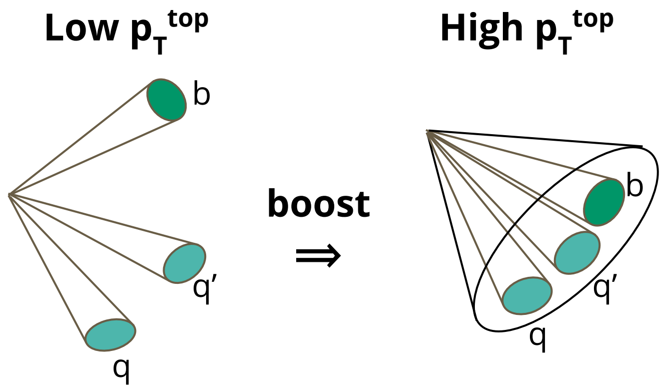
Measurements involving tagging techniques in ATLAS and CMS

Trisha Farooque (ATLAS),
Benedikt Maier (CMS)

Introduction

- Standard Model measurements at the LHC are reaching unprecedented levels of precision and extending their reach to previously unexplored regions of phase space
- Measurements in the high-pT regime are especially interesting because of sensitivity to new physics
  - Highly boosted heavy resonances (top quarks, H/W/Z bosons) appear in final states
  - Require employment of boosted object taggers with good **signal efficiency** and **background rejection**
- Collimation of final state objects can simplify combinatorics in event reconstruction compared to traditional, resolved topologies

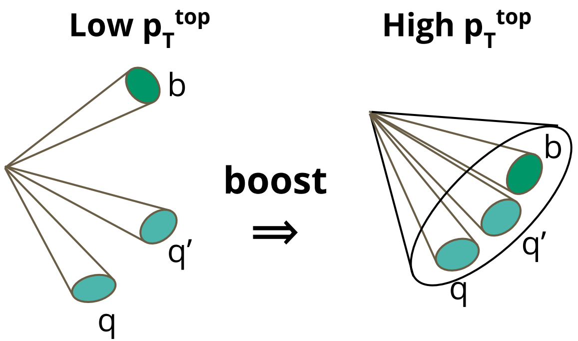


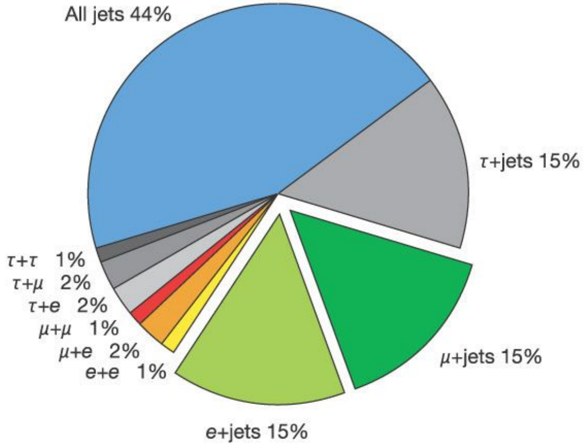
ATLAS Measurements

Measurements involving boosted top quarks

ttbar measurements in boosted topologies

- ttbar events characterised by large multiplicity of objects in the final state (multiple jets, b-jets, leptons, MET)
    - Challenging event reconstruction
- Boosted topologies reduce combinatorics and simplify top quark reconstruction
    - Especially important in the all-hadronic channel
- Boosted tagging techniques also essential to access high-pT regime
    - Study tails sensitive to new physics effects

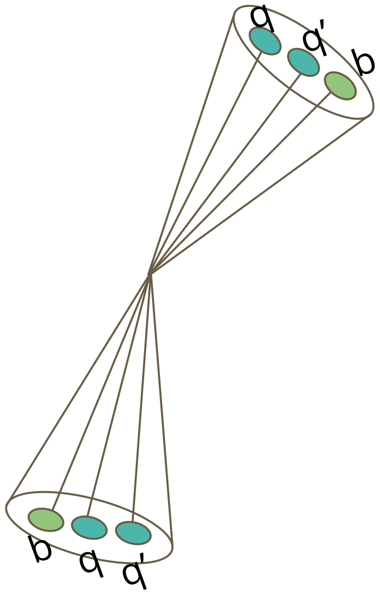


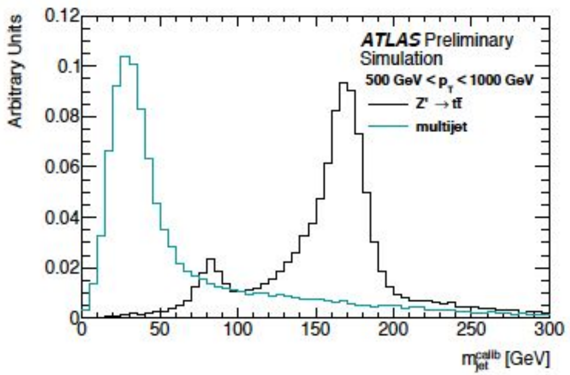


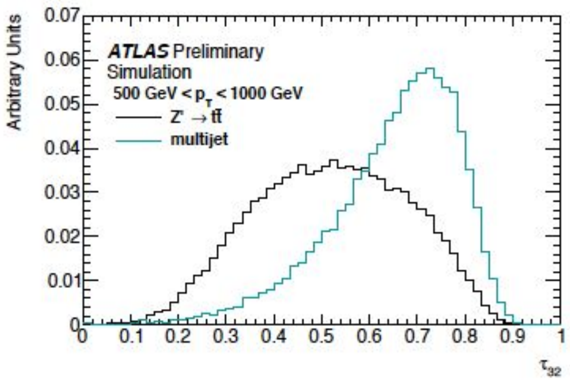
Measurement of ttbar production (0-lepton)

arXiv:1801.0205

- All-hadronic ttbar channel
  - Largest tt branching fraction (when including hadronic τ decays)
  - Swamped by multijet background
  - Large combinatorics in resolved state (6 jets, 2 b-jets)
- Boosted top quark reconstruction can be used to improve combinatorial background
- Tag boosted top quark jets with substructure observables
  - anti-kT R=1.0 jets built from locally calibrated topoclusters in calorimeter
  - pT-dependent cuts on mass and τ_{32}



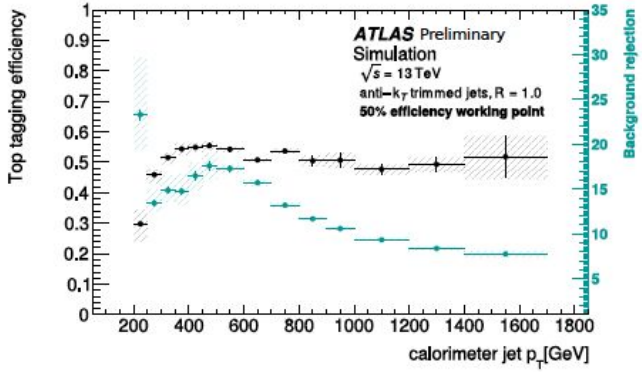




$\tau_{32} = \tau_3/\tau_2$

- N-subjettiness ratio
- Measure of 3-prongedness of jet

Measurement of tt production (0-lepton)



- All-hadronic channel suffers from very large multijet background
  - Essential to have high-purity tagger
- 50% efficiency working point chosen for top tagger
- 70% efficiency MVA-based b-tagger

- Events selected with >=2 anti-kT R=1.0 jet with pT>350 GeV
  - pTlead > 500 GeV, |mJ - mtop| < 50 GeV
- >=2 anti-kT R=0.4 jets
- Main background: multijet production
  - Estimated with 2D sideband (extended ABCD) technique based on top-tagging and b-tagging state of two leading jets in event
  - Weak correlations between tagging states measured in data

2nd large-R jet				
1t1b	J (7.6%)	K (21%)	L (42%)	S
0t1b	B (2.2%)	D (5.8%)	H (13%)	N (47%)
1t0b	E (0.7%)	F (2.4%)	G (6.4%)	M (30%)
0t0b	A (0.2%)	C (0.8%)	I (2.2%)	O (11%)
	0t0b	1t0b	0t1b	1t1b

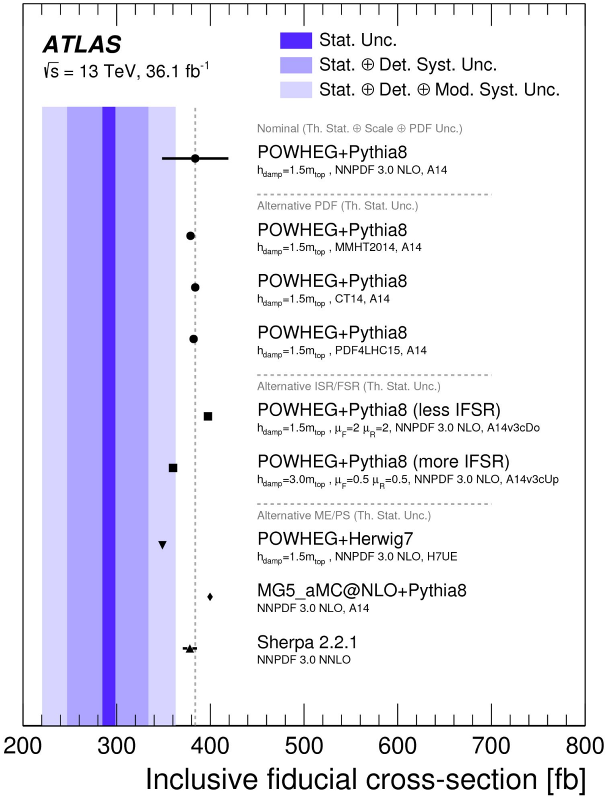
Leading large-R jet

$$S = \frac{J \times O}{A} \cdot \frac{D \times A}{B \times C} \cdot \frac{G \times A}{E \times I} \cdot \frac{F \times A}{E \times C} \cdot \frac{H \times A}{B \times I},$$
$$= \frac{J \times O \times H \times F \times D \times G \times A^3}{(B \times E \times C \times I)^2},$$

arXiv:1801.0205

arXiv:1801.0205

Measurement of ttbar production (0-lepton)



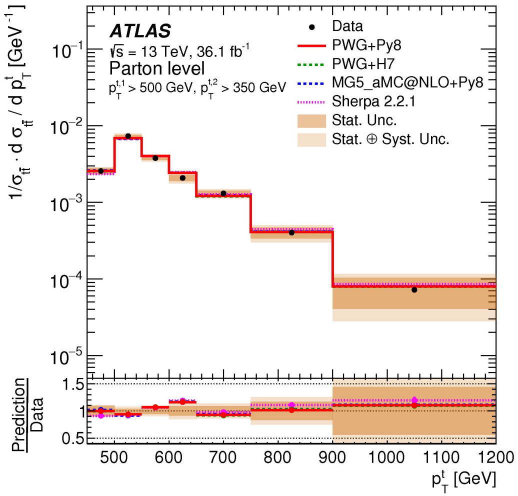
- Modelling uncertainties on ttbar production and boosted top tagging calibration uncertainties are the dominant systematics

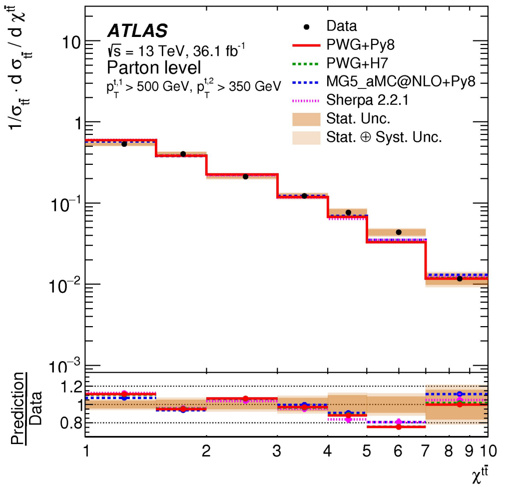
Source	Percentage
Large-R jet energy scale	5.9
Large-R jet mass calibration	1.4
Large-R jet top-tagging	12
Small-R jets	0.3
Pileup	0.6
Flavor tagging	8.3
Background	0.9
Luminosity	2.0
Monte Carlo statistical uncertainty	0.9
Alternative hard-scattering model	11
Alternative parton-shower model	14
ISR/FSR + scale	1.1
Total systematic uncertainty	24
Data statistical uncertainty	2.3
Total uncertainty	24

- Inclusive fiducial cross section compared to several different ttbar models
- Sensitive to variations in PDF, fragmentation, generator choice

Measurement of ttbar production (0-lepton)

- Differential cross sections measured with respect to several observables
  - Unfolded to parton level
- Good agreement between data and simulation across top pT range
- Some differences observed in angular observables such as χ_{tt}
  - Measure of rapidity difference between top quarks in event
  - Sensitive to new physics effects, e.g. contact interactions





$$y^{\star} = 1/2\left(y^{t,1} - y^{t,2}\right)$$

$$\chi^{t\bar{t}} = \exp 2|y^{\star}|$$

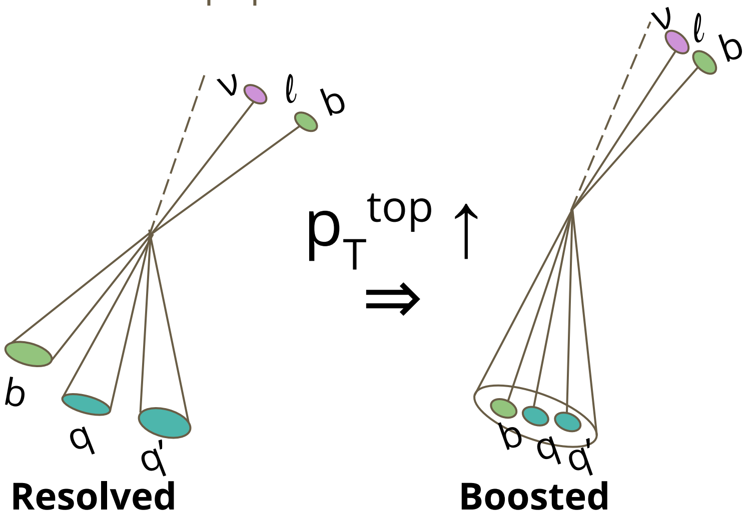
arXiv:1801.0205

Measurement of ttbar production (1-lepton)

arXiv:1908.07305

- tt production in ℓ+jets channel
  - Presence of lepton, MET, b-jets can be exploited to greatly improve signal purity
  - Lower combinatorial background makes resolved channel more accessible
- Boosted channel still essential to probe phase space with high-pT top quarks
  - Sensitive to effects from new physics
  - Complementary sensitivity to resolved channel
  - Provides alternative reconstruction method for hadronic top quark

- Measurement of inclusive and differential cross sections measured separately in both channels

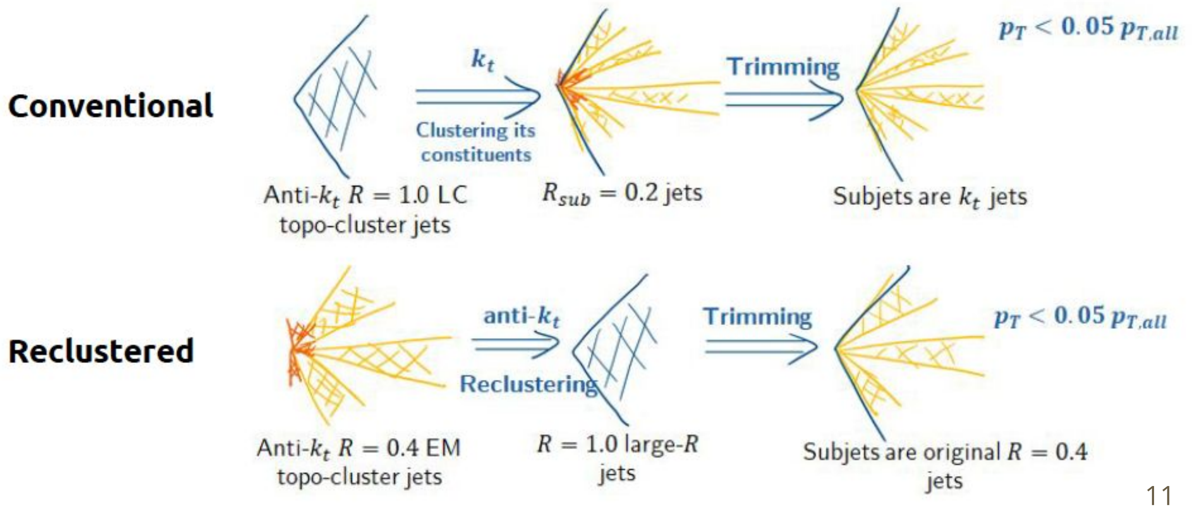


Measurement of ttbar production (1-lepton)

arXiv:1908.07305

- Low background level in ℓ + jets channel allows use of more efficient boosted top tagger
- **Jet re-clustering** employed to build large-radius R=1.0 jets using calibrated, R=0.4 jets directly as inputs
  - Calibration and uncertainties propagated directly from R=0.4 jets
  - Smooth transition between resolved and boosted channels without efficiency loss

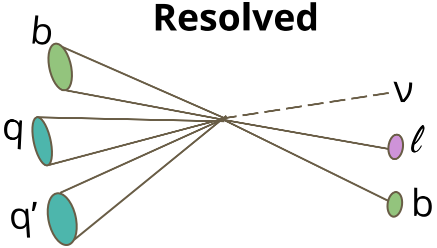
- Tag boosted top quark using simple mass window cut: $120 < m_J < 220$ GeV
- Tagging efficiency: 60%



Measurement of ttbar production (1-lepton)

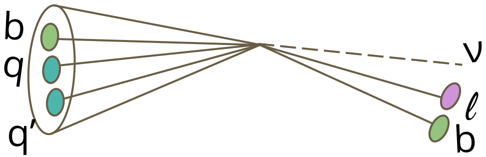
arXiv:1908.07305

Resolved



- 1 e/μ
- >=4 R=0.4 jets
- >=2 b-jets (70% WP)
- Fail boosted selection

Boosted



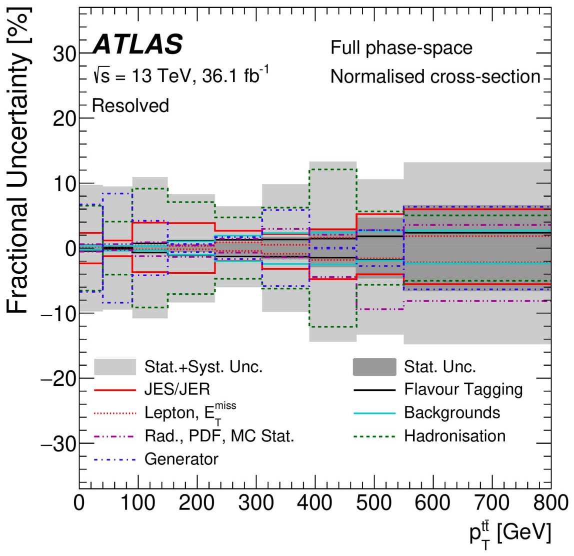
- 1 e/μ
- >=1 re-clustered R=1.0 jet
  - pT>350 GeV, 120 < m < 220 GeV
  - $\Delta\phi(\ell, J) > 1.0$

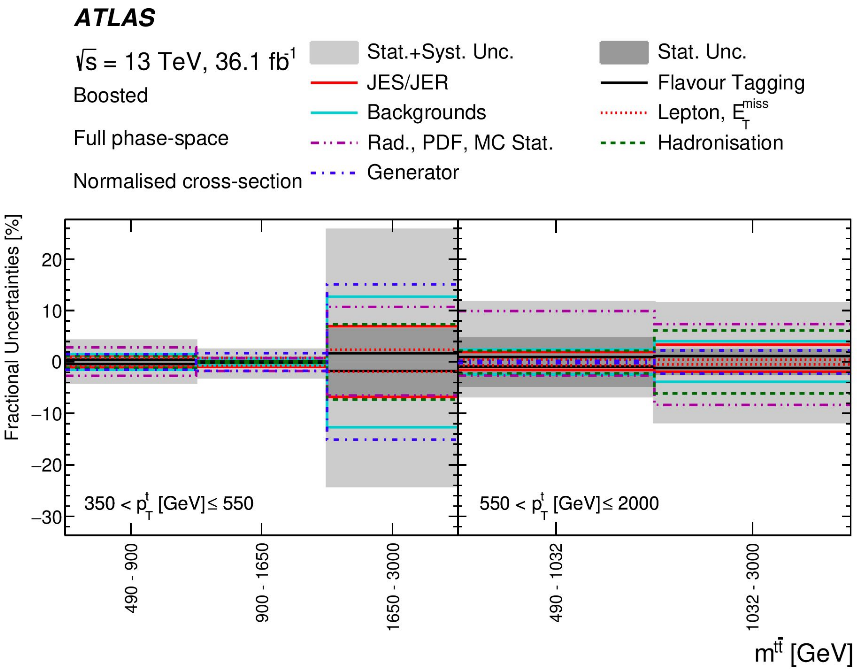
  } **Top tagging**
- >=1 b-jet (70% WP)
- >=1 R=0.4 jet with $\Delta R(\ell, j)<2.0$, $\Delta R(j, J) > 1.5$
- MET>20 GeV, MET+m_T^W > 60 GeV

Differential ttbar cross section (1-lepton)

arXiv:1908.07305

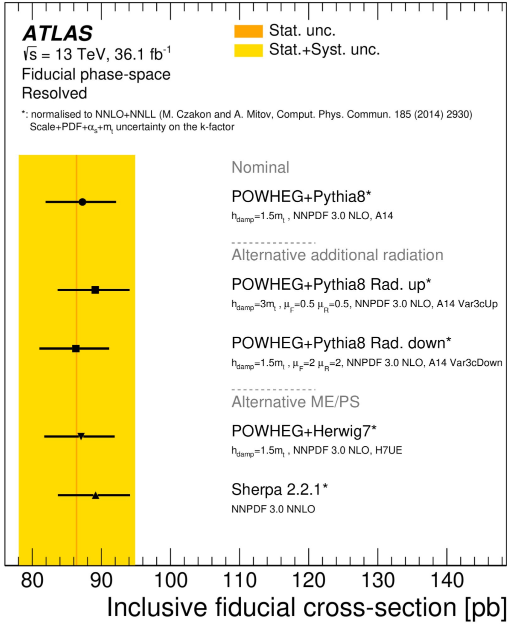
- Modelling uncertainties on tt production are dominant in both resolved and boosted channels
- Measurement precision varies between channels:
  - Total uncertainty in resolved channel **~10-15%**
  - Up to **40%** uncertainty at parton level in boosted channel



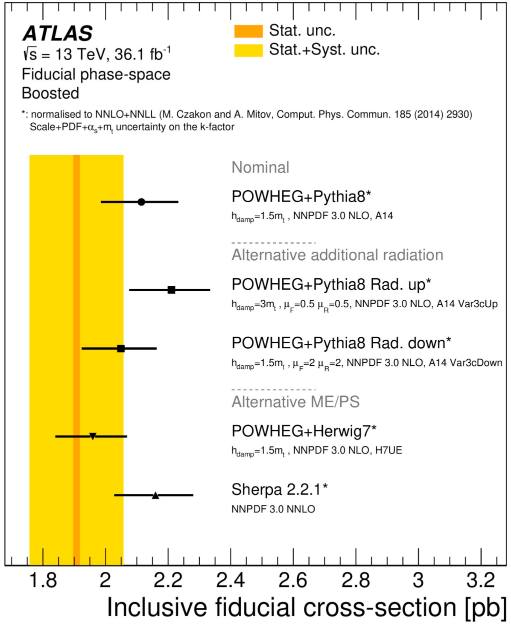


Differential ttbar cross section (1-lepton)

arXiv:1908.07305



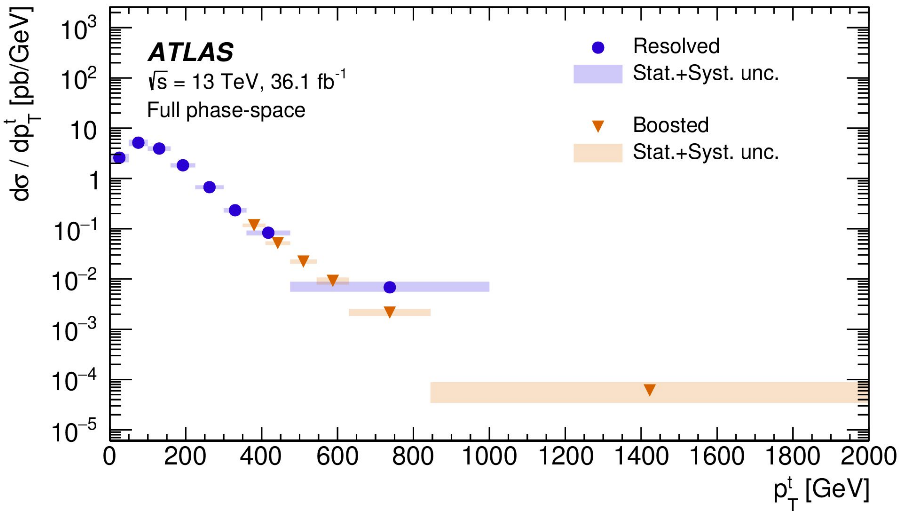
- Measured fiducial cross sections compared to different tt models
- Good agreement between models in resolved channel
- More variation between models seen in boosted phase space

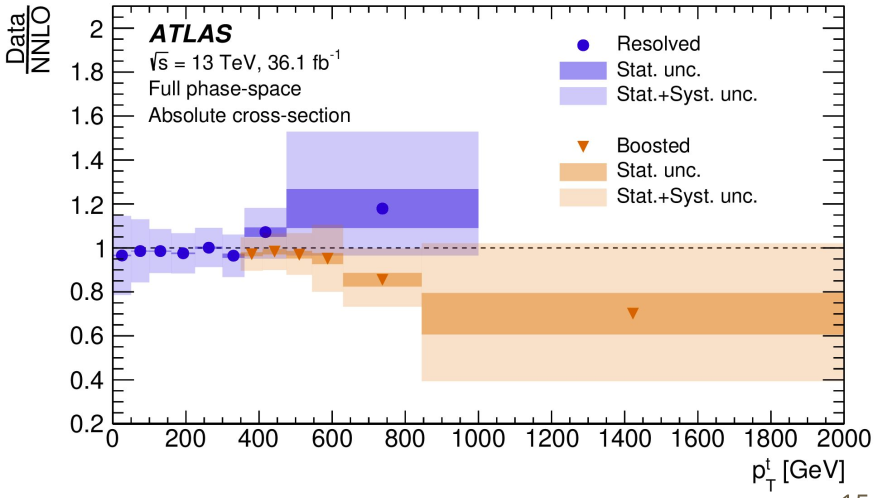


Differential ttbar cross section (1-lepton)

arXiv:1908.07305

- Cross sections unfolded to parton level are compared to NNLO predictions
    - Data agrees with predictions within systematic uncertainties
    - More tensions observed at high-pT
    - Results in resolved boosted channels agree in their regime of overlap





Charge asymmetry in boosted ttbar events

ATLAS-CONF-2019-026

- Top quark pair production is charge symmetric at leading order in Standard Model
  - tt production is LHC dominated by gluon fusion
  - Small asymmetry expected from qq→tt production channel due to difference in pT of valence quarks vs sea anti-quarks
- Charge asymmetric ttbar production expected in many BSM theories
  - Anomalous vector / axial couplings, heavy Z', interference with SM production
  - Charge asymmetry expected especially at high m_{tt}, β_{tt} (longitudinal boost of tt system)
- Measurement performed in boosted and resolved channels and combined to maximise sensitivity

$$A_C = \frac{N(\Delta|y| > 0) - N(\Delta|y| < 0)}{N(\Delta|y| > 0) + N(\Delta|y| < 0)}$$

$$\Delta|y| = |y_t| - |y_{\bar{t}}|$$

Charge asymmetry in boosted ttbar events

ATLAS-CONF-2019-026

- Both resolved and boosted channels required to contain exactly 1 e/μ
  - MET > 30 GeV, m_T^W > 30 GeV (e channel)
  - MET + m_T^W > 30 GeV (μ channel)
  - >=1 b-jets (1b and >=2b regions divided)

Resolved

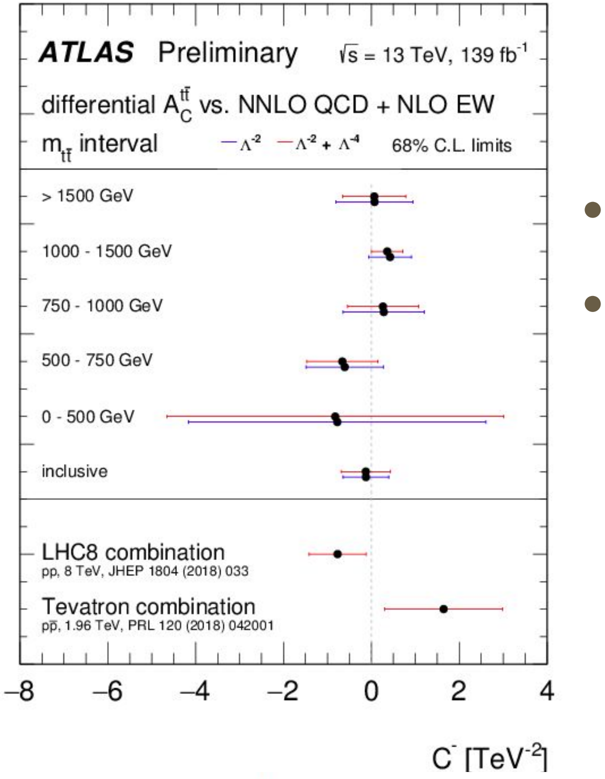
- Remove events passing boosted channel
- >=4 jets (R=0.4)
- Reconstruct ttbar system using BDT with 13 input variables

Boosted

- >=1 R=1.0 jet (J)
  - $\Delta\varphi(\ell, J) > 2.3$
  - pT > 350 GeV
  - Tagged as top quark based on mass and τ_{32} variables
  - Tagger cuts chosen to produce 80% signal efficiency
- >=1 R=0.4 jet (j)
  - $\Delta R(\ell, j) < 1.5$, $\Delta R(j, J) > 1.5$
- m_{tt} > 500 GeV

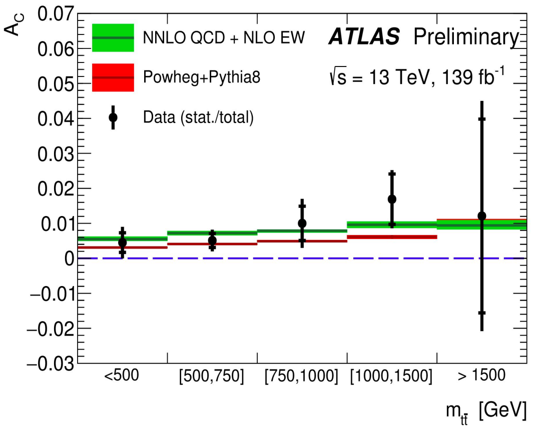
Charge asymmetry in boosted ttbar events

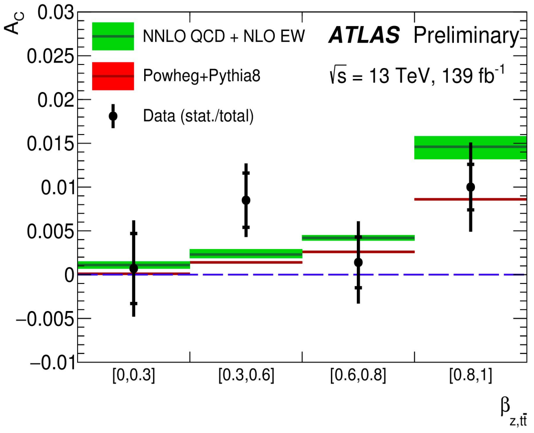
ATLAS-CONF-2019-026



$$\mathcal{L}_{eff} = \mathcal{L}_{SM} + \frac{1}{\Lambda^2}\sum_i C_i O_i + O\left(\Lambda^{-4}\right)$$

- Data in agreement with Standard Model prediction
- Results also interpreted in a Standard Model Effective Field Theory (SMEFT)
  - Limits set on Wilson coefficients for dim-6 operators
  - Significant improvement over previous measurements

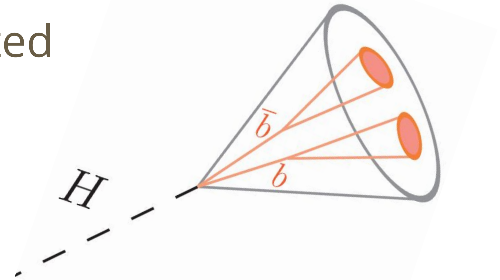




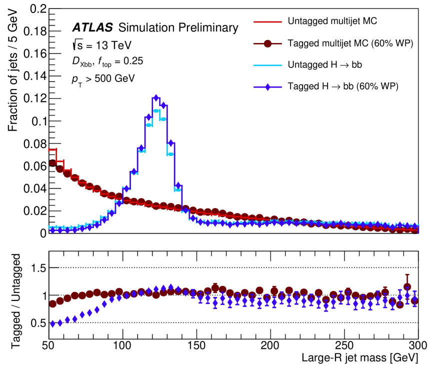
Measurements involving boosted Higgs bosons

Measurements involving boosted Higgs bosons

- Measurement of Higgs boson in H→bb decay channel motivated by large branching fraction
- Direct probe of Yukawa coupling to down-type quarks
- Sensitive to dim-6 operators in EFT
  - Effects grow with p_T^{Higgs}
- Collimated decays at high-pT can be tagged by using **jet substructure** and **b-tagging** information
  - Reconstruct R=1.0 jets from locally calibrated topoclusters
    - pT > 250 GeV
    - m_J > 50 / 60 GeV (analysis-dependent)
  - Reconstruct variable-R jets from charged tracks (0.02 < R < 0.4)
    - Match to R=1.0 jets by ghost association
    - Apply multivariate b-tagging algorithm to ghost associated track jets
- Candidate R=1.0 Higgs jets required to have >=2 ghost-associated track jets
  - Defined as Higgs tagged if leading 2 associated track jets are **b-tagged**
- Tagger based on flavour tagging:
  - Minimal use of jet substructure information reduces mass sculpting of background

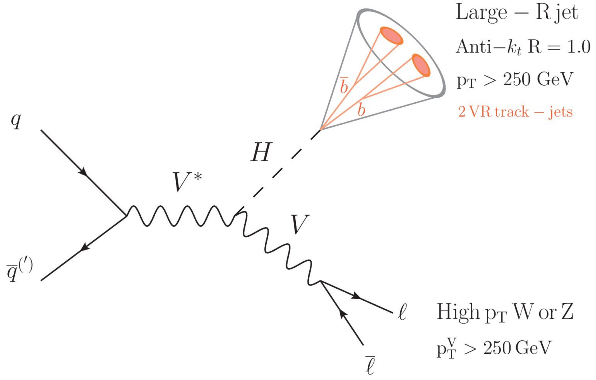


ATL-PHYS-PUB-2020-019



Mass spectra after tagging with newest X->bb tagger (not used in presented results)

Measurement of VH(bb) production



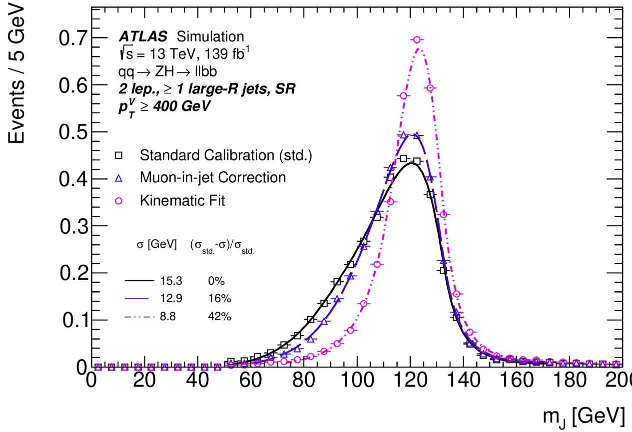
- Measurement of VH(bb) production with leptonic decays of Z/W and boosted H→bb decays
  - High-pT regime particularly sensitive to new physics
- Separate 0,1,2-lepton channels targeting Z(νν), W(ℓν), Z(ℓℓ) decays, respectively
  - >=1 Higgs candidate required in all regions
  - p_T^V > 250 GeV required in all channels (0lep: MET, 1lep: pT(ℓ + MET), 2lep: pT(ℓℓ))
  - Additional selection cuts used to suppress background (see backup)
- Main backgrounds: ttbar, tW, V+jets
- Events categorised into high-purity (HP) and low-purity (LP) signal regions, and ttbar control regions

Channel	Categories					
	$250 < p_T^V < 400$ GeV			$p_T^V \geq 400$ GeV		
	0 add. b-track-jets		≥ 1 add. b-track-jets	0 add. b-track-jets		≥ 1 add. b-track-jets
	0 add. small-R jets	≥ 1 add. small-R jets		0 add. small-R jets	≥ 1 add. small-R jets	
0-lepton	HP SR	LP SR	CR	HP SR	LP SR	CR
1-lepton	HP SR	LP SR	CR	HP SR	LP SR	CR
2-lepton	SR			SR		

arXiv:2008.0250

Measurement of VH(bb) production

- Dedicated corrections applied to large-R jet to account for semileptonic b-decays inside it
  - Add 4-momentum of closest reconstructed non-isolated muon within $\Delta R = \min(0.4, 0.04 + 10\ \mathrm{GeV}/p_T^{\mu})$
  - Remove associated calorimeter clusters
- Improve scale and resolution of jet energy and mass further in 2-lep channel by kinematic fit
  - Require transverse momentum in event to fully balance



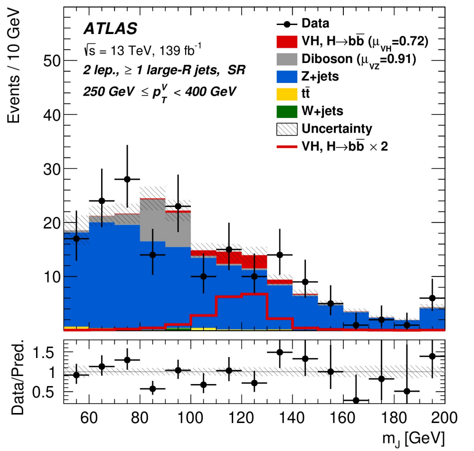
arXiv:2008.0250

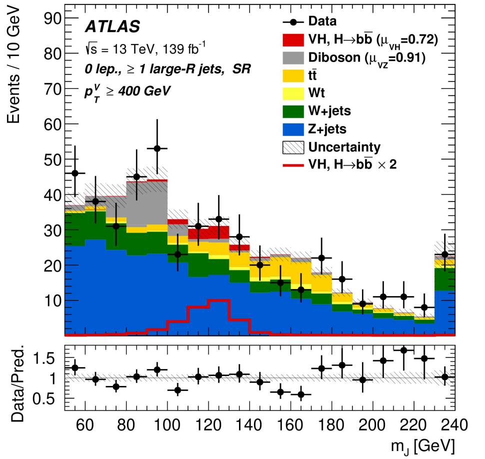
Measurement of VH(bb) production

Source of uncertainty		Avg. impact
Total		0.372
Statistical		0.283
Systematic		0.240
Experimental uncertainties		
Small-R jets		0.038
Large-R jets		0.133
$E_{\mathrm{T}}^{\mathrm{miss}}$		0.007
Leptons		0.010
b-tagging	b-jets	0.016
	c-jets	0.011
	light-flavour jets	0.008
	extrapolation	0.004
Pile-up		0.001
Luminosity		0.013
Theoretical and modelling uncertainties		
Signal		0.038
Backgrounds		0.100
↪ Z + jets		0.048
↪ W + jets		0.058
↪ $t\bar{t}$		0.035
↪ Single top quark		0.027
↪ Diboson		0.032
↪ Multijet		0.009
MC statistical		0.092

arXiv:2008.0250

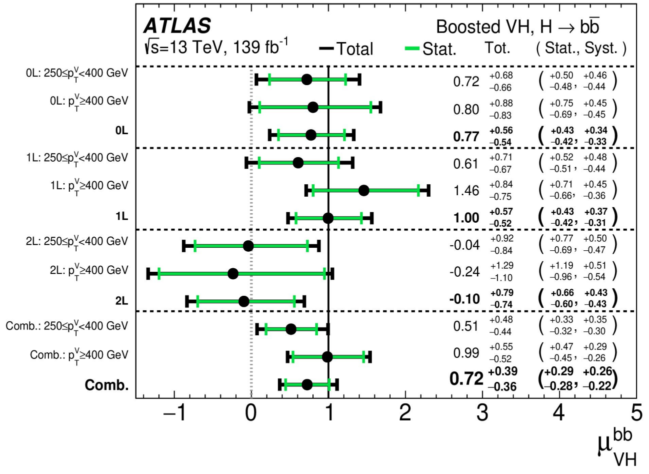
- Results obtained with a binned profile likelihood fit to mass of candidate Higgs jet
  - Excess over background observed with **2.1σ** significance after combining all channels
- Experimental uncertainties dominated by large-R jet calibration uncertainties

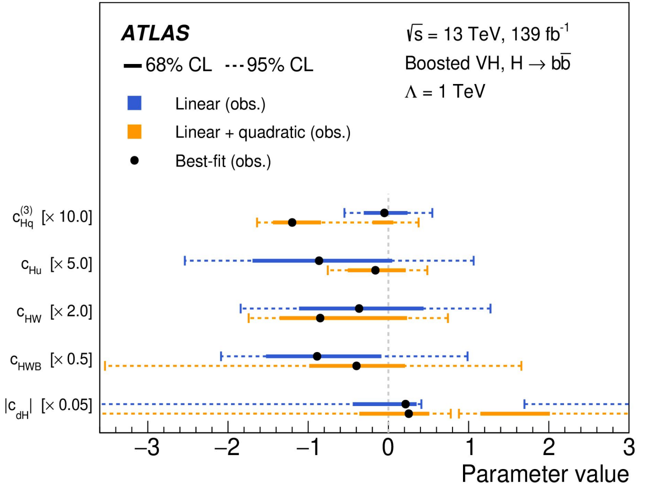




Measurement of VH->bb production

- Combination across all channels significantly improves measurement precision
- Measured signal strength compatible with Standard Model
- Cross sections measured within the simplified template cross section (STXS) framework and used to constrain dim-6 operators in an SMEFT

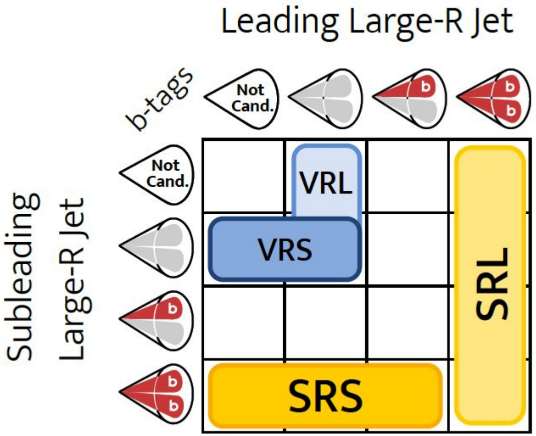




arXiv:2008.0250

ATLAS-CONF-2021-010

Measurement of H->bb production



- Measurement targeting inclusive production of Higgs bosons in Standard Model
  - Focus on high-pT regime which is sensitive to BSM physics
  - Using H→bb decays preserves largest branching fraction
- No restrictions placed on production channel
  - Measurement sensitive to **ggF / VBF / VH / ttH** production

Process	Jet p_T Range [GeV] 250–450	450–650	650–1000	> 1000
SRL				
ggF	–	0.56	0.50	0.39
VBF	–	0.17	0.16	0.17
VH	–	0.14	0.18	0.25
$t\bar{t}H$	–	0.13	0.16	0.19
SRS				
ggF	0.28	0.46	0.43	–
VBF	0.07	0.19	0.21	–
VH	0.26	0.24	0.26	–
$t\bar{t}H$	0.39	0.11	0.10	–

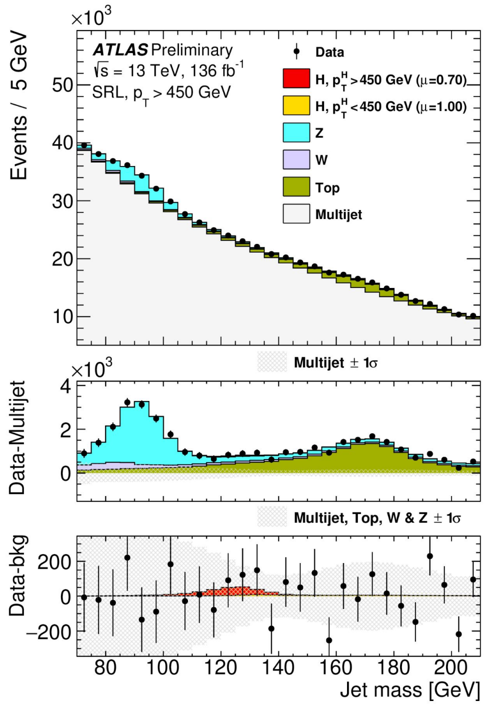
- Events categorised into signal and validation categories according to number of b-jets associated to Higgs candidate
  - Validation regions used to check quality of background modelling
- Dedicated single-muon control region for ttbar measurement

Jet	N track-jets	N b-tags	Angular Selection	Jet Mass [GeV]
J_b	≥ 1	1	$0.04 + 10/p_T^\mu < \Delta R(\mu, J^b) < 1.5$	–
J_t	≥ 3	1	$\Delta\phi(J^b, J^t) > \frac{2\pi}{3}$	140–200

[ATLAS-CONF-2021-010]

Measurement of H->bb production

- Results obtained with profile likelihood fit to Higgs candidate jet mass
- Multijet background dominates in the signal regions
  - Smooth spectrum modelled as exponential of polynomials, and measured simultaneously with signal extraction
- Main systematic uncertainty comes from jet mass resolution and scale
  - Strongly correlated to measured Z+jets normalisation
  - Z/W resonance widths directly measured in dedicated regions in data to reduce impact of JMS/JMR uncertainties

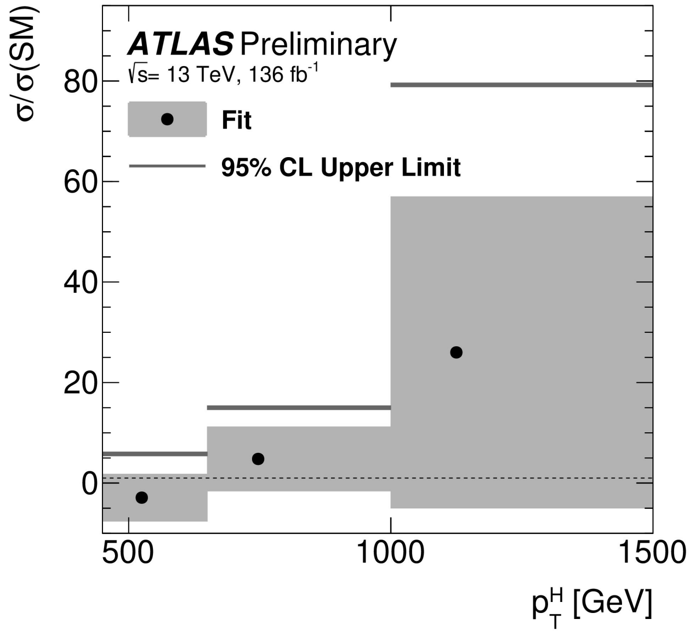


Uncertainty Contribution	$p_T^H > 450$ GeV	$p_T^H > 1$ TeV
Total	3.3	31
Statistical	2.8	30
Jet Systematics	1.2	7
Modeling and Theory Systs.	1.0	1
Flavor Tagging Systs.	0.5	3
Total Systematics	1.7	8

ATLAS-CONF-2021-010

Measurement of H->bb production

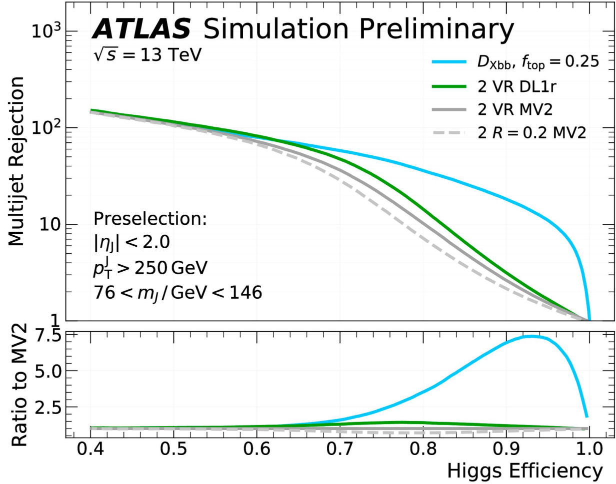
- Signal strengths measured inclusively as well as in exclusive pT bins
  - Upper limits also set on production cross sections
- All results compatible with SM predictions
- Trend in pT? *

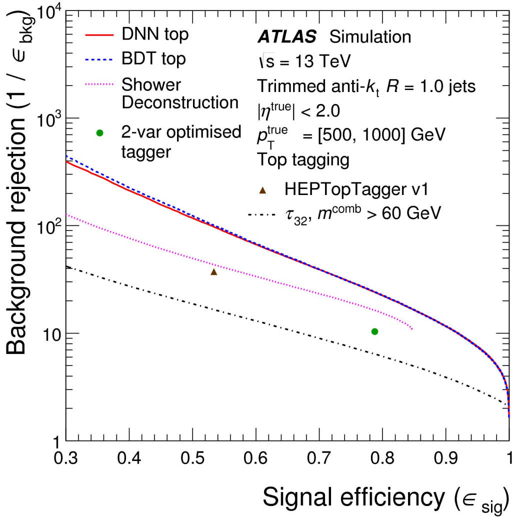


* see also slide 45

Perspectives on prospectives

- Newer generation of boosted object taggers already available with better performance
  - Machine learning techniques used in many cases to improve tagger purity and efficiency
  - Exploit correlations from a large set of input variables
- Expected to significantly improve measurement precision and search sensitivity for next round of analyses
- Uncertainties on large-R jet / tagger calibration can have large impact in boosted regime
  - Usually dominated by modelling uncertainties (2-point generator comparisons)
  - Major limiting factor for measurement precision

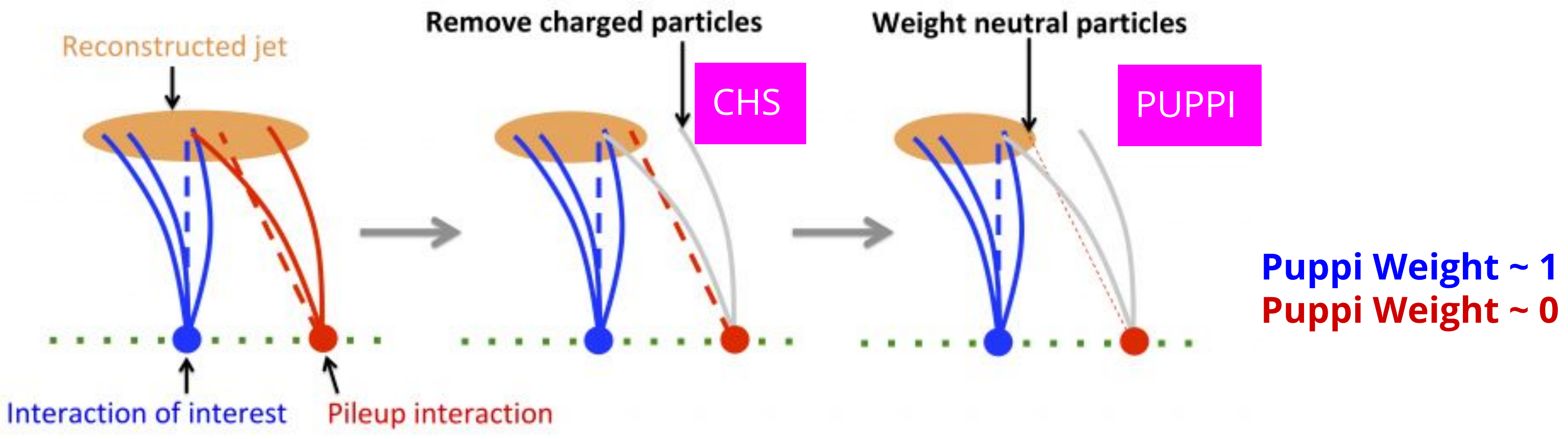




ATL-PHYS-PUB-2020-019
arXiv:1808.07858

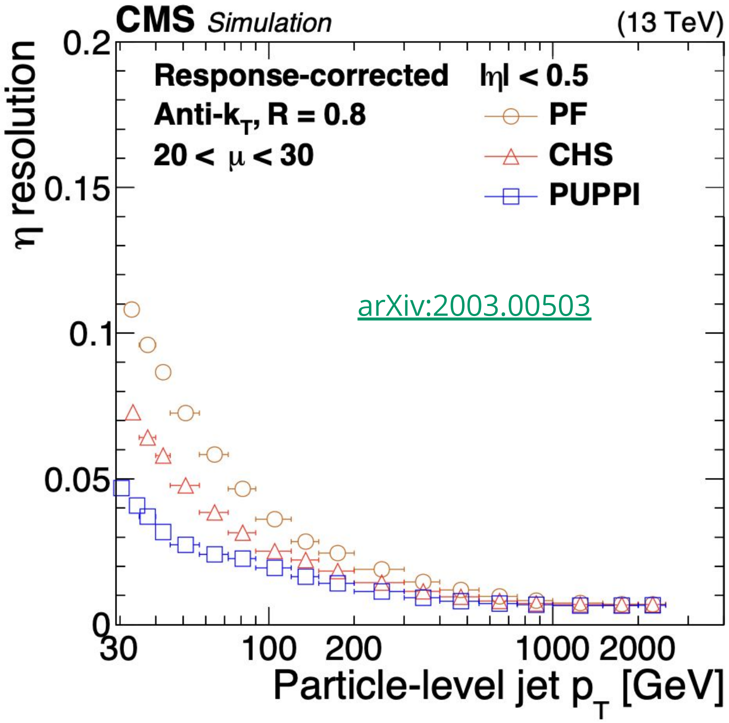
CMS Measurements

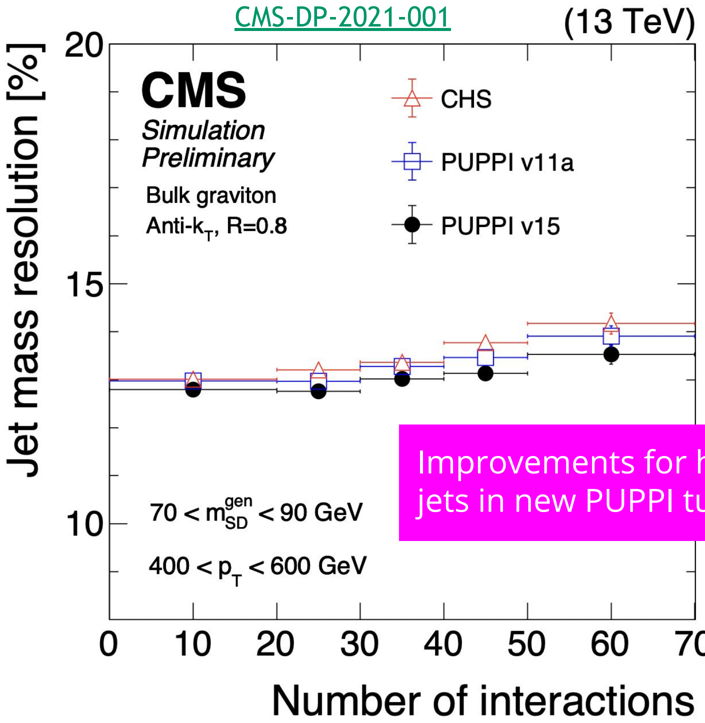
Pileup treatment in CMS when tagging high-pT objects



- Large-radius jets, if untreated for pileup, show a much degraded resolution due to particles from pileup interactions
- First remove charged hadrons which are assigned to a PU vertex (==CHS)
- Then compute PUPPI (PileUp Per Particle Identification) weight for neutrals by considering particles in cone around particle of interest and comparing to density of charged particles.

Pileup treatment in CMS when tagging high-pT objects





Improvements for high-pT jets in new PUPPI tune

Overview of tagging techniques in CMS

arXiv:2004.08262

Algorithm	Subsection	jet p_T [GeV]	t quark	W boson	Z boson	H boson
$m_{SD} + \tau_{32}$	6.1	400	✓			
$m_{SD} + \tau_{32} + b$	6.1	400	✓			
$m_{SD} + \tau_{21}$	6.1	200	✓	✓		
HOTVR	6.2	200	✓			
N_3-BDT (CA15)	6.3	200	✓			
$m_{SD} + N_2$	6.3	200		✓	✓	✓
BEST	6.5	500	✓	✓	✓	✓
ImageTop	6.6	600	✓			
DeepAK8$^{(*)}$	6.7	200	✓	✓	✓	✓
Jet mass decorrelated algorithms						
$m_{SD} + N_2^{DDT}$	6.3	200		✓	✓	✓
double-b	6.4	300			✓	✓
ImageTop-MD	6.6	600	✓			
DeepAK8-MD$^{(*)}$	6.7	200	✓	✓	✓	✓

Classical (“cut-based”); very successful historically, theory-inspired; baseline to improve on with ML, or to use in conjunction with ML (e.g. N2 + Double-B tagger for H bosons)

Shallow machine learning algos using high-level observables

More advanced/deep machine learning algos using also low-level information like Particle Flow objects

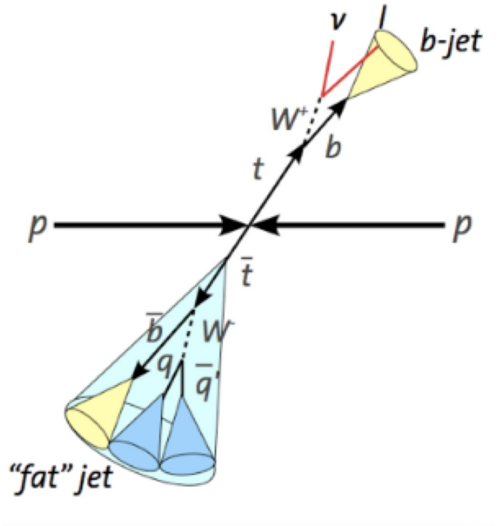
Plus analysis-specific algos/solutions not listed in this table

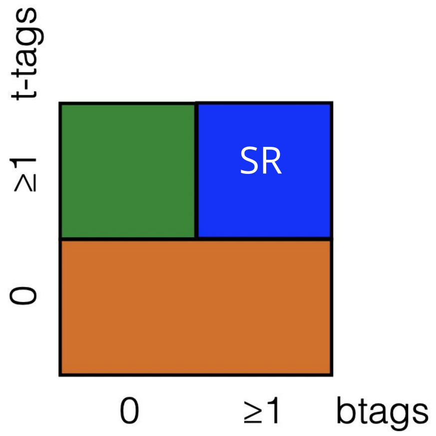
Boosted tops

Boosted tops in ttbar: 1-lepton channel

arXiv:2008.07860

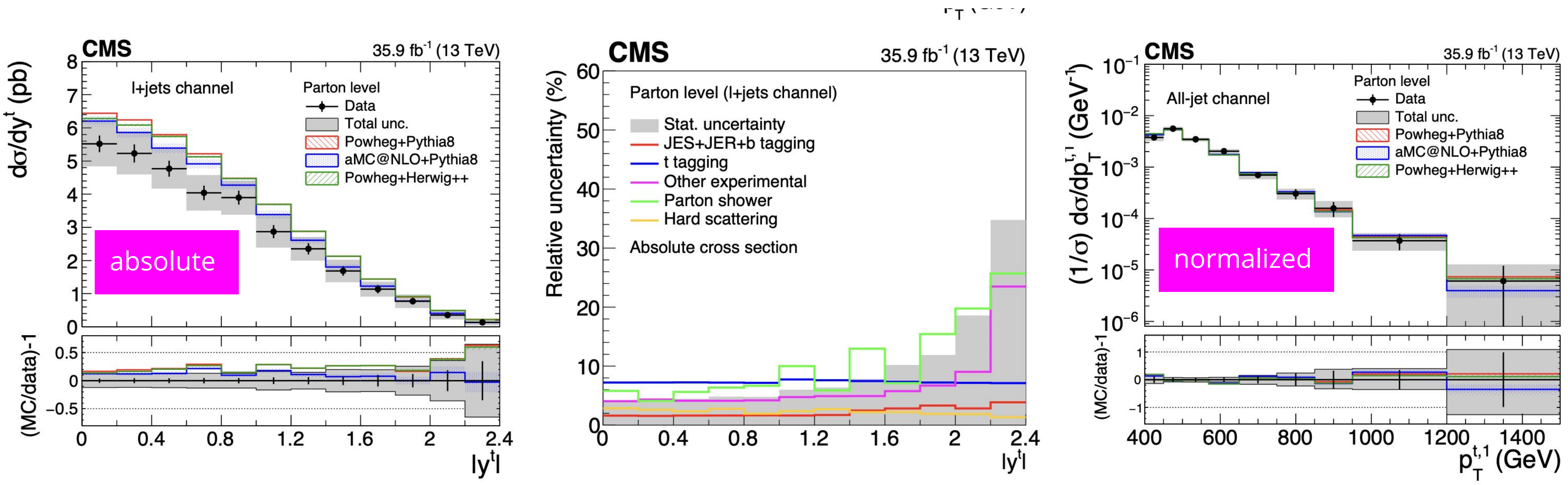
- Event selection:
  - 1 clean lepton with pT>50 GeV
  - 1 AK4 b jet with pT>50 GeV identified with shallow b tagging network, close to lepton: $0.3 < \Delta R(l,j) < \pi/2$
  - 1 AK8 jet with pT>400 GeV away from lepton: $\Delta R(l,j) > \pi/2$
- Classify events based on whether t jet and b jet pass tagging requirements
- Cut-based top tagging:
  - 105 < soft-drop mass < 220 GeV && tau_32 < 0.81 && b-tagging on subjets
- Simultaneous fit in three regions to extract top tag scale factor and background normalizations
  - Fit distributions of AK4 jet η in 0t and in 1t0b and soft-drop mass in 1t1b





Boosted tops in ttbar: 1-lepton channel

arXiv:2008.07860

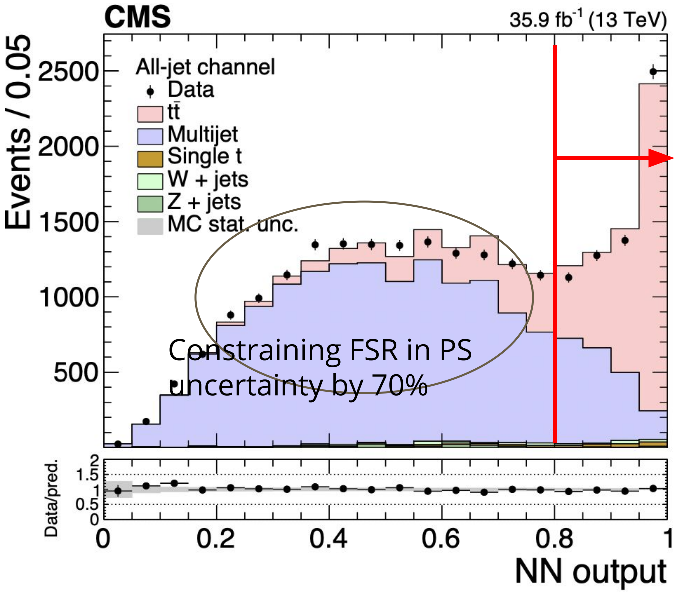


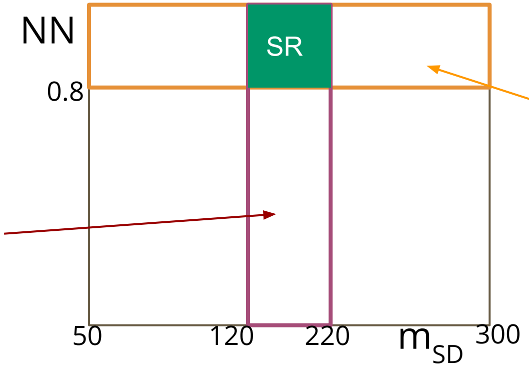
- Unfold background-subtracted data in 1t1b to parton level
- Theory somewhat overpredicts for all models, but describes shapes well
- aMC@NLO predicts slightly more central y distribution than Powheg; data favors aMC@NLO
- Dominant uncertainties from parton shower (entire spectrum) and from statistics (high y)

Boosted tops in ttbar: 0-lepton channel

arXiv:2008.07860

- Event selection:
  - Two fat jets with pT>400 GeV and $120<m_{SD}<220$ GeV
  - Deep neural network event classifier trained with N-subjettiness variables τ_1, τ_2, τ_3 from both fat jets
  - Both fat jets also have to contain a b-tagged subjet
- Background mostly QCD, others negligible
- QCD normalization and nuisance parameters constrained with extended signal regions and control regions



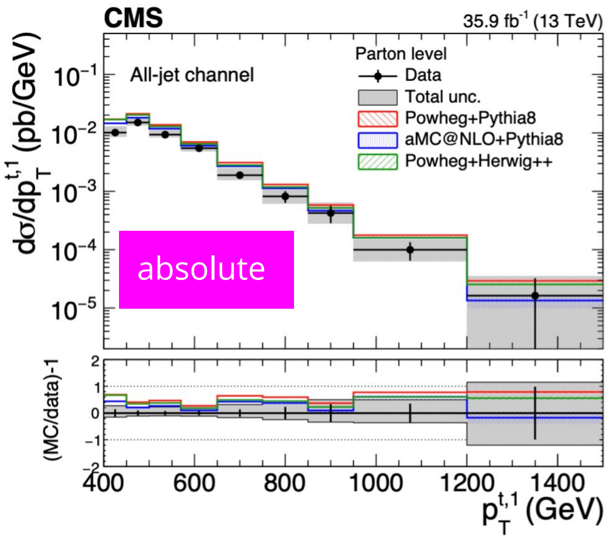


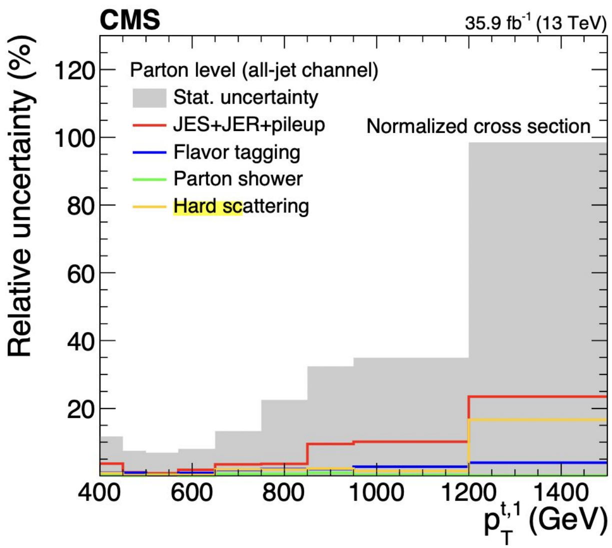
Using extended SR to constrain signal-modelling uncertainties in template fit

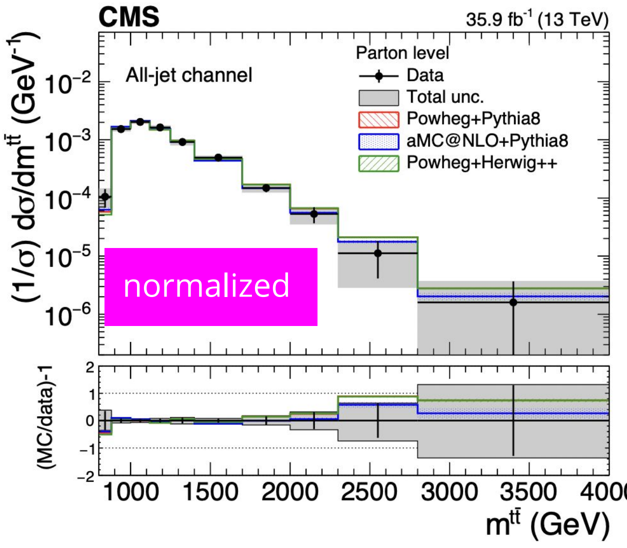
Using the extended SR together with identical CR (inverted subjet b tagging) to constrain shape and norm. of QCD in template fit

Boosted tops in ttbar: 0-lepton channel

arXiv:2008.07860







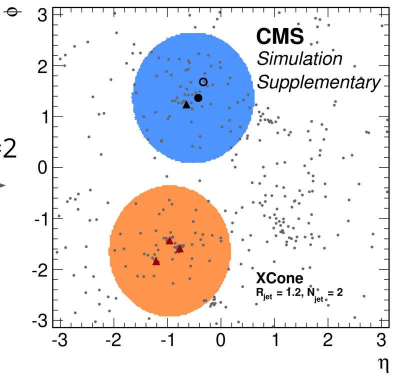
- Unfold background-subtracted SR data parton level
- Theory overpredicts by 20%
- Shapes look good, mtt has discrepancies in tails, more data needed
- Dominant uncertainties from jet energy scale and resolution

Boosted tops in ttbar: top quark mass

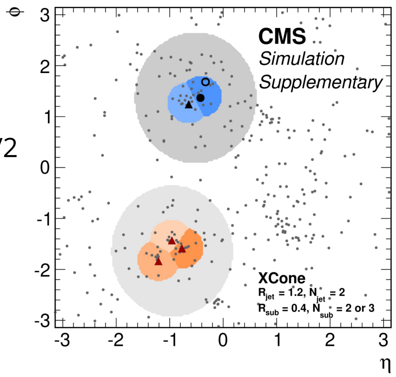
arXiv:1911.03800

- Idea: use semileptonic events; go into a boosted regime to measure the jet mass; unfold to particle level; extract m_{top}
- Using XCone algorithm to identify hadronically decaying top quarks
  - Exclusive jet algorithm, returns exactly N jets → expected event signature defines clustering; jet axes found by minimizing N-subjettiness; smooth transition between boosted and resolved regimes
  - Strategy: Cluster particles with N=2 and R=1.2 to obtain two large-radius jets
  - Recluster constituents with N=3 and R=0.4 for hadronic jet and N=2 for the leptonic jet
  - Mass resolution 6%, compared to 14% for Cambridge-Aachen R1.2

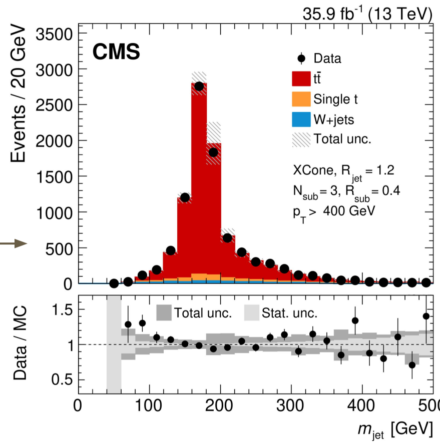
XCone 1.2, N=2 →



XCone 0.4, N=3/2 →

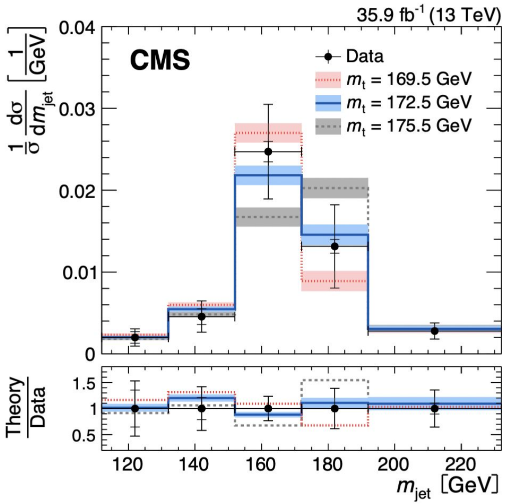


→



Boosted tops in ttbar: top quark mass

arXiv:1911.03800

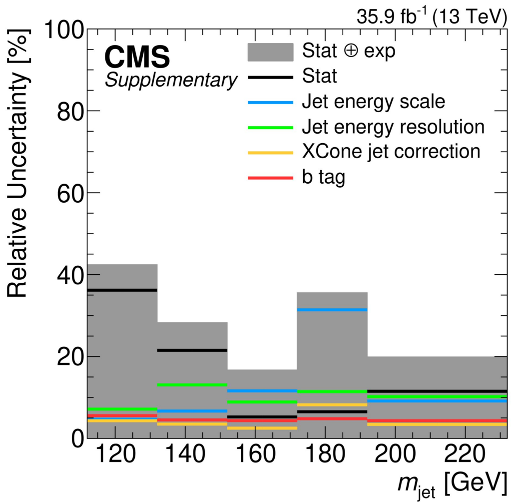


$$\chi^2 = d^T V^{-1} d$$

Distance vector between per-bin measurements and predictions

Covariance matrix encoding all uncertainties

Bulk most affected by JES and parton shower uncs



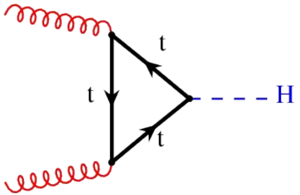
$$m_t = 172.6 \pm 0.4\,(\text{stat}) \pm 1.6\,(\text{exp}) \pm 1.5\,(\text{model}) \pm 1.0\,(\text{theo})\,\text{GeV}$$

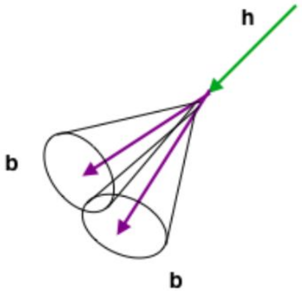
- Unfolding to particle level after background subtraction includes sideband region to constrain migrations in and out of measurement phase space
- Much improved result compared to 8 TeV measurement

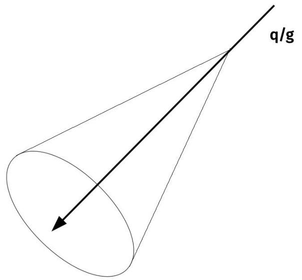
Boosted Higgs

Measurement of H->bb production [arXiv:2006.13251]

- Tail of inclusive H → bb spectrum highly sensitive to new physics in loop
- H → bb to retain most signal, but also incurs large QCD background
- Use N2 variable targeting substructure to select large-radius jets with 2 prongs
- Employ machine learning to identify flavor content
  - Use a deep neural network with 1D convolutions among two input sets: charged particles (8 features per particle) and secondary vertices (2 features per SV)
  - Mass decorrelation achieved by computing difference in mass distributions of tagged and untagged QCD jets and adding the difference to the loss function the network tries to minimize

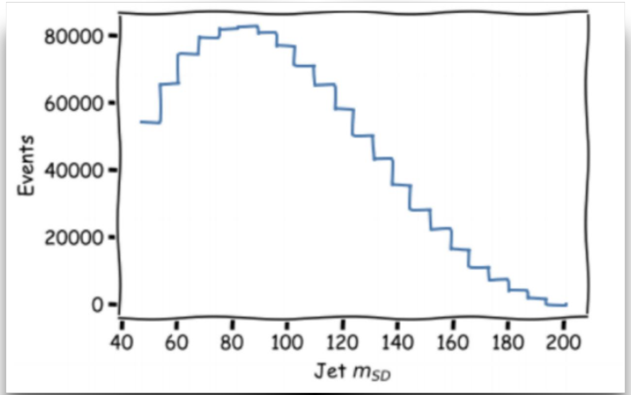






Measurement of H->bb production arXiv:2006.13251

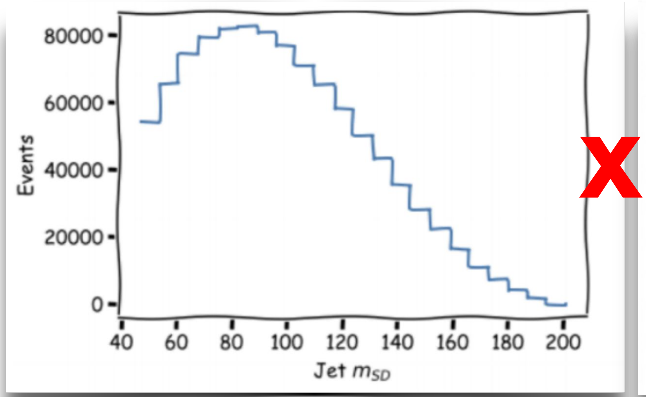
- Main background in signal region comes from QCD multijet production with non-trivial, pT-dependent jet mass shape → hard to model
- Data-driven background estimation, starting from high-stat background-enriched region *failing* the deep-double-B tagger



prediction in region failing H(bb) tagger

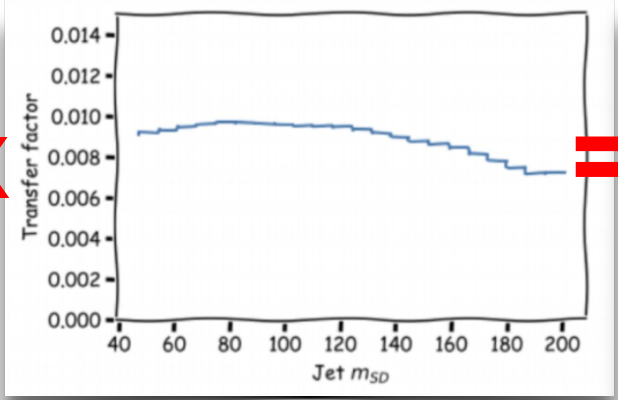
Measurement of H->bb production [arXiv:2006.13251]

- The pass/fail ratio, obtained from MC simulation and interpolated with a polynomial, acts as a transfer factor to propagate estimation to signal region



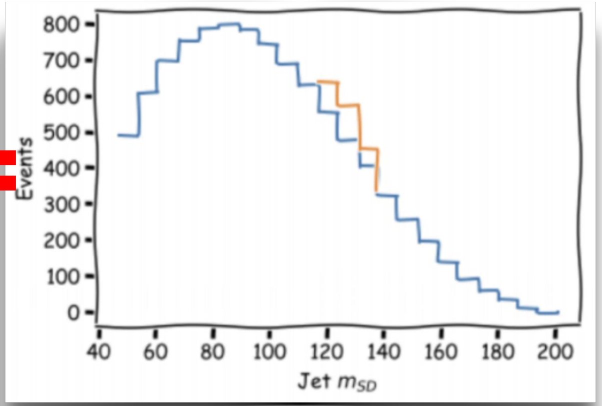
prediction in region failing H(bb) tagger

x



pass/fail ratios parametrized as 2nd order Bernstein polynomial, fit to MC simulation + uncertainties

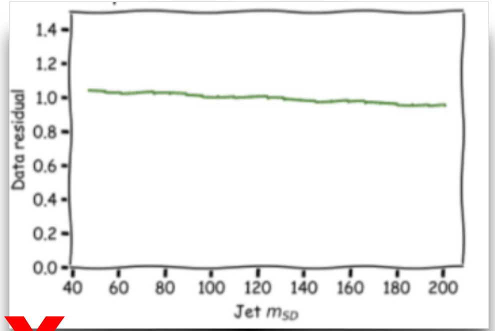
=

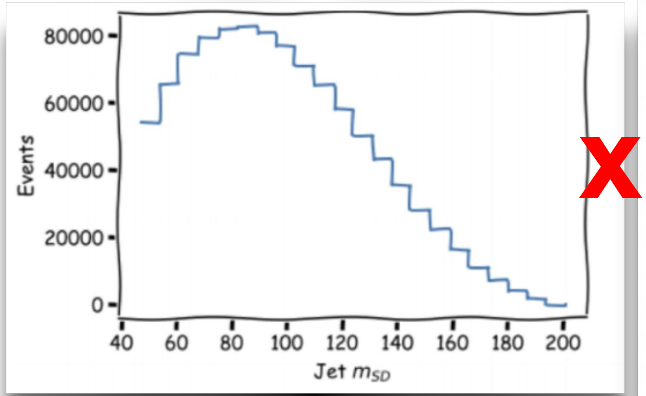


prediction in signal region (i.e., passing tagger)

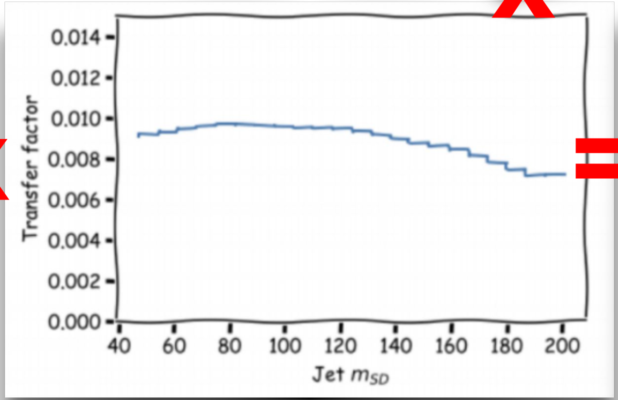
Measurement of H->bb production

fit also for parameters of Bernstein polynomial to take care of residual data/simulation difference → in-situ tagger calibration



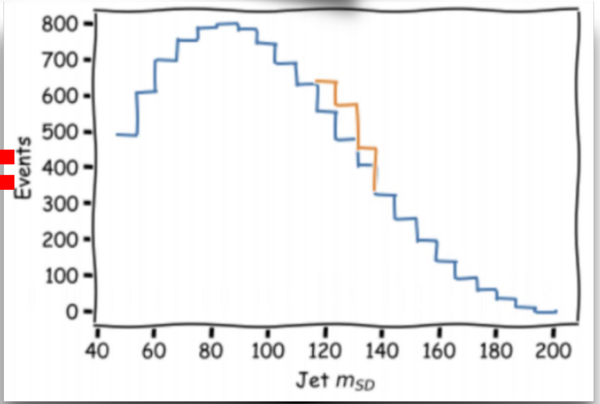


x



x

=



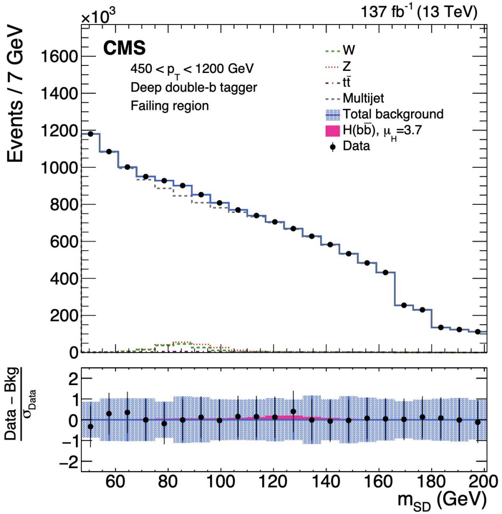
prediction in region failing H(bb) tagger

pass/fail ratios parametrized as 2nd order Bernstein polynomial, fit to MC simulation + uncertainties

prediction in signal region (i.e., passing tagger)

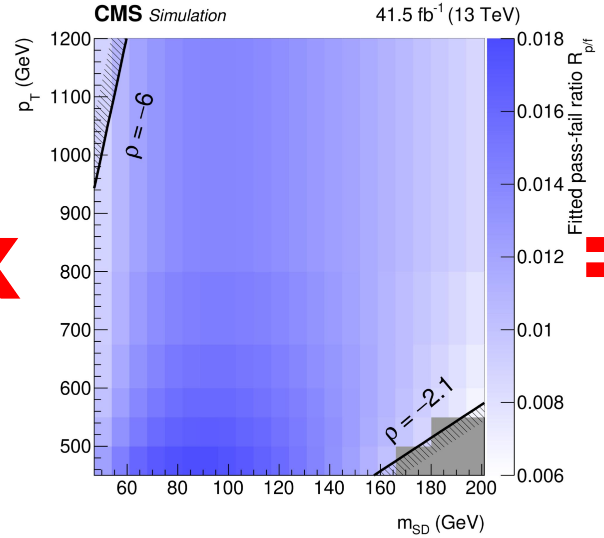
Measurement of H->bb production [arXiv:2006.13251]

- Leading uncertainties from fitted polynomial parameters determining pass/fail transfer factor



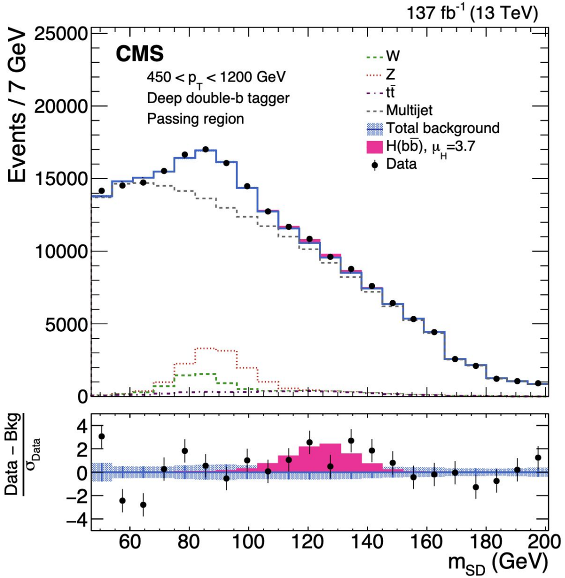
prediction in region failing H(bb) tagger

x



Fitted pass/fail ratio in 22 mSD bins and 6 pT bins

=

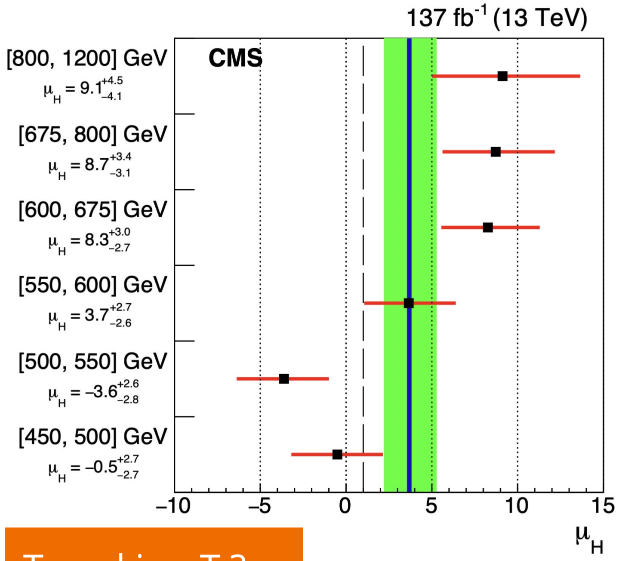


prediction in signal region (i.e., passing tagger)

Measurement of H->bb production arXiv:2006.13251

- Cross-check with Z → bb gives μ_Z=1.01(+0.24/-0.20); fix Z hereafter to expectation + unc, serving as additional constraint when extracting H → bb

μ_H=3.7(+1.6/-1.5)



	2016	2017	2018	Combined
Expected μ_Z	$1.00^{+0.38}_{-0.28}$	$1.00^{+0.42}_{-0.29}$	$1.00^{+0.43}_{-0.29}$	$1.00^{+0.23}_{-0.19}$
Observed μ_Z	$0.86^{+0.32}_{-0.24}$	$1.11^{+0.48}_{-0.33}$	$0.91^{+0.37}_{-0.26}$	$1.01^{+0.24}_{-0.20}$
HJ-MINLO [32, 33]				
Expected μ_H	$1.0^{+3.3}_{-3.5}$	1.0 ± 2.5	$1.0^{+2.3}_{-2.4}$	1.0 ± 1.4
Observed μ_H	$7.9^{+3.4}_{-3.2}$	$4.8^{+2.6}_{-2.5}$	1.7 ± 2.3	$3.7^{+1.6}_{-1.5}$
Expected H significance ($\mu_H = 1$)	$0.3\,\sigma$	$0.4\,\sigma$	$0.4\,\sigma$	$0.7\,\sigma$
Observed H significance	$2.4\,\sigma$	$1.9\,\sigma$	$0.7\,\sigma$	$2.5\,\sigma$
Expected UL μ_H ($\mu_H = 0$)	<6.8	<5.0	<4.7	<2.9
Observed UL μ_H	<13.9	<9.3	<5.9	<6.4
Ref. [23] H p_T spectrum				
Expected μ_H	1.0 ± 1.5	$1.0^{+1.1}_{-1.0}$	$1.0^{+1.1}_{-1.0}$	$1.0^{+0.7}_{-0.6}$
Observed μ_H	$4.0^{+1.9}_{-1.6}$	$2.2^{+1.4}_{-1.2}$	1.1 ± 1.1	$1.9^{+0.9}_{-0.7}$
Expected H significance ($\mu_H = 1$)	$0.7\,\sigma$	$0.9\,\sigma$	$1.0\,\sigma$	$1.7\,\sigma$
Observed H significance	$2.6\,\sigma$	$1.8\,\sigma$	$1.1\,\sigma$	$2.9\,\sigma$
Expected UL μ_H ($\mu_H = 0$)	<3.4	<2.4	<2.3	<1.4
Observed UL μ_H	<7.4	<4.6	<3.2	<3.4

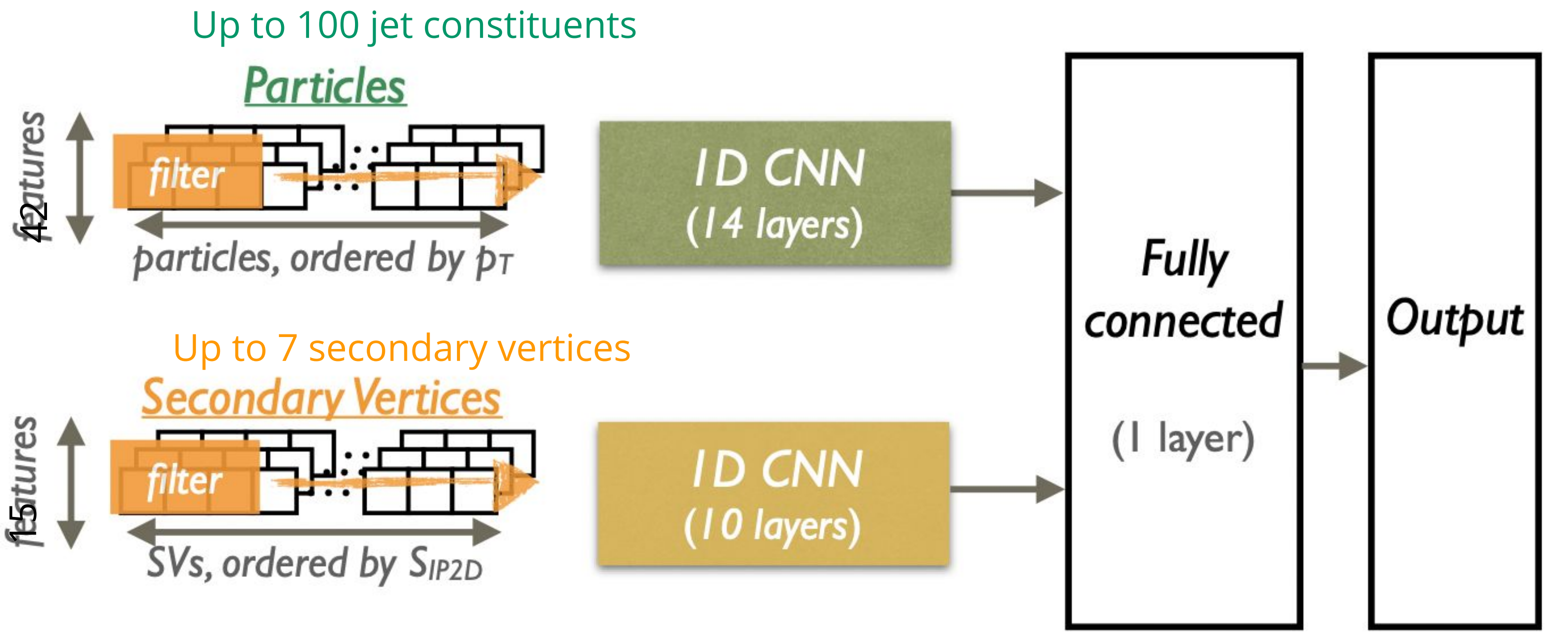
Trend in pT ?

Next-generation taggers & objectives

DeepAK8

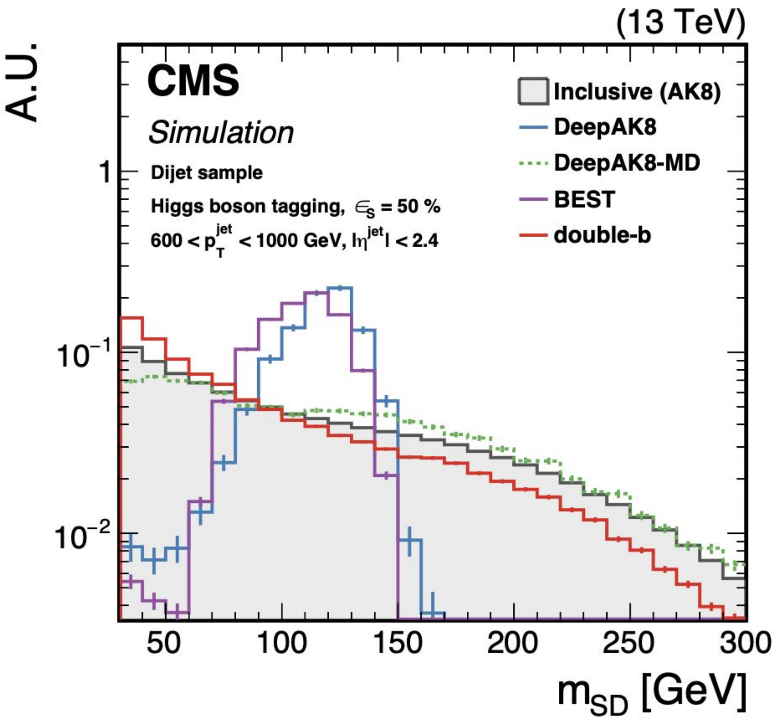
arXiv:2004.08262

- Similar to deep neural network in Hbb measurement, but blown up!
- Different output nodes, e.g. for Higgs decay modes (bb, cc, ...)

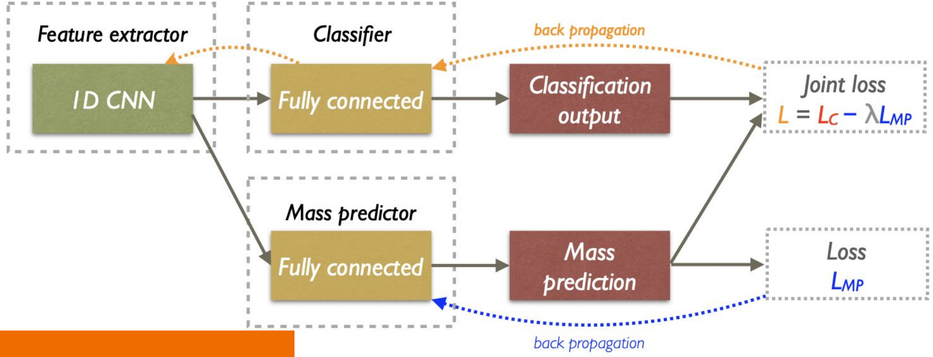


DeepAK8

arXiv:2004.08262



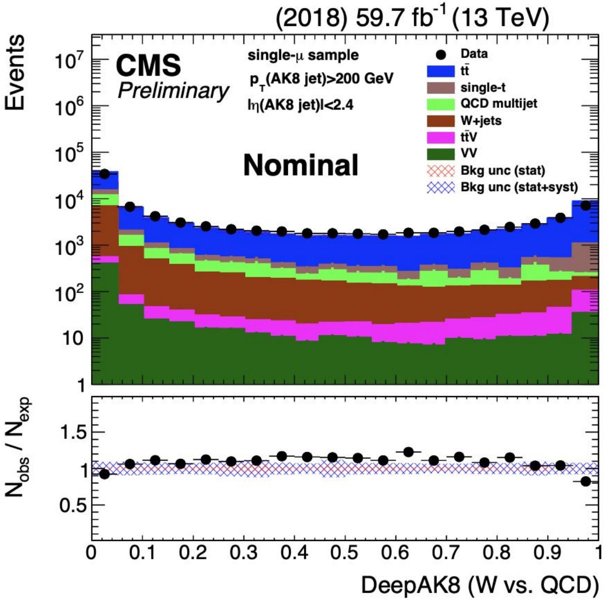
- Strong sculpting of the mass in QCD events after cut on the Higgs tagger
- Undesirable to perform a bump hunt on top of a bump
- Use adversarial debiasing to get rid of mass dependence

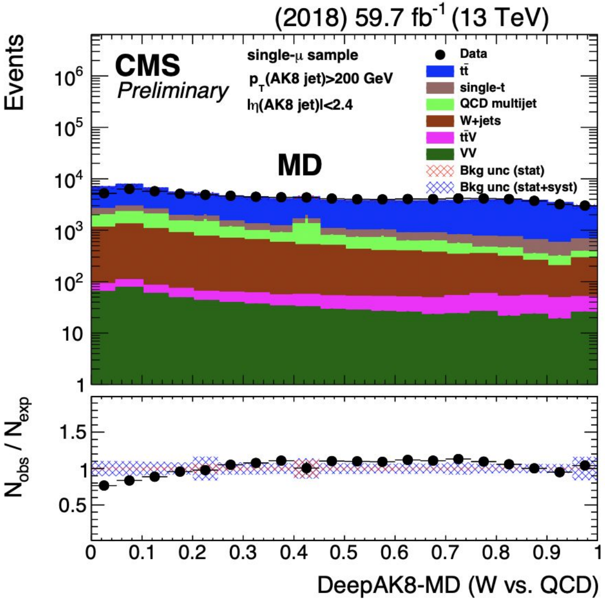


DeepAK8-MD

DeepAK8

arXiv:2004.08262

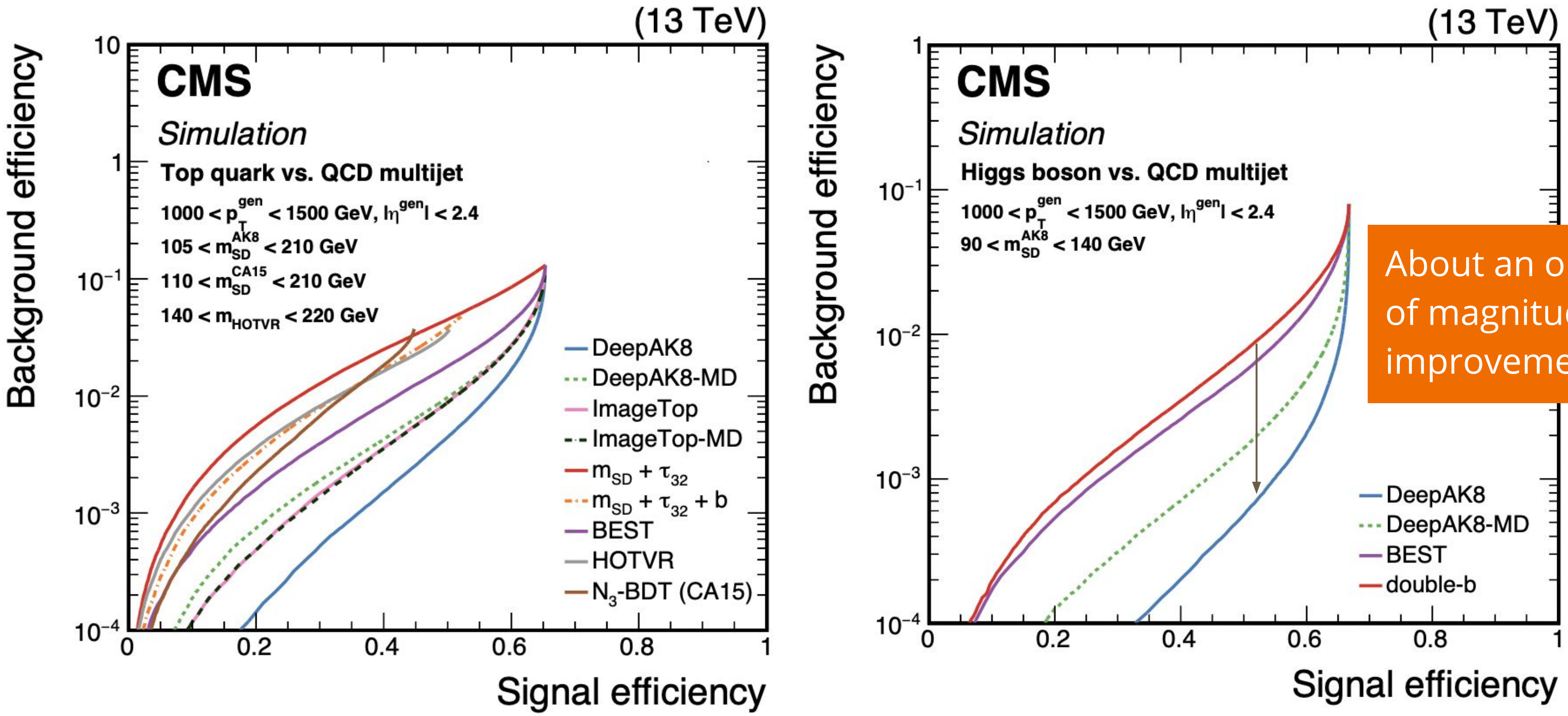




- Even though mass-decorrelated version needs to exploit more subtle differences between signal and background, which may not be very well modelled, the data/MC agreement does not suffer compared to DeepAK8

DeepAK8

arXiv:2004.08262

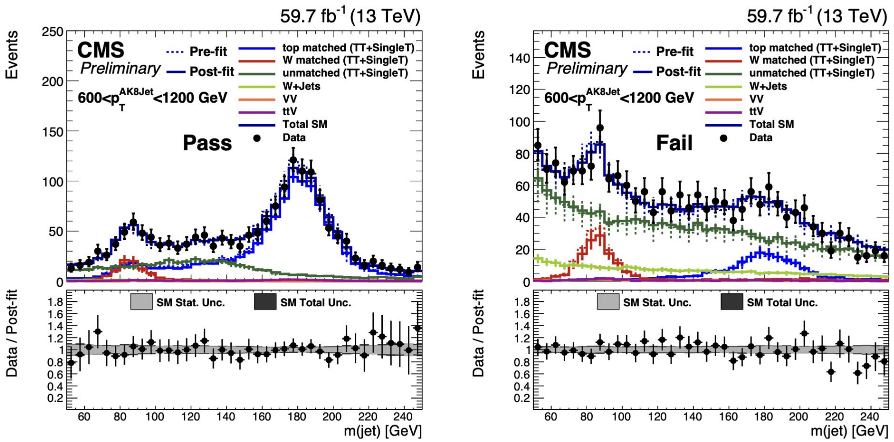


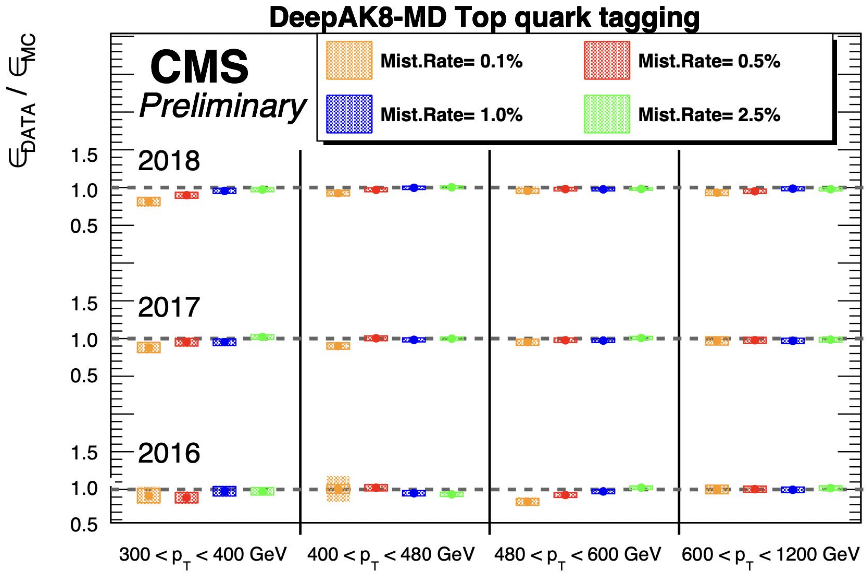
About an order of magnitude improvement

DeepAK8

CMS DP2020/025

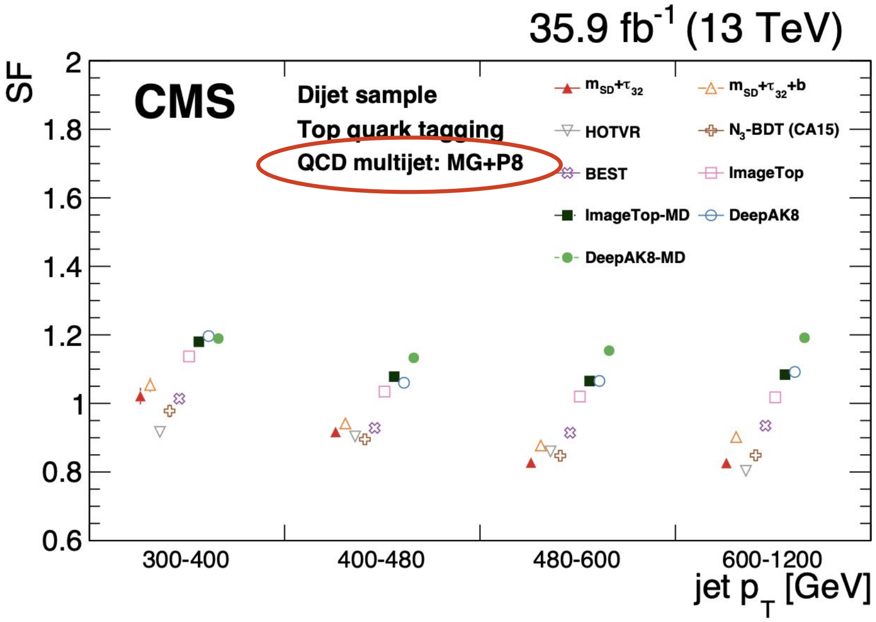
- Calibration of top tagger by fitting for pass/fail ratio in control sample

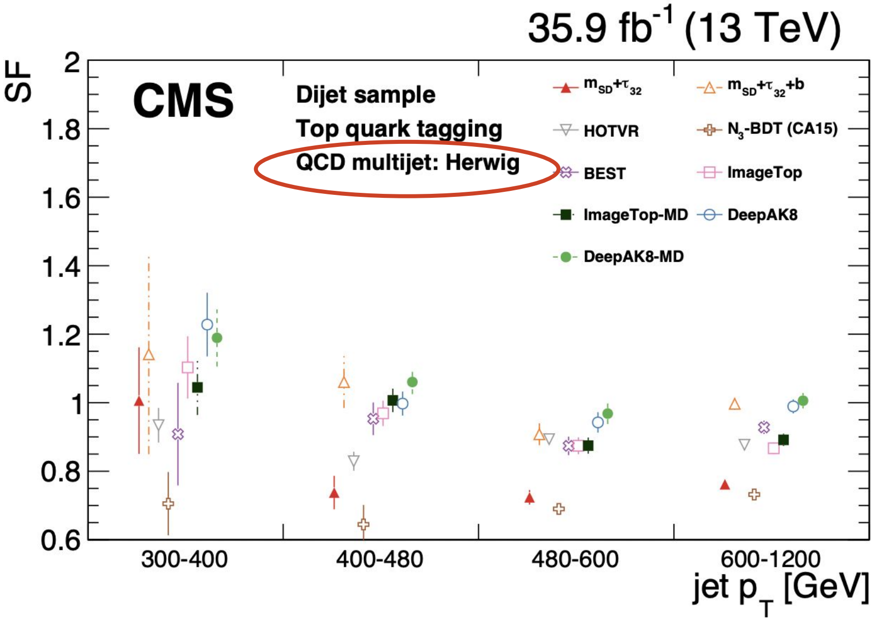




- Scale factors obtained are under control for all years and for different mistag rates
- Parton shower uncertainties typically drive the SF uncertainty
- Mass-decorrelated versions typically have smaller uncs

DeepAK8 - mistag scale factors

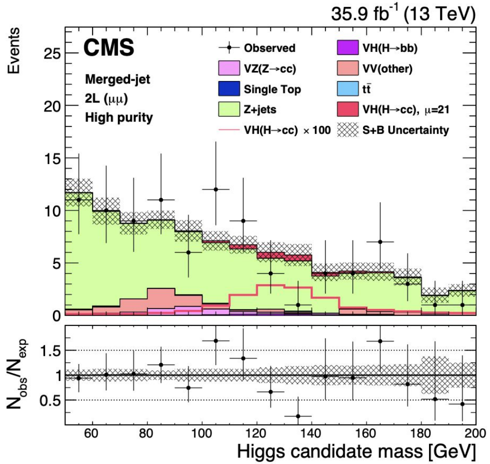




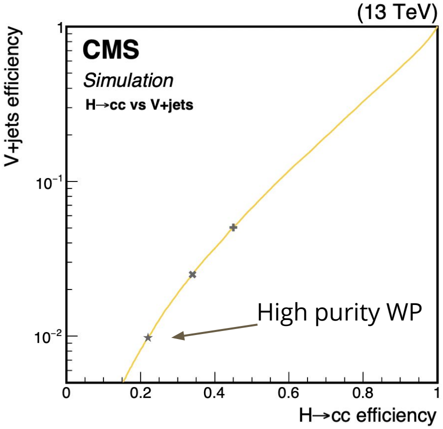
- Again, different parton shower models (MG+Pythia8 vs. Herwig++) introduce a large uncertainty

DeepAK15 used for VH, H ⟶ cc arXiv:1912.01662

- Perform search in 0L,1L,2L channels and in resolved and boosted channels
- Divide phase space into resolved and boosted analysis based on vector boson pT (</> 300 GeV) to make use of beneficial S/B
- Extract signal from simultaneous fit on m_{SD} in signal and control regions defined by different purities of cc tagger, event-level BDT, #leptons



Leading uncertainties from charm tagging and limited statistics in control samples



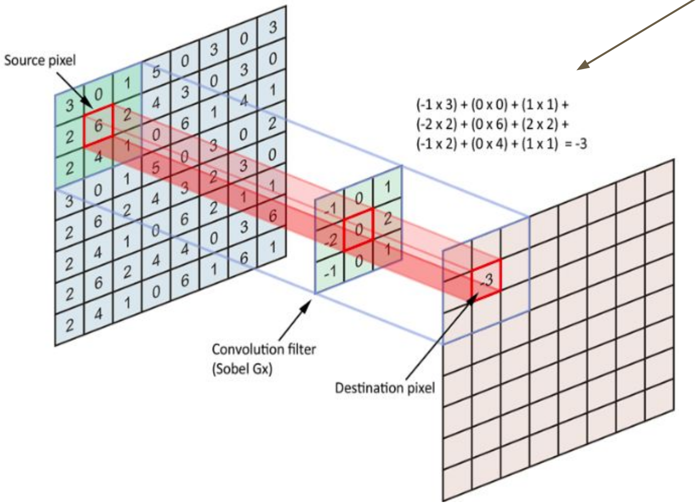
95% CL exclusion limit on $\mu_{VH(H\to c\bar{c})}$

	Resolved-jet	Merged-jet	Combination			
	$(p_T(V) < 300\,GeV)$	$(p_T(V) \geq 300\,GeV)$	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

Jets as particle clouds in CMS

- Convolutional networks highly successful in image recognition by relating region of interest to its surroundings, and DeepAK8 is very performant

Kernel operation



- But: *uniform* pixels in an image vs. *irregular* distribution of jet constituents

Jets as particle clouds in CMS

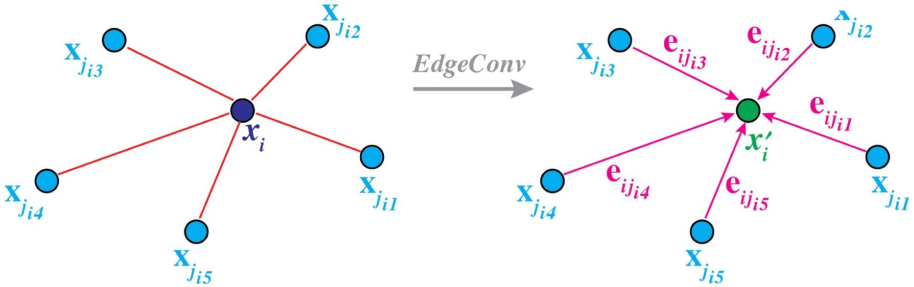
- So-called EdgeConv kernel can still be thought of as a convolution of a local patch in vicinity of particle
- Acts on permutation-invariant set of jet constituents (unlike CNNs, RNNs)

$$x'_i = \mathop{\square}_{j=1}^{k} h_{\Theta}(x_i, x_{i_j})$$

Nearest neighbors (k)

Function parametrized by Learnable parameters (==Dense Neural Network) (h_{Θ})

Symmetric aggregation operation: mean 1/k Σ ($\square$)

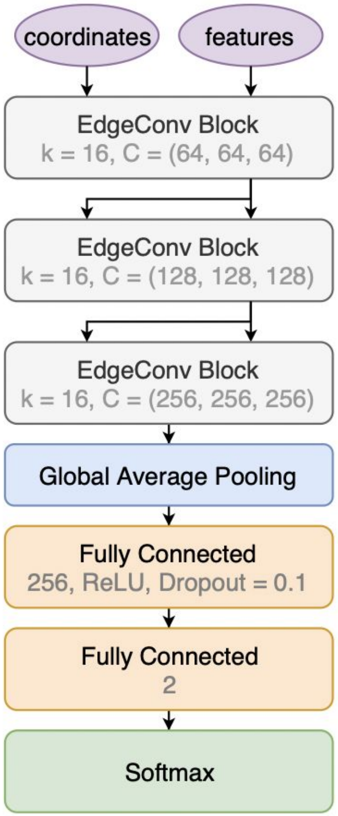


Input space

Latent space

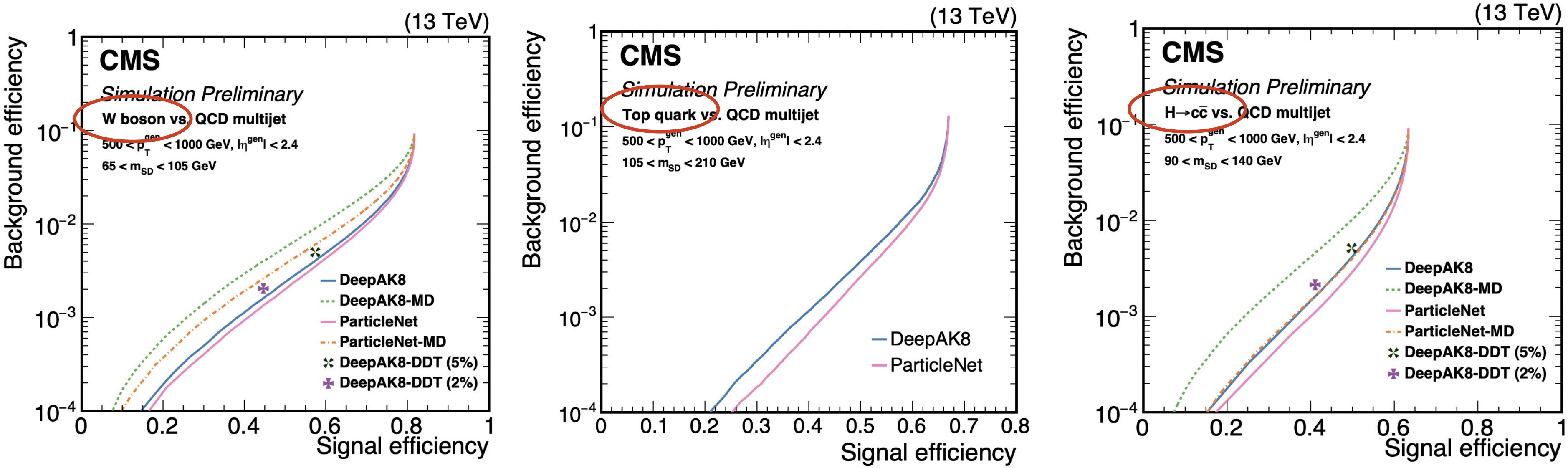
arXiv:1801.07829

ParticleNet arXiv:1902.08570



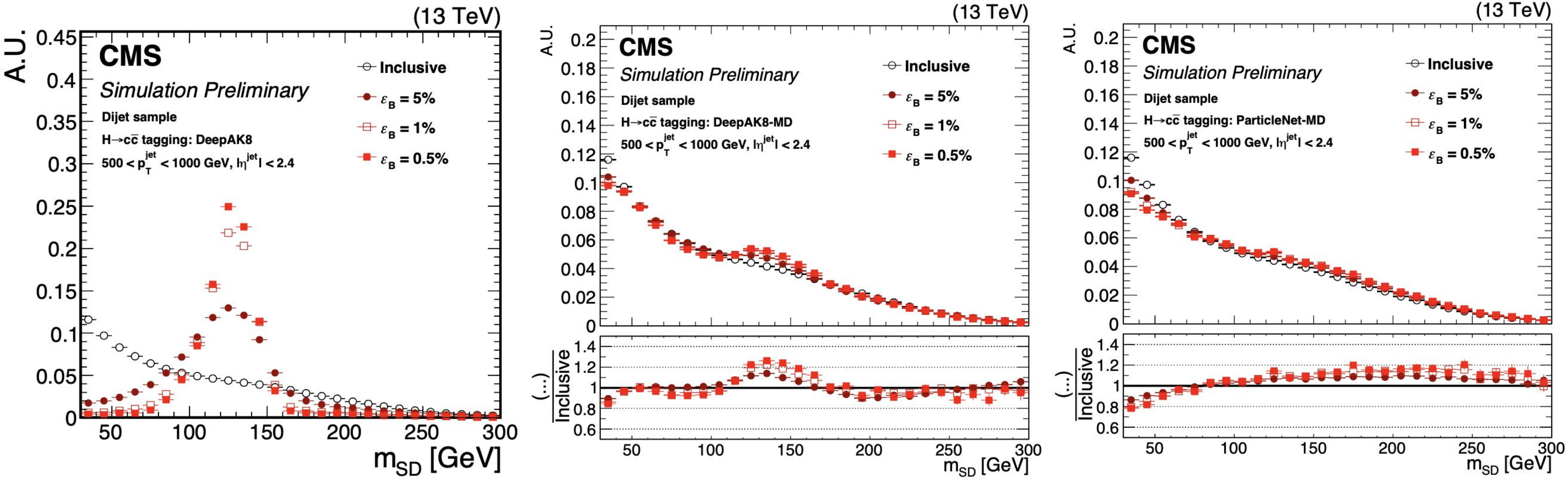
- First EdgeConv block takes edges obtained from kNN with spatial coordinates
- Subsequent EdgeConv layers compute kNNs from latent space, i.e., from a *dynamically learned* graph
- Trained with the same inputs as DeepAK8

ParticleNet [arXiv:1902.08570]



- ParticleNet(-MD) outperforming DeepAK8(-MD) in all cases
- Graph neural networks seem to be the state-of-the-art ML technique for jets

H → cc



- Training with signal samples flat in mass seems to show better decorrelation than adversarial term in loss
- ParticleNet will be essential tool to probe for H → cc

Conclusions

- Boosted object tagging is a useful tool in critical Standard Model measurements
  - Probe high-pT regime; enhance sensitivity to new physics
  - Event reconstruction in boosted topologies is often simpler than resolved topologies
  - Reduce combinatorics in final states with high object multiplicities
- Uncertainties on large-R jet calibration and tagging efficiencies are often dominant
  - Latest round of calibrations and techniques **reduce these uncertainties**
  - Improved **background rejection** from new machine learning-based taggers
  - Expect this to translate to **improved precision** in upcoming round of measurements and searches