Probing Higgs Boson Properties With Boosted Objects at the CMS Experiment

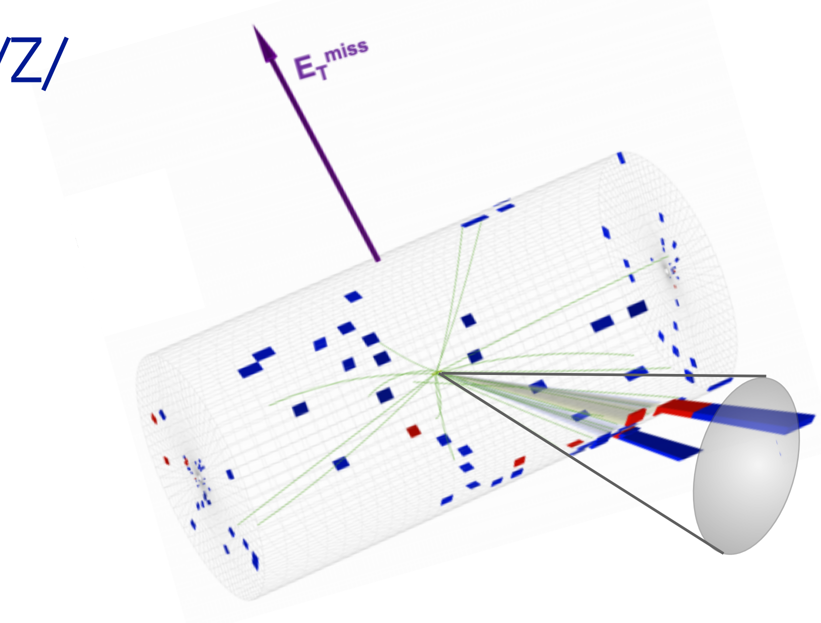
Huilin Qu (CERN)

Higgs potential and BSM opportunity
August 29, 2021

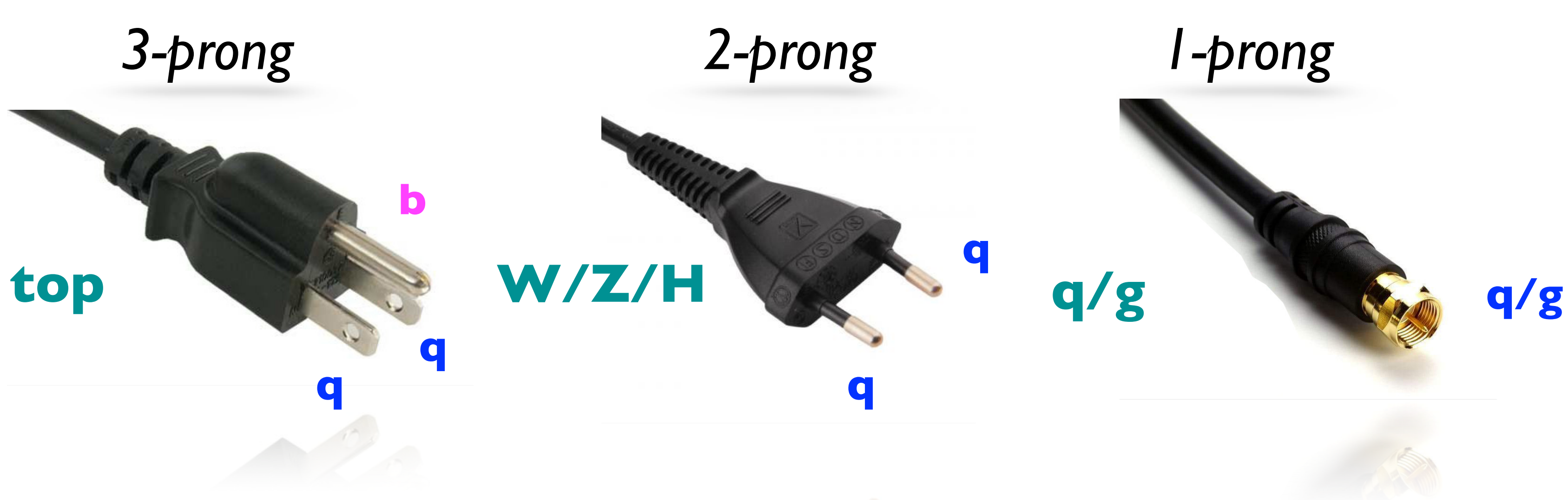
Introduction

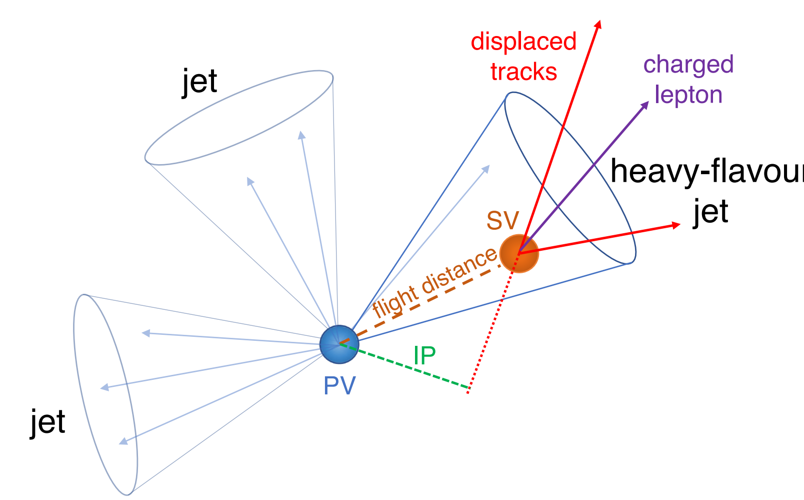
- Boosted objects: powerful tools for new physics searches and standard model measurements at the LHC
- Hadronic decays of highly boosted heavy particles (Higgs/W/Z/top) lead to large-R jets with distinctive characteristics:
  - different radiation patterns ("**substructure**")
    - 3-prong (top), 2-prong (W/Z/H) vs 1-prong (gluon/light quark jet)
  - different **flavor** content: existence of one or more b-/c-quarks
  - simultaneously exploiting both **substructure** and **flavor** to maximize the performance
  - significant performance leap thanks to new **machine learning** (ML) techniques

$pp \to Z(\nu\nu)H(cc)$



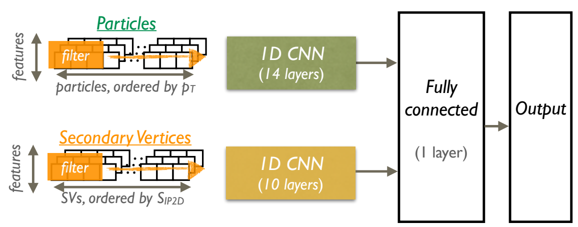
$$\Delta R \sim \frac{2m}{p_T}$$



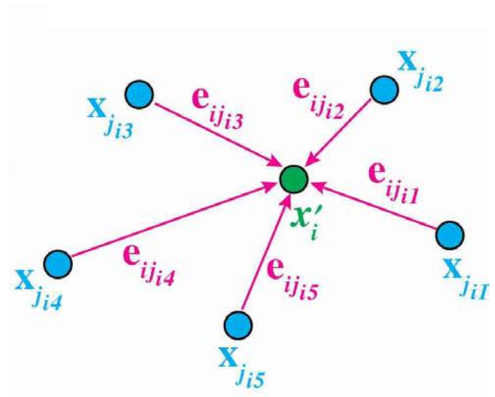


OUTLINE

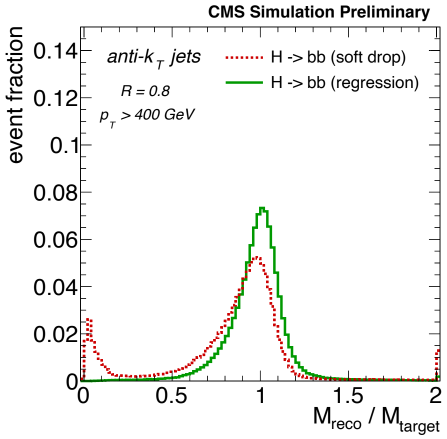
The tool box …



DeepAK8



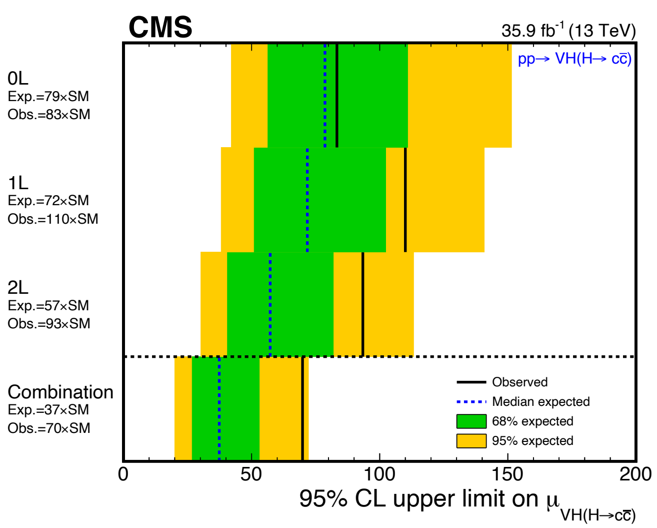
ParticleNet



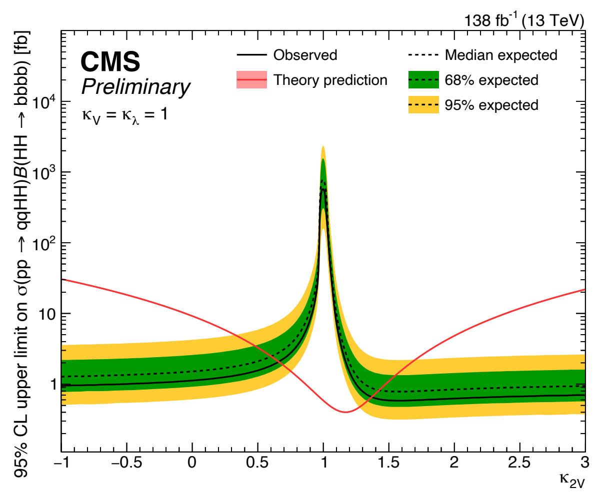
Mass regression NEW

…

… in action!



VH(H→cc)



VBF HH(→4b) NEW

…

See also [Qiang's talk] for more applications targeting BSM scenarios.

DeepAK8

- Advanced deep learning-based algorithm for boosted object tagging, using AK8 (anti-k_T R=0.8) jets
  - multi-class classifier for top quark and W, Z, Higgs boson tagging
    - sub-classes based on decay modes (e.g., H→bb, H→cc, H→VV*→4q)
    - output scores can be aggregated/transformed for different tasks -> highly versatile tagger
  - directly uses jet constituents (particle-flow candidates / secondary vertices)
  - 1D convolutional neural network (CNN) based on the ResNet [arXiv: 1512.03385] architecture
  - significant performance improvement

Inputs

Substructure – Particles

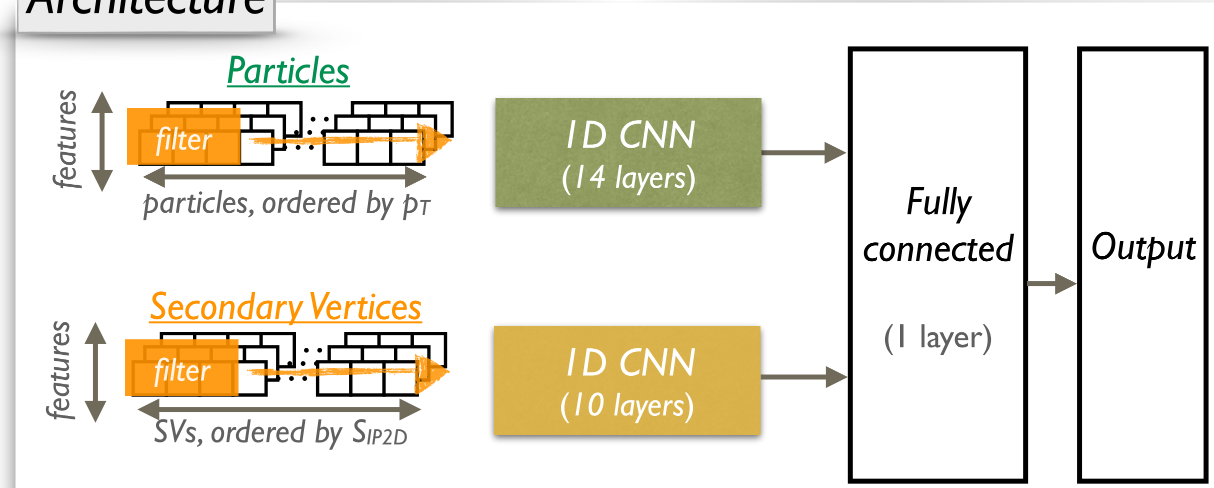
- Up to 100 PF candidates(*)
- Sorted in descending p_T order
- Uses basic kinematic variables, Puppi weights, and track properties (quality, covariance, displacement, etc.)

Flavour – Secondary vertices

- Up to 7 SVs(*) (inside jet cone)
- Sorted in descending S_{IP2D} order
- Uses SV kinematics and properties (quality, displacement, etc.)

(*) Number chosen to include all candidates for ≥ 90% of the events

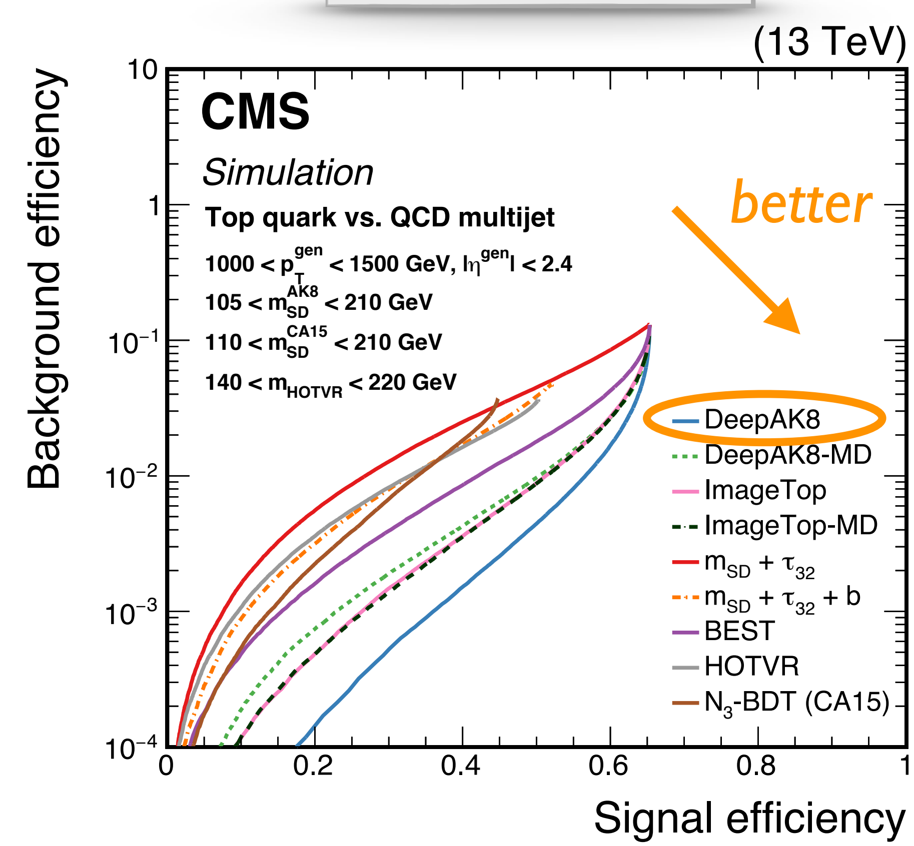
Architecture



Output

Category	Label
Higgs	H (bb)
	H (cc)
	H (VV*→qqqq)
Top	top (bcq)
	top (bqq)
	top (bc)
	top (bq)
W	W (cq)
	W (qq)
Z	Z (bb)
	Z (cc)
	Z (qq)
QCD	QCD (bb)
	QCD (cc)
	QCD (b)
	QCD (c)
	QCD (others)

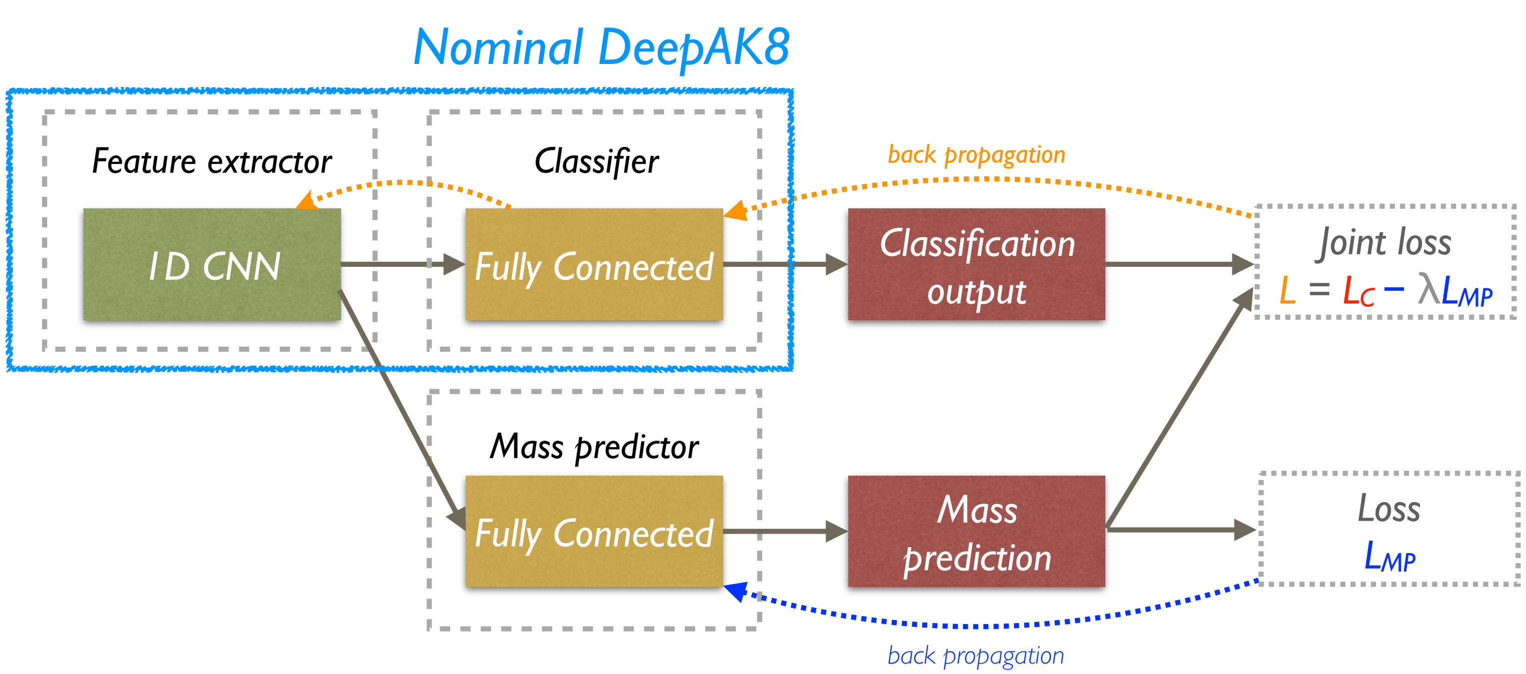
Top quark tagging

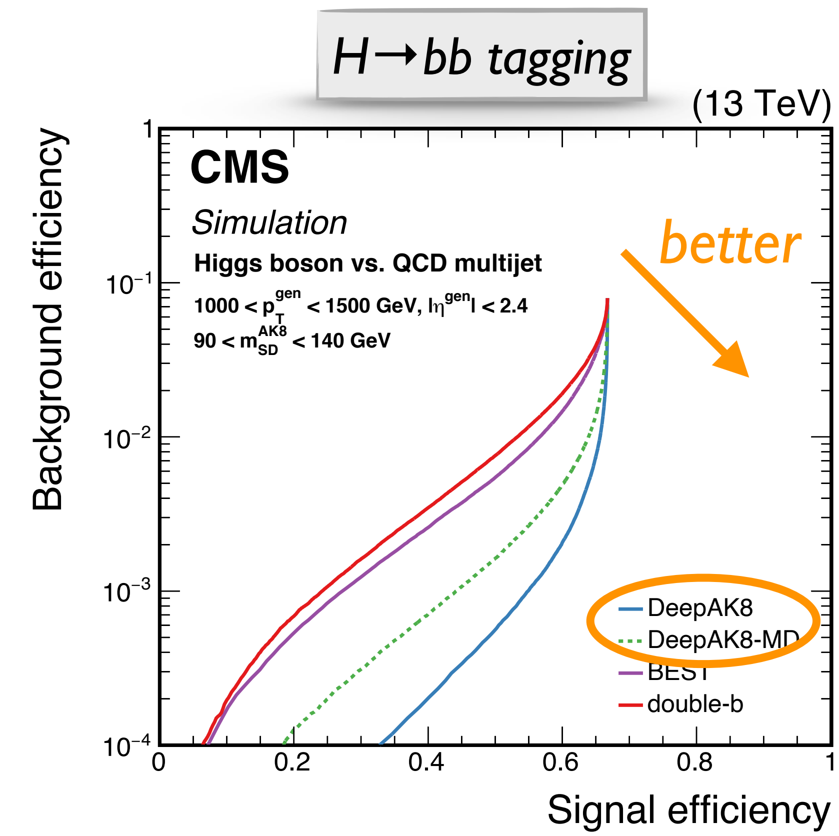


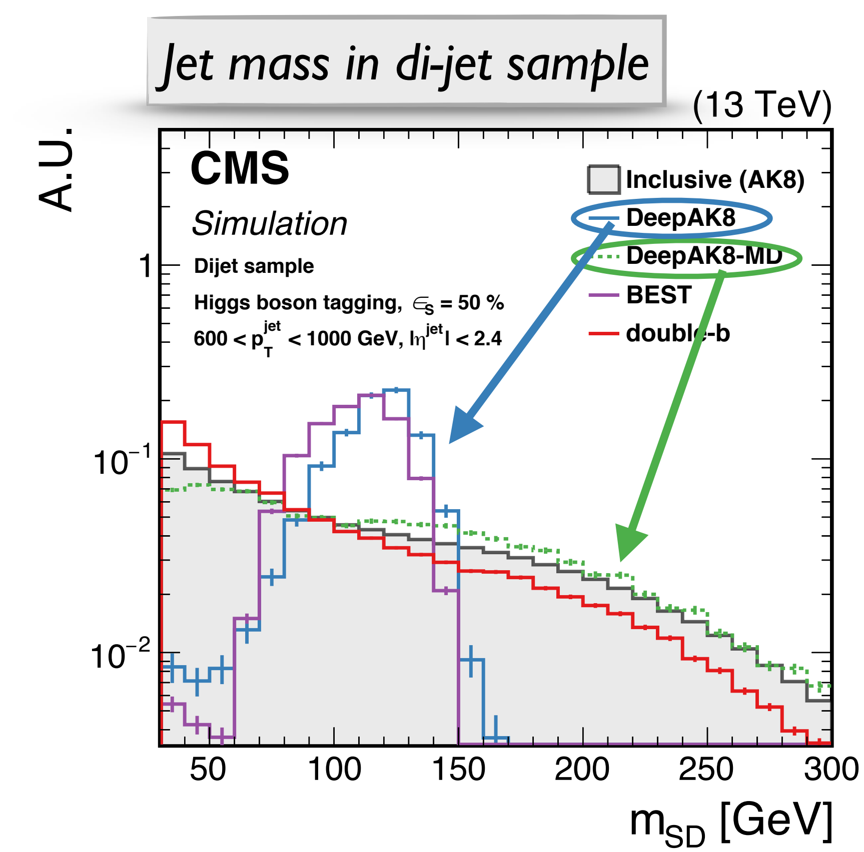
JINST 15 (2020) P06005

DEEPAK8-MD

- The nominal version of DeepAK8 shows significantly improved performance, but also features strong "mass sculpting"
  - i.e., jet mass shape of the background becomes similar to that of the signal after selection with the tagger
- Mass-decorrelated tagger: "DeepAK8-MD"
  - mitigate mass sculpting using "adversarial training" [arXiv: 1611.01046]
  - added a mass prediction network to predict the jet mass from the learned features
    - higher mass prediction accuracy -> stronger correlation w/ the jet mass
  - accuracy of the mass prediction included in the loss function as a penalty
    - minimizing the joint loss -> improving classification accuracy while preventing mass correlation
  - significantly reduced mass sculpting yet still strong performance



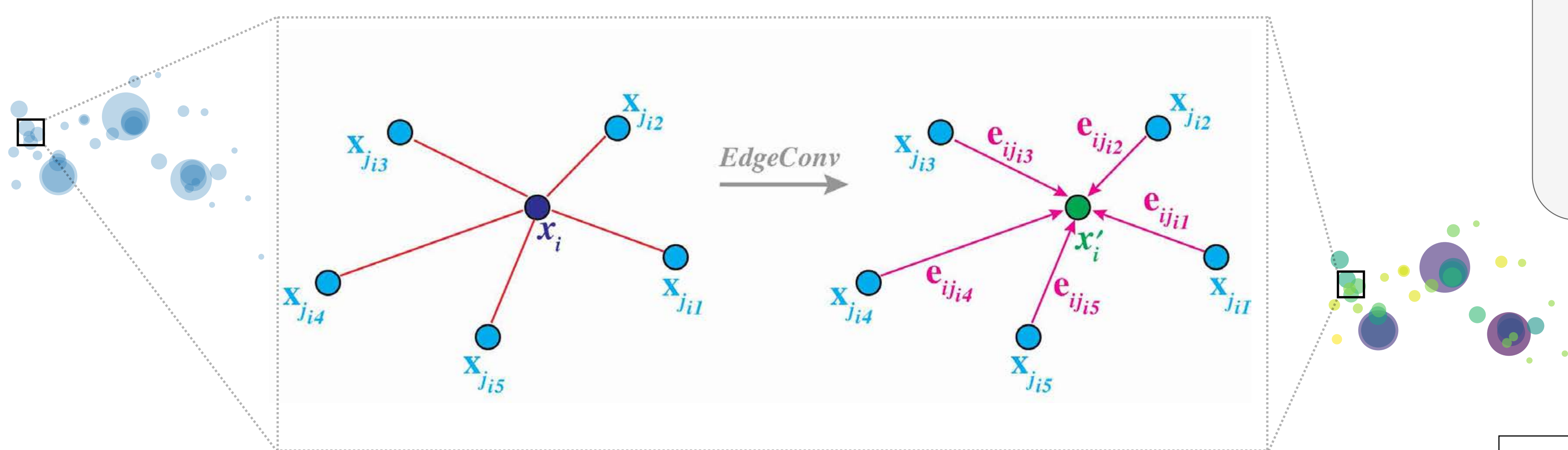




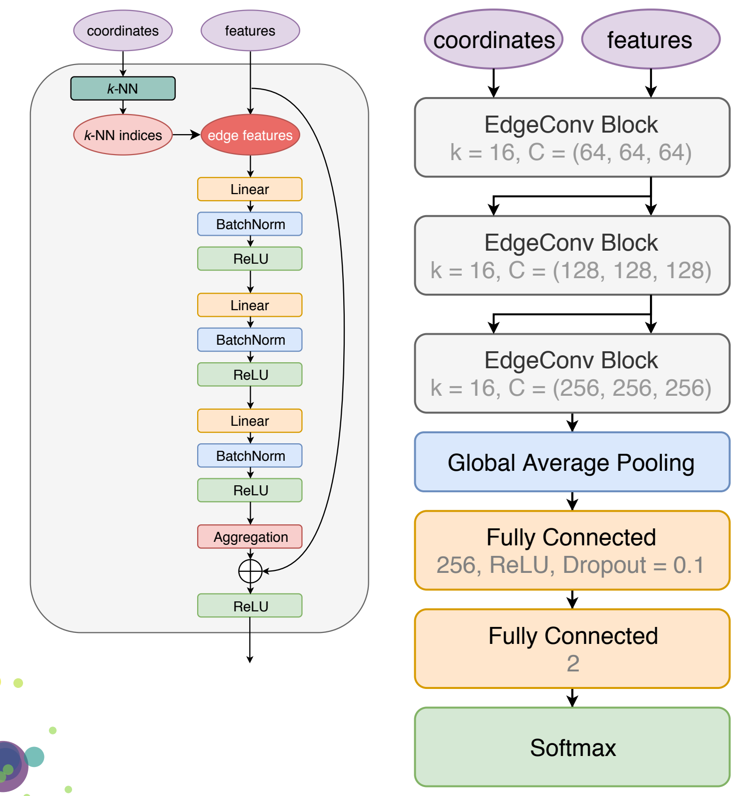
JINST 15 (2020) P06005

ParticleNet

- ParticleNet [Phys. Rev. D 101, 056019 (2020)]
  - treating a jet as an **unordered set of particles** in space
  - using permutation-invariant **graph neural networks**

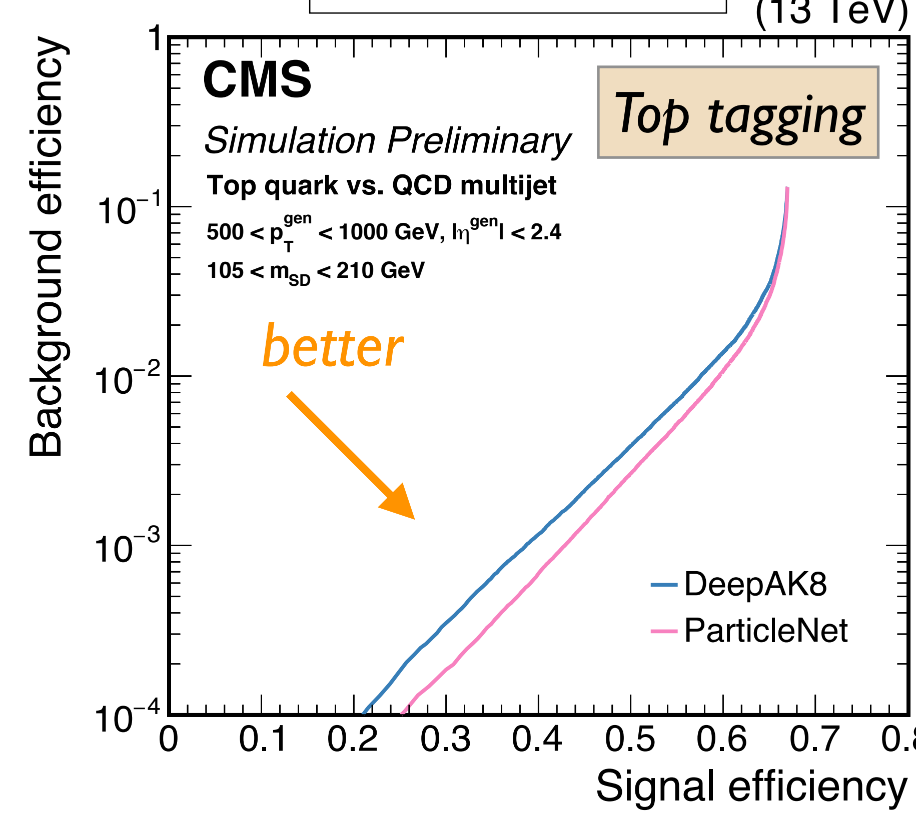


ParticleNet architecture



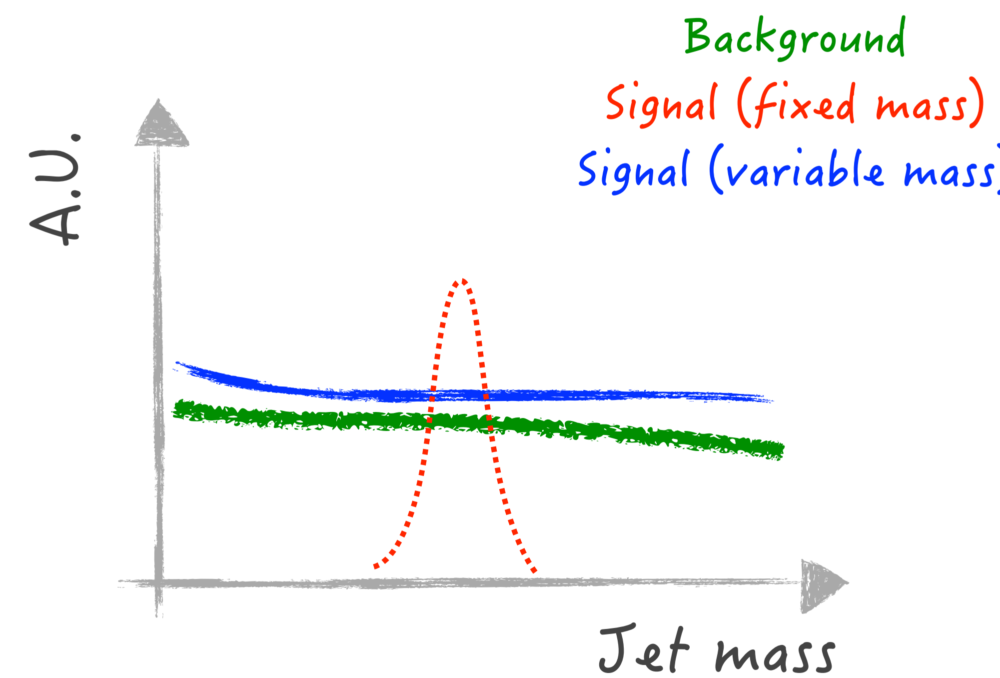
- ParticleNet for boosted jet tagging in CMS
  - multi-class tagger for t/W/Z/H tagging
  - same inputs as DeepAK8 (PF candidates + secondary vertices)
  - significant performance improvement

CMS DP-2020/002

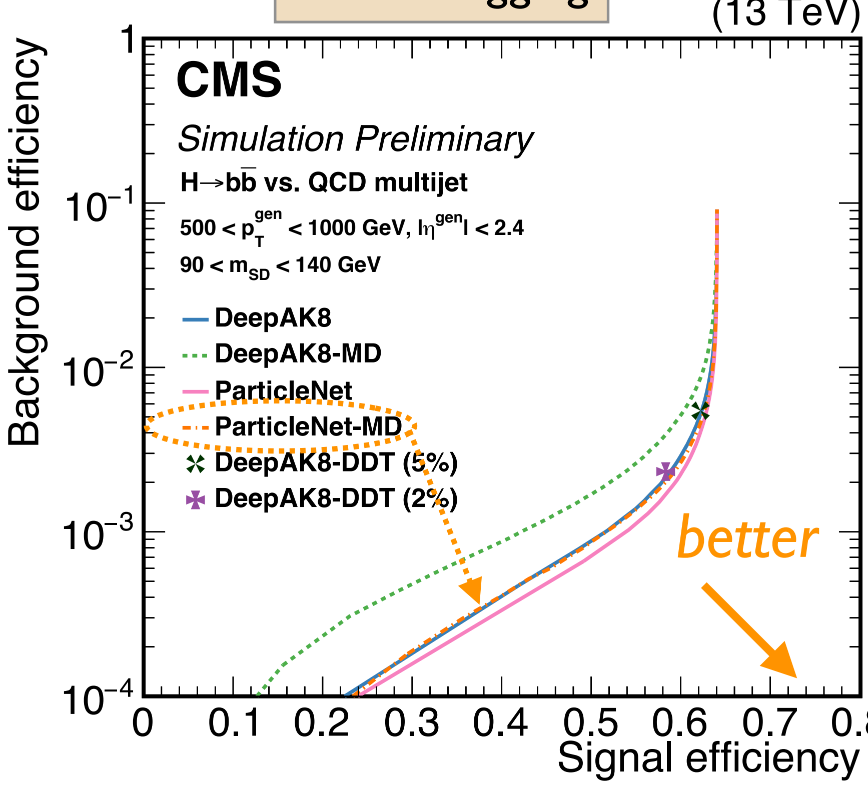


ParticleNet-MD

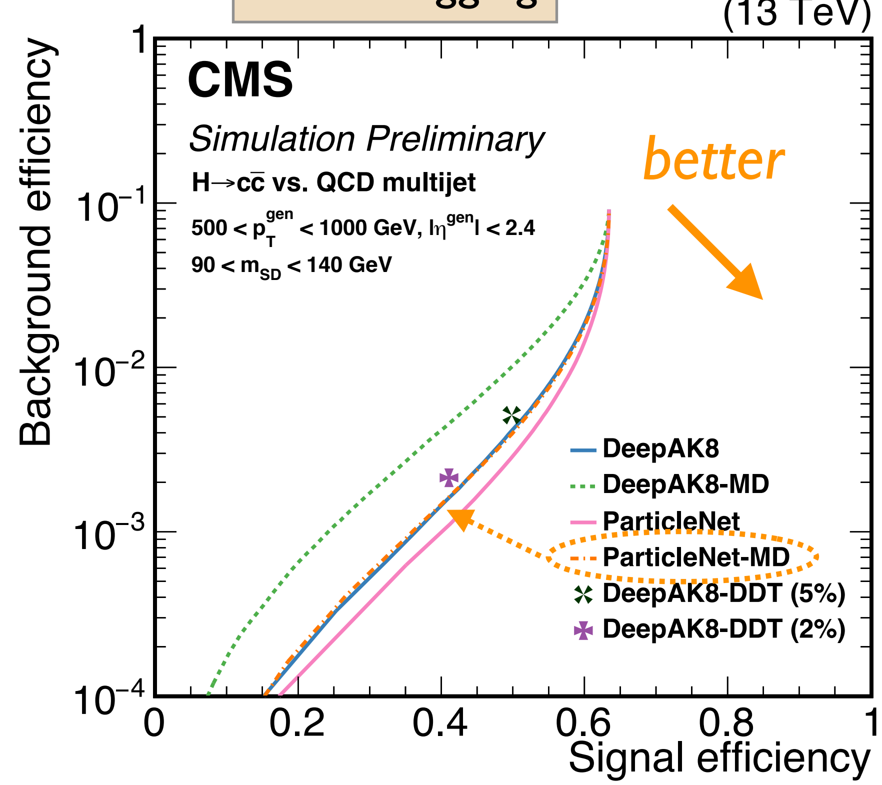
- ParticleNet-MD
  - exploiting a dedicated signal sample for training:
    - hadronic decays of a spin-0 particle X
      - $X \to b\bar{b}, X \to c\bar{c}, X \to q\bar{q}$
    - **flat mass spectrum**: $m_X \in [15, 250]$ GeV
  - in addition: signal/background samples reweighted to a ~flat (p_T, m_{SD}) distribution to aid the training
  - both signal and background have the same mass spectrum, thus no sculpting can form during the training



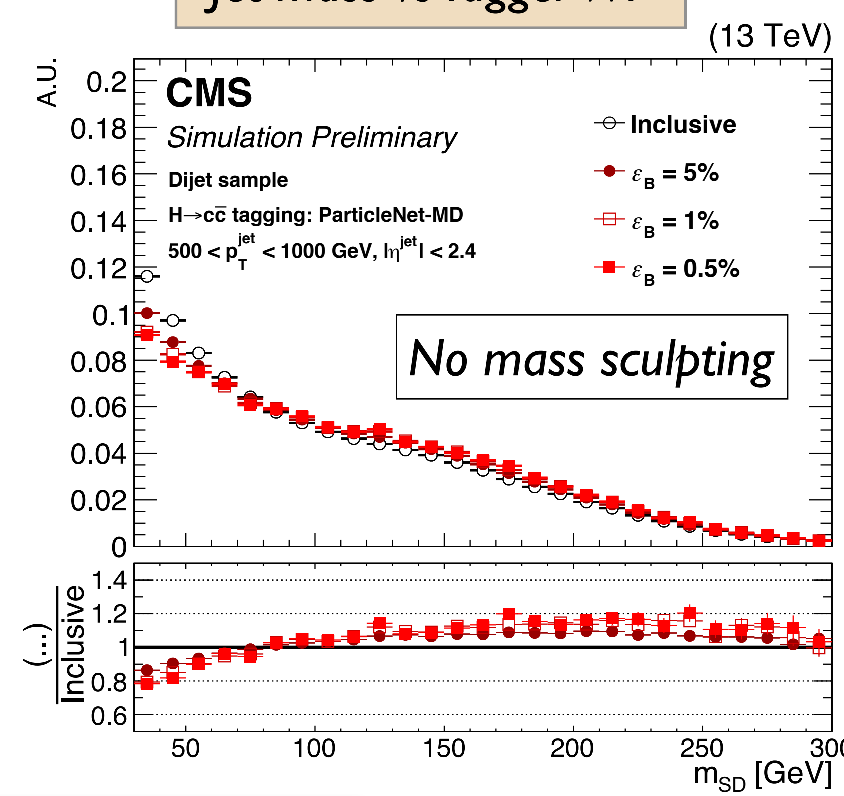
H→bb tagging



H→cc tagging



Jet mass vs Tagger WP



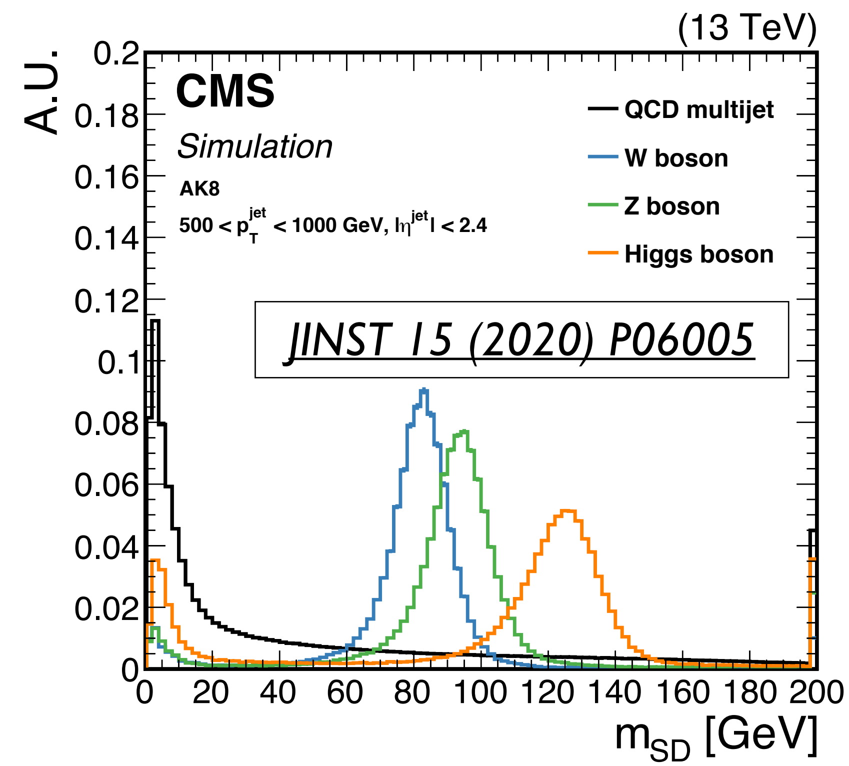
ParticleNet-MD: ~3-4x better background rejection than DeepAK8-MD

MASS REGRESSION

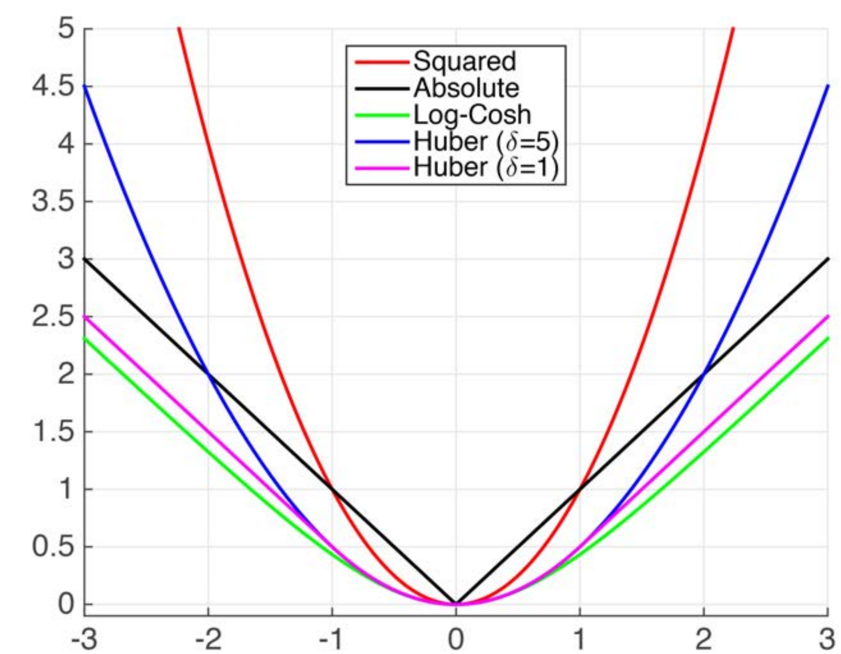
NEW

CMS DP-2021/017

- Jet mass: one of the most powerful observables for boosted jet tagging
  - characteristic mass peak for top/W/Z/H jets v.s. continuum for QCD jets
  - grooming technique (e.g., soft drop) typically used to reduce sensitivity to unrelated radiations (initial-state radiation, underlying event, pileup, etc.)
- Mass regression
  - exploit the ParticleNet architecture to predict the jet mass directly from jet constituents
  - similar setup as the ParticleNet-MD tagger (inputs, training samples, etc.)
  - regression target
    - signal: generated particle mass (pole mass) of X [ranging from 15−250 GeV]
    - background: soft drop mass of the particle-level jet
  - loss function
    - LogCosh: $$L(y, y^p) = \sum_{i=1}^{n} \log(\cosh(y_i^p - y_i))$$
  - focus on Higgs (or generally 2-prong jets) for now



JINST 15 (2020) P06005



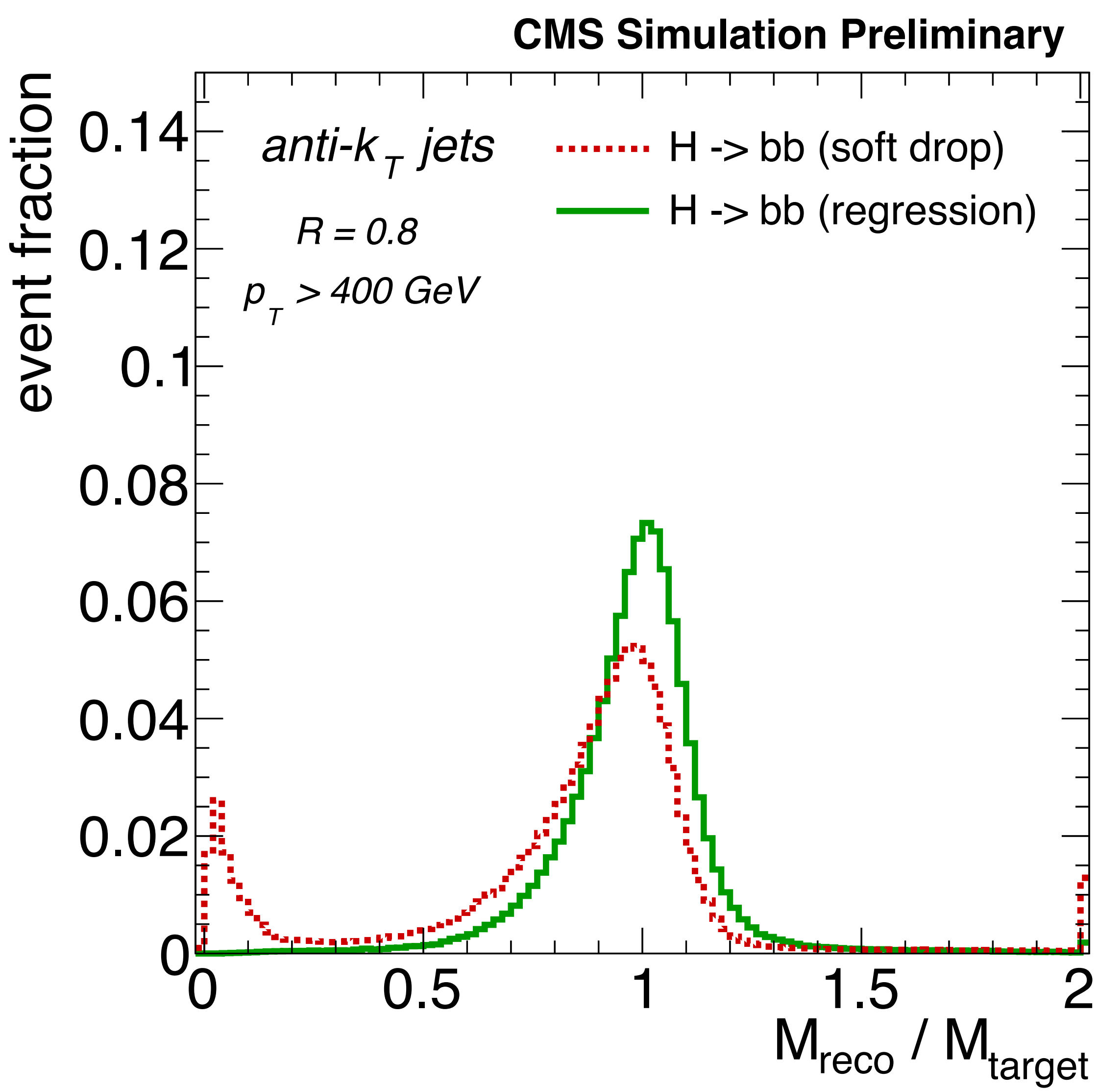
https://www.cs.cornell.edu/courses/cs4780/2015fa/web/lecturenotes/lecturenote10.html

Mass Regression: Performance

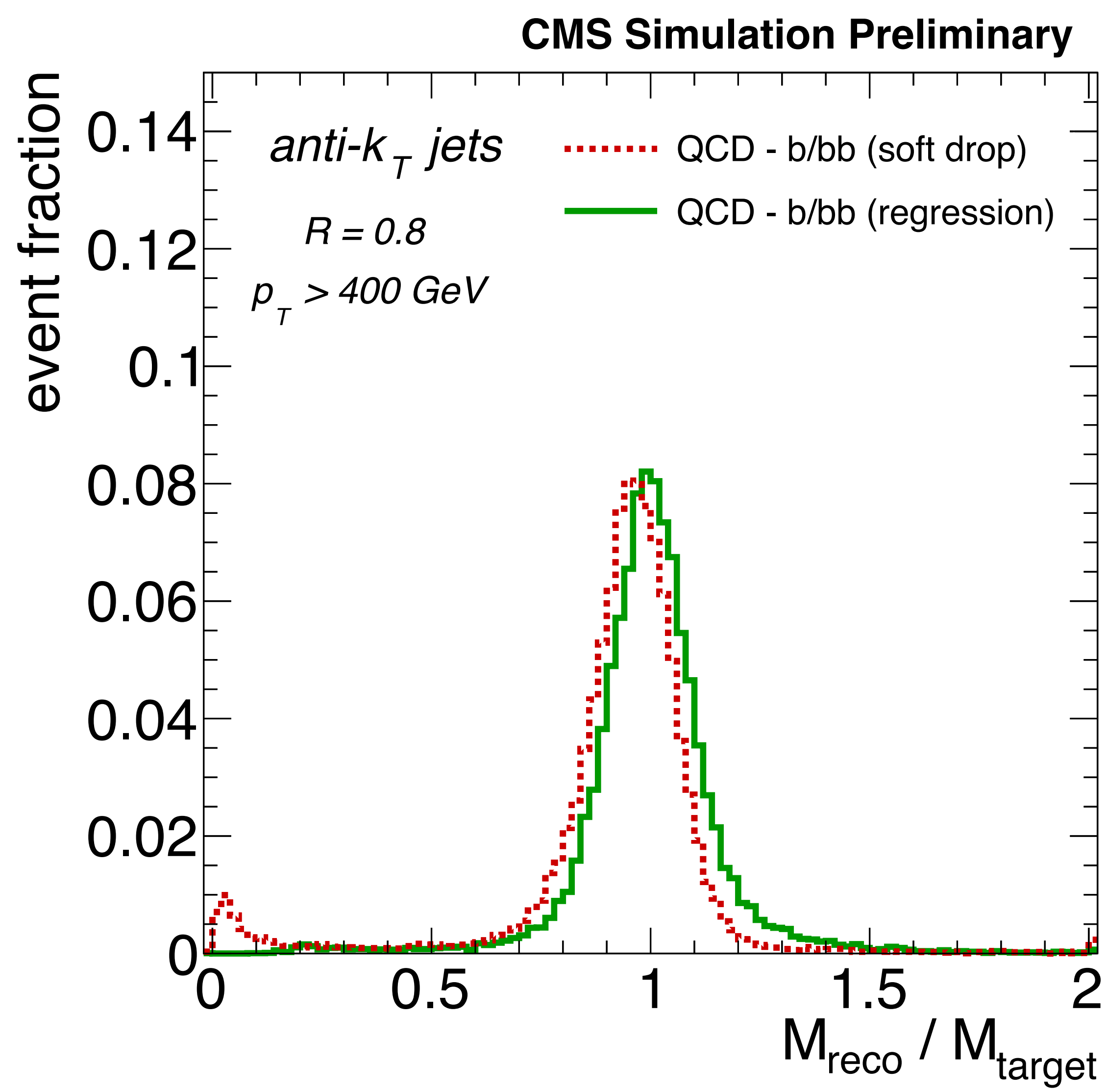
NEW

CMS DP-2021/017

Signal jets: H→bb



Background jets: QCD

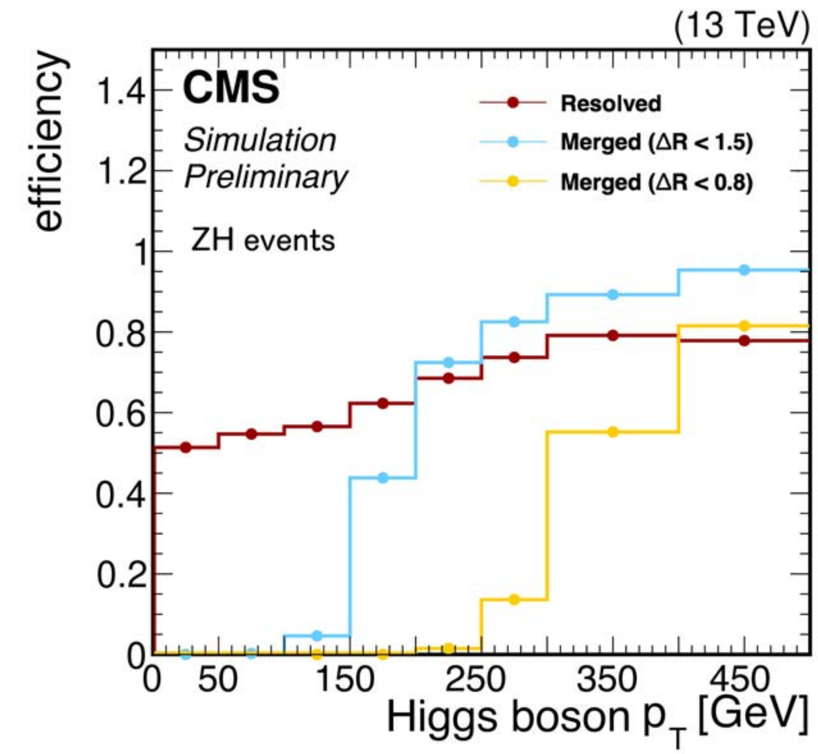


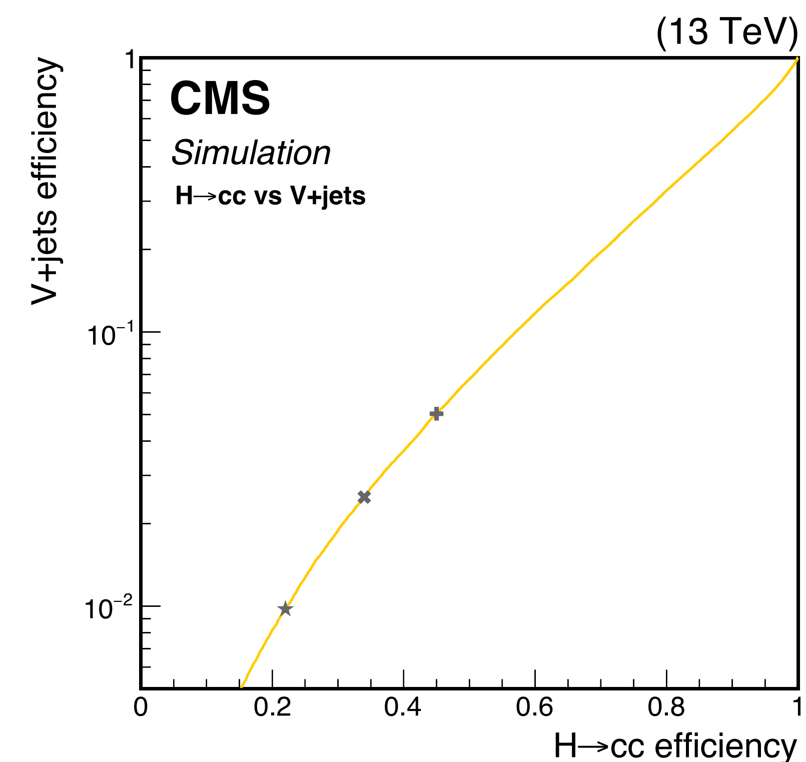
- Substantial improvement in both mass scale and resolution, especially for signal jets
- Tails in m_{SD} also significantly reduced
- Up to ~20-25% improvement in analysis sensitivity with H->bb/cc

VH(H→cc)

JHEP 03 (2020) 131

- First direct search for H→cc in CMS
  - VH channel: V (W, Z) → ll, lν, νν
  - two complimentary approaches to fully explore the H→cc decay topologies
- Resolved-jet topology
  - H→cc decay reconstructed with two resolved jets (R=0.4)
  - charm quark jets identified with DNN-based DeepCSV algorithm
  - analysis strategy similar to the VH(H→bb) analysis [PRL 121, 121801 (2018)]
    - fit to BDT shapes to extract the VH(H→cc) signal
- Merged-jet topology
  - H→cc decay reconstructed with one large-R jets
    - using **R=1.5** (instead of R=0.8) to increase acceptance at lower p_T (~200–300 GeV)
  - the **DeepAK8-MD** algorithm adapted to select cc-jet and suppress light-/bb-flavor jets
  - **fit to the mass of the large-R jet** (Higgs boson candidate) to extract the VH(H→cc) signal

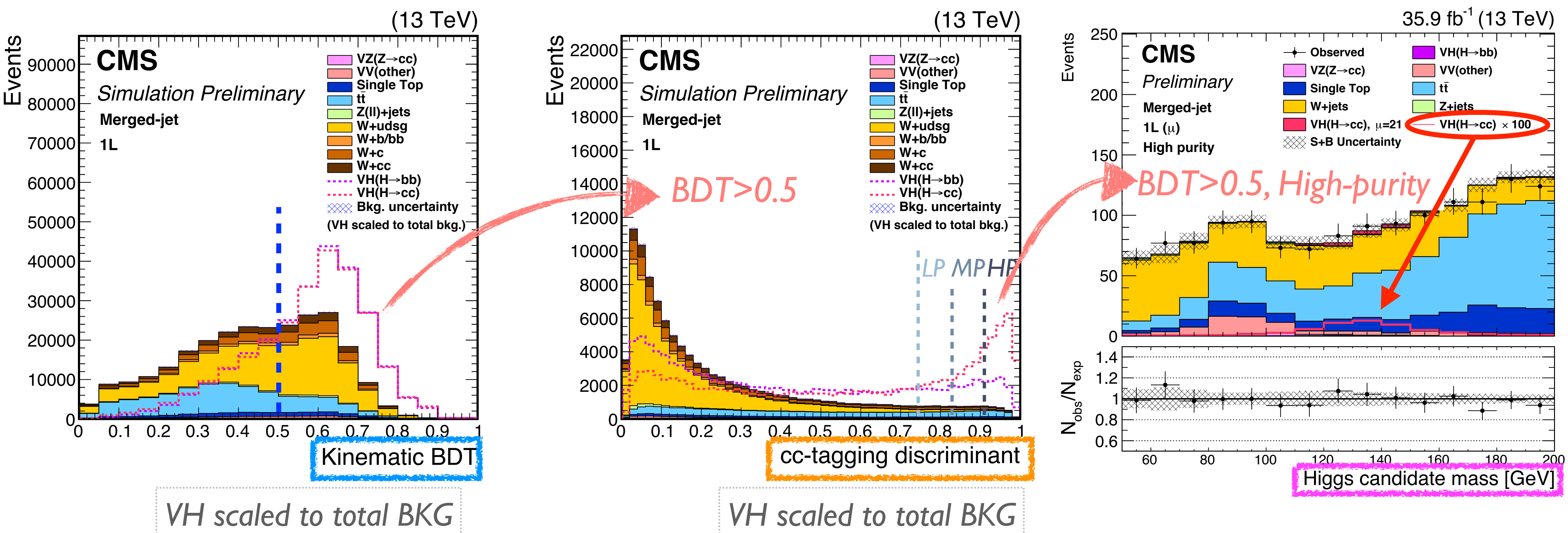




H→cc: Analysis Strategy

JHEP 03 (2020) 131

- Analysis strategy of the merged-jet topology
  - event-level kinematic BDT developed in each channel to better suppress the dominant backgrounds (V+jets, ttbar)
    - using only event kinematics, NOT the intrinsic properties (e.g., flavor/mass) of the Higgs candidate (H_{cand})
  - cc-tagging discriminant used to select cc-flavor jets and reject light/bb-flavor jets
  - distinct m(H_{cand}) shapes between signal and V+jets/ttbar background:
    - fit the m(H_{cand}) shape to extract the H→cc signal
- Kinematic BDT, cc-tagging discriminant and m(H_{cand}) largely independent of each other
  - allowing for a simple and robust strategy for background estimation and signal extraction



H→cc: Results

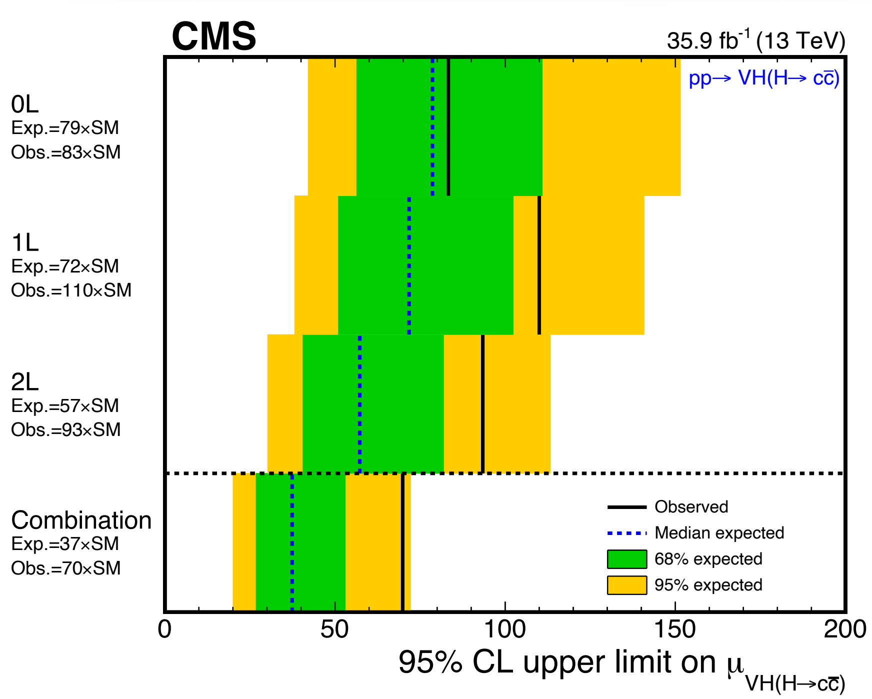
JHEP 03 (2020) 131

- Results from the two approaches combined for the final results
  - **resolved-jet topology** : $p_T(V)$ < 300 GeV; **merged-jet topology** : $p_T(V)$ > 300 GeV
  - using **35.9 fb^{-1}** data (2016)

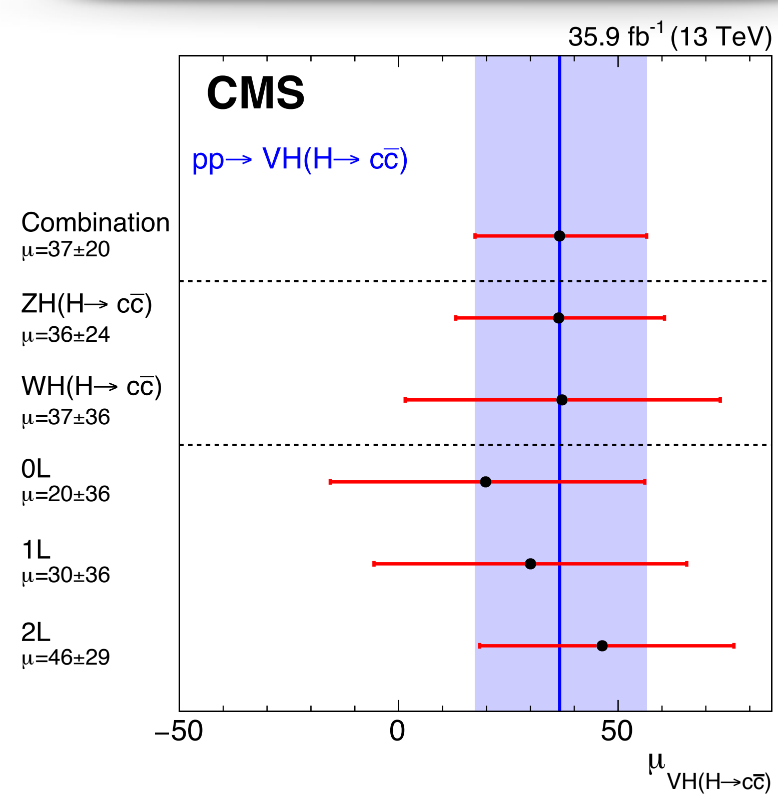
95% CL exclusion limit on $\mu_{VH(H\to c\bar{c})}$						
	Resolved-jet	Merged-jet	Combination			
	$(p_T(V) < 300\,\mathrm{GeV})$	$(p_T(V) \geq 300\,\mathrm{GeV})$	0L	1L	2L	All channels
Expected	45^{+18}_{-13}	73^{+34}_{-22}	79^{+32}_{-22}	72^{+31}_{-21}	57^{+25}_{-17}	37^{+16}_{-11}
Observed	86	75	83	110	93	70

- cf. ATLAS [ATLAS-CONF-2021-021, **139 fb^{-1}**]: $\mu_{VH(H\to cc)}$ < 26 (31) obs. (exp.)

Upper limit on $\mu_{VH(H\to cc)}$ at 95% CL



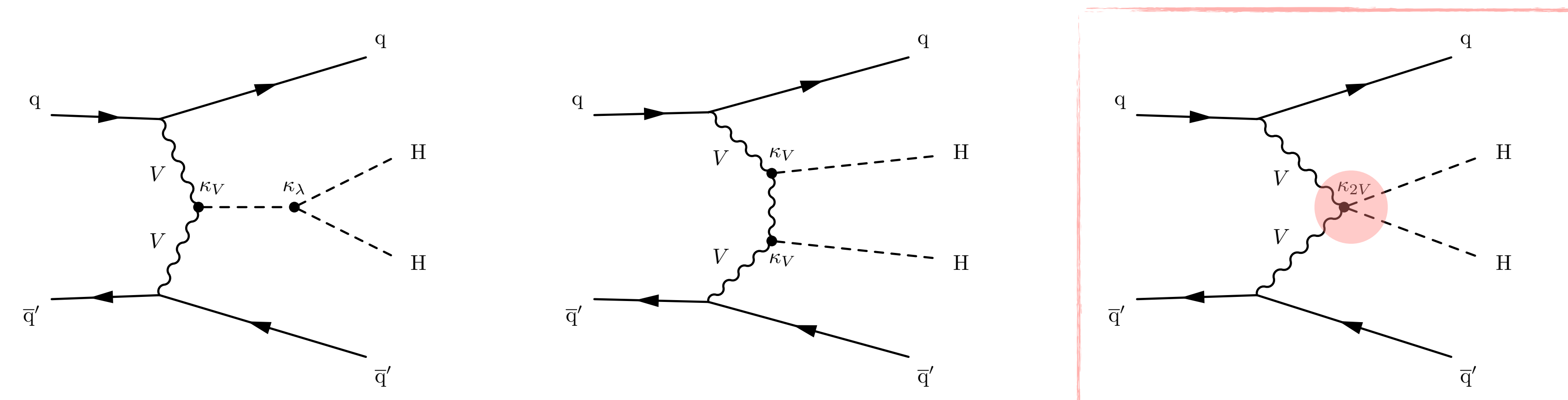
Best fit signal strength $\mu_{VH(H\to cc)}$



VBF HH(→4b)

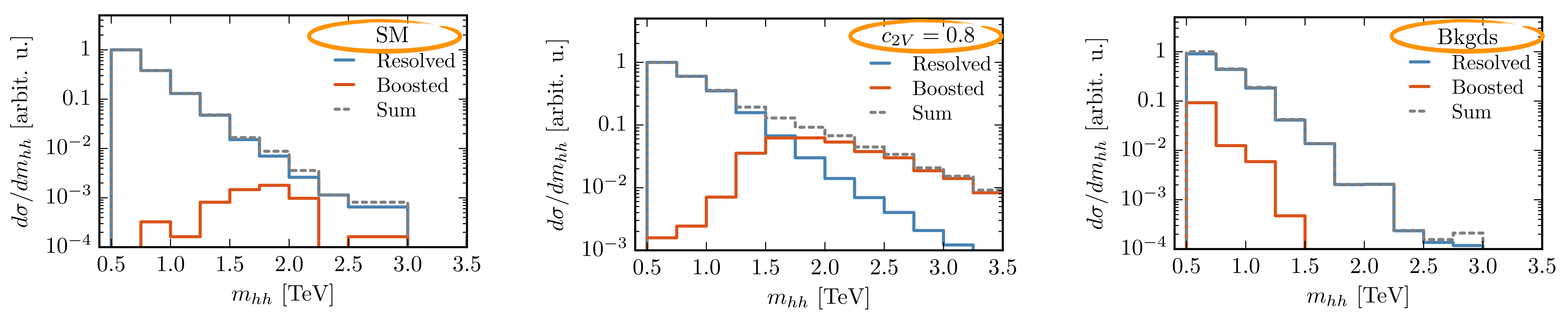
CMS-PAS-B2G-21-001

- VBF di-Higgs production: a unique channel to probe the hhVV quartic coupling (κ_{2V})



  - very rare process in SM: σ ~ 1.7 fb
  - however, if the hhVV coupling deviates from the SM ($\kappa_{2V} \neq 1$), the cross section can be enhanced
    - meanwhile, a significant fraction of signal becomes highly boosted -> enhanced sensitivity using boosted objects for Higgs boson reconstruction

Eur.Phys.J.C 77 (2017) 7, 481

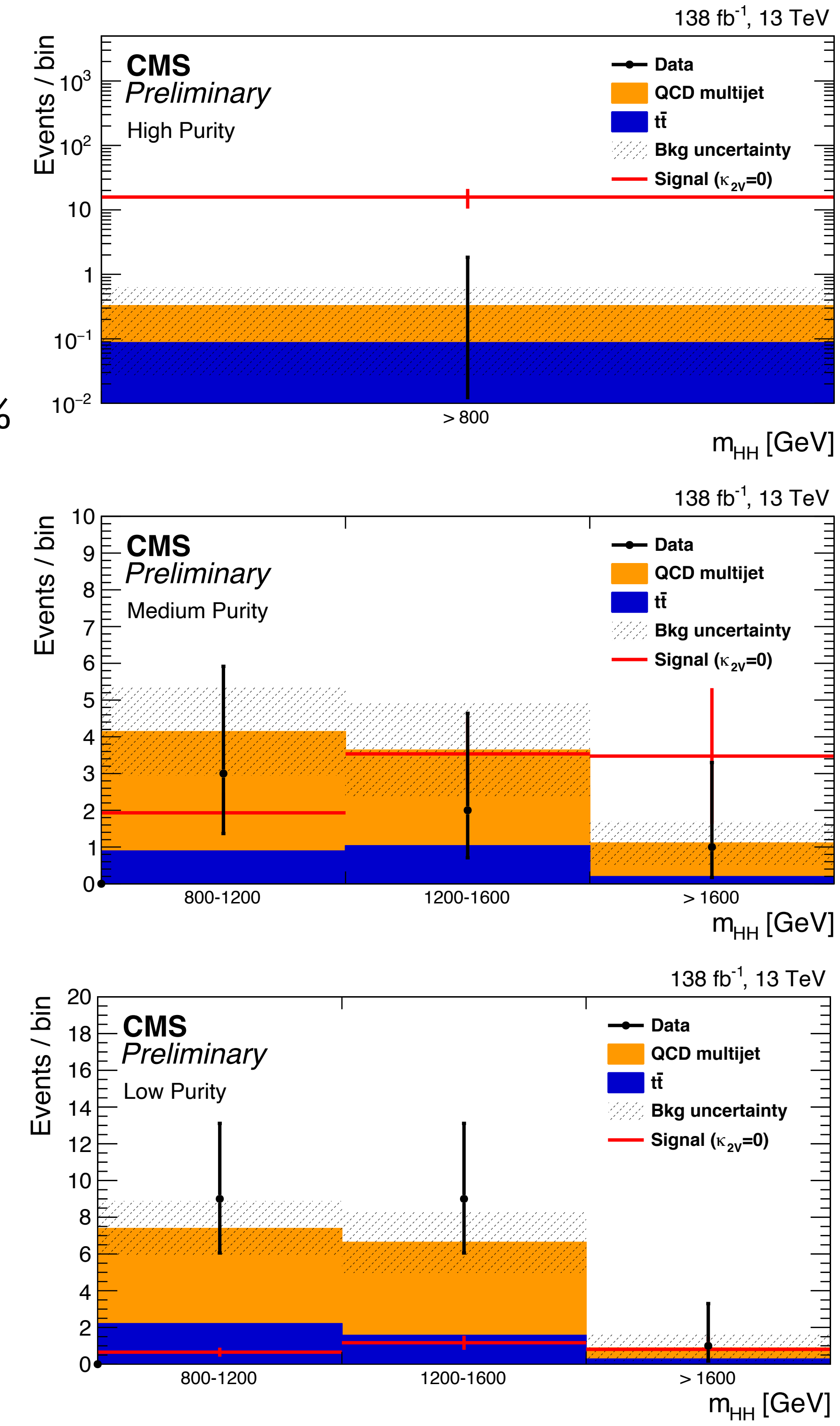


VBF HH(→4b): Analysis Strategy

NEW

- First search for non-resonant VBF HH production **in the boosted topology**
- Analysis strategy
  - Higgs bosons reconstructed as two **high p_T AK8 jets** (p_T > 500/400 GeV)
  - **H→bb tagging** with the ParticleNet algorithm
    - 3 WPs: signal efficiencies ~ 60%, 80%, 90% at QCD mis-id. rate ~0.3%, 1% and 2%
  - Higgs jet mass reconstructed with the **ParticleNet mass regression** (~20% improvement w.r.t soft drop algorithm)
  - selection of **VBF topology**: two AK4 jets with dijet mass > 500 GeV and |Δη| > 4
- Background estimation
  - ttbar background estimated from simulation, with corrections derived from a top-enriched region
  - QCD multijet background estimated with a data-driven method
    - using QCD-enriched "fail" region by inverting the ParticleNet bb-tagging selections
- Signal extraction
  - by fitting to m_{HH} in three search categories of increasing purity

CMS-PAS-B2G-21-001

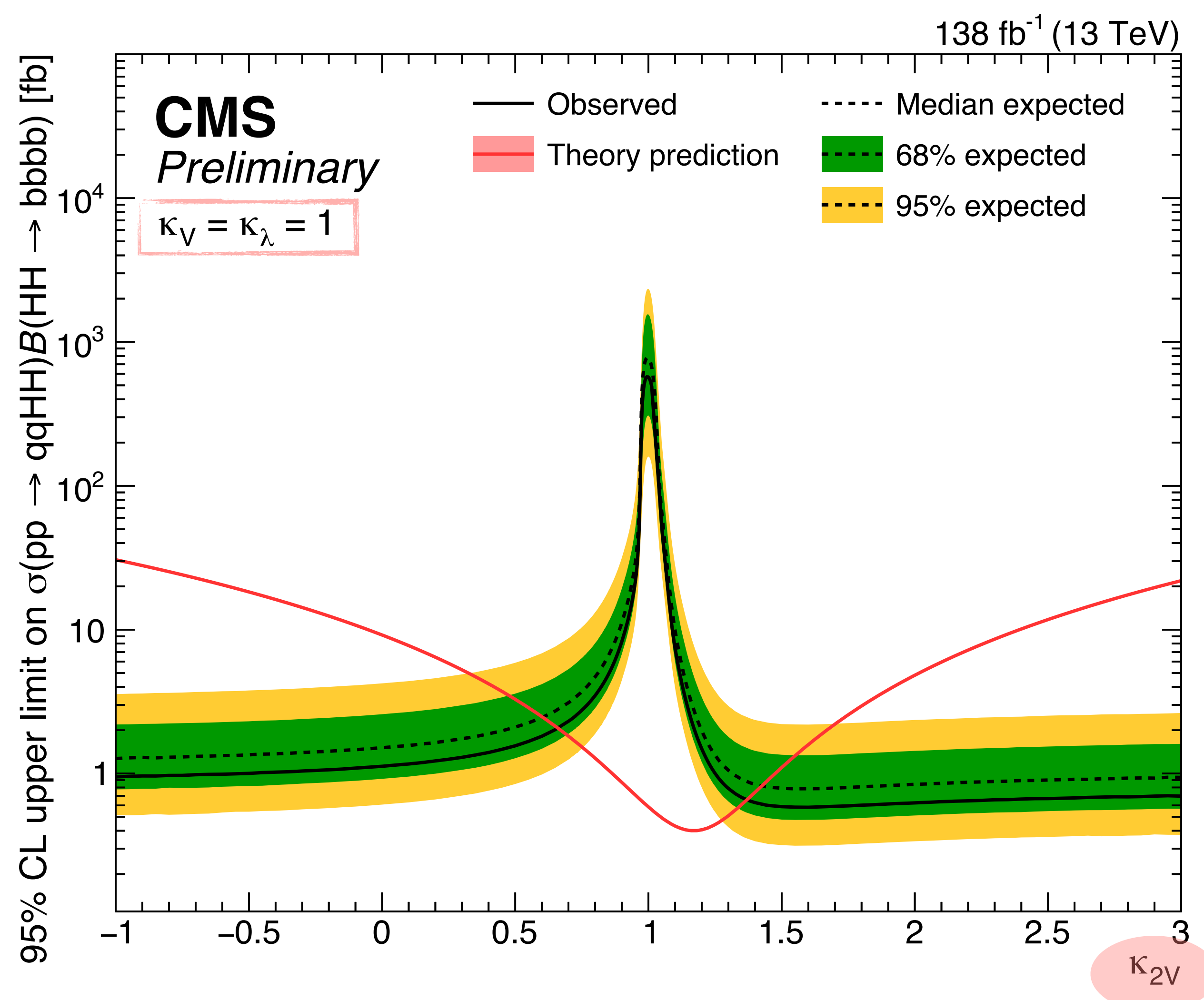


VBF HH(→4b): RESULTS

NEW

CMS-PAS-B2G-21-001

Upper limits on σ(VBF HH) × B(HH→bbbb)



- Most stringent constraint on κ_{2V} to date: **$0.6 < \kappa_{2V} < 1.4$**
  - $\kappa_{2V} = 0$ **excluded for the first time!**
  - cf. ATLAS [JHEP 07 (2020) 108]: -0.43 (-0.55) < κ_{2V} < 2.56 (2.72) obs. (exp.)

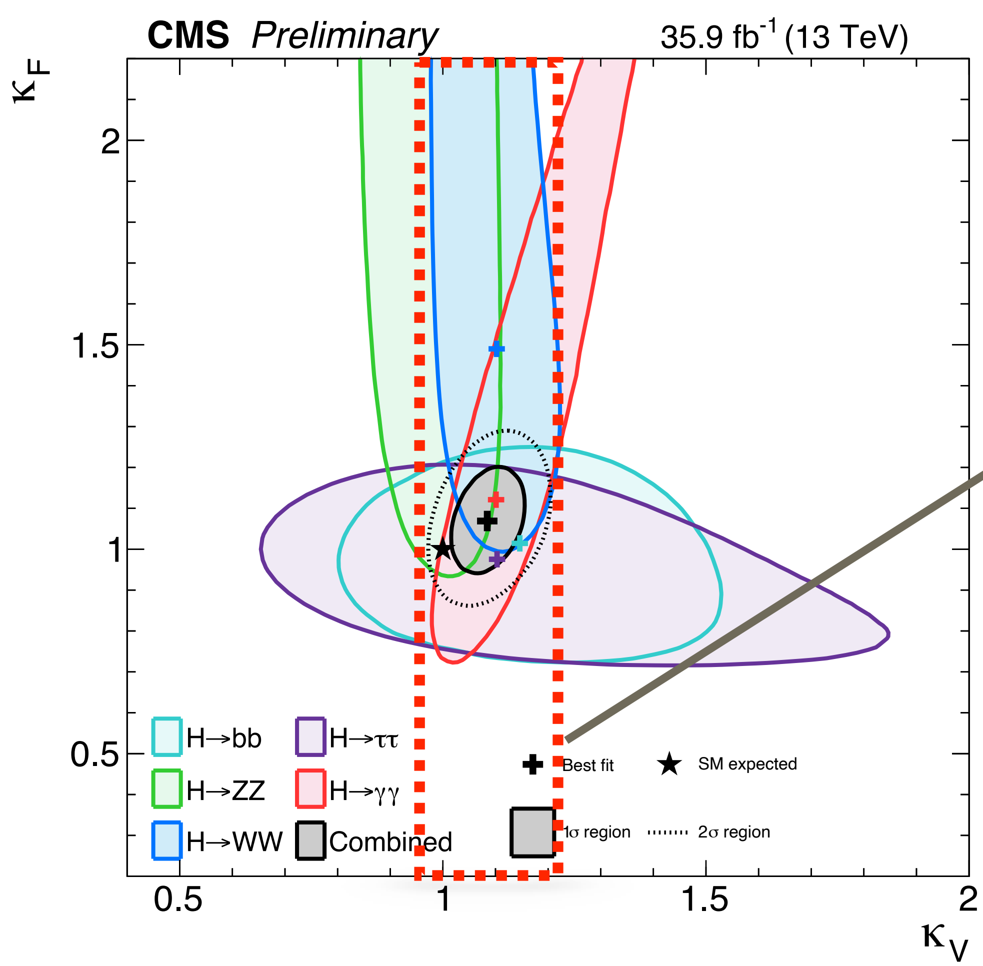
SUMMARY

- Lots of progress in boosted object techniques in recent years
  - substantial performance improvements with the introduction of novel machine learning-based approaches
  - performance gains confirmed in real data, and led to significantly increased sensitivity in relevant analyses
- Advances in boosted object techniques brought new opportunities for Higgs physics
  - measurement of the Higgs couplings, complementary to the resolved-jet approach
    - VH(H→cc) [JHEP 03 (2020) 131]
    - VBF HH(→4b) [CMS-PAS-B2G-21-001]
  - probing Higgs boson production in the boosted regime [JHEP 12 (2020) 085]
  - search for new resonances decaying into Higgs bosons [CMS-PAS-B2G-20-007, CMS-PAS-B2G-20-004, …]
  - … and more to come!

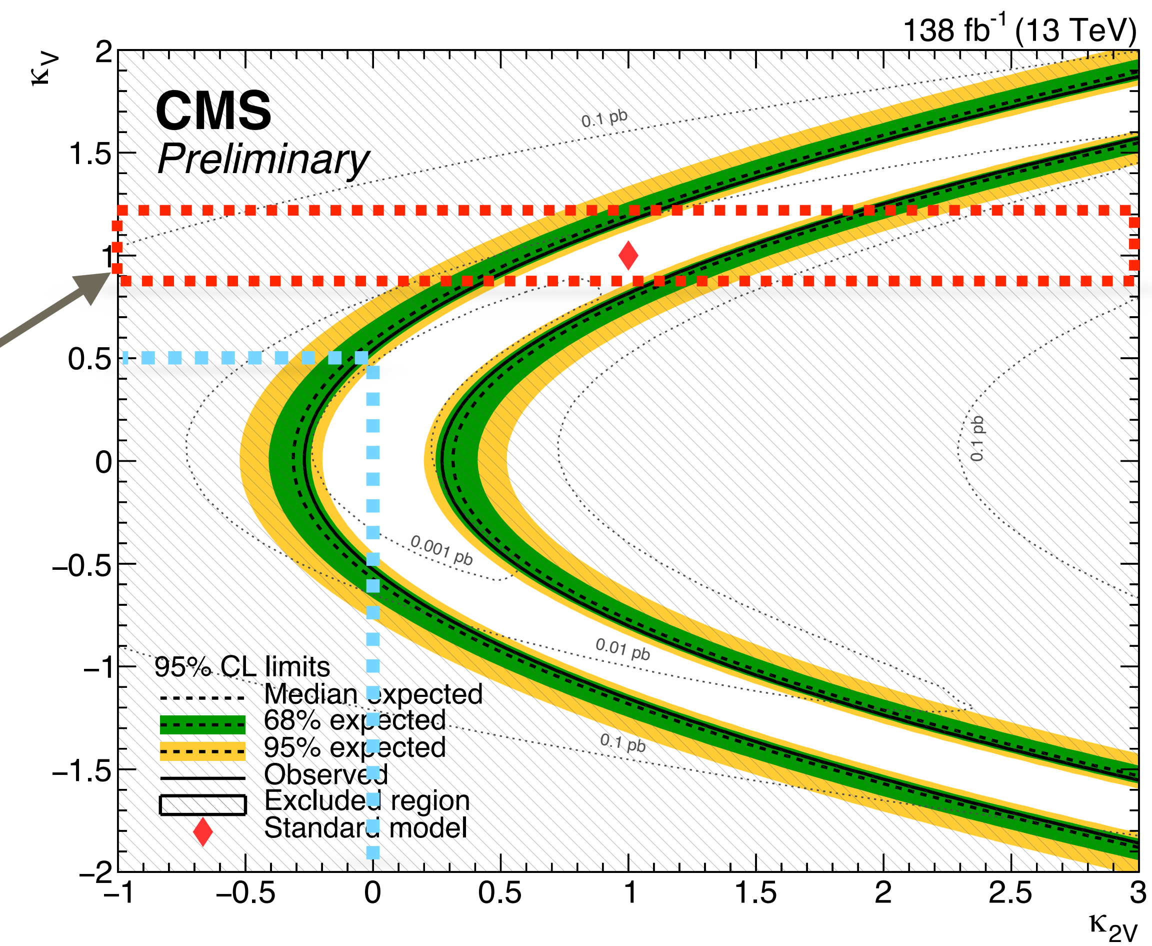
Backups

VBF HH(→4b): RESULTS

Eur. Phys. J. C 79 (2019) 421

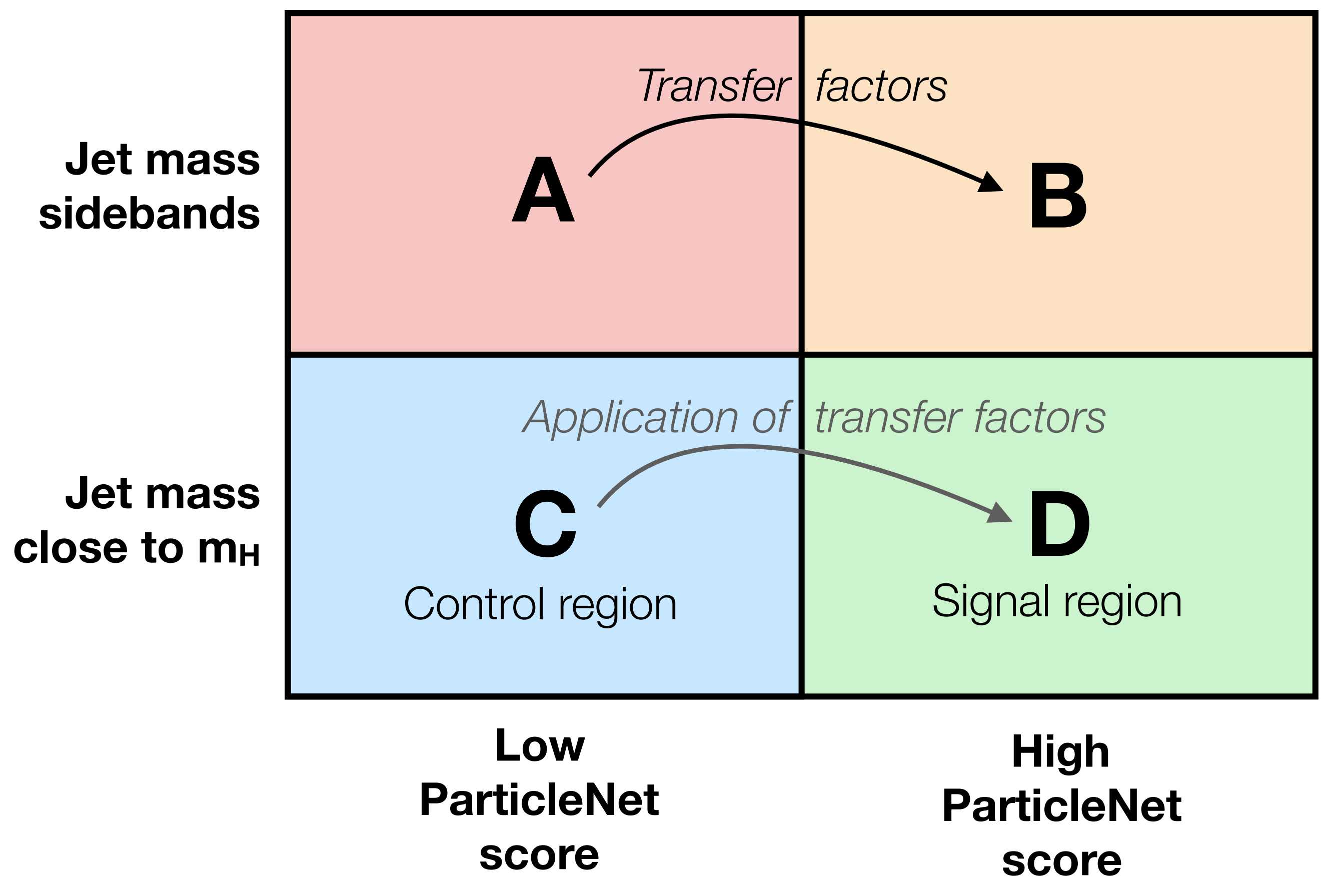


CMS-PAS-B2G-21-001



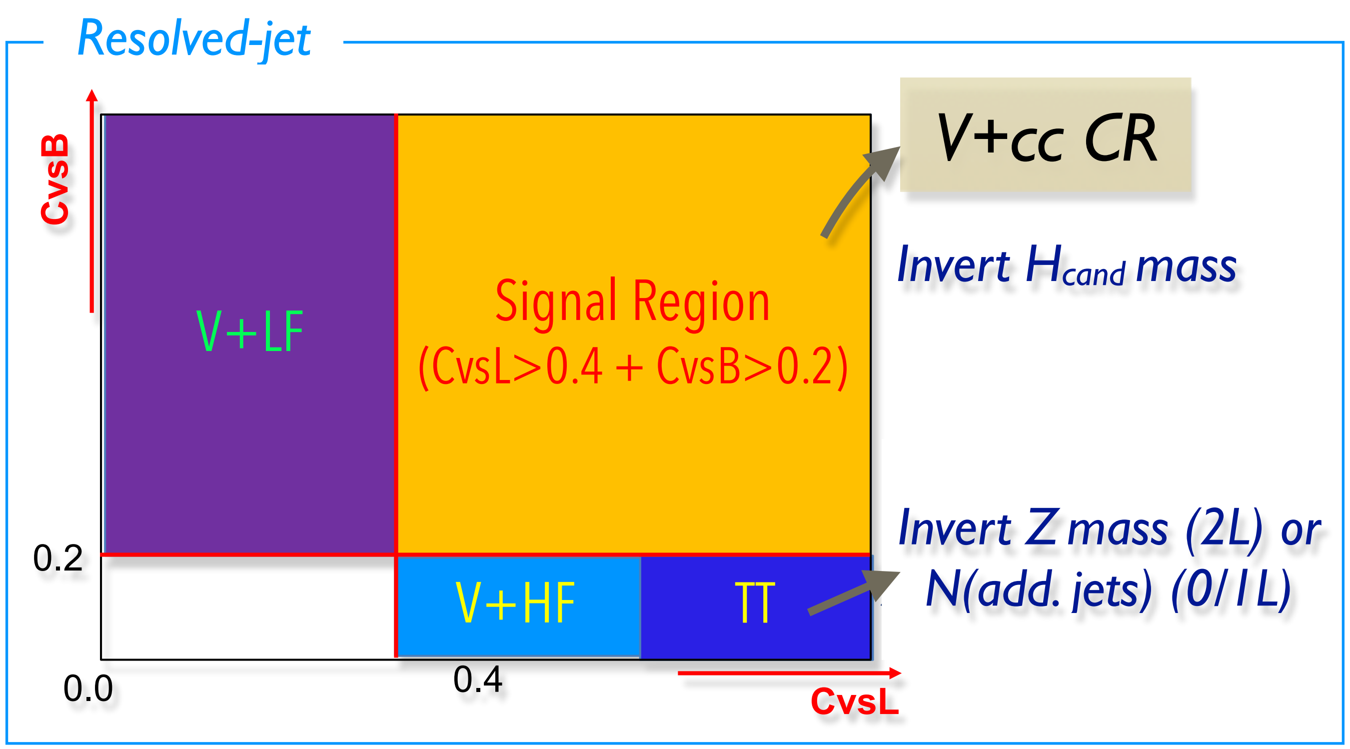
2D limit scan excludes $\kappa_{2V} = 0$ for $\kappa_V > 0.5$
(with other couplings fixed to SM values)

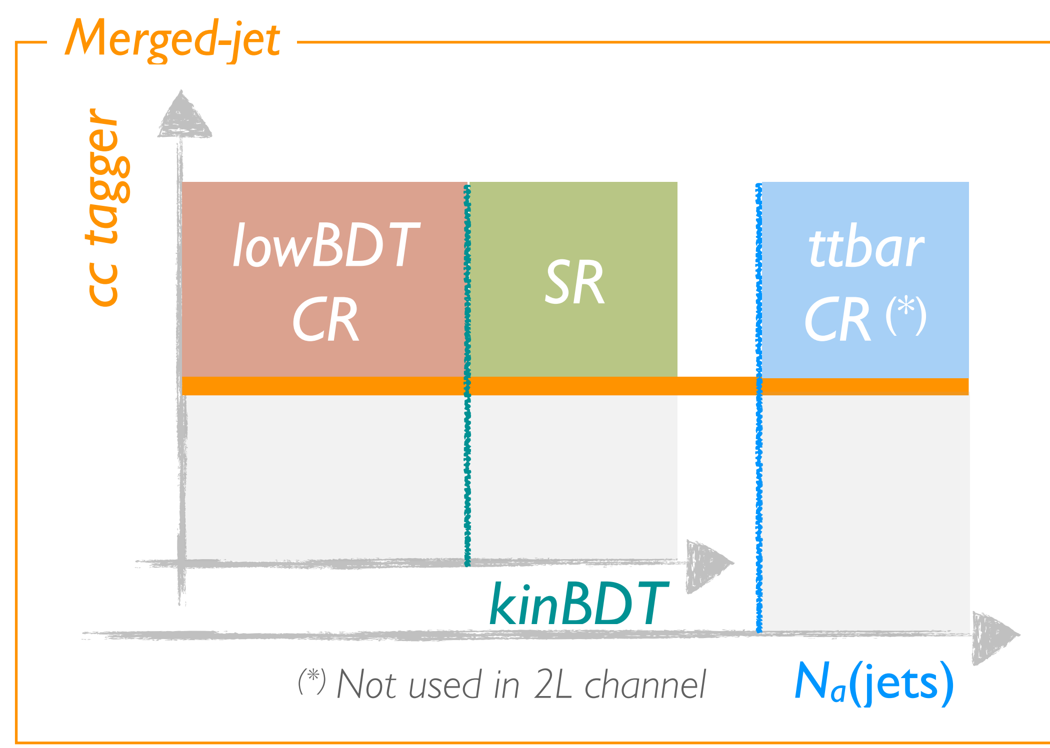
VBF HH(→4b): Background Estimation



H→cc: Fit Strategy

- Dedicated control regions are set up to measure the normalizations of major backgrounds (W/Z+jets, ttbar)
  - simultaneous fit of signal regions and control regions to constrain BKGs and extract the signal





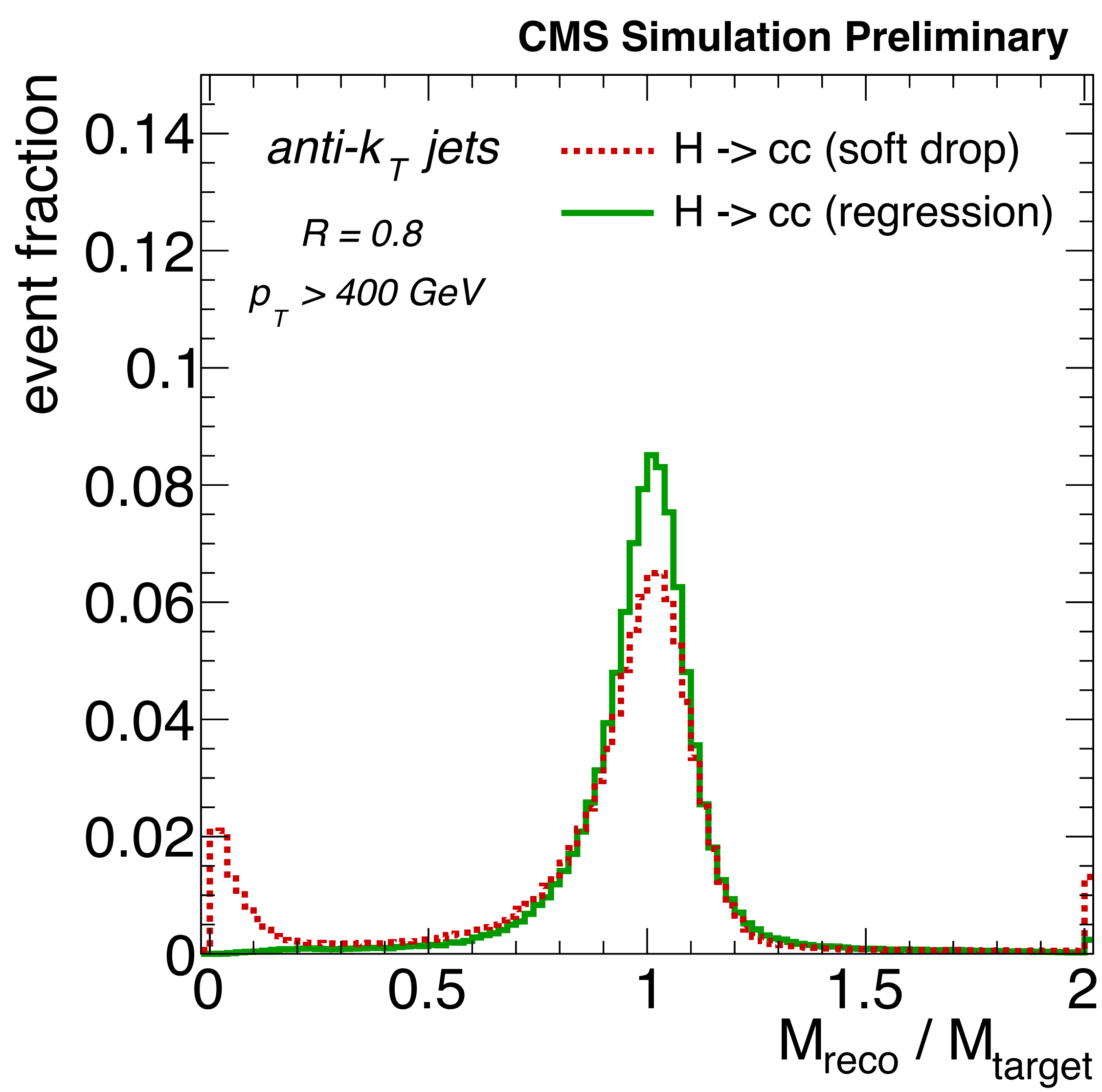
- Dominant sources of uncertainties:
  - size of the MC simulation / data control samples
  - charm tagging efficiencies
  - simulation modeling

Uncertainty source	$\Delta\mu \mid \mu = 37$	
Statistical	+17.3	−17.1
Background normalisations	+10.1	−10.2
Experimental	+7.6	−8.2
Charm tagging efficiencies	+5.6	−4.8
Simulation modeling	+4.2	−5.1
Jet energy scale and resolution	+2.4	−2.8
Lepton identification efficiencies	+0.4	−1.8
Luminosity	+1.6	−1.7
Statistics of the simulated samples	+0.5	−1.9
Theory	+6.5	−4.6
Signal	+5.0	−2.5
Backgrounds	+4.3	−3.9
Total	+20.0	−19.5

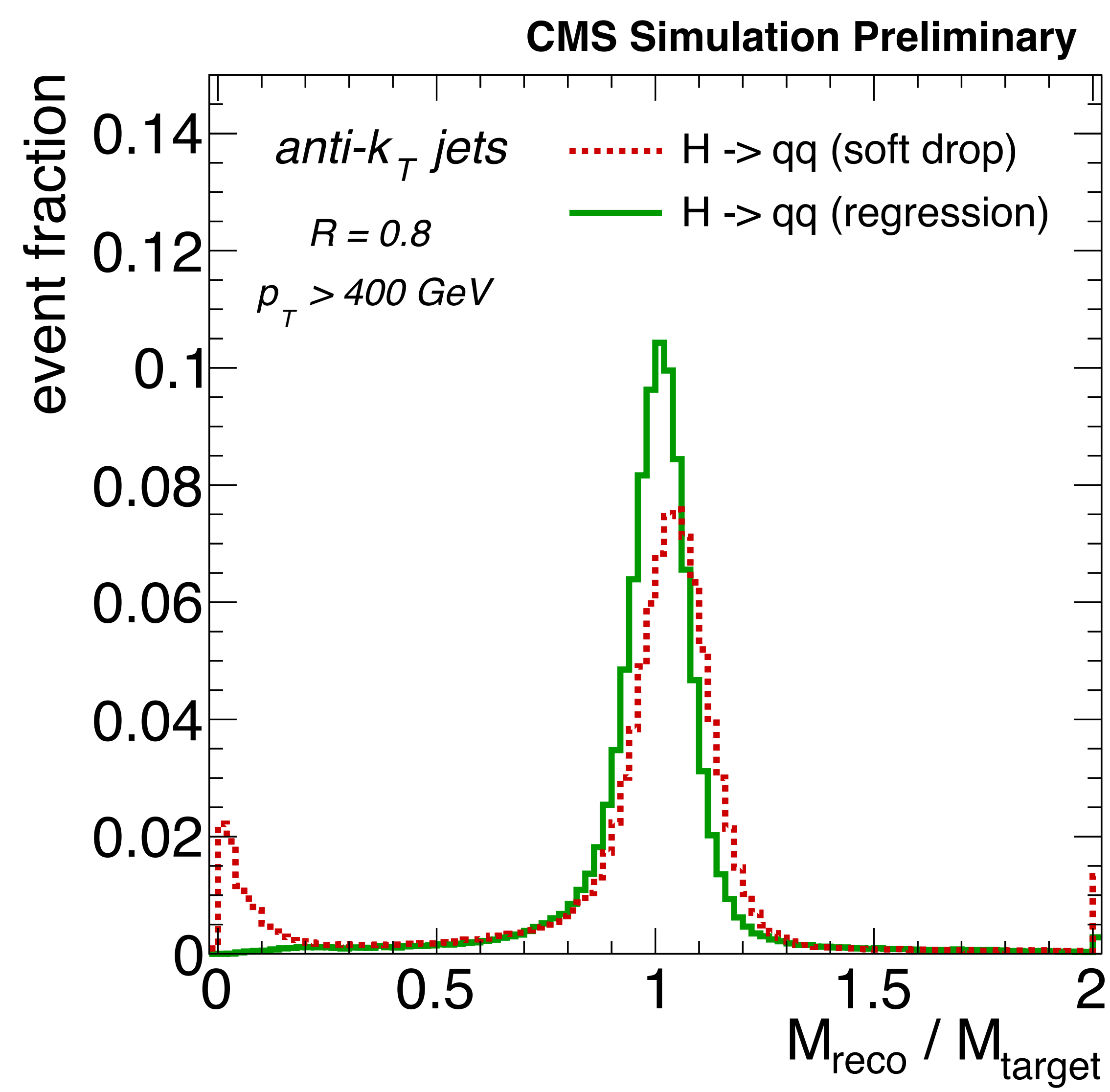
Mass Regression: Performance (II)

CMS DP-2021/017

Signal jets: H→cc



Signal jets: H→qq

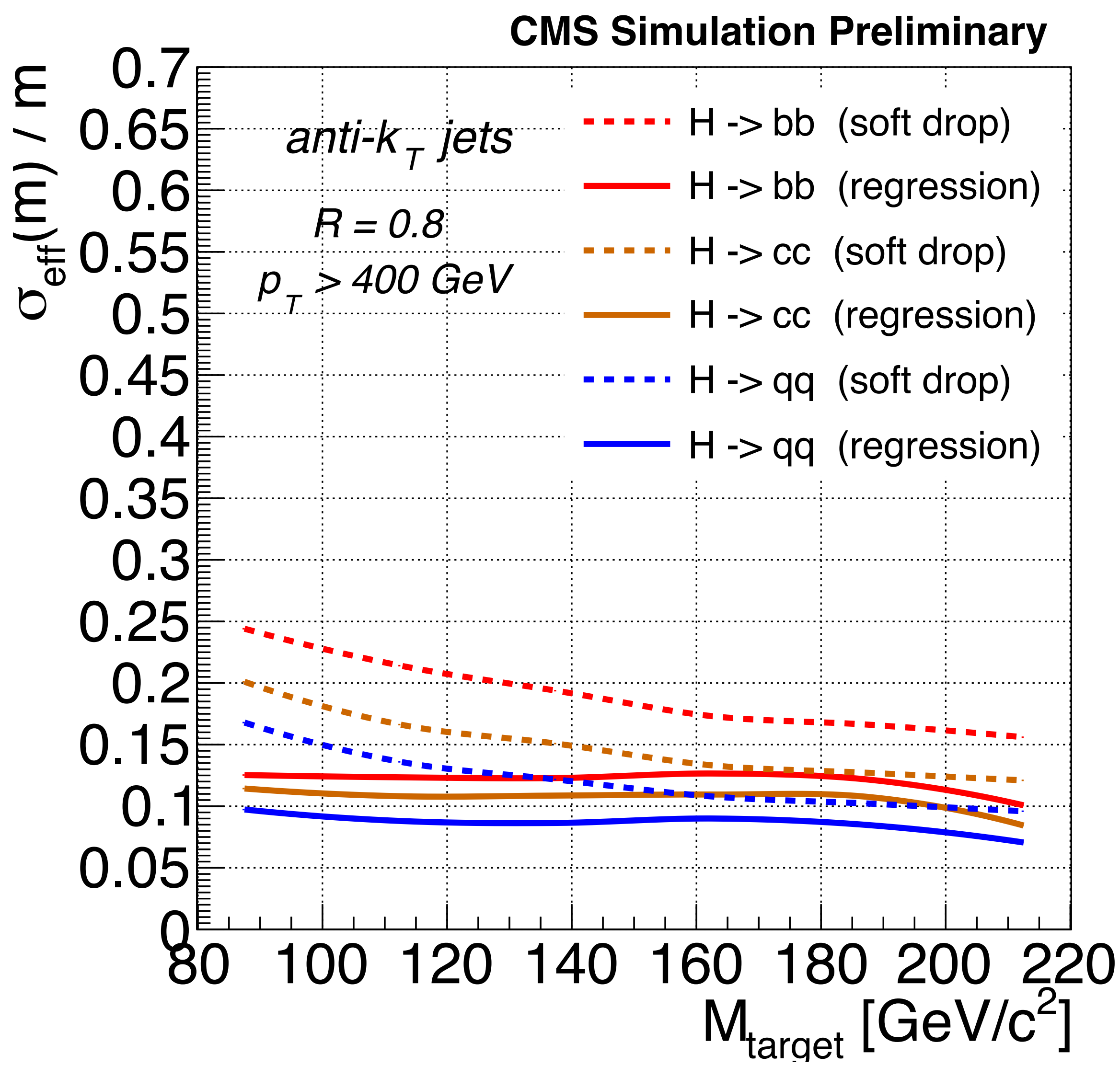


- Consistent improvements in all jet flavors

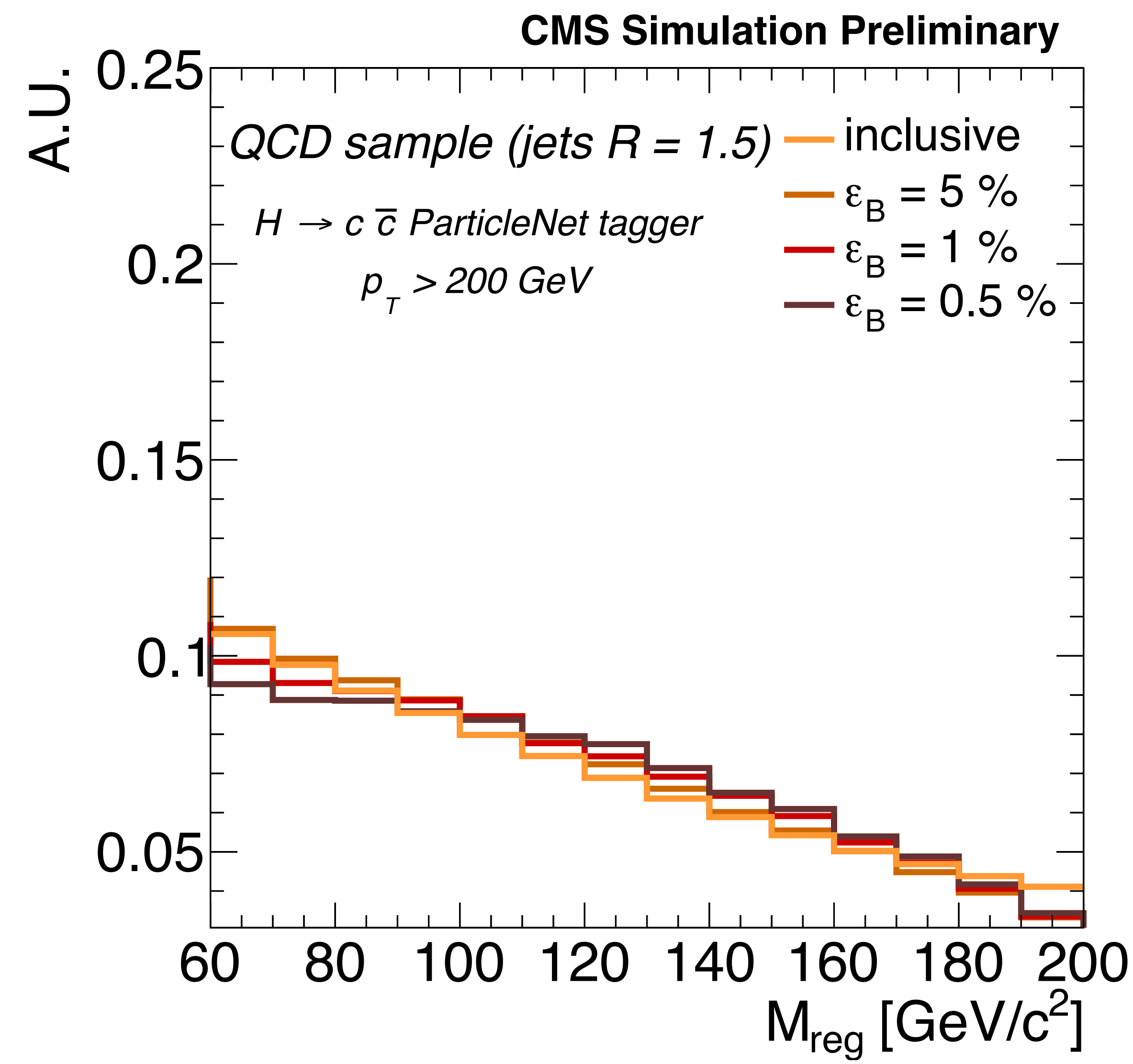
Mass Regression: Performance (III)

CMS DP-2021/017

Mass resolution vs m_X



Regressed mass vs Tagger WP



- Mass resolution more stable vs m_X compared to soft drop
- No signs of mass sculpting – even for very tight tagger selections
- Up to ~20-25% improvement in analysis sensitivity with H->bb/cc

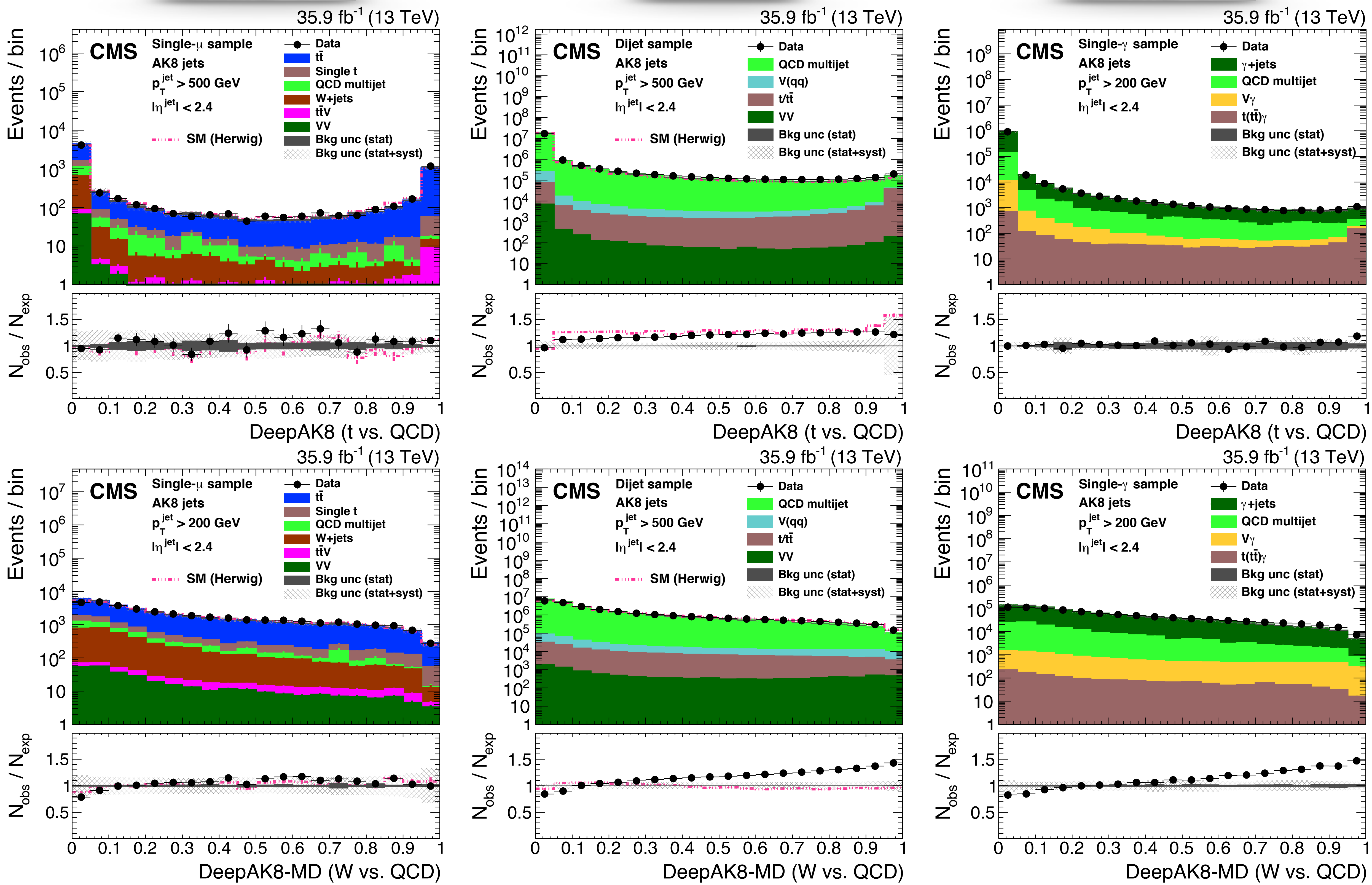
PERFORMANCE IN DATA

JINST 15 (2020) P06005

Single-μ sample

Dijet sample

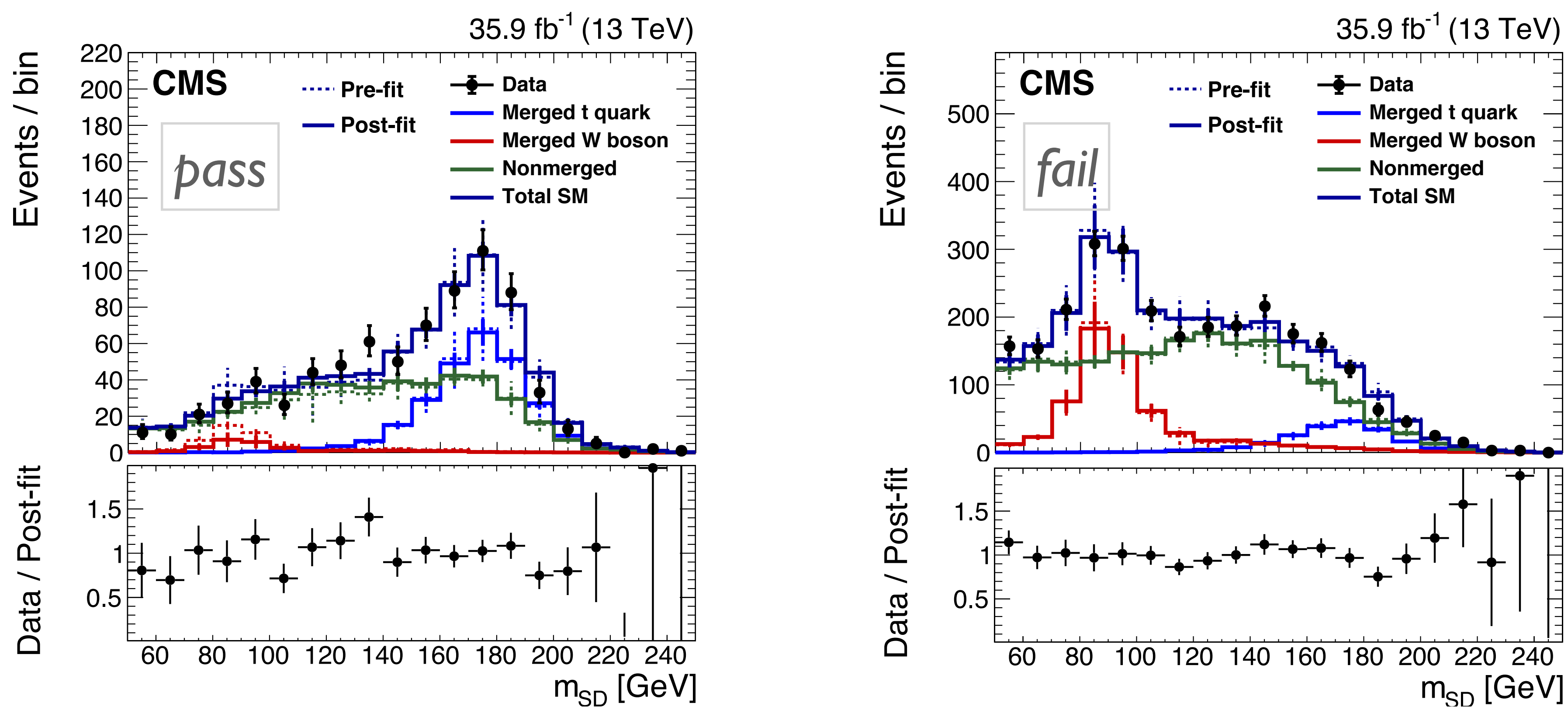
Single-γ sample



TAGGER CALIBRATION IN DATA

- Crucial to calibrate these taggers in real data for them to be used in analyses
- Top/W tagging efficiency

JINST 15 (2020) P06005



  - measured using the single-μ sample enriched in semi-leptonic ttbar events
  - fit jet mass templates in the "pass" and "fail" categories simultaneously to extract efficiency in data
    - simulation-to-data scale factors SF := eff(data) / eff(MC) derived to correct the simulation
  - jet mass scale and resolution scale factors can also be extracted
- H->bb/H->cc tagging efficiency: measured via proxy jets, gluon->bb/cc, using a di-jet sample
- Mistag rates of background jet typically derived directly from analysis-specific control regions