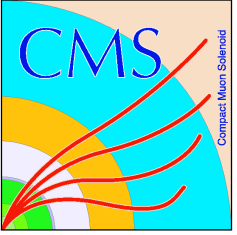


Boosted Higgs, b tagging and other tools/techniques (Part 2)

Dinko Ferenček
Rutgers, The State University of New Jersey

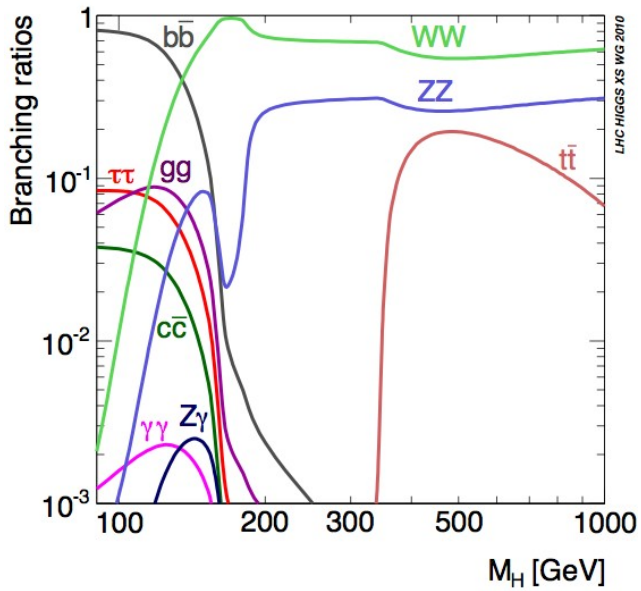
BSM Higgs Workshop @ LPC
November 3–5, 2014
Fermi National Accelerator Laboratory
Batavia, IL, USA

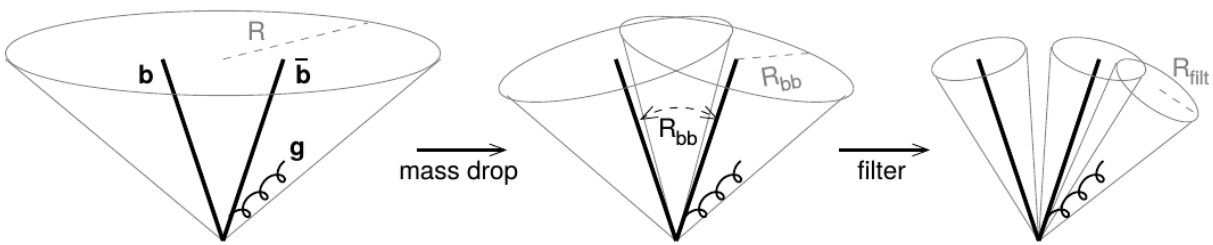
Outline

- Boosted Higgs and jet substructure
- b tagging in boosted topologies
- Calibration
- Higgs tagger
- Run 2 challenges
- Summary and outlook

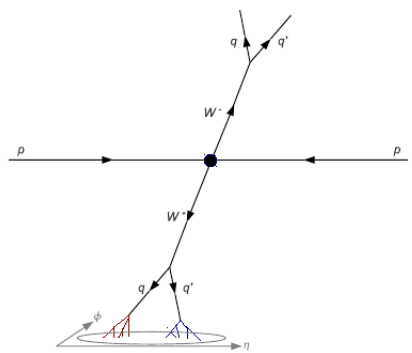
Boosted Higgs and jet substructure

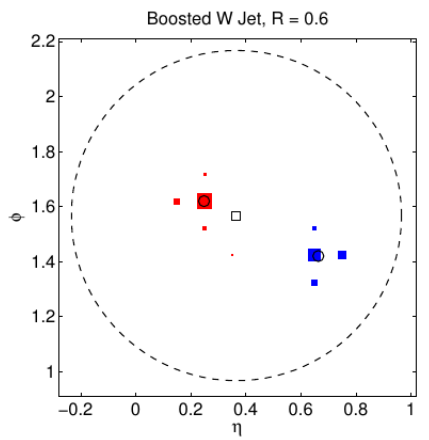
- Recently discovered boson with m≈125 GeV consistent with predictions for the Standard Model Higgs boson
- Dominant decay mode H→b$\bar{b}$ (Br(H→ b$\bar{b}$)≈57% [1])
- Since the BDRS paper (arXiv:0802.2470) proposing to use boosted H→b$\bar{b}$ decays, various jet substructure tools and techniques have been proposed (see Part 1)





- Two-prong decay, in many respects similar to boosted hadronically decaying W/Z bosons
  → Can rely on well established 2-prong tagging algorithms to tag boosted H→b$\bar{b}$ decays

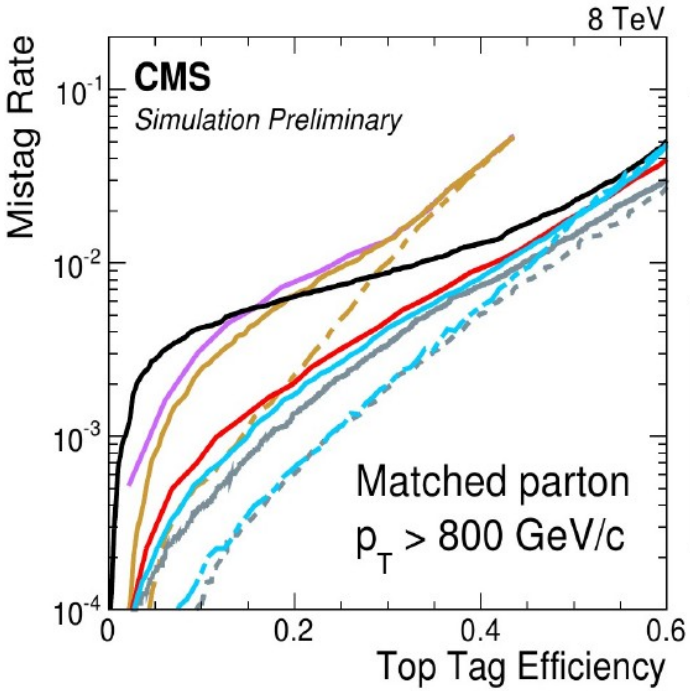


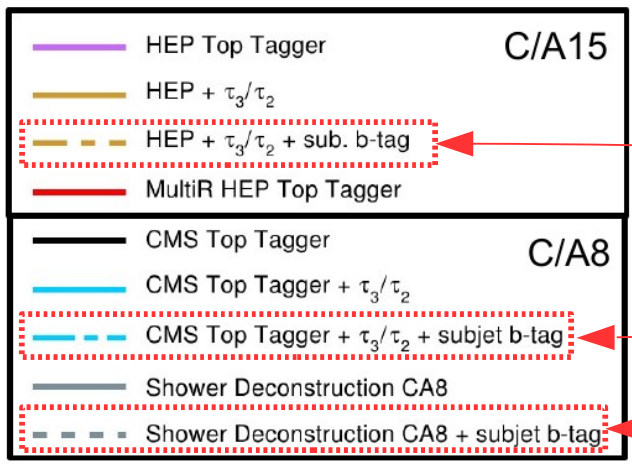


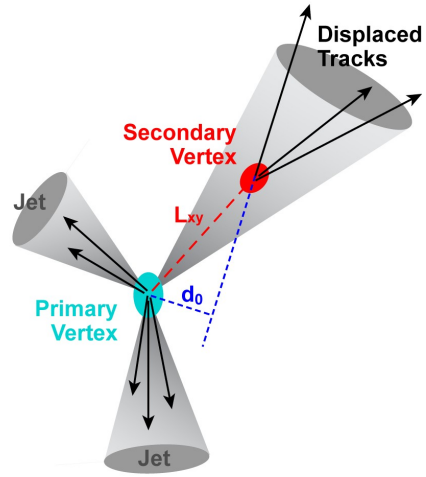
[1] https://twiki.cern.ch/twiki/bin/view/LHCPhysics/CERNYellowReportPageBR2

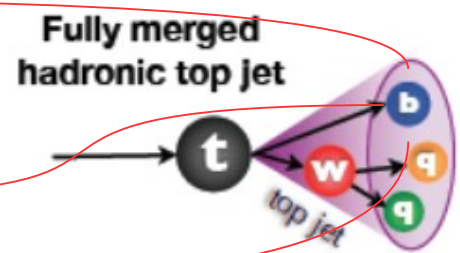
Boosted Higgs and jet substructure (cont'd)

- Distinct feature in this case is b hadrons and their long lifetime
  → Displaced tracks and secondary vertices
- More traditional 2-prong tagging algorithms do not explicitly exploit this information
- Example of top tagging algorithms:



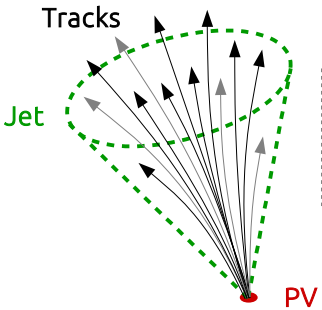




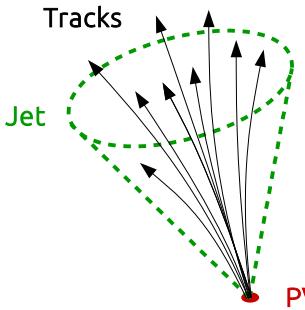


- b tagging, being largely complementary to 2-prong tagging, could significantly improve the sensitivity of a dedicated Higgs tagging algorithm

General b-tagging workflow



CMS: fixed-size association cone ΔR(track,jet)<0.3
ATLAS: shrinking-cone association



Jet-track association

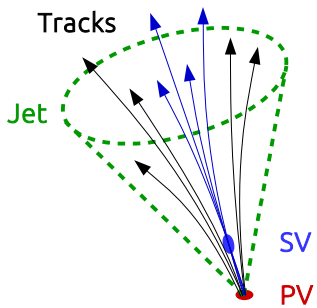
Track selection

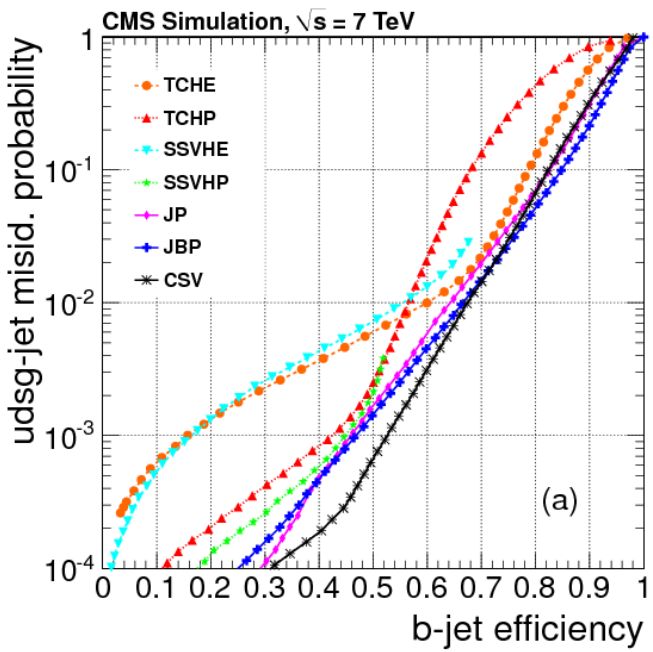
Track-based tagging algorithms

Combined tagging algorithms

Secondary vertex reconstruction

SV-based tagging algorithms





Operating points for CMS taggers:

L = loose (≈10% light-flavor mistag rate)

M = medium (≈1% light-flavor mistag rate)

T = tight (≈0.1% light-flavor mistag rate)

Boosted b tagging

- Possible b-tagging strategies:
  - Fat jet b tagging
  - Subjet b tagging
  - b tagging of standard (R=0.4) jets and matching them to fat jets (using some ΔR requirement)
  - b tagging of smaller-size jets and matching them to fat jets and/or subjets

Boosted b tagging: CMS

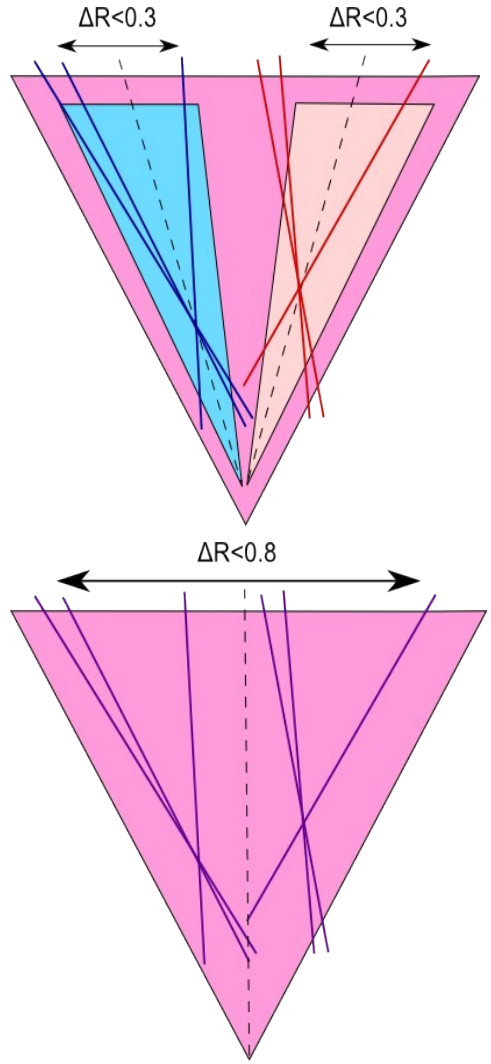
- Using Combined Secondary Vertex (CSV) algorithm
- b-tagging scenarios considered:

Subjet b tagging

- Standard CSV applied to pruned subjets of Higgs candidate fat jets
- Standard jet-track association $\Delta R<0.3$
- No dedicated algorithm retraining performed

Fat jet b tagging

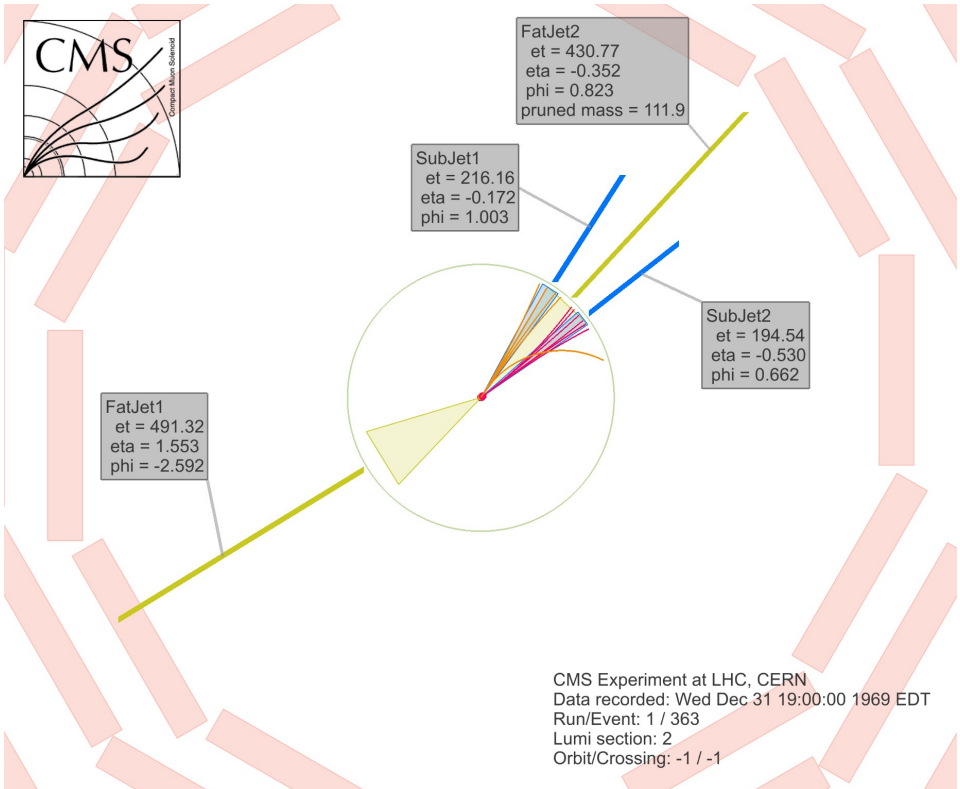
- Standard CSV applied to Higgs candidate fat jets
- Extended jet-track association $\Delta R<R_{jet}$ (0.8 or 1.2)
- No dedicated algorithm retraining performed

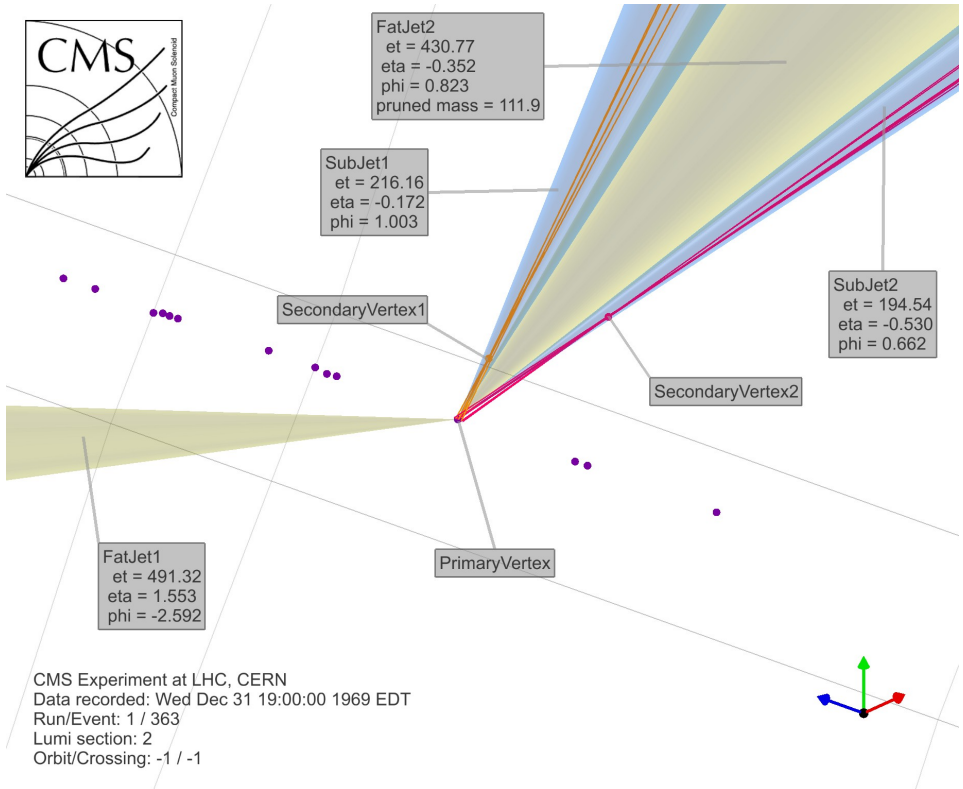


For more information: CMS-PAS-BTV-13-001
https://twiki.cern.ch/twiki/bin/view/CMSPublic/BoostedBTaggingPlots2014

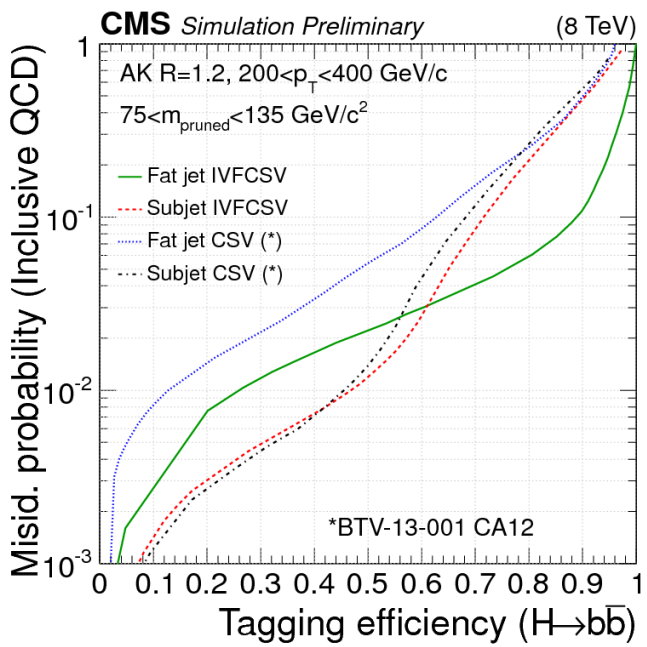
Subjet b tagging

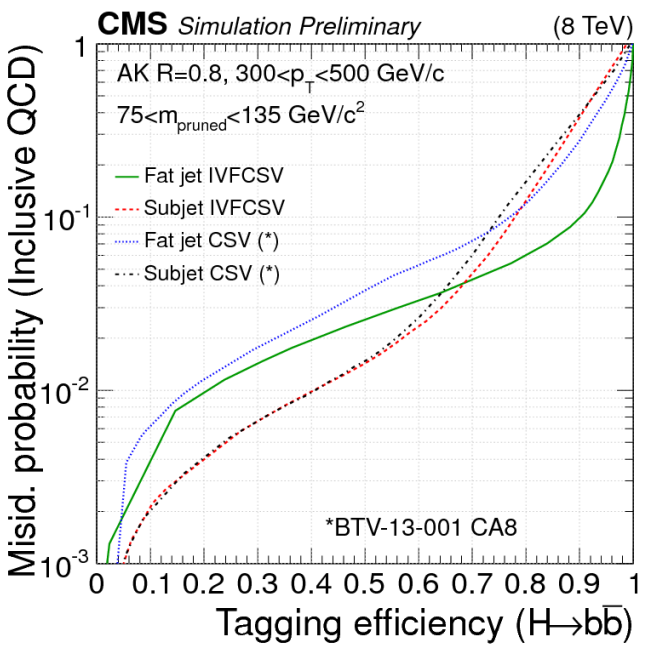
- Boosted $H \rightarrow b\bar{b}$ (simulation)

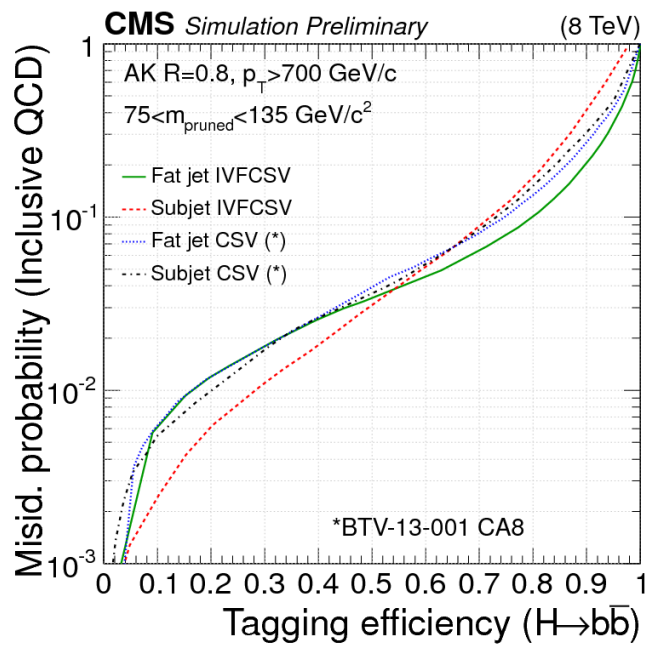




Boosted H→$b\bar{b}$ (inclusive QCD as background)



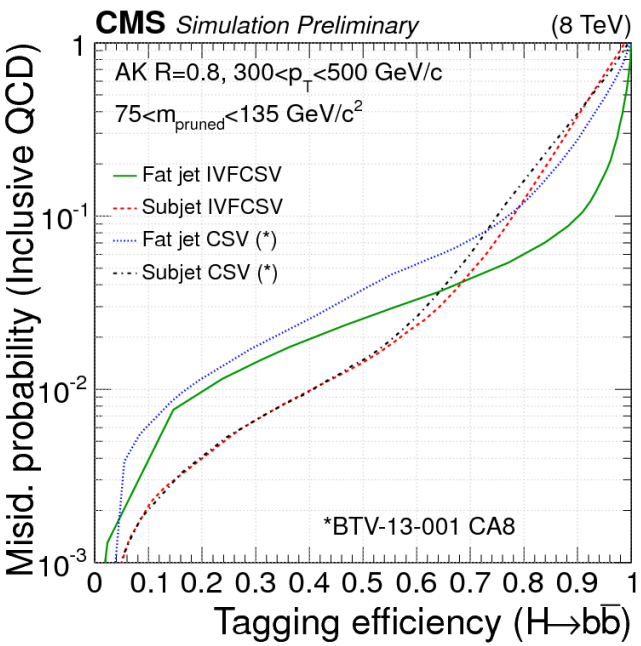


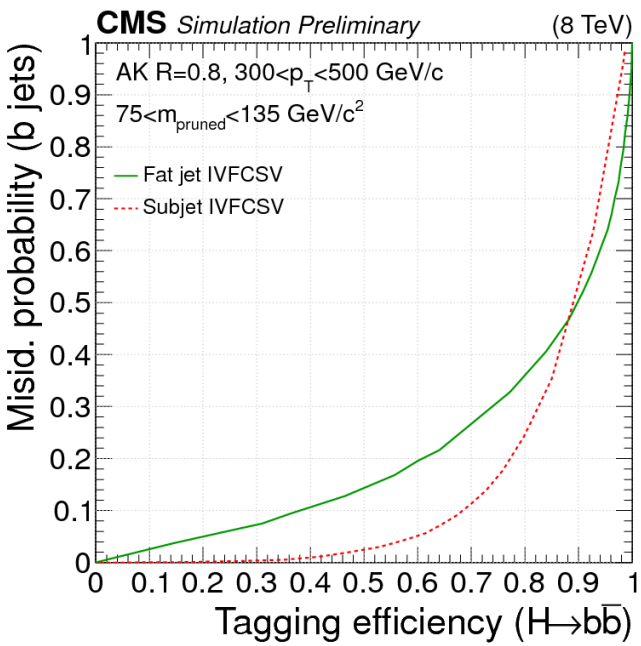


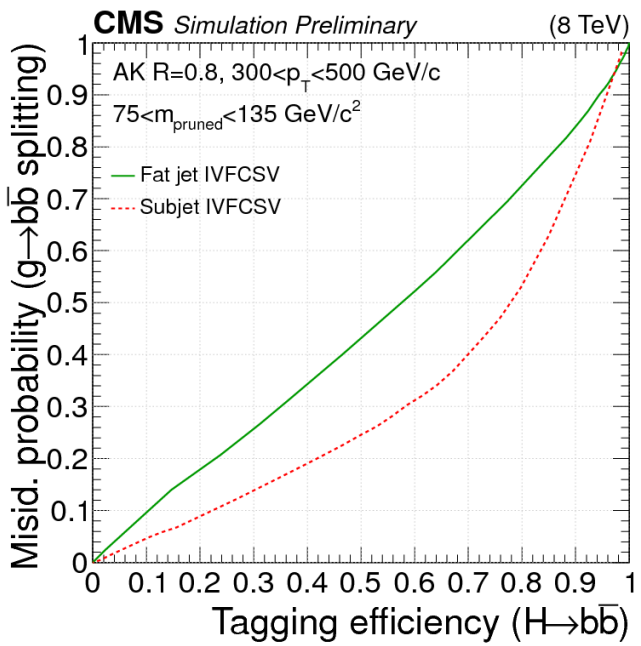
Subjet tagging efficiency refers to tagging both subjets

- AK R=0.8 or 1.2 (depending on the p_T range) fat jets and pruned subjets (z_{cut}=0.1 and R_{cut}=0.5), IVFCSV includes Run 2 developments (see backup)
- Improved CSV algorithm with IVF vertices performs better than the older generation CSV
  - Older CSV algorithm applied to CA jets but the choice of the clustering algorithm found to have negligible impact on the b-tagging performance
- Subjet and fat jet performance curves cross each other with fat jet b tagging performing better at high tagging efficiencies

Boosted H→$b\bar{b}$ (different backgrounds)







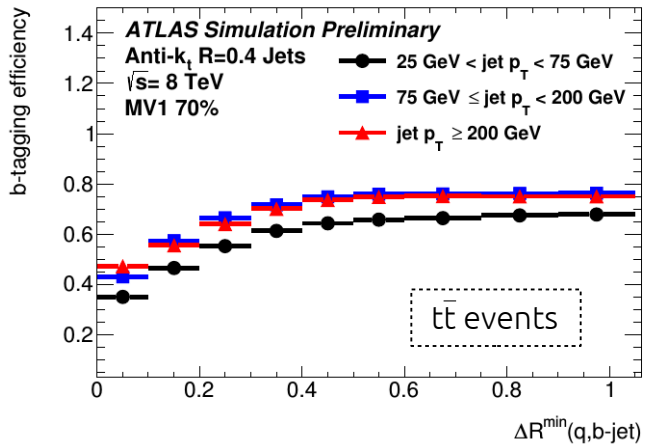
- Some level of complementarity between fat jet and subjet b tagging present (depends on the background composition)
- Dedicated re-training expected to improve the performance

Boosted b tagging: ATLAS

- Just like CMS, using b-tagging algorithms that combine displaced track and secondary vertex information
- b-tagging scenarios considered:

b tagging standard (R=0.4) jets

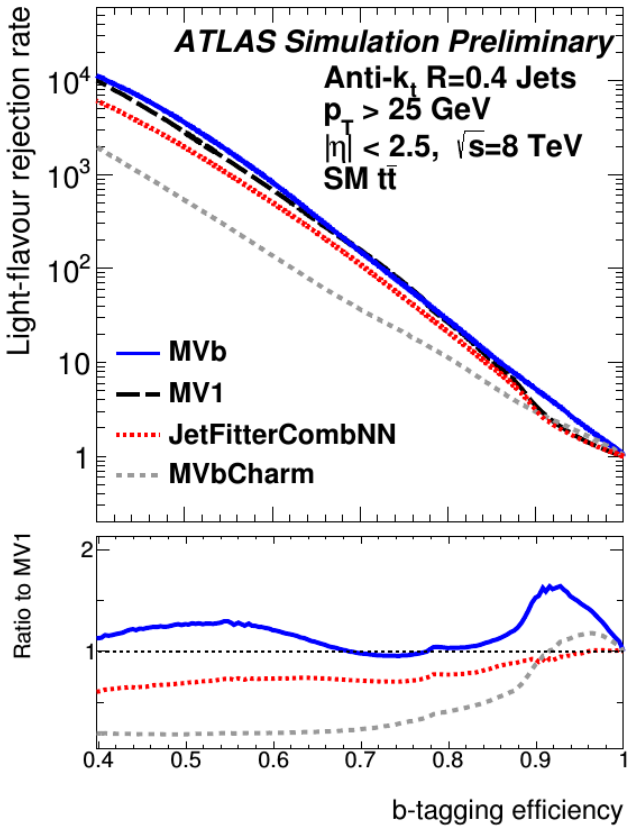
  - Performance of the standard MV1 algorithm degrades in dense environments
  - MV1 extended (by incorporating additional more robust variables) and retrained (MVb and MVbCharm)

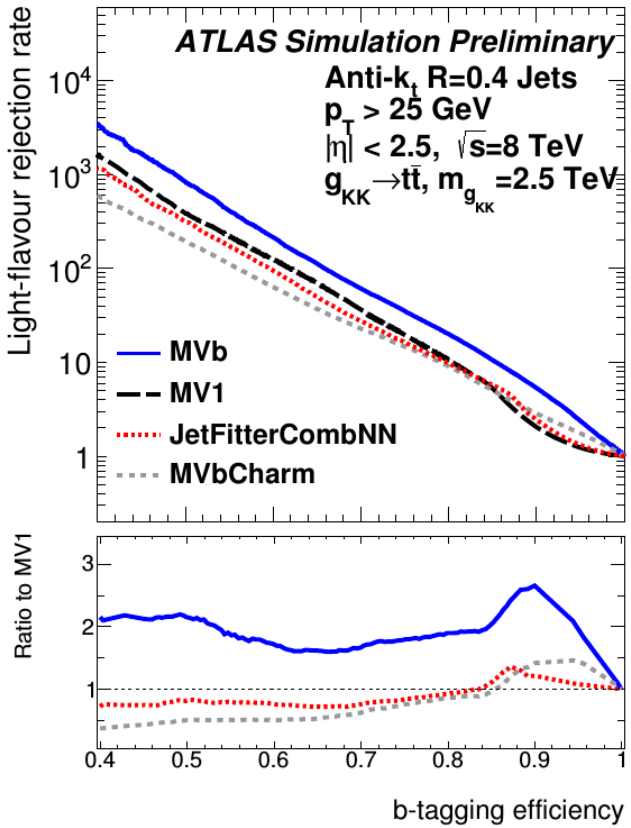


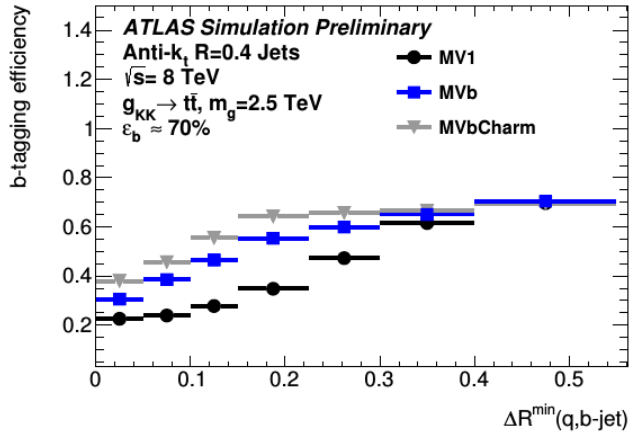
- **b tagging smaller-size track jets**
  - Using standard MV1 applied to smaller-size track jets
  - Track jets associated to fat jet and/or subjets using “ghost” clustering procedure
  - No dedicated algorithm retraining performed

For more information: ATL-PHYS-PUB-2014-014
ATL-PHYS-PUB-2014-013

b tagging of standard (R=0.4) jets

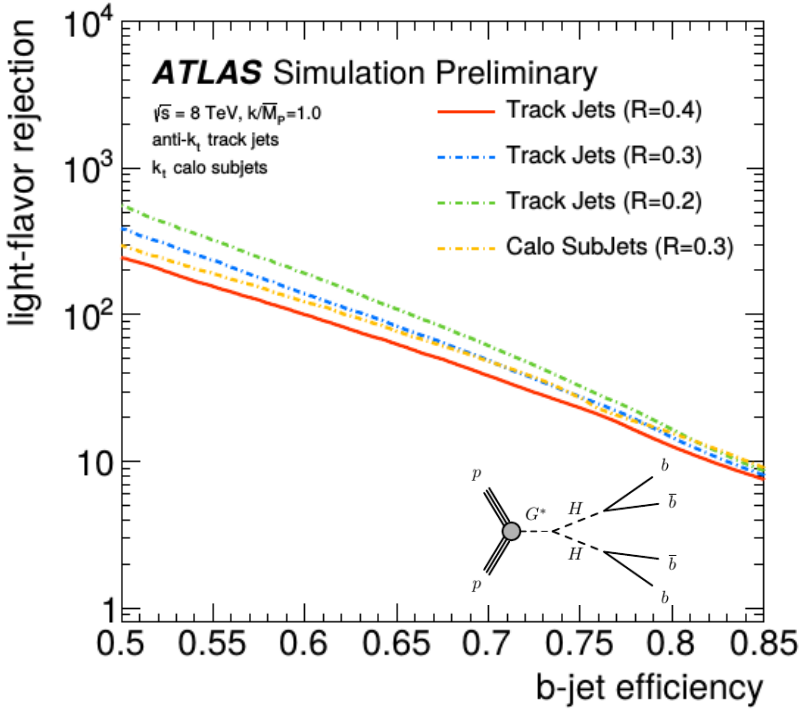


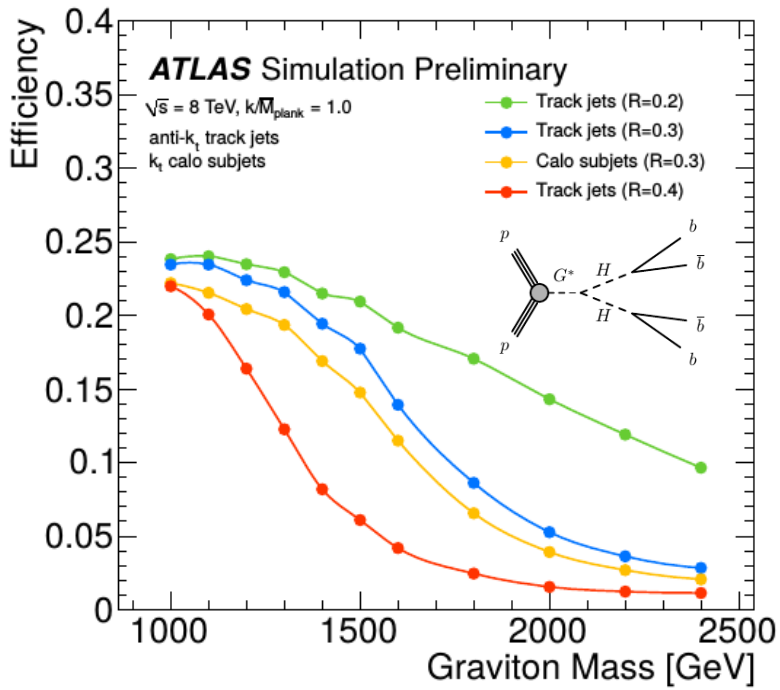




Dedicated and re-trained algorithm perform better than the standard one (improvement up to 160%)

b tagging of smaller-size track jets





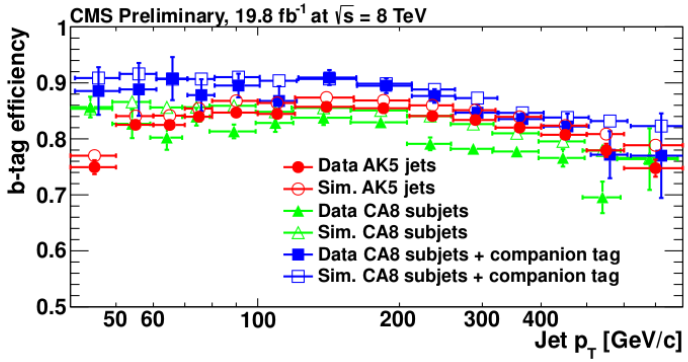
- b-tagged jets defined independently of calorimeter objects
- Very flexible, can be associated to any calorimeter-based object (only one data/MC calibration needed)
- Can better resolve individual subjets than standard (R=0.4) jets

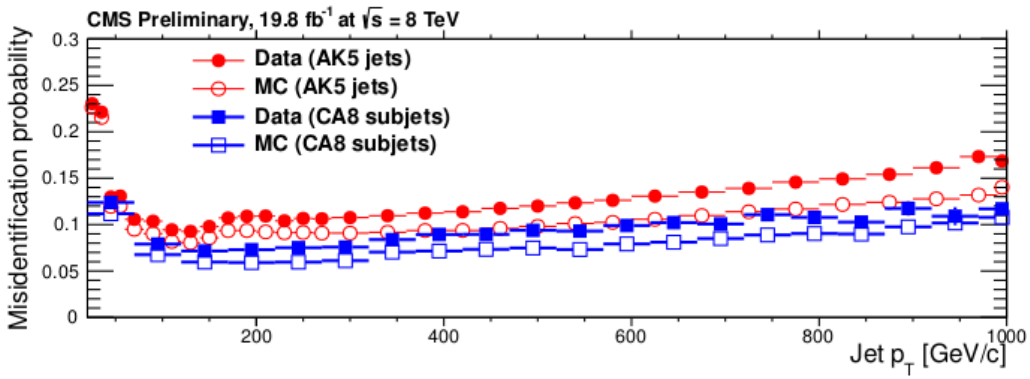
Calibration

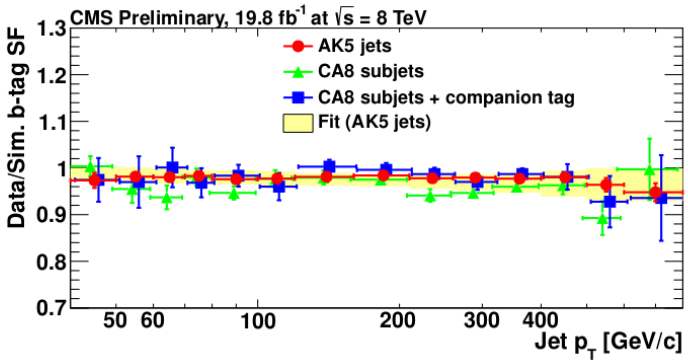
- Simulation does not perfectly reproduce b-tagging performance in data → Scale factors derived and applied to simulated events

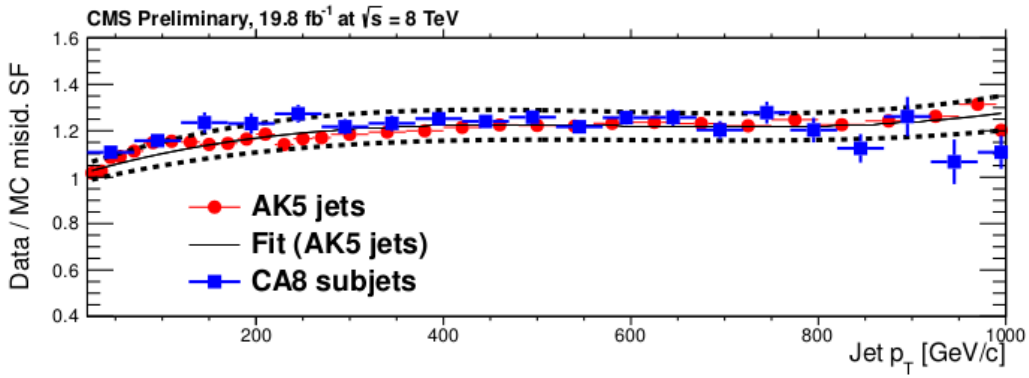
$$\mathrm{SF} = \frac{\varepsilon_{\mathrm{DATA}}}{\varepsilon_{\mathrm{MC}}}$$

- Subjet b-tagging efficiency measured in a sample enriched in gluon splitting jets, likely to contain two b hadrons inside a single fat jet







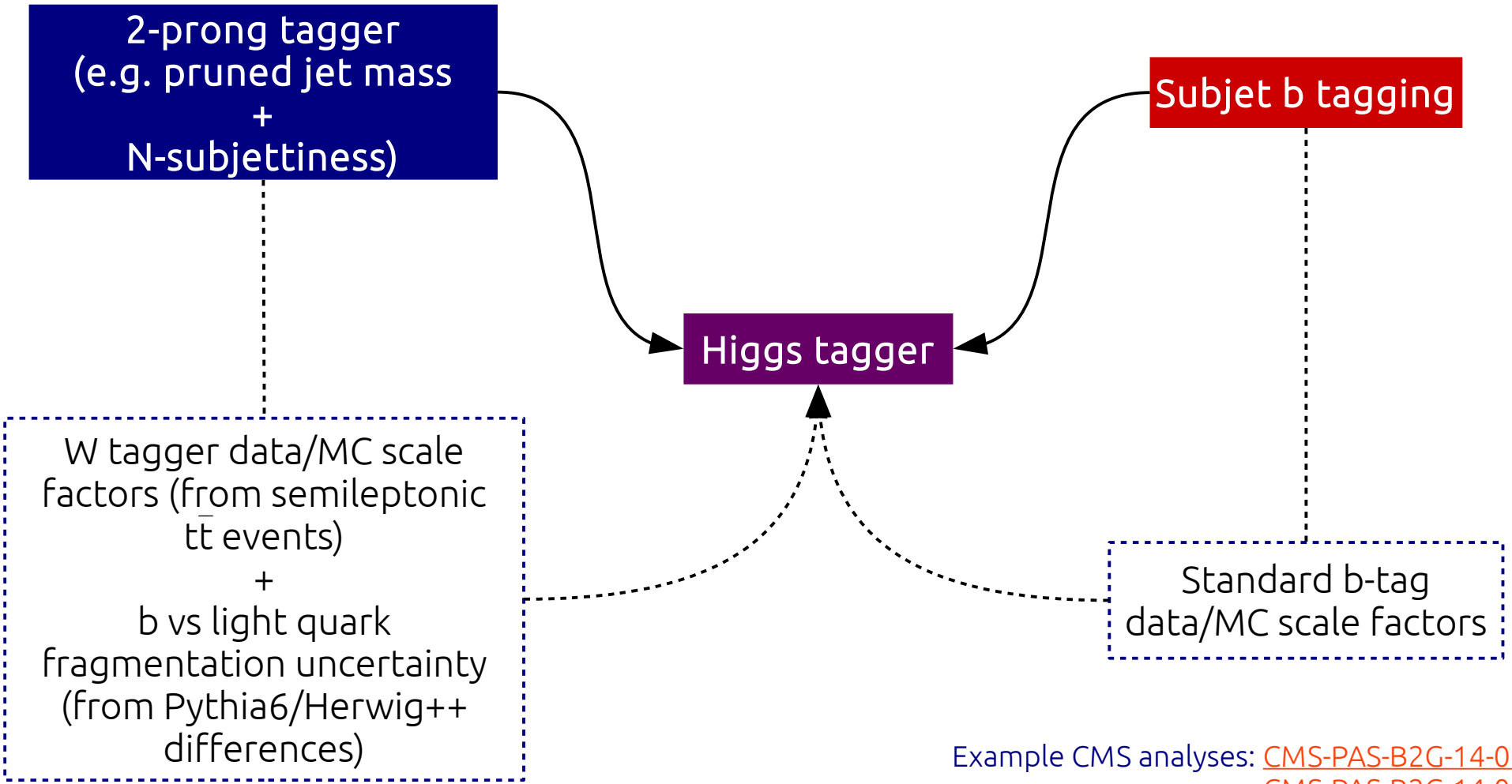


Calibration (cont'd)

- Scale factors measured on subjets in good agreement with those measured on the standard (R=0.5) jets
  - Benefiting from the fact that the same setup is being used for both subjets and standard jets
  - Even though not fully optimal, using the same setup facilitated commissioning studies and early adoption in physics analyses
- CMS analyses recommended to use the same scale factors for the standard jets and subjets
  - **Caveat:** When subjets get close to each other ($\Delta R<0.4$), analyses recommended to switch to standard b-tagging applied to fat jets (to avoid dealing with correlated subjet tags caused by shared tracks) → Addressed by Run 2 developments (see backup)
- ATLAS working on calibrating both of their boosted b tagging approaches using Run 1 data

Putting it all together: Higgs tagger

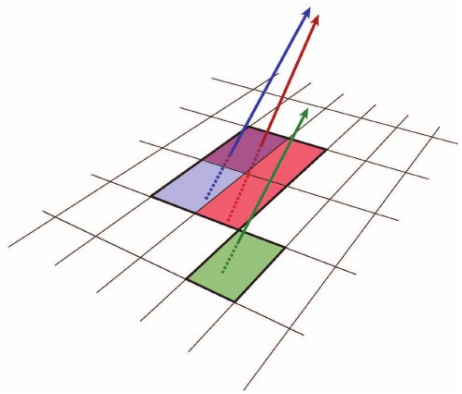
- Example of CMS Run 1 Higgs tagger:

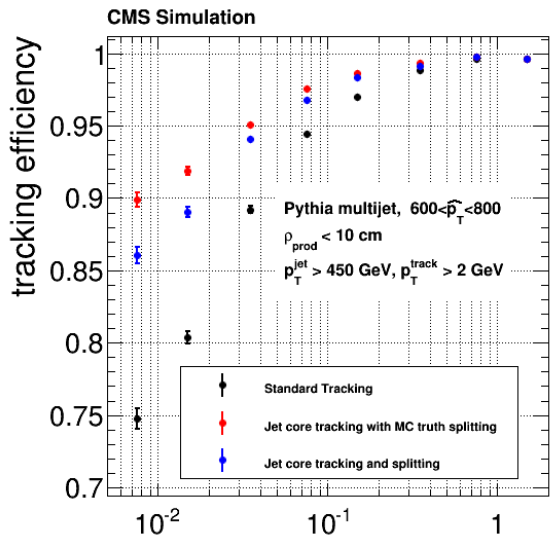


Example CMS analyses: CMS-PAS-B2G-14-001
CMS-PAS-B2G-14-002

Run 2 challenges

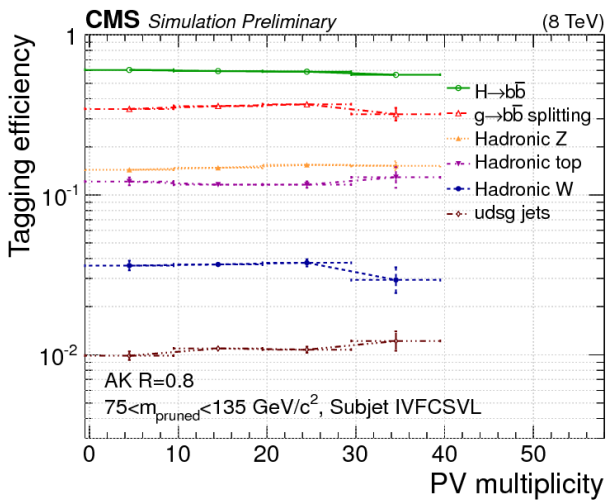
- More energy → More high-p_T jets
  - Dense environment in the core of high-p_T jet leads to overlapping tracks and merged pixel clusters → Challenge for track reconstruction
  - ATLAS and CMS have developed cluster splitting algorithms → Improved jet substructure and b tagging at high p_T

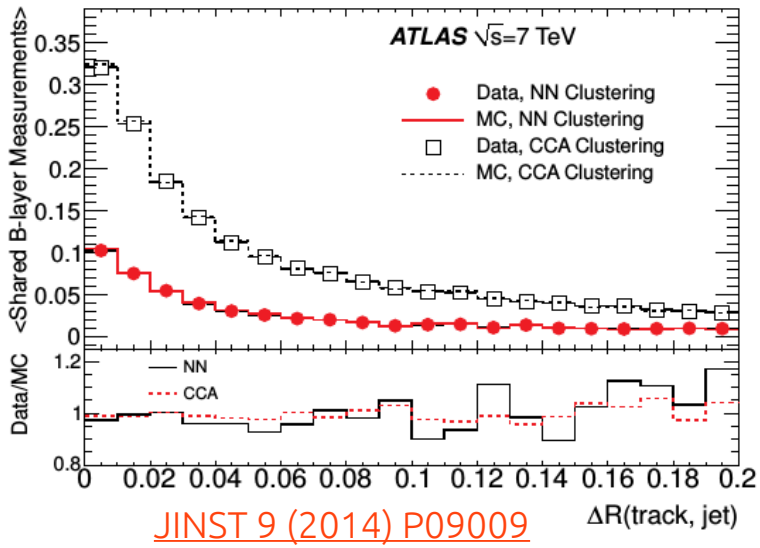




TWiki: CMSPublic/HighPtTrackingDP

- Higher pileup
  - Performance stable up to ~50 PU events





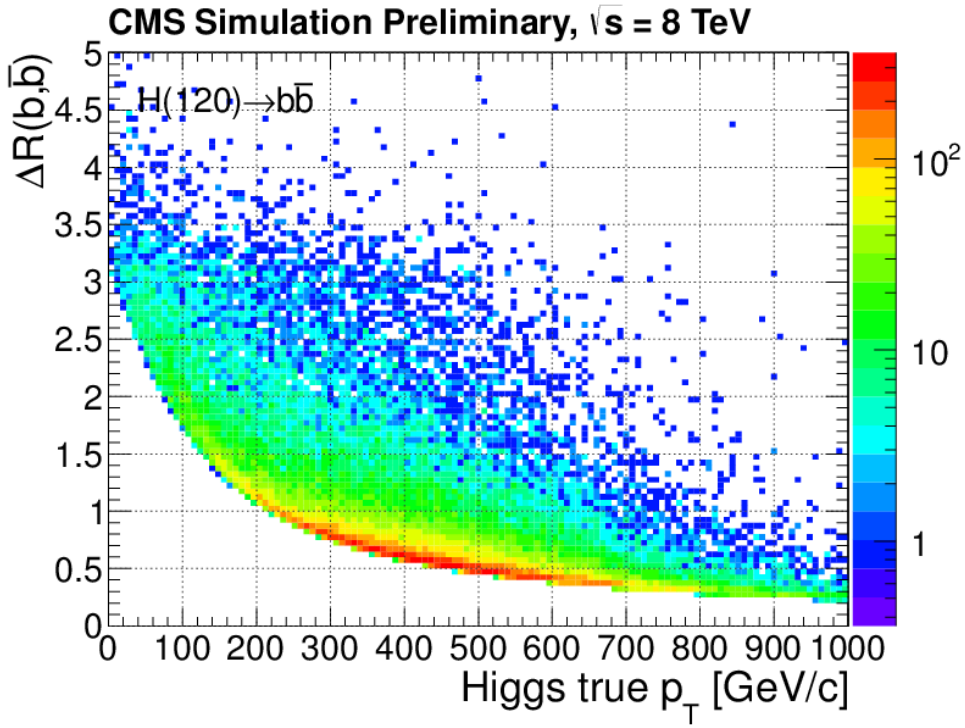
JINST 9 (2014) P09009

Summary and outlook

- Several complementary strategies for b tagging in boosted topologies studied in Run 1
- Subjet b tagging successfully commissioned and being applied in Run 1 analyses
- Further performance improvements possible from dedicated algorithm developments and re-training
- Strategies to deal with Run 2 challenges being developed
- Higgs tagging firmly established and added to our jet toolbox
- Looking forward to Run 2 data

Backup Slides

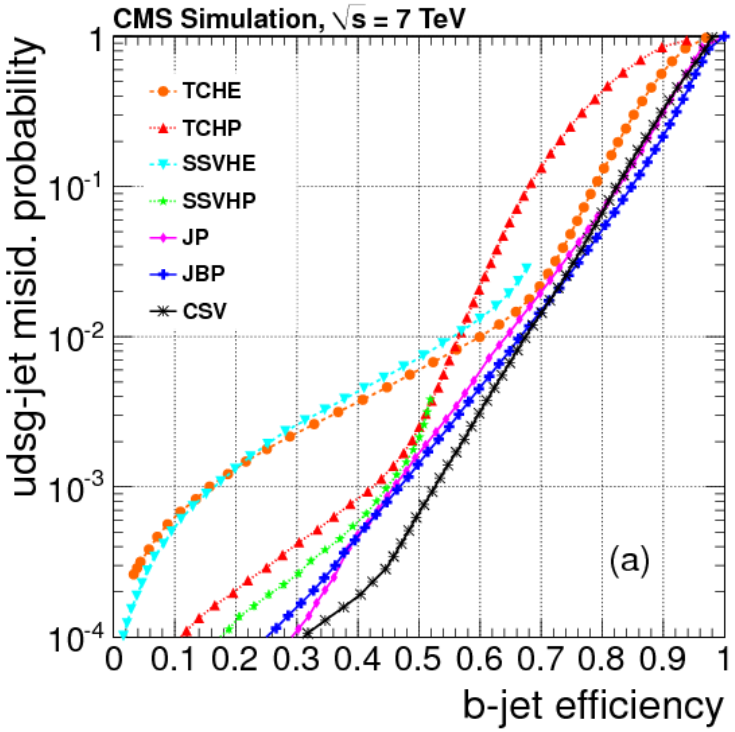
Boosted H→bb̄ decays



$$\Delta R(\mathrm{b},\overline{\mathrm{b}}) \gtrsim \frac{2m}{p_{\mathrm{T}}}$$

CMS b-tagging algorithms

Tagging Algorithm	Operating points	Supported at 7 TeV	Supported at 8 TeV
Track Counting High Efficiency	TCHEL	✔	✘
	TCHEM	✔	✘
	TCHET	✘	✘
Track Counting High Purity	TCHPL	✘	✘
	TCHPM	✔	✘
	TCHPT	✔	✔
Jet Probability	JPL	✔	✔
	JPM	✔	✔
	JPT	✔	✔
Jet B Probability	JBPL	✔	✘
	JBPM	✔	✘
	JBPT	✔	✘
Simple Secondary Vertex High Efficiency	SSVHEM	✔	✘
	SSVHET	✘	✘
Simple Secondary Vertex High Purity	SSVHPT	✔	✘
Combined Secondary Vertex	CSVL	✔	✔
	CSVM	✔	✔
	CSVT	✔	✔



From JINST 8 (2013) P04013

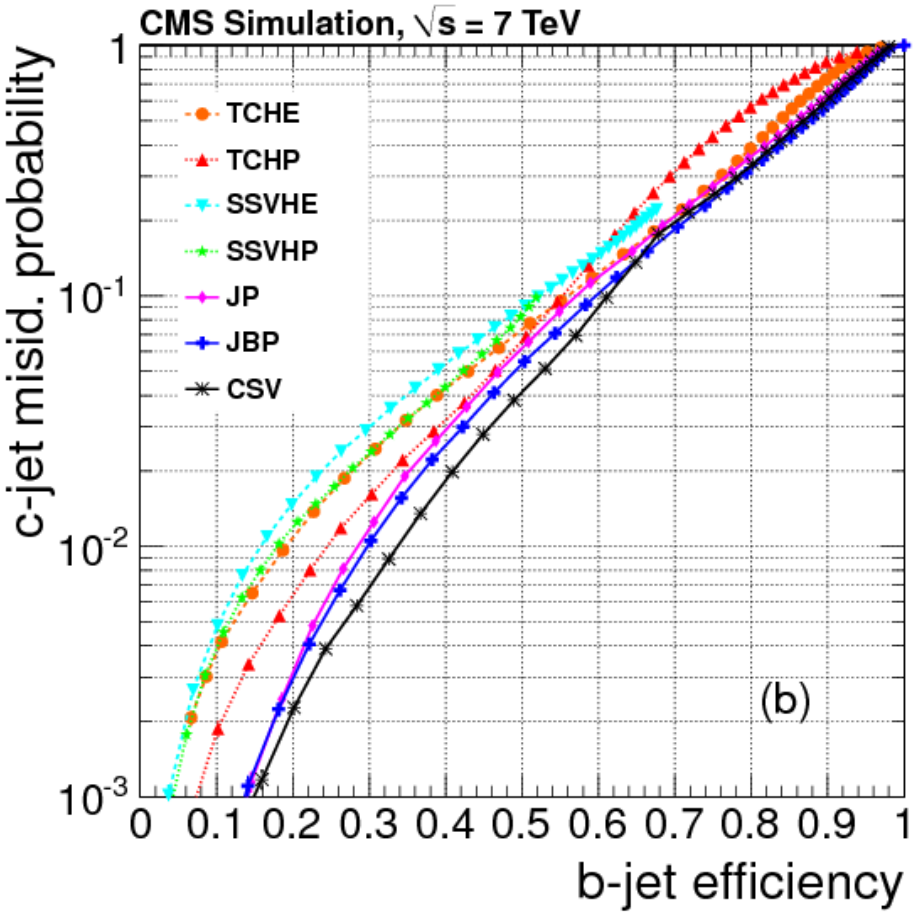
Tagger operating points:

L = loose (≈10% light-flavor mistag rate)

M = medium (≈1% light-flavor mistag rate)

T = tight (≈0.1% light-flavor mistag rate)

CMS b-tagging algorithms (cont'd)



From JINST 8 (2013) P04013

CSV algorithms

Legacy CSV:

- Likelihood-ratio-based discriminator
- Based on the variables listed below

Variable	Vertex category		
	RecoVertex	PseudoVertex	NoVertex
trackSip3dSig	✔	✔	✔
trackSip2dSigAboveCharm	✔	✔	✘
trackEtaRel	✔	✔	✘
vertexMass	✔	✔	✘
vertexNTracks	✔	✔	✘
vertexEnergyRatio	✔	✔	✘
flightDistance2dSig	✔	✘	✘

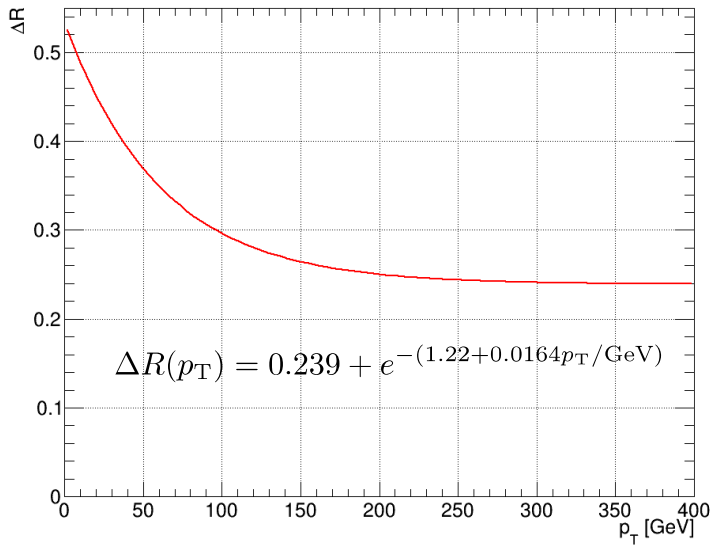
CSVv2:

- MLP-based discriminator
- Based on the variables listed below

Variable	Vertex category		
	RecoVertex	PseudoVertex	NoVertex
trackSip3dSig	✔	✔	✔
trackSip2dSigAboveCharm	✔	✔	✔
jetNTracks	✔	✔	✔
trackEtaRel	✔	✔	✘
vertexMass	✔	✔	✘
vertexNTracks	✔	✔	✘
vertexEnergyRatio	✔	✔	✘
vertexJetDeltaR	✔	✔	✘
flightDistance2dSig	✔	✘	✘
jetNSecondaryVertices	✔	✘	✘

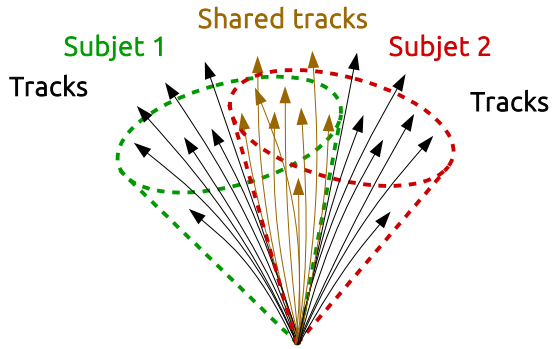
ATLAS shrinking-cone jet-track association

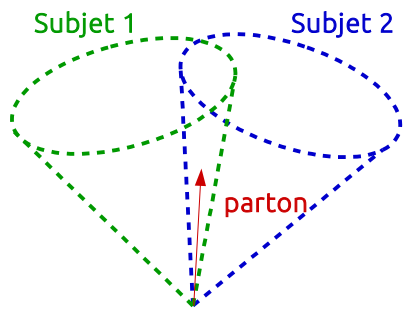
- ATLAS uses a shrinking cone jet-track association with built-in association ambiguity resolution (tracks associated to the closest jet)



Limitations of CMS Run 1 b-tagging setup

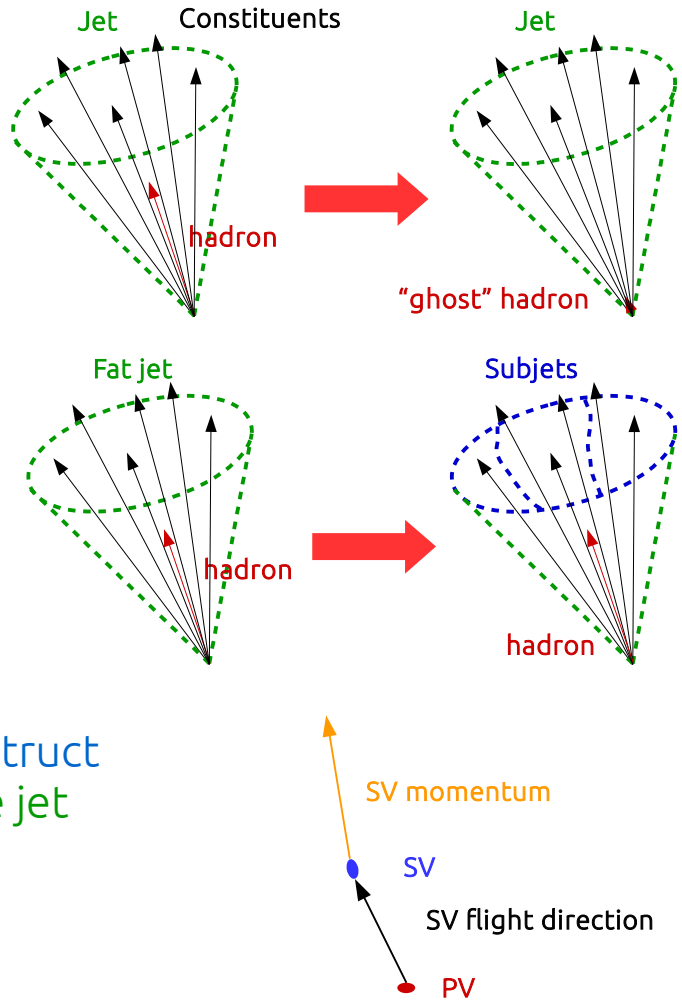
- Current boosted b-tagging setup based on the software framework and tagging algorithms designed for R=0.5 jets
  - Facilitated commissioning studies and early adoption in physics analyses
  - Certain aspects suboptimal for boosted topologies
- Jet-track association:
  - Based on a fixed-size cone
  - Can lead to double-counting of tracks at high p_T and subjet tag correlations (problematic for the application of data/MC scale factors)
  - Default cone size also not optimal for fat jet b tagging
- Jet flavor assignment:
  - Also based on a fixed-size cone ($\Delta R<0.3$)
  - Can lead to subjet flavor ambiguities
- Secondary vertex reconstruction:
  - Using tracks associated to jets (not optimal when the fraction of shared tracks becomes significant)
  - Using a fixed-size cone for SV-jet matching ($\Delta R<0.5$)





CMS Run 2 developments

- Improved (sub)jet flavor definition
  - Using b and c hadrons instead of b and c quarks
  - Based on clustering "ghost" hadrons/partons instead of ΔR matching → Subjet flavor ambiguities eliminated
- Explicit jet-track association
  - Uses tracks linked to charged constituents of particle-flow jets
  - Eliminates the problem of shared tracks
- Inclusive Vertex Finder (IVF) secondary vertices
  - Does not require jets and instead uses all tracks to reconstruct secondary vertices → By construction independent of the jet size and reduces track sharing
  - Jet clustering used to assign SV's to (sub)jets
- Improved CSV algorithm (CSVv2)



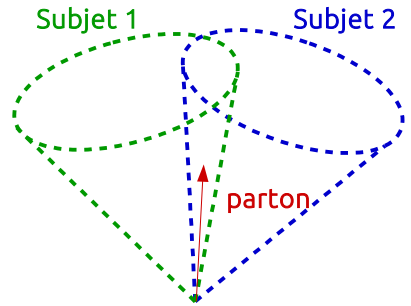
Inclusive Vertex Finder SV reconstruction

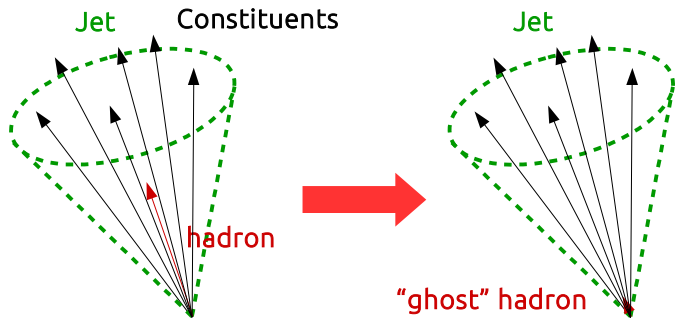
1. Coarse track pre-clustering around displaced seed tracks
   - Based on track distances and angles
2. Vertex reconstruction/fitting from the track clusters obtained in step 1 (using "adaptive vertex fit")
3. Vertex merging
   - Check vertices for shared tracks
   - Remove vertex if shared fraction >0.7 and distance significance <2
4. Track-vertex arbitration
   - Trade off tracks between PV and SV based on their compatibility with vertices
   - Refit vertices with new track selection
5. Vertex merging
   - Same as step 3 with max. shared fraction of 0.2 and min. distance significance of 10

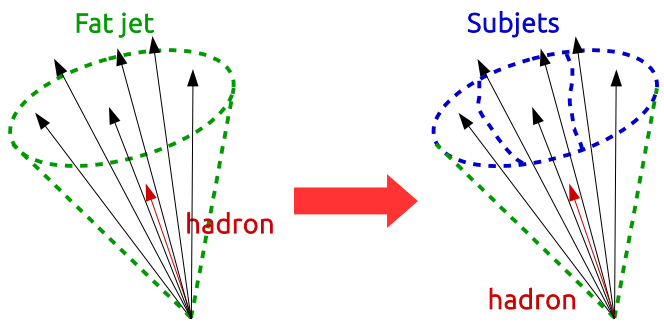
Algorithm employed in JHEP 03 (2011) 136

Jet flavor

- Jet flavor tools:
  - Problem: Written specifically for Pythia6 → Not fully compatible with newer MC event generators. ΔR<0.3 cone used for matching generator partons and reconstructed jets → Not optimal for jets of different sizes and can lead to flavor ambiguities for nearby subjets
  - Solution: Use b and c hadrons instead of partons and assign jet flavor using jet clustering instead of simple ΔR matching
    - Rescale the hadron momenta to make them extremely soft (turn them into "ghosts") and then recluster "ghost" hadrons and regular jet constituents
    - "Ghost" hadrons clustered inside a fat jet later assigned to the closest subjet
  - More information available in a dedicated TWiki [1]



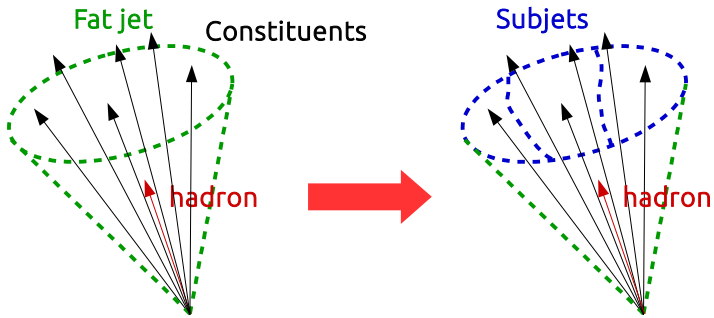




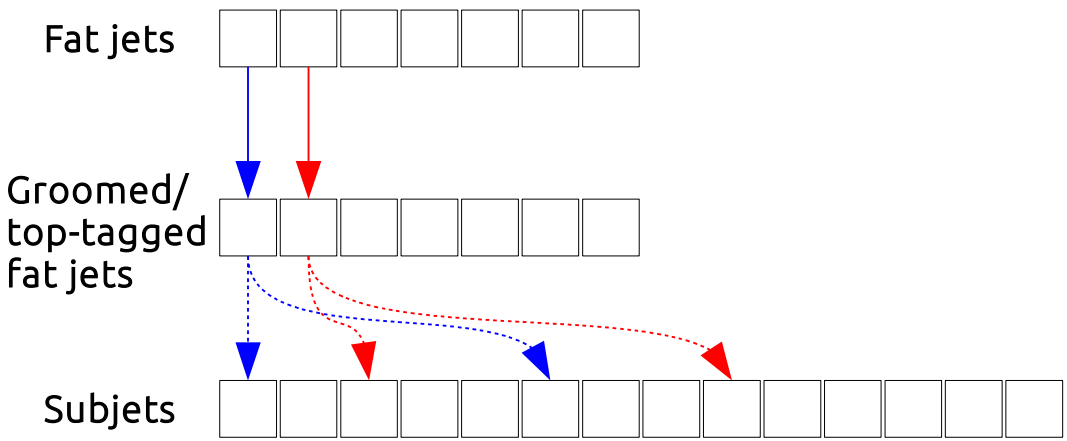
[1] https://twiki.cern.ch/twiki/bin/view/CMSPublic/SWGuideBTagMCTools

Subjet flavor

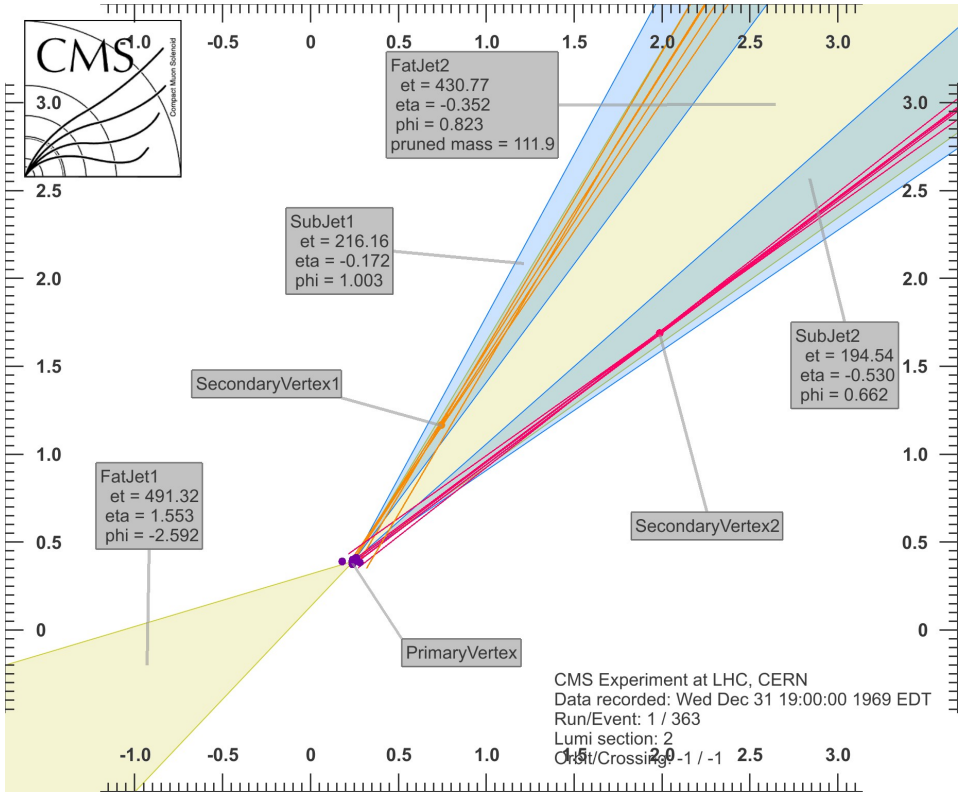
- Subjet flavor definition:
  - "Ghost" hadrons/partons clustered inside a fat jet later assigned to the closest subjet in rapidity-based ΔR



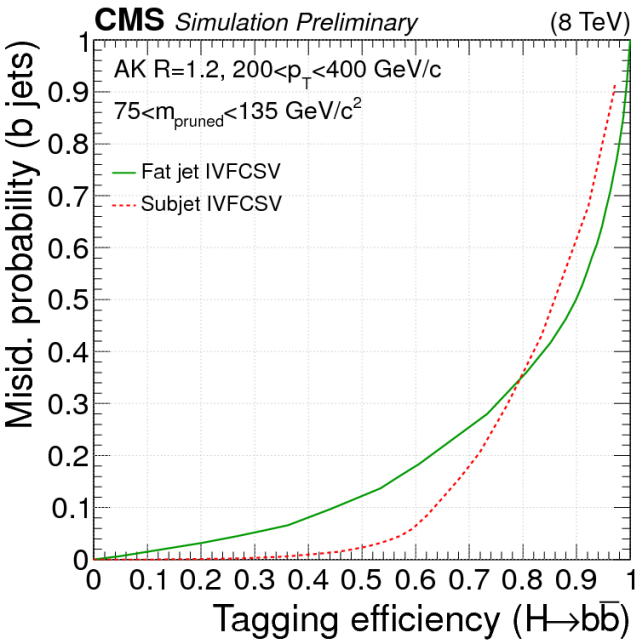
→ In order to assign subjet flavor, need external fat jet collections (to avoid flavor inconsistencies between subjets and fat jets)

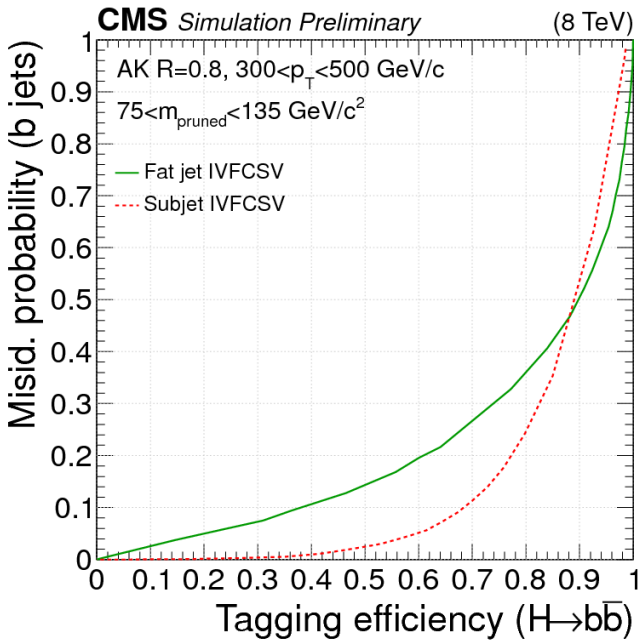


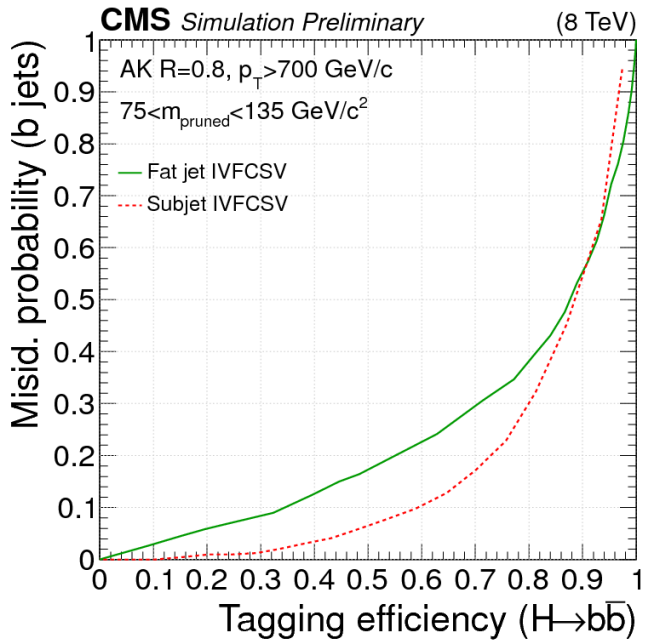
Boosted H→$b\bar{b}$ (simulation)



Boosted H→$b\bar{b}$ (b jets as background)

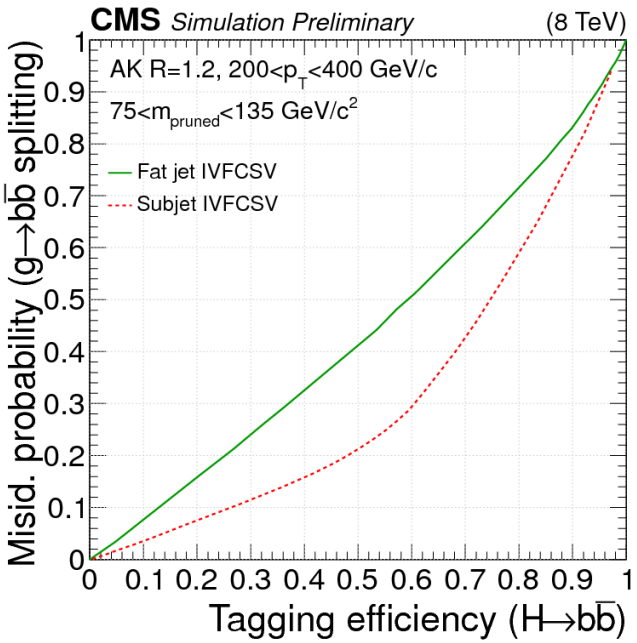


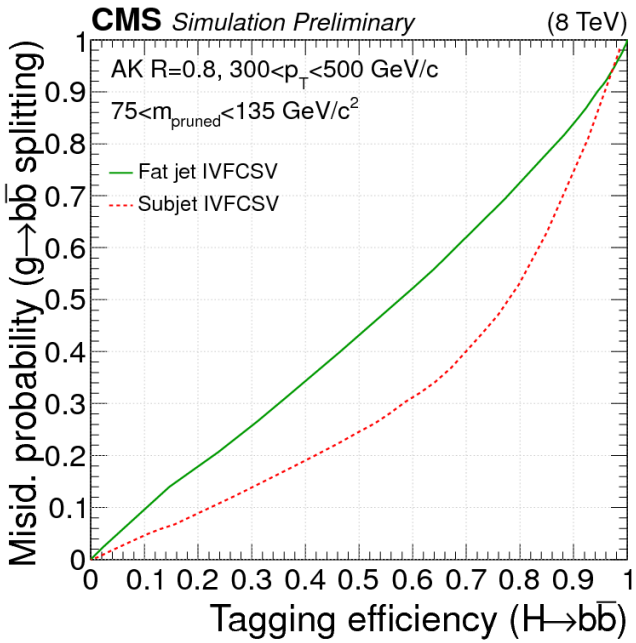


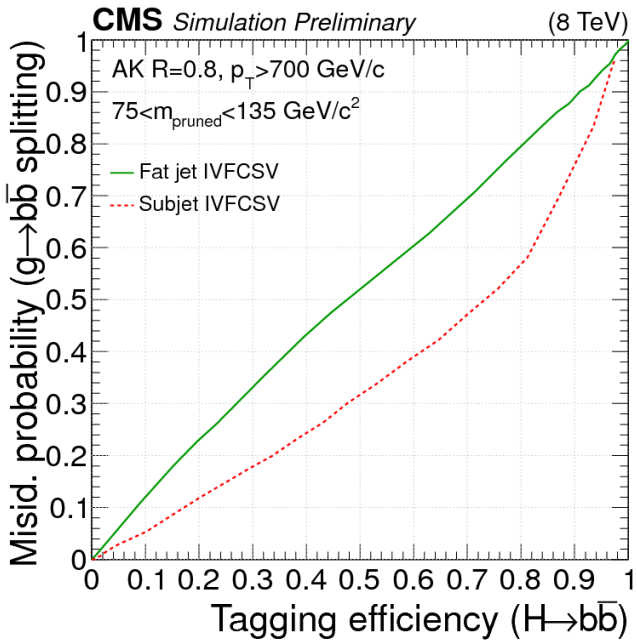


- Subjet b tagging generaly outperforms fat jet b tagging except at high tagging efficiencies for lower p_T

Boosted H→$b\bar{b}$ (Gluon splitting as background)

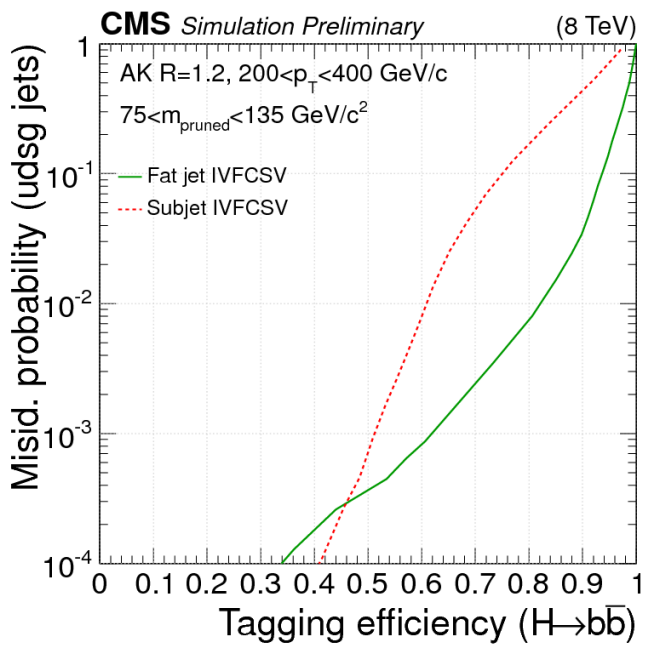


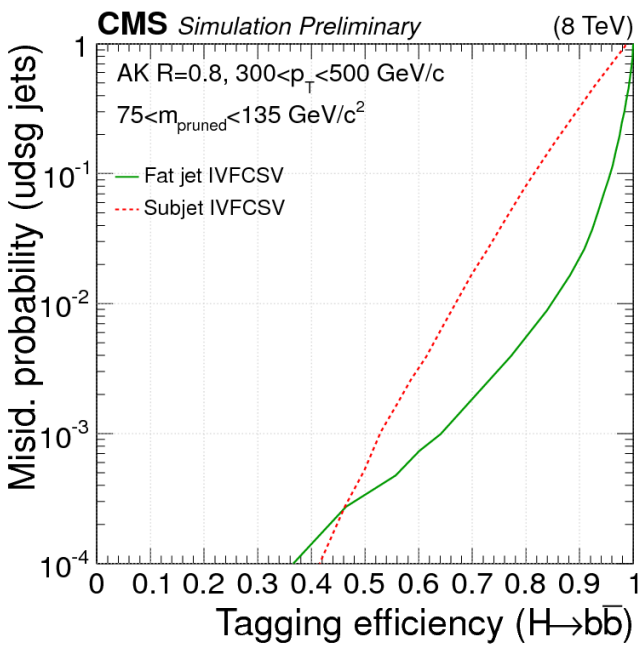


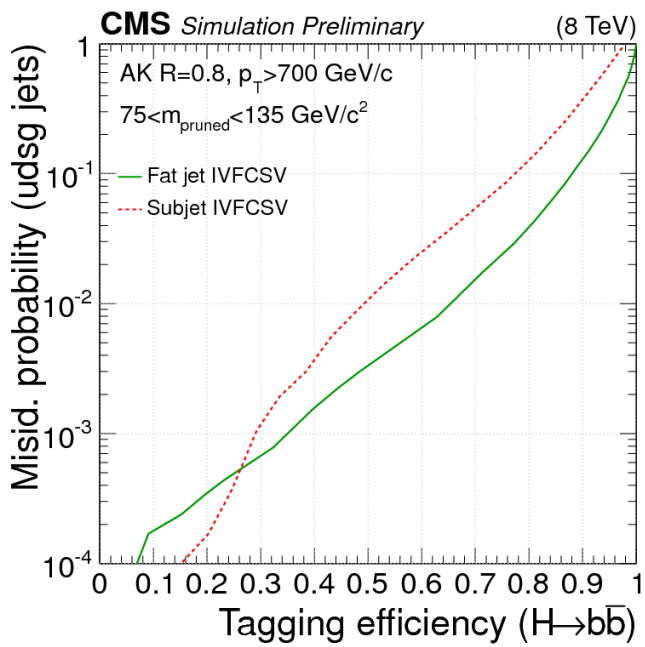


- Subjet b tagging outperforms fat jet b tagging in the entire p_T range considered

Boosted H→$b\bar{b}$ (udsg jets as background)

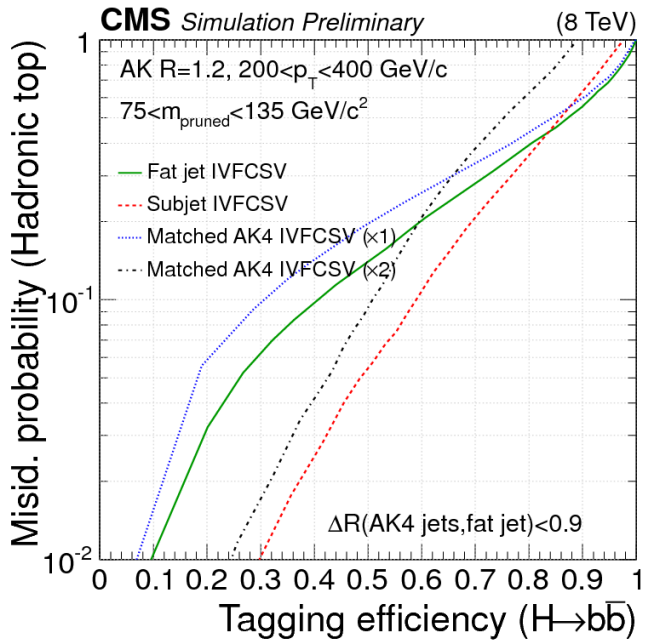


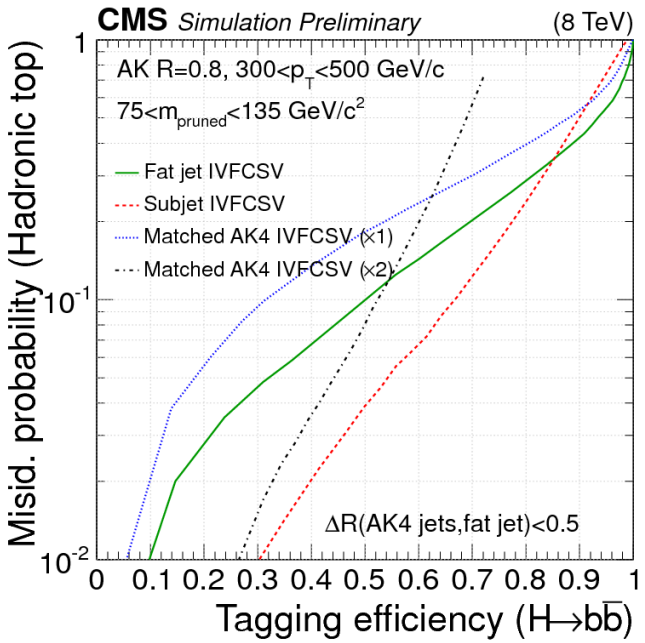


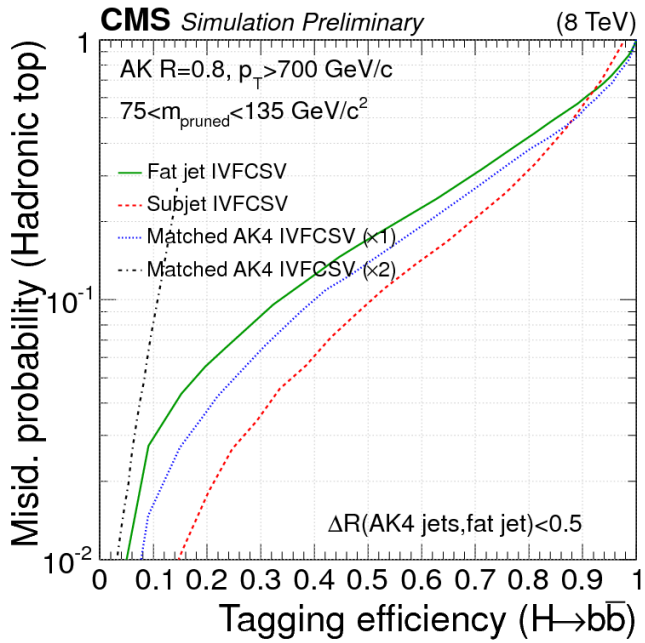


- Fat jet b tagging generally outperforms subjet b tagging in the entire p_T range considered, expect at low tagging efficiencies

Boosted H→$b\bar{b}$ (Hadronic top as background)





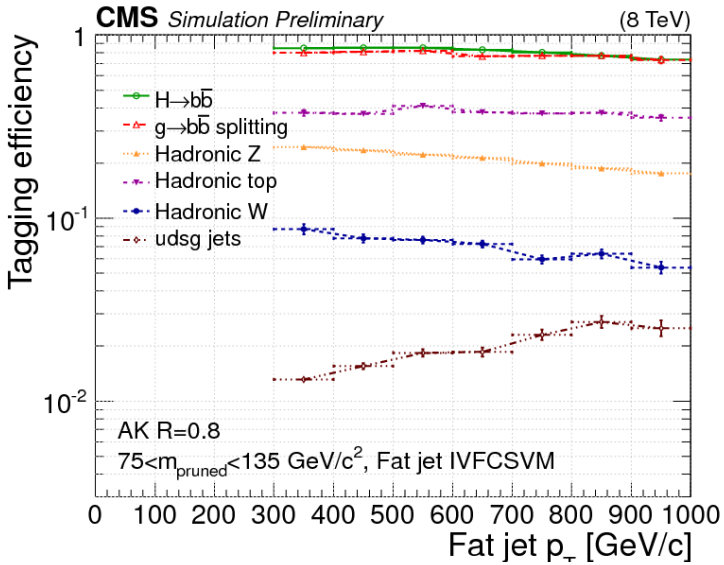


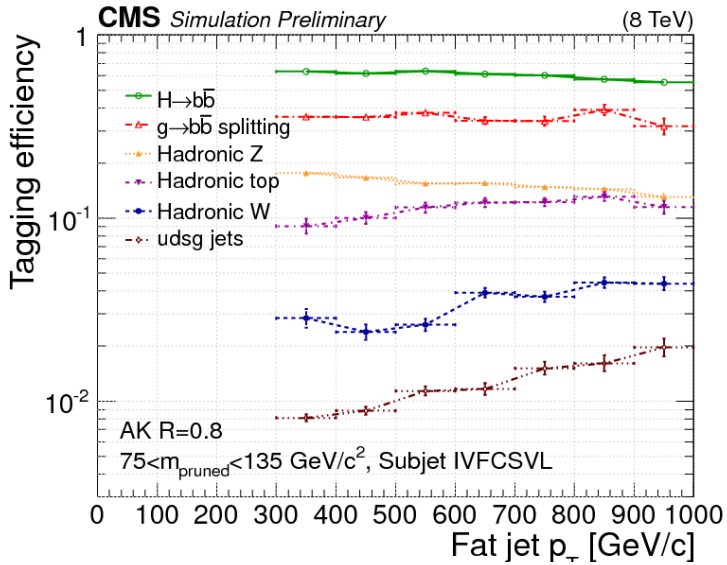
- Subjet b tagging outperforms both fat jet b tagging and matched b-tagged AK4 jets in the entire p_T range considered

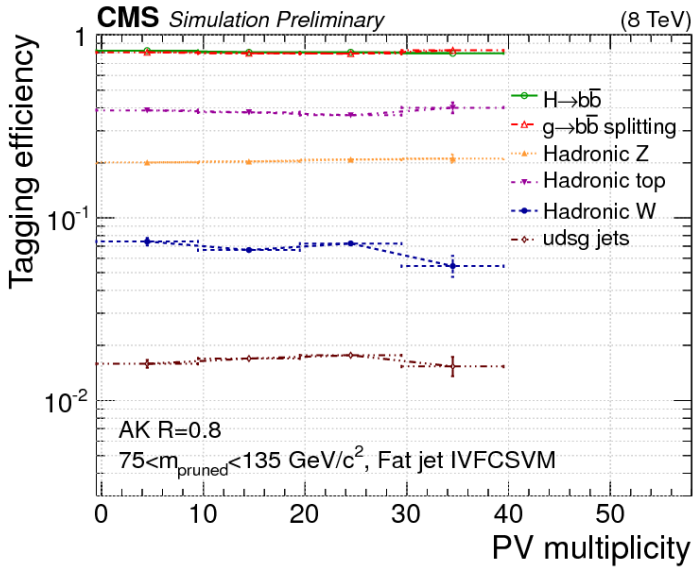
p_T and pileup dependence

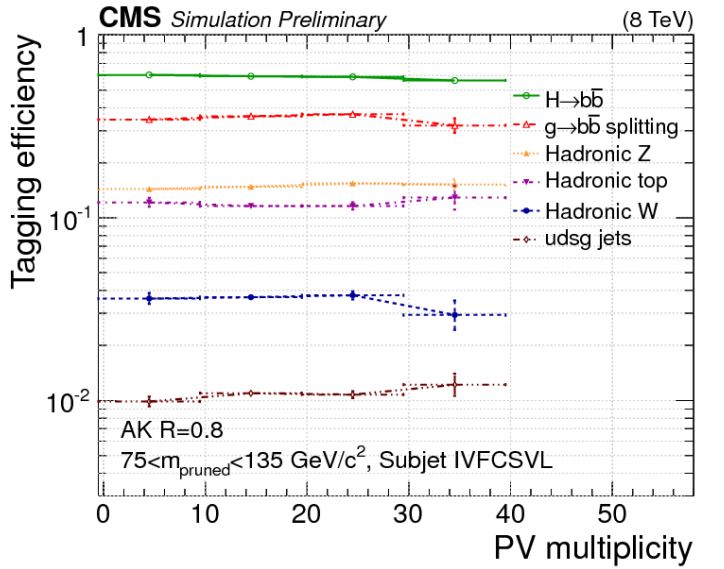
Boosted H→bb̄

Boosted H→bb̄

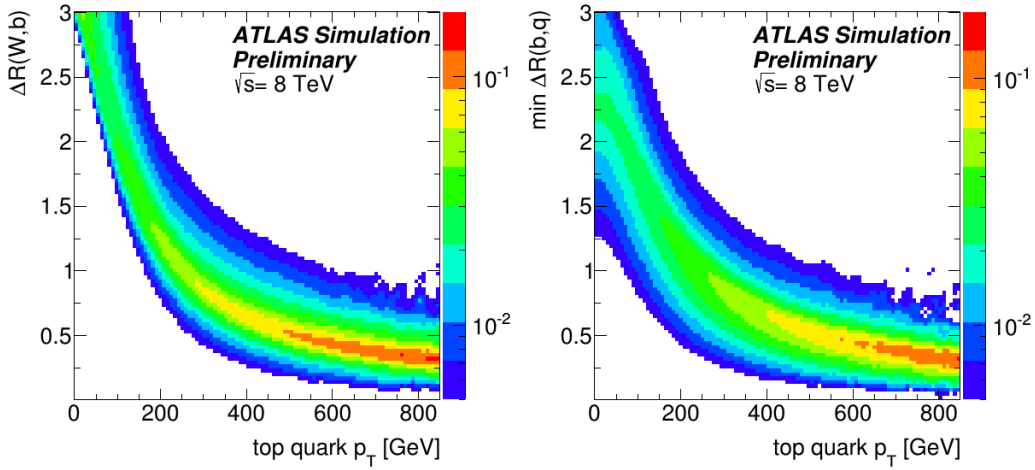




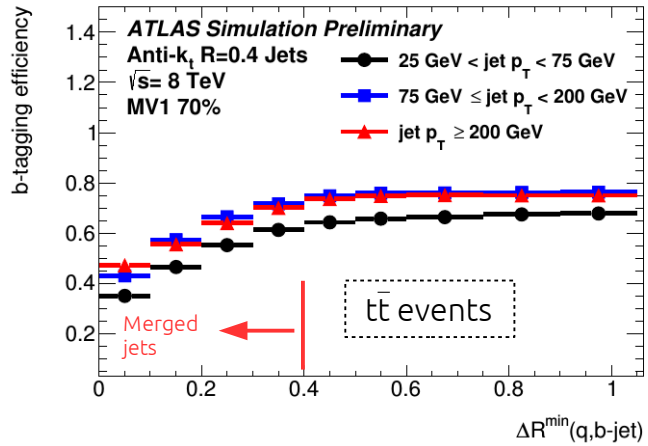




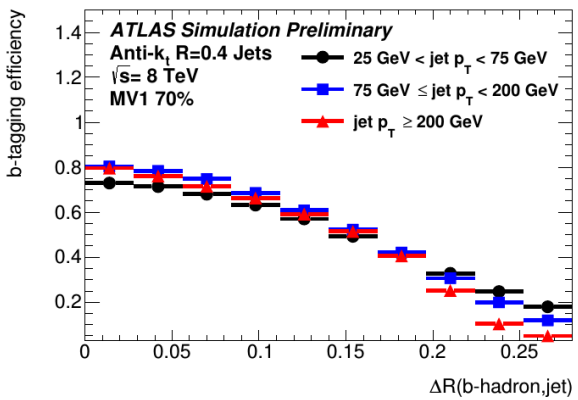
b tagging of standard (R=0.4) jets

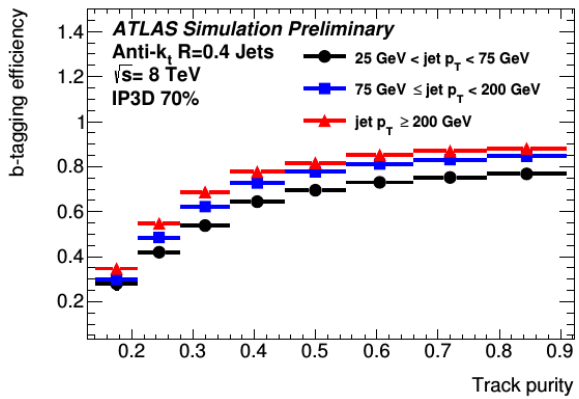


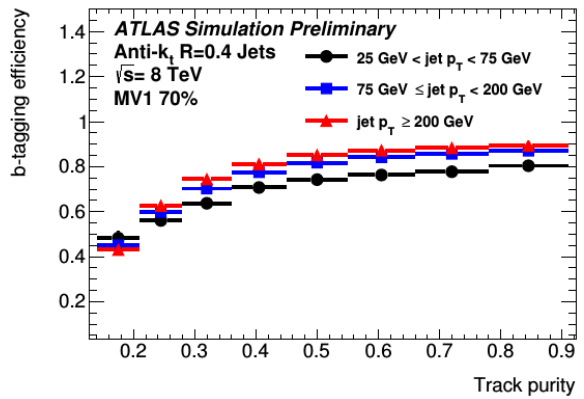
Angular separation between top decay products



Performance degrades as decay products get closer

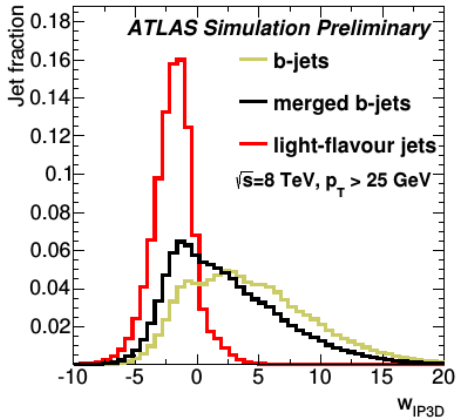




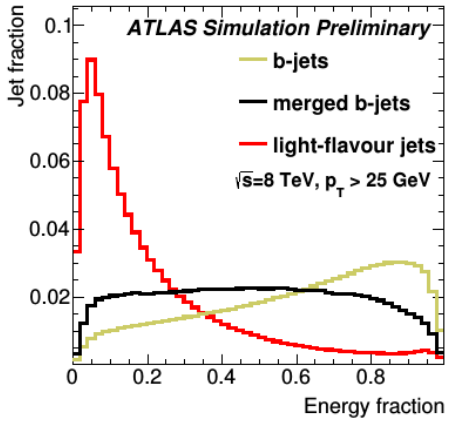


Degradation attributed to the two main effects: 1) shifted jet axis (not necessarily aligned with the b hadron flight direction), 2) light-flavor contamination

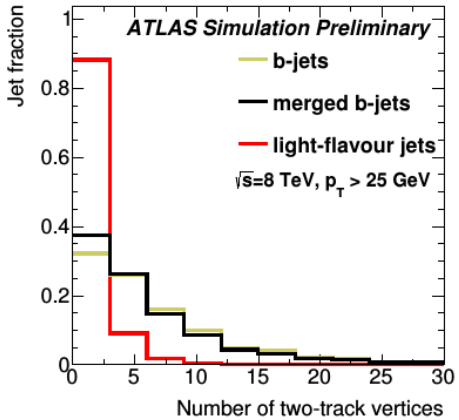
b tagging of standard (R=0.4) jets (cont'd)



(a) IP3D weight



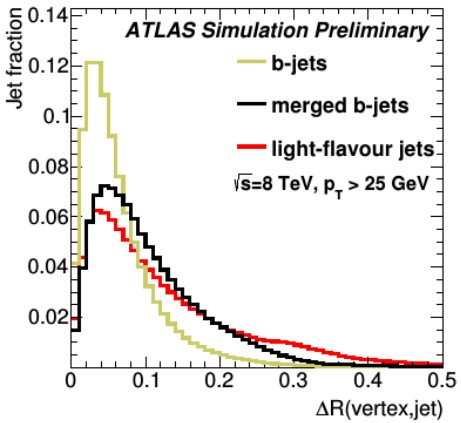
(b) energy fraction (JetFitter based)



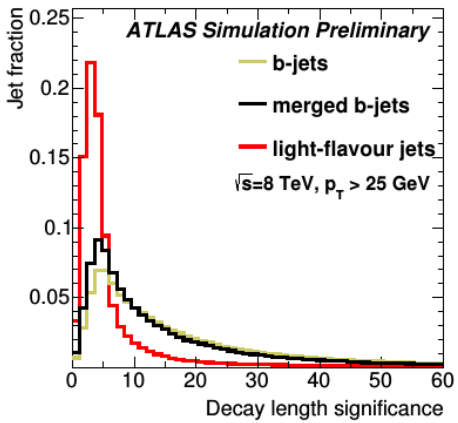
(c) number of two-track vertices (JetFitter based)

ATLAS performed a systematic study of the sensitivity of various track- and SV-related variables used in their b-tagging algorithms

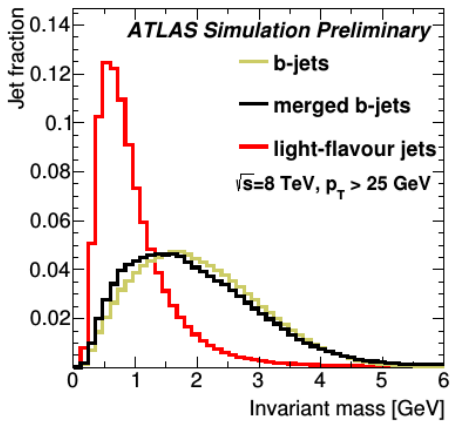
Performance degrades because of reduced discrimination power and distributions dissimilar to those from the training sample



(d) ΔR(vertex, jet) (SV1 based)

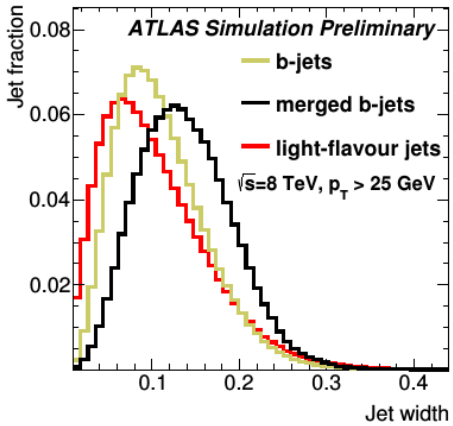


(e) decay length significance (SV1 based)

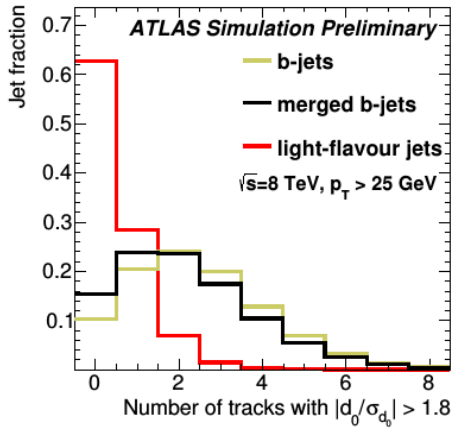


(f) invariant mass (SV1 based)

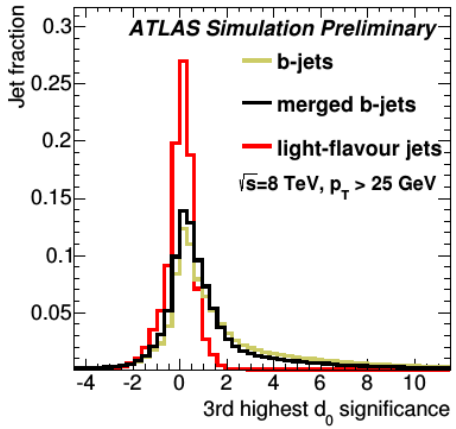
b tagging of standard (R=0.4) jets (cont'd)



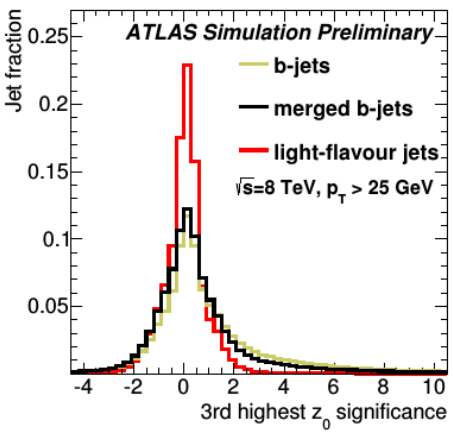
(a) jet shape based



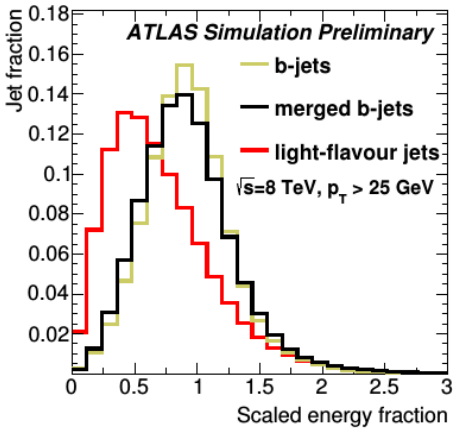
(b) IP based



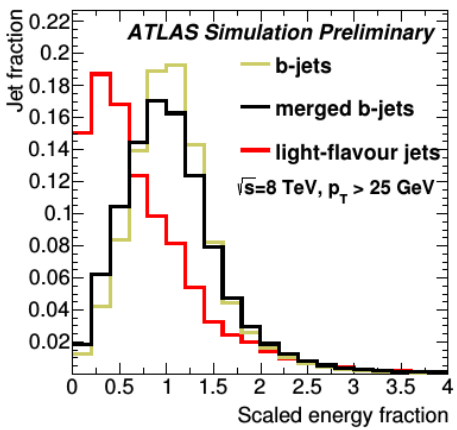
(c) IP based



(d) IP based



(e) SV1 based



(f) JetFitter based

Additional variables studied for their robustness

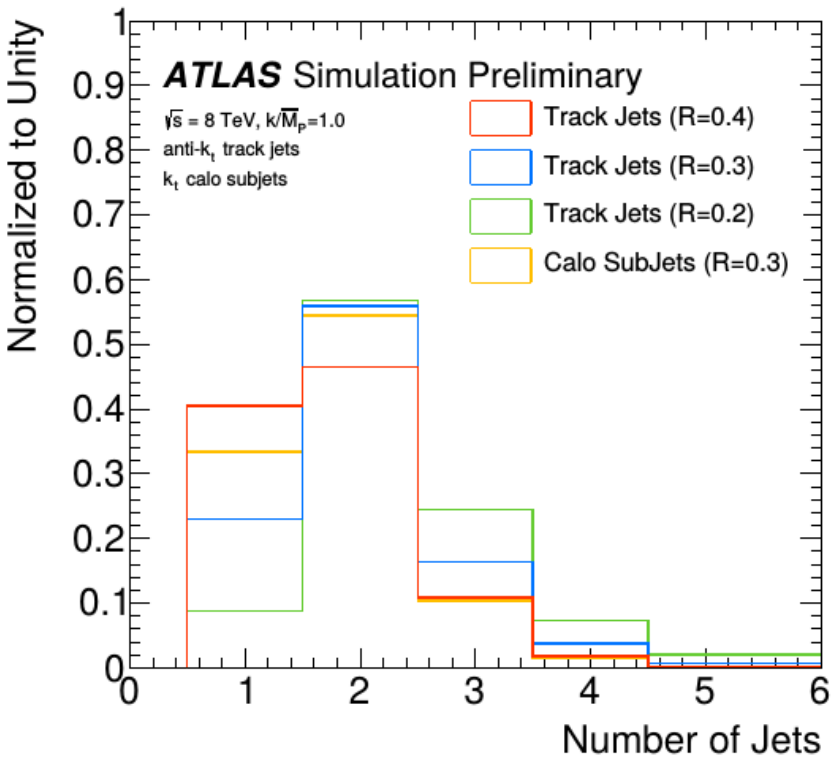
Dedicated algorithm defined by introducing additional variables less sensitive to jet overlaps (effectively an extension of MV1)

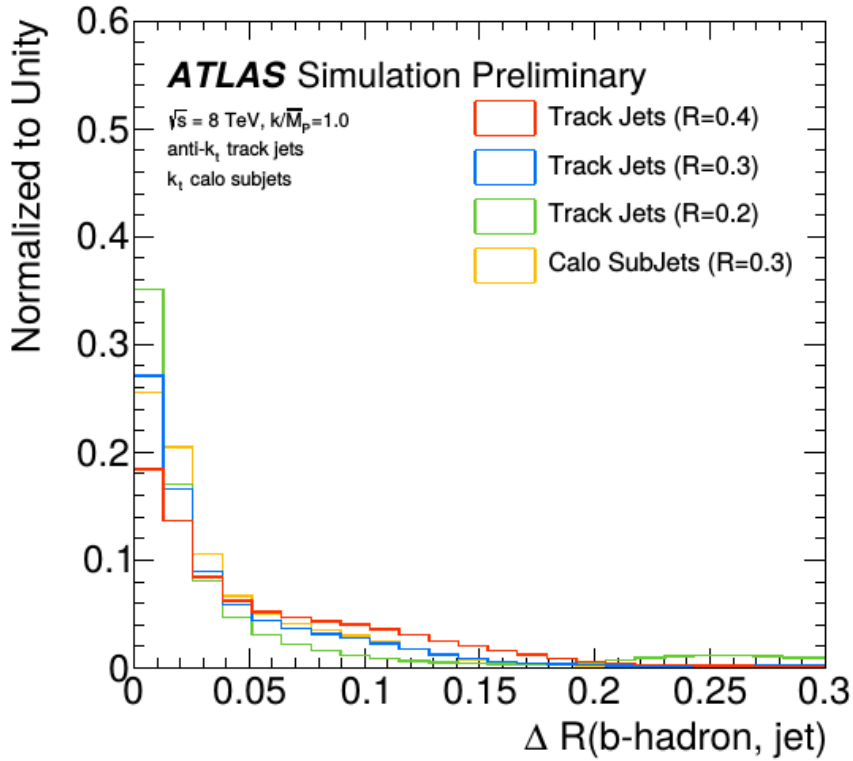
Using 23 input variables and putting them into a BDT

MVb (trained for b vs light)

MVbCharm (trained for b vs c)

b tagging of smaller-size track jets





Using $G_{RS} \rightarrow hh \rightarrow 4b$ as a benchmark process