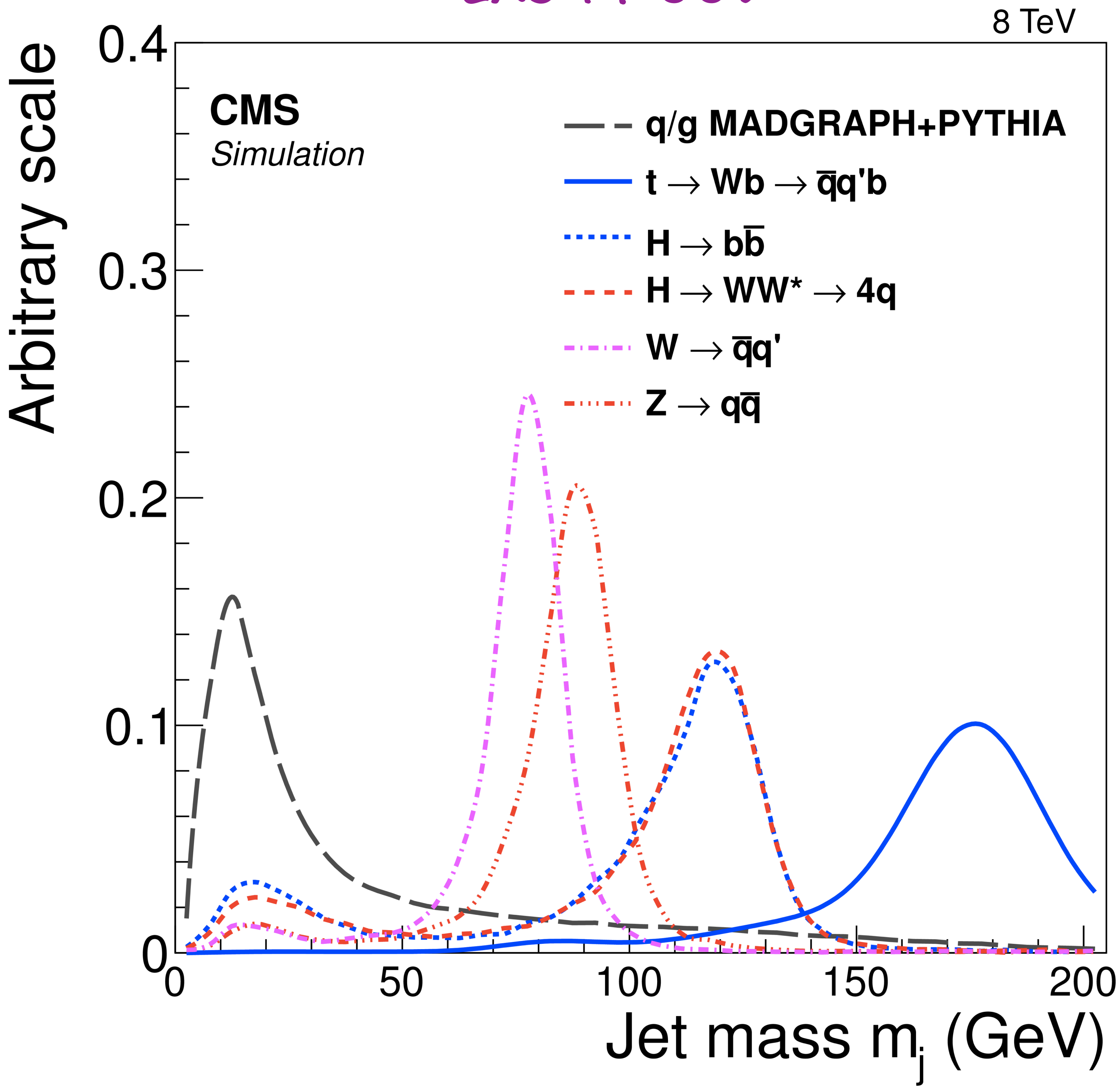


CMS Jet Mass Selection

See C. Mantilla's talk

- anti-kt **0.8** jets
- **Pruned jet mass** is required to be 105 < m < 135 GeV
- Lower bound optimized in the context of VV searches
  - to avoid overlap with VV, HH and VH searches are using this mass window
  - **60%** efficiency on h-jets

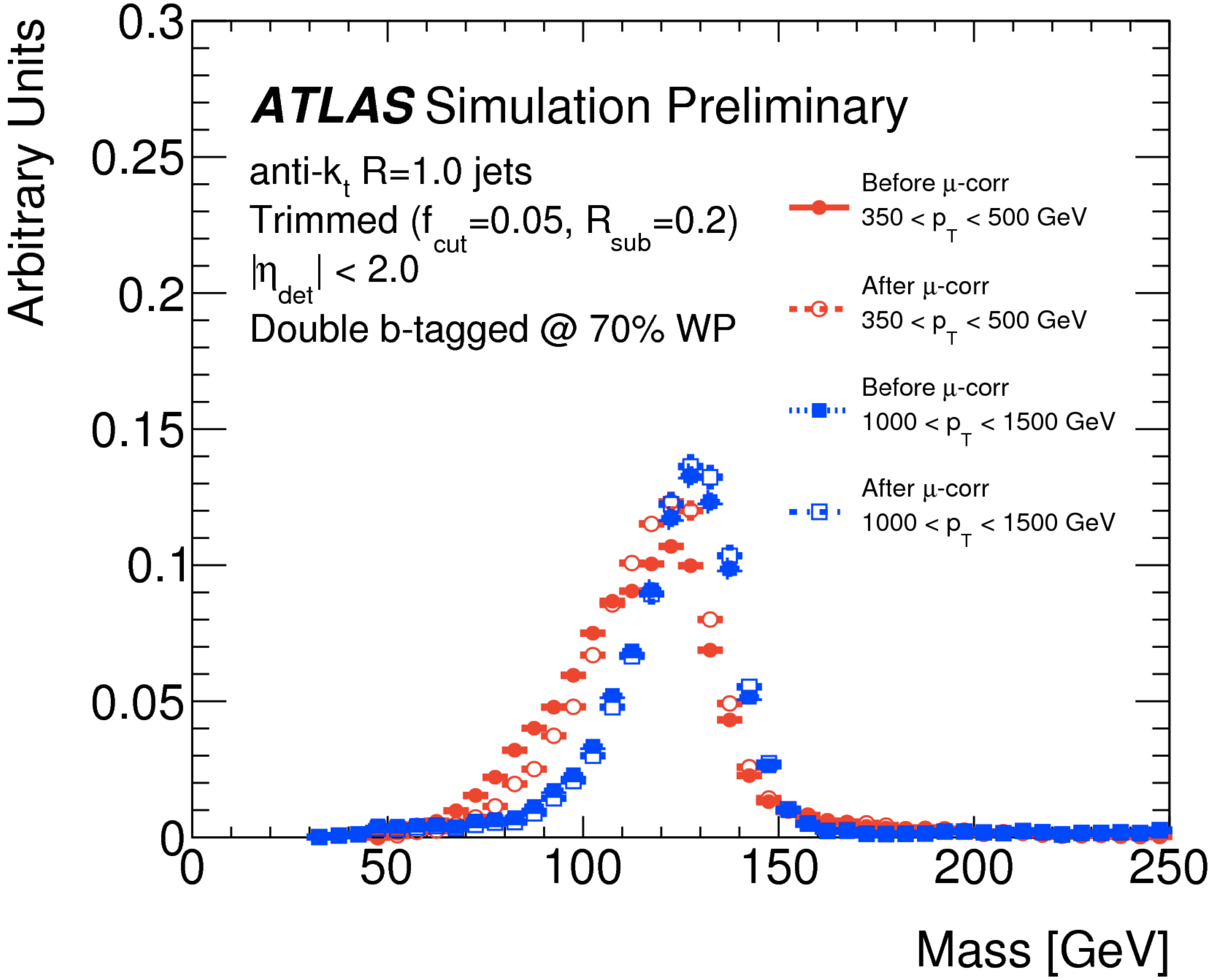
EXO-14-009



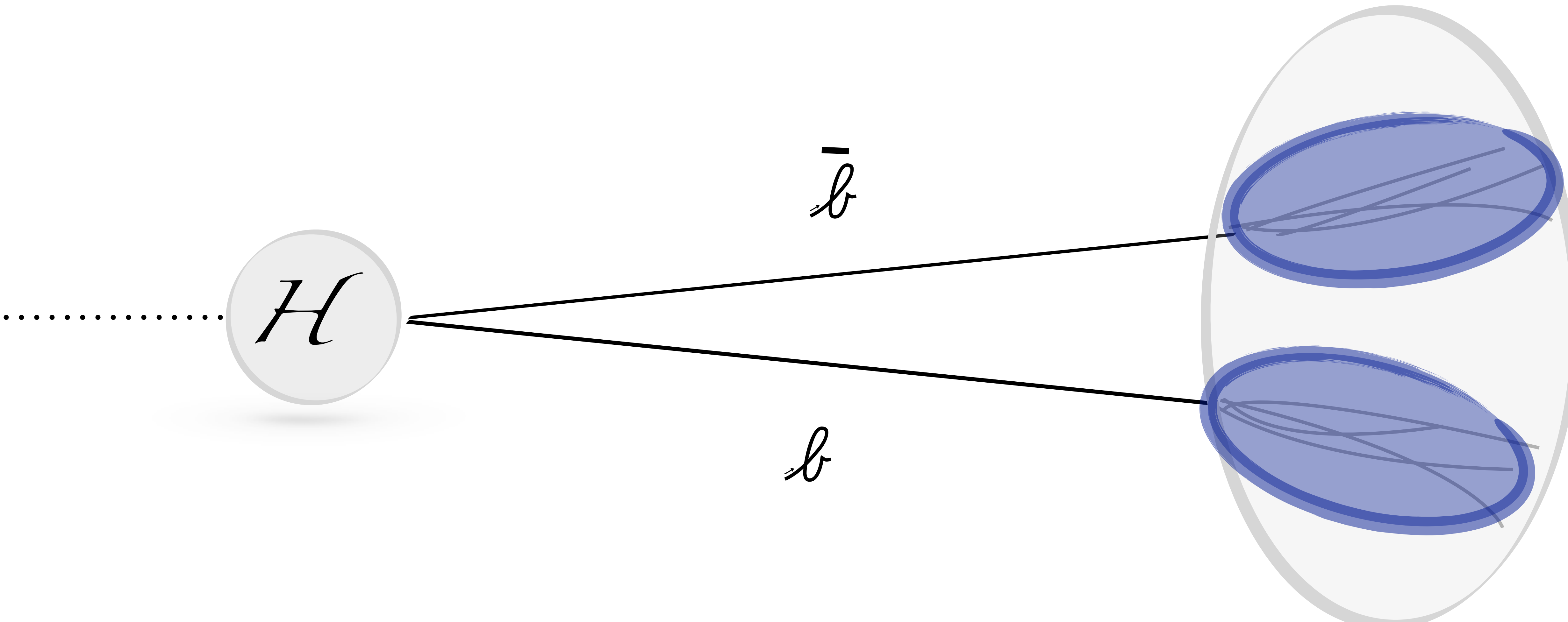
ATLAS Jet Mass Selection

- anti-kt **1.0** jets
- **Trimming**
- Muon-in-b-jet correction correcting for semi-leptonic b hadron decays
- Mass windows used are such that the efficiency on h-jets is
  - **90%** (76<m<146)
  - **68%** (93<m<134)

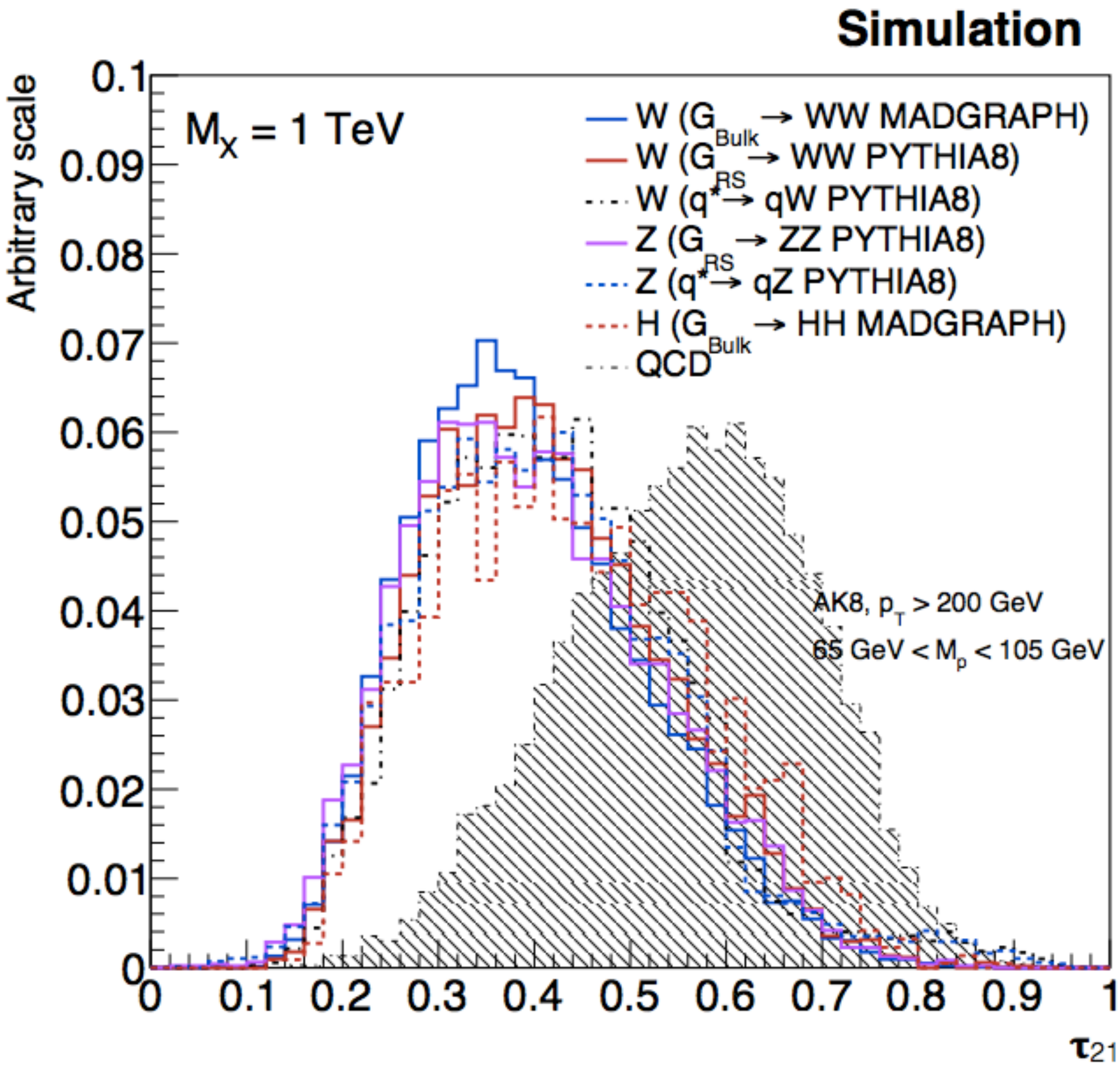
ATL-PHYS-PUB-2015-035



2. SUB-STRUCTURE



CMS, n-jettiness

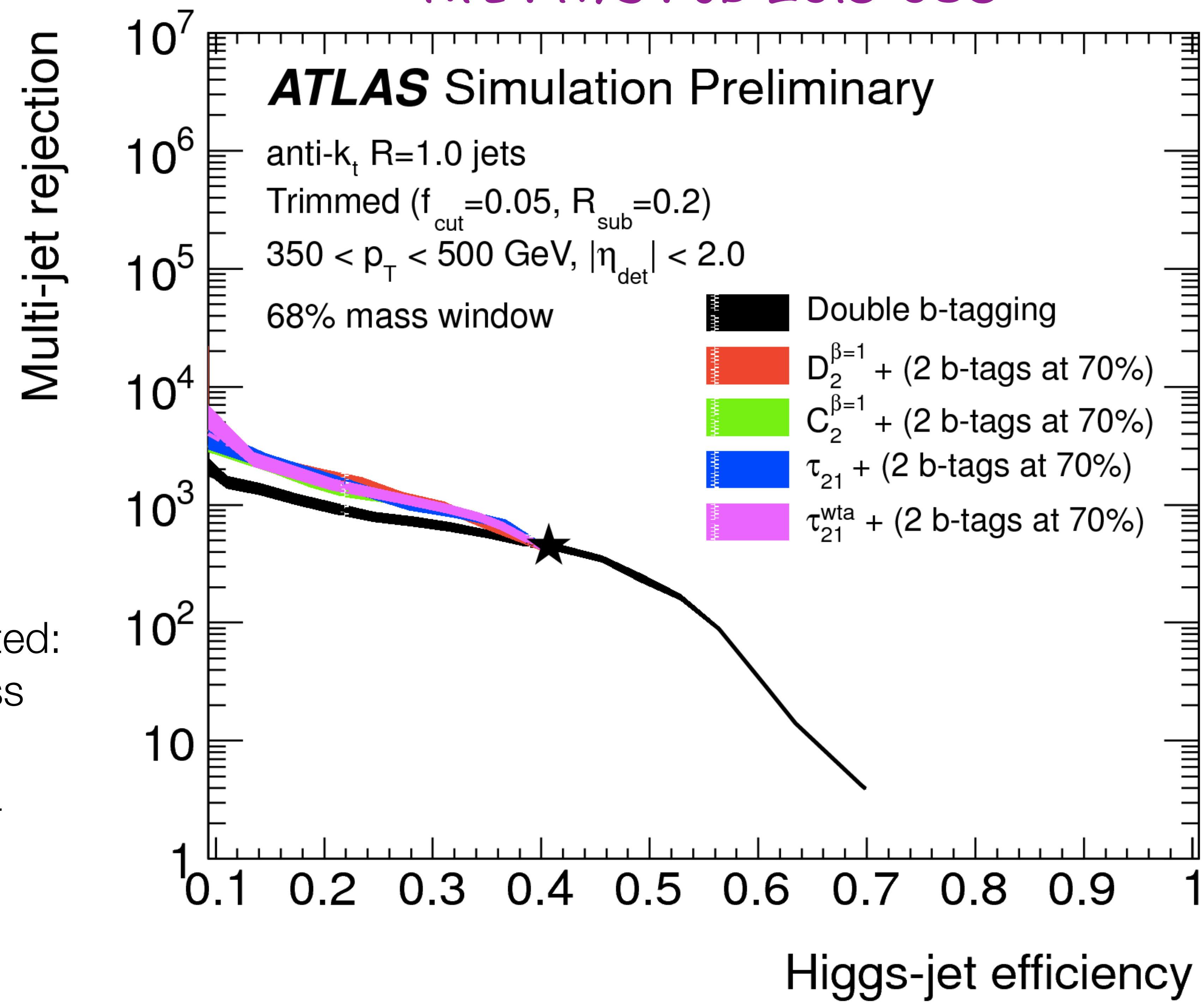


- **τ2/τ1=τ21** measures how consistent the jet is with having 2 sub-jets,
- In 2015 CMS analyses used a loose selection
  - **90%** efficiency on h-jet
- One of the dominant systematic uncertainty
  - after b-tagging it doesn't add much discrimination
  - some searches use **only mass +b-tagging** B2G-16-003, EXO-16-020

ATLAS, D2 substructure

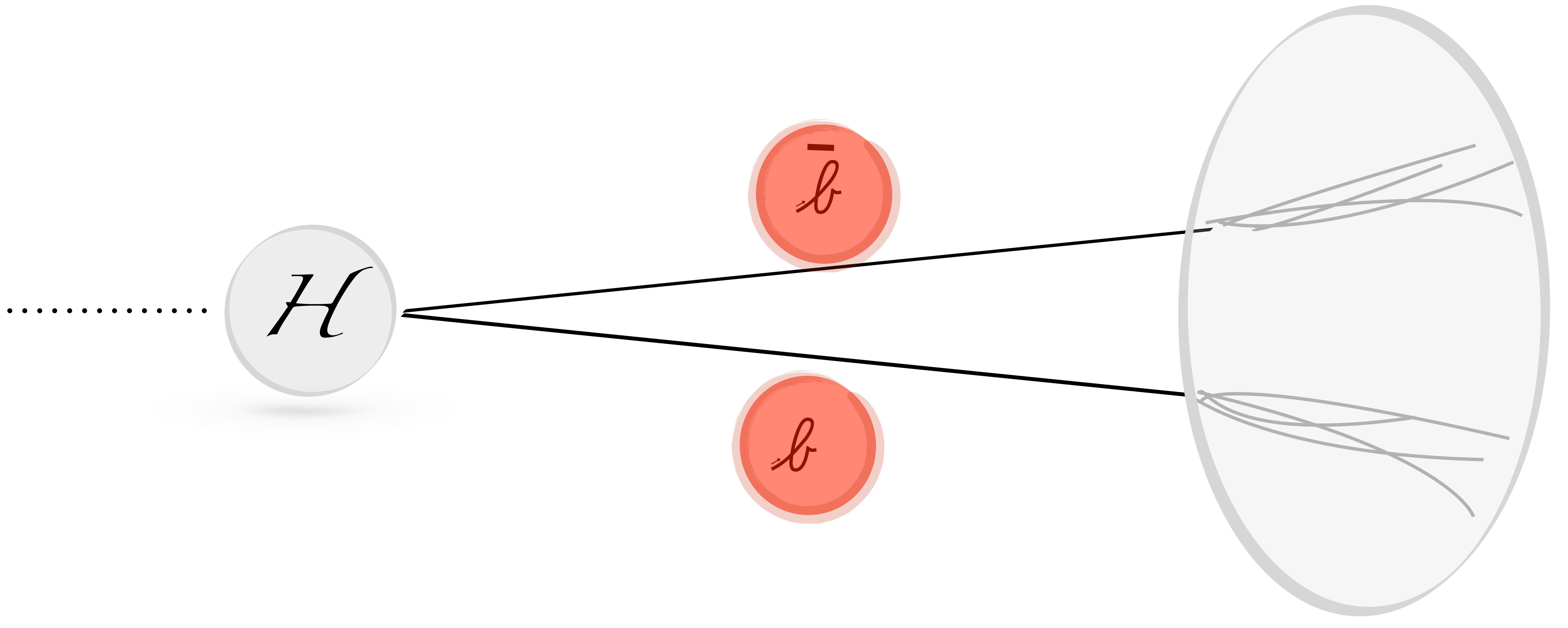
See C. Mantilla's talk

ATL-PHYS-PUB-2015-035

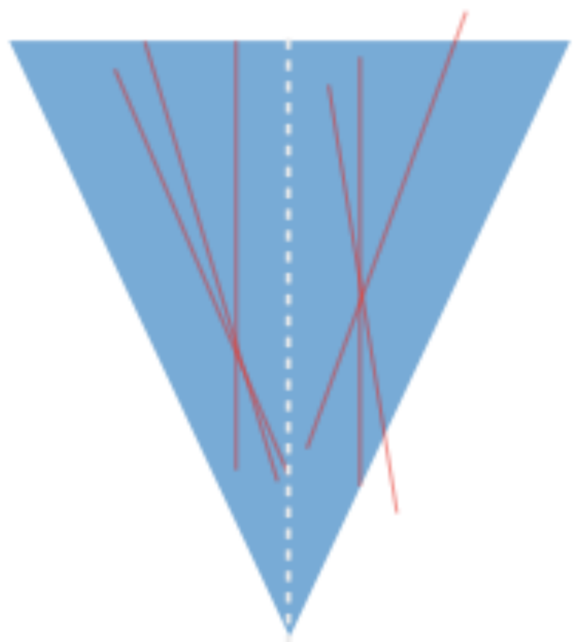


- Several variables investigated:
- Similar performance across
  - **D2** is chosen due to better modeling in data

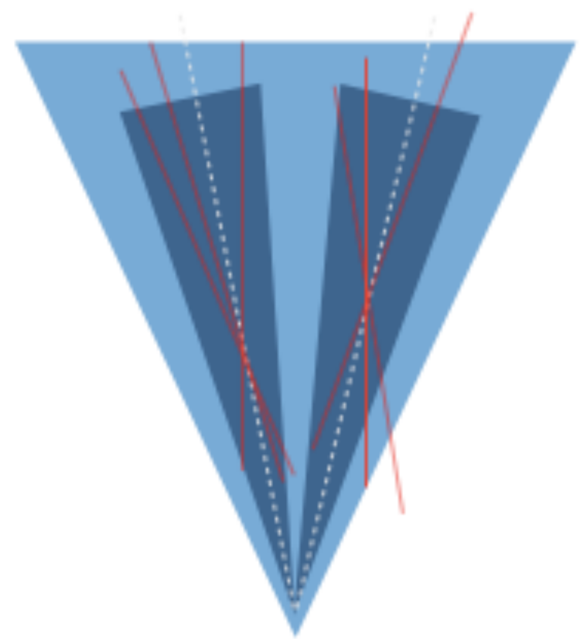
3. B-TAGGING



b-tagging, multiple approaches



fatjet



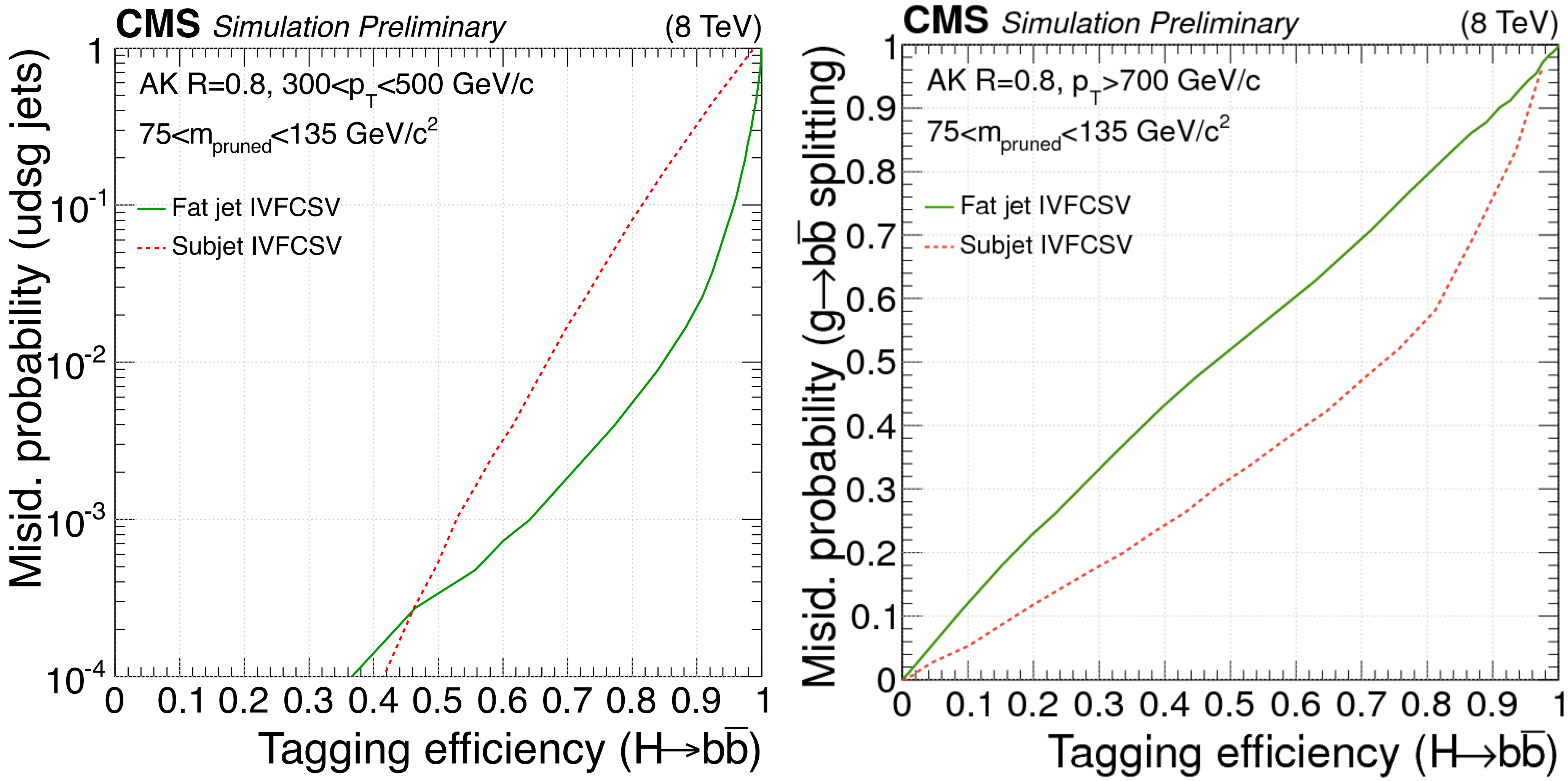
subjets

observables from SV and tracks associated to the fat-jet

b-tagging observables for each sub-jet

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

CMS DP-2014/031



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

sub-jet b tagging b-tagging applied to each sub-jet

Fat-jet b-tagging works well against udsg but not $g(b\bar{b})$

Sub-jet b-tagging improves discrimination against $g(b\bar{b})$ but it depends on p_T

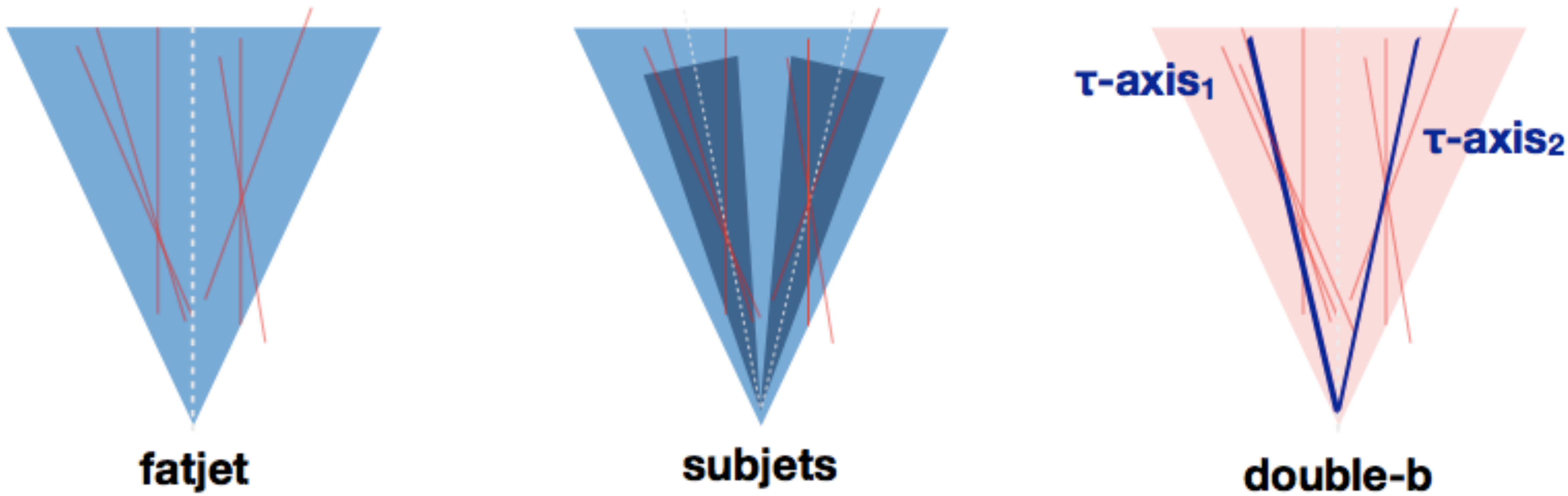
At very high p_T jets from $h(b\bar{b})$ get too close

sub-jet b-tagging loose discrimination against $g(b\bar{b})$

b-tagging, multiple approaches

BTV-15-002

NEW



observables from SV and tracks associated to the fat-jet

b-tagging observables for each sub-jet

observables from SV and tracks associated to each τ-axis

- **b-tagging** two approaches
  - AK0.4/0.2 sub-jets b-tagging (CMS/ATLAS)
  - double-b-tag (CMS)

CMS, double-b tagger

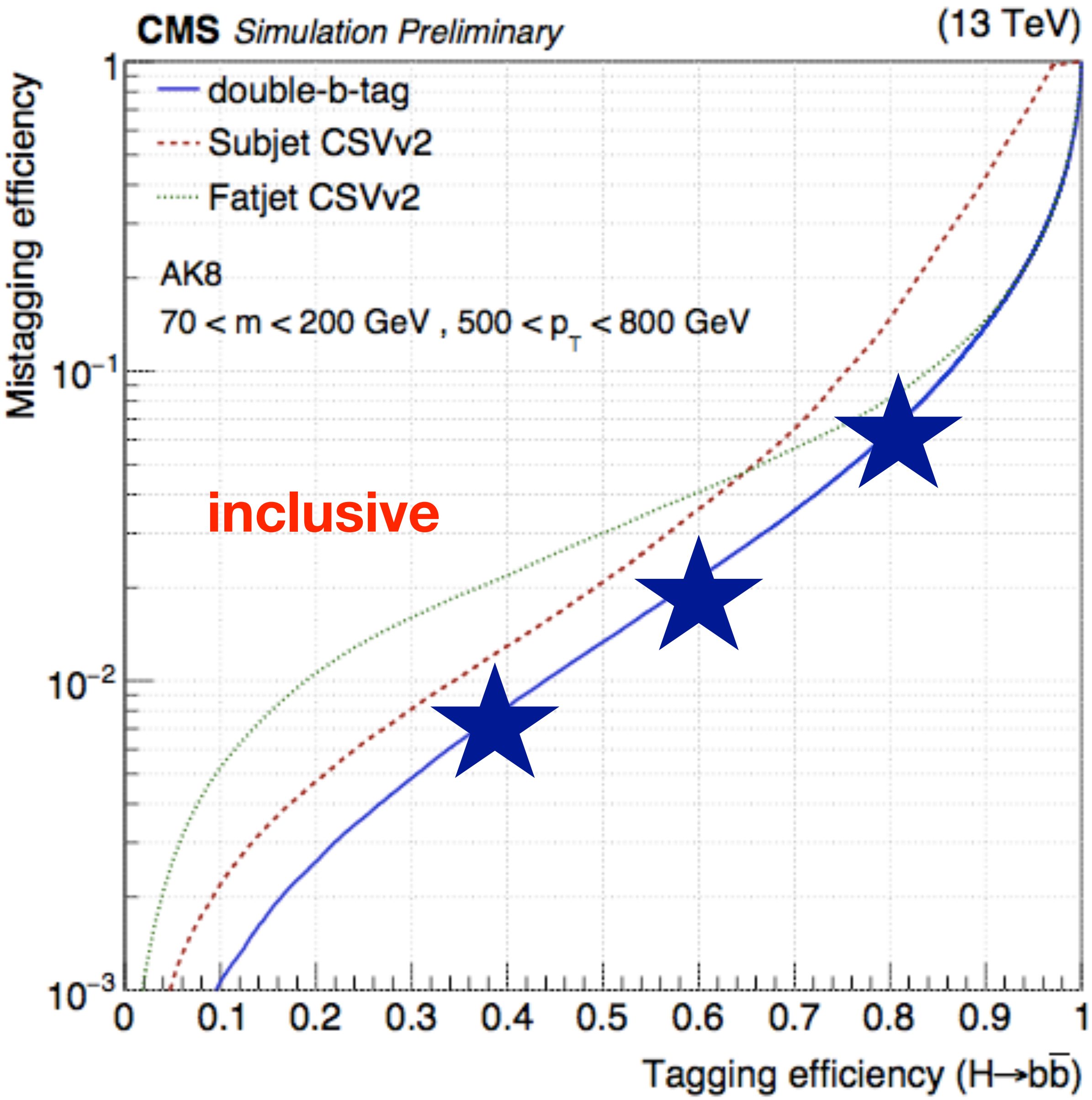
General strategy:

- Exploit b-tagging to identify both the b and $\bar{b}$ within the same fat jet
- Use the Inclusive Vertex Finder algorithm (**IVF**) which identifies secondary vertex independently of jet reconstruction

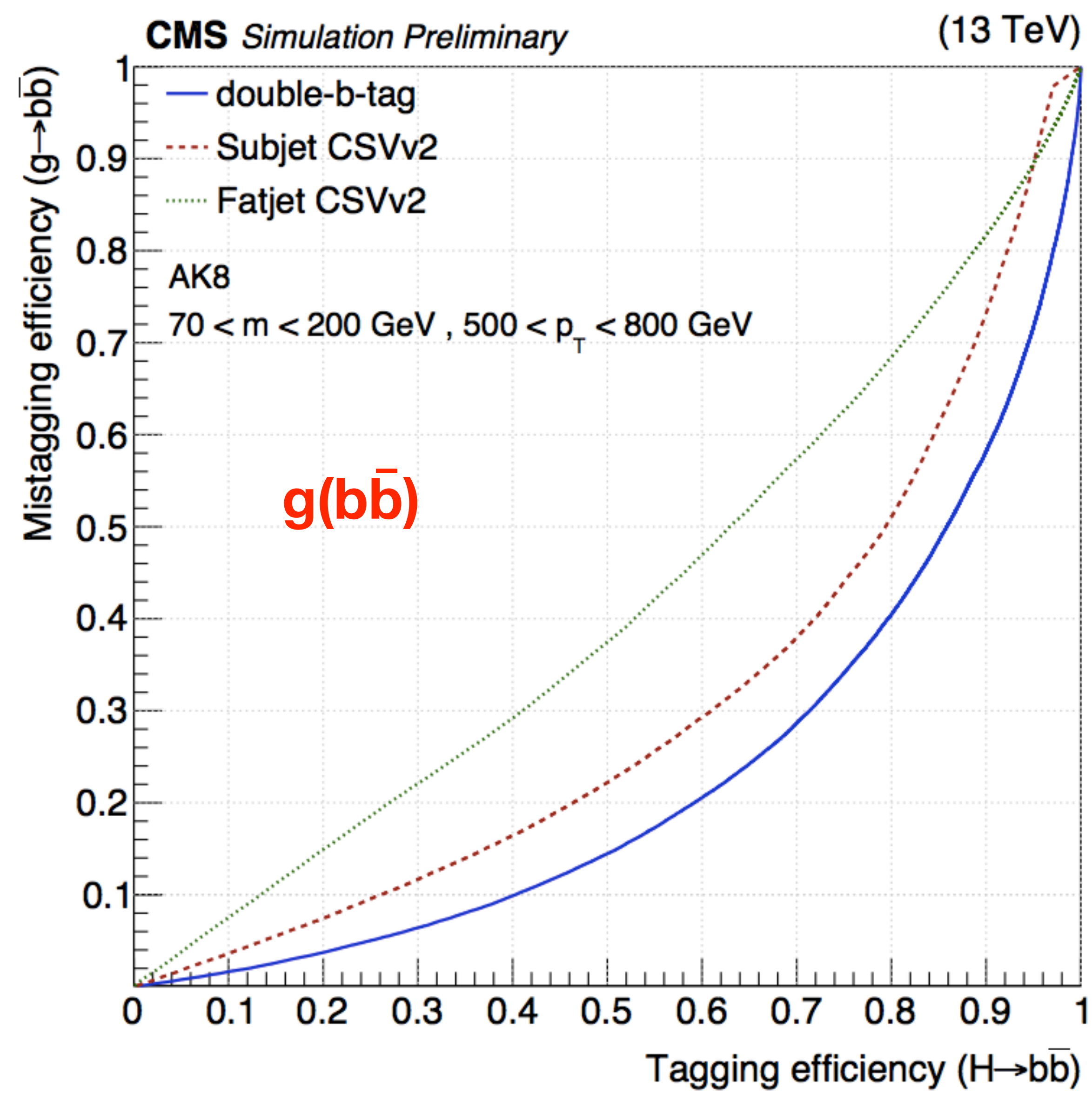
- The **double-b** combines tracking and vertexing information in an MVA
- It targets the $b\bar{b}$ signal from a resonance, not just the Higgs boson, aiming to be:
  - **mass independent**
    - easier to validate
    - can be applied to Z to $b\bar{b}$ as well as any BSM particles decaying to $b\bar{b}$
  - **p_T independent**, to better adapt to different kinematic regimes
    - training is performed using a very wide p_T range for both signal and background
    - inputs are chosen in order to not have any strong p_T correlation

CMS, Performance

BTV-15-002



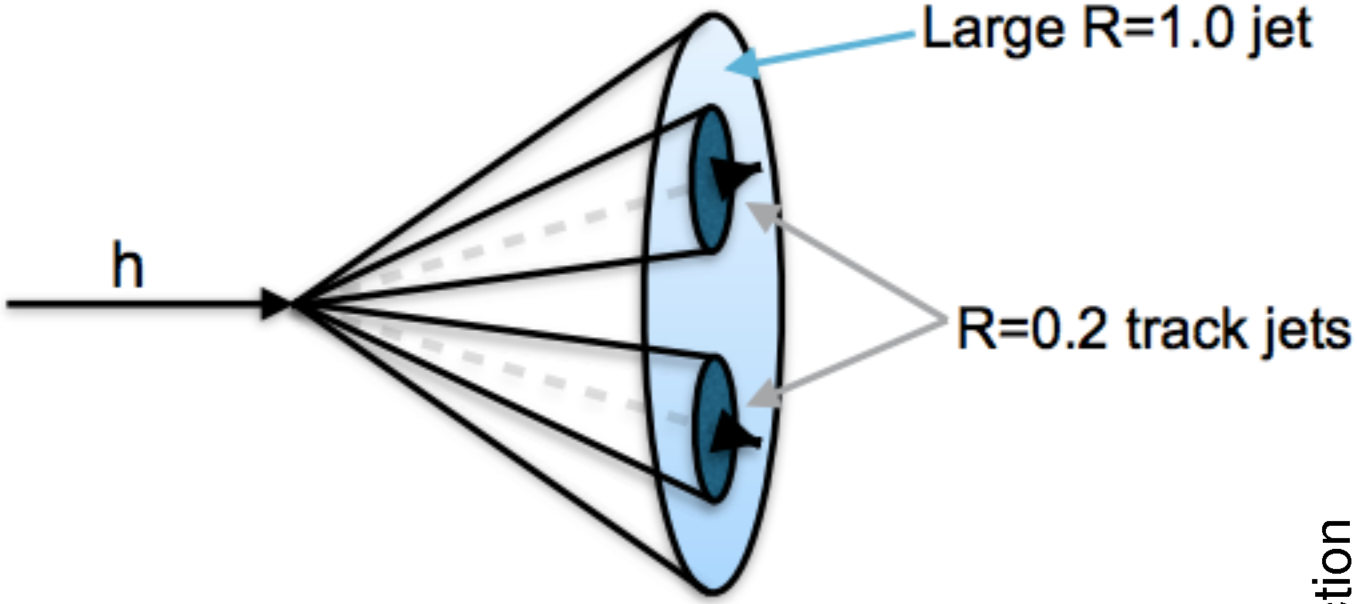
BTV-15-002



Improved performance with respect to both fat and sub-jet b-tagging

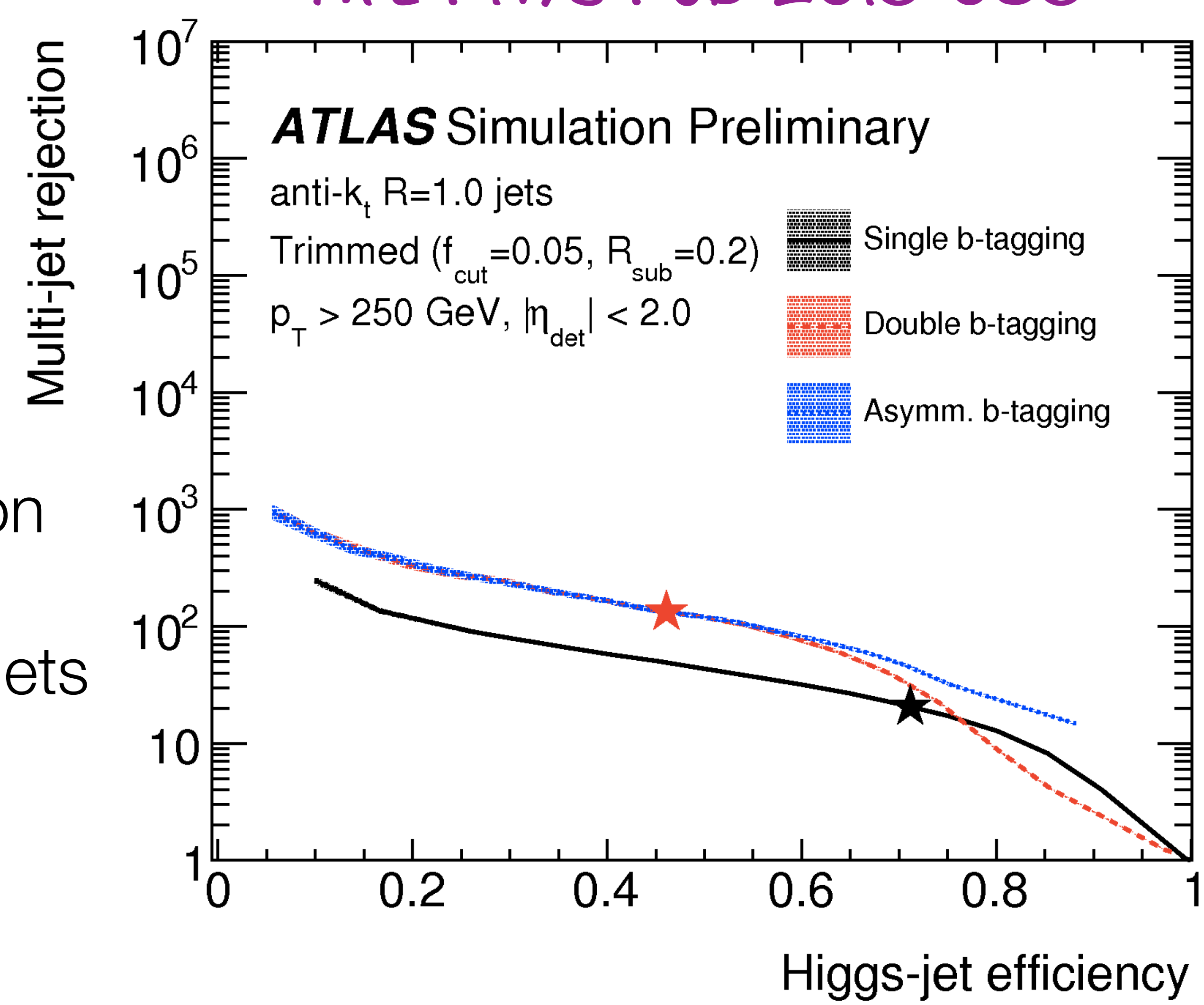
At high p_T larger improvement as planned

ATLAS, sub-jet b-tagging



Use **small radius (R=0.2) track jets** to resolve close-by B-hadrons

ATL-PHYS-PUB-2015-035



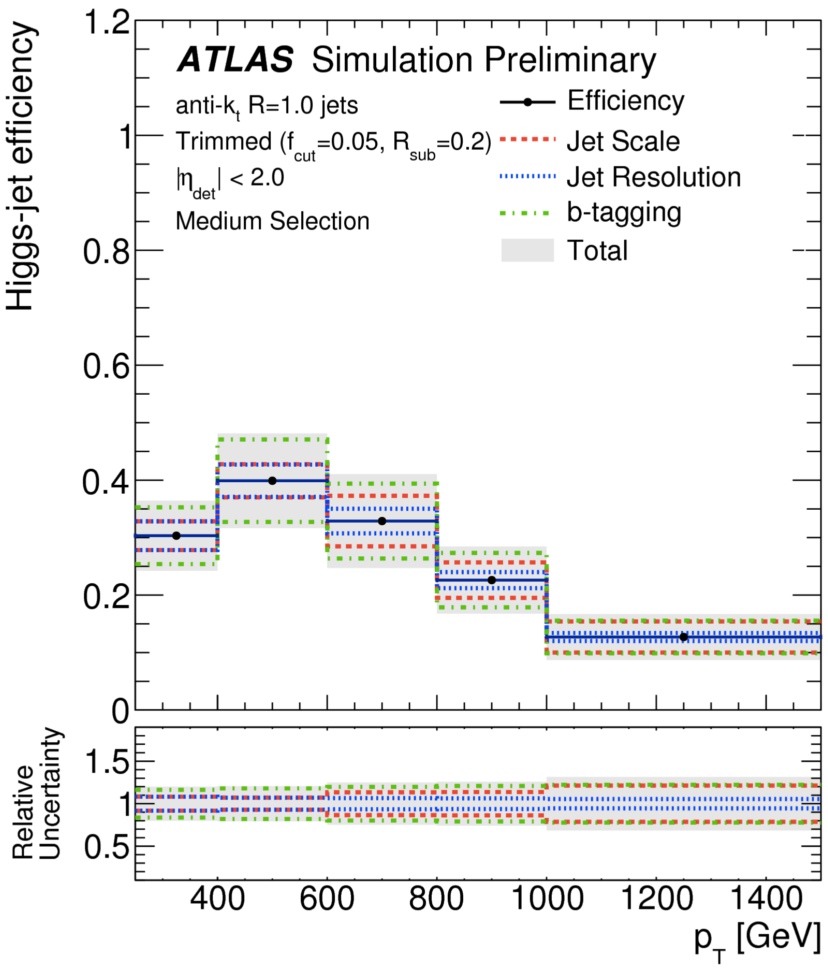
- Advantage of track jets
  - Better estimate b-hadron flight direction
  - Pile-up resistant
  - b-tagging independent of calorimeter jets

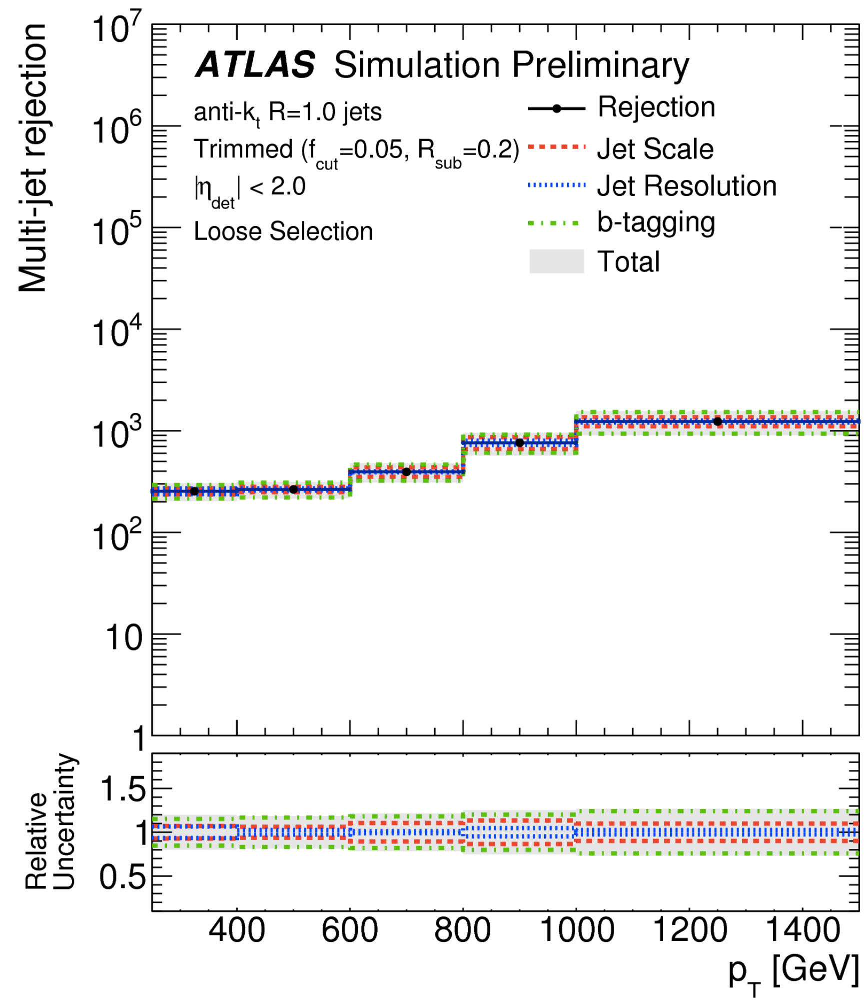
Performance

- Three working points (WP) defined

Selection	double b-tagging	jet Mass	D2
Loose	70% WP	90% window, $m \in [76, 146]$ GeV	-
Medium	70% WP	68% window, $m \in [93, 134]$ GeV	-
Tight	70% WP	68% window, $m \in [93, 134]$ GeV	p_{T}-dependent cut

ATL-PHYS-PUB-2015-035





- Systematic uncertainties:
  - b-tagging largest for loose selection
  - Jet energy/mass scale & resolution
  - larger for tight selection

VALIDATION IN DATA

- We don't have a enough $h \rightarrow b\bar{b}$ in data — not even observed yet…
  - boosted approach could help SM searches too…
- **Substrucure** and **jet mass** are monitored using W-jet from tt semileptonic events
  - we extrapolate for the h case what is measured for W-jets
  - an additional uncertainty to propagate from W-jets to h-jets by comparing PYTHIA 8 and HERWIG++ hadronizers
- **b-tagging** is calibrated using an enriched sample of $g \rightarrow b\bar{b}$
  - close-by b-jets topology

- **Substrucure, jet mass** and **b-tagging** are all monitored using an enriched sample of $g \rightarrow b\bar{b}$

CMS, Validation in data

double-b

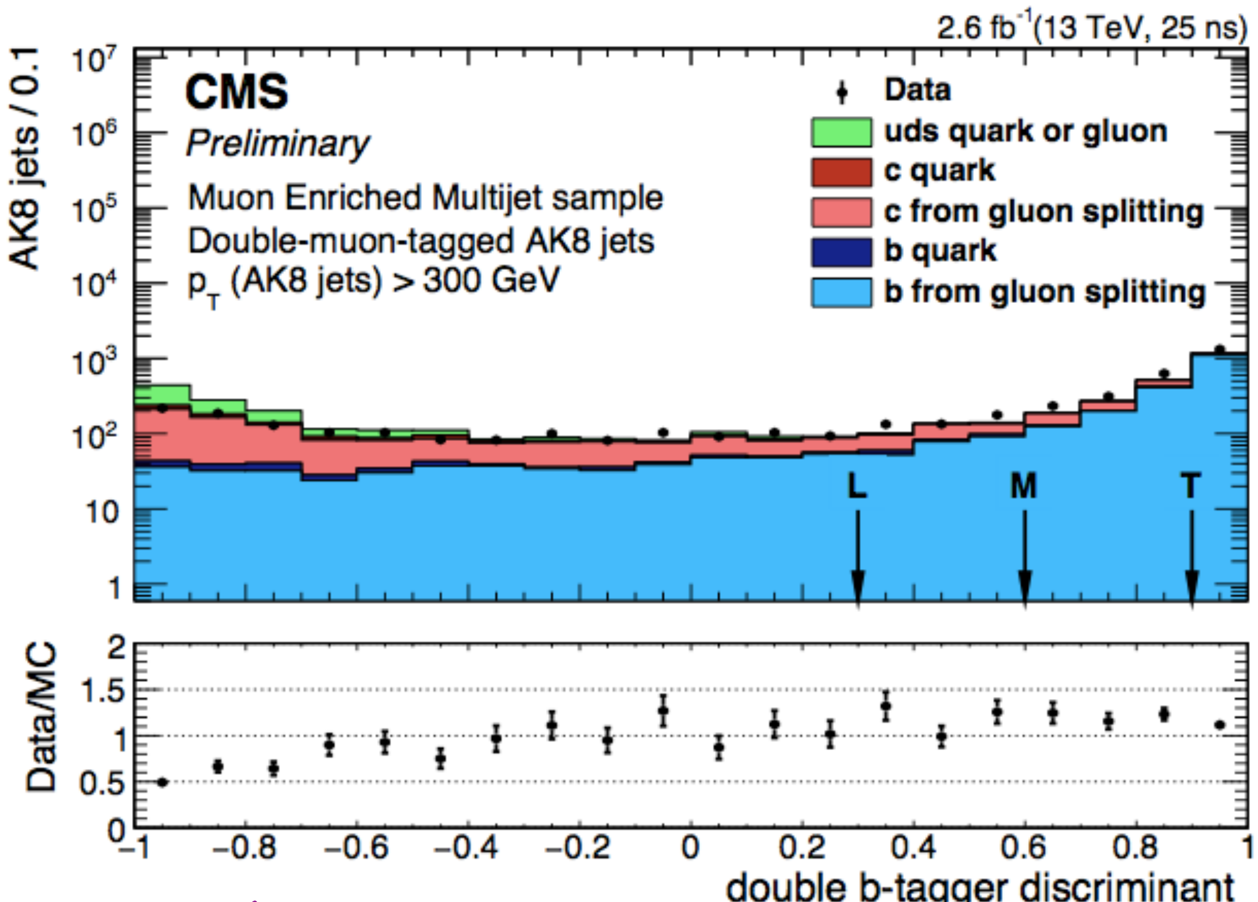
Double Muon tagged AK8 jets with p_T >300 GeV and jet mass > 50 GeV

– at least 2 muon matched to the fat jet to select QCD gluon splitting more signal-like

sub-jet b-tagging

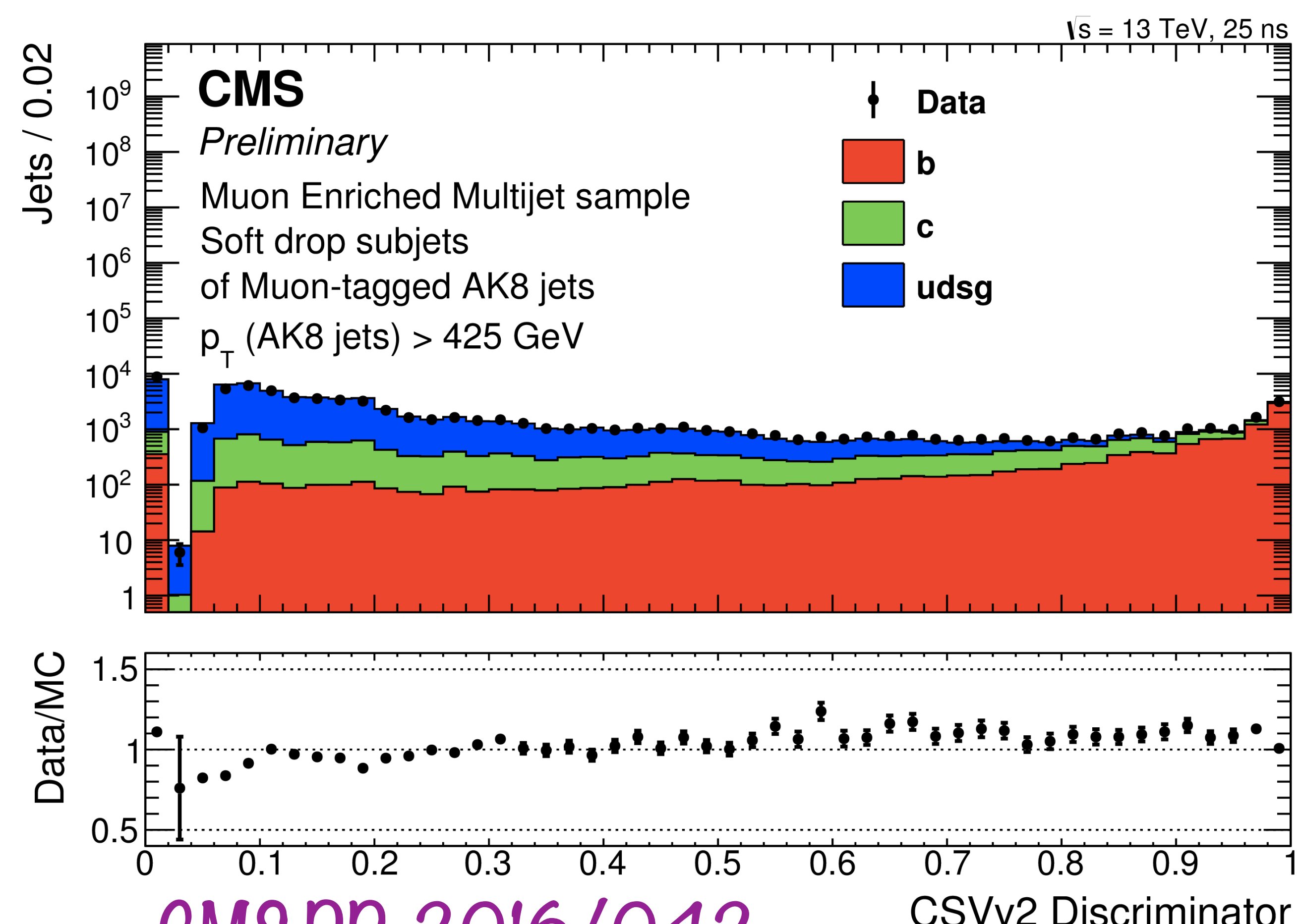
At least one muon matched to the AK8 jet with p_T >425 GeV

DOUBLE-B



BTV-15-002

SUBJET-B-TAG

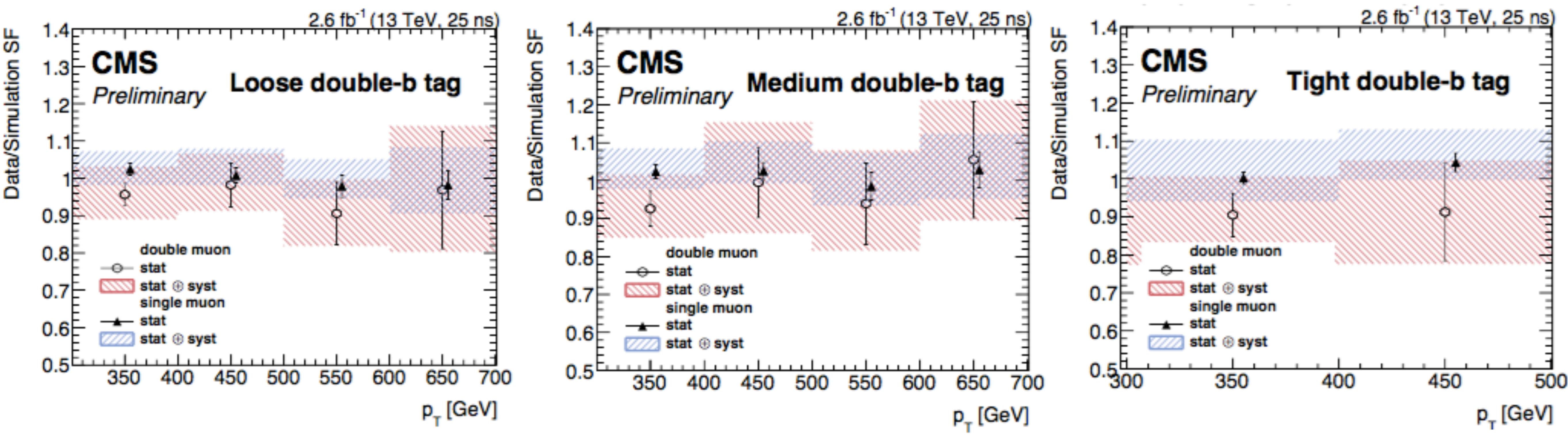


CMS DP-2016/042

Data/MC agrees reasonably well

CMS, data/MC SF for the double-b

BTV-15-002



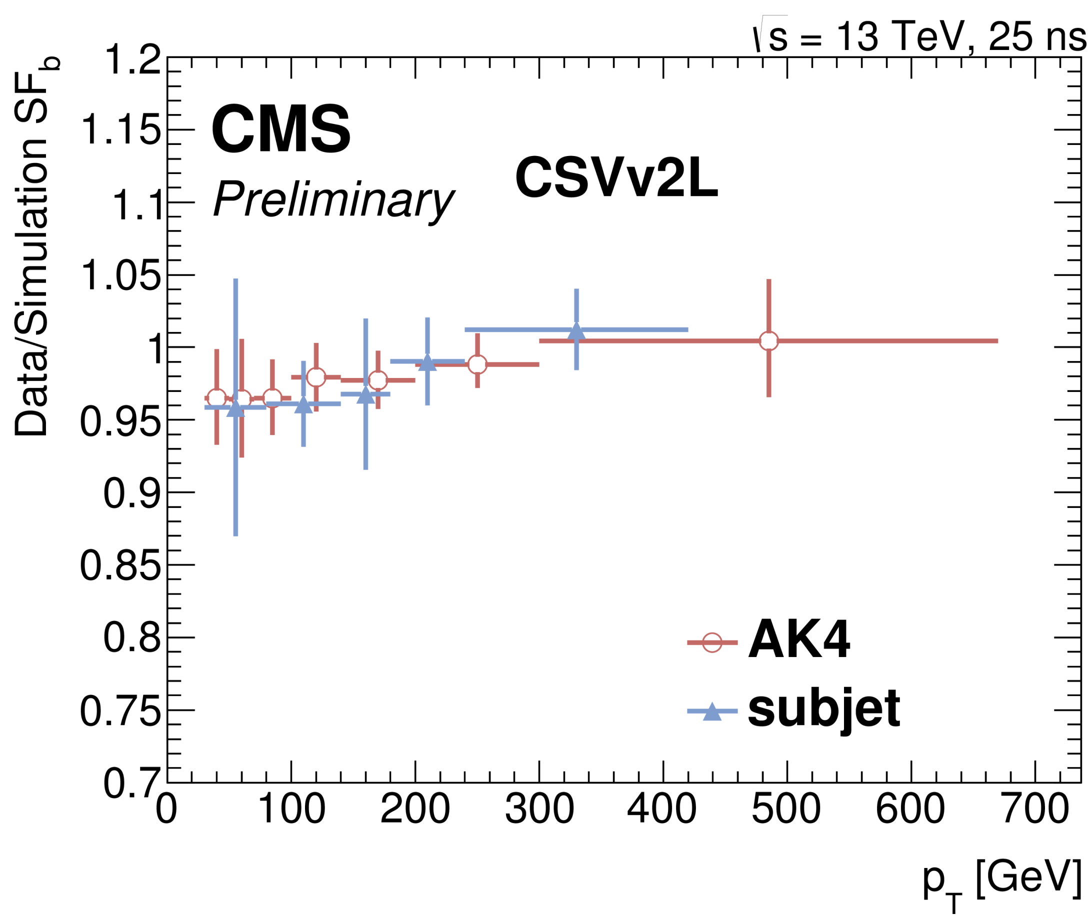
Loose 0.3 (80%), **Medium** 0.6 (70%), **Tight** 0.9 (35%)

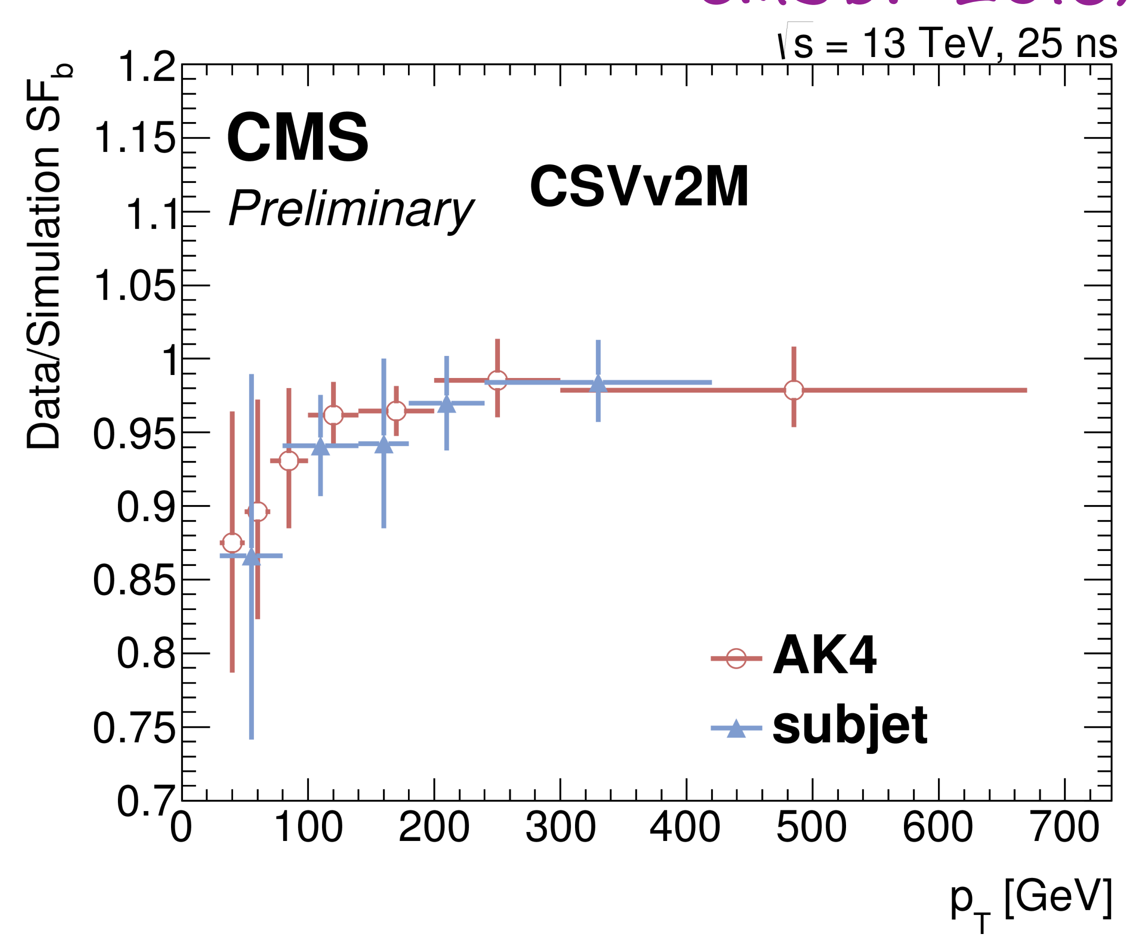
Data/MC close to 1 and 7-25% associated uncertainty per AK8 jet
Statistical uncertainty is dominant
With 2016 dataset it will be reduced

CMS, Sub-jet b-tagging SF

- Validated using **single muon tagged** AK8 jets with p_T > 425 GeV ($\tau 21<0.5$)
  - Sub-jet validation approach doesn't take into account of two b-tagged sub-jet within the same jet topology
  - two close sub-jets from bb resonance can get close
- Good agreement with SFs for the "standard" anti-k_T R=0.4 jets

CMS DP-2016/042





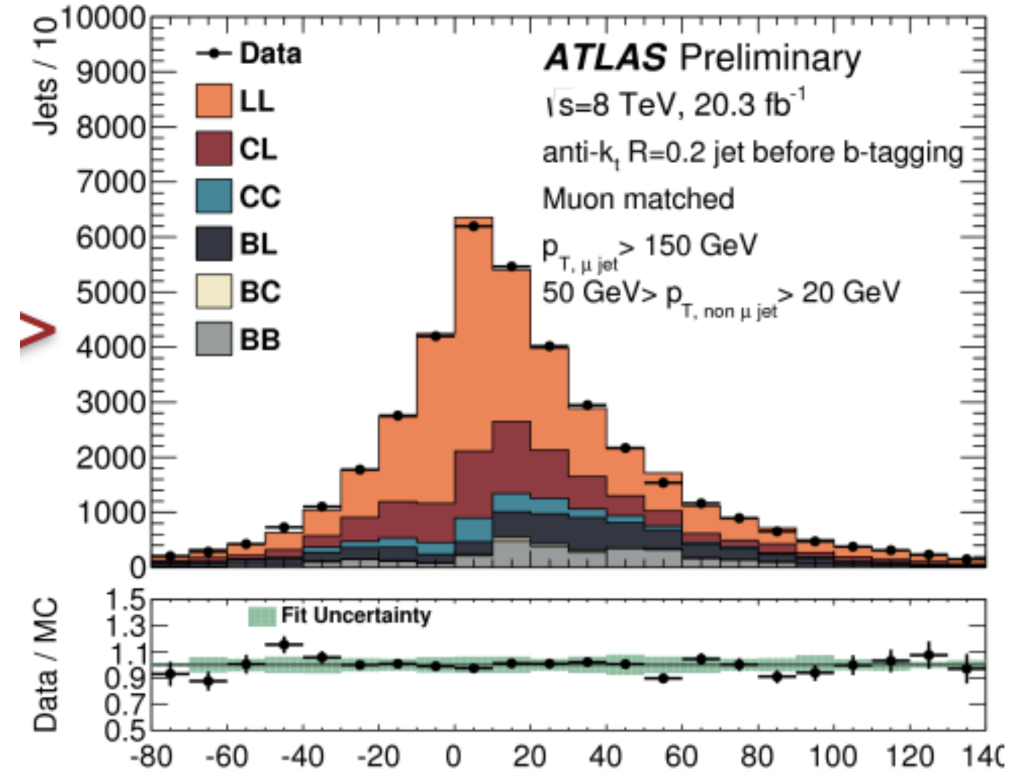
5-15% uncertainty to be applied once per each b-tagged sub-jet, twice for h to $b\bar{b}$

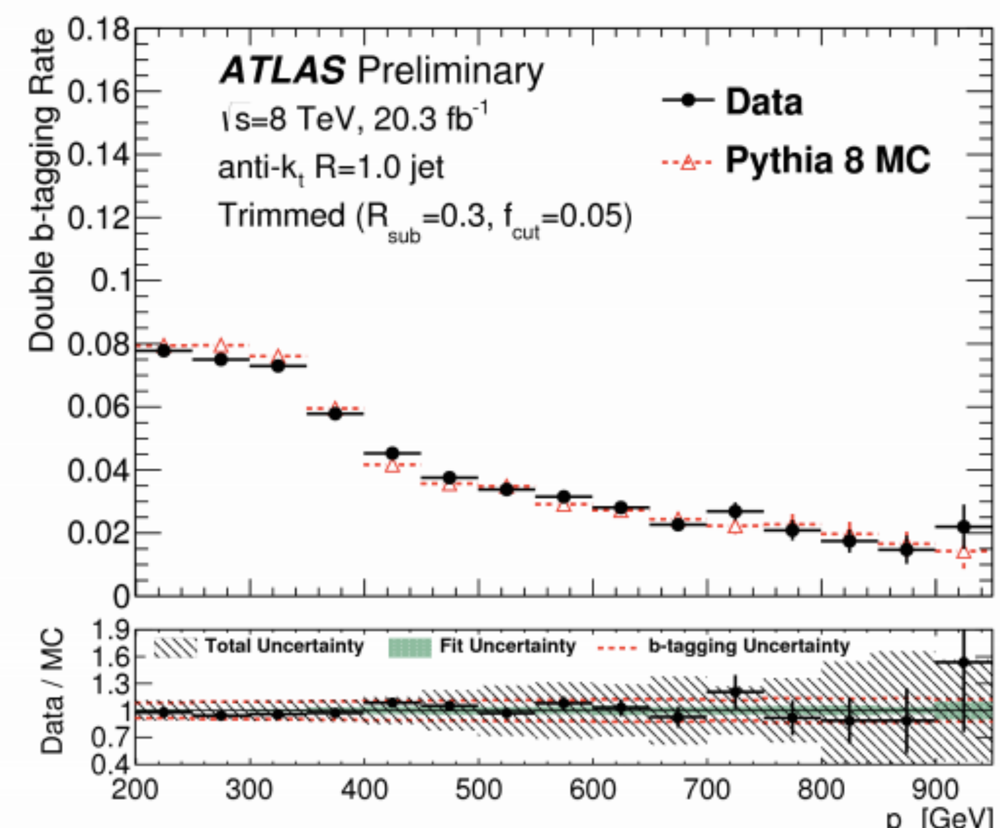
8–30% for h jet in the 2b-category (B2G-16-003)

ATLAS, Validation in data

- g→ b$\bar{b}$ provides close-by b-jets
  - Double b-tagging systematics for track jets
  - Check modeling of large-R jet substructure variables
  - Cross-check large-R jet energy scale (JES) / jet mass scale
- **Strategy**
  - At least one of small radius track jets should be matched to a muon
  - Further double b-tagging on small-R track jets to obtain high purity g→ b$\bar{b}$ samples
  - **Flavor Fraction Correction to MC**
    - MC does not model the heavy flavor content in data (especially at low pt)
    - Fit variable sensitive to flavor composition to data
    - Largest track impact parameter significance inside track jet

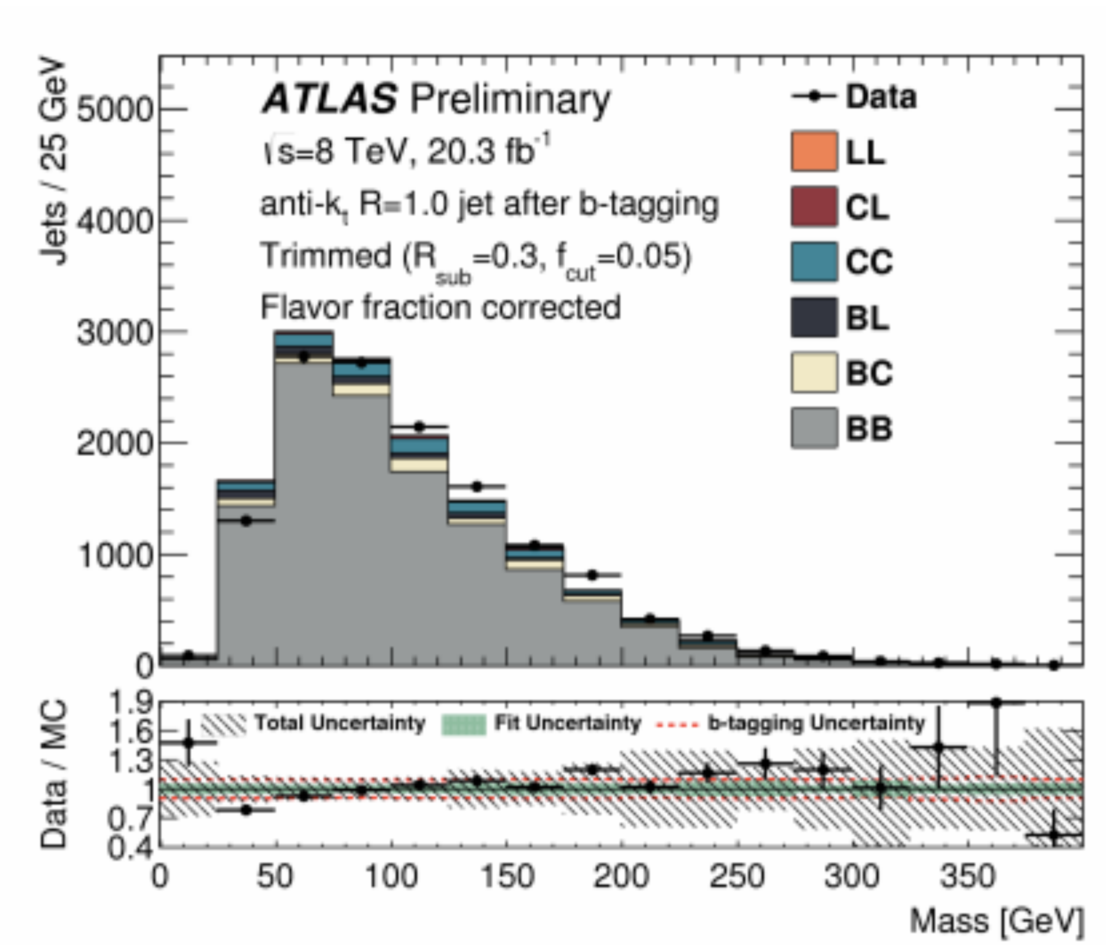
ATLAS-CONF-2016-002



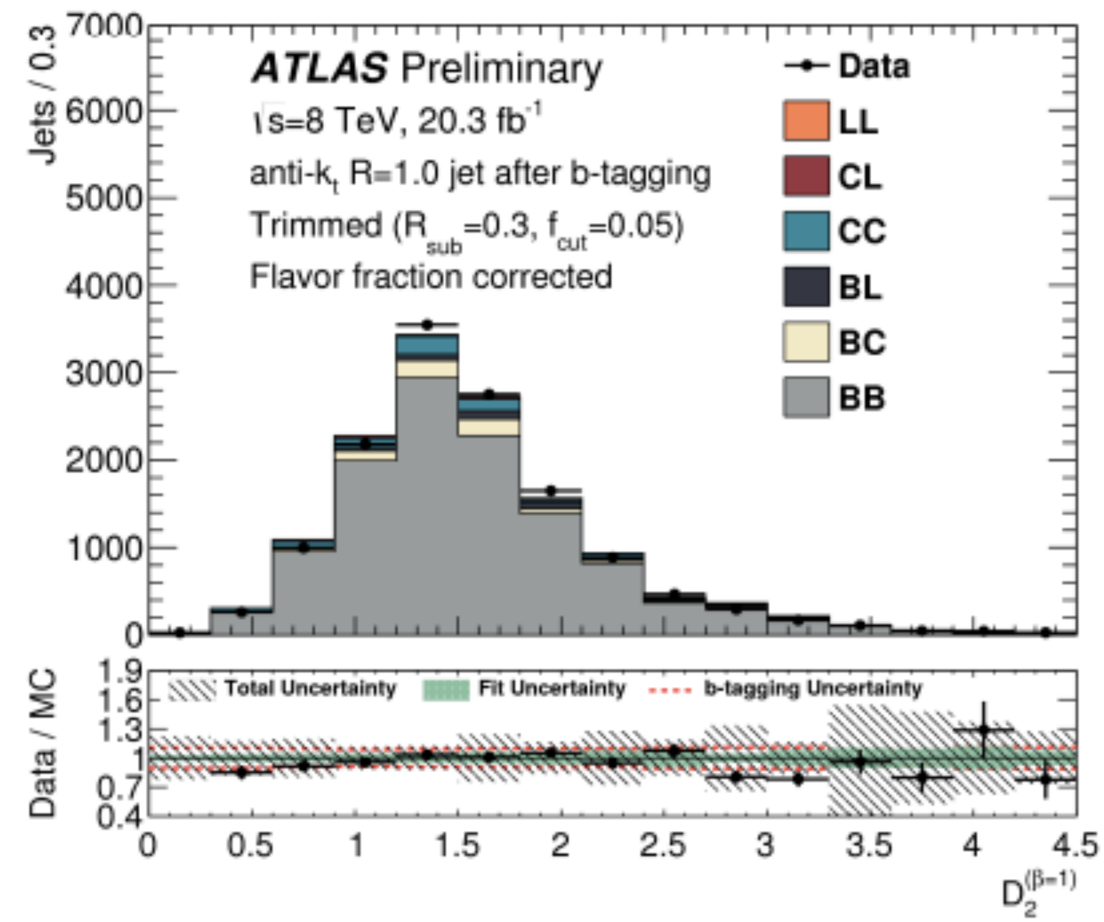


ATLAS, Validation in data

ATLAS-CONF-2016-002



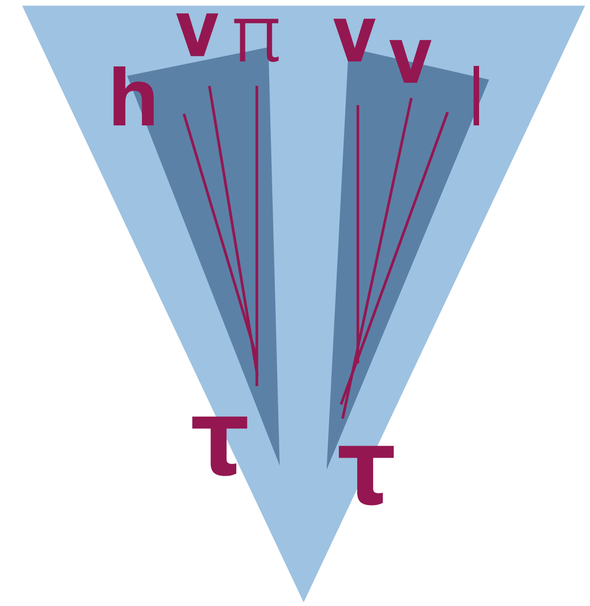
ATLAS-CONF-2016-002



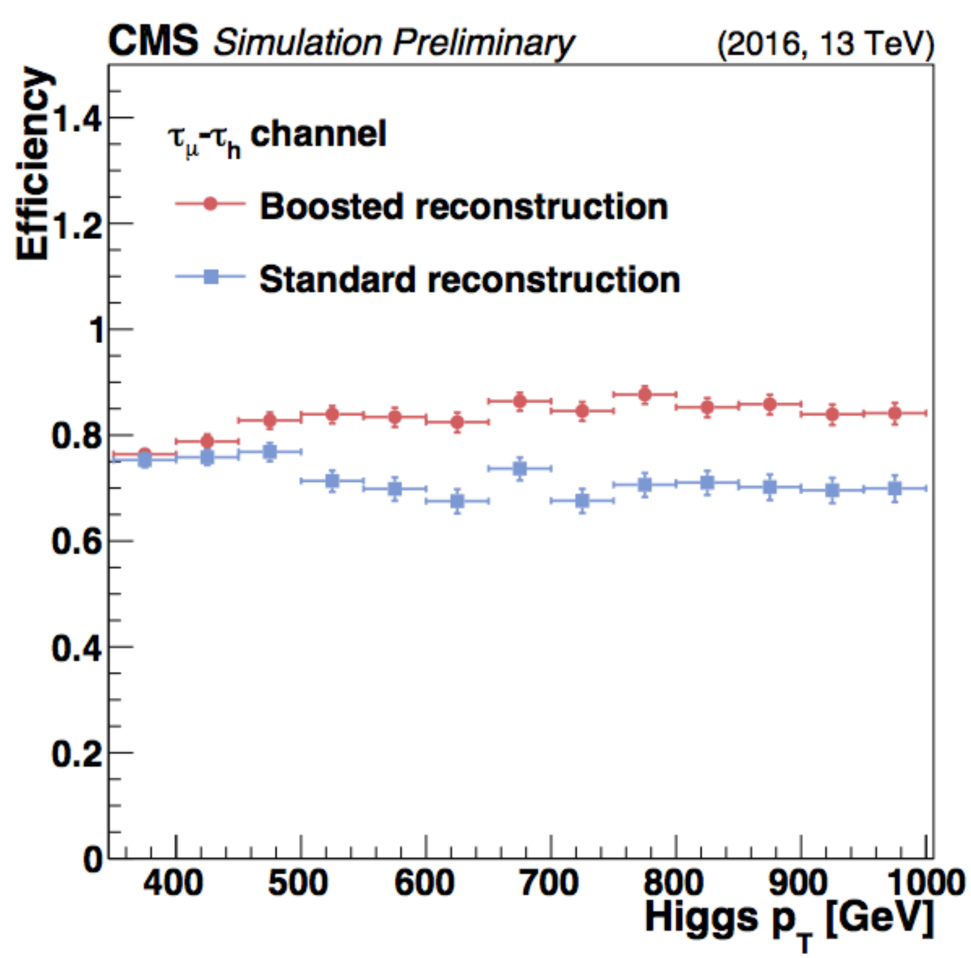
Modeling is in good agreement within uncertainties

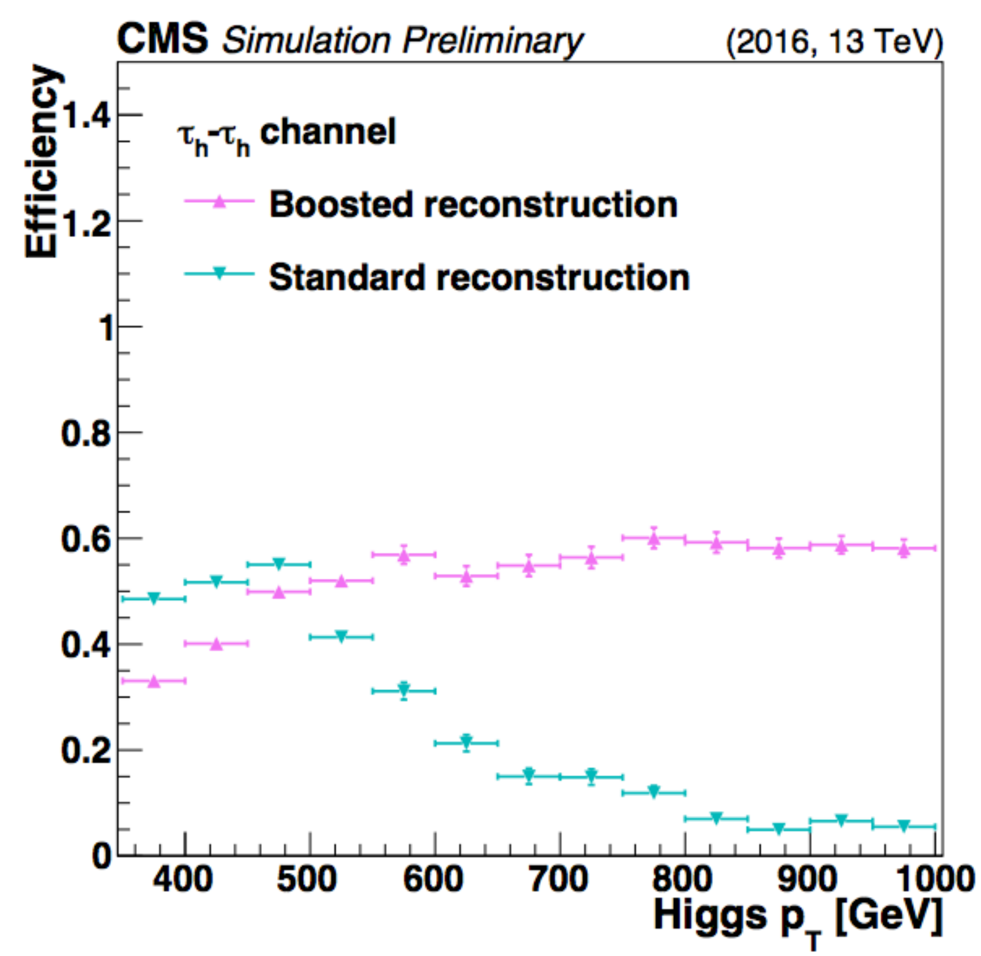
CMS, **h→ττ**

- High-p_T h→ττ reconstruction is quite challenging
- New developments in Run II:
  - Requiring two sub-jets and then reconstruct the τ within each of them



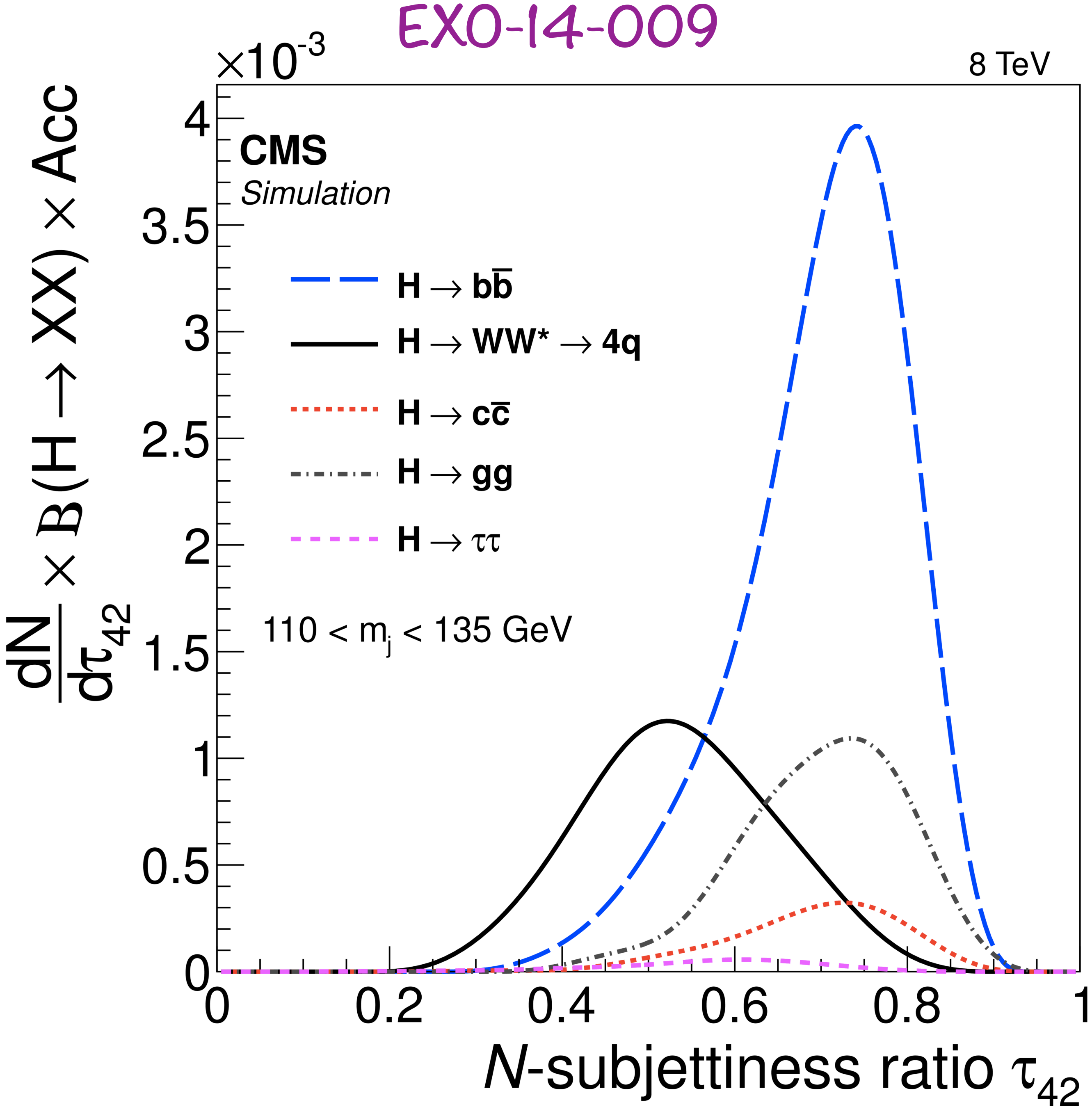
CMS-DP-2016/038





CMS, **h→WW(4q)**

EXO-14-009



- τ42 works best to discriminate between four-pronged h → WW(4q) and QCD jets.
- The τ42 distribution of HWW jets tends to peak around 0.55.
- By contrast, τ42 distributions of multijet background and W/Z jets have a larger fraction of events at large values of τ42

Summary

- $h(b\bar{b})$ as tool to discover new physics
  - produced at high p_T
- Higgs tagging as a collection of criteria to identify boosted jets
  - AK8/AK1.0 jet
  - **pruned/trimmed mass** to reduce pileup and soft radiation effects
  - **τ21/D2** to satisfy the two prongs hypothesis
    - jet p_T and mass dependency fixed with appropriate renormalization
  - **b-tagging** as key element to reduce QCD background
  - **Validation in data** is done using an enriched gluon splitting sample
    - gluon splitting works very well as proxy for the h to $b\bar{b}$ in data
    - CMS: Double muon tagged jets for double-b, while for sub-jet b-tagging only single muon selection is used
    - ATLAS: single muon tagged jet + flavor fraction correction
      - Same data sample used to monitor also substructure and jet mass/p_T

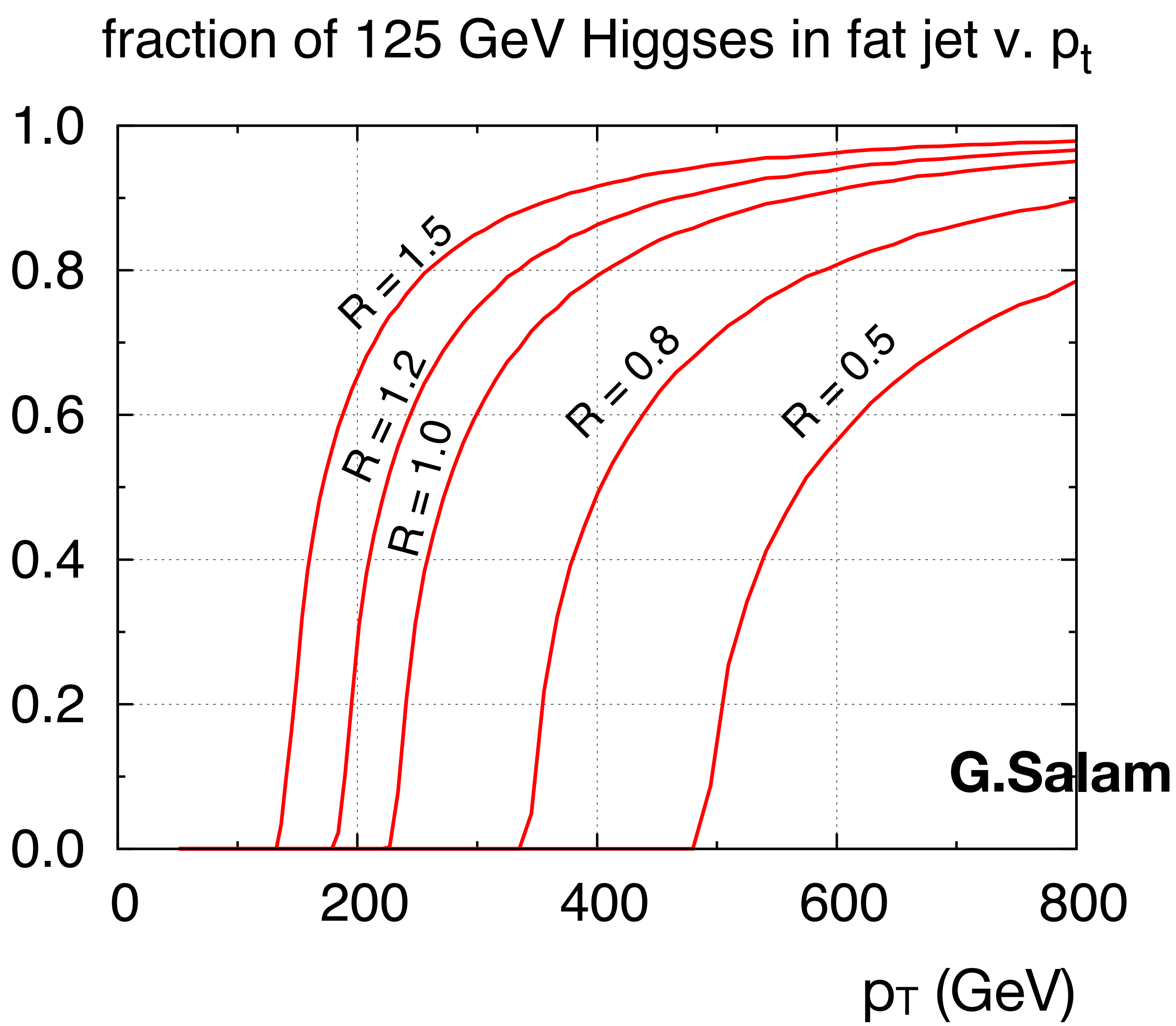
Outlook

- Besides h($b\bar{b}$) also h($\tau\tau$) and h(WW4q) are promising tools for investigating BSM scenarios involving the Higgs boson
  - CMS has made some progress on h($\tau\tau$) reconstruction in boosted topology
  - h(WW4q) is a promising addition
    - already investigated in Run I in CMS
    - With more statistic available also $WW^* \to lqq$

–BACKUP–

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

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



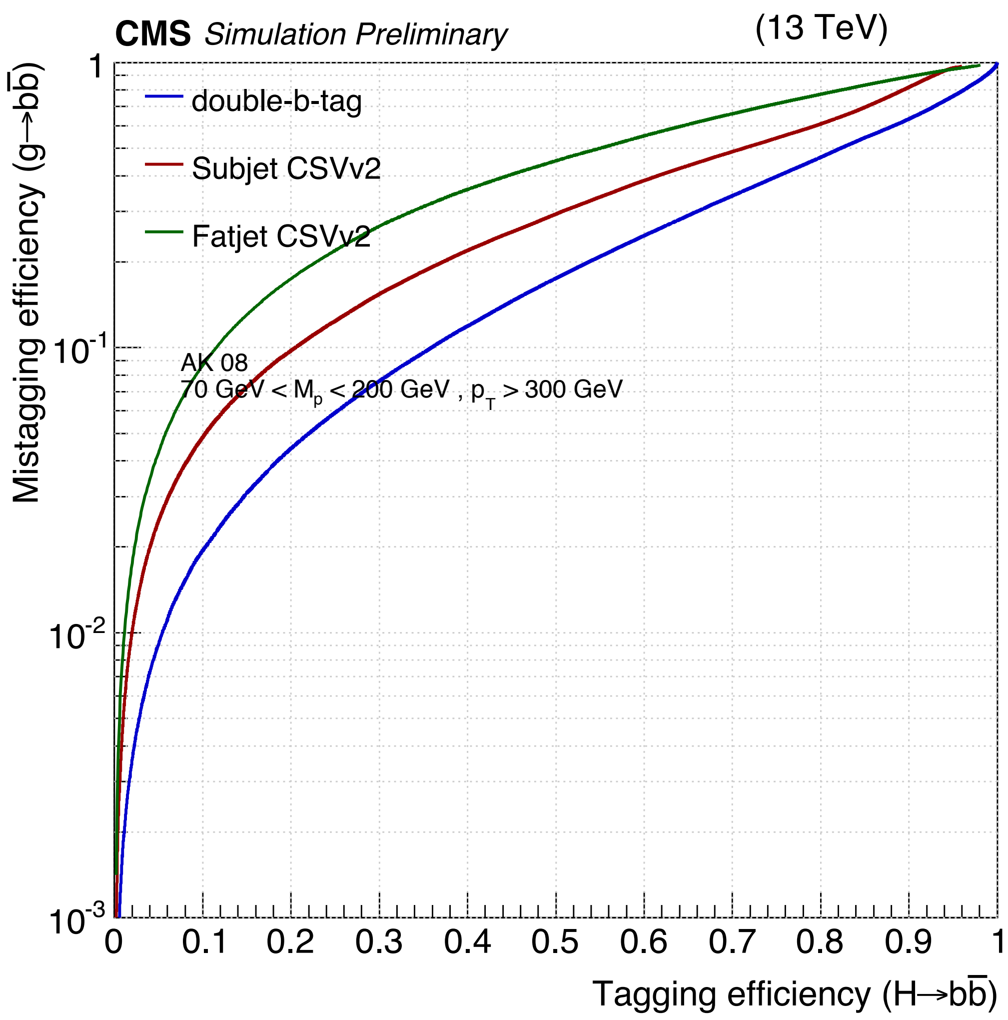
$dR(b\bar{b}) \sim 2m/p_T$

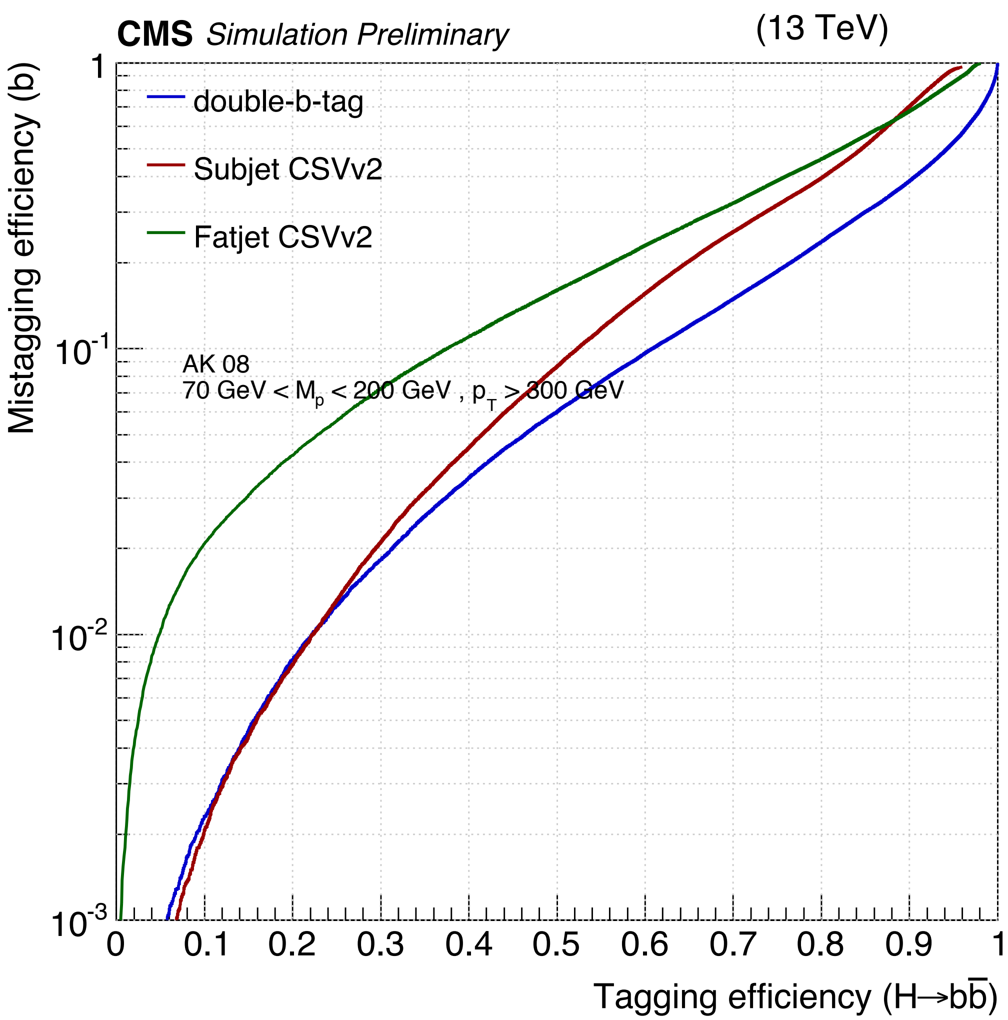
The boosted h($b\bar{b}$) signal is expected to be a **single "fat" jet**

Fully contained in a jet of radius :

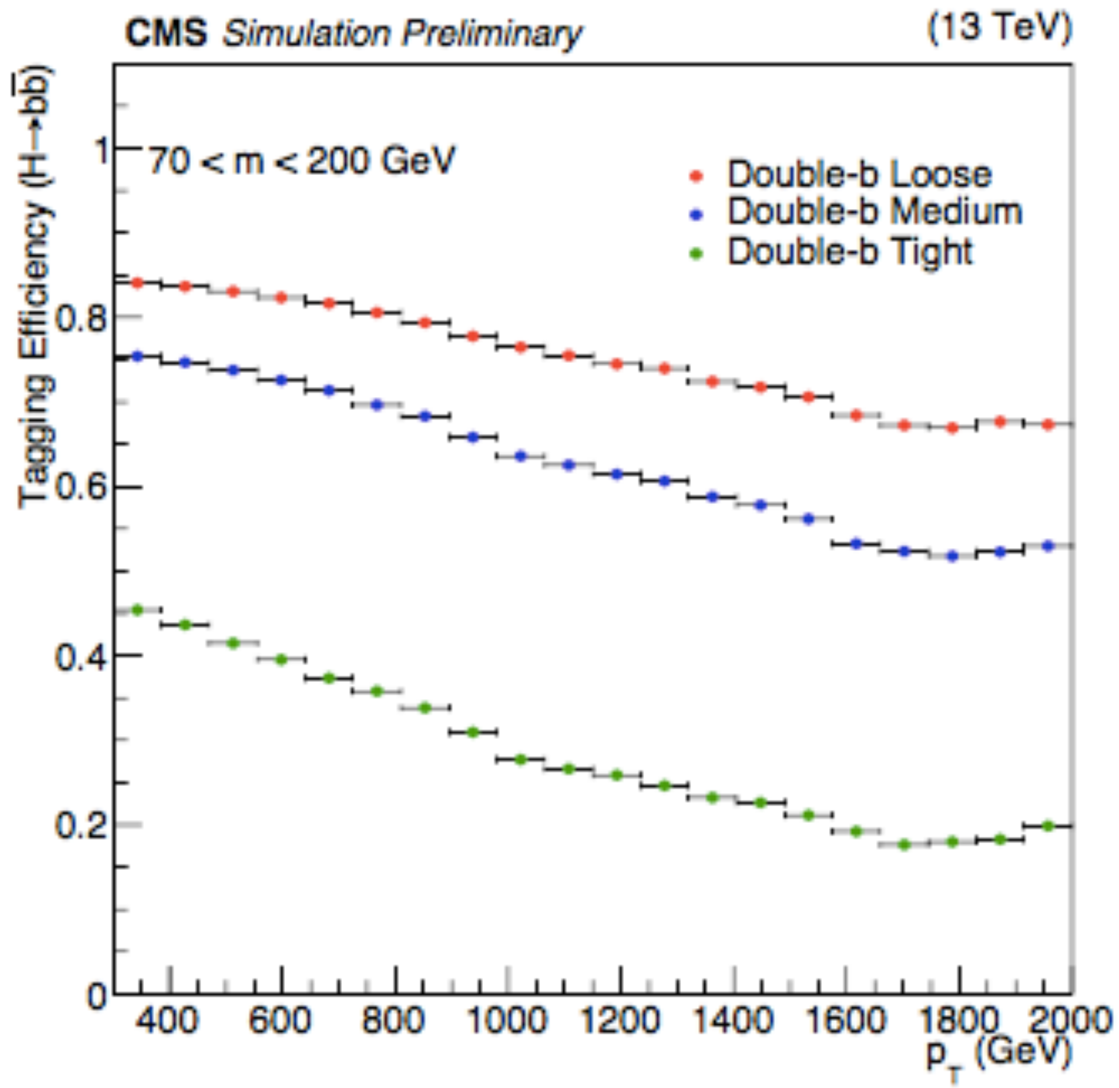
- R =1.5 for H p_T ~ 200 GeV
- R = 0.8 for H p_T ~ 500 GeV

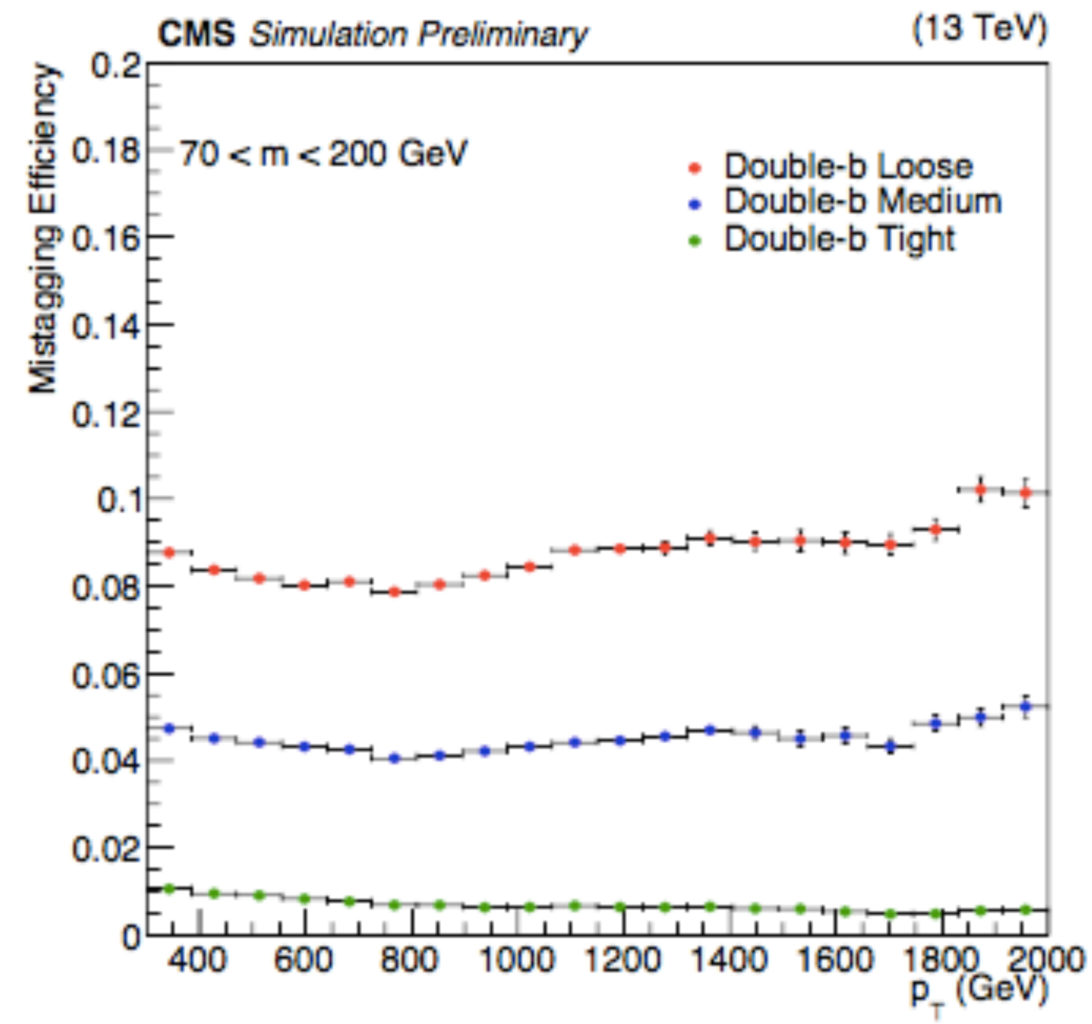
Performance/II





Performance against QCD





Improved performance with respect to both fat and sub-jet b-tagging

At high p_T larger improvement as aimed

Scale Factors

BTV-15-002

DoubleB L

p_T (GeV)	300 - 400	400 - 500	500 - 600	600 - 700
ϵ (Data)	0.791 ± 0.072	0.778 ± 0.095	0.699 ± 0.139	0.663 ± 0.173
ϵ (MC)	0.827 ± 0.009	0.792 ± 0.011	0.771 ± 0.009	0.685 ± 0.008
SF	0.956 ± 0.088	0.983 ± 0.121	0.906 ± 0.181	0.969 ± 0.253

DoubleB M

p_T (GeV)	300 - 400	400 - 500	500 - 600	600 - 700
ϵ (Data)	0.692 ± 0.069	0.694 ± 0.089	0.607 ± 0.128	0.584 ± 0.120
ϵ (MC)	0.748 ± 0.010	0.698 ± 0.012	0.647 ± 0.013	0.553 ± 0.010
SF	0.926 ± 0.093	0.995 ± 0.129	0.939 ± 0.198	1.055 ± 0.218

DoubleB T

p_T (GeV)	300 - 400	400 - 500
ϵ (Data)	0.425 ± 0.045	0.358 ± 0.055
ϵ (MC)	0.469 ± 0.011	0.392 ± 0.013
SF	0.905 ± 0.099	0.913 ± 0.142